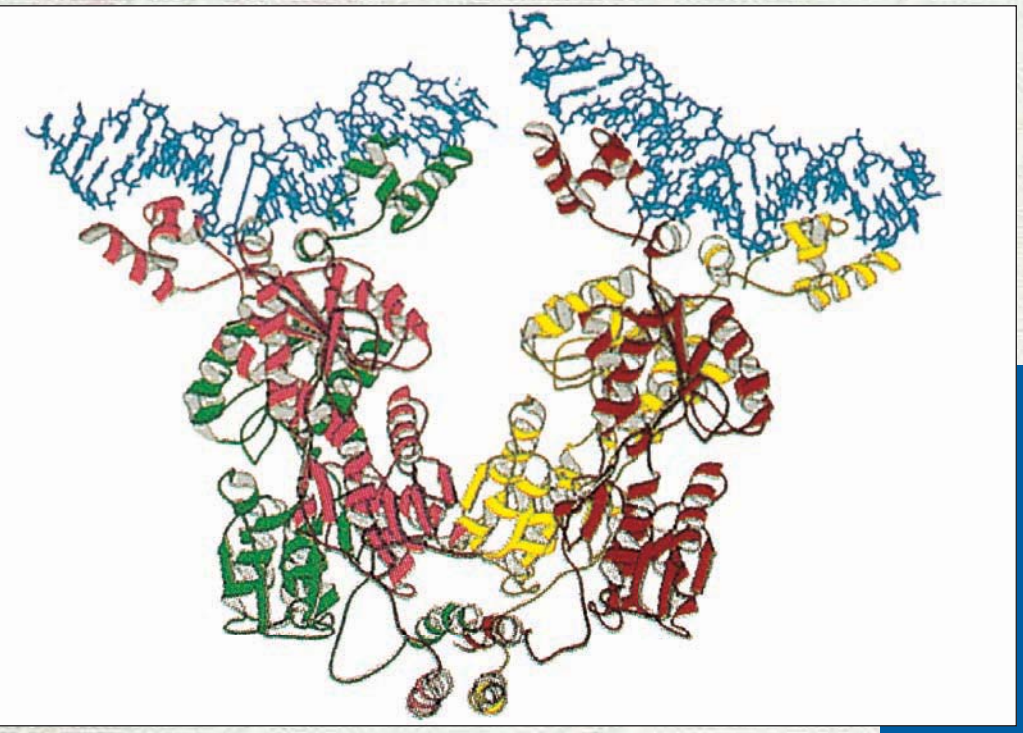
CHAPTER 7

Operons: Fine Control of Bacterial Transcription



X-ray crystal structure of the *lac* repressor tetramer bound to two operator fragments. *Lewis et al, Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. Science 271 (1 Mar 1996), f. 6, p. 1251. © AAAS*

The *E. coli* genome contains over 3000 genes. Some of these are active all the time because their products are in constant demand. But some of them are turned off most of the time because their products are rarely needed. For example, the enzymes required for the metabolism of the sugar arabinose would be useful only when arabinose is present and when the organism's favorite energy source, glucose, is absent. Such conditions are not common, so the genes encoding these enzymes are usually turned off. Why doesn't the cell just leave all its genes on all the time, so the right enzymes are always there to take care of any eventuality? The reason is that gene expression is an expensive process. It takes a lot of energy to produce RNA and protein. In fact, if all of an *E. coli* cell's genes were turned on all the time, production of RNAs and proteins would drain the cell of so much energy

that it could not compete with more efficient organisms. Thus, control of gene expression is essential to life. In this chapter we will explore one strategy bacteria employ to control the expression of their genes: by grouping functionally related genes together so they can be regulated together easily. Such a group of contiguous, coordinately controlled genes is called an *operon.*

7.1 The *lac* Operon

The first operon to be discovered has become the prime example of the operon concept. It contains three genes that code for the proteins that allow *E. coli* cells to use the sugar **lactose,** hence the name ***lac* operon.** Consider a flask of *E. coli* cells growing on a medium containing the sugars **glucose** and **lactose** (Figure 7.1). The cells exhaust the glucose and stop growing. Can they adjust to the new nutrient source? For a short time it appears that they cannot; but then, after a lag period of about an hour, growth resumes. During the lag, the cells have been turning on the *lac* operon and beginning to accumulate the enzymes they need to metabolize lactose. The growth curve in Figure 7.1 is called "diauxic" from the Latin *auxilium,* meaning help, because the two sugars help the bacteria grow.

What are these enzymes? First, the bacteria need an enzyme to transport the lactose into the cells. The name of this enzyme is **galactoside permease.** Next, the cells need an enzyme to break the lactose down into its two component sugars: galactose and glucose. Figure 7.2 shows this reaction. Because lactose is composed of two simple sugars, we call it a *disaccharide*. These six-carbon sugars, galactose and glucose, are joined together by a linkage called a β-galactosidic bond. Lactose is therefore called a β-galactoside, and the enzyme that cuts it in half is called **β-galactosidase.** The genes for these two enzymes, galactoside permease and β-galactosidase, are found side by side in the *lac* operon, along with another structural gene—for **galactoside transacetylase**—whose function in lactose metabolism is still unclear.

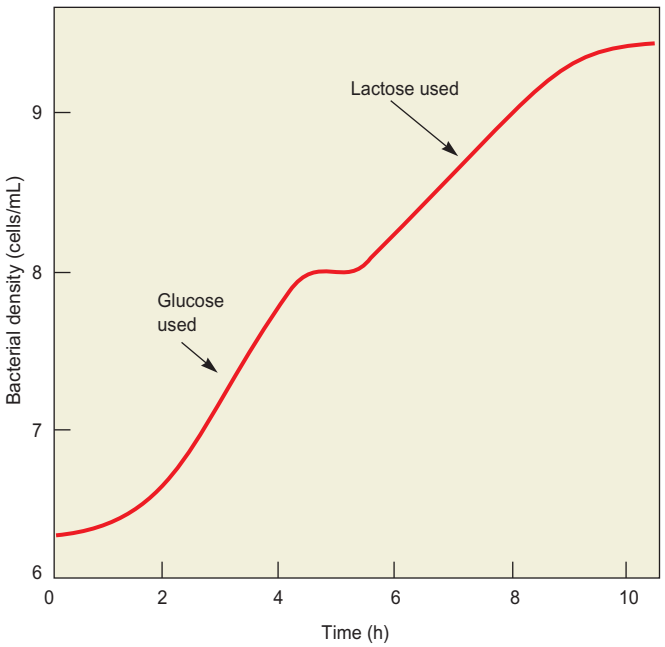


Figure 7.1 Diauxic growth. *E. coli* cells are grown on a medium containing both glucose and lactose, and the bacterial density (number of cells/mL) is plotted versus time in hours. The cells grow rapidly on glucose until that sugar is exhausted, then growth levels off while the cells induce the enzymes needed to metabolize lactose. As those enzymes appear, growth resumes.

The three genes coding for enzymes that carry out lactose metabolism are grouped together in the following order: β-galactosidase ***(lacZ),*** galactoside permease ***(lacY),*** galactoside transacetylase ***(lacA).*** They are all transcribed together to produce one messenger RNA, called a **polycistronic message,** starting from a single promoter. Thus, they can all be controlled together simply by controlling that promoter. The term *polycistronic* comes from **cistron,** which is a synonym for *gene*. Therefore, a polycistronic message is simply a message with information from more than one gene. Each cistron in the mRNA has its own ribosome binding site, so each cistron can be translated by separate ribosomes that bind independently of each other.

As mentioned at the beginning of this chapter, the *lac* operon (like many other operons) is tightly controlled.

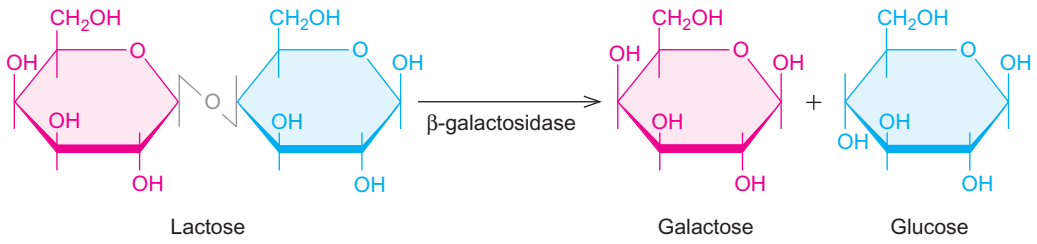


Figure 7.2 The β-galactosidase reaction. The enzyme breaks the β-galactosidic bond (gray) between the two sugars, galactose (pink) and glucose (blue), that compose lactose.

In fact, two types of control are operating. First is **negative control,** which is like the brake of a car: You need to release the brake for the car to move. The "brake" in negative control is a protein called the ***lac* repressor,** which keeps the operon turned off (or repressed) as long as lactose is absent. That is economical; it would be wasteful for the cell to produce enzymes that use an absent sugar.

If negative control is like the brake of a car, **positive control** is like the accelerator pedal. In the case of the *lac* operon, removing the repressor from the operator (releasing the brake) is not enough to activate the operon. An additional positive factor called an **activator** is needed. We will see that the activator responds to low glucose levels by stimulating transcription of the *lac* operon, but high glucose levels keep the concentration of the activator low, so transcription of the operon cannot be stimulated. The advantage of this positive control system is that it keeps the operon turned nearly off when the level of glucose is high. If there were no way to respond to glucose levels, the presence of lactose alone would suffice to activate the operon. But that is inappropriate when glucose is still available, because *E. coli* cells metabolize glucose more easily than lactose; it would therefore be wasteful for them to activate the *lac* operon in the presence of glucose.

SUMMARY Lactose metabolism in *E. coli* is carried out by two enzymes, with possible involvement by a third. The genes for all three enzymes are clustered together and transcribed together from one promoter, yielding a polycistronic message. These three genes, linked in function, are therefore also linked in expression. They are turned off and on together. Negative control keeps the *lac* operon repressed in the absence of lactose, and positive control keeps the operon relatively inactive in the presence of glucose, even when lactose is present.

Negative Control of the *lac* Operon

Figure 7.3 illustrates one aspect of *lac* operon regulation: the classical version of negative control. We will see later in this chapter and in Chapter 9 that this classical view is oversimplified, but it is a useful way to begin consideration of the operon concept. The term "negative control" implies that the operon is turned on unless something intervenes to stop it. The "something" that can turn off the *lac* operon is the ***lac* repressor.** This repressor, the product of a regulatory gene called the ***lacI* gene** shown at the extreme left in Figure 7.3, is a tetramer of four identical polypeptides; it binds to the **operator** just to the right of the promoter. When the repressor is bound to the operator, the operon is **repressed.** That is because the operator and promoter are contiguous, and when the repressor occupies the operator, it appears to prevent RNA polymerase from binding to the promoter and transcribing the operon. Because its genes are not transcribed, the operon is off, or repressed.

The *lac* operon is repressed as long as no lactose is available. On the other hand, when all the glucose is gone and lactose is present, a mechanism should exist for removing the repressor so the operon can be derepressed to take advantage of the new nutrient. How does this mechanism work? The repressor is a so-called **allosteric protein:** one in which the binding of one molecule to the protein changes the shape of a remote site on the protein and alters its interaction with a second molecule (Greek: *allos,* meaning other + *stereos,* meaning shape). The first molecule in this case is called the **inducer** of the *lac* operon because it binds to the repressor, causing the protein to change to a conformation that favors dissociation from the operator (the second molecule), thus inducing the operon (Figure 7.3b).

What is the nature of this inducer? It is actually an alternative form of lactose called **allolactose** (again, Greek: *allos,* meaning other). When β-galactosidase cleaves lactose to galactose plus glucose, it rearranges a small fraction of the lactose to allolactose. Figure 7.4 shows that allolactose is just galactose linked to glucose in a different way than in lactose. (In lactose, the linkage is through a β-1,4 bond; in allolactose, the linkage is β-1,6.)

You may be asking yourself: How can lactose be metabolized to allolactose if no permease is present to get it into the cell and no β-galactosidase exists to perform the metabolizing because the *lac* operon is repressed? The answer is that repression is somewhat leaky, and a low basal level of the *lac* operon products is always present. This is enough to get the ball rolling by producing a little inducer. It does not take much inducer to do the job, because only about 10 tetramers of repressor are present per cell. Furthermore, the derepression of the operon will snowball as more and more operon products are available to produce more and more inducer.

Discovery of the Operon

The development of the operon concept by François Jacob and Jacques Monod and their colleagues was one of the classic triumphs of the combination of genetic and biochemical analysis. The story begins in 1940, when Monod began studying the inducibility of lactose metabolism in *E. coli.* Monod learned that an important feature of lactose metabolism was β-galactosidase, and that this enzyme was inducible by lactose and by other galactosides. Furthermore, he and Melvin Cohn had used an anti-β-galactosidase antibody to detect β-galactosidase protein, and they

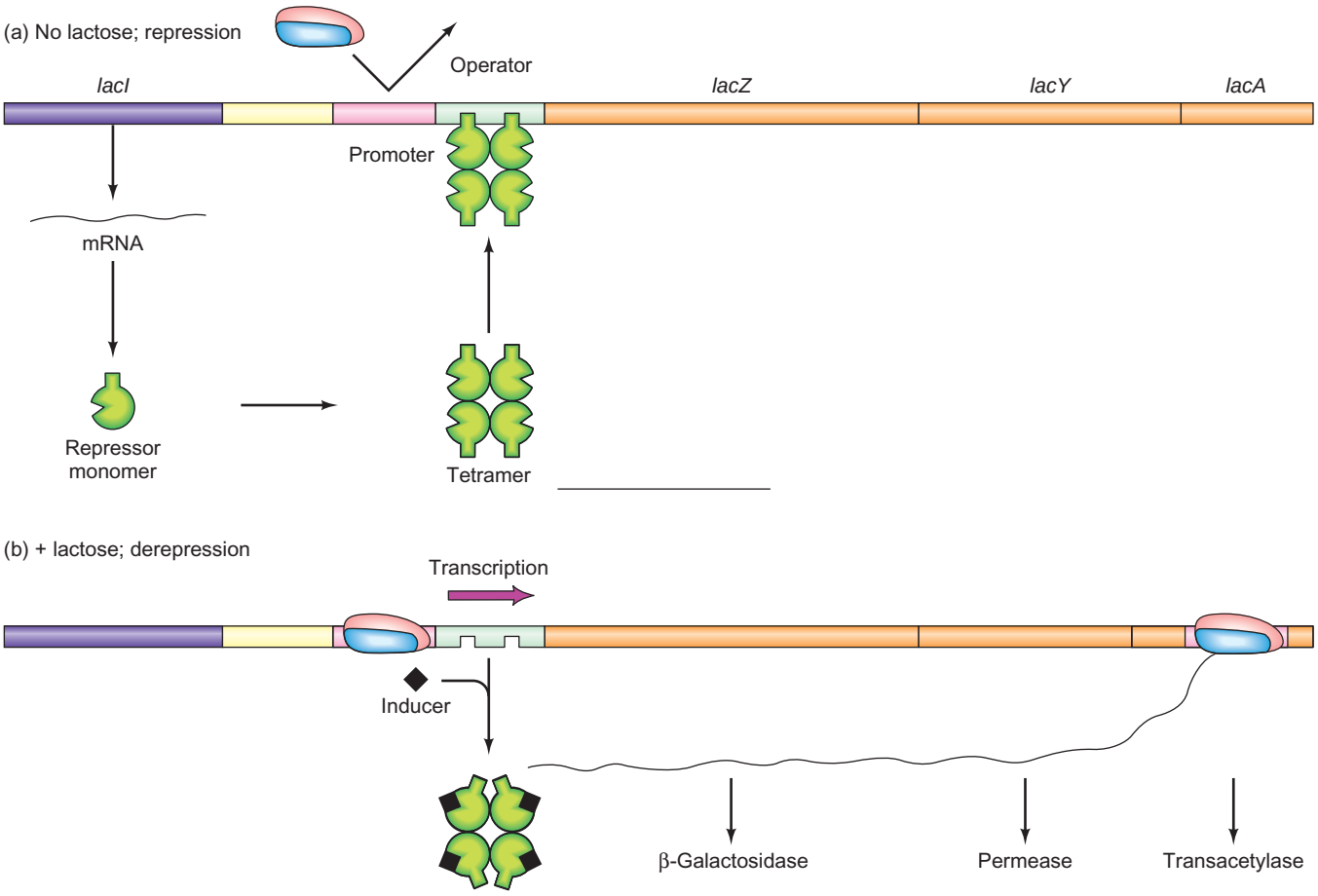


Figure 7.3 Negative control of the *lac* operon. **(a)** No lactose; repression. The *lacI* gene produces repressor (green), which binds to the operator and blocks RNA polymerase from transcribing the *lac* genes. **(b)** Presence of lactose, derepression. The inducer (black) binds to repressor, changing it to a form (bottom) that no longer binds well to the operator. This removes the repressor from the operator, allowing RNA polymerase to transcribe the structural genes. This produces a polycistronic mRNA that is translated to yield β-galactosidase, permease, and transacetylase.

showed that the amount of this protein increased on induction. Because more gene product appeared in response to lactose, the β-galactosidase gene itself was apparently being induced.

To complicate matters, certain mutants (originally called "cryptic mutants") were found that could make β-galactosidase but still could not grow on lactose. What was missing in these mutants? To answer this question, Monod and his coworkers added a radioactive galactoside to wild-type and mutant bacteria. They found that uninduced wild-type cells did not take up the galactoside, and neither did the mutants, even if they were induced. Induced

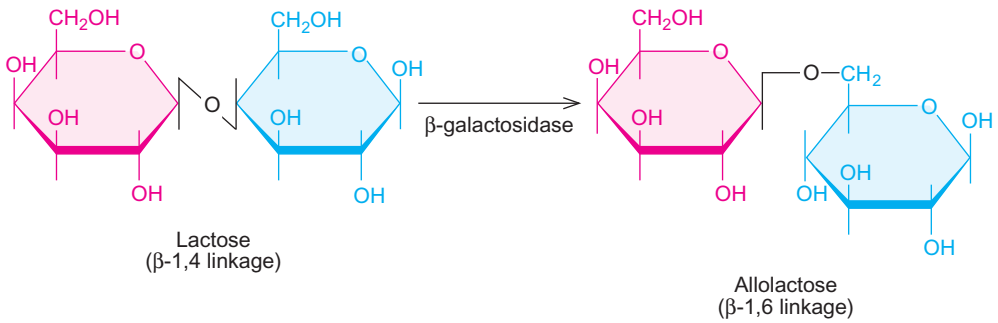


Figure 7.4 Conversion of lactose to allolactose. A side reaction carried out by β-galactosidase rearranges lactose to the inducer, allolactose. Note the change in the galactosidic bond from β-1,4 to β-1,6.

wild-type cells did accumulate the galactoside. This revealed two things: First, a substance (galactoside permease) is induced along with β-galactosidase in wild-type cells and is responsible for transporting galactosides into the cells; second, the mutants seem to have a defective gene (Y^-) for this substance (Table 7.1).

