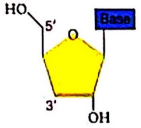
Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

Group Discussion Problems

1. The 5′ cap and poly (A) tail that are attached to eukaryotic mRNAs appear to help the message last longer by protecting it from degradation by ribonucleases. But why is an enzyme such as a ribonuclease in the cell in the first place? What function would an enzyme that destroys messages serve?
2. Look back at Figure 16.1, which shows the formation of a phosphodiester bond during RNA polymerization. Then study the structure of the nucleotide shown below, called cordycepin. If cordycepin triphosphate, which has three phosphate groups bonded to the 5′ hydroxyl group in the figure, is added to a cell-free transcription reaction, the nucleotide is added onto the growing RNA chain. This observation confirms that synthesis occurs by the addition of monomers in the form of triphosphates to the 3′ end of the growing chain. Briefly explain why. Be sure to describe the expected result if synthesis occurred at the 5′ end of the polymer.



3. Carl Woese has determined the nucleotide sequence of rRNA molecules from a diverse array of organisms and compared them with each other. According to his data, certain portions of the rRNAs in the large subunit are very similar in all organisms. To make sense of this result, Woese suggests that the conserved sequences have an important functional role. His logic is that these conserved sequences are so important to cell function that any changes in the sequence cause death. Woese also claims that the existence of the conserved sequences supports other data indicating that peptide bond formation is catalyzed by the large rRNA, not by ribosomal proteins. Explain the thinking behind his claim.
4. Recent work has revealed why α-amanitin inhibits transcription by RNA pol II. Structural models show that α-amanitin binds to a site inside the enzyme, but not to the active site itself. Based on the model of bacterial RNA polymerase in Figure 16.2, predict where α-amanitin binds and why it inhibits transcription.

Answers to Multiple-Choice Questions 1. c; 2. b; 3. d; 4. c; 5. c; 6. d

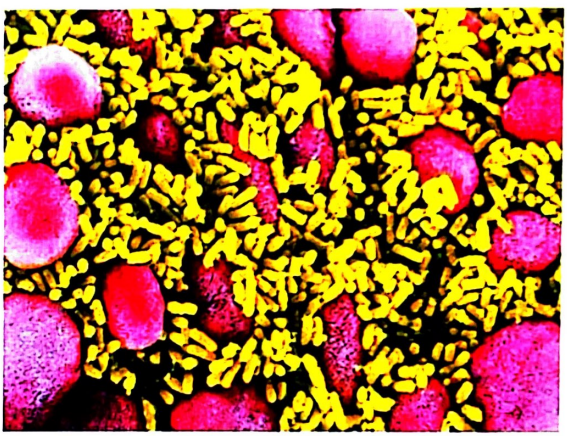
www.prenhall.com/freeman is your resource for the following: Web Tutorials; Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Caption Questions and Exercises; and Additional Readings and Research.

Motif Exhibit 1043, Page 26 of 65

17 Control of Gene Expression in Bacteria

KEY CONCEPTS

- Changes in gene expression allow bacterial cells to respond to environmental changes.
- Gene expression can be controlled at three levels: transcription, translation, or post-translation (protein activation).
- Transcriptional control can be positive or negative. In the case of the genes for lactose metabolism in *E. coli*, positive control occurs when a regulatory protein binds to DNA and increases the transcription rate. Negative control occurs because another regulatory protein must be released before transcription can occur.
- Many regulatory proteins bind to specific sites in DNA. Because different regulatory proteins have different amino acid sequences, they bind to different DNA sequences.



The red mounds in this micrograph are human intestinal cells; the yellow structures are bacteria. In the intestine, the nutrients available to bacteria constantly change. This chapter explores how changes in gene expression help bacteria respond to environmental changes.

Imagine waiting anxiously to hear the opening lines of a wonderfully melodic symphony, played by a renowned orchestra. The crowd applauds as the celebrated conductor comes onstage, then hushes as he takes the podium. He cocks the baton; the musicians raise their instruments. As the baton comes down, every instrument begins blaring a different tune at full volume. A tuba plays "Dixie"; a violinist renders "In-A-Gadda-Da-Vida"; and a cellist begins Mexico's national anthem. A snare drum lays down beats for "Baby Got Back" while the bass drum simulates the cannons in the "1812 Overture." Instead of music, there is pandemonium. The conductor staggers off stage, clutching his heart.

A cacophony like this would result if a bacterial cell "played" all its genes at full volume all the time. The *Escherichia coli* cells living in your gut right now have over 4300 genes. If all of those genes were expressed at the fastest possible rate at all times, the *E. coli* cells would stagger off the stage, too. But this does not happen. Cells are extremely selective about which genes are expressed, in what amounts, and when.

This chapter explores how bacterial cells control the activity of their genes. **Gene expression** is said to occur when a protein or other gene product is synthesized and is active in the cell.

Understanding how cells regulate gene expression is a fundamental issue in biological science, for two reasons. First, bacteria are the most abundant organisms on Earth, and they occupy virtually every habitat known. In many cases, bacterial cells are able to grow and reproduce successfully because they respond rapidly to dramatic changes in temperature, pH, light, competitors, and nutrients. Changes in gene expression give bacteria the ability to cope with these types of environmental change. Second, questions about bacterial gene expression have enormous practical significance. For example, biologists are trying

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

to understand which genes are expressed when disease-causing bacteria colonize a person and start an infection in the hope that drugs can be developed that alter the expression of key genes. Bacterial cells are also used commercially to manufacture important products such as human insulin and growth hormone. Efficient use of bacteria depends on a solid understanding of how transcription and translation are regulated.

Understanding gene expression is fundamental to our understanding of life. This chapter's goal is to introduce key concepts in the regulation of gene expression. Let's begin by reviewing some of the environmental challenges that bacterial cells face and then explore how these organisms meet them.

17.1 Gene Regulation and Information Flow

Bacteria are found in virtually every habitat on Earth, from boiling hot springs to alpine snowfields and from the open ocean to crevices miles underground. The millions of bacterial species existing today have evolved a bewildering variety of ways to solve a central problem of living: obtaining the materials and energy required for growth and reproduction. Although some bacteria specialize by using just one type of food, the vast majority of species are able to switch among several different sources of carbon and energy, according to which nutrients are available in the environment at the time. The fundamental question addressed in this chapter is *how* this switching occurs. Each type of nutrient requires a different membrane transport protein to bring the nutrient molecule into the cell and a different suite of enzymes to process it. How does a bacterial cell turn some genes on and others off so that it can take advantage of alternative food sources? More generally, how do bacteria regulate gene expression so that a cell makes only the products that it needs?

As a case study, we explore the strategies of *Escherichia coli*. These cells can use a wide array of carbohydrates to supply the carbon and energy they need. The carbohydrates are in your food. As your diet changes from day to day, the availability of different sugars in your intestines varies. Precise control of gene expression gives *E. coli* the ability to respond to these changes in its environment and use the different sugars.

To appreciate why precise control over gene expression is so important to these cells, it is critical to realize that bacteria can be packed an inch thick along your intestinal walls. The organisms represent many different species that are competing for space and nutrients. For a cell to survive and reproduce in this environment, it must use resources efficiently—particularly resources that provide energy and carbon compounds. For example, recall from Chapter 4 that carbohydrates are polymers made up of simple sugars and from Chapter 9 that carbohydrates provide cells with energy and chemical building blocks. *Escherichia coli* is capable of metabolizing a wide variety of carbohydrates. But it would be energetically wasteful for *E. coli* to produce all of the enzymes required to process all of the various carbohydrates all of the time. Instead, it is logical to predict that the enzymes being produced match the sugars that are available at a given time. Efficient use of resources, via tight control over gene expression, is critical if cells are going to be able to compete successfully for space and nutrients.

On the basis of this reasoning, transcription and translation of individual genes in bacteria should be triggered by specific signals from the environment, such as the presence of specific sugars. Did you drink milk at your last meal, or eat french fries and a candy bar? Each type of food contains different sugars. Each sugar should induce a different response by the *E. coli* cells in your intestine. Just as a conductor needs to regulate the orchestra's musicians, cells need to regulate which proteins they produce at any given time.

Mechanisms of Regulation—an Overview

Gene expression can be controlled at any step between the synthesis of RNA and the activation of the final gene product. Three steps occur as information flows from DNA to proteins, represented by arrows in the following expression:

$$\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein} \rightarrow \text{activated protein}$$

The arrow from DNA to RNA represents transcription—the making of messenger RNA (mRNA). The arrow from RNA to protein represents translation, in which ribosomes read the information in mRNA and use that information to synthesize a protein. The arrow from protein to activated protein represents post-translational modifications—including folding, addition of carbohydrate or lipid groups, or perhaps phosphorylation.

How can a bacterial cell conserve ATP and amino acids by avoiding the production of proteins that are not needed at a particular time? A look at the flow of information from DNA to protein suggests three possible mechanisms:

1. The cell could avoid making the mRNAs for particular enzymes. If there is no mRNA, then ribosomes cannot make the gene product. For example, various regulatory proteins affect the ability of RNA polymerase to bind to a promoter and initiate transcription. Genes that are controlled in this way are said to undergo **transcriptional control**:

$$\text{DNA} \not\rightarrow \text{mRNA} \rightarrow \text{protein} \rightarrow \text{activated protein}$$

2. If the mRNA for an enzyme has been transcribed, the cell might have a way to prevent the mRNA from being translated into protein. Mechanisms that alter the length of time an mRNA survives before it is degraded by ribonucleases, that affect translation initiation, or that affect elongation factors and other proteins during the translation process are forms of **translational control**:

$$\text{DNA} \rightarrow \text{mRNA} \not\rightarrow \text{protein} \rightarrow \text{activated protein}$$

Motif Exhibit 1043, Page 28 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

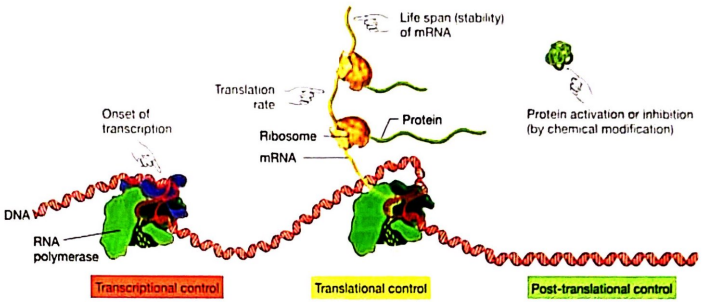


FIGURE 17.1 Gene Expression Can Be Regulated at Three Levels
Although this chapter focuses on how regulatory molecules affect the ability of RNA polymerase to initiate transcription, translational and post-translational controls also occur in bacteria.

3. Chapter 16 pointed out that some proteins are manufactured in an inactive form and have to be activated by chemical modification, such as the addition of a phosphate group. This type of regulation is **post-translational control**:

$$\text{DNA} \rightarrow \text{mRNA} \rightarrow \text{protein} \rightarrow\!\!\!\!\times\!\!\!\!\leftarrow \text{activated protein}$$

Which of these three forms of control occur in bacteria? The short answer to this question is "all of the above." As **Figure 17.1** shows, many factors affect how much active protein is produced from a particular gene. Transcriptional control is particularly important due to its efficiency—it saves the most energy for the cell, because it stops the process at the earliest possible point. Translational control is advantageous because it allows a cell to change quickly which proteins are produced. Post-translational control is significant as well, and it provides the most rapid response of all three mechanisms. Among these mechanisms of gene regulation, there is a clear trade-off between the speed of response and the conservation of ATP, amino acids, and other resources. Transcriptional control is slow but efficient in resource use. Post-translational control is fast but energetically expensive.

Although this chapter focuses almost exclusively on mechanisms of transcriptional control in bacteria, it is important to keep in mind that both translational and post-translational controls occur in these organisms. It's also important to recognize that some genes—such as those that code for the enzymes required for glycolysis—are transcribed all the time, or **constitutively**. But the expression of other genes is regulated, meaning that they may be induced or repressed. Finally, it is critical to realize that gene expression is not an all-or-none proposition. Genes are not just "on" or "off"— instead, the level of expression is highly variable. What factors determine how much a particular gene is expressed at a particular time?

