THE OPERON

3.10.1 Introduction

Key Terms

A ***trans*-acting** product can function on any copy of its target DNA. This implies that it is a diffusible protein or RNA.

A ***cis*-acting** site affects the activity only of sequences on its own molecule of DNA (or RNA); this property usually implies that the site does not code for protein.

A **structural gene** codes for any RNA or protein product other than a regulator.

A **regulator gene** codes for a product (typically protein) that controls the expression of other genes (usually at the level of transcription).

Gene expression can be controlled at any of several stages, which we divide broadly into transcription, processing, and translation:

- Transcription often is controlled at the stage of initiation. Transcription is not usually controlled at elongation, but may be controlled at termination to determine whether RNA polymerase is allowed to proceed past a terminator to the gene(s) beyond.
- In eukaryotic cells, processing of the RNA product may be regulated at the stages of modification, splicing, transport, or stability. In bacteria, an mRNA is in principle available for translation as soon as (or even while) it is being synthesized, and these stages of control are not available.
- Translation may be regulated, usually at the stages of initiation and termination (like transcription). Regulation of initiation is formally analogous to the regulation of transcription: the circuitry can be drawn in similar terms for regulating initiation of transcription on DNA or initiation of translation on RNA.

The basic concept for how transcription is controlled in bacteria was provided by the classic formulation of the model for control of gene expression by Jacob and Monod in 1961 (508). They distinguished between two types of sequences in DNA: sequences that code for ***trans*-acting** products; and ***cis*-acting** sequences that function exclusively within the DNA. Gene activity is regulated by the specific interactions of the *trans*-acting products (usually proteins) with the *cis*-acting sequences (usually sites in DNA). In more formal terms:

- A gene is a sequence of DNA that codes for a diffusible product. This product may be protein (as in the case of the majority of genes) or may be RNA (as in the case of genes that code for tRNA and rRNA). *The crucial feature is that the product diffuses away from its site of synthesis to act elsewhere.* Any gene product that is free to diffuse to find its target is described as *trans*-acting.
- The description *cis*-acting applies to any sequence of DNA that is not converted

into any other form, but that functions exclusively as a DNA sequence *in situ*, affecting only the DNA to which it is physically linked. (In some cases, a *cis*-acting sequence functions in an RNA rather than in a DNA molecule.)

To help distinguish between the components of regulatory circuits and the genes that they regulate, we sometimes use the terms structural gene and regulator gene. A **structural gene** is simply any gene that codes for a protein (or RNA) product. Structural genes represent an enormous variety of protein structures and functions, including structural proteins, enzymes with catalytic activities, and regulatory proteins. A **regulator gene** simply describes a gene that codes for a protein (or an RNA) involved in regulating the expression of other genes.

The simplest form of the regulatory model is illustrated in **Figure 10.1**: *a regulator gene codes for a protein that controls transcription by binding to particular site(s) on DNA.* This interaction can regulate a target gene in either a positive manner (the interaction turns the gene on) or in a negative manner (the interaction turns the gene off). The sites on DNA are usually (but not exclusively) located just upstream of the target gene.

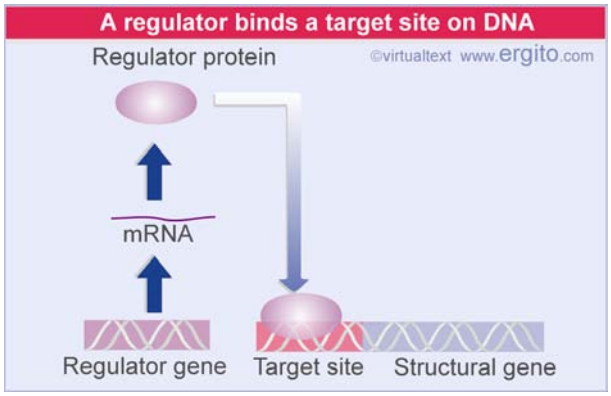

Figure 10.1 A regulator gene codes for a protein that acts at a target site on DNA.