Monod named this substance galactoside permease, and then endured criticism from his colleagues for naming a protein before it had been isolated. He later remarked, "This attitude reminded me of that of two traditional English gentlemen who, even if they know each other well by name and by reputation, will not speak to each other before having been formally introduced." In their efforts to purify galactoside permease, Monod and his colleagues identified another protein, galactoside transacetylase, which is induced along with β-galactosidase and galactoside permease.

Thus, by the late 1950s, Monod knew that three enzyme activities (and therefore presumably three genes) were induced together by galactosides. He had also found some mutants, called **constitutive mutants,** that needed no induction. They produced the three gene products all the time. Monod realized that further progress would be greatly accelerated by genetic analysis, so he teamed up with François Jacob, who was working just down the hall at the Pasteur Institute.

In collaboration with Arthur Pardee, Jacob and Monod created **merodiploids** (partial diploid bacteria) carrying both the wild-type (inducible) and constitutive alleles. The inducible allele proved to be dominant, demonstrating that wild-type cells produce some substance that keeps the *lac* genes turned off unless they are induced. Because this substance turned off the genes from the constitutive as well as the inducible parent, it made the merodiploids inducible. Of course, this substance is the *lac* repressor. The constitutive mutants had a defect in the gene *(lacI)* for this repressor. These mutants are therefore $lacI^-$ (Figure 7.5a).

The existence of a repressor required that some specific DNA sequence exists to which the repressor would bind. Jacob and Monod called this the operator. The specificity of this interaction suggested that it should be subject to genetic mutation; that is, some mutations in the operator should abolish its interaction with the repressor. These would also be constitutive mutations, so how can they be distinguished from constitutive mutations in the repressor gene?

Jacob and Monod realized that they could make this distinction by determining whether the mutation was dominant or recessive. Because the repressor gene produces a repressor protein that can diffuse throughout the cell, it can bind to both operators in a merodiploid. We call such a gene ***trans*-acting** because it can act on loci on both DNA molecules in the merodiploid (Latin: *trans,* meaning across). A mutation in one of the repressor genes will still leave the other repressor gene undamaged, so its wild-type product can still diffuse to both operators and turn them off. In other words, both *lac* operons in the merodiploid would still be repressible. Thus, such a mutation should be recessive (Figure 7.5a), and we have already observed that it is.

Table 7.1 Effect of Cryptic Mutant ($lacY^-$) on Accumulation of Galactoside

Genotype	Inducer	Accumulation of Galactoside
Z^+Y^+	–	–
Z^+Y^+	+	+
Z^+Y^- (cryptic)	–	–
Z^+Y^- (cryptic)	+	–

On the other hand, because an operator controls only the operon on the same DNA molecule, we call it ***cis*-acting** (Latin: *cis,* meaning here). Thus, a mutation in one of the operators in a merodiploid should render the operon on that DNA molecule unrepressable, but should not affect the operon on the other DNA molecule. We call such a mutation ***cis*-dominant** because it is dominant only with respect to genes on the same DNA (***in cis***), not on the other DNA in the merodiploid (***in trans***). Jacob and Monod did indeed find such *cis*-dominant mutations, and they proved the existence of the operator. These mutations are called O^c, for **operator constitutive.**

What about mutations in the repressor gene that render the repressor unable to respond to inducer? Such mutations should make the *lac* operon uninducible and should be dominant both *in cis* and *in trans* because the mutant repressor will remain bound to both operators even in the presence of inducer or of wild-type repressor (Figure 7.5c). Monod and his colleagues found two such mutants, and Suzanne Bourgeois later found many others. These are named I^s to distinguish them from constitutive repressor mutants (I^-), which make a repressor that cannot recognize the operator.

Both of the common kinds of constitutive mutants (I^- and O^c) affected all three of the *lac* genes (*Z, Y,* and *A*) in the same way. The genes had already been mapped and were found to be adjacent on the chromosome. These findings strongly suggested that the operator lay near these three genes.

We now recognize yet another class of repressor mutants, those that are constitutive and dominant (I^{-d}). This kind of mutant gene (Figure 7.5d) makes a defective product that can still form tetramers with wild-type repressor monomers. However, the defective monomers spoil the activity of the whole tetramer so it cannot bind

to the operator. Hence the dominant nature of this mutation. These mutations are not just *cis*-dominant because the "spoiled" repressors cannot bind to either operator in a merodiploid. This kind of "spoiler" mutation is widespread in nature, and it is called by the generic name **dominant-negative.**

Thus, Jacob and Monod, by skillful genetic analysis, were able to develop the operon concept. They predicted the existence of two key control elements: the repressor gene and the operator. Deletion mutations revealed a third element (the promoter) that was necessary for expression of all three *lac* genes. Furthermore, they could conclude that all three *lac* genes (*lacZ*, *Y*, and *A*) were clustered into a single control unit: the *lac* operon. Subsequent biochemical studies have amply confirmed Jacob and Monod's beautiful hypothesis.

SUMMARY Negative control of the *lac* operon occurs as follows: The operon is turned off as long as the repressor binds to the operator, because the repressor keeps RNA polymerase from transcribing the three *lac* genes. When the supply of glucose is exhausted and lactose is available, the few molecules of *lac* operon enzymes produce a few molecules of allolactose from the lactose. The allolactose acts as an inducer by binding to the repressor and causing a conformational shift that encourages dissociation from the operator. With the repressor removed, RNA polymerase is free to transcribe the three *lac* genes. A combination of genetic and biochemical experiments revealed the two key elements of negative control of the *lac* operon: the operator and the repressor.

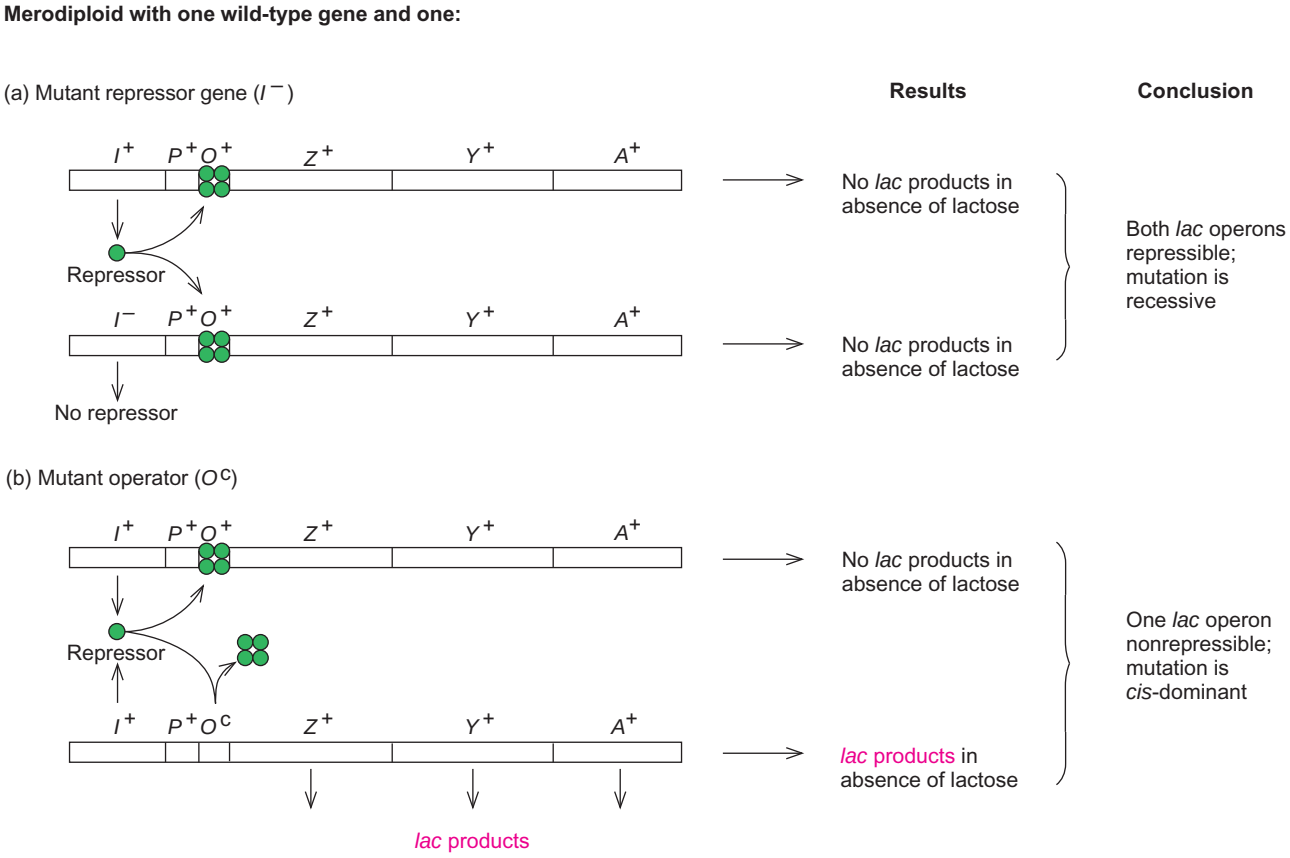


Figure 7.5 Effects of regulatory mutations in the *lac* operon in merodiploids. Jacob, Monod, and others created merodiploid *E. coli* strains as described in panels (a)–(d) and tested them for *lac* products in the presence and absence of lactose. **(a)** This merodiploid has one wild-type operon (top) and one operon (bottom) with a mutation in the repressor gene (I^-). The wild-type repressor gene (I^+) makes enough normal repressor (green) to repress both operons, so the I^- mutation is recessive. **(b)** This merodiploid has one wild-type operon (top) and one operon (bottom) with a mutation in the operator (O^c) that makes it defective in binding repressor (green). The wild-type operon remains repressible, but the mutant operon is not; it makes *lac* products even in the absence of lactose. Because only the operon connected to the mutant operator is affected, this mutation is *cis*-dominant.

(continued)

Repressor–Operator Interactions

After the pioneering work of Jacob and Monod, Walter Gilbert and Benno Müller-Hill succeeded in partially purifying the *lac* repressor. This work is all the more impressive, considering that it was done in the 1960s, before the advent of modern gene cloning. Gilbert and Müller-Hill's challenge was to purify a protein (the *lac* repressor) that is present in very tiny quantities in the cell, without an easy assay to identify the protein. The most sensitive assay available to them was binding a labeled synthetic inducer (isopropylthiogalactoside, or IPTG) to the repressor. But, with a crude extract of wild-type cells, the repressor was in such low concentration that this assay could not detect it. To get around this problem, Gilbert and Müller-Hill used a mutant *E. coli* strain with a repressor mutation ($lacI^t$) that causes the repressor to bind IPTG more tightly than normal. This tight binding allowed the mutant repressor to bind enough inducer that the protein could be detected even in very impure extracts. Because they could detect the protein, Gilbert and Müller-Hill could purify it.

Melvin Cohn and his colleagues used repressor purified by this technique in operator-binding studies. To assay repressor–operator binding, Cohn and colleagues used the nitrocellulose filter-binding assay we discussed in Chapters 5 and 6. If repressor–operator interaction worked normally, we would expect it to be blocked by inducer. Indeed, Figure 7.6 shows a typical saturation curve for repressor–operator binding in the absence of inducer, but no binding in the presence of the synthetic inducer, IPTG. In another binding experiment (Figure 7.7), Cohn and coworkers showed that DNA containing the constitutive mutant operator ($lacO^c$) required a higher concentration of repressor to achieve full binding than did the wild-type operator. This was an important

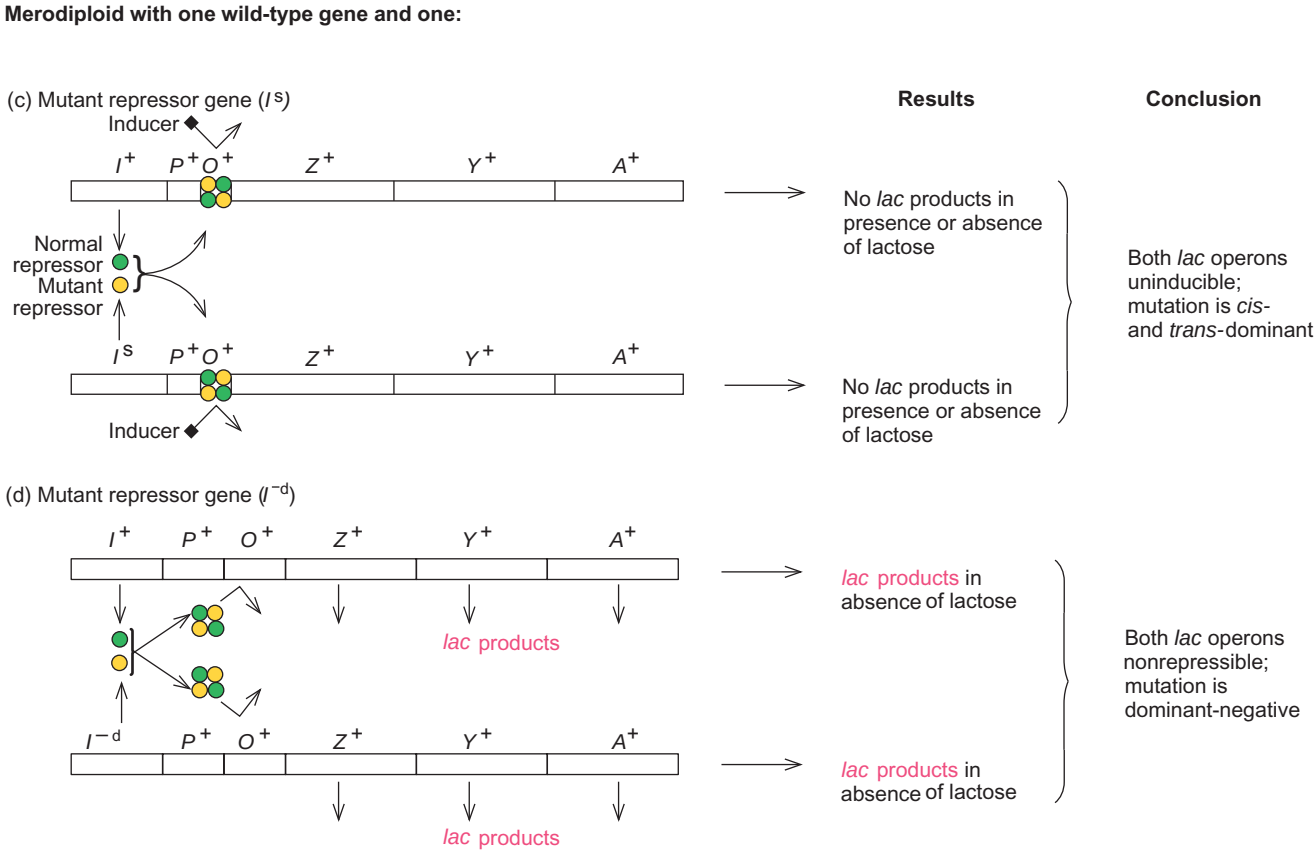


Figure 7.5 ***(continued)*** **(c)** This merodiploid has one wild-type operon (top) and one operon (bottom) with a mutant repressor gene (*I*s) whose product (yellow) cannot bind inducer. The mutant repressor therefore binds irreversibly to both operators and renders both operons uninducible. This mutation is therefore dominant. Notice that these repressor tetramers containing some mutant and some wild-type subunits behave as mutant proteins. That is, they remain bound to the operator even in the presence of inducer. **(d)** This merodiploid has one wild-type operon (top) and one operon (bottom) with a mutant repressor gene (I^{-d}) whose product (yellow) cannot bind to the *lac* operator. Moreover, mixtures (heterotetramers) composed of both wild-type and mutant repressor monomers still cannot bind to the operator. Thus, because the operon remains derepressed even in the absence of lactose, this mutation is dominant. Furthermore, because the mutant protein poisons the activity of the wild-type protein, we call the mutation dominant-negative.

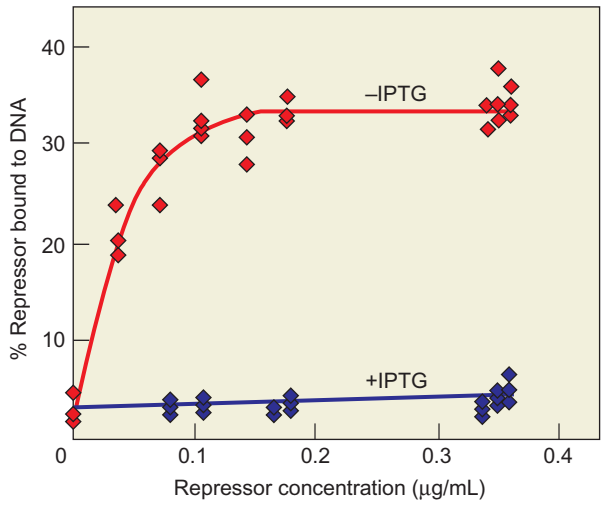


Figure 7.6 Assaying the binding between *lac* operator and *lac* repressor. Cohn and colleagues labeled *lacO*-containing DNA with ^{32}P and added increasing amounts of *lac* repressor. They assayed binding between repressor and operator by measuring the radioactivity attached to nitrocellulose. Only labeled DNA bound to repressor would attach to nitrocellulose. Red: repressor bound in the absence of the inducer IPTG. Blue: repressor bound in the presence of 1 mM IPTG, which prevents repressor–operator binding. (*Source:* Adapted from Riggs, A.D., et al.,1968. DNA binding of the *lac* repressor, *Journal of Molecular Biology,* Vol. 34: 366.)

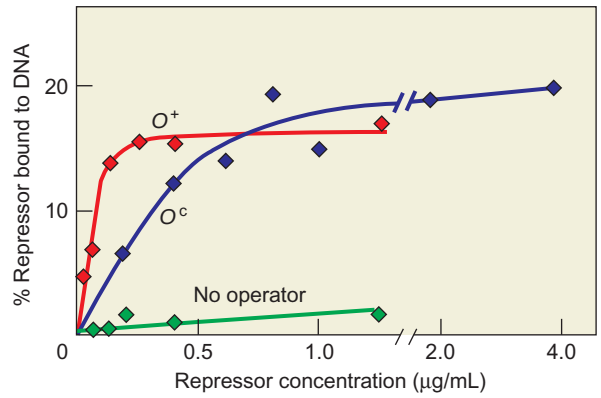


Figure 7.7 The O^c *lac* operator binds repressor with lower affinity than does the wild-type operator. Cohn and colleagues performed a *lac* operator–repressor binding assay as described in Figure 7.6, using three different DNAs as follows: red, DNA containing a wild-type operator (O^+); blue, DNA containing an operator-constitutive mutation (O^c) that binds repressor with a lower affinity; green, control, λϕ80 DNA, which does not have a *lac* operator. (*Source:* Adapted from Riggs, A.D., et al. 1968. DNA binding of the *lac* repressor. *Journal of Molecular Biology,* Vol. 34: 366.)

demonstration: What Jacob and Monod had defined genetically as the operator really was the binding site for repressor. If it were not, then mutating it should not have affected repressor binding.

