Many genes respond to combinations of signals: one signal that turns a gene on, for example, may be overridden by another signal that turns it off. And, in more complex organisms in particular, a gene might be turned on (or off) only if multiple signals converge, a matter referred to as “signal integration.”

**GENE REGULATION AND EVOLUTION**

Changes in gene regulation can contribute dramatically to morphological diversity. For example, the most striking morphological difference between the plant teosinte and maize, its domesticated form, is accounted for by the change in expression pattern of a single gene. To take another example, the disparate uses of a forelimb in two species of Crustacea—feeding in one case and locomotion in the other—is determined by a difference in the expression pattern of a single gene. And although at the moment we cannot be precise about this, it is generally believed that mammals—humans and mice, for example—contain to a large extent the same genes; it is the differences in how these genes are expressed that account for the distinctive features of the animals.

As of some 500 million years ago, representatives of most animal phyla had appeared, including chordates, arthropods, and echinoderms. At that time, these phyla underwent a dramatic period of diversification called the Cambrian explosion, and enormous phenotypic changes have also occurred since then. According to one popular view, however, changes in patterns of gene expression (rather than evolution of new genes) have had an important, perhaps even determinative, role in generating much of that diversity. The idea that changes in gene regulation have important roles in evolution was formulated more than 20 years ago by, among others, Mary-Claire King and Alan Wilson and by François Jacob. These ideas have been extended and supported by the work of many evolutionary and developmental biologists since then.

The following surmise, offered with more evocative than literal intent, summarizes these matters: a relatively small number of genes and signals have generated an astounding panoply of organisms. Thus, the regulatory machinery must be such that it readily throws up variations—new patterns of gene expression—for selection to work on. And these variants must be produced without destroying what has already been selected. In other words, the underlying mechanisms of gene regulation must be highly “evolvable.”

**GENE EXPRESSION AND ITS REGULATION**

Two steps of gene expression are essentially the same in all organisms: the typical gene is transcribed into messenger RNA and that mRNA is then translated into protein. All cells contain at least one form of RNA polymerase—the enzyme that transcribes DNA into RNA—and the machinery that translates the mRNA into protein. There are additional steps in the process of gene expression found in some, particularly more complex, organisms. For example, for many genes, the RNA transcript must be “spliced,” a process that removes unwanted sequences; it must be chemically modified at one or both of its ends; and it must be transported out of the nucleus for translation. We know of examples of regulation that affect each of these steps and additional steps besides.

But the most pervasive form of gene regulation—from bacteria to higher eukaryotes—involves the initiation of transcription by RNA polymerase. Regulation of that step, as we mentioned earlier, is the main subject of this book: we are concerned with mechanisms that ensure transcription will be initiated at one rather than another gene under a specified set of conditions. The first critical insights were reported some 40 years ago, as we now recount.

The modern study of gene regulation was initiated in the 1950s by François Jacob, Jacques Monod, Andre Lwoff, and many coworkers, at the Institute Pasteur in Paris. They recognized that bacteria, like higher organisms, regulate expression of their genes. Thus, for example, although the bacterium *E. coli* does not undergo a developmental process analogous to that of higher organisms, it nevertheless does not express all of its genes all of the time.

Jacob and colleagues studied extensively two examples. The first involves the ability of *E. coli* to grow on a wide array of different sugars including lactose; and the second concerns the lifestyle of a bacterial virus (a bacteriophage) called λ. These two apparently unrelated cases have something in common: in each, genes that previously were silent (turned off) are expressed (turned on or activated) in response to specific environmental signals.

The gene encoding the enzyme β-galactosidase is silent until its substrate—lactose—is added to the medium; the gene is then transcribed
(turned on) and the enzyme is synthesized. In the λ example, a set of about 50 phage genes (the so-called lytic genes) can be maintained in a bacterium in a dormant (unexpressed) state called lysogeny. Those genes are transcribed in response to ultraviolet (UV) irradiation.

A key insight of the Paris group was that gene regulation—the decision as to which gene to express—can be separated from the process of gene expression per se. They isolated mutants of the bacterium and the phage that were specifically deficient in the regulatory process. In so doing, they identified a class of genes—regulatory genes—whose products have as their sole function the regulation of expression of other genes. Thus, in those early experiments, they isolated mutants of E. coli that expressed the lacZ gene—which encodes β-galactosidase—whether or not lactose is present in the medium. In these mutants, lacZ expression has become “constitutive”: the gene is expressed independently of the environmental signal that ordinarily is required to switch it on. Analogous mutants of λ express the phage lytic genes constitutively.

These constitutive mutants, in both the bacterial and phage cases, were of two classes. One class defined the regulatory genes whose products (called the Lac and λ repressors, respectively) keep the target genes turned off when the inducing signal is absent. The second class defined DNA sites near the target genes, called operators, where the repressors act. In the late 1960s and early 1970s, it was shown explicitly that the Lac and λ repressors are proteins that bind specifically to their respective operators on DNA. Operators typically overlap the sequence recognized by polymerase, called the promoter. Repressor bound at the operator excludes binding of the polymerase and hence prevents transcription of the gene.

such that, in the absence of lactose, the repressor binds DNA, but in its presence does not. The site on the repressor that binds the sugar is distinct from that which binds DNA; and because the protein changes shape, the phenomenon was called allostery (other shape). The site to which (in this case) the sugar binds was called the allosteric site.

Allostery is a profoundly important mechanism by which signals can be interpreted by organisms. Some enzymes, for example, are also controlled by allostery. Because the allosteric site need not, and typically does not, resemble the active site, any signal can, in principle, be used to control the function of any appropriately designed protein.

GENE ACTIVATION

The early description of the lac and λ systems was incomplete in one important way: the lac genes, and certain λ genes, are subject to activation as well as to repression. That is, even in the absence of the repressors, these genes are not fully expressed unless they are activated. Specific DNA-binding proteins effect this “positive control” just as specific binding proteins effect repression. As with repressors, the typical activator senses a signal that determines whether or not it binds DNA and activates transcription.

For reasons that we hope will become clear, a coherent view of gene regulation requires that we first understand molecular mechanisms of gene activation.

