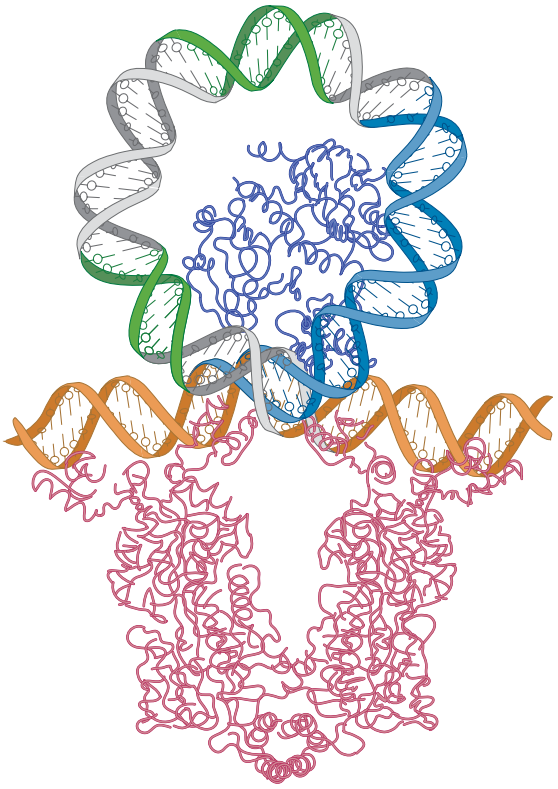

Model showing how the *lac* repressor (red) and catabolite-activating protein (dark blue in center of DNA loop) bind to the *lac* operon promoter, creating a 93-base-pair repression loop in the *lac* regulatory DNA.

16

Regulation of Gene Expression in Prokaryotes

CHAPTER CONCEPTS

- In bacteria, regulation of gene expression is often linked to the metabolic needs of the cell.
- Efficient expression of genetic information in bacteria is dependent on intricate regulatory mechanisms that exert control over transcription.
- Mechanisms that regulate transcription are categorized as exerting either positive or negative control of gene expression.
- Prokaryotic genes that encode proteins with related functions tend to be organized in clusters and are often under coordinated control. Such clusters, including their adjacent regulatory sequences, are called operons.
- Transcription of genes within operons is either inducible or repressible.
- Often, the metabolic end product of a biosynthetic pathway induces or represses gene expression in that pathway.

Previous chapters have discussed how DNA is organized into genes, how genes store genetic information, and how this information is expressed through the processes of transcription and translation. We now consider one of the most fundamental questions in molecular genetics: *How is genetic expression regulated?* It is clear that not all genes are expressed at all times in all situations. For example, detailed analysis of proteins in *E. coli* shows that concentrations of the 4000 or so polypeptide chains encoded by the genome vary widely. Some proteins may be present in as few as 5 to 10 molecules per cell, whereas others, such as ribosomal proteins and the many proteins involved in the glycolytic pathway, are present in as many as 100,000 copies per cell. Although most prokaryotic gene products are present continuously at a basal level (a few copies), the concentration of these products can increase dramatically when required. Clearly, fundamental regulatory mechanisms must exist to control the expression of the genetic information.

In this chapter, we will explore regulation of gene expression in prokaryotes specifically. As we have seen in a number of previous chapters, these organisms served as excellent research organisms in many seminal investigations in molecular genetics. Bacteria have been especially useful research organisms in genetics for a number of reasons. For one thing, they have extremely short reproductive cycles. Literally hundreds of generations, giving rise to billions of genetically identical bacteria, can be produced in overnight cultures. In addition, they can be studied in "pure culture," allowing mutant strains of genetically unique bacteria to be isolated and investigated separately.

Relevant to our current topic, bacteria also serve as an excellent model system for studies involving the induction of genetic transcription in response to changes in environmental conditions. Our focus will be on regulation at the level of the gene. Keep in mind that posttranscriptional regulation also occurs in bacteria. However, we will defer discussion of this level of regulation to Chapter 17, in which we consider eukaryotic regulation as well.

16.1 Prokaryotes Regulate Gene Expression in Response to Environmental Conditions

Regulation of gene expression has been extensively studied in prokaryotes, particularly in *E. coli*. Geneticists have learned that highly efficient genetic mechanisms have evolved in these organisms to turn transcription of specific genes on and off, depending on the cell's metabolic need for the respective gene products. Not only do bacteria respond to changes in their environment, but they also regulate gene activity associated with a variety of nonenvironmentally regulated cellular activities (including the replication, recombination, and repair of their DNA) and with cell division.

The idea that microorganisms regulate the synthesis of their gene products is not a new one. As early as 1900, it was shown that when lactose (a galactose and glucose-containing disaccharide) is present in the growth medium of yeast, the organisms synthesize enzymes required for lactose metabolism. When lactose is absent, the enzymes are not manufactured. Soon thereafter, investigators were able to generalize that bacteria also adapt to their environment, producing certain enzymes only when specific chemical substrates are present. These were thus referred to as **adaptive enzymes** (sometimes also called *facultative*). In contrast, enzymes that are produced continuously, regardless of the chemical makeup of the environment, were called **constitutive enzymes**. Since then, the term *adaptive* has been replaced with the more accurate term **inducible**, reflecting the role of the substrate, which serves as the **inducer** in enzyme production.

More recent investigation has revealed a contrasting system, whereby the presence of a specific molecule inhibits gene expression. Such molecules are usually end products of anabolic biosynthetic pathways. For example, the amino acid tryptophan can be synthesized by bacterial cells. If a sufficient supply of tryptophan is present in the environment or culture medium, then there is no reason for the organism to expend energy in synthesizing the enzymes necessary for tryptophan production. A mechanism has therefore evolved whereby tryptophan plays a role in repressing the transcription of mRNA needed for producing tryptophan-synthesizing enzymes. In contrast to the inducible system controlling lactose metabolism, the system governing tryptophan expression is said to be **repressible**.

Regulation, whether of the inducible or repressible type, may be under either **negative** or **positive control**. Under negative control, genetic expression occurs *unless it is shut off by some form of a regulator molecule.* In contrast, under positive control, transcription occurs *only if a regulator molecule directly stimulates RNA production.* In theory, either type of control or a combination of the two can govern inducible or repressible systems. Our discussion in the ensuing sections of this chapter will help clarify these contrasting systems of regulation. The enzymes involved in lactose digestion and tryptophan synthesis are under negative control.

16.2 Lactose Metabolism in *E. coli* Is Regulated by an Inducible System

Beginning in 1946 with the studies of Jacques Monod and continuing through the next decade with significant contributions by Joshua Lederberg, François Jacob, and

André Lwoff, genetic and biochemical evidence concerning lactose metabolism was amassed. Research provided insights into the way in which the gene activity is repressed when lactose is absent but induced when it is available. In the presence of lactose, the concentration of the enzymes responsible for its metabolism increases rapidly from a few molecules to thousands per cell. The enzymes responsible for lactose metabolism are thus inducible, and lactose serves as the inducer.

In prokaryotes, genes that code for enzymes with related functions (for example, the set of genes involved with lactose metabolism) tend to be organized in clusters on the bacterial chromosome, and transcription of these genes is often under the coordinated control of a single regulatory region. The location of this regulatory region is almost always upstream (5′) of the gene cluster it controls. Because the regulatory region is on the same strand as those genes, we refer to it as a ***cis*-acting site**. *Cis*-acting regulatory regions bind molecules that control transcription of the gene cluster. Such molecules are called ***trans*-acting elements**. Events at the regulatory site determine whether the genes are transcribed into mRNA and thus whether the corresponding enzymes or other protein products may be synthesized from the genetic information in the mRNA. Binding of a *trans*-acting element at a *cis*-acting site can regulate the gene cluster either negatively (by turning off transcription) or positively (by turning on transcription of genes in the cluster). In this section, we discuss how transcription of such bacterial gene clusters is coordinately regulated.

The discovery of a regulatory gene and a regulatory site that are part of the gene cluster was paramount to the understanding of how gene expression is controlled in the system. Neither of these regulatory elements encodes enzymes necessary for lactose metabolism—the function of the three genes in the cluster. As illustrated in Figure 16–1, the three structural genes and the adjacent regulatory site constitute the **lactose**, or ***lac*, operon**. Together, the entire gene cluster functions in an integrated fashion to provide a rapid response to the presence or absence of lactose.

Structural Genes

Genes coding for the primary structure of an enzyme are called **structural genes**. There are three structural genes in

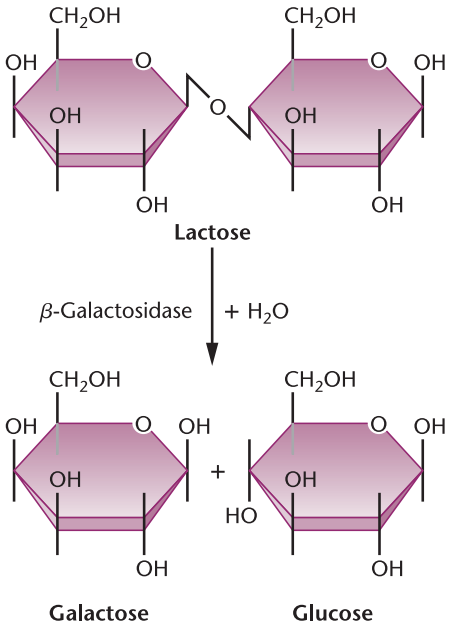

FIGURE 16–2 The catabolic conversion of the disaccharide lactose into its monosaccharide units, galactose and glucose.

the *lac* operon. The *lacZ* gene encodes ***β*-galactosidase**, an enzyme whose primary role is to convert the disaccharide lactose to the monosaccharides glucose and galactose (Figure 16–2). This conversion is essential if lactose is to serve as the primary energy source in glycolysis. The second gene, *lacY*, specifies the primary structure of **permease**, an enzyme that facilitates the entry of lactose into the bacterial cell. The third gene, *lacA*, codes for the enzyme **transacetylase**. While its physiological role is still not completely clear, it may be involved in the removal of toxic by-products of lactose digestion from the cell.

To study the genes coding for these three enzymes, researchers isolated numerous mutations that lacked the function of one or the other enzyme. Such *lac*$^-$ mutants were first isolated and studied by Joshua Lederberg. Mutant cells that fail to produce active *β*-galactosidase (*lacZ*$^-$) or permease (*lacY*$^-$) are unable to use lactose as an energy source. Mutations were also found in the transacetylase gene. Mapping studies by Lederberg established that all three genes are

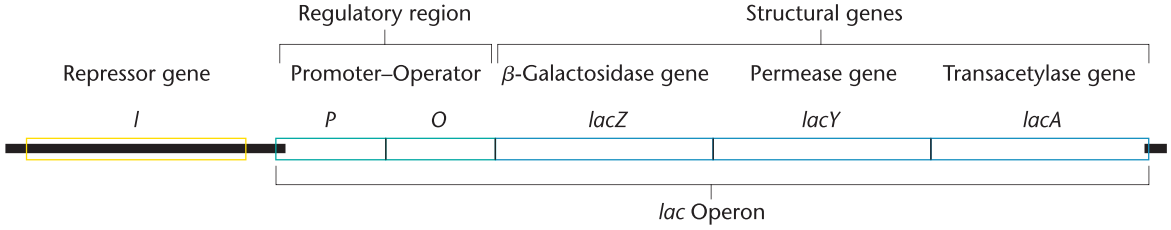

FIGURE 16–1 A simplified overview of the genes and regulatory units involved in the control of lactose metabolism. (The regions within this stretch of DNA are not drawn to scale.) A more detailed model will be developed later in this chapter. (See Figure 16–10.)

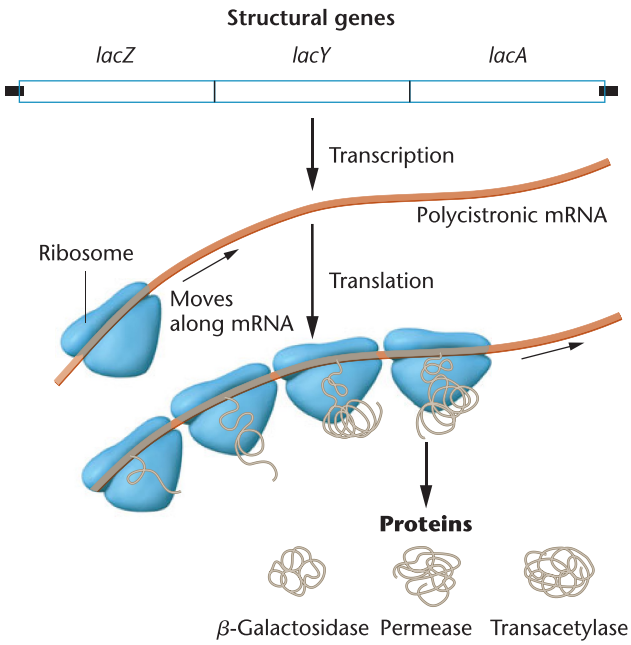

FIGURE 16–3 The structural genes of the *lac* operon are transcribed into a single polycistronic mRNA, which is translated simultaneously by several ribosomes into the three enzymes encoded by the operon.

closely linked or contiguous to one another on the bacterial chromosome, in the order *Z–Y–A* (see Figure 16–1).