Metabolizing Lactose—A Model System

As Chapters 13 through 16 have shown, many of the fundamental advances in genetics have been achieved through the analysis of a model system. Studying the inheritance of seed shape in garden peas revealed the fundamental patterns of gene transmission. Exploring transcription in viruses and *E. coli* led to the discovery of RNA polymerase, transcription factors (such as the sigma proteins), and promoters. In studies of gene regulation, the key model system has been the metabolism of the sugar lactose in *E. coli*.

Jacques Monod, François Jacob, and many colleagues introduced lactose metabolism in *E. coli* as a model system during the 1950s and 1960s. Although they worked with a single species of bacterium, their results had a profound effect on thinking about gene regulation in all organisms. Some details turned out to be specific to the *E. coli* genes responsible for lactose metabolism, but many of Monod and Jacob's results are universal.

Escherichia coli can use a wide variety of sugars for ATP production, via glycolysis and the electron transport chain or via fermentation. These sugars also serve as a raw material in the synthesis of amino acids, vitamins, and other complex compounds. Glucose is *E. coli*'s preferred carbon source, however, meaning that it is the source of energy and carbon atoms that the organism uses most efficiently. This observation is logical, because glycolysis begins with glucose and is the main pathway for the production of ATP. Lactose, the sugar found in milk, is also used by *E. coli*, but only when glucose supplies are depleted. Recall from Chapter 5 that lactose is a disaccharide made up of one molecule of glucose and one molecule of galactose.

To use lactose, *E. coli* must first transport the sugar into the cell. Once lactose is inside the cell, *E. coli* can break the sugar down into glucose and galactose by means of the enzyme β-galactosidase. The glucose released by this reaction is used

Motif Exhibit 1043, Page 29 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

directly via glycolysis; other enzymes convert the galactose to an intermediate in the glycolytic pathway.

In the early 1950s, biologists discovered that *E. coli* produces high levels of β-galactosidase only when lactose is present in the environment. If lactose is absent, little β-galactosidase is synthesized. Because lactose appears to induce the production of the enzyme that breaks itself down, researchers proposed that lactose itself regulates the gene for β-galactosidase. More formally, the hypothesis was that lactose acts as an inducer. An **inducer** is a molecule that stimulates the expression of a specific gene.

In the late 1950s Jacques Monod investigated how the presence of glucose affects the regulation of the β-galactosidase gene. Would *E. coli* produce high levels of β-galactosidase when both glucose and lactose were present in the surrounding environment? Figure 17.2 summarizes the experimental approach by which Monod found that the answer was no. This meant that β-galactosidase production is induced by the availability of lactose, but only when glucose is unavailable. The key observations can be summarized as follows:

- Cells grown on glucose only: no β-galactosidase production
- Cells grown on glucose + lactose: no β-galactosidase production
- Cells grown on lactose only: β-galactosidase production

Monod teamed up with François Jacob to investigate exactly how lactose and glucose regulate the genes responsible for lactose metabolism. Research on how these genes are regulated opened a window on how genes in all organisms are controlled. Research on this system is still continuing, some 50 years later.

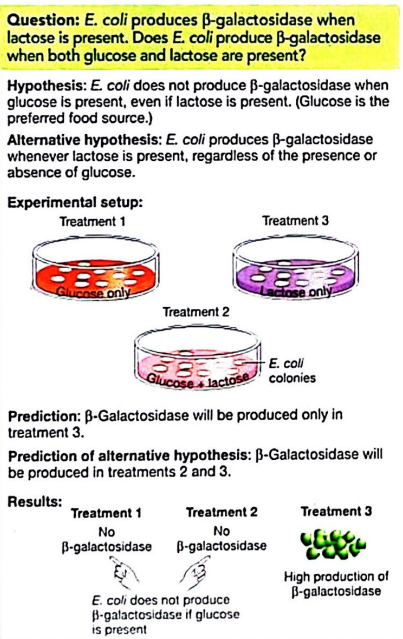


FIGURE 17.2 Glucose Affects the Regulation of the β-Galactosidase Gene

17.2 Identifying the Genes Involved in Lactose Metabolism

To understand how *E. coli* controls production of β-galactosidase and the membrane transport protein that brings lactose into the cell, Monod and Jacob first had to find the genes that code for these proteins. To do this, they employed the same tactic used in the pioneering studies of DNA replication, transcription, and translation reviewed in earlier chapters: They isolated and analyzed mutant individuals. In this case, their goal was to find *E. coli* cells that were *not* capable of metabolizing lactose. Cells that can't use lactose must lack either β-galactosidase or the lactose-transporter protein.

Finding mutants with respect to a particular trait was a two-step process. The researchers' first step was to generate a large number of individuals with mutations at random locations in their genomes. To produce mutant cells, Monod and others exposed *E. coli* populations to X-rays, UV light, or chemicals that damage DNA and increase mutation rates. The second step is to screen the mutants to find individuals with defects in the process or biochemical pathway in question—in this case, defects in lactose metabolism. A technique that allows researchers to identify individuals with a particular type of mutation is called a **screen**.

Monod and colleagues were looking for cells that cannot grow in an environment that contains only lactose as an energy source. Normal cells grow well in this environment. How could the researchers select cells on the basis of *lack* of growth?

Screening Mutants—Replica Plating and Indicator Plates

Replica plating and growth on indicator plates were key techniques in the search for mutants with defects in lactose metabolism. **Replica plating** begins with spreading mutagenized

Motif Exhibit 1043, Page 30 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

bacteria on a plate that is filled with gelatinous agar containing glucose, as illustrated in **Figure 17.3**. This plate is known as the **master plate**. The bacteria are then allowed to grow, so that each cell produces a single colony. (A bacterial colony consists of a large number of identical cells that are descended from a single cell.) Next, a block covered with a piece of sterilized velvet is pressed onto the master plate. Because of the contact, cells from each colony on the master plate are transferred to the velvet. Then the velvet is pressed onto a plate containing a medium that differs from that of the master plate by a single component. For instance, in Figure 17.3, the second medium has only lactose as a source of carbon and energy. Cells from the velvet stick to the plate's surface, producing an exact copy of the colonies on the master plate. This copy is called the **replica plate**. After these cells grow, an investigator can compare the colonies that thrive on the replica plate's medium with those on the master plate. In this case, colonies that grow on the master plate but that are missing on the replica plate represent mutants deficient in lactose metabolism. By picking these particular colonies from the master plate, researchers build a collection of lactose mutants.

Monod also used an alternative strategy based on **indicator plates**, where mutants with metabolic deficiencies are observed directly. Consider what happens when mutagenized bacteria are grown on agar plates containing lactose and are then sprayed with a solution containing the colorless compound ONPG (o-nitrophenyl-β-D-galactoside). ONPG consists of two subunits—a galactose molecule and a molecule called o-nitrophenol—that are joined by the same β-1,4 glycosidic linkage that joins the galactose and glucose subunits of lactose. Consequently, ONPG acts as an alternative substrate for β-galactosidase. When the enzyme cleaves ONPG, galactose and o-nitrophenol are released. This is significant, because o-nitrophenol is intensely yellow. When colonies exposed to ONPG turn yellow, it means that they have β-galactosidase activity. Colonies that stay white are unable to cleave ONPG, meaning they have a defect in the enzyme or its production.

Different Classes of Lactose Metabolism Mutants

The initial mutant screen yielded the three types of mutants summarized in **Table 17.1**. In one class, the mutant cells were unable to cleave ONPG even if lactose was present inside the cells to induce production of the β-galactosidase protein. The investigators concluded that these mutants must lack a functioning version of the β-galactosidase protein—meaning the gene that encodes β-galactosidase is defective. This gene was designated *lacZ*, and the mutant allele *lacZ*$^-$.

In the second class of mutants, the cells failed to accumulate lactose inside the cell. In contrast, the concentration of lactose in wild-type cells is about 100 times that of lactose in the surrounding environment. To interpret this result, Jacob and Monod hypothesized that the mutant cells had defective copies of the membrane protein responsible for transporting lactose into the cell. This protein was identified and named **galactoside permease**; the gene that encodes it was designated *lacY*.

The third and most surprising class of mutants did not regulate expression of β-galactosidase and galactoside permease normally. For example, when these mutant cells were grown on lactose alone and then sprayed with ONPG, they turned yellow just as normal cells do. But if they were grown on a medium that contained glucose but no lactose, they still turned

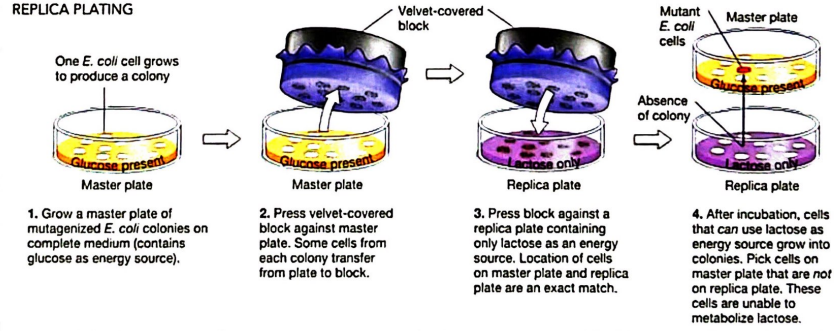


FIGURE 17.3 Replica Plating Is a Technique for Identifying Mutant Cells
Replica plating, used here to isolate mutant *E. coli* cells with a deficiency in lactose metabolism. There are two keys to this technique: (1) The location of colonies on the master plate and replica plate must match exactly, and (2) the media in the two plates must differ by just one type of component. QUESTION How would you alter this protocol to isolate mutant cells with a deficiency in the enzymes required to synthesize tryptophan?

Motif Exhibit 1043, Page 31 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

TABLE 17.1 Three Distinct Types of Mutants in Lactose Metabolism of *E. coli*

Mutant Phenotype	Interpretation	Inferred Genotype
Cells cannot cleave ONPG even if lactose is present as an inducer.	No β-galactosidase; gene for β-galactosidase is defective. Call this gene *lacZ*.	*lacZ*⁻
Cells cannot accumulate lactose inside the cell.	No membrane protein (galactoside permease) required for import of lactose; gene for galactoside permease is defective. Call this gene *lacY*.	*lacY*⁻
ONPG is cleaved even if lactose is absent (no inducer).	Constitutive expression of *lacZ* and *lacY*; gene for regulatory protein that shuts down *lacZ* and *lacY* is defective—it does not need to be induced by lactose. Call this gene *lacI*.	*lacI*⁻

yellow after being sprayed with ONPG. Normal cells, in contrast, do not produce β-galactosidase under these conditions. Normal cells remain white when grown on glucose and then sprayed with ONPG.

Cells that are abnormal because they produce a product at all times are called **constitutive mutants.** Unlike normal cells, their product is always part of the cell's constitution. The gene that mutated to produce constitutive β-galactosidase expression was named *lacI*. The use of "I" was appropriate because these mutants did not need an inducer to express β-galactosidase or galactoside permease. Recall that, in normal cells, the expression of these genes is induced by the presence of lactose. But in cells with a mutant form of *lacI* (*lacI*⁻ mutants), gene expression occurred with or without lactose. This meant that *lacI*⁻ mutants have a defect in gene regulation. In these mutants, the gene remains "on" when it should be turned off.

Based on these observations, it is logical to infer that the normal product of the *lacI* gene prevents the transcription of *lacZ* and *lacY* when lactose is absent. Because lactose acts as an inducer for the production of β-galactosidase, it is reasonable to expect that the *lacI* gene or gene product interacts with lactose in some way. (Later work showed that the inducer is a derivative of lactose called *allolactose*. For the sake of historical accuracy and simplicity, however, this discussion refers to lactose itself as the inducer.)

Several Genes Are Involved in Metabolizing Lactose

Jacob and Monod had succeeded in identifying three genes involved in lactose metabolism: *lacZ*, *lacY*, and *lacI*. They had concluded that *lacZ* and *lacY* code for proteins involved in the metabolism and import of lactose, while *lacI* is responsible for some sort of regulatory function. The *lacI* gene or gene product repressed the expression of *lacZ* and *lacY* in the absence of lactose, but when lactose was present, the opposite occurred—transcription of *lacZ* and *lacY* was induced.

