CHAPTER

16

Gene Regulation in Prokaryotes

In Chapter 12 we saw how DNA is transcribed into RNA by the enzyme RNA polymerase. We also described the sequence elements that constitute a promoter—the region at the start of a gene where the enzyme binds and initiates transcription. In bacteria the most common form of RNA polymerase (that bearing σ^{70}) recognizes promoters formed from three elements—the "−10", "−35", and "UP" elements—and we saw that the strength of any given promoter is determined by which of these elements it possesses and how well they match optimum "consensus" sequences. In the absence of regulatory proteins, these elements determine the efficiency with which polymerase binds to the promoter and, once bound, how readily it initiates transcription.

Now we turn to mechanisms that regulate expression—that is, mechanisms that increase or decrease expression of a given gene as the requirement for its product varies. There are various stages at which expression of a gene can be regulated. The most common is transcription initiation, and the bulk of this chapter focuses on the regulation of that step in bacteria. We start with an overview of general mechanisms and principles and proceed to some well-studied examples that demonstrate how the basic mechanisms are used in various combinations to control genes in specific biological contexts. We also consider mechanisms of gene regulation that operate at steps after transcription initiation, including transcriptional antitermination and the regulation of translation. ■

OUTLINE

- Principles of Transcriptional Regulation (p. 1)
- Regulation of Transcription Initiation: Examples from Bacteria (p. 5)
- Two Examples of Gene Regulation at Steps after Transcription Initiation (p. 21)
- The Case of Phage λ: Layers of Regulation Underlie a Complex Program of Gene Expression (p. 27)

PRINCIPLES OF TRANSCRIPTIONAL REGULATION

Gene Expression Is Controlled by Regulatory Proteins

As we described in the introduction to this section, genes are very often controlled by extracellular signals—in the case of bacteria, this typically means molecules present in the growth medium. These signals are communicated to genes by regulatory proteins, which come in two types: positive regulators, or **activators;** and negative regulators, or **repressors.** Typically these regulators are DNA binding proteins that recognize specific sites at or near the genes they control. An activator increases transcription of the regulated gene; repressors decrease or eliminate that transcription.

How do these regulators work? Recall the steps in transcription initiation described in Chapter 12. First, RNA polymerase binds to the promoter in a closed complex (in which the DNA strands remain together). The polymerase-promoter complex then undergoes a transition to an open complex in which the DNA at the start site of transcription is

unwound and the polymerase is positioned to initiate transcription. This is followed by promoter escape, or clearance, the step in which polymerase leaves the promoter and starts transcribing. Which steps are stimulated by activators and inhibited by repressors? That depends on the promoter and regulators in question. We consider two general cases.

Many Promoters Are Regulated by Activators That Help RNA Polymerase Bind DNA and by Repressors That Block That Binding

At many promoters, in the absence of regulatory proteins, RNA polymerase binds only weakly. This is because one or more of the promoter elements discussed above is imperfect. When polymerase does occasionally bind, however, it spontaneously undergoes a transition to the open complex and initiates transcription. This gives a low level of **constitutive** expression called the **basal** level. Binding of RNA polymerase is the rate limiting step in this case (Figure 16-1a).

To control expression from such a promoter, a repressor need only bind to a site overlapping the region bound by polymerase. In that way repressor blocks polymerase binding to the promoter, thereby preventing transcription (Figure 16-1b), although it is important to note that repression can work in other ways as well. The site on DNA where a repressor binds is called an **operator.** To activate transcription, an activator just helps polymerase bind the promoter. Typically this is achieved as follows: The activator uses one surface to bind to a site on the DNA near the promoter; with another surface, the activator simultaneously interacts with polymerase, bringing the enzyme to the promoter (Figure 16-1c). This mechanism, often called **recruitment,** is an example of **cooperative binding** of proteins to DNA. The interac-

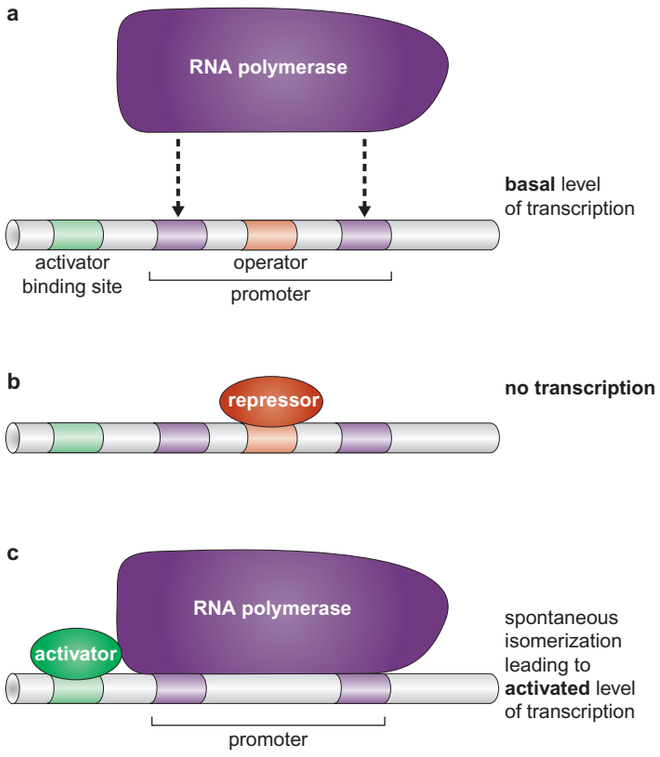


FIGURE 16-1 Activation by Recruitment of RNA Polymerase. (a) In the absence of both activator and repressor, RNA polymerase occasionally binds the promoter spontaneously and initiates a low level (basal level) of transcription. (b) Binding of the repressor to the operator sequence blocks binding of RNA polymerase and so inhibits transcription. (c) Recruitment of RNA polymerase by the activator gives high levels of transcription. RNA polymerase is shown recruited in the closed complex. It then spontaneously isomerizes to the open complex and initiates transcription.

tions between the activator and polymerase, and between activator and DNA, serve merely "adhesive" roles: the enzyme is active and the activator simply brings it to the nearby promoter. Once there, it spontaneously isomerizes to the open complex and initiates transcription.

The *lac* genes of *E. coli* are transcribed from a promoter that is regulated by an activator and a repressor working in the simple ways outlined. We will describe this case in detail later in the chapter.

Some Activators Work by Allostery and Regulate Steps after RNA Polymerase Binding

Not all promoters are limited in the same way. Thus, consider a promoter at the other extreme from that described above. In this case, RNA polymerase binds efficiently unaided and forms a stable closed complex. But that closed complex does not spontaneously undergo transition to the open complex (Figure 16-2a). At this promoter, an activator must stimulate the transition from closed to open complex, since that transition is the rate-limiting step.

Activators that stimulate this kind of promoter work by triggering a conformational change in either RNA polymerase or DNA. That is, they interact with the stable closed complex and induce a conformational change that causes transition to the open complex (Figure 16-2b). This mechanism is an example of **allostery.** In Chapter 5 we encountered allostery as a general mechanism for controlling the activities of proteins. One of the examples we considered there was a protein (a cyclin) binding to, and activating, a kinase (cdk) involved in cell cycle regulation. The cyclin does this by inducing a conformational change in the kinase, switching it from an inactive to an active state. In this chapter we will see two examples of transcriptional activators working by allostery. In one case (at the *glnA* promoter), the activator (NtrC) interacts with the RNA polymerase bound in a closed complex at the promoter, stimulating transition to open complex. In the other example (at the *merT* promoter), the activator (MerR) induces a conformational change in the promoter DNA.

There are variations on these themes: some promoters are inefficient at more than one step and can be activated by more than one mechanism. Also, repressors can work in ways other than just blocking the binding of RNA polymerase. For example, some repressors inhibit transition to the open complex, or promoter escape. We will consider examples of these later in the chapter.

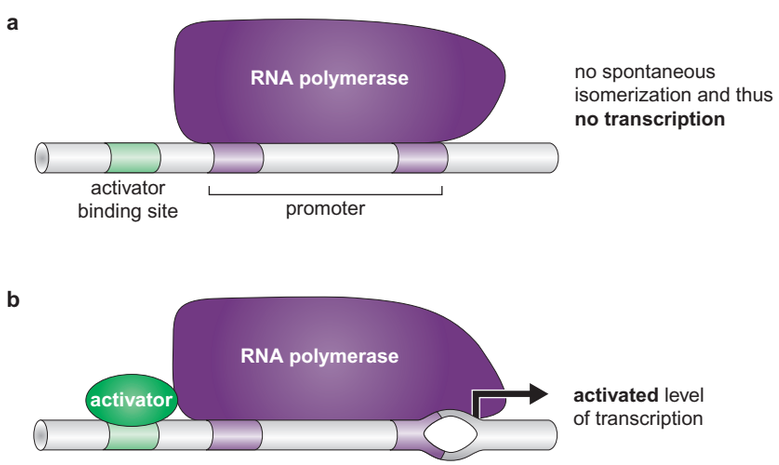


FIGURE 16-2 Allosteric Activation of RNA Polymerase. (a) Binding of RNA polymerase to the promoter in a stable closed complex. (b) Activator interacts with polymerase to trigger transition to the open complex and high levels of transcription. The representations of the closed and open complexes are shown only diagrammatically; for more accurate illustrations of those states see Chapter 12.

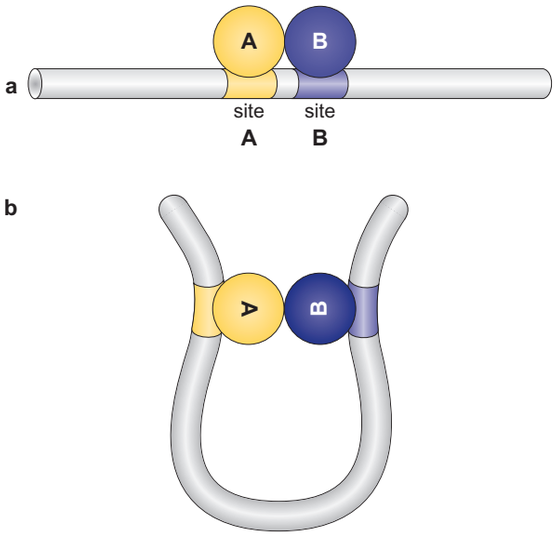


FIGURE 16-3 Interactions between Proteins Bound to DNA. (a) Cooperative binding of proteins to adjacent sites. (b) Cooperative binding of proteins to separated sites.

Action at a Distance and DNA Looping

Thus far we have tacitly assumed that DNA-binding proteins that interact with each other bind to adjacent sites (e.g., RNA polymerase and activator in Figures 16-1 and 16-2). Often this is the case. But some proteins interact even when bound to sites well separated on the DNA. To accommodate this interaction, the DNA between the sites loops out, bringing the sites into proximity with one another (Figure 16-3).

We will encounter examples of this in bacteria. Indeed, one of the activators we have already mentioned (NtrC) activates "from a distance": its binding sites are normally located about 150 bps upstream of the promoter, and the activator works even when those sites are placed further away (a kb or more). We will also consider repressors that interact to form loops of up to 3 kb. In the next chapter—on eukaryotic gene regulation—we will be faced with more numerous and more dramatic examples.

One way to help bring distant DNA sites closer together (and so help looping) is by the binding of other proteins to sequences between those sites. In bacteria there are cases in which a protein binds between an activator binding site and the promoter and helps the activator interact with polymerase by bending the DNA (Figure 16-4). Such "architectural" proteins facilitate interactions between proteins in other processes as well (e.g., site-specific recombination; see Chapter 11).

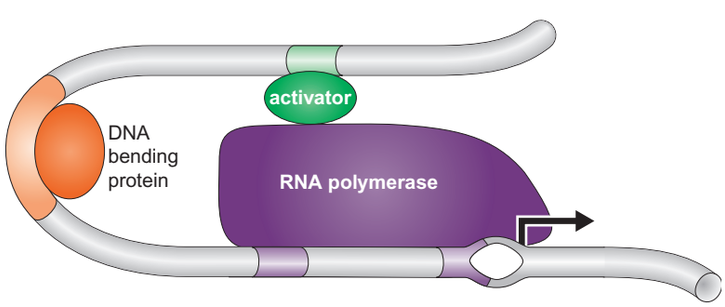


FIGURE 16-4 DNA-Bending Protein Can Facilitate Interaction between DNA-Binding Proteins. A protein that bends DNA binds to a site between the activator binding site and the promoter. This brings the two sites closer together in space and thereby helps the interaction between the DNA-bound activator and polymerase.

Cooperative Binding and Allostery Have Many Roles in Gene Regulation

We have already pointed out that gene activation can be mediated by simple cooperative binding: the activator interacts simultaneously with DNA and with polymerase and so recruits the enzyme to the promoter. And we have described how activation can, in other cases, be mediated by allosteric events: an activator interacts with polymerase already bound to the promoter and, by inducing a conformational change in the enzyme or the promoter, stimulates transcription initiation. Both cooperative binding and allostery have additional roles in gene regulation as well.

For example, groups of regulators often bind DNA cooperatively. That is, two or more activators and/or repressors interact with each other and with DNA, and thereby help each other bind near a gene they all regulate. As we will see, this kind of interaction can produce sensitive switches that allow a gene to go from completely off to fully on in response to only small changes in conditions. Cooperative binding of activators can also serve to integrate signals; that is, some genes are activated only when multiple signals (and thus multiple regulators) are simultaneously present. A particularly striking and well-understood example of cooperativity in gene regulation is provided by bacteriophage λ. We consider the basic mechanism and consequences of cooperative binding in more detail when we discuss that example later in the chapter, and also in Box 16-4.

Allostery, for its part, is not only a mechanism of gene activation, it is also often the way regulators are controlled by their specific signals. Thus, a typical bacterial regulator can adopt two conformations—in one it can bind DNA; in the other it cannot. Binding of a signal molecule locks the regulatory protein in one or another conformation, thereby determining whether or not it can act. We saw an example of this in Chapter 5, where we also considered the basic mechanism of allostery in some detail; in this and the next chapter we will see several examples of allosteric control of regulators by their signals.

Antitermination and Beyond: Not All of Gene Regulation Targets Transcription Initiation

As stated in the introduction to this section, the bulk of gene regulation takes place at the initiation of transcription. This is true in eukaryotes just as it is in bacteria. But regulation is certainly not restricted to that step in either class of organism. In this chapter we will see examples of gene regulation that involve transcriptional elongation, RNA processing, and translation of the mRNA into protein.

REGULATION OF TRANSCRIPTION INITIATION: EXAMPLES FROM BACTERIA

Having outlined basic principles of transcriptional regulation, we turn to some examples that show how these principles work in real cases. First, we consider the genes involved in lactose metabolism in *E. coli*—those of the *lac* operon. Here we will see how an activator and a repressor regulate expression in response to two signals. We also

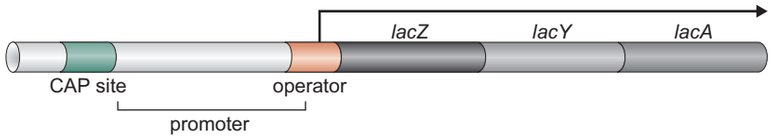


FIGURE 16-5 The *lac* Operon. The three genes (*lacZ, Y* and *A*) are transcribed as a single mRNA from the promoter (as indicated by the arrow). The CAP site and the operator are each about 20 bp. The operator lies within the region bound by RNA polymerase at the promoter, and the CAP site lies just upstream of the promoter (see Figure 16-8 for more details of the relative arrangements of these binding sites). The picture is simplified in that there are two additional, weaker, *lac* operators located nearby.

describe some of the experimental approaches that reveal how these regulators work.

An Activator and a Repressor Together Control the *lac* Genes

The three *lac* genes—*lacZ, lacY,* and *lacA*—are arranged adjacently on the *E. coli* genome and are called the *lac* operon (Figure 16-5). The *lac* promoter, located at the 5′ end of *lacZ,* directs transcription of all three genes as a single mRNA (called a **polycistronic** message) which is translated to give the three protein products. *LacZ* encodes the enzyme β-galactosidase, which cleaves the sugar lactose into galactose and glucose, both of which are used by the cell as energy sources. The *lacY*

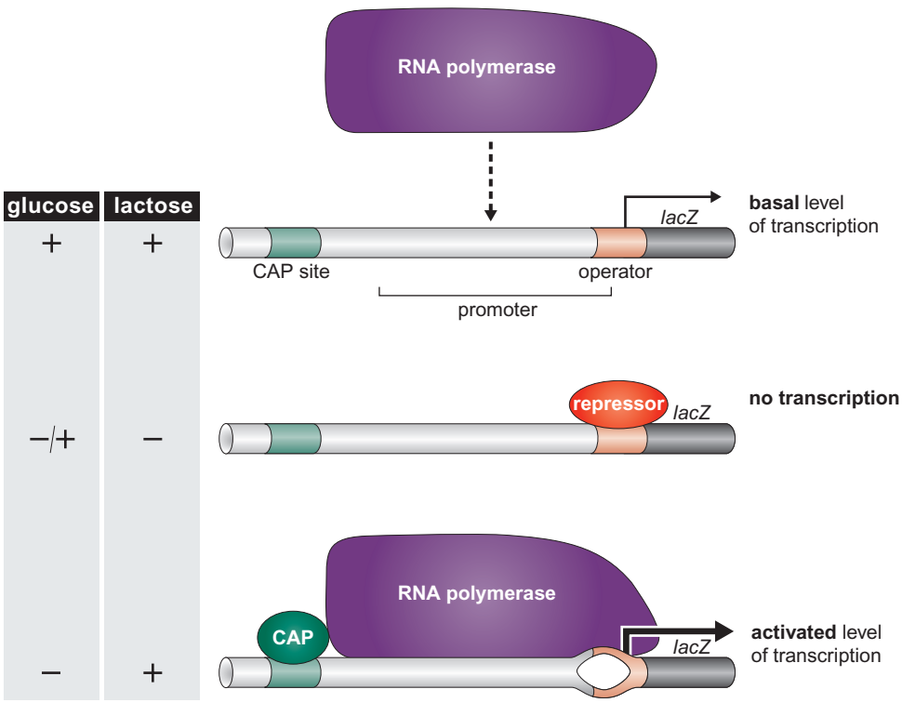


FIGURE 16-6 The Three States of the *lac* Genes. When bound to the operator, repressor excludes polymerase whether or not active CAP is present. CAP and Lac repressor are shown as single units, but CAP actually binds DNA as a dimer, and Lac repressor binds as a tetramer (see Figure 16-13). CAP recruits polymerase to the *lac* promoter where it spontaneously undergoes isomerization to the open complex (the state shown in the bottom line).

gene encodes the lactose permease, a protein that inserts into the membrane and transports lactose into the cell. The *lacA* gene encodes thiogalactoside transacetylase, which rids the cell of toxic thiogalactosides that also get transported in by *lacY.*

These genes are expressed at high levels only when lactose is available, and glucose—the preferred energy source—is not. Two regulatory proteins are involved: one is an activator called CAP, the other a repressor called the Lac repressor. Lac repressor is encoded by the *lacI* gene, which is located near the other *lac* genes, but transcribed from its own (constitutively expressed) promoter. The name CAP stands for Catabolite Activator Protein, but this activator is also known as CRP (for cAMP Receptor Protein, for reasons we will explain later). The gene encoding CAP is located elsewhere on the bacterial chromosome, not linked to the *lac* genes. Both CAP and Lac repressor are DNA-binding proteins and each binds to a specific site on DNA at or near the *lac* promoter (see Figure 16-5).

Each of these regulatory proteins responds to one environmental signal and communicates it to the *lac* genes. Thus, CAP mediates the effect of glucose, whereas Lac repressor mediates the lactose signal. This regulatory system works in the following way. Lac repressor can bind DNA and repress transcription only in the absence of lactose. In the presence of that sugar, the repressor is inactive and the genes de-repressed (expressed). CAP can bind DNA and activate the *lac* genes only in the absence of glucose. Thus, the combined effect of these two regulators ensures that the genes are expressed at significant levels only when lactose is present and glucose absent (see Figure 16-6).

