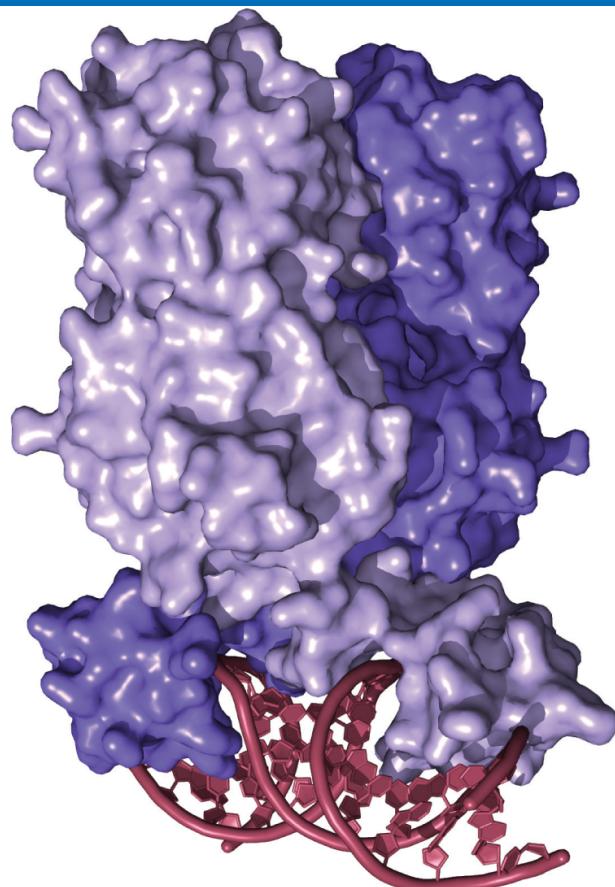


CHAPTER 11

Regulation of Gene Expression in Bacteria and Their Viruses



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

The control of gene expression is governed primarily by DNA-binding proteins that recognize specific control sequences of genes. Here, the binding of the Lac repressor protein to the lac operator DNA is modeled.

CHAPTER OUTLINE AND LEARNING OBJECTIVES

11.1 GENE REGULATION

LO 11.1 Illustrate how both positive and negative regulation control the activity of the *lac* operon.

11.2 DISCOVERY OF THE LAC SYSTEM: NEGATIVE REGULATION

LO 11.2 Infer the components of genetic switches from experimental data and predict the effect of mutations in the different components on gene expression.

11.3 CATABOLITE REPRESSION OF THE LAC OPERON: POSITIVE REGULATION

11.4 DUAL POSITIVE AND NEGATIVE REGULATION: THE ARABINOSE OPERON

LO 11.3 Illustrate and compare the mechanisms that coordinate expression of sets of genes in bacteria and bacteriophage.

11.5 METABOLIC PATHWAYS AND ADDITIONAL LEVELS OF REGULATION: ATTENUATION

11.6 BACTERIOPHAGE LIFE CYCLES: MORE REGULATORS, COMPLEX OPERONS

LO 11.4 Explain the roles of sequence-specific DNA-binding proteins and DNA regulatory sequences in coordinating the expression of sets of genes in bacteria and bacteriophage.

11.7 ALTERNATIVE SIGMA FACTORS REGULATE LARGE SETS OF GENES

CHAPTER OBJECTIVE

Bacteria and their viruses use a straightforward logic of positive and negative regulation to coordinately control the expression of genes in response to environmental conditions. The broad objective for this chapter is to learn how this regulatory logic can be uncovered using genetic approaches in bacteria and their viruses.

In December 1965, the king of Sweden presented the Nobel Prize in Physiology or Medicine to François Jacob, Jacques Monod, and André Lwoff of the Pasteur Institute for their discoveries of how gene expression is regulated (**Figure 11-1**). The prizes were the fruit of an exceptional collaboration among three superb scientists. They were also triumphs over great odds. The chances were slim that each of these three men would have lived to see that day, let alone earn such honors.

Pioneers of gene regulation



The Pasteur Institute.

FIGURE 11-1 François Jacob, Jacques Monod, and André Lwoff were awarded the 1965 Nobel Prize in Physiology or Medicine for their pioneering work on how gene expression is regulated.

Twenty-five years earlier, Monod had been a doctoral student at the Sorbonne in Paris, working on a phenomenon in bacteria called “enzymatic adaptation” that seemed so obscure to some that the director of the zoological laboratory where he worked stated, “What Jacques Monod is doing is of no interest whatever to the Sorbonne.” Jacob was a 19-year-old medical student intent on becoming a surgeon. Lwoff was by that time a well-established member of the Pasteur Institute in Paris, chief of its department of microbial physiology.

Then came World War II.

As France was invaded and quickly defeated, Jacob raced for the coast to join the Free French forces assembling in England. He served as a medic in North Africa and in Normandy until badly wounded. Monod joined the French Resistance while continuing his work. After a Gestapo raid on his Sorbonne laboratory, Monod decided that working there was too dangerous (his predecessor in the Resistance was arrested and executed), and André Lwoff offered him space at the Pasteur. Monod, in turn, connected Lwoff with the Resistance.

After the liberation of Paris, Monod served in the French army and happened on an article by Oswald Avery and colleagues demonstrating that DNA is the hereditary material in bacteria (see [Chapter 7](#)). His interest in genetics was rekindled, and he rejoined Lwoff after the war. Meanwhile, Jacob’s injuries were too severe for him to pursue a career in surgery. Inspired by the enormous impact of antibiotics introduced late in the war, Jacob eventually decided to pursue scientific research. Jacob approached Lwoff several times for a position in his laboratory but was declined. He made one last try and caught Lwoff in a jovial mood. The senior scientist told Jacob, “You know, we have just found the induction of the prophage. Would you be interested in working on the phage?” Jacob had no idea what Lwoff was talking about. He stammered, “That’s just what I would like to do.”

The cast was set. What unfolded in the subsequent decade was one of the most creative and productive collaborations in the history of genetics, whose discoveries still reverberate throughout biology today.

One of the most important insights arrived not in the laboratory but in a movie theater. Struggling with a lecture that he had to prepare, Jacob opted instead to take his wife, Lise, to a Sunday matinee. Bored and daydreaming, Jacob drew a connection between the work he had been doing on the induction of prophage and that of Monod on the induction of enzyme synthesis. Jacob became “involved by a sudden excitement mixed with a vague pleasure.... Both experiments ...

on the phage ... and that done with Pardee and Monod on the lactose system ... are the same! Same situation. Same result ... In both cases, a gene governs the formation ... of a repressor blocking the expression of other genes and so preventing either the synthesis of the galactosidase or the multiplication of the virus.... Where can the repressor act to stop everything at once? The only simple answer ... is on the DNA itself!”¹

And so was born the concept of a repressor acting on DNA to repress the induction of genes. It would take many years before the hypothesized repressors were isolated and characterized biochemically. The concepts worked out by Jacob and Monod and explained in this chapter—messenger RNA, promoters, operators, regulatory genes, operons, and allosteric proteins—were deduced entirely from genetic evidence, and these concepts shaped the future field of molecular genetics.

Walter Gilbert, who isolated the first repressor and was later awarded a Nobel Prize in Chemistry for co-inventing a method of sequencing DNA, explained the effect of Jacob and Monod’s work at that time: “Most of the crucial discoveries in science are of such a simplifying nature that they are very hard even to conceive without actually having gone through the experience involved in the discovery.... Jacob’s and Monod’s suggestion made things that were utterly dark, very simple.”²

The concepts that Jacob and Monod illuminated went far beyond bacterial enzymes and viruses. They understood, and were able to articulate with exceptional eloquence, how their discoveries about gene regulation pertained to the general mysteries of cell differentiation and embryonic development in animals. The two men once quipped, “anything found to be true of *E. coli* must also be true of Elephants.”³ In the next three chapters, we will see to what degree that assertion is true. We’ll start in this chapter with bacterial examples that illustrate key themes and mechanisms in the regulation of gene expression. We will largely focus on single regulatory proteins and the genetic “switches” on which they act. Then, in [Chapter 12](#), we’ll tackle gene regulation in eukaryotic cells, which entails more complex biochemical and genetic machinery. Finally, in [Chapter 13](#), we’ll examine the role of gene regulation in the development of multicellular animals. There we will see how sets of regulatory proteins act on arrays of genetic switches to control gene expression in time and space and choreograph the building of bodies and body parts.

11.1 GENE REGULATION

LO 11.1 Illustrate how both positive and negative regulation control the activity of the *lac* operon.

Despite their simplicity of form, bacteria have in common with larger and more complex organisms the need to regulate expression of their genes. One of the main reasons is that they are nutritional opportunists. Consider how bacteria obtain the many important compounds, such as sugars, amino acids, and nucleotides, needed for metabolism. Bacteria swim in a sea of potential nutrients. They can either acquire the compounds that they need from the environment or synthesize them by enzymatic pathways. But synthesizing these compounds also requires expending energy and cellular resources to produce the necessary enzymes for these pathways. Thus, given the choice, bacteria will take compounds from the environment instead. Natural selection favors efficiency and selects against the waste of resources and energy. To be economical, bacteria will synthesize the enzymes necessary to produce compounds only when there is no other option—in other words, when compounds are unavailable in their local environment.

Bacteria have evolved regulatory systems that couple the expression of gene products to sensor systems that detect the relevant compound in a bacterium's local environment. The regulation of enzymes taking part in sugar metabolism provides an example. Sugar molecules can be broken down to provide energy, or they can be used as building blocks for a great range of organic compounds. However, there are many different types of sugar that bacteria could use, including lactose, glucose, galactose, and xylose. A different import protein is required to allow each of these sugars to enter the cell. Further, a different set of enzymes is required to process each of the sugars. If a cell were to simultaneously synthesize all the enzymes that it might possibly need, the cell would expend much more energy and materials to produce the enzymes than it could ever derive from breaking down prospective carbon sources. The cell has devised mechanisms to shut down (repress) the transcription of all genes encoding enzymes that are not needed at a given time and to turn on (activate) those genes encoding enzymes that are needed. For example, if only lactose is in the environment, the cell will shut down the transcription of the genes encoding enzymes needed for the import and metabolism of glucose, galactose, xylose, and other sugars. Conversely, *E. coli* will initiate the transcription of the genes encoding enzymes needed for the import and metabolism of lactose. In sum, cells need mechanisms that fulfill two criteria:

1. They must be able to recognize environmental conditions in which they should activate or repress the transcription of the relevant genes.
 2. They must be able to toggle on or off, like a switch, the transcription of each specific gene or group of genes.
-

KEY CONCEPT Cells must be able both to recognize environmental conditions and to respond to those conditions by activating or repressing particular genes.

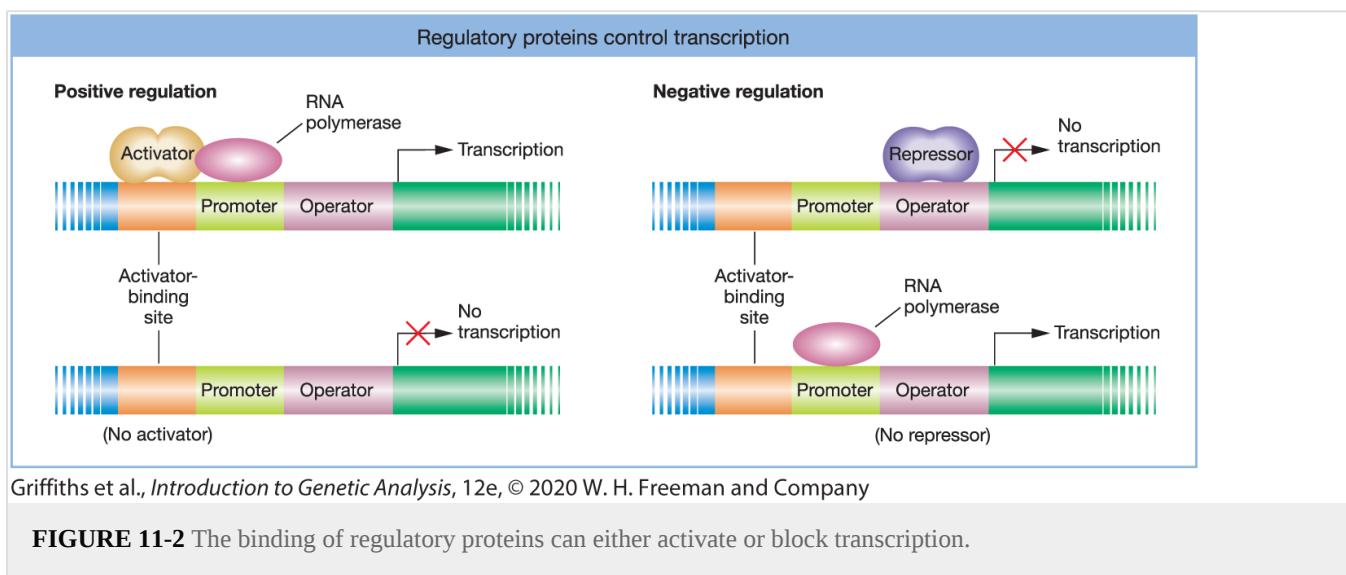
Let's preview the current model for bacterial transcriptional regulation and then use a well-understood example—the regulation of the genes in the metabolism of the sugar lactose—to examine it in detail. In particular, we will focus on how this regulatory system was dissected with the use of the tools of classical genetics and molecular biology.

The basics of bacterial transcriptional regulation: genetic switches

The regulation of transcription depends mainly on two types of protein–DNA interactions. Both take place near the site at which gene transcription begins.

One of these DNA–protein interactions determines where transcription begins. The DNA that participates in this interaction is a DNA segment called the **promoter** ([Chapter 8, Section 8.2](#)), and the protein that binds to this site is RNA polymerase. When RNA polymerase binds to the promoter DNA, transcription can start a few bases away from the promoter site. Every gene must have a promoter or it cannot be transcribed.

The other type of DNA–protein interaction determines whether promoter-driven transcription takes place. DNA segments near the promoter serve as binding sites for sequence-specific regulatory proteins called **activators** and **repressors**. In bacteria, most binding sites for repressors are termed **operators**. For some genes, an activator protein must bind to its target DNA site as a necessary prerequisite for transcription to begin. Such instances are sometimes referred to as **positive regulation** because the *presence* of the bound protein is required for transcription ([Figure 11-2](#)). For other genes, a repressor protein must be prevented from binding to its target site as a necessary prerequisite for transcription to begin. Such cases are sometimes termed **negative regulation** because the *absence* of the bound repressor allows transcription to begin.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-2 The binding of regulatory proteins can either activate or block transcription.

In How do activators and repressors regulate transcription? Often, a DNA-bound activator protein physically helps tether RNA polymerase to its nearby promoter so that polymerase may begin transcribing. A DNA-bound repressor protein typically acts either by physically interfering with the binding of RNA polymerase to its promoter (blocking transcription initiation) or by impeding the movement of RNA polymerase along the DNA chain (blocking transcription). Together, these regulatory proteins and their binding sites constitute **genetic switches** that control the efficient changes in gene expression that occur in response to environmental conditions.

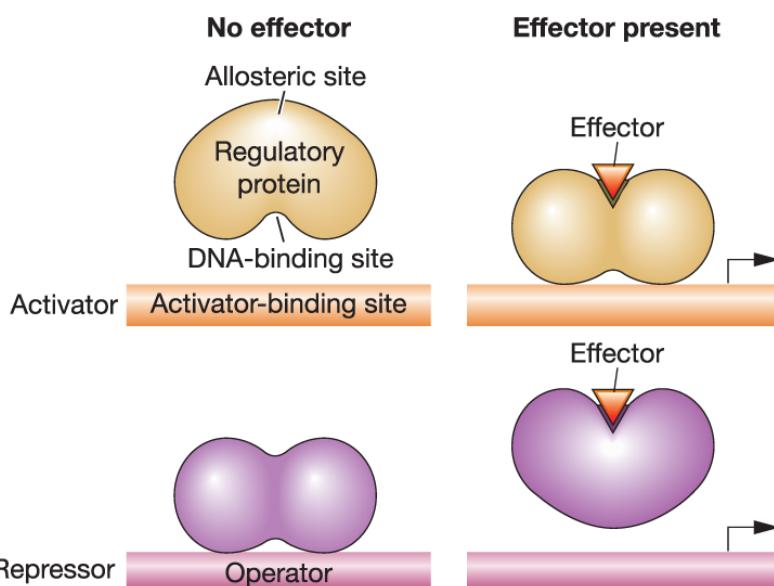
KEY CONCEPT Genetic switches are proteins and DNA sequences that control gene transcription. Activator or repressor proteins bind to operator sequences in the vicinity of the promoter to control its accessibility to RNA polymerase.

Both activator and repressor proteins must be able to recognize when environmental conditions are appropriate for their actions and act accordingly. Thus, for activator or repressor proteins to do their job, each must be able to exist in two states: one that can bind its DNA targets and another that cannot. The binding state must be appropriate to the set of physiological conditions present in the cell and its environment. For many regulatory proteins, DNA binding is effected through the interaction of two different sites in the three-dimensional structure of the protein. One site is the **DNA-binding domain**. The other site, the **allosteric site**, acts as a sensor that sets the DNA-binding domain in one of two modes: functional or nonfunctional. The allosteric site interacts with small molecules called **allosteric effectors**.

In lactose metabolism, it is actually an isomer of the sugar lactose (called allolactose) that is an allosteric effector: the sugar binds to a regulatory protein that inhibits the expression of genes

needed for lactose metabolism. In general, an allosteric effector binds to the allosteric site of the regulatory protein in such a way as to change its activity. In this case, allolactose changes the shape and structure of the DNA-binding domain of a regulatory protein. Some activator or repressor proteins must bind to their allosteric effectors before they can bind DNA. Others can bind DNA only in the absence of their allosteric effectors. Two of these situations are shown in **Figure 11-3**.

Allosteric effectors bind to regulatory proteins



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

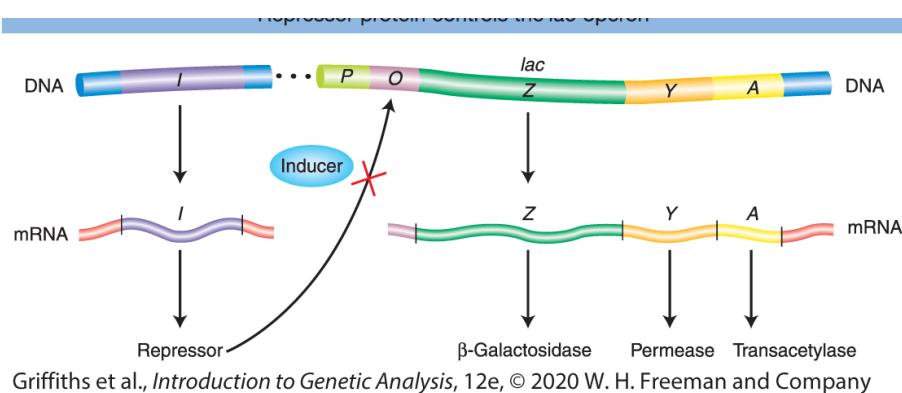
FIGURE 11-3 Allosteric effectors influence the DNA-binding activities of activators and repressors.

KEY CONCEPT Allosteric effectors are small molecules that bind to activator or repressor proteins and control their ability to bind to their DNA target sites.

A first look at the *lac* regulatory circuit

The pioneering work of François Jacob and Jacques Monod in the 1950s showed how lactose metabolism is genetically regulated. Let's examine the system under two conditions: the presence and the absence of lactose. **Figure 11-4** is a simplified view of the components of this system. The cast of characters for *lac* operon regulation includes protein-coding genes and sites on the DNA that are targets for DNA-binding proteins.