When Jacob and Monod followed up on these experiments by mapping the physical location of the three genes on *E. coli*'s circular chromosome, they found that the genes are close together (Figure 17.4). This was a crucial result, because it suggested that both *lacZ* and *lacY* might be controlled by *lacI*. Could one regulatory gene manage more than one protein-encoding gene? If so, how does *lacI* actually work? And why do lactose and glucose have opposite effects on it?

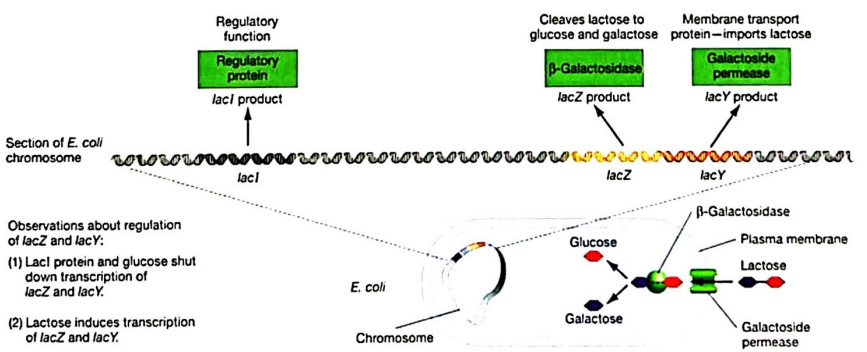


FIGURE 17.4 The *lac* Genes Are in Close Physical Proximity
Early experiments on lactose use in *E. coli* identified the *lacI*, *lacZ*, and *lacY* genes, documented the function of their protein products, and mapped their locations on the *E. coli* chromosome.

Motif Exhibit 1043, Page 32 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

17.3 The Discovery of the Repressor

In principle, there are two ways that transcription can be regulated: via positive control or negative control. **Positive control** is in effect when something must be added for transcription to occur; **negative control** is in effect when something must be taken away for transcription to occur. In starting up a car, positive control occurs when you step on the gas pedal, meaning that you add fuel; negative control occurs when the parking brake is set. It turned out that the *lacZ* and *lacY* genes in *E. coli* are controlled by both a gas pedal and a parking brake—that is, they are under positive control *and* negative control.

The hypothesis that the *lacZ* and *lacY* genes might be under negative control originated with Leo Szilard in the late 1950s. Szilard suggested to Jacques Monod that the *lacI* gene codes for a product that represses transcription of the *lacZ* and *lacY* genes. Recall from Section 17.2 that, in *lacI*$^-$ mutants, the production of β-galactosidase and galactoside permease is constitutive, even if lactose is not available. Szilard proposed that regulation fails because the LacI$^+$ protein—which normally prevents enzyme synthesis—is inactive. Stated another way, the *lacI* gene produces an inhibitor that exerts negative control over the *lacZ* and *lacY* genes. This transcription inhibitor was called a **repressor** and was thought to bind directly to DNA near or on the promoter for the *lacZ* and *lacY* genes (**Figure 17.5a**). To explain how the presence of lactose triggers transcription in normal cells, Szilard and Monod proposed that lactose interacts with the repressor in a way that makes the repressor release from its binding site (**Figure 17.5b**). The idea was that lactose induces transcription by removing negative control. Lactose releases the parking brake. As **Figure 17.5c** shows, it was logical to observe that constitutive transcription occurs in *lacI*$^-$ mutants because a functional repressor is absent—the parking brake is broken.

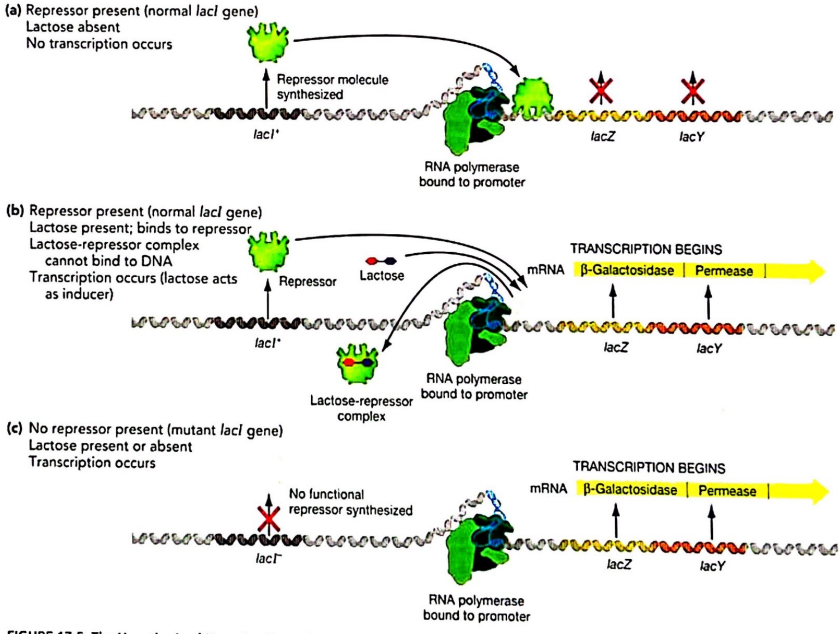


FIGURE 17.5 The Hypothesis of Negative Control
The negative-control hypothesis maintains **(a)** that transcription of genes involved in lactose use is normally blocked by a repressor molecule that binds to DNA on or near the promoter for *lacZ* and *lacY*, **(b)** that lactose induces transcription of *lacZ* and *lacY* by interacting with the repressor, and **(c)** that, when a functional repressor is absent, transcription proceeds.

Motif Exhibit 1043, Page 33 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

How could researchers assess these hypotheses? This was no trivial task. But in a set of brilliant experiments, each of the predictions of the negative-control hypothesis was tested and confirmed. The key studies were based on creating *E. coli* cells that were diploid for the genes involved in lactose metabolism. For example, consider an experiment that begins with cells that are haploid and have the genotype $I^- lacZ^+lacY^-$. These cells lack the hypothesized repressor but have normal copies of the genes for β-galactosidase or galactoside permease. As predicted by the hypothesis of negative control, they produce β-galactosidase constitutively. But what happens when these cells are made diploid by the addition of DNA with the genotype $I^+ lacZ^-lacY^-$? The second set of genes codes for a functioning copy of the repressor, but not of β-galactosidase or galactoside permease. Once the cells are made diploid, β-galactosidase production declines and then stops. This result supports the hypothesis that the repressor codes for a protein that shuts down transcription. As predicted, the repressor is the "parking brake" on transcription. But if an inducer such as lactose is added to these diploid cells, β-galactosidase activity continues. This result supports the hypothesis that lactose removes the repressor—it releases the parking brake. Stated another way, lactose acts as an inducer by removing negative control.

The *lac* Operon

Jacob and Monod summarized the results of their experiments with a comprehensive model of negative control that was published in 1961. In essence, their experimental data had confirmed the hypotheses illustrated in Figure 17.5. One of their key conclusions was that the genes for β-galactosidase and galactoside permease are controlled together. To capture this point, they coined the term **operon** for a set of coordinately regulated bacterial genes that are transcribed together into one mRNA. Logically enough, the group of genes involved in lactose metabolism was termed the **lac operon.**

Three hypotheses are central to the Jacob-Monod model of *lac* operon regulation:

1. The *lacZ* and *lacY* genes are adjacent and are transcribed into one mRNA initiated from the single promoter of the *lac* operon. As a result, the expression of these two genes is coordinated.
2. The repressor is a protein encoded by *lacI* that binds to DNA and prevents transcription of *lacZ* and *lacY*. The binding site for the repressor is a section of DNA in the *lac* operon called the **operator.** Jacob and Monod proposed that *lacI* is expressed constitutively and that its product (the repressor) prevents transcription by blocking RNA polymerase from making contact with the promoter.
3. The inducer (lactose) interacts directly with the repressor by binding to it. As a result, the repressor changes shape in a way that causes it to drop off the DNA strand. Transcription is then able to proceed because RNA polymerase can access the promoter. Recall from Chapter 3 that this form of control over protein function is called *allosteric regulation*. When allosteric regulation occurs, a small molecule binds directly to a protein and causes it to change its shape and activity. When the inducer binds to the repressor, negative control ends.

Testing the *lac* Operon Model

Did Jacob and Monod "get it right"? After their model of negative control was published, Jacob and Monod confirmed the existence of the operator. They did this by finding *E. coli* mutants that expressed *lacZ* constitutively yet had normal forms of the repressor. In these new mutants, the repressor protein was unable to function because the nucleotide sequence of the operator was altered. In 1967 Walter Gilbert and Benno Müller-Hill were able to tag copies of the repressor protein with a radioactive atom. Their experiments showed that the repressor physically binds to the DNA sequences of the operator. This result confirmed that the operator is not a protein or an RNA product, but instead is part of the DNA sequence of the *lac* operon.

The only major modification of the original *lac* operon model is based on the recent recognition that the repressor does not physically block RNA polymerase from contacting the promoter, as Jacob and Monod had proposed. Instead, the repressor prevents the RNA polymerase holoenzyme from opening the DNA helix once that enzyme has bound to the promoter. As a result, the template strand cannot be threaded through the enzyme's active site so that transcription can begin.

In addition, a third gene of the *lac* operon, *lacA*, was discovered after the original model of negative control was published. The *lacA* gene is tightly linked to *lacY* and *lacZ* and is transcribed as part of the same mRNA strand that carries *lacZ* and *lacY*. The gene codes for the enzyme transacetylase, which transfers acetyl (CH_3COO^-) groups to β-galactoside sugars. The protein's function is protective in nature: The acetylation step allows sugars that cannot be metabolized to be exported from the cell instead of accumulating to dangerous levels within the cell.

An mRNA like the one produced by the *lac* operon is known as a polycistronic mRNA. **Polycistronic mRNAs** contain more than one protein-encoding segment. In most cases, bacterial genes that are transcribed as a polycistronic mRNA code for products devoted to a common end, such as the *lacZ* and *lacY* products required for lactose metabolism. Similarly, the polycistronic *trp* (pronounced "trip") operon in *E. coli* consists of five genes, each of which is required for different steps in the synthesis of the amino acid tryptophan. The *trp* operon is also under negative control—but with a twist explained in Box 17.1, page 370. Biologists interpret the polycistronic structure of many bacterial genes as an adaptation

Motif Exhibit 1043, Page 34 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

Almost Parallel Worlds: Similarities and Contrasts in Control of the *trp* and *lac* Operons

The genes in the *lac* operon code for proteins that are responsible for breaking down a compound in the cell, lactose. Logically enough, the expression of these genes depends on the presence of lactose, the compound that is to be broken down. When the inducer, a lactose derivative, binds to the repressor, the repressor comes off the operator and transcription begins.

Now imagine how a cell would regulate an operon that is responsible for synthesizing a molecule instead of breaking it down. In this case, the cell should stimulate gene expression when the molecule is absent but should reduce transcription when the molecule is present. To see how cells accomplish this task, let's explore how *E. coli* regulates the **trp operon**—a set of five cotranscribed genes, each of which is required for the synthesis of the amino acid tryptophan.

If RNA polymerase binds to the promoter for the *trp* operon and begins transcription, all five *trp* operon genes are transcribed and translated together. Once translation is complete, the synthesis of tryptophan commences.

The *trp* operon is under negative control. It has an operator sequence that overlaps the promoter, and a repressor protein that binds to the operator to prevent transcription—as in the *lac* operon.

In the *trp* operon, tryptophan binds to the repressor to regulate transcription, just as allolactose binds to the *lac* operon's repressor to regulate transcription. There is a crucial difference in the two systems, however. If tryptophan binds to the repressor, it means that tryptophan is present in the cell and transcription should not proceed. This is exactly what biologists observe: As Figure 17.6a shows, the *trp* repressor binds operator DNA and exerts negative control only when the repressor is complexed with tryptophan. When tryptophan is absent, the repressor does not bind to the operator and the operon is expressed constitutively (Figure 17.6b). This is exactly the opposite of how the presence of lactose affects the *lac* repressor. And it makes perfect sense for the *trp* operon: Bacterial cells will have higher fitness if the operon is transcribed when tryptophan levels are low but shut down when tryptophan is abundant.

