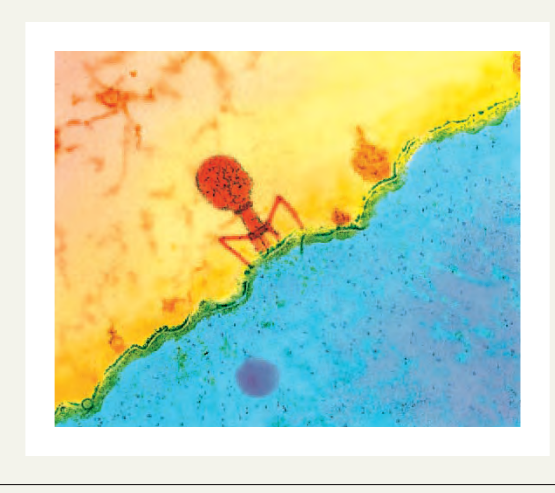
14

GENE EXPRESSION

Control in Prokaryotes and Phages



Artificially colored transmission electron micrograph of a T4 bacteriophage attached to an *Escherichia coli* bacterium. (© Biozentrum, University of Basel/SPL/Photo Researchers, Inc.)

STUDY OBJECTIVES

1. To study the way in which inducible and repressible operons work 406
2. To examine attenuator control in bacteria 415
3. To analyze the control of the life cycle of phage λ 418
4. To determine the way in which transposable genetic elements transpose and control gene expression in bacteria 425
5. To look at other transcriptional and posttranscriptional mechanisms of control of gene expression in bacteria and phages 430

STUDY OUTLINE

The Operon Model 406
***Lac* Operon (Inducible System)** 406
- Lactose Metabolism 406
- The Regulator Gene 406
- The Operator 408
- Induction of the *Lac* Operon 409
- *Lac* Operon Mutants 409

Catabolite Repression 412
***Trp* Operon (Repressible System)** 413
- Tryptophan Synthesis 413
- Operator Control 414

***Trp* Operon (Attenuator-Controlled System)** 415
- Leader Transcript 415
- Leader Peptide Gene 416
- TRAP Control 417
- Redundant Controls 418

Lytic and Lysogenic Cycles in Phage λ 418
- Phage λ Operons 418
- Early and Late Transcription 420
- Repressor Transcription 421
- Maintenance of Repression 421
- Lysogenic Versus Lytic Response 423

Transposable Genetic Elements 425
- IS Elements 425
- Composite Transposons 427
- Mechanism of Transposition 427
- Phenotypic and Genotypic Effects of Transposition 429

Other Transcriptional Control Systems 430
- Transcription Factors 430
- Promotor Efficiency 430

Translational Control 430
Posttranslational Control 433
- Feedback Inhibition 433
- Protein Degradation 433

Summary 434
Solved Problems 435
Exercises and Problems 436
Critical Thinking Questions 438

Genes are transcribed into RNA, which, for the most part, is then translated into protein. Control mechanisms are exercised along the way. Without some control of gene expression, an *Escherichia coli* cell, for example, would produce all its proteins in large quantities all the time, and all the cells in a eukaryotic organism would be identical. Although most control mechanisms are negative (preventing something from happening), controls can also be positive (causing some action to occur or enhancing some action). This chapter is devoted to analyzing control processes in prokaryotes and phages; in chapter 16, we examine control processes in eukaryotes.

In the process leading from a sequence of nucleotides in DNA to a protein, control is exerted in many places. In general, control of gene expression can take place at the levels of transcription, translation, or protein functioning. The most efficient place to control gene expression is at the level of transcription.

One of the best-understood mechanisms exerts control of transcription, regulating the production of messenger RNA according to need. *E. coli* messenger RNAs are short-lived in vivo: They degrade enzymatically within about two minutes. A complete turnover (degradation and resynthesis) in the cell's messenger RNA occurs rapidly and continually, and this rapid turnover is a prerequisite for transcriptional control, a central feature of the regulation of prokaryotic gene expression.

THE OPERON MODEL

Not all of the proteins prokaryotes can produce are needed in all circumstances in the same quantities. For example, some metabolites, such as sugars, which the cell breaks down for energy and as a carbon source, may not always be present in the cell's environment. If a given metabolite is not present, enzymes for its breakdown are not useful, and synthesizing these enzymes is wasteful. If the cell produces enzymes for the degradation of a particular carbon source only when this carbon source is present in the environment, the enzyme system is known as an **inducible system.** Inducible enzymes are synthesized when the environment includes a substrate for those enzymes. The enzymes will then catabolize (break down) the substrate.

On the other hand, the enzymes in many synthetic pathways are in low concentration or absent when an adequate quantity of the end product of the pathway is already available to the cell. That is, if the cell encounters an abundance of the amino acid tryptophan in the environment or if it is overproducing tryptophan, the cell stops the manufacture of tryptophan until a need arises again. A **repressible system** is a system of enzymes whose presence is repressed, stopping the production of the end product when it is no longer needed. Repressible systems are repressed by an excess of the end product of their synthetic (anabolic) pathway.

The best-studied inducible system is the ***lac* operon** in *E. coli.* Since the term *operon* refers to the control mechanism, we will defer a definition until we describe the mechanism.

LAC OPERON (INDUCIBLE SYSTEM)

Lactose Metabolism

Lactose (milk sugar—a disaccharide) is a β-galactoside that *E. coli* can use for energy and as a carbon source after it is broken down into glucose and galactose. The enzyme that performs the breakdown is **β-galactosidase** (fig. 14.1). (The enzyme can additionally convert lactose to allolactose, which, as we will see, is also important.) There are very few molecules of β-galactosidase in a wild-type *E. coli* cell grown in the absence of lactose. Within minutes after adding lactose to the medium, however, this enzyme appears in quantity within the bacterial cell. When the synthesis of β-galactosidase (encoded by the *lacZ,* or *z* gene) is induced, the production of two additional enzymes is also induced: **β-galactoside permease** (encoded by the *lacY,* or *y* gene) and **β-galactoside acetyltransferase** (encoded by the *lacA,* or *a* gene). The permease is involved in transporting lactose into the cell. The transferase is believed to protect the cell from the buildup of toxic products created by β-galactosidase acting on other galactosides. By acetylating galactosides other than lactose, the transferase prevents β-galactosidase from cleaving them.

The Regulator Gene

Not only are the three *lac* genes (*z, y, a*) induced together, but they are adjacent to one another in the *E. coli* chromosome; they are, in fact, transcribed on a single, polycistronic messenger RNA (fig. 14.2). Induction involves the protein product of another gene, called the **regulator gene,** or *i* gene (*lacI*). Although the regulator gene is located adjacent to the three other *lac* genes, it is a totally independent transcriptional entity. The regulator specifies a protein, called a **repressor,** that interferes with the transcription of the genes involved in lactose metabolism.

© The McGraw–Hill Companies, 2001

Figure 14.1 The enzyme β-galactosidase hydrolytically cleaves lactose into glucose and galactose (*a*). The enzyme can also convert lactose to allolactose (*b*).

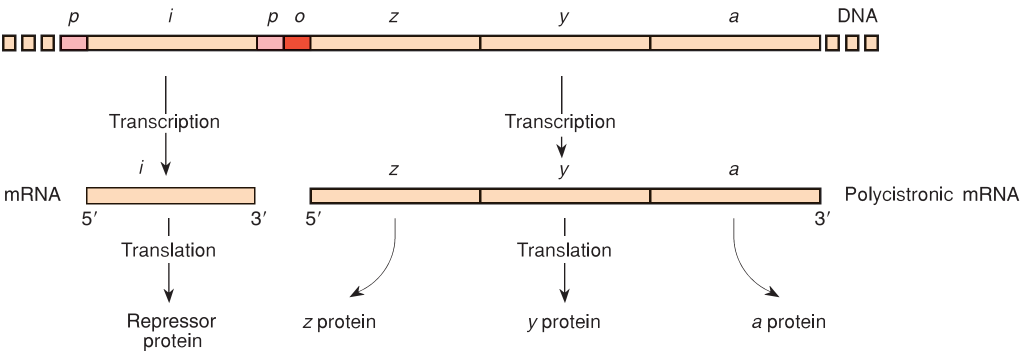


Figure 14.2 The *lac* operon is transcribed as a multigenic (polycistronic) mRNA. The *z, y,* and *a* indicate the *lacZ, lacY,* and *lacA* loci. The mRNA transcript is then translated as individual proteins. The *lac* operon regulator gene is denoted as *i*; the *o* stands for operator and the *p* for promoter. Both the operon and the regulator gene have their own promoters. (Source: Data from R. C. Dickson, et al., "Genetic regulation: The *lac* control region," *Science,* 187:27–35, January 10, 1975.)

The Operator

For the repressor protein to exert its influence over transcription, there must be a control element (receptor site) located near the beginning of the β-galactosidase (*lacZ*) gene. This control element is a region referred to as the **operator,** or operator site (fig. 14.2). The operator site is a sequence of DNA that the product of the regulator gene, the repressor, recognizes. When the repressor is bound to the operator, it either interferes with RNA polymerase binding or prevents the RNA polymerase from achieving the open complex (see chapter 10). In either case, transcription of the operon is prevented (fig. 14.3). The repressor is released when it combines with an *inducer,* a derivative of lactose called allolactose (see fig. 14.1).

Note that the promoter not only is recognized by RNA polymerase but also has other controlling elements in the immediate vicinity of the initiation site of transcription. We can now define an **operon** as a sequence of adjacent genes all under the transcriptional control of the same promoter and operator.

The nucleotide sequence of the *lac* operator region is shown in figure 14.4. The operator in figure 14.3 is referred to as the primary operator, o_1, centered at +11. Two other operator sequences have been found. One, o_2, is centered at +412. The third overlaps the C-terminal end of the *i* gene, is centered at −82, and is referred to as o_3. The structure of the repressor and its interaction with the operator sites was worked out recently with X-ray crystallography. The functional repressor is a homotetramer of the protein product of

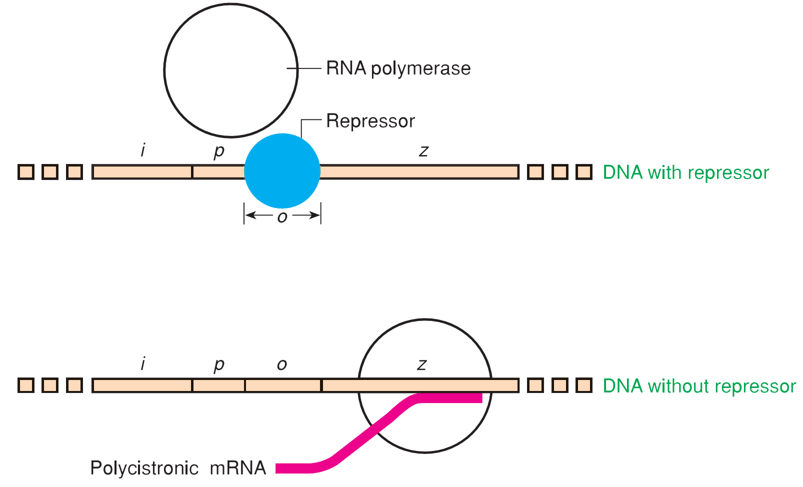


Figure 14.3 The repressor. By binding to the operator, the repressor either prevents RNA polymerase from binding to the promoter and transcribing the *lac* operon as shown, or prevents the polymerase from achieving the open configuration. In either case, transcription of the *lac* operon is prevented. When the repressor is not present, transcription takes place. The functional repressor is a tetramer.

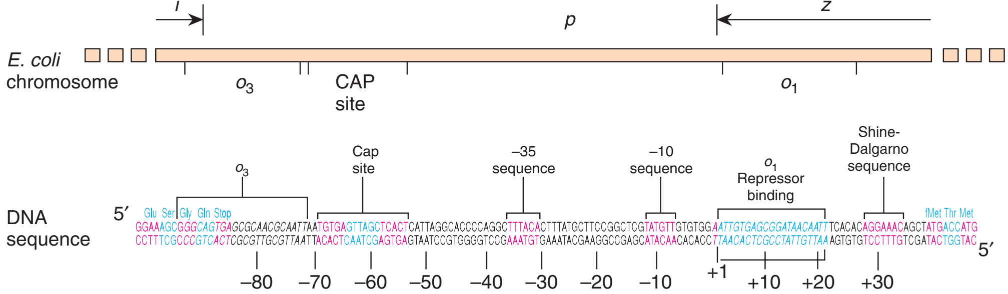


Figure 14.4 The *lac* operon promoter and operator regions. The CAP site is described later. The base sequence corresponds to the diagram above it. The terminal amino acids of the *i* gene are shown, as well as the initial amino acids of the *lacZ* gene. In addition, we picture the Shine-Dalgarno sequence of the DNA, the repressor-binding region (centered at around +10 of the gene), the −10 and −35 sequences of the promoter, and primary (o_1) and secondary (o_3) operator sites (see text). (Data from R. C. Dickson, et al., "Genetic regulation: The *lac* control region," *Science,* 187:27–35, January 10, 1975.)

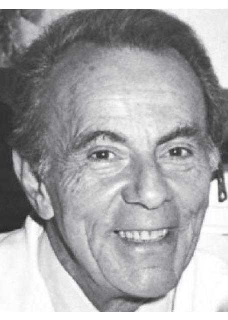
the *i* gene; that is, it is formed from four identical copies of the repressor protein. Since each operator site has twofold symmetry, two repressor monomer proteins bind to each operator site. The monomer is shaped so that it fits into the major groove of the DNA to locate the exact base sequence of the operator; it then binds at that point through electrostatic forces. A tetramer can bind to two of the operator sites at the same time, presumably o_1 and o_3 or o_1 and o_2. In the process, the DNA is formed into a loop (fig. 14.5).

Induction of the Lac *Operon*

Under conditions of repression, before the operon can be "turned on" to produce lactose-utilizing enzymes, the repressor will have to be removed from the operator. The repressor is an **allosteric protein;** when it binds with one particular molecule, it changes the shape of the protein, which changes its ability to react with a second particular molecule. Here the first molecule is the inducer allolactose and the second molecule is the operator DNA. When allolactose is bound to the repressor, it causes the repressor to change shape and lose its affinity for operator sequences (fig. 14.5).

With allolactose bound to the repressor, the ability of the repressor to bind to the operator is greatly reduced, by a factor of 10^3. Since no covalent bonds are involved, the repressor simply dissociates from the operator. After the repressor releases from the operator, RNA polymerase can now begin transcription. The three *lac* operon genes are then transcribed and subsequently translated into their respective proteins.

This system of control is very efficient. The presence of the lactose molecule permits transcription of the genes of the *lac* operon, which act to break down the lactose. After all the lactose is metabolized, the repressor returns to its original shape and can again bind to the operator. The system is "turned off." Using very elegant genetic analysis, details of this system were worked out by François Jacob and Jacques Monod, who subsequently won 1965 Nobel prizes for their efforts.



François Jacob (1920–). (Courtesy of Dr. François Jacob.)

Jacques Monod (1910–1976). (Archives Photographiques, Musée Pasteur.)

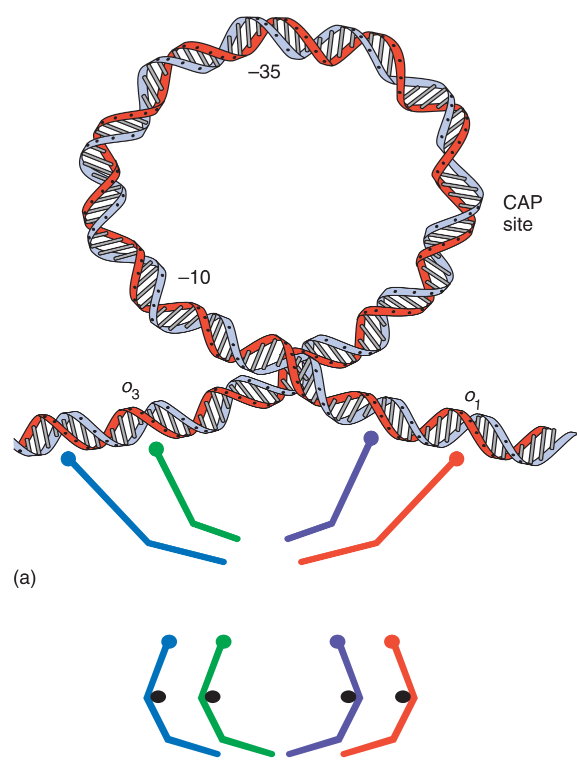


Figure 14.5 Because the *lac* operator DNA sequences are palindromes, each half can bind one repressor subunit. (*a*) The tetrameric repressor binds to o_1 and o_3, causing the DNA in between to form a loop. Each of the subunits is shown in a different *color.* The round portion of the subunit in touch with the DNA is the N-terminal end of the repressor subunit; the C-terminal ends form tails that bind the subunits together. Also indicated are the CAP site and the −10 and −35 sequences. (*b*) When each of the subunits binds an allolactose molecule (*black circles*), the shape of the middle portion of the subunit changes, causing the subunit to fall free of the operators.

Lac *Operon Mutants*

Merozygote Formation

Discovery and verification of the *lac* operon system came about through the use of mutants and partial diploids of the *lac* operon well before DNA sequencing techniques had been developed. The structural (enzyme-specifying) genes of the *lac* operon, *z, y,* and *a,* all have known mutant forms in which the particular enzyme does not perform its function. These mutant forms are designated z^-, y^-, and a^-. The alleles for normal forms of the enzymes are z^+, y^+, and a^+.

Partial diploids in *E. coli* can be created through sexduction (chapter 7) because some strains of *E. coli* have

the *lac* operon incorporated into an F′ factor. Since F^+ strains can pass the F′ particle into F^- strains, *lac* operon diploids (also called merozygotes, or partial diploids) can be formed. By careful manipulation, various combinations of mutations can be looked at in the diploid state.