CAP and Lac Repressor Have Opposing Effects on RNA Polymerase Binding to the *lac* Promoter

The site bound by Lac repressor is called the ***lac* operator.** This 21-bp sequence is twofold symmetric and is recognized by two subunits of Lac repressor, one binding to each half-site (see Figure 16-7). We will look at that binding in more detail later in this chapter, in the section "CAP and Lac repressor bind DNA using a common structural motif." How does repressor, when bound to the operator, repress transcription?

The *lac* operator overlaps the promoter, and so repressor bound to the operator physically prevents RNA polymerase from binding to the promoter and thus initiating RNA synthesis (see Figure 16-8). Protein binding sites in DNA can be identified, and their location mapped, using DNA footprinting and gel mobility assays described in Box 16-1.

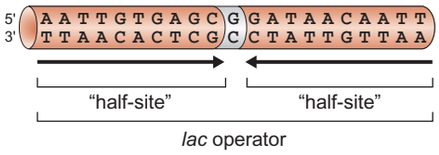


FIGURE 16-7 The Symmetric Half-Sites of the *lac* Operator.

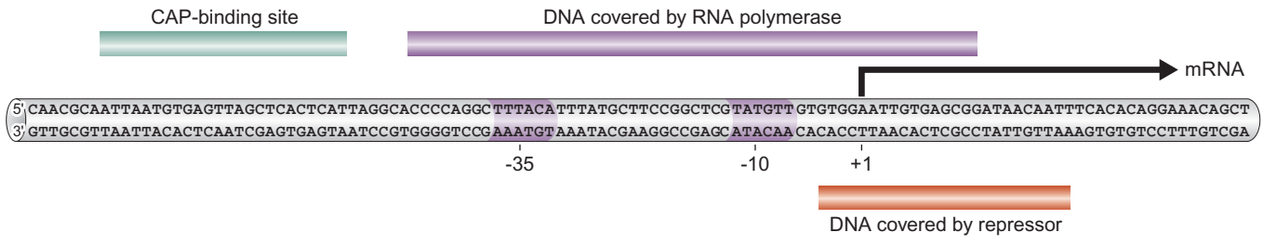


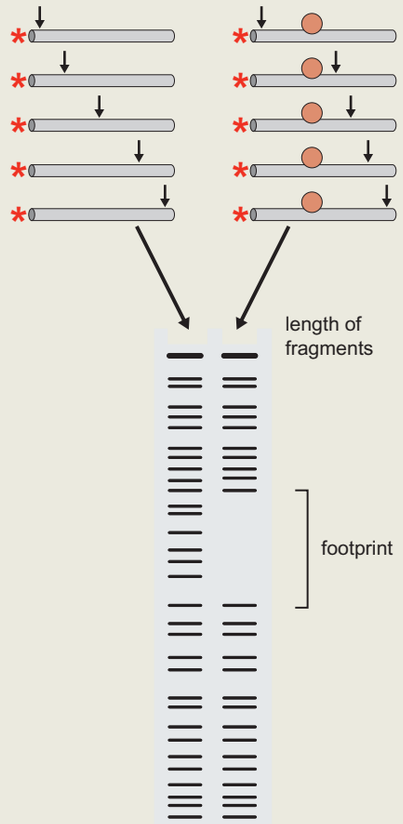
FIGURE 16-8 The Control Region of the *lac* Operon. The nucleotide sequence and organization of the *lac* operon control region. This shows that Lac repressor covers more DNA than that sequence defined as the minimal operator binding site.

Box 16-1 Detecting DNA Binding Sites

DNA Footprinting.

How can a protein binding site in DNA, such as an operator, be identified? A series of powerful approaches allows identification of the sites where proteins act and which chemical groups in DNA (methyl, amino, or phosphate) a protein contacts. The basic principle that underlies these methods, similar to the chemical method of DNA sequencing (see Chapter 20 on Techniques of Molecular Biology), is as follows. If a DNA fragment is labeled with a radioactive atom only at one end of one strand, the location of any break in this strand can be deduced from the size of the labeled fragment that results. The size, in turn, can be determined by high-resolution electrophoresis in a polyacrylamide gel. In the **nuclease protection footprinting method** the binding site is marked by internucleotide bonds that are shielded from the cutting action of a nuclease by the binding protein (Box 16-1 Figure 1). The resulting "footprint" is revealed by the absence of bonds of particular sizes. The related **chemical protection footprinting method** relies on the ability of a bound protein to modify the reactivity of bases in the binding site to those base-specific reagents that (after a further reaction) give rise to backbone cuts in the standard chemical DNA-sequencing method.

By changing the order of the first two steps, a third method **(chemical interference footprinting)** determines which features of the DNA structure are *necessary* for the protein to bind. An average of one chemical change per DNA is made, and then protein-DNA complexes are isolated. If a modification at a particular site does not prevent binding, DNA isolated from the complex will contain the modified chemical group, and the harmless modification allows the DNA to be broken at this site by further chemical treatment. If, on the other hand, a modification blocks DNA binding, then no DNA modified at the site will be found complexed to the binding protein and the isolated fragments will not be broken at this site by subsequent chemical treatment. By using all three methods, we can learn where a protein makes specific contacts both with bases and with the phosphates in the sugar-phosphate backbone of DNA.



BOX 16-1 FIGURE 1 Footprinting Method. The stars represent the radioactive labels at the ends of the DNA fragments, arrows indicate sites where DNase cuts, and red circles represent Lac repressor bound to operator. On the left, DNA molecules cut at random by DNase are separated by size by gel electrophoresis. On the right DNA molecules are first bound to repressor then subjected to DNase treatment.

RNA polymerase binds the *lac* promoter poorly in the absence of CAP, even when there is no active repressor present. This is because the sequence of the −35 region of the *lac* promoter is not optimal for its binding, and the promoter lacks an UP-element (see Chapter 12 and Figure 16-8). This is typical of promoters that are controlled by activators.

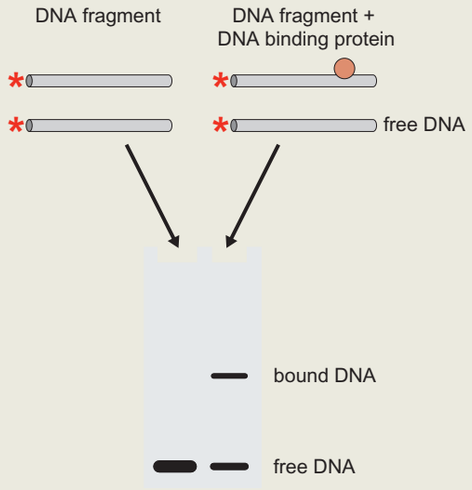
CAP binds as a dimer to a site similar in length to the *lac* operator, but different in sequence. This site is located some 60 bps upstream of the start site of transcription (see Figure 16-8). When CAP binds to that site, the activator helps polymerase bind to the promoter by interacting with the enzyme and recruiting it to the promoter (see Figure 16-6). This cooperative binding stabilizes the binding of polymerase to the promoter.

Box 16-1 ***(Continued)***

Gel Mobility Shift Assay.

As we have seen, how far a DNA molecule migrates during gel electrophoresis varies with size: the smaller the molecule the more easily it moves through the gel, and so the further it gets in a given time. If a DNA molecule has a protein bound to it migration through the gel is retarded. This forms the basis of an assay to detect specific DNA binding activities. The general approach is as follows: A short DNA fragment containing the sequence of interest is radioactively labeled so it can be detected in small quantities by polyacrylamide gel electrophoresis and autoradiography. This DNA "probe" is then mixed with the protein of interest and the mixture is run on a gel. If the protein binds to the probe, a band appears higher up the gel than bands formed from free DNA (see Box 16-1 Figure 2).

This method can be used to identify multiple proteins in a crude extract. Thus, if the probe has sites for a number of proteins found in a given cell type, and the probe is mixed with an extract of that cell type, multiple bands can be resolved. This is because proteins of different size will migrate to different extents—the larger the protein the slower the migration. In this way, for example, the various transcriptional regulators that bind to the regulatory region of a given gene can be identified.



BOX 16-1 FIGURE 2 Gel Mobility Shift Assay. The principle of the mobility shift assay is shown schematically. A protein is mixed with radiolabeled probe DNA containing a binding site for that protein. After incubation, the mixture is resolved by acrylamide gel electrophoresis and visualized using autoradiography.

CAP Has Separate Activating and DNA-Binding Surfaces

Various experiments support the view that CAP activates the *lac* genes by recruitment. Mutant versions of CAP have been isolated that bind DNA but do not activate transcription. The existence of these so-called **positive control** mutants demonstrates that, to activate transcription, the activator must do more than simply bind DNA near the promoter. Thus, activation is not caused by, for example, the activator changing local DNA structure. The amino acid substitutions in the positive control mutants identify the region of CAP that touches polymerase, called the **activating region.**

Where does the activating region of CAP touch RNA polymerase when activating the *lac* genes? This site is revealed by mutant forms of polymerase that can transcribe most genes normally, but cannot be activated by CAP at the *lac* genes. These mutants have amino acid substitutions in the **C-terminal domain** (CTD) of the **α subunit** of RNA polymerase. As we saw in Chapter 12, this domain is attached to the N-terminal domain (NTD) of α by a flexible linker. The NTD is embedded in the body of the enzyme, but the CTD extends out from it and binds the UP-element of the promoter (when that element is present). At the *lac* promoter, where there is no UP-element, CTD binds to CAP

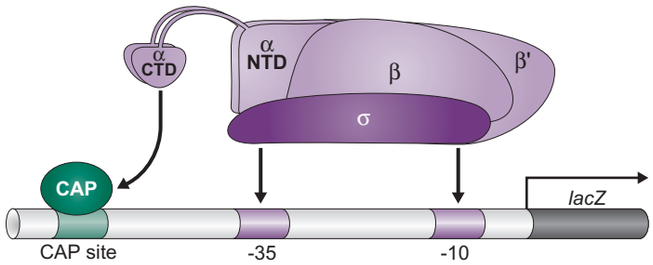


FIGURE 16-9 Activation of the *lac* Promoter by CAP. RNA polymerase binding at the *lac* promoter with the help of CAP. CAP is recognized by the CTDs of the α subunits. The CTDs also contact DNA, adjacent to the CAP site, when interacting with CAP.

and adjacent DNA instead (Figure 16-9). This picture is supported by a crystal structure of a complex containing CAP, α CTD, and a DNA oligonucleotide duplex containing a CAP site and an adjacent UP element (Figure 16-10). In Box 2 we describe an experiment, called an activator bypass, showing that activation of the *lac* promoter requires no more than polymerase recruitment.

Having seen how CAP activates transcription at the *lac* operon—and how Lac repressor counters that effect—we now look more closely at how these regulators recognize their DNA binding sites.

CAP and Lac Repressor Bind DNA Using a Common Structural Motif

X-ray crystallography has been used to determine the structural basis of DNA binding for a number of bacterial activators and repressors, including CAP and the Lac repressor. Although the details differ, the basic mechanism of DNA recognition is similar for most bacterial regulators, as we now describe.

In the typical case, the protein binds as a homodimer to a site that is an inverted repeat (or near repeat). One monomer binds each half-site, with the axis of symmetry of the dimer lying over that of the binding site (as we saw for Lac repressor, Figure 16-7). Recognition

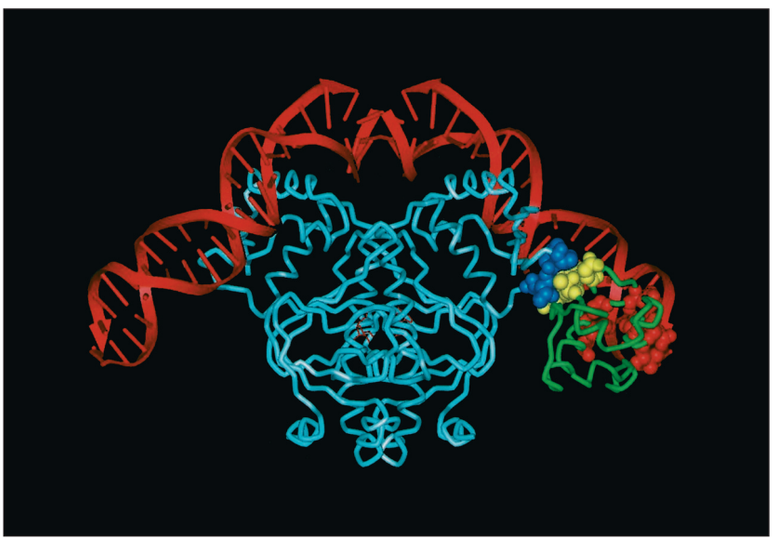
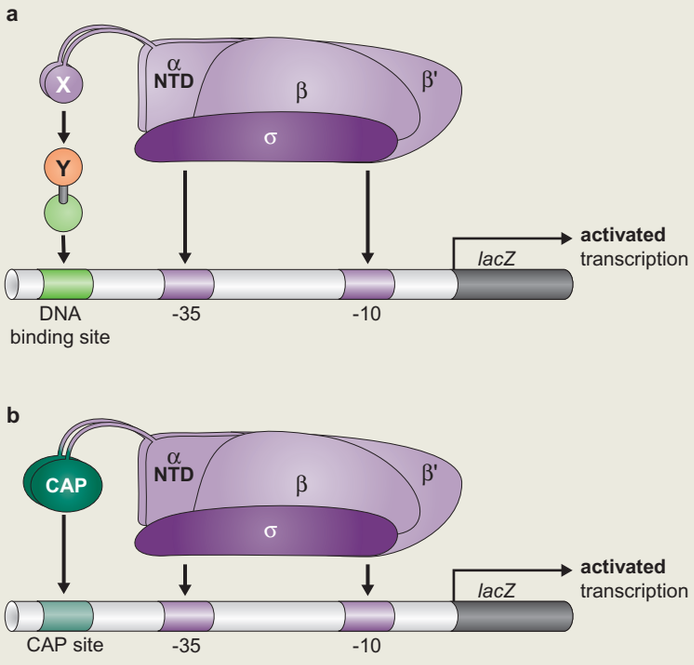


FIGURE 16-10 Structure of CAP-αCTD-DNA Complex. The crystal structure of CAP bound to its site on DNA interacting with the αCTD of RNA polymerase. DNA and cAMP (in the center of CAP) are shown in red; CAP is represented in light blue with its "activating region 1" shown in dark blue; αCTD is shown in green with its point of CAP contact in yellow and its point of DNA contact shown in red. (Source: Reproduced, with permission, from Benoff et al. 2002. *Science* **297:** 1562–1566, Figure 2A.)

Box 16-2 Activator Bypass Experiments

If an activator has only to recruit polymerase to the gene, then other ways of bringing the polymerase to the gene should work just as well. This turns out to be true of the *lac* genes, as shown by the following experiments (Box 16-2 Figure 1).

In one experiment, another protein : protein interaction is used in place of that between CAP and polymerase. This is done by taking two proteins known to interact with each other, attaching one to a DNA-binding domain, and, with the other, replacing the C-terminal domain of the polymerase α subunit (αCTD). The modified polymerase can be activated by the makeshift "activator" as long as the appropriate DNA-binding site is introduced near the promoter. In another experiment, the αCTD of polymerase is replaced with a DNA-binding domain (for example, that of CAP). This modified polymerase efficiently initiates transcription from the *lac* promoter in the absence of any activator, as long as the appropriate DNA-binding site is placed nearby. A third experiment is even simpler: polymerase can transcribe the *lac* genes at high levels in the absence of any activator if the enzyme is present at high concentration. So we see that either recruiting polymerase artificially or supplying it at a high concentration is sufficient to produce activated levels of expression of the *lac* genes. This would not be the case if the activator had to induce a specific allosteric change in polymerase to activate transcription.



BOX 16-2 FIGURE 1 Two Activator Bypass Experiments. (a) The αCTD is replaced by a protein X, which interacts with protein Y. Protein Y is fused to a DNA-binding domain, and the site recognized by that domain is shown placed near the *lac* genes. (b) The αCTD is replaced by the DNA-binding portion of CAP.

of specific DNA sequences is achieved using a conserved region of secondary structure called a **helix-turn-helix** (Figure 16-11). This domain is composed of two α helices, one of which—the **recognition helix**—fits into the major groove of the DNA. As we discussed in Chapter 5, an α helix is just the right size to fit into the major groove, allowing amino acid residues on its outer face to interact with chemi-

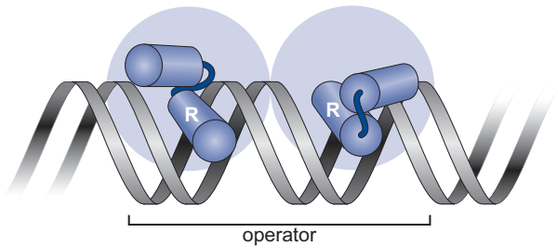


FIGURE 16-11 Binding of Dimeric Protein with Helix-Turn-Helix Domain to DNA. The shaded circles represent two identical subunits of a DNA-binding protein bound to an operator. The helix-turn-helix motif on each monomer is indicated; the "recognition helix" is labeled R.

cal groups on the edges of base pairs. Recall that in Chapter 6 we saw how each base pair presents a characteristic pattern of hydrogen bonding acceptors and donors. Thus, a protein can distinguish different DNA sequences in this way without unwinding the DNA duplex (Figure 16-11).

The contacts made between the amino acid side chains protruding from the recognition helix and the edges of the bases can be mediated by direct H-bonds, indirect H-bonds (bridged by water molecules), or by Van der Waals forces, as discussed in detail in Chapters 3, 5, and 6. Figure 16-12 illustrates an example of the interactions made by a given recognition helix and its DNA binding site.

The second helix of the helix-turn-helix domain sits across the major groove and makes contact with the DNA backbone, ensuring proper presentation of the recognition helix, and at the same time adding binding energy to the overall protein–DNA interaction.

This description is essentially true for not only CAP and Lac repressor, but for many other bacterial regulators as well, including the phage λ repressor and Cro proteins we will encounter in a later section; there are differences in detail, as the following examples illustrate.

- Lac repressor binds as a tetramer, not a dimer. Nevertheless, each operator is contacted by only two of these subunits. Thus, the different oligomeric form does not alter the mechanism of DNA recognition. The other two monomers within the tetramer can bind one of two other *lac* operators, located 400 bps downstream and 90

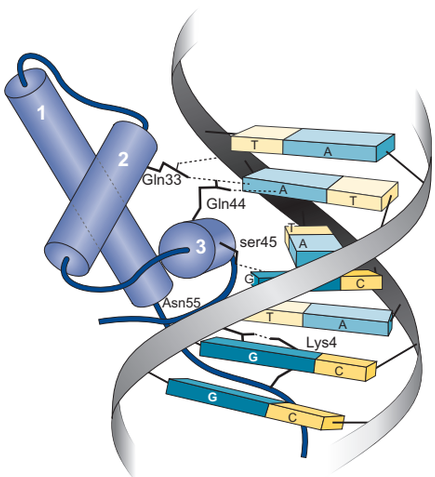


FIGURE 16-12 Hydrogen Bonds between λ Repressor and Base Pairs in the Major Groove of the Operator. Diagram of the repressor-operator complex, showing hydrogen bonds (in dotted lines) between amino acid side chains and bases in the consensus half-site. Only the important amino acid side chains are shown. In addition to Gln44 and Ser45 in the recognition helix, Asn55 in the loop following the recognition helix also makes contact with a specific base. Furthermore (and unusual to this case, see text) Lys4 in the N-terminal arm of the protein makes a contact in the major groove on the opposite face of the DNA helix. Gln33 contacts the backbone. (Source: Redrawn from Jordan, S. and Pabo, C. *Science* **242:** 896, Fig. 3B.)

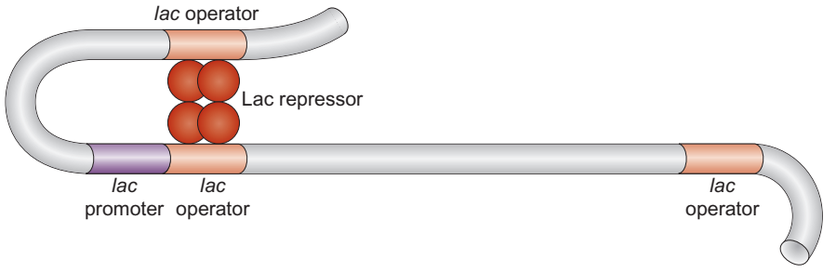


FIGURE 16-13 Lac Repressor Binds as a Tetramer to Two Operators. The loop shown is between the Lac repressor bound at the primary operator and the upstream auxiliary one. A similar loop can alternatively form with the downstream operator. In this figure, each repressor dimer is shown as two circles, rather than as a single oval (as used in earlier figures).

bps upstream of the primary operator. In such cases, the intervening DNA loops out to accommodate the reaction (Figure 16-13).