In addition to offering a new twist on negative control, the *trp* operon has provided other fascinating insights into gene regulation. Experiments have revealed that negative control in the *trp* operon is supplemented by a regulatory mechanism called **attenuation**. Attenuation works like a dimmer switch, because it superimposes fine-tuned control on what would otherwise be a simple on-off system. In the case of the *trp* operon, the dimmer is sensitive to tryptophan levels. When tryptophan levels are high, attenuation causes transcription to terminate early in the operon. When tryptophan levels are low, transcription proceeds through the entire operon. Attenuation that is based on amino acid concentration is seen in many other operons devoted to amino acid synthesis.

After considering how the *lac* and *trp* operons are regulated in *E. coli*, it's difficult not to have new respect for "simple" forms of life. Bacteria may be small and unicellular, but they perform fantastically complex, finely controlled chemistry. Even the simplest cells are marvelous machines.

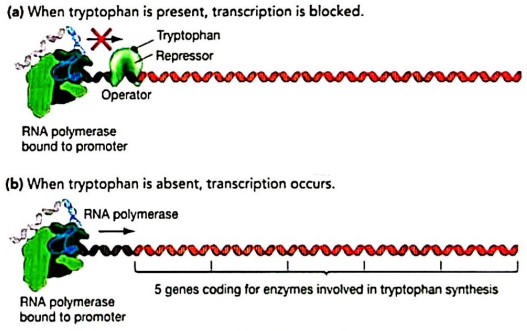


FIGURE 17.6 The *trp* Operon Is under Negative Control
(a) When tryptophan is present, it binds to a repressor protein. The tryptophan-repressor complex binds to the operator and shuts down transcription, which ends the synthesis of tryptophan. **(b)** The *trp* operon contains five coding genes.

that increases the efficiency of gene expression. By binding to one promoter, regulatory proteins can control the expression of several genes that are required for the same process. Polycistronic mRNAs are common in bacteria but rare or nonexistent in most eukaryotes.

The Impact of the *lac* Operon Model

Regulation of the *lac* operon provided an immensely important model system in genetics, partly because aspects of the model turned out to be important in many other species, genes, and operons. For example, follow-up work has

Motif Exhibit 1043, Page 35 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

shown that numerous bacterial genes are under negative control via repressor proteins. Even more important, the *lac* operon model introduced the idea that gene expression is regulated by physical contact between specific regions in regulatory proteins and specific regulatory sites in DNA. Stated another way, the model introduced the idea that gene expression could be modified by proteins that bind directly to regulatory sequences in DNA. Publication of the *lac* operon model was a watershed event in the history of biological science. Its impact on thinking about gene regulation is hard to overestimate.

As elegant as the model presented in Figure 17.5 is, however, something is missing. Where does glucose fit in? This is the final element of the regulatory system for lactose metabolism. As it turns out, it is also the element that is based on positive control of transcription.

✓CHECK YOUR UNDERSTANDING

Negative control occurs when something must be taken away for transcription to occur. The *lac* operon repressor exerts negative control over three protein-coding genes by binding to the operator site in DNA near the promoter. For transcription to occur, an inducer molecule (a derivative of lactose) must bind to the repressor, causing it to release from the operator. You should be able to (1) explain why it is logical to observe that a molecule derived from lactose induces transcription of the *lac* operon; (2) diagram the *lac* operon, showing the relative positions of the operator, the promoter, and the three protein-coding genes; and (3) show what that operon looks like in the presence and absence of lactose.

17.4 Catabolite Repression and Positive Control

When glucose is present in the environment, transcription of the *lac* operon is reduced. This is true even when lactose is available to induce β-galactosidase expression. Given that glucose is the preferred carbon source, this aspect of control over the *lac* operon is logical. Glucose is produced when β-galactosidase cleaves lactose. When glucose is present, it is not necessary for the cell to cleave lactose and produce still more glucose.

Recall from Chapter 9 that biologists use the term *catabolism* to refer to reactions that break larger molecules into simpler subunits. The hydrolysis of lactose into its glucose and galactose subunits is an example of catabolism (Figure 17.7a). In many cases, operons that encode catabolic enzymes are inhibited when the end product of the reaction, the **catabolite**, is abundant. This form of end-product inhibition is **catabolite repression**. In the case of the *lac* operon, glucose is the catabolite. When glucose is abundant in the cell, transcription of the *lac* operon is decreased by catabolite repression (Figure 17.7b).

How does glucose prevent expression of the *lac* operon? An answer to this question began to emerge when researchers discovered a second major control element in the *lac* operon. This regulatory switch consists of a DNA sequence known as the **CAP binding site**, which is located just upstream of the *lac* promoter, and a regulatory protein called **catabolite activator protein (CAP)**, which binds to this DNA sequence.

To understand how CAP works, notice that not all promoters are created equal. Strong promoters allow efficient initiation of transcription by RNA polymerase; weak ones support much less efficient initiation of transcription. The *lac* promoter is weak. But when the CAP regulatory protein is bound to the

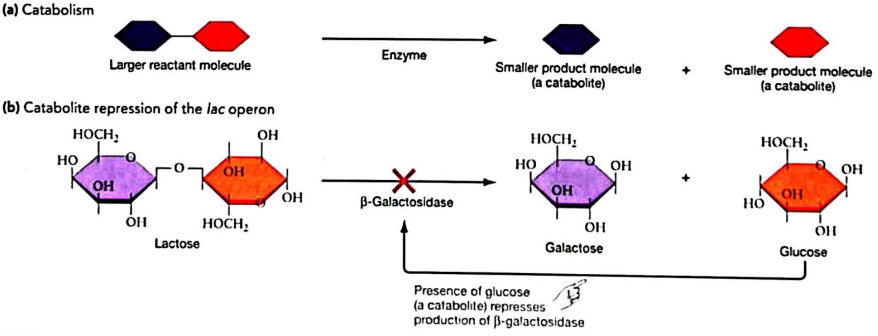


FIGURE 17.7 Catabolite Repression Is a Mechanism of Gene Regulation
(a) A generalized example of catabolism. **(b)** Catabolite repression occurs when one of the small product molecules represses the production of the enzyme responsible for the reaction. In the case of lactose metabolism, the production of β-galactosidase is suppressed when glucose is present.

Motif Exhibit 1043, Page 36 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

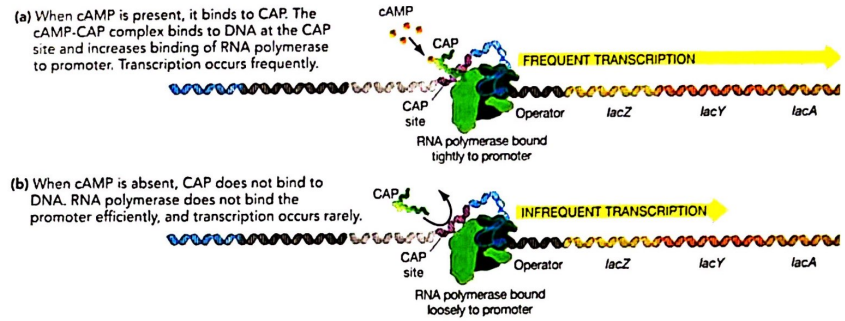


FIGURE 17.8 Positive Control of the lac Operon
(a) When glucose levels in an *E. coli* cell are low, cAMP is produced. cAMP then interacts with CAP to increase transcription of the *lac* operon. **(b)** When glucose is abundant, cAMP is rare in the cell and positive control does not occur.

CAP site just upstream of the *lac* promoter, the protein interacts with RNA polymerase in a way that allows transcription to begin much more frequently. CAP binding greatly strengthens the *lac* promoter. In this way, CAP exerts positive control of the *lac* operon. When CAP is active, transcription increases.

Researchers also discovered that CAP, like the repressor protein, is allosterically regulated. CAP changes shape when the regulatory molecule **cyclic AMP (cAMP)** binds to it. Only when cAMP is attached to it can CAP bind to DNA (Figure 17.8a). The same type of allosteric regulation goes on with the repressor and allolactose during negative control, except that the repressor binds to DNA only in the *absence* of the inducer. During positive control, the CAP-cAMP complex binds to the CAP binding site. As long as a repressor isn't bound to the operator, the complex increases the efficiency of transcription. If no cAMP is bound to CAP, then CAP has a conformation that does not allow binding to the CAP site (Figure 17.8b).

How Does Glucose Influence Formation of the CAP-cAMP Complex?

Where does the elusive molecule glucose fit into this scheme? Glucose's role in positive control of the *lac* operon is indirect. More specifically, it is mediated by the influence that glucose has on the concentration of cAMP. Glucose levels outside the cell and cAMP levels inside the cell are inversely related. When extracellular glucose concentrations are high, intracellular cAMP concentrations are low; when extracellular glucose concentrations are low, intracellular cAMP concentrations are high. This seesaw is driven by the enzyme **adenylyl cyclase**, which produces cAMP from ATP (Figure 17.9a). Adenylyl cyclase's activity is inhibited by extracellular glucose. To see the consequences of this fact, imagine a situation in which glucose is abundant outside the cell (Figure 17.9b). In this state, adenylyl cyclase activity is low. Therefore, cAMP levels inside the cell are low. CAP is not in a CAP-cAMP complex, so it does not have the conformation that allows it to bind to the CAP site and stimulate *lac* operon transcription. Conversely, when the extracellular concentration of glucose is low, there is an increase in the intracellular level of cAMP. In this case, the CAP-cAMP complex forms, binds to the CAP site, and allows RNA polymerase to initiate transcription efficiently. As long as the repressor is not bound to the operator, transcription of the *lac* operon is initiated frequently and lactose can be employed as an alternative energy source.

It is important to note that the CAP-cAMP system influences many genes in addition to the *lac* operon. CAP sites are found adjacent to the promoters for several operons that are required for the metabolism of sugars other than glucose. When glucose levels fall and cAMP concentrations rise, the effect on gene regulation is similar to ringing an alarm bell. In response to the alarm, genes that encode enzymes required for the use of lactose, maltose, glycerol, and other food sources may be turned on—but only if the metabolite of a particular operon is available to the cell. Conversely, cAMP levels are low when glucose supplies are adequate. In this case, the genes for these enzymes are expressed only at low levels even when the appropriate metabolite is present.

Figure 17.10 summarizes how positive control and negative control combine to regulate the *lac* operon. The general message of this figure is that interactions among regulatory elements produce finely tuned control over gene expression. Because positive and negative control elements are superimposed, *E. coli* fully activates the genes for lactose metabolism only when lactose is available and glucose is scarce or absent. In this way, control over gene expression increases the ability of these cells to compete, grow, and reproduce efficiently.