Knowledge of their close linkage led to another discovery relevant to what later became known about the regulation of structural genes: All three genes are transcribed as a single unit, resulting in a so-called *polycistronic mRNA* (Figure 16–3; recall that *cistron* refers to the part of a nucleotide sequence coding for a single gene). This results in the coordinate regulation of all three genes, since a single-message RNA is simultaneously translated into all three gene products.

The Discovery of Regulatory Mutations

How does lactose stimulate transcription of the *lac* operon and induce the synthesis of the enzymes for which it codes? A partial answer came from studies using **gratuitous inducers**, chemical analogs of lactose such as the sulfur-containing analog **isopropylthiogalactoside (IPTG)**, shown in Figure 16–4.

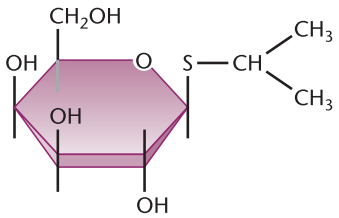

FIGURE 16–4 The gratuitous inducer isopropylthiogalactoside (IPTG).

Gratuitous inducers behave like natural inducers, but they do not serve as substrates for the enzymes that are subsequently synthesized. Their discovery provides strong evidence that the primary induction event does *not* depend on the interaction between the inducer and the enzyme.

What, then, is the role of lactose in induction? The answer to this question required the study of another class of mutations described as **constitutive mutations**. In cells bearing these types of mutations, enzymes are produced regardless of the presence or absence of lactose. Studies of the constitutive mutation *lacI*$^-$ mapped the mutation to a site on the bacterial chromosome close to, but distinct from, the structural genes *lacZ, lacY,* and *lacA*. This mutation led researchers to discover the *lacI* gene, which is appropriately called a **repressor gene**. A second set of constitutive mutations producing effects identical to those of *lacI*$^-$ is present in a region immediately adjacent to the structural genes. This class of mutations, designated *lac* O^C, is located in the **operator region** of the operon. In both types of constitutive mutants, the enzymes are produced continually, inducibility is eliminated, and gene regulation has been lost.

The Operon Model: Negative Control

Around 1960, Jacob and Monod proposed a hypothetical mechanism involving negative control that they called the **operon model**, in which a group of genes is regulated and expressed together as a unit. As we saw in Figure 16–1, the *lac* operon they proposed consists of the *Z, Y,* and *A* structural genes, as well as the adjacent sequences of DNA referred to as the *operator region.* They argued that the *lacI* gene regulates the transcription of the structural genes by producing a **repressor molecule**, and that the repressor is **allosteric**, meaning that the molecule reversibly interacts with another molecule, undergoing both a conformational change in three-dimensional shape and a change in chemical activity. Figure 16–5 illustrates the components of the *lac* operon as well as the action of the *lac* repressor in the presence and absence of lactose.

Jacob and Monod suggested that the repressor normally binds to the DNA sequence of the operator region. When it does so, it inhibits the action of RNA polymerase, effectively repressing the transcription of the structural genes [Figure 16–5(b)]. However, when lactose is present, this sugar binds to the repressor and causes an allosteric (conformational) change. The change alters the binding site of the repressor, rendering it incapable of interacting with operator DNA [Figure 16–5(c)]. In the absence of the repressor–operator interaction, RNA polymerase transcribes the structural genes, and the enzymes necessary for lactose metabolism are produced. Because transcription occurs only when the repressor *fails* to bind to the operator region, regulation is said to be under *negative control.*

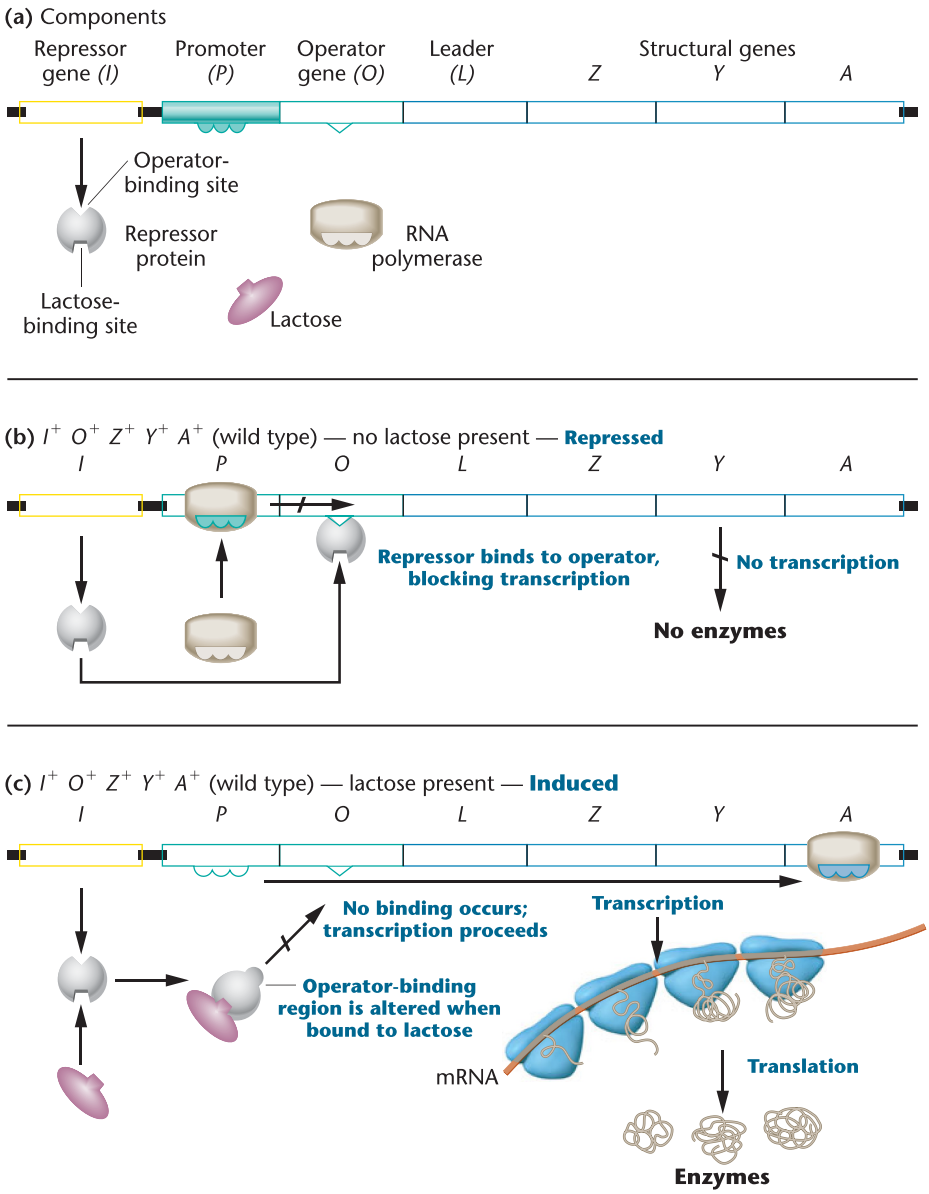

FIGURE 16–5 The components of the wild-type *lac* operon and the response in the absence and presence of lactose.

To summarize, the operon model invokes a series of molecular interactions between proteins, inducers, and DNA to explain the efficient regulation of structural gene expression. In the absence of lactose, the enzymes encoded by the genes are not needed, and expression of genes encoding these enzymes is repressed. When lactose is present, it indirectly induces the activation of the genes by binding with the repressor.* If all lactose is metabolized, none is available to bind to the repressor, which is again free to bind to operator DNA and to repress transcription.

Both the I^- and O^C constitutive mutations interfere with these molecular interactions, allowing continuous transcription of the structural genes. In the case of the I^- mutant, seen in Figure 16–6(a), the repressor protein is altered or absent and cannot bind to the operator region, so the structural genes are always turned on. In the case of the O^C mutant [Figure 16–6(b)], the nucleotide sequence of the operator DNA is altered and will not bind with a normal repressor molecule. The result is the same: The structural genes are always transcribed.

*Technically, the inducer is allolactose, an isomer of lactose. When lactose enters the bacterial cell, some of it is converted to allolactose by the β-galactosidase enzyme.

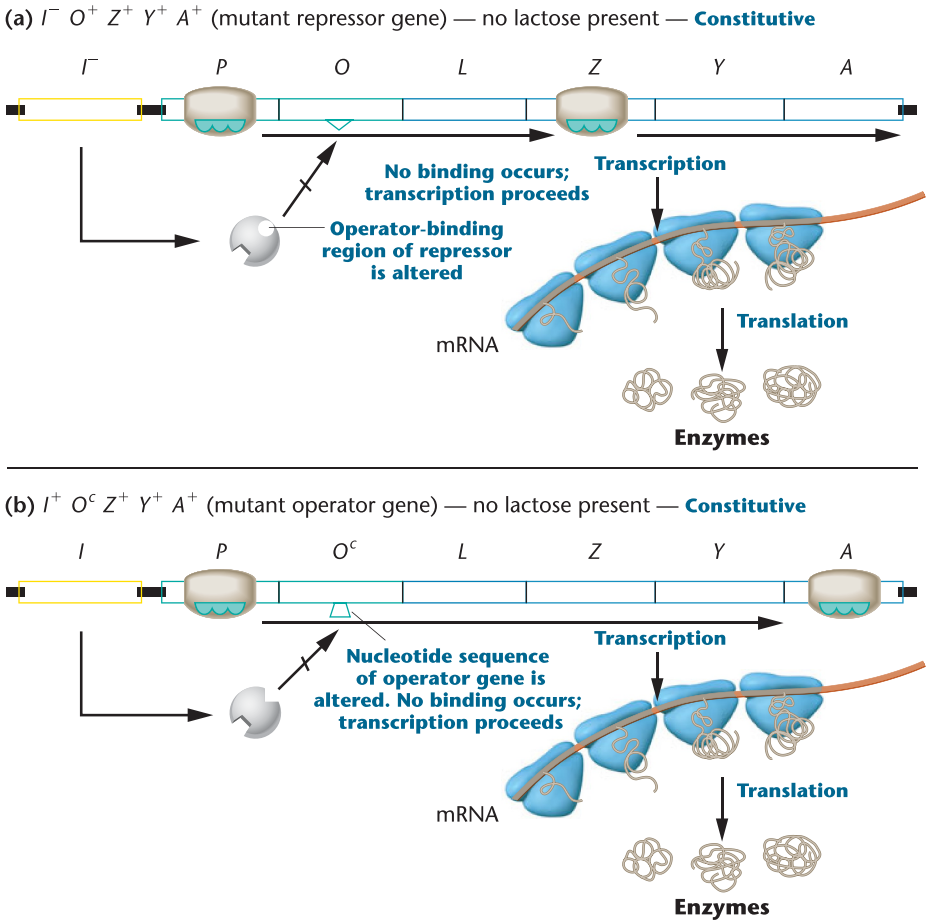

FIGURE 16–6 The response of the *lac* operon in the absence of lactose when a cell bears either the I^- or the O^C mutation.

Genetic Proof of the Operon Model

The operon model is a good one because it leads to three major predictions that can be tested to determine its validity. The major predictions to be tested are that (1) the *I* gene produces a diffusible product (that is, a *trans*-acting product); (2) the *O* region is involved in regulation but does not produce a product (it is *cis*-acting); and (3) the *O* region must be adjacent to the structural genes in order to regulate transcription.

The creation of partially diploid bacteria allows us to assess these assumptions, particularly those that predict the presence of *trans*-acting regulatory elements. For example, the F plasmid may contain chromosomal genes (Chapter 6), in which case it is designated F′. When an F^- cell acquires such a plasmid, it contains its own chromosome plus one or more additional genes present in the plasmid. This host cell is thus a **merozygote**, a cell that is diploid for certain added genes (but not for the rest of the chromosome). The use of such a plasmid makes it possible, for example, to introduce an I^+ gene into a host cell whose genotype is I^-, or to introduce an O^+ region into a host cell of genotype O^C. The Jacob–Monod operon model predicts how regulation should be affected in such cells. Adding an I^+ gene to an I^- cell should restore inducibility, because the normal wild-type repressor, which is a *trans*-acting factor, would be produced by the inserted I^+ gene. In contrast, adding an O^+ region to an O^C cell should have no effect on constitutive enzyme production, since regulation depends on an O^+ region being located immediately adjacent to the structural genes—that is, O^+ is a *cis*-acting regulator.

Results of these experiments are shown in Table 16.1, where *Z* represents the structural genes (and the inserted genes are listed after the designation F′). In both cases described above, the Jacob–Monod model is upheld (part B of Table 16.1). Part C of the table shows the reverse experiments, where either an I^- gene or an O^C region is added to cells of normal inducible genotypes. As the model predicts, inducibility is maintained in these partial diploids.

Another prediction of the operon model is that certain mutations in the *I* gene should have the opposite effect of I^-.