- In some cases, other regions of the protein, outside the helix-turn-helix domain, also interact with the DNA. λ repressor, for example, makes additional contacts using N-terminal arms. These reach around the DNA and interact with the minor groove on the back face of the helix (see Figure 16-12).
- In many cases, binding of the protein does not alter the structure of the DNA. In some cases, however, various distortions are seen in the protein-DNA complex. For example, CAP induces a dramatic bend in the DNA, partially wrapping it around the protein. This is caused by other regions of the protein, outside the helix-turn-helix domain, interacting with sequences outside the operator. In other cases, binding results in twisting of the operator DNA.

Not all prokaryotic repressors bind using a helix-turn-helix. A few have been described that employ quite different approaches. A striking example is the Arc repressor from phage P22 (a phage related to λ but one which infects *Salmonella*). The Arc repressor binds as a dimer to an inverted repeat operator, but instead of an alpha helix, it recognizes its binding site using two antiparallel β-strands inserted into the major groove.

The Activities of Lac Repressor and CAP Are Controlled Allosterically by Their Signals

When lactose enters the cell, it is converted to allolactose. It is the enzyme β-galactosidase that triggers this conversion. But as β-galactosidase is itself the product of one of the *lac* genes, how is this possible? The answer is that expression of the *lac* genes is leaky: even when they are repressed, an occasional transcript gets made. That happens because every so often RNA polymerase will manage to bind the promoter in place of Lac repressor. This leakiness ensures there is a low level of β-galactosidase in the cell even in the absence of lactose, and so there is enzyme poised to catalyze the conversion of lactose to allolactose.

Allolactose binds to Lac repressor and triggers a change in the shape (conformation) of that protein. In the absence of allolactose, repressor is present in a form that binds its site on DNA (and so keeps the *lac* genes switched off). Once allolactose has altered its shape, repressor can no longer bind DNA, and so the *lac* genes are no longer repressed. In Chapter 5 we described in detail the structural basis of this allosteric change in Lac repressor. An important point to emphasize is that allolactose binds to a part of Lac repressor distinct from its DNA binding domain.

CAP activity is regulated in a similar manner. Glucose lowers the intracellular concentration of a small effector molecule, cAMP. This molecule is the allosteric effector for CAP: only when CAP is complexed with cAMP does the protein adopt a conformation that binds DNA. Thus, only when glucose levels are low (and cAMP levels high) does CAP bind DNA and activate the *lac* genes. The part of CAP that binds the effector, cAMP, is separate from the part of the protein that binds DNA.

The *lac* operon of *E. coli* is one of the two systems used by French biologists François Jacob and Jacques Monod in formulating the early ideas about gene regulation. In Box 16-3 we give a brief description of those early studies and why the ideas they generated have proved so influential.

Combinatorial Control: CAP Controls Other Genes As Well

The *lac* genes provide an example of **signal integration**: their expression is controlled by two signals, each of which is communicated to the genes via a single regulator—the Lac repressor and CAP, respectively.

Consider another set of *E. coli* genes, the *gal* genes. These encode enzymes involved in galactose metabolism. As is the case with the *lac* genes, the *gal* genes are only expressed when their substrate sugar—in this case galactose—is present, and the preferred energy source, glucose, is absent. Again, analogous to *lac,* the two signals are communicated to the genes via two regulators—an activator and a repressor. The repressor, encoded by the gene *galR,* mediates the effects of the inducer galactose, but the activator of the *gal* genes is again CAP. Thus a regulator (CAP) works together with different repressors at different genes. This is an example of **combinatorial control.** In fact, CAP acts at more than 100 genes in *E. coli,* working with an array of partners.

Combinatorial control is a characteristic feature of gene regulation: when the same signal controls multiple genes, it is typically communicated to each of those genes by the same regulatory protein. That regulator will be communicating just one of perhaps several signals involved in regulating each gene; the other signals, different in most cases, will each be mediated by a separate regulator. More complex organisms—higher eukaryotes in particular—tend to have more signal integration, and there we will see greater and more elaborate examples of combinatorial control.

Alternative σ Factors Direct RNA Polymerase to Alternative Sets of Promoters

Recall from Chapter 12 that it is the σ subunit of RNA polymerase that recognizes the promoter sequences. The *lac* promoter we have been discussing, along with the bulk of other *E. coli* promoters, is recognized by RNA polymerase bearing the σ^{70} subunit. *E. coli* encodes several other σ subunits that can replace σ^{70} under certain circumstances and direct the polymerase to alternative promoters. One of these alternatives is the heat shock σ factor, σ^{32}. Thus, when *E. coli* is subject to heat shock, the amount of this new σ factor increases in the cell, it displaces σ^{70} from a proportion of RNA polymerases, and directs those enzymes to transcribe genes whose products protect the cell from the effects of heat shock. The level of σ^{32} is increased by two mechanisms: first, its translation is stimulated—that is, its mRNA is translated with greater efficiency after heat shock than it was before;

Box 16-3 Jacob, Monod, and the Ideas Behind Gene Regulation

The idea that the expression of a gene can be controlled by the product of another gene—that there exist regulatory genes the sole function of which is regulating the expression of other genes—was one of the great insights from the early years of molecular biology. It was proposed by a group of scientists working in Paris in the 1950s and early 1960s, in particular François Jacob and Jacques Monod. They sought to explain two apparently unrelated phenomena: the appearance of β-galactosidase in *E. coli* grown in lactose, and the behavior of the bacterial virus (bacteriophage) λ upon infection of *E. coli.* Their work culminated in publication of their operon model in 1961 (and the 1965 Nobel Prize for medicine, which they shared with their colleague, Andre Lwoff).

It is difficult to appreciate the magnitude of their achievement now that we are so familiar with their ideas and have such direct ways of testing their models. To put it in perspective, consider what was known at the time they began their classic experiments: β-galactosidase activity appeared in *E. coli* cells only when lactose was provided in the growth medium. It was not clear that the appearance of this enzyme involved switching on expression of a gene. Indeed, one early explanation was that the cell contained a general (generic) enzyme, and that enzyme took on whatever properties were required by the circumstances. Thus, when lactose was present, the generic enzyme took on the appropriate shape to metabolize lactose, using the sugar itself as a template!

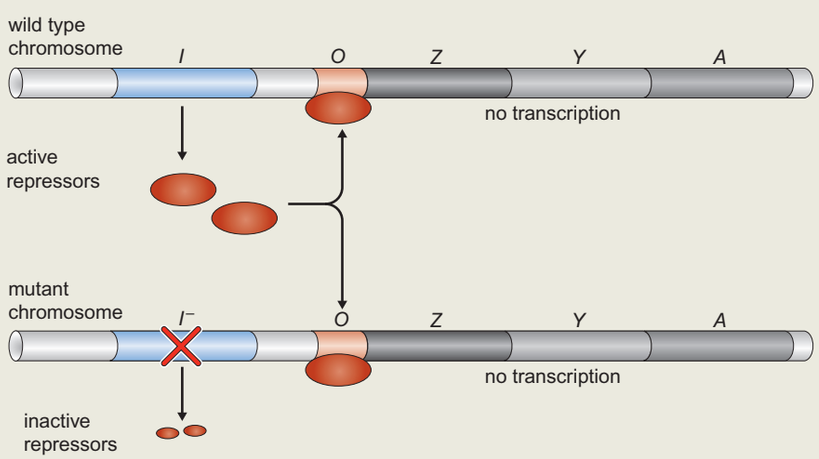
Jacob, Monod, and their coworkers dissected the problem genetically. We will not go through their experiments in any detail, but a brief summary gives a taste of their ingenuity.

First, they isolated mutants of *E. coli* that made β-galactosidase irrespective of whether lactose was present—that is, mutants in which the enzyme was produced **constitutively.** These mutants came in two classes: in one, the gene encoding the Lac repressor was inactivated; in the other, the operator site was defective. These two classes could be distinguished using a *cis-trans* test, as we now describe.

Jacob and Monod constructed partially diploid cells in which a section of the chromosome from a wild type cell carrying the *lac* genes (that is, the Lac repressor gene, *LacI,* the genes of the *lac* operon, and their regulatory elements) was introduced (on a plasmid called an F′) into a cell carrying a mutant version of the *lac* genes on its chromosome. This transfer resulted in the presence of two copies of the *lac* genes in the cell, making it possible to test whether the wild type copy could complement various mutant copies. When the chromosomal genes were expressed constitutively because of a mutation in the *lacI* gene (encoding repressor), the wild type copy on the plasmid restored repression (and inducibility)—i.e., β-galactosidase was once again only made when lactose was present (Box 16-3 Figure 1). This is because the repressor made from the wild type *lacI* gene on the plasmid could diffuse to the chromosome—that is, it could act in *trans*.

When the mutation causing constitutive expression of the chromosomal genes was in the *lac* operator, it could not be complemented in *trans* by the wild type genes (Box 16-3 Figure 2). The operator functions only in *cis* (that is, it only acts on the genes directly linked to it on the same DNA molecule).

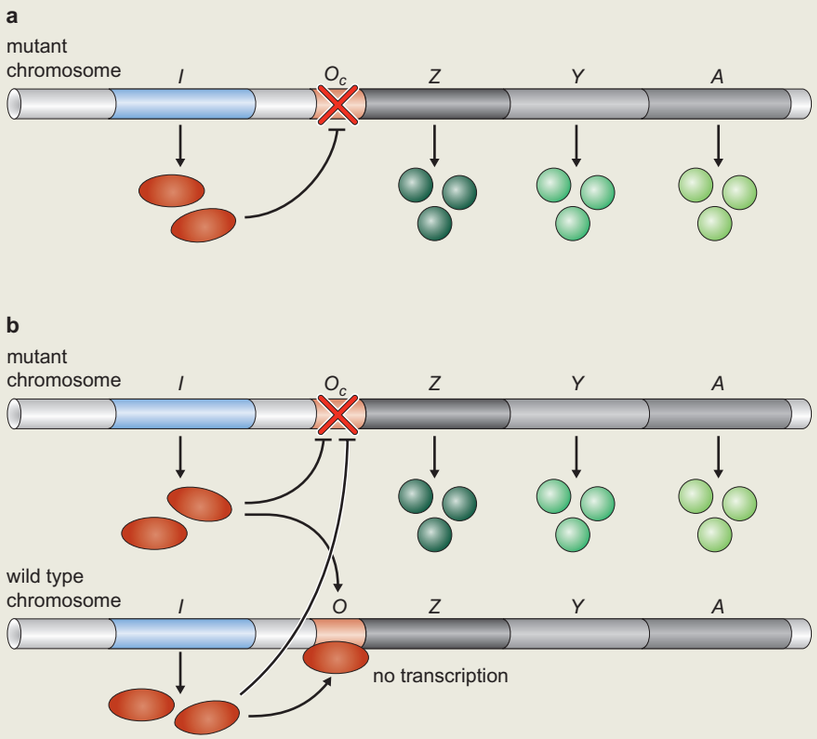
These and other results led Jacob and Monod to propose that genes were expressed from specific sites called promoters found at the start of the gene and that this expression was regulated by repressors that act in *trans* through operator sites located on the DNA beside the promoter.



BOX 16-3 FIGURE 1 Partial Diploid Cells Show That Functional Repressors Are Dominant over Inactive Repressors. In the absence of externally added β-galactosides, no significant amounts of β-galactosidase molecules will be produced in these cells.

Box 16-3 (*continued*)

But these experiments with the *lac* system were not carried out in isolation; in parallel, Jacob and Monod did similar experiments on phage λ (a system we consider in detail later in this chapter). The phage can propogate through either of two life cycles. Which is chosen depends on which of the relevant phage genes are expressed. The French scientists found they could isolate mutants defective in controlling gene expression in this system just as they had in the *lac* case. These mutations again defined a repressor that acted in *trans* through *cis* acting operator sites. The similarity of these two regulatory systems convinced Jacob and Monod that they had identified a fundamental mechanism of gene regulation and that their model would apply throughout nature. As we will see, although their description was not complete—most noticeably, they did not include activators (such as CAP) in their scheme—the basic model they proposed of *cis* regulatory sites recognized by *trans* regulatory factors has dominated the vast majority of subsequent thinking about gene regulation.



BOX 16-3 FIGURE 2 The Control of Specific mRNA Synthesis by Normal and Mutant Operators. (a) Haploid cell containing mutant operator (O_c). (b) Partially diploid cell containing a normal operator (O) and a mutant operator (O_c). The O_c is dominant over the O form.

and second, the protein is transiently stabilized. Another example of an alternative σ factor, σ^{54}, is considered in the next section. σ^{54} is associated with a small fraction of the polymerase molecules in the cell and directs that enzyme to genes involved in nitrogen metabolism.

Sometimes a series of alternative sigmas directs a particular program of gene expression. Two examples are found in the bacterium *B. subtilis.* We consider the most elaborate of these, which controls sporulation in that organism, in Chapter 18. The other we describe briefly here.

Bacteriophage SPO1 infects *B. subtilis,* where it grows lytically to produce progeny phage. This process requires that the phage express its genes in a carefully controlled order. That control is imposed on polymerase by a series of alternative σ factors. Thus, upon infection, the bacte-

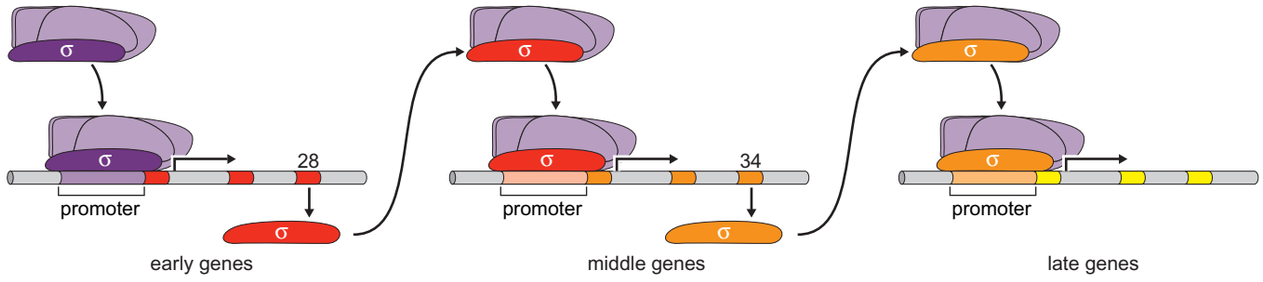


FIGURE 16-14 Alternative σ Factors Control the Ordered Expression of Genes in a Bacterial Virus. The bacterial phage SPO1 uses three σ factors in succession to regulate expression of its genome. This ensures that viral genes are expressed in the order in which they are needed.

rial RNA polymerase (bearing the *B. subtilis* version of σ^{70}) recognizes so-called "early" phage promoters, which direct transcription of genes that encode proteins needed early in infection. One of these genes (called gene 28) encodes an alternative σ. This displaces the bacterial σ factor and directs the polymerase to a second set of promoters in the phage genome, those associated with the so-called "middle" genes. One of these genes in turn encodes the σ factor for the phage "late" genes (Figure 16-14).

NtrC and MerR: Transcriptional Activators That Work by Allostery Rather Than by Recruitment

Although the majority of activators work by recruitment, there are exceptions. Two examples of activators that work not by recruitment but by allosteric mechanisms are NtrC and MerR. Recall what we mean by an allosteric mechanism. Activators that work by recruitment simply bring an active form of RNA polymerase to the promoter. In the case of activators that work by allosteric mechanisms, polymerase initially binds the promoter in an inactive complex. To activate transcription, the activator triggers an allosteric change in that complex.

NtrC controls expression of genes involved in nitrogen metabolism, such as the *glnA* gene. At the *glnA* gene, RNA polymerase is prebound to the promoter in a stable closed complex. The activator NtrC induces a conformational change in the enzyme, triggering transition to the open complex. Thus the activating event is an allosteric change in RNA polymerase (see Figure 16-2).

MerR controls a gene called *merT,* which encodes an enzyme that makes cells resistant to the toxic effects of mercury. MerR also acts on an inactive RNA polymerase–promoter complex. Like NtrC, MerR induces a conformational change that triggers open complex formation. In this case, however, the allosteric effect of the activator is on the DNA rather than the polymerase.

NtrC Has ATPase Activity and Works from DNA Sites Far from the Gene

As with CAP, NtrC has separate activating and DNA-binding domains and binds DNA only in the presence of a specific signal. In the case of NtrC, that signal is low nitrogen levels. Under those conditions, NtrC is

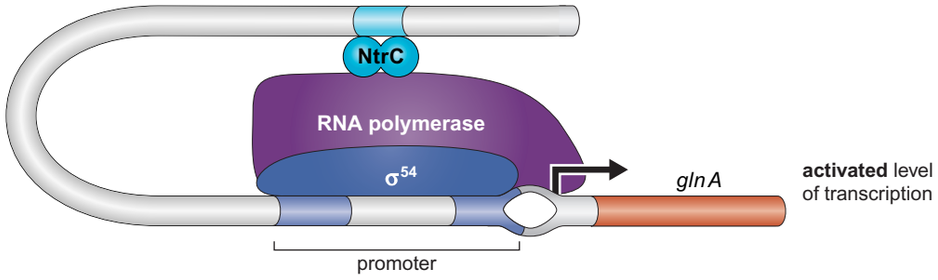


FIGURE 16-15 Activation by NtrC. The promoter sequence recognized by σ^{54}-containing holoenzyme is different from that recognized by σ^{70}-containing holoenzyme (see Chapter 12). Although not specified in the figure, NtrC contacts the σ^{54} subunit of polymerase. NtrC is shown as a dimer, but in fact forms a higher order complex on DNA.

phosphorylated by a kinase, NtrB, and as a result undergoes a conformational change that reveals the activator's DNA-binding domain. Once active, NtrC binds four sites located some 150 base pairs upstream of the promoter. NtrC binds to each of its sites as a dimer, and, through protein:protein interactions between the dimers, binds to the four sites in a highly cooperative manner.

The form of RNA polymerase that transcribes the *glnA* gene contains the σ^{54} subunit. This enzyme binds to the *glnA* promoter in a stable, closed complex in the absence of NtrC. Once active, NtrC (bound to its sites upstream) interacts directly with σ^{54}. This requires that the DNA between the activator binding sites and the promoter form a loop to accommodate the interaction. If the NtrC binding sites are moved further upstream (as much as 1 to 2 kb) the activator can still work (Figure 16-15).

NtrC itself has an enzymatic activity—it is an ATPase; this activity provides the energy needed to induce a conformational change in polymerase. That conformational change triggers polymerase to initiate transcription. Specifically, it stimulates conversion of the stable, inactive, closed complex to an active, open complex.

At some genes controlled by NtrC, there is a binding site for another protein, called IHF, located between the NtrC binding sites and the promoter. Upon binding, IHF bends DNA; when the IHF binding site—and hence the DNA bend—are in the correct register, this event increases activation by NtrC. The explanation is that, by bending the DNA, IHF brings the DNA-bound activator closer to the promoter, helping the activator interact with the polymerase bound there (see Figure 16-4).

MerR Activates Transcription by Twisting Promoter DNA

When bound to a single DNA binding site, in the presence of mercury, MerR activates the *merT* gene. As shown in Figure 16-16, MerR binds to a sequence located between the −10 and −35 regions of the *merT* promoter (this gene is transcribed by σ^{70}-containing polymerase). MerR binds on the opposite face of the DNA helix from that bound by RNA polymerase, and so polymerase can (and does) bind to the promoter at the same time as MerR.