The sequences that mark the beginning and end of the transcription unit, the promoter and terminator, are examples of *cis*-acting sites. *A promoter serves to initiate transcription only of the gene or genes physically connected to it on the same stretch of DNA.* In the same way, a terminator can terminate transcription only by an RNA polymerase that has traversed the preceding gene(s). In their simplest forms, promoters and terminators are *cis*-acting elements that are recognized by the same *trans*-acting species, that is, by RNA polymerase (although other factors also participate at each site).

Additional *cis*-acting regulatory sites are often juxtaposed to, or interspersed with, the promoter. A bacterial promoter may have one or more such sites located close by, that is, in the immediate vicinity of the startpoint. A eukaryotic promoter is likely to have a greater number of sites, spread out over a longer distance.

References

508. Jacob, F. and Monod, J. (1961). *Genetic regulatory mechanisms in the synthesis of proteins*. J. Mol. Biol. 3, 318-389.

THE OPERON

3.10.2 Regulation can be negative or positive

Key Terms

A **repressor** is a protein that inhibits expression of a gene. It may act to prevent transcription by binding to an operator site in DNA, or to prevent translation by binding to RNA.

The **operator** is the site on DNA at which a repressor protein binds to prevent transcription from initiating at the adjacent promoter.

A **transcription factor** is required for RNA polymerase to initiate transcription at specific promoter(s), but is not itself part of the enzyme.

Key Concepts

- In negative regulation a repressor protein binds to an operator to prevent a gene from being expressed.
- In positive regulation a transcription factor is required to bind at the promoter in order to enable RNA polymerase to initiate transcription.

A classic mode of control in bacteria is *negative:* a **repressor** protein prevents a gene from being expressed. **Figure 10.2** shows that the "default state" for such a gene is to be expressed via the recognition of its promoter by RNA polymerase. Close to the promoter is another *cis*-acting site called the **operator**, which is the target for the repressor protein. When the repressor binds to the operator, RNA polymerase is prevented from initiating transcription, and *gene expression is therefore turned off.*

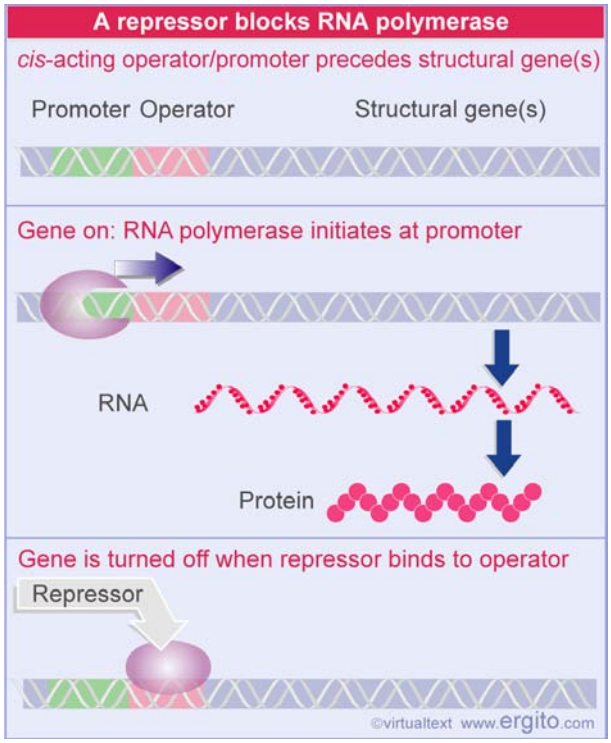

Figure 10.2 In negative control, a trans-acting repressor binds to the *cis*-acting operator to turn off transcription.