Motif Exhibit 1043, Page 37 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

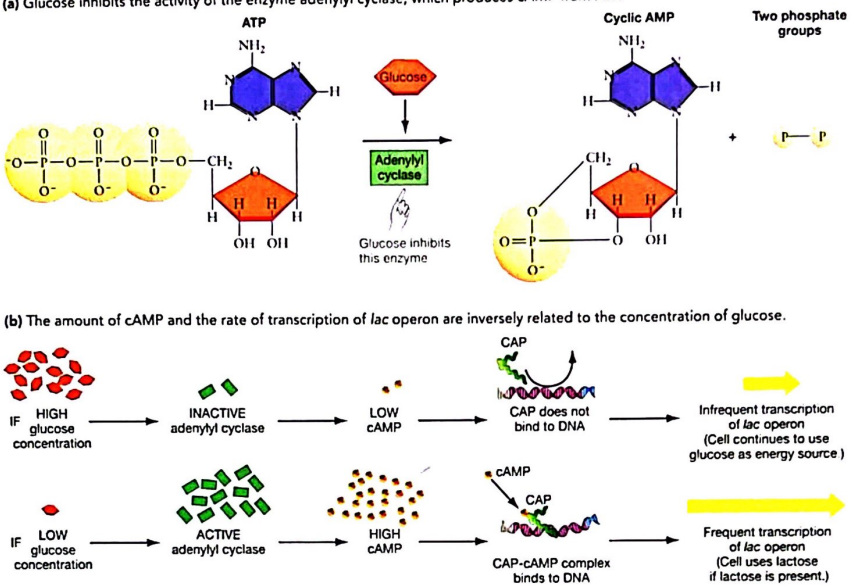


FIGURE 17.9 Cyclic AMP (cAMP) Is Synthesized When Glucose Levels Are Low

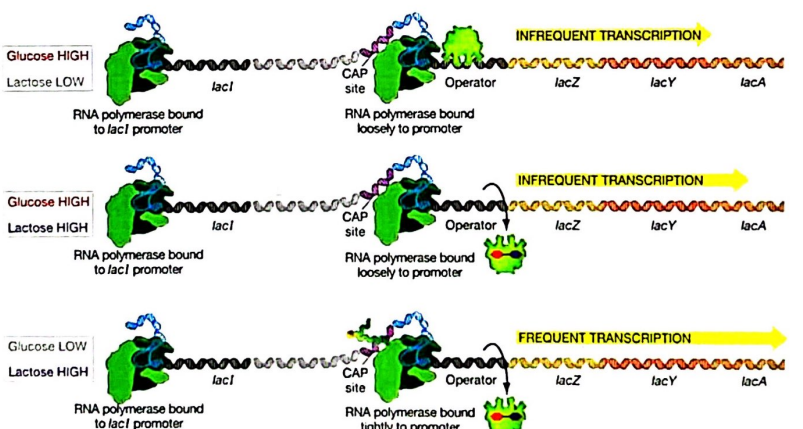


FIGURE 17.10 An Overview of Lactose Use

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

✓CHECK YOUR UNDERSTANDING

Positive control occurs when something must be added for transcription to occur. CAP exerts positive control over the *lac* operon by binding to the CAP site and increasing the transcription rate. CAP can bind only when it is complexed with cAMP—a signaling molecule indicating that glucose levels are low. You should be able to (1) diagram what the operon looks like when cAMP levels are low and high and (2) explain how positive control and negative control interact, based on the presence or absence of glucose and lactose.

17.5 The Operator and the Repressor—an Introduction to DNA-Binding Proteins

The insights that biologists gained from studying the *lac* operon proved to be relevant to many other genes and species. For example, transcription of bacterial and eukaryotic genes is either constitutive or controlled. When regulation occurs, it is negative, positive, or based on a combination of negative and positive factors. Negative control is based on repressor proteins, such as the *lacI* product, whereas positive control depends on transcription activator proteins, such as CAP. In both bacteria and eukaryotes, as in the *lac* operon, it is particularly common for genes to be regulated by a combination of several control systems. Catabolite repression is also extremely common.

Currently, biologists who are interested in gene regulation are trying to understand how DNA-binding proteins such as CAP and the repressor can control transcription. Precisely how do these proteins interact with DNA? How do small regulatory molecules such as cAMP and lactose induce changes in the conformation of these large DNA-binding proteins? Once again, investigators are using the *lac* operon as a model system that offers insights applicable to all cells.

Before taking a detailed look at the repressor and operator, let's step back and think about what is required for a DNA-binding protein to control transcription. First, the protein must bind to a specific DNA sequence associated with a specific gene or operon. Often an investigator's first task is to identify what that sequence is. Second, the protein's structure must allow it to bind DNA, interact with a regulatory molecule such as cAMP or lactose, and affect RNA polymerase. What structural features make each of these functions possible? As it turns out, the *lac* operon's operator DNA sequence and repressor protein structure have features that are common to many regulatory elements in bacteria and eukaryotes.

Finding the Operator

To understand how the *lac* repressor controls transcription, a logical first step is to identify the DNA sequences it targets: the operator. By mapping mutants with defects in the operator, researchers were able to confirm that the target site is just downstream from the promoter. But what is the actual DNA sequence?

Investigators use a technique called DNA footprinting to find and characterize the sequences targeted by DNA-binding proteins. DNA footprinting was used to characterize the first promoter ever discovered (see Chapter 16); the technique is explained in detail in **Box 17.2.**

Footprinting experiments with the *lac* repressor yielded two new discoveries. First, the *lac* operon actually contains three sites where the repressor protein may bind. As **Figure 17.12a** shows, these sites are called O_1, O_2, and O_3. When the repressor protein is bound to DNA, the protein simultaneously binds O_1 and one of the two other operators. This finding required a modification of the original Jacob-Monod model, which proposed the existence of a single operator. Second, all three

BOX 17.2 DNA Footprinting

DNA footprinting is used to characterize DNA sequences that are bound by regulatory proteins. Researchers begin a footprinting study by obtaining DNA from a region of interest—the *lac* operon of *E. coli*, for example. Then they generate many copies of the sequence, using techniques introduced in Chapter 19. The identical DNA fragments are then radioactively labeled at one end (see step 1 of **Figure 17.11**). Once labeling is complete, half of the fragments are mixed with the DNA-binding protein being studied—in this case, the repressor—and half are left without the DNA-binding protein (step 2). Both sets of fragments are then treated with a nuclease—an enzyme that cleaves DNA in random locations. The concentration of nuclease is so low and the exposure time is so short, however, that most fragments are never cut; those that are cut are cut just once. As step 3 shows, the nuclease treatment produces a population of DNA fragments of different lengths. In the sample containing the DNA-binding protein, the binding site is protected from the nuclease.

The DNA fragments produced by the nuclease treatment are separated by gel electrophoresis and visualized via autoradiography (see Chapter 3) or other techniques. As step 4 shows, the sample with the repressor lacks any cuts and thus any fragments in the segment that was protected by that protein. The gap in the series of fragments is the "footprint" of a DNA-binding protein. By aligning the footprinted region with the gene's DNA sequence, researchers can identify the nucleotide sequence of the binding site.

Motif Exhibit 1043, Page 39 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

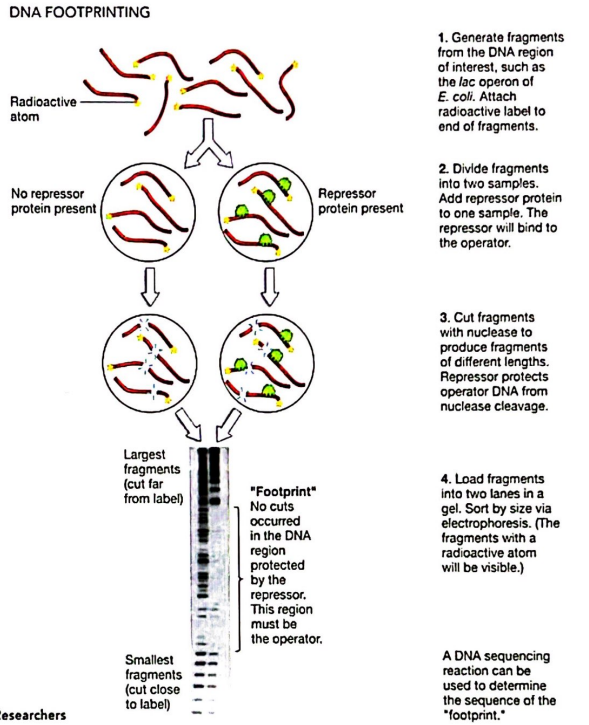


FIGURE 17.11 DNA Footprinting Allows Researchers to Identify Sites Where Proteins Bind

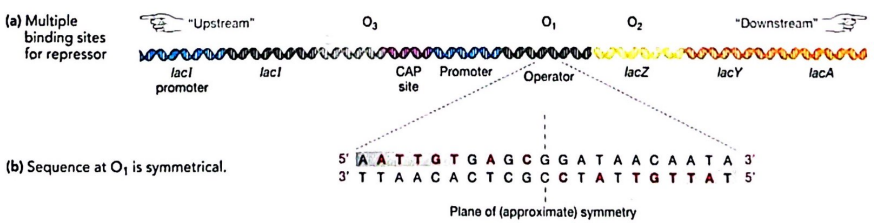


FIGURE 17.12 Operator Sequences
(a) The lac operon contains three operator sequences, called O_1, O_2, and O_3, where the repressor can bind. A repressor simultaneously binds O_1 and either O_2 or O_3. **(b)** Many sequences targeted by DNA-binding proteins are symmetrical. In this example, all but three bases are identical on the two sides of the plane of symmetry.

Motif Exhibit 1043, Page 40 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

operators have a similar DNA sequence that has an unusual characteristic: It contains an axis of symmetry. To understand what an axis of symmetry in DNA is, look at the O_1 sequence shown in Figure 17.12b (page 375). Note that if the sequence were rotated 180°, it would be little changed. This type of symmetry is known as **two-fold rotational symmetry** or **dyad symmetry.** Follow-up work showed that the nucleotide sequences recognized by many DNA-binding proteins have two-fold rotational symmetry.

Now that we've examined the operator, we need to consider the repressor. How does this DNA-binding protein recognize specific sequences along the double helix? In particular, why is two-fold rotational symmetry important?

DNA Binding via the Helix-Turn-Helix Motif

Like other proteins, the *lac* operon's repressor resembles a hand tool in one important respect: Both have distinct regions or parts. Consider a screwdriver, which consists of a handle, a shaft, and a blade. Each of these parts has a distinctive three-dimensional structure and function. Proteins also have regions or parts, called **domains,** that have a distinctive three-dimensional structure and function. Just as screwdrivers of various shapes and sizes all have a handle, shaft, and blade, it is common to observe the same domain in proteins of various shapes and sizes. A domain that is observed in many different proteins is known as a **motif.** Each discrete screwdriver has a handle domain, a shaft domain, and a blade domain. As a group, screwdrivers are characterized by a handle motif, a shaft motif, and a blade motif. The handle and shaft motifs are also found in hammers and socket wrenches, but the blade motif is unique to screwdrivers. Hundreds of different motifs have been characterized in proteins.

When investigators analyzed repressor proteins from different bacteria, they discovered that many of the bacteria contained the domain illustrated in **Figure 17.13a.** This structure is a **helix-turn-helix motif.** As its name indicates, it consists of two α-helices connected by a short stretch of amino acids that forms a turn. Physical models that were constructed during the 1980s suggested that the helix-turn-helix region is the section of the repressor that binds to DNA. This hypothesis was later supported by structural studies based on X-ray crystallography.

To see how a helix-turn-helix domain works, recall from Chapter 4 that the exterior of the DNA double helix has a minor groove and a major groove (**Figure 17.13b**). A protein that has a helix-turn-helix domain can bind to DNA because one of the helices interacts with the sugar-phosphate backbone of a DNA strand while the other helix binds to the base pairs in the major groove (**Figure 17.13c**). DNA does not have to become single stranded for the helix-turn-helix domain of the repressor to bind to the operator. This observation holds for nearly all proteins that recognize specific sequences of double-stranded DNA.

The section of the helix-turn-helix domain that binds inside the major groove is called the **recognition sequence.** The amino acid sequence of this section is crucial, because specific sequences of amino acids bind to specific sequences of DNA. Each type of regulatory protein with a helix-turn-helix motif has a unique sequence of amino acids in its recognition sequence. As a result, each of these regulatory proteins binds to a unique regulatory sequence in DNA. In many cases, precise control over gene expression is based on a precise chemical interactions between

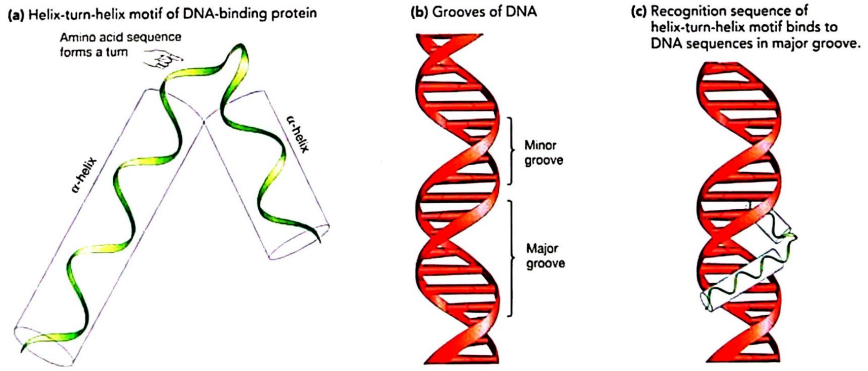


FIGURE 17.13 Helix-Turn-Helix Motifs in DNA-Binding Proteins
QUESTION The [illegible] motif that interacts directly with DNA is called the recognition sequence. Why?