The *merT* promoter is unusual. The distance between the −10 and −35 elements is 19 bp instead of the 15 to 17 bp typically found in a σ^{70} promoter (see Chapter 12). As a result, these two sequence elements recognized by σ are neither optimally separated nor aligned;

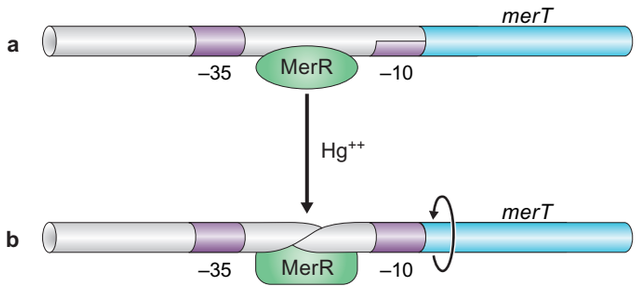


FIGURE 16-16 Activation by MerR. The −10 and −35 elements of the *merT* promoter lie on nearly opposite sides of the helix. (a) In the absence of mercury, MerR binds and stabilizes the inactive form of the promoter. (b) In the presence of mercury, MerR twists the DNA so as to properly align the promoter elements.

they are somewhat rotated around the face of the helix in relation to each other. Furthermore, the binding of MerR (in the absence of Hg^{2+}) locks the promoter in this unpropitious conformation: polymerase can bind, but not in a manner that allows it to initiate transcription. Therefore, there is no basal transcription.

When MerR binds Hg^{2+}, however, the protein undergoes a conformational change that causes the DNA in the center of the promoter to twist. This structural distortion restores the disposition of the −10 and −35 regions to something close to that found at a strong σ^{70} promoter. In this new configuration, RNA polymerase can efficiently initiate transcription. The structures of promoter DNA in the 'active' and 'inactive' states have been determined (for another promoter regulated in this manner) and are shown in Figure 16-17.

It is important to note that in this example the activator does not interact with RNA polymerase to activate transcription, but instead alters the conformation of the DNA in the vicinity of the prebound enzyme. Thus, unlike the earlier cases, there is no separation of DNA binding and activating regions: for MerR, DNA binding is intimately linked to the activation process.

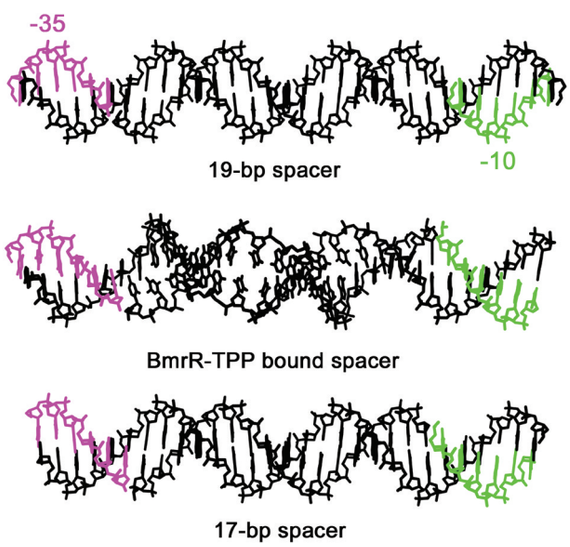


FIGURE 16-17 Structure of a *merT*-like Promoter. (a) Promoter with 19-bp spacer. (b) Promoter with a 19-bp spacer when in complex with active activator. (c) Promoter with a 17-bp spacer. The promoter shown is from the *bmr* gene of *Bacillus subtilis,* which is controlled by the regulator BmrR. BmrR works as an activator when complexed with the drug tetraphenylphosphonium (TPP). The −35 (TTGACT) and −10 (TACAGT) elements of one strand are shown in pink and green, respectively. (Source: Adapted, with permission, from Zheleznova Heldwein, E. E. and Brennan, R. G. 2001. *Nature* **409:**378; Figure 3 b, c, d.)

Some Repressors Hold RNA Polymerase at the Promoter Rather Than Excluding It

Lac repressor works in the simplest possible way: by binding to a site overlapping the promoter, it blocks RNA polymerase binding. Most repressors work in that same way. In the MerR case, we saw a different form of repression; in that case the protein holds the promoter in a conformation incompatible with transcription initiation. There are other ways repressors can work, one of which we now consider.

Some repressors work from binding sites that do not overlap the promoter. Those repressors do not block polymerase binding—rather they bind to sites beside a promoter, interact with polymerase bound at that promoter, and inhibit initiation. One is the *E. coli* Gal repressor, which we mentioned earlier. The Gal repressor controls genes that encode enzymes involved in galactose metabolism; in the absence of galactose the repressor keeps the genes off. In this case, the repressor interacts with the polymerase in a manner that inhibits transition from the closed to open complex.

Another example is provided by the P_4 protein from a bacteriophage (ϕ29) that grows on the bacterium *B. subtilis.* This regulator binds to a site adjacent to one promoter—a weak promoter called P_{A3}—and, by interacting with polymerase, serves as an activator. The interaction is with the α CTD, just as we saw with CAP. But this activator also binds at another promoter—a strong promoter called P_{A2c}. Here it makes the same contact with polymerase as at the weak promoter, but the result is repression. It seems that whereas in the former case the extra binding energy helps recruit polymerase, and hence activates the gene, in the latter case, the overall binding energy—provided by the strong interactions between the polymerase and the promoter and the additional interaction provided by the activator—is so strong that the polymerase is unable to escape the promoter.

AraC and Control of the *araBAD* Operon by Anti-Activation

The promoter of the *araBAD* operon from *E. coli* is activated in the presence of arabinose and the absence of glucose and directs expression of genes encoding enzymes required for arabinose metabolism. Two activators work together here: AraC and CAP. When arabinose is present, AraC binds that sugar and adopts a configuration that allows it to bind DNA as a dimer to the adjacent half sites, $araI_1$ and $araI_2$ (Figure 16-18a). Just upstream of these is a CAP site: in the absence of glucose, CAP binds here and helps activation.

In the absence of arabinose the *araBAD* genes are not expressed. Under these conditions AraC adopts a different conformation and binds DNA in a different way: one monomer still binds the $araI_1$ site, but the other monomer binds a distant half site called $araO_2$, as shown in Figure 16-18b. As these two half sites are 194 bp apart, when AraC binds in this fashion the DNA between the two sites forms a loop. When bound in this way there is no monomer of AraC bound at $araI_2$, and as that is the position from which activation of *araBAD* promoter is mediated, there is no activation in this configuration.

The magnitude of induction of the *araBAD* promoter by arabinose is very large, and for this reason the promoter is often used in **expression vectors.** Expression vectors are DNA constructs in which efficient synthesis of any protein can be ensured by fusing its gene to a strong promoter (see Chapter 20). In this case, fusing a gene to the

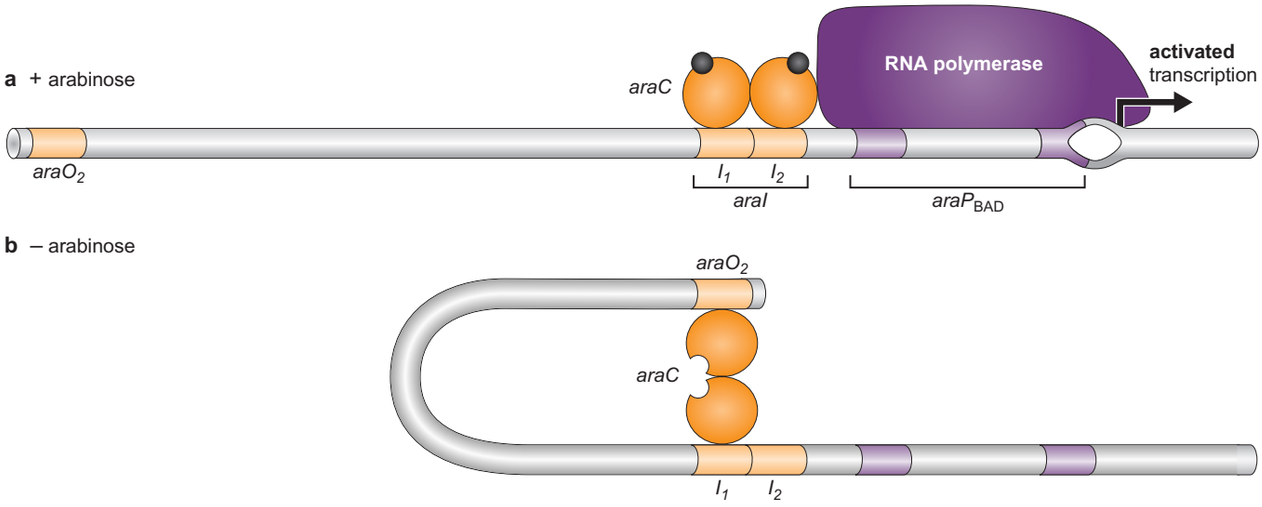


FIGURE 16-18 Control of the *araBAD* Operon. (a) Arabinose binds to AraC, changing the shape of that activator so it binds as a dimer to $araI_1$ and $araI_2$. This places one monomer of AraC close to the promoter from which it can activate transcription. (b) In the absence of arabinose, the AraC dimer adopts a different conformation and binds to *araO* and $araI_1$. In this position there is no monomer at site $araI_2$, and so the protein cannot activate the *araBAD* promoter. This promoter is also controlled by CAP, but that is not shown in this figure.

araBAD promoter allows expression of the gene to be controlled by arabinose. In that way, the gene can be kept off until expression is desirable, and then "induced" when its product is wanted. This allows expression of even those genes with products that are toxic to the bacterial cells.

TWO EXAMPLES OF GENE REGULATION AT STEPS AFTER TRANSCRIPTION INITIATION

Amino Acid Biosynthetic Operons Are Controlled by Premature Transcription Termination

In *E. coli* the five contiguous *trp* genes encode enzymes that synthesize the amino acid tryptophan. These genes are expressed efficiently only when tryptophan is limiting (Figure 16-19). The genes are controlled by a repressor, just as the *lac* genes are, but in this case the ligand that controls the activity of that repressor (tryptophan) acts not as an **inducer** but as a **corepressor.** That is, when tryptophan is present, it binds the Trp repressor and induces a conformational change in that protein, enabling it to bind the *trp* operator and prevent transcription. When the tryptophan concentration is low, the Trp repressor is free of its corepressor and vacates its operator, allowing the synthesis of *trp* mRNA to commence from the adjacent promoter. Surprisingly, however, once polymerase has initiated a *trp* mRNA molecule it does not always complete the full transcript. Indeed, most messages are terminated prematurely before they include even the first *trp* gene (*trpE*), unless a second and novel device confirms that little tryptophan is available to the cell.

This second mechanism overcomes the premature transcription termination, called **attenuation.** When tryptophan levels are high, RNA polymerase that has initiated transcription pauses at a specific site, and then terminates before getting to *TrpE,* as we just described. When

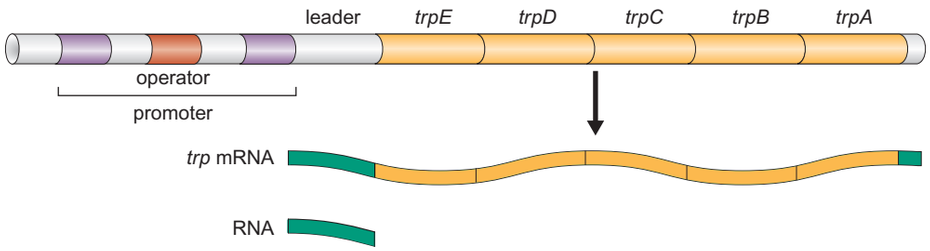


FIGURE 16-19 The *trp* Operon. The tryptophan operon of *E. coli,* showing the relation of the leader to the structural genes that code for the trp enzymes. The gene products are anthranilate synthetase (product of *trpE*), phosphoribosyl anthranilate transferase (*trpD*), phosphoribosyl anthranilate isomerase-indole glycerol phosphate synthetase (*trpC*), tryptophan synthetase β (*trpB*), and tryptophan synthetase α (*trpA*).

tryptophan is limiting, however, that termination does not occur and polymerase reads through the *trp* genes. Attenuation, and the way it is overcome, rely on the close link between transcription and translation in bacteria, and on the ability of RNA to form alternative structures through intramolecular base pairing, as we now describe.

The key to understanding attenuation came from examining the sequence of the 5′ end of *trp* operon mRNA. This analysis revealed that 161 nucleotides of RNA are made from the *tryptophan* promoter before RNA polymerase encounters the first codon of *trpE* (Figures 16-19 and 16-20). Near the end of the sequence, and before *trpE,* is a transcription terminator, composed of a characteristic hairpin loop in the RNA (made from sequences in regions 3 and 4 of Figure 16-20), followed by eight uridine residues. At this so-called **attenuator,** RNA synthesis usually stops (and, we might have thought, should always stop), yielding a leader RNA 139 nucleotides long.

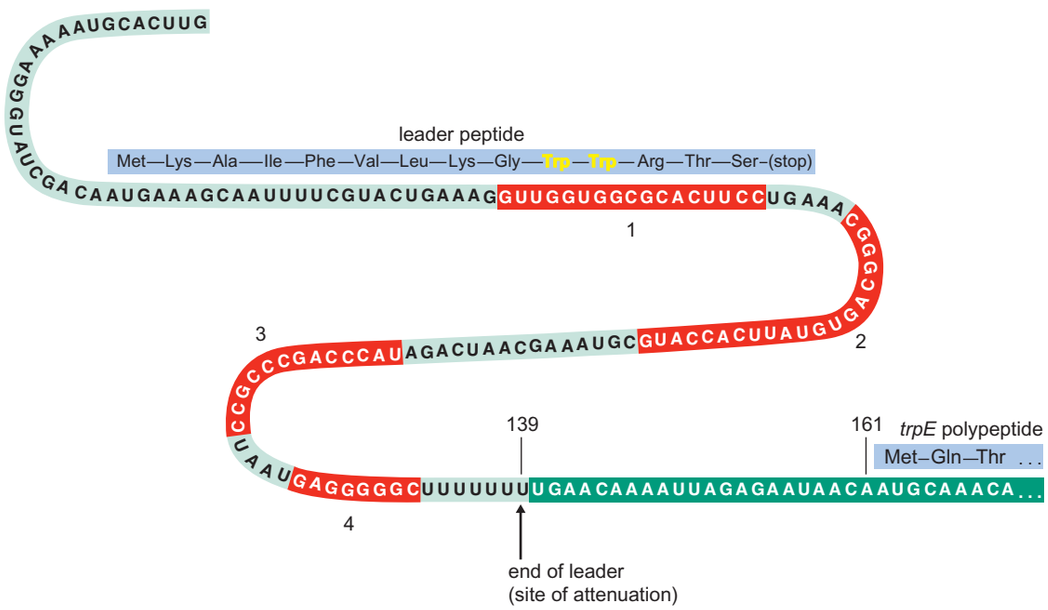


FIGURE 16-20 *Trp* Operator Leader RNA. Features of the nucleotide sequence of the *trp* operon leader RNA.

How, then, can mRNA for the whole operon ever be made? Three features of the leader sequence allow the attenuator to be passed by RNA polymerase when the cellular concentration of tryptophan is low. First, there is a second hairpin (besides the terminator hairpin) that can form between regions 1 and 2 of the leader (see Figure 16-20). Second, region 2 also is complementary to region 3; thus, yet another hairpin consisting of regions 2 and 3 can form, and when it does it prevents the terminator hairpin (3, 4) from forming. Third, the leader RNA codes for a short leader peptide of 14 amino acids that is preceded by a strong ribosome binding site (see Figure 16-20). The sequence encoding the leader peptide has a striking feature of two tryptophan codons in a row. Their importance is underscored by corresponding sequences found in similar leader peptides of other operons encoding enzymes that make amino acids (see Table 16-1). Thus, the leucine operon leader peptide has four adjacent leucine codons, and the histidine operon leader peptide has seven histidine codons in a row. In each case these operons are controlled by attenuation.

The function of these codons is to stop a ribosome attempting to translate the leader peptide; thus, when tryptophan is scarce, little charged tryptophan tRNA is available, and the ribosome stalls when it reaches the tryptophan codons. Thus, RNA around the tryptophan codons is within the ribosome and cannot be part of a hairpin loop. Figure 16-21 shows the consequence. A ribosome caught at the tryptophan codons (part b) masks region 1, leaving region 2 free to pair with region 3; thus the terminator hairpin (formed by regions 3 and 4) cannot be made, and RNA polymerase passes the attenuator and moves on into the operon, allowing Trp enzyme expression. If, on the other hand, there is enough tryptophan (and therefore enough charged Trp tRNA) for the ribosome to proceed through the tryptophan codons, the ribosome blocks sequence 2 by the time RNA containing regions 3 and 4 has been made. (Recall that transcription and translation proceed simultaneously in bacteria.) Ribosome blocking region 2 allows formation of the terminator hairpin (from regions 3 and 4), aborting transcription at the end of the leader RNA. The leader peptide itself has no function and is in fact immediately destroyed by cellular proteases.

The use of both repression and attenuation to control expression allows a finer tuning of the level of intracellular tryptophan. It provides a two-stage response to progressively more stringent tryptophan starvation—the initial response being the cessation of repressor binding, with greater starvation leading to relaxation of attenuation. But attenuation alone can provide robust regulation: other amino acid operons like *his* and *leu* have no repressors; instead, they rely entirely on attenuation for their control.

Ribosomal Proteins Are Translational Repressors of Their Own Synthesis

Regulation of translation often works in a manner analogous to transcriptional repression: a "repressor" binds to the translation start site and blocks initiation of that process. In some cases, this binding involves recognition of specific secondary structures in the mRNA. We consider here the regulation of the genes that encode ribosomal proteins.

Correct expression of ribosomal protein genes poses an interesting regulatory problem for the cell. Each ribosome contains some 50 distinct proteins that must be made at the same rate. Furthermore, the

TABLE 16-1 Leader Peptides of Attenuator-Controlled Operons Containing Genes for Amino Acid Biosynthesis*

Operon	Amino Acid Sequence of Leader Peptides
Tryptophan	Met Lys Ala Ile Phe Val Leu Lys Gly Trp Trp Arg Thr Ser
Threonine	Met Lys Arg Ile Ser Thr Thr Ile Thr Thr Thr Ile Thr Ile Thr Thr Gly Asn Gly Ala Gly
Histidine	Met Thr Arg Val Gln Phe Lys His His His His His His His Pro Asp
Isoleucine-valine GEDA	Met Thr Ala Leu Leu Arg Val Ile Ser Leu Val Val Ile Ser Val Val Val Ile Ile Ile . . .
	. . . Pro Pro Cys Gly Ala Ala Leu Gly Arg Gly Lys Ala
Leucine	Met Ser His Ile Val Arg Phe Thr Gly Leu Leu Leu Leu Asn Ala Phe Ile Val Arg Gly Arg Pro . . .
	. . . Val Gly Gly Ile Gln His
Phenylalanine	Met Lys His Ile Pro Phe Phe Phe Ala Phe Phe Phe Thr Phe Pro
Isoleucine-valine B	Met Thr Thr Ser Met Leu Asn Ala Lys Leu Leu Pro Thr Ala Pro Ser Ala Ala Val Val Val . . .
	. . . Val Arg Val Val Val Val Val Gly Asn Ala Pro

*The biosynthesis of isoleucine and valine is complex: The genes are encoded in several operons, and the pathway to leucine synthesis is a branch of the valine pathway. Thus, isoleucine, valine, and leucine are all involved in attenuation of the *isoleucine-valine* operons. (*Source:* After Bauer, C., Carey, J., Kasper, L., Lynn, S., Waechter, D., and Gardner, J. *Gene Function in Prokaryotes.* 1983. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. p. 68, with permission.)