An alternative mode of control is *positive*. This is used in bacteria (probably) with about equal frequency to negative control, and it is the most common mode of control in eukaryotes. A **transcription factor** is required to assist RNA polymerase in initiating at the promoter. **Figure 10.3** shows that the typical default state of a eukaryotic gene is inactive: RNA polymerase cannot by itself initiate transcription at the promoter. Several *trans*-acting factors have target sites in the vicinity of the promoter, and *binding of some or all of these factors enables RNA polymerase to initiate transcription.*

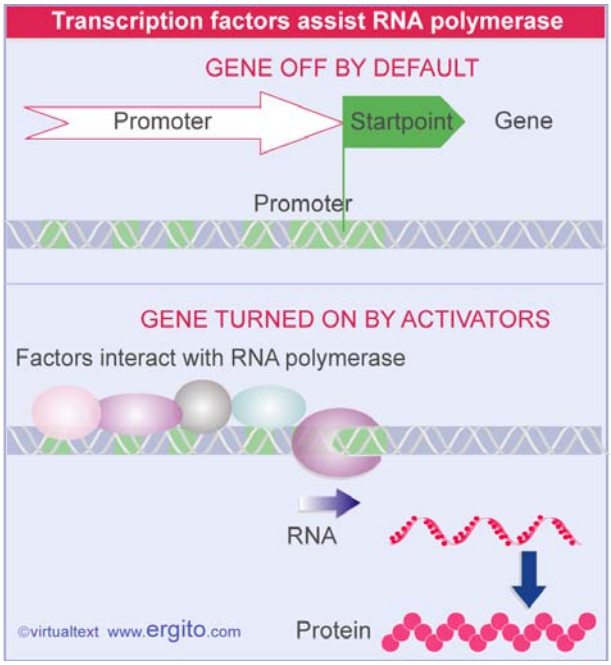

Figure 10.3 In positive control, *trans*-acting factors must bind to *cis*-acting sites in order for RNA polymerase to initiate transcription at the promoter.

The unifying theme is that regulatory proteins are *trans*-acting factors that recognize *cis*-acting elements (usually) upstream of the gene. The consequences of this recognition are to activate or to repress the gene, depending on the individual type of regulatory protein. A typical feature is that the protein functions by recognizing a very short sequence in DNA, usually <10 bp in length, although the protein actually binds over a somewhat greater distance of DNA. The bacterial promoter is an example: although RNA polymerase covers >70 bp of DNA at initiation, the crucial sequences that it recognizes are the hexamers centered at –35 and –10.

A significant difference in gene organization between prokaryotes and eukaryotes is that structural genes in bacteria are organized in clusters, while those in eukaryotes occur individually. Clustering of structural genes allows them to be coordinately controlled by means of interactions at a single promoter: as a result of these interactions, the entire set of genes is either transcribed or not transcribed. In this chapter, we discuss this mode of control and its use by bacteria (for review see 82). The means employed to coordinate control of dispersed eukaryotic genes are discussed in *Molecular Biology 5.22 Activating transcription*.

Reviews

82. Miller, J. and Reznikoff, W. (1978). *The Operon.* Cold Spring Harbor Symp. Quant. Biol..

THE OPERON

3.10.3 Structural gene clusters are coordinately controlled

Key Terms

An **operon** is a unit of bacterial gene expression and regulation, including structural genes and control elements in DNA recognized by regulator gene product(s).

Key Concepts

- Genes coding for proteins that function in the same pathway may be located adjacent to one another and controlled as a single unit that is transcribed into a polycistronic mRNA.

Bacterial structural genes are often organized into clusters that include genes coding for proteins whose functions are related. It is common for the genes coding for the enzymes of a metabolic pathway to be organized into such a cluster. In addition to the enzymes actually involved in the pathway, other related activities may be included in the unit of coordinate control; for example, the protein responsible for transporting the small molecule substrate into the cell.