Constitutive Mutants

Constitutive mutants are mutants in which the three *lac* operon genes are transcribed at all times—that is, they are not turned off even in the absence of lactose. Inspection of figure 14.3 shows that constitutive production of the enzymes can come about in several ways. A defective repressor, produced by a mutant regulator gene, will not turn the system off, nor will a mutant operator that will no longer bind the normal repressor. The regulator constitutive mutants are designated i^-; the operator constitutive mutants are designated o^c. Both types of mutants produce the same phenotype: constitutive expression of the three *lac* operon genes.

When a new mutant is isolated, it is possible to determine whether it is caused by a regulator or operator mutation. For example, we can determine the exact location of a mutation on the bacterial chromosome by standard mapping techniques (see chapter 7) or, more recently, by DNA sequencing (see chapter 13). Alternatively, the Jacob and Monod model predicts different modes of action for the two types of mutations. In merozygotes, a constitutive operator mutation affects only the operon it is physically a part of. Operator muta-

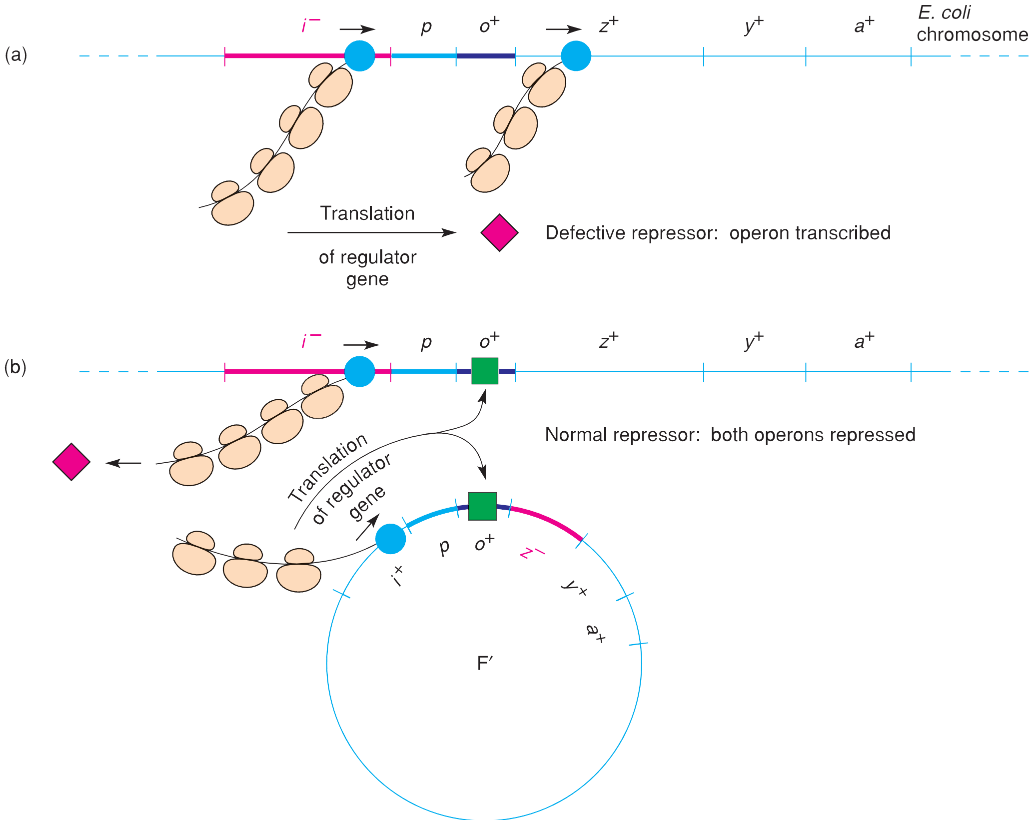


Figure 14.6 (*a*) A *lac* operon in *E. coli* with a mutation of the regulator gene (i^-). Transcription and translation of this gene yield a defective repressor; the cell thus has constitutive production of the *lac* operon. In (*b*), the wild-type regulator gene is introduced in an F′ factor; there is both a bacterial chromosome and an F′ factor, each containing a regulator gene. (The F′ operon carries a mutant *z* allele, allowing us to keep track of the transcriptional control of the chromosomal operon only.) In this case, the phenotype is now normal (inducible) because enough repressor is produced by the F′ allele (i^+), by transcription and translation, to bind to both operators. RNA polymerase is shown as *solid spheres* on the DNA; the wild-type repressor is shown as a *green square;* the mutant repressor, which cannot bind to the operator, is shown as a *red diamond.*

tions are therefore called ***cis*-dominant.** However, a constitutive *i*-gene mutation, since it works through an altered protein, is recessive to a wild-type regulator gene in the same cell, regardless of which operon (chromosomal or F′ factor) the mutation is on. Constitutive regulator mutations are, therefore, ***trans*-acting.** (If two mutations are on the same piece of DNA, they are in the *cis* configuration. If they are on different pieces of DNA, they are in the *trans* configuration.) *Trans-acting* mutations usually work through a protein product that diffuses through the cytoplasm. *Cis-acting* mutants are changes in recognition sequences on the DNA.

In figure 14.6*a,* the bacterium has a regulator constitutive mutation (i^-); the cell has constitutive production of the operon. If the wild-type regulator is introduced in an F′ plasmid (fig. 14.6*b*), the normal (inducible) phenotype is restored because the F′ i^+ allele is dominant to the chromosomal mutation—the i^+ regulates both the chromosomal and F′ operons. Hence, both operons are inducible. We don't need to be concerned about the other components of the F′ plasmid because it carries a z^- allele; only the activity of the chromosomal operon will be observed. In figure 14.7*a*, however, the chromosomal operon carries an operator constitutive mutation; the cell also has constitutive production of the operon. When a wild-type operator is introduced into the cell in an F′ plasmid (fig. 14.7*b*), the cell still has the constitutive phenotype because the operator allele on the F′

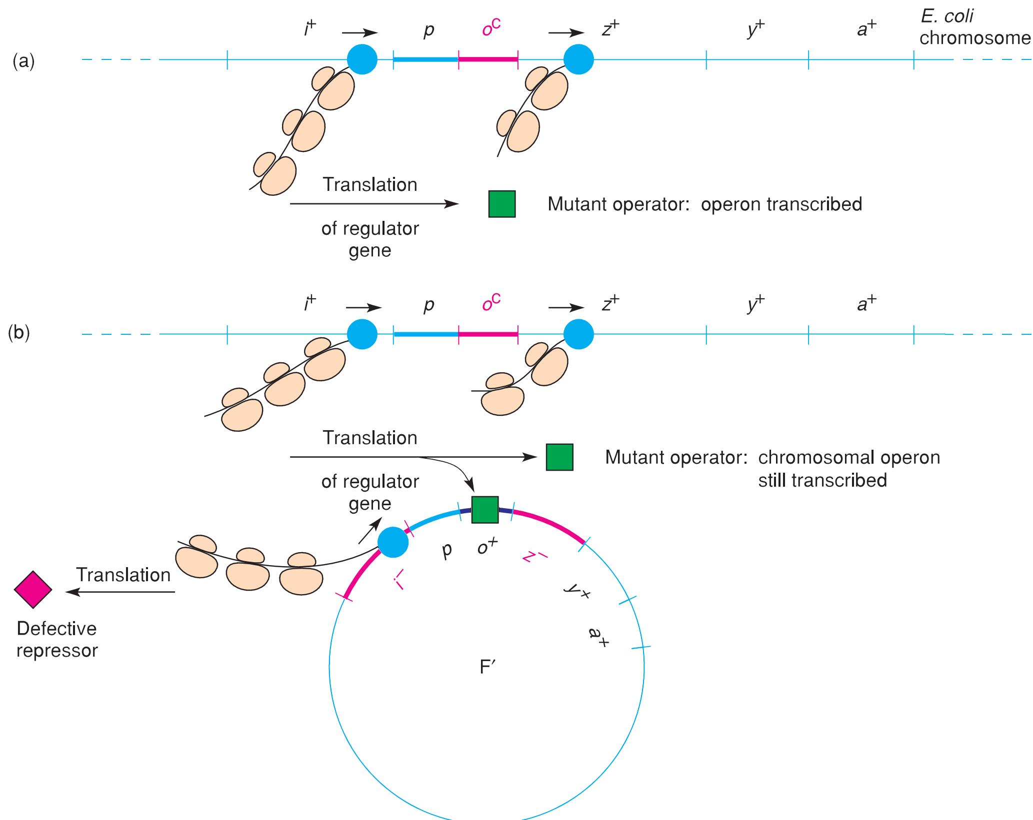


Figure 14.7 (*a*) A *lac* operon in *E. coli* with a mutation of the operator (o^c). The cell has a constitutive phenotype; the operator cannot bind the wild-type repressor protein, and thus transcription is continuous, even in the absence of lactose. The phenotype is unchanged even when a wild-type operator is introduced into the cell in an F′ factor (*b*); there is both a bacterial chromosome and an F′ factor, each containing an operator. (The F′ operon carries mutant regulator and *z* alleles, allowing us to keep track of the transcriptional control of the chromosomal operon only.) The F′ operator does not change the phenotype of the cell because the wild-type operator exerts no control over the chromosomal operator, which exerts a *cis-dominant* effect; another operator on another operon has no effect. RNA polymerase is shown as *solid spheres* on the DNA; the wild-type repressor is shown as a *green square;* the mutant repressor, which cannot bind to the operator, is shown as a *red diamond.*

plasmid does not control the bacterial operon; the *lac* operon on the bacterial chromosome will be continually transcribed. The chromosomal operon has a *cis-dominant* operator mutation that has a constitutive phenotype. Note, too, that only the bacterial chromosome determines the phenotype because the introduced F′ plasmid has a z^- allele.

Other Lac *Operon Control Mutations*

Other mutations have also been discovered that support the Jacob and Monod operon model. A superrepressed mutation, i^s, was located. This mutation represses the operon even in the presence of large quantities of the inducer. Thus, the repressor seems to have lost the ability to recognize the inducer. Basically, the *i*-gene product is acting as a constant repressor rather than as an allosteric protein. In an i^s/i^+ merozygote, both operons are repressed because the i^s repressor binds to both operators. Another mutation, i^Q, produces much more of the repressor than normal and presumably represents a mutation of the promoter region of the *i* gene.

In 1966, W. Gilbert and B. Müller-Hill isolated the *lac* repressor and thereby provided the final proof of the validity of the model. At about the same time, M. Ptashne and his colleagues isolated the repressor for phage λ operons. Control of gene expression in phage λ is discussed later in this chapter.

Mark Ptashne (1940–). (Courtesy of Dr. Mark Ptashne.)

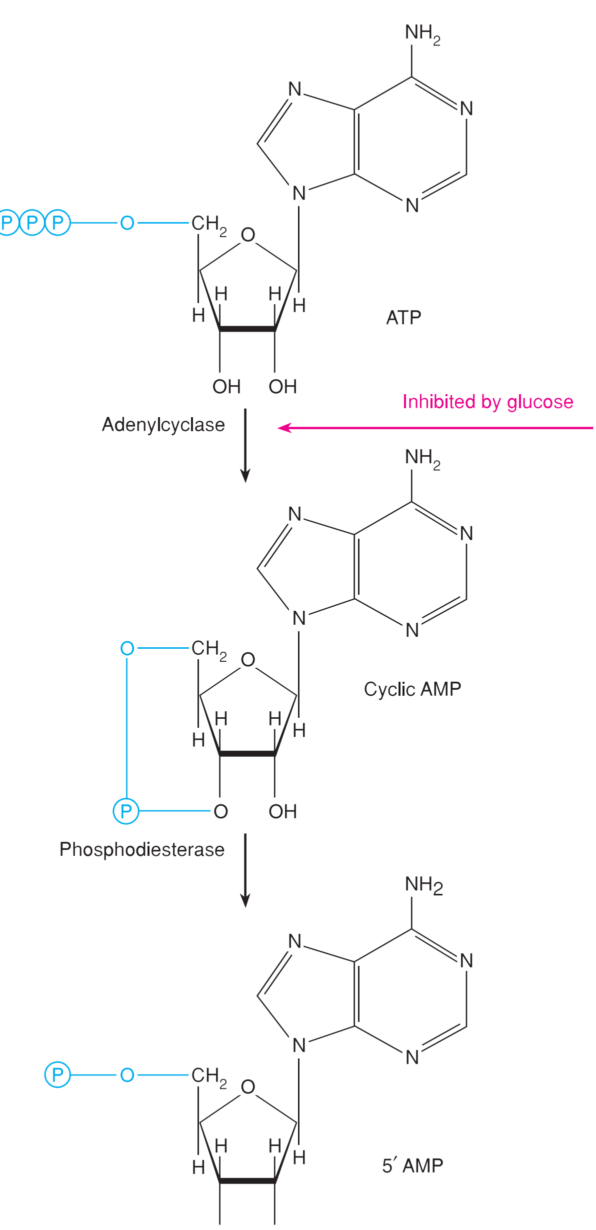


Figure 14.8 Structure of cyclic AMP (cAMP). Glucose uptake lowers the quantity of cyclic AMP in the cell by inhibiting the enzyme adenylcyclase, which converts ATP to cAMP.

CATABOLITE REPRESSION

An interesting property of the *lac* operon and other operons that code for enzymes that catabolize certain sugars (e.g., arabinose, galactose) is that they are all repressed in the presence of glucose. That is, glucose is catabolized in preference to other sugars; the mechanism **(catabolite repression)** involves **cyclic AMP** (cAMP; fig. 14.8). In eukaryotes, cAMP acts as a *second messenger,* an intracellular messenger regulated by certain extracellular hormones. Geneticists were surprised to discover cAMP in *E. coli,* where it works in conjunction with another regulatory protein, the **catabolite activator protein (CAP),** to control the transcription of certain operons.

In the absence of glucose, cAMP combines with CAP, and the CAP-cAMP complex binds to a distal part of the promoter of operons with CAP sites (e.g., the *lac* operon; see fig. 14.4). This binding apparently enhances the affinity of RNA polymerase for the promoter, because without the binding of the CAP-cAMP complex to the promoter, the transcription rate is very low. The uptake of glucose by *E. coli* cells causes the loss of cAMP from the cell, probably by inhibiting adenylcyclase (fig. 14.8), and thus lowers the CAP-cAMP level. The transcription rate of operons with CAP sites will, therefore, be reduced (fig. 14.9). The same reduction of transcription rates is noticed in mutant strains of *E. coli* when this part of the distal end of the promoter is deleted. The binding of CAP-cAMP to the CAP site causes the DNA to bend more than 90 degrees (fig. 14.10). This bending, by itself, may enhance transcription, making the DNA more available to RNA polymerase.

In addition, at some point in the process of initiation of transcription, the CAP is in direct contact with RNA polymerase. This was shown by photo cross-linking studies in which the CAP was treated with a cross-linking agent that bound the α subunit of RNA polymerase when irradiated with UV light. For the two proteins to cross-link, they must be in direct contact during the initiation of transcription.

Catabolite repression is an example of positive regulation: Binding of the CAP-cAMP complex at the CAP site enhances the transcription rate of that transcriptional unit. Thus, the *lac* operon is both positively and negatively regulated; the repressor exerts negative control, and the CAP-cAMP complex exerts positive control of transcription.

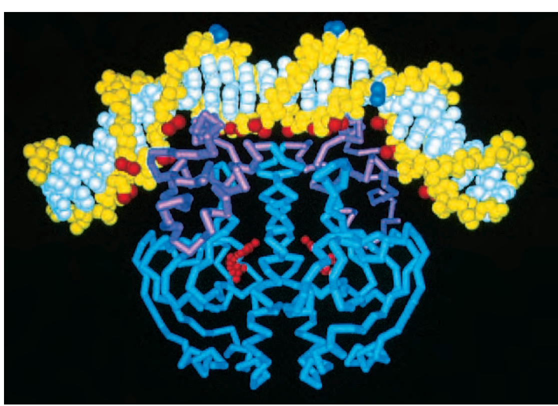
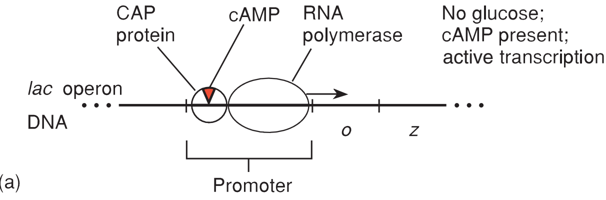


Figure 14.10 CAP-DNA interaction: model of cap protein and DNA. The cap site has twofold symmetry, like the operator. The cAMP-binding domain is *dark blue,* the DNA-binding domain is *purple,* and the cyclic AMP molecules within the protein are *red.* The DNA sugar-phosphate backbones are shown in *yellow,* the bases in *light blue.* DNA phosphates in *red* (on the double helix) are those whose modification interfere with CAP binding. DNA phosphates in *dark blue* (also on the double helix) are those especially prone to nuclease attack because of the bending of the DNA. (Courtesy of Thomas A. Steitz.)



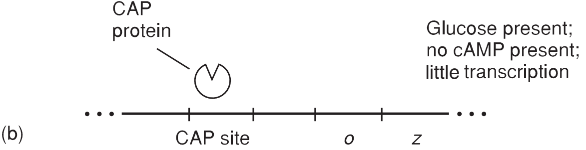


Figure 14.9 Catabolite repression. When cAMP is present in the cell (no glucose is present), it binds with CAP protein, and together they bind to the CAP site in various sugar-metabolizing operons, such as the *lac* operon shown here. The CAP-cAMP complex enhances the transcription of the operon. When glucose is present, it inhibits the formation of cAMP. Thus no CAP-cAMP complex forms, and transcription of the same operons is reduced.