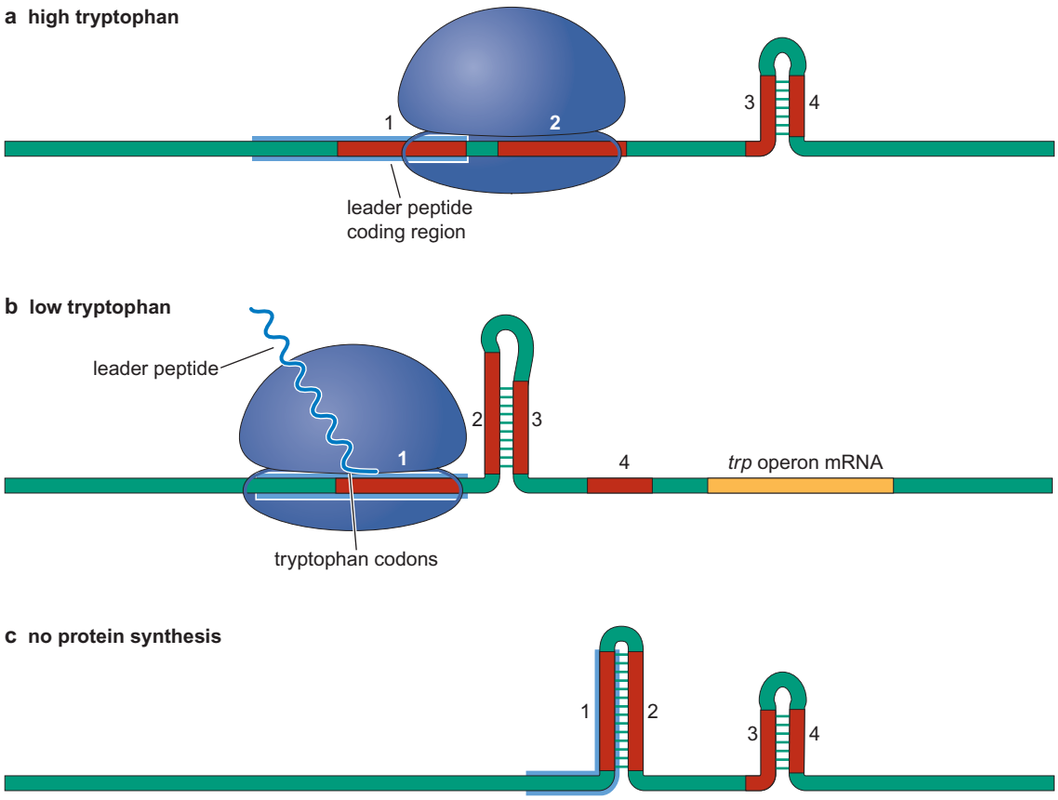


FIGURE 16-21 Transcription Termination at the *trp* Attenuator. How transcription termination at the *trp* operon attenuator is controlled by the availability of tryptophan. In (a) (conditions of high tryptophan), sequence 3 can pair with sequence 4 to form the transcription termination hairpin. In (b) (conditions of low tryptophan), the ribosome stalls at adjacent tryptophan codons, leaving sequence 2 free to pair with sequence 3, thereby preventing formation of the 3, 4, termination hairpin. In (c) (no protein synthesis), if no ribosome begins translation of the leader peptide AUG, the hairpin forms by pairing of sequences 1 and 2, preventing formation of the 2, 3, hairpin, and allowing formation of the hairpin at sequences 3, 4. The Trp enzymes are not expressed.

rate at which a cell makes protein, and thus the number of ribosomes it needs, is tied closely to the cell's growth rate; a change in growth conditions quickly leads to an increase or decrease in the rate of synthesis of all ribosomal components. How is all this coordinated regulation accomplished?

Control of ribosomal protein genes is simplified by their organization into several different operons, each containing genes for up to 11 ribosomal proteins (Figure 16-22). Some nonribosomal proteins that also are required according to growth rate are contained in these operons, including RNA polymerase subunits α, β, and β'. As with other operons, these are sometimes regulated at the level of RNA synthesis. But, the primary control of ribosomal protein synthesis is at the level of *translation* of the mRNA, not transcription. This distinction is shown by a simple experiment. When extra copies of a ribosomal protein operon are introduced into the cell, the amount of mRNA increases correspondingly, but synthesis of the proteins stays nearly the same. Thus, the cell compensates for extra mRNA by curtailing its activity as a template. This happens because ribosomal proteins are repressors of their own translation.

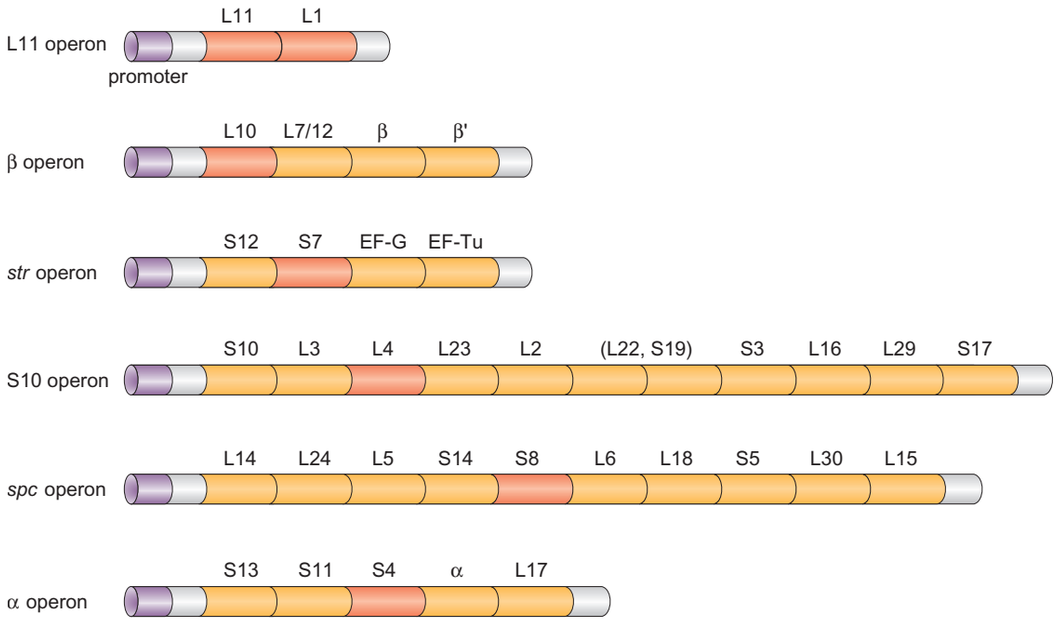


FIGURE 16-22 *E. coli* Ribosomal Protein Operons. Ribosomal protein operons of *E. coli.* The protein that in each case acts as a translational repressor of the other proteins is shaded red. (Source: After Nomura, M., Gourse, R., and Baughman, G. 1984. *Ann. Rev. Biochem.* **53**:82, with permission.)

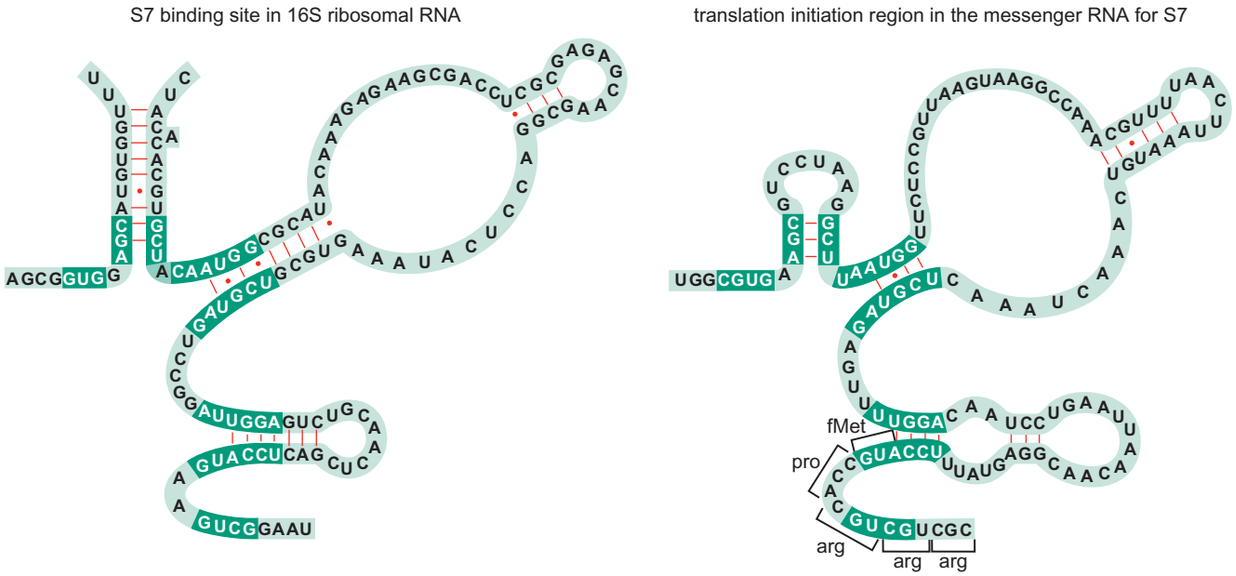


FIGURE 16-23 Ribosomal Protein S7 Binds 16S rRNA. A comparison of the region where ribosomal protein S7 (encoded by the *str* operon; Figure 16-22) binds 16S RNA in the ribosome, with the translation initiation site in its mRNA. Similar sequences are shaded in dark green. (Source: After Nomura, M., Yates, J., Dean, D., and Post, L. 1980. *Proc. Nat. Acad. Sci.* **77**:7086, with permission.)

For each operon, one (or a complex of two) of the ribosomal proteins binds the messenger near the translation initiation sequence of one of the first genes of the operon, preventing ribosomes from binding and initiating translation. Repressing translation of the first gene also prevents expression of some or all of the rest. This strategy is very sensitive. A few unused molecules of protein L4, for example, will shut down synthesis of that protein, as well as synthesis of the other ten ribosomal proteins in its operon. In this way, these proteins are made just at the rate they are needed for assembly into ribosomes (see Chapter 14).

How one protein can function both as a ribosomal component and as a regulator of its own translation is shown by comparing the sites where that protein binds to ribosomal RNA and to its messenger RNA. These sites are similar both in sequence and in secondary structure (Figure 16-23). The comparison suggests a precise mechanism of regulation. Since the binding site in the messenger includes the initiating AUG, mRNA bound by excess protein S7 (in this example) cannot attach to ribosomes to initiate translation. (This is analogous to Lac repressor binding to the *lac* promoter and thereby blocking access to RNA polymerase.) Binding is stronger to ribosomal RNA than to mRNA, so translation is repressed only when all need for the protein in ribosome assembly is satisfied.

THE CASE OF PHAGE λ: LAYERS OF REGULATION UNDERLIE A COMPLEX PROGRAM OF GENE EXPRESSION

Bacteriophage λ is a virus that infects *E. coli.* Upon infection, the phage can propagate in either of two ways: **lytically** or **lysogenically,** as illustrated in Figure 16-24. Lytic growth requires replication of the phage DNA and synthesis of new coat proteins. These components combine to form new phage particles that are released by lysis of the host cell. Lysogeny—the alternative propagation pathway—involves integration of the phage DNA into the bacterial chromosome where it is passively replicated at each cell division—just as though it were a legitimate part of the bacterial genome. A lysogen is extremely stable under normal circumstances, but the phage dormant within it—the **prophage**—can efficiently switch to lytic growth if the cell is exposed to agents that damage DNA (and thus threaten the host cell's continued existence). This switch from lysogenic to lytic growth is called **lysogenic induction.**

The choice of developmental pathway depends on which of two alternative programs of gene expression is adopted in that cell. The program responsible for the lysogenic state can be maintained stably for many generations, but then, upon induction, switch over to the lytic program with great efficiency.

Alternative Patterns of Gene Expression Control Lytic and Lysogenic Growth

λ has a 50-kb genome and some 50 genes. Most of these encode coat proteins, proteins involved in DNA replication, recombination and lysis (Figure 16-25). The products of these genes are important in making new phage particles during the lytic cycle, but our concern here is restricted to the regulatory proteins, and where they act. We

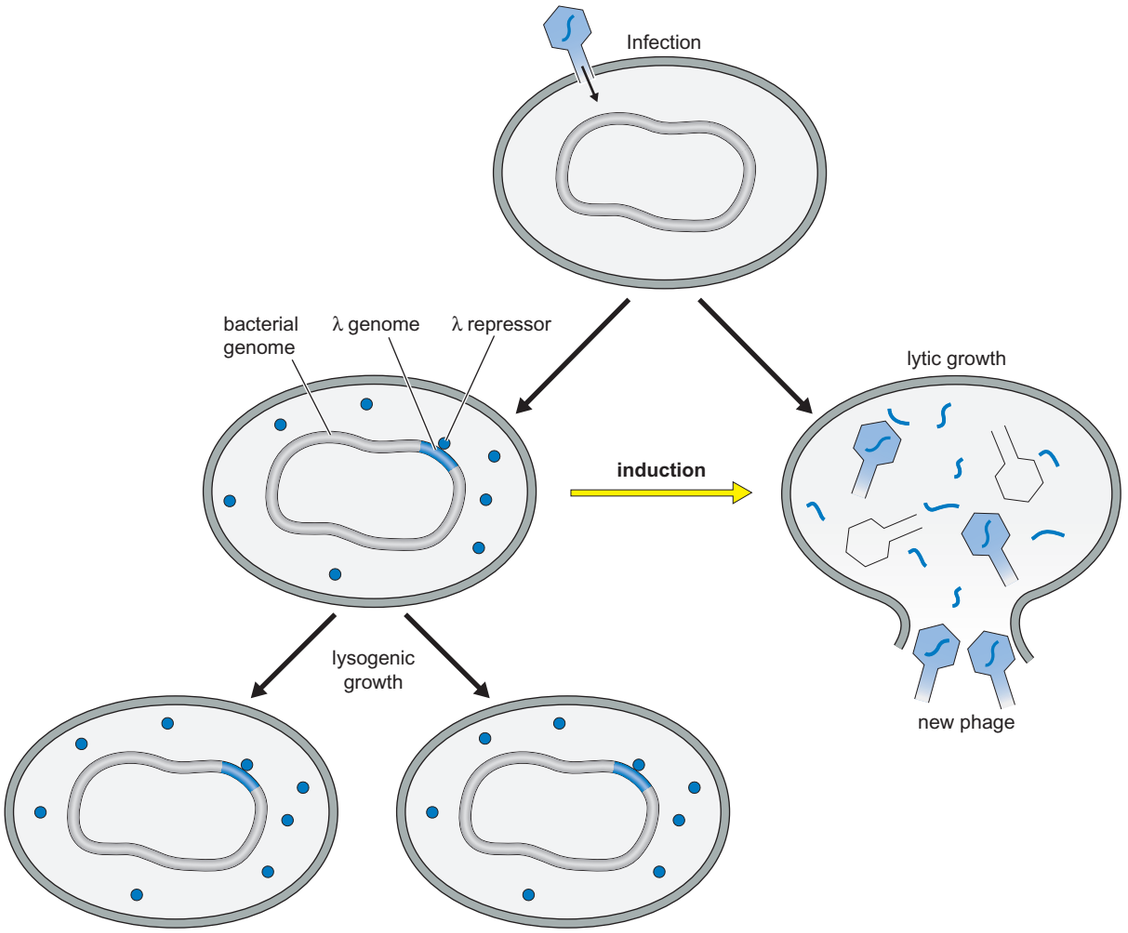


FIGURE 16-24 Growth and Induction of λ Lysogen. Upon infection, λ can grow either lytically or lysogenically. A lysogen can be propogated stably for many generations, or it can be induced. Following induction, sets of the lytic genes are expressed sequentially, leading to the production of new phage particles.

can therefore concentrate on just a few of them, and start by considering a very small area of the genome, shown in Figure 16-26.

The depicted region contains two genes (*c*I and *cro*) and three promoters (P_R, P_L, and P_{RM}). All the other phage genes (except one minor one) are outside this region and are transcribed directly from P_R and P_L (which stand for rightward and leftward promoter, respectively), or from other promoters whose activities are controlled by products of genes transcribed from P_R and P_L. P_{RM} (promoter for repressor maintenance) transcribes only the *cI* gene. P_R and P_L are strong, constitutive promoters—that is, they have the elements required to bind RNA polymerase efficiently and direct transcription without help from an activator. P_{RM}, in contrast, is a "weak" promoter and only directs efficient transcription when an activator is bound just upstream. Thus, P_{RM} resembles the *lac* promoter.

There are two arrangements of gene expression depicted in Figure 16-27: one renders growth lytic, the other lysogenic. Lytic growth proceeds when P_L and P_R remain switched on, while P_{RM} is kept off. Lysogenic growth, in contrast, is a consequence of P_L and P_R being switched off, and P_{RM} switched on. How are these promoters controlled?

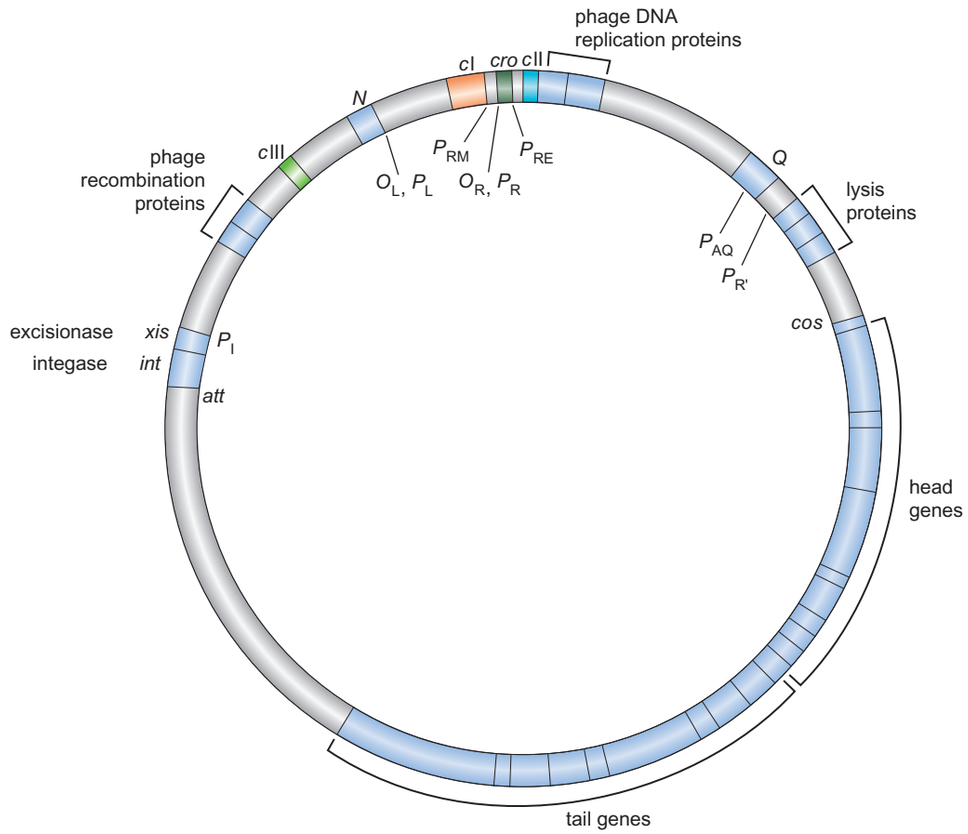


FIGURE 16-25 Map of Phage λ in the Circular Form. λ genome is linear in the phage head, but, upon infection, circularizes at the *cos* site. When integrated into the bacterial chromosome it is in a linear form, with ends at the *att* site.

Regulatory Proteins and Their Binding Sites

The *cI* gene encodes λ repressor, a protein of two domains joined by a flexible linker region (Figure 16-28). The N-terminal domain contains the DNA binding region (a helix-turn-helix domain, as we saw earlier). As with the majority of DNA binding proteins, λ repressor binds DNA as a dimer; the main dimerization contacts are made between the C-terminal domains. A single dimer recognizes a 17-bp DNA sequence, each monomer recognizing one half-site, again just as we saw in the *lac* system.

Despite its name, λ repressor can both activate and repress transcription. When functioning as a repressor, it works in the same way as does Lac repressor—it binds to sites that overlap the promoter and excludes RNA polymerase. As an activator, λ repressor works like CAP, by recruitment. λ repressor's activating region is in the N-terminal domain of the protein. Its target on polymerase is a region of the σ subunit adjacent to the part of σ that recognizes the −35 region of the promoter (see Chapter 12).

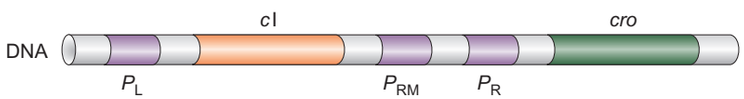


FIGURE 16-26 Promoters in the Right and Left Control Regions of Phage λ.