TABLE 16.1

A Comparison of Gene Activity (+ or −) in the Presence or Absence of Lactose for Various *E. coli* Genotypes

	Genotype	Presence of β-Galactosidase Activity: Lactose Present	Lactose Absent
	$I^+O^+Z^+$	+	−
A.	$I^+O^+Z^-$	−	−
	$I^-O^+Z^+$	+	+
	$I^+O^CZ^+$	+	+
B.	$I^-O^+Z^+/F'I^+$	+	−
	$I^+O^CZ^+/F'O^+$	+	+
C.	$I^+O^+Z^+/F'I^-$	+	−
	$I^+O^+Z^+/F'O^C$	+	−
D.	$I^SO^+Z^+$	−	−
	$I^SO^+Z^+/F'I^+$	−	−

Note: In parts B to D, most genotypes are partially diploid, containing an F factor plus attached genes (F′).

That is, instead of being constitutive because the repressor can't bind the operator, mutant repressor molecules should be produced that cannot interact with the inducer, lactose. As a result, these repressors would always bind to the operator sequence, and the structural genes would be permanently repressed. In cases like this, the presence of an additional I^+ gene would have little or no effect on repression.

In fact, such a mutation, I^S, was discovered wherein the operon, as predicted, is "superrepressed," as shown in part D of Table 16.1 (and depicted in Figure 16–7). An additional I^+ gene does not effectively relieve repression of gene activity. These observations are consistent with the idea that the repressor contains separate DNA-binding domains and inducer-binding domains.

Isolation of the Repressor

Although Jacob and Monod's operon theory succeeded in explaining many aspects of genetic regulation in prokaryotes, the nature of the repressor molecule was not known when their landmark paper was published in 1961. While they had assumed that the allosteric repressor was a protein, RNA was also a candidate because the activity of the molecule required the ability to bind to DNA. Despite many attempts to isolate and characterize the hypothetical repressor molecule, no direct chemical evidence was forthcoming. A single *E. coli* cell contains no more than ten or so copies of the *lac* repressor, and direct chemical identification of ten molecules in a population of millions of proteins and RNAs in a single cell presented a tremendous challenge.

In 1966, Walter Gilbert and Benno Müller-Hill reported the isolation of the *lac* repressor in partially purified form. To achieve the isolation, they used a *regulator quantity* (I^q) mutant strain that contains about ten times as much repressor as do wild-type *E. coli* cells. Also instrumental in their success was the use of the gratuitous inducer IPTG, which binds to the repressor, and the technique of **equilibrium dialysis**. In this technique, extracts of I^q cells were placed in a dialysis bag and allowed to attain equilibrium with an external solution of radioactive IPTG, a molecule small enough to diffuse freely in and out of the bag. At equilibrium, the

NOW SOLVE THIS

16–1 Even though the *lac Z*, *Y*, and *A* structural genes are transcribed as a single polycistronic mRNA, each gene contains the initiation and termination signals essential for translation. Predict what will happen when a cell growing in the presence of lactose contains a deletion of one nucleotide (a) early in the *Z* gene and (b) early in the *A* gene.

HINT: *This problem requires you to combine your understanding of the genetic expression of the lac operon with that of the genetic code, frameshift mutations, and termination of transcription. The key to its solution is to consider the effect of the loss of one nucleotide within a polycistronic mRNA.*

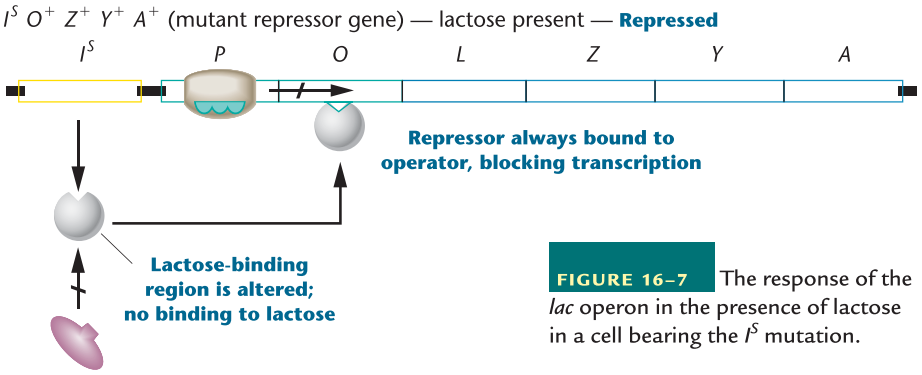

FIGURE 16–7 The response of the *lac* operon in the presence of lactose in a cell bearing the I^S mutation.

concentration of radioactive IPTG was higher inside the bag than in the external solution, indicating that an IPTG-binding material was present in the cell extract and was too large to diffuse across the wall of the bag.

Ultimately, the IPTG-binding material was purified and shown to have various characteristics of a protein. In contrast, extracts of I^- constitutive cells having no *lac* repressor *activity* did not exhibit IPTG binding, strongly suggesting that the isolated protein was the repressor molecule.

To confirm this thinking, Gilbert and Müller-Hill grew *E. coli* cells in a medium containing radioactive sulfur and then isolated the IPTG-binding protein, which was labeled in its sulfur-containing amino acids. Next, this protein was mixed with DNA from a strain of phage lambda (λ) carrying the $lacO^+$ gene. When the two substances are not mixed together, the DNA sediments at 40*S*, while the IPTG-binding protein sediments at 7*S*. However, when the DNA and protein were mixed and sedimented in a gradient, using ultracentrifugation, the radioactive protein sedimented at the same rate as did DNA, indicating that the protein binds to the DNA. Further experiments showed that the IPTG-binding, or repressor, protein binds only to DNA containing the *lac* region and does not bind to *lac* DNA containing an operator-constitutive O^C mutation.

16.3 The Catabolite-Activating Protein (CAP) Exerts Positive Control over the *lac* Operon

As described in the preceding discussion of the *lac* operon, the role of β-galactosidase is to cleave lactose into its components, glucose and galactose. Then, in order to be used by the cell, the galactose, too, must be converted to glucose. What if the cell found itself in an environment that contained ample amounts of both lactose *and* glucose? Given that glucose is the preferred carbon source for *E. coli,* it would not be energetically efficient for a cell to induce transcription of the *lac* operon, since what it really needs—glucose—is already present. As we will see next, still another molecular component, called the **catabolite-activating protein (CAP)**, is involved in effectively repressing the expression of the *lac* operon when glucose is present. This inhibition is called **catabolite repression**.

To understand CAP and its role in regulation, let's backtrack for a moment to review the system depicted in Figure 16–5. When the *lac* repressor is bound to the inducer (lactose), the *lac* operon is activated, and RNA polymerase transcribes the structural genes. As stated in Chapter 14, transcription is initiated as a result of the binding that occurs between RNA polymerase and the nucleotide sequence of the **promoter region**, found upstream (5′) from the initial coding sequences. Within the *lac* operon, the promoter is found between the *I* gene and the operator region (*O*) (see Figure 16–1). Careful examination has revealed that polymerase binding is never very efficient unless CAP is also present to facilitate the process.

The mechanism is summarized in Figure 16–8. In the absence of glucose and under inducible conditions, CAP exerts positive control by binding to the CAP site, facilitating RNA-polymerase binding at the promoter, and thus transcription. Therefore, for maximal transcription of the structural genes, the repressor must be bound by lactose (so as not to repress operon expression), *and* CAP must be bound to the CAP-binding site.

This leads to the central question about CAP: How does the presence of glucose inhibit CAP binding? The answer involves still another molecule, **cyclic adenosine monophosphate (cAMP)**, upon which CAP binding is dependent. *In order to bind to the promoter, CAP must be bound to cAMP.* The level of cAMP is itself dependent on an enzyme, **adenyl cyclase**, which catalyzes the conversion of ATP to cAMP (see Figure 16–9).* The role of glucose in catabolite repression is now clear. It inhibits the activity of adenyl cyclase, causing a decline in the level of cAMP in the cell. Under this condition, CAP cannot form the CAP–cAMP complex essential to the positive control of transcription of the *lac* operon.

Like the *lac* repressor, CAP and cAMP–CAP have been examined by X-ray crystallography. CAP is a dimer that inserts into adjacent regions of a specific nucleotide sequence of the DNA making up the *lac* promoter. The cAMP–CAP complex, when bound to DNA, bends it, causing it to assume a new conformation.

Binding studies in solution further clarify the mechanism of gene activation. Alone, neither cAMP–CAP nor RNA polymerase has a strong tendency to bind to *lac* promoter DNA, nor does either molecule have a strong affinity for the other. However, when both are together in the presence of the *lac* promoter DNA, a tightly bound complex is formed, an example of what is called **cooperative binding**. In the case of cAMP–CAP and the *lac* operon, this phenomenon illustrates the high degree of specificity that is involved in the genetic regulation of just one small group of genes.

Regulation of the *lac* operon by catabolite repression results in efficient energy use, because the presence of glucose will

* Because of its involvement with cAMP, CAP is also called *cyclic AMP receptor protein (CRP),* and the gene encoding the protein is named *crp*. Since the protein was first named CAP, we will adhere to the initial nomenclature.

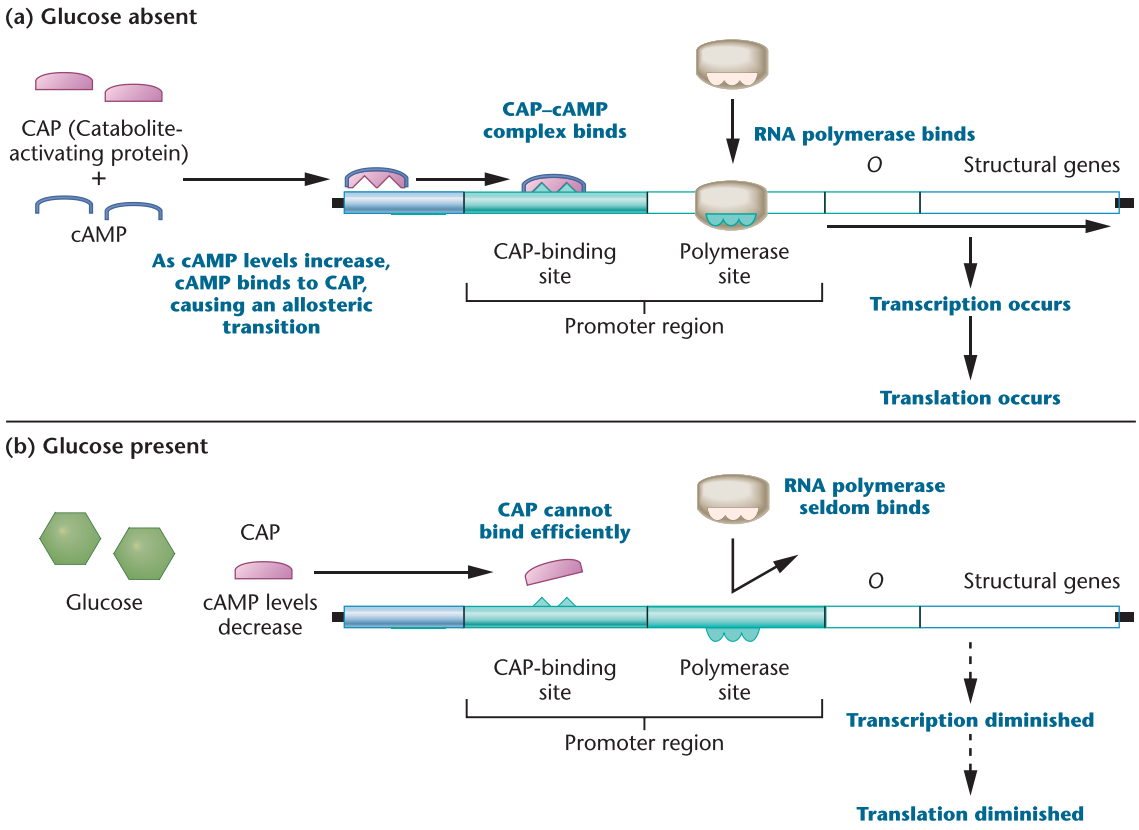

FIGURE 16–8 Catabolite repression. (a) In the absence of glucose, cAMP levels increase, resulting in the formation of a CAP–cAMP complex, which binds to the CAP site of the promoter, stimulating transcription. (b) In the presence of glucose, cAMP levels decrease, CAP–cAMP complexes are not formed, and transcription is not stimulated.

override the need for the metabolism of lactose, should the lactose also be available to the cell. In contrast to the negative regulation conferred by the *lac* repressor, the action of cAMP–CAP constitutes positive regulation. Thus, a combination of positive and negative regulatory mechanisms determines transcription levels of the *lac* operon. Catabolite repression involving CAP has also been observed for other inducible operons, including those controlling the metabolism of galactose and arabinose.