Repressor protein controls the *lac* operon



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

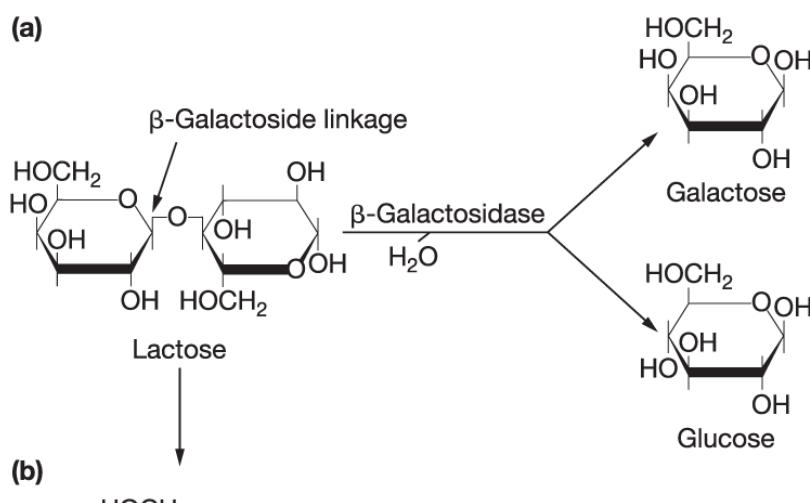
FIGURE 11-4 A simplified *lac* operon model. Coordinate expression of the *Z*, *Y*, and *A* genes is under negative regulation by the product of the *I* gene, the repressor. When the inducer binds the repressor, the operon is fully expressed.

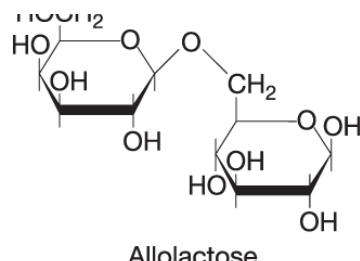


The *lac* structural genes

The metabolism of lactose requires two enzymes: (1) a permease to transport lactose into the cell, and (2) β -galactosidase to modify lactose into allolactose and to cleave the lactose molecule to yield glucose and galactose (Figure 11-5). The structures of the β -galactosidase and permease proteins are encoded by two adjacent sequences, *Z* and *Y*, respectively. A third contiguous sequence, *A*, encodes an additional enzyme, termed *transacetylase*, which is not required for lactose metabolism. We will call *Z*, *Y*, and *A* *structural genes*—in other words, segments encoding proteins—while reserving judgment on this categorization until later. We will focus mainly on the *Z* and *Y* genes. All three genes are transcribed into a single messenger RNA molecule. Regulation of the production of this mRNA coordinates the synthesis of all three enzymes. That is, either all or none of the three enzymes are synthesized. Genes whose transcription is controlled by a common means are said to be coordinately controlled genes.

Lactose is broken down into two sugars and modified into allolactose





Allolactose

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020

W. H. Freeman and Company

FIGURE 11-5 The metabolism of lactose. (a) The enzyme β -galactosidase catalyzes a reaction in which water is added to the β -galactoside linkage to break lactose into separate molecules of glucose and galactose. (b) The enzyme also modifies a smaller proportion of lactose into allolactose, which acts as an inducer of the *lac* operon.

KEY CONCEPT If the genes encoding proteins constitute a single transcription unit, the expression of all these genes will be coordinately regulated.

Regulatory components of the *lac* system

Key regulatory components of the lactose metabolic system include a gene encoding a transcription regulatory protein and two binding sites on DNA: one site for the regulatory protein and another site for RNA polymerase.

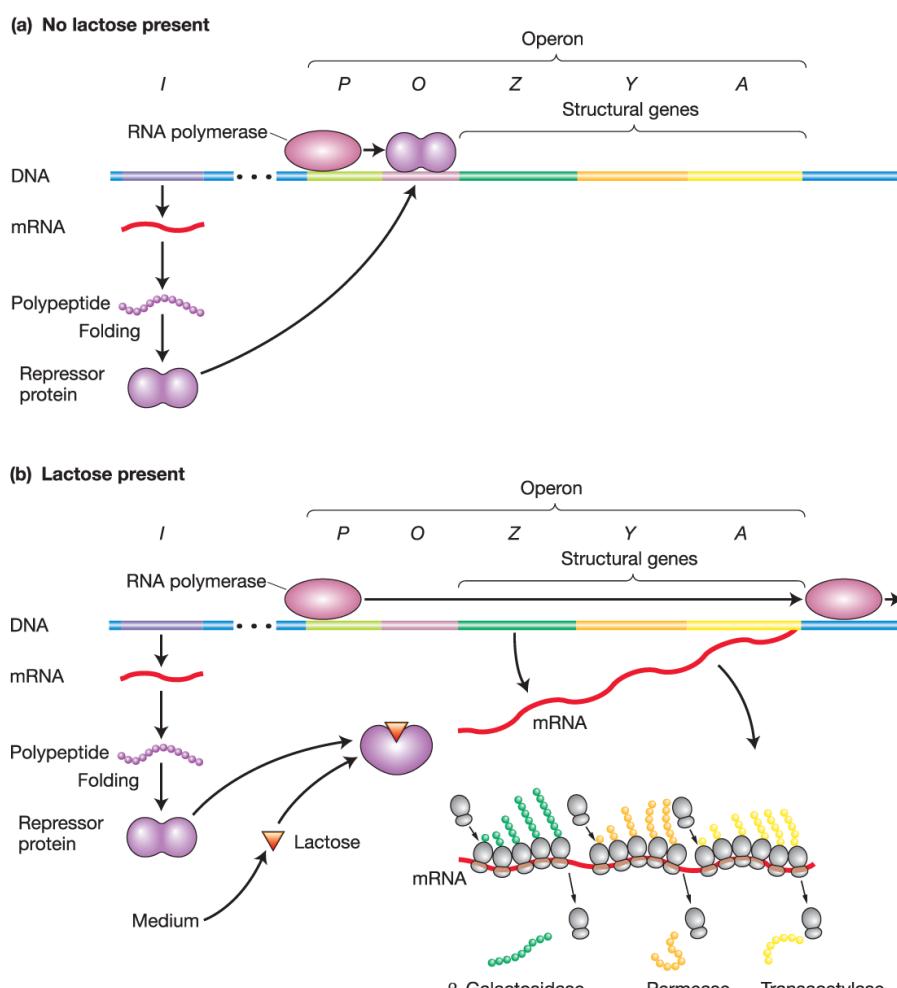
1. *The gene for the Lac repressor.* A fourth gene (besides the structural genes *Z*, *Y*, and *A*), the *I* gene, encodes the Lac repressor protein. It is so named because it can block the expression of the *Z*, *Y*, and *A* genes. The *I* gene happens to map close to the *Z*, *Y*, and *A* genes, but this proximity is not important to its function because it encodes a diffusible protein.
2. *The lac promoter site.* The promoter (*P*) is the site on the DNA to which RNA polymerase binds to initiate transcription of the *lac* structural genes (*Z*, *Y*, and *A*).
3. *The lac operator site.* The operator (*O*) is the site on the DNA to which the Lac repressor binds. It is located between the promoter and the *Z* gene near the point at which transcription of the multigenic mRNA begins.

The induction of the *lac* system

The *P*, *O*, *Z*, *Y*, and *A* segments (shown in [Figure 11-6](#)) together constitute an **operon**, defined as a segment of DNA that encodes a multigenic mRNA as well as an adjacent common promoter and regulatory region. The *lacI* gene, encoding the Lac repressor, is *not* considered part of the *lac*

operon itself, but the interaction between the Lac repressor and the *lac* operator site is crucial to proper regulation of the *lac* operon. The Lac repressor has a *DNA-binding* site that can recognize the operator DNA sequence and an *allosteric site* that binds allolactose or analogs of lactose that are useful experimentally. The repressor will bind tightly only to the *O* site on the DNA near the genes that it is controlling and not to other sequences distributed throughout the chromosome. By binding to the operator, the repressor prevents transcription by RNA polymerase that has bound to the adjacent promoter site; the *lac* operon is switched “off.”

The *lac* operon is transcribed only in the presence of lactose



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-6 Regulation of the *lac* operon. The *I* gene continually makes repressor. (a) In the absence of lactose, the repressor binds to the *O* (operator) region and blocks transcription. (b) The binding of lactose changes the shape of the repressor so that the repressor no longer binds to *O* and falls off the DNA. The RNA polymerase is then able to transcribe the *Z*, *Y*, and *A* structural genes, and so the three enzymes are produced.

ANIMATED ART Sapling Plus

Assaying lactose presence or absence through the Lac repressor

When allolactose or its analogs bind to the repressor protein, the protein undergoes an **allosteric transition**, a change in shape. This slight alteration in shape in turn alters the DNA-binding site so that the repressor no longer has high affinity for the operator. Thus, in response to binding allolactose, the repressor falls off the DNA, allowing RNA polymerase to proceed (transcribe the gene): the *lac* operon is switched “on.” The repressor’s response to allolactose satisfies one requirement for such a control system—that the presence of lactose stimulates the synthesis of genes needed for its processing. The relief of repression for systems such as *lac* is termed **induction**. Allolactose and its analogs that allosterically inactivate the repressor, leading to the expression of the *lac* genes, are termed **inducers**.

Let’s summarize how the *lac* switch works ([Figure 11-6](#)). In the absence of an inducer (allolactose or an analog), the Lac repressor binds to the *lac* operator site and prevents transcription of the *lac* operon by blocking the movement of RNA polymerase. In this sense, the Lac repressor acts as a roadblock on the DNA. Consequently, all the structural genes of the *lac* operon (the Z, Y, and A genes) are repressed, and there are very few molecules of β -galactosidase, permease, or transacetylase in the cell. In contrast, when an inducer is present, it binds to the allosteric site of each Lac repressor subunit, thereby inactivating the site that binds to the operator. The Lac repressor falls off the DNA, allowing the transcription of the structural genes of the *lac* operon to begin. The enzymes β -galactosidase, permease, and transacetylase now appear in the cell in a coordinated fashion. So, when lactose is present in the environment of a bacterial cell, the cell produces the enzymes needed to metabolize it. But when no lactose is present, resources are not wasted.

11.2 DISCOVERY OF THE *LAC* SYSTEM: NEGATIVE REGULATION

LO 11.1 Illustrate how both positive and negative regulation control the activity of the *lac* operon.

LO 11.2 Infer the components of genetic switches from experimental data and predict the effect of mutations in the different components on gene expression.

To study gene regulation, ideally we need three ingredients: a biochemical assay that lets us measure the amount of mRNA or expressed protein or both, reliable conditions in which the levels of expression differ in a wild-type genotype, and genetic mutations that perturb the levels of expression. In other words, we need a way of describing wild-type gene regulation, and we need mutations that can disrupt the wild-type regulatory process. With these elements in hand, we can analyze the expression in mutant genotypes, treating the mutations singly and in combination, to unravel any kind of gene-regulation event. The classical application of this approach was used by Jacob and Monod, who performed the definitive studies of bacterial gene regulation.

Jacob and Monod used the lactose metabolism system of *E. coli* (see [Figure 11-4](#)) to genetically dissect the process of enzyme induction—that is, the appearance of a specific enzyme only in the presence of its substrates. This phenomenon had been observed in bacteria for many years, but how could a cell possibly “know” precisely which enzymes to synthesize? How could a particular substrate induce the appearance of a specific enzyme?

In the *lac* system, the presence of lactose causes cells to produce more than 1000 times as much of the enzyme β -galactosidase as they produced when grown in the absence of lactose. What role did lactose play in the induction phenomenon? When Monod and co-workers followed the fate of radioactively labeled amino acids added to growing cells either before or after the addition of an inducer, they found that induction resulted in the synthesis of new enzyme molecules, as indicated by the presence of the radioactive amino acids in the enzymes. These new molecules could be detected as early as three minutes after the addition of an inducer. Additionally, withdrawal of lactose brought about an abrupt halt in the synthesis of the new enzyme. Therefore, it became clear that the cell has a rapid and effective mechanism for turning gene expression on and off in response to environmental signals.

Genes controlled together

When Jacob and Monod induced β -galactosidase, they found that they also induced the enzyme permease, which is required to transport lactose into the cell. The analysis of mutants indicated that each enzyme was encoded by a different gene. The enzyme transacetylase (with a dispensable and as yet unknown function) also was induced together with β -galactosidase and permease and was later shown to be encoded by a separate gene. Therefore, Jacob and Monod could identify three coordinately controlled genes. Recombination mapping showed that the Z, Y, and A genes were very closely linked on the chromosome (see [Section 6.2](#)).

Genetic evidence for the operator and repressor

Now we come to the heart of Jacob and Monod's work: How did they deduce the mechanisms of gene regulation in the *lac* system? Their strategy was a classic genetic approach: to examine the physiological consequences of mutations. Thus, they induced mutations in the structural genes and regulatory elements of the *lac* operon. As we will see, the properties of mutations in these different components of the *lac* operon are quite different, providing important clues for Jacob and Monod.

Natural inducers, such as allolactose, are not optimal for these experiments because they are broken down by β -galactosidase. The inducer concentration decreases during the experiment, and so the measurements of enzyme induction become quite complicated. Instead, for such experiments, Jacob and Monod used synthetic inducers, such as isopropyl- β -D-thiogalactoside (IPTG; [Figure 11-7](#)). IPTG is not hydrolyzed by β -galactosidase, but it still induces β -galactosidase enzyme expression.

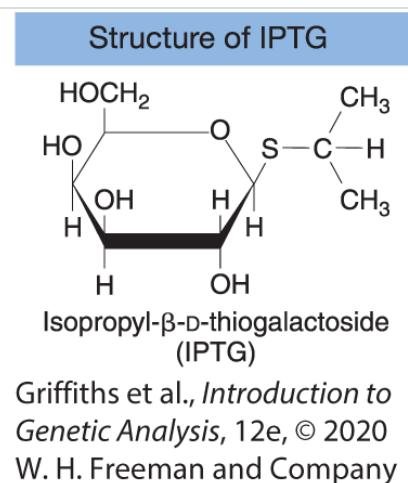


FIGURE 11-7 IPTG is an inducer of the *lac* operon.

Jacob and Monod found that several different classes of mutations can alter the expression of the structural genes of the *lac* operon. They were interested in assessing the interactions between the new alleles, such as which alleles exhibited dominance. But to perform such tests, one needs diploids, and bacteria are haploid. However, Jacob and Monod were able to produce bacteria that are partially diploid by inserting F' factors carrying the *lac* region of the genome. (An F' factor is a plasmid that carries one or more bacterial genes and that can be transferred from one bacteria to another through a process known as conjugation; see [Section 6.2](#).) They could then create strains that were heterozygous for selected *lac* mutations, but still haploid for the rest of the genome. These **partial diploids** allowed Jacob and Monod to distinguish mutations in the regulatory DNA site (the *lac* operator) from mutations in the regulatory protein (the Lac repressor encoded by the *I* gene).

We begin by examining mutations that inactivate the structural genes for β-galactosidase and permease (designated Z^- and Y^- , respectively). The first thing that we learn is that Z^- and Y^- are recessive to their respective wild-type alleles (Z^+ and Y^+). For example, strain 2 in [Table 11-1](#) can be induced to synthesize β-galactosidase (like the wild-type haploid strain 1 in this table), even though it is heterozygous for mutant and wild-type *Z* alleles. This demonstrates that the Z^+ allele is dominant over its Z^- counterpart.

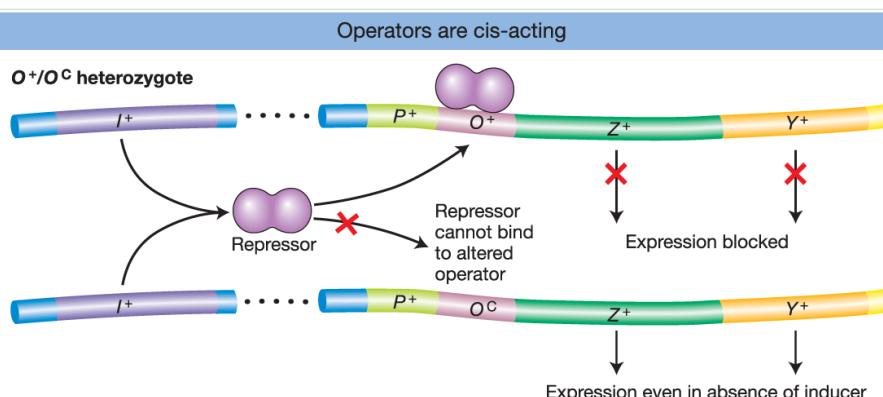
TABLE 11-1 Synthesis of #-Galactosidase and Permease in Haploid and Heterozygous Diploid Operator Mutants

Strain	Genotype	β-Galactosidase (Z)		Permease (Y)		Conclusion
		Noninduced	Induced	Noninduced	Induced	
1	$O^+ Z^+ Y^+$	-	+	-	+	Wild type is inducible
2	$O^+ Z^+ Y^+/F' O^+ Z^- Y^+$	-	+	-	+	Z^+ is dominant to Z^-
3	$O^C Z^+ Y^+$	+	+	+	+	O^C is constitutive
4	$O^+ Z^- Y^+/F' O^C Z^+ Y^-$	+	+	-	+	Operator is cis-acting

Note: Bacteria were grown in glycerol (no glucose present) with and without the inducer IPTG. Expression of maximal enzyme levels is indicated by +. Absence or very low levels of enzyme activity is indicated by -. All strains are I^+ .

Jacob and Monod first identified two classes of regulatory mutations, called O^C and I^- . These were called **constitutive mutations** because they caused the *lac* operon structural genes to be

expressed regardless of whether inducer was present. Jacob and Monod identified the existence of the operator on the basis of their analysis of the O^C mutations. These mutations make the operator incapable of binding to repressor; they damage the switch such that the operon is always “on” ([Table 11-1](#), strain 3). Importantly, the constitutive effects of O^C mutations were restricted solely to those *lac* structural genes *on the same chromosome* as the O^C mutation. For this reason, the operator mutant was said to be **cis-acting**, as demonstrated by the phenotype of strain 4 in [Table 11-1](#). Here, because the wild-type permease (Y^+) gene is cis to the wild-type operator, permease is expressed only when lactose or an analog is present. In contrast, the wild-type β -galactosidase (Z^+) gene is cis to the O^C mutant operator; hence, β -galactosidase is expressed constitutively. This unusual property of cis action suggested that the operator is a segment of DNA that influences only the expression of the structural genes linked to it ([Figure 11-8](#)). The operator thus acts simply as a protein-binding site and makes *no* gene product.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-8 O^+/O^C heterozygotes demonstrate that operators are cis-acting. Because a repressor cannot bind to O^C operators, the *lac* structural genes linked to an O^C operator are expressed even in the absence of an inducer. However, the *lac* genes adjacent to an O^+ operator are still subject to repression.