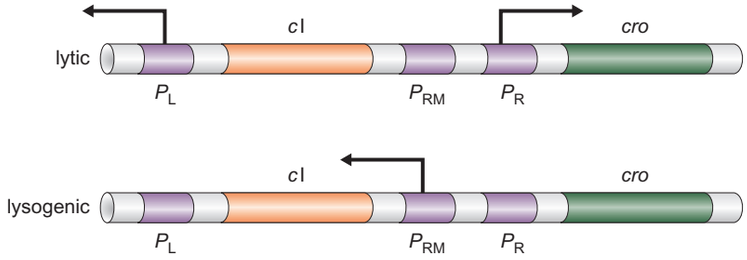


FIGURE 16-27 Transcription in the λ Control Regions in Lytic and Lysogenic Growth.

Cro (which stands for control of repressor and other things) only represses transcription, like Lac repressor. It is a single domain protein and again binds as a dimer to 17-bp DNA sequences.

λ repressor and Cro can each bind to any one of six operators. These sites, which are shown in an expansion of our picture of the control region (Figure 16-29), are recognized with different affinities by each of the proteins. We will focus on the three operators on the right of the *c*I gene, but binding of repressor and Cro to the three operators on the left follows the same pattern.

The three binding sites in the right operator are called O_{R1}, O_{R2}, and O_{R3}; these sites are similar in sequence, but not identical, and each one—if isolated from the others and examined separately—can bind either a dimer of repressor or a dimer of Cro. The affinities of these various interactions, however, are not all the same. Thus, repressor binds O_{R1} tenfold better than it binds O_{R2}. In other words, ten times more repressor—a tenfold higher concentration—is needed to bind O_{R2} than O_{R1}. O_{R3} binds repressor with about the same affinity as does O_{R2}. Cro, on the other hand, binds O_{R3} with highest affinity, and only binds O_{R2} and O_{R1} when present at tenfold higher concentration.

λ Repressor Binds to Operator Sites Cooperatively

λ repressor binds DNA cooperatively. This is critical to its function and occurs as follows. Consider repressor binding to sites in O_R. In addition to providing the dimerization contacts, the C-terminal domain of λ repressor mediates interactions *between* dimers (the point of contact is the patch marked "tetramerization" in Figure 16-28). In this way, two dimers of repressor can bind cooperatively to adjacent sites on DNA. For example, repressor at O_{R1} helps repressor bind to the lower affinity site O_{R2} by cooperative binding. Repressor thus binds both sites simultaneously and does so at a concentration that would be sufficient to bind only O_{R1} were the two sites tested sepa-

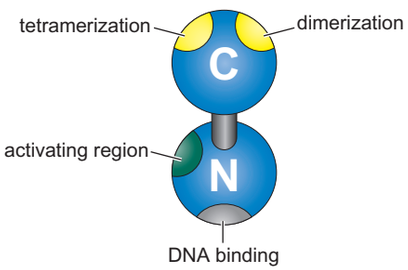


FIGURE 16-28 λ Repressor. N indicates the amino domain, C the carboxy domain. "Tetramerization" denotes the region where two dimers interact when binding cooperatively to sites on DNA. These patches mediate octamerization as well.

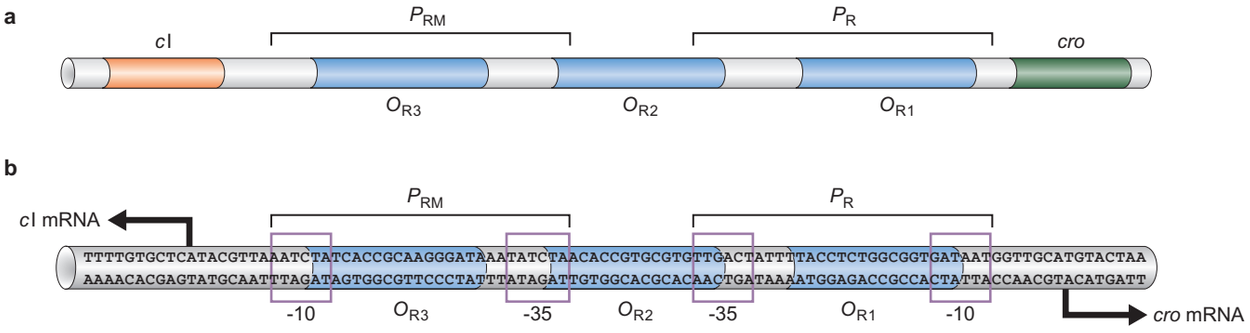


FIGURE 16-29 Relative Positions of Promoter and Operator Sites in O_R. Note that O_{R2} overlaps the −35 region of P_R by three base pairs, and that of P_{RM} by two. This difference is enough for P_R to be repressed and P_{RM} activated by repressor bound at O_{R2}. (Source: Part b redrawn, with permission, from Ptashne, M. 1992.)

rately (Figure 16-30). (Recall that, without cooperativity, a tenfold higher concentration of repressor would be needed to bind O_{R2}). O_{R3} is not bound: repressor bound cooperatively at O_{R1} and O_{R2} cannot simultaneously make contact with a third dimer at that adjacent site.

We have already discussed the idea of cooperative binding and seen an example: activation of the *lac* genes by CAP. As in that case, cooperative binding of repressors is a simple consequence of their touching each other while simultaneously binding to sites on the same DNA molecule.

A more detailed discussion of the causes and effects of cooperative binding is given in Box 16-4 on Concentration, Affinity, and Cooperative Binding. Cooperative binding of regulatory proteins is used to ensure that changes in the level of expression of a given gene can be dramatic even in response to small changes in the level of a signal that controls that gene. The lysogenic induction of λ, discussed below, provides an excellent example of this sensitive aspect of control. In some systems, cooperative binding between activators is also the basis of signal integration (see the discussion on β-interferon in Chapter 17).

Repressor and Cro Bind in Different Patterns to Control Lytic and Lysogenic Growth

How do repressor and Cro control the different patterns of gene expression associated with the different ways λ can grow?

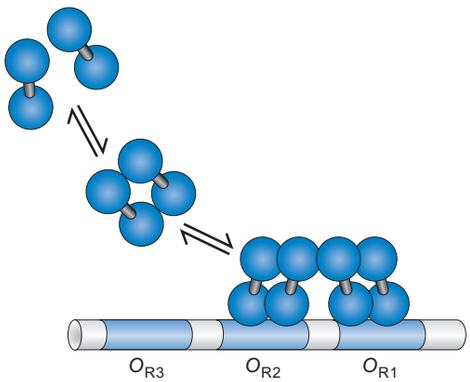


FIGURE 16-30 Cooperative Binding of λ Repressor to DNA. The λ repressor monomers interact to form dimers, and those dimers interact to form tetramers. These interactions ensure that binding of repressor to DNA is cooperative. That cooperative binding is helped further by interactions between repressor tetramers at O_R interacting with others at O_L (see later in text and Figure 16-32).

Box 16-4 Concentration, Affinity, and Cooperative Binding

What do we mean when we talk about "strong" and "weak" binding sites? When we say two molecules recognize each other, or interact with each other—such as a protein and its site on DNA—we mean they have some affinity for each other. Whether they are actually found bound together at any given time depends on two things: 1) how high that affinity is—i.e., how tightly they interact, and 2) the concentration of the molecules. As we emphasized in Chapter 5, the molecular interactions that underpin regulation in biological systems are reversible: when interacting molecules find each other, they stick together for a period of time and then separate. The higher the affinity, the tighter the two molecules stick together, and in general the longer they remain together before parting. The higher the concentration, the more often they will find each other in the first place. Thus, higher affinity or higher concentration have similar effects: they both result in the two molecules, in general, spending more time bound to each other.

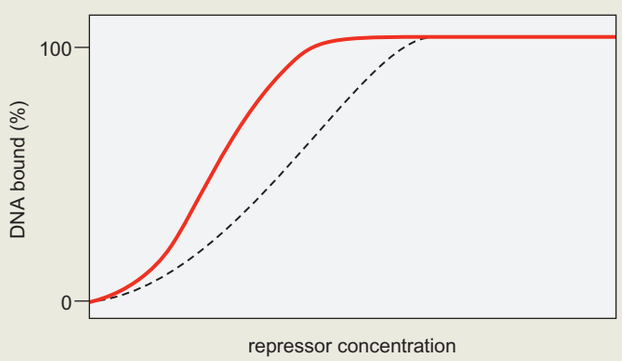
Cooperativity Visualized

Cooperativity can be expressed in terms of increased affinity. Repressor has a higher affinity for O_{R1} than for O_{R2}. But once repressor is bound to O_{R1}, repressor can bind O_{R2} more tightly because it interacts with not only O_{R2}, but with repressor bound at O_{R1} as well. Neither of these interactions is very strong alone, but when combined they substantially increase the affinity of binding of that second repressor. As we saw in Chapter 4, the relationship between binding energy and equilibrium is an exponential one (See Table 4-1). Thus, increasing the binding energy as little as twofold increases affinity by an order of magnitude.

Another way to picture how cooperativity works is to think of it as increasing the local concentration of repressor. Picture repressor bound cooperatively at O_{R1} and O_{R2}. Although repressor at O_{R2} periodically lets go of DNA, it is holding on to repressor at O_{R1} and so remains in the proximity of O_{R2}. This effectively increases the local concentration of repressor in the vicinity of that site and ensures repressor rebinds frequently. If you dispense with cooperativity and just increase the concentration of repressor in the cell, when repressor falls off O_{R2} it will not be held nearby by repressor at O_{R1} and will usually drift away before it can rebind O_{R2}. But at the higher concentrations of repressor, another molecule of repressor will likely be close to O_{R2} and bind there. Thus even if each repressor dimer only sits on O_{R2} for a short time, by either holding it nearby or increasing the number of possible replacements, you increase the likelihood of repressor being bound at any given time.

Yet another way of thinking about cooperative binding is as an entropic effect. When a protein goes from being free in solution to being constrained on a DNA binding site, the entropy of the system decreases. But repressor held close to O_{R2} by interaction with repressor at O_{R1} is already constrained compared to its free state. Rebinding of that constrained repressor has less entropic cost than does binding of free repressor.

Thus we see three ways in which cooperativity can be pictured. We should also consider some of the consequences of cooperative binding that make it so useful in biology. For example, cooperativity not only enables a weak site to be filled at a lower concentration of protein than its inherent affinity would predict, it also changes the steepness of the curve describing the filling of that site with changes in concentration. To understand what is meant by that, consider as an example a protein binding cooperatively to two weak sites, A and B. These sites will go from essentially completely empty to almost completely filled over a much narrower range of protein concentration than would a single site (see Box 16-4 Figure 1). In fact, the cooperativity in the λ system is even greater than you might expect because a large fraction of free repressor (i.e., that not bound to DNA) is



BOX 16-4 FIGURE 1 Cooperative Binding Reaction.
The dashed line shows the curve that describes binding of a protein to a single site. The steeper sigmoid curve shows cooperative binding of, for example, repressor to the λ operator sites. (Source: Modified with permission, from Ptashne, M. 1992.)

Box 16-4 (*continued*)

found as monomer in the cell; thus it is in essence a cooperative binding of four monomers rather than two stable dimers, adding to the concerted nature of complex formation on DNA, and so adding to the steepness of the curve. But why does cooperativity make the binding curve steeper?

We have already seen how the site is filled at a lower concentration of repressor than its affinity would suggest; but how is it that, as repressor concentration decreases, binding falls away so quickly? Consider interactions between components of any system: as the concentration of the components is reduced, any given interaction between two of them will occur less frequently. If the system requires multiple interactions between several different components, this will become very rare at lower concentrations. Thus, binding of four monomers of a protein to two sites requires several (in fact, seven) interactions; the chance of the individual components coming together is drastically reduced as their individual concentrations decrease.

Cooperativity and DNA Binding Specificity

A final important aspect of cooperative binding is that it imposes specificity on DNA binding. CAP activation of *lac* promoter shows this. CAP brings RNA polymerase to promoters that bear CAP sites specifically (as opposed to other promoters of comparable affinity that lack CAP sites). Likewise, λ repressor at O_{R1} directs another molecule of repressor to bind to the weak site adjacent to it, not some other site of equal affinity elsewhere in the cell. In fact, cooperativity is vital to ensuring that proteins can bind with sufficient specificity for life to work as we know it.

To illustrate this, consider a protein binding to a site on DNA. This protein has a high affinity for its correct site. But the DNA within the cell represents a huge number of potential (but incorrect) binding sites for that protein. What is important, therefore, is not simply the absolute affinity of the protein for its correct site, but its affinity for that site compared to its affinity for all the other, incorrect sites. And remember, those incorrect sites are at a much higher concentration than the correct site (representing, as they do, all the DNA in the cell except the correct site). So even if the affinity for the incorrect sites is lower than for the correct site, the higher concentration of the former ensures the protein will often sample them while attempting to reach its correct site.

What is needed is a strategy that increases affinity for the correct site without aiding interactions with the incorrect sites. Increasing the number of contacts between the protein and its DNA site (for example by making the protein larger) does not necessarily help because it also tends to increase binding to the incorrect sites. Once affinity for the incorrect sites gets too high, the protein essentially never finds its correct site; it spends too long sampling incorrect sites. Thus a kinetic problem replaces the specificity one and it can be just as disruptive.

Cooperativity solves the problem. By binding to two adjacent sites cooperatively, a protein increases dramatically its affinity for those sites, without increasing affinity for other sites. The reason it does not increase affinity for the incorrect sites is simply because the chance of two molecules of protein binding incorrect sites close together at the same time (allowing cooperativity to stabilize that binding) is extremely remote. Only when they find the correct sites do they remain bound long enough to give a second protein a chance to turn up.

Cooperativity and Allostery

Although in this chapter we use the term *cooperativity* to refer to a particular mechanism of cooperative binding, the term is also used in other contexts where different mechanisms apply. In general we might say that cooperativity describes any situation in which two ligands bind to a third molecule in such a way that the binding of one of those ligands helps the binding of the other. Thus, for the DNA-binding proteins we considered here, cooperativity is mediated by simple adhesive interactions, but in other situations cooperativity can be mediated by allosteric events. Perhaps the best example of that is the binding of oxygen molecules to hemoglobin.

Hemoglobin is a homotetramer, and each subunit binds one molecule of oxygen. That binding is cooperative: when the first oxygen binds, it causes a conformational change which fixes the binding site for the next oxygen in a conformation of higher affinity. Thus, in this case there is no direct interaction between the ligands, but by triggering an allosteric transition one ligand increases affinity for a second.

For lytic growth, a single Cro dimer is bound to O_{R3}; this site overlaps P_{RM} and so Cro represses that promoter (which would only work at a low level anyway in the absence of activator because the promoter is weak) (Figure 16-31). As neither repressor nor Cro is bound to O_{R1} and O_{R2}, P_R binds RNA polymerase and directs transcription of lytic genes; P_L does likewise. Recall that both P_R and P_L are strong promoters that need no activator.

During lysogeny, P_{RM} is on, while P_R (and P_L) are off. Repressor bound cooperatively at O_{R1} and O_{R2} blocks RNA polymerase binding at P_R, repressing transcription from that promoter. But repressor bound at O_{R2} *activates* transcription from P_{RM}.

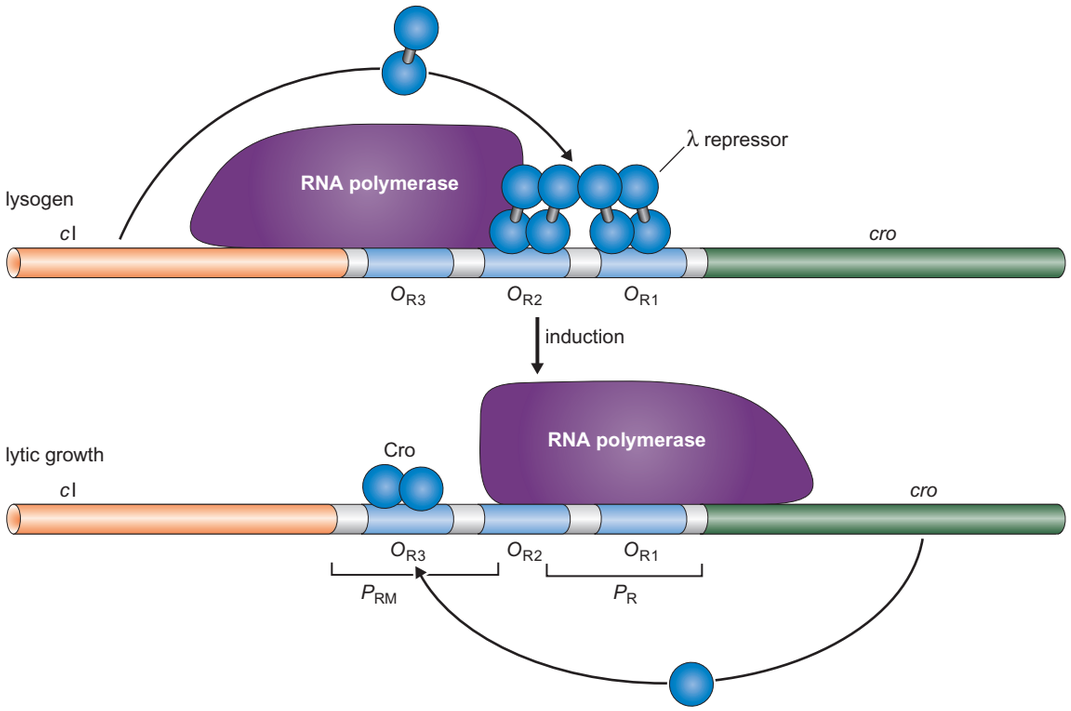


FIGURE 16-31 The Action of λ Repressor and Cro. Repressor bound to O_{R1} and O_{R2} turns off transcription from P_R. Repressor bound at O_{R2} contacts RNA polymerase at P_{RM}, activating expression of the *c*I (repressor) gene. O_{R3} lies within P_{RM}; Cro bound there represses transcription of *c*I.

We return to the question of how the phage chooses between these alternative pathways shortly. But first we consider induction—how the lysogenic state outlined above switches to the alternative lytic one when the cell is threatened.

Lysogenic Induction Requires Proteolytic Cleavage of λ Repressor

E. coli senses and responds to DNA damage. It does this by activating the function of a protein called RecA. This enzyme is involved in recombination (which accounts for its name; see Chapter 10) but it has another function. That is, it stimulates the proteolytic autocleavage of certain proteins. The primary substrate for this activity is a bacterial repressor protein called LexA that represses genes encoding DNA repair enzymes. Activated RecA stimulates autocleavage of LexA, releasing repression of those genes. This is called the SOS response (see Chapter 9).

If the cell is a lysogen, it is in the best interests of the prophage to escape under these threatening circumstances. To this end, λ repressor has evolved to resemble LexA, ensuring that λ repressor too undergoes autocleavage in response to activated RecA. The cleavage reaction removes the C-terminal domain of repressor, and so dimerization and cooperativity are immediately lost. As these functions are critical for repressor binding to O_{R1} and O_{R2} (at concentrations of repressor found in a lysogen), loss of cooperativity ensures that repressor dissociates from those sites (as well as from O_{L1} and O_{L2}). Loss of repression

triggers transcription from P_R and P_L leading to lytic growth. This switch from lysogenic to lytic growth is called **induction.**

For induction to work efficiently, the level of repressor in a lysogen must be tightly regulated. If levels were to drop too low, the lysogen might spontaneously induce; if levels rose too high, appropriate induction would be inefficient. The reason for the latter is that more repressor would have to be inactivated (by RecA) for the concentration to drop enough to vacate O_{R1} and O_{R2}. We have already seen how repressor ensures that its level never drops too low: it activates its own expression, an example of **positive autoregulation.** But how does it ensure levels never get too high? Repressor also regulates itself negatively.

This **negative autoregulation** works as follows. As drawn, Figure 16-31 shows P_{RM} being activated by repressor (at O_{R2}) to make more repressor. But if the concentration gets too high, repressor will bind to O_{R3} as well, and repress P_{RM} (in a manner analogous to Cro binding O_{R3} and repressing P_{RM} during lytic growth). This prevents synthesis of new repressor until its concentration falls to a level at which it vacates O_{R3}.