TRP OPERON (REPRESSIBLE SYSTEM)

The inducible operons are activated when the substrate that is to be catabolized enters the cell. Anabolic operons function in the reverse manner: They are turned off (repressed) when their end product accumulates beyond the needs of the cell. Two entirely different, although not mutually exclusive, mechanisms seem to control the transcription of repressible operons. The first mechanism follows the basic scheme of inducible operons and involves the end product of the pathway. The second mechanism involves secondary structure in messenger RNA transcribed from an attenuator region of the operon.

Tryptophan Synthesis

One of the best-studied repressible systems is the tryptophan, or *trp,* operon in *E. coli.* The *trp* operon contains

the five genes that code for the synthesis of the enzymes that build tryptophan, starting with chorismic acid (fig. 14.11). It has a promoter-operator sequence (*p, o*) as well as its own regulator gene (*trpR*).

Operator Control

In this repressible system, the product of the *trpR* gene, the repressor, is inactive by itself; it does not recognize the operator sequence of the *trp* operon. The repressor only becomes active when it combines with tryptophan. Thus, when tryptophan builds up, enough is available to bind with and activate the repressor. Tryptophan is thus referred to as the **corepressor.** The corepressor-repressor complex then recognizes the operator, binds to it, and prevents transcription by RNA polymerase.

After the available tryptophan in the cell is used up, the diffusion process causes tryptophan to leave the repressor, which then detaches from the *trp* operator. The transcription process no longer is blocked and can proceed normally (the operon is now **derepressed**). Transcription continues until enough of the various enzymes have been synthesized to again produce an excess of tryptophan. Some becomes available to bind to the repressor and make a functional complex, and the operon is again shut off and the process repeated, ensuring that tryptophan is being synthesized as needed (fig. 14.12). This regulation is modified, however, by the existence of the second mechanism for regulating repressible operons—attenuation.

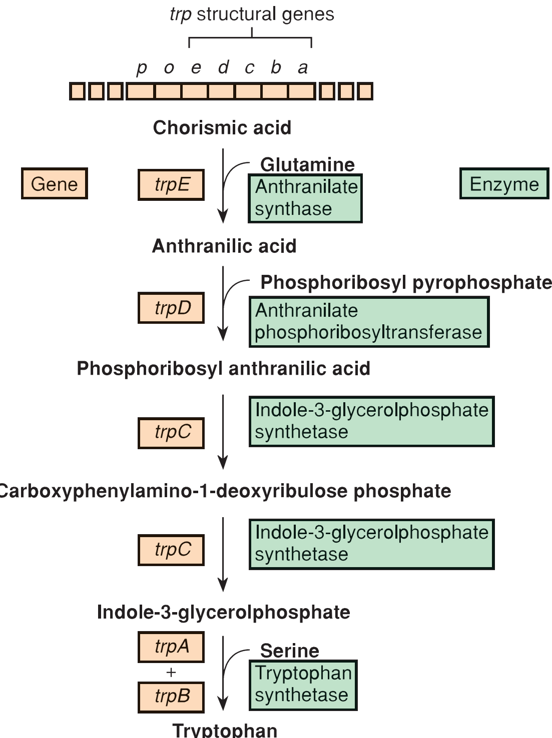


Figure 14.11 Genes of the tryptophan operon in *E. coli.* The enzymes they produce control the conversion of chorismic acid to tryptophan. The symbol *o* on the chromosome refers to the *trp* operator, which has its own repressor, the product of the *trpR* gene.

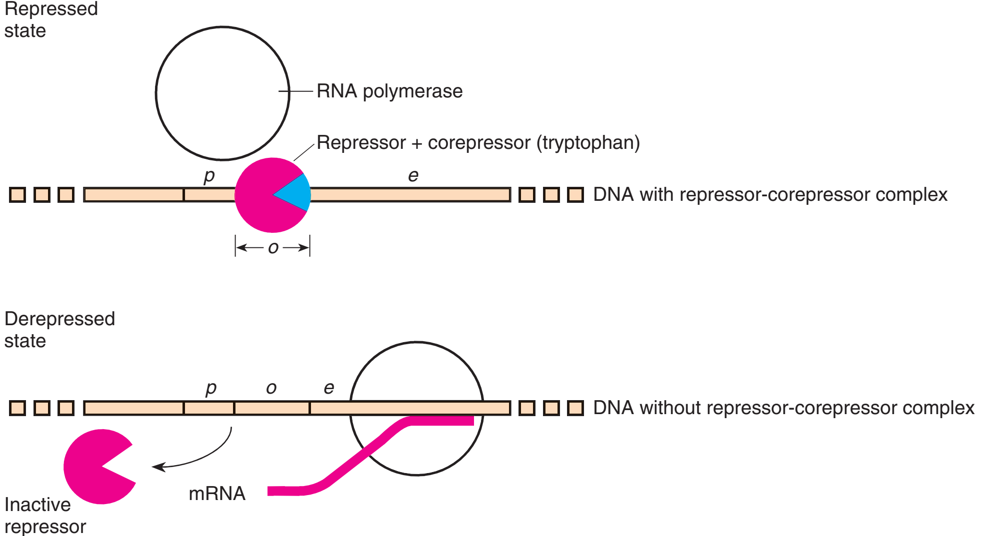


Figure 14.12 The repressor-corepressor complex binds at the operator and prevents the transcription of the *trp* operon in *E. coli.* Without the corepressor, the repressor cannot bind, and therefore transcription is not prevented. The *blue* wedge is the corepressor (two tryptophan molecules), and the partial *red* circle is the repressor.

© The McGraw−Hill Companies, 2001

TRP OPERON (ATTENUATOR-CONTROLLED SYSTEM)

Details of the second control mechanism of repressible operons have been elucidated primarily by C. Yanofsky and his colleagues, who worked with the tryptophan operon in *E. coli.* This type of operon control, control by an **attenuator region,** has been demonstrated for at least five other amino acid-synthesizing operons, including the leucine and histidine operons. This regulatory mechanism may be the same for most operons involved in the synthesis of an amino acid.

Charles Yanofsky (1925–).
(Courtesy of Dr. Charles Yanofsky.)

Leader Transcript

In the *trp* operon, an attenuator region lies between the operator and the first structural gene (fig. 14.13). The messenger RNA transcribed from the attenuator region, termed the **leader transcript,** has been sequenced, revealing two surprising and interesting facts. First, four subregions of the messenger RNA have base sequences that are complementary to each other so that three different stem-loop structures can form in the messenger RNA (fig. 14.14). Depending on circumstances, regions 1–2 and 3–4 can form two stem-loop structures, or region 2–3 can form a single stem-loop. When one stem-loop structure is formed, the others are preempted. As

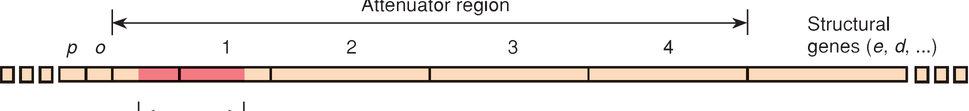


Figure 14.13 Attenuator region of the *trp* operon, which contains the leader peptide gene (*red*). This region is transcribed into the leader transcript.

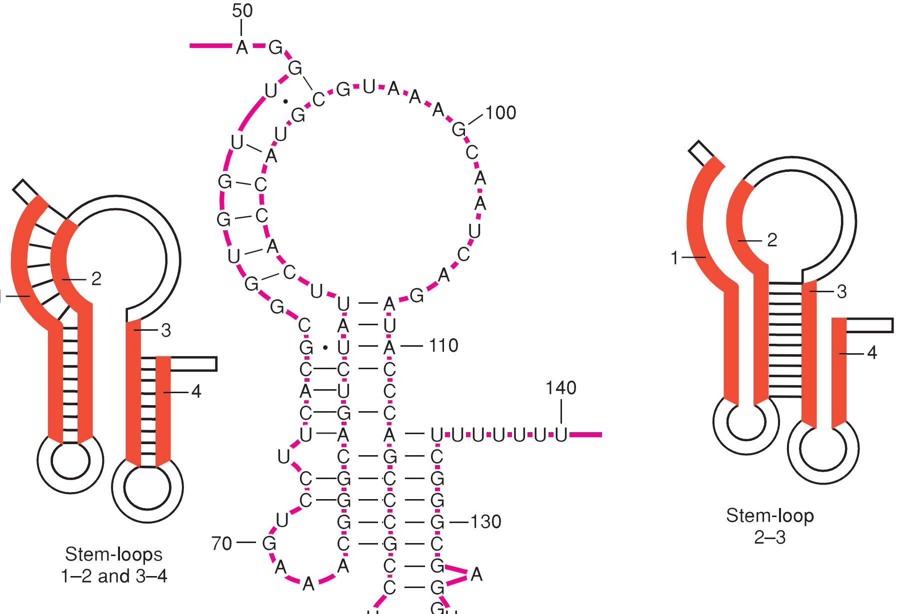


Figure 14.14 Nucleotide sequence of part of the leader transcript of the *trp* attenuator region (bases 50 to 140). Stem-loops 1–2 and 3–4, or stem-loop 2–3, can form because of complementarity of the nucleotides. All possible base pairings are shown in the middle of the figure. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

we will see, the particular combination of stem-loop structures determines whether transcription continues.

Leader Peptide Gene

The second fact obtained by sequencing the leader transcript is that there is a small gene coding information for a peptide from bases 27 to 68 (fig. 14.15). The gene for this peptide is referred to as the **leader peptide gene.** It codes for fourteen amino acids, including two adjacent tryptophans. These adjacent tryptophan codons are critically important in attenuator regulation. The proposed mechanism for this regulation follows.

Excess Tryptophan

Assuming that the operator site is available to RNA polymerase, transcription of the attenuator region will begin. As soon as the 5′ end of the messenger RNA for the leader peptide gene has been transcribed, a ribosome attaches and begins translating this messenger RNA. Depending on the levels of amino acids in the cell, three different outcomes can take place. If the concentration of tryptophan in the cell is such that abundant tryptophanyl-tRNAs exist, translation proceeds down the leader peptide gene. The moving ribosome overlaps regions 1 and 2 of the transcript and allows stem-loop 3–4 to form, as shown in the configuration at the far left of figure 14.16. This stem-loop structure, referred to as the **terminator,** or **attenuator, stem,** causes transcription to be terminated. Note that stem-loop 3–4, the terminator stem, followed by a series of uracil-containing bases, is a rho-independent transcription terminator (see chapter 10). Hence, when existing quantities of tryptophan, in the form of tryptophanyl-tRNA, are adequate for translation of the leader peptide gene, transcription is terminated.

Tryptophan Starvation

If the quantity of tryptophanyl-tRNA is lowered, the ribosome must wait at the first tryptophan codon until it acquires a Trp-tRNA^{Trp}. This is shown in the configuration in the middle part of figure 14.16. The stalled ribosome will

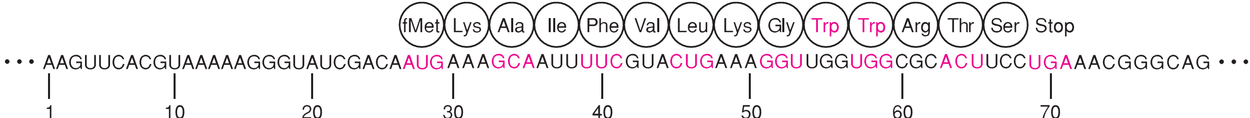


Figure 14.15 Base sequence of the *trp* leader transcript and the amino acids these nucleotides code. Note the presence of adjacent tryptophan codons. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

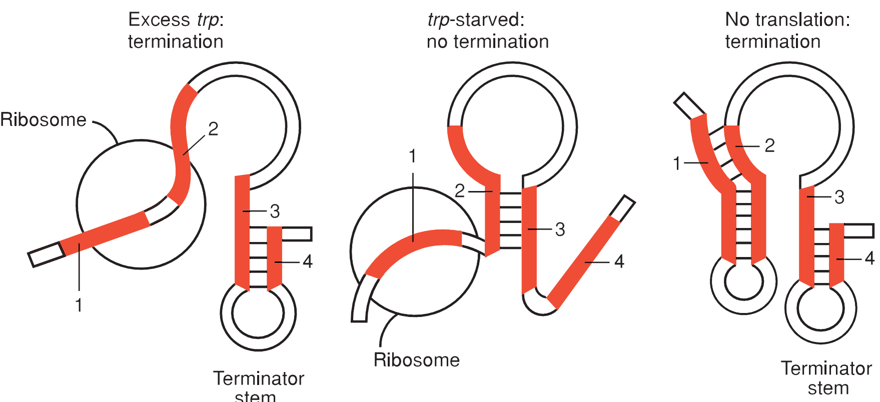


Figure 14.16 Model for attenuation in the *E. coli trp* operon. The *circle* represents the ribosome attempting to translate the leader transcript of figure 14.14. Under conditions of excess tryptophan, the 3–4 stem-loop forms (the terminator stem), terminating transcription. Under conditions of tryptophan starvation, the ribosome is stalled, and stem-loop 2–3 forms, allowing continued transcription. Under general starvation, there is no translation, resulting in the formation of stem-loops 1–2 and 3–4, which again results in the termination of transcription. (From D. L. Oxender, et al., "Attenuation in the *Escherichia coli* tryptophan operon: Role of RNA secondary structure involving the tryptophan codon region," *Proceedings of the National Academy of Sciences,* 76:5524–28, 1979. Reprinted by permission.)

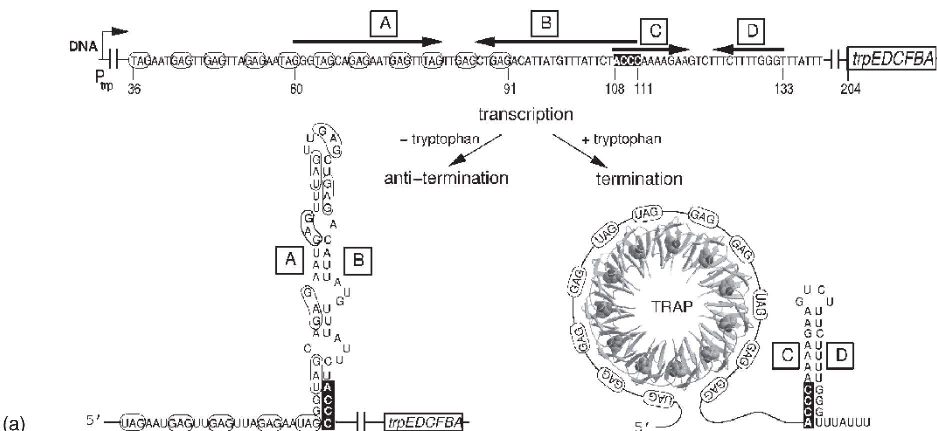
permit stem-loop 2–3 to form, which precludes the formation of the terminator stem-loop (3–4). In this configuration, transcription is not terminated, so that eventually, the whole operon is transcribed and translated, raising the level of tryptophan in the cell. The stem-loop 2–3 structure is referred to as the **preemptor stem.** Note that the preemptor stem is not a rho-independent transcription terminator and thus, without the rho protein present, will not terminate transcription (see chapter 10).

General Starvation

A final configuration is possible, as shown on the far right in figure 14.16. Here, no ribosome interferes with stem formation and, presumably, stem-loops 1–2 and 3–4 (terminator) form. This configuration also terminates transcription because of the terminator stem. It is believed that this configuration occurs if the ribosome is stalled on the 5′ side of the *trp* codons, which happens when the cell is starved for other amino acids. Presumably, it makes no sense to manufacture tryptophan when other amino acids are in short supply. Hence, the cell can carefully bring up the levels of the various amino acids in the most efficient manner.

TRAP Control

The tryptophan operon in bacilli such as *Bacillus subtilis* is also controlled by attenuation, but secondary structure in the mRNA transcript is induced by binding not the ribosome, but a ***trp*** **RNA-binding attenuation protein (TRAP).** This protein attaches to the nascent messenger RNA only after the protein binds tryptophan molecules; the result is a terminator stem that forms in the messenger RNA. In the absence of excess tryptophan, TRAP does not bind to the messenger RNA, a preemptor (also called an antiterminator) stem, not the terminator stem, forms, and transcription continues (fig. 14.17*a*). Recently, the



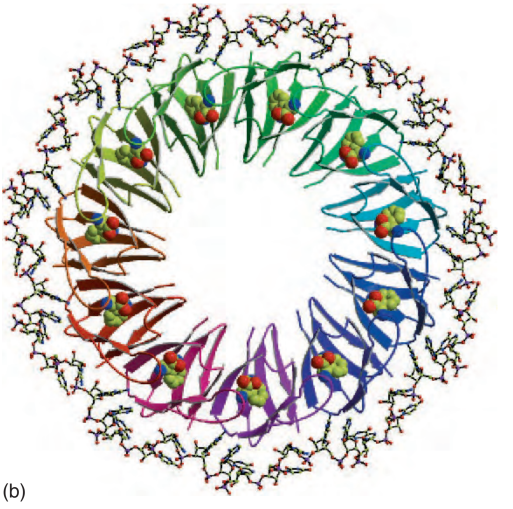


Figure 14.17 The *trp* operon control by attenuation in *Bacillus subtilis.* (*a*) The top of the figure shows the leader region of the DNA with the two parts of the antitermination stem (A, B) and the termination stem (C, D). The triplets (GAG and TAG in DNA, or GAG and UAG in the messenger RNA) that the *trp* RNA-binding attenuation protein (TRAP) binds to are circled. The label *trpEDCFBA* refers to the structural genes of the *trp* operon. Nucleotides 108–133 (C, D) form the terminator stem, and nucleotides 60–111 (A, B) form the antiterminator stem. The arrows below the boxed letters A–D indicate the inverted repeat sequences forming the stems. Without TRAP, the antiterminator stem forms; with TRAP, the terminator stem forms as TRAP is bound by nucleotides 36–91 of the messenger RNA. Part (*b*) is a close-up of the mRNA (ball-and-stick model) wrapped around TRAP (ribbon diagram with subunits in different colors) bound by tryptophan molecules (spheres). (From Alfred A. Antson, et al., "Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA" in *Nature*, Vol.40, September 16, 1999, fig. 1 p. 234 and fig. 2a p. 237. Reprinted by permission of Macmillan Ltd.)

structure of the protein was worked out; it has eleven symmetrical loops, each of which can bind a tryptophan molecule. When TRAP is bound to tryptophan molecules, it can attach to triplets in the messenger RNA transcript, triplets of GAG or UAG. The TRAP wraps the mRNA around itself, forming an elegant pinwheel (fig. 14.17*b*).