ANIMATED ART Sapling Plus

O^C *lac* operator mutations

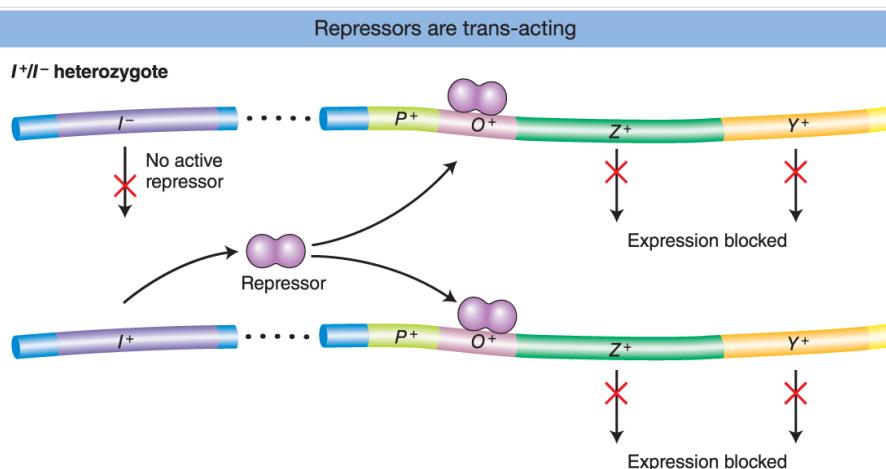
Jacob and Monod did comparable genetic tests with the I^- mutations ([Table 11-2](#)). A comparison of the inducible wild-type I^+ (strain 1) with I^- strains shows that I^- mutations are constitutive (strain 2). That is, they cause the structural genes to be expressed at all times. Strain 3 demonstrates that the inducible phenotype of I^+ is dominant over the constitutive phenotype of I^- . This finding showed Jacob and Monod that the amount of wild-type protein encoded by one copy

of the gene is sufficient to regulate both copies of the operator in a diploid cell. Most significantly, strain 4 showed them that the I^+ gene product is **trans-acting**, meaning that the gene product can regulate *all* structural *lac* operon genes, whether residing on the same DNA molecule or on different ones (in *cis* or in *trans*, respectively). Unlike the operator, the *I* gene behaves like a standard protein-coding gene. The protein product of the *I* gene is able to diffuse throughout a cell and act on both operators in the partial diploid (**Figure 11-9**).

TABLE 11-2 Synthesis of β -Galactosidase and Permease in Haploid and Heterozygous Diploid Strains Carrying I^+ and I^-

Strain	Genotype	β -Galactosidase (Z)		Permease (Y)		Conclusion
		Noninduced	Induced	Noninduced	Induced	
1	$I^+ Z^+ Y^+$	–	+	–	+	I^+ is inducible
2	$I^- Z^+ Y^+$	+	+	+	+	I^- is constitutive
3	$I^+ Z^- Y^+/F' I^- Z^+ Y^+$	–	+	–	+	I^+ is dominant to I^-
4	$I^- Z^- Y^+/F' I^+ Z^+ Y^-$	–	+	–	+	I^+ is trans-acting

Note: Bacteria were grown in glycerol (no glucose present) with and without the inducer IPTG. Expression of maximal enzyme levels is indicated by +. Absence or very low levels of enzyme activity is indicated by –. All strains are O^+ .



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-9 The recessive nature of I^- mutations demonstrates that the repressor is trans-acting. Although no active repressor is synthesized from the I^- gene, the wild-type (I^+) gene provides a functional repressor that binds to both operators in a diploid cell and blocks *lac* operon expression (in the absence of an inducer).

ANIMATED ART  Sapling Plus

I^- Lac repressor mutations

KEY CONCEPT Operator mutations reveal that such a site is cis-acting; that is, it regulates the expression of an adjacent transcription unit on the same DNA molecule. In contrast, mutations in the gene encoding a repressor protein reveal that this protein is trans-acting; that is, it can act on any copy of the target DNA.

Genetic evidence for allostery

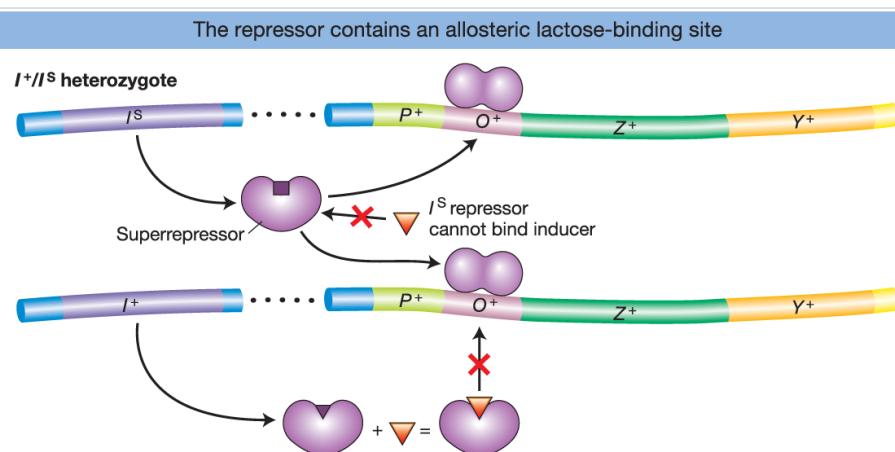
Finally, Jacob and Monod were able to demonstrate allostery through the analysis of another class of repressor mutations. Recall that the Lac repressor inhibits transcription of the *lac* operon in the absence of an inducer but permits transcription when the inducer is present. This regulation is accomplished through a second site on the repressor protein, the allosteric site, which binds to the inducer. When bound to the inducer, the repressor undergoes a change in overall structure such that its DNA-binding site can no longer function.

Jacob and Monod isolated another class of repressor mutation, called superrepressor (I^S) mutations. I^S mutations cause repression to persist even in the presence of an inducer (compare strain 2 in [Table 11-3](#) with the inducible wild-type strain 1). Unlike I^- mutations, I^S mutations are dominant over I^+ (see [Table 11-3](#), strain 3). This key observation led Jacob and Monod to speculate that I^S mutations alter the allosteric site so that it can no longer bind to an inducer. As a consequence, I^S -encoded repressor protein continually binds to the operator—preventing transcription of the *lac* operon even when the inducer is present in the cell. On this basis, we can see why I^S is dominant over I^+ . Mutant I^S protein will bind to both copies of the operator in the partial diploid cell, even in the presence of an inducer and regardless of the fact that I^+ -encoded protein may be present in the same cell ([Figure 11-10](#)).

TABLE 11-3 Synthesis of β -Galactosidase and Permease by the Wild Type and by Strains Carrying Different Alleles of the *I* Gene

Strain	Genotype	β -Galactosidase (Z)		Permease (Y)		Conclusion
		Noninduced	Induced	Noninduced	Induced	
1	$I^+ Z^+ Y^+$	–	+	–	+	I^+ is inducible
2	$I^S Z^+ Y^+$	–	–	–	–	I^S is always repressed
3	$I^S Z^+ Y^+/F' I^+ Z^+ Y^+$	–	–	–	–	I^S is dominant to I^+

Note: Bacteria were grown in glycerol (no glucose present) with and without the inducer IPTG. Expression of maximal enzyme levels is indicated by +. Absence or very low levels of enzyme activity is indicated by –.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

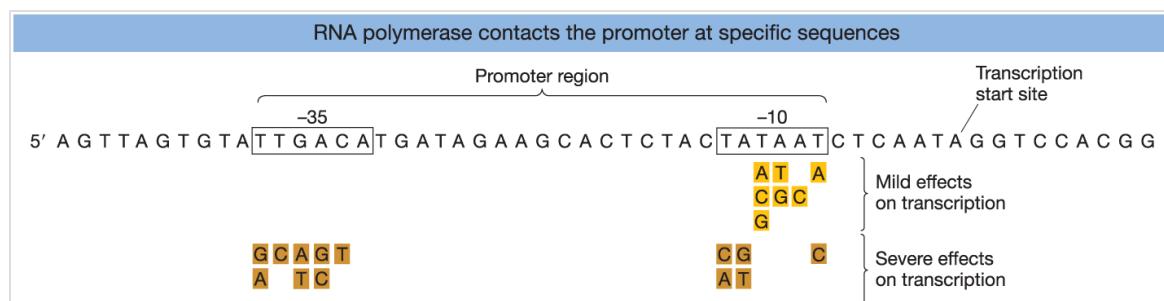
FIGURE 11-10 The dominance of the I^S mutation is due to the inactivation of the allosteric site on the Lac repressor. In an I^S/I^+ diploid cell, none of the *lac* structural genes are transcribed. The I^S repressor lacks a functional allolactose-binding site (the allosteric site) and thus is not inactivated by an inducer. Therefore, even in the presence of an inducer, the I^S repressor binds irreversibly to all operators in a cell, thereby blocking transcription of the *lac* operon.

ANIMATED ART Sapling Plus

I^S Lac superrepressor mutations

Genetic analysis of the *lac* promoter

Mutational analysis also demonstrated that an element essential for *lac* transcription is located between the gene for the repressor *I* and the operator site *O*. This element, termed the *promoter* (*P*), serves as the initiation site for transcription by RNA polymerase (see [Chapter 8](#)). There are two binding regions for RNA polymerase in a typical bacterial promoter, shown in [Figure 11-11](#) as the two highly conserved regions at -35 and -10 . Promoter mutations are *cis*-acting in that they affect the transcription of all adjacent structural genes in the operon. Like operators and other *cis*-acting elements, promoters are sites on the DNA molecule that are bound by proteins and themselves produce no protein product.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-11 Specific DNA sequences are important for the efficient transcription of *E. coli* genes by RNA polymerase. Only the coding strand (non-template strand) is shown here (see [Figure 8-5](#)). Transcription would proceed from left to right (5' to 3'), and the mRNA transcript would be homologous to the sequence shown. The boxed sequences are highly conserved in all *E. coli* promoters, an indication of their role as contact sites on the DNA for RNA polymerase binding. Mutations in these regions have mild (gold) and severe (brown) effects on transcription. [Data from J. D. Watson, M. Gilman, J. Witkowski, and M. Zoller, *Recombinant DNA*, 2nd ed.]

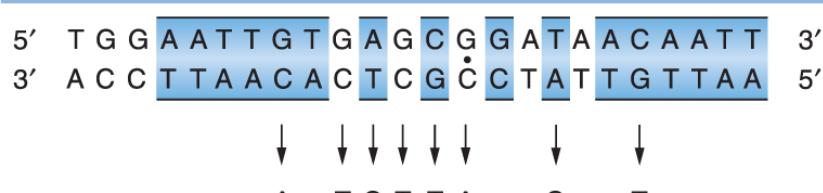
Th

Molecular characterization of the Lac repressor and the *lac* operator

Walter Gilbert and Benno Müller-Hill provided a decisive demonstration of the *lac* system in 1966 by monitoring the binding of the radioactively labeled inducer IPTG to purified repressor protein. They showed that in the test tube, repressor protein binds to DNA containing the operator and comes off the DNA in the presence of IPTG. (A more detailed description of how the repressor and other DNA-binding proteins work is given later, at the end of [Section 11.6](#).)

Gilbert and his co-workers showed that the repressor can protect specific bases in the operator from chemical reagents. This information allowed them to isolate the DNA segment constituting the operator and to determine its sequence. They took operon DNA to which repressor was bound and treated it with the enzyme DNase, which breaks up DNA. They were able to recover short DNA strands that had been shielded from the enzyme activity by the repressor molecule. These short strands presumably constituted the operator sequence. The base sequence of each strand was determined, and each operator mutation was shown to be a change in the sequence ([Figure 11-12](#)). These results showed that the operator locus is a specific sequence of 17 to 25 nucleotides situated just before (5' to) the structural *Z* gene. They also showed the incredible specificity of repressor-operator recognition, which can be disrupted by a single base substitution. When the sequence of bases in the *lac* mRNA (transcribed from the *lac* operon) was determined, the first 21 bases on the 5' initiation end proved to be complementary to the operator sequence that Gilbert had determined, showing that the operator sequence is transcribed.

The operator is a specific DNA sequence



O^c mutations A I G I I A U I
 T A C A A T G A

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 11-12 The DNA base sequence of the lactose operator and the base changes associated with eight O^C mutations.

Regions of twofold rotational symmetry are indicated by color and by a dot at their axis of symmetry. [Data from W. Gilbert, A. Maxam, and A. Mirzabekov, in N. O. Kjeldgaard and O. Malløe, eds., *Control of Ribosome Synthesis*. Academic Press, 1976.]

In The results of these experiments provided crucial confirmation of the mechanism of repressor action formulated by Jacob and Monod.

KEY CONCEPT The function of cis-acting elements such as promoters and operators is determined by their DNA sequence.

11.3 CATABOLITE REPRESSION OF THE *LAC* OPERON: POSITIVE REGULATION

LO 11.1 Illustrate how both positive and negative regulation control the activity of the *lac* operon.

Through a long evolutionary process, the existing *lac* system has been selected to operate for the optimal energy efficiency of the bacterial cell. Presumably to maximize energy efficiency, two environmental conditions have to be satisfied for the lactose metabolic enzymes to be expressed.

One condition is that lactose must be present in the environment. This condition makes sense because it would be inefficient for the cell to produce the lactose metabolic enzymes if there is no lactose to metabolize. We have already seen that the cell is able to respond to the presence of lactose through the action of a repressor protein.

The other condition is that glucose cannot be present in the cell's environment. Because the cell can capture more energy from the breakdown of glucose than it can from the breakdown of other sugars, it is more efficient for the cell to metabolize glucose rather than lactose. Thus, mechanisms have evolved that prevent the cell from synthesizing the enzymes for lactose metabolism when both lactose and glucose are present together. The repression of the transcription of lactose-metabolizing genes in the presence of glucose is an example of **catabolite repression** (glucose is a breakdown product, or a **catabolite**, of lactose). The transcription of genes encoding proteins necessary for the metabolism of many different sugars is similarly repressed in the presence of glucose. We will see that catabolite repression works through an *activator protein*.

The basics of *lac* catabolite repression: choosing the best sugar to metabolize

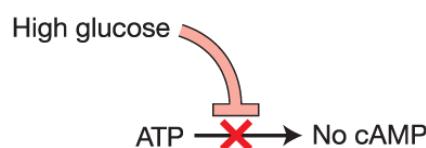
If both lactose and glucose are present, the synthesis of β -galactosidase is not induced until all the glucose has been metabolized. Thus, the cell conserves its energy by metabolizing any existing glucose before going through the energy-expensive process of creating new machinery to metabolize lactose. There are multiple mechanisms that bacteria have evolved to ensure the

preferential use of a carbon source and optimal growth. One mechanism is to exclude lactose from the cell. A second mechanism is to regulate operon expression via catabolites.

The results of studies indicate that a breakdown product of glucose prevents activation of the *lac* operon by lactose—the catabolite repression just mentioned. The glucose breakdown product is known to modulate the level of an important cellular constituent—cyclic adenosine monophosphate (cAMP), which is synthesized from the major energy source within the cell: adenosine triphosphate (ATP). When glucose is present in high concentrations, it inhibits the conversion of ATP to cAMP, so the cell's cAMP concentration is low. As the glucose concentration decreases, the conversion of ATP to cAMP is no longer inhibited, and the cell's concentration of cAMP increases correspondingly ([Figure 11-13a](#)). A high concentration of cAMP is necessary for activation of the *lac* operon. Mutants that cannot convert ATP into cAMP cannot be induced to produce β -galactosidase because the concentration of cAMP is not great enough to activate the *lac* operon.

Glucose levels control the *lac* operon

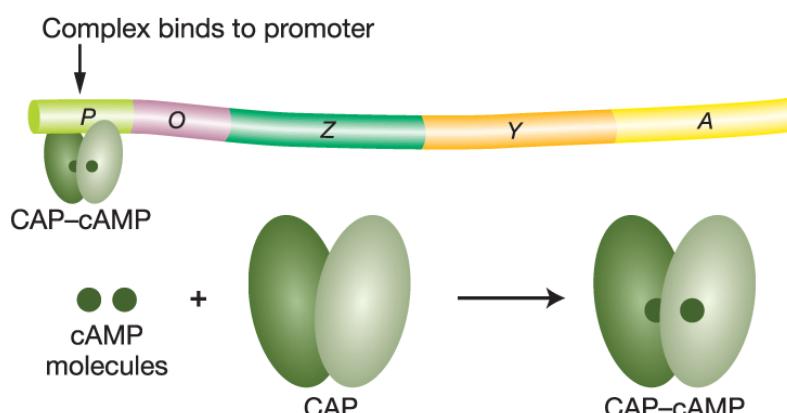
(a) Glucose levels regulate cAMP levels



Low glucose



(b) cAMP-CAP complex activates transcription



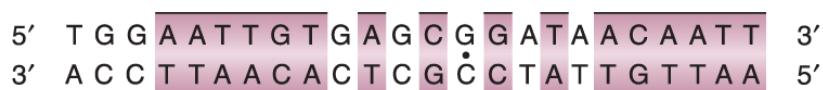
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

FIGURE 11-13 Catabolite control of the *lac* operon. (a) Only under conditions of low glucose is cAMP (cyclic adenosine monophosphate) formed from ATP. (b) When cAMP is present, it forms a complex with CAP (catabolite activator protein) that activates transcription by binding to a region within the *lac* promoter.

What is the role of cAMP in *lac* activation? A study of a different set of mutants provided an answer. These mutants make cAMP but cannot activate the Lac enzymes because they lack yet another protein, called **catabolite activator protein (CAP)**, encoded by the *crp* gene. CAP binds to a specific DNA sequence of the *lac* operon (the CAP-binding site; see [Figure 11-14b](#)). The DNA-bound CAP is then able to interact physically with RNA polymerase and increases that enzyme's affinity for the *lac* promoter. By itself, CAP cannot bind to the CAP-binding site of the *lac* operon. However, by binding to cAMP, its allosteric effector, CAP is able to bind to the CAP-binding site and activate transcription by RNA polymerase ([Figure 11-13b](#)). By inhibiting CAP when glucose is available, the catabolite-repression system ensures that the *lac* operon will be activated only when glucose is scarce.