OVERVIEW OF THE BOOK

In Chapter 1 we describe three different mechanisms found in E. coli for activating genes. We described one of these at the beginning of this Introduction, and it is disarmingly straightforward: the activator recruits RNA polymerase to a specific gene where the transcription reaction—a complicated affair—then proceeds spontaneously. “Regulated recruitment,” as we call this mechanism, readily lends itself to the use of repressors, together with activators, to control specific genes, and to the use of regulators in different combinations (combinatorial control). Bacteria provide us with two additional mechanisms for gene activation and with experimental strategies for distinguishing between all three.
Lessons from Bacteria

"He, like the rest of us, had many impressions which saved him the trouble of distinct ideas."  
GEORGE ELIOT

"Classical models tell us more than we at first can know."  
KARL POPPER

We begin with a brief description of the enzyme RNA polymerase and a summary of the three mechanisms of gene activation found in *Escherichia coli*. We then describe the mechanisms in more detail, using examples of each, and include a description of the role of repressors where appropriate. We will also see, for each case, how signals are transmitted from environment to gene.

We pay particular attention to the nature of the molecular interactions required in each case. We will encounter experimental approaches that distinguish between the mechanisms. As shown in subsequent chapters, several of these tests can be applied to analyzing gene regulation in eukaryotes as well.

**RNA POLYMERASE**

Four subunits comprise the core of RNA polymerase (see Figure 1.1). When examined in vitro, this complex transcribes DNA into RNA, but initiates at nonspecific sites on the DNA.

In the bacterial cell, the core is typically found to be associated with one other essential subunit, called Sigma (σ), and the complex is referred to as the holoenzyme. σ imposes a level of specificity; it restricts initiation of transcription to promoter sequences. There are six or seven different σ
Polymerase Activation: RNA polymerase bearing the $\sigma^{54}$ subunit binds to specific genes in a stable but inactive state. Activation of a particular gene requires that the enzyme attached to that gene literally be "activated," an effect that requires an allosteric change in that enzyme: the activator works by inducing that change. The genes we will consider are involved in nitrogen metabolism.

Promoter Activation: RNA polymerase bearing $\sigma^{70}$ (the same form as that which transcribes the lac and $\lambda$ genes) also binds a small set of genes at which it forms a stable, inactive complex. Activation of a specific gene requires that the promoter of that gene undergo a conformational change; that change in the promoter is induced by the activator. Genes in this category encode proteins that render the cell resistant to one or another poisonous metal (e.g., mercury).

As we shall see, for regulated recruitment, simple "adhesive" interactions of the activator with DNA and with polymerase suffice, whereas in the other two cases, more specialized kinds of interactions are required for gene activation.

**REGULATED RECRUITMENT: THE lac GENES**

The enzyme $\beta$-galactosidase, the product of the lacZ gene, cleaves lactose, the first step in metabolism of that sugar. The gene is transcribed if, and only if, lactose is present in the medium. But that physiological signal is almost entirely overridden by the simultaneous presence of glucose, a more efficient energy source than lactose. Only after having exhausted the supply of glucose does the bacterium fully turn on expression of lacZ.

The effects of these two signals (lactose and glucose) are mediated by two DNA-binding regulatory proteins, each of which senses one of the sugars. Lac repressor binds the operator only in the absence of lactose, and CAP, an activator, binds DNA only in the absence of glucose.

The arrangement of regulator binding sites in front of the lacZ gene is shown in Figure 1.2. The states of the lac genes as found under different environmental conditions are shown in Figure 1.3 and are as follows.

1. In the presence of both sugars, neither regulatory protein binds DNA, and RNA polymerase transcribes the lac genes at a low level. That
level—called the basal level—is determined by the frequency with which the enzyme spontaneously binds the promoter and initiates transcription.

2. In the presence of lactose and the absence of glucose, CAP is bound to its DNA site, but repressor is not. Transcription of the genes is activated: that is, they are transcribed at a level some 40-fold higher than the basal level. CAP has this effect by recruiting RNA polymerase to the promoter. Polymerase binding and transcription initiation are the same as in the basal case—they simply happen more frequently.

3. In the absence of lactose, repressor is bound to the operator, polymerase is excluded from the promoter, and transcription is essentially abolished. This is true whether or not glucose is present, and therefore whether or not CAP is active.

In sum, we have a constitutively active enzyme (RNA polymerase) that, alone, works with a certain frequency. The activator increases this frequency by recruiting the enzyme to the gene, and the repressor decreases the frequency by excluding the enzyme.

We now turn to some of the molecular details of these reactions and describe experiments that support the description we have given.

**FIGURE 1.3.** Three states of the lac genes. When bound to the operator, repressor excludes polymerase whether or not active CAP is present. Polymerase is shown in a more simplified form than in Figure 1.1: the carboxyl domains of the α subunits, which project from the left end of the polymerase as drawn here, are not shown. CAP actually binds DNA as a dimer, and Lac repressor as a tetramer.

**Protein-DNA Interactions**

Both CAP and Lac repressor dock with their DNA sites using a similar structural motif, the so-called helix-turn-helix (HTH). The schematic representation of Figure 1.4 shows a protein dimer bound to DNA. Each HTH bears one α-helix (the “recognition helix”) that inserts into the major groove of DNA. The side chains of amino acids exposed along the
FIGURE 1.4. DNA binding by a dimeric protein bearing a helix-turn-helix domain. The dotted circles represent two identical subunits of a DNA-binding protein complexed with an operator. The HTH motif on each monomer is indicated, with the "recognition helix" labeled R. A single recognition helix can contact functional groups displayed on the edges of approximately 6–8 bp. A single simultaneous use of two HTH motifs renders the binding far more specific than was a single HTH to be used. (Modified, with permission, from Ptashne 1992.)

recognition helix make sequence-specific contacts with edges of base pairs. A second helix lies across the DNA; it helps position the recognition helix and strengthens the binding. Differences in the residues along the outside of recognition helices largely account for differences in the DNA-binding specificities of regulators.

The HTH motif is the predominant DNA-recognition module found among E. coli transcriptional regulatory proteins. A somewhat modified form is found in eukaryotes in so-called homeodomain proteins.