Motif Exhibit 1043, Page 41 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

the amino acid sequence in a recognition sequence and the DNA sequence in the major groove of a regulatory site.

Biologists used X-ray crystallography to document the precise way that the *lac* repressor binds to the DNA of the operator. Consider the structure of the repressor alone, shown in Figure 17.14a. Note that the protein consists of four identical polypeptide chains, and that four helix-turn-helix domains project from the main body of the molecule. When researchers analyzed the structure of the repressor while it was bound to the operator sequence, they confirmed that each of these helix-turn-helix motifs binds operator DNA inside the major groove.

Now recall that the operator essentially is composed of two symmetrical halves. Each half of the operator sequence is bound by one of the repressor's helix-turn-helix motifs. The major grooves in the operator and the recognition helices of the repressor can pair up in this way because the operator is symmetrical and because the repressor consists of identical polypeptides that are flipped 180° with respect to each other. In addition, because the repressor contains four helix-turn-helix domains and two of the four bind per operator, a total of two operators can be bound simultaneously by one repressor. When this occurs, a loop of DNA is created between O_1 and whichever other operator (O_2 or O_3) is bound (Figure 17.14b). This type of DNA looping is observed at many other bacterial promoters and at virtually all eukaryotic promoters.

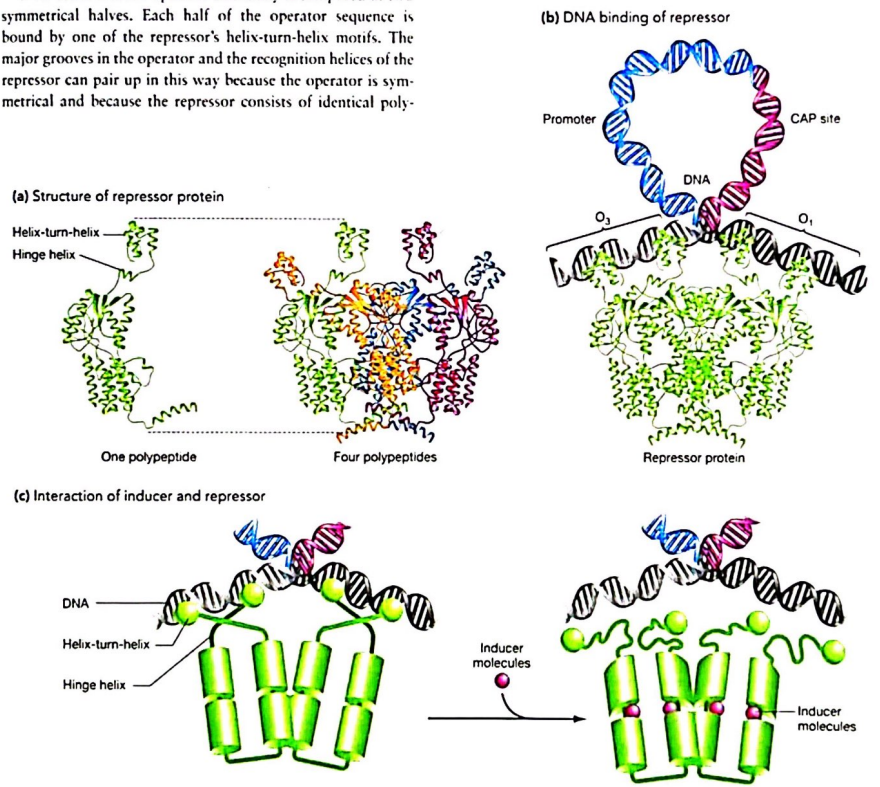


FIGURE 17.14 How Does the Repressor Interact with the Operator and the Inducer?
(a) The repressor is a complex of four polypeptides, each a product of the *lacI* locus. **(b)** The repressor protein has four helix-turn-helix motifs to bind DNA at O_1 and either O_2 or O_3. **(c)** "Hinge helices" in the repressor exist only when the inducer is not present (left). When the inducer binds to the repressor, the hinge helices relax and cause the protein to fall off the operator DNA (right).

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

How Does the Inducer Change the Repressor's Affinity for DNA?

Negative control of the *lac* operon occurs when the repressor binds to the operator. In this state, the helix-turn-helix motifs of the repressor are attached to DNA. Even though RNA polymerase can bind to the promoter nearby, the enzyme cannot open the double helix and begin transcription. This physical interference is typical of how negative regulators work. Positive regulators, in contrast, work by stabilizing the interaction between RNA polymerase and the promoter.

Negative control continues as long as the repressor is bound to DNA. But when lactose is present, the repressor dissociates from the operator. How does the inducer affect the repressor in a way that disrupts DNA binding?

To answer this question, researchers analyzed the three-dimensional structure of the repressor when it was bound to a molecule called IPTG, which acts as an inducer in the *lac* system. Their data showed that IPTG binds at locations in the middle of the repressor and that it produces an important change in the shape of the protein near the helix-turn-helix domains. To see this change in shape, compare the left and right sides of **Figure 17.14c** (page 377). The diagram on the left summarizes the repressor's shape without the inducer. Pay attention to the structures labeled **hinge helices** near the helix-turn-helix motifs. The hinge helices lock the DNA-binding motifs into place when they contact the operator. As the diagram on the right shows, however, the hinge helices uncoil when IPTG or the inducer binds to the repressor. As a result, the two DNA-binding regions can't bind to the operator at the same time, and the protein releases from the site.

The "unhinging" of the hinge helices that occurs in the *lac* repressor turns out to be a unique mechanism for changing the activity of a regulatory protein. Still, several general lessons emerge from the study of this repressor. First, the shape and activity of a protein may change when a small molecule binds to it. Allosteric shifts like that observed in the *lac* repressor occur in thousands of different proteins. Second, key elements in transcriptional control are regulated at the **post-translational level**. For example, the repressor protein is always present because it is transcribed and translated constitutively at low levels. When a rapid change in *lac* operon activity is required, it does not occur via changes in the transcription or translation of the repressor. Rather, the activity of the already translated protein is altered. In virtually all cases, the activity of key regulatory proteins is controlled by post-translational modifications.

Future Directions

Now that biologists understand the physical basis of negative control and induction, what does the future hold? One promising line of research is focused on how the activity of many genes and operons is coordinated. For example, suppose that an *E. coli* cell in your gut is faced with starvation. Does it adjust the activity of hundreds or even thousands of its genes in order to survive? If so, how are changes in gene expression coordinated throughout the genome? Because biologists have a fairly solid understanding of how individual genes and operons are regulated, the current challenge is to explore how multiple genes and operons are controlled in a way that increases the fitness of bacterial cells. Answering these questions may be important to our understanding of how certain bacteria are able to sustain infections that threaten human health and how some bacterial species are able to detoxify pollutants by using them as food. Chapter 19 introduces some techniques that biologists use to study coordinated changes in gene expression.

ESSAY Control of Gene Expression in Bacterial Pathogens

A staggering number of bacteria are living in and on your body right now. To give just one example, the number of bacterial cells in your mouth at this moment far exceeds the number of people who have ever lived, including the world's current population of more than 6.4 billion. Virtually all of these bacteria are **commensal**—meaning they live in association with their host without causing harm.

As you know, however, a few bacterial species can cause disease. How are certain bacteria able to invade and harm us? How are disease-causing bacteria, or **pathogens**, transmitted from person to person? And how can bacterial infections be controlled or prevented? In many cases, the answers depend on an understanding of how the expression of key genes is controlled. To drive this point home, consider *Vibrio cholerae*, the causative agent of cholera (**Figure 17.15**).

Cholera is a disease characterized by watery diarrhea that, if untreated, can lead to dehydration and death. The disease is currently most prevalent in Africa and Asia, where 5 million cases occur each year. Because *V. cholerae* is transmitted from person to person when diarrhea from an infected individual contaminates drinking water, cholera occurs primarily in regions with poor sanitation.

A staggering number of bacteria are living in and on your body right now.

Vibrio cholerae is extraordinarily flexible regarding the habitats it occupies. For example, near the confluence of the Ganges and Brahmaputra Rivers in the Bengal region of India and Bangladesh, *V. cholerae* can be found as free-swimming cells, inside two different eukaryotes, in dense layers coating rocks and other underwater surfaces, and inside humans. In the Bengal, most incidences of human infection occur during an outbreak that coincides with the end of

Motif Exhibit 1043, Page 43 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492



FIGURE 17.15 ***Vibrio cholerae*, the Bacterium That Causes Cholera**
[©Dr. Dennis Kunkel/Visuals Unlimited]

the monsoon rains in October and November each year. This observation suggests that environmental signals trigger *V. cholerae*'s yearly emergence as a human pathogen.

What genes allow *V. cholerae* to act as a human pathogen, and how are these genes regulated? Once it is ingested by a person, *V. cholerae* confronts an environment that is markedly different from its other habitats. In fact, most of the ingested bacteria are killed in the harsh, acidic environment of the stomach. The few that survive enter the small intestine, where they encounter a thick layer of mucus that protects the intestinal cells. Then, in response to a still-unknown signal in the environment, major changes in *V. cholerae* gene expression occur. The new proteins that are produced allow *V. cholerae* cells to move through the mucus layer and make contact with the intestinal cells. Each invading cell then produces a short, hairlike projection called a pilus, which allows the *V. cholerae* cell to attach to an intestinal cell. The proteins required to create this structure result from the coordinately regulated transcription of pilus-specific genes.

Once the bacterial cells have attached to intestinal cells, *V. cholerae* begins to manufacture and secrete a toxin composed of two different polypeptide chains. When intestinal cells take up the toxin, a devastating chain of events occurs. The toxin activates a host protein that stimulates human adenylyl cyclase. Massive amounts of cAMP are produced in response, resulting in a flow of Cl^- ions from the affected intestinal cells into the center (lumen) of the small intestine. Water follows by osmosis. The result is the watery diarrhea symptomatic of cholera.

Although the diarrhea can be fatal to the host, it increases the fitness of *V. cholerae* cells in environments where human sanitation is bad. Because *V. cholerae* cells are shed in diarrhea, they can enter drinking-water supplies if systems for transporting and processing human sewage are in poor repair or nonexistent. In this way, *V. cholerae* cells that induce diarrhea are efficiently transmitted to new hosts. *Vibrio cholerae* cells that have alleles responsible for diarrhea have higher fitness than do *V. cholerae* cells that lack these alleles.

The genes that encode the pilus proteins and cholera toxin are under the control of regulatory genes called the *Tox* (for toxin) loci. As **Figure 17.16** shows, *ToxR* produces a transcription factor that responds to a signal in the intestinal environment. When the ToxR protein is activated, it stimulates transcription of the *ToxT* gene. The *ToxT* product, ToxT, is a DNA-binding protein that activates transcription of several genes required for infection, including the genes that encode the cholera toxin subunits. These genes are called *ctxA* and *ctxB*. As you might suspect, they are part of a polycistronic operon.

Biologists are currently trying to translate research on gene regulation into better tools for fighting cholera. For example, blocking the activity of the ToxR or ToxT proteins should prevent disease. This hypothesis has been supported by the observation that *ToxR*$^-$ mutants are harmless. Researchers are now attempting to develop drugs that prevent ToxR from binding the signal molecule, that stop ToxR from activating *ToxT*, or that prevent ToxT from contacting DNA.

The axiom "knowledge is power" applies here: Although drugs like those proposed earlier are not yet available, research on gene regulation has provided several novel and promising drug targets.