SUMMARY Cohn and colleagues demonstrated with a filter-binding assay that *lac* repressor binds to *lac* operator. Furthermore, this experiment showed that a genetically defined constitutive *lac* operator has lower than normal affinity for the *lac* repressor, demonstrating that the sites defined genetically and biochemically as the operator are one and the same.

The Mechanism of Repression

For years it was assumed that the *lac* repressor acted by denying RNA polymerase access to the promoter, in spite of the fact that Ira Pastan and his colleagues had shown as early as 1971 that RNA polymerase could bind tightly to the *lac* promoter, even in the presence of repressor. Pastan's experimental plan was to incubate polymerase with DNA containing the *lac* operator in the presence of repressor, then to add inducer (IPTG) and rifampicin together. As we will see later in this chapter, rifampicin will inhibit transcription unless an open promoter complex has already formed. (Recall from Chapter 6 that an open promoter complex is one in which the RNA polymerase has caused local DNA melting at the promoter and is tightly bound there.) In this case, transcription did occur, showing that the *lac* repressor had not prevented the formation of an open promoter complex. Thus, these results suggested that the repressor does not block access by RNA polymerase to the *lac* promoter. Susan Straney and Donald Crothers reinforced this view in 1987 by showing that polymerase and repressor can bind together to the *lac* promoter.

If we accept that RNA polymerase can bind tightly to the promoter, even with repressor occupying the operator, how do we explain repression? Straney and Crothers suggested that repressor blocks the formation of an open promoter complex, but that would be hard to reconcile with the rifampicin resistance of the complex observed by Pastan. Barbara Krummel and Michael Chamberlin proposed an alternative explanation: Repressor blocks the transition from the initial transcribing complex state (Chapter 6) to the elongation state. In other words, repressor traps the polymerase in a nonproductive state in which it spins its wheels making abortive transcripts without ever achieving promoter clearance.

Jookyung Lee and Alex Goldfarb provided some evidence for this idea. First, they used a run-off transcription assay (Chapter 5) to show that RNA polymerase is already engaged on the DNA template, even in the presence of repressor. The experimental plan was as follows: First, they incubated repressor with a 123-bp DNA fragment containing the *lac* control region plus the beginning of the *lacZ* gene. After allowing 10 min for the repressor to bind to the operator, they added polymerase. Then they added

heparin—a polyanion that binds to any RNA polymerase that is free or loosely bound to DNA and keeps it from binding to DNA. They also added all the remaining components of the RNA polymerase reaction except CTP. Finally, they added labeled CTP with or without the inducer IPTG. The question is this: Will a run-off transcript be made? If so, the RNA polymerase has formed a heparin-resistant (open) complex with the promoter even in the presence of the repressor. In fact, as Figure 7.8 shows, the run-off transcript did appear, just as if repressor had not been present. Thus, under these conditions in vitro, repressor does not seem to inhibit tight binding between polymerase and the *lac* promoter.

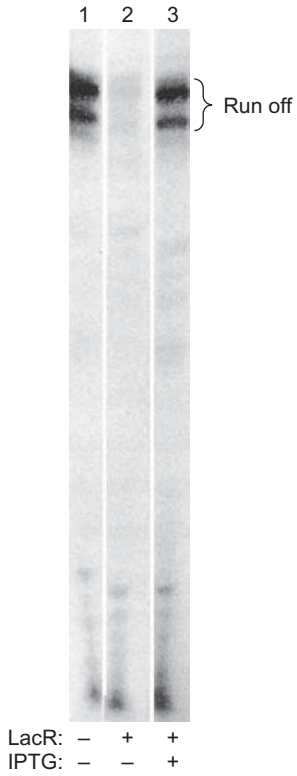


Figure 7.8 RNA polymerase forms an open promoter complex with the *lac* promoter even in the presence of *lac* repressor in vitro. Lee and Goldfarb incubated a DNA fragment containing the *lac* UV5 promoter with (lanes 2 and 3) or without (lane 1) *lac* repressor (LacR). After repressor–operator binding had occurred, they added RNA polymerase. After allowing 20 min for open promoter complexes to form, they added heparin to block any further complex formation, along with all the other reaction components except CTP. Finally, after 5 more minutes, they added [α-^{32}P]CTP alone or with the inducer IPTG. They allowed 10 more minutes for RNA synthesis and then electrophoresed the transcripts. Lane 3 shows that transcription occurred even when repressor bound to the DNA before polymerase could. Thus, repressor did not prevent polymerase from binding and forming an open promoter complex. (*Source:* Lee J., and Goldfarb A., *lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* 66 (23 Aug 1991) f. 1, p. 794. Reprinted by permission of Elsevier Science.)

If it does not inhibit transcription of the *lac* operon by blocking access to the promoter, how would the *lac* repressor function? Lee and Goldfarb noted the appearance of shortened abortive transcripts (Chapter 6), only about 6 nt long, in the presence of repressor. Without repressor, the abortive transcripts reached a length of 9 nt. The fact that *any* transcripts—even short ones—were made in the presence of repressor reinforced the conclusion that, at least under these conditions, RNA polymerase really can bind to the *lac* promoter in the presence of repressor. This experiment also suggested that repressor may limit *lac* operon transcription by locking the polymerase into a nonproductive state in which it can make only abortive transcripts. Thus, extended transcription cannot get started.

One problem with the studies of Lee and Goldfarb and the others just cited is that they were performed in vitro under rather nonphysiological conditions. For example, the concentrations of the proteins (RNA polymerase and repressor) were much higher than they would be in vivo. To deal with such problems, Thomas Record and colleagues performed kinetic studies in vitro under conditions likely to prevail in vivo. They formed RNA polymerase/*lac* promoter complexes, then measured the rate of abortive transcript synthesis by these complexes alone, or after addition of either heparin or *lac* repressor. They measured transcription by using a UTP analog with a fluorescent tag on the γ-phosphate (*pppU). When UMP was incorporated into RNA, tagged pyrophosphate (*pp) was released, and the fluorescence intensity increased. Figure 7.9 demonstrates that the rate of abortive transcript synthesis continued at a high level in the absence of competitor, but rapidly leveled off in the presence of either heparin or repressor.

Record and colleagues explained these results as follows: The polymerase–promoter complex is in equilibrium with free polymerase and promoter. Moreover, in the absence of competitor (curve 1), the polymerases that dissociate go right back to the promoter and continue making abortive transcripts. However, both heparin (curve 2) and repressor (curve 3) prevent such reassociation. Heparin does so by binding to the polymerase and preventing its association with DNA. But the repressor presumably does so by binding to the operator adjacent to the promoter and blocking access to the promoter by RNA polymerase. Thus, these data support the old hypothesis of a competition between polymerase and repressor.

We have seen that the story of the *lac* repressor mechanism has had many twists and turns. Have we seen the last twist? The latest results suggest that the original, competition hypothesis is correct, but we may not have heard the end of the story yet.

Another complicating factor in repression of the *lac* operon is the presence of not one, but three operators: one *major*

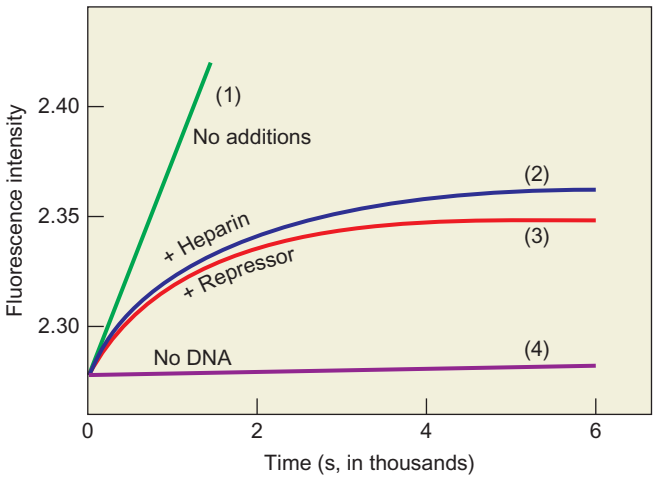


Figure 7.9 Effect of *lac* repressor on dissociation of RNA polymerase from the *lac* promoter. Record and colleagues made complexes between RNA polymerase and DNA containing the *lac* promoter–operator region. Then they allowed the complexes to synthesize abortive transcripts in the presence of a UTP analog fluorescently labeled in the γ-phosphate. As the polymerase incorporates UMP from this analog into transcripts, the labeled pyrophosphate released increases in fluorescence intensity. The experiments were run with no addition (curve 1, green), with heparin to block reinitiation by RNA polymerase that dissociates from the DNA (curve 2, blue), and with a low concentration of *lac* repressor (curve 3, red). A control experiment was run with no DNA (curve 4, purple). The repressor inhibited reinitiation of abortive transcription as well as heparin, suggesting that it blocks dissociated RNA polymerase from reassociating with the promoter. (*Source:* Adapted from Schlax, P.J., Capp, M.W., and M.T. Record, Jr. Inhibition of transcription initiation by *lac* repressor, *Journal of Molecular Biology* 245: 331–50.)

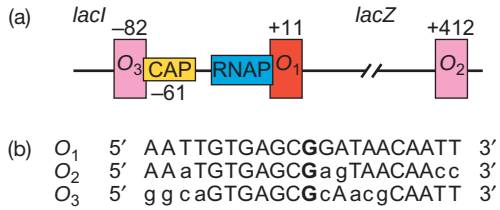


Figure 7.10 The three *lac* operators. (a) Map of the *lac* control region. The major operator (O_1) is shown in red; the two auxiliary operators are shown in pink. The CAP and RNA polymerase binding sites are in yellow and blue, respectively. CAP is a positive regulator of the *lac* operon discussed in the next section of this chapter. **(b)** Sequences of the three operators. The sequences are aligned, with the central G of each in boldface. Sites at which the auxiliary operator sequences differ from the major operator are lower case in the O_2 and O_3 sequences.

operator near the transcription start site and two *auxiliary operators* (one upstream and one downstream). Figure 7.10 shows the spatial arrangement of these operators, the classical (major) operator O_1, centered at position +11, the downstream auxiliary operator O_2, centered at position +412, and the upstream auxiliary operator O_3, centered at position –82. We have already discussed the classical operator, and the role investigators have traditionally ascribed to it alone. But Müller-Hill and others have more recently investigated the auxiliary operators and have discovered that they are not just trivial copies of the major operator. Instead, they play a significant role in repression. Müller-Hill and colleagues demonstrated this role by showing that removal of either of the auxiliary operators decreased repression only slightly, but removal of both auxiliary operators decreased repression about 50-fold. Figure 7.11 outlines the results of these experiments and shows that all three operators together repress transcription 1300-fold, two operators together repress from 440- to 700-fold, but the classical operator by itself represses only 18-fold.

In 1996, Mitchell Lewis and coworkers provided a structural basis for this cooperativity among operators. They determined the crystal structure of the *lac* repressor and its complexes with 21-bp DNA fragments containing operator sequences. Figure 7.12 summarizes their findings. We can see that the two dimers in a repressor tetramer are independent DNA-binding entities that interact with the major groove of the DNA. It is also clear that the two dimers within the tetramer are bound to separate operator sequences. It is easy to imagine these two operators as part of a single long piece of DNA.

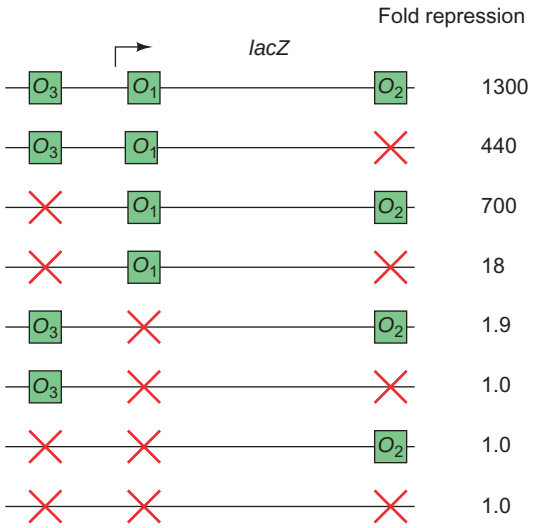


Figure 7.11 Effects of mutations in the three *lac* operators. Müller-Hill and colleagues placed wild-type and mutated *lac* operon fragments on λ phage DNA and allowed these DNAs to lysogenize *E. coli* cells (Chapter 8). This introduced these *lac* fragments, containing the three operators, the *lac* promoter, and the *lacZ* gene, into the cellular genome. The cell contained no other *lacZ* gene, but it had a wild-type *lacI* gene. Then Müller-Hill and coworkers assayed for β-galactosidase produced in the presence and absence of the inducer IPTG. The ratio of activity in the presence and absence of inducer is the repression given at right. For example, the repression observed with all three operators was 1300-fold. λ Ewt 123 (top) was wild-type in all three operators (green). All the other phages had one or more operators deleted (red X). *Source:* Adapted from Oehler, S., E.R. Eismann, H. Krämer, and B. Müller-Hill. 1990. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* 9:973–79.

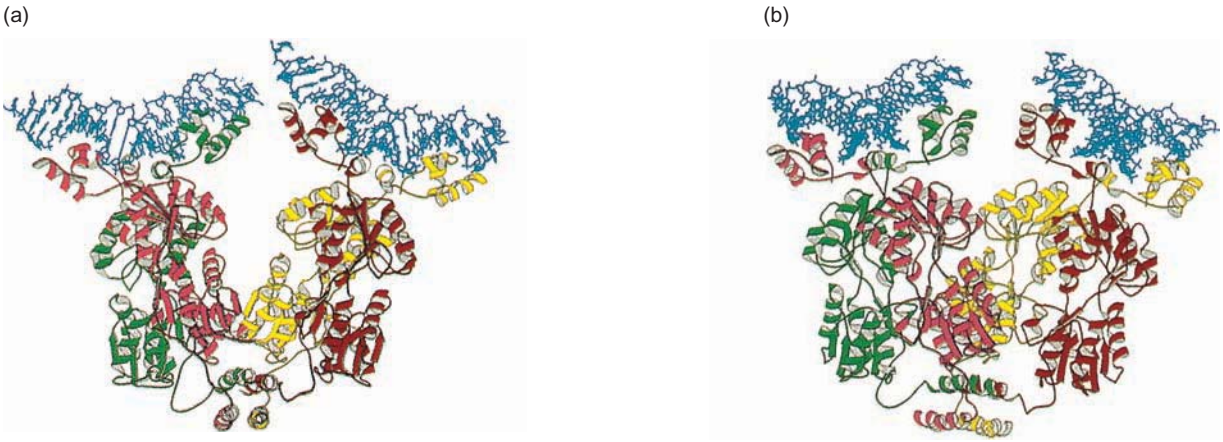


Figure 7.12 Structure of the *lac* repressor tetramer bound to two operator fragments. Lewis, Lu, and colleagues performed x-ray crystallography on *lac* repressor bound to 21-bp DNA fragments containing the major *lac* operator sequence. The structure presents the four repressor monomers in pink, green, yellow, and red, and the DNA fragments in blue. Two repressor dimers interact with each other at bottom to form tetramers. Each of the dimers contains two DNA-binding domains that can be seen interacting with the DNA major grooves at top. The structure shows clearly that the two dimers can bind independently to separate *lac* operators. Panels **(a)** and **(b)** are "front" and "side" views of the same structure. (*Source:* Lewis et al., Crystal structure of the lactose operon processor and its complexes with DNA and inducer. *Science* 271 (1 Mar 1996), f. 6, p. 1251. © AAAS.)

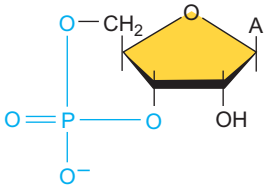
SUMMARY Two competing hypotheses seek to explain the mechanism of repression of the *lac* operon. One is that the RNA polymerase can bind to the *lac* promoter in the presence of the repressor, but the repressor inhibits the transition from abortive transcription to processive transcription. The other is that the repressor, by binding to the operator, blocks access by the polymerase to the adjacent promoter. The latest evidence supports the latter hypothesis. In addition to the classical (major) *lac* operator adjacent to the promoter, two auxiliary *lac* operators exist: one each upstream and downstream. All three operators are required for optimum repression, two work reasonably well, but the classical operator by itself produces only a modest amount of repression.

Positive Control of the *lac* Operon

As we learned earlier in this chapter, *E. coli* cells keep the *lac* operon in a relatively inactive state as long as glucose is present. This selection in favor of glucose metabolism and against use of other energy sources has long been attributed to the influence of some breakdown product, or *catabolite*, of glucose. It is therefore known as **catabolite repression.**

The ideal positive controller of the *lac* operon would be a substance that sensed the lack of glucose and responded by activating the *lac* promoter so that RNA polymerase could bind and transcribe the *lac* genes (assuming, of course, that lactose is present and the repressor is therefore not bound to the operator). One substance that responds to glucose concentration is a nucleotide called **cyclic-AMP** (**cAMP**) (Figure 7.13). As the level of glucose drops, the concentration of cAMP rises.

Catabolite Activator Protein Ira Pastan and his colleagues demonstrated that cAMP, added to bacteria, could overcome catabolite repression of the *lac* operon and a number of other operons, including the *gal* and *ara* operons. The latter two govern the metabolism of the sugars galactose and arabinose, respectively. In other words, cAMP rendered these genes active, even in the presence of glucose. This finding implicated cAMP strongly in the positive control of the *lac* operon. Does this mean that cAMP is the positive effector? Not exactly. The positive controller of the *lac* operon is a complex composed of two parts: cAMP and a protein factor.



Cyclic-AMP

Figure 7.13 Cyclic-AMP. Note the cyclic 5′-3′ phosphodiester bond (blue).

Geoffrey Zubay and coworkers showed that a crude cell-free extract of *E. coli* would make β-galactosidase if supplied with cAMP. This finding led the way to the discovery of a protein in the extract that was necessary for the stimulation by cAMP. Zubay called this protein **catabolite activator protein,** or **CAP.** Later, Pastan's group found the same protein and named it **cyclic-AMP receptor protein,** or **CRP.** To avoid confusion, we will refer to this protein from now on as CAP, regardless of whose experiments we are discussing. However, the gene encoding this protein has been given the official name *crp*.

Pastan and colleagues found that the dissociation constant for the CAP–cAMP complex was 1–2 × 10^{-6} M. However, they also isolated a mutant whose CAP bound about 10 times less tightly to cAMP. If CAP–cAMP really is important to positive control of the *lac* operon, we would expect reduced production of β-galactosidase by a cAMP-supplemented cell-free extract of these mutant cells. Figure 7.14 shows that this is indeed the case. To make the point even more strongly, Pastan showed that β-galactosidase synthesis by this mutant extract (plus cAMP) could be stimulated about threefold by the addition of wild-type CAP.