Detecting Physiological Signals

As we noted in the Introduction, Lac repressor undergoes a conformational (allosteric) change upon binding inducer (alloolactose). This change greatly decreases DNA binding of the repressor. In contrast, the allosteric change undergone by CAP upon binding a small molecule (cyclic AMP or cAMP) increases its ability to bind DNA. Glucose exerts its effect by (somehow) decreasing synthesis of cyclic AMP. The name CAP stands for catabolite activator protein; the same protein is often called CRP, for cyclic AMP receptor protein.

Promoter Recognition and Transcription by RNA Polymerase

The lac promoter contains two sequence elements that are recognized by the σ70 subunit of polymerase. The sequence in the −35 region (counting back from the transcription start site) is recognized by an HTH domain; the other sequence, i.e. the region around −10, is recognized by a different part of σ.

All promoters recognized by a σ70-containing holoenzyme bear a sequence related to the canonical −10 region shown in Figure 1.5. Some, like the lac promoter, also bear sequences related to the canonical −35 region. Some, unlike lac, also bear a so-called UP-element, positioned upstream of the −35 region, which is recognized by the carboxyl domain of the σ-subunit.

All three elements are found at some promoters, and where both upstream elements are lacking, the −10 region is extended by one or two characteristic base pairs. In general, the more elements present at a pro-
moter, and the more closely they resemble the canonical sequences, the more efficiently it works. Some promoters, as we will see, work at very high levels without the aid of an activator.

Polymerase initially binds to promoter DNA in its ordinary double-stranded (or "closed") state. That complex then undergoes a structural transition to the "open" form in which approximately 14 base pairs (bp) around the transcription start site are melted ("opened") to expose the template strand. Whereas formation of the closed complex is readily reversible, formation of the open complex is generally irreversible and, in the presence of the appropriate RNA precursors (nucleoside triphosphates), leads to transcription.

As we have discussed, despite all these complexities associated with transcriptional initiation, CAP activates transcription simply by recruiting the polymerase to the promoter. We turn now to the molecular interactions required for that recruitment.

**Switching the Genes On: Activation by CAP**

Polymerase recruitment, as shown in Figure 1.3, requires formation of a tripartite complex comprising CAP, polymerase, and DNA. The formation of that complex is an example of cooperative binding of proteins to DNA.

The term cooperative binding can have connotations other than those we intend here (see Appendix 1, More on Cooperativity), but the form of cooperativity we now describe has a crucial role in many aspects of gene regulation in many organisms.

**Cooperative Binding of Proteins to DNA**

In the example shown (Figure 1.6a), neither site A nor site B is efficiently occupied when protein A or B is present alone, i.e., were a series of snapshots taken of site A (in the presence of protein A alone), protein A might be found bound in, say, 1% of those pictures. This frequency would be increased were the concentration of protein A to be increased.

When both proteins are present, however, both sites are much more likely to be occupied even at the lower protein concentrations. This "cooperative" effect depends on the simultaneous interaction of each protein with its DNA site and with the other protein. We say the proteins cooper-
ate, in that they help each other bind to DNA. The following are four important aspects of this reaction.

- The effect requires neither the utilization of energy (e.g., hydrolysis of ATP) nor conformational change in either protein: the proteins need only “touch” one another. To understand this, consider the frequency with which each protein falls off the DNA when the proteins are bound cooperatively. One protein might dissociate from the DNA, but its continued interaction with the other, DNA-bound, protein would hold it nearby, ensuring that it readily rebinds to its DNA site. For the protein to completely dissociate, it would have to let go of the DNA and the other protein simultaneously, or both proteins would have to simultaneously let go of the DNA. It is sometimes said that one protein helps another bind by increasing the latter’s local concentration in the vicinity of its site.\(^4\)

- The interaction between the proteins is weak in energetic terms (1 or 2 kcal). Interactions of this magnitude require very few specific contacts, but they can have important physiological consequences. A kilocalorie of interaction energy between proteins binding to separate DNA sites roughly corresponds, in terms of site occupancy, to increasing the concentration of either protein alone approximately 10-fold. And a change of 10–100-fold in binding can determine the difference between a site being essentially vacant or essentially fully occupied.

- For cooperative binding to be used as a control mechanism, the concentrations of the interacting partners must be held below a specified level—that at which they bind unaided—but high enough that they bind DNA in the presence of their respective partner. Precisely what these relevant concentrations are depends on the affinities of the proteins for their DNA sites, the strengths of the interactions between them, and so on.

- If the binding sites for A and B are not adjacent, cooperative binding would require that the intervening DNA “loop out” to accommodate the reaction as shown in Figure 1.6b.\(^3\)

For further discussion of these matters, see Appendix 1, More on Cooperativity.

---

**Cooperative Binding and Gene Activation by CAP**

To apply the reactions depicted in Figure 1.6 to the case at hand (i.e., the stimulatory effect of CAP on transcription of the lac genes), we replace protein A with RNA polymerase and protein B with CAP (Figure 1.7). CAP recruits polymerase to the promoter by binding cooperatively with it. CAP holds polymerase there until the complex isomerizes to the open form. At that stage, the reaction becomes essentially irreversible, and initiation of transcription proceeds.

The physiological consequences of the CAP-polymerase interaction are dramatic—a 40-fold increase in expression—but the interaction is small in energetic terms. Indeed, it probably involves no more than a kilocaloric or so of interaction energy and has not been detected in the absence of DNA. Here are some of the key experimental findings that support these conclusions concerning the mechanism of action of CAP at the lac genes.

**In vivo footprinting.** A variety of techniques, lumped under the term “footprinting,” can reveal whether DNA molecules bearing a binding site
for a protein actually bind that protein at any given moment. In particular, footprinting can be used to assay polymerase occupancy of the lac promoter in a population of bacterial cells. It is found that, in the absence of functional CAP and Lac repressor, RNA polymerase can be detected at the lac promoter in only a very few cells, consistent with the low level of expression seen under those circumstances. But in the presence of functional CAP (and in the absence of repressor), virtually all of the cells bear polymerase at their lac promoters. Thus, CAP “recruits” polymerase to the promoter.