Many DNA binding sites are symmetrical

(a) *lac* operator



(b) CAP-binding site



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

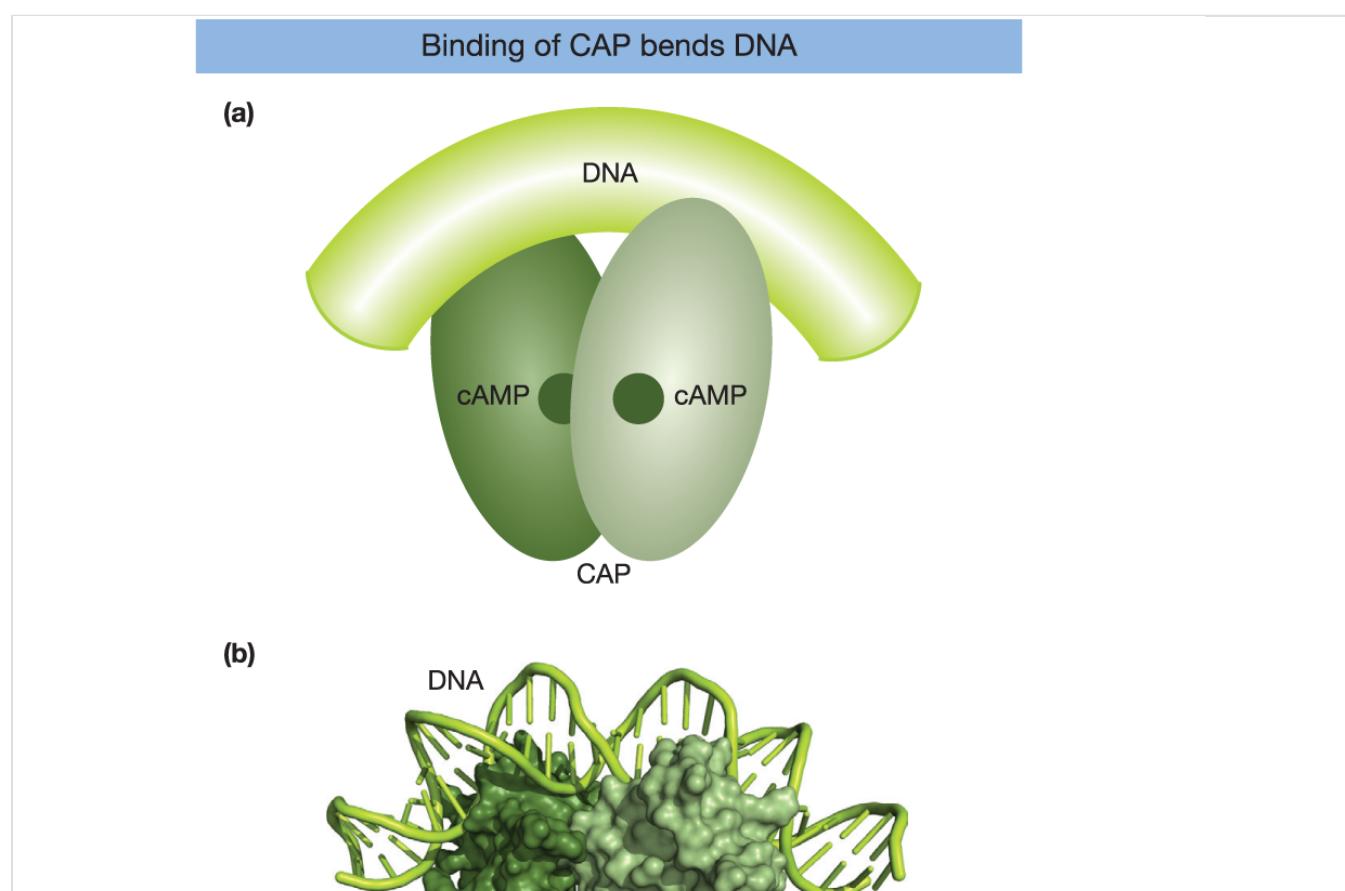
FIGURE 11-14 The DNA base sequences of (a) the *lac* operator, to which the Lac repressor binds, and (b) the CAP-binding site, to which the CAP–cAMP complex binds. Sequences exhibiting twofold rotational symmetry are indicated by the colored boxes and by a dot at the center point of symmetry. [(a) Data from W. Gilbert, A. Maxam, and A. Mirzabekov, in N. O. Kjeldgaard and O. Malløe, eds., *Control of Ribosome Synthesis*. Academic Press, 1976.]

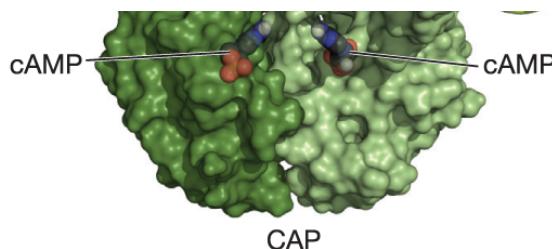
In **KEY CONCEPT** Operons that metabolize a nutrient, such as the *lac* operon, often have an added level of control so that the operon is inactive in the presence of its catabolic breakdown product (i.e., glucose) even if the nutrient (i.e., lactose) is present.

The structures of target DNA sites

The DNA sequences to which the CAP–cAMP complex binds (see [Figure 11-14](#)) are different from the sequences to which the Lac repressor binds. These differences underlie the specificity of DNA binding by these very different regulatory proteins. One property that these sequences do have in common and that is common to many other DNA-binding sites is rotational twofold symmetry. In other words, if we rotate the DNA sequence shown in [Figure 11-14](#) by 180 degrees within the plane of the page, the sequence of the highlighted bases of the binding sites will be identical. The highlighted bases are thought to constitute the important contact sites for protein–DNA interactions. This rotational symmetry corresponds to symmetries within the DNA-binding proteins, many of which are composed of two or four identical subunits. We will consider the structures of some DNA-binding proteins later in the chapter.

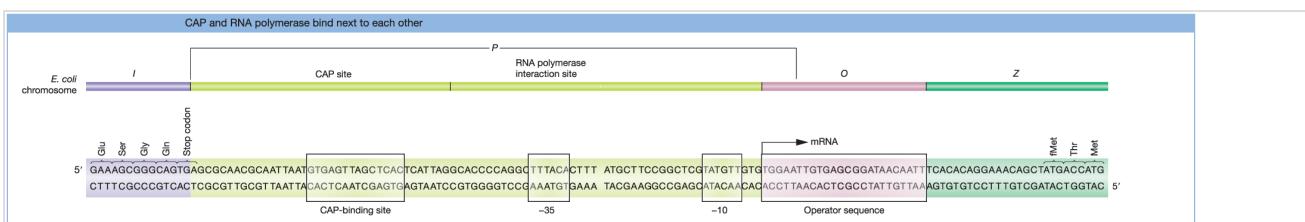
How does the binding of the cAMP–CAP complex to the operon further the binding of RNA polymerase to the *lac* promoter? In [Figure 11-15](#), the DNA is shown as being bent when CAP is bound. This bending of DNA may aid the binding of RNA polymerase to the promoter. There is also evidence that CAP makes direct contact with RNA polymerase. The base sequence shows that CAP and RNA polymerase bind directly adjacent to each other on the *lac* promoter ([Figure 11-16](#)).





Griffiths et al., *Introduction to Genetic Analysis*, 12e, ©
2020 W. H. Freeman and Company
(b) PDB ID 1cgp.

FIGURE 11-15 (a) When CAP binds the promoter, it creates a bend greater than 90 degrees in the DNA. (b) Image derived from the structural analysis of two subunits of CAP bound to the CAP-binding site.



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-16 The control region of the *lac* operon. The base sequence and the genetic boundaries of the control region of the *lac* operon, with partial sequences for the structural genes. Note that the *lac* operon promoter sequences at the -35 and the -10 sites differ from the consensus in [Figure 11-11](#). [Data from R. C. Dickson, J. Abelson, W. M. Barnes, and W. S. Reznikoff, “Genetic Regulation: The Lac Control Region,” *Science* 187, 1975, 27.]

KEY CONCEPT Generalizing from the *lac* operon model, regulatory proteins bind to DNA in the operator sites in the operons that they control. The exact pattern of binding in an operon will depend on physiological signals and whether activators or repressors regulate particular operons.

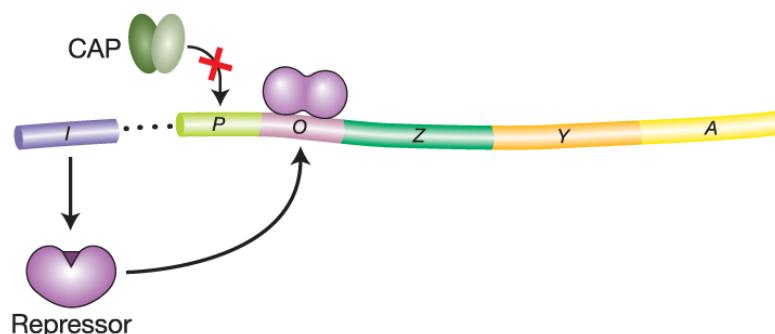
A summary of the *lac* operon

We can now fit the CAP–cAMP- and RNA-polymerase-binding sites into the detailed model of the *lac* operon, as shown in [Figure 11-17](#). The presence of glucose prevents lactose metabolism because a glucose breakdown product inhibits maintenance of the high cAMP levels necessary for formation of the CAP–cAMP complex, which in turn is required for the RNA polymerase to attach at the *lac* promoter site (see [Figure 11-17a, b](#)). Even when there is a shortage of glucose catabolites and CAP–cAMP forms, the mechanism for lactose metabolism will be implemented only if lactose is present (see [Figure 11-17c](#)). Only two or three molecules of β -galactosidase are present per cell in the absence of lactose or in the presence of lactose and glucose. These few molecules of β -galactosidase are likely due to a very low level of spurious transcription that

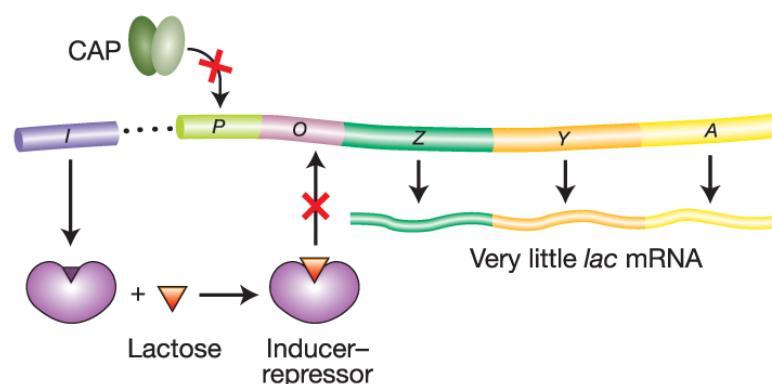
results because the repressor can briefly dissociate from the DNA. This increases to approximately 3000 molecules of enzyme when lactose is present and glucose is absent. Thus, the cell conserves its energy and resources by producing the lactose-metabolizing enzymes only when they are both needed and useful.

Negative and positive regulation of the *lac* operon

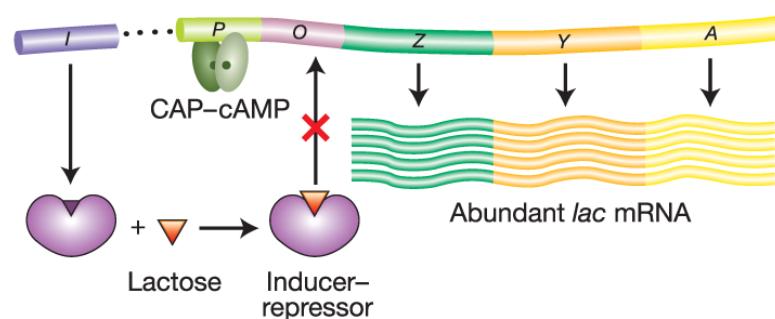
(a) Glucose present (cAMP low); no lactose; no *lac* mRNA



(b) Glucose present (cAMP low); lactose present



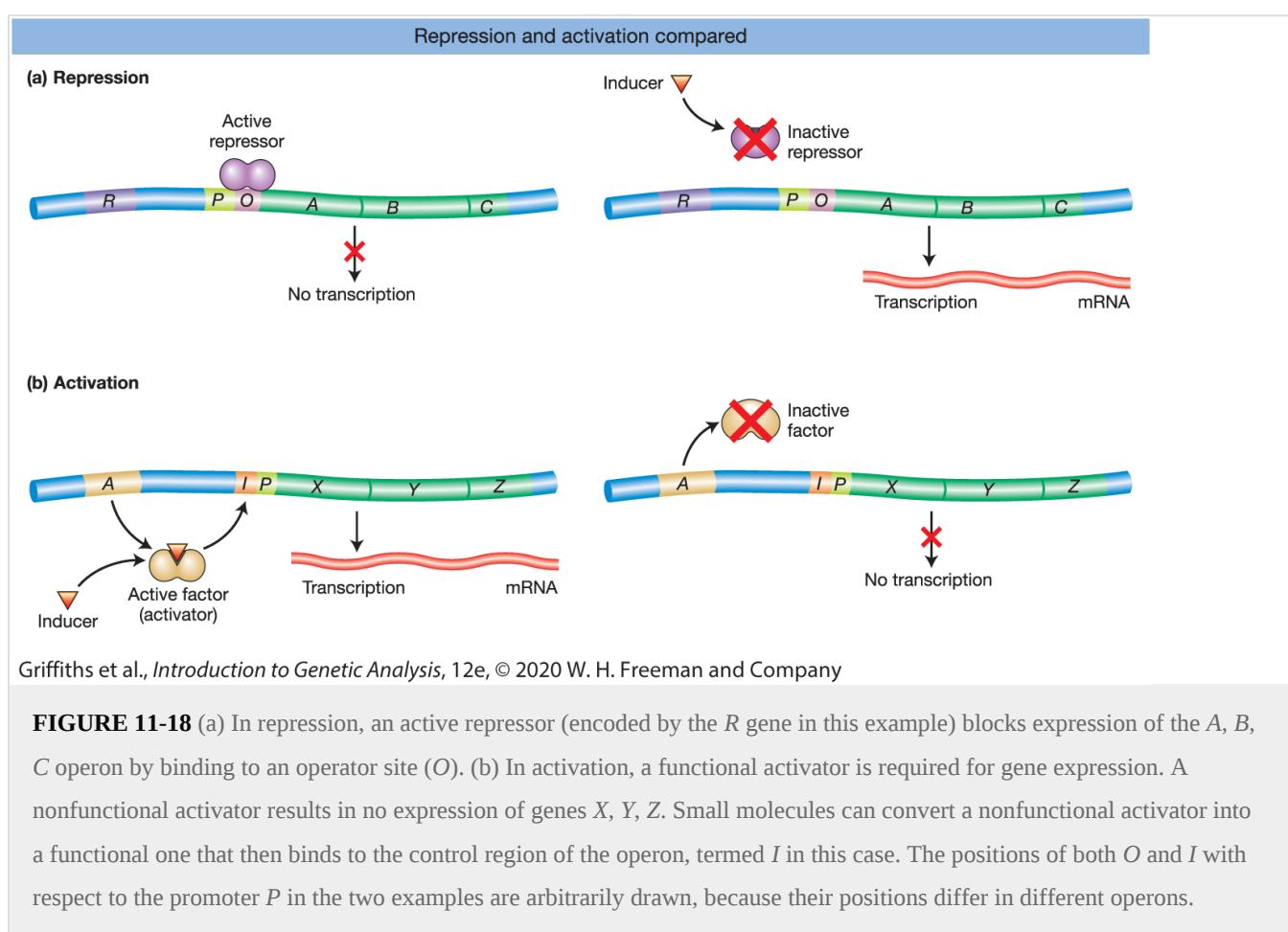
(c) No glucose present (cAMP high); lactose present



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-17 The *lac* operon is controlled jointly by the Lac repressor (negative regulator) and the catabolite activator protein (CAP; positive regulator). Large amounts of mRNA are produced only when lactose is present to inactivate the repressor, and low glucose levels promote the formation of the CAP-cAMP complex, which positively regulates transcription.

Inducer-repressor control of the *lac* operon is an example of repression, or negative regulation, in which expression is normally blocked. In contrast, the CAP-cAMP system is an example of activation, or positive regulation, because it acts as a signal that activates expression—in this case, the activating signal is the interaction of the CAP-cAMP complex with the CAP-binding site on DNA. **Figure 11-18** outlines these two basic types of control systems.



KEY CONCEPT Negative regulation promotes gene expression in the absence of the repressor, and positive regulation promotes gene expression in the presence of an activator.

11.4 DUAL POSITIVE AND NEGATIVE REGULATION: THE ARABINOSE OPERON

LO 11.3 Illustrate and compare the mechanisms that coordinate expression of sets of genes in bacteria and bacteriophage.

As with the *lac* system, the control of transcription in bacteria is neither purely positive nor purely negative; rather, both positive and negative regulation may govern individual operons. The regulation of the arabinose operon provides an example in which a single DNA-binding protein may act as *either* a repressor *or* an activator—a twist on the general theme of transcriptional regulation by DNA-binding proteins.

The structural genes *araB*, *araA*, and *araD* encode the metabolic enzymes that break down the sugar arabinose. The three genes are transcribed in a unit as a single mRNA. [Figure 11-19](#) shows a map of the *ara* operon. Transcription is activated at *araI*, the **initiator** region, which contains a binding site for an activator protein. The *araC* gene, which maps nearby, encodes an activator protein. When bound to arabinose, this protein binds to the *araI* site and activates transcription of the *ara* operon, perhaps by helping RNA polymerase bind to the promoter. In addition, the same CAP–cAMP catabolite repression system that prevents *lac* operon expression in the presence of glucose also prevents expression of the *ara* operon.

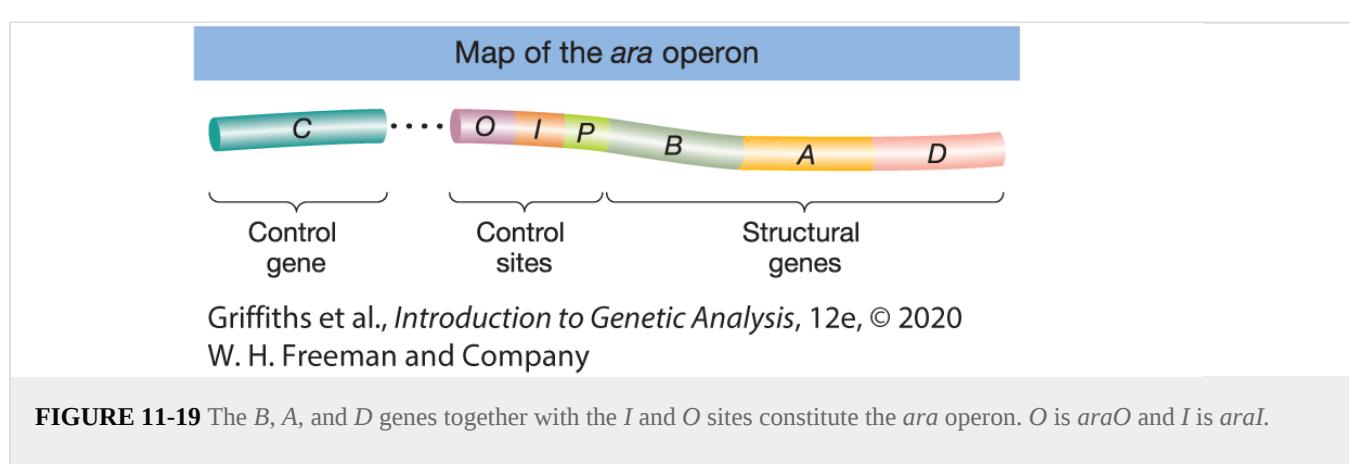


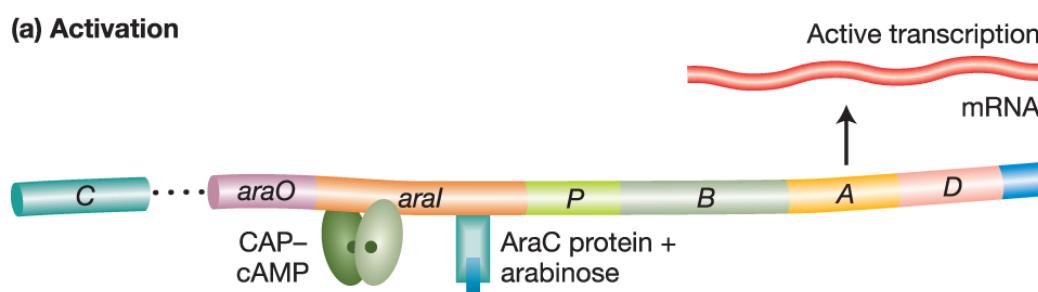
FIGURE 11-19 The *B*, *A*, and *D* genes together with the *I* and *O* sites constitute the *ara* operon. *O* is *araO* and *I* is *araI*.