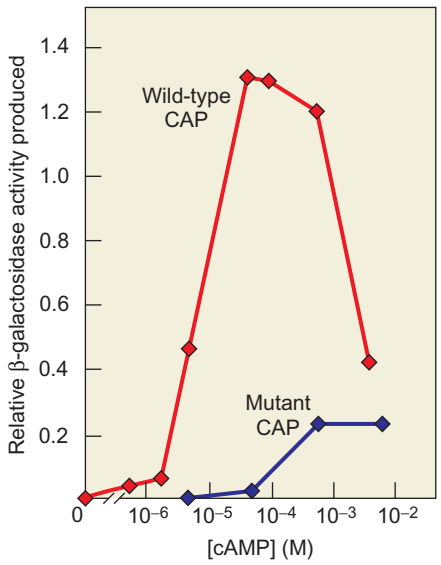


Figure 7.14 Stimulation of β-galactosidase synthesis by cAMP with wild-type and mutant CAP. Pastan and colleagues stimulated cell-free bacterial extracts to make β-galactosidase in the presence of increasing concentrations of cAMP with a wild-type extract (red), or an extract from mutant cells that have a CAP with reduced affinity for cAMP (blue). This mutant extract made much less β-galactosidase, which is what we expect if the CAP–cAMP complex is important in *lac* operon transcription. Too much cAMP obviously interfered with β-galactosidase synthesis in the wild-type extract. This is not surprising because cAMP has many effects, and some may indirectly inhibit some step in expression of the *lacZ* gene in vitro. (*Source:* Adapted from Emmer, M., et al., Cyclic AMP receptor protein of *E. coli:* Its role in the synthesis of inducible enzymes, *Proceedings of the National Academy of Sciences* 66(2): 480–487, June 1970.)

SUMMARY Positive control of the *lac* operon, and certain other inducible operons that code for sugar-metabolizing enzymes, is mediated by a factor called catabolite activator protein (CAP), which, in conjunction with cyclic-AMP, stimulates transcription. Because cyclic-AMP concentration is depressed by glucose, this sugar prevents stimulation of transcription. Thus, the *lac* operon is activated only when glucose concentration is low and therefore a need arises to metabolize an alternative energy source.

The Mechanism of CAP Action

How do CAP and cAMP stimulate *lac* transcription? Zubay and colleagues discovered a class of *lac* mutants in which CAP and cAMP could not stimulate *lac* transcription. These mutations mapped to the *lac* promoter, suggesting that the binding site for the CAP–cAMP complex lies in the promoter. Later molecular biological work, which we will discuss shortly, has shown that the CAP–cAMP binding site (the **activator-binding site**) lies just upstream of the promoter. Pastan and colleagues went on to show that this binding of CAP and cAMP to the activator site helps RNA polymerase to form an open promoter complex. The role of cAMP is to change the shape of CAP to increase its affinity for the activator-binding site.

Figure 7.15 shows how this experiment worked. First, Pastan and colleagues allowed RNA polymerase to bind to the *lac* promoter in the presence or absence of CAP and cAMP. Then they challenged the promoter complex by adding nucleotides and rifampicin simultaneously to see if an open promoter complex had formed. If not, transcription should be rifampicin-sensitive because the DNA melting step takes so much time that it would allow the antibiotic to inhibit the polymerase before initiation could occur. However, if it was an open promoter complex, it would be primed to polymerize nucleotides. Because nucleotides reach the polymerase before the antibiotic, the polymerase has time to initiate transcription. Once it has initiated an RNA chain, the polymerase becomes resistant to rifampicin until it completes that RNA chain. In fact, Pastan and colleagues found that when the polymerase– promoter complex formed in the absence of CAP and cAMP it was still rifampicin-sensitive. Thus, it had not formed an open promoter complex. On the other hand, when CAP and cAMP were present when polymerase associated with the promoter, a rifampicin-resistant open promoter complex formed.

Figure 7.15b presents a dimer of CAP–cAMP at the activator site on the left and polymerase at the promoter

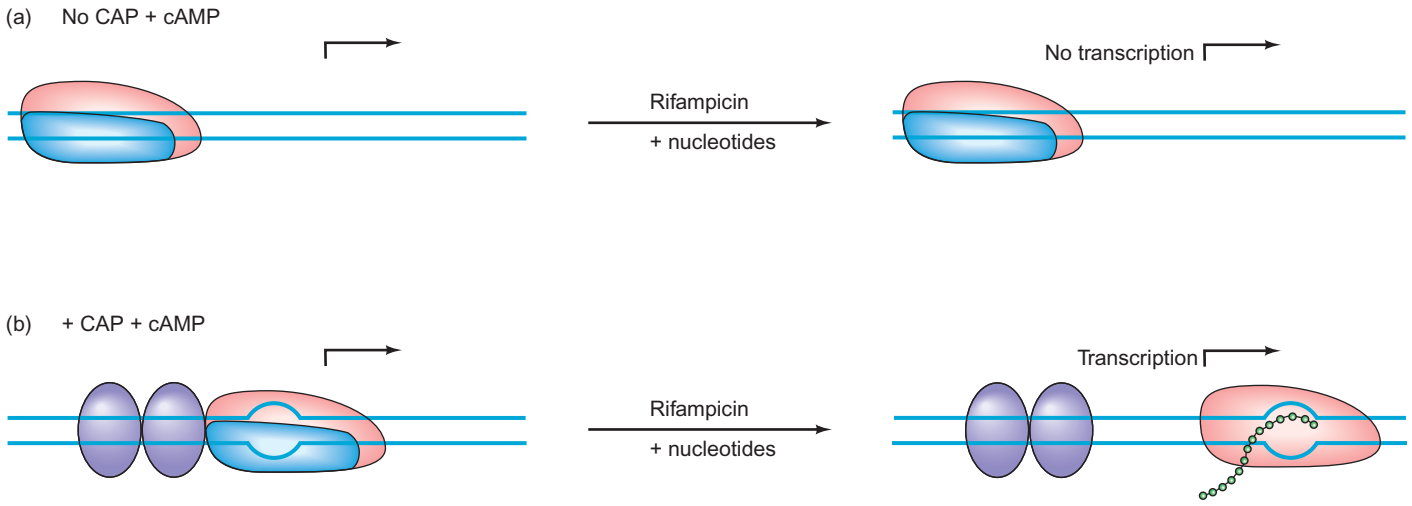


Figure 7.15 CAP plus cAMP allow formation of an open promoter complex. (a) When RNA polymerase binds to DNA containing the *lac* promoter without CAP, it binds randomly and weakly to the DNA. This binding is susceptible to inhibition when rifampicin is added along with nucleotides, so no transcription occurs. **(b)** When RNA polymerase binds to the *lac* promoter in the presence of CAP and cAMP (purple), it forms an open promoter complex. This is not susceptible to inhibition when rifampicin and nucleotides are added at the same time because the open promoter complex is ready to polymerize the nucleotides, which reach the polymerase active site before the antibiotic. Once the first few phosphodiester bonds form, the polymerase is resistant to rifampicin inhibition until it reinitiates. Thus, transcription occurs under these conditions, demonstrating that CAP and cAMP facilitate formation of an open promoter complex. The RNA is shown as a green chain.

on the right. How do we know that is the proper order? The first indication came from genetic experiments. Mutations to the left of the promoter prevent stimulation of transcription by CAP and cAMP, but still allow a low level of transcription. An example is a deletion called L1, whose position is shown in Figure 7.16. Because this deletion completely obliterates positive control of the *lac* operon by CAP and cAMP, the CAP-binding site must lie at least partially within the deleted region. On the other hand, since the L1 deletion has no effect on CAP-independent transcription, it has not encroached on the promoter, where RNA polymerase binds. Therefore, the right-hand end of this deletion serves as a rough dividing line between the activator-binding site and the promoter.

The CAP-binding sites in the *lac, gal,* and *ara* operons all contain the sequence TGTGA. The conservation of this sequence suggests that it is an important part of the CAP-binding site, and we also have direct evidence for this notion. For example, footprinting studies show that binding of the CAP–cAMP complex protects the G's in this sequence against methylation by dimethyl sulfate, suggesting that the CAP–cAMP complex binds tightly enough to these G's that it hides them from the methylating agent.

The *lac* operon, and other operons activated by CAP and cAMP, have remarkably weak promoters. Their −35 boxes are particularly unlike the consensus sequences; in fact, they are scarcely recognizable. This situation is actually not surprising. If the *lac* operon had a strong promoter, RNA polymerase could form open promoter complexes readily without help from CAP and cAMP, and it would therefore be active even in the presence of glucose. Thus, this promoter has to be weak to be dependent on CAP and

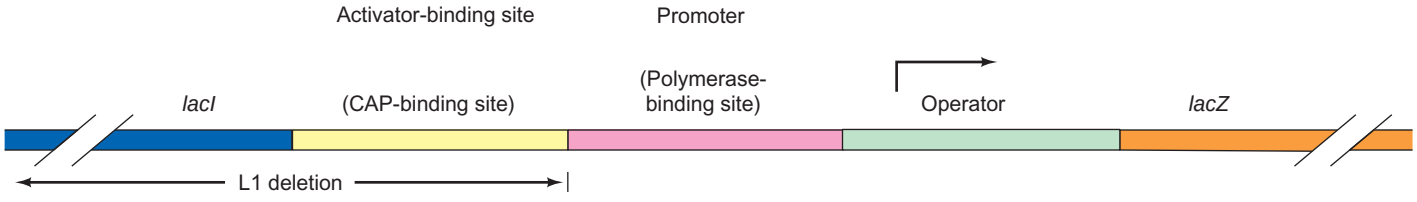


Figure 7.16 The *lac* control region. The activator–promoter region, just upstream of the operator, contains the activator-binding site, or CAP-binding site, on the left (yellow) and the promoter, or polymerase-binding site, on the right (pink). These sites have been defined by footprinting experiments and by genetic analysis. An example of the latter approach is the L1 deletion, whose right-hand end is shown. The L1 mutant shows basal transcription of the *lac* operon, but no stimulation by CAP and cAMP. Thus, it still has the promoter, but lacks the activator-binding site.

cAMP. In fact, strong mutant *lac* promoters are known (e.g., the *lac*UV5 promoter) and they do not depend on CAP and cAMP.

SUMMARY The CAP–cAMP complex stimulates transcription of the *lac* operon by binding to an activator-binding site adjacent to the promoter and helping RNA polymerase bind to the promoter.

Recruitment How does CAP–cAMP recruit polymerase to the promoter? Such **recruitment** has two steps: (1) Formation of the closed promoter complex, and (2) conversion of the closed promoter complex to the open promoter complex. William McClure and his colleagues summarized these two steps in the following equation:

$$R + P \underset{K_B}{\rightleftarrows} RP_c \xrightarrow{k_2} RP_o$$

where R is RNA polymerase, P is the promoter, RP_c is the closed promoter complex, and RP_o is the open promoter complex. McClure and coworkers devised kinetic methods of distinguishing between the two steps and determined that CAP–cAMP acts directly to stimulate the first step by increasing K_B. CAP–cAMP has little if any effect on k_2, so the second step is not accelerated. Nevertheless, by increasing the rate of formation of the closed promoter complex, CAP–cAMP provides more raw material (closed promoter complex) for conversion to the open promoter complex. Thus, the net effect of CAP–cAMP is to increase the rate of open promoter complex formation.

How does binding CAP–cAMP to the activator-binding site facilitate binding of polymerase to the promoter? One long-standing hypothesis is that CAP and RNA polymerase actually touch as they bind to their respective DNA target sites and therefore they bind cooperatively.

This hypothesis has much experimental support. First, CAP and RNA polymerase cosediment on ultracentrifugation in the presence of cAMP, suggesting that they have an affinity for each other. Second, CAP and RNA polymerase, when both are bound to their DNA sites, can be chemically cross-linked to each other, suggesting that they are in close proximity. Third, DNase footprinting experiments (Chapter 5) show that the CAP–cAMP footprint lies adjacent to the polymerase footprint. Thus, the DNA binding sites for these two proteins are close enough that the proteins could interact with each other as they bind to their DNA sites. Fourth, several CAP mutations decrease activation without affecting DNA binding (or bending), and some of these mutations alter amino acids in the region of CAP (**activation region I [ARI]**) that is thought to interact with polymerase. Fifth, the polymerase site that is presumed to interact with ARI on CAP is the carboxyl terminal domain of the α-subunit (the αCTD), and deletion of the αCTD prevents activation by CAP–cAMP.

Sixth, Richard Ebright and colleagues performed x-ray crystallography in 2002 on a complex of DNA, CAP– cAMP, and the αCTD of RNA polymerase. They showed that the ARI site on CAP and the αCTD do indeed touch in the crystal structure, although the interface between the two proteins is not large. They arranged for the αCTD to bind on its own to the complex by changing the sequences flanking the CAP-binding site to A–T-rich sequences (5′-AAAAAA-3′) that are attractive to the αCTD. Figure 7.17a presents the crystal structure they determined. One molecule of αCTD (αCTD^{DNA}) binds to DNA alone; the other molecule ($\alpha CTD^{CAP,DNA}$) binds to both DNA and CAP. The latter αCTD clearly contacts the part of CAP identified as ARI, and detailed analysis of the structure showed exactly which amino acids in each protein were involved in the interaction. The fact that only one monomer of αCTD binds to a monomer of CAP reflects the situation in vivo; the other monomer of αCTD does not contact CAP either in the crystal structure or in vivo.

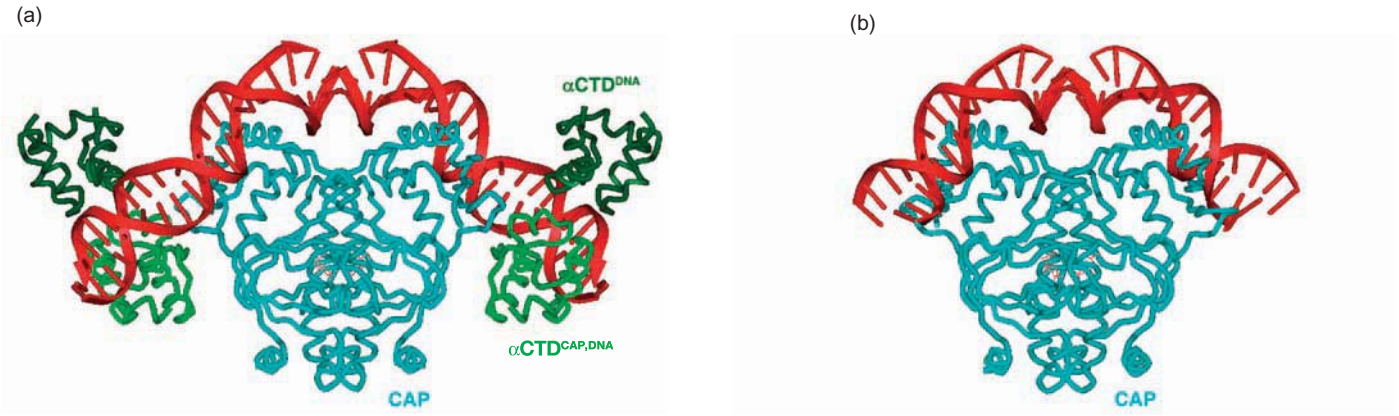


Figure 7.17 Crystal structures of the CAP–cAMP–αCTD–DNA complex and the CAP–cAMP–DNA complex. (a) The CAP–cAMP–αCTD–DNA complex. DNA is in red, CAP is in cyan, with cAMP represented by thin red lines, αCTD^{DNA} is in dark green, and $\alpha CTD^{CAP,DNA}$ is in light green. **(b)** CAP–cAMP–DNA complex. Same colors as in panel (a). (*Source:* Benoff et al., *Science* 297 © 2002 by the AAAS.)

Another thing to notice about Figure 7.17a is that binding of CAP–cAMP to its DNA target bends the DNA considerably—about 100 degrees. This bend had been noticed before in the crystal structure of the CAP–cAMP–DNA complex in the absence of αCTD, determined by Thomas Steitz and colleagues in 1991, and can be seen again in an equivalent crystal structure determined in this study (Figure 7.17b). It is interesting that the structure of the DNA and CAP in the CAP–cAMP–DNA complex and in the CAP–cAMP–DNA–αCTD complex are superimposable. This means that the αCTD did not perturb the structure.

The DNA bend observed in the crystallography studies had been detected as early as 1984 by Hen-Ming Wu and Donald Crothers, using electrophoresis (Figure 7.18). When a piece of DNA is bent, it migrates more slowly during electrophoresis. Furthermore, as Figure 7.18b and c illustrate, the closer the bend is to the middle of the DNA, the more slowly the DNA electrophoreses. Wu and Crothers took advantage of this phenomenon by preparing DNA fragments of the *lac* operon, all the same length, with the CAP-binding site located at different positions in each. Next, they bound CAP–cAMP to each fragment and electrophoresed the DNA–protein complexes. If CAP binding really did bend the DNA, then the different fragments should have migrated at different rates. If the DNA did not bend, they all should have migrated at the same rate. Figure 7.18d demonstrates that the fragments really did migrate at different rates. Moreover, the more pronounced the DNA bend, the greater the difference in electrophoretic rates should be. In other words, the shape of the curve in Figure 7.18 should give us an estimate of the degree of bending of DNA by CAP–cAMP. In fact the bending seems to be about 90 degrees, which agrees reasonably well with

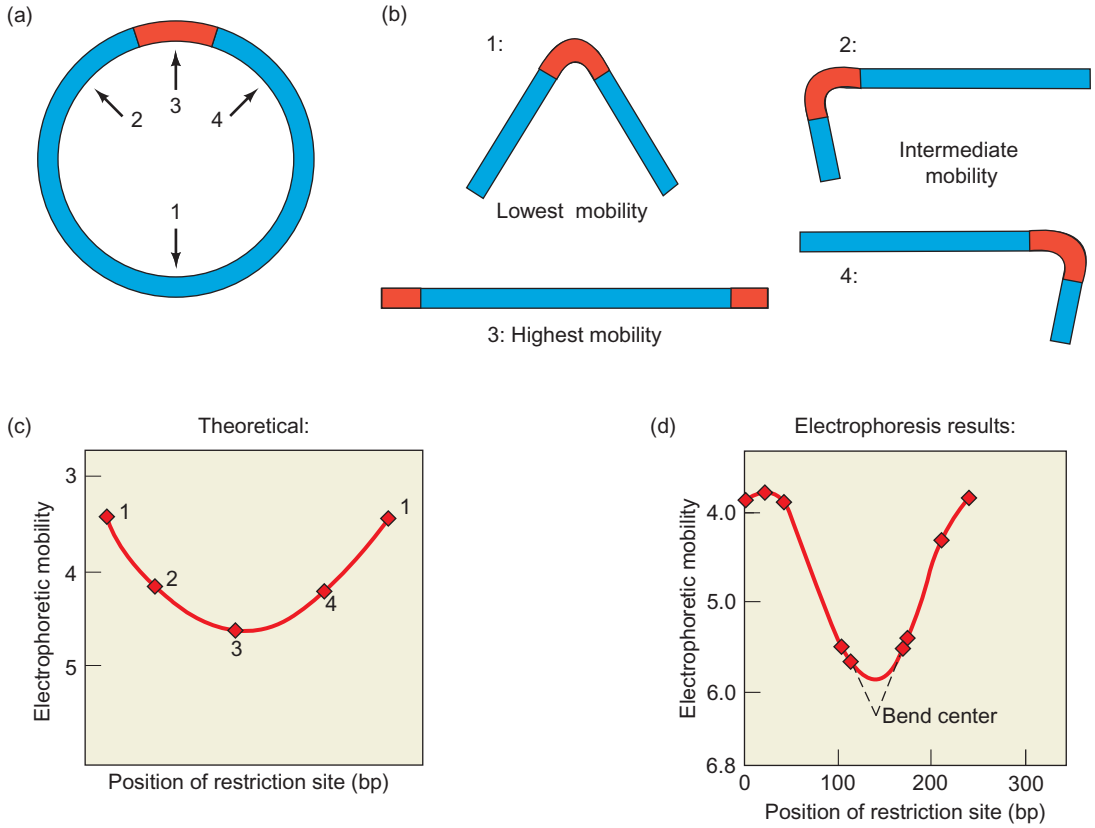


Figure 7.18 Electrophoresis of CAP–cAMP–promoter complexes. **(a)** Map of a hypothetical DNA circle, showing a protein-binding site at center (red), and cutting sites for four different restriction enzymes (arrows). **(b)** Results of cutting DNA in panel (a) with each restriction enzyme, then adding a DNA-binding protein, which bends DNA. Restriction enzyme 1 cuts across from the binding site, leaving it in the middle; restriction enzymes 2 and 4 place the binding site off center; and restriction enzyme 3 cuts within the binding site, allowing little if any bending of the DNA. **(c)** Theoretical curve showing the relationship between electrophoretic mobility and bent DNA, with the bend at various sites along the DNA. Note that the mobility is lowest when the bend is closest to the middle of the DNA fragment (at either end of the curve). Note also that mobility increases in the downward direction on the *y* axis. **(d)** Actual electrophoresis results with CAP–cAMP and DNA fragments containing the *lac* promoter at various points in the fragment, depending on which restriction enzyme was used to cut the DNA. The symmetrical curve allowed Wu and Crothers to extrapolate to a bend center that corresponds to the CAP–cAMP-binding site in the *lac* promoter. (*Source:* Wu, H.M., and D.M. Crothers, The locus of sequence-directed and protein-induced DNA bending. *Nature* 308:511, 1984.)