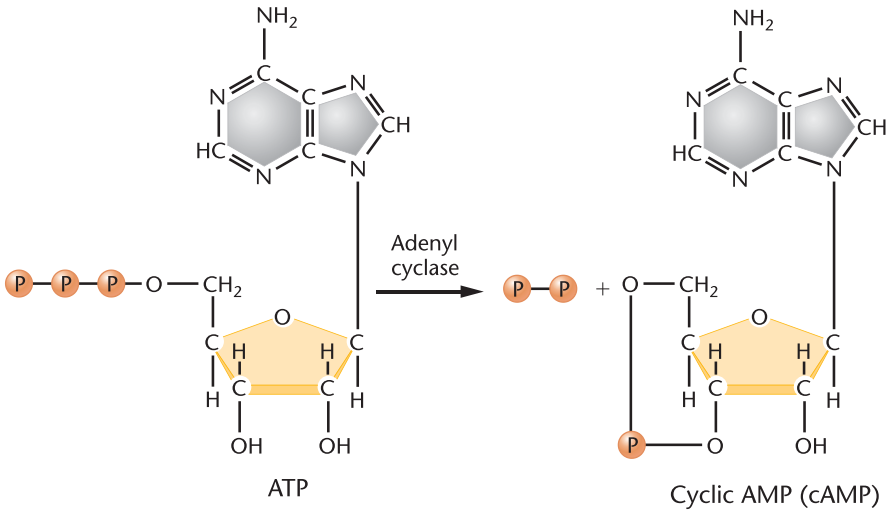

FIGURE 16–9 The formation of cAMP from ATP, catalyzed by adenyl cyclase.

NOW SOLVE THIS

16–2 Predict the level of genetic activity of the *lac* operon as well as the status of the *lac* repressor and the CAP protein under the cellular conditions listed in the accompanying table.

	Lactose	Glucose
(a)	–	–
(b)	+	–
(c)	–	+
(d)	+	+

HINT: *This problem asks you to combine your knowledge of the Jacob–Monod model of the regulation of the lac operon with your understanding of how catabolite repression impacts on this model. The key to its solution is to keep in mind that regulation involving lactose is a negative control system, while regulation involving glucose and catabolite repression is a positive control system.*

16.4 Crystal Structure Analysis of Repressor Complexes Has Confirmed the Operon Model

We now have thorough knowledge of the biochemical nature of the regulatory region of the *lac* operon, including the precise locations of its various components relative to one another (Figure 16–10). In 1996, Mitchell Lewis, Ponzy Lu, and their colleagues succeeded in determining the crystal structure of the *lac* repressor, as well as the structure of the repressor bound to the inducer and to operator DNA. As a result, previous information that was based on genetic and biochemical data has now been complemented with the missing structural interpretation. Together, these contributions provide a nearly complete picture of the regulation of the operon.

The repressor, as the gene product of the *I* gene, is a monomer consisting of 360 amino acids. Within this monomer, the region of inducer binding has been identified [Figure 16–11(a)]. While dimers [Figure 16–11(b)] can also bind the inducer, the functional repressor is a homotetramer (that is, it contains four copies of the monomer). The tetramer can be cleaved with a protease under controlled conditions to yield five fragments. Four are derived from the N-terminal ends of the tetramer subunits, and they bind to operator DNA. The fifth fragment is the remaining core of the tetramer, derived from the COOH-terminus ends; it binds to lactose and gratuitous inducers such as IPTG. Analysis has revealed that each tetramer can bind to two symmetrical operator DNA helices at a time.

The operator DNA that was previously defined by mutational studies ($lacO^C$) and confirmed by DNA-sequencing analysis is located just upstream from the beginning of the actual coding sequence of the *lacZ* gene. Crystallographic studies show that the actual region of repressor binding of this primary operator, O_1, consists of 21 base pairs. Two other auxiliary operator regions have been identified, as shown in Figure 16–10. One, O_2, is 401 base pairs downstream from the primary operator, within the *lacZ* gene. The other, O_3, is 93 base pairs upstream from O_1, just beyond the CAP site. *In vivo*, all three operators must be bound for maximum repression.

Binding by the repressor simultaneously at two operator sites distorts the conformation of DNA, causing it to bend away from the repressor. When a model is created to demonstrate dual binding of operators O_1 and O_3 [Figure 16–11(c)], the 93 base pairs of DNA that intervene must jut out, forming what is called a **repression loop**. This model positions the promoter region that binds RNA polymerase on the inside of the loop, which prevents access by the polymerase during repression. In addition, the repression loop positions the CAP-binding site in a way that facilitates CAP interaction with RNA polymerase upon subsequent induction. The DNA looping caused by repression in this model is similar to configuration changes that are predicted to occur in eukaryotic systems (see Chapter 17).

Studies have also defined the three-dimensional conformational changes that accompany the interactions with the inducer molecules. Taken together, the crystallographic studies have brought a new level of understanding of the regulatory process occurring within the *lac* operon, confirming the findings and predictions of Jacob and Monod in their model set forth over 40 years ago, which was based strictly on genetic observations.

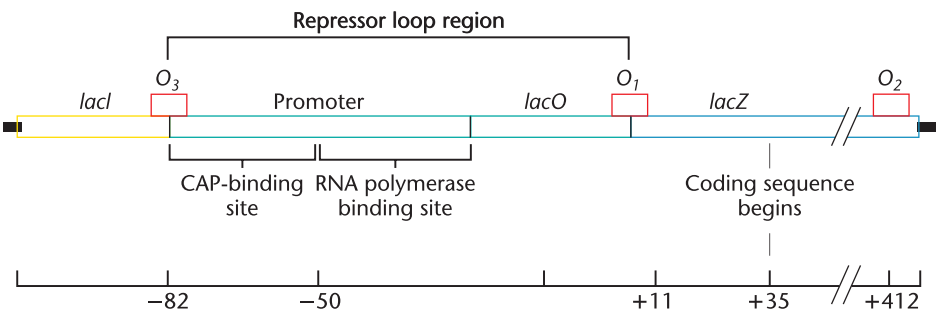

FIGURE 16–10 The various regulatory regions involved in the control of genetic expression of the *lac* operon, as described in the text. The numbers on the bottom scale represent nucleotide sites upstream and downstream from the initiation of transcription.

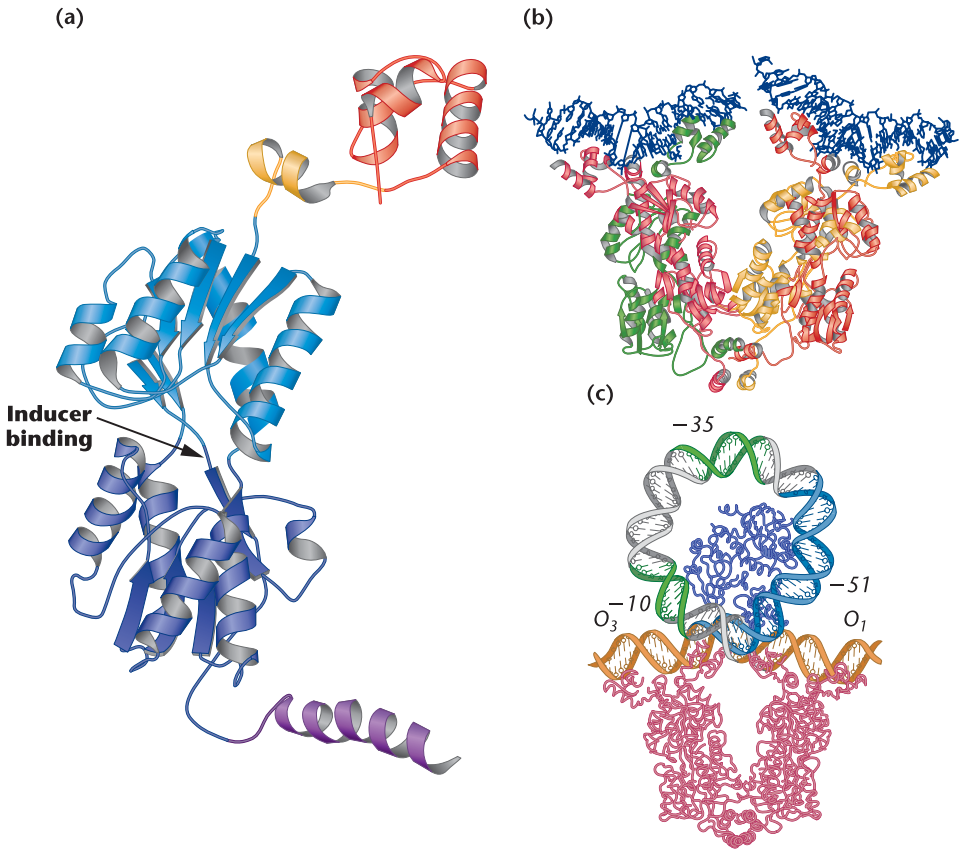

FIGURE 16–11 **Models of the *lac* repressor and its binding** to operator sites in DNA, as generated from crystal structure analysis. (a) The repressor monomer, showing the inducer-binding site. The DNA-binding region is shown in red. (b) The repressor dimer bound to two 21-base-pair segments of operator DNA (shown in dark blue). (c) The repressor (shown in pink) and CAP (shown in dark blue) bound to the *lac* DNA. Binding to operator regions O_1 and O_3 creates a 93-base-pair repression loop of promoter DNA.

16.5 The Tryptophan (*trp*) Operon in *E. coli* Is a Repressible Gene System

Although the process of induction had been known for some time, it was not until 1953 that Monod and colleagues discovered a repressible operon. Wild-type *E. coli* are capable of producing the enzymes necessary for the biosynthesis of amino acids as well as other essential macromolecules. Focusing his studies on the amino acid tryptophan and the enzyme **tryptophan synthetase**, Monod discovered that if tryptophan is present in sufficient quantity in the growth medium, the enzymes necessary for its synthesis are not produced. It is energetically advantageous for bacteria to repress expression of genes involved in tryptophan synthesis when ample tryptophan is present in the growth medium.

Further investigation showed that a series of enzymes encoded by five contiguous genes on the *E. coli* chromosome are involved in tryptophan synthesis. These genes are part of an operon, and in the presence of tryptophan, all are coordinately repressed, and none of the enzymes is produced. Because of the great similarity between this repression and the induction of enzymes for lactose metabolism, Jacob and Monod proposed a model of gene regulation analogous to the *lac* system (the updated version is shown in Figure 16–12).

To account for repression, Jacob and Monod suggested the presence of a *normally inactive repressor* that alone cannot interact with the operator region of the operon. However, the repressor is an allosteric molecule that can bind to tryptophan. When tryptophan is present, the resultant complex of repressor and tryptophan attains a new conformation that binds to the operator, repressing transcription. Thus, when tryptophan, the end product of this anabolic pathway, is present, the system is repressed and enzymes are not made. Since the regulatory complex inhibits transcription of the operon, this repressible system is under negative control. And as tryptophan participates in repression, it is referred to as a **corepressor** in this regulatory scheme.

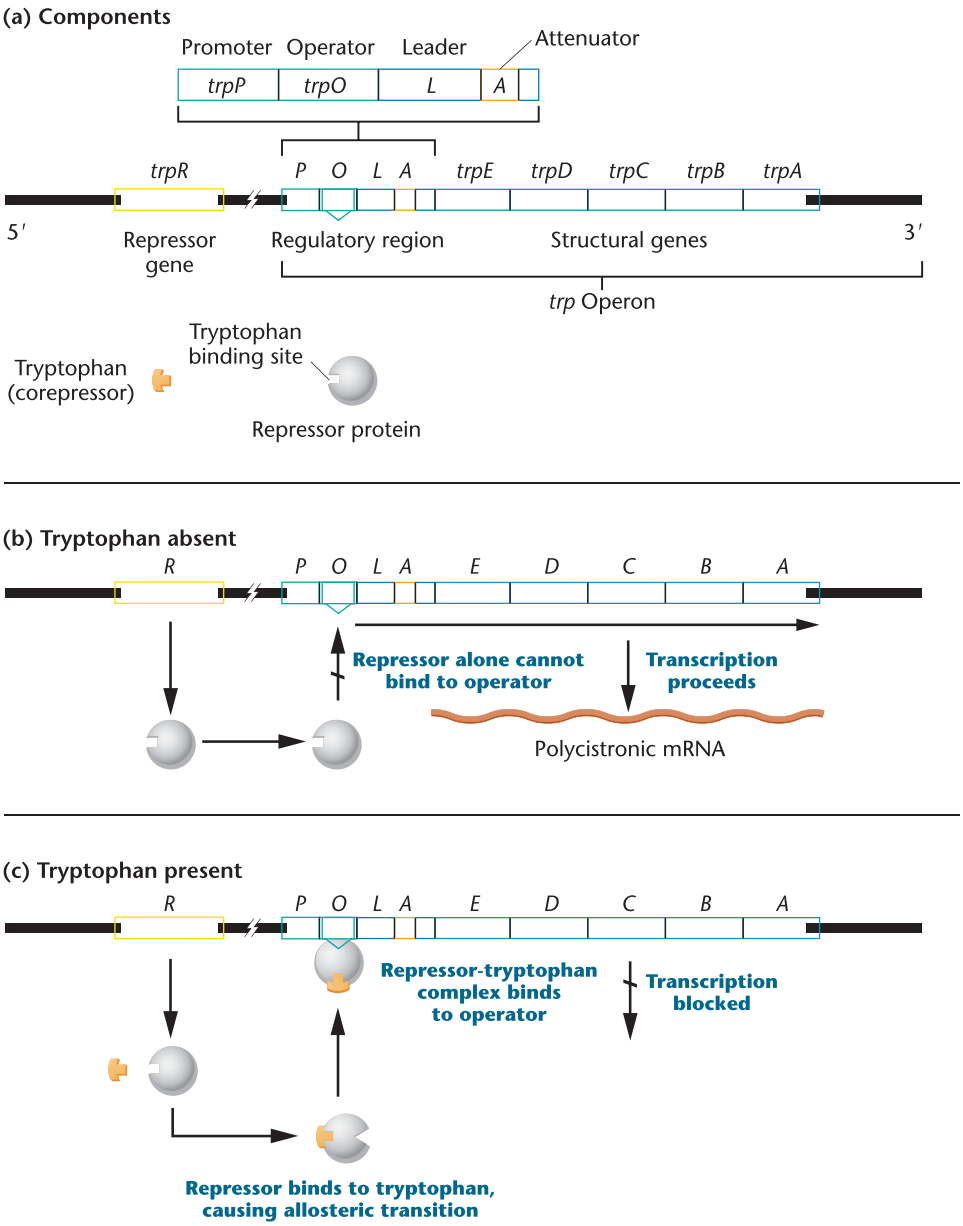

FIGURE 16–12 A repressible operon. (a) The components involved in regulation of the tryptophan operon. (b) In the absence of tryptophan, an inactive repressor is made that cannot bind to the operator (*O*), thus allowing transcription to proceed. (c) When tryptophan is present, it binds to the repressor, causing an allosteric transition to occur. This complex binds to the operator region, leading to repression of the operon.

