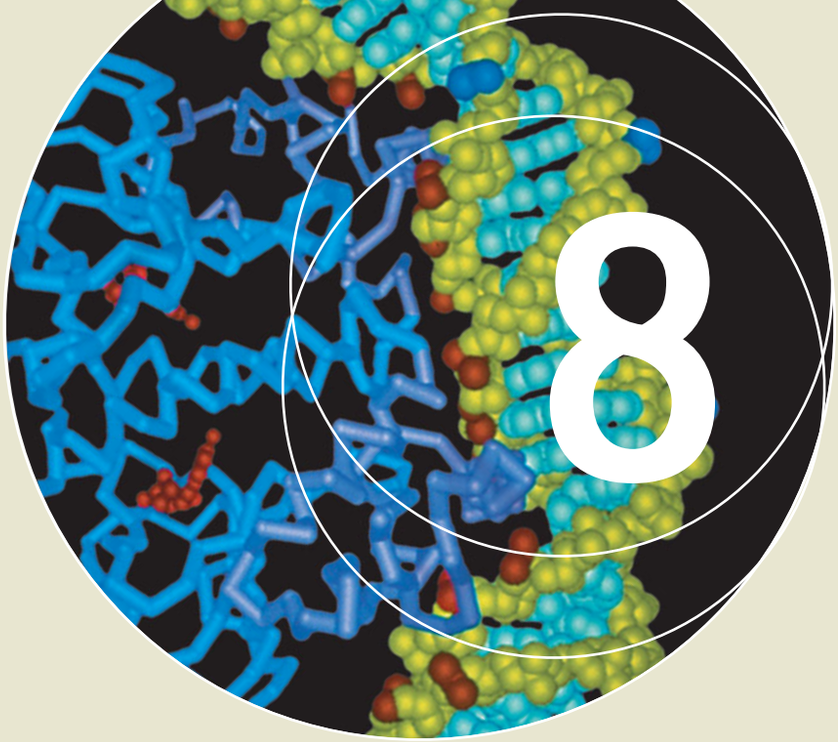

8

Regulation of Gene Expression

Regulation at the transcriptional level, a common mechanism for controlling gene expression in prokaryotes, is triggered by the attachment or release of DNA-binding proteins to specific genes on the DNA.

After the genetic information stored as DNA is transcribed into RNA, the information is translated to yield specific proteins. Collectively, these processes are called **gene expression**. Most proteins are enzymes that carry out the hundreds of different biochemical reactions needed for cell growth. To efficiently orchestrate the numerous reactions in a cell and to make maximal use of available resources, cells must *regulate* the kinds and amounts of proteins and other macromolecules they make. Such regulation is the focus of this chapter.

I Overview of Regulation

Some proteins and RNA molecules are needed in the cell at about the same level under all growth conditions. The expression of these molecules is said to be *constitutive*. However, more often a particular macromolecule is needed under some conditions but not others. For instance, enzymes required for using the sugar lactose are useful only if lactose is available. Microbial genomes encode many more proteins than are actually present in the cell under any particular condition. Thus, regulation is a major process in all cells and helps to conserve energy and resources.

Cells use two major approaches to regulating protein function. One controls the *activity* of an enzyme or other protein and the other controls the *amount* of an enzyme. The activity of a protein can be regulated only after it has been synthesized (that is, post-translationally). Regulating the activity of an enzyme in the cell is typically very rapid (taking seconds or less), whereas synthesizing an enzyme is relatively slow (taking several minutes). After synthesis of an enzyme begins, it takes some time before it is present in amounts sufficient to affect metabolism. Conversely, after synthesis of an enzyme stops, a considerable time may elapse before the enzyme is sufficiently diluted that it no longer affects metabolism. However, working together, regulation of enzyme activity and of enzyme synthesis efficiently controls cell metabolism.

8.1 Major Modes of Regulation

Most bacterial genes are transcribed into messenger RNA (mRNA), which in turn is translated into protein, as we discussed in Chapter 6. The components of a typical gene, with the corresponding mRNA and protein (the gene product), are summarized in **Figure 8.1**. The structural gene encodes the gene product and its expression is controlled by sequences in the upstream region. The amount of protein synthesized can be regulated at either the level of transcription, by varying the amount of mRNA made, or, less often, at the level of translation, by translating or not translating the mRNA. Occasionally the amount of protein may be regulated by degradation of the protein. Note that the sequences that determine the beginning and end of transcription are distinct from those that determine the beginning and end of translation. They are separated by small spacer regions, the 5′ and 3′ untranslated regions (5′-UTR and 3′-UTR).

Systems that control the level of expression of particular genes are varied, and genes are often regulated by more than one system. The processes that regulate the activity of enzymes have already been discussed (↻ Section 4.16). Here we consider how the synthesis of RNA and proteins is controlled.

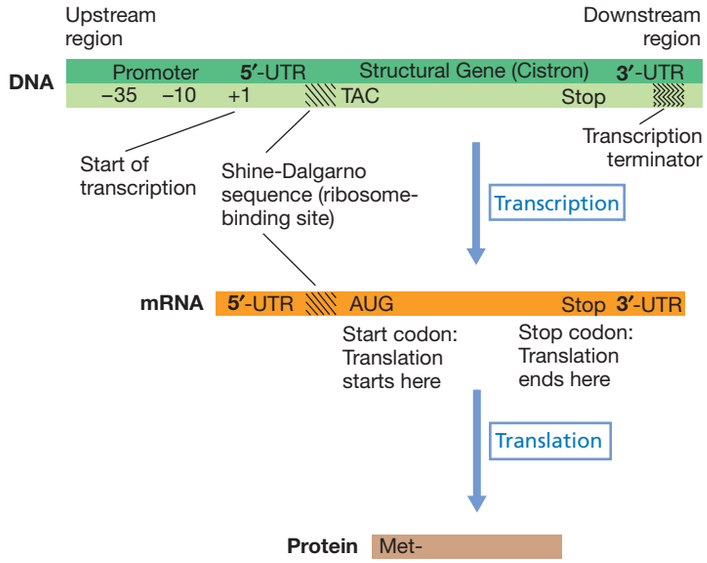

Figure 8.1 Components of a bacterial gene. The promoter, consisting of −35 and −10 regions, lies upstream of the gene. The 5′ untranslated region (5′-UTR) is a short region between the start of transcription and the start of translation. The 3′ untranslated region (3′-UTR) is a short region between the stop codon and the transcription terminator. The synthesis of the gene product (protein) may be regulated at the level of transcription or of translation or both.

MiniQuiz

- What steps in the synthesis of protein might be subject to regulation?
- Which is likely to be more rapid, the regulation of activity or the regulation of synthesis? Why?

II DNA-Binding Proteins and Regulation of Transcription

The amount of a protein present in a cell may be controlled at the level of transcription, at the level of translation, or, occasionally, by protein degradation. Our discussion begins with control at the level of transcription because this is the major means of regulation in prokaryotes.

The half-life of a typical mRNA in prokaryotes is short, only a few minutes at best. This allows prokaryotes to respond quickly to changing environmental parameters. Although there are energy costs in resynthesizing mRNAs that have been translated only a few times before being degraded, there are benefits to removing mRNAs rapidly when they are no longer needed, as this prevents the production of unneeded proteins. Thus, transcription and mRNA degradation coexist in the growing cell.

For a gene to be transcribed, RNA polymerase must recognize a specific promoter on the DNA and begin functioning (↻ Section 6.12). Regulation of transcription typically requires proteins that can bind to DNA. Thus, before discussing specific regulatory mechanisms, we must consider DNA-binding proteins.

8.2 DNA-Binding Proteins

Small molecules often take part in regulating transcription. However, they rarely do so directly. Instead, they typically influence the binding of certain proteins, called *regulatory proteins*, to specific sites on the DNA. It is these proteins that actually regulate transcription.

Interaction of Proteins with Nucleic Acids

Interactions between proteins and nucleic acids are central to replication, transcription, and translation, and also to the regulation of these processes. Protein–nucleic acid interactions may be nonspecific or specific, depending on whether the protein attaches anywhere along the nucleic acid or whether it recognizes a specific sequence. Histones (↻ Section 7.5) are good examples of nonspecific binding proteins. Histones are universally present in *Eukarya* and are also present in many *Archaea*. Because they are positively charged, histones combine strongly and relatively nonspecifically with negatively charged DNA. If the DNA is covered with histones, RNA polymerase cannot bind and the DNA cannot be transcribed. However, removal of histones does not automatically lead to transcription, but simply leaves the DNA accessible to other proteins that control gene expression.

Most DNA-binding proteins interact with DNA in a sequence-specific manner. Specificity is provided by interactions between specific amino acid side chains of the proteins and specific chemical groups on the nitrogenous bases and the sugar–phosphate backbone of the DNA. Because of its size, the *major groove* of DNA is the main site of protein binding. Figure 6.2 identified atoms of the bases in the major groove that are known to interact with proteins. To achieve high specificity, the binding protein must interact simultaneously with several nucleotides. In practice, this means that a specific binding protein binds only to DNA containing a specific base sequence.

We have already described a structure in DNA called an *inverted repeat* (↻ Figure 6.6). Such inverted repeats are frequently the locations at which regulatory proteins bind specifically to DNA (**Figure 8.2**). Note that this interaction does not involve the formation of stem–loop structures in the DNA. DNA-binding proteins are often homodimeric, meaning they are composed of two identical polypeptide subunits, each subdivided into **domains**; that is, regions of a protein with a specific structure and function. Each subunit has a domain that interacts specifically with a region of DNA in the major groove. When protein dimers interact with inverted repeats on DNA, each subunit binds to one of the inverted repeats. The dimer as a whole thus binds to both DNA strands (Figure 8.2). The DNA-binding protein recognizes base sequences by making a series of molecular contacts that are specific for that particular sequence.

Structure of DNA-Binding Proteins

DNA-binding proteins in both prokaryotes and eukaryotes possess several classes of protein domains that are critical for proper binding to DNA. One of the most common is the *helix-turn-helix* structure (**Figure 8.3**). This consists of two segments of polypeptide chain that have α-helix secondary structure connected by a short sequence forming the "turn." The first helix is the *recognition helix* that interacts specifically with DNA. The second helix, the *stabilizing helix*, stabilizes the first helix by interacting hydrophobically with it. The turn linking the two helices consists of three amino acid residues, the first of which is typically a glycine. Sequences are recognized by noncovalent interactions, including hydrogen bonds and van der Waals contacts, between the recognition helix of the protein and specific chemical groups in the sequence of base pairs on the DNA.

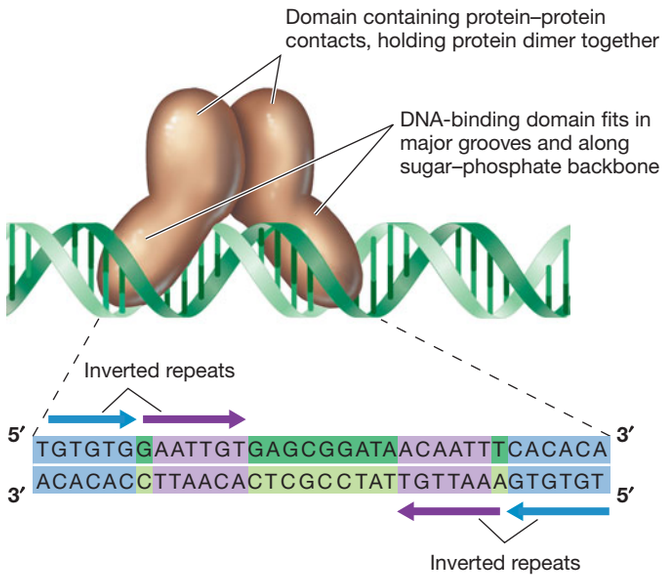

Figure 8.2 DNA-binding proteins. Many DNA-binding proteins are dimers that combine specifically with two sites on the DNA. The specific DNA sequences that interact with the protein are inverted repeats. The nucleotide sequence of the operator gene of the lactose operon is shown, and the inverted repeats, which are sites at which the *lac* repressor makes contact with the DNA, are shown in shaded boxes.

Many different DNA-binding proteins from *Bacteria* contain the helix-turn-helix structure. These include many repressor proteins, such as the *lac* and *trp* repressors of *Escherichia coli* (Section 8.3), and some proteins of bacterial viruses, such as the bacteriophage lambda repressor (Figure 8.3*b*). Indeed, over 250 different known proteins with this motif bind to DNA to regulate transcription in *E. coli*.

Two other types of protein domains are commonly found in DNA-binding proteins. One of these, the *zinc finger*, is frequently found in regulatory proteins in eukaryotes (**Figure 8.4*a***). The zinc finger is a protein structure that, as its name implies, binds a zinc ion. Part of the "finger" of amino acids that is created forms an α-helix, and this recognition helix interacts with DNA in the major groove. There are usually at least two or three zinc fingers on proteins that use them for DNA binding.

The other protein domain commonly found in DNA-binding proteins is the *leucine zipper* (Figure 8.4*b*). These are regions in which leucine residues are spaced every seven amino acids, somewhat resembling a zipper. Unlike the helix-turn-helix structure and the zinc finger, the leucine zipper does not interact with DNA itself but functions to hold two recognition helices in the correct orientation to bind DNA.

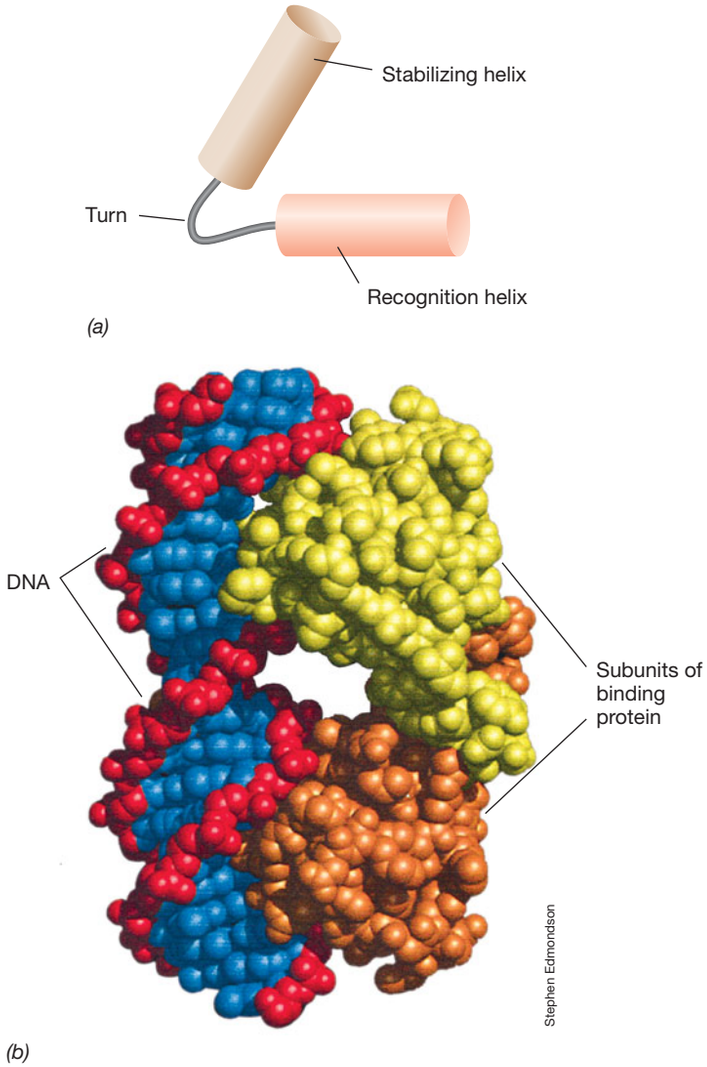

Figure 8.3 The helix-turn-helix structure of some DNA-binding proteins. *(a)* A simple model of the helix-turn-helix structure within a single protein subunit. *(b)* A computer model of both subunits of the bacteriophage lambda repressor, a typical helix-turn-helix protein, bound to its operator. The DNA is red and blue. One subunit of the dimeric repressor is shown in brown and the other in yellow. Each subunit contains a helix-turn-helix structure. The coordinates used to generate this image were downloaded from the Protein Data Base, Brookhaven, NY (http://www.pdb.org/pdb/home/).

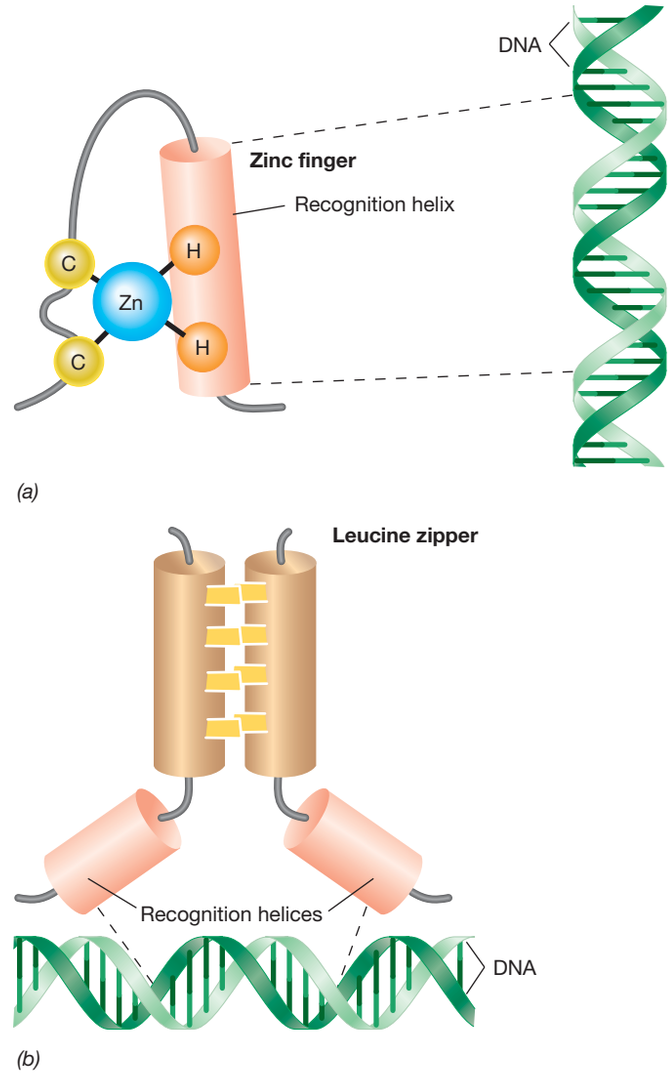

Figure 8.4 Simple models of protein substructures found in eukaryotic DNA-binding proteins. Cylinders represent α-helices. Recognition helices are the domains that bind DNA. *(a)* The zinc finger structure. The amino acids holding the Zn^{2+} ion always include at least two cysteine residues (C), with the other residues being histidine (H). *(b)* The leucine zipper structure. The leucine residues (yellow) are spaced exactly every seven amino acids. The interaction of the leucine side chains helps hold the two helices together.