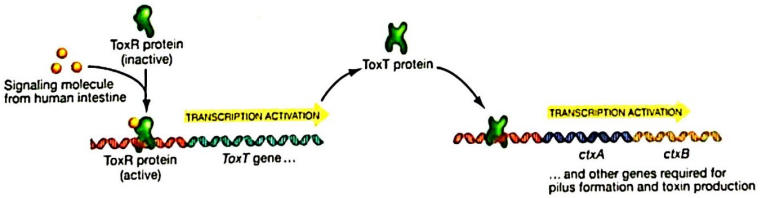


FIGURE 17.16 In the Cholera Bacterium, the Genes Required for Infection Are Regulated
QUESTION Biologists have not been able to identify the signaling molecule that triggers the cascade of events shown here. What properties must this molecule have?

Motif Exhibit 1043, Page 44 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

CHAPTER REVIEW

Summary of Key Concepts

- **Changes in gene expression allow bacterial cells to respond to environmental changes.**

  Transcription in bacteria is constitutive or inducible. Constitutive expression occurs in genes whose products are required at all times, such as genes that encode glycolytic enzymes. Expression of most genes is induced by environmental signals, however, so gene products are produced on demand.

  Precise regulation over gene expression is important to bacterial cells. For example, some species switch between different food sources to take advantage of the nutrients currently available in their environment. Efficient use of nutrients is beneficial because cells are usually competing for access to resources. In the intestinal bacterium *E. coli*, for example, changes in gene expression allow cells to use lactose as a food source when glucose is not available.

- **Gene expression can be controlled at three levels: transcription, translation, or post-translation (protein activation).**

  Among the three levels of gene regulation, there is a trade-off between the speed of response to changed conditions and the efficient use of resources. In post-translational control, for example, the change in gene expression is extremely fast but demands a considerable investment of resources to make and regulate the protein in question.

- **Transcriptional control can be positive or negative. In the case of the genes for lactose metabolism in *E. coli*, positive control occurs when a regulatory protein binds to DNA and increases the transcription rate. Negative control occurs because another regulatory protein must be released before transcription can occur.**

  Protein-coding genes involved in lactose metabolism are normally transcribed at low levels because a repressor protein binds to a DNA sequence—the operator—near their promoter. But when present, lactose binds to the repressor and induces a conformational change that causes the repressor to fall off the operator. Rapid transcription of the *lac* loci does not begin, however, unless glucose is also lacking. When glucose is scarce, cAMP levels rise in the cell. This is important because the regulatory protein CAP undergoes a conformational change when it is bound to cAMP. The CAP-cAMP complex binds to a control sequence near the *lac* promoter and increases transcription by facilitating binding by RNA polymerase.

  Web Tutorial 17.1 The *Lac* Operon

- **Many regulatory proteins bind to specific sites in DNA. Because different regulatory proteins have different amino acid sequences, they bind to different DNA sequences.**

  The regulatory proteins involved in negative control and positive control are often allosterically regulated, meaning that they have the ability to switch between two different conformations in response to binding a small molecule—often a nutrient or product affected by the gene being regulated. The change in conformation affects the protein's ability to bind to DNA.

Questions

Content Review

1. Genes for enzymes in glycolysis are expressed constitutively. Why?
   a. Their expression is controlled at three levels: transcriptional, translational, and post-translational.
   b. Glycolysis occurs at all times, so they need to synthesize ATP-producing enzymes constantly.
   c. Transcription is activated only in response to signals indicating that ATP supplies are low.
   d. They do not need to be expressed when fermentation instead of cellular respiration is occurring.

2. Why do researchers frequently begin a search for mutants by exposing organisms to UV light or X-rays?
   a. The treatment causes mutants to turn yellow.
   b. The treatment triggers gene expression.
   c. The treatment exposes constitutive mutants.
   d. The treatment increases the mutation rate by damaging DNA.

3. What is the goal of a genetic screen?
   a. to identify individuals that have mutations in genes that are required for a particular process
   b. to produce a large number of mutant individuals, which can later be analyzed in detail
   c. to determine whether the mutations carried by particular individuals are in protein-coding genes or regulatory elements
   d. to determine the fitness or viability of mutant individuals

4. Why are the genes involved in lactose metabolism considered to be an operon?
   a. They occupy adjacent locations on the *E. coli* chromosome.
   b. They have a similar function.
   c. They are all required for normal cell function.
   d. They are under the control of the same promoter.

5. What is catabolite repression?
   a. a mechanism that turns off the synthesis of enzymes responsible for catabolic reactions when the product is present
   b. a mechanism that turns off the synthesis of enzymes responsible for catabolic reactions when the product is absent
   c. repression that occurs because of allosteric changes in a regulatory protein
   d. repression that occurs because of allosteric changes in a DNA sequence

6. What is a helix-turn-helix motif?
   a. a protein domain involved in folding into the active conformation
   b. a protein domain involved in induction
   c. a protein domain involved in DNA binding
   d. a protein domain involved in catalysis

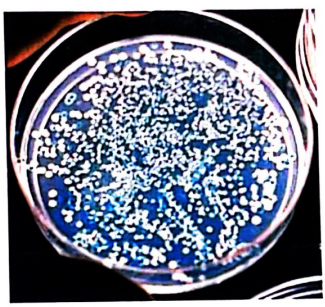
Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

Conceptual Review

1. Explain the difference between positive and negative control over transcription. Why is it advantageous for the *lac* operon in *E. coli* to be under both positive control and negative control? What would happen if only negative control occurred? What would happen if only positive control occurred?
2. In *E. coli*, rising levels of cAMP can be a considered a starvation signal. Explain.
3. Explain the role of post-translational control in the *lac* operon.
4. CAP is also known as the cAMP-receptor protein. Why?
5. Explain how the *lac* repressor binds to DNA. Then explain the relationship between the two-fold rotational symmetry of the operator and the four-part structure of the repressor protein.
6. The galactose released when β-galactosidase cleaves lactose enters the glycolytic pathway in *E. coli* once a series of enzyme-catalyzed reactions has converted the galactose to glucose-6-phosphate. Why, then, is glucose, not lactose, the preferred sugar in *E. coli*?

Group Discussion Problems

1. You are interested in using bacteria to metabolize wastes at an old chemical plant and convert them into harmless compounds. You find bacteria that are able to tolerate high levels of the toxic compounds toluene and benzene, and you suspect that it is due to the ability of the bacteria to break these compounds into less-toxic products. If that is true, these toluene- and benzene-resistant strains will be valuable for cleaning up toxic sites. How could you find out whether these bacteria have enzymes that allow them to metabolize toluene?
2. Assuming that the bacteria you examined in Group Discussion Problem 1 do have an enzymatic pathway to break down toluene, would you predict that the genes involved are constitutively expressed, under positive control, or under negative control? Why? What experiments could you conduct to test your hypothesis?
3. The *lacI* mutants are constitutive because the repressor fails to recognize and bind to the operator region. Other repressor mutants have been isolated that are called $LacI^S$ mutants. These repressor proteins continue to bind to the operator, even in the presence of the inducer. How would this mutation affect the function of the *lac* operon? Specifically, how well would $LacI^S$ mutants do in an environment that has lactose as its sole sugar?
4. X-gal is a colorless, lactose-like molecule that can be split into two fragments by β-galactosidase. One of these product molecules is blue. The following photograph is a close-up of *E. coli* colonies growing in a medium that contains lactose.



Draw a line to three colonies whose cells have functioning copies of β-galactosidase. Draw a line to three colonies whose cells have mutations in the *lacZ* locus or in one of the genes involved in regulation of the *lacZ*. Suppose you could analyze the sequence of the β-galactosidase gene from each of the mutant colonies. How would these data help you distinguish which cells are structural mutants and which are regulatory mutants?

Answers to Multiple-Choice Questions 1. b; 2. d; 3. a; 4. d; 5. a; 6. c

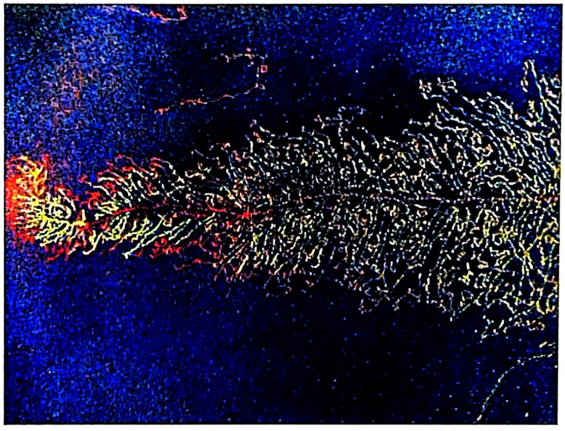
www.prenhall.com/freeman is your resource for the following: Web Tutorials, Online Quizzes and other Online Study Guide materials; Answers to Conceptual Review Questions; Solutions to Group Discussion Problems; Answers to Figure Capture Questions and Exercises; and Additional Readings and Research.

Motif Exhibit 1043, Page 46 of 65

18 Control of Gene Expression in Eukaryotes

KEY CONCEPTS

- Changes in gene expression allow eukaryotic cells to respond to changes in the environment and cause distinct cell types to develop.
- In eukaryotes, DNA is packaged with proteins into complex structures that must be opened before transcription can occur.
- In eukaryotes, transcription can be initiated only when specific proteins bind to the promoter and to regulatory sequences that may be close to the promoter or far from it.
- Alternative splicing allows a single gene to code for several different products.
- Cancer can develop when mutations disable genes that regulate cell-cycle control genes.



A eukaryotic cell's DNA being transcribed. The horizontal strands in the middle of the micrograph is DNA; the yellow strands above and below it are mRNAs.

In bacteria, precise regulation of genes is essential for cells to respond to changes in their environment, as Chapter 17 indicated. In *Escherichia coli*, the genes that are required to import and cleave lactose are rapidly expressed only when the cells have to rely on lactose as a source of energy—meaning glucose is absent and lactose is present. *Vibrio cholerae* can live in a wide array of environments, but the genes that allow the cells to infect a human are expressed only when the cells find themselves inside someone's small intestine.

Unicellular eukaryotes face similar challenges. Consider the yeast *Saccharomyces cerevisiae*, which is used extensively in the production of beer, wine, and bread. In nature, the cells live on the skins of grapes and other fruits. In the wild, variation in temperature and humidity can be high. In addition, the sugars that are available to the cells change dramatically in type and concentration as the fruit that they are living on ripens, falls, and rots. For yeast cells to grow and reproduce efficiently, gene expression has to be modified in response to these environmental changes. Further, the offspring of a cell that lived on a grape may be blown to a nearby orchard and take up residence on an apple. The genes that are expressed in the cells of grapes versus those expressed in the cells of apples vary markedly.

The cells that make up multicellular eukaryotes face an even greater challenge. Consider that, at one time, your body consisted of a single cell—a fertilized egg. Now you have trillions of cells, each of which has a specialized structure and function. You have heart muscle cells, lung cells, nerve cells, skin cells, and so on. Even though these cells look and behave very differently, each of them contains the same genes. Your bone cells are different from your blood cells not because they contain different genes, but because they express different genes. Your bone cells have blood cell genes—they just don't transcribe them.

How is it possible for the cells inside your body to express unique sets of genes? The answer is that your cells respond to changes in their environment, just as bacteria and unicellular eukaryotes do. But instead of responding to the presence of mole-

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

cules such as glucose or lactose in the external environment, your cells respond to the presence of signals from other cells—from an internal environment. As a human being or an oak tree develops, cells that are located in different parts of the organism are exposed to different signals. As a result, they express different genes. Differential gene expression is responsible for creating different cell types, arranging them into tissues, and coordinating their activity to form the multicellular society we call an individual.

How does all this happen? Unit 4 explores the nature of the signals that trigger the formation of muscle, bone, leaf, and flower cells in multicellular organisms. Here we need to focus on what happens once a eukaryotic cell receives a signal. Let's begin by surveying the various ways that gene expression can be controlled and then plunge into a detailed look at key points in gene regulation. The chapter closes with a look at how breakdowns in the control of gene expression can help trigger cancer.