**Positive control mutants.** CAP contains two surfaces required for activation of the lac genes: one is the DNA-binding surface, and the other is its “activating region,” which contacts RNA polymerase. The functions of these two surfaces can be distinguished by mutation. Mutants of CAP defective in their polymerase-binding surface, but which retain the ability to bind DNA, cannot activate transcription. These are called pc (for “positive control”) mutants. Thus, activation is not solely a consequence of DNA binding by the activator (as it would be were the effect mediated through changes in local DNA structure). Rather, the pc mutants suggest that a protein-protein interaction between the DNA-bound CAP and polymerase is required for activation.

**Chemical cross-linking.** Chemical cross-linking experiments reveal that the activating region of CAP, as defined by the pc mutants, interacts with the carboxyl domain of the α-subunit of polymerase. The cross-linking is observed when both CAP and polymerase are bound to DNA.

**Polymerase mutants.** Deletion of the carboxyl domain of the α-subunit of polymerase eliminates activation by CAP as assayed in vitro. Point mutations in the αCTD can also abolish activation by CAP.

**DNA-binding mutants.** Mutants of CAP that cannot bind DNA do not activate transcription. This is the result expected if CAP must recruit polymerase to DNA.

The idea that CAP recruits polymerase to the promoter—but has no additional effect—makes a further striking prediction: Alternative ways of bringing polymerase to the promoter should activate the gene. The following “activator bypass” experiments show that this is the case. Each of these, with the exception of the last, has been performed both in vivo and in vitro.

---

**Activator bypass 1: Activation through heterologous protein:protein interactions.** The CAP-polymerase interaction can be replaced by another protein-protein interaction. The experiment is performed as follows (Figure 1.8): one of two interacting proteins (Y) is attached to a DNA-binding domain, and the other interacting protein (X) is attached to RNA polymerase. Binding of the first hybrid protein to a site near the lacZ gene activates transcription in the presence of the modified polymerase. The key point is that the “heterologous” interaction (between X and Y) is not normally involved in transcriptional activation but nevertheless substitutes very well for the ordinary CAP-polymerase interaction in this experiment. Some examples of X and Y are given in the legend to Figure 1.8.

**Activator bypass 2: Direct tethering of polymerase.** The protein-protein interaction between CAP and polymerase can be replaced by a protein-DNA interaction. In these experiments, a CAP DNA-binding domain is
Repression by Lac Repressor

As we have already noted, the lac operator overlaps the lac promoter, and the repressor bound there blocks binding of polymerase to the promoter. In principle, any other protein bound equally tightly at this site would have a similar effect.9

Our description has been simplified by omitting the fact that there are actually two Lac repressor-binding sites in addition to the primary one we have described. These can contribute to repression, as we now explain.

The two additional lac operator sites are positioned about 90 bp upstream and 400 bp downstream from the primary operator. Deletion of either site decreases repression some 2—3-fold, and deletion of both reduces repression about 50-fold (i.e., from a factor of ~10,000 to a factor of ~200). A single Lac repressor—a stable tetramer—can simultaneously contact the primary operator and one or the other of the secondary operators, with the DNA looping out to accommodate the binding. Simultaneous binding to multiple operators strengthens the binding, and perhaps the looped structure itself contributes to repression (see Appendix 1, More on Cooperativity). These secondary operators, in the absence of the primary site, bind the repressor weakly and mediate virtually no repression.

Interim Summary and Extension

According to our formulation of the action of CAP, the term “gene activation” might be misleading in the following sense: the term “activation” might suggest to some that either the gene or the polymerase is being switched from an inactive to an active state—literally “activated.” But CAP does neither of these. Rather, it simply recruits the polymerase to the promoter—it binds cooperatively with polymerase at the gene—and transcription then proceeds. In the absence of any regulator, polymerase occasionally spontaneously binds to the lac promoter and initiates transcription. This “basal” level of transcription results from the same enzymatic events and involves the same enzymatic machinery, as does activated transcription; the only difference is that the former occurs less frequently (~40-fold) than the latter, and so fewer transcripts are made.

Regulated recruitment is thus a quantitative affair: CAP increases the probability of polymerase-promoter interaction and, to that extent, “acti-
vates" the gene; Lac repressor decreases the likelihood of that interaction and, to that extent, "represses" the gene.

In general, cooperative binding of proteins to DNA, while stabilizing the protein-DNA interaction, is reversible. If one of the binding proteins is RNA polymerase, however, the reaction becomes essentially irreversible when the polymerase-promoter complex undergoes the transition to the open form.

Each of the protein-protein and protein-DNA interactions encountered in this example involves a patch on the surface of the relevant molecule that we might characterize as "glue-like." This term accurately conveys the image of simple binding (adhesive) interactions. But the interactions are often highly specific and they span a broad range of affinities. Thus, Lac repressor and CAP bind to their respective sites tightly and with high specificities, whereas the interaction of CAP's activating region with polymerase is orders of magnitude weaker.

The signals that govern transcription of the lac genes (glucose and lactose) are first detected by the changes (allosteric) they induce in the shapes of the regulatory proteins, promoting DNA binding for CAP and preventing it for Lac repressor. The meaning (specificity) of those signals in terms of gene regulation—which genes are transcribed in response to them—is determined by the DNA-binding "address" of those proteins, i.e., where the regulatory proteins bind. Thus, the Lac repressor could be used to bring any gene under the control of lactose merely by inserting its binding site in the promoter of that gene. In nature, it is used at the lac genes because these are the genes usefully regulated by lactose.

CAP, in contrast, works on some 200 genes in addition to the lac genes. For example, CAP works in conjunction with the Gal repressor to control transcription of the gal genes, the products of which metabolize galactose. The ability of CAP to work with disparate regulators (in these cases two different repressors) is an example of combinatorial control.

MORE REGULATED RECRUITMENT: THE BACTERIOPHAGE λ

We noted in the Introduction that the bacterial virus λ can establish dormant residency in an E. coli cell. In such a cell, the phage chromosome is

integrated into—and passively replicated along with—the host chromosome. One phage gene, called cl, is expressed: the product of that gene, the bacteriophage λ repressor, keeps other phage genes switched off. Upon exposure to UV light, the repressor function is abolished, with two consequences: the previously inert phage genes (called lytic genes) are switched on, and the repressor gene itself is switched off. Some 45 minutes later, the cell lyses to release a crop of about 100 new phage particles (Figure 1.10).