Once a protein binds at a specific site on the DNA, various outcomes are possible. Some DNA-binding proteins are enzymes that catalyze a specific reaction on the DNA, such as transcription by RNA polymerase. In other cases, however, the binding event can either block transcription (negative regulation, Section 8.3) or activate it (positive regulation, Section 8.4).

MiniQuiz

- What is a protein domain?
- Why are some interactions of proteins with DNA specific to certain DNA sequences?

8.3 Negative Control of Transcription: Repression and Induction

Transcription is the first step in biological information flow; because of this, it is simple and efficient to control gene expression at this point. If one gene is transcribed more frequently than another, there will be more of its mRNA available for translation and therefore a greater amount of its protein product in the cell. Several different mechanisms for controlling gene expression are known in bacteria, and all of them are greatly influenced by the environment in which the organism is growing, in particular by the presence or absence of specific small

molecules. These molecules can interact with specific proteins such as the DNA-binding proteins just described. The result is the control of transcription or, more rarely, translation.

We begin by describing repression and induction, simple forms of regulation that govern gene expression at the level of transcription. In this section we deal with **negative control** of transcription, control that prevents transcription.

Enzyme Repression and Induction

Often the enzymes that catalyze the synthesis of a specific product are not made if the product is already present in the medium in sufficient amounts. For example, the enzymes needed to synthesize the amino acid arginine are made only when arginine is absent from the culture medium; an excess of arginine decreases the synthesis of these enzymes. This is called enzyme **repression**.

As can be seen in **Figure 8.5**, if arginine is added to a culture growing exponentially in a medium devoid of arginine, growth continues at the previous rate, but production of the enzymes for arginine synthesis stops. Note that this is a specific effect, as the synthesis of all other enzymes in the cell continues at the previous rate. This is because the enzymes affected by a particular repression event make up only a tiny fraction of the entire complement of proteins in the cell. Enzyme repression is widespread in bacteria as a means of controlling the synthesis of enzymes required for the production of amino acids and the nucleotide precursors purines and pyrimidines. In most cases, the final product of a particular biosynthetic pathway represses the enzymes of the pathway. This ensures that the organism does not waste energy and nutrients synthesizing unneeded enzymes.

Enzyme **induction** is conceptually the opposite of enzyme repression. In enzyme induction, an enzyme is made only when its substrate is present. Enzyme repression typically affects biosynthetic (anabolic) enzymes. In contrast, enzyme induction usually affects degradative (catabolic) enzymes.

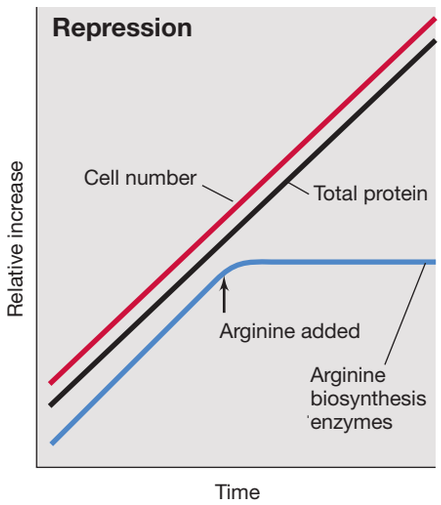

Figure 8.5 **Enzyme repression.** The addition of arginine to the medium specifically represses production of enzymes needed to make arginine. Net protein synthesis is unaffected.

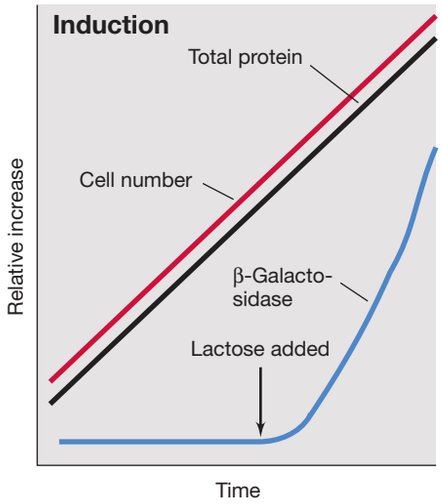

Figure 8.6 **Enzyme induction.** The addition of lactose to the medium specifically induces synthesis of the enzyme β-galactosidase. Net protein synthesis is unaffected.

Consider, for example, the utilization of the sugar lactose as a carbon and energy source by *Escherichia coli*. **Figure 8.6** shows the induction of β-galactosidase, the enzyme that cleaves lactose into glucose and galactose. This enzyme is required for *E. coli* to grow on lactose. If lactose is absent, the enzyme is not made, but synthesis begins almost immediately after lactose is added. The three genes in the *lac* operon encode three proteins, including β-galactosidase, that are induced simultaneously upon adding lactose. This type of control mechanism ensures that specific enzymes are synthesized only when needed.

Inducers and Corepressors

The substance that induces enzyme synthesis is called an *inducer* and a substance that represses enzyme synthesis is called a *corepressor*. These substances, which are normally small molecules, are collectively called *effectors*. Interestingly, not all inducers and corepressors are actual substrates or end products of the enzymes involved. For example, structural analogs may induce or repress even though they are not substrates of the enzyme. Isopropylthiogalactoside (IPTG), for instance, is an inducer of β-galactosidase even though IPTG cannot be hydrolyzed by this enzyme. In nature, however, inducers and corepressors are probably normal cell metabolites. Detailed studies of lactose utilization in *E. coli* have shown that the actual inducer of β-galactosidase is not lactose, but its isomer allolactose, which is made from lactose. www.microbiologyplace.com **Online Tutorial 8.1: Negative Control of Transcription and the *lac* Operon**

Mechanism of Repression and Induction

How can inducers and corepressors affect transcription in such a specific manner? They do this indirectly by binding to specific DNA-binding proteins, which, in turn, affect transcription. For an example of a repressible enzyme, we consider the arginine operon. **Figure 8.7*a*** shows transcription of the arginine genes, which proceeds when the cell needs arginine. When arginine is

UNIT 3

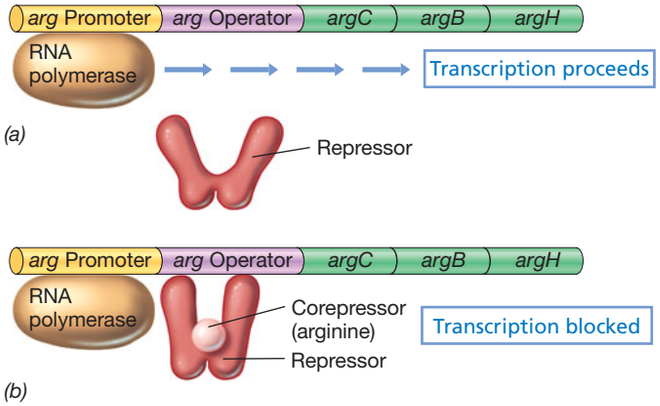

Figure 8.7 **Enzyme repression in the arginine operon.** *(a)* The operon is transcribed because the repressor is unable to bind to the operator. *(b)* After a corepressor (small molecule) binds to the repressor, the repressor binds to the operator and blocks transcription; mRNA and the proteins it encodes are not made. For the *argCBH* operon, the amino acid arginine is the corepressor that binds to the arginine repressor.

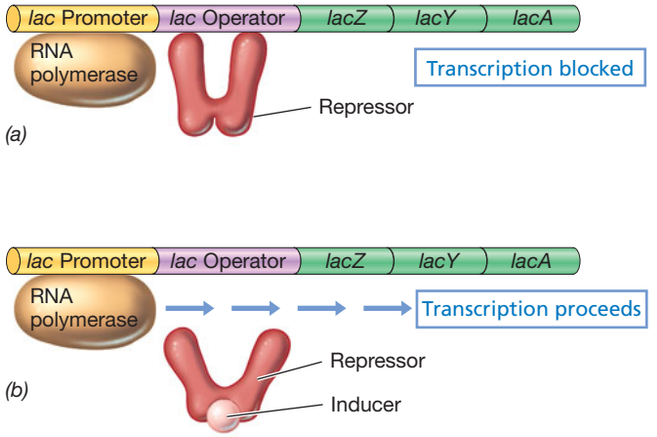

Figure 8.8 **Enzyme induction in the lactose operon.** *(a)* A repressor protein bound to the operator blocks the binding of RNA polymerase. *(b)* An inducer molecule binds to the repressor and inactivates it so that it no longer can bind to the operator. RNA polymerase then transcribes the DNA and makes an mRNA for that operon. For the *lac* operon, the sugar allolactose is the inducer that binds to the lactose repressor.

plentiful it acts as corepressor. As Figure 8.7*b* shows, arginine binds to a specific **repressor protein**, the arginine repressor, present in the cell. The repressor protein is allosteric (↻ Section 4.16); that is, its conformation is altered when the corepressor binds to it.

By binding its effector, the repressor protein becomes active and can then bind to a specific region of the DNA near the promoter of the gene, known as the *operator*. This region gave its name to the **operon**, a cluster of genes arranged in a linear and consecutive fashion whose expression is under the control of a single operator (↻ Section 6.5). All of the genes in an operon are transcribed as a single unit yielding a single mRNA. The operator is located downstream of the promoter where synthesis of mRNA is initiated (Figure 8.7). If the repressor binds to the operator, transcription is physically blocked because RNA polymerase can neither bind nor proceed. Hence, the polypeptides encoded by the genes in the operon cannot be synthesized. If the mRNA is polycistronic (↻ Section 6.15), all the polypeptides encoded by this mRNA will be repressed.

Enzyme induction may also be controlled by a repressor. In this case, the repressor protein is active in the absence of the inducer, completely blocking transcription. When the inducer is added, it combines with the repressor protein and inactivates it; inhibition is overcome and transcription can proceed (**Figure 8.8**).

All regulatory systems employing repressors have the same underlying mechanism: inhibition of mRNA synthesis by the activity of specific repressor proteins that are themselves under the control of specific small effector molecules. And, as previously noted, because the repressor's role is inhibitory, regulation by repressors is called *negative control*. One point to note is that genes are not turned on and off completely like light switches. DNA-binding proteins vary in concentration and affinity and thus control is quantitative. Even when a gene is "fully repressed" there is often a very low level of basal transcription.

MiniQuiz

- Why is "negative control" so named?
- How does a repressor inhibit the synthesis of a specific mRNA?

8.4 Positive Control of Transcription

Negative control relies on a repressor protein to bring about repression of mRNA synthesis. By contrast, in **positive control** of transcription the regulatory protein is an **activator** that activates the binding of RNA polymerase to DNA. An excellent example of positive regulation is the catabolism of the sugar maltose in *Escherichia coli*.

Maltose Catabolism in *Escherichia coli*

The enzymes for maltose catabolism in *E. coli* are synthesized only after the addition of maltose to the medium. The expression of these enzymes thus follows the pattern shown for β-galactosidase in Figure 8.6 except that maltose rather than lactose is required to induce gene expression. However, the synthesis of maltose-degrading enzymes is not under negative control as in the *lac* operon, but under positive control; transcription requires the binding of an **activator protein** to the DNA.

The maltose activator protein cannot bind to the DNA unless it first binds maltose, the inducer. When the maltose activator protein binds to DNA, it allows RNA polymerase to begin transcription (**Figure 8.9**). Like repressor proteins, activator proteins bind specifically only to certain sequences on the DNA. However, the region on the DNA that is the binding site of the activator is not called an operator (Figures 8.7 and 8.8), but instead an *activator-binding site* (Figure 8.9). Nevertheless, the genes controlled by this activator-binding site are still called an operon.

Binding of Activator Proteins

The promoters of positively controlled operons have nucleotide sequences that bind RNA polymerase weakly and are poor

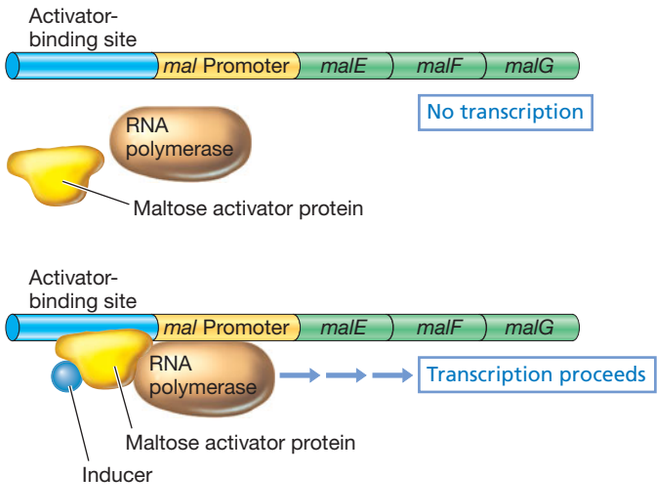

Figure 8.9 **Positive control of enzyme induction in the maltose operon.** *(a)* In the absence of an inducer, neither the activator protein nor the RNA polymerase can bind to the DNA. *(b)* An inducer molecule (for the *malEFG* operon it is the sugar maltose) binds to the activator protein (MalT), which in turn binds to the activator-binding site. This allows RNA polymerase to bind to the promoter and begin transcription.

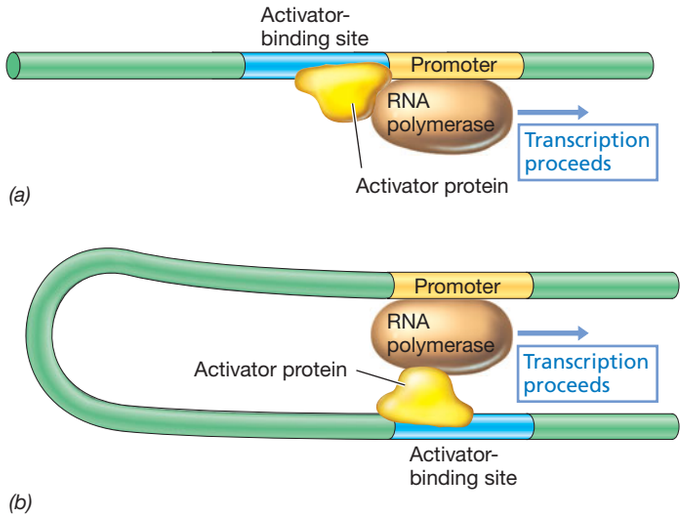

Figure 8.11 **Activator protein interactions with RNA polymerase.** *(a)* The activator-binding site is near the promoter. *(b)* The activator-binding site is several hundred base pairs from the promoter. In this case, the DNA must be looped to allow the activator and the RNA polymerase to contact.

matches to the consensus sequence (↻ Section 6.13). Thus, even with the correct sigma (σ) factor, the RNA polymerase has difficulty binding to these promoters. The role of the activator protein is to help the RNA polymerase recognize the promoter and begin transcription.

For example, the activator protein may modify the structure of the DNA by bending it (**Figure 8.10**), allowing the RNA polymerase to make the correct contacts with the promoter to begin transcription. Alternatively, the activator protein may interact directly with the RNA polymerase. This can happen either when the activator-binding site is close to the promoter (**Figure 8.11*a***) or when it is several hundred base pairs away from the promoter, a situation in which DNA looping is required to make the necessary contacts (Figure 8.11*b*).

Many genes in *E. coli* have promoters under positive control and many have promoters under negative control. In addition, many operons have promoters with multiple types of control and some have more than one promoter, each with its own control system! Thus, the simple picture outlined above is not typical of all operons. Multiple control features are common in the operons of virtually all prokaryotes, and thus their overall regulation can be very complex.

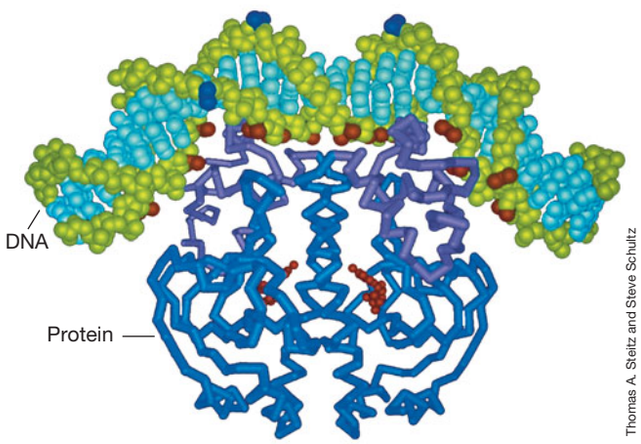

Figure 8.10 **Computer model of a positive regulatory protein interacting with DNA.** This model shows the cyclic AMP receptor protein (CRP), a regulatory protein that controls several operons. The α-carbon backbone of this protein is shown in blue and purple. The protein is binding to a DNA double helix (green and light blue). Note that binding of the CRP protein to DNA has bent the DNA.