Table 7.2 Activation of *lac* P1 Transcription by CAP–cAMP

	Transcripts (cpm)						
	−cAMP–CAP		+cAMP–CAP		P1/UV5 (%)		
Enzyme	P1	UV5	P1	UV5	−cAMP–CAP	+cAMP–CAP	Activation (fold)
α-WT	46	797	625	748	5.8	83.6	14.4
α-256	53	766	62	723	6.9	8.6	1.2
α-235	51	760	45	643	6.7	7.0	1.0

the 100 degrees determined later by x-ray crystallography. This bending is presumably necessary for optimal interaction among the proteins and DNA in the complex.

All of the studies we have cited point to the importance of protein–protein interaction between CAP and RNA polymerase—the αCTD of polymerase, in particular. This hypothesis predicts that mutations that remove the αCTD should prevent transcription stimulation by CAP–cAMP. In fact, Kazuhiko Igarashi and Akira Ishihama have provided such genetic evidence for the importance of the αCTD of RNA polymerase in activation by CAP–cAMP. They transcribed cloned *lac* operons in vitro with RNA polymerases reconstituted from separated subunits. All the subunits were wild-type, except in some experiments, in which the α-subunit was a truncated version lacking the CTD. One of the truncated α-subunits ended at amino acid 256 (of the normal 329 amino acids); the other ended at amino acid 235. Table 7.2 shows the results of run-off transcription (Chapter 5) from a CAP–cAMP-dependent *lac* promoter (P1) and a CAP–cAMP-independent *lac* promoter (*lac*UV5) with reconstituted polymerases containing the wild-type or truncated α-subunits in the presence and absence of CAP–cAMP. As expected, CAP–cAMP did not stimulate transcription from the *lac*UV5 promoter because it is a strong promoter that is CAP–cAMP-insensitive. Also as expected, transcription from the *lac* P1 promoter was stimulated over 14-fold by CAP–cAMP. But the most interesting behavior was that of the polymerases reconstituted with truncated α-subunits. These enzymes were just as good as wild-type in transcribing from either promoter in the absence of CAP–cAMP, but they could not be stimulated by CAP–cAMP. Thus, the αCTD, missing in these truncated enzymes, is not necessary for reconstitution of an active RNA polymerase, but it is necessary for stimulation by CAP–cAMP.

Figure 7.19 illustrates the hypothesis of activation we have been discussing, in which the CAP–cAMP dimer binds to its activator site and simultaneously binds to the carboxyl-terminal domain of the polymerase α-subunit (αCTD), facilitating binding of polymerase to the promoter. This would be the functional equivalent of the αCTD binding to an UP element in the DNA (Chapter 6), thereby enhancing polymerase binding to the promoter.

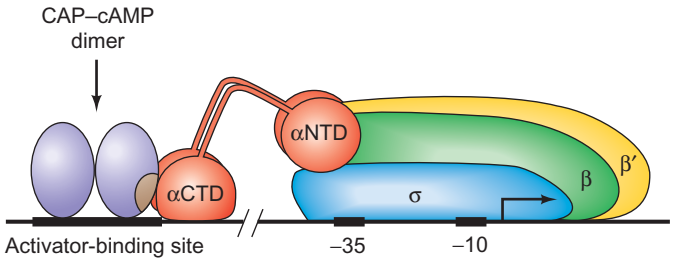


Figure 7.19 Hypothesis for CAP–cAMP activation of *lac* transcription. The CAP–cAMP dimer (purple) binds to its target site on the DNA, and the αCTD (red) interacts with a specific site on the CAP protein (brown). This strengthens binding between polymerase and promoter. (*Source:* Adapted from Busby, S. and R.H. Ebright, Promoter structure, promoter recognition, and transcription activation in prokaryotes, *Cell* 79:742, 1994.)

CAP stimulates transcription at over 100 promoters, and it is just one of a growing number of bacterial transcription activators. We will examine more examples in Chapter 9.

SUMMARY CAP–cAMP binding to the *lac* activator-binding site recruits RNA polymerase to the adjacent *lac* promoter to form a closed promoter complex. This closed complex then converts to an open promoter complex. CAP–cAMP causes recruitment through protein–protein interaction with the αCTD of RNA polymerase. CAP–cAMP also bends its target DNA by about 100 degrees when it binds.

7.2 The *ara* Operon

We have already mentioned that the *ara* operon of *E. coli*, which codes for the enzymes required to metabolize the sugar arabinose, is another catabolite-repressible operon. It has several interesting features to compare with the *lac* operon. First, two *ara* operators exist: *ara*O_1 and *ara*O_2. The former regulates transcription of a control gene called *araC*.

The other operator is located far upstream of the promoter it controls (P_{BAD}), between positions −265 and −294, yet it still governs transcription. Second, the CAP-binding site is about 200 bp upstream of the *ara* promoter, yet CAP can still stimulate transcription. Third, the operon has another system of negative regulation, mediated by the AraC protein.

The *ara* Operon Repression Loop

How can *ara*O_2 control transcription from a promoter over 250 bp downstream? The most reasonable explanation is that the DNA in between these remote sites (the operator and the promoter) loops out as illustrated in Figure 7.20a. Indeed, we have good evidence that DNA looping is occurring. Robert Lobell and Robert Schleif found that if they inserted DNA fragments containing an integral number of double-helical turns (multiples of 10.5 bp) between the operator and the promoter, the operator still functioned. However, if the inserts contained a nonintegral number of helical turns (e.g., 5 or 15 bp), the operator did not function. This is consistent with the general notion that a double-stranded DNA can loop out and bring two protein-binding sites together as long as these sites are located on the same face of the double helix. However, the DNA cannot twist through the 180 degrees required to bring binding sites on opposite faces around to the same face so they can interact with each other through looping (see Figure 7.20). In this respect, DNA resembles a piece of stiff coat hanger wire: It can be bent relatively easily, but it resists twisting.

The simple model in Figure 7.20 assumes that proteins bind first to the two remote binding sites, then these proteins interact to cause the DNA looping. However, Lobell and Schleif found that the situation is more subtle than that. In fact, the *ara* control protein (**AraC**), which acts as both a positive and a negative regulator, has three binding sites, as illustrated in Figure 7.21a. In addition to the far upstream site, *ara*O_2, AraC can bind to *ara*O_1, located between positions −106 and −144, and to *araI,* which really includes two half-sites: *araI*$_1$ (−56 to −78) and *araI*$_2$ (−35 to −51), each of which can bind one monomer of AraC. The *ara* operon is also known as the *araCBAD* operon, for its four genes, *araA–D*. Three of these genes, *araB, A,* and *D,* encode the arabinose metabolizing enzymes; they are transcribed rightward from the promoter *ara*P_{BAD}. The other gene, *araC,* encodes the control protein AraC and is transcribed leftward from the *ara*P_C promoter.

In the absence of arabinose, when no *araBAD* products are needed, AraC exerts negative control, binding to *ara*O_2 and *araI*$_1$, looping out the DNA in between and repressing the operon (Figure 7.21b). On the other hand, when arabinose is present, it apparently changes the conformation of AraC so that it no longer binds to *ara*O_2, but occupies *araI*$_1$ and *araI*$_2$ instead. This breaks the repression loop, and the operon is derepressed (Figure 7.21c). As in the *lac* operon, however, derepression isn't the whole story. Positive control mediated by CAP and cAMP also occurs, and Figure 7.21c shows this complex attached to its binding site upstream of the *araBAD* promoter. DNA looping presumably explains how binding of CAP–cAMP at a site remote from the *araBAD* promoter can control transcription. The looping would allow CAP to contact the polymerase and thereby stimulate its binding to the promoter.

Evidence for the *ara* Operon Repression Loop

What is the evidence for the looping model of *ara* operon repression? First, Lobell and Schleif used electrophoresis to show that AraC can cause loop formation in the absence

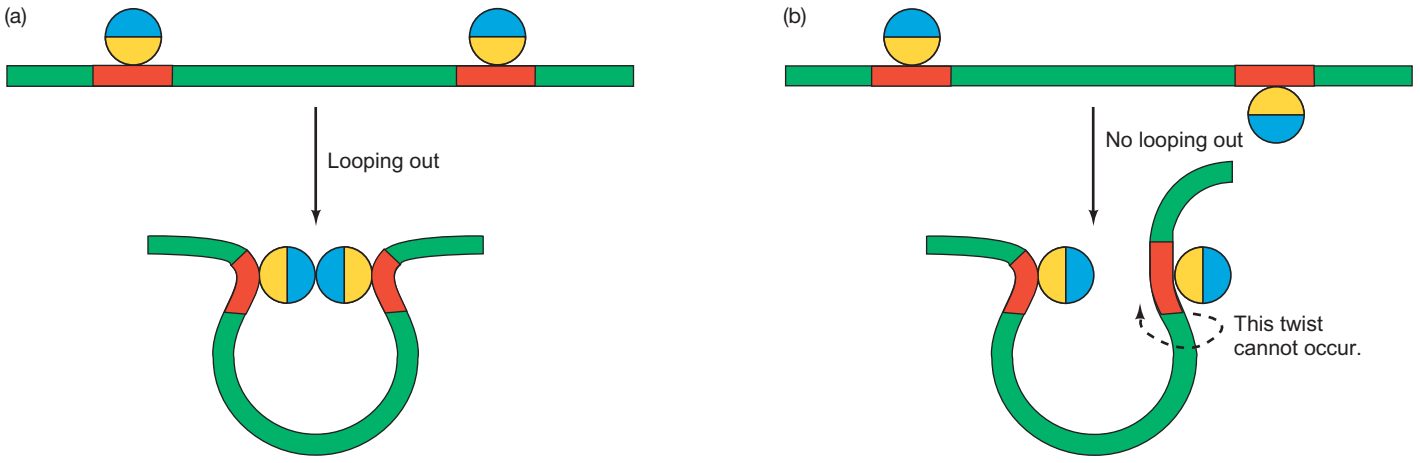


Figure 7.20 Proteins must bind to the same face of the DNA to interact by looping out the DNA. (a) Two proteins with DNA-binding domains (yellow) and protein–protein interaction domains (blue) bind to sites (red) on the same face of the DNA double helix. These proteins can interact because the intervening DNA can loop out without twisting. **(b)** Two proteins bind to sites on opposite sides of the DNA duplex. These proteins cannot interact because the DNA is not flexible enough to perform the twist needed to bring the protein interaction sites together.

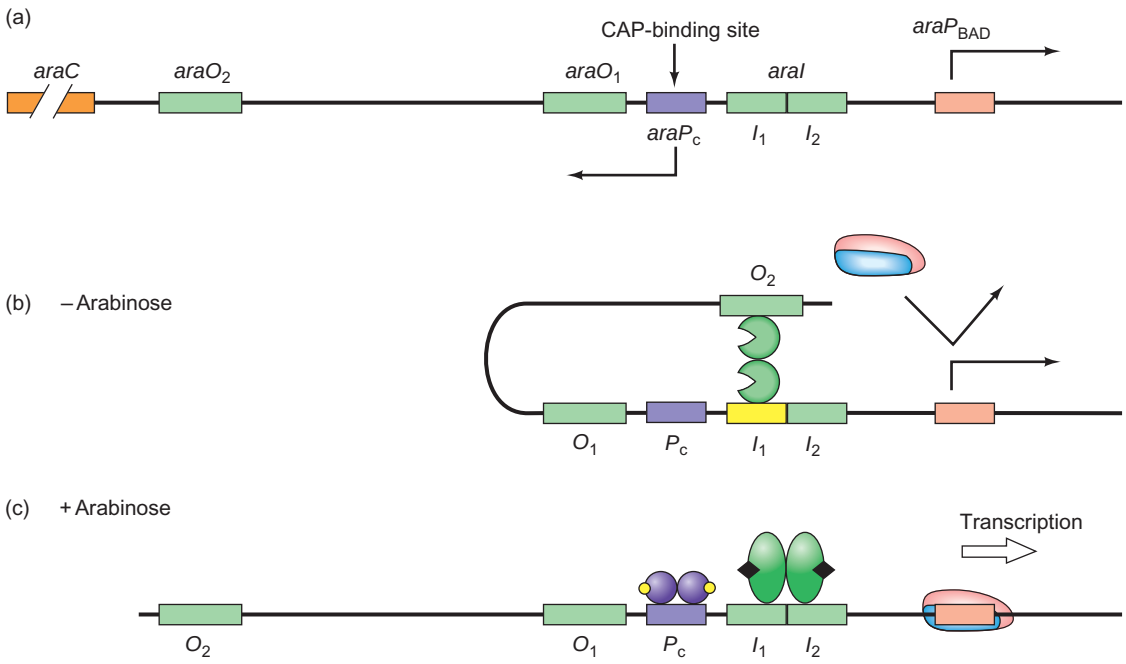


Figure 7.21 Control of the *ara* operon. (a) Map of the *ara* control region. There are four AraC-binding sites (*araO*$_1$, *araO*$_2$, *araI*$_1$, and *araI*$_2$), which all lie upstream of the *ara* promoter, *araP*$_{BAD}$. The *araP*$_c$ promoter drives leftward transcription of the *araC* gene at far left. **(b)** Negative control. In the absence of arabinose, monomers of AraC (green) bind to O_2 and I_1, bending the DNA and blocking access to the promoter by RNA polymerase (red and blue). **(c)** Positive control. Arabinose (black) binds to AraC, changing its shape so it prefers to bind as a dimer to I_1 and I_2 and not to O_2. This opens up the promoter (pink) to binding by RNA polymerase. If glucose is absent, the CAP–cAMP complex (purple and yellow) is in high enough concentration to occupy the CAP-binding site, which stimulates polymerase binding to the promoter. Now active transcription can occur.

of arabinose. Instead of the entire *E. coli* DNA, they used a small (404-bp) supercoiled circle of DNA, called a *minicircle*, that contained the *araO*$_2$ and *araI* sites, 160 bp apart. They then added AraC and measured looping by taking advantage of the fact that looped supercoiled DNAs have a higher electrophoretic mobility than the same DNAs that are unlooped. Figure 7.22 shows one such assay. Comparing lanes 1 and 2, we can see that the addition of AraC causes the appearance of a new, high-mobility band that corresponds to the looped minicircle.

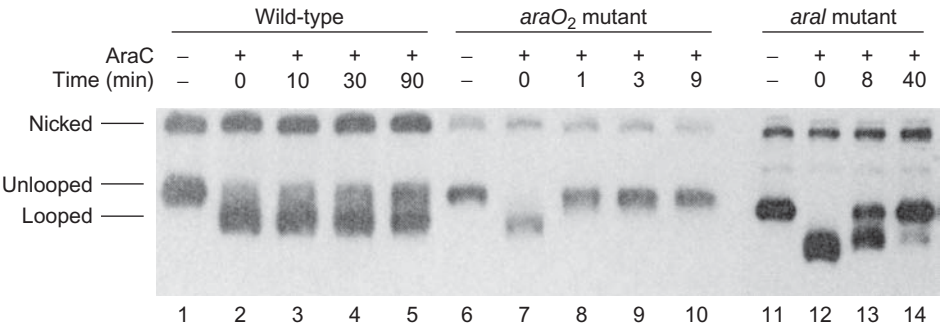


Figure 7.22 Effects of mutations in *araO*$_2$ and *araI* on the stability of looped complexes with AraC. Lobell and Schleif prepared labeled minicircles (small DNA circles) containing either wild-type or mutant AraC binding sites, as indicated at top. Then they added AraC to form a complex with the labeled DNA. Next they added an excess of unlabeled DNA containing an *araI* site as a competitor, for various lengths of time. Finally they electrophoresed the protein–DNA complexes to see whether they were still in looped or unlooped form. The looped DNA was more supercoiled than the unlooped DNA, so it migrated faster. The wild-type DNA remained in a looped complex even after 90 min in the presence of the competitor. By contrast, dissociation of AraC from the mutant DNAs, and therefore loss of the looped complex, occurred much faster. It lasted less than 1 min with the *araO*$_2$ mutant DNA and was half gone in less than 10 min with the *araI* mutant DNA. (*Source:* Lobell, R.B. and Schleif, R.F., DNA looping and unlooping by AraC protein. *Science* 250 (1990), f. 2, p. 529. © AAAS.)