It is interesting to note that the term "induction" is used to describe both the switch from lysogenic to lytic growth in λ, and the switching on of the *lac* genes in response to lactose. This common usage stems from the fact that both phenomena were studied in parallel by Jacob and Monod (see Box 16-3). It is also worth noting that, just as lactose induces a conformational change in Lac repressor to relieve repression of the *lac* genes, so too the inducing signals of λ work by causing a structural change (in this case proteolytic cleavage) in λ repressor.

Negative Autoregulation of Repressor Requires Long-Distance Interactions and a Large DNA Loop

We have discussed cooperative binding of repressor dimers to adjacent operators such as O_{R1} and O_{R2}. There is yet another level of cooperative binding seen in the prophage of a lysogen, one critical to proper negative autoregulation control. Repressor dimers at O_{R1} and O_{R2} interact with repressor dimers bound cooperatively at O_{L1} and O_{L2}. These interactions produce an octomer of repressor; each dimer within the octamer is bound to a separate operator.

To accommodate the long-distance interaction between repressors at O_R and O_L, the DNA between those operator regions—some 3.5 kb, including the *c*I gene itself—must form a loop (Figure 16-32). When the loop is formed, O_{R3} is held close to O_{L3}. This allows another two dimers of repressor to bind cooperatively to these two sites. This cooperativity means O_{R3} binds repressor at a lower concentration than

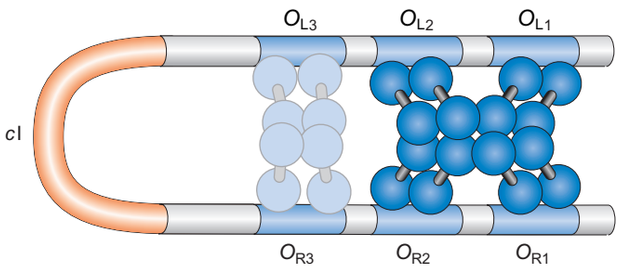


FIGURE 16-32 Interaction of Repressors at O_R and O_L. Repressors at O_R and O_L interact as shown. This interaction stabilizes binding and increases repression at O_R (and presumably at O_L).

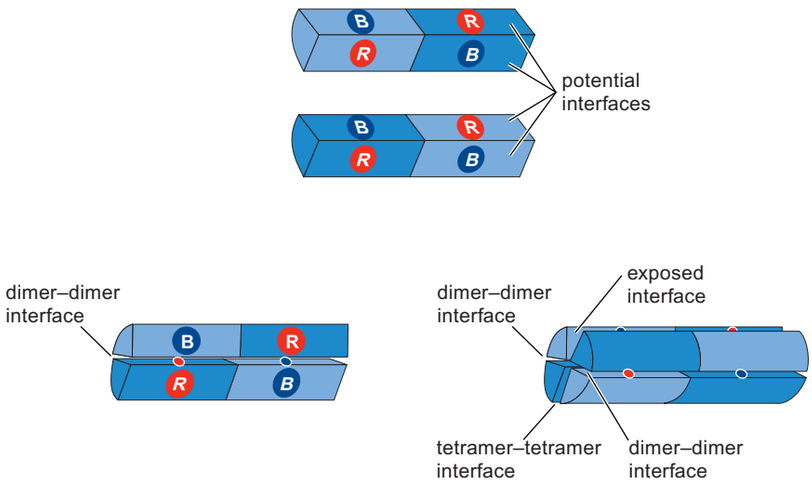


FIGURE 16-33 Interactions between the CTD of the λ Repressors. Model of the λ repressor CTD dimer, showing tetramer formation and octamer formation. Once the octamer has formed, there is no space left for a further dimer to enter the complex, and so the octamer is the highest order structure that forms. [Source: Modified, with permission, from Bell et al. 2000. *Cell* **101:**801–811, Figures 4 (parts a, b) and 5 (Parts a, b, c)].

it otherwise would—indeed, at a concentration only just a little higher than that required to bind O_{R1} and O_{R2}. Thus, repressor concentration is very tightly controlled indeed—small decreases are compensated for by increased expression of its gene and increases by switching the gene off. This explains why lysogeny can be so stable while also ensuring induction is very efficient.

The structure of the C-terminal domain of λ repressor, interpreted in the light of earlier genetic studies, reveals the basis of dimer formation. But it also shows how two dimers interact to form the tetrameric form (as occurs when repressor is bound cooperatively to O_{R1} and O_{R2}). Moreover, the structure reveals the basis for the octomer form—and shows that this is the highest order oligomer repressor can form (Figure 16-33).

Another Activator, λcII, Controls the Decision Between Lytic and Lysogenic Growth upon Infection of a New Host

We have seen how λ repressor and Cro control lysogenic and lytic growth, and the switch from one to the other upon induction. Now we turn to the early events of infection, those that determine which pathway the phage chooses in the first place. Critical to this choice are the products of two other λ genes, *c*II and *c*III. We need only expand slightly our map of the regulatory region of λ to see where *c*II and *c*III lie: *c*II is on the right of *c*I and is transcribed from P_R; *c*III, on the left of

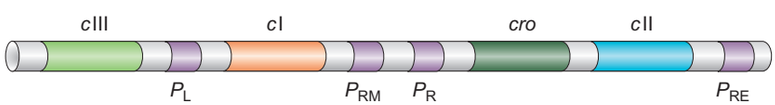


FIGURE 16-34 Genes and Promoters Involved in the Lytic/Lysogenic Choice. Not shown here is the gene *N* which lies between P_L and *c*III (see Figure 16-25).

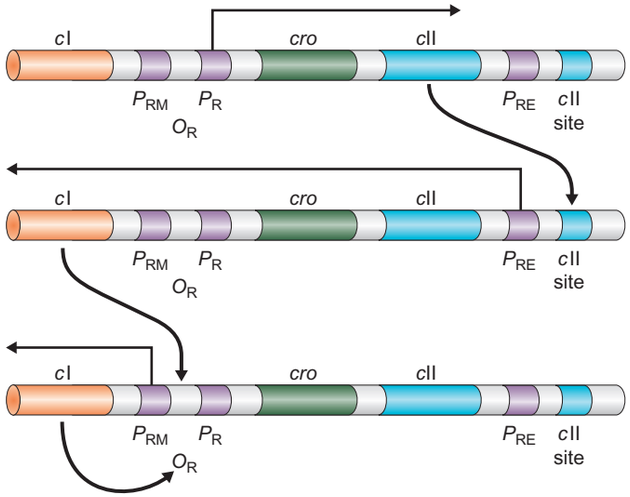


FIGURE 16-35 Establishment of Lysogeny. The *c*I gene is transcribed from P_{RE} when establishing lysogeny and from P_{RM} when maintaining that state. Repressor bound at O_{R1} and O_{R2} turns off the establishment mode of expression (which depends on transcription from P_R) and at the same time activates the maintenance mode (transcription from P_{RM}). As shown in this figure, P_R controls not only lytic genes (as indicated in the text), but also *c*II, which is required to establish lysogeny. Similarly, though not shown in the figure, P_L, which controls many lytic genes, also controls a few genes which help establish lysogeny.

*c*I, is transcribed from P_L (Figure 16-34). These and other genes critical to the lytic/lysogenic choice were isolated in clever genetic screens outlined in Box 16-5 on Genetic Approaches that Identified Genes Involved in the Lytic/Lysogenic Choice.

Like λ repressor, CII is a transcriptional activator. It binds to a site upstream of a promoter called P_{RE} (for repressor establishment) and stimulates transcription of the *c*I (repressor) gene from that promoter. Thus the repressor gene can be transcribed from two different promoters (P_{RE} and P_{RM}).

P_{RE} is a weak promoter because it has a very poor −35 sequence. CII protein binds to a site that overlaps the −35 region but is located on the opposite face of the DNA helix; by directly interacting with polymerase, CII helps polymerase bind to the promoter.

Only once sufficient repressor has been made from P_{RE} can that repressor bind to O_{R1} and O_{R2} and direct its own synthesis from P_{RM}. Thus we see that repressor synthesis is **established** by transcription from one promoter (stimulated by one activator) and then **maintained** by transcription from another (under its own control—positive autoregulation).

We can now see in summary how CII orchestrates the choice between lytic and lysogenic development. Upon infection, transcription is immediately initiated from the two constitutive promoters P_R and P_L. P_R directs synthesis of both Cro and CII. Cro expression favors lytic development: once Cro reaches a certain level it will bind O_{R3} and block P_{RM}. CII expression, on the other hand, favors lysogenic growth by directing transcription of the repressor gene (Figure 16-35). For successful lysogeny, repressor must then bind to O_{R1} and O_{R2} and activate P_{RM} before Cro can inhibit that promoter.

Growth Conditions of the *E. coli* Cell Control the Stability of CII Protein and Thus the Lytic/Lysogenic Choice

The efficiency with which CII directs transcription of the *c*I gene—and hence the rate at which repressor is made—is the critical step in deciding how λ will develop. What determines how efficiently CII works in any given infection?

When the phage infects a population of bacterial cells that are healthy and growing vigorously, it tends to propagate lytically, releasing progeny into an environment rich in fresh host cells. When conditions are poor for bacterial growth, however, the phage is more likely to form lysogens and sit tight; there will likely be few host cells in the vicinity for any progeny to infect. These different growth conditions impinge on CII as follows.

CII is a very unstable protein in *E. coli*; it is degraded by a specific protease called FtsH (HflB), encoded by the *hfl* gene (see Box 16-5). The speed with which CII can direct synthesis of repressor is thus determined by how quickly it is being degraded by FtsH. Cells lacking the *hfl* gene (and thus FtsH) almost always form lysogens upon infection by λ: in the absence of the protease, CII is stable and directs synthesis of ample repressor. FtsH activity is itself regulated by the growth conditions of the bacterial cell, though it is not understood exactly how that is achieved. Nevertheless, if growth is good, FtsH is very active, CII is destroyed efficiently, repressor is not made, and the phage tend to grow lytically. In poor growth conditions the opposite happens: low FtsH activity, slow degradation of CII, repressor accumulation, and a tendency toward lysogenic development. Levels of CII

Box 16-5 Genetic Approaches That Identified Genes Involved in the Lytic/Lysogenic Choice

Genes involved in lytic/lysogenic choice were identified by screening for λ mutants that grow efficiently either only lytically or only lysogenically. To understand how these mutants were found, we need to consider how phage are grown in the lab. Bacterial cells can be grown as a confluent, opaque lawn across an agar plate. A lytic phage, grown on that lawn, produces clear plaques, or holes. Each plaque is typically initiated by a single phage infecting a bacterial cell. The progeny phage from that infection then infect surrounding cells, and so on, killing off (lysing) the bacterial cells in the vicinity of the original infected cell and causing a clear cell-free zone in the otherwise opaque lawn of bacterial cells.

Phage λ forms plaques too, but they are turbid (or cloudy)—that is, the region within the plaque is clearer than the uninfected lawn, but only marginally so. The reason for this is that a phage that grows only lytically makes clear plaques because it kills all the cells it infects; λ, on the other hand, kills only a proportion of those it infects, while the others survive as lysogens. Lysogens are resistant to subsequent infection and so can grow within the plaque unharmed by the mass of phage particles found there. The reason for this "immunity" is quite simple: in a lysogen, the integrated phage DNA (the prophage) continues making repressor from P_{RM}. Any new λ genome entering that cell will at once be bound by repressor, giving no chance of lytic growth.

In one classic study, mutants of λ that formed clear plaques were isolated. These mutant phage are unable to form lysogens but still grow lytically. The λ clear mutations identified the three phage genes, called *c*I, *c*II, and *c*III (for clear I, II and III). In other studies, so called virulent (*vir*) mutations were isolated. These define the operator sites where λ repressor binds, and were isolated by virtue of the fact that such phage can grow on lysogens. By analogy to the *lac* system, the *c*I mutants are comparable to the Lac repressor (*lacI*) mutants, *vir* mutants are the equivalent of the *lac* operator (*lacO*) mutants (see Box 16-3). Another revealing mutation was identified in a different experiment, this one a mutation in a host gene. The mutant is called *hfl* for high frequency of lysogeny. When infected with wild type λ, this strain almost always forms lysogens, very rarely allowing the phage to grow lytically.

are also modulated by the phage protein CIII. CIII stabilizes CII by acting as an alternative (and thus competing) substrate for FtsH.

A second *c*II protein-dependent promoter, P_I, has a nucleotide sequence similar to that of P_{RE} and is located in front of the phage gene *int* (see Figure 16-25); this gene encodes the integrase enzyme that catalyzes site-specific recombination of λ DNA into the bacterial chromosome to form the prophage (see Chapter 11). A third *c*II-dependent promoter, P_{AQ}, located in the middle of gene *Q,* acts to retard lytic development and thus to promote lysogenic development. This is because the P_{AQ} RNA acts as an antisense message, binding to the Q message and promoting its degradation. Q is another regulator, one that promotes the late stages of lytic growth, as we will see in the next section.

Transcriptional Antitermination in λ development

We earlier saw examples of gene regulation that operated at stages after transcription initiation. Two more examples are found in λ development, as we now describe, starting with a type of positive transcriptional regulation called **antitermination.**

The transcripts controlled by λ N and Q proteins are initiated perfectly well in the absence of those regulators. But the transcripts terminate a few hundred to a thousand nucleotides downstream of the promoter unless RNA polymerase has been modified by the regulator; λ N and Q proteins are therefore called antiterminators.

N protein regulates early gene expression by acting at three terminators: one to the left of the *N* gene itself, one to the right of *cro,* and one between genes *P* and *Q* (Figures 16-25 and 16-36). Q protein has one known target, a terminator that is 200 nucleotides downstream from the late gene promoter, $P_{R'}$, located between genes *Q* and *S* (Figure 16-36). The late gene operon of λ, transcribed from $P_{R'}$, is remarkably large for a prokaryotic transcription unit: about 26 kb, a distance that takes about 10 minutes for RNA polymerase to traverse. Possibly in this great expanse of DNA there are other, unidentified terminators at which Q protein activity is also required.

Our understanding of how antiterminators work is incomplete. Like other regulatory proteins, N and Q only work on genes that carry particular sequences. Thus, N protein prevents termination in the early operons of λ, but not in other bacterial or phage operons.

The specific recognition sequences for antiterminators are not found in the terminators where they act, but instead occur in the operons well before the terminators. N protein requires sites named *nut* (for *N* utilization) that are 60 and 200 nucleotides downstream from P_L and P_R (Figure 16-36). But N does not bind to these sequences within DNA. Rather, it binds to RNA transcribed from DNA containing a *nut* sequence.

Thus, once RNA polymerase has passed a *nut* site, N binds to the RNA and from there is loaded on to the polymerase itself. In this state, the polymerase is resistant to the terminators found just beyond the *N* and *cro* genes. λ N works along with the products of the bacterial genes *nusA, nusB, nusE,* and *nusG.* The NusA protein is an important cellular transcription factor. NusE is the small ribosomal subunit protein S10, but its role in N protein function is unknown. No cellular function of NusB protein is known. These proteins form a complex with N at the *nut* site, but N can work in their absence if present at high concentration, suggesting that it is N itself that promotes antitermination.

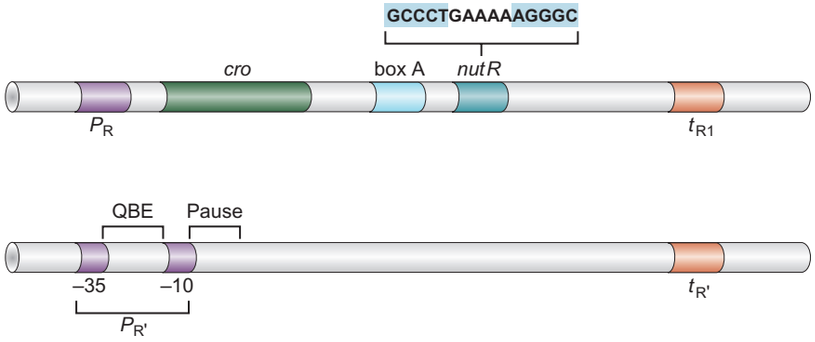


FIGURE 16-36 Recognition Sites and Sites of Action of the λ N and Q Protein Transcription Antiterminators. The upper line shows the early rightward promoter P_R that precedes the *cro* gene. Box A is a seven-nucleotide sequence associated with *nut* sites that is required for efficient N protein function. The sequence of the RNA-like strand of *nutR* is shown above. The lower line shows the promoter $P_{R'}$, the sequences essential for Q protein function, and the terminator at which Q protein acts.

Unlike N protein, the λ Q protein recognizes DNA sequences (QBE) between the −10 and −35 regions of the late gene promoter ($P_{R'}$) (see Figure 16-36). In the absence of Q, polymerase binds $P_{R'}$ and initiates transcription, only to pause after a mere 16 or 17 nucleotides; it then commences again but terminates when it reaches the terminator ($t_{R'}$) some 200 bps downstream. If Q is present, it binds to QBE once the polymerase has left the promoter, and transfers from there to the nearby paused polymerase. With Q on board, the polymerase is then able to transcribe through $t_{R'}$.

Retroregulation: An Interplay of Controls on RNA Synthesis and Stability Determines *int* Gene Expression

The CII protein activates the promoter P_I that directs expression of the *int* gene, as well as the promoter P_{RE} responsible for repressor synthesis (see Figure 16-25). The Int protein is the enzyme which integrates the phage genome into that of the host cell during formation of a lysogen (see Chapter 11). Therefore, upon infection, conditions favoring CII protein activity give rise to a burst of both repressor and integrase enzyme.

However, the *int* gene is transcribed from P_L as well as from P_I, so that we would expect integrase to be made even in the absence of cII protein. This does not happen. The reason is that *int* messenger RNA initiated at P_L is degraded by cellular nucleases, whereas messenger RNA initiated at P_I is stable and can be translated into integrase protein. This occurs because the two messages have different structures at their 3′ ends.

RNA initiated at P_I stops at a terminator about 300 nucleotides after the end of the *int* gene; it has a typical stem-and-loop structure followed by six uridine nucleotides at the end (Figure 16-37). When RNA synthesis is initiated at P_L, on the other hand, RNA polymerase is modified by the N protein antiterminator and thus goes through and beyond the terminator. This longer mRNA can form a stem that is a substrate for nucleases. Because the site responsible for this negative regulation is downstream of the gene it affects, and because degrada-

tion proceeds backward through the gene, this process is called **retroregulation.**

The biological function of retroregulation is clear. When CII activity is low and lytic development is favored, there is no need for integrase enzyme; thus, its mRNA is destroyed. But when CII activity is high and lysogeny is favored, the *int* gene is expressed to promote recombination of the repressed phage DNA into the bacterial chromosome.

There is yet a further subtlety in this regulatory device. When a prophage is induced, it needs to make integrase (together with another enzyme, called excisionase; see Chapter 11) to catalyze reformation of free phage DNA by recombination out of the bacterial DNA; and it must do this whether or not CII activity is high. Thus, under these circumstances, the phage must make stable integrase mRNA from P_L despite the antitermination activity of N protein. How is this achieved?

When the phage genome is integrated into the bacterial chromosome during the establishment of lysogeny, the phage attachment site at which recombination occurs is *between* the end of the *int* gene and those sequences encoding the extended stem from which mRNA degradation is begun (see Figure 16-25). Thus, in the integrated form, the site causing degradation is removed from the end of the *int* gene, and so *int* mRNA made from P_L is stable.