In the presence of arabinose, both the CAP–cAMP complex and the AraC–arabinose complex must bind to *araI* in order for RNA polymerase to bind to the promoter and transcribe the *ara* operon ([Figure 11-20a](#)). In the absence of arabinose, the AraC protein assumes a different

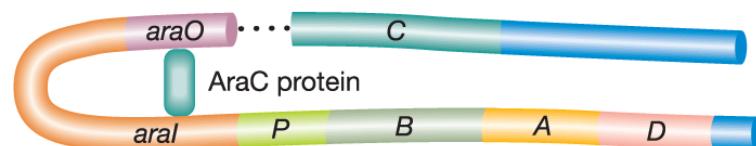
conformation and represses the *ara* operon by binding both to *araI* and to a second distant site, *araO*, thereby forming a loop (Figure 11-20b) that prevents transcription. Thus, the AraC protein has two conformations, one that acts as an activator and another that acts as a repressor. The on/off switch of the operon is “thrown” by arabinose. The two conformations, dependent on whether the allosteric effector arabinose has bound to the protein, differ in their abilities to bind a specific target site in the *araO* region of the operon.

AraC serves as an activator and as a repressor

(a) Activation



(b) Repression



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

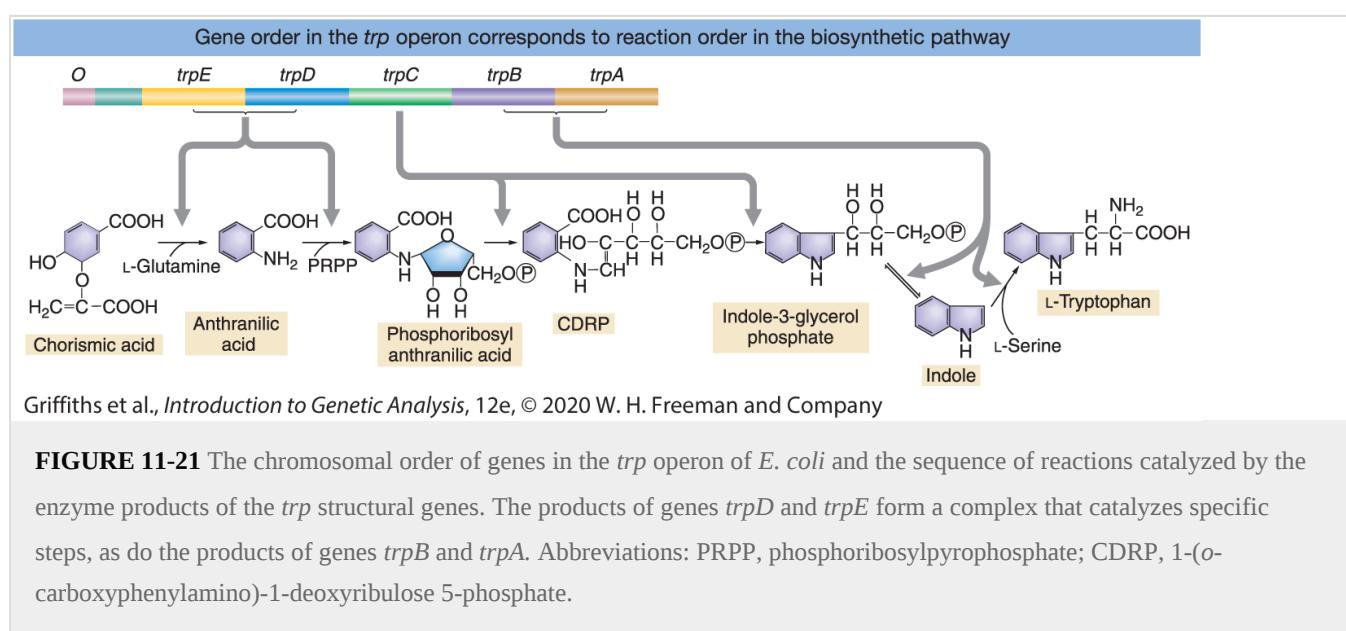
FIGURE 11-20 Dual control of the *ara* operon. (a) In the presence of arabinose, the AraC protein binds to the *araI* region. The CAP–cAMP complex binds to a site adjacent to *araI*. This binding stimulates the transcription of the *araB*, *araA*, and *araD* genes. (b) In the absence of arabinose, the AraC protein binds to both the *araI* and the *araO* regions, forming a DNA loop. This binding prevents transcription of the *ara* operon.

KEY CONCEPT Operon transcription is commonly regulated by both activation and repression. However, the specific mechanisms regulating the expression of operons that control the metabolism of similar compounds, such as sugars, can be quite different.

11.5 METABOLIC PATHWAYS AND ADDITIONAL LEVELS OF REGULATION: ATTENUATION

LO 11.3 Illustrate and compare the mechanisms that coordinate expression of sets of genes in bacteria and bacteriophage.

Coordinate control of genes in bacteria is widespread. Just as we saw in the preceding sections, there is a need for the cell to regulate pathways for the breakdown of specific sugars, depending upon the availability of that sugar. Similarly, the pathways that synthesize essential molecules, like amino acids, must be regulated so that the enzymes needed for their synthesis are produced by the bacteria only when amino acids are not available from the environment. In pathways that synthesize essential molecules, the genes that encode the enzymes are also organized into operons, complete with multigenic mRNAs. Furthermore, in cases for which the sequence of catalytic activity is known, there is a remarkable congruence between the order of operon genes on the chromosome and the order in which their products act in the metabolic pathway. This congruence is strikingly illustrated by the organization of the tryptophan operon in *E. coli* ([Figure 11-21](#)). The tryptophan operon contains five genes (*trpE*, *trpD*, *trpC*, *trpB*, *trpA*) that encode enzymes that contribute to the synthesis of the amino acid tryptophan.



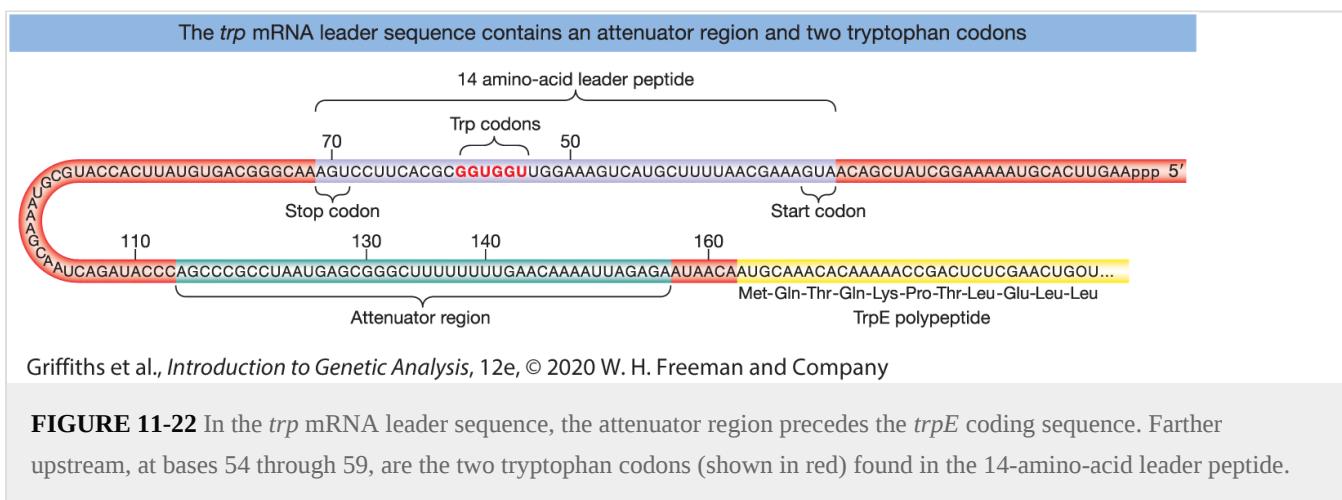
KEY CONCEPT In bacteria, genes that encode enzymes that are in the same metabolic pathways are generally organized into operons.

There are two mechanisms for regulating transcription of the tryptophan operon and some other operons functioning in amino acid biosynthesis. One provides global control of operon mRNA expression, and the other provides fine-tuned control.

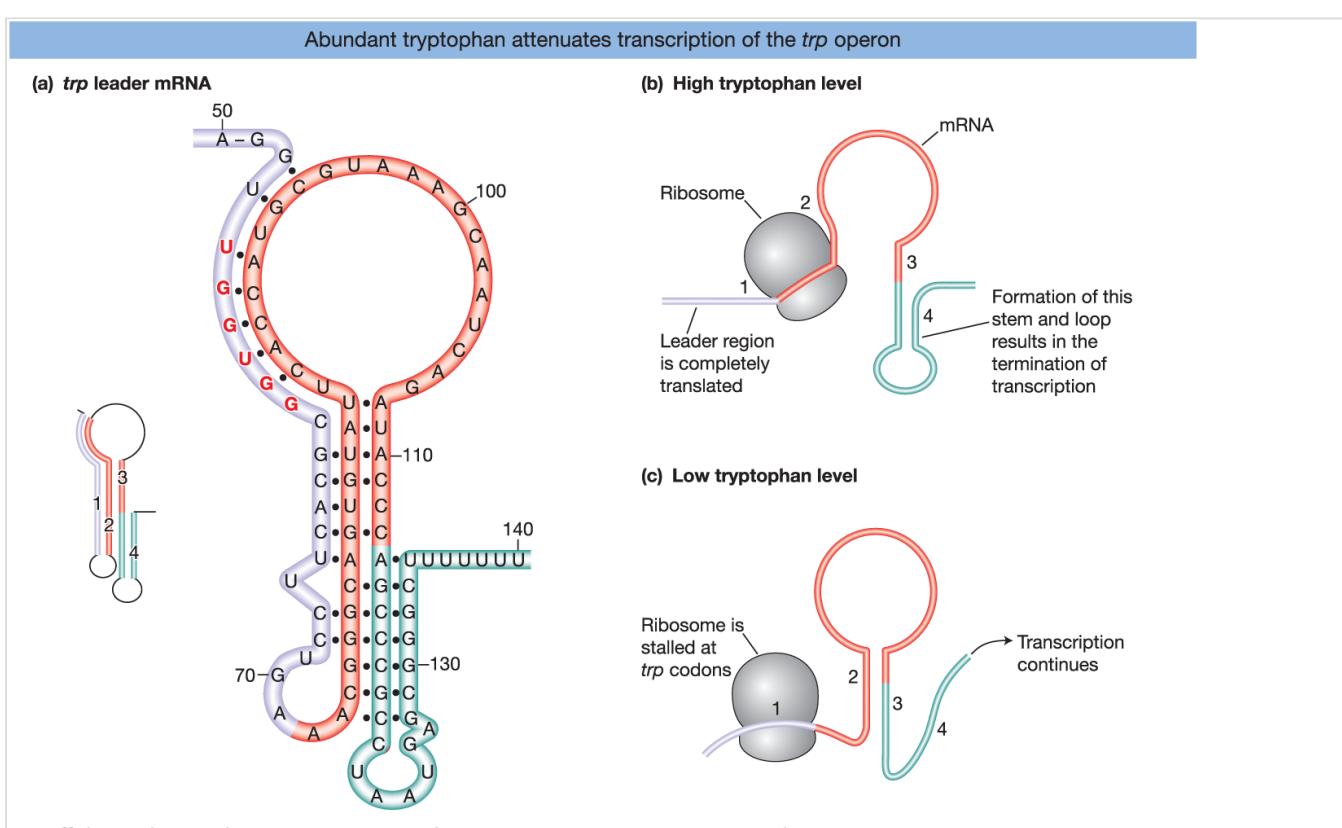
The level of *trp* operon gene expression is governed by the level of tryptophan. When tryptophan is absent from the growth medium, *trp* gene expression is high; when levels of tryptophan are high, the *trp* operon is repressed. One mechanism for controlling the transcription of the *trp* operon is similar to the mechanism of negative regulation that we have already seen controls the *lac* operon: a repressor protein binds an operator, preventing the initiation of transcription. This repressor is the Trp repressor, the product of the *trpR* gene. The Trp repressor binds tryptophan when adequate levels of the amino acid are present, and only after binding tryptophan will the Trp repressor bind to the operator and switch off transcription of the operon. This simple mechanism ensures that the cell does not waste energy producing tryptophan when the amino acid is sufficiently abundant in the environment. *E. coli* strains with mutations in *trpR* continue to express the *trp* mRNA and thus continue to produce tryptophan when the amino acid is abundant.

In studying these *trpR* mutant strains, Charles Yanofsky discovered that, when tryptophan was removed from the medium, the production of *trp* mRNA further increased several-fold. This finding was evidence that, in addition to the Trp repressor, a second control mechanism existed to negatively regulate transcription. This mechanism is called **attenuation** because mRNA production is normally *attenuated*, meaning “decreased,” when tryptophan is plentiful. Unlike the other bacterial control mechanisms described thus far, attenuation acts at a step *after* transcription initiation.

The mechanisms governing attenuation were discovered by identifying mutations that reduced or abolished attenuation. Strains with these mutations produce *trp* mRNA at maximal levels even in the presence of tryptophan. Yanofsky mapped the mutations to a region between the *trp* operator and the *trpE* gene; this region, termed the **leader sequence**, is at the 5' end of the *trp* operon mRNA before the first codon of the *trpE* gene (**Figure 11-22**). The *trp* leader sequence is unusually long for a bacterial mRNA, 160 bases, and detailed analyses have revealed how a part of this sequence works as an **attenuator** that governs the further transcription of *trp* mRNA.



The key observations are that, in the absence of the TrpR repressor protein, the presence of tryptophan halts transcription after the first 140 bases or so, whereas, in the absence of tryptophan, transcription of the operon continues. The mechanism for terminating or continuing transcription consists of two key elements. First, the *trp* mRNA leader sequence encodes a short, 14-amino-acid peptide that includes two adjacent tryptophan codons. Tryptophan is one of the least abundant amino acids in proteins, and it is encoded by a single codon. This pair of tryptophan codons is therefore an unusual feature. Second, the *trp* mRNA leader sequence consists of four segments that form stem-and-loop RNA structures that are able to alternate between two conformations. One of these conformations favors the termination of transcription, while the other favors the continuation of transcription ([Figure 11-23](#)).



Griffiths et al., Introduction to Genetic Analysis, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-23 (a) Proposed secondary structures in the conformation of *trp* leader mRNA that favors termination of transcription. Four regions can base pair to form three stem-and-loop structures, but only two regions base pair with one another at a given time. (b) When tryptophan is abundant, segment 1 of the *trp* mRNA is translated. Segment 2 enters the ribosome, enabling segments 3 and 4 to base pair to form a stem-loop that causes RNA polymerase to terminate transcription. (c) In contrast, when tryptophan is scarce, the ribosome is stalled at the codons of segment 1. Thus, segment 2 can interact with segment 3, and so segments 3 and 4 cannot pair. Consequently, transcription continues. [Data from D. L. Oxender, G. Zurawski, and C. Yanofsky, Proc. Natl. Acad. Sci. USA 76, 1979, 5524.]

The regulatory logic of the operon pivots on the abundance of tryptophan. When tryptophan is abundant, there is a sufficient supply of aminoacyl-tRNA^{Trp} to allow translation of the 14-amino-acid leader peptide. Recall that transcription and translation in bacteria are coupled; so ribosomes can engage mRNA transcripts and initiate translation before transcription is complete. The engagement of the ribosome alters *trp* mRNA conformation to the form that favors termination of transcription. Because segment 1 of the *trp* leader mRNA is translated in the presence of tryptophan, segment 2 of the leader mRNA will enter the ribosome. This allows base pairing between the attenuator region found in segments 3 and 4 of the leader mRNA, which leads to the termination of transcription by RNA polymerase (Figure 11-23b). However, when tryptophan is scarce, the ribosome is stalled at the tryptophan codons in segment 1, such that segments 2 and 3 base pair, and transcription is able to continue (Figure 11-23c). This mechanism is exquisitely sensitive to the levels of tryptophan in the environment because the number of transcripts produced will be directly related to the number of stalled ribosomes, which in turn is related to the amount of tryptophan in the cell. Thus, attenuation provides a way for the bacteria to rapidly fine-tune the synthesis of tryptophan depending upon environmental conditions.

Other operons for enzymes in biosynthetic pathways have similar attenuation controls. One signature of amino acid biosynthesis operons is the presence of multiple codons for the amino acid being synthesized in a separate peptide encoded by the 5' leader sequence. For instance, the *phe* operon has seven phenylalanine codons in a leader peptide and the *his* operon has seven tandem histidine codons in its leader peptide (Figure 11-24).

Leader peptides of amino acid biosynthesis operons

(a) <i>trp</i> operon	Met - Lys - Ala - Ile - Phe - Val - Leu - Lys - Gly - Trp - Trp - Arg - Thr - Ser - Stop
	5' AUG - AAA - GCA - AUU - UUC - GUA - CUG - AAA - GGU - UGG - UGG - CGC - ACU - UCC - UGA 3'
(b) <i>phe</i> operon	Met - Lys - His - Ile - Pro - Phe - Phe - Phe - Ala - Phe - Phe - Phe - Thr - Phe - Pro - Stop
	5' AUG - AAA - CAC - AUA - CCG - UUU - UUU - UUU - UUC - GCA - UUC - UUU - UUU - ACC - UUC - CCC - UGA 3'
(c) <i>his</i> operon	Met - Thr - Arg - Val - Gln - Phe - Lys - His - His - His - His - His - His - Pro - Asp
	5' AUG - ACA - CGC - GUU - CAA - UUU - AAA - CAC - CAC - CAU - CAU - CAC - CAU - CCU - GAC 3'

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-24 (a) The translated part of the *trp* leader region contains two consecutive tryptophan codons, (b) the *phe* leader sequence contains seven phenylalanine codons, and (c) the *his* leader sequence contains seven consecutive histidine codons.

KEY CONCEPT A second level of regulation in operons that control amino acid biosynthesis is attenuation of transcription mediated by the abundance of the amino acid and translation of a leader peptide.

11.6 BACTERIOPHAGE LIFE CYCLES: MORE REGULATORS, COMPLEX OPERONS

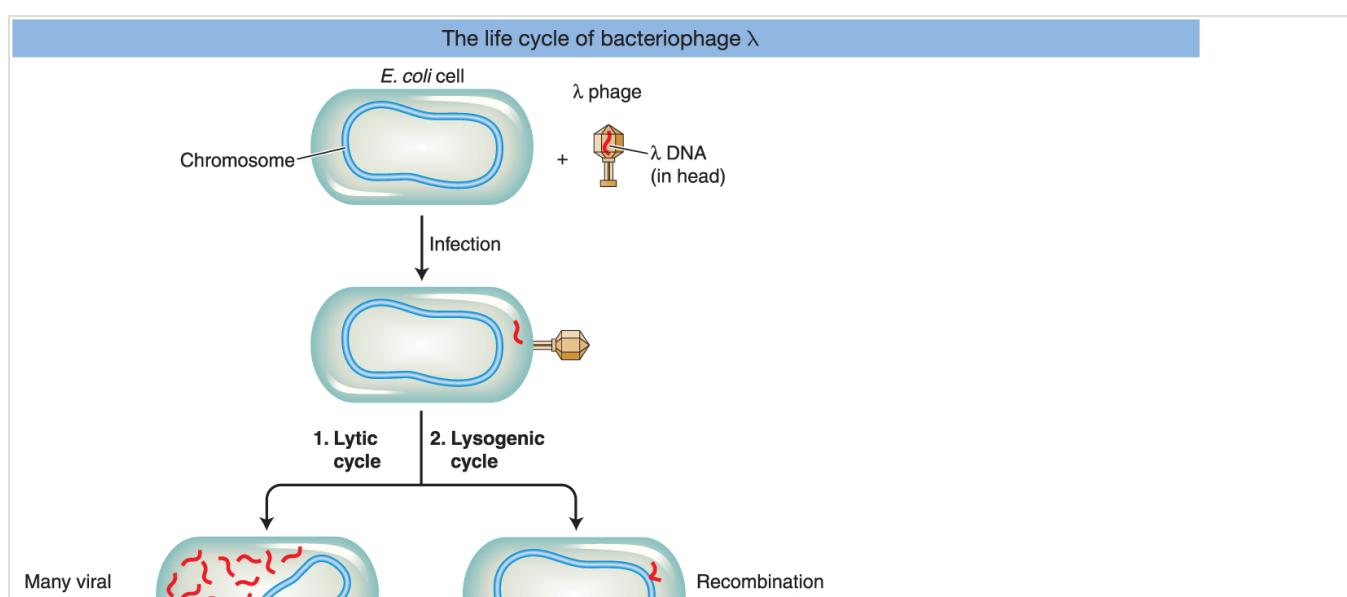
LO 11.3 Illustrate and compare the mechanisms that coordinate expression of sets of genes in bacteria and bacteriophage.