This experiment also shows that the stability of the loop depends on binding of AraC to both *araO*$_2$ and *araI*. Lobell and Schleif made looped complexes with a wild-type minicircle, with a minicircle containing a mutant *araO*$_2$ site, and with a minicircle containing mutations in both *araI* sites. They then added an excess of unlabeled wild-type minicircles and observed the decay of each of the looped complexes. Lanes 3–5 show only about 50% conversion of the looped to unlooped wild-type minicircle in 90 min. Thus, the half-time of dissociation of the wild-type looped complex is about 100 min. In contrast, the *araO*$_2$ mutant minicircle's conversion from looped to unlooped took less than 1 min (compare lanes 7 and 8). The *araI* mutant's half-time of loop breakage is also short—less than 10 min. Thus, both *araO*$_2$ and *araI* are involved in looping by AraC because mutations in either one greatly weaken the DNA loop.

Next, Lobell and Schleif demonstrated that arabinose breaks the repression loop. They did this by showing that arabinose added to looped minicircles immediately before electrophoresis eliminates the band corresponding to the looped DNA. Figure 7.23 illustrates this phenomenon. In a separate experiment, Lobell and Schleif showed that a broken loop could re-form if arabinose was removed. They used arabinose to prevent looping, then diluted the DNA into buffer containing excess competitor DNA, either with or without arabinose. The buffer with arabinose maintained the broken loop, but the buffer without arabinose diluted the sugar to such an extent that the loop could re-form.

What happens to the AraC monomer bound to *araO*$_2$ when the loop opens up? Apparently it binds to *araI*$_2$. To demonstrate this, Lobell and Schleif first showed by **methylation interference** that AraC contacts *araI*$_1$, but not *araI*$_2$, in the looped state. The strategy was to partially methylate the minicircle DNA, bind AraC to loop the DNA, separate looped from unlooped DNA by electrophoresis, and then break the looped and unlooped DNAs at their methylated sites. Because methylation at important sites blocks looping, those sites that are important for looping will be unmethylated in the looped DNA, but methylated in the unlooped DNA. Indeed, two *araI*$_1$ bases were heavily methylated in the unlooped DNA, but only lightly methylated in the looped DNA. In contrast, no *araI*$_2$ bases showed this behavior. Thus, it appears that AraC does not contact *araI*$_2$ in the looped state.

Lobell and Schleif confirmed this conclusion by showing that mutations in *araI*$_2$ have no effect on AraC binding in the looped state, but have a strong effect on binding in the unlooped state. We infer that *araI*$_2$ is necessary for AraC binding in the unlooped state and is therefore contacted by AraC under these conditions.

These data suggest the model of AraC–DNA interaction that was depicted in Figure 7.21b and c. A dimer of AraC causes looping by simultaneously interacting with *araI*$_1$ and *araO*$_2$. Arabinose breaks the loop by changing the conformation of AraC so the protein loses its affinity for *araO*$_2$ and binds instead to *araI*$_2$.

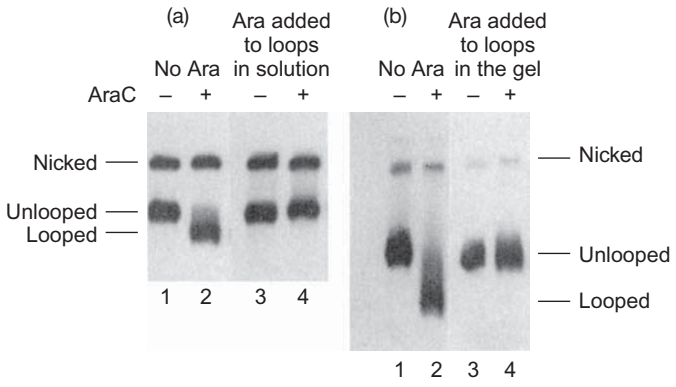


Figure 7.23 Arabinose breaks the loop between *araO*$_2$ and *araI*. **(a)** Lobell and Schleif added arabinose to preformed loops before electrophoresis. In the absence of arabinose, AraC formed a DNA loop (lane 2). In the presence of arabinose, the loop formed with AraC was broken (lane 4). **(b)** This time the investigators added arabinose to the gel after electrophoresis started. Again, in the absence of arabinose, looping occurred (lane 2). However, in the presence of arabinose, the loop was broken (lane 4). The designation Ara at top refers to arabinose. (*Source:* Lobell R.B., and Schleif R.F., DNA looping and unlooping by AraC protein. *Science* 250 (1990), f. 4, p. 530. © AAAS.)

Autoregulation of *araC*

So far, we have only briefly mentioned a role for *araO*$_1$. It does not take part in repression of *araBAD* transcription; instead it allows AraC to regulate its own synthesis. Figure 7.24 shows the relative positions of *araC*, P_c, and *araO*$_1$. The *araC* gene is transcribed from P_c in the leftward direction, which puts *araO*$_1$ in a position to control this transcription. As the level of AraC rises, it binds to *araO*$_1$ and inhibits leftward transcription, thus preventing an accumulation of too much repressor. This kind of mechanism, where a protein controls its own synthesis, is called **autoregulation.**

SUMMARY The *ara* operon is controlled by the AraC protein. AraC represses the operon by looping out the DNA between two sites, *araO*$_2$ and *araI*$_1$, that are 210 bp apart. Arabinose can derepress the operon by causing AraC to loosen its attachment to *araO*$_2$ and to bind to *araI*$_2$ instead. This breaks the loop and allows transcription of the operon. CAP and cAMP further stimulate transcription by binding to a site upstream of *araI*. AraC controls its own synthesis by binding to *araO*$_1$ and preventing leftward transcription of the *araC* gene.

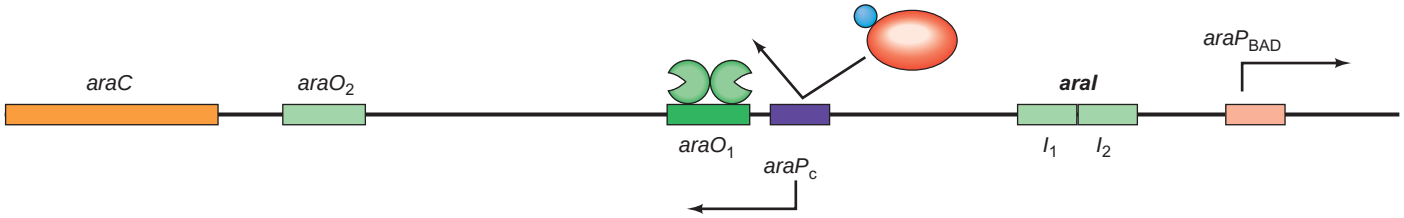


Figure 7.24 Autoregulation of *araC*. AraC (green) binds to $araO_1$ and prevents transcription leftward from P_c through the *araC* gene. This can presumably happen whether or not arabinose is bound to AraC, that is, with the control region either unlooped or looped.

7.3 The *trp* Operon

The *E. coli* ***trp*** (pronounced "trip") **operon** contains the genes for the enzymes that the bacterium needs to make the amino acid tryptophan. Like the *lac* operon, it is subject to negative control by a repressor. However, there is a fundamental difference. The *lac* operon codes for **catabolic** enzymes—those that break down a substance. Such operons tend to be turned on by the presence of that substance, lactose in this case. The *trp* operon, on the other hand, codes for **anabolic** enzymes—those that build up a substance. Such operons are generally turned off by that substance. When the tryptophan concentration is high, the products of the *trp* operon are not needed any longer, and we would expect the *trp* operon to be repressed. That is what happens. The *trp* operon also exhibits an extra level of control, called attenuation, not seen in the *lac* operon.

Tryptophan's Role in Negative Control of the *trp* Operon

Figure 7.25 shows an outline of the structure of the *trp* operon. Five genes code for the polypeptides in the enzymes that convert a tryptophan precursor, chorismic acid, to tryptophan. In the *lac* operon, the promoter and operator precede the genes, and the same is true in the *trp* operon. However, the *trp* operator lies wholly within the *trp* promoter, whereas the two loci are merely adjacent in the *lac* operon.

In the negative control of the *lac* operon, the cell senses the presence of lactose by the appearance of tiny amounts of its rearranged product, allolactose. In effect, this inducer causes the repressor to fall off the *lac* operator and derepresses the operon. In the case of the *trp* operon, a plentiful supply of tryptophan means that the cell does not need to spend any more energy making this amino acid. In other words, a high tryptophan concentration is a signal to turn off the operon.

How does the cell sense the presence of tryptophan? In essence, tryptophan helps the *trp* repressor bind to its operator. Here is how that occurs: In the absence of tryptophan, no *trp* repressor exists—only an inactive protein called the **aporepressor.** When the aporepressor binds tryptophan, it changes to a conformation with a much higher affinity for the *trp* operator (Figure 7.25b). This is another allosteric

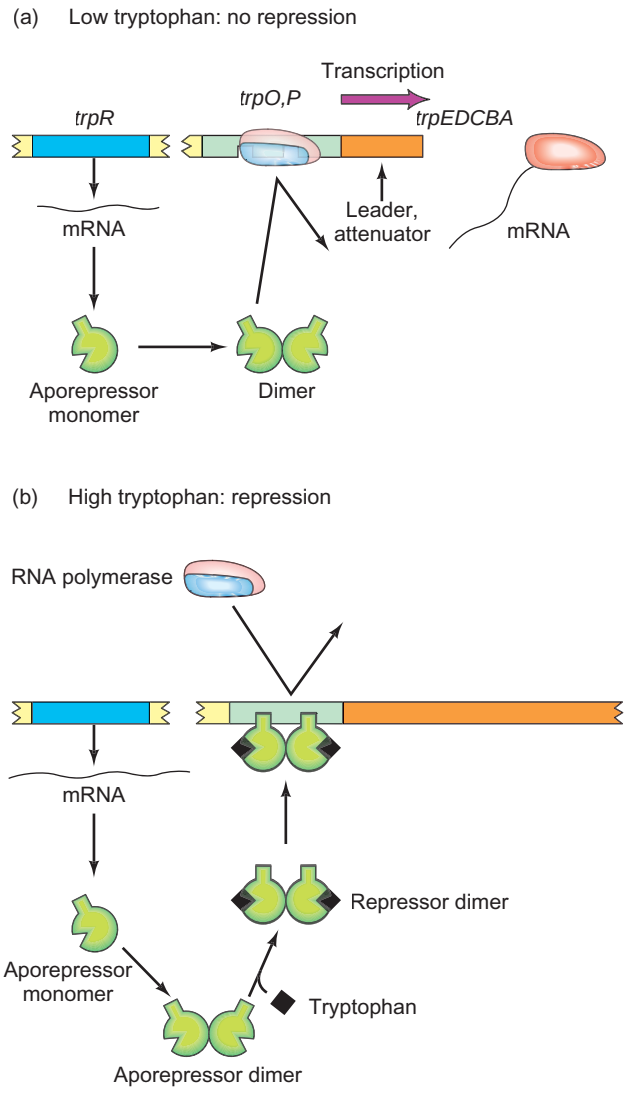


Figure 7.25 Negative control of the *trp* operon. (a) Derepression. RNA polymerase (red and blue) binds to the *trp* promoter and begins transcribing the *trp* genes (*trpE, D, C, B,* and *A*). Without tryptophan, the aporepressor (green) cannot bind to the operator. **(b)** Repression. Tryptophan, the corepressor (black), binds to the inactive aporepressor, changing it to repressor, with the proper shape for binding successfully to the *trp* operator. This prevents RNA polymerase from binding to the promoter, so no transcription occurs.

transition like the one we encountered in our discussion of the *lac* repressor. The combination of aporepressor plus tryptophan is the ***trp* repressor;** therefore, tryptophan is called a **corepressor.** When the cellular concentration of tryptophan is high, plenty of corepressor is available to bind and form the active *trp* repressor. Thus, the operon is repressed. When the tryptophan level in the cell falls, the amino acid dissociates from the aporepressor, causing it to shift back to the inactive conformation; the repressor–operator complex is thus broken, and the operon is derepressed. In Chapter 9, we will examine the nature of the conformational shift in the aporepressor that occurs on binding tryptophan and see why this is so important in operator binding.

SUMMARY The negative control of the *trp* operon is, in a sense, the mirror image of the negative control of the *lac* operon. The *lac* operon responds to an inducer that causes the repressor to dissociate from the operator, derepressing the operon. The *trp* operon responds to a repressor that includes a corepressor, tryptophan, which signals the cell that it has made enough of this amino acid. The corepressor binds to the aporepressor, changing its conformation so it can bind better to the *trp* operator, thereby repressing the operon.

Control of the *trp* Operon by Attenuation

In addition to the standard, negative control scheme we have just described, the *trp* operon employs another mechanism of control called **attenuation.** Why is this extra control needed? The answer probably lies in the fact that repression of the *trp* operon is weak—much weaker, for example, than that of the *lac* operon. Thus, considerable transcription of the *trp* operon can occur even in the presence of repressor. In fact, in attenuator mutants where only repression can operate, the fully repressed level of transcription is only 70-fold lower than the fully derepressed level. The attenuation system permits another 10-fold control over the operon's activity. Thus, the combination of repression and attenuation controls the operon over a 700-fold range, from fully inactive to fully active: (70-fold [repression] × 10-fold [attenuation] = 700-fold). This is valuable because synthesis of tryptophan requires considerable energy.

Here is how attenuation works. Figure 7.25 lists two loci, the ***trp* leader** and the ***trp* attenuator,** in between the operator and the first gene, *trpE*. Figure 7.26 gives a closer view of the leader–attenuator, whose purpose is to attenuate, or weaken, transcription of the operon when tryptophan is relatively abundant. The attenuator operates by causing premature termination of transcription. In other words, transcription that gets started, even though the tryptophan concentration is high, stands a 90% chance of terminating in the attenuator region.

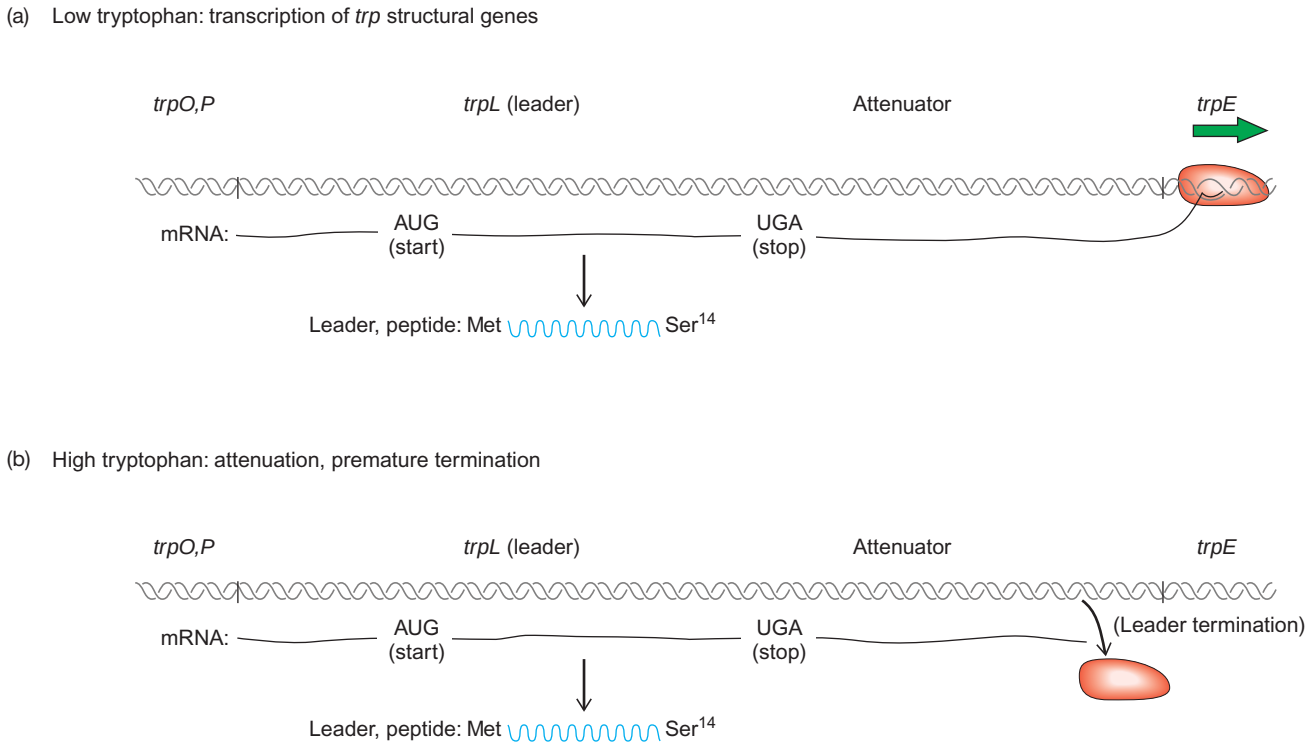


Figure 7.26 Attenuation in the *trp* operon. (a) Under low tryptophan conditions, the RNA polymerase (red) reads through the attenuator, so the structural genes are transcribed. **(b)** In the presence of high tryptophan, the attenuator causes premature termination of transcription, so the *trp* genes are not transcribed.

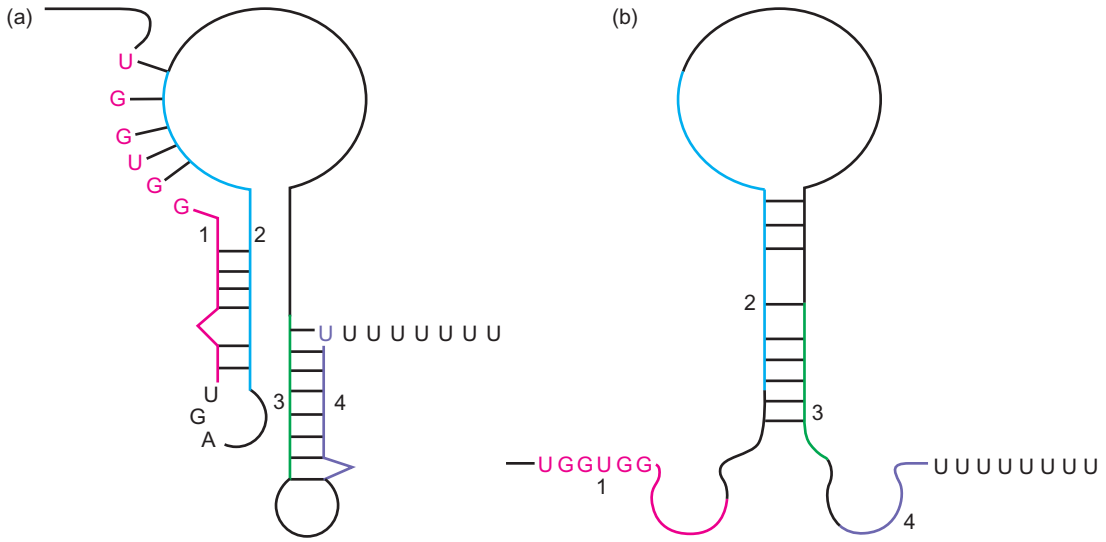


Figure 7.27 Two structures available to the leader–attenuator transcript. (a) The more stable structure, with two hairpin loops. **(b)** The less stable structure, containing only one hairpin loop. The curved shape of the RNA at the bottom is not meant to suggest a shape for the molecule—it is drawn this way simply to save space. The base-paired segments (1–4) in (a) are colored, and these same regions are colored the same way in (b) so they can be recognized.