Operons versus Regulons

In *E. coli*, the genes required for maltose utilization are spread out over the chromosome in several operons, each of which has an activator-binding site to which a copy of the maltose activator protein can bind. Therefore, the maltose activator protein actually controls the transcription of more than one operon. When more than one operon is under the control of a single regulatory protein, these operons are collectively called a **regulon**. Therefore, the enzymes for maltose utilization are encoded by the maltose regulon.

Regulons are known for operons under negative control as well. For example, the arginine biosynthetic enzymes (Section 8.3) are encoded by the arginine regulon, whose operons are all under the control of the arginine repressor protein (only one of the arginine operons was shown in Figure 8.7). In regulon control a specific DNA-binding protein binds only at those operons it controls regardless of whether it is functioning as an activator or repressor; other operons are not affected.

MiniQuiz

- Compare and contrast the activities of an activator protein and a repressor protein.
- Distinguish between an operon and a regulon.

UNIT 3

8.5 Global Control and the *lac* Operon

An organism often needs to regulate many unrelated genes simultaneously in response to a change in its environment. Regulatory mechanisms that respond to environmental signals by regulating the expression of many different genes are called *global control systems*. Both the lactose operon and the maltose regulon respond to global controls in addition to their own controls discussed in Sections 8.3 and 8.4. We begin our consideration of global regulation with the *lac* operon and the choice between different sugars.

Catabolite Repression

We have not yet considered the possibility that bacteria might be confronted with several different carbon sources that could be used. For example, *Escherichia coli* can use many different sugars. When faced with several sugars, including glucose, do cells of *E. coli* use them simultaneously or one at a time? The answer is that glucose is used first. It would be wasteful to induce enzymes for using other sugars when glucose is available, because *E. coli* grows faster on glucose than on other carbon sources. **Catabolite repression** is a mechanism of global control that decides between different available carbon sources if more than one is present.

When cells of *E. coli* are grown in a medium that contains glucose, the synthesis of enzymes needed for the breakdown of other carbon sources (such as lactose or maltose) is repressed, even if those other carbon sources are present. Thus, the presence of a favored carbon source overrules the induction of pathways that catabolize other carbon sources. Catabolite repression is sometimes called the "glucose effect" because glucose was the first substance shown to cause this response. The key point is that the favored substrate is a better carbon and energy source for the organism. Thus, catabolite repression ensures that the organism uses the *best* available carbon and energy source first.

Why is catabolite repression called *global* control? In *E. coli* and other organisms for which glucose is the best energy source, catabolite repression prevents expression of most other catabolic operons as long as glucose is present. Dozens of catabolic operons are affected, including those for lactose, maltose, a host of other sugars, and most other commonly used carbon and energy sources for *E. coli*. In addition, genes for the synthesis of flagella are controlled by catabolite repression because if bacteria have a good carbon source available, there is no need to swim around in search of nutrients.

One consequence of catabolite repression is that it may lead to two exponential growth phases, a situation called *diauxic growth*. If two usable energy sources are available, the cells grow first on the better energy source. Growth stops when the better source is depleted, but then following a lag period, it resumes on the other energy source. Diauxic growth is illustrated in **Figure 8.12** for *E. coli* on a mixture of glucose and lactose. The cells grow more rapidly on glucose than on lactose. Although glucose and lactose are both excellent energy sources for *E. coli*, glucose is superior, and growth is faster.

The proteins of the *lac* operon, including the enzyme β-galactosidase, are required for using lactose and are induced in its presence (Figures 8.6 and 8.8). But the synthesis of these proteins is also subject to catabolite repression. As long as glucose is present, the *lac* operon is not expressed and lactose is not used. However, when glucose is depleted, catabolite repression is abolished, the *lac* operon is expressed, and the cells grow on lactose.

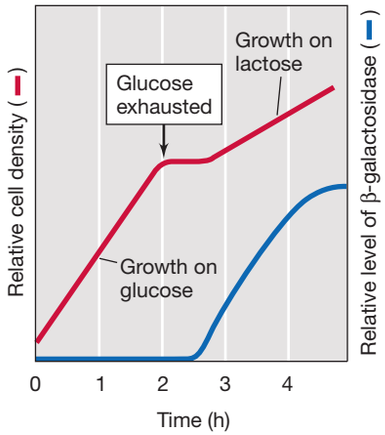

Figure 8.12 Diauxic growth of *Escherichia coli* on a mixture of glucose and lactose. The presence of glucose represses the synthesis of β-galactosidase, the enzyme that cleaves lactose into glucose and galactose. After glucose is depleted, there is a lag during which β-galactosidase is synthesized. Growth then resumes on lactose but at a slower rate.

Cyclic AMP and Cyclic AMP Receptor Protein

Despite its name, catabolite repression relies on an activator protein and is actually a form of positive control (Section 8.4). The activator protein is called the *cyclic AMP receptor protein* (*CRP*). A gene that encodes a catabolite-repressible enzyme is expressed only if CRP binds to DNA in the promoter region. This allows RNA polymerase to bind to the promoter. CRP is an allosteric protein and binds to DNA only if it has first bound a small molecule called *cyclic adenosine monophosphate* (*cyclic AMP* or *cAMP*) (**Figure 8.13**). Like many DNA-binding proteins (Section 8.2), CRP binds to DNA as a dimer.

Cyclic AMP is a key molecule in many metabolic control systems, both in prokaryotes and eukaryotes. Because it is derived from a nucleic acid precursor, it is a **regulatory nucleotide**. Other regulatory nucleotides include cyclic guanosine monophosphate (cyclic GMP; important mostly in eukaryotes), cyclic di-GMP (important in biofilm formation; Section 23.4), and guanosine tetraphosphate (ppGpp; Section 8.10). Cyclic AMP is synthesized from ATP by an enzyme called *adenylate cyclase*. However, glucose inhibits the synthesis of cyclic AMP and also stimulates cyclic AMP transport out of the cell. When glucose enters the cell, the cyclic AMP level is lowered, CRP protein cannot bind DNA, and RNA polymerase fails to bind to the promoters of operons subject to catabolite repression. Thus, catabolite repression is an indirect result of the presence of a better energy source (glucose). The direct cause of catabolite repression is a low level of cyclic AMP.

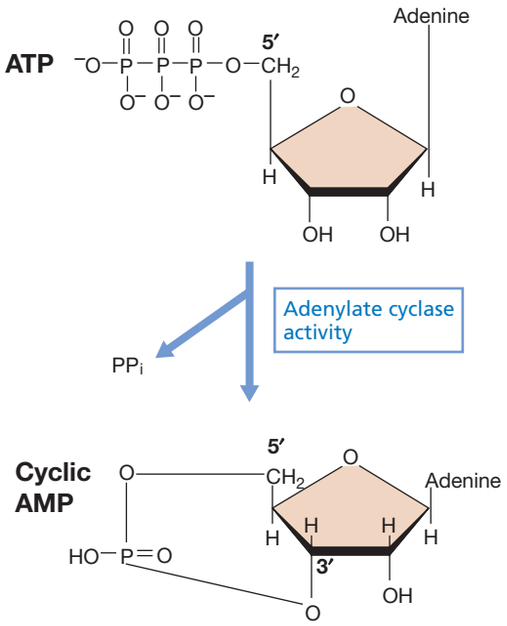

Figure 8.13 **Cyclic AMP.** Cyclic adenosine monophosphate (cyclic AMP) is made from ATP by the enzyme adenylate cyclase.

Let us return to the *lac* operon and include catabolite repression. The entire regulatory region of the *lac* operon is diagrammed in **Figure 8.14**. For *lac* genes to be transcribed, two requirements must be met: (1) The level of cyclic AMP must be high enough for the CRP protein to bind to the CRP-binding site (positive control), and (2) lactose or another molecule capable of acting as inducer must be present so that the lactose repressor (LacI protein) does not block transcription by binding to the operator (negative control). If these two conditions are met, the cell is signaled that glucose is absent and lactose is present; then and only then does transcription of the *lac* operon begin.

MiniQuiz

- Explain how catabolite repression depends on an activator protein.
- What role does cyclic AMP play in glucose regulation?
- Explain how the *lac* operon is both positively and negatively controlled.

8.6 Control of Transcription in *Archaea*

There are two alternative approaches to regulating the activity of RNA polymerase. One strategy, common in *Bacteria*, is to use DNA-binding proteins that either block RNA polymerase activity (repressor proteins) or stimulate RNA polymerase activity (activator proteins). The alternative, common in eukaryotes, is to transmit signals to the protein subunits of the RNA polymerase itself. Given the greater overall similarity between the mechanism of transcription in *Archaea* and *Eukarya* (Chapter 7), it is perhaps surprising that the regulation

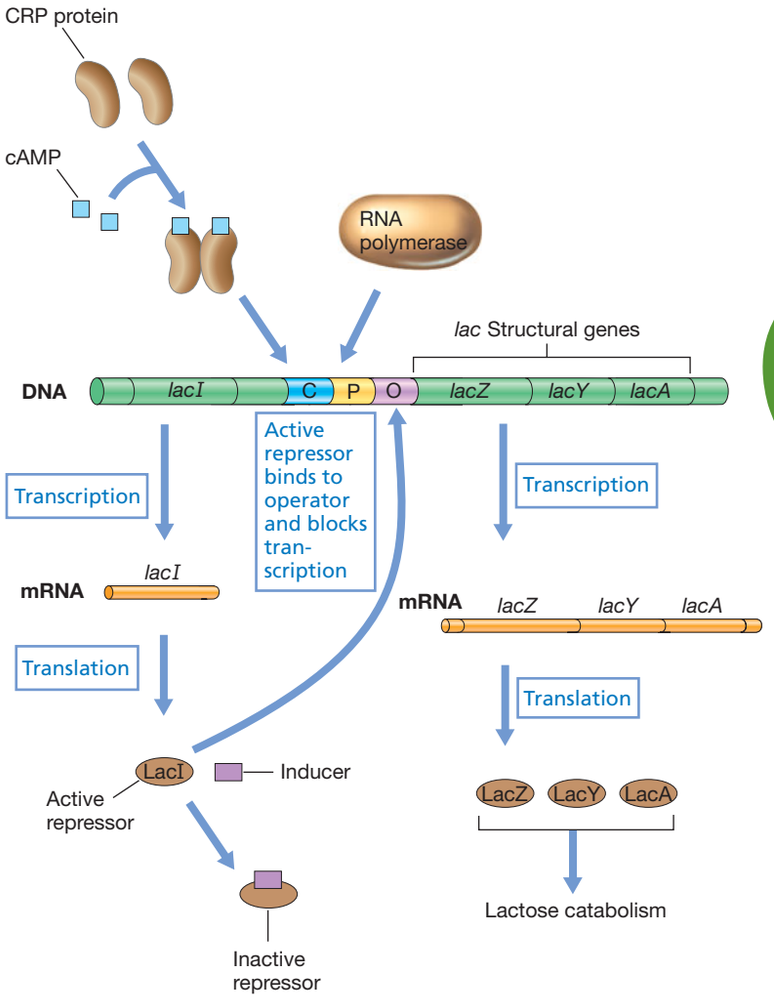

Figure 8.14 **Overall regulation of the *lac* system.** The *lac* operon consists of *lacZ*, encoding β-galactosidase, which breaks down lactose, plus two other genes, *lacY*, encoding lactose permease, and *lacA*, encoding lactose acetylase. The LacI repressor protein is encoded by a separate gene, *lacI*. LacI binds to the operator (O) unless the inducer is present. RNA polymerase binds to the promoter (P). CRP binds to the C site when activated by cyclic AMP. For the *lac* operon to be transcribed by RNA polymerase, the LacI repressor must be absent (that is, inducer must be present) and cyclic AMP levels must be high (due to the absence of glucose), thus allowing CRP to bind.

of transcription in *Archaea* more closely resembles that of *Bacteria*.

Few repressor or activator proteins from *Archaea* have yet been characterized in detail, but it is clear that *Archaea* have both types of regulatory proteins. Archaeal repressor proteins either block the binding of RNA polymerase itself or block the binding of TBP (TATA-binding protein) and TFB (transcription factor B), which are required for RNA polymerase to bind to the promoter in *Archaea* (↻ Section 7.2). At least some archaeal activator proteins function in just the opposite way, by recruiting TBP to the promoter, thereby facilitating transcription.

A good example of an archaeal repressor is the NrpR protein from the methanogen *Methanococcus maripaludis;* this protein

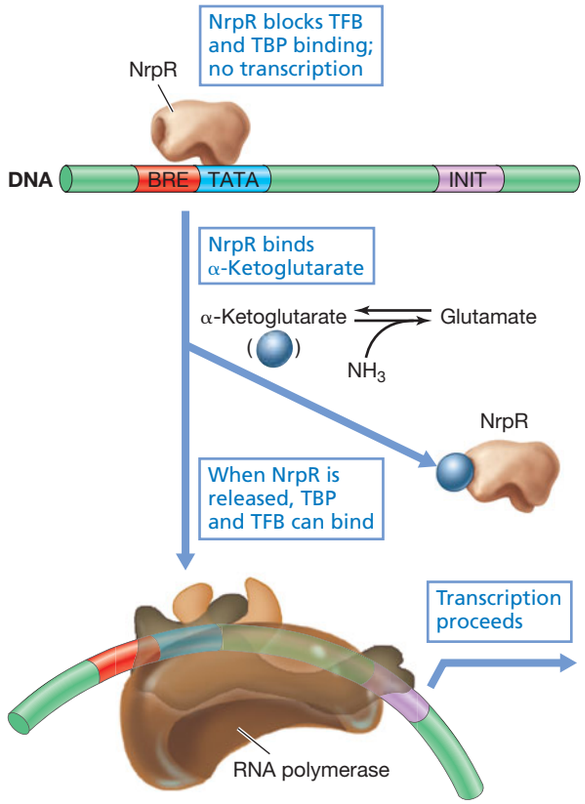

Figure 8.15 Repression of genes for nitrogen metabolism in *Archaea*. The NrpR protein of *Methanococcus maripaludis* acts as a repressor. It blocks the binding of the TFB and TBP proteins, which are required for promoter recognition, to the BRE site and TATA box, respectively. If there is a shortage of ammonia, α-ketoglutarate is not converted to glutamate. The α-ketoglutarate accumulates and binds to NrpR, releasing it from the DNA. Now TBP and TFB can bind. This in turn allows RNA polymerase to bind and transcribe the operon.

represses genes active in nitrogen assimilation (**Figure 8.15**), such as those for nitrogen fixation (↻ Section 13.15) and glutamine synthesis (↻ Section 4.16). When organic nitrogen is plentiful in the *M. maripaludis* cell, NrpR represses nitrogen assimilation genes. However, if the level of nitrogen becomes limiting, α-ketoglutarate accumulates to high levels. This occurs because α-ketoglutarate, a citric acid cycle intermediate, is also a major acceptor of ammonia during nitrogen assimilation.

When levels of α-ketoglutarate rise, this signals that ammonia is limiting and that additional pathways need to be activated for obtaining ammonia, such as nitrogen fixation or the high-affinity nitrogen assimilation enzyme glutamine synthetase. Elevated levels of α-ketoglutarate function as an inducer by binding to the NrpR protein. In this state, NrpR loses its affinity for the promoter regions of its target genes and no longer blocks transcription from these promoters. In this respect, the NrpR protein resembles the LacI repressor and similar proteins of *Bacteria* (Section 8.3).

Other archaeal proteins regulate transcription in a positive manner. Thus their binding in the promoter region increases transcription. Some of these transcription activators are related to bacterial proteins whereas others appear to be unique to the *Archaea*. The SurR protein of *Pyrococcus furiosus* is an example of a regulatory protein that functions either as an activator or as a repressor, depending on the location of its binding site within the promoter region. SurR controls the response of *Pyrococcus furiosus* to elemental sulfur and its conversion to hydrogen sulfide.

MiniQuiz

- What is the major difference between transcriptional regulation in *Archaea* and eukaryotes?
- How do transcriptional activators in *Archaea* often differ in mechanism from those in *Bacteria*?

III Sensing and Signal Transduction

Prokaryotes regulate cell metabolism in response to environmental fluctuations, including temperature changes, changes in pH and oxygen availability, changes in the availability of nutrients, and even changes in the number of other cells present. Therefore, there must be mechanisms by which cells receive signals from the environment and transmit them to the specific target to be regulated.

Some signals are small molecules that enter the cell and function as effectors. However, in many cases the external signal is not transmitted directly to the regulatory protein but instead is detected by a sensor that transmits it to the rest of the regulatory machinery, a process called **signal transduction**.

8.7 Two-Component Regulatory Systems

Because most signal transduction systems contain two parts, they are called **two-component regulatory systems**. Characteristically, such systems consist of a specific **sensor kinase protein** usually located in the cytoplasmic membrane and a **response regulator protein** present in the cytoplasm.

A kinase is an enzyme that phosphorylates compounds, typically using phosphate from ATP. Sensor kinases detect a signal from the environment and phosphorylate themselves (a process called autophosphorylation) at a specific histidine residue (**Figure 8.16**). Sensor kinases thus belong to the class of enzymes called *histidine kinases*. The phosphate is then transferred to another protein inside the cell, the response regulator. This is typically a DNA-binding protein that regulates transcription, in either a positive or a negative fashion depending on the system. In the example shown in Figure 8.16, regulation is negative; the response regulator is a repressor that binds DNA, blocking transcription, until the transfer of the phosphate releases it, permitting transcription.