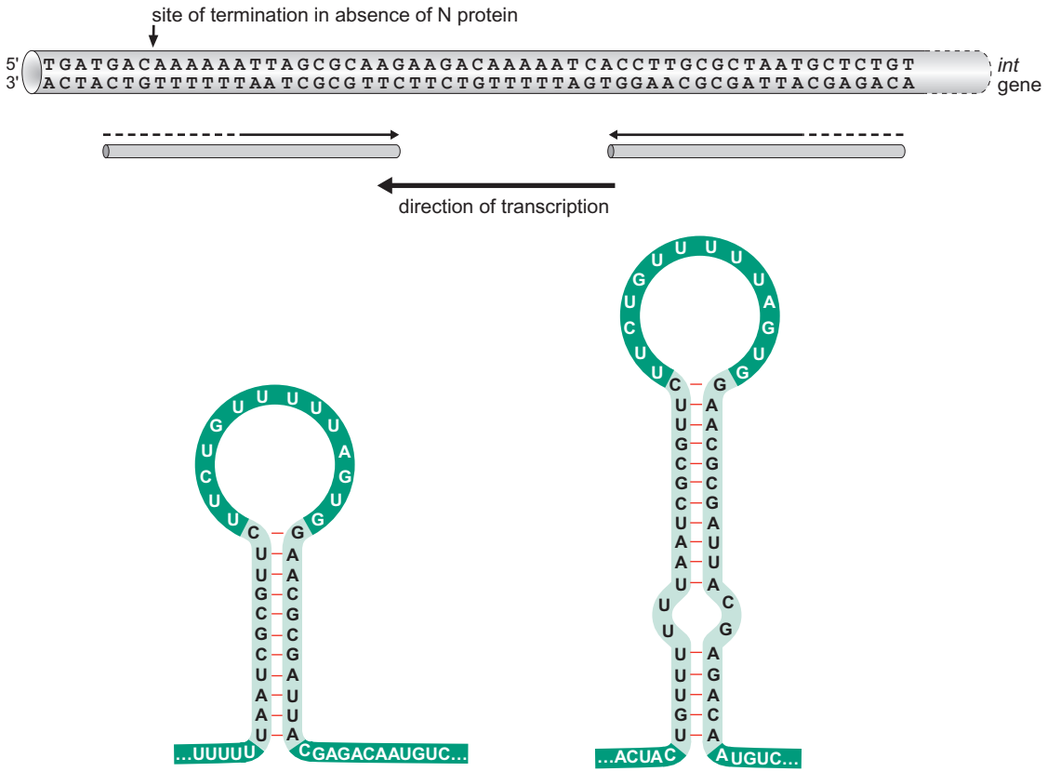


FIGURE 16-37 DNA Site and Transcribed RNA Structures Active in Retroregulation of Int Expression. Below, the small cylinders show the symmetric sequences that form hairpins in RNA. The structure on the left shows the terminator formed in RNA transcribed from $P_{I'}$ without antitermination by N protein, which is resistant to degradation by nucleases. The structure on the right shows an extended loop formed in RNA transcribed from P_L under the influence of N protein antiterminator, which is a target for cleavage by RNase III and degradation by nucleases.

SUMMARY

A typical gene is switched on and off in response to the need for its product. This regulation is predominantly at the level of transcription initiation. Thus, for example, in *E. coli,* a gene encoding the enzyme that metabolizes lactose is transcribed at high levels (and the enzyme is made) only when lactose is available in the growth medium. Furthermore, when glucose (a better energy source) is also available, the gene is not expressed even when lactose is also present.

Signals, such as the presence of a specific sugar, are communicated to genes by regulatory proteins. These are of two types: *activators,* positive regulators that switch genes on; and *repressors,* negative regulators that switch genes off. Typically these regulators are DNA-binding proteins that recognize specific sites at or near the genes they control.

Activators, in the simplest (and most common) cases, work on promoters that are inherently weak. That is, RNA polymerase binds to the promoter (and thus initiates transcription) poorly in the absence of any regulator. An activator binds to DNA with one surface and with another surface binds polymerase and recruits it to the promoter. This process is an example of cooperative binding, and is sufficient to stimulate transcription.

Repressors can inhibit transcription by binding to a site that overlaps the promoter, thereby blocking RNA polymerase binding. Repressors can work in other ways as well, for example by binding to a site beside the promoter and, by interacting with polymerase bound at the promoter, inhibiting initiation.

The *lac* genes of *E. coli* are controlled by an activator and a repressor that work in the simplest way just outlined. CAP, in the absence of glucose, binds DNA near the *lac* promoter and, by recruiting polymerase to that promoter, activates expression of those genes. The Lac repressor binds a site that overlaps the promoter and shuts off expression in the absence of lactose.

Another way in which RNA polymerase is recruited to different genes is by the use of alternative σ factors. Thus, different σ factors can replace the most prevalent one (σ^{70} in *E. coli*) and direct the enzyme to promoters of different sequences. Examples include σ^{32}, which directs transcription of genes in response to heat shock, and σ^{54}, which directs transcription of genes involved in nitrogen metabolism. Phage SPO1 uses a series of alternative σ to control the ordered expression of its genes during infection.

There are, in bacteria, examples of other kinds of transcriptional activation as well. Thus, at some promoters, RNA polymerase binds efficiently unaided, and forms a stable, but inactive, closed complex. That closed complex does not spontaneously undergo transition to the open complex and initiate transcription. At such a promoter, an activator must stimulate the transition from closed to open complex.

Activators that stimulate this kind of promoter work by allostery: they interact with the stable closed complex and induce a conformational change that causes transition to the open complex. In this chapter we saw two examples of transcriptional activators working by allostery. In one case the activator (NtrC) interacts with the RNA polymerase (bearing σ^{54}) bound in a stable closed complex at the *glnA* promoter, stimulating transition to open complex. In the other example the activator (MerR) induces a conformational change in the *merT* promoter DNA.

In all the cases we have considered, the regulators themselves are controlled allosterically by signals. That is, the shape of the regulator changes in the presence of its signal; in one state it can bind DNA, in the other it cannot. Thus, for example, the Lac repressor is controlled by the ligand allolactose (a product made from lactose). When allolactose binds repressor it induces a change in the shape of that protein; in that state the protein cannot bind DNA.

Thus we saw that two fundamental mechanisms—cooperative binding and allostery—are used in various ways in the regulation of transcription. Allostery is used to control the activities of regulators. In some cases it is also the mechanism used to activate a gene. Cooperative binding is the mechanism of activation in other cases.

Gene expression can be regulated at steps after transcription initiation. For example, regulation can be at the level of transcriptional elongation. Three cases were discussed here: attenuation at the *trp* genes and antitermination by the N and Q proteins of phage λ. The *trp* genes encode enzymes required for the synthesis of the amino acid tryptophan. These genes are only transcribed when the cell lacks tryptophan. One way that amino acid controls expression of these genes is attenuation: a transcript initiated at the *trp* promoter aborts before it transcribes the structural genes if there is tryptophan (in the form of Trp tRNAs) available in the cell. The λ proteins N and Q load on to RNA polymerases initiating transcription at certain promoters in the phage genome. Once modified in this way, the enzyme can pass through certain transcriptional terminator sites that would otherwise block expression of downstream genes. Beyond transcription, we saw an example of gene regulation that operated at the level of translation of mRNA (the case we described was that of the ribosomal protein genes).

PROBLEMS

1. What are the essential characteristics of promoters that are activated by recruitment (cooperative binding) and those activated by allostery?

2. In an activator bypass experiment, RNA polymerase is brought to a promoter in the absence of a traditional activator. Describe three ways this experiment might be

done. How does this experiment distinguish between the two classes of promoter described above?

3. When various strains of λ phage are seeded on a lawn of *E. coli,* they can form clear or turbid plaques.

a. Explain the difference between the two types of plaques. Can all bacteriophage form clear and turbid plaques?

b. For mutant λ phages that can only form clear plaques, give two different types of mutation in the phage that can explain the clear plaque phenotype.

c. How could you distinguish the two mutations using a simple plate assay?

4. *a.* Why can a wild-type λ phage not grow on a λ lysogen?

b. Will a *c*I$^-$ mutant (that is, a λ repressor mutant) of λ grow in a λ lysogen? Why or why not?

c. Mutations that allow λ to grow in a λ lysogen are called *virulent* (*vir*) mutations. To which type of *lac* operon mutation are virulent mutations similar?

d. It turns out that virulent mutations of λ are very hard to isolate because they are extremely rare, occurring in about one of every 10^{15} phages. Can you think of a reason why such mutants are so rare?

e. P22 is a virus very similar to λ. However, it gives rise to virulent mutants at much higher frequencies, about one in 10^5. The reason for this difference lies in the fact that P22 encodes a protein called *anti-repressor* which, when expressed, complexes with and inactivates the phage repressor protein. In light of this information, suggest an explanation for the higher rate of P22 virulence mutations.

5. When glucose is present, the *lac* genes are not fully expressed, even in the presence of inducer. This is called *catabolite repression.*

a. Why does it make biological sense to have the lactose operon under negative control by Lac repressor?

b. Why does it make biological sense to have the *lactose* operon controlled by catabolite repression?

• It is commonly stated that lactose induces the *lac* operon. However, allolactose, which is a product of basal β-galactosidase activity on lactose, is the actual inducer molecule.

c. Devise an experiment to prove this.

• You grow cells under each of the following conditions:

	(1)	(2)	(3)	(4)
glucose	+	+	−	−
lactose	+	−	+	−

You measure the levels of β-galactosidase (lacZ) and get the following values:

	(1)	(2)	(3)	(4)
no. of units	7000	10	18,000	10

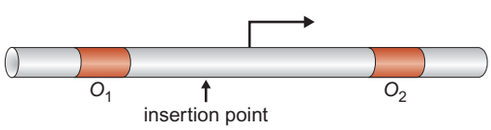
d. What does the fact that the same results are obtained under conditions (2) and (4) tell you about the role of positive control in the repressed state? Can you suggest a molecular interpretation of this result?

• CAP is necessary to turn on several sugar operons (including the arabinose, lactose, maltose, and galactose operons). Cells with mutations in CAP cannot efficiently metabolize any of these sugars. On plates that contain a sugar and tetrazolium (an indicator dye), colonies are white if that sugar is metabolized and red if it is not. This kind of plate is often used to screen for cells which cannot metabolize a particular sugar.

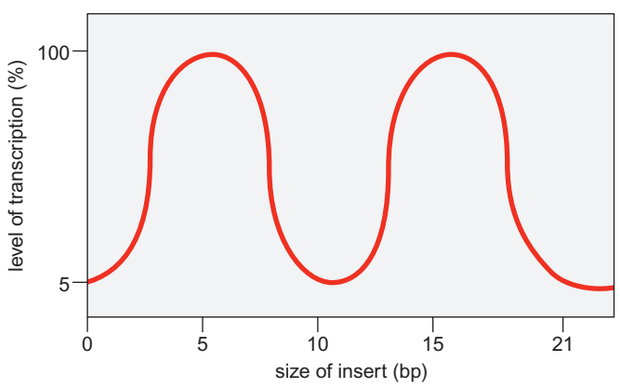
e. How could you use these plates to isolate CAP mutants?

f. You find that you obtain two classes of mutants with this screen. The first class of mutants are CAP mutants. What do you think the second class could be?

6. An operon in *E. coli* is controlled by a repressor that binds at the two operator sites (O_1 and O_2) diagrammed in the figure below. In the presence of the appropriate inducer, a transcription rate of 100 is observed, but in the absence of inducer, the transcription rate falls to 5. If either of the two sites is mutated so that the repressor cannot bind, then the transcription rate is observed to be 100. Additionally, if base pairs are inserted at the arrow, the level of transcription is found to vary with the size of the insert as shown in the graph. Briefly explain this data.



***E. coli* Operon Controlled by Repressor.**



7. NtrC and MerR are two activators that work by an allosteric mechanism rather than by recruitment. Describe each system briefly. In which is there no repressor? Why is no repressor needed in that case? In one of these systems mutant derivatives of the activator were isolated. They had the following characteristics: One mutant activator could bind DNA but could not activate transcription, and the other could not bind DNA but could activate transcription when that mutant activator was present at high concentration. In which system could these mutants arise? Explain the behavior of each.

8. Describe two examples of regulators that acted on transcription at steps after initiation. Of the two examples, one must be from a phage system and the other from a bacterial biosynthetic operon.

9. The *c*I, *c*II, and *c*III genes of phage λ are named for the fact that a mutation in any one of them leads to a phage that produces clear plaques when grown on a lawn of wild-type *E. coli* cells. Each gene encodes a protein involved in regulating the choice between lytic and lysogen growth, and when absent, the phage grows almost exclusively lytically upon infection. The *E. coli hfl* gene encodes a protease that degrades the λ CII protein. When a wild type λ phage infects a host cell mutant for the *hfl* gene, it almost always grows lysogenically.

Two λ phage can infect the same cell at the same time. During such a mixed infection, gene products that can act in *trans* do so. Under growth conditions that favor lysogeny of a wild-type phage, which of the following infections produce lysogens with reasonable efficiency?

λ*c*I/λ*c*I infecting a wild type cell

λ*c*I/λ*c*I infecting *hfl* cell

λ*c*I/λ*c*II infecting wt cell

λ*c*I/λ*c*II infecting an *hfl* cell

λ*c*II*c*III (that is, a phage mutant in both genes) infecting wt cell

λ*c*II*c*III infecting *hfl* cell

λ*c*III/λ*c*III infecting a wt cell

λ*c*III/λ*c*III infecting an *hfl* cell

Which of these cases was the hardest to decide about and why?

10. Three classes of mutants were found that affect expression of the *lac* operon. Two of these resulted in constitutive expression of the operon, and the other eliminated all expression, even in the presence of lactose.

a. Describe the gene or regulatory element in which each type of mutation is likely found.

b. Describe how the two categories of constitutive mutants can be distinguished.

c. Which of these two types of constitutive mutant do you think was more commonly isolated?

d. You identify another rare class of constitutive repressor mutants that fall within the repressor coding sequence, yet are dominant and cannot be rescued by a wild type copy of the gene acting in trans. Based on what you know about the binding of repressor to the *lac* operator, speculate as to how these mutations might alter the structure of the protein.

11. *a.* What are the two levels of regulation controlling the expression of enzymes involved in tryptophan biosynthesis?

b. Describe the role of the following elements within the leader sequence of the *trp* mRNA: the leader peptide coding sequence, region 1, region 2, region 3, region 4. What is the role of the tryptophan residues within the leader peptide sequence?

c. You create a mutant strain of bacteria that constitutively initiates *trp* mRNA production, and are examining the effect of mutations within the mRNA leader sequence on attenuation. You make a single nucleotide substitution within region four of the leader, and, as you expected, this mutation interferes with attenuation and allows a constant level of expression of the entire transcript, independent of tryptophan levels. In an attempt to isolate revertants, you mutagenize this strain, and isolate a strain in which a second mutation is present in the leader sequence. This double mutant has restored attenuation activity, but now the transcript is constantly attenuated, and can no longer be induced by low tryptophan levels. Provide a molecular explanation for this double mutant.

12. Describe three ways in which cooperativity helps DNA-binding proteins work. Why does elimination of the cooperative interactions between λ repressor molecules suffice to induce a lysogen, despite the fact that the domain bearing the DNA-binding and activating regions of that protein remaining intact in the cell.

BIBLIOGRAPHY

Books

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. 2002. *Molecular Biology of the Cell,* 4th edition. New York, N.Y.: Garland Science.

Baumberg, S., ed. 1999. *Prokaryotic Gene Expression.* Oxford, United Kingdom: Oxford University Press.

Beckwith, J., Davies, J., and Gallant, J., eds. 1983. *Gene Function in Prokaryotes.* Cold Spring Harbor, NY.: Cold Spring Harbor Laboratory.

Cold Spring Harbor Symposia on Quantitative Biology. 1998. Volume 63: Mechanisms of transcription. Cold Spring Harbor, NY.: Cold Spring Harbor Laboratory Press.

Müller-Hill, B. 1996. *The* lac *Operon.* Berlin: de Gruyter.

Ptashne, M. 1992. *A Genetic Switch: Phage* λ *and Higher Organisms,* 2nd edition. Malden, Mass.: Blackwell Science, and Cambridge, Mass.: Cell Press.

Ptashne, M. and Gann, A. 2002. *Genes & Signals.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Weaver, R. F. 2002. *Molecular Biology,* 2nd edition. New York, N.Y.: McGraw-Hill Higher Education.

Activation and Repression

Adhya, S., Geanacopoulos, M., Lewis, D. E., Roy, S., and Aki, T. 1998. Transcription regulation by repressosome and by RNA polymerase contact. *Cold Spring Harbor Symp. Quant. Biol.* **63**:1–9.

Buck, M., Gallegos, M. T., Studholme, D. J., Guo, Y., and Gralla, J. D. 2000. The bacterial enhancer-dependent σ^{54} (σ^{N}) transcription factor. *J. Bacteriol.* **182**:4129–4136.

Busby, S. and Ebright, R. H. 1999. Transcription activation by catabolite activator protein (CAP). *J. Mol. Biol.* **293**:199–213.

Hochschild, A. and Dove, S. L. 1998. Protein-protein contacts that activate and repress prokaryotic transcription. *Cell* **92**:597–600.

Jacob, F. and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**:318–356.

Lloyd, G., Landini, P., and Busby, S. 2001. Activation and repression of transcription initiation in bacteria. *Essays Biochem.* **37**:17–31.

Magasanik, B. 2000. Global regulation of gene expression. *Proc. Natl. Acad. Sci.* **97**:14044–14045.

Müller-Hill, B. 1998. Some repressors of bacterial transcription. *Curr. Opin. Microbiol.* **1**:145–151.

O'Halloran, T. V. 1993. Transition metals in control of gene expression. *Science* **261**:715–725.

Ptashne, M. and Gann, A. 1997. Transcriptional activation by recruitment. *Nature* **386**:569–577.

Rojo, F. 2001. Mechanisms of transcriptional repression. *Curr. Opin. Microbiol.* **4**:145–151.

Rombel, I. North, A., Hwang, I., Wyman, C., and Kustu, S. 1998. The bacterial enhancer-binding protein NtrC as a molecular machine. *Cold Spring Harbor Symp. Quant. Biol.* **63**:157–166.

Roy, S., Garges, S., and Adhya, S. 1998. Activation and repression of transcription by differential contact: Two sides of a coin. *J. Biol. Chem.* **273**:14059–14062.

Xu, H. and Hoover, T. R. 2001. Transcriptional regulation at a distance in bacteria. *Curr. Opin. Microbiol.* **4**:138–144.

DNA Binding, Cooperativity, and Allostery

Bell, C. E. and Lewis, M. 2001. The Lac repressor: A second generation of structural and functional studies. *Curr. Opin. Struct. Biol.* **11**:19–25.

Hochschild, A. 2002. The switch: *c*I closes the gap in autoregulation. *Curr. Biol.* **12**:R87–R89.

Luscombe, N. M., Austin, S. E., Berman, H. M., and Thornton, J. M. 2000. An overview of the structures of protein-DNA complexes. *Genome Biol.* **1:** REVIEWS001.

Monod, J. 1966. From enzymatic adaptation to allosteric transitions. *Science* **154**:475–483.

Regulation at Steps After Transcription Initiation

Bauer, C., Carey, J., Kasper, L., Lynn, S., Waechter, D., and Gardner, J. 1983. Attenuation in bacterial operons. In *Gene Function in Prokaryotes* (Beckwith, J., Davies, J., and Gallant, J., eds.), pp 65–89. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Friedman, D. I. and Court, D. L. 2001. Bacteriophage λ: Alive and well and still doing its thing. *Curr. Opin. Microbiol.* **4**:201–207.

Gottesman, M. 1999. Bacteriophage λ: The untold story. *J. Mol. Biol.* **293**:177–180.

Greenblatt, J., Mah, T. F., Legault, P., Mogridge, J., Li, J., and Kay, L. E. 1998. Structure and mechanism in transcriptional antitermination by the bacteriophage λ N protein. *Cold Spring Harbor Symp. Quant. Biol.* **63**: 327–336.

Nomura, M. 1999. Regulation of ribosome biosynthesis in *Escherichia coli* and *Saccharomyces cerevisiae*: Diversity and common principles. *J. Bacteriol.* **181**: 6857–6864.

Nomura, M., Gourse, R., and Baughman, G. 1984. Regulation of the synthesis of ribosomes and ribosomal components. *Ann. Rev. Biochem.* **53**:75–117.

Nomura, M., Yates, J., Dean, D., and Post, L. 1980. Feedback regulation of ribosomal protein gene expression in *Escherichia coli:* Structural homology of ribosomal RNA and ribosomal protein mRNA. *Proc. Nat. Acad. Sci.* **77**:7084–7088.

Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H., and Roberts, C. W. 1998. Antitermination by bacteriophage λ Q protein. *Cold Spring Harbor Symp. Quant. Biol.* **63**:319–325.

Weisberg, R. A. and Gottesman, M. E. 1999. Processive antitermination. *J. Bacteriol.* **181**:359–367.

Yanofsky, C. 2000. Transcription attenuation: Once viewed as a novel regulatory strategy. *J. Bacteriol.* **182**:1–8.