A bacterium carrying a dormant phage chromosome is called a λ lysogen; the switch to lytic growth is called induction.
The Switch

To understand how the switch works, we need consider two regulatory genes (Cl and Cro) and the regulatory region called OR (right operator), which are shown in Figure 1.11. In a lysogen, Cl is on and Cro is off, and vice versa when lytic growth ensues. Figure 1.12 shows this regulatory region in more detail. The operator comprises three binding sites—OR1, OR2, and OR3—that overlap two opposing promoters. One of these, Prm, directs transcription of lytic genes and the other, Prp, directs transcription of the Cl gene.

Figure 1.13 shows that in a lysogen, the Cl repressor (the product of the Cl gene as we have said), at OR, is bound mainly at the two adjacent sites OR1 and OR2. At these positions, it performs two functions: it represses rightward transcription from the promoter Pr, thereby turning off expression of Cro and other lytic genes; simultaneously (and despite its name) it activates transcription of its own gene from the promoter Prm.

Upon induction, repressor vacates the operator and transcription from Pr—an inherently much stronger promoter than Prm—commences spontaneously. The first newly made protein is Cro: This protein binds first to OR3, apparently helping to abolish repressor synthesis.

Establishing Lysogeny

According to the description given thus far, the Cl repressor is required to activate transcription of its own gene. This raises the question of how expression of that gene is turned on when the virus first infects a bacterium to establish lysogeny. Figure 1.14 shows the answer: the repressor gene is initially transcribed from a promoter called Pr (promoter for repressor establishment).

Transcription from Pr is activated by the product of another phage gene, cII. The newly made repressor (CI) activates transcription of its own gene from Prm (promoter for repressor maintenance) as it turns off transcription of other phage genes, including that of cII. The lysogenic system thus established is then self-perpetuating in the absence of an inducing signal.
Bacteriophage λ does not always lysogenize when it infects—sometimes it grows lytically in newly infected bacteria. The choice between establishing lysogeny and growing lytically is, like the process of induction, influenced by extracellular signals (see below, Detecting Physiological Signals).

**FIGURE 1.13.** The action of λ repressor and Cro. O₅ and O₆ overlap P₈ and repressor bound to those sites turns off cro and the other lytic genes. O₂ is positioned such that repressor bound there contacts polymerase at P₉, and thereby activates the cI (repressor) gene. O₂ lies within P₉, and Cro bound there abolishes transcription of cI. (Modified, with permission, from Ptashne 1992.)

**FIGURE 1.14.** Establishment of lysogeny. The same gene, cI, is transcribed from two different promoters: from P₉ to establish lysogeny and from P₁₅ to maintain that state. Repressor bound at O₅ and O₆ turns off the establishment mode of expression (which depends on transcription from P₉) as it activates the maintenance mode (transcription from P₁₅). As implied by this figure, P₉ controls not only lytic genes (as indicated in the text), but also cII, which is required to establish lysogeny. Similarly, P₁₅, which controls many lytic genes, also controls a few genes which help establish lysogeny.

**Analogy with lac**

The essential features of gene regulation as described for the lac case are also represented in the λ switch.
Promoters

Of the three promoters in the \( \lambda \) switch (\( P_{R}^l \), \( P_{R}^m \), and \( P_{R}^e \)), two, \( P_{R}^m \) and \( P_{R}^e \), resemble the wild-type lac promoter: they are weak and require activators to work efficiently. Promoter \( P_{R}^l \) in contrast, functions at a high level in the absence of any activator.

Protein-DNA Interactions

\( \lambda \) repressor, Cro, and CI all bind to DNA using the familiar HTH motif. In each case, residues along the recognition helix, as well as certain other residues, direct the protein to its specific site or sites on DNA.

Repression

Repression is effected by a mechanism essentially the same as that encountered in the lac case. Thus, in a lysogen, \( \lambda \) repressor excludes polymerase from \( P_{R}^l \) and upon induction—as lytic growth begins—Cro excludes polymerase from \( P_{R}^m \).

Activation

Again, as in the lac case, a protein-protein interaction mediates gene activation. Thus, repressor bound at \( O_{R}^2 \) contacts polymerase at \( P_{R}^m \) to stimulate repressor synthesis in a lysogen. The same rule applies for CI: this protein, bound to a site adjacent to \( P_{R}^e \), contacts polymerase at that promoter to stimulate repressor synthesis.

Experiments similar to those that we described for lac show that polymerase need only be recruited to \( P_{R}^m \) to achieve high levels of transcription. For example, activator bypass experiments work at \( P_{R}^m \) as they do at the lac promoter. The following are a few additional experiments characterizing activation in this case.

- **Polymerase target mutants.** pc mutants of \( \lambda \) repressor, like pc mutants of CAP, bind DNA but do not activate transcription. Starting with such a pc mutant, a mutant of RNA polymerase was found that restored activation. The mutation results in an amino acid change in \( \sigma \) and is believed to identify the surface contacted by repressor's activating region.

- **Creating a new activator.** Cro works only as a repressor: even when positioned adjacent to a promoter, it fails to activate transcription because it lacks an activating region. Residues on one surface of Cro were changed so as to resemble the activating region of the \( \lambda \) repressor. When bound near \( P_{R}^m \), the modified Cro activated transcription from that promoter. Thus, an activating region can work when presented in different protein contexts, provided the DNA-binding address is appropriate.

- **Activating region variants.** Many variants of the \( \lambda \) repressor's activating region have been generated artificially. Among those that retain function, some work weakly and others work up to about fivefold better than the wild type. The ease with which these variants are found reinforces the notion that an activating region need only provide a simple binding surface, rather than performing some more specialized function.

- **CAP activation of \( P_{R}^m \).** On a DNA template bearing \( P_{R}^m \) and a suitably positioned CAP-binding site (see Figure 1.2), CAP activates \( P_{R}^m \). Thus, the interaction that works at the lac promoter also works at \( P_{R}^m \).

Detecting Physiological Signals

In the lac case, the relevant physiological signals are communicated by the binding of two small molecules (cAMP and allolactose) to the regulatory proteins CAP and Lac repressor, respectively. In the \( \lambda \) case, there are two stages at which physiological signals impinge. In both cases, the signals trigger proteolysis of regulatory proteins, \( \lambda \) repressor in one case and CI in the other.