The reason for this premature termination is that the attenuator contains a transcription stop signal (terminator): an inverted repeat followed by a string of eight A–T pairs in a row. Because of the inverted repeat, the transcript of this region would tend to engage in intramolecular base pairing, forming a "hairpin". As we learned in Chapter 6, a hairpin followed by a string of U's in a transcript destabilizes the binding between the transcript and the DNA and thus causes termination.

SUMMARY Attenuation imposes an extra level of control on an operon, over and above the repressor–operator system. It operates by causing premature termination of transcription of the operon when the operon's products are abundant.

Defeating Attenuation

When tryptophan is scarce, the *trp* operon must be activated, and that means that the cell must somehow override attenuation. Charles Yanofsky proposed this hypothesis: Something preventing the hairpin from forming would destroy the termination signal, so attenuation would break down and transcription would proceed. A look at Figure 7.27a reveals not just one potential hairpin near the end of the leader transcript, but two. However, the terminator includes only the second hairpin, which is adjacent to the string of U's in the transcript. Furthermore, the two-hairpin arrangement is not the only one available; another, containing only one hairpin, is shown in Figure 7.27b. Note that this alternative hairpin contains elements from each of the two hairpins in the first structure. Figure 7.27 illustrates this concept by labeling the sides of the original two hairpins 1, 2, 3, and 4. If the first of the original hairpins involves elements 1 and 2 and the second involves 3 and 4, then the alternative hairpin in the second structure involves 2 and 3. This means that the formation of the alternative hairpin (Figure 7.27b) precludes formation of the other two hairpins, including the one adjacent to the string of U's, which is a necessary part of the terminator (Figure 7.27a).

The two-hairpin structure involves more base pairs than the alternative, one-hairpin structure; therefore, it is more stable. So why should the less stable structure ever form? A clue comes from the base sequence of the leader region shown in Figure 7.28. One very striking feature of this sequence is that two codons for tryptophan (UGG) occur in a row in element 1 of the first potential hairpin. This

Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser Stop
pppA---AUGAAAGCAAUUUUCGUACUGAAAGGUUGGUGGCGCACUUCCUGA

Figure 7.28 Sequence of the leader. The sequence of part of the leader transcript is presented, along with the leader peptide it encodes. Note the two Trp codons in tandem (blue).

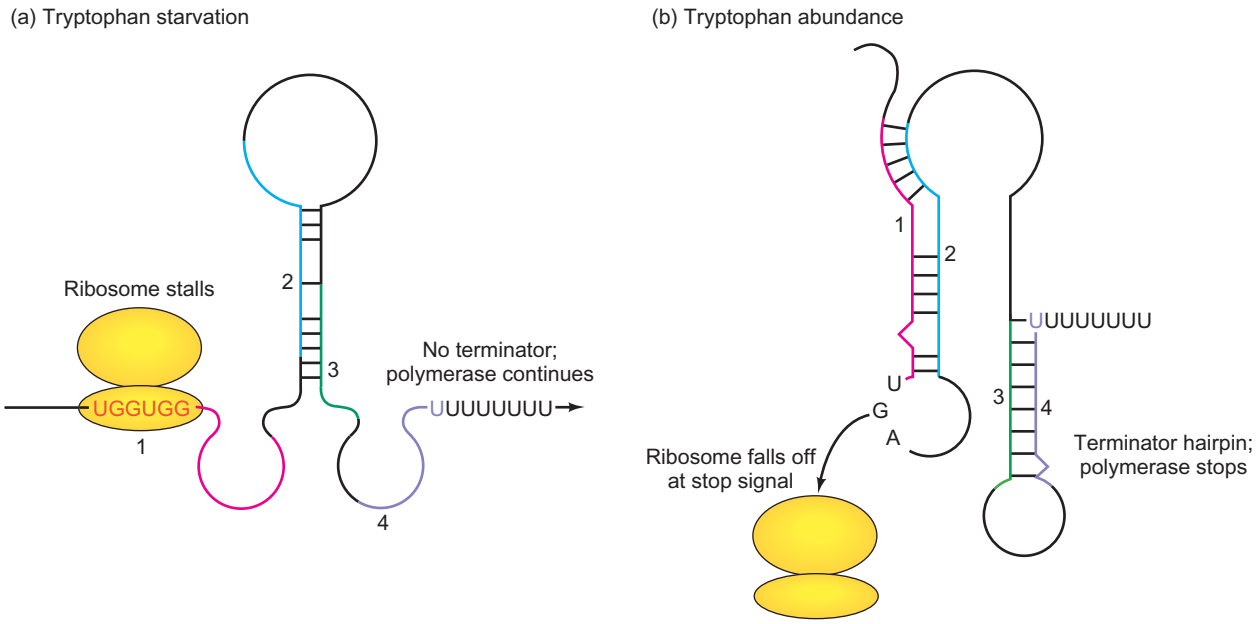


Figure 7.29 Overriding attenuation. (a) Under conditions of tryptophan starvation, the ribosome (yellow) stalls at the Trp codons and prevents element 1 (red) from pairing with element 2 (blue). This forces the one-hairpin structure, which lacks a terminator, to form, so no attenuation should take place. **(b)** Under conditions of tryptophan abundance, the ribosome reads through the two tryptophan codons and falls off at the translation stop signal (UGA), so it cannot interfere with base pairing in the leader transcript. The more stable, two-hairpin structure forms; this structure contains a terminator, so attenuation occurs.

may not seem unusual, but tryptophan (Trp) is a rare amino acid in most proteins; it is found on average only once in every 100 amino acids. So the chance of finding two Trp codons in a row *anywhere* is quite small, and the fact that they are found in the *trp* operon is very suspicious.

In bacteria, transcription and translation occur simultaneously. Thus, as soon as the *trp* leader region is transcribed, ribosomes begin translating this emerging mRNA. Think about what would happen to a ribosome trying to translate the *trp* leader under conditions of tryptophan starvation (Figure 7.29a). Tryptophan is in short supply, and here are two demands in a row for that very amino acid. In all likelihood, the ribosome will not be able to satisfy those demands immediately, so it will pause at one of the Trp codons. And where does that put the stalled ribosome? Right on element 1, which should be participating in formation of the first hairpin. The bulky ribosome clinging to this RNA site effectively prevents its pairing with element 2, which frees 2 to pair with 3, forming the one-hairpin alternative structure. Because the second hairpin (elements 3 and 4) cannot form, transcription does not terminate and attenuation has been defeated. This is desirable, of course, because when tryptophan is scarce, the *trp* operon should be transcribed.

Notice that this mechanism involves a coupling of transcription and translation, where the latter affects the former. It would not work in eukaryotes, where transcription and translation take place in separate compartments. It also depends on transcription and translation occurring at about the same rate. If RNA polymerase outran the ribosome, it might pass through the attenuator region before the ribosome had a chance to stall at the Trp codons.

You may be wondering how the polycistronic mRNA made from the *trp* operon can be translated if ribosomes are stalled in the leader at the very beginning. The answer is that each of the genes represented on the mRNA has its own translation start signal (AUG). Ribosomes recognize each of these independently, so translation of the *trp* leader does not affect translation of the *trp* genes.

On the other hand, consider a ribosome translating the leader transcript under conditions of abundant tryptophan (Figure 7.29b). Now the dual Trp codons present no barrier to translation, so the ribosome continues through element 1 until it reaches the stop signal (UGA) between elements 1 and 2 and falls off. With no ribosome to interfere, the two hairpins can form, completing the transcription termination signal that halts transcription before it reaches the *trp* genes. Thus, the attenuation system responds to the presence of adequate tryptophan and prevents wasteful synthesis of enzymes to make still more tryptophan.

Other *E. coli* operons besides *trp* use the attenuation mechanism. The most dramatic known use of consecutive codons to stall a ribosome occurs in the *E. coli* histidine *(his)* operon, in which the leader region contains seven histidine codons in a row!

SUMMARY Attenuation operates in the *E. coli trp* operon as long as tryptophan is plentiful. When the supply of this amino acid is restricted, ribosomes stall at the tandem tryptophan codons in the *trp* leader. Because the *trp* leader is being synthesized just as stalling occurs, the stalled ribosome will influence the way this RNA folds. In particular, it prevents the formation of a hairpin, which is part of the transcription termination signal that causes attenuation. Therefore, when tryptophan is scarce, attenuation is defeated and the operon remains active. This means that the control exerted by attenuation responds to tryptophan levels, just as repression does.

7.4 Riboswitches

We have just seen an example of controlling gene expression by manipulating the structure of the 5′-untranslated region (UTR) of an mRNA (the *trp* mRNA of *E. coli*). In this case, a macromolecular assembly (the ribosome) senses the concentration of a small molecule (tryptophan) and binds to the *trp* 5′-UTR, altering its shape, thereby controlling its continued transcription. So this is an example of a group of macromolecules mediating the effect of a small molecule (or *ligand*) on gene expression.

We also have a growing number of examples of small molecules acting *directly* on mRNAs (usually on their 5′-UTRs) to control their expression. The regions of these mRNAs that are capable of altering their structures to control gene expression in response to ligand binding are called **riboswitches.** Riboswitches are responsible for 2–3% of gene expression control in bacteria, and they are also found in archaea, fungi, and plants. Later in this section we will learn of a possible example in animals.

The region of a riboswitch that binds to the ligand is called an **aptamer.** Aptamers were first discovered by scientists studying evolution in a test tube, who exploited rapidly replicating RNAs to select for short RNA sequences that bind tightly and specifically to ligands. As the RNAs replicate, they make mistakes, producing new RNA sequences, and those that bind best to a particular ligand are selected. Experimenters found many such aptamers in these in vitro experiments and wondered why living things did not take advantage of them. Now we know that they do.

A classic example of a riboswitch is the *ribD* operon in *B. subtilis*. This operon controls the synthesis and transport of the vitamin riboflavin and one of its products, flavin mononucleotide (FMN). Bacterial *rib* operons contain a conserved element in their 5′-UTRs known as the ***RFN* element.** Mutations in this region abolish normal control of the *ribD* operon by FMN, which led to the hypothesis that this *RFN* element interacts with a protein that responds to FMN or, perhaps, with FMN itself.

To test the hypothesis that the *RFN* element is an aptamer that binds directly to FMN, Ronald Breaker and colleagues used a technique called **in-line probing.** This method relies on the fact that efficient hydrolysis (breakage) of a phosphodiester bond in RNA needs a 180-degree ("in-line") arrangement among the attacking nucleophile (water), the phosphorus atom in the phosphodiester bond, and the leaving hydroxyl group at the end of one of the RNA fragments created by the hydrolysis. Unstructured RNA can easily assume this in-line conformation, but RNA that is constrained by secondary structure (intramolecular base pairing) or by binding to a ligand cannot. Thus, spontaneous cleavage of linear, unstructured RNA will occur much more readily than will cleavage of a structured RNA with lots of base pairing or with a ligand bound to it.

Thus, Breaker and colleagues incubated a labeled RNA fragment containing the *RFN* element in the presence and absence of FMN. Figure 7.30a shows that the patterns of spontaneous hydrolysis of the RNA were different in the presence and absence of FMN, suggesting that FMN binds directly to the RNA and causes it to shift its conformation. This is what we would expect of an aptamer bound to its ligand.

In particular, Breaker and colleagues found that FMN binding rendered certain phosphodiester bonds less susceptible to cleavage, whereas others retained their normal susceptibility (Figure 7.30b). Furthermore, the changes in susceptibility were half-maximal at an FMN concentration of only 5 nM. This indicates high affinity between the RNA and its ligand.

The patterns of decreased susceptibility to cleavage in the presence of FMN suggested the two alternative conformations of the *RFN* element depicted in Figure 7.30c. In the absence of FMN, the element should form an antiterminator, with the hairpin remote from the string of six U's. But FMN would cause the conformation of the element to shift such that it forms a terminator, blocking expression of the operon. This makes sense because, with abundant FMN, there is no need to express the *ribD* operon, so the proposed attenuation by FMN would save the cell energy.

To test this hypothesis, Breaker and colleagues performed an in vitro transcription assay with a cloned DNA template containing both the *RFN* element and the proposed terminator. They found that transcription terminated about 10% of the time at the terminator even in the absence of FMN, but FMN raised the frequency of termination to 30%. They mapped the termination site with a run-off transcription assay (Chapter 5) and showed that transcription terminated right at the end of the string of U's. Next, they used a mutant version of the DNA template that encoded fewer than six U's in the putative terminator. In this case, FMN caused no change

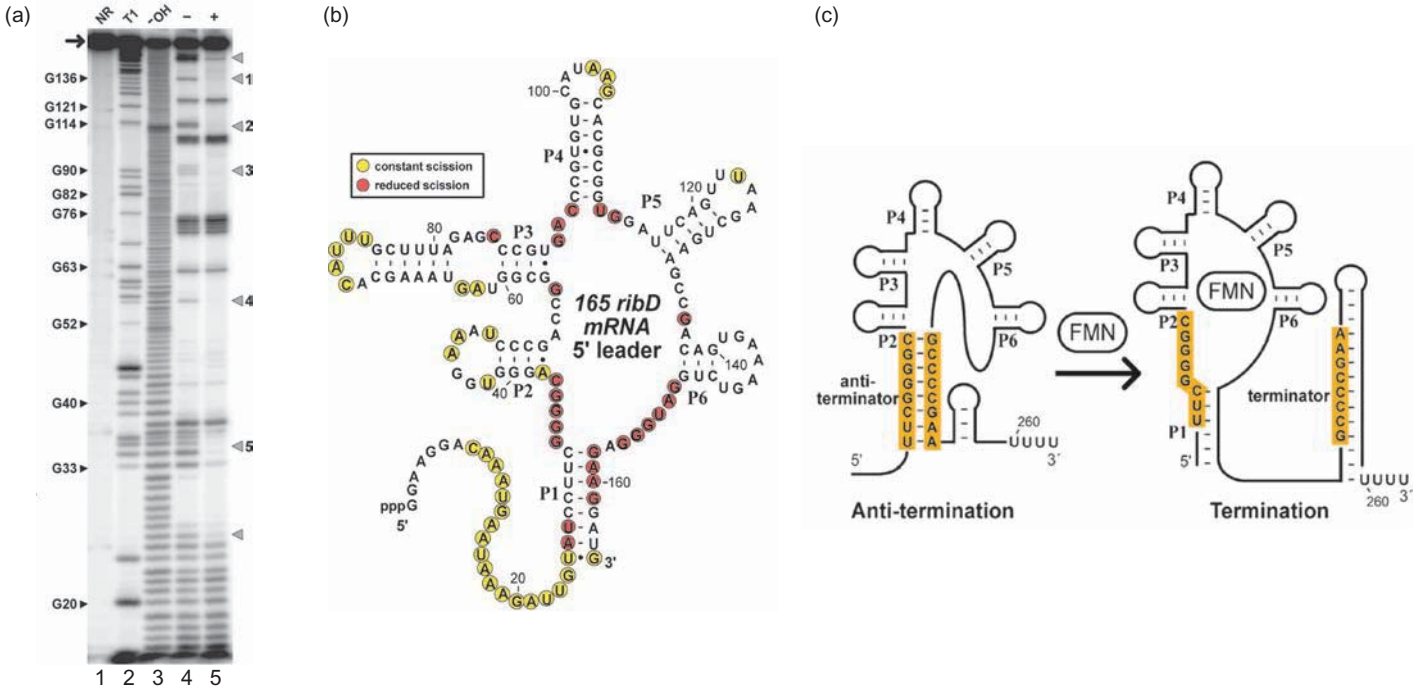


Figure 7.30 Results of in-line probing of *RFN* element and model for the action of the *ribD* riboswitch. (a) Gel electrophoresis results of in-line probing. Lane 1, no RNA; lane 2, RNA cut with RNase T1; lane 3, RNA cut with base; lanes 4 and 5, RNAs subjected to spontaneous cleavage in the absence (−) and presence (+) of FMN for 40 h at 25°C. Arrows at right denote regions of the RNA that became less susceptible to cleavage in the presence of FMN. **(b)** Sequence of part of the 5′-UTR of the *B. subtilis ribD* mRNA, showing the internucleotide linkages that became less susceptible to spontaneous cleavage upon FMN binding (red), and those that showed constant susceptibility (yellow). The secondary structure of the element is based on comparisons of sequences of many *RFN* elements. **(c)** Proposed change in structure of the riboswitch upon FMN binding. In the absence of FMN, base pairing between the two yellow regions forces the riboswitch to assume an antiterminator conformation, with the hairpin remote from the string of U's. Conversely, binding of FMN to the growing mRNA allows the GCCCCGAA sequence to base-pair with another part of the riboswitch, creating a terminator that stops transcription. (*Source: (a-c)* © 2002 National Academy of Science. Proceedings of the National Academy of Sciences, vol. 99, no. 25, December 10, 2002, pp. 15908–15913 "An mRNA structure that controls gene expression by binding FMN," Chalamish, and Ronald R. Breaker, fig.1, p. 15909 & fig. 3, p. 15911.)

in the frequency of termination, presumably because the shorter string of U's considerably lowered the efficiency of the terminator, even with FMN. Thus, with the wild-type gene, FMN really does appear to force more of the growing transcripts to form terminators that halt transcription.

Breaker and colleagues discovered another riboswitch in a conserved region in the 5′-untranslated region (5′-UTR) of the *glmS* gene of *Bacillus subtilis* and at least 17 other Gram-positive bacteria. This gene encodes an enzyme known as glutamine-fructose-6-phosphate amidotransferase, whose product is the sugar glucosamine-6-phosphate (GlcN6P). Breaker and colleagues found that the riboswitch in the 5′-UTR of the *glmS* mRNA is a ribozyme (an RNase) that can cleave the mRNA molecule itself. It does this at a low rate when concentrations of GlcN6P are low. However, when the concentration of GlcN6P rises, the sugar binds to the riboswitch in the mRNA and changes its conformation to make it a much better RNase (about 1000-fold better). This RNase destroys the mRNA, so less of the enzyme is made, so the GlcN6P concentration falls.

This riboswitch mechanism may not be confined to bacteria. In 2008, Harry Noller, William Scott, and colleagues discovered a very active **hammerhead ribozyme** in the 3′-UTRs of rodent C-type lectin type II (*Clec2*) mRNAs. Hammerhead ribozymes are so named because their secondary structure loosely resembles a hammer, with three base-paired stems constituting the "handle," "head," and "claw" of the hammer. At the junction of these three stems is a highly conserved group of 17 nucleotides that make up the RNase and the cleavage site, which lies at the bottom of the hammerhead where it joins the handle. Presumably, the hammerhead ribozyme in the *Clec2* mRNA responds to some cellular cue by cleaving itself and thus reducing *Clec2* gene expression, but it is not yet known what that cue is.