Although the term is rarely used, a *one-component regulatory system* consists of a single protein that both detects a signal and carries out a regulatory response. Examples include the LacI repressor, the MalT activator, and the Crp protein. All three bind a small molecule (the signal) and then bind to DNA to regulate transcription.

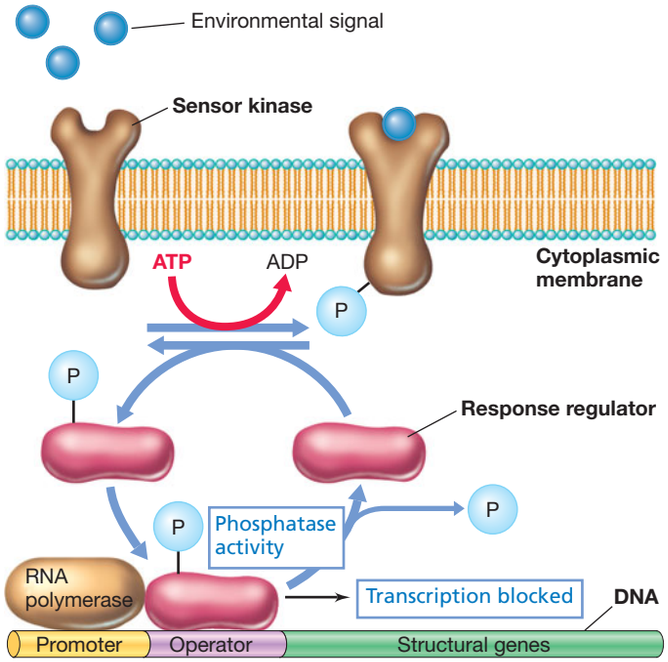

Figure 8.16 The control of gene expression by a two-component regulatory system. One component is a sensor kinase in the cytoplasmic membrane that phosphorylates itself in response to an environmental signal. The phosphoryl group is then transferred to the second component, a response regulator. The phosphorylated form of the response regulator then binds to DNA. In the system shown here, the phosphorylated response regulator is a repressor protein. The phosphatase activity of the response regulator slowly releases the phosphate from the response regulator and resets the system.

A balanced regulatory system must have a feedback loop, that is, a way to complete the regulatory circuit and terminate the response. This resets the system for another cycle. This feedback loop involves a phosphatase, an enzyme that removes the phosphate from the response regulator at a constant rate. This reaction is often carried out by the response regulator itself, although in some cases separate proteins are involved (Figure 8.16). Phosphatase activity is typically slower than phosphorylation. However, if phosphorylation ceases due to reduced sensor kinase activity, phosphatase activity eventually returns the response regulator to the fully nonphosphorylated state.

Examples of Two-Component Regulatory Systems

Two-component systems regulate a large number of genes in many different bacteria. Interestingly, two-component systems are rare or absent in *Archaea* and in *Bacteria* that live as parasites of higher organisms. A few key examples of two-component systems include those that respond to phosphate limitation, nitrogen limitation, and osmotic pressure. In *Escherichia coli* almost 50 different two-component systems are present, and several are listed in **Table 8.1**. For example, the osmolarity of the environment controls the relative levels of the proteins OmpC and OmpF in the *E. coli* outer membrane. OmpC and OmpF are porins, proteins that allow metabolites to cross the outer membrane of gram-negative bacteria (↻ Section 3.7). If osmotic pressure is low, the synthesis of OmpF, a porin with a larger pore, increases; if osmotic pressure is higher, OmpC, a porin with a smaller pore, is made in larger amounts. The response regulator of this system is OmpR. When OmpR is phosphorylated, it activates transcription of the *ompC* gene and represses transcription of the *ompF* gene. The *ompF* gene in *E. coli* is also controlled by antisense RNA (Section 8.14).

Some signal transduction systems have multiple regulatory elements. For instance, in the Ntr regulatory system, which regulates nitrogen assimilation in many *Bacteria*, including *E. coli*, the response regulator is the activator protein nitrogen regulator I (NRI). NRI activates transcription from promoters recognized by RNA polymerase using the alternative sigma factor σ^{54} (RpoN) (↻ Section 6.13). The sensor kinase in the Ntr system, nitrogen regulator II (NRII), fills a dual role as both kinase and phosphatase. The activity of NRII is in turn regulated by the addition or removal of uridine monophosphate groups from another protein, known as PII.

The Nar regulatory system (Table 8.1) controls a set of genes that allow the use of nitrate or nitrite or both as alternative electron acceptors during anaerobic respiration (↻ Section 14.7). The Nar system contains two different sensor kinases and two different response regulators. In addition, all of the genes regulated by this system are also controlled by the FNR protein (fumarate nitrite regulator), a global regulator for genes of anaerobic respiration (see Table 8.3). This type of multiple regulation is common for systems of central importance to cellular metabolism.

Genomic analyses allow easy detection of genes encoding two-component regulatory systems because the histidine kinases show significant amino acid sequence conservation. Two-component

Table 8.1 *Examples of two-component systems that regulate transcription in* Escherichia coli

System	*Environmental signal*	*Sensor kinase*	*Response regulator*	*Activity of response regulator*[a]
Arc system	Oxygen	ArcB	ArcA	Repressor/activator
Nitrate and nitrite respiration (Nar)	Nitrate and nitrite	NarX	NarL	Activator/repressor
		NarQ	NarP	Activator/repressor
Nitrogen utilization (Ntr)	Shortage of organic nitrogen	NRII (= GlnL)	NRI (= GlnG)	Activator of promoters requiring RpoN/σ^{54}
Pho regulon	Inorganic phosphate	PhoR	PhoB	Activator
Porin regulation	Osmotic pressure	EnvZ	OmpR	Activator/repressor

[a]Note that many response regulator proteins act as both activators and repressors depending on the genes being regulated. Although ArcA can function as either an activator or a repressor, it functions as a repressor on most operons that it regulates.

systems closely related to those in *Bacteria* are also present in microbial eukaryotes, such as the yeast *Saccharomyces cerevisiae*, and even in plants. However, most eukaryotic signal transduction pathways rely on phosphorylation of serine, threonine, and tyrosine residues of proteins that are unrelated to those of bacterial two-component systems.

MiniQuiz

- What are kinases and what is their role in two-component regulatory systems?
- What are phosphatases and what is their role in two-component regulatory systems?

8.8 Regulation of Chemotaxis

We have previously seen that some prokaryotes can move toward attractants or away from repellents, a behavior called *chemotaxis* (Section 3.15). We noted that prokaryotes are too small to sense spatial gradients of a chemical, but they can respond to temporal gradients. That is, they can sense the *change* in concentration of a chemical over time rather than the absolute concentration of the chemical stimulus. Prokaryotes use a modified two-component system to sense temporal changes in attractants or repellents and process this information to regulate flagellar rotation. Note that chemotaxis uses a two-component system to directly regulate the activity of preexisting flagella rather than the transcription of the genes encoding the flagella.

Step One: Response to Signal

The mechanism of chemotaxis is complex and depends upon multiple proteins. Several sensory proteins reside in the cytoplasmic membrane and sense the presence of attractants and repellents. These sensor proteins are not themselves sensor kinases but interact with cytoplasmic sensor kinases. These sensory proteins allow the cell to monitor the concentration of various substances over time.

The sensory proteins are called *methyl-accepting chemotaxis proteins* (*MCPs*). *Escherichia coli* possesses five different MCPs. Each MCP is a transmembrane protein that can sense certain compounds. For example, the Tar MCP of *E. coli* senses the attractants aspartate and maltose and the repellents cobalt and nickel. MCPs bind attractants or repellents directly or in some cases indirectly through interactions with periplasmic binding proteins. Binding of an attractant or repellent triggers interactions with cytoplasmic proteins that affect flagellar rotation.

MCPs make contact with the cytoplasmic proteins CheA and CheW (**Figure 8.17**). CheA is the sensor kinase for chemotaxis. When an MCP binds a chemical, it changes conformation and, with help from CheW, affects the autophosphorylation of CheA to form CheA-P. Attractants *decrease* the rate of autophosphorylation, whereas repellents *increase* this rate. CheA-P then passes the phosphate to CheY (forming CheY-P); this is the response regulator that controls flagellar rotation. CheA-P can also pass the phosphate to CheB (a response regulator active in step three). This phosphorylation is much slower than that of CheY, and is discussed later.

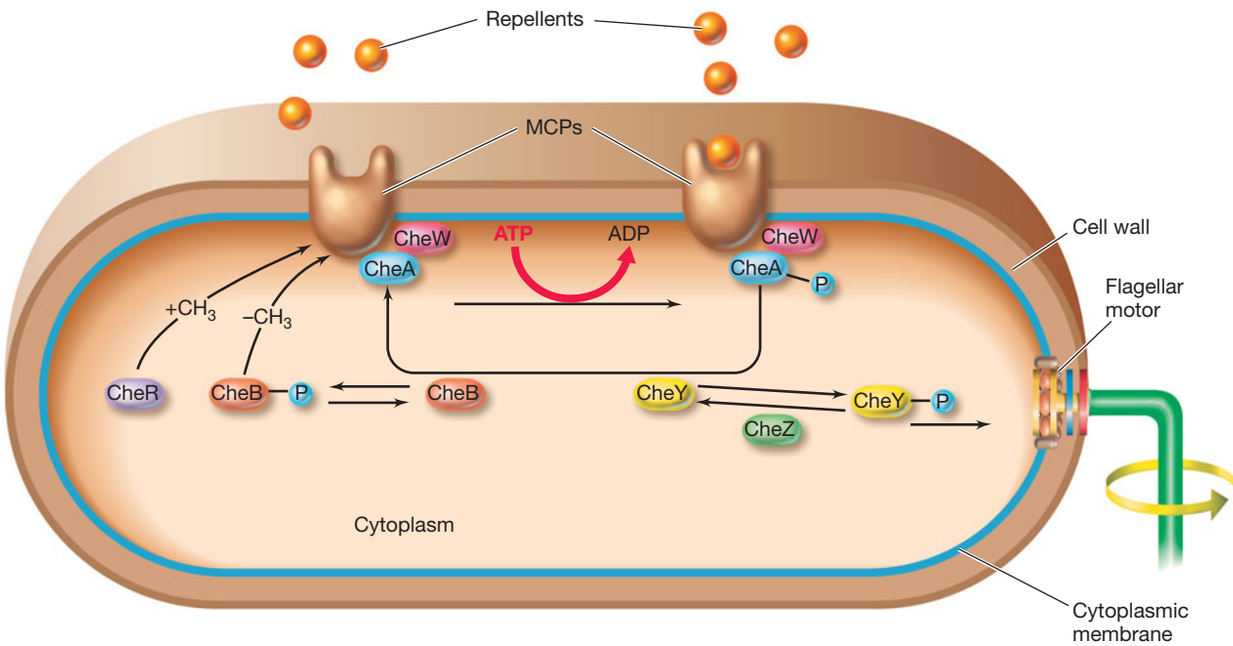

Figure 8.17 Interactions of MCPs, Che proteins, and the flagellar motor in bacterial chemotaxis. The methyl-accepting chemotaxis protein (MCP) forms a complex with the sensor kinase CheA and the coupling protein CheW. This combination triggers autophosphorylation of CheA to CheA-P. CheA-P can then phosphorylate the response regulators CheB and CheY. Phosphorylated CheY (CheY-P) binds to the flagellar motor switch. CheZ dephosphorylates CheY-P. CheR continually adds methyl groups to the MCP. CheB-P (but not CheB) removes them. The degree of methylation of the MCPs controls their ability to respond to attractants and repellents and leads to adaptation.

Step Two: Controlling Flagellar Rotation

CheY is a key protein in the system because it governs the direction of rotation of the flagellum. Recall that if rotation of the flagellum is counterclockwise, the cell will continue to move in a run, whereas if the flagellum rotates clockwise, the cell will tumble (↻Section 3.15). CheY-P interacts with the flagellar motor to induce clockwise flagellar rotation, which causes tumbling. When unphosphorylated, CheY cannot bind to the flagellar motor and the flagellum rotates counterclockwise; this causes the cell to run. Another protein, CheZ, dephosphorylates CheY, returning it to the form that allows runs instead of tumbles. Because repellents increase the level of CheY-P, they lead to tumbling, whereas attractants lead to a lower level of CheY-P and smooth swimming (runs).

Step Three: Adaptation

Once an organism has successfully responded to a stimulus, it must stop responding and reset the sensory system to await further signals. This is known as *adaptation*. During adaptation of the chemotaxis system, a feedback loop resets the system. This relies on the response regulator CheB, mentioned earlier.

As their name implies, MCPs can be methylated. The cytoplasmic protein CheR (Figure 8.17) continually adds methyl groups to the MCPs at a slow rate using *S*-adenosylmethionine as a methyl donor. The response regulator CheB is a demethylase that removes methyl groups from the MCPs. Phosphorylation of CheB greatly increases its rate of activity. The changes in methylation of the MCPs cause conformational changes similar to those due to binding of attractant or repellent. When MCPs are fully methylated they no longer respond to attractants, but are more sensitive to repellents. Conversely, when MCPs are unmethylated they respond highly to attractants, but are insensitive to repellents. Varying the methylation level thus allows adaptation to sensory signals.

If the level of an attractant remains high, CheY and CheB are not phosphorylated. Consequently, the cell swims smoothly. Methylation of the MCPs increases during this period because CheB-P is not present to rapidly demethylate them. However, MCPs no longer respond to the attractant when they become fully methylated. Therefore, if the level of attractant remains high but constant, the cell begins to tumble. Eventually, CheB becomes phosphorylated and CheB-P demethylates the MCPs. This resets the receptors and they can once again respond to further increases or decreases in level of attractants. Therefore the cell stops swimming if the attractant concentration is constant. It only continues to swim if even higher levels of attractant are encountered.

The course of events is just the opposite for repellents. Fully methylated MCPs respond best to an increasing gradient of repellents and send a signal for cell tumbling to begin. The cell then moves off in a random direction while MCPs are slowly demethylated. With this mechanism for adaptation, chemotaxis successfully achieves the ability to monitor small changes in the concentrations of both attractants and repellents over time.

Other Types of Taxis

In addition to chemotaxis, several other forms of taxis are known, for example, *phototaxis* (movement toward light) and *aerotaxis* (movement toward oxygen) (↻ Section 3.15). Many of the cytoplasmic Che proteins that function in chemotaxis also play a role in these. In phototaxis, a light sensor protein replaces the MCPs of chemotaxis, and in aerotaxis, a redox protein monitors levels of oxygen. These sensors then interact with cytoplasmic Che proteins to direct runs or tumbles. Thus several different kinds of signals converge on the same flagellar control system.

MiniQuiz

- What are the primary response regulator and the primary sensor kinase for regulating chemotaxis?
- Why is adaptation during chemotaxis important?
- How does the response of the chemotaxis system to an attractant differ from its response to a repellent?

UNIT 3

8.9 Quorum Sensing

Many prokaryotes respond to the presence in their surroundings of other cells of their species. Some prokaryotes have regulatory pathways that are controlled by the density of cells of their own kind. This is called **quorum sensing** (the word "quorum" in this sense means "sufficient numbers").

Mechanism of Quorum Sensing

Quorum sensing is a mechanism to assess population density. Many bacteria use this approach to ensure that sufficient cell numbers are present before starting activities that require a certain cell density to work effectively. For example, a pathogenic (disease-causing) bacterium that secretes a toxin will have no effect as a single cell; production of toxin by one cell alone would merely waste resources. However, if a sufficiently large population of cells is present, the coordinated expression of the toxin may successfully cause disease.

Quorum sensing is widespread among gram-negative bacteria but is also found in gram-positive bacteria. Each species that employs quorum sensing synthesizes a specific signal molecule called an **autoinducer**. This molecule diffuses freely across the cell envelope in either direction. Because of this, the autoinducer reaches high concentrations inside the cell only if there are many cells nearby, each making the same autoinducer. Inside the cell, the autoinducer binds to a specific activator protein and triggers transcription of specific genes (**Figure 8.18*b***).

There are several different classes of autoinducers (**Table 8.2**). The first to be identified were the *acyl homoserine lactones* (AHLs) (Figure 8.18*a*). Several different AHLs, with acyl groups of different lengths, are found in different species of gram-negative bacteria. In addition, many gram-negative bacteria make autoinducer 2 (AI-2; a cyclic furan derivative). This is apparently used as a common autoinducer between many species of bacteria. Gram-positive bacteria generally use certain short peptides as autoinducers.

Quorum sensing was first discovered as the mechanism of regulating light emission in bioluminescent bacteria. Several bacterial species can emit light, including the marine bacterium *Aliivibrio fischeri* (↻ Section 17.12). **Figure 8.19** shows bioluminescent colonies of *A. fischeri*. The light is generated by an

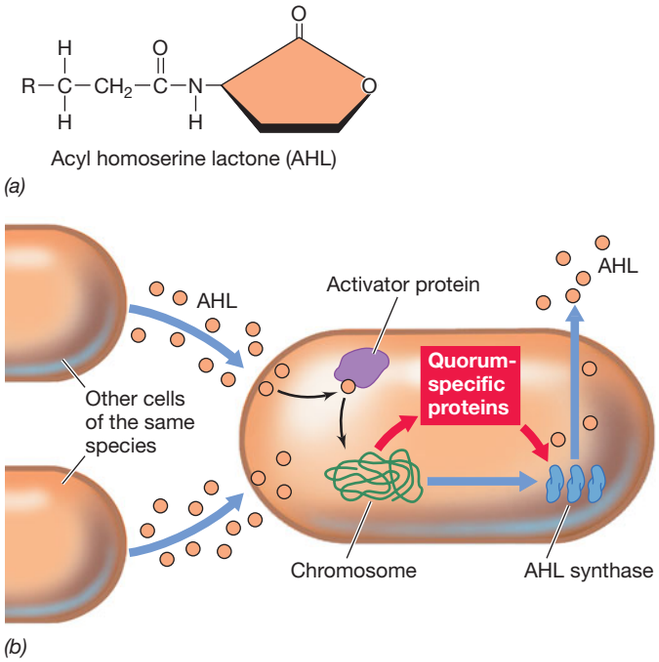

Figure 8.18 Quorum sensing. *(a)* General structure of an acyl homoserine lactone (AHL). Different AHLs are variants of this parent structure. R = alkyl group (C_1–C_{17}); the carbon next to the R group is often modified to a keto group (C=O). *(b)* A cell capable of quorum sensing expresses AHL synthase at basal levels. This enzyme makes the cell's specific AHL. When cells of the same species reach a certain density, the concentration of AHL rises sufficiently to bind to the activator protein, which activates transcription of quorum-specific genes.