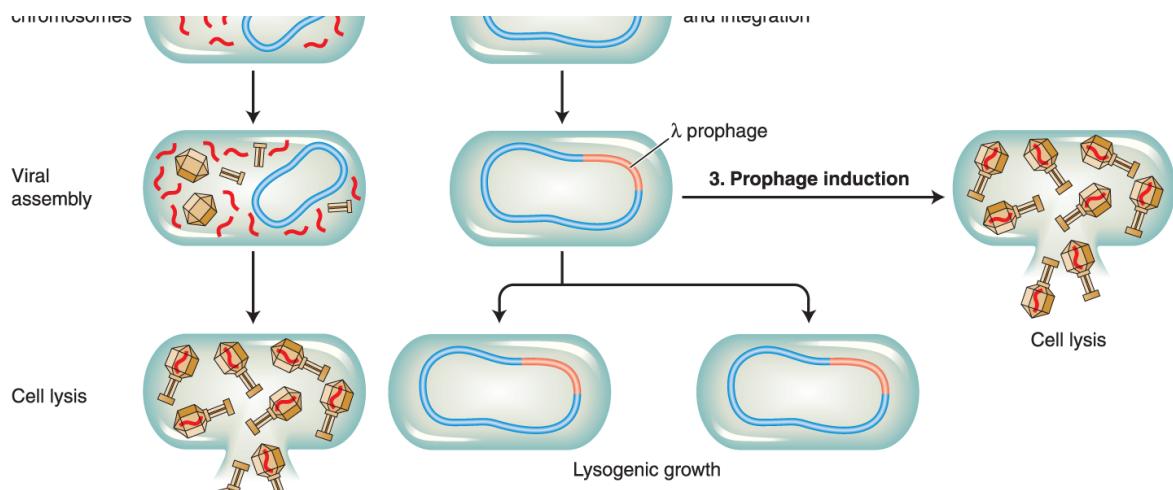
LO 11.4 Explain the roles of sequence-specific DNA-binding proteins and DNA regulatory sequences in coordinating the expression of sets of genes in bacteria and bacteriophage.

In that Paris movie theater, François Jacob had a flash of insight that the phenomenon of prophage induction might be closely analogous to the induction of β -galactosidase synthesis. He was right. Here, we are going to see how the life cycle of the bacteriophage λ is regulated. Although its regulation is more complex than that of individual operons, it is controlled by now-familiar modes of gene regulation.

Regulation of the bacteriophage # life cycle

Bacteriophage λ is a so-called temperate phage that has two alternative life cycles ([Figure 11-25](#)). When a normal bacterium is infected by a wild-type λ phage, two possible outcomes may follow: (1) the phage may replicate and eventually lyse the cell (the [lytic cycle](#)) or (2) the phage genome may be integrated into the bacterial chromosome as an inert prophage (the [lysogenic cycle](#)). In the lytic state, most of the phage's 71 genes are expressed at some point, whereas in the lysogenic state, most genes are inactive.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

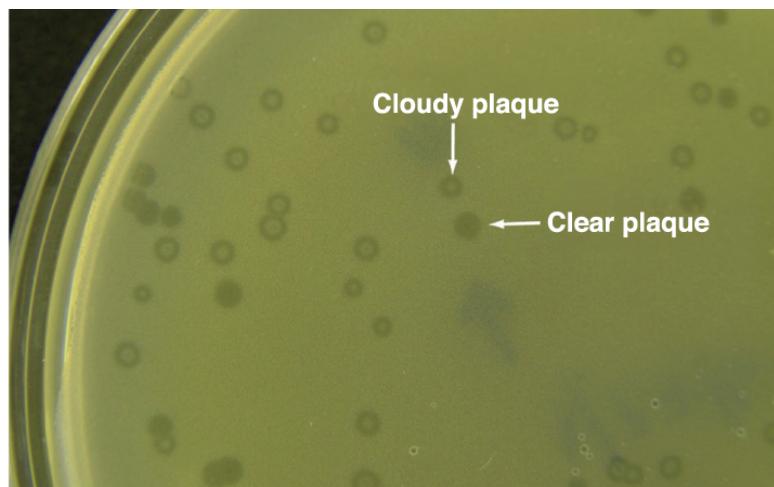
FIGURE 11-25 Whether bacteriophage λ enters the lytic cycle immediately or enters the lysogenic cycle depends on the availability of resources. The lysogenic virus inserts its genome into the bacterial chromosome, where it remains quiescent until conditions are favorable.

What decides which of these two pathways is taken? The physiological control of the decision between the lytic or lysogenic pathway depends on the resources available in the host bacterium. If resources are abundant, the lytic cycle is preferred because then there are sufficient nutrients to make many copies of the virus. If resources are limited, the lysogenic pathway is taken. The virus then remains present as a *prophage* until conditions improve. The inert prophage can be induced by ultraviolet light to enter the lytic cycle—the phenomenon studied by Jacob. The lytic and lysogenic states are characterized by very distinct programs of gene expression that must be regulated. Which alternative state is selected is determined by a complex genetic switch comprising several DNA-binding regulatory proteins and a set of operator sites.

Just as they were for the *lac* and other regulatory systems, genetic analyses of mutants were sources of crucial insights into the components and logic of the λ genetic switch. Jacob used simple phenotypic screens to isolate mutants that were defective in either the lytic or the lysogenic pathway. Mutants of each type could be recognized by the appearance of infected plaques on a lawn of bacteria. When wild-type phage particles are placed on a lawn of sensitive bacteria, clearings (called “plaques”) appear where bacteria are infected and lysed, but these plaques are cloudy because bacteria that are lysogenized grow within them ([Figure 11-26](#)). Mutant phages that are unable to lysogenize cells form clear plaques.

Clear and cloudy bacteriophage plaques on a lawn of *E. coli* host bacteria





From Microbiology An Evolving Science 1e, Figure 10.22.
©John Foster.

FIGURE 11-26 Plaques are clear where host cell lysis has occurred; they are cloudy where cells have survived infection and continued to grow as a lysogen. [From Microbiology: An Evolving Science 1e, [Figure 10.22](#). © John Foster.]

Such *clear* mutants (designated by *c*) turn out to be analogous to the *I* and *O* mutants of the *lac* system. These mutants were often isolated as temperature-sensitive mutants that had *clear* phenotypes at higher temperatures but wild-type phenotypes at lower temperatures. Three classes of mutants led to the identification of the key regulatory features of phage λ . In the first class, mutants for the *cI*, *cII*, and *cIII* genes form clear plaques; that is, they are unable to establish lysogeny. Mutants in the second class were isolated that do not lysogenize cells but can replicate and enter the lytic cycle in a lysogenized cell. These mutants turn out to be analogous to the operator-constitutive mutants of the *lac* system. A third key mutant can lysogenize but is unable to lyse cells. The mutated gene in this case is the *cro* gene (for control of repressor and other things). The decision between the lytic and the lysogenic pathways hinges on the activity of the proteins encoded by the four genes *cI*, *cII*, *cIII*, and *cro*, three of which are DNA-binding proteins.

KEY CONCEPT Wild-type λ phage can induce the lysogenic cycle in *E. coli*, resulting in cloudy plaques. Because clear plaques indicate that only the lytic cycle has occurred, a genetic screen for clear plaques can be used to identify mutations in genes required for the lysogenic cycle of λ phage.

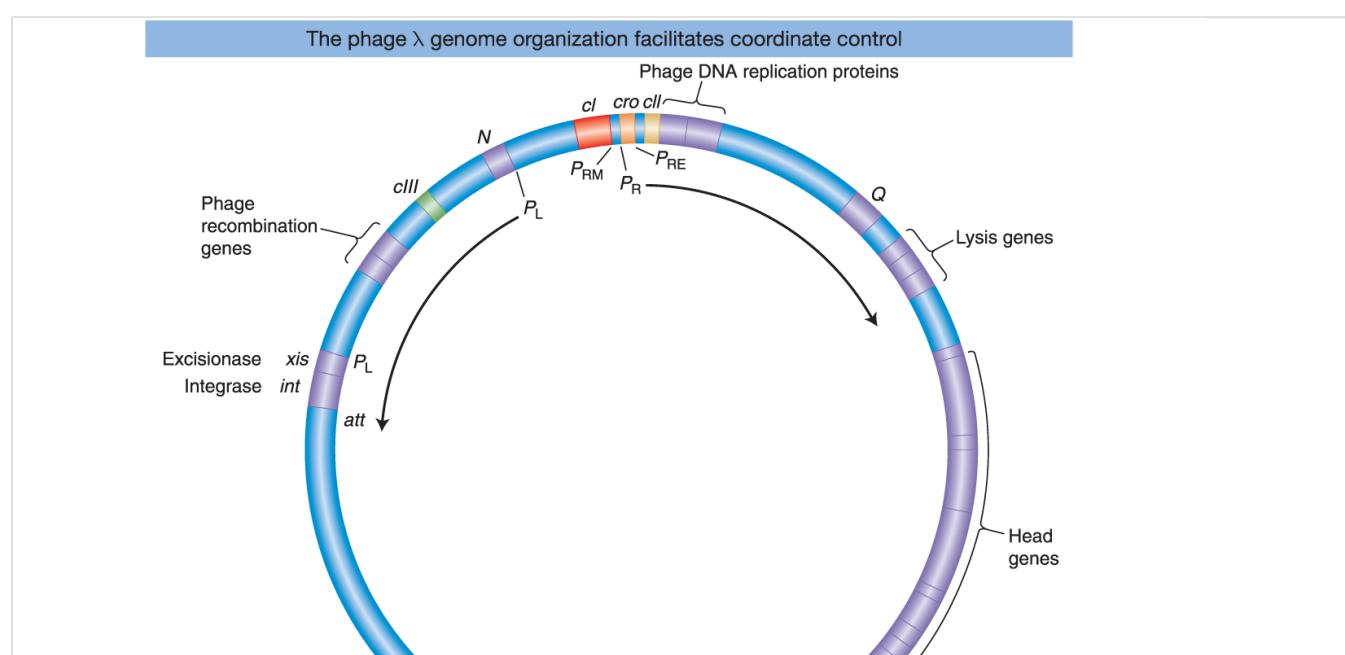
We will first focus on the two genes *cI* and *cro* and the proteins that they encode ([Table 11-4](#)). The *cI* gene encodes a repressor, often referred to as λ repressor, which represses lytic growth and promotes lysogeny. The *cro* gene encodes a repressor that represses lysogeny, thereby permitting lytic growth. The genetic switch controlling the two λ phage life cycles has two states: in the lysogenic state, *cI* is on but *cro* is off, and in the lytic cycle, *cro* is on but *cI* is off. Therefore, λ

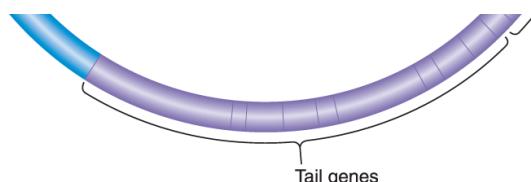
repressor and Cro are in competition, and whichever repressor prevails determines the state of the switch and of the expression of the λ genome.

TABLE 11-4 Major Regulators of Bacteriophage λ Life Cycle

Gene	Protein	Promotes
<i>cI</i>	λ repressor	lysogenic pathway
<i>cro</i>	Cro repressor	lytic pathway
<i>N</i>	positive regulator	cII, cIII expression
<i>cII</i>	activator	cI expression
<i>cIII</i>	protease inhibitor	cII activity

The race between λ repressor and Cro is initiated when phage λ infects a normal bacterium. The sequence of events in the race is critically determined by the organization of genes in the λ genome and of promoters and operators between the *cI* and the *cro* genes. The roughly 50-kb λ genome encodes proteins having roles in DNA replication, recombination, assembly of the phage particle, and cell lysis (Figure 11-27). These proteins are expressed in a logical sequence such that copies of the genome are made first, these copies are then packaged into viral particles, and, finally, the host cell is lysed to release the virus and begin the infection of other host cells (see Figure 11-25). The order of viral gene expression flows from the initiation of transcription at two promoters, P_L and P_R (for leftward and rightward promoter with respect to the genetic map). On infection, RNA polymerase initiates transcription at both promoters. Looking at the genetic map (Figure 11-27), we see that from P_R , *cro* is the first gene transcribed, and from P_L , *N* is the first gene transcribed.



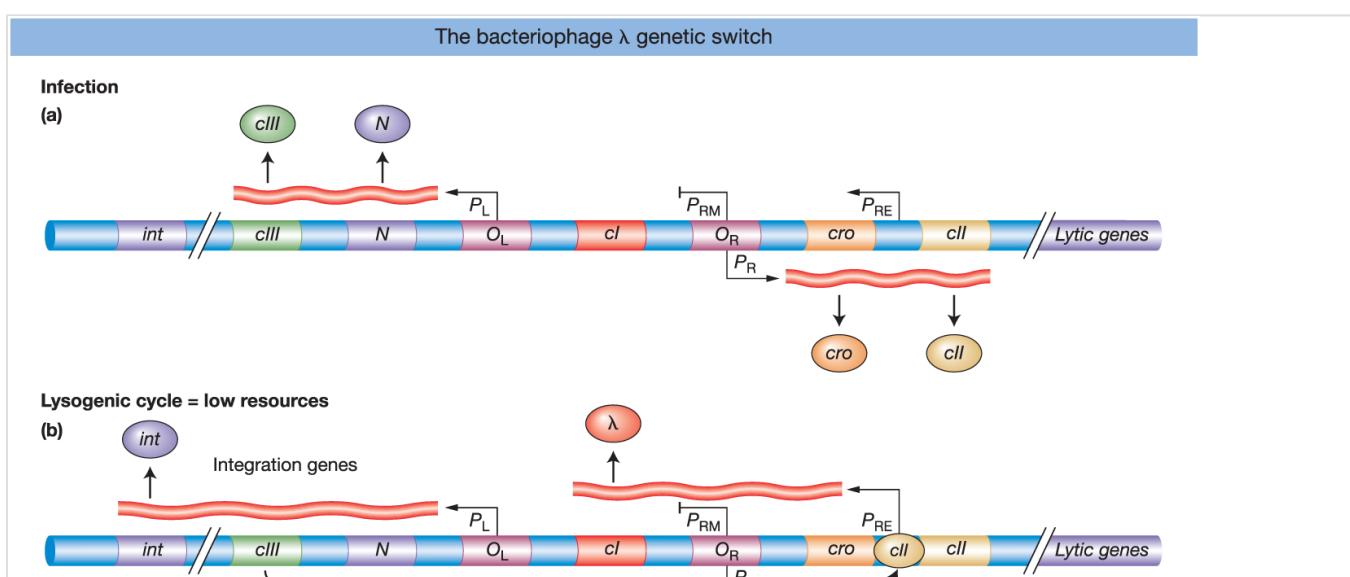


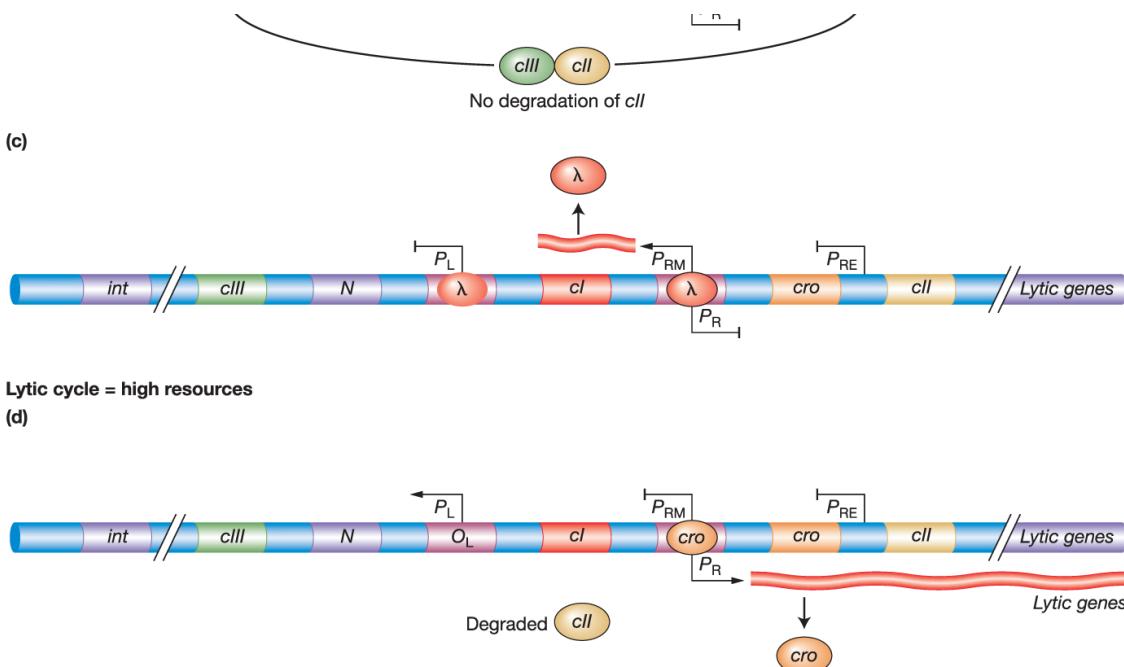
Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-27 Map of phage λ in the circular form. The genes for recombination, integration and excision, replication, head and tail assembly, and cell lysis are clustered together and coordinately regulated. Transcription of the right side of the genome begins at P_R , and that of the leftward genes begins at P_L . Key regulatory interactions governing the lysogenic-versus-lytic decision take place at operators between the *cro* and the *cI* genes.

The *N* gene encodes a positive regulator, but the mechanism of this protein differs from those of other regulators that we have considered thus far. Protein N works by enabling RNA polymerase to continue to transcribe through regions of DNA that would otherwise cause transcription to terminate. A regulatory protein such as N that acts by preventing transcription termination is called an **antiterminator**. Thus, *N* allows the transcription of *cIII* and other genes to the left of *N*, as well as *cII* and other genes to the right of *cro*. The *cII* gene encodes an activator protein that binds to a site that promotes transcription leftward from a different promoter, P_{RE} (for promoter of repressor establishment), which activates transcription of the *cI* gene. Recall that the *cI* gene encodes λ repressor, which will prevent lytic growth.

Before the expression of the rest of the viral genes takes place, a “decision” must be made—whether to continue with viral-gene expression and lyse the cell, or to repress that pathway and lysogenize the cell (**Figure 11-28**). The decision whether to lyse or lysogenize a cell pivots on the activity of the *cII* protein. The *cII* protein is unstable because it is sensitive to bacterial proteases—enzymes that degrade proteins. These proteases respond to environmental conditions: they are more active when resources are abundant, but less active when cells are starved.