18.1 Mechanisms of Gene Regulation—an Overview

Like bacteria, eukaryotes can control gene expression at the levels of transcription, translation, and post-translation (protein activation or inactivation). But as Figure 18.1 shows, two additional levels of control occur in eukaryotes as genetic information flows from DNA to proteins. The first level involves the DNA-protein complex at the top of the figure. In eukaryotes, DNA is wrapped around proteins to create a protein-DNA complex called **chromatin**. Eukaryotic genes have promoters, just as bacterial genes do. Before transcription can begin in eukaryotes, DNA near the promoter must be released from tight interactions with proteins so that RNA polymerase can make contact with the promoter. To capture this point, biologists say that **chromatin remodeling** must occur prior to transcription.

The second level of regulation that is unique to eukaryotes involves **RNA processing**—the steps required to produce a mature, processed mRNA from a **primary RNA transcript**, the preliminary result of transcription. Recall from Chapter 16 that introns have to be spliced out of primary transcripts. In some cases, carefully orchestrated alterations in the splicing pattern occur. When the splicing events in a particular primary RNA transcript change, a different message emerges. The altered message, in turn, leads to the production of a different product.

How many of the five potential control points shown in Figure 18.1 are actually used? In all cells, every one of them is employed at certain times. The primary focus of this chapter is to explore three of these key transition points: (1) control of packing DNA into chromatin (control over chromatin remodeling), (2) control of transcription, and (3) control of RNA processing.

To appreciate the breadth and complexity of gene regulation in eukaryotes, let's follow the series of events that occurs as an embryonic cell responds to a developmental signal. Suppose a molecule arrives that specifies the production of a muscle-specific protein. What happens next?

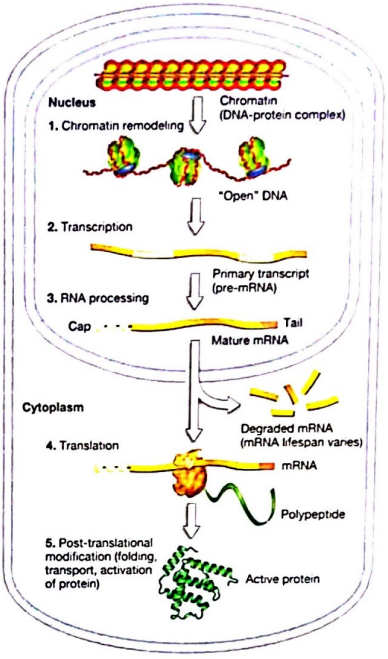


FIGURE 18.1 In Eukaryotes, Gene Expression Can Be Controlled at Many Different Steps
Numerous steps occur during protein synthesis in eukaryotes. Each step provides an opportunity for regulation of gene expression.
EXERCISE Place and label a horizontal line such that mechanisms of transcriptional control fall above the line and mechanisms of post-transcriptional control fall below it.

18.2 Eukaryotic DNA and the Regulation of Gene Expression

If the arrival of a signaling molecule from outside the cell is going to result in the transcription of a particular gene, the DNA around the target gene must be drastically remodeled. To appreciate why, consider that a typical cell in your body contains about six billion base pairs of DNA. Lined up end to end, these nucleotide pairs would form a double helix about 2 m (6.5 feet) long. But the nucleus that holds this DNA is thinner than a sheet of paper. In eukaryotes, DNA is packed inside the nucleus so tightly that RNA polymerase can't access it. Part of this packing is done by supercoiling—meaning the DNA double helix is twisted on itself

Motif Exhibit 1043, Page 48 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

many times, just as in bacterial chromosomes (see Chapter 7). The supercoiled DNA found in bacteria does not need to be altered extensively before it can be transcribed, however. Eukaryotic DNA is different—it has to undergo a series of significant physical changes before transcription can take place. The reason is that supercoiling is just part of the packing system found in eukaryotes.

Chromatin Structure

The first data on the physical nature of eukaryotic DNA were published in the early 1900s, when chemical analyses established that eukaryotic DNA is intimately associated with proteins. Later work documented that the most abundant DNA-associated proteins belong to a group called the **histones**. In the 1970s electron micrographs like the one in **Figure 18.2a** revealed that the protein-DNA complex, or chromatin, has a regular structure. In some preparations for electron microscopy, chromatin actually looked like beads on a string. The "beads" came to be called **nucleosomes**. More details emerged in 1984 when researchers determined the three-dimensional structure of eukaryotic DNA by using X-ray crystallography (a technique introduced in Chapter 4). The X-ray crystallographic data indicated that each nucleosome consists of DNA wrapped almost twice around a core of eight histone proteins. As **Figure 18.2b** indicates, between each pair of nucleosomes there is a "linker" stretch of DNA associated with a histone called **H1**.

The intimate association between DNA and histones occurs in part because DNA is negatively charged and histones are positively charged. DNA has a negative charge because of its phosphate groups; histones are positively charged because they contain many lysine and/or arginine residues (see Chapter 3).

More recent work has shown that there is another layer of complexity in eukaryotic DNA. As **Figure 18.2c** indicates, H1 histones interact with each other and with histones in other nucleosomes to produce a tightly packed structure. Based on its width, this structure is called the **30-nanometer fiber**. (Recall that a nanometer is one-billionth of a meter and is abbreviated nm.) Often, 30-nm fibers are packed into still larger structures, which organize DNA inside the nucleus.

The elaborate structure of chromatin does more than just package DNA into the nucleus. Chromatin structure has profound implications for the control of gene expression. To drive this point home, consider the 30-nanometer fiber illustrated in Figure 18.2c. If this tightly packed stretch of DNA contains a promoter, how can RNA polymerase bind to it and initiate transcription?

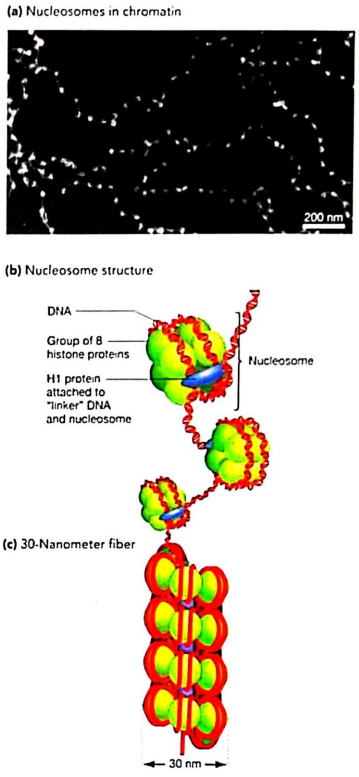


FIGURE 18.2 Chromatin Has Several Levels of Structure
(a) Electron micrograph of chromatin, showing that bead-like structures, called nucleosomes, are linked by short stretches of DNA. **(b)** X-ray crystallographic data revealed that nucleosomes consist of DNA wrapped around a core of eight histone proteins. The linker DNA is associated with a particular histone called H1. **(c)** A model of the structure of the 30-nanometer fiber. It proposes that H1 proteins form a core in the center of the fiber. Each ball here represents a group of eight histones, to make the overall structure of the 30-nanometer fiber clearer.

Evidence that Chromatin Structure Is Altered in Active Genes

Once the nucleosome-based structure of chromatin was established, biologists hypothesized that the close physical interaction between DNA and histones must be altered for RNA polymerase to make contact with DNA. More specifically, biologists hypothesized that a gene could not be transcribed until the chromatin near its promoter was remodeled. The idea was that the chromatin must be relaxed or decondensed for RNA polymerase to bind to the promoter. If so, then chromatin remodeling would represent the first step in the control of eukaryotic gene expression. Two types of studies have provided strong support for this hypothesis.

The first type of evidence comes from studies with the enzyme DNase, which cuts DNA at random locations. DNase cannot cut DNA efficiently if the molecule is tightly complexed with his-

Motif Exhibit 1043, Page 49 of 65

Case No.: IPR2023-00322
U.S. Patent No. 10,273,492

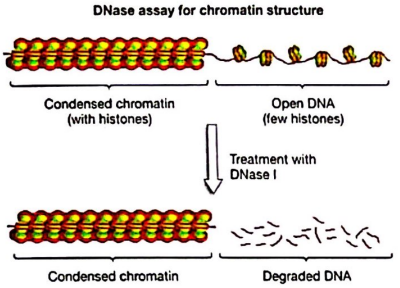


FIGURE 18.3 DNase Cannot Cut Intact Condensed Chromatin
DNase is an enzyme that cuts DNA at random locations.

tones. As Figure 18.3 shows, the enzyme works effectively only if DNA is in the "open" configuration. Harold Weintraub and Mark Groudine used this observation to test the hypothesis that the DNA of actively transcribed genes is in an open configuration. They performed the test by comparing the structure of chromatin in two genes of the same cell type: the β-globin and ovalbumin genes of chicken blood cells. β-Globin is a protein that is part of the hemoglobin found in red blood cells; ovalbumin is a major protein of egg white. In blood cells, the β-globin gene is transcribed at high levels. The ovalbumin gene, in contrast, is not transcribed at all. After treating blood cells with DNase and then analyzing the state of the β-globin and ovalbumin genes, the researchers found that DNase cut up the β-globin gene much more readily than the ovalbumin gene. They interpreted this finding as evidence that chromatin in blood cells was in an open configuration at the β-globin gene but closed at the ovalbumin gene. Analogous studies using DNase on different genes and in different cell types yielded similar results.

The second type of evidence in support of the chromatin-remodeling hypothesis comes from studies of mutant brewer's yeast cells that do not produce the usual complement of histones. Researchers found that many yeast genes that are normally never transcribed are instead transcribed at high levels at all times in these mutant cells. To interpret this finding, biologists hypothesized that the lack of histone proteins prevented the assembly of normal chromatin. If the absence of normal histone-DNA interactions promotes transcription, then the presence of normal histone-DNA interactions must prevent it.

Taken together, the data suggest that the normal or default state of eukaryotic genes is to be turned off. If so, then gene expression depends on chromatin being opened up in the promoter region.

How Is Chromatin Altered?

Although much remains to be learned about how chromatin is opened up prior to transcription, researchers have made important progress recently. One key finding is that two major types of protein are involved in modifying chromatin structure. One group of proteins creates structures called **chromatin-remodeling complexes**, which reshape chromatin through a series of reactions that are dependent on ATP. It is not known exactly how these ATP-dependent enzymes work, however. The second type of chromatin-modifying protein works by adding small molecules such as acetyl (CH_3COOH) or methyl (CH_3) groups to histones. These processes are known as **acetylation** and **methylation**, respectively.

Figure 18.4 shows how the enzymes called **histone acetyl transferases (HATs)** affect chromatin structure. HATs acetylate the positively charged lysine residues in histones. When a HAT adds an acetyl group to selected histones, the number of positive charges on the histones is reduced. The result is less electrostatic attraction between histones and the negatively charged DNA. The association between nucleosomes and DNA is loosened and the chromatin decondenses.

Figure 18.4 also shows how chromatin is "recondensed" by a group of enzymes called **histone deacetylases (HDACs)**, which remove the acetyl groups added by HATs. Histone deacetylase activity reverses the effects of acetylation, returning chromatin to its default condensed state. If HATs are an on switch for transcription, HDACs are the off switch.

What happens once a section of DNA is opened up and exposed to RNA polymerase? In bacteria, the protein sigma binds to RNA polymerase and allows it to contact the promoter. Then transcription begins if any required positive regulators, such as CAP, are in place and negative regulators, such as the repressor, are absent. Does anything similar happen in eukaryotes?

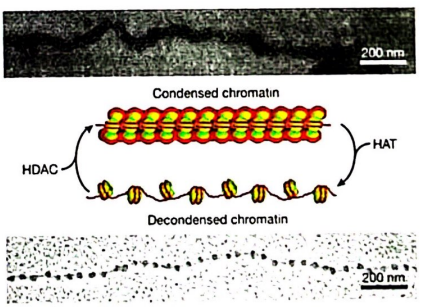


FIGURE 18.4 Chromatin Decondensation and Condensation Are Driven by Histone Acetyl Transferases and Histone Deacetylases
Histone acetyl transferases (HATs) cause chromatin to decondense, and histone deacetylases (HDACs) cause it to condense. EXERCISE In the condensed chromatin, each nucleosome has an acetyl group attached to it. Indicate this by adding "Ac" to several nucleosomes in the bottom part of the drawing.

Motif Exhibit 1043, Page 50 of 65