Evidence for the *trp* Operon

Support for the concept of a repressible operon was soon forthcoming, based primarily on the isolation of two distinct categories of constitutive mutations. The first class, *trpR*$^-$, maps at a considerable distance from the structural genes. This locus represents the gene coding for the repressor. Presumably, the mutation inhibits either the repressor's interaction with tryptophan or repressor formation entirely. Whichever the case, no repression ever occurs in cells with the *trpR*$^-$ mutation. As expected, if the *trpR*$^+$ gene encodes a functional repressor molecule, the presence of a copy of this gene will restore repressibility.

The second constitutive mutation is analogous to that of the operator of the lactose operon, because it maps immediately adjacent to the structural genes. Furthermore, the addition of a wild-type operator gene into mutant cells (as an external element) does not restore repression. This is what would be predicted if the mutant operator no longer interacts with the repressor–tryptophan complex.

The entire *trp* operon has now been well defined, as shown in Figure 16–12. The five contiguous structural genes (*trp E*, *D*, *C*, *B*, and *A*) are transcribed as a polycistronic message directing translation of the enzymes that catalyze the biosynthesis of tryptophan. As in the *lac* operon, a promoter region (*trpP*) represents the binding site for RNA polymerase, and an operator region (*trpO*) binds the repressor. In the absence of binding, transcription is initiated within the *trpP–trpO* region and proceeds along a **leader sequence** 162 nucleotides prior to the first structural gene (*trpE*). Within that leader sequence, still another regulatory site has been demonstrated, called an *attenuator*—the subject of the next section. As we will see, this regulatory unit is an integral part of this operon's control mechanism.

16.6 Attenuation Is a Process Critical to the Regulation of the *trp* Operon in *E. coli*

The preceding section has established that the genes within the *trp* operon are expressed only when the quantity of tryptophan is severely limiting in the cell, but repressed in the presence of tryptophan. However, in the presence of low levels of tryptophan, transcription of the upstream region of the operon may nevertheless be initiated. This discovery was made by Charles Yanofsky, Kevin Bertrand, and their colleagues, who realized that once transcription is initiated, an independent mechanism exists that effectively represses the expression of the operon if tryptophan is present. They named this mechanism **attenuation**, a word derived from the verb *attenuate,* which means "to weaken or impair." After many years of work, this group discovered that following initiation of transcription of the *trp* operon, mRNA synthesis may be terminated at a point about 140 nucleotides along the transcript, prior to encountering the DNA template representing the structural genes. This upstream RNA molecule is referred to as the *5′-leader sequence*. Obviously, in the absence of tryptophan, attenuation is somehow overcome, allowing transcription of the entire operon.

Identification of the site involved in attenuation was made possible by the isolation of various deletion mutations in a 25-nt region 115 to 140 nucleotides into the leader sequence. Such mutations abolish attenuation. This site is referred to as the *attenuator.* An explanation of how attenuation occurs and how it is overcome, as presented by Yanofsky and colleagues, is summarized in Figure 16–13. The operon and the initial DNA sequence that is transcribed is shown in Figure 16–13(a). Transcription initially gives rise to the leader sequence [Figure 16–13(b)], which has the potential to fold into either of two mutually exclusive stem-loop structures described as *hairpins.* In the presence of tryptophan, the hairpin that is formed behaves as a *terminator,* causing transcription to be terminated prematurely. On the other hand, if tryptophan is limiting, the alternative hairpin structure, referred to as the *antiterminator,* is formed. In this case, transcription proceeds beyond the antiterminator region, and the mRNA representing the entire operon is subsequently produced. The hairpins are examples of *secondary RNA structures* and are shown in Figure 16–13(c).

The major issue remains as to how the presence of tryptophan leads to attenuation and how the absence of tryptophan circumvents it. The Yanofsky model, a revelation in the study of regulatory genetics, relies on two key points: (1) Bacterial ribosomes may attach to mRNA molecules and initiate *translation* before *transcription* of a gene (or operon) is complete (see Figure 13–15); and (2) the leader region of mRNA includes several tryptophan codons. Yanofsky proposed that if the tryptophan codons are *translated,* the *terminator hairpin* is formed. However, when tryptophan is very scarce or absent in the cell, inadequate charged $tRNA^{Trp}$ is available and translation stalls at the tryptophan codons. This translational stalling induces the *antiterminator hairpin* to form, allowing *transcription* to proceed throughout the operon.

The above model represents a unique approach to gene regulation, which can be termed *translational regulation of transcription.* The bottom line is that availability of $tRNA^{Trp}$, the formation of which is dependent on tryptophan levels, is critical to this mechanism that fine tunes the regulation of the operon. This model is supported experimentally by a number of genetic studies, whereby various regions of the leader sequence representing regulatory stems have been mutated or deleted. In each case, the predictions of such genetic alterations have been upheld.

Because regulation by attenuation is dependent on simultaneous transcription and translation, a linkage that is impossible in eukaryotes, the process is limited to prokaryotes. It has now been established in other bacterial operons in *E. coli,* including those responsible for the biosynthesis of amino acids, including threonine, histidine, leucine, and phenylalanine. As with the *trp* operon, attenuators in these operons contain multiple codons calling for regulation of the amino acid. When the relevant amino acid is unavailable, translation "stalls," allowing transcription to proceed. For example, the leader sequence in the histidine operon encodes seven contiguous histidine residues.

TRAP and AT Proteins Govern Attenuation in *B. subtilis*

Keep in mind that not all organisms, even those of the same type, solve problems such as gene regulation in exactly the same way. Thus, it is not unusual for geneticists to discover that new strategies have arisen during evolution. Often, a

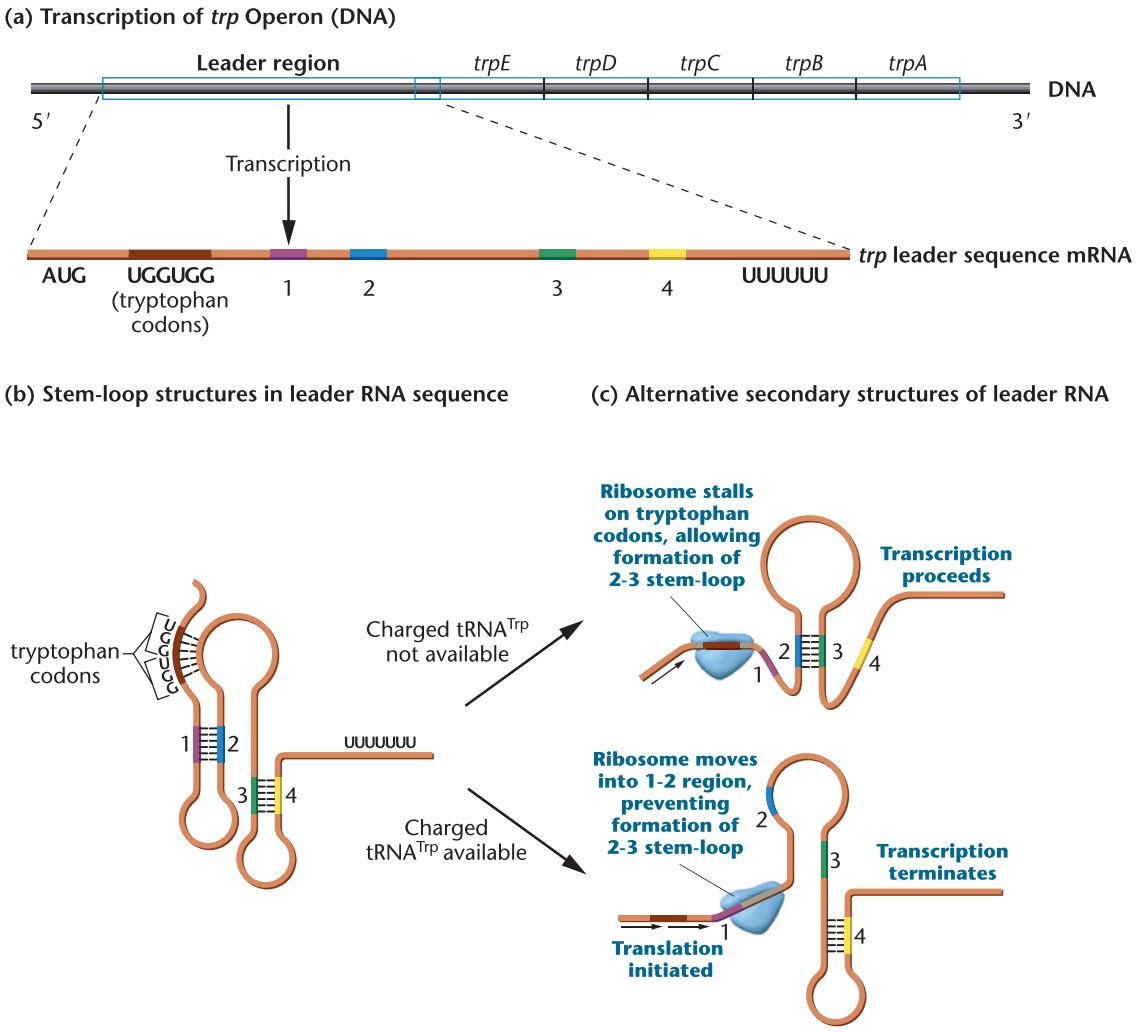

FIGURE 16–13 Attenuation in the tryptophan operon. (a) Transcription of the *trp* operon, showing the leader region followed by the five *trp* genes. The leader region in the mRNA is enlarged to show the translation start codon (AUG), the two tryptophan codons (UGGUGG), and the string of U residues at the end of the leader. The four regions marked 1, 2, 3, and 4 indicate the location of four sequences that have the potential to form stems by base pairing. (b) Two possible stem-loop structures that can form within the leader sequence mRNA. (c) Two alternative secondary structures that can be formed within the leader mRNA sequence during translation of the mRNA, depending on the availability of tryptophan. The first (at the top) occurs in the absence of tryptophan and thus in the absence of its charged cognate tRNA. As a result, a stem-loop involving regions 2 and 3 forms. During translation, the ribosome stalls at the trp codons prior to region 1, forming an antiterminator conformation. As a result, transcription by RNA polymerase continues. At the bottom, in the presence of tryptophan, and thus its charged cognate tRNA, the ribosome continues translation, moving past region 1 and the tryptophan codons, which inhibits the formation of the antiterminator region 2-3 stem-loop. As a result, the antiterminator signal is disrupted and transcription terminates. It is this process that is referred to as attenuation.

new approach is a variation on a well-established theme. Such is the case with the regulation of the *trp* operon in different bacteria.

As we saw previously, *E. coli*, a Gram-negative bacterium, uses charged $tRNA^{Trp}$ and a terminator hairpin in the leader sequence of the transcript as machinery for attenuating transcription of its *trp* operon. The Gram-positive bacterium *Bacillus subtilis* also uses attenuation and hairpins to regulate its *trp* operon. In fact, *B. subtilis* relies on attenuation as its sole mechanism of regulation because that bacterium lacks a mechanism that represses transcription entirely in this operon, as is present in *E. coli*.

However, the molecular signals that cause attenuation in *B. subtilis* do not use a process of translation and stalling, as *E. coli* does, to induce the hairpin that terminates transcription. Instead, a specific protein, isolated in the 1990s by Charles Yanofsky and coworkers, either binds or does not bind to the attenuator leader sequence, thereby inducing the alternative terminator or antiterminator configurations, respectively.