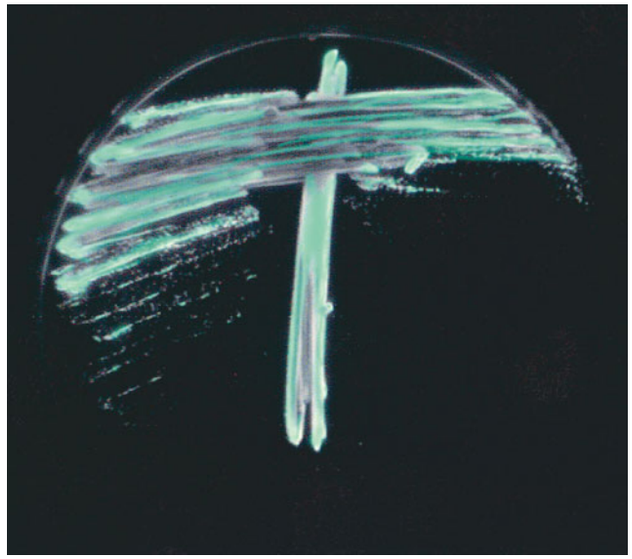

Figure 8.19 Bioluminescent bacteria producing the enzyme luciferase. Cells of the bacterium *Aliivibrio fischeri* were streaked on nutrient agar in a Petri dish and allowed to grow overnight. The photograph was taken in a darkened room using only the light generated by the bacteria.

enzyme called luciferase. The *lux* operons encode the proteins needed for bioluminescence. They are under control of the activator protein LuxR and are induced when the concentration of the specific *A. fischeri* AHL, *N*-3-oxohexanoyl homoserine lactone, becomes high enough. This AHL is synthesized by the enzyme encoded by the *luxI* gene.

Examples of Quorum Sensing

Various genes are controlled by quorum sensing, including some in pathogenic bacteria. For example, pseudomonads use 4-hydroxyalkyl quinolines as autoinducers to induce genes involved in virulence. In *Pseudomonas aeruginosa*, for instance, quorum sensing triggers the expression of a large number of unrelated genes when the population density becomes sufficiently high. These genes assist cells of *P. aeruginosa* in the transition from growing freely suspended in liquid to growing in a semisolid matrix called a *biofilm* (↻ Section 23.4). The biofilm, formed by specific polysaccharides produced by *P. aeruginosa*, increases the pathogenicity of this organism and prevents the penetration of antibiotics.

The pathogenesis of *Staphylococcus aureus* (↻ Section 33.9) involves, among many other things, the production and secretion of small extracellular peptides that damage host cells or that interfere with the immune system. The genes encoding these virulence factors are under the control of a quorum-sensing system that uses a small peptide as autoinducer. The regulation of these virulence genes is quite complex and requires a regulatory RNA molecule as well as regulatory proteins that form a signal transduction system.

Quorum sensing also occurs in microbial eukaryotes. For example, in the yeast *Saccharomyces cerevisiae*, specific aromatic

Table 8.2 *Examples of quorum sensing and autoinducers*

Organism	*Autoinducer*	*Receptor*	*Process regulated*
Proteobacteria	Acyl homoserine lactones	LuxR protein	Diverse processes
Many diverse bacteria	AI-2 (furanone ± borate)[a]	LuxQ protein	Diverse processes
Pseudomonads	4-Hydroxyalkyl quinolines	PqsR protein	Virulence; biofilms
Streptomyces	Gamma-butyrolactones	ArpA repressor	Antibiotic synthesis; sporulation
Gram-positive bacteria	Oligopeptides (linear or cyclic)	Two-component systems	Diverse processes
Yeast	Aromatic alcohols	?	Filamentation

[a]The AI-2 autoinducer exists in several slightly different structures, some of which have an attached borate group.

alcohols are produced as autoinducers and control the transition between growth of *S. cerevisiae* as single cells and as elongated filaments. Similar transitions are seen in other fungi, some of which cause disease in humans. An example is *Candida*, whose quorum sensing is mediated by the long-chain alcohol farnesol.

Some eukaryotes produce molecules that interfere with bacterial quorum sensing. Most of those known so far are furanone derivatives with halogens attached. These mimic the AHLs or AI-2 and disrupt bacterial behavior that relies on quorum sensing. Quorum-sensing disruptors have been suggested to have possible future applications in dispersing bacterial biofilms and preventing the expression of virulence genes.

MiniQuiz

- What properties are required for a molecule to function as an autoinducer?
- How do the autoinducers used in quorum sensing by gram-negative bacteria differ from those used by gram-positive bacteria?

8.10 The Stringent Response

Nutrient levels in the natural environments of bacterial cells often change significantly, even if only briefly. Such changing conditions can easily be simulated in the laboratory, and much work has been done with *Escherichia coli* and other bacteria on the regulation of gene expression following a "shift down" or "shift up" in nutrient status. These include, in particular, the regulatory events triggered by starvation for amino acids or energy.

As a result of a shift down from amino acid excess to limitation, as occurs when a culture is transferred from a rich complex medium to a defined medium with a single carbon source, the synthesis of rRNA and tRNA ceases almost immediately (**Figure 8.20*a***). No new ribosomes are produced. Protein and DNA synthesis is curtailed, but the biosynthesis of new amino acids is activated. Following such a shift, new proteins must be made to synthesize the amino acids no longer available in the environment; these are made by existing ribosomes. After a while, rRNA synthesis (and hence, the production of new ribosomes) begins again but at a new rate commensurate with the cell's reduced growth rate (Figure 8.20*a*). This course of events is called the **stringent response** (or stringent control) and is another example of global control.

The stringent response is triggered by a mixture of two regulatory nucleotides, *guanosine tetraphosphate* (ppGpp) and *guanosine pentaphosphate* (pppGpp); this mixture is often written as (p)ppGpp (Figure 8.20*b*). In *E. coli*, these nucleotides, which are also called *alarmones*, rapidly accumulate during a shift down from amino acid excess to amino acid starvation. Alarmones are synthesized by a specific protein, called RelA, using ATP as a phosphate donor (Figure 8.20*b,c*). RelA adds two phosphate groups from ATP to GTP or GDP, thus producing pppGpp or ppGpp, respectively. RelA is associated with the 50S subunit of the ribosome and is activated by a signal from the ribosome during amino acid limitation. When the growth of the cell is limited by a shortage of amino acids, the pool of *uncharged* tRNAs increases relative to *charged* tRNAs. Eventually, an uncharged

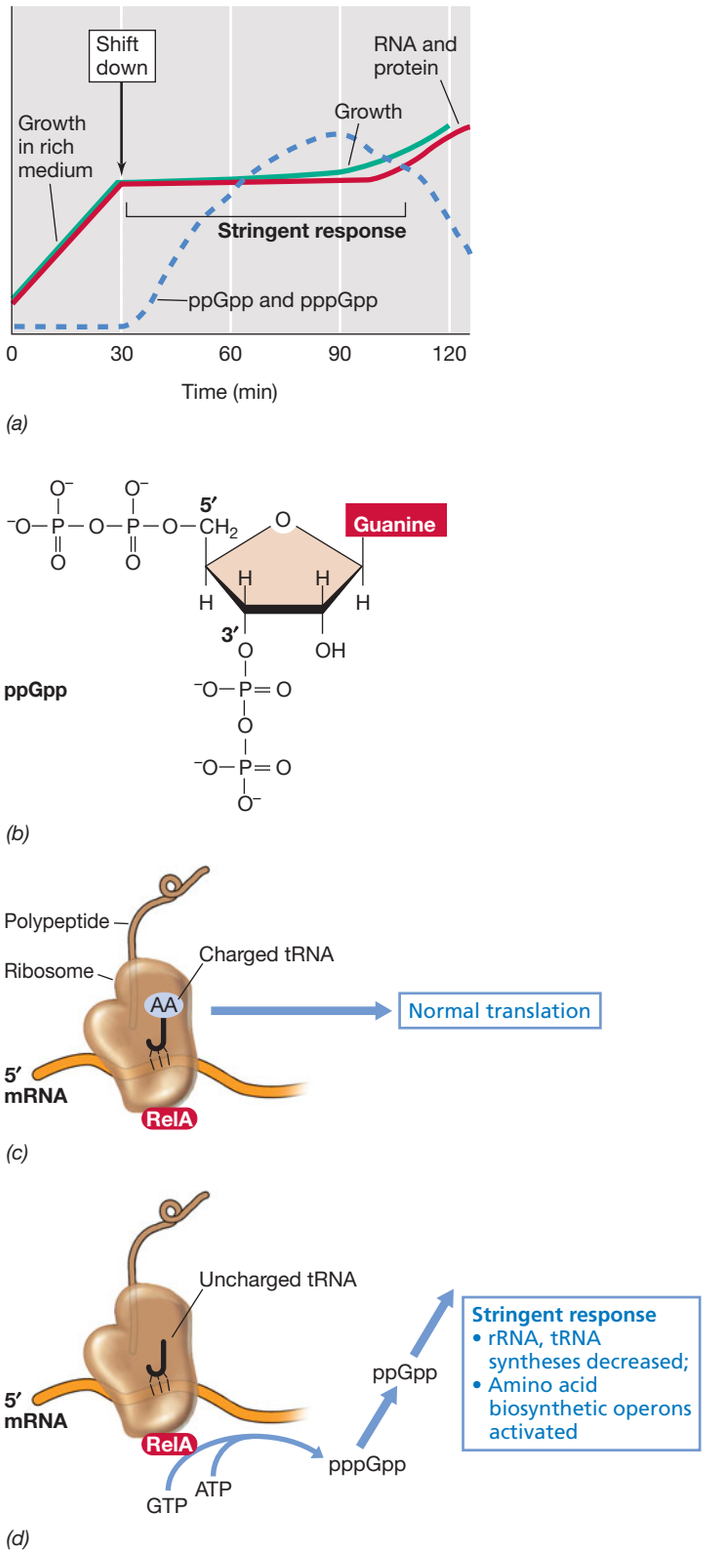

Figure 8.20 The stringent response. *(a)* Upon nutrient downshift, rRNA, tRNA, and protein syntheses temporarily cease. Sometime later, growth resumes at a new (decreased) rate. *(b)* Structure of guanosine tetraphosphate (ppGpp), a trigger of the stringent response. *(c)* Normal translation, which requires charged tRNAs. *(d)* Synthesis of ppGpp. When cells are starved for amino acids, an uncharged tRNA can bind to the ribosome, which stops ribosome activity. This event triggers the RelA protein to synthesize a mixture of pppGpp and ppGpp.

tRNA is inserted into the ribosome instead of a charged tRNA during protein synthesis. When this happens, the ribosome stalls, and this leads to (p)ppGpp synthesis by RelA (Figure 8.20*d*). The protein Gpp converts pppGpp to ppGpp so that ppGpp is the major overall product.

The alarmones ppGpp and pppGpp have global control effects. They strongly inhibit rRNA and tRNA synthesis by binding to RNA polymerase and preventing initiation of transcription of genes for these RNAs. On the other hand, alarmones activate the biosynthetic operons for certain amino acids as well as catabolic operons that yield precursors for amino acid synthesis. By contrast, operons that encode biosynthetic proteins whose amino acid products are present in sufficient amounts remain shut down. The stringent response also inhibits the initiation of new rounds of DNA synthesis and cell division and slows down the synthesis of cell envelope components, such as membrane lipids. Efficient binding of (p)ppGpp to RNA polymerase requires the protein DksA, which is needed to position the (p)ppGpp correctly in the channel that normally allows substrates (that is, nucleoside triphosphates) into the RNA polymerase active site.

In addition to RelA, another protein, SpoT, helps trigger the stringent response. The SpoT protein can either make (p)ppGpp or degrade it. Under most conditions, SpoT is responsible for degrading (p)ppGpp; however, SpoT synthesizes (p)ppGpp in response to certain stresses or when there is a shortage of energy. Thus the stringent response results not only from the absence of precursors for protein synthesis, but also from the lack of energy for biosynthesis.

The stringent response can be thought of as a mechanism for adjusting the cell's biosynthetic machinery to the availability of the required precursors and energy. By so doing, the cell achieves a new balance between anabolism and catabolism. In many natural environments, nutrients appear suddenly and are consumed rapidly. Thus a global mechanism such as the stringent response that balances the metabolic state of a cell with the availability of precursors and energy likely improves its ability to compete in nature.

The RelA/(p)ppGpp system is found only in *Bacteria* and in the chloroplasts of plants. *Archaea* and eukaryotes do not make (p)ppGpp in response to resource shortages. Although *Archaea* display an overall response similar to the stringent response of *Bacteria* when faced with carbon and energy shortages, they use regulatory mechanisms different from those described here to deal with these situations.

MiniQuiz

- Which genes are activated during the stringent response and why?
- Which genes are repressed during the stringent response and why?
- How are the alarmones ppGpp and pppGpp synthesized?

8.11 Other Global Control Networks

Catabolite repression and the stringent response are both examples of global control. There are several other global control systems in *Escherichia coli* (and probably in all prokaryotes), and a few of these are listed in **Table 8.3**. Global control systems regulate many genes comprising more than one regulon (Section 8.4). Global control networks may include activators, repressors, signal molecules, two-component regulatory systems (Section 8.7), regulatory RNA (Sections 8.14 and 8.15), and alternative sigma (σ) factors (↻ Section 6.13).

An example of a global response that is widespread in all three domains of life is the response to high temperature. In many bacteria this **heat shock response** is largely controlled by alternative σ factors.

Heat Shock Proteins

Most proteins are relatively stable. Once made, they continue to perform their functions and are passed along at cell division. However, some proteins are less stable at elevated temperatures and tend to unfold. Improperly folded proteins are recognized by protease enzymes and are degraded. Consequently, cells that are heat stressed induce the synthesis of a set of proteins, the **heat shock proteins**, that help counteract the damage. Heat shock proteins assist the cell in recovering from stress. They are induced not only by heat, but also by several other stress factors that the cell can encounter. These include exposure to high levels of certain chemicals, such as ethanol, and exposure to high doses of ultraviolet (UV) radiation.

Table 8.3 *Examples of global control systems known in* Escherichia coli[a]

System	*Signal*	*Primary activity of regulatory protein*	*Number of genes regulated*
Aerobic respiration	Presence of O_2	Repressor (ArcA)	>50
Anaerobic respiration	Lack of O_2	Activator (FNR)	>70
Catabolite repression	Cyclic AMP level	Activator (CRP)	>300
Heat shock	Temperature	Alternative sigmas (RpoH and RpoE)	36
Nitrogen utilization	NH_3 limitation	Activator (NR_I)/alternative sigma RpoN	>12
Oxidative stress	Oxidizing agents	Activator (OxyR)	>30
SOS response	Damaged DNA	Repressor (LexA)	>20

[a]For many of the global control systems, regulation is complex. A single regulatory protein can play more than one role. For instance, the regulatory protein for aerobic respiration is a repressor for many promoters but an activator for others, whereas the regulatory protein for anaerobic respiration is an activator protein for many promoters but a repressor for others. Regulation can also be indirect or require more than one regulatory protein. Many genes are regulated by more than one global system.

In *E. coli* and in most prokaryotes examined, there are three major classes of heat shock protein, Hsp70, Hsp60, and Hsp10. We have encountered these proteins before, although not by these names (↻ Section 6.21). The Hsp70 protein of *E. coli* is DnaK, which prevents aggregation of newly synthesized proteins and stabilizes unfolded proteins. Major representatives of the Hsp60 and Hsp10 families in *E. coli* are the proteins GroEL and GroES, respectively. These are molecular chaperones that catalyze the correct refolding of misfolded proteins. Another class of heat shock proteins includes various proteases that degrade denatured or irreversibly aggregated proteins.

The heat shock proteins are very ancient and highly conserved. Molecular sequencing of heat shock proteins, especially Hsp70, has been used to help unravel the phylogeny of eukaryotes. Heat shock proteins are present in all cells, although the regulatory system that controls their expression varies greatly in different groups of organisms.

Heat Shock Response

In many bacteria, such as *E. coli*, the heat shock response is controlled by the alternative σ factors RpoH (σ^{32}) and RpoE (**Figure 8.21**). RpoH controls expression of heat shock proteins in the cytoplasm, and RpoE regulates the expression of a different set of heat shock proteins in the periplasm and cell envelope. RpoH is normally degraded within a minute or two of its synthesis. However, when cells suffer a heat shock, degradation of RpoH is inhibited and its level therefore increases. Consequently, transcription of those operons whose promoters are recognized by RpoH increases too. The rate of degradation of RpoH depends on the level of free DnaK protein, which inactivates RpoH. In unstressed cells the level of free DnaK is relatively high and the level of intact RpoH is correspondingly low. However, if heat stress unfolds proteins, DnaK binds preferentially to the unfolded proteins and so is no longer free to promote degradation of RpoH. Thus, the more denatured proteins there are, the lower the level of free DnaK and the higher the level of RpoH; the result is heat shock gene expression.