The cluster of the three *lac* structural genes, *lacZYA*, is typical. **Figure 10.4** summarizes the organization of the structural genes, their associated *cis*-acting regulatory elements, and the *trans*-acting regulatory gene. *The key feature is that the cluster is transcribed into a single polycistronic mRNA from a promoter where initiation of transcription is regulated.*

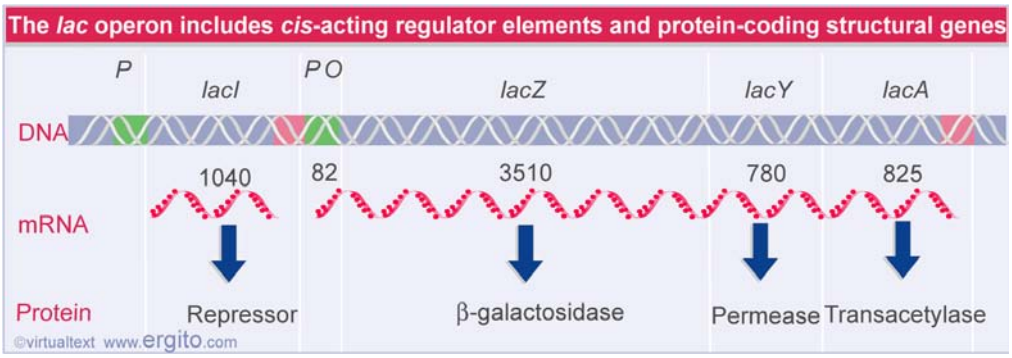

Figure 10.4 The *lac* operon occupies ~6000 bp of DNA. At the left the *lacI* gene has its own promoter and terminator. The end of the *lacI* region is adjacent to the promoter, *P*. The operator, *O*, occupies the first 26 bp of the transcription unit. The long *lacZ* gene starts at base 39, and is followed by the *lacY* and *lacA* genes and a terminator.

The protein products enable cells to take up and metabolize β-galactosides, such as lactose. The roles of the three structural genes are:

- *lacZ* codes for the enzyme β-galactosidase, whose active form is a tetramer of ~500 kD. The enzyme breaks a β-galactoside into its component sugars. For example, lactose is cleaved into glucose and galactose (which are then further metabolized).
- *lacY* codes for the β-galactoside permease, a 30 kD membrane-bound protein constituent of the transport system. This transports β-galactosides into the cell.
- *lacA* codes for β-galactoside transacetylase, an enzyme that transfers an acetyl group from acetyl-CoA to β-galactosides.

Mutations in either *lacZ* or *lacY* can create the *lac* genotype, in which cells cannot utilize lactose. (The genotypic description *"lac"* without a qualifier indicates loss-of-function.) The *lacZ* mutations abolish enzyme activity, directly preventing metabolism of lactose. The *lacY* mutants cannot take up lactose from the medium. (No defect is identifiable in *lacA* cells, which is puzzling. It is possible that the acetylation reaction gives an advantage when the bacteria grow in the presence of certain analogs of β-galactosides that cannot be metabolized, because the modification results in detoxification and excretion.)

The entire system, including structural genes and the elements that control their expression, forms a common unit of regulation; this is called an **operon**. The activity of the operon is controlled by regulator gene(s), whose protein products interact with the *cis*-acting control elements.

THE OPERON

3.10.4 The *lac* genes are controlled by a repressor

Key Terms

The default state of genes that are controlled by **negative regulation** is to be expressed. A specific intervention is required to turn them off.

Key Concepts

- Transcription of the *lacZYA* gene cluster is controlled by a repressor protein that binds to an operator that overlaps the promoter at the start of the cluster.
- The repressor protein is a tetramer of identical subunits coded by the gene *lacI*.

We can distinguish between structural genes and regulator genes by the effects of mutations. A mutation in a structural gene deprives the cell of the particular protein for which the gene codes. But a mutation in a regulator gene influences the expression of all the structural genes that it controls. The consequences of a regulatory mutation reveal the type of regulation (508; for review see 84).

Transcription of the *lacZYA* genes is controlled by a regulator protein synthesized by the *lacI* gene. It happens that *lacI* is located adjacent to the structural genes, but it comprises an independent transcription unit with its own promoter and terminator. Since *lacI* specifies a diffusible product, in principle it need not be located near the structural genes; it can function equally well if moved elsewhere, or carried on a separate DNA molecule (the classic test for a *trans*-acting regulator).