Attenuation is accomplished in *B. subtilis* in a unique way. The protein, ***trp* RNA-binding attenuation protein** (**TRAP**), binds to tryptophan if it is present in the cell. TRAP consists of 11 subunits, forming a symmetrical quaternary protein structure. Each subunit can bind one molecule of

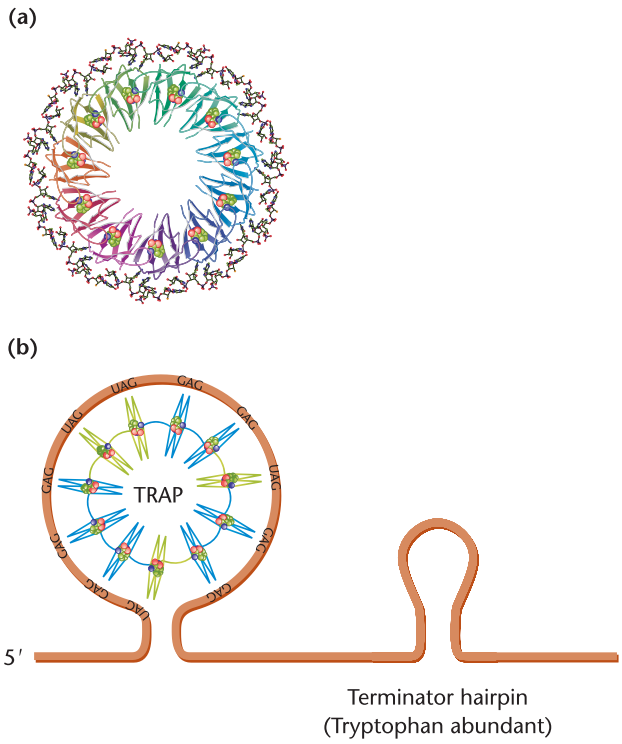

FIGURE 16–14 Model of the *trp* RNA-binding attenuation protein (TRAP). The symmetrical molecule consists of 11 subunits, each capable of binding to one molecule of tryptophan that is embedded within the subunit. (b) The interaction of a tryptophan-bound TRAP molecule with the leader sequence of the *trp* operon of *B. subtilis*. This interaction induces the formation of the terminator hairpin, attenuating expression of the *trp* operon.

tryptophan, incorporating it into a deep pocket within the protein [Figure 16–14(a)]. When fully saturated with tryptophan, this protein can bind to the 5′-leader sequence of the RNA transcript, which contains 11 triplet repeats of either GAG or UAG, each separated by several spacer nucleotides. Each triplet is linked to one of the subunits, which contains a binding pocket for the triplet. When bound to the transcript, the TRAP is encircled by a belt of RNA [Figure 16–14(b)], and this conformation prevents the antiterminator hairpin from forming. Instead, the terminator configuration results, leading to premature termination of transcription and, consequently, attenuation of expression of the operon.

The attenuation strategy of *B. subtilis* involves an interesting mechanism of regulation. Two observations had suggested to Yanofsky and his colleagues that the attenuation strategy in *B. subtilis* might be even more complex than had originally appeared. First, regulation of the *trp* operon in *B. subtilis* is extremely sensitive to a wide range of tryptophan concentrations. Such a fine tuning suggests that something more than the simple on–off mechanism attributed to TRAP might be at work. Second, a mutation in the gene encoding tryptophanyl-tRNA synthetase leads to the overexpression of the *trp* operon, even in the presence of excess tryptophan. This enzyme is responsible for charging $tRNA^{Trp}$; in its absence, uncharged $tRNA^{Trp}$ molecules accumulate, and regulation is interrupted. The interruption of regulation would suggest that uncharged $tRNA^{Trp}$ plays an essential role in attenuation, leading Yanofsky and his coworker Angela Valbuzzi to ask how $tRNA^{Trp}$ might be involved with TRAP. What they found extends our knowledge of the system of regulation.

They discovered that there exists still another protein, **anti-TRAP (AT)**, which provides a metabolic signal that $tRNA^{Trp}$ is uncharged, thereby indicating that tryptophan is very scarce in the cell. Yanofsky and Valbuzzi theorized that uncharged $tRNA^{Trp}$ induces a separate operon to express the *AT* gene. The AT protein then associates with TRAP, specifically when it is in the tryptophan-activated state, and inhibits the binding of TRAP to its target leader RNA sequence.

Such a finding not only adds yet another dimension to our understanding of this system of gene regulation, but it also explains the original observation of the mutation in the $tRNA^{Trp}$ synthetase gene that leads to overexpression of the *trp* operon. The mutation prevents charging of $tRNA^{Trp}$, even in the presence of tryptophan. As uncharged $tRNA^{Trp}$ accumulates, it induces AT, which binds to TRAP. The TRAP is then unable to bind to the leader RNA sequence; as a result, the *trp* operon is overinduced.

The preceding description demonstrates the complexity of regulatory mechanisms in bacteria. That such intricate strategies have resulted from the evolutionary process attests to the critical importance of carefully regulating gene expression in bacteria.

NOW SOLVE THIS

16–3 Consider the regulation of the *trp* operon in *B. subtilis*. For each of the following cases, indicate whether the structural genes in the operon are being expressed or not expressed:

(a) TRAP present, $tRNA^{Trp}$ abundant, tryptophan abundant.
(b) TRAP present, $tRNA^{Trp}$ abundant, tryptophan scarce.
(c) TRAP present, $tRNA^{Trp}$ scarce, tryptophan abundant.
(d) TRAP absent, $tRNA^{Trp}$ abundant, tryptophan abundant.

In each case, also indicate whether AT is present. Describe the role of TRAP and AT in attenuation.

HINT: *This problem concerns the regulation of the B. subtilis trp operon, which involves the TRAP protein and requires you to understand the concept of attenuation. You are asked to predict whether or not the structural genes are expressed under various conditions. The key to its solution is to realize that unlike E. coli, which uses tryptophan as a repressor along with the process of attenuation in regulating the trp operon, B. subtilis relies solely on the process of attenuation and the TRAP protein for regulation of this operon.*

16.7 Riboswitches Utilize Metabolite-sensing RNAs to Regulate Gene Expression

Since the elucidation of attenuation in the *trp* operon, numerous cases of gene regulation that also depend on alternative forms of mRNA secondary structure have been documented. These involve what are called **riboswitches,** which are mRNA sequences (or elements), present in the 5′-untranslated region (5′-UTR) upstream from the coding sequences. These elements are capable of binding with small molecule ligands, such as metabolites, whose synthesis or activity is controlled by the genes encoded by the mRNA. Such binding causes a conformational change in one domain of the riboswitch element, which induces another change at a second RNA domain, most often creating a *terminator structure.* This terminator region interfaces directly with the transcriptional machinery and shuts it down. Thus, the mRNA is able to respond to the environment of the cell and regulate its own expression. Common to all riboswitches, then, is the presence of a *metabolite-sensing RNA sequence* that can adopt one of two secondary structures that allows transcription of that RNA to either proceed or not to proceed. The decision depends on the concentration of a molecular component of the cell whose existence is influenced by the genes encoded by that RNA.

Many classes of riboswitches recognize a broad range of ligands, including amino acids, purines, vitamin cofactors, aminosugars, and metal ions, among others. They are widespread in bacteria. In *Bacillus subtilis,* for example, approximately 5 percent of this bacterium's genes are regulated by riboswitches. They are also found in archaea, fungi, and plants, and may prove to be present in animals as well.

The two important domains within a riboswitch are called the **aptamer**, which binds to the ligand, and the **expression platform**, which is capable of forming the terminator structure. Figure 16–15 illustrates the principles involved in riboswitch control. The 5′-UTR of an mRNA is shown on the left side of the figure in the absence of the ligand (metabolite). RNA polymerase has transcribed the unbound aptamer domain, and in the *default conformation,* the expression domain adopts an *antiterminator conformation.* Thus, transcription continues through the expression platform and into the coding region. On the right side of the figure, in the presence of the ligand, it binds to the aptamer domain, inducing an alternative conformation in the aptamer, which in turn induces a change in the expression domain, creating the *terminator conformation.* RNA polymerase is effectively blocked and transcription ceases.

While this description fits the majority of riboswitches that have been studied, there are two variations from this model that complete our coverage of the topic. First, it is possible for the default position of a riboswitch to be in the *terminator conformation.* For example, consider the 5′UTRs representing genes encoding aminoacyl tRNA synthetases, the enzymes that charge tRNAs with their cognate amino acid (see Chapter 14). In these cases, uncharged tRNAs serve as the ligand and bind to the aptamer of the 5′UTR, inducing the *antiterminator conformation.* This allows the transcription of the genes, leading to the subsequent translation of the enzymes that are required to charge tRNAs. When charged tRNAs are in abundance and uncharged tRNAs are

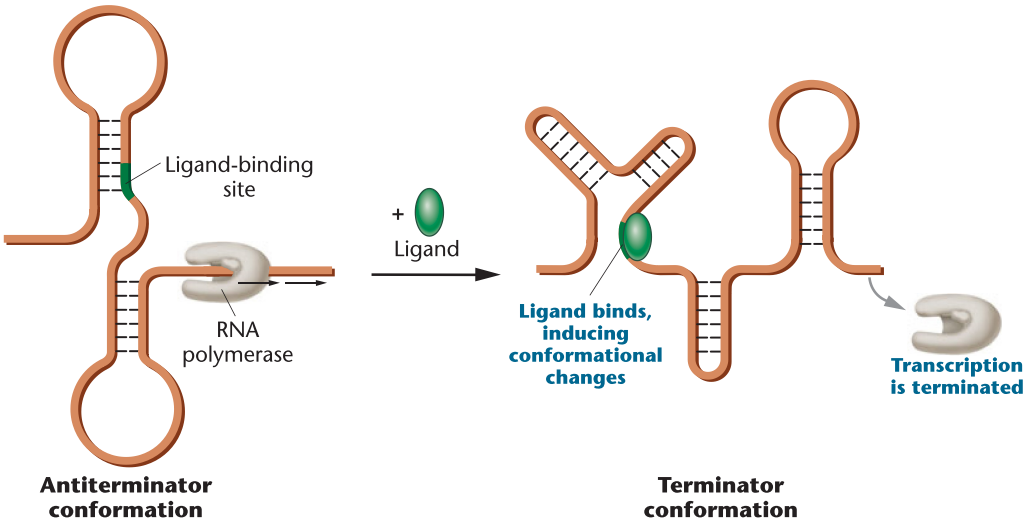

FIGURE 16–15 Illustration of the mechanism of riboswitch regulation of gene expression, where the default position (left) is in the antiterminator conformation and, upon binding by the ligand, adopts the terminator conformation.

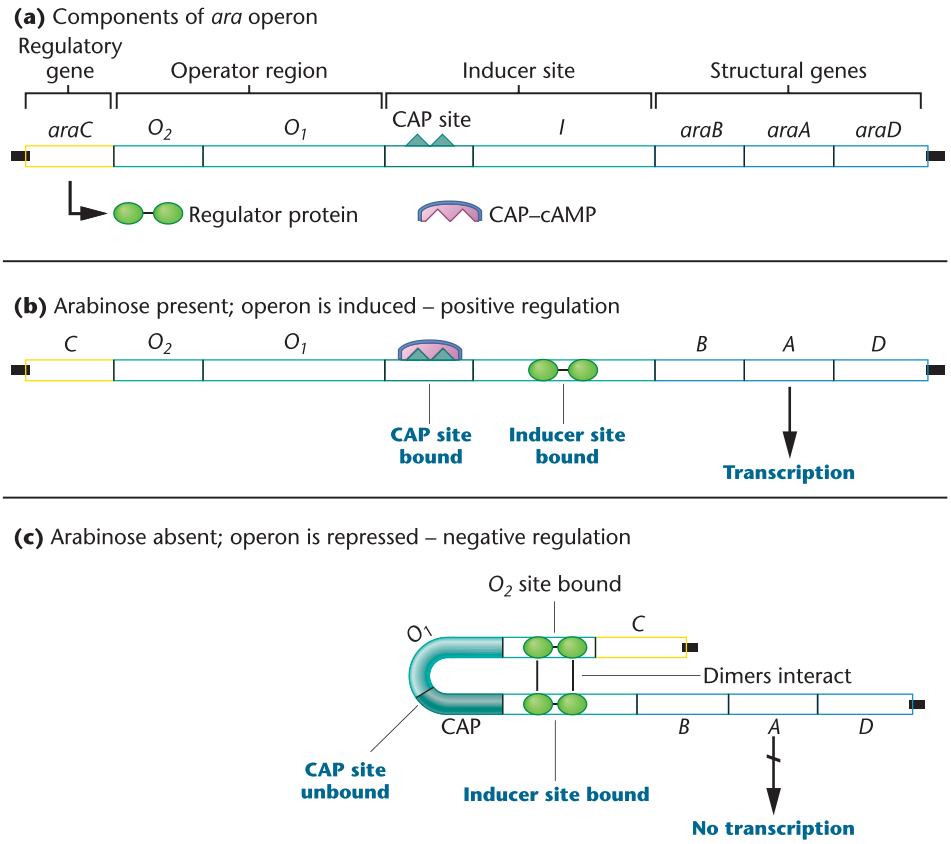

FIGURE 16–16 Genetic regulation of the *ara* operon. The regulatory protein of the *araC* gene acts as either an inducer (in the presence of arabinose) or a repressor (in the absence of arabinose).

not present to bind to the aptamer, the riboswitch, in the default condition, adapts the *terminator conformation,* effectively shutting down transcription. As predicted, charged tRNAs are unable to bind to the aptamer.

Finally, there are examples of bacterial riboswitches where, like attenuation of the *trp* operon discussed earlier, the alternative 5′UTR conformations allow transcription to be completed, but ligand binding to the aptamer induces a conformation that inhibits translation by ribosomes. In such cases, regulation of gene expression has been invoked at the level of translation.