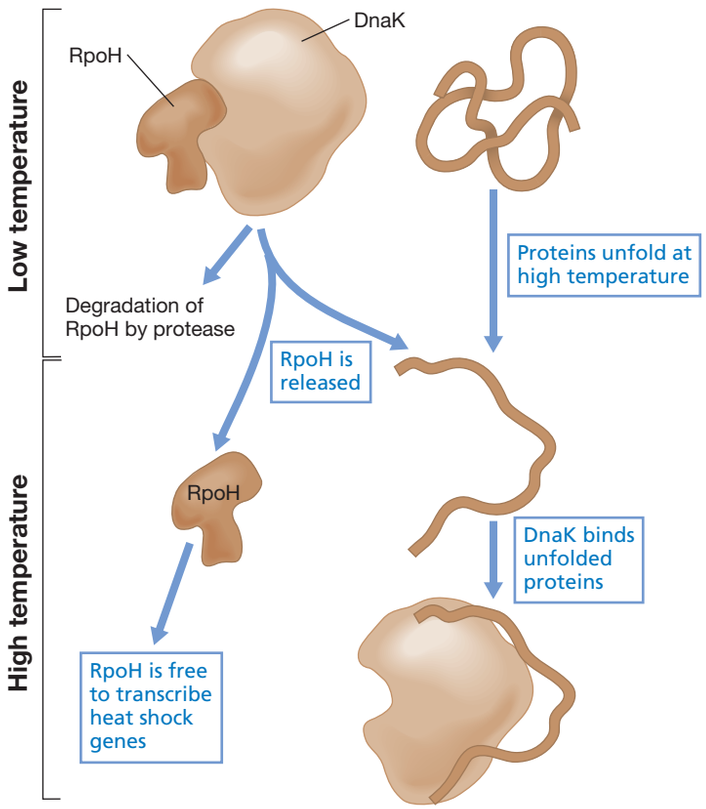

Figure 8.21 Control of heat shock in *Escherichia coli*. The RpoH alternative sigma factor is broken down rapidly by proteases at normal temperatures. This is stimulated by binding of the DnaK chaperonin to RpoH. At high temperatures, some proteins are denatured, and DnaK recognizes and binds to the unfolded polypeptide chains. This removes DnaK from RpoH, which slows the degradation rate. The level of RpoH rises, and the heat shock genes are transcribed.

When the stress situation has passed, for example, upon a temperature downshift, RpoH is rapidly inactivated by DnaK and the synthesis of heat shock proteins is greatly reduced. Because heat shock proteins perform vital functions in the cell, there is always a low level of these proteins present, even under optimal conditions. However, the rapid synthesis of heat shock proteins in stressed cells emphasizes how important they are in surviving excessive heat, chemicals, or physical agents. Such stresses can generate large amounts of inactive proteins that need to be refolded (and in the process, reactivated) or degraded to release free amino acids for the synthesis of new proteins.

There is also a heat shock response in *Archaea*, even in species that grow best at very high temperatures. An analog of the bacterial Hsp70 is found in many *Archaea* and is structurally quite similar to those found in gram-positive species of *Bacteria*. Hsp70 is also present in eukaryotes. In addition, other types of heat shock proteins are present in *Archaea* that are unrelated to stress proteins of *Bacteria*.

One problem faced by all cells during cold shock is that RNA, including mRNA, tends to form stable secondary structures, especially stem–loop structures, that may interfere with translation. Cold shock proteins include several RNA-binding proteins. Some of these prevent secondary structure formation and others (RNA helicases) unwind base-paired regions in RNA.

MiniQuiz

- What triggers the heat shock response?
- Why do cells have more than one type of σ factor?
- Why might the proteins induced during heat shock not be needed during cold shock?

IV Regulation of Development in Model *Bacteria*

Differentiation and development are largely characteristics of multicellular organisms. Because most prokaryotic microorganisms grow as single cells, few show differentiation. Nonetheless, occasional examples among single-celled prokaryotes illustrate the basic principle of differentiation, namely that one cell gives rise to two genetically identical descendants that perform different roles and must therefore express different sets of

genes. Here we discuss two well-studied examples, the formation of endospores in the gram-positive bacterium *Bacillus* and the formation of two cell types, motile and stationary, in the gram-negative bacterium *Caulobacter*.

Although forming just two different cell types may seem superficially simple, the regulatory systems that control these processes are highly complex. There are three major phases for the regulation of differentiation: (1) triggering the response, (2) asymmetric development of two sister cells, and (3) reciprocal communication between the two differentiating cells.

8.12 Sporulation in *Bacillus*

Many microorganisms, both prokaryotic and eukaryotic, respond to adverse conditions by forming spores (↻ Section 3.12). Once favorable conditions return, the spore germinates and the microorganism returns to its normal lifestyle. Among the *Bacteria*, the genus *Bacillus* is well known for the formation of endospores, that is, spores formed inside a mother cell. Prior to endospore formation, the cell divides asymmetrically. The smaller cell develops into the endospore, which is surrounded by the larger mother cell. Once development is complete, the mother cell bursts, releasing the endospore.

Endospore formation in *Bacillus subtilis* is triggered by unfavorable conditions, such as starvation, desiccation, or growth-inhibitory temperatures. Multiple aspects of the environment are monitored by a group of five sensor kinases. These function via a phosphotransfer relay system whose mechanism resembles that of a two-component regulatory system (Section 8.7), but is considerably more complex (**Figure 8.22**). The net result of multiple adverse conditions is the successive phosphorylation of several proteins called *sporulation factors*, culminating with sporulation factor Spo0A. When Spo0A is highly phosphorylated, sporulation proceeds. Spo0A controls the expression of several genes. The product of one of these, SpoIIE, is responsible for removing the phosphate from SpoIIAA. This allows SpoIIAA in turn to remove the anti-sigma factor, SpoIIAB, and liberate the σ factor, σF, as discussed below.

Once triggered, endospore development is controlled by four different σ factors, two of which, σF and σG, activate genes needed inside the developing endospore itself, and two of which, σE and σK, activate genes needed in the mother cell surrounding the endospore (Figure 8.22b). The sporulation signal, transmitted via Spo0A, activates σF in the smaller cell that is destined to become the endospore. σF is already present, but is inactive, as it is bound by an anti-σ factor. The signal from Spo0A activates a protein that binds to the anti-σ factor and inactivates it, so liberating σF. Once free, σF binds to RNA polymerase and promotes transcription (inside the spore) of genes whose products are needed for the next stage of sporulation. These include the gene for the sigma factor σG and the genes for proteins that cross into the mother cell and activate σE. Active σE is required for transcription inside the mother cell of yet more genes, including the gene for σK. The sigma factors σG (in the endospore) and σK (in the mother cell) are required for transcription of genes needed even later in the sporulation process.

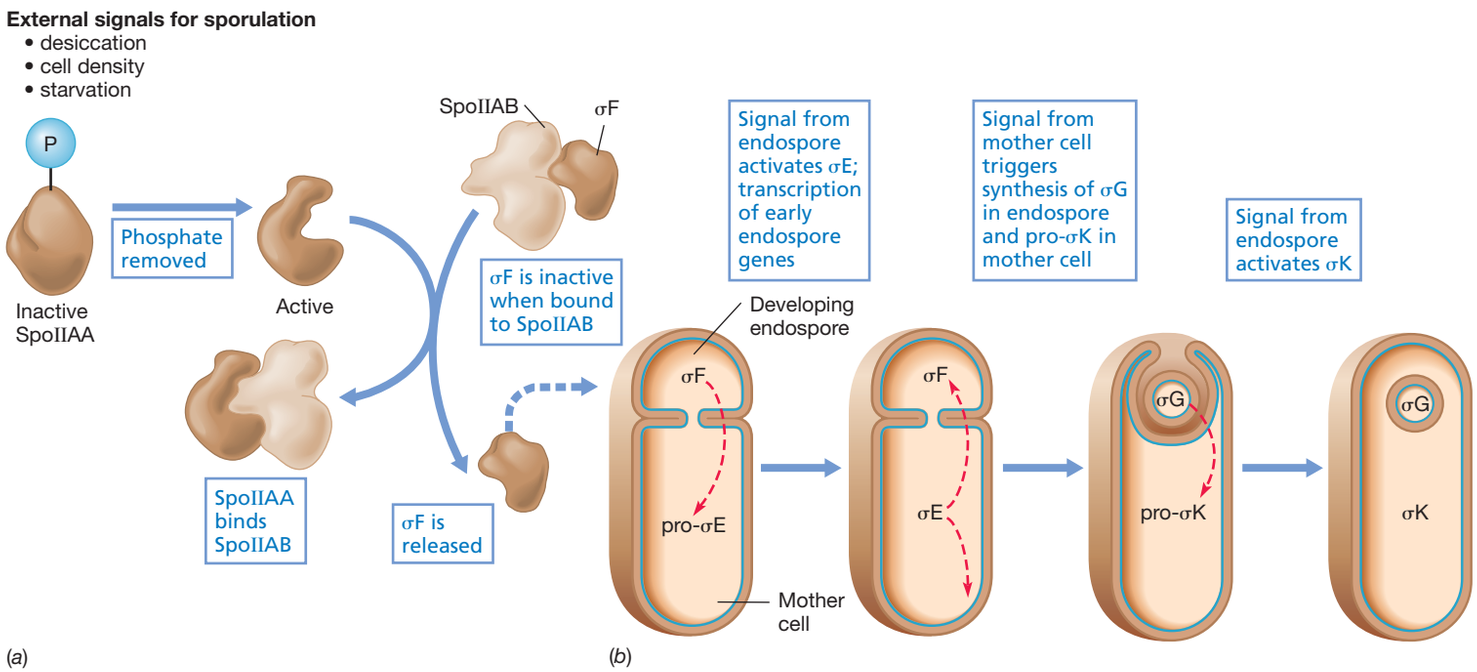

Figure 8.22 **Control of endospore formation in *Bacillus*.** After an external signal is received, a cascade of sigma (σ) factors controls differentiation. *(a)* Active SpoIIAA binds the anti-σ factor SpoIIAB, thus liberating the first σ factor, σF. *(b)* σF initiates a cascade of sigma factors, some of which already exist and need to be activated, others of which are not yet present and whose genes must be expressed. These σ factors then promote transcription of genes needed for endospore development.

One fascinating aspect of endospore formation is that it is preceded by what is in effect cellular cannibalism. Those cells in which Spo0A has already become activated secrete a protein that lyses nearby cells of the same species whose Spo0A protein has not yet become activated. This toxic protein is accompanied by a second protein that delays sporulation of neighboring cells. Cells committed to sporulation also make an antitoxin protein to protect themselves against the effects of their own toxin. Their sacrificed sister cells are used as a source of nutrients for developing endospores. Shortages of certain nutrients, such as phosphate, increase the expression level of the toxin-encoding gene.

MiniQuiz

- How are different sets of genes expressed in the developing endospore and the mother cell?
- What is an anti-σ factor and how can its effect be overcome?

8.13 *Caulobacter* Differentiation

Caulobacter provides another example in which a cell divides into two genetically identical daughter cells that perform different roles and express different sets of genes. *Caulobacter* is a species of *Proteobacteria* that is common in aquatic environments, typically in waters that are nutrient-poor (↻ Section 17.16). In the *Caulobacter* life cycle, free-swimming (swarmer) cells alternate with cells that lack flagella and are attached to surfaces by a stalk with a holdfast at its end. The role of the swarmer cells is dispersal, as swarmers cannot divide or replicate their DNA. Conversely, the role of the stalked cell is reproduction.

The *Caulobacter* cell cycle is controlled by three major regulatory proteins whose concentrations oscillate in succession (**Figure 8.23**). Two of these are the transcriptional regulators, GcrA and CtrA. The third is DnaA, a protein that functions both in its normal role in initiating DNA replication and also as a transcriptional regulator. Each of these regulators is active at a specific stage in the cell cycle, and each controls many other genes that are needed at that particular stage in the cycle.

CtrA is activated by phosphorylation in response to external signals. Once phosphorylated, CtrA-P activates genes needed for the synthesis of the flagella and other functions in swarmer cells. Conversely, CtrA-P represses the synthesis of GcrA and also inhibits the initiation of DNA replication by binding to and blocking the origin of replication (Figure 8.23). As the cell cycle proceeds, CtrA is degraded by a specific protease; as a consequence, levels of DnaA rise. The absence of CtrA-P allows access to the chromosomal origin of replication, and, as in all *Bacteria*, DnaA binds to the origin and triggers the initiation of DNA replication (↻ Section 6.9). In addition, in *Caulobacter* DnaA activates several other genes needed for chromosomal replication. The level of DnaA then falls due to protease degradation, and the level of GcrA rises. The GcrA regulator promotes the elongation phase of chromosome replication, cell division, and the growth of the stalk on the immobile daughter cell. Eventually, GcrA levels fall and high levels of CtrA reappear (in the daughter cell destined to swim away) (Figure 8.23).

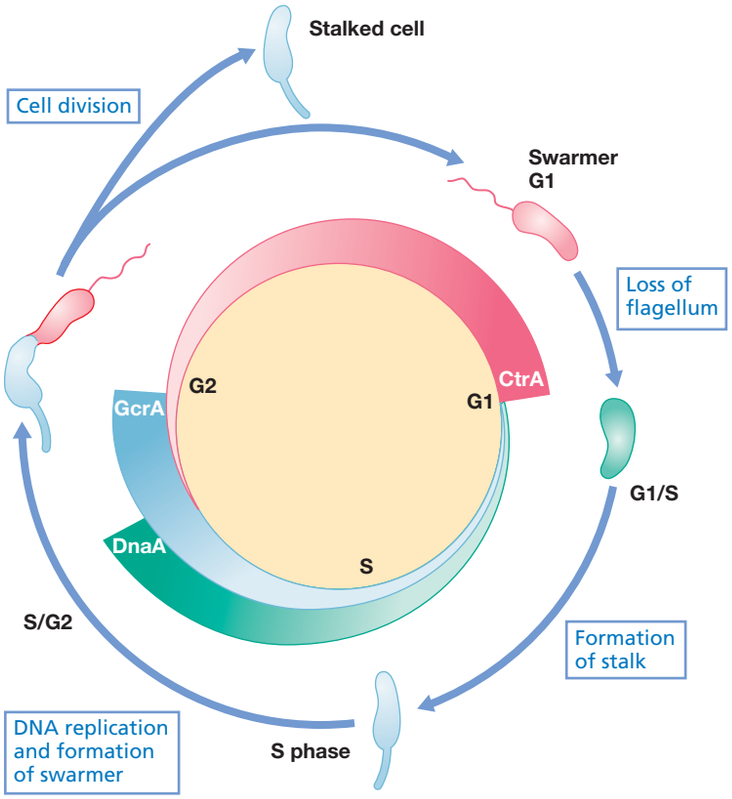

Figure 8.23 **Cell cycle regulation in *Caulobacter*.** Three global regulators, CtrA, DnaA, and GcrA, oscillate in levels through the cycle as shown. G1 and G2 are the two growth phases and S is the synthesis (of DNA) phase. In G1 swarmer cells, CtrA represses initiation of DNA replication and expression of GcrA. At the G1/S transition, CtrA is degraded and DnaA levels rise. DnaA binds to the origin of replication and initiates replication. GcrA also rises and activates genes for cell division and DNA synthesis. At the S/G2 transition, CtrA levels begin to rise again and shut down GcrA expression. GcrA levels slowly decline in the stalked cell but are rapidly degraded in the swarmer. CtrA is degraded in the stalked cell.

Many of the details of the regulation of the *Caulobacter* cell cycle are still uncertain. Both external stimuli and internal factors such as nutrient and metabolite levels affect the cycle, but how this information is integrated into the overall control system is only partly understood. However, since its genome has been sequenced and good genetic systems for gene transfer and analysis are available, differentiation in *Caulobacter* has been used as a model system for studying cell developmental processes.

MiniQuiz

- Why are the levels of DnaA protein controlled during the *Caulobacter* cell cycle?
- When do the regulators CtrA and GcrA carry out their main roles during the *Caulobacter* life cycle?

V RNA-Based Regulation

Thus far we have focused on mechanisms in which regulatory proteins sense signals or bind to DNA. In some cases a single protein does both; in other cases separate proteins carry out these two activities. Nonetheless, all of these mechanisms rely on *regulatory proteins*. However, RNA itself may regulate gene expression, both at the level of transcription and the level of translation of mRNA to produce proteins.

RNA molecules that are not translated to give proteins are collectively known as **noncoding RNA (ncRNA)**. This category includes the rRNA and tRNA molecules that take part in protein synthesis and the RNA present in the signal recognition particle that is involved in protein secretion (↻ Section 6.21). Noncoding RNA also includes small RNA molecules necessary for RNA processing, especially the splicing of mRNA in eukaryotes (↻ Section 7.8). In addition, small RNA (sRNA) molecules that range from approximately 40–400 nucleotides long and regulate gene expression are widely distributed in both prokaryotes and eukaryotes. In *Escherichia coli*, for example, a number of sRNA molecules have been found to regulate various aspects of cell physiology by binding to other RNAs or in some cases even to other small molecules.

8.14 RNA Regulation and Antisense RNA

The most frequent way in which regulatory RNA molecules exert their effects is by base pairing with other RNA molecules, usually mRNA, that have regions of complementary sequence. These double-stranded regions tie up the mRNA and prevent its translation (**Figure 8.24**). Small RNAs (sRNAs) that show this activity are called *antisense RNA*, because the sRNA has a sequence complementary to the coding sense of the mRNA.