Redundant Controls

Some amino acid operons are controlled only by attenuation, such as the *his* operon in *E. coli,* in which the leader peptide gene contains seven histidine codons in a row, or the *trp* operon in *B. subtilis.* Redundant control (repression and attenuation) of tryptophan biosynthesis in *E. coli* allows the cell to test both the tryptophan levels (tryptophan is the corepressor) and the tryptophanyl-tRNA levels (in the attenuator control system). The attenuator system also allows the cell to regulate tryptophan synthesis on the basis of the shortage of other amino acids. For example, when there is a shortage of both tryptophan and arginine, operator control allows transcription to begin, but attenuator control terminates transcription because stem-loops 1–2 and 3–4 form (fig. 14.16).

LYTIC AND LYSOGENIC CYCLES IN PHAGE λ

When a bacteriophage infects a cell, it must express its genes in an orderly fashion; some gene products are needed early in infection, and other products are not needed until late in infection. Early genes usually control phage DNA replication; late genes usually determine phage coat proteins and the lysis of the bacterial cell. A phage is most efficient if it expresses the early genes first and the late genes last in the infection process. Also, temperate phages have the option of entering into lysogeny with the cell; here, too, control processes determine which path is taken. One generalization that holds true for most phages is that their genes are clustered into early and late operons, with separate transcriptional control mechanisms for each.

Phage λ is perhaps the best-studied bacteriophage. It has a chromosome of about 48,500 base pairs. Since it is a temperate phage, it can exist either vegetatively or as a prophage, integrated into the host chromosome. This phage warrants our attention because of the interesting and complex way that its life-cycle choice is determined. It is a model system of operon controls. The complexity results from having two conflicting life-cycle choices.

Briefly, the expression of one of the two life-cycle alternatives, lysogenic or lytic cycles, depends on whether two repressors, CI and Cro*,* have access to operator sites. The CI repressor acts to favor lysogeny: it represses the lytic cycle. The Cro repressor favors the lytic cycle and represses lysogeny. The operator sites, when bound by either CI or Cro, can either enhance or repress transcription. Other control mechanisms are also involved in determining aspects of the λ life cycle, including antitermination and multiple promoters for the same genes.

Phage λ Operons

Phage λ (see fig. 7.21) exhibits a complex system of controls of both early and late operons, as well as controls for the decision of lytic infection versus lysogenic integration. The λ genes are grouped into four operons: left, right, late, and repressor (fig. 14.18). The left and right operons contain the genes for DNA replication and recombination and phage integration. The late operon contains the genes that determine phage head and tail proteins and lysis of the host cell. The sequence of events following phage infection is relatively well known.

The map of λ (fig. 14.18) is a circle, but the λ chromosome has two linear stages in its life cycle (fig. 14.19). It is packed within the phage head in one linear form, and it integrates into the host chromosome to form a prophage in another linear form (fig. 14.19). Those two linear forms do not have the same ends (figs. 14.18 and 14.19*b*). The mature DNA, which is packed within the phage heads before lysis of the cells, is flanked by *cos* sites (chapter 13). It results from a break in the circular map between the *A* and *R* loci. The prophage is integrated at the *att* site, and the circular map is thus broken there at integration.

The homologous integration sites on both λ and the *E. coli* chromosome consist of a 15 bp core sequence (called "O" in both), flanked by different sequences on both sides in both the bacterium and the phage (fig. 14.20). In the phage, the region is referred to as POP′, where P and P′ (P for phage) are two different regions flanking the O core on the phage DNA. In the bacterium, the region is called BOB′, where B and B′ (B for bacterium) are two different regions flanking the O core on the *E. coli* chromosome. Integration, which is a part of the lysogenic life cycle, requires the product of the λ *int* gene, a protein known as *integrase,* and is referred to as **site-specific recombination.** Later excision of the prophage, during induction, when the phage leaves the host chromosome to enter the lytic cycle, requires both the integrase and the protein product of the neighboring *xis* gene, *excisionase.*

After infection of the *E. coli* cell by a λ phage, the phage DNA circularizes, using the complementarity of the *cos* sites. Transcription begins, and within a very short time the phage is guided toward either entering the lytic cycle and producing virus progeny or entering

 © The McGraw–Hill Companies, 2001

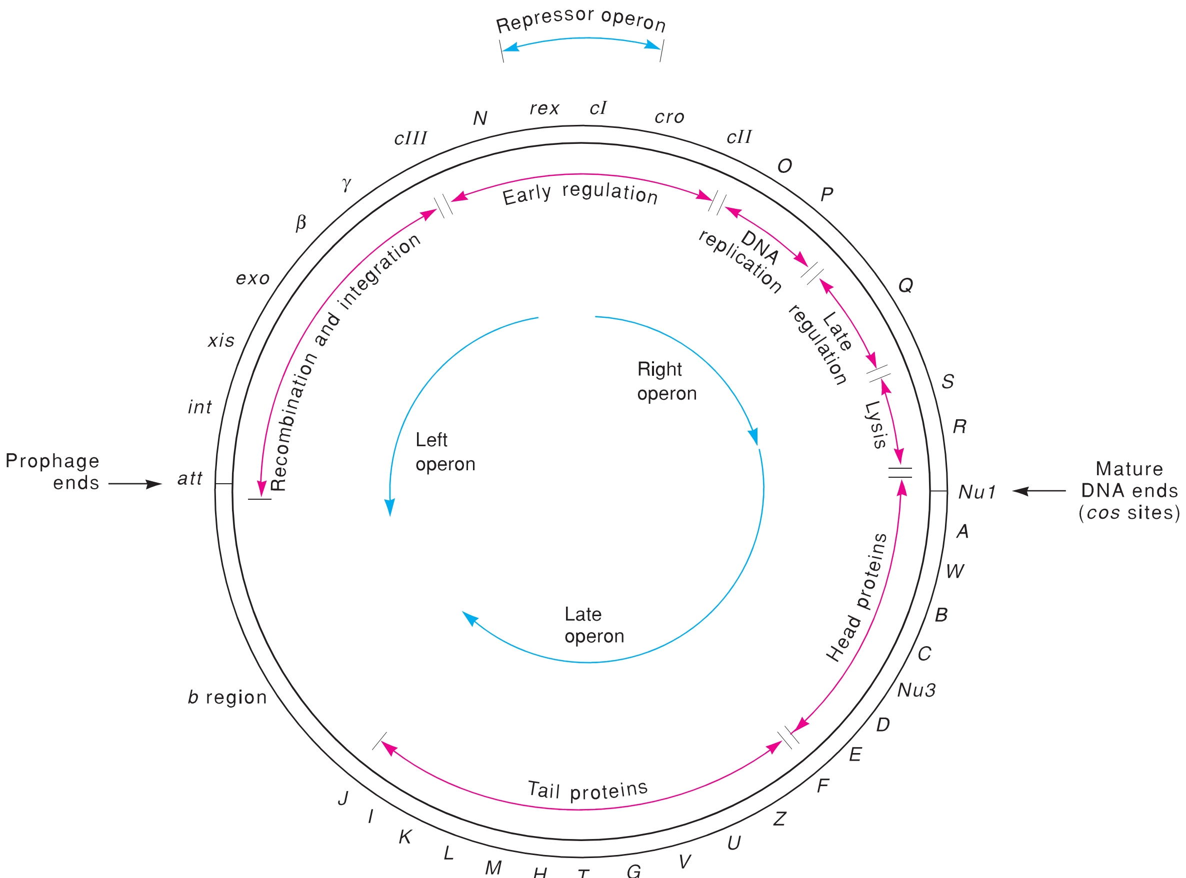


Figure 14.18 Genetic map of phage λ. There are four operons present: the repressor, left, right, and late. The prophage, a linear form integrated into the bacterial chromosome, begins and ends at *att.* The mature phage, another linear form found packed into the phage heads, begins and ends at *Nu1* (*cos* sites).

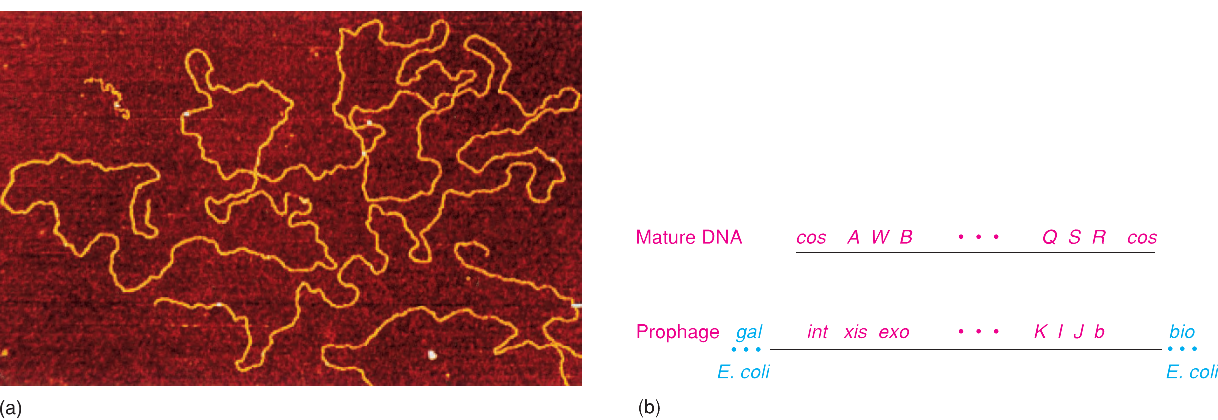


Figure 14.19 The two linear forms of λ phage. (*a*) False color electron micrograph of the λ chromosome, approximately 16 μm in length. This is the linear form of the phage chromosome found within the phage heads. (*b*) The mature linear DNA (found within phage protein coats) is flanked by *cos* sites. The prophage is flanked by *E. coli* DNA (*bio* and *gal* loci). ([*a*] Courtesy of Martin Guthold and Carlos Bustamante, Institute of Molecular Biology and HHMI, University of Oregon.)

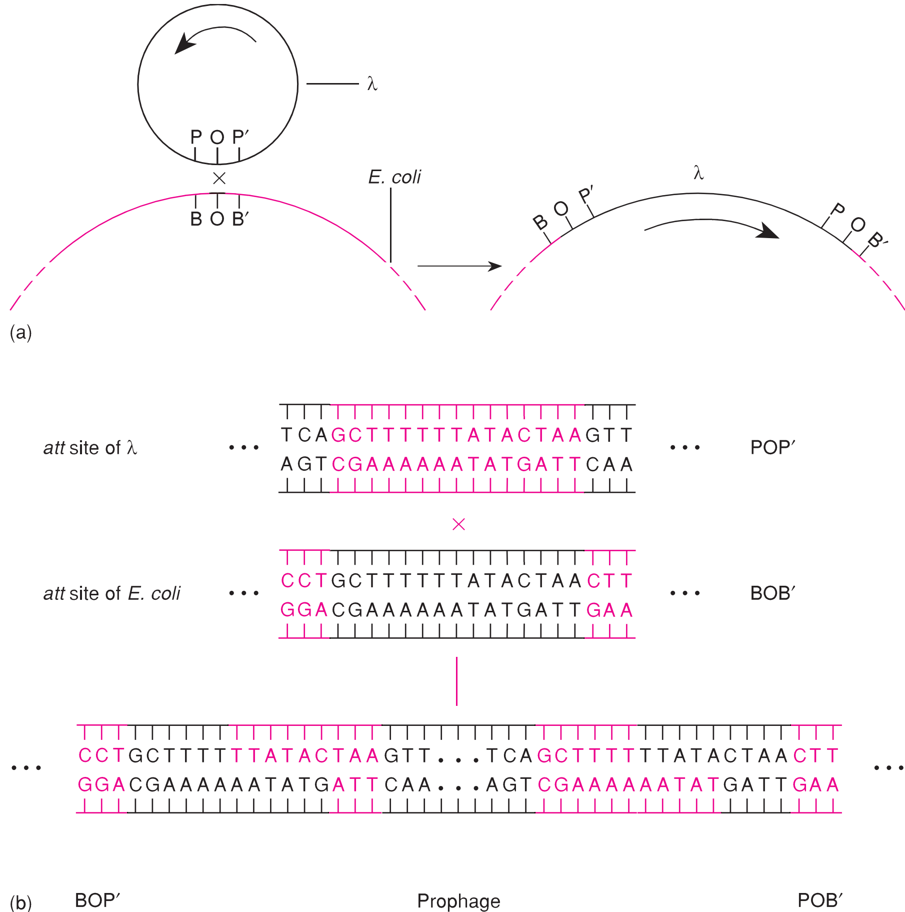


Figure 14.20 Integration of the λ phage into the *E. coli* chromosome requires a crossover between the two attach sites, called POP′ (phage) and BOB′ (bacteria). (*a*) General pattern of this site-specific attachment. (*b*) Nucleotide sequences of the various components.

the lysogenic cycle and integrating into the host chromosome. What events lead up to this "decision" on which path to take?

Early and Late Transcription

When the phage first infects an *E. coli* cell, transcription of the left and right operons begins at the left (p_L) and right (p_R) promoters, respectively: The *N* (left) and *cro* (right) genes are transcribed (fig. 14.21) and then translated into their respective proteins. Transcription then stops on both operons at rho-dependent terminators (t_{R1}, t_{L1}). Transcription cannot continue until the protein product of the *N* gene is produced. This protein is called an **antiterminator protein.** When it binds at sites upstream from the terminators, called *nutL* and *nutR* (*nut* stands for *N utilization*; *L* and *R* stand for left and right), the polymerase reads through the terminators and continues on to transcribe the left and right operons. (Although it is not completely clear why antitermination

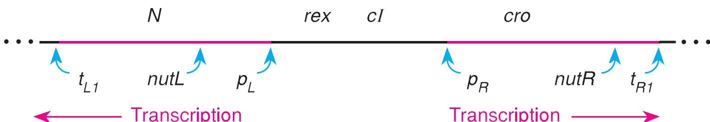


Figure 14.21 Transcription begins at the left and right promoters (p_L, p_R) and proceeds to the left and right terminators (t_{L1}, t_{R1}). Transcription continues through these terminators when the protein product of the *N* gene binds to the *nutL* and *nutR* sites.

has evolved here, it seems to give the phage better control over the timing of events.)

Transcription then continues along the left and right operons through the *cII* and *cIII* genes (see fig. 14.18). Later, if the lytic response is followed, the *Q* gene, which codes for a second antiterminator protein in the right operon, has the same effect on the late operon as the *N* gene did on the two early operons: Without the *Q*-gene product, transcription of the late operon proceeds about two hundred nucleotides and then terminates. With the *Q*-gene product, the late operon is transcribed. Hence, in phage λ, proteins that allow RNA polymerase to proceed past termination signals mediate general control of transcription. If only the previously described events were to transpire, the lytic cycle pathway would always be followed. However, a complex series of events can also take place in the repressor region that may lead to a "decision" to follow the lysogenic cycle instead.

Repressor Transcription

The *cIII*-protein product inhibits a host cell protease, called FtsH, that would break down the *cII*-gene product. The *cII*-gene product binds at two promoters, enhancing their availability to RNA polymerase, just as the CAP-cAMP product enhances the transcription of the *lac* operon. The *cII* protein binds at the promoters for *cI* transcription and for *int* transcription (fig. 14.22). At this point, the phage can still "choose" between either the lytic or the lysogenic cycles. Integrase (the product of the *int* gene) and *cI* (repressor) proteins are now produced, favoring lysogeny, as well as the *cro*-gene product, the ***antirepressor,*** which is a repressor of *cI* and therefore favors the lytic pathway. (*Cro* stands for *c*ontrol of *r*epression and *o*ther things; the *c* of *cI,* the repressor, stands for "clear," which is the appearance of λ plaques that have *cI* mutations. These mutants can only undergo lysis without the possibility of lysogeny. Normal λ infections produce turbid plaques, accounted for by lysogenic bacterial growth within the plaques.) We now focus further on the repressor region with its operators and promoters.

Maintenance of Repression

The *cI* gene, with the aid of the *cII*-gene product, is transcribed from a promoter known as p_{RE}, the *RE* standing for *r*epression *e*stablishment (fig. 14.23). Once *cI* is transcribed, it is translated into a protein called the λ ***repressor,*** which interacts at the left and right operators, o_L and o_R, of the left and right operons. When these operators are bound by *cI* protein, transcription of the left and right operons (and therefore also the late operon) ceases. There are several ramifications of the repression. First, lysogeny can be initiated because the *int* gene has been transcribed at the early stage of infection. Second, since CII and CIII are no longer being synthesized, *cI* transcription from the p_{RE} promoter stops. However, *cI* can still be transcribed because there is a second promoter, p_{RM} (*RM* stands for *r*epression *m*aintenance), that allows low levels of transcription of the *cI* gene.