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-28 On infection (a), host RNA polymerase initiates transcription at P_L and P_R , expressing the *N* and *cro* genes, respectively. Antiterminator protein *N* enables transcription of the *cIII* gene and recombination genes (see [Figure 11-27](#), left), and the *cII* gene and other genes. Next, (b) the *cII* protein, protected by the *cIII* protein, turns on *cI* by activating transcription at P_{RE} and also activates the transcription of *int*. If resources and proteases are not abundant, *cII* remains active, *cI* transcription proceeds at a high level, and the *Int* protein integrates the phage chromosome. Eventually (c), the *cI* protein (λ repressor) shuts off all genes except itself. The phage will then remain in the lysogenic state. However, if resources and proteases are abundant (d), the *cII* protein is degraded, *Cro* represses transcription of *cI* from P_{RM} and activates transcription of *Cro* and lytic genes from P_L and P_R , and the lytic cycle continues.

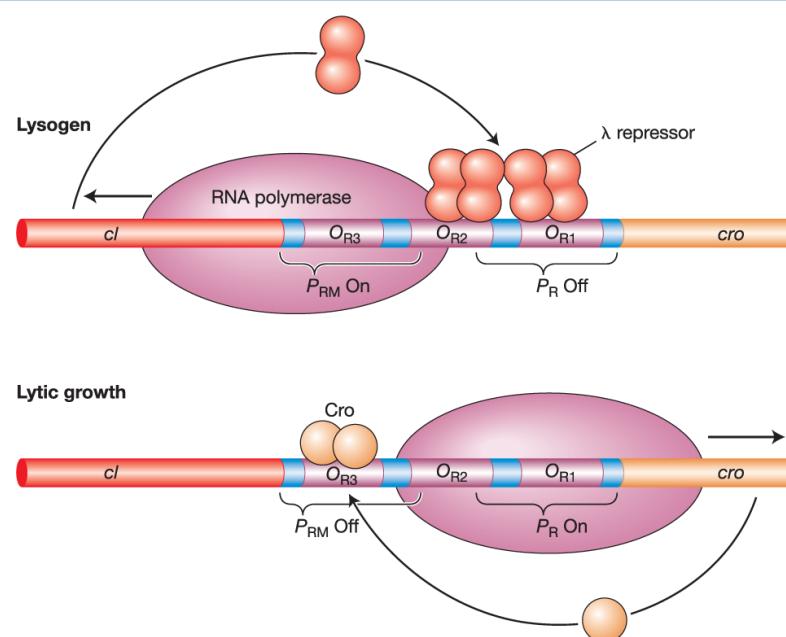
When resources are abundant, *cII* is degraded and little λ repressor is produced. The genes transcribed from P_L and P_R continue to be expressed, and the lytic cycle prevails. However, if resources are limited, *cII* is more active and more λ repressor is produced. In this case, the genes transcribed from P_L and P_R are repressed by the λ repressor and the lysogenic cycle is entered. The *cII* protein is also responsible for activating the transcription of *int*, a gene that encodes an additional protein required for lysogeny—an integrase required for the λ genome to integrate into the host chromosome. The *cIII* protein shields *cII* from degradation; so it, too, contributes to the lysogenic decision.

Molecular anatomy of the genetic switch

To see how the decision is executed at the molecular level, let's turn to the activities of λ repressor and Cro. The O_R operator lies between the two genes encoding these proteins and contains three

sites, O_{R1} , O_{R2} , and O_{R3} , that overlap two opposing promoters: P_R , which promotes transcription of lytic genes, and P_{RM} (for *repressor maintenance*), which directs transcription of the *cI* gene (see Figures 11-27 and 11-28). Recall that the *cI* gene encodes the λ repressor. The three operator sites are similar but not identical in sequence, and although Cro and λ repressor can each bind to any one of the operators, they do so with different affinities: λ repressor binds to O_{R1} with the highest affinity, whereas Cro binds to O_{R3} with the highest affinity. The λ repressor's occupation of O_{R1} blocks transcription from P_R and thus blocks the transcription of genes for the lytic cycle. Cro's occupation of O_{R3} blocks transcription from P_{RM} and thus blocks maintenance of *cI* transcription. Hence, no λ repressor is produced, and transcription of genes for the lytic cycle can continue. The occupation of the operator sites therefore determines the lytic-versus-lysogenic patterns of λ gene expression (Figure 11-29).

The lysogenic-versus-lytic cycle is determined by repressor occupancy on the O_R operators



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-29 Lysogeny is promoted by λ repressor binding to O_{R1} and O_{R2} , which prevents transcription from P_R . On induction or in the lytic cycle, the binding of Cro to O_{R3} prevents transcription of the *cI* gene from P_{RM} . [Data from M. Ptashne and A. Gann, *Genes and Signals*, p. 30, Fig. 1-13.]

After a lysogen has been established, it is generally stable. But the lysogen can be induced to enter the lytic cycle by various environmental changes. Ultraviolet light induces the expression of host genes. One of the host genes encodes a protein, RecA, that stimulates cleavage of the λ repressor, thus crippling maintenance of lysogeny and resulting in lytic growth. Prophage induction, just as Jacob and Monod surmised, requires the release of a repressor from DNA. The physiological role

of ultraviolet light in lysogen induction makes sense in that this type of radiation damages host DNA and stresses the bacteria; the phage replicates and leaves the damaged, stressed cell for another host.

KEY CONCEPT The phage λ genetic switch illustrates that the regulatory logic that underlies the control of gene expression in response to physiological signals is conserved. Just as in the *lac*, *ara*, *trp*, and other systems, the alternative states of gene expression in λ phage are determined by the interaction of a few key DNA-binding regulatory proteins with control sites on the DNA. The order and orientation of these genetic elements in the genome is important to the switch function in all of these systems.

Sequence-specific binding of regulatory proteins to DNA

How do λ repressor and Cro recognize different operators with different affinities? This question directs our attention to a fundamental principle in the control of gene transcription—the regulatory proteins bind to specific DNA sequences. For individual proteins to bind to certain sequences and not others requires specificity in the interactions between the side chains of the protein's amino acids and the chemical groups of DNA bases. Detailed structural studies of λ repressor, Cro, and other bacterial regulators have revealed how the three-dimensional structures of regulators and DNA interact, and how the arrangement of particular amino acids enables them to recognize specific base sequences.

Crystallographic analysis has identified a common structural feature of the DNA-binding domains of λ and Cro. Both proteins make contact with DNA through a *helix-turn-helix* domain that consists of two helices joined by a short flexible linker region ([Figure 11-30](#)). One helix, the recognition helix, fits into the major groove of DNA. In that position, amino acids on the helix's outer face are able to interact with chemical groups on the DNA bases. The specific amino acids in the recognition helix determine the affinity of a protein for a specific DNA sequence.

Helix-turn-helix is a common DNA-binding motif

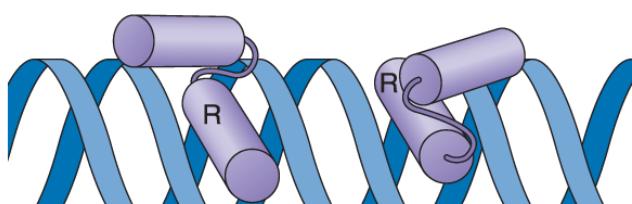
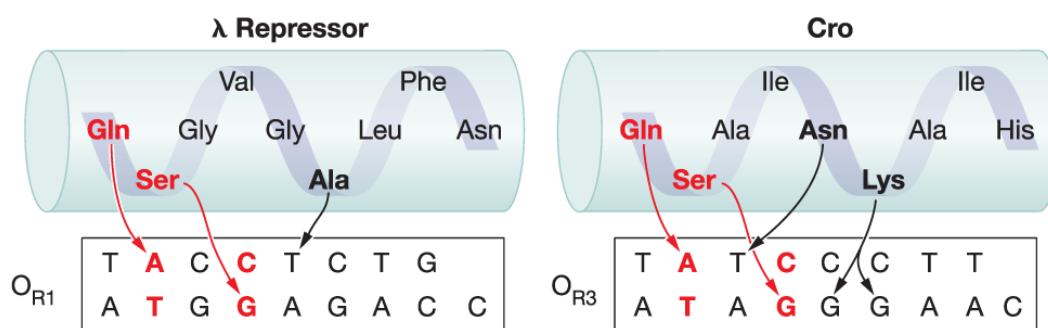




FIGURE 11-30 The binding of a helix-turn-helix motif to DNA. The purple cylinders are alpha helices. Many regulatory proteins bind as dimers to DNA. In each monomer, the recognition helix (R) makes contact with bases in the major groove of DNA.

The recognition helices of the λ repressor and Cro have similar structures and some identical amino acid residues. Differences between the helices in key amino acid residues determine their DNA-binding properties. For example, in the λ repressor and Cro proteins, glutamine and serine side chains contact the same bases, but an alanine residue in the λ repressor and lysine and asparagine residues in the Cro protein impart different binding affinities for sequences in O_{R1} and O_{R3} (Figure 11-31).

Amino acid side chains determine the specificity of DNA binding



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-31 Interactions between amino acids and bases determine the specificity and affinity of DNA-binding proteins. The amino acid sequences of the recognition helices of the λ repressor and Cro proteins are shown. Interactions between the glutamine (Gln), serine (Ser), and alanine (Ala) residues of the λ repressor and bases in the O_{R1} operator determine the strength of binding. Similarly, interactions between the glutamine, serine, asparagine (Asn), and lysine (Lys) residues of the Cro protein mediate binding to the O_{R3} operator. Each DNA sequence shown is that bound by an individual monomer of the respective repressor; it is half of the operator site occupied by the repressor dimer. [Data from M. Ptashne, *A Genetic Switch: Phage λ and Higher Organisms*, 2nd ed.]

The Lac and TrpR repressors, as well as the AraC activator and many other proteins, also bind to DNA through helix-turn-helix motifs of differing specificities, depending on the primary amino acid sequences of their recognition helices. In general, other domains of these proteins, such as those that bind their respective allosteric effectors, are dissimilar.

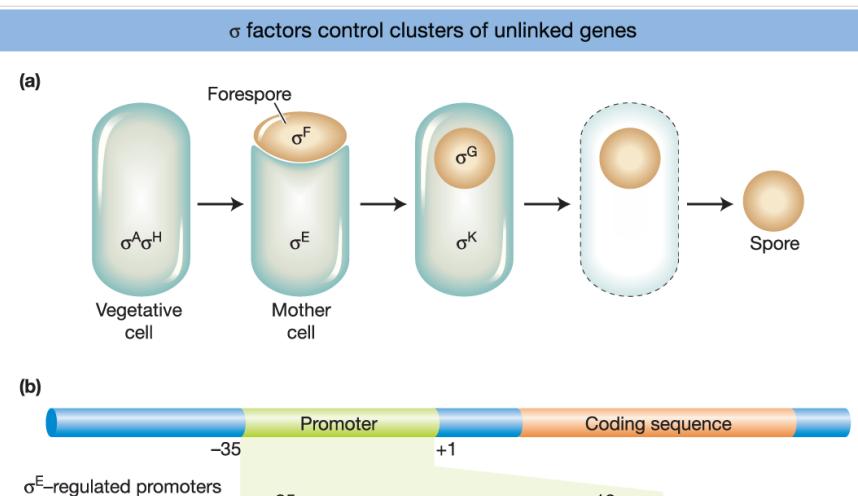
KEY CONCEPT The biological specificity of gene regulation is due to the chemical specificity of amino acid–base interactions between individual regulatory proteins and discrete DNA sequences.

11.7 ALTERNATIVE SIGMA FACTORS REGULATE LARGE SETS OF GENES

LO 11.4 Explain the roles of sequence-specific DNA-binding proteins and DNA regulatory sequences in coordinating the expression of sets of genes in bacteria and bacteriophage.

Thus far, we have seen how single switches can control the expression of single operons or two operons containing as many as a couple of dozen genes. Some physiological responses to changes in the environment require the coordinated expression of large sets of unlinked genes located throughout the genome to bring about dramatic physiological and even morphological changes. Analyses of these processes have revealed another twist in bacterial gene regulation: the control of large numbers of genes by alternative sigma (σ) factors of RNA polymerase. One such example, the process of sporulation in *Bacillus subtilis*, has been analyzed in great detail in the past few decades. Under stress, the bacterium forms spores that are remarkably resistant to heat and desiccation.

Early in the process of sporulation, the bacterium divides asymmetrically, generating two components of unequal size that have very different fates. The smaller compartment, the forespore, develops into the spore. The larger compartment, the mother cell, nurtures the developing spore and lyses when spore morphogenesis is complete to liberate the spore ([Figure 11-32a](#)). Genetic dissection of this process has entailed the isolation of many mutants that cannot sporulate. Detailed investigations have led to the characterization of several key regulatory proteins that directly regulate programs of gene expression that are specific to either the forespore or the mother cell. Four of these proteins are alternative σ factors.



<i>ybaN</i>	TCGG TTATATT CAATTGT-CCATGCT CATAAGAT ...
<i>ydcC</i>	GTCT GCATATT AGGGAAA-CCCACT CATATATT ...
<i>ydcA</i>	TACG TACTATT AAATGG-TTTTCT CATAAACG ...
 σ^F -regulated promoters	
<i>yrR</i>	ATCT GTTTA GCAGCGAAACACCTC GTCCACAATG ...
<i>ytT</i>	CCGG GTTTATT TTTTTTT-AGGAATT GGCGATAATG ...
<i>yuiC</i>	TTTT GAATA ATGCTCTCCACTT GGGAACAATG ...

Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020 W. H. Freeman and Company

FIGURE 11-32 Sporulation in *Bacillus subtilis* is regulated by cascades of σ factors. (a) In vegetative cells, σ^A and σ^H are active. On initiation of sporulation, σ^F is active in the forespore and σ^E is active in the mother cell. These σ factors are then superseded by σ^G and σ^K , respectively. The mother cell eventually lyses and releases the mature spore. (b) Factors σ^E and σ^F control the regulons of many genes (*ybaN*, and so forth, in this illustration). Three examples of the large number of promoters regulated by each σ factor are shown. Each σ factor has a distinct sequence-specific binding preference at the -35 and -10 sequences of target promoters. [Data from P. Eichenberger et al., J. Mol. Biol. 327, 2003, 945–972; and S. Wang et al., J. Mol. Biol. 358, 2006, 16–37.]

Recall from [Chapter 8](#) that transcription initiation in bacteria includes the binding of the σ subunit of RNA polymerase to the -35 and -11 regions of gene promoters. The σ factor dissociates from the complex when transcription begins and is recycled. In *B. subtilis*, two σ factors, σ^A and σ^H , are active in vegetative cells. During sporulation, a different σ factor, σ^F , becomes active in the forespore and activates a group of more than 40 genes. One gene activated by σ^F is a secreted protein that in turn triggers the proteolytic processing of the inactive precursor pro- σ^E , a distinct σ factor in the mother cell. The σ^E factor is required to activate sets of genes in the mother cell. Two additional σ factors, σ^K and σ^G , are subsequently activated in the mother cell and forespore, respectively ([Figure 11-32a](#)). The expression of distinct σ factors allows for the coordinated transcription of different sets of genes, or **regulons**, by a single RNA polymerase.

New approaches for characterizing the expression of all genes in a genome (see [Section 14.7](#)) have made it possible to monitor the transcription of each *B. subtilis* gene during vegetative growth and spore formation and in different compartments of the spore. Several hundred genes have been identified in this fashion that are transcriptionally activated or repressed during spore formation.

How are the different sets of genes controlled by each σ factor? Each σ factor has different sequence-specific DNA-binding properties. The operons or individual genes regulated by particular σ factors have characteristic sequences in the -35 and -11 regions of their promoters that are bound by one σ factor and not others ([Figure 11-32b](#)). For example, σ^E binds to at least 121 promoters, within 34 operons and 87 individual genes, to regulate more than 250 genes, and σ^F binds to at least 36 promoters to regulate 48 genes.

KEY CONCEPT Sequential expression of alternative σ factors that recognize alternative promoter sequences enables the coordinated expression of large numbers of independent operons and unlinked genes during the developmental program of sporulation.

Alternative σ factors also play important roles in the virulence of human pathogens. For example, bacteria of the genus *Clostridium* produce potent toxins that are responsible for severe diseases such as botulism, tetanus, and gangrene. Key toxin genes of *C. botulinum*, *C. tetani*, and *C. perfringens* have recently been discovered to be controlled by related, alternative σ factors that recognize similar sequences in the -35 and -10 regions of the toxin genes. Understanding the mechanisms of toxin-gene regulation may lead to new means of disease prevention and therapy.

SUMMARY

Gene regulation is often mediated by proteins that react to physiological signals from within and around the cell. The proteins respond by raising or lowering the transcription rates of specific genes. The logic of this regulation is straightforward. For regulation to operate appropriately, the regulatory proteins have built-in sensors that continually monitor cellular conditions. The activities of these proteins would then depend on the right set of environmental conditions.

In bacteria and their viruses, the control of several structural genes may be coordinated by clustering the genes together into operons on the chromosome so that they are transcribed into multigenic mRNAs. Coordinated control simplifies the task for bacteria because one cluster of regulatory sites per operon is sufficient to regulate the expression of all the operon's genes. Alternatively, coordinate control can also be achieved through discrete σ factors that regulate dozens of independent promoters simultaneously.

In negative regulatory control, a repressor protein blocks transcription by binding to DNA at the operator site. Negative regulatory control is exemplified by the *lac* system. Negative regulation is one very straightforward way for the *lac* system to shut down metabolic genes in the absence of appropriate sugars in the environment. In positive regulatory control, protein factors are required to activate transcription. Positive regulatory control is exemplified by repression of the *lac* system in the presence of its catabolite breakdown product, glucose. By contrast, repression of operons that synthesize amino acids is often controlled by attenuation.

Many regulatory proteins are members of families of proteins that have very similar DNA-binding motifs, such as the helix-turn-helix domain. Other parts of the proteins, such as their protein–protein interaction domains, tend to be less similar. The specificity of gene regulation depends on chemical interactions between the side chains of amino acids and chemical groups on DNA bases.

The mechanisms of the regulatory control of gene expression can be inferred from analyzing the physiological effects of genetic mutations.