16.8 The *ara* Operon Is Controlled by a Regulator Protein That Exerts Both Positive and Negative Control

We conclude this chapter with a brief discussion of the **arabinose (*ara*) operon** as analyzed in *E. coli.* This inducible operon is unique because the same regulatory protein is capable of exerting both positive and negative control, and as a result, at some times induces and at others represses gene expression. Figure 16–16(a) shows the operon components, and Figure 16–16(b) and (c) show, respectively, the conditions under which the operon is active or inactive.

1. The metabolism of the sugar arabinose is governed by the enzymatic products of three structural genes, *ara B*, *A*, and *D* [Figure 16–16(a)]. Their transcription is controlled by the regulatory protein AraC, encoded by the *araC* gene, which interacts with two regulatory regions, *araI* and *araO*$_2$.* These sites can be bound individually or coordinately by the AraC protein.

2. The *I* region bears that designation because when it is the only region bound by AraC, the system is induced [Figure 16–16(b)]. For this binding to occur, both arabinose and cAMP must be present. Thus, like induction of the *lac* operon, a CAP-binding site is present in the promoter region that modulates catabolite repression in the presence of glucose.

*An additional operator O_1 region is also present, but is not involved in the regulation of the *ara* structural gene.

3. In the absence of both arabinose and cAMP, the AraC protein binds coordinately to both the *I* site *and* the O_2 site (so named because it was the second operator region to be discovered in this operon). When both *I* and O_2 are bound by AraC, a conformational change occurs in the DNA to form a tight loop [Figure 16–16(c)], and the structural genes are repressed.
4. The O_2 region is located about 200 nucleotides upstream from the *I* region. Binding at either regulatory region involves a dimer of AraC. When both regions are bound, the dimers interact, producing the loop that causes repression. Presumably, this loop inhibits the access of RNA polymerase to the promoter region. The region of DNA that loops out (the stretch between *I* and O_2) is critical to the formation of the repression complex. Genetic alteration here by insertion or deletion of even a few nucleotides is sufficient to interfere with loop formation and, therefore, repression.

Our observations involving the *ara* operon illustrate the degree of complexity that may be seen in the regulation of a group of related genes. As we mentioned at the beginning of this chapter, the development of regulatory mechanisms has provided evolutionary advantages that allow bacterial systems to adjust to a range of natural environments. Without question, these systems are well equipped genetically not only to survive under varying physiological conditions but to do so with great biochemical efficiency.

CASE STUDY | Food poisoning and bacterial gene expression

At a Midwestern university, several undergraduates reported to the health service with fever, muscle aches, and vomiting. Blood tests showed they had bacterial food poisoning, caused by *Listeria monocytogenes*. One student, a biology major, asked how the bacteria were able to get from the intestine into the bloodstream. The physician explained that *Listeria* cells that express a cell-surface protein, InlA, are able to use receptors on the surface of intestinal cells to gain entry and spread from there to the blood. Once in the blood, the bacteria reach the liver, where they use another cell-surface protein InlB, to enter liver cells, which are the main sites of infection. She further explained that transcriptional levels of the *inlA* and *inlB* genes determine the infective ability of different *Listeria* strains.

Later, the student discussed the incident with his genetics instructor, and together they pondered several questions about how bacteria are able to hijack the normal functions of receptors on human cells as part of their life cycle.

1. How can we tell if the *inlA* and *inlB* genes are expressed constitutively or are induced in the presence of the human receptors?
2. Might these genes be normally expressed at low levels, but undergo increased transcription once in contact with intestinal epithelial cells? How could this be determined?
3. Are the *inlA* and *inlB* genes part of an operon, or do they use a common promoter? What experiments might show this?
4. How might an understanding of the regulation of these bacterial genes be used to treat infections?

Summary Points

For activities, animations, and review quizzes, go to the study area at www.masteringgenetics.com

1. Research on the *lac* operon in *E. coli* pioneered our understanding of gene regulation in bacteria.
2. Genes involved in the metabolism of lactose are coordinately regulated by a negative control system that responds to the presence or absence of lactose.
3. The catabolite-activating protein (CAP) exerts positive control over *lac* gene expression by interacting with RNA polymerase at the *lac* promoter and by responding to the levels of cyclic AMP in the bacterial cell.
4. The *lac* repressor of *E. coli* has been isolated and studied. Crystal structure analysis has shown how it interacts with the DNA of the operon as well as with inducers, revealing conformational changes in DNA leading to the formation of a repression loop that inhibits binding between RNA polymerase and the promoter region of the operon.
5. Unlike the inducible *lac* operon, the *trp* operon is repressible. In the presence of tryptophan, the repressor binds to the regulatory region of the *trp* operon and represses transcription initiation.
6. The process of attenuation, which regulates operons based on the presence of an end product, involves alterations to mRNA secondary structure, leading to premature termination of transcription.
7. Riboswitches regulate gene expression by virtue of conformational changes in the secondary structure of the 5'UTR leader region of mRNAs.
8. The *ara* operon in *E. coli* is unique in that the regulator protein exerts both positive and negative control over expression of the genes specifying the enzymes that metabolize the sugar arabinose.

GENETICS, TECHNOLOGY, AND SOCIETY

Quorum Sensing: Social Networking in the Bacterial World

For decades, scientists regarded bacteria as independent single-celled organisms, incapable of cell-to-cell communication. However, recent research is revealing that many bacteria can regulate gene expression and coordinate group behavior through a form of communication termed *quorum sensing*. Through this process, bacteria send and receive chemical signals called autoinducers that relay information about population size. When the population size reaches a "quorum," defined in the business world as the minimum number of members of an organization that must be present to conduct business, the autoinducers regulate gene expression in a way that benefits the group as a whole. Quorum sensing has been described in more than 70 species of bacteria, and its uses range from controlling bioluminescence in marine bacteria to regulating the expression of virulence factors in pathogenic bacteria. Our understanding of quorum sensing has altered our perceptions of prokaryotic gene regulation and is leading to the development of practical applications, including new antibiotic drugs.

Quorum sensing was discovered in the 1960s, during research on the bioluminescent bacterium *Vibrio fischeri* which lives in a symbiotic relationship with the Hawaiian bobtail squid *Euprymna scolopes*. While hunting for food at night, the squid uses light emitted by the *V. fischeri* present in its light organ to illuminate the ocean floor and to counter the shadows created by moonlight that normally act as a beacon for the squid's predators. In return, the bacteria gain a protected, nutrient-rich environment in the squid's light organ. During the day, the bacteria do not glow as a result of the squid's ability to reduce the concentration of bacteria in its light organ, which in turn prevents expression of the bacterial luciferase (*lux*) operon.

What turns on the bacteria's *lux* genes in response to high cell density and off in response to low cell density? In *V. fischeri*, the responsible autoinducer is a secreted homoserine lactone molecule. At a critical population size, these molecules accumulate, are taken up by bacteria within the population, and regulate the *lux* operon by binding directly to transcription factors that stimulate *lux* gene expression.

Since the discovery of quorum sensing in *Vibrio fischeri*, scientists have identified similar microbial communication systems in other bacteria, including significant human pathogens such as pseudomonal, staphylococcal, and streptococcal species. The expression of as many as 15 percent of bacterial genes may be regulated by quorum sensing.

Quorum sensing molecules may also mediate communication among members of different species. In 1994, Bonnie Bassler and her colleagues at Princeton University discovered an autoinducer molecule in the marine bacterium *Vibrio harveyi* that was also present in many diverse types of bacteria. This molecule, autoinducer-2 (AI-2), has the potential to mediate "quorum-sensing cross talk" between species and thus serve as a universal language for bacterial communication. Because the accumulation of AI-2 is proportional to cell number, and because the structure of AI-2 may vary slightly between different species, the current hypothesis is that AI-2 can transmit information about both the cell density and species composition of a bacterial community.

Pathogenic bacteria use quorum sensing to regulate the expression of genes whose products help these bacteria invade a host and avoid immune system detection. For example, *Vibrio cholerae*, the causative agent of cholera, uses AI-2 and an additional species-specific autoinducer to activate the genes controlling the production of cholera toxin. *Pseudomonas aeruginosa*, the Gram-negative bacterium that often affects cystic fibrosis patients, uses quorum sensing to regulate the production of elastase, a protease that disrupts the respiratory epithelium and interferes with ciliary function. *P. aeruginosa* also uses autoinducers to control the production of biofilms, tough protective shells that resist host defenses and make treatment with antibiotics nearly impossible. Other bacteria determine cell density through quorum sensing to delay the production of toxic substances until the colony is large enough to overpower the host's immune system and establish an infection. Because many bacteria rely on quorum sensing to regulate disease-causing genes, therapeutics that block quorum sensing may help combat infections. Research into these potential therapies is now in progress, and several are now approaching the clinical trial phase. Thus, what began as a fascinating observation in the glowing squid has launched an exciting era of research in bacterial genetics that may one day prove of great clinical significance.

Your Turn

Take time, individually or in groups, to answer the following questions. Investigate the references and links to help you understand the mechanisms and potential uses of quorum sensing in bacteria.

1. Inhibitors of quorum sensing molecules have potential as antibacterial agents. What are some ways in which quorum sensing inhibitors could work to combat bacterial infections? Have any of these therapeutics reached clinical trials?

A recent review of quorum sensing therapeutics can be found in Njoroge, J. and Sperandio, V. 2009. Jamming bacterial communication: New approaches for the treatment of infectious diseases. *EMBO Mol. Med.* 1(4): 201–210.

2. Regulation of bacterial gene expression by autoinducer molecules involves several different mechanisms. Describe these mechanisms and how each could be used as a target for the control of bacterial infections.

The mechanisms by which autoinducers regulate gene expression are summarized in Asad, S. and Opal, S.M. 2008. Bench-to-bedside review: Quorum sensing and the role of cell-to-cell communication during invasive bacterial infection. *Critical Care* 12: 236–247.

3. Quorum sensing systems are also capable of detecting and responding to chemical signals given off by host cells. Explain how this works and how this might benefit pathogenic bacteria.

A review article dealing with interkingdom communication and quorum sensing can be found at Wagner, V.E. et al. 2006. Quorum sensing: dynamic response of Pseudomonas aeruginosa to external signals. *Trends Microbiol.* 14(2): 55–58.

INSIGHTS AND SOLUTIONS

1. A hypothetical operon (*theo*) in *E. coli* contains several structural genes encoding enzymes that are involved sequentially in the biosynthesis of an amino acid. Unlike the *lac* operon, in which the repressor gene is separate from the operon, the gene encoding the regulator molecule is contained within the *theo* operon. When the end product (the amino acid) is present, it combines with the regulator molecule, and this complex binds to the operator, repressing the operon. In the absence of the amino acid, the regulatory molecule fails to bind to the operator, and transcription proceeds.

   Categorize and characterize this operon, then consider the following mutations, as well as the situation in which the wild-type gene is present along with the mutant gene in partially diploid cells (F′):

   (a) Mutation in the operator region.

   (b) Mutation in the promoter region.

   (c) Mutation in the regulator gene.

   In each case, will the operon be active or inactive in transcription, assuming that the mutation affects the regulation of the *theo* operon? Compare each response with the equivalent situation for the *lac* operon.

   Solution: The *theo* operon is repressible and under negative control. When there is no amino acid present in the medium (or the environment), the product of the regulatory gene cannot bind to the operator region, and transcription proceeds under the direction of RNA polymerase. The enzymes necessary for the synthesis of the amino acid are produced, as is the regulator molecule. If the amino acid *is* present, either initially or after sufficient synthesis has occurred, the amino acid binds to the regulator, forming a complex that interacts with the operator region, causing repression of transcription of the genes within the operon.

   The *theo* operon is similar to the tryptophan system, except that the regulator gene is within the operon rather than separate from it. Therefore, in the *theo* operon, the regulator gene is itself regulated by the presence or absence of the amino acid.

   (a) As in the *lac* operon, a mutation in the *theo* operator gene inhibits binding with the repressor complex, and transcription occurs constitutively. The presence of an F′ plasmid bearing the wild-type allele would have no effect, since it is not adjacent to the structural genes.

   (b) A mutation in the *theo* promoter region would no doubt inhibit binding to RNA polymerase and therefore inhibit transcription. This would also happen in the *lac* operon. A wild-type allele present in an F′ plasmid would have no effect.

   (c) A mutation in the *theo* regulator gene, as in the *lac* system, may inhibit either its binding to the repressor or its binding to the operator gene. In both cases, transcription will be constitutive, because the *theo* system is repressible. Both cases result in the failure of the regulator to bind to the operator, allowing transcription to proceed. In the *lac* system, failure to bind the corepressor lactose would permanently repress the system. The addition of a wild-type allele would restore repressibility, provided that this gene was transcribed constitutively.

Problems and Discussion Questions

For instructor-assigned tutorials and problems, go to www.masteringgenetics.com

HOW DO WE KNOW?