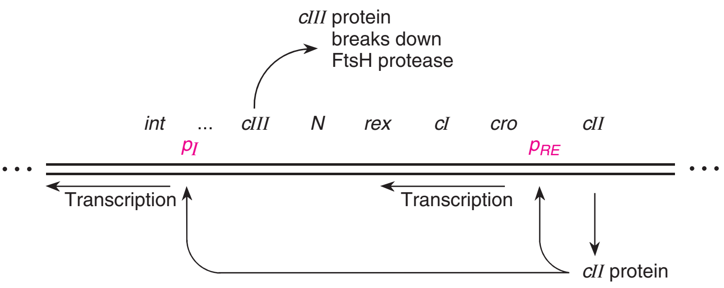


Figure 14.22 The *cII*-gene product of phage λ binds to the *cI* promoter (p_{RE}) and the *int* promoter (p_I), enhancing transcription of those genes. The *cIII* protein breaks down the FtsH protease that would normally break down the *cII* protein.

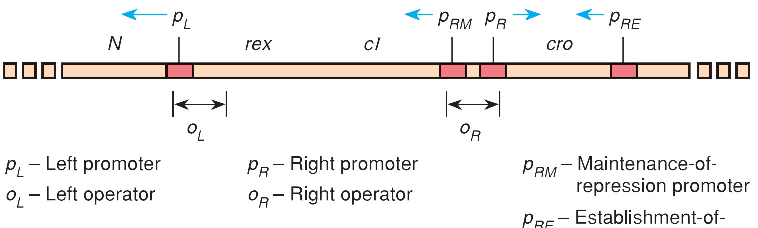


Figure 14.23 Early regulation region of phage λ. Two promoters, p_{RE} and p_{RM}, transcribe the *cI* and *rex* genes. The left operator overlaps the left promoter, and the right operator overlaps both the right promoter and the maintenance-of-repression promoter.

The *cI* gene can further control its own concentration in the cell. When the right and left operators were sequenced, each was discovered to have three sites of repressor recognition (fig. 14.24). On the right operator, for example, the right-most site (o_{R1}) was found to be most efficient at binding repressor. When repressor was bound at this site, the right operon was repressed, and transcription of *cI* was enhanced (again, in a way similar to enhancement of transcription by the binding of CAP-cAMP at the CAP site in the *lac* operon). Excess repressor, when present, however, was also bound by the other two sites within o_R. The foregoing process results in the repression of the *cI* gene itself. Hence, maintenance levels of CI can be kept within very narrow limits.

A third ramification of repression is the prevention of superinfection. That is, bacteria lysogenic for λ phage are protected from further infection by other λ phages because repressor is already present in the cell. Thus, the excess of repressor controls new invading λ phages. (We say that bacterial cells lysogenic for phage λ are immune from infection by additional λ phage.) These bacteria are also protected from infection by T4 phage with *rII* mutants. The *rex*-gene product, the product of the other gene in the repressor operon, controls this protection.

The promoters for maintenance and establishment of repression differ markedly in their control of repressor gene expression. When p_{RE} is active, a very high level of repressor is present, whereas p_{RM} produces only a low level of repressor. The level of repressor is due to the length of the leader RNA transcribed on the 5′ side of the *cI* gene. The p_{RE} promoter transcribes a very long leader RNA and is very efficient at translation of the *cI* region. In contrast, the p_{RM} promoter begins transcription at the initiation codon of the protein. This leaderless messenger RNA is translated very inefficiently into CI.

The λ repressor is a dimer of two identical subunits (fig. 14.25). Each subunit is composed of two domains, or

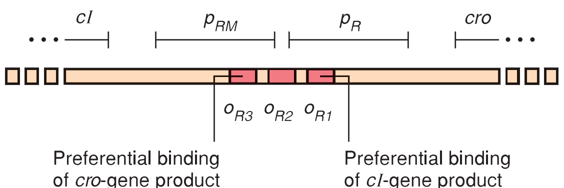


Figure 14.24 The right operator on the phage λ chromosome overlaps the p_{RM} and p_R promoters. There are three repressor recognition sites within the operator: o_{R1}, o_{R2}, and o_{R3}. Preferential binding by the Cro repressor to o_{R3} and the CI repressor to o_{R1} determines whether transcription occurs to the left or the right.

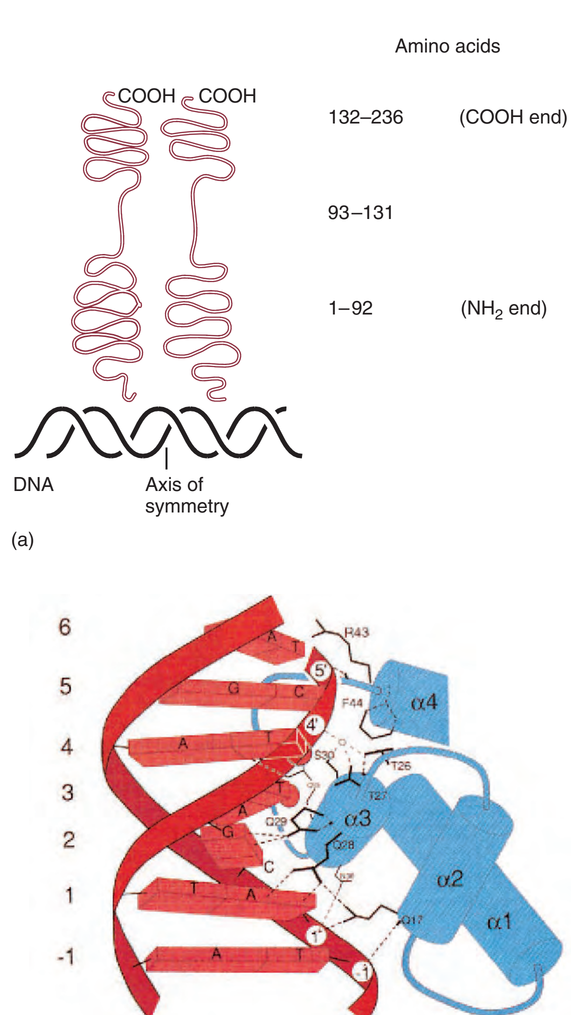


Figure 14.25 The λ repressor. (*a*) The λ repressor is a dimer with each subunit having helical amino- and carboxyl-terminal ends. The helical structure of each amino-terminal end binds in the major DNA groove. (*b*) Diagram of the interaction of amino acid residues 1–69 (*blue*) with o_{R1} (*red*) in the closely related phage 434. *Dashed lines* are hydrogen bonds. Numbers −1 to 6 and 4′ and 5′ are phosphate numbers. Amino acids are designated by the single-letter code (fig. 11.1). Small *red circles* are water molecules. (From D. W. Rodgers and S. C. Harrison, "The complex between phage 434 repressor DNA-binding domain and operator site O_{R3}: Structural differences between consensus and non-consensus half-sites," *Structure,* 1:227–40, Dec. 15, 1993. © Current Biology Ltd.)

"ends." The carboxyl- and amino-terminal ends are separated by a relatively open region, susceptible to protease attack. The alpha-helical regions of the amino-terminal ends interdigitate into the major groove of the DNA to locate the specific sequences making up the left and right operator sequences. As described earlier for the *lac* operator, o_{R1}, o_{R2}, and o_{R3} each have twofold symmetry.

The binding of the λ repressor in o_{R1} enhances the binding of another molecule of repressor into o_{R2}. Together, they enhance p_{RM} transcription, presumably through contact with RNA polymerase. The repressors also block p_R transcription (see fig. 14.24).

Lysogenic Versus Lytic Response

We have described the mechanism by which λ establishes lysogeny. How then does λ turn toward the lytic cycle? Here, control is exerted by the *cro*-gene product, another repressor molecule that works at the left and right operators in a manner antagonistic to the way the CI repressor works. In other words, using the right operator as an example, *cro*-gene product binds preferentially to the leftmost of the three sites within o_R and represses *cI* but enhances the transcription of *cro* (see fig. 14.24).

The *cro*-gene product can direct the cell toward a lytic response if it occupies the o_R and o_L sites before the λ repressor, or if the λ repressor is removed. From the point of view of phage λ, when would be a good time for the CI repressor to be removed? Thinking in evolutionary terms, we would expect that a prophage might be at an advantage if it left a host's chromosome and began the lytic cycle when it "sensed" damage to the host. In fact, one of the best ways to induce a prophage to enter the lytic cycle is to direct ultraviolet (UV) light at the host bacterium. (Actually, this was how lysogeny was discovered, by French geneticist André Lwoff.) UV light causes damage to DNA and induces several repair systems. One, called SOS repair (see chapter 12), makes use of the protein product of the *recA* gene. Among the activities of this enzyme is to cleave the λ repressor in the susceptible region between domains. The cleaved repressor falls free of the DNA, making the operator sites available for the *cro*-gene product. The lytic cycle then follows.

Initially, however, when the phage first infects an *E. coli* cell, the "decision" for lytic versus lysogenic growth is probably determined by the *cII*-gene product. This protein, as we mentioned, is susceptible to a bacterial protease, which, in turn, is an indicator of cell growth. When *E. coli* growth is limited, its proteases tend to be limited, a circumstance that would favor lysogeny for the phage. It is the *cII* protein that, when active, favors lysogeny and when inactive favors the lytic cycle. Thus, under active bacterial growth, the *cII* protein is more readily destroyed, it thus fails to enhance *cI* transcription, and lysis follows. When bacteria are not growing actively, the *cII* protein is not readily destroyed, it enhances *cI* transcription, and lysogeny results. Thus, under initial infection, the choice between lysogeny or the lytic cycle depends primarily on the *cII* protein, which gauges the health and activity of the host. After lysogeny is established, it can be reversed by processes that inactivate the *cI* protein, indicating genetic damage to the bacterium (the SOS response) or an abundance of other hosts in the environment (zygotic induction, see chapter 7). In zygotic induction, the lytic cycle is induced during conjugation, presumably when an Hfr cell sends a copy of the λ prophage into an F^- cell. At that point, without repressor present, the prophage can reassess whether to continue lysogeny or enter the lytic cycle.

Not all the details regarding the CI-Cro competition are known, but an understanding of the relationship of lytic and lysogenic life cycles and the nature of DNA-protein recognition has emerged (fig. 14.26 and table 14.1).

Table 14.1 Elements in Phage λ Infection

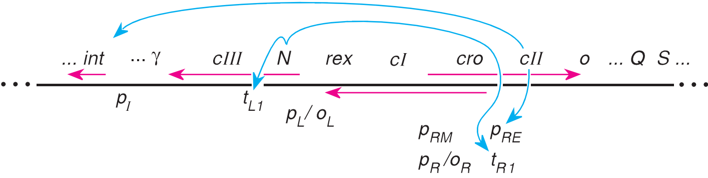
Gene Products	
cI	Repressor protein whose function favors lysogeny
cII	Enhances transcription at the p_I and p_{RE} promoters
cIII	Inhibits the *FtsH* protease
cro	Antirepressor protein that favors lytic cycle
N	Antiterminator acting at *nutR* and *nutL*
rex	Protects bacterium from infection by T4 *rII* mutants
int	Integrase for prophage integration
Q	Antiterminator of late operon
FtsH	Bacterial protease that degrades *cII* protein
Promoters of	
p_R	Right operon
p_L	Left operon
p_{RE}	Establishment of repression at repressor region
p_{RM}	Maintenance of repression at repressor region
p_R	Late operon
p_I	*int* gene
Terminators	
t_{R1}	Terminates after *cro* gene
t_{L1}	Terminates after *N* gene
Antiterminators	
nutR	In *cro* gene
nutL	In *N* gene

(1) Initial infection. Transcription from p_R and p_L through *cro* and *N*. Termination at t_{R1} and t_{L1}.

... *int* ... γ *cIII* *N* *rex* *cI* *cro* *cII* *o* ... *Q S* ...

p_I t_{L1} p_L/o_L p_{RM} p_{RE} p_R/o_R t_{R1}

(2) *N* protein allows antitermination at t_{L1} and t_{R1}. Transcription continues through *cII* and *cIII*. Protein product of *cII* allows transcription at p_I and p_{RE}.



(3) Repressor and antirepressor (*cI*– and *cro*-gene products) compete for o_R and o_L sites.

Lytic growth

Antirepressor (*cro* protein) gains access to o_{R3}, o_{R2}, o_{L3}, and o_{L2}. Right, left, and late operons transcribed. Repressor region (*cI*, *rex*) repressed.

Lysogeny

Repressor (*cI* protein) gains access to o_{R1}, o_{R2}, o_{L1}, and o_{L2}. Right, left, and late operons repressed. Transcription at p_{RM} enhanced.

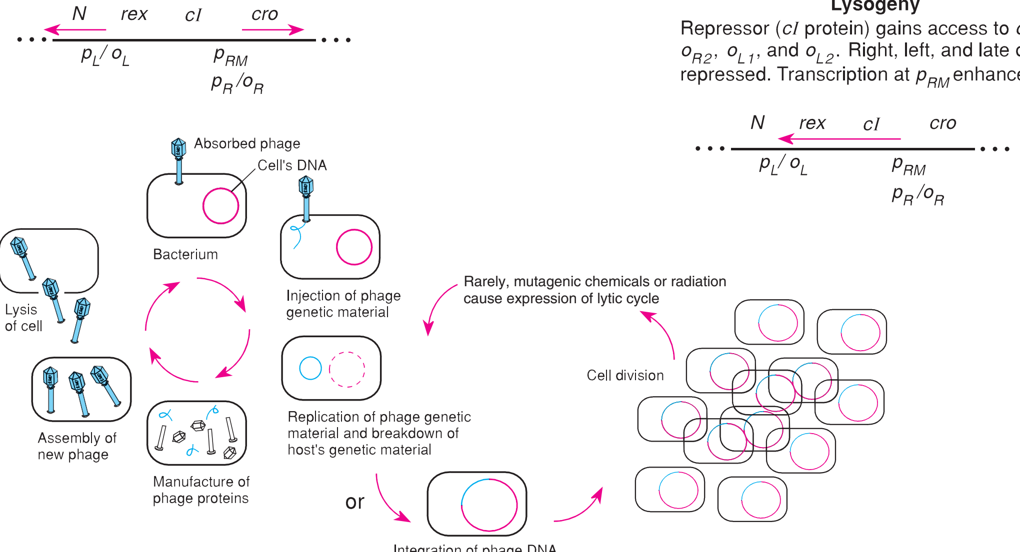


Figure 14.26 Summary of regulation of phage λ life cycles. (*1*) In the initial infection, transcription begins in *cro* and *N* but terminates shortly thereafter at left and right terminators. (*2*) The product of the *N* gene allows transcription through the initial terminators; in essence, all genes can now be transcribed. (*3*) Lysogeny will occur if the *cI* protein gains access to the right and left operators; the lytic cycle will prevail if the *cro*-gene product gains access to those two operators.

TRANSPOSABLE GENETIC ELEMENTS

Up until this point, we have thought of the genome in fairly conservative terms. If we map a gene today, we expect to see it in the same place tomorrow. However, our discovery of mobile genetic elements has modified that view to some extent. We now know that some segments of the genome can move readily from one place to another. Their moving can have an effect on the phenotype of the organism, primarily at the transcriptional level. We thus begin our discussion of mobile genetic elements here, and we conclude it in chapter 16, because mobile elements also affect the phenotypes of eukaryotes.

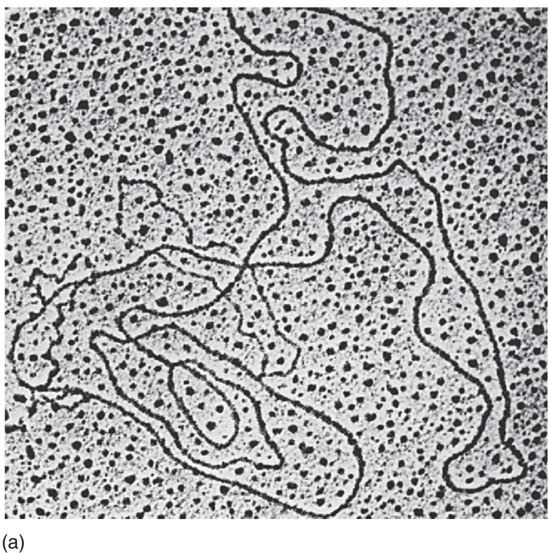
IS Elements

Transposable genetic elements, transposons, or even *jumping genes,* are regions of the genome that can move from one place to another. In some cases, transposition is *conservative:* the transposons move without copying themselves. They are liberated from the donor site by double-strand breaks in the DNA. In other cases, transposition is *replicative:* a copy of the transposon is inserted while the original stays in place. This mechanism involves only single-strand breaks of the DNA at the donor site.

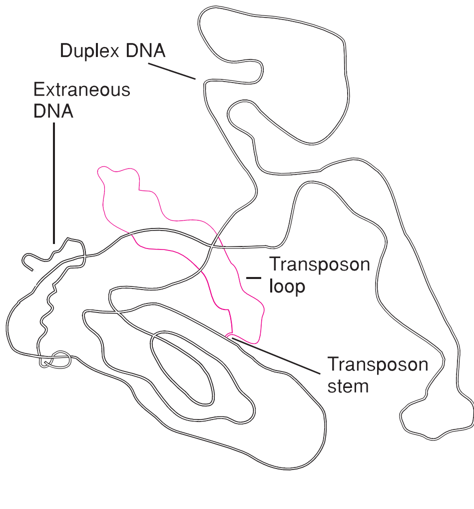
Barbara McClintock first discovered transposable elements in corn in the 1940s (see chapter 16); they were discovered in prokaryotes in 1967, where they first showed up as **polar mutants** in the galactose operon of *E. coli.* No genes of the operon were expressed past the point of the polar mutation. This effect was explained by assuming that the transposon brought with it a transcription stop signal. The presence of an inserted piece of DNA in these polar mutants was verified by heteroduplex analysis (fig. 14.27).