Upon infection, the decision of whether or not to establish lysogeny is determined by the activity of a host-encoded protease that attacks CI. When the cells are growing vigorously, the protease is particularly active. CI is destroyed and repressor synthesis does not initiate, and so the infecting phage usually chooses lytic growth. If the cells are growing poorly, the protease is relatively inactive and so, because CII accumulates to a high level, the infecting phages readily establish repressor synthesis and usually lysogenize the host.

Inducing agents, such as UV light, damage DNA. The damaged DNA
binds to and activates the protein RecA. Activated RecA interacts with and stimulates cleavage of repressor, resulting in lytic growth.11

Making an Efficient Switch

The λ switch is highly efficient. In the absence of an inducing signal, spontaneous induction is rare (less than once in a million cell divisions). But, when exposed to an appropriate signal, virtually every cell in the lysogenic population produces phage. Various features contribute to this efficiency; none requires interactions fundamentally different from those we have already encountered, as we now show.

Cooperative Binding of λ Repressor to DNA

λ repressor binds cooperatively to the operator sites O_{R1} and O_{R2} as shown in Figure 1.15. Monomers first interact to form dimers, and then two dimers interact to bind to the operator sites as shown.12

FIGURE 1.15. Cooperative binding of λ repressor to DNA. The λ repressor amino-domain (N) is separated from the carboxyl domain (C) by a linker of 40 amino acids. Repressor binding to these sites would be further helped by the interaction between repressor at O_{R2} and polymerase at P_{RM}.

FIGURE 1.16. Interaction of repressors at O_{R} and O_{L}. Interaction between repressor at O_{R} and O_{L} stabilizes binding and increases repression at O_{L} (and presumably at O_{R}). The crystallographic structure of the carboxyl domains of repressor suggests that it could form an octamer, and such octamers have been observed at high repressor concentrations in vitro.

The repressor tetramer bound at O_{R} as in Figure 1.15 undergoes a further interaction: an octamer is formed by the interaction with another pair of dimers bound to the second major λ operator O_{L} (left operator). This operator is positioned about 3 kb away from O_{R} and controls the promotor P_{L} which, in turn, directs synthesis of another group of λ genes. The interaction of repressors at O_{R} and O_{L} is accompanied by formation of a large DNA loop as observed in vitro and shown in Figure 1.16.

This long-range interaction between repressors at O_{L} and O_{R} has two effects: it increases repression at P_{L} and P_{R} a fewfold and, as we describe below, increases the negative autogenous (self) regulation of repressor.
A consequence of these interactions is that the curve describing the binding of repressor to the two operator sites, as a function of the repressor concentration, is highly inflected, or “sigmoid.” This means that, around a certain value, small changes in repressor concentration can have large effects on site occupancy.\(^\text{13}\)

Cooperativity has a crucial role in the λ switch, as revealed by the mechanism of induction: the inducing agent simply eliminates cooperative binding of repressor monomers to DNA. To see this, we must describe the positions of the various functional patches on the repressor’s surface, and the fate of repressor upon induction.

As shown in Figure 1.17, both the DNA-binding surface and the activating region are carried on the amino domain of the repressor. In contrast, the cooperativity functions of repressor—those surfaces that mediate formation of dimers and interaction between dimers—lie on the carboxyl domain. Cleavage of λ repressor upon induction, mediated by RecA, results in separation of the amino from the carboxyl domain.

The separated amino domain is capable of binding DNA and activating transcription, but as it cannot bind cooperatively, its concentration in the cell is too low to maintain lysogeny. When artificially expressed at high concentrations, the separated amino domains bind DNA and activate transcription.

Elimination of cooperativity between λ repressor monomers has the same effect as decreasing the concentration of repressor monomers some 100-fold. Because only a 5–10-fold decrease in concentration is required for repressor to completely vacate the operator, induction proceeds efficiently.

**Autogenous Control by Repressor**

To ensure that binding of repressor depends on cooperativity—and the distinct “all-or-none” nature of the switch is maintained—the concentration of repressor in the lysogen must be held at a suitably low level. The design of the switch ensures that repressor concentration never exceeds that level by the following form of autogenous negative control. In addition to sites O\(_{R1}\) and O\(_{R2}\), repressor can also bind to the third site, O\(_{R3}\). But O\(_{R3}\) is an intrinsically weak binding site for repressor and, moreover, repressors bound at O\(_{R1}\) and O\(_{R2}\) do not bind cooperatively with repressor at O\(_{R3}\).\(^\text{14}\)

Site O\(_{R3}\) is thus filled only if the repressor dimer concentration increases to a level higher than that required to fill O\(_{R1}\) and O\(_{R2}\). When bound to site O\(_{R3}\), the repressor shuts off synthesis of further repressor by excluding polymerase from P\(_{R}\). This prevents the synthesis of new repressor until the existing repressor concentration falls to the appropriate level (a consequence of dilution through cell division).

In the absence of this form of autogenous negative control (e.g., if O\(_{R3}\) is mutated so that repressor cannot bind there), the level of repressor in the lysogen increases some threefold, and induction is severely impaired.

Repressor binding to O\(_{R3}\) is improved by interaction of repressors bound at O\(_{R}\) with those bound at O\(_{L}\). A likely explanation for this effect is found in Figure 1.16. There we see a loop formed by an octamer of repressor bound simultaneously at O\(_{R1}\), O\(_{R2}\), O\(_{L1}\), and O\(_{L2}\). Under these circumstances, O\(_{L3}\) and O\(_{R3}\) are held in close proximity. This configuration evidently helps another pair of repressor dimers bind cooperatively to O\(_{L3}\) and O\(_{R3}\) (as indicated by the dotted repressors in the figure).\(^\text{15}\)
Induction is further facilitated by the loss of autogenous positive control by repressor. As repressor occupancy of the operator site $O_r$ falls, the rate of synthesis of new repressor also falls, because repressor no longer stimulates transcription of its own gene. The drop in repressor concentration is thus more dramatic than it would otherwise be.