KEY TERMS

activator

allosteric effector

allosteric site
allosteric transition
antiterminator
attenuation
attenuator
catabolite
catabolite activator protein (CAP)
catabolite repression
cis-acting
constitutive mutation
coordinately controlled genes
cyclic adenosine monophosphate (cAMP)
DNA-binding domain
genetic switch
inducer
induction
initiator
leader sequence
lysogenic cycle
lytic cycle
negative regulation
operator
operon
partial diploid
positive regulation
promoter
regulon
repressor
trans-acting

SOLVED PROBLEMS

This set of four solved problems, which are similar to Problem 15 in the Basic Problems at the end of this chapter, is designed to test understanding of the operon model. Here, we are given several diploids and are asked to determine whether Z and Y gene products are made in the presence or absence of an inducer. Use a table similar to the one in Problem 15 as a basis for your answers, except that the column headings will be as follows:

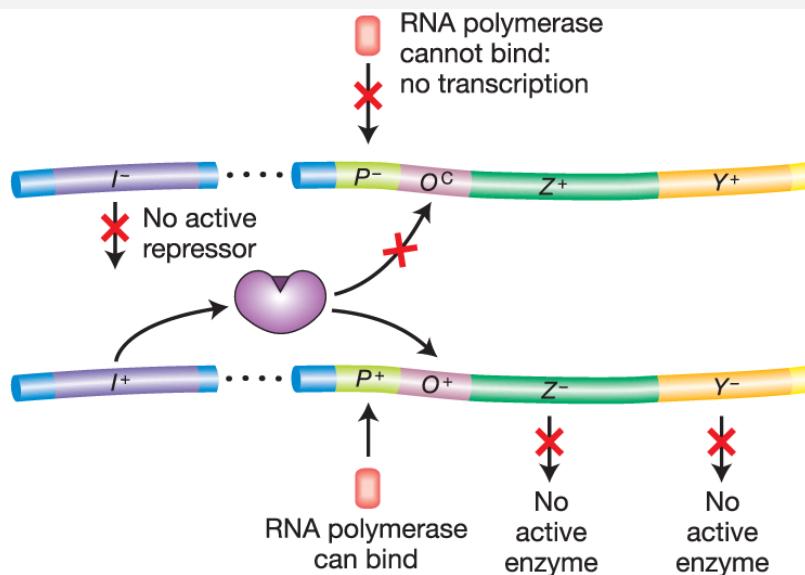
	Z gene		Y gene	
Genotype	No inducer	Inducer	No inducer	Inducer

SOLVED PROBLEM 1

$$\begin{array}{c} I^- P^- O^C Z^+ Y^+ \\ \hline I^- P^- O^- Z^+ Y^+ + I^+ P^+ O^+ Z^- Y^- \end{array}$$

SOLUTION

One way to approach these problems is first to consider each chromosome separately and then to construct a diagram. The following illustration diagrams this diploid:



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

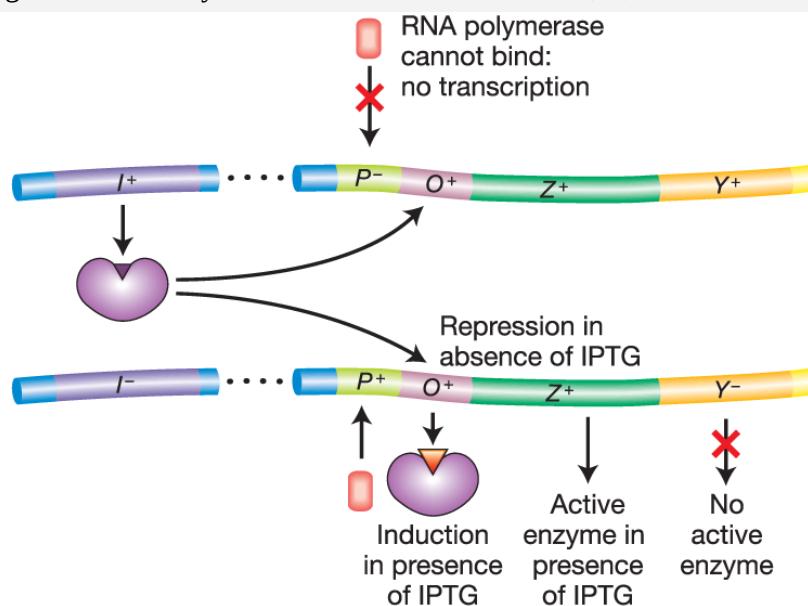
In The first chromosome is P^- , and so transcription is blocked and no Lac enzyme can be synthesized from it. The second chromosome (P^+) can be transcribed, and thus transcription is repressible (O^+). However, the structural genes linked to the good promoter are defective; thus, no active Z product or Y product can be generated. The symbols to add to your table are “-, -, -, -.”

SOLVED PROBLEM 2

$$\begin{array}{c} I^+ P^- O^+ Z^+ Y^+ \\ \hline I^+ P^- O^- Z^+ Y^+ + I^- P^+ O^+ Z^- Y^- \end{array}$$

SOLUTION

The first chromosome is P^- , and so no enzyme can be synthesized from it. The second chromosome is O^+ , and so transcription is repressed by the repressor supplied from the first chromosome, which can act in trans through the cytoplasm. However, only the Z gene from this chromosome is intact. Therefore, in the absence of an inducer, no enzyme is made; in the presence of an inducer, only the Z gene product, β -galactosidase, is generated. The symbols to add to the table are “ $-$, $+$, $-$, $-$.”



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

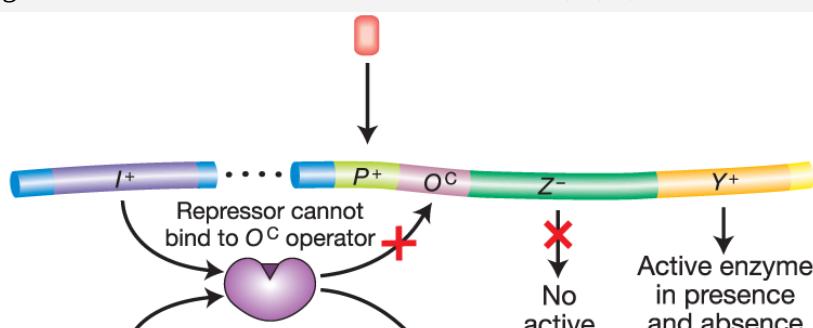
In

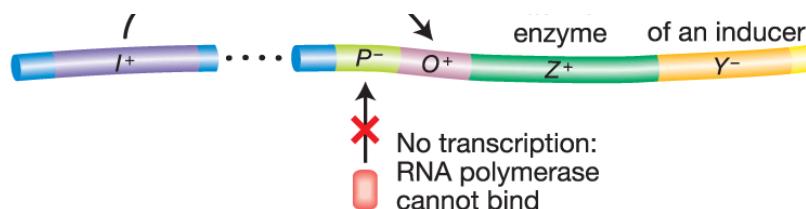
SOLVED PROBLEM 3

$$\frac{I^+ P^+ O^C Z^- Y^+}{I^+ P^+ O^C Z^+ Y^-}$$

SOLUTION

Because the second chromosome is P^- , we need consider only the first chromosome. This chromosome is O^C , and so enzyme is made in the absence of an inducer, although, because of the Z^- mutation, only active permease (Y) is generated. The entries in the table should be “ $-$, $-$, $+$, $+$.”





Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

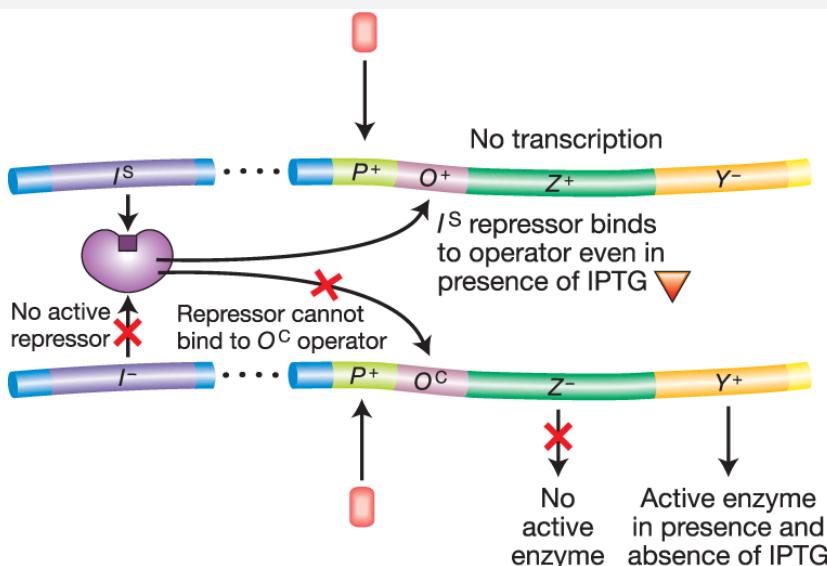
In

SOLVED PROBLEM 4

$$\frac{I^S P^+ O^+ Z^+ Y^-}{IS P^+ O^+ Z^+ Y^- I^- P^+ O^C Z^- Y^+}$$

SOLUTION

In the presence of an I^S repressor, all wild-type operators are shut off, both with and without an inducer. Therefore, the first chromosome is unable to produce any enzyme. However, the second chromosome has an altered (O^C) operator and can produce enzyme in both the absence and the presence of an inducer. Only the Y gene is wild type on the O^C chromosome, and so only permease is produced constitutively. The entries in the table should be “–, –, +, +.”



Griffiths et al., *Introduction to Genetic Analysis*, 12e, © 2020
W. H. Freeman and Company

In

PROBLEMS

Visit SaplingPlus for supplemental content. Problems with the  icon are available for review/grading.

Problems with the  icon have a Problem Solving Video. Problems with the  icon have an Unpacking the Problem exercise.

WORKING WITH THE FIGURES

(The first 10 questions require inspection of text figures.)

1. Based on [Figure 11-6](#), why does the binding of the repressor protein to the operator sequence in the absence of lactose prevent expression of the structural genes? Why does the absence of binding of the repressor protein to the operator sequence in the presence of lactose allow expression of the structural genes?
2. Compare the structure of IPTG shown in [Figure 11-7](#) with the structure of galactose shown in [Figure 11-5](#). Why is IPTG bound by the Lac repressor but not broken down by β -galactosidase?
3. Looking at [Figure 11-9](#), why were partial diploids essential for establishing the trans-acting nature of the Lac repressor? Could one distinguish cis-acting from trans-acting genes in haploids?
4. Why do promoter mutations cluster at positions -10 and -35 as shown in [Figure 11-11](#)? Which protein-DNA interaction is disrupted by these mutations?
5. Comparing [Figures 11-14](#), [11-15](#), and [11-30](#), why are many regulatory DNA-binding sites symmetrical?
6. Looking at [Figure 11-16](#), note the large overlap between the operator and the region of the *lac* operon that is transcribed. Which protein binds specifically to this overlapping sequence, and what effect does it have on transcription?
7. Looking at [Figure 11-20b](#), why do you think binding of the AraC protein to the *araO* and *araI* DNA sites leads to repression of the *ara* operon?
8. Examining [Figure 11-21](#), what effect do you predict *trpA* mutations will have on tryptophan levels? What effect do you predict *trpA* mutations will have on *trp* mRNA expression?
9. Based on [Figure 11-29](#), why does binding of λ repressor prevent expression from the P_R promoter? Why does binding of Cro prevent expression from the P_{RM} promoter? 

10. On the basis of the sequences shown in [Figure 11-32b](#), would you expect all point mutations in -35 or -10 regions to affect gene expression?

BASIC PROBLEMS

11. Which of the following molecules is an inducer of the *lac* operon:
- Galactose
 - Allolactose
 - Glucose
 - Isothiocyanate
12. Explain why I^- alleles in the *lac* system are normally recessive to I^+ alleles and why I^+ alleles are recessive to I^S alleles.
13. What do we mean when we say that O^C mutations in the *lac* system are cis-acting?
14. The symbols *a*, *b*, and *c* in the table below represent the *E. coli* *lac* system genes for the repressor (*I*), the operator (*O*) region, and the β -galactosidase (*Z*), although not necessarily in that order. Furthermore, the order in which the symbols are written in the genotypes is not necessarily the

actual sequence in the *lac* operon.



Activity (+) or inactivity (-) of *Z* gene

Genotype	Inducer absent	Inducer present
$a^- b^+ c^+$	+	+
$a^+ b^+ c^-$	+	+
$a^+ b^- c^-$	-	-
$a^+ b^- c^+ / a^- b^+ c^-$	+	+
$a^+ b^+ c^+ / a^- b^- c^-$	-	+
$a^+ b^+ c^- / a^- b^- c^+$	-	+
$a^- b^+ c^+ / a^+ b^- c^-$	+	+

- a. Which symbol (*a*, *b*, or *c*) represents each of the *lac* genes *I*, *O*, and *Z*?
- b. In the table, a superscript minus sign on a gene symbol merely indicates a mutant, but some mutant behaviors in this system are given special mutant designations. Using the conventional gene symbols for the *lac* operon, designate each genotype in the table.
15. The map of the *lac* operon is

POZY

The promoter (*P*) region is the start site of transcription through the binding of the RNA polymerase molecule before actual mRNA production. Mutationally altered promoters (P^-) apparently cannot bind the RNA polymerase molecule. Certain predictions can be made about the effect of P^- mutations. Use your predictions and your knowledge of the lactose system to complete the following table. Insert a “+” where an enzyme is produced and a “-” where no enzyme is produced.

The first one has been done as an example.

Genotype	β-Galactosidase		Permease	
	No lactose	Lactose	No lactose	Lactose
$I^+ P^+ O^+ Z^+ Y^+ / I^+ P^+ O^+ Z^+ Y^+$	-	+	-	+
a. $I^- P^+ O^C Z^+ Y^- / I^+ P^+ O^+$				
b. $I^+ P^- O^C Z^- Y^+ / I^- P^+ O^C Z^+ Y^-$				
c. $I^S P^+ O^+ Z^+ Y^- / I^+ P^+ O^+$				
d. $I^S P^+ O^+ Z^+ Y^+ / I^- P^+ O^+$				
e. $I^- P^+ O^C Z^+ Y^- / I^- P^+ O^+$				
f. $I^- P^- O^+ Z^+ Y^+ / I^- P^+ O^C Z^+ Y^-$				

g. $I^+ P^+ O^+ Z^- Y^+/I^- P^+ O^+$
 $Z^+ Y^-$

16. Explain why it makes sense for the cell to synthesize β -galactosidase only when levels of lactose are high and levels of glucose are low.
17. Explain the fundamental differences between negative regulation and positive regulation of transcription in bacteria.
18. Which molecule regulates both the *lac* operon and the *ara* operon?
19. Compare the mechanisms of negative and positive regulation in the *lac* operon with those in the *ara* operon. 
20. Mutants that are *lacY*⁻ retain the capacity to synthesize β -galactosidase. However, even though the *lacI* gene is still intact, β -galactosidase can no longer be induced by adding lactose to the medium. Explain.
21. What is the function of the two tryptophan codons in the 14-amino-acid leader peptide in the regulation of *trp* operon? 
22. Could the attenuation mechanism found in the *trp* operon regulate gene expression in eukaryotic cells?
23. What are the similarities between the mechanisms controlling the *lac* operon and those controlling bacteriophage λ genetic switches?
24. Compare the arrangement of cis-acting sites in the control regions of the *lac* operon and bacteriophage λ .
25. Which regulatory protein induces the lytic phase genes of the bacteriophage λ life cycle?
 - a. cI
 - b. Cro
 - c. Int
 - d. cIII
26. What protein in bacteriophage λ serves as a readout of the level of resources in the cell?
27. What is the function of the cIII protein in the bacteriophage λ genetic switch?

28. Predict the effect of a mutation that eliminates the DNA-binding activity of the σ^E protein on spore formation in *Bacillus subtilis*. 

CHALLENGING PROBLEMS

29. An interesting mutation in *lacI* results in repressors with 110-fold increased binding to both operator and nonoperator DNA. These repressors display a “reverse” induction curve, allowing β -galactosidase synthesis in the absence of an inducer (IPTG) but partly repressing β -galactosidase expression in the presence of IPTG. How can you explain this? (Note that, when IPTG binds a repressor, it does not completely destroy operator affinity; rather, it reduces affinity 110-fold. Additionally, as cells divide and new operators are generated by the synthesis of daughter strands, the repressor must find the new operators by searching along the DNA, rapidly binding to nonoperator sequences and dissociating from them.)
30. Certain *lacI* mutations eliminate operator binding by the Lac repressor but do not affect the aggregation of subunits to make a tetramer, the active form of the repressor. These mutations are partly dominant over wild type. Can you explain the partly dominant I^-/I^+ phenotype of the I^-/I^+ heterodiploids?
31. You are examining the regulation of the lactose operon in the bacterium *Escherichia coli*. You isolate seven new independent mutant strains that lack the products of all three structural genes. You suspect that some of these mutations are *lacI^S* mutations and that other mutations are alterations that prevent the binding of RNA polymerase to the promoter region. Using whatever haploid and partial diploid genotypes that you think are necessary, describe a set of genotypes that will permit you to distinguish between the *lacI* and *lacP* classes of uninducible mutations. 
32. You are studying the properties of a new kind of regulatory mutation of the lactose operon. This mutation, called *S*, leads to the complete repression of the *lacZ*, *lacY*, and *lacA* genes, regardless of whether lactose is present. The results of studies of this mutation in partial diploids demonstrate that this mutation is completely dominant over wild type. When you treat bacteria of the *S* mutant strain with a mutagen and select for mutant bacteria that can express the enzymes encoded by *lacZ*, *lacY*, and *lacA* genes in the presence of lactose, some of the mutations map to the *lac* operator region and others to the *lac* repressor gene. On the basis of your knowledge of the lactose operon, provide a molecular genetic explanation for all these properties of the *S* mutation. Include an explanation of the constitutive nature of the “reverse mutations.”
33. The *trp* operon in *E. coli* encodes enzymes essential for the biosynthesis of tryptophan. The general mechanism for controlling the *trp* operon is similar to that observed with the *lac* operon: when the repressor binds to the operator, transcription is prevented; when the repressor does not bind to the operator, transcription proceeds. The regulation of the *trp* operon differs from the regulation of the

lac operon in the following way: the enzymes encoded by the *trp* operon are synthesized not when tryptophan is present, but rather when it is absent. In the *trp* operon, the repressor has two binding sites: one for DNA, and the other for the effector molecule, tryptophan. The *trp* repressor must first bind to a molecule of tryptophan before it can bind effectively to the *trp* operator. 

- Draw a map of the tryptophan operon, indicating the promoter (*P*), the operator (*O*), and the first structural gene of the tryptophan operon (*trpA*). In your drawing, indicate where on the DNA the repressor protein binds when it is bound to tryptophan.
 - The *trpR* gene encodes the repressor; *trpO* is the operator; *trpA* encodes the enzyme tryptophan synthetase. A *trpR*⁻ repressor cannot bind tryptophan, a *trpO*⁻ operator cannot be bound by the repressor, and the enzyme encoded by a *trpA*⁻ mutant gene is completely inactive. Do you expect to find active tryptophan synthetase in each of the following mutant strains when the cells are grown in the presence of tryptophan? In its absence?
 - R*⁺ *O*⁺ *A*⁺ (wild type)
 - R*⁻ *O*⁺ *A*⁺/*R*⁺ *O*⁺ *A*⁻
 - R*⁺ *O*⁻ *A*⁺/*R*⁺ *O*⁺ *A*⁻
34. The activity of the enzyme β -galactosidase produced by wild-type cells grown in media supplemented with different carbon sources is measured. In relative units, the following levels of activity are found:  
- | Glucose | Lactose | Lactose + glucose |
|---------|---------|-------------------|
| 0 | 100 | 1 |
- Predict the relative levels of β -galactosidase activity in cells grown under similar conditions when the cells are *lacI*⁻, *lacI*^S, *lacO*^C, and *crp*⁻.
- A bacteriophage λ is found that is able to lysogenize its *E. coli* host at 30°C but not at 42°C. What genes may be mutant in this phage? 
 - What would happen to the ability of bacteriophage λ to lyse a host cell if it acquired a mutation in the *O_R* binding site for the Cro protein? Why?
 - Sketch the effects of exposure of host cells to UV on the bacteriophage λ genetic switch.

38. Contrast the effects of mutations in genes encoding sporulation-specific σ factors with mutations in the -35 and -10 regions of the promoters of genes in their regulons. Would functional mutations in the σ -factor genes or in the individual promoters have the greater effect on sporulation? 

GENETICS AND SOCIETY

How might an understanding of the regulation of gene expression in bacteria be important for the treatment or prevention of human disease?