The first transposable elements discovered in bacteria were called **insertion sequences** or **IS elements.** It turns out that these are the simplest transposons. The IS elements consist of a central region of about 700 to 1,500 base pairs surrounded by an inverted repeat of about 10 to 30 base pairs, the numbers depending on the specific IS element. Presumably, the inverted repeats signal the transposing enzyme that it is at the ends of the IS element. The central region of the IS element contains a gene or genes for the transposing event (usually genes for transposase and resolvase enzymes); the relatively small IS elements carry no bacterial genes (fig. 14.28).

The target site that the transposable element moves to is not a specific sequence, as with the *att* site of λ. It becomes a direct repeat flanking the IS element only after insertion, giving rise to a model of insertion (fig. 14.29). The target site is cut in a staggered fashion, leaving single-stranded ends. The IS element is then inserted between the single-stranded ends. Repair processes convert the two single-stranded tails to double-stranded segments and, hence, to direct flanking repeats. When DNA is sequenced, the pattern of a direct flanking repeat surrounding an inverted repeat, with a segment in the middle, signals the existence of a transposable element. Currently, we know of more than fifteen families, including a total of over five hundred known members, of IS elements.



(a)



(b)

Figure 14.27 Heteroduplex analysis revealing a transposon. (*a*) Two plasmids were hybridized, one with and one without a transposon. (*b*) The transposon is seen as a single-stranded loop (*red*); it has nothing to pair with in the heteroduplex. ([*a*] Courtesy of Richard P. Novick, M.D.)

Chromosomal target site

— TTAG —
— AATC —

Transposon (IS*5*)

GGAAGGTGCGAATAAG ... CTTGTTCGCACCTTCC
CCTTCCACGCTTATTC ... GAACAAGCGTGGAAGG

Inverted repeats

Insertion

... — TTAGGGAAGGTGCGAATAAG ... CTTGTTCGCACCTTCCTTAG — ...
— AATCCCTTCCACGCTTATTC ... GAACAAGCGTGGAAGGAATC —

Inverted repeats

Direct repeats

Figure 14.28 An IS element (IS*5*) inserted into a target site in a bacterial chromosome creates a direct repeat on either side of the IS element. The explanation is shown in figure 14.29. An inverted repeat (*red*) is seen as the same sequence read inwards from outside on the upper strand (*left*) and the lower strand (*right*). (Reprinted with permission from *Nature,* Vol. 297, M. Kroger and G. Hoborn, "Structural Analysis of Insertion Sequence IS5." Copyright © 1982 Macmillan Magazines Limited.)

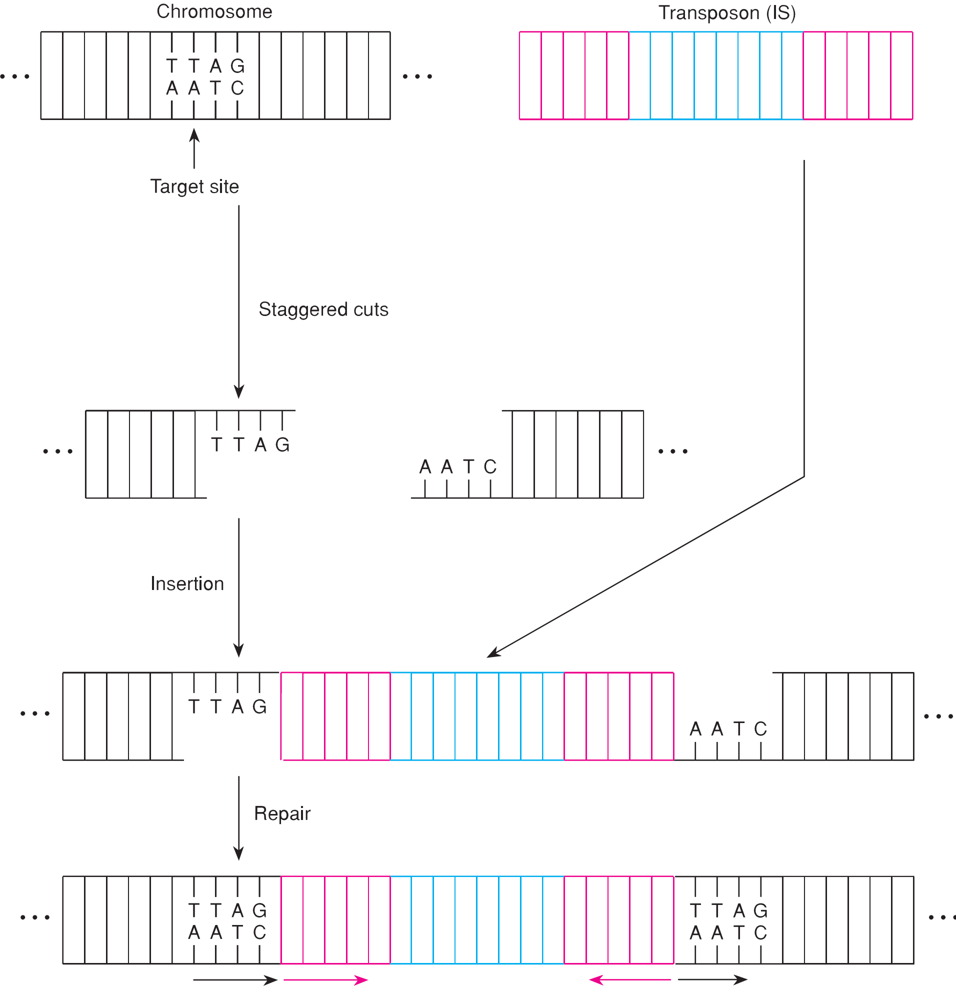


Figure 14.29 Insertion of an IS element (IS*5* of fig. 14.28) results in a direct flanking repeat surrounding the transposon in the host chromosome. This occurs because the insertion takes place at a point in which a staggered cut is made in the host DNA, leaving complementary regions on either side of the transposon. Repair replication results in two copies of the flanking sequence.

Composite Transposons

After the discovery of IS elements, a more complex type of transposable element, a **composite transposon,** was discovered. A composite transposon consists of a central region surrounded by two IS elements. The central region usually contains bacterial genes, frequently antibiotic resistance loci. For example the composite transposon Tn*10* contains the genes for transposase and resolvase, as well as the bacterial gene for β-lactamase, which confers resistance to ampicillin (fig. 14.30). Arrangements of composite transposons can vary quite a bit. The IS elements at the two ends can be identical or different; they can be in the same or different orientations; they can be similar to known IS elements or different from any freely existing IS elements. In the latter case, they are called IS-like elements.

Two IS elements can transpose virtually any region between them. In fact, composite transposons most likely came into being when two IS elements became located near each other. We can see this very clearly in a simple experiment. In figure 14.31, there is a small plasmid constructed with transposon Tn*10* in it. The "reverse" transposon, consisting of the two IS elements and the plasmid genes, or the normal transposon, could each transpose.

Mechanism of Transposition

Transposition can come about by several mechanisms; however, it does not use the normal recombination machinery of the cell (see chapter 12). One model, by J. Shapiro, explains the fact that many transposons in the

J. A. Shapiro (1943–). (Courtesy of Dr. J. A. Shapiro.)

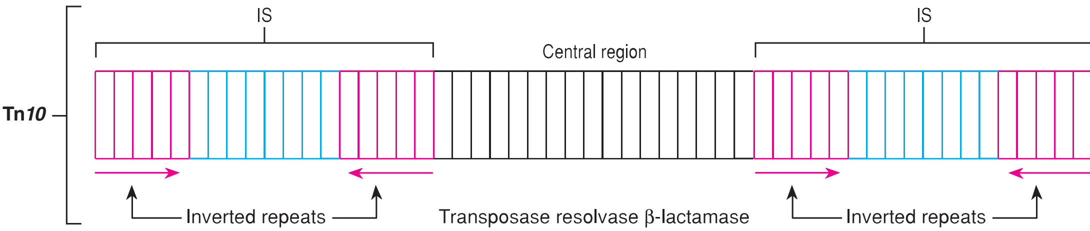


Figure 14.30 A composite transposon consists of a central region flanked by two IS elements. Transposon Tn*10* contains the transposase and resolvase enzyme genes as well as the bacterial gene β-lactamase, which protects the cell from the antibiotic ampicillin.

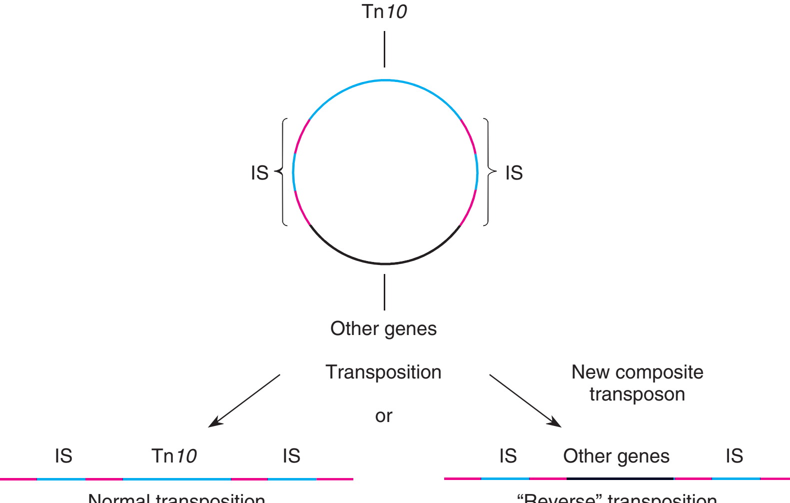


Figure 14.31 Two IS elements in a plasmid can transpose either of the two regions between them. In the case shown, either the Tn*10* transposon or the "reverse" transposon ("other genes") is transposed.

process of transposition go through a **cointegrate** state (fig. 14.32), in which there is a fusion of two elements. During the process of transposition (in this case from one plasmid to another), an intermediate cointegrate stage is formed, made up of both plasmids and two copies of the transposon. Then, through a process called *resolution,* the cointegrate is reduced back to the two original plasmids, each now containing a copy of the transposon.

A diagram of Shapiro's mechanism appears in figure 14.33. At first, staggered cuts are made in the donor and recipient DNA molecules (fig. 14.33*a* and *b*). Then nonhomologous ends are joined so that only one strand of the transposon connects them (fig. 14.33*c*). This process is presumably controlled by the transposon-coded *transposase* enzyme. Repair-DNA replication now takes place to fill in the single-stranded segments. The result is a cointegrate of the two plasmids with two copies of the transposon. The last step is a recombination event at a homologous site within the two transposons. This is catalyzed by a *resolvase* enzyme, which resolves the cointegrate into the original two plasmids, each with a copy of the transposon (fig. 14.33*e*).

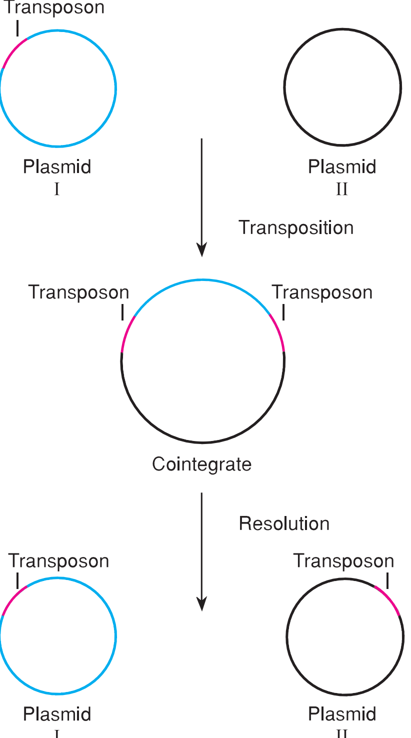


Figure 14.32 Transposition frequently goes through an intermediate cointegrate stage. In this case, the transposon is copied from one plasmid to another, with an intermediate stage consisting of a single large plasmid.

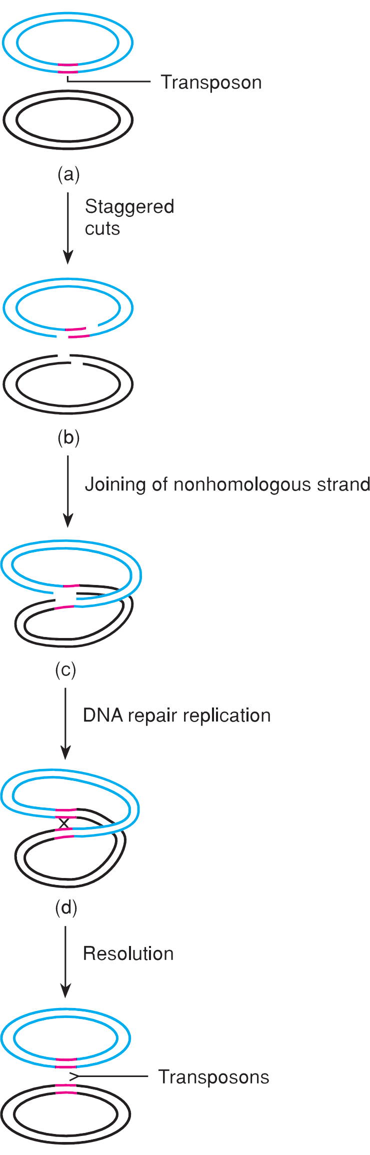
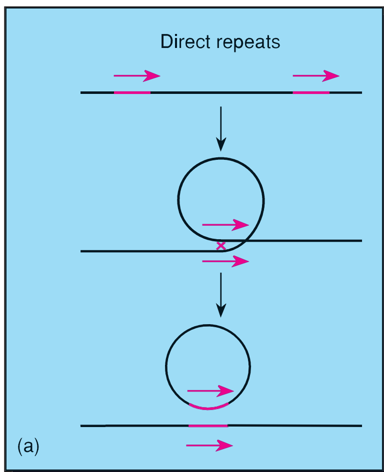


Figure 14.33 The Shapiro mechanism of transposition. Staggered cuts are made at the site of transposon insertion and at either side of the transposon itself (*a* and *b*). Nonhomologous single strands join, resulting in two single-stranded copies of the transposon in the cointegrate (*c*). Repair replication of these single strands produces two copies of the transposon (*d*). A crossover at the transposon resolves the cointegrate into two plasmids, each with a copy of the transposon (*e*).

Phenotypic and Genotypic Effects of Transposition

Transposition can have several effects on the phenotype and genotype of an organism. If transposition takes place into a gene or its promoter, it can disrupt the expression of that gene. Depending on the orientation of a transposon, it can prevent the expression of genes. A transposon can also cause deletions and inversions.

Direct repeats on a chromosome can come about, for example, by the sequential transposition of the same IS or transposon, in the same orientation. Pairing followed by recombination results in a deletion of the section between the repeats (fig. 14.34). In the case of inverted repeats, pairing followed by recombination results in an inversion of the section between the repeats.



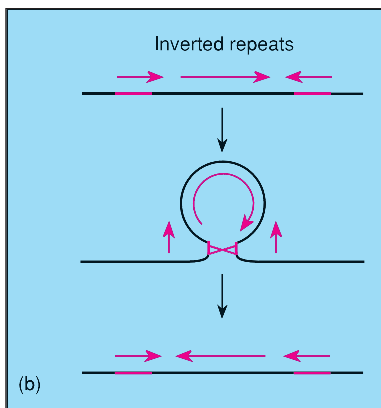


Figure 14.34 Pairing and recombination in DNA repeats. (*a*) Direct repeats can result in deletion (in the form of a *circle*) due to a single crossover. (*b*) Inverted repeats can result in an inversion of the region between the repeats due to a crossover.

A well-known case of transposon orientation controlling a phenotype in bacteria occurs in *Salmonella typhimurium*. The flagella of this bacterium occur in two types. Any particular bacterium has either type 1 or type 2 flagella (called phase 1 or phase 2 flagella). The difference is in the flagellin protein the flagella are composed of. Phase 1 flagella are determined by the *H1* gene and phase 2 flagella are determined by the *H2* gene. The change from one phase to another occurs at a rate of about 10^{-4} per cell division. After extensive genetic analysis, the following scheme was suggested and later verified using recombinant DNA techniques.

The *H1* and *H2* genes are at separate locations on the bacterial chromosome (fig. 14.35). *H2* is part of an operon that also contains the *rH1* gene, the repressor of *H1*. The promoter of this operon lies within a transposon upstream of the operon. When the promoter is in the proper orientation, the *H2* operon is expressed, resulting in phase 2 flagella. The *rH1* gene product represses the *H1* gene (fig. 14.35*a*). If the inverted repeat ends of the transposon undergo recombination, the transposon is inverted (see fig. 14.34), moving the promoter into an incorrect orientation for the transcription of the *H2* operon. No *H1* repressor is made, so the *H1* gene is expressed (fig. 14.35*b*).