Antisense RNAs bind to their mRNA complements to form double-stranded RNAs that cannot be translated and are soon degraded by specific ribonucleases (Figure 8.24). This removes the mRNA and thus prevents the synthesis of new protein molecules from mRNAs already present in the cell but whose gene products are no longer needed because of a change in conditions.

Theoretically, antisense RNA could be made by transcribing the nontemplate strand of the same gene that yielded the target mRNA. Instead, a distinct "anti-gene" is used to form the antisense RNA. Only a relatively short piece of antisense RNA is needed to block transcription of mRNA, and therefore the "anti-gene" that encodes the antisense RNA is much shorter than the gene that encodes the original message. Typically, antisense RNAs are around 100 nucleotides long and bind to a target region approximately 30 nucleotides long. In addition, each antisense RNA can usually regulate several different mRNAs, all of which share the same target sequence for antisense RNA binding.

Transcription of antisense RNA is enhanced under conditions in which its target genes need to be turned off. For example, the RyhB antisense RNA of *Escherichia coli* is transcribed when iron is limiting for growth. RyhB antisense RNA binds to a dozen or more target mRNAs that encode proteins needed for iron metabolism or that use iron as cofactors. Binding of RyhB antisense RNA blocks translation of the mRNA. The base-paired RyhB/mRNA molecules are then degraded by ribonucleases, in particular, ribonuclease E. This forms part of the mechanism by which *E. coli* and related bacteria respond to a shortage of iron. Other responses to iron limitation in *E. coli* include transcriptional controls involving repressor and activator proteins (Sections 8.3 and 8.4) that increase the capacity of cells to take up iron and to tap into intracellular iron stores.

The binding of many antisense RNAs to their targets depends on a small protein, Hfq, that binds not only to both RNA molecules but also to ribonuclease E (**Figure 8.25**). Hfq protein forms hexameric rings with RNA-binding sites on both surfaces. Hfq and similar proteins are known as *RNA chaperones*, as they help small RNA molecules, including many antisense RNAs, maintain their correct structure.

Although antisense RNA usually blocks translation of mRNA, occasional examples are known in which antisense RNA does

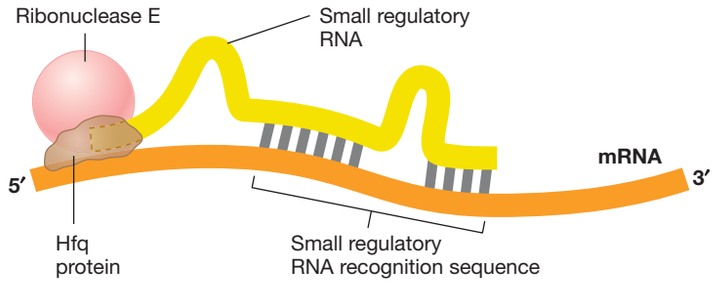

Figure 8.25 The RNA chaperone Hfq holds RNAs together. Binding of antisense RNA to mRNA often requires the Hfq protein. Antisense RNA molecules usually have several stem–loop structures. One consequence is that the complementary base sequence that recognizes the mRNA is noncontiguous. The antisense RNA blocks the ribosome-binding site on the mRNA and prevents its translation. Ribonuclease E, also bound by Hfq, then begins to degrade the mRNA.

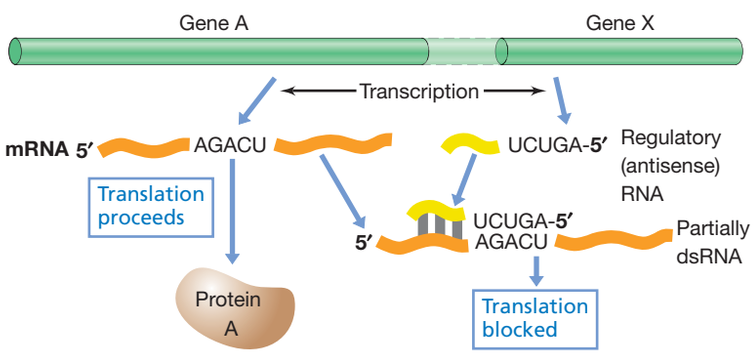

Figure 8.24 Regulation by antisense RNA. Gene A is transcribed from its promoter to yield an mRNA that can be translated to form protein A. Gene X is a small gene with a sequence identical to that of part of Gene A but with its promoter at the opposite end. Therefore, if it is transcribed, the resulting antisense RNA will be complementary to the mRNA of gene A. If these two RNAs base-pair, forming double-stranded RNA, translation will be blocked.

The CRISPR Antiviral Defense System

Ever since RNA interference (RNAi) was discovered in eukaryotes (↻ Section 7.10), scientists have wondered whether bacteria have an equivalent system to protect themselves against virus attack. Recent discoveries have revealed that although bacteria do not have RNAi, they do have another RNA-based defense program to destroy invading virus genomes. In fact, the bacterial CRISPR system tackles both RNA and DNA viruses, unlike RNAi, which only works against RNA viruses.

CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats. The CRISPR region on the bacterial chromosome is essentially a memory bank of hostile virus sequences. It consists of many different segments of virus sequence alternating with identical repeated sequences (Figure 1). The CRISPR system provides resistance to any viruses that contain the same or very closely related sequences.

The proteins of the CRISPR system (CRISPR associated proteins, or CAS proteins) perform two roles. Some use the stored sequence information to recognize intruding virus genomes and destroy them. Others are involved in obtaining and storing segments of virus sequence, a process that remains obscure. The CAS proteins are encoded by genes that lie upstream of the CRISPR sequences (Figure 1).

The CRISPR region is transcribed as a whole into a long RNA molecule that is then cleaved by CAS proteins in the middle of each of the repeated sequences. This converts it into individual virus-specific segments. If one of these segments base-pairs with the nucleic acid of an invading virus, then the virus DNA or RNA is destroyed by other CAS proteins.

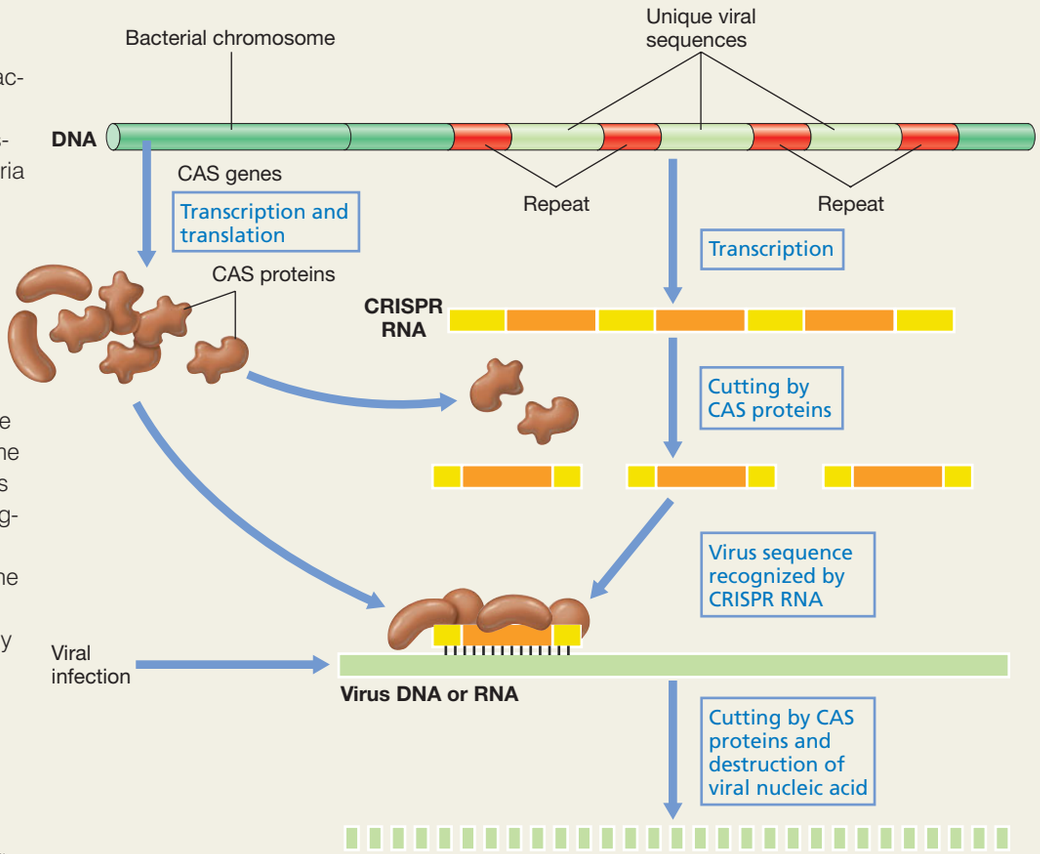

Figure 1 Operation of the CRISPR system. The CRISPR region on the bacterial chromosome is transcribed into a long RNA molecule that is then cut into segments by some of the CAS proteins. Each segment carries a single virus-specific sequence. If one of these short CRISPR RNA molecules recognizes a virus nucleic acid by base pairing, other CAS proteins destroy the virus DNA or RNA.

The CRISPR system is widely distributed in both *Archaea* and *Bacteria*. Approximately 90% of the sequenced genomes of *Archaea* and 70% of those of *Bacteria* possess the CRISPR system. However, many occurrences of the CRISPR system detected by genomic sequencing appear to be incomplete or defective.

just the opposite and actually enhances the translation of its target mRNA. It is hypothesized that in these cases the native mRNA forms a secondary structure that prevents translation. The antisense RNA is thought to bind to a short region of the mRNA and unfold it, thereby allowing access to the ribosome.

Antisense RNA does not always work via an effect on mRNA. For example, the replication of the high copy number plasmid ColE1 is regulated by an sRNA that primes DNA synthesis and its antisense partner that blocks initiation of DNA synthesis. The level of the antisense RNA determines how often replication is initiated.

Regulation by antisense RNA usually modulates the expression of genes that are also controlled by other systems, and many complex examples are known in higher organisms. For example, in the mold *Neurospora* the time of day controls growth via a complex mechanism that uses antisense RNA. The levels of sense and

antisense transcripts for a biological-clock gene are found to cycle out of step with each other in response to day and night cycles.

Fragments of antisense RNA are also used to detect and destroy viral intruders in the CRISPR defense mechanism (Microbial Sidebar, "The CRISPR Antiviral Defense System").

MiniQuiz

- Why are antisense RNAs much shorter than the mRNA molecules to which they bind?
- How do cells synthesize antisense RNA molecules?
- What happens to mRNA molecules following binding of their antisense RNAs?

8.15 Riboswitches

Recently it has become clear that RNA can carry out many roles once thought to be limited to proteins. In particular, RNA can specifically recognize and bind other molecules, including low-molecular-weight metabolites. It is important to emphasize that such binding does not involve complementary base pairing (as does binding of the antisense RNA described in the previous section) but results from the folding of the RNA into a specific three-dimensional structure that recognizes the target molecule, much as a protein enzyme recognizes its substrate. RNA molecules that are catalytically active are called ribozymes. Other RNA molecules resemble repressors and activators in binding metabolites such as amino acids or vitamins and regulating gene expression; these are the **riboswitches**.

Certain mRNAs contain regions upstream of the coding sequences that can fold into specific three-dimensional structures that bind small molecules (**Figure 8.26**). These recognition domains are riboswitches and exist as two alternative structures, one with the small molecule bound and the other without. Alternation between the two forms of the riboswitch thus depends on the presence or absence of the small molecule, which in turn controls expression of the mRNA. Riboswitches have been found that control the synthesis of enzymes in biosynthetic pathways for various enzymatic cofactors, such as the vitamins thiamine, riboflavin, and cobalamin (B_{12}), for a few amino acids, for the purine bases adenine and guanine, and for glucosamine 6-phosphate, a precursor in peptidoglycan synthesis.

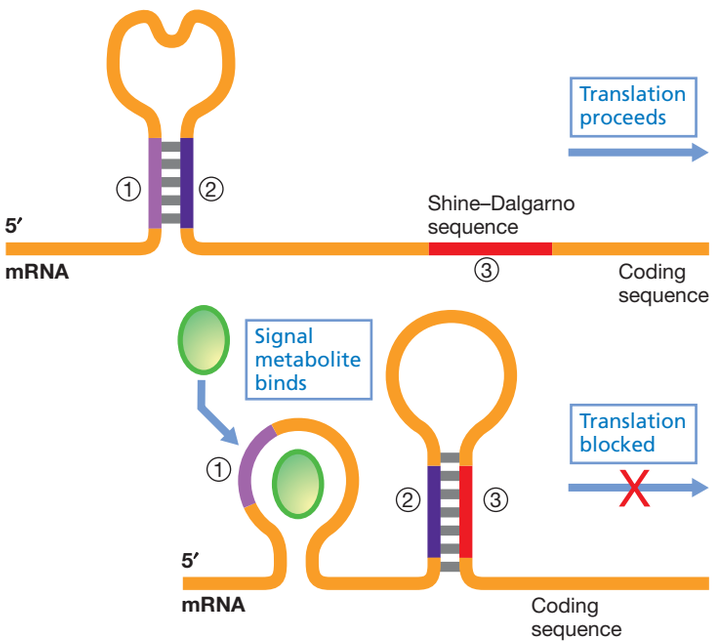

Figure 8.26 **Regulation by a riboswitch.** Binding of a specific metabolite alters the secondary structure of the riboswitch domain, which is located in the 5′ untranslated region of the mRNA, preventing translation. The Shine–Dalgarno site is where the ribosome binds the RNA.

Mechanism of Riboswitches

Earlier in this chapter we discussed the regulation of gene expression by negative control of transcription (Section 8.3). The presence of a specific metabolite often shuts down the transcription of genes encoding enzymes for the corresponding biosynthetic pathway. In our example of the arginine biosynthetic pathway this is performed by a protein repressor. In a riboswitch, there is no regulatory protein. Instead, the metabolite binds directly to the riboswitch domain at the 5′ end of the mRNA. Riboswitches usually exert their control after the mRNA has already been synthesized. Therefore, most riboswitches control *translation* of the mRNA, rather than its *transcription*.

The metabolite that is bound by the riboswitch is typically the product of a biosynthetic pathway whose constituent enzymes are encoded by the mRNAs that carry the corresponding riboswitches. For example, the thiamine riboswitch that binds thiamine pyrophosphate is upstream of the coding sequences for enzymes that participate in the thiamine biosynthetic pathway. When the pool of thiamine pyrophosphate is sufficient in the cell, this metabolite binds to its specific riboswitch mRNA. The new secondary structure of the riboswitch blocks the ribosome-binding site on the mRNA (Section 6.19) and prevents the mRNA from binding to the ribosome; this prevents translation (Figure 8.26). If the concentration of thiamine pyrophosphate drops sufficiently low, this molecule can dissociate from its riboswitch mRNA. This unfolds the mRNA and exposes the ribosome-binding site, allowing the mRNA to bind to the ribosome and be translated.

The thiamine analog pyrithiamine blocks the synthesis of thiamine and, hence, inhibits bacterial growth. Until the discovery of riboswitches, the site of action of pyrithiamine remained mysterious. It now appears that pyrithiamine is converted by cells to pyrithiamine pyrophosphate, which then binds to the thiamine riboswitch. Thus the biosynthetic pathway is shut off even when no thiamine is available. Bacterial mutants selected for resistance to pyrithiamine have alterations in the sequence of the riboswitch that result in failure to bind both pyrithiamine pyrophosphate and thiamine pyrophosphate.

In *Bacillus subtilis*, where about 2% of the genes are under riboswitch control, the same riboswitch is present on several mRNAs that together encode the proteins for a particular pathway. For example, over a dozen genes in six operons are controlled by the thiamine riboswitch.

Despite being part of the mRNA, some riboswitches nevertheless do control transcription. The mechanism is similar to that seen in attenuation (Section 8.16) where a conformational change in the riboswitch causes premature termination of the synthesis of the mRNA that carries it.

Riboswitches and Evolution

How widespread are riboswitches and how did they evolve? Thus far riboswitches have been found only in some bacteria and a few plants and fungi. Some scientists believe that riboswitches are remnants of the RNA world, a period eons ago before cells, DNA, and protein, when it is hypothesized that catalytic RNAs were the only self-replicating life forms (↻ Section 16.2). In such an environment, riboswitches may have been a primitive mechanism of metabolic control—a simple means by which RNA life forms could have controlled the synthesis of other RNAs. As proteins evolved, riboswitches might have been the first control mechanisms for their synthesis as well. If this is true, the riboswitches that remain today may be the last vestiges of this simple form of control because, as we have seen in this chapter, metabolic regulation is almost exclusively carried out by way of regulatory proteins.

MiniQuiz

- What happens when a riboswitch binds the small metabolite that regulates it?
- What are the major differences between using a repressor protein versus a riboswitch to control gene expression?

8.16 Attenuation

Attenuation is a form of transcriptional control that functions by premature termination of mRNA synthesis. That is, in attenuation, control is exerted *after* the initiation of transcription, but *before* its completion. Consequently, the number of completed transcripts from an operon is reduced, even though the number of initiated transcripts is not.

The basic principle of attenuation is that the first part of the mRNA to be made, called the *leader region*, can fold up into two alternative secondary structures. In this respect, the mechanism of attenuation resembles that of riboswitches. In attenuation, one mRNA secondary structure allows continued synthesis of the mRNA, whereas the other secondary structure causes premature termination. Folding of the mRNA depends either on events at the ribosome or on the activity of regulatory proteins, depending on the organism. The best examples of attenuation are the regulation of genes controlling the biosynthesis of certain amino acids in gram-negative *Bacteria*. The first to be described was in the tryptophan operon in *Escherichia coli*, and we focus on it here. Attenuation control has been documented in several other species of *Bacteria*, and genomic analyses of *Archaea* suggest that the mechanism is present in this domain as well. However, because the processes of transcription and translation are spatially separated in eukaryotes, attenuation control is absent from *Eukarya*.