An important form of gene regulation that does not involve control of transcriptional initiation is also found in $\lambda$ and is briefly described in the Antitermination panel.

### Antitermination

In the text we stress the action of regulators that activate or repress the initiation of transcription. Here we note that a subsequent step—elongation of the transcript—can be regulated as well. Thus the polymerase, having initiated transcription, pauses when it encounters certain DNA sequences, and terminates (and falls off the DNA) at others. We know of two kinds of “antiterminators” encoded by bacteriophage $\lambda$ that cause the polymerase to read through such signals.

#### $\lambda$ Q

The product of the phage Q gene is a DNA-binding protein that recognizes a DNA sequence (called $qut$: Q utilization) positioned between the $-10$ and $-35$ regions of a promoter called $P_r$\textsuperscript{K}. This promoter is responsible for directing transcription of genes expressed late in the life cycle of the virus. In the absence of Q, polymerase initiates transcription at the strong promoter $P_r$, pauses 17 bp downstream, and then terminates (and falls off the DNA) at a terminator positioned some 200 bp downstream. Q, made midway in the lytic cycle, binds to $qut$ and interacts with the paused polymerase. Polymerase now reads through the pause site and—evidently because Q has joined the polymerase—reads through the downstream terminator. In fact, the transcript extends more than 26,000 bp, ignoring numerous other termination sites on route.

#### $\lambda$ N

The product of the phage N gene is an RNA-binding protein that recognizes the nut (N utilization) site present in two phage mRNAs. These mRNAs emanate from $P_r$ and $P_r$; in the absence of N protein, the former terminates after transcribing $cro$, and the latter terminates after transcribing N. N binds nut and polymerase and evidently joins the enzyme to make it impervious to the terminators. N is helped in this regard by four other bacterial proteins, including one called NusA. nut (as a DNA sequence) was inserted between other promoters and terminators and, when provided, N caused antitermination in those cases as well.

### Interim Summary and Extensions

In the $\lambda$ switch, we encounter several features not found in the lac case. These include a regulator ($\lambda$ repressor) that works as an activator and as a repressor; cooperative binding of that regulator to adjacent operator sites; binding of different regulators (repressor and Cro) to the same operator sites but with different orders of affinities and with different effects; autogenous positive ($\lambda$ repressor) and negative ($\lambda$ repressor and Cro) control; and the use of two promoters ($P_{RE}$ and $P_{RM}$), one transiently, to direct transcription of the same gene (the repressor gene). However, none of these added features requires molecular interactions other than those of the kind we encountered at lac.

For example, there is nothing particularly surprising about the ability of the $\lambda$ repressor to work both as an activator and as a repressor. The repressor-binding sites are arranged so that, in a lysogen, the repressor is positioned to exclude polymerase from the promoter of lytic genes and to touch, with its activating region, polymerase at $P_{RM}$. CAP, which at lac works only as an activator, also works as a repressor at certain other promoters, depending on the location of its binding site on DNA.

The cooperativity between $\lambda$ repressors, which helps render the switch so efficient, has been imposed in a simple way. Instead of having one site to which the activator—$\lambda$ repressor—binds, there are two sites to which it binds cooperatively. That cooperative binding in turn depends on adhesive interactions between the repressor molecules. Induction separates the cooperativity domain from the DNA-binding domain, and that loss of cooperativity is sufficient to flip the switch.

It is apparent that the same kinds of interactions—binding interactions that we have characterized as adhesive—assorted and reiterated in different ways, can be used to solve quite distinct gene regulatory problems.

### Activation: A Closer Look

The two activators that we have described—CAP working at the lac genes, and $\lambda$ repressor working at the $\lambda$ ci gene—recruit polymerase by contacting different parts of it. Thus, CAP contacts the carboxyl domain of $\sigma$ and $\lambda$ repressor contacts $\sigma$. There is nothing inherent in the promoters that requires these specific interactions: CAP will activate the $\lambda$ ci gene if its binding site is suitably positioned.
and activates transcription from \( P_{N}, \) In a construct in which \( O_{c} \) is moved one or a few turns of the DNA helix upstream, activation is abolished. In this case, repressor still binds cooperatively to the two sites (with the intervening DNA forming a small loop). Evidently, interaction between repressor at \( O_{c} \), with repressor at the upstream site subtly alters the way repressor sits at \( O_{c} \), and eliminates its interaction with polymerase.

CytR is an example of a repressor that normally works by inhibiting an activator, in this case CAP. At the cytR promoter, it binds to a site adjacent to, and downstream from, the CAP site and covers the CAP-activating region.

Inhibiting Polymerase

The P4 protein, encoded by a phage that grows on the bacterium *Bacillus subtilis*, binds at the P-A2c promoter and prevents polymerase escape. The polymerase binds strongly to this promoter, and the additional interaction between the regulator and polymerase is sufficient to prevent escape of the polymerase from the promoter. The same regulator, making the same contact with polymerase, activates transcription from another, weaker, promoter.

These effects are explained by the fact that promoters differ in the ease with which polymerase escapes the open complex as transcription initiates. Where that step is difficult, an additional protein-protein interaction between activator and polymerase can prevent escape. At other promoters, escape occurs readily, and interaction with the regulator activates transcription by recruitment of polymerase.

Experiments with the Gal repressor of *E. coli* suggest that this protein too can repress by interacting with polymerase. In this case, it is believed that the inhibition is caused by the repressor interacting with polymerase in such a fashion as to prevent the transition from the closed to the open complex at the gal promoter. In contrast to the action of the P4 protein, this inhibitory effect of the Gal repressor is highly sensitive to the precise positioning of the repressor-binding site in relation to the promoter.

### Footnotes

1. Another protein, called \( \sigma \), is associated with RNA polymerase. \( \sigma \) holds together the amino and carboxyl termini of \( \beta' \), and thereby helps that protein fold. Once the folded \( \beta' \) has been incorporated into the polymerase, \( \sigma \) can be removed without affecting enzyme activity in vitro. In vivo, elimination of \( \sigma \) slows down formation of active polymerase but is not lethal.