We will see another example of a riboswitch in Chapter 17, when we study the control of translation. We will learn that a ligand can bind to a riboswitch in an mRNA's 5′-UTR, and can control translation of that mRNA by changing the conformation of the 5′-UTR to hide the ribosome-binding site.

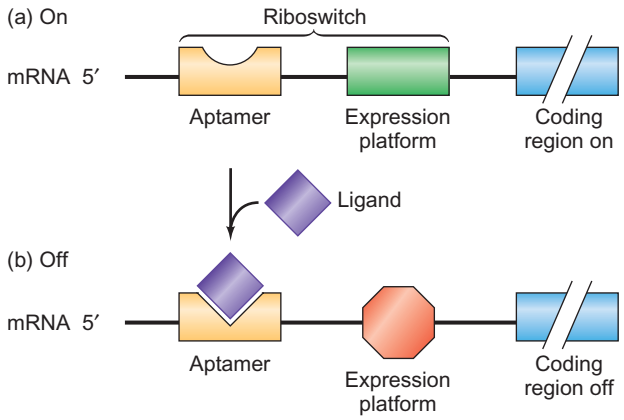


Figure 7.31 A model for riboswitch action. (a) Absence of the ligand. Gene expression is turned on. **(b)** Presence of the ligand. The ligand has bound to the aptamer in the riboswitch, causing a change in the conformation of the riboswitch, including the expression platform. This turns gene expression off.

These examples of riboswitches both operate by depressing gene expression: one at the transcriptional level, and one at the translational level. Indeed, all riboswitches studied to date work that way, although there is no reason why a riboswitch could not work by stimulating gene expression. These examples, among others, also lead to a general model for riboswitches (Figure 7.31). They are regions in the 5′-UTRs of mRNAs that contain two modules: an aptamer and another module, which Breaker and colleagues call an *expression platform*. The expression platform can be a terminator, a ribosome-binding site, or another RNA element that affects gene expression. By binding to its aptamer and changing the conformation of the riboswitch, a ligand can affect an expression platform, and thereby control gene expression.

Note that a riboswitch is another example of allosteric control, that is, one in which a ligand causes a conformational change in a large molecule that in turn affects the ability of the large molecule to interact with something else. We encountered an allosteric mechanism earlier in this chapter in the context of the *lac* operon, where a ligand (allolactose) bound to a protein (*lac* repressor) and interfered with its ability to bind to the *lac* operator. In fact, many examples of allosteric control are known, but up until recently they all involved allosteric proteins. Riboswitches work similarly, except that the large molecule is an RNA, rather than a protein.

Finally, riboswitches may provide a window on the "RNA world," a hypothetical era early in the evolution of life, in which proteins and DNA had not yet evolved. In this world, genes were made of RNA, not DNA, and enzymes were made of RNA, not protein. (We will see modern examples of catalytic RNAs in Chapters 14, 17, and 19.) Without proteins to control their genes, life forms in the RNA world would have had to rely on small molecules interacting directly with their genes. If this hypothesis is true, riboswitches are relics of one of the most ancient forms of genetic control.

SUMMARY A riboswitch is a region, usually in the 5′-UTR of an mRNA, that contains two modules: an aptamer that can bind a ligand, and an expression platform whose change in conformation can cause a change in expression of the gene. For example, FMN can bind to an aptamer in a riboswitch called the *RFN* element in the 5′-UTR of the *ribD* mRNA. Upon binding FMN, the base pairing in the riboswitch changes to create a terminator that attenuates transcription. This saves the cell energy because FMN is one of the products of the *ribD* operon. In another example, the *glmS* mRNA of *B. subtilis* contains a riboswitch that responds to the product of the enzyme encoded by the mRNA. When this product builds up, it binds to the riboswitch, changing the conformation of the RNA to stimulate an inherent RNase activity in the RNA so it cleaves itself.

SUMMARY

Lactose metabolism in *E. coli* is carried out by two proteins, β-galactosidase and galactoside permease. The genes for these two, and one additional enzyme, are clustered together and transcribed together from one promoter, yielding a polycistronic message. These functionally related genes are therefore controlled together.

Control of the *lac* operon occurs by both positive and negative control mechanisms. Negative control appears to occur as follows: The operon is turned off as long as repressor binds to the operator, because the repressor prevents RNA polymerase from binding to the promoter to transcribe the three *lac* genes. When the supply of glucose is exhausted and lactose is available, the few molecules of *lac* operon enzymes produce a few molecules of allolactose from the lactose. The allolactose acts as an inducer by binding to the repressor and causing a conformational shift that encourages dissociation from the operator. With the repressor removed, RNA polymerase is free to transcribe the three *lac* genes. A combination of genetic and biochemical experiments revealed the two key elements of negative control of the *lac* operon: the operator and the repressor. DNA sequencing revealed the presence of two auxiliary *lac* operators: one upstream, and one downstream of the major operator. All three are required for optimal repression.

Positive control of the *lac* operon, and certain other inducible operons that code for sugar-metabolizing enzymes, is mediated by a factor called catabolite activator protein (CAP), which, in conjunction with cyclic-AMP (cAMP), stimulates transcription. Because cAMP concentration is depressed by glucose, this sugar prevents positive control from operating. Thus, the *lac* operon is activated only when glucose concentration is low and a corresponding need arises to metabolize an alternative energy source. The CAP–cAMP complex stimulates expression of the *lac* operon by binding to an activator site adjacent to the promoter. CAP–cAMP binding helps RNA polymerase form an open promoter complex. It does this by recruiting polymerase to form a closed promoter complex, which then converts to an open promoter complex. Recruitment of polymerase occurs through protein–protein interactions between CAP and the αCTD of RNA polymerase.

The *ara* operon is controlled by the AraC protein. AraC represses the operon by looping out the DNA between two sites, *araO*$_2$ and *araI*$_1$, that are 210 bp apart. Arabinose can induce the operon by causing AraC to loosen its attachment to *araO*$_2$ and to bind to *araI*$_1$ and *araI*$_2$ instead. This breaks the loop and allows transcription of the operon. CAP and cAMP further stimulate transcription by binding to a site upstream of *araI*. AraC controls its own synthesis by binding to *araO*$_1$ and preventing leftward transcription of the *araC* gene.

The *trp* operon responds to a repressor that includes a corepressor, tryptophan, which signals the cell that it has made enough of this amino acid. The corepressor binds to the aporepressor, changing its conformation so it can bind better to the *trp* operator, thereby repressing the operon.

Attenuation operates in the *E. coli trp* operon as long as tryptophan is plentiful. When the supply of this amino acid is restricted, ribosomes stall at the tandem tryptophan codons in the *trp* leader. Because the *trp* leader is being synthesized just as this is taking place, the stalled ribosome will influence the way this RNA folds. In particular, it prevents the formation of a hairpin, which is part of the transcription termination signal that causes attenuation. When tryptophan is scarce, attenuation is therefore defeated and the operon remains active. This means that the control exerted by attenuation responds to tryptophan levels, just as repression does.

A riboswitch is a region in the 5′-UTR of an mRNA that contains two modules: an aptamer that can bind a ligand, and an expression platform whose change in conformation can cause a change in expression of the gene. For example, FMN can bind to an aptamer in a riboswitch called the *RFN* element in the 5′-UTR of the *ribD* mRNA. Upon binding FMN, the base pairing in the riboswitch changes to create a terminator that attenuates transcription.

REVIEW QUESTIONS

1. Draw a growth curve of *E. coli* cells growing on a mixture of glucose and lactose. What is happening in each part of the curve?
2. Draw diagrams of the *lac* operon that illustrate (a) negative control and (b) positive control.
3. What are the functions of β-galactosidase and galactoside permease?
4. Why are negative and positive control of the *lac* operon important to the energy efficiency of *E. coli* cells?
5. Describe and give the results of an experiment that shows that the *lac* operator is the site of repressor binding.
6. Describe and give the results of an experiment that shows that RNA polymerase can bind to the *lac* promoter, even if repressor is already bound at the operator.
7. Describe and give the results of an experiment that shows that *lac* repressor prevents RNA polymerase from binding to the *lac* promoter.
8. How do we know that all three *lac* operators are required for full repression? What are the relative effects of removing each or both of the auxiliary operators?
9. Describe and give the results of an experiment that shows the relative levels of stimulation of β-galactosidase synthesis by cAMP, using wild-type and mutant extracts, in which the mutation reduces the affinity of CAP for cAMP.
10. Present a hypothesis for activation of *lac* transcription by CAP–cAMP. Include the C-terminal domain of the polymerase α-subunit (the αCTD) in the hypothesis. What evidence supports this hypothesis?
11. Describe and give the results of an electrophoresis experiment that shows that binding of CAP–cAMP bends the *lac* promoter region.
12. What other data support DNA bending in response to CAP–cAMP binding?
13. Explain the fact that insertion of an integral number of DNA helical turns (multiples of 10.5 bp) between the *araO*$_2$ and *araI* sites in the *araBAD* operon permits repression by AraC, but insertion of a nonintegral number of helical turns prevents repression. Illustrate this phenomenon with diagrams.
14. Use a diagram to illustrate how arabinose can relieve repression of the *araBAD* operon. Show where AraC is located (a) in the absence of arabinose, and (b) in the presence of arabinose.
15. Describe and give the results of an experiment that shows that arabinose can break the repression loop formed by AraC.
16. Describe and give the results of an experiment that shows that both *araO*$_2$ and *araI* are involved in forming the repression loop.
17. Briefly outline evidence that shows that *araI*$_2$ is important in binding AraC when the DNA is in the unlooped, but not the looped, form.

18. Present a model to explain negative control of the *trp* operon in *E. coli*.
19. Present a model to explain attenuation in the *trp* operon in *E. coli*.
20. Why does translation of the *trp* leader region not simply continue into the *trp* structural genes (*trpE*, etc.) in *E. coli*?
21. How is *trp* attenuation overridden in *E. coli* when tryptophan is scarce?
22. What is a riboswitch? Illustrate with an example.
23. Describe what is meant by "in-line probing."

ANALYTICAL QUESTIONS

1. The table below gives the genotypes (with respect to the *lac* operon) of several partial diploid *E. coli* strains. Fill in the phenotypes, using a "+" for β-galactosidase synthesis and "−" for no β-galactosidase synthesis. Glucose is absent in all cases. Give a brief explanation of your reasoning.

	Phenotype for β-galactosidase Production	
Genotype	*No Inducer*	*Inducer*
a. $I^+O^+Z^+/I^+O^+Z^+$		
b. $I^+O^+Z^-/I^+O^+Z^+$		
c. $I^-O^+Z^+/I^+O^+Z^+$		
d. $I^sO^+Z^+/I^+O^+Z^+$		
e. $I^+O^cZ^+/I^+O^+Z^+$		
f. $I^+O^cZ^-/I^+O^+Z^+$		
g. $I^sO^cZ^+/I^+O^+Z^+$		

2. (*a*) In the genotype listed in the following table, the letters *A*, *B*, and *C* correspond to the *lacI*, and *lacO*, *lacZ* loci, though not necessarily in that order. From the mutant phenotypes exhibited by the first three genotypes listed in the table, deduce the identities of *A*, *B*, and *C* as they correspond to the three loci of the *lac* operon. The minus superscripts (e.g., A^-) can refer to the following aberrant functions: Z^-, O^c, or I^-.
   (*b*) Determine the genotypes, in conventional *lac* operon genetic notation, of the partial diploid strains shown in lines 4 and 5 of the table. Here, I^+, I^-, and I^s are all possible.

	Phenotype for β-galactosidase Production	
Genotype	*No Inducer*	*Inducer*
1. $A^+B^+C^-$	+	+
2. $A^-B^+C^+$	+	+
3. $A^+B^+C^-/A^+B^+C^+$	+	+
4. $A^-B^+C^+/A^+B^+C^+$	−	−
5. $A^-B^+C^+/A^+B^+C^+$	−	+

3. Consider *E. coli* cells, each having one of the following mutations:
   a. a mutant *lac* operator (the O^c locus) that cannot bind repressor
   b. a mutant *lac* repressor (the I^- gene product) that cannot bind to the *lac* operator
   c. a mutant *lac* repressor (the I^s gene product) that cannot bind to allolactose
   d. a mutant *lac* promoter region that cannot bind CAP plus cAMP

   What effect would each mutation have on the function of the *lac* operon (assuming no glucose is present)?
4. You are studying a new operon in *E. coli* involved in phenylalanine biosynthesis.
   a. How would you predict this operon is regulated (inducible or repressible by phenylalanine, positive or negative)? Why?
   b. You sequence the operon and discover that it contains a short open reading frame near the 5′-end of the operon that contains several codons for phenylalanine. What prediction would you make about this leader sequence and the peptide that it encodes?
   c. What would happen if the sequence of this leader were changed so that the phenylalanine codons (UUU, UUU) were changed to leucine codons (UUA, UUG)?
   d. What is this kind of regulation called and would it work in a eukaryotic cell? Why or why not?
5. You suspect that the mRNA from gene *X* of *E. coli* contains an aptamer that binds to a small molecule, Y. Describe an experiment to test this hypothesis.
6. The *aim* operon includes sequences A, B, C, and D. Mutations in these sequences have the following effects, where a plus sign (+) indicates that a functional enzyme is produced and a minus sign (−) indicates that a functional enzyme is not produced. X is a metabolite.

	X present		X absent	
Mutation in sequence:	*Enzyme 1*	*Enzyme 2*	*Enzyme 1*	*Enzyme 2*
A	−	−	−	−
B	+	+	+	+
C	+	−	−	−
D	−	+	−	−
Wild-Type	+	+	−	−

   a. Do the structural gene products from the *aim* operon participate in an anabolic or catabolic process?
   b. Is the repressor protein associated with the *aim* operon produced in an initially active or inactive form?
   c. What does sequence D encode?
   d. What does sequence B encode?
   e. What is sequence A?

SUGGESTED READINGS

General References and Reviews

Beckwith, J.R. and D. Zipser, eds. 1970. *The Lactose Operon.* Plainview, NY: Cold Spring Harbor Laboratory Press.

Corwin, H.O. and J.B. Jenkins. *Conceptual Foundations of Genetics: Selected Readings.* 1976. Boston: Houghton Mifflin Co.

Jacob, F. 1966. Genetics of the bacterial cell (Nobel lecture). *Science* 152:1470–78.

Matthews, K.S. 1996. The whole lactose repressor. *Science* 271:1245–46.

Miller, J.H. and W.S. Reznikoff, eds. 1978. *The Operon.* Plainview, NY: Cold Spring Harbor Laboratory Press.

Monod, J. 1966. From enzymatic adaptation to allosteric transitions (Nobel lecture). *Science* 154:475–83.

Ptashne, M. 1989. How gene activators work. *Scientific American* 260 (January):24–31.

Ptashne, M. and W. Gilbert. 1970. Genetic repressors. *Scientific American* 222 (June):36–44.

Vitreschak, A.G., D.A. Rodionov, A.A. Mironov, and M.S. Gelfand. 2004. Riboswitches: The oldest mechanism for the regulation of gene expression? *Trends in Genetics* 20:44–50.

Winkler, W.C. and R.R. Breaker. 2003. Genetic control by metabolite-binding riboswitches. *Chembiochem* 4:1024–32.

Research Articles

Adhya, S. and S. Garges. 1990. Positive control. *Journal of Biological Chemistry* 265:10797–800.

Benoff, B., H. Yang, C.L. Lawson, G. Parkinson, J. Liu, E. Blatter, Y.W. Ebright, H.M. Berman, and R.H. Ebright. 2002. Structural basis of transcription activation: The CAP–αCTD–DNA complex. *Science* 297:1562–66.

Busby, S. and R.H. Ebright. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* 79:743–46.

Chen, B., B. deCrombrugge, W.B. Anderson, M.E. Gottesman, I. Pastan, and R.L. Perlman. 1971. On the mechanism of action of *lac* repressor. *Nature New Biology* 233:67–70.

Chen, Y., Y.W. Ebright, and R.H. Ebright. 1994. Identification of the target of a transcription activator protein by protein–protein photocrosslinking. *Science* 265:90–92.

Emmer, M., B. deCrombrugge, I. Pastan, and R. Perlman. 1970. Cyclic-AMP receptor protein of *E. coli:* Its role in the synthesis of inducible enzymes. *Proceedings of the National Academy of Sciences USA* 66:480–87.

Gilbert, W. and B. Müller-Hill. 1966. Isolation of the *lac* repressor. *Proceedings of the National Academy of Sciences USA* 56:1891–98.

Igarashi, K. and A. Ishihama. 1991. Bipartite functional map of the *E. coli* RNA polymerase α subunit: Involvement of the C-terminal region in transcription activation by cAMP–CRP. *Cell* 65:1015–22.

Jacob, F. and J. Monod. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3:318–56.

Krummel, B. and M.J. Chamberlin. 1989. RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in the early ternary complexes. *Biochemistry* 28:7829–42.

Lee, J. and A. Goldfarb. 1991. *Lac* repressor acts by modifying the initial transcribing complex so that it cannot leave the promoter. *Cell* 66:793–98.

Lewis, M., G. Chang, N.C. Horton, M.A. Kercher, H.C. Pace, M.A. Schumacher, R.G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science* 271:1247–54.

Lobell, R.B. and R.F. Schleif. 1991. DNA looping and unlooping by AraC protein. *Science* 250:528–32.

Malan, T.P. and W.R. McClure. 1984. Dual promoter control of the *Escherichia coli* lactose operon. *Cell* 39:173–80.

Oehler, S., E.R. Eismann, H. Krämer, and B. Müller-Hill. 1990. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal* 9:973–79.

Riggs, A.D., S. Bourgeois, R.F. Newby, and M. Cohn. 1968. DNA binding of the *lac* repressor. *Journal of Molecular Biology* 34:365–68.

Schlax, P.J., M.W. Capp, and M.T. Record, Jr. 1995. Inhibition of transcription initiation by *lac* repressor. *Journal of Molecular Biology* 245:331–50.

Schultz, S.C., G.C. Shields, and T.A. Steitz. 1991. Crystal structure of a CAP–DNA complex: The DNA is bent by 90 degrees. *Science* 253:1001–7.

Straney, S. and D.M. Crothers. 1987. *Lac* repressor is a transient gene-activating protein. *Cell* 51:699–707.

Winkler, W.C., S. Cohen-Chalamish, and R.R. Breaker. 2002. An mRNA structure that controls gene expression by binding FMN. *Proceedings of the National Academy of Sciences, USA* 99:15908–13.

Wu, H.-M. and D.M. Crothers. 1984. The locus of sequence-directed and protein-induced DNA bending. *Nature* 308:509–13.

Yanofsky, C. 1981. Attenuation in the control of expression of bacterial operons. *Nature* 289:751–58.

Zubay, G., D. Schwartz, and J. Beckwith. 1970. Mechanism of activation of catabolite-sensitive genes: A positive control system. *Proceedings of the National Academy of Sciences USA* 66:104–10.