As N. Kleckner has summarized, transposons can have marked effects on the phenotype by their actions in transposition and by the fact that they may carry genes valuable to the cell. However, they can also exist without any noticeable consequences. This fact has led some evolutionary geneticists to suggest that transposons are an evolutionary accident that, once created, are self-maintaining. Since they may exist without a noticeable benefit to the host's phenotype, transposons have been

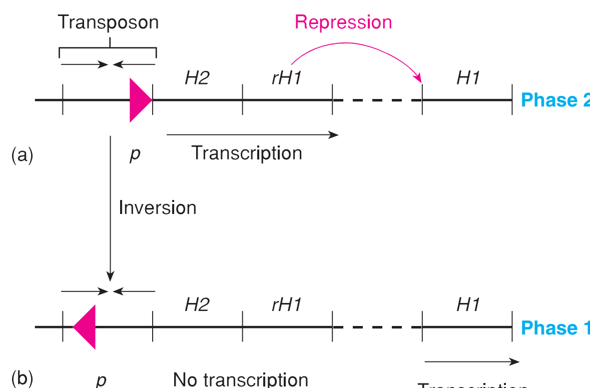


Figure 14.35 Arrangement of flagellin genes on the *Salmonella* chromosome. The promoter (*p*) is within a transposon. In one orientation (*a*), the *H2* operon is transcribed, which results in *H2* flagellin and *rH1* protein, the repressor of the *H1* gene. In the second orientation (*b*), the *H2* operon is not transcribed, resulting in uninhibited transcription of the *H1* gene.

Nancy Kleckner (1947–). (Courtesy of Nancy Kleckner. Photo by Stu Rosner.)

referred to as **selfish DNA.** In recent theoretical and experimental studies, however, some scientists have suggested that transposons improve the evolutionary fitness of the bacteria that have them (see chapter 20).

OTHER TRANSCRIPTIONAL CONTROL SYSTEMS

Transcription Factors

Phage T4

Phage T4, a relatively large phage with seventy-three genes, has transcription controlled by particular RNA polymerase specificity factors. Like phage λ, phage T4 has early, middle, and late genes, genes that need to be expressed in a particular order. Early T4 genes have promoters whose specificity of recognition depends on the sigma factor of the host (σ^{70} of *E. coli*). Two products of early phage genes are the AsiA and MotA proteins. AsiA binds to the −35 recognition region of σ^{70}, preventing transcription from both host genes and early phage genes. AsiA is thus called an **anti-sigma factor,** a protein that interferes with a sigma factor. Middle phage promoters do not have −35 recognition regions but do bind MotA. Host RNA polymerase bound with the σ^{70}-AsiA complex recognizes these promoters. Finally, late phage genes have promoters that depend on the phage-encoded sigma factor σ^{gp55}. Some proteins are needed both early and late in the infection process; they are specified by genes that have promoters recognized by different specificity factors.

Heat Shock Proteins

A response to elevated temperature, found in both prokaryotes and eukaryotes, is the production of heat shock proteins (see chapter 10). In *E. coli,* elevated temperatures cause the general shutdown of protein synthesis concomitant with the appearance of at least seventeen heat shock proteins. These proteins help protect the cell against the consequences of elevated temperature; some are molecular chaperones (see chapter 11). The production of these proteins is the direct result of the gene product of the *htpR* gene, which codes for σ^{32}. The normal sigma factor is σ^{70}, the product of the *rpoD* gene; the heat shock genes have promoters recognized by σ^{32} rather than σ^{70}. Heat causes the *htpR* gene to become active, as well as stabilizing σ^{32}. From DNA sequence data, the difference in promoters between normal genes and heat shock genes seems to lie in the −10 consensus sequence (Pribnow box). In normal genes, it is TATAAT; in heat shock protein genes, it is CCCCATXT, in which *X* is any base.

Promoter Efficiency

In addition to the mechanisms previously described, there are other ways to regulate the transcription of messenger RNA. One is to control the efficiency of various processes. For example, we know that the promoter sequence of different genes in *E. coli* differs. Since the affinity for RNA polymerase is different for the different sequences, the rate of initiation of transcription for these genes also varies. The more efficient promoters are transcribed at a greater rate than the less efficient promoters. An example is the promoter of the *i* gene of the *lac* operon. This promoter is for a constitutive gene that usually produces only about one messenger RNA per cell cycle. However, mutants of the promoter sequence are known that produce up to fifty messenger RNAs per cell cycle. Here, then, the transcriptional rate is controlled by the efficiency of the promoter in binding RNA polymerase. Efficiency can be controlled by the direct sequence of nucleotides (i.e., differences from the consensus sequence) or by the distance between consensus regions. For example, promoters vary in the number of bases between the −35 and −10 sequences. Seventeen seems to be the optimal number of bases separating the two. Presumably, more or fewer than seventeen reduces the efficiency of transcription.

TRANSLATIONAL CONTROL

When considering control of gene expression, it is important to remember that all control mechanisms are aimed at exerting an influence on either the amount, or the activity, of the gene product. Therefore, in addition to transcriptional controls, which influence the amount of messenger RNA produced, there are also translational controls affecting how efficiently the messenger RNA is translated. (Attenuator control—see fig. 14.16—can also be viewed as translational control because the environment is tested by translation even though attenuation results in the cessation of transcription.) In prokaryotes, translational control is of lesser importance than transcriptional control for two reasons. First, messenger RNAs are extremely unstable; with a lifetime of only

about two minutes, there is little room for controlling the rates of translation of existing messenger RNAs because they simply do not last very long. Second, although there are some indications of translational control in prokaryotes, such control is inefficient—energy is wasted synthesizing messenger RNAs that may never be used.

Translational control can be exerted on a gene if the gene occurs distally from the promoter in a polycistronic operon. The genes that are transcribed last appear to be translated at a lower rate than the genes transcribed first. The three *lac* operon genes, for instance, are translated roughly in a ratio of 10:5:2. This ratio is due to the polarity of the translation process. That is, in prokaryotes, translation is directly tied to transcription—a messenger RNA can have ribosomes attached to it well before transcription ends. Thus, genes at the beginning of the operon are available for translation before genes at the end. In addition, exonucleases seem to degrade messenger RNA more efficiently from the 3′ end. Presumably, natural selection has ordered the genes within operons so that those producing enzymes needed in greater quantities will be at the beginning of an operon.

Translation can also be regulated by RNA-RNA hybridization. RNA complementary to the 5′ end of a messenger RNA can prevent the translation of that messenger RNA. The regulating RNA is called **antisense RNA.** In figure 14.36, the messenger RNA from the *ompF* gene in *E. coli* is prevented from being translated by complementary base pair binding with an antisense RNA called *micF* RNA (*mic* stands for *m*RNA-*i*nterfering *c*omplementary RNA). The *ompF* gene codes for a membrane component called a *porin,* which, as the name suggests, provides pores in the cell membrane for transport of materials. Surprisingly, a second porin gene, *ompC,* seems to be the source of the *micF* RNA. Transcription of the opposite DNA strand (the one not normally transcribed) near the promoter of the *ompC* gene yields the antisense RNA. One porin gene thus seems to regulate the expression of another porin gene, for reasons that are not completely understood. Antisense RNA has also been implicated in such phenomena as the control of plasmid number and the control of transposon Tn*10* transposition. Control by antisense RNA is a fertile field for gene therapy because antisense RNA can be artificially synthesized and then injected into eukaryotic cells.

A third translational control mechanism consists of the efficiency with which the messenger RNA binds to the ribosome. This is related to some extent to the sequence of nucleotides at the 5′ end of the messenger RNA that is complementary to the 3′ end of the 16S ribosomal RNA segment in the ribosome (the Shine-Dalgarno sequence). Variations from the consensus sequences demonstrate different efficiencies of binding and, therefore, the initiation of translation occurs at different rates.

The redundancy in the genetic code can also play a part in translational control of some proteins since different transfer RNAs occur in the cell in different quantities. Genes with abundant protein products may have codons that specify the more common transfer RNAs, a concept called **codon preference.** In other words, certain codons are preferred; they specify transfer RNAs that are abundant. Genes that code for proteins not needed in abundance could have several codons specifying the rarer transfer RNAs, which would slow down the rate of translation for these genes. The codon distribution of the phage MS2 in table 14.2 shows that every codon is used except the UGA stop codon. (The numbers in the table refer to the incidence of a particular codon in the phage genome.) However, the distribution is not random for all amino acids. For example, the amino acid glycine has two common codons and two rarer codons. The same holds for arginine but not, for example, valine.

Finally, translational control can be exerted at the ribosome. When an uncharged transfer RNA finds its way into

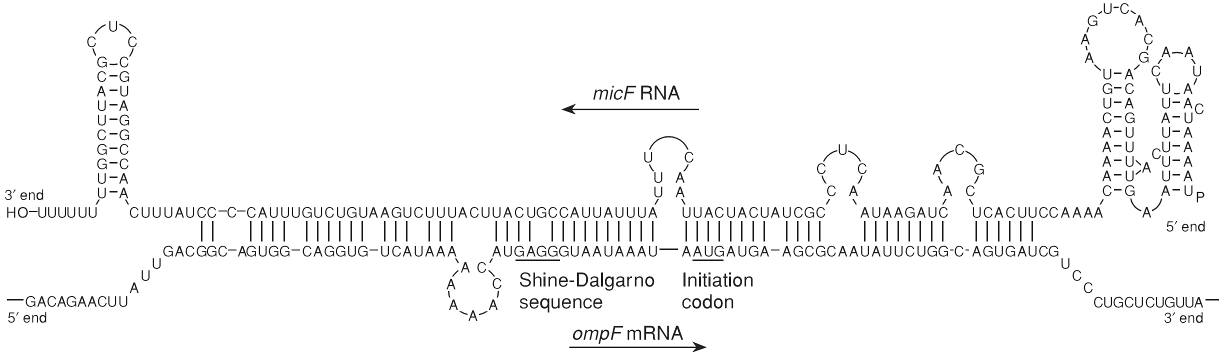


Figure 14.36 Complementarity between the RNA of the *ompF* gene and antisense RNA, or *micF* RNA. The region of overlap includes the Shine-Dalgarno sequence and the initiation codon, effectively preventing ribosome binding and translation of the *ompF* RNA. Notice the stem-loops on each side of the overlap. (Reproduced, with permission, from the *Annual Review of Biochemistry,* Volume 55, ©1986 by Annual Reviews, Inc.)

the A site of the ribosome, a likely event under amino acid starvation, it causes an **idling reaction** in the ribosome, which entails the production of the nucleotide guanosine tetraphosphate (5′-ppGpp-3′; fig. 14.37). This is part of a control mechanism called the **stringent response.** A protein called the **stringent factor,** the product of the *relA* gene, produces guanosine tetraphosphate (ppGpp), originally called *magic spot* because of its sudden appearance on chromatograms. (The gene is called *rel* from the **relaxed mutant,** which does not have the stringent response.) The stringent factor is associated with the ribosome, where ppGpp is synthesized, although it is not one of the structural proteins of the ribosome. The SpoT protein, the product of the *spoT1* gene, breaks down ppGpp; thus, the concentration of ppGpp is regulated.

The ppGpp interacts with RNA polymerase, causing an almost complete cessation of the transcription of ribosomal RNA; thus, no energy is wasted synthesizing ribosomes when translation is not possible. However, many amino acid-synthesizing operons require ppGpp for transcription; ppGpp thus inhibits ribosomal RNA production and enhances the production of enzymes to synthesize amino acids, all when the cell is starved for amino acids.

One other thing a ribosome can do when faced with amino acid shortages is to slide past *hungry* codons, codons for which a charged transfer RNA is not available. At that point, the growing peptide chain will be attached to the last charged transfer RNA to enter the ribosome, the one now in the peptidyl (P) site. The complex of the transfer RNA and the ribosome slides down the messenger RNA until it encounters the next codon for the transfer RNA. At this point, it is hoped, the next codon en-

Figure 14.37 The idling reaction. The stringent factor catalyzes the conversion of GDP to 5′-ppGpp-3′. The added pyrophosphate groups come from ATP.

Table 14.2 Codon Distribution in MS2, an RNA Virus

	Second Position								
First Position	**U**		**C**		**A**		**G**		**Third Position**
U	Phe	10	Ser	13	Tyr	8	Cys	7	**U**
	Phe	13	Ser	10	Tyr	13	Cys	4	**C**
	Leu	11	Ser	10	*stop*	1	*stop*	0	**A**
	Leu	4	Ser	13	*stop*	1	Trp	14	**G**
C	Leu	10	Pro	7	His	4	Arg	13	**U**
	Leu	14	Pro	3	His	4	Arg	11	**C**
	Leu	13	Pro	6	Gln	10	Arg	6	**A**
	Leu	6	Pro	5	Gln	16	Arg	4	**G**
A	Ile	8	Thr	14	Asn	11	Ser	4	**U**
	Ile	16	Thr	10	Asn	23	Ser	8	**C**
	Ile	7	Thr	8	Lys	12	Arg	8	**A**
	Met	15	Thr	5	Lys	17	Arg	6	**G**
G	Val	13	Ala	19	Asp	18	Gly	17	**U**
	Val	12	Ala	12	Asp	11	Gly	11	**C**
	Val	11	Ala	14	Glu	9	Gly	4	**A**
	Val	10	Ala	8	Glu	14	Gly	4	**G**

countered in the aminoacyl (A) site will code for a charged transfer RNA present.

POSTTRANSLATIONAL CONTROL

Feedback Inhibition

Even after a gene has been transcribed and the messenger RNA translated, a cell can still exert some control over the functioning of the enzymes produced if the enzymes are allosteric proteins. We have discussed the activation and deactivation of operon repressors (e.g., *lac, trp*) owing to their allosteric properties. Similar effects occur with other proteins. The need for posttranslational control is apparent because of the relative longevity of proteins as compared with RNA. When an operon is repressed, it no longer transcribes messenger RNA; however, the messenger RNA that was previously transcribed has been translated into protein, and this protein is still functioning. Thus, during operon repression, it would also be efficient for the cell to control the activity of existing proteins.

An example of posttranslational control occurs with the enzyme aspartate transcarbamylase, which catalyzes the first step in the pathway of pyrimidine biosynthesis in *E. coli* (fig. 14.38). An excess of one of the end products of the pathway, cytidine triphosphate (CTP), inhibits the functioning of aspartate transcarbamylase. This method of control is called **feedback inhibition** because a product of the pathway is the agent that turns the pathway off.

Aspartate transcarbamylase is an allosteric enzyme. Its active site is responsible for the condensation of carbamyl phosphate and L-aspartate (fig. 14.38). However, it also has regulatory sites that have an affinity for CTP. When CTP is bound in a regulatory site, the conformation of the enzyme changes, and the enzyme has a lowered affinity for its normal substrates; recognition of CTP inhibits the condensation reaction the enzyme normally carries out (fig. 14.39). Thus, allosteric enzymes provide a mechanism for control of protein function after the protein has been synthesized.

Protein Degradation

A final control affecting the amount of gene product in a cell is control of the rate at which proteins degrade. The normal life spans of proteins vary greatly. For example, some proteins last longer than a cell cycle, whereas others may be broken down in minutes. Several models have been suggested for control of protein degradation, including the **N-end rule** and the **PEST hypothesis.**

According to the N-end rule, the amino acid at the amino-, or N-terminal, end of a protein is a signal to proteases that control the average length of life of a protein. In recent experiments, the life span of the β-galactosidase protein was determined with almost complete predictability based on its modified N-terminal amino acid. Protein life spans range from two minutes for those with N-terminal arginine to greater than twenty hours for those with N-terminal methionine or five other amino acids (table 14.3).

According to the PEST hypothesis, protein degradation is determined by regions rich in one of four amino acids: proline, glutamic acid, serine, and threonine. (The one-letter abbreviations of these four amino acids are P, E, S, and T, respectively.) Proteins that have these regions tend to degrade in less than 2 hours. In one study of thirty-five proteins with half-lives of between 20 and 220 hours, only three contained a PEST region. We see that not only are different proteins programmed to survive for varying lengths of time in the cell, but that programming seems to be based on the N-terminal amino acid as well as various regions rich in the PEST amino acids.

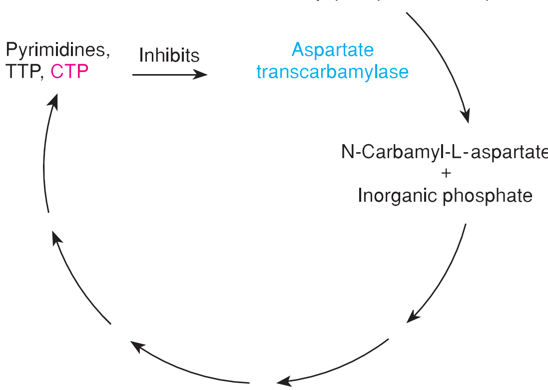


Figure 14.38 Aspartate transcarbamylase catalyzes the first step in pyrimidine biosynthesis. An end product, cytidine triphosphate (CTP), inhibits the enzyme.

Table 14.3 Relationship Between N-Terminal Amino Acid and Half-Life of *E. coli* β-Galactosidase Proteins with Modified N-Terminal Amino Acids

N-Terminal Amino Acid	Half-Life
Met, Ser, Ala, Thr, Val, Gly	>20 hours
Ile, Glu	30 minutes
Tyr, Gln	10 minutes
Pro	7 minutes
Phe, Leu, Asp, Lys	3 minutes
Arg	2 minutes

Source: Data from Bachmair, et al., *Science,* 234:179–86, 1986.