The two \( \sigma \) factors considered in the text (\( \sigma^{70} \) and \( \sigma^{14} \)) are made constitutively. Some \( \sigma \) factors are induced by environmental signals. For example, heat shock induces expression of a form of \( \sigma \) that directs the enzyme to a set of genes, expression of which helps to overcome the shock. Changes of \( \sigma \) factors can drive developmental programs. For example, the bacterium *Bacillus subtilis* undergoes a series of changes to form spores, and each stage is characterized by the synthesis of a new \( \sigma \) factor that directs the enzyme to a new set of genes.

2. The operator sequence is approximately twofold rotationally symmetric, and the center of that symmetry lies under that of the two binding domains of the protein. Parts of the protein in addition to the HTH motif can contribute to both the energy and specificity of the binding. Proteins such as the Lac repressor that bind as tetramers or even high-order oligomers typically contact an operator using a pair of subunits as shown in Figure 1.4. The orientation of the HTH motif is inverted in the Lac repressor compared with its orientation in CAP and other DNA-binding proteins we will encounter, e.g., \( \lambda \) repressor and Cso. Most DNA-protein recognition is effected using a "recognition \( \alpha\) helix;" but other ways, including use of a \( \beta \)-sheet, are known.

3. In the Lac repressor, the HTH motifs are presented on small domains, called the headpieces, attached to the main body of the protein. In the absence of inducer, the headpieces are rigidly held apart at the same distance that separates successive segments of the major groove along one face of the DNA. In the induced conformation, however, the headpieces are free to wiggle independently of each other. The added entropic cost of "fixing" the headpieces greatly reduces specific binding.

4. The phrase "increasing local concentration," a useful shorthand, can be misleading. Merely increasing the local concentration of a protein at a given site by introducing a second site for that same protein nearby will not, per se, increase the occupancy of the first site. Rather, the effect requires a simultaneous interaction between the proteins bound to the two sites. Likewise, if there is an interaction between the two proteins, but that interaction covers the DNA-binding surface of one of them, again the local concentration might be increased, but DNA binding will not be.

5. If site B in Figure 1.6 were modified so as to bind its protein considerably more tightly than site A binds protein A, binding of A to its site would depend on the presence of B, but binding of B might not require A. The discussion here and in the text also applies if A and B are two identical proteins that interact with each other while binding separate sites. More than two proteins can bind cooperatively to DNA. All of these variations are found in nature.

6. CAP has three activating regions. The two others touch the main body of \( \alpha \) and \( \sigma \), respectively. These additional activating regions have no role at the lac genes, but they do at other genes where the CAP-binding site is found at different positions relative to the polymerase-binding site.
One might imagine two ways that transcription would proceed once polymerase was tethered to DNA as in these experiments. In the first, the polymerase would maintain contact with the upstream tethering site and the DNA would be threaded past the immobilized polymerase. In the second, the upstream polymerase-DNA contacts would be broken as the polymerase moved away from the promoter. The latter is evidently the rule: a variety of experiments indicate that if the upstream polymerase-DNA interaction becomes too strong, that interaction can repress rather than activate transcription (see panel on More on Repression in Bacteria).

The following experiment emphasizes the simple adhesive roles of the -35 and UP elements. A truncated form of σ70, lacking its -35 region recognition domain, was generated. Polymerase bearing the truncated σ does not work at a promoter bearing the ordinary -10 and -35 sequences. It does work, however, if a CAP site is introduced and active CAP is present. The CAP-σ interaction compensates for the loss of the interaction between σ and the -35 region on DNA.

It has been reported that the Lac repressor and polymerase form a stable tripartite complex with DNA at the lac promoter. But other studies indicate that formation of this complex occurs only under nonphysiological conditions in vitro. And in vivo (as measured in footprinting experiments), polymerase is not found at the promoter in conjunction with repressor. So we have adopted the view that repressor works by excluding binding of RNA polymerase, rather than by forming a stable inactive complex.

Later in the λ life cycle, the concentration of Cro reaches a point at which it binds to O5 and O2 and turns down transcription of lytic phage genes as well.

We say "stimulates cleavage" because RecA does not bear a typical protease active site. Rather, it holds the repressor in a conformation that stimulates an auto-cleavage reaction.

Both of these protein-protein interactions are sufficiently weak that, at the concentration of repressor found in the cell, a significant fraction of the unbound protein is present as monomer.

If the λ system worked as does at lac, where the repressor is a stable tetramer whether interacting with DNA or not, then a much larger fraction of repressor molecules would need to be inactivated to achieve a similar level of induction.

Repressor at O5 can bind cooperatively either with another dimer at O5 or with another dimer at O2. The phenomenon is called "alternate pairwise cooperation." Because O5 binds repressor significantly more tightly than does O2, the former reaction predominates.

Only recently has it been demonstrated that autogenous negative control by repressor, in a lysogen, is physiologically important for allowing efficient induction. The degree of self-repression was previously underestimated for the following reason, one that raises a common methodological problem. As described in A Genetic Switch, the workings of the switch were deciphered (in part) by attaching regulatory elements (e.g., O5) to reporter genes (e.g., the lacZ gene), and supplying the regulators (e.g., λ repressor) from independently manipulated promoters. This "reductionist" approach could not reveal the effects of proteins binding to distal elements—in this case repressor binding to O5—and so the degree of autogenous self-regulation was significantly underestimated.

Consistent with these ideas are the two findings alluded to above, namely, that CAP can work at P5, and that activator bypass experiments work at both the lac and λ promoters. In addition, a single-amino-acid change in polymerase alters the effect of λ repressor at P5 from one on isomerization to one (largely) on initial polymerase binding.

Were both activators to contact the polymerase simultaneously, the total energy would be the sum of the two interactions considered separately. Because, in a simple Mass Action formulation, the energies of interaction contribute exponentially to the binding constant, the effect on recruitment would be synergistic, as defined here.

To our knowledge, activator bypass experiments, other than providing high concentrations of polymerase in vitro, have not been attempted for this case.

All of these activators typically work when bound well upstream of their target genes. Some genes bear separate binding sites for two such activators, and those genes can be controlled by either activator. For example, certain genes activated by NtrC are also activated—individually —by the related activator NifA, which senses oxygen and ammonia levels.

**BIBLIOGRAPHY**

**Books**


**RNA Polymerase, Promoters, and DNA Binding**


Activation and Repression


Antitermination