Attenuation and the Tryptophan Operon

The tryptophan operon contains structural genes for five proteins of the tryptophan biosynthetic pathway plus the usual promoter and regulatory sequences at the beginning of the operon (**Figure 8.27**). Like many operons, the tryptophan operon has more than one type of regulation. The first enzyme in the pathway, anthranilate synthase (a multi-subunit enzyme encoded by *trpD* and *trpE*), is subject to feedback inhibition by tryptophan (↻ Section 4.16). Transcription of the entire tryptophan operon

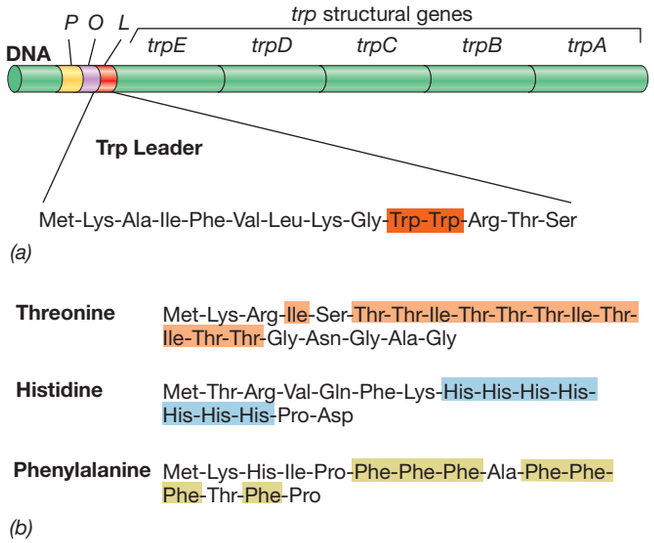

Figure 8.27 Attenuation and leader peptides in *Escherichia coli*. Structure of the tryptophan (*trp*) operon and of the tryptophan leader peptide and other leader peptides in *E. coli*. *(a)* Arrangement of the *trp* operon. Note that the leader (*L*) encodes a short peptide containing two tryptophan residues near its terminus (there is a stop codon following the Ser codon). The promoter is labeled *P*, and the operator is labeled *O*. The genes labeled *trpE* through *trpA* encode the enzymes needed for tryptophan synthesis. *(b)* Amino acid sequences of leader peptides of some other amino acid synthetic operons. Because isoleucine is made from threonine, it is an important constituent of the threonine leader peptide.

is also under negative control (Section 8.3). However, in addition to the promoter and operator regions needed for negative control, there is a sequence in the operon called the *leader sequence* that encodes a short polypeptide, the *leader peptide*. The leader sequence contains tandem tryptophan codons near its terminus and functions as an attenuator (Figure 8.27).

The basis of control of the tryptophan attenuator is as follows. If tryptophan is plentiful in the cell, there will be plenty of charged tryptophan tRNAs and the leader peptide will be synthesized. Synthesis of the leader peptide results in termination of transcription of the remainder of the *trp* operon, which includes the structural genes for the biosynthetic enzymes. On the other hand, if tryptophan is scarce, the tryptophan-rich leader peptide will not be synthesized. If synthesis of the leader peptide is halted by a lack of tryptophan, the rest of the operon is transcribed. www.microbiologyplace.com **Online Tutorial 8.2: Attenuation and the Tryptophan Operon**

Mechanism of Attenuation

How does translation of the leader peptide regulate transcription of the tryptophan genes downstream? Consider that in prokaryotic cells transcription and translation are simultaneous processes; as mRNA is released from the DNA, the ribosome binds to it and translation begins (↻ Section 6.19). That is, while transcription of downstream DNA sequences is still proceeding, translation of already transcribed sequences is under way (**Figure 8.28**).

Transcription is attenuated because a portion of the newly formed mRNA folds into a unique stem–loop that inhibits RNA polymerase activity. The stem–loop structure forms in the mRNA

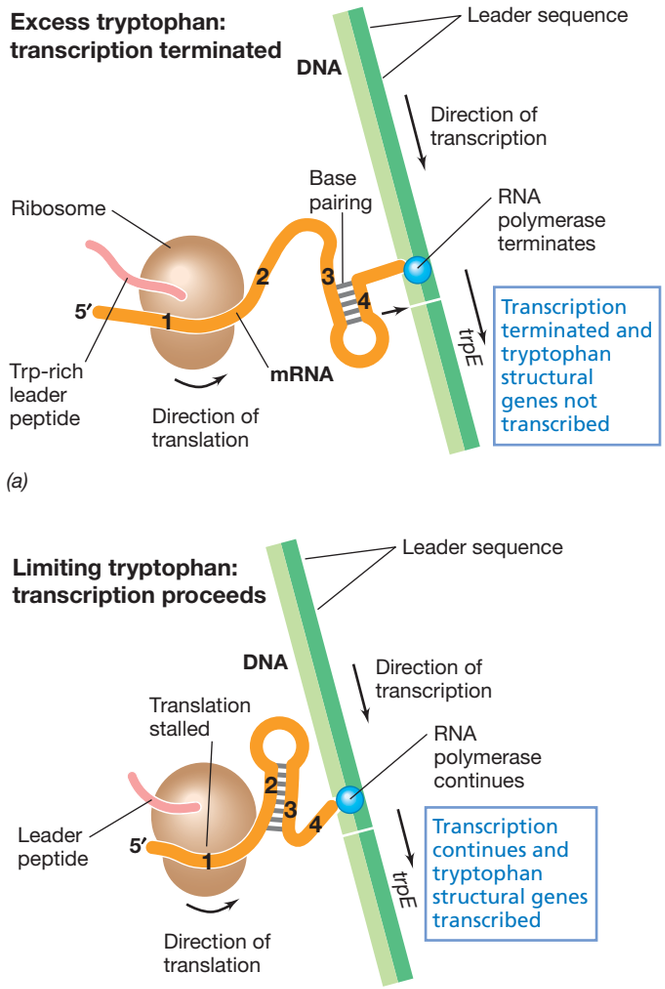

Figure 8.28 **Mechanism of attenuation.** Control of transcription of tryptophan (*trp*) operon structural genes by attenuation in *Escherichia coli*. The leader peptide is encoded by regions 1 and 2 of the mRNA. Two regions of the growing mRNA chain are able to form double-stranded loops, shown as 3:4 and 2:3. *(a)* When there is excess tryptophan, the ribosome translates the complete leader peptide, and so region 2 cannot pair with region 3. Regions 3 and 4 then pair to form a loop that terminates transcription. *(b)* If translation is stalled because of tryptophan starvation, a loop forms by pairing of region 2 with region 3, loop 3:4 does not form, and transcription proceeds past the leader sequence.

because two stretches of nucleotides near each other are complementary and can thus base-pair. If tryptophan is plentiful, the ribosome translates the leader sequence until it comes to the leader stop codon. The remainder of the leader sequence then forms a stem–loop, a transcription pause site, which is followed by a uracil-rich sequence that actually causes termination (Figure 8.28*a*).

If tryptophan is in short supply, transcription of genes encoding tryptophan biosynthetic enzymes is obviously desirable. During transcription of the leader, the ribosome pauses at a tryptophan codon because of a shortage of charged tryptophan tRNAs. The presence of the stalled ribosome at this position allows a stem–loop to form (sites 2 and 3 in Figure 8.28*b*) that differs from the terminator stem–loop. This alternative stem–loop is not a transcription termination signal. Instead, it prevents the terminator stem–loop (sites 3 and 4 in Figure 8.28*a*) from forming. This allows RNA polymerase to move past the termination site and begin transcription of tryptophan structural genes. Thus, in attenuation control, the rate of transcription is influenced by the rate of translation.

Attenuation also occurs in *Escherichia coli* in the biosynthetic pathways for histidine, threonine–isoleucine, phenylalanine, and several other amino acids and essential metabolites. As shown in Figure 8.27*b*, the leader peptide for each of these amino acid biosynthetic operons is rich in that particular amino acid. The *his* operon is dramatic in this regard because its leader peptide contains seven histidines in a row near the end of the peptide (Figure 8.27*b*). This long stretch of histidines gives attenuation a major effect in regulation, which may compensate for the fact that unlike the *trp* operon, the *his* operon in *E. coli* is not also under negative control by a protein repressor.

Translation-Independent Attenuation Mechanisms

Gram-positive *Bacteria*, such as *Bacillus*, also use attenuation of transcription to regulate certain amino acid biosynthetic operons. And, as in gram-negative *Bacteria*, the mechanism relies on alternative mRNA secondary structures, which in one configuration lead to termination. However, the mechanism is independent of translation and requires an RNA-binding protein.

In the *Bacillus subtilis* tryptophan operon, the binding protein is called the *trp* attenuation protein. In the presence of sufficient amounts of the amino acid tryptophan, this regulatory protein binds to the leader sequence in the mRNA and causes transcription termination. By contrast, if tryptophan is limiting, the protein does not bind to the leader sequence. This allows the favorable secondary structure to form and transcription proceeds.

Attenuation also occurs with genes unrelated to amino acid biosynthesis. These mechanisms obviously do not rely on amino acid levels. Some of the operons for pyrimidine biosynthesis (the *pyr* operons) in *E. coli* are regulated by attenuation, and the same is true for *Bacillus*. The mechanisms in the two organisms are, however, quite different, although each employs a system to assess the level of pyrimidines in the cell. In *E. coli* the mechanism monitors the rate of transcription, not translation. If pyrimidines are plentiful, RNA polymerase moves along and transcribes the leader DNA at a normal rate; this allows a terminator stem–loop to form in the mRNA. By contrast, if pyrimidines are scarce, the RNA polymerase pauses at pyrimidine-rich sequences, which leads to formation of a nonterminator stem–loop that allows further transcription.

In *Bacillus*, a different mechanism is employed. For *pyr* attenuation, an RNA-binding protein controls the alternative stem–loop structures of the *pyr* mRNA, terminating transcription when pyrimidines are in excess. In this way the cell can maintain levels of pyrimidines, compounds that require significant cell resources to biosynthesize (Section 4.14), at levels needed to balance biosynthetic needs.

MiniQuiz

- Explain how the formation of one stem–loop in the RNA can block the formation of another.
- How does attenuation of the tryptophan operon differ between *Escherichia coli* and *Bacillus subtilis*?

Big Ideas

8.1
Most genes encode proteins and most proteins are enzymes. Expression of an enzyme-encoding gene is regulated by controlling the activity of the enzyme or controlling the amount of enzyme produced.

8.2
Certain proteins bind to DNA when specific domains of the proteins bind to specific regions of the DNA molecule. In most cases the interactions are sequence-specific. Proteins that bind to DNA are often regulatory proteins that affect gene expression.

8.3
The amount of a specific enzyme in the cell can be controlled by increasing (inducing) or decreasing (repressing) the amount of messenger RNA that encodes the enzyme. This transcriptional regulation is carried out by allosteric regulatory proteins that bind to DNA. In negative control of transcription, the regulatory protein is called a repressor and it functions by inhibiting mRNA synthesis.

8.4
Positive regulators of transcription are called activator proteins. They bind to activator-binding sites on the DNA and stimulate transcription. Inducers modify the activity of activating proteins. In positive control of enzyme induction, the inducer promotes the binding of the activator protein and thus stimulates transcription.

8.5
Global control systems regulate the expression of many genes simultaneously. Catabolite repression is a global control system that helps cells make the most efficient use of available carbon sources. The *lac* operon is under the control of catabolite repression as well as its own specific negative regulatory system.

8.6
Archaea resemble *Bacteria* in using DNA-binding activator and repressor proteins to regulate gene expression at the level of transcription.

8.7
Signal transduction systems transmit environmental signals to the cell. In prokaryotes, signal transduction is typically carried out by a two-component regulatory system that includes a membrane-integrated sensor kinase and a cytoplasmic response regulator. The activity of the response regulator depends on its state of phosphorylation.

8.8
Chemotactic behavior responds in a complex manner to attractants and repellents. The regulation of chemotaxis affects the activity of proteins rather than their synthesis. Adaptation by methylation allows the system to reset itself to the continued presence of a signal.

8.9
Quorum sensing allows cells to monitor their environment for cells of their own kind. Quorum sensing depends on the sharing of specific small molecules known as autoinducers. Once a sufficient concentration of the autoinducer is present, specific gene expression is triggered.

8.10
The stringent response is a global control mechanism triggered by amino acid starvation. The alarmones ppGpp and pppGpp are produced by RelA, a protein that monitors ribosome activity. Within the cell the stringent response achieves balance between protein production and amino acid requirements.

8.11
Cells can control sets of genes by employing alternative sigma factors. These recognize only certain promoters and thus allow transcription of a select category of genes that is appropriate under certain environmental conditions. Cells respond to both heat and cold by expressing sets of genes whose products help the cell overcome stress.

8.12
Sporulation in *Bacillus* during adverse conditions is triggered via a complex phosphotransfer relay system that monitors multiple aspects of the environment. The sporulation factor Spo0A then sets in motion a cascade of regulatory responses under the control of several alternative sigma factors.

8.13
Differentiation in *Caulobacter* consists of the alternation between motile cells and those that are attached to surfaces. Three major regulatory proteins—CtrA, GcrA, and DnaA—act in succession to control the three phases of the cell cycle. Each in turn controls many other genes needed at specific times in the cell cycle.

8.14
Cells can control genes in several ways by employing regulatory RNA molecules. One way is to take advantage of base pairing and use antisense RNA to form a double-stranded RNA that cannot be translated.

8.15
Riboswitches are RNA domains at the 5′ ends of mRNA that recognize small molecules and respond by changing their three-dimensional structure. This, in turn, affects the translation of the mRNA or, sometimes, premature termination of transcription. Riboswitches are mostly used to control biosynthetic pathways for amino acids, purines, and a few other metabolites.

8.16
Attenuation is a mechanism whereby transcription is controlled after initiation of mRNA synthesis. Attenuation mechanisms depend upon alternative stem–loop structures in the mRNA.

Review of Key Terms

Activator protein a regulatory protein that binds to specific sites on DNA and stimulates transcription; involved in positive control

Attenuation a mechanism for controlling gene expression that terminates transcription after initiation but before a full-length messenger RNA is produced

Autoinducer small signal molecule that takes part in quorum sensing

Catabolite repression the suppression of alternative catabolic pathways by a preferred source of carbon and energy

Cyclic AMP a regulatory nucleotide that participates in catabolite repression

Gene expression transcription of a gene followed by translation of the resulting mRNA into protein

Heat shock proteins proteins induced by high temperature (or certain other stresses) that protect against high temperature, especially by refolding partially denatured proteins or by degrading them

Heat shock response response to high temperature that includes the synthesis of heat shock proteins together with other changes in gene expression

Induction production of an enzyme in response to a signal (often the presence of the substrate for the enzyme)

Negative control a mechanism for regulating gene expression in which a repressor protein prevents transcription of genes

Noncoding RNA RNA that is not translated into protein; examples include ribosomal RNA, transfer RNA, and small regulatory RNAs

Operon one or more genes transcribed into a single RNA and under the control of a single regulatory site

Positive control a mechanism for regulating gene expression in which an activator protein functions to promote transcription of genes

Quorum sensing a regulatory system that monitors the population level and controls gene expression based on cell density

Regulatory nucleotide a nucleotide that functions as a signal rather than being incorporated into RNA or DNA

Regulon a series of operons controlled as a unit

Repression prevention of the synthesis of an enzyme in response to a signal

Repressor protein a regulatory protein that binds to specific sites on DNA and blocks transcription; involved in negative control

Response regulator protein one of the members of a two-component regulatory system; a protein that is phosphorylated by a sensor kinase and then acts as a regulator, often by binding to DNA

Riboswitch an RNA domain, usually in a messenger RNA molecule, that can bind a specific small molecule and alter its secondary structure; this, in turn, controls translation of the mRNA

Sensor kinase protein one of the members of a two-component regulatory system; a protein that phosphorylates itself in response to an external signal and then transfers the phosphoryl group to a response regulator protein

Signal transduction *see* two-component regulatory system

Stringent response a global regulatory control that is activated by amino acid starvation or energy deficiency

Two-component regulatory system a regulatory system consisting of two proteins: a sensor kinase and a response regulator

Review Questions

1. Describe why a protein that binds to a specific sequence of double-stranded DNA is unlikely to bind to the same sequence if the DNA is single-stranded (Section 8.2).
2. Most biosynthetic operons need only be under negative control for effective regulation, whereas most catabolic operons need to be under both negative and positive control. Why (Sections 8.2 to 8.4)?
3. What is the difference between an operon and a regulon (Section 8.4)?
4. Describe the mechanism by which cAMP receptor protein (CRP), the regulatory protein for catabolite repression, functions. Use the lactose operon as an example (Section 8.4).
5. What are the two components that give their name to a signal transduction system in prokaryotes? What is the function of each of the components (Section 8.7)?
6. Adaptation allows the mechanism controlling flagellar rotation to be reset. How is this achieved (Section 8.8)?
7. How can quorum sensing be considered a regulatory mechanism for conserving cell resources (Section 8.9)?
8. What events trigger the stringent response? Why are the events in the stringent response a logical consequence of the trigger of the response (Section 8.10)?
9. Describe the proteins produced when cells of *Escherichia coli* experience a heat shock. Of what value are they to the cell (Section 8.11)?
10. Explain how alternative sigma factors control sporulation in *Bacillus* (Section 8.12).
11. What role does the DnaA protein play in differentiation in *Caulobacter* (Section 8.13)?
12. How does regulation by antisense RNA differ from that of riboswitches (Sections 8.14 and 8.15)?
13. Describe how transcriptional attenuation works. What is actually being “attenuated” (Section 8.16)?