1. **In this chapter, we focused on the regulation of gene expression in prokaryotes. Along the way, we found many opportunities to consider the methods and reasoning by which much of this information was acquired. From the explanations given in the chapter, what answers would you propose to the following fundamental questions?**
   (a) How do we know that bacteria regulate the expression of certain genes in response to the environment?
   (b) What evidence established that lactose serves as the inducer of a gene whose product is related to lactose metabolism?
   (c) What led researchers to conclude that a repressor molecule regulates the lac operon?
   (d) How do we know that the *lac* repressor is a protein?
   (e) How do we know that the *trp* operon is a repressible control system, in contrast to the *lac* operon, which is an inducible control system?
2. Contrast the need for the enzymes involved in lactose and tryptophan metabolism in bacteria when lactose and tryptophan, respectively, are (a) present and (b) absent.
3. Contrast positive versus negative control of gene expression.
4. Contrast the role of the repressor in an inducible system and in a repressible system.
5. For the *lac* genotypes shown in the accompanying table, predict whether the structural genes (*Z*) are constitutive, permanently repressed, or inducible in the presence of lactose.

Genotype	Constitutive	Repressed	Inducible
$I^+O^+Z^+$			×
$I^-O^+Z^+$			
$I^-O^cZ^+$			
$I^-O^cZ^+/F'O^+$			
$I^+O^cZ^+/F'O^+$			
$I^sO^+Z^+$			
$I^sO^+Z^+/F'I^+$			

6. For the genotypes and conditions (lactose present or absent) shown in the following table, predict whether functional enzymes, nonfunctional enzymes, or no enzymes are made.

Genotype	Condition	Functional Enzyme Made	Nonfunctional Enzyme Made	No Enzyme Made
$I^+O^+Z^+$	No lactose			×
$I^+O^cZ^+$	Lactose			
$I^-O^+Z^-$	No lactose			
$I^-O^+Z^-$	Lactose			
$I^-O^+Z^+/F'I^+$	No lactose			
$I^+O^cZ^+/F'O^+$	Lactose			
$I^+O^+Z^-/F'I^+O^+Z^+$	Lactose			
$I^-O^+Z^-/F'I^+O^+Z^+$	No lactose			
$I^sO^+Z^+/F'O^+$	No lactose			
$I^+O^cZ^+/F'O^+Z^+$	Lactose			

7. The locations of numerous *lacI*$^-$ and *lacI*^{S} mutations have been determined within the DNA sequence of the *lacI* gene. Among these, *lacI*$^-$ mutations were found to occur in the 5′-upstream region of the gene, while *lacI*^{S} mutations were found to occur farther downstream in the gene. Are the locations of the two types of mutations within the gene consistent with what is known about the function of the repressor that is the product of the *lacI* gene?
8. Describe the experimental rationale that allowed the *lac* repressor to be isolated.
9. What properties demonstrate the *lac* repressor to be a protein? Describe the evidence that it indeed serves as a repressor within the operon system.
10. Predict the effect on the inducibility of the *lac* operon of a mutation that disrupts the function of (a) the *crp* gene, which encodes the CAP protein, and (b) the CAP-binding site within the promoter.
11. Erythritol, a natural sugar abundant in fruits and fermenting foods, is about 65 percent as sweet as table sugar and has about 95 percent less calories. It is "tooth friendly" and generally devoid of negative side effects as a human consumable product. Pathogenic *Brucella* strains that catabolize erythritol contain four closely spaced genes, all involved in erythritol metabolism. One of the four genes (*eryD*) encodes a product that represses the expression of the other three genes. Erythritol catabolism is stimulated by erythritol. Present a simple regulatory model to account for the regulation of erythritol catabolism in *Brucella*. Does this system appear to be under inducible or repressible control?
12. Describe the role of attenuation in the regulation of tryptophan biosynthesis.
13. Attenuation of the *trp* operon was viewed as a relatively inefficient way to achieve genetic regulation when it was first discovered in the 1970s. Since then, however, attenuation has been found to be a relatively common regulatory strategy. Assuming that attenuation is a relatively inefficient way to achieve genetic regulation, what might explain its widespread occurrence?
14. Neelaredoxin is a 15-kDa protein that is a gene product common in anaerobic prokaryotes. It has superoxide-scavenging activity, and it is *constitutively expressed*. In addition, its expression is not further *induced* during its exposure to O_2 or H_2O_2 (Silva, G., et al. 2001. *J. Bacteriol.* 183: 4413–4420). What do the terms *constitutively expressed* and *induced* mean in terms of neelaredoxin synthesis?
15. Milk products such as cheeses and yogurts are dependent on the conversion by various anaerobic bacteria, including several *Lactobacillus* species, of lactose to glucose and galactose, ultimately producing lactic acid. These conversions are dependent on both permease and β-galactosidase as part of the *lac* operon. After selection for rapid fermentation for the production of yogurt, one *Lactobacillus* subspecies lost its ability to regulate *lac* operon expression (Lapierre, L., et al. 2002. *J. Bacteriol.* 184: 928–935). Would you consider it likely that in this subspecies the *lac* operon is "on" or "off"? What genetic events would likely contribute to the loss of regulation as described above?
16. Assume that the structural genes of the *lac* operon have been fused, through recombinant DNA techniques, to the regulatory apparatus of the *ara* operon. If arabinose is provided in a minimal medium to *E. coli* carrying this gene fusion, would you expect β-galactosidase to be produced at induced levels? Explain.
17. Both attenuation of the *trp* operon in *E. coli* and riboswitches in *B. subtilis* rely on changes in the secondary structure of the leader regions of mRNA to regulate gene expression. Compare and contrast the specific mechanisms in these two types of regulation.
18. Keeping in mind the life cycle of bacteriophages discussed in Chapter 6, consider the following problem: During the reproductive cycle of a temperate bacteriophage, the viral DNA inserts into the bacterial chromosome where the resultant prophage behaves much like a Trojan horse. It can remain quiescent, or it can become lytic and initiate a burst of progeny viruses. Several operons maintain the prophage state by interacting with a repressor that keeps the lytic cycle in check. Insults (ultraviolet light, for example) to the bacterial cell lead to a partial breakdown of the repressor, which in turn causes the production of enzymes involved in the lytic cycle. As stated in this simple form, would you consider this system of regulation to be operating under positive or negative control?

Extra-Spicy Problems

For instructor-assigned tutorials and problems, go to www.masteringgentics.com

19. Bacterial strategies to evade natural or human-imposed antibiotics are varied and include membrane-bound efflux pumps that export antibiotics from the cell. A review of efflux pumps (Grkovic, S., et al., 2002) states that, because energy is required to drive the pumps, activating them in the absence of the antibiotic has a selective disadvantage. The review also states that a given antibiotic may play a role in the regulation of efflux by interacting with either an activator protein or a repressor protein, depending on the system involved. How might such systems be categorized in terms of *negative control* (*inducible* or *repressible*) or *positive control* (*inducible* or *repressible*)?

20. In a theoretical operon, genes *A*, *B*, *C*, and *D* represent the repressor gene, the promoter sequence, the operator gene, and the structural gene, *but not necessarily in the order named*. This operon is concerned with the metabolism of a theoretical molecule (tm). From the data provided in the accompanying table, first decide whether the operon is inducible or repressible. Then assign *A*, *B*, *C*, and *D* to the four parts of the operon. Explain your rationale. (AE = active enzyme; IE = inactive enzyme; NE = no enzyme)

Genotype	tm Present	tm Absent
$A^+B^+C^+D^+$	AE	NE
$A^-B^+C^+D^+$	AE	AE
$A^+B^-C^+D^+$	NE	NE
$A^+B^+C^-D^+$	IE	NE
$A^+B^+C^+D^-$	AE	AE
$A^-B^+C^+D^+/F'A^+B^+C^+D^+$	AE	AE
$A^+B^-C^+D^+/F'A^+B^+C^+D^+$	AE	NE
$A^+B^+C^-D^+/F'A^+B^+C^+D^+$	AE + IE	NE
$A^+B^+C^+D^-/F'A^+B^+C^+D^+$	AE	NE

21. A bacterial operon is responsible for the production of the biosynthetic enzymes needed to make the hypothetical amino acid tisophane (tis). The operon is regulated by a separate gene, *R*, deletion of which causes the loss of enzyme synthesis. In the wild-type condition, when tis is present, no enzymes are made; in the absence of tis, the enzymes are made. Mutations in the operator gene (O^-) result in repression regardless of the presence of tis. Is the operon under positive or negative control? Propose a model for (a) repression of the genes in the presence of tis in wild-type cells and (b) the mutations.
22. A marine bacterium is isolated and shown to contain an inducible operon whose genetic products metabolize oil when it is encountered in the environment. Investigation demonstrates that the operon is under positive control and that there is a *reg* gene whose product interacts with an operator region (*o*) to regulate the structural genes, designated *sg*.

    In an attempt to understand how the operon functions, a constitutive mutant strain and several partial diploid strains were isolated and tested with the results shown here:

Host Chromosome	F′ Factor	Phenotype
wild type	none	inducible
wild type	*reg* gene from mutant strain	inducible
wild type	operon from mutant strain	constitutive
mutant strain	*reg* gene from wild type	constitutive

    Draw all possible conclusions about the mutation as well as the nature of regulation of the operon. Is the constitutive mutation in the *trans*-acting *reg* element or in the *cis*-acting *o* operator element?
23. The SOS repair genes in *E. coli* (discussed in Chapter 13) are negatively regulated by the *lexA* gene product, called the LexA repressor. When a cell sustains extensive damage to its DNA, the LexA repressor is inactivated by the *recA* gene product (RecA), and transcription of the SOS genes is increased dramatically.

    One of the SOS genes is the *uvrA* gene. You are a student studying the function of the *uvrA* gene product in DNA repair. You isolate a mutant strain that shows constitutive expression of the UvrA protein. Naming this mutant strain $uvrA^C$, you construct the following diagram showing the *lexA* and *uvrA* operons:

    (a) Describe two different mutations that would result in a *uvrA* constitutive phenotype. Indicate the actual genotypes involved.
    (b) Outline a series of genetic experiments that would use partial diploid strains to determine which of the two possible mutations you have isolated.
24. A fellow student considers the issues in Problem 23 and argues that there is a more straightforward, nongenetic experiment that could differentiate between the two types of mutations. The experiment requires no fancy genetics and would allow you to easily assay the products of the other SOS genes. Propose such an experiment.
25. Figure 16–13 depicts numerous critical regions of the leader sequence of mRNA that play important roles during the process of attenuation in the *trp* operon. A closer view of the leader sequence, which begins at about position 30 downstream from the 5′ end, is given below (running across both columns of this page):

AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGAAACGGGCAGUGUAUUCACCAUGCGUAAAGCAAUCAGAUACCCAGCCCGCCUAAUGAGCGGGCUUUUUUUU

    Within this molecule are the sequences that cause the formation of the alternative hairpins. It also contains the successive triplets that encode tryptophan, where stalling during translation occurs. Take a large piece of paper (such as manila wrapping paper) and, along with several other students from your genetics class, work through the base sequence to identify the *trp* codons and the parts of the molecule representing the base-pairing regions that form the terminator and antiterminator hairpins shown in Figure 16–13.
26. One of the most prevalent sexually transmitted diseases is caused by the bacterium *Chlamydia trachomatis* and leads to blindness if left untreated. Upon infection, metabolically inert cells differentiate, through gene expression, to become metabolically active cells that divide by binary fission. It has been proposed that release from the inert state is dependent on heat-shock proteins that both activate the reproductive cycle and facilitate the binding of chlamydiae to host cells. Researchers made the following observations regarding the heat-shock regulatory system in *Chlamydia trachomatis*: (1) a regulator protein (call it R) binds to a *cis*-acting DNA element (call it D); (2) R and D function as a repressor–operator pair; (3) R functions as a negative regulator of transcription; (4) D is composed of an inverted-repeat sequence; (5) repression by R is dependent on D being supercoiled (Wilson & Tan, 2002).
    (a) Based on this information, devise a model to explain the heat-dependent regulation of metabolism in *Chlamydia trachomatis*.
    (b) Some bacteria, like *E. coli*, use a heat-shock sigma factor to regulate heat-shock transcription. Are the above findings in *Chlamydia* compatible with use of a heat-sensitive sigma factor?
27. A reporter gene—that is, a gene whose activity is relatively easy to detect—is commonly used to study promoter activity under a variety of investigative circumstances. Recombinant

DNA technology is used to attach such a reporter gene to the promoter of interest, after which the recombinant molecule is inserted by transformation or transfection into the cell or host organism of choice. Green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* fluoresces green when exposed to blue light. Assume that you wished to study the regulatory apparatus of the *lac* operon and you developed a plasmid (p) that contained all the elements of the *lac* operon except that you replaced the *lac* structural genes with the *GFP* gene (your reporter gene). You used this plasmid to transform *E. coli.* Considering the following genotypes, under which of the following conditions would you expect β-galactosidase production and/or green colonies when examined under blue light?

	Medium Condition			
Genotype	**Lactose**	**Glucose**	**β-galactosidase**	**Green Colonies**
$I^+O^+Z^+/pI^+O^+GFP$	−	+		
	+	+		
	+	−		
$I^+O^+Z^-/pI^+O^cGFP$	−	+		
	+	−		
$I^+O^+Z^+/pI^sO^cGFP$	−	+		
	+	−		
$I^+O^cZ^+/pI^sO^+GFP$	−	+		
	+	−		