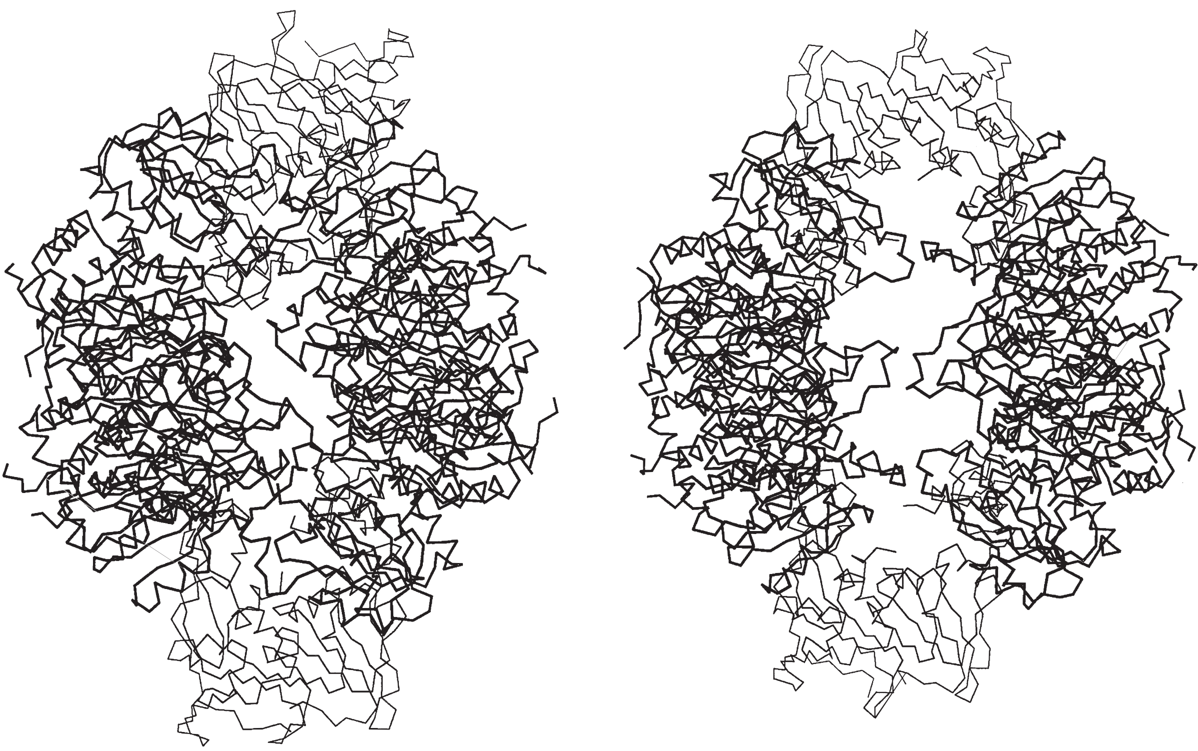


Figure 14.39 Aspartate transcarbamylase. *Left:* the enzyme bound with cytidine triphosphate (CTP). *Right:* the enzyme without CTP. Notice the difference in shape with and without CTP. With CTP, the enzyme literally closes up. (Courtesy of Evan R. Kantrowitz.)

SUMMARY

STUDY OBJECTIVE 1: To study the way in which inducible and repressible operons work 406–414

Most bacterial genes are organized into operons, which can either be repressed or induced. Transcription begins in inducible operons, such as *lac,* when the metabolite that the operon enzymes act upon appears in the environment. The metabolite (or a derivative), the inducer, combines with the repressor (the product of the independent regulator gene) and renders the repressor nonfunctional. In the absence of the inducer, the repressor binds to the operator, a segment between the promoter and the first gene of the operon. When in place, the repressor blocks transcription. After combining with the inducer, the repressor diffuses from the operator, and transcription proceeds.

All operons responsible for the breakdown of sugars in *E. coli* are inducible. In the presence of glucose, other inducible sugar operons (such as the arabinose and galactose operons) are repressed, even if their sugars appear in the environment. This process is called *catabolite repression.* Cyclic AMP and a catabolite activator protein (CAP) enhance the transcription of the nonglucose sugar operons. Glucose lowers the level of cyclic AMP in the cell and thus prevents the enhancement of transcription of these other operons.

Repressible operons, such as the *trp* operon in *E. coli,* have the same basic traits as an inducible operon—polycistronic transcription controlled by an operator site between the promoter and the first structural gene. However, the repressor protein, controlled by an independent regulator gene, is functional in blocking transcription only after it has combined with the corepressor. This corepressor is the end product of the operon's pathway or some form of the end product (tryptophan in the *trp* operon).

STUDY OBJECTIVE 2: To examine attenuator control in bacteria 415–418

Amino acid-synthesizing operons often have an attenuator region. The ability of a ribosome to translate a leader peptide gene determines the secondary structure of the messenger RNA transcript. If the ribosome can translate the leader peptide gene, there must be adequate quantities of

the amino acid present, and a terminator stem and loop form in the messenger RNA, causing termination of transcription. In the *trp* operon of bacilli, a protein that binds tryptophans serves the same purpose.

STUDY OBJECTIVE 3: To analyze the control of the life cycle of phage λ 418–424

Control of gene expression in λ phage is complex. The "decision" for lytic versus lysogenic response is determined by competition between two repressors, CI and Cro.

STUDY OBJECTIVE 4: To determine the way in which transposable genetic elements transpose and control gene expression in bacteria 425–430

Transposons are mobile genetic elements; copies of them can be inserted at other places in the genome. Their ends are inverted repeats. Upon insertion, they are flanked by short, direct repeats. They can be simple (IS elements) or complex. Their presence can cause inversions or deletions. The flagellar phase in *Salmonella* is controlled by the orientation of a transposon.

STUDY OBJECTIVE 5: To look at other transcriptional and posttranscriptional mechanisms of control of gene expression in bacteria and phages 430–434

Affinity of early and late operons in phages for different sigma factors are another form of transcriptional control, as seen in phage T4 transcription and the transcription of heat shock proteins. Translational control can be exercised through a gene's position in an operon (genes at the beginning are transcribed most frequently), through redundancy of the genetic code, or through a stringent response that shuts down most transcription during starvation. Posttranslational control is primarily regulated by feedback inhibition. The N-terminal amino acid or particular regions within the proteins program the rate of protein degradation.

SOLVED PROBLEMS

PROBLEM 1: How could you determine whether the genes for the breakdown of the sugar arabinose are under inducible control in *E. coli*?

Answer: *Inducible* means that the genes to break down the substrate—arabinose, a five-carbon sugar, in this case—are not active in the absence of the inducer (again, arabinose). Therefore, in the absence of arabinose in the cells' environment, the arabinose utilization enzymes should not be active within the bacterial cells, but after arabinose is added to the medium, the enzymes should be present. We thus need to assay the contents of the cells before and after arabinose is added to the medium, performing the assay after the cells are broken open and the DNA destroyed so as not to confound the experiment. Using a standard biochemical analysis for arabinose, we should find that the bacterial cell is incapable of metabolizing arabinose before induction but capable of metabolizing it afterward. If the cells were capable of metabolizing arabinose in both cases, we would say that arabinose utilization is constitutive. If the cells were incapable of utilizing arabinose in both cases, we would conclude that the bacterium is incapable of using the sugar arabinose as an energy source. (In fact, arabinose utilization is inducible.)

PROBLEM 2: Why would the RecA protein of *E. coli* cleave the λ repressor?

Answer: Since the cleaving of the λ repressor is a signal to begin the lytic phase of the life cycle of the phage, it seems odd that the lysogenized bacterial cell would be an accomplice to its own destruction. However, the phenomenon makes much more sense if we realize that the RecA protein has several other functions critically important to the bacterial cell (see chapter 12). The λ phage has evolved the ability to take advantage of the existence of the RecA protein by evolving a repressor sensitive to it. Evolutionary biologists view this as "coevolution," two interacting organisms evolving to take advantage of or minimize properties of the other. The bacterium, however, might be at a disadvantage. Since RecA has many functions involving interactions with other proteins, it may be highly limited in how it can change. This is one plausible explanation as to why RecA liberates phage λ.

PROBLEM 3: What are the differences in action of the λ promoters p_{RE} and p_{RM}?

Answer: The promoters p_{RE} and p_{RM} are both promoters of the repressor operon of phage λ. Transcription from these promoters allows production of the ***cI*** repressor protein, the repressor that favors lysogeny. Initially, the promoter p_{RE} is activated. For it to be a transcription site, it must be activated by the product of the ***cII*** gene, which lies in the right operon. This promoter, p_{RE}, produces a messenger RNA with a long leader that is translated very efficiently. Once the repressor binds at the operators of the left and right operons, the ***cII*** gene is no longer transcribed, and therefore p_{RE} is no longer a site for transcription. However, the repressor gene can still be transcribed from the p_{RM} promoter, which does not need the product

of the *cII* gene. This promoter produces a transcript with no leader and thus is translated very inefficiently. At that point, however, only a very small quantity of repressor is needed to maintain lysogeny. Thus, the two promoters are the sites for the initiation of the repressor operon under different circumstances: one early in the infection stage and one after lysogeny is under way.

PROBLEM 4: What are the phenotypes of the following partial diploids for the *lac* operon in *E. coli* in the presence and absence of lactose?

a. (F′) $i^+ \ o^+ \ p^+ \ z^-/i^- \ o^+ \ p^+ \ z^+$ (chromosome)

b. (F′) $i^+ \ o^c \ p^+ \ z^+/i^+ \ o^+ \ p^- \ z^+$ (chromosome)

Answer: Consider one DNA molecule at a time. If one DNA molecule can never make the enzyme, it can be ignored. In (*a*), the plasmid DNA (F′) will never make enzyme (it is z^-), and the chromosomal DNA will never make repressor (it is i^-). The functional repressor in the plasmid (i^+) will bind to both DNAs, and hence the chromosomal operon will not be transcribed in the absence of lactose and will be induced to transcribe in the presence of lactose. In (*b*), the plasmid DNA (F′) will always be transcribing (operator constitutive) because the repressor can never bind the operator (o^c); hence, the operon will be transcribed all the time. The chromosomal DNA can never make RNA (it is p^-).

EXERCISES AND PROBLEMS*

LAC OPERON (INDUCIBLE SYSTEM)

1. Are the following *E. coli* cells constitutive or inducible for the *z* gene?

a. $i^+ \ o^+ \ z^+$
b. $i^- \ o^+ \ z^+$
c. $i^- \ o^c \ z^+$
d. $i^+ \ o^c \ z^+$
e. $i^s \ o^+ \ z^+$
f. $i^Q \ o^+ \ z^+$

2. Determine whether the following *lac* operon merozygotes are inducible or constitutive for the *z* gene.

a. $i^+ \ o^+ \ z^+/F' \ i^+ \ o^+ \ z^+$
b. $i^- \ o^+ \ z^+/F' \ i^+ \ o^+ \ z^-$
c. $i^+ \ o^+ \ z^+/F' \ i^- \ o^+ \ z^-$
d. $i^- \ o^c \ z^-/F' \ i^+ \ o^+ \ z^+$
e. $i^- \ o^+ \ z^-/F' \ i^+ \ o^c \ z^+$

3. You have isolated a repressor for an inducible operon and have determined that it has two different binding sites, one for the inducer and one for the operator. Mutants of the repressor result in three different phenotypes as far as binding is concerned. What are these phenotypes?

4. An *E. coli* strain is isolated that produces β-galactosidase (*lac z*) and permease (*lac y*) constitutively. Provide two possible mutations that could cause this phenotype, and then describe how each mutation would behave in a partial diploid in which the second operon is wild-type for the entire *lac* system.

5. You have isolated two *E. coli* mutants that synthesize β-galactosidase constitutively.

a. If these mutants affect different functions, in what two functions could they be defective?

b. You can make a partial diploid of the mutants with the wild-type. What result do you expect for each mutant?

6. A hypothetical operon has a sequence of sites, Q R S T U, in the promoter region, but the exact location of the operator and promoter consensus sequences have not been identified. Various deletions of this operator region are isolated and mapped. Their locations appear as follows, with a "/" representing a deleted region.

Deletion	Q	R	S	T	U
1	/////////				
2		/////////			
3			/////////		
4				//////////////////	
5					///////

Deletions 3 and 4 are found to produce constitutive levels of RNA of the operon, and deletion 1 is found to never make RNA. Where are the operator and promoter consensus sequences probably located?

CATABOLITE REPRESSION

7. Describe the role of cyclic AMP in transcriptional control in *E. coli.*

8. Operon systems exert negative control by acting through inhibition. The CAP system exerts positive control because it acts through enhancement of transcription. Describe how an operon could work if it were dependent only upon positive control.

* Answers to selected exercises and problems are on page A-16.

9. J. Beckwith isolated point mutations that were simultaneously uninducible for the *lac, ara, mal,* and *gal* operons, even in the absence of glucose. Provide two different functions that could be missing in these mutants.

TRP OPERON (REPRESSIBLE SYSTEM)

10. Construct a merozygote of the *trp* operon in *E. coli* with two forms of the first gene (*e* gene: e_1, e_2) in the operon. Describe the types of *cis* and *trans* effects that are possible, given mutants of any component of the operon. Can this repressible system work for any type of operon other than those that control amino acid synthesis?
11. The tryptophan operon is under negative control; it is on (transcribing) in the presence of low levels of tryptophan and off in the presence of excess tryptophan. The symbols *a, b,* and *c* represent the gene for tryptophan synthetase, the operator region, and the repressor—but not necessarily in that order. From the following data, in which superscripts denote wild-type or defective, determine which letter is the gene, the repressor, and the operator (+ is tryptophan synthetase activity; − is no activity).

Strain	Genotype	Tryptophan Absent	Tryptophan Present
1	$a^- b^+ c^+$	+	+
2	$a^+ b^+ c^-$	+	+
3	$a^+ b^- c^+$	−	−
4	$a^+ b^- c^+/a^- b^+ c^-$	+	+
5	$a^+ b^+ c^+/a^- b^- c^-$	+	−
6	$a^+ b^+ c^-/a^- b^- c^+$	+	−
7	$a^- b^+ c^+/a^+ b^- c^-$	+	+

12. The histidine operon is a repressible operon. The corepressor is charged $tRNA^{His}$, and its gene is not part of the operon. For the following mutants, tell whether the enzymes of the operon will be made; then tell whether each mutant would be *cis*-dominant in a partial diploid.
    - **a.** RNA polymerase cannot bind the promoter.
    - **b.** The repressor-corepressor complex cannot bind operator DNA (the operator has the normal sequence).
    - **c.** The repressor cannot bind charged $tRNA^{His}$.

TRP OPERON (ATTENUATOR-CONTROLLED SYSTEM)

13. Describe the interaction of the attenuator and the operator control mechanisms in the *trp* operon of *E. coli* under varying concentrations of tryptophan in the cell. How does attenuator control react to shortages of other amino acids?

LYTIC AND LYSOGENIC CYCLES IN PHAGE λ

14. What is the fate of a λ phage entering an *E. coli* cell that contains quantities of λ repressor? What is the fate of the same phage entering an *E. coli* cell that contains quantities of the *cro*-gene product?
15. Describe the fate of λ phages during the infection process with mutants in the following genes: *cI, cII, cIII, N, cro, att, Q.*
16. What is the fate of λ phages during the infection process with mutants in the following areas: o_{R1}, o_{R3}, p_L, p_{RE}, p_{RM}, p_R, t_{L1}, t_{R1}, *nutL, nutR?*
17. What are the three different physical forms that the phage λ chromosome can take?
18. How does ultraviolet light (UV) damage induce the lytic life cycle in phage λ?
19. The λ prophage is sometimes induced into the lytic life cycle when an Hfr lysogen (lysogenic cell) conjugates with a nonlysogenic F^- cell. How might induction come about in this instance?
20. A temperature-sensitive mutant of the λ *cI* gene has been isolated. At 30° C the *cI* repressor binds λ DNA, but it cannot bind DNA at 42° C (it denatures). What is the consequence of incubating *E. coli* that are lysogenic for this λ mutant at 42° C?
21. The mutant in problem 20 is heated to 42° C for five minutes, cooled to 30° C, and grown for one hour so that the cells divide several times. The temperature is then raised to 42° C, and you wait for lysis. Many of the cells are not lysed and are in fact able to form colonies. Explain these results.

TRANSPOSABLE GENETIC ELEMENTS

22. Why are IS elements sometimes referred to as "selfish DNA"?
23. What are the differences among an IS element, a transposon, an intron, a plasmid, and a cointegrate?
24. Describe the Shapiro model of transposition. What are the roles of transposase, DNA polymerase I, ligase, and resolvase?
25. Why are transposons flanked by direct repeats?
26. How do transposons induce deletions? inversions?
27. Describe how a transposon controls the expression of the flagellar phase in *Salmonella.*
28. What is a polar mutation? What can cause it?

OTHER TRANSCRIPTIONAL CONTROL SYSTEMS

29. List the steps from transcription through translation to enzyme function, noting all the points at which control could be exerted. (*See also* TRANSLATIONAL CONTROL and POSTTRANSLATIONAL CONTROL)

30. What are the advantages of transcriptional control over translational control? (*See also* TRANSLATIONAL CONTROL)

31. How are heat shock proteins induced?

32. In phage T4, the genes *rIIA* and *rIIB* lie adjacent to each other on the T4 chromosome. During the early phase of infection, *rIIA* and *rIIB* products are present in equimolar amounts. In the late phase of infection, the amount of *rIIB* protein is ten to fifteen times higher than that of *rIIA* protein. Nonsense mutations (mutations to a stop codon) in *rIIA* eliminate early but not late *rIIB* transcription. In the mutants that contain small deletions near the end of *rIIA,* the amount of *rIIA* product is always equal to the amount of *rIIB* product, regardless of the time of infection. Based on this information, devise a map of the *rII* region. Include the location(s) of the promoter(s).

TRANSLATIONAL CONTROL

33. What is the stringent response? How does it work? What is an idling reaction?

34. What is antisense RNA? How does it work? What is the obvious source of this regulatory RNA? How could this RNA be used to treat a disease clinically?

POSTTRANSLATIONAL CONTROL

35. What is feedback inhibition? What other roles do allosteric proteins play in regulating gene expression?

36. What controls the rate of degradation of proteins?

CRITICAL THINKING QUESTIONS

1. From an evolutionary perspective, why do you think *Escherichia coli* evolved a CAP system of positive control of gene expression? Why not just metabolize any and all sugars in the environment as they appear?

2. Why might some proteins and messenger RNAs produced in *Escherichia coli* be degraded so quickly?

Suggested Readings for chapter 14 are on page B-13.