The *lac* genes are controlled by **negative regulation**: *they are transcribed unless turned off by the regulator protein.* A mutation that inactivates the regulator causes the structural genes to remain in the expressed condition. The product of *lacI* is called the *Lac repressor*, because its function is to prevent the expression of the structural genes.

The repressor is a tetramer of identical subunits of 38 kD each. There are ~10 tetramers in a wild-type cell. The regulator gene is not controlled by the availability of lactose. It is transcribed into a monocistronic mRNA at a rate that appears to be governed simply by the affinity of its promoter for RNA polymerase.

The repressor functions by binding to an operator (formally denoted O_{lac}) at the start of the *lacZYA* cluster. The operator lies between the promoter (P_{lac}) and the structural genes (*lacZYA*). *When the repressor binds at the operator, it prevents RNA polymerase from initiating transcription at the promoter*. **Figure 10.5** expands our view of the region at the start of the *lac* structural genes. The operator extends from position -5 just upstream of the mRNA startpoint to position +21 within the transcription unit. So it overlaps the right end of the promoter. We discuss the relationship between repressor and RNA polymerase in more detail in *Molecular Biology 3.10.10 Repressor protein binds to the operator* and *Molecular*

Biology 3.10.16 Repressor binds to three operators and interacts with RNA polymerase (for review see 83; 85; 86; 87).

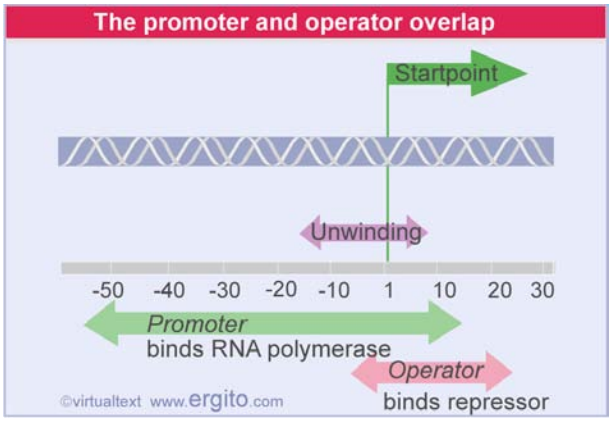

Figure 10.5 Repressor and RNA polymerase bind at sites that overlap around the transcription startpoint of the *lac* operon.

Reviews

83. Barkleya, M. D. and Bourgeois, S. (1978). *Repressor recognition of operator and effectors*. The Operon, 177-220.

84. Beckwith, J. (1978). *lac: the genetic system*. The Operon, 11-30.

85. Beyreuther, K. (1978). *Chemical structure and functional organization of lac repressor from E. coli*. The Operon, 123-154.

86. Miller, J. H. (1978). *The lacI gene: its role in lac operon control and its use as a genetic system*. The Operon, 31-88.

87. Weber, K. and Geisler, N. (1978). *Lac repressor fragments produced in vivo and in vitro: an approach to the understanding of the interaction of repressor and DNA*. The Operon, 155-176.

References

508. Jacob, F. and Monod, J. (1961). *Genetic regulatory mechanisms in the synthesis of proteins*. J. Mol. Biol. 3, 318-389.

THE OPERON

3.10.5 The *lac* operon can be induced

Key Terms

Induction refers to the ability of bacteria (or yeast) to synthesize certain enzymes only when their substrates are present; applied to gene expression, it refers to switching on transcription as a result of interaction of the inducer with the regulator protein.

The level of response from a system in the absence of a stimulus is its **basal level.** (The basal level of transcription of a gene is the level that occurs in the absence of any specific activation.)

Repression describes the ability of bacteria to prevent synthesis of certain enzymes when their products are present; more generally, refers to inhibition of transcription (or translation) by binding of repressor protein to a specific site on DNA (or mRNA).

An **inducer** is a small molecule that triggers gene transcription by binding to a regulator protein.

A **corepressor** is a small molecule that triggers repression of transcription by binding to a regulator protein.

Key Concepts

- Small molecules that induce an operon are identical with or related to the substrate for its enzymes.
- β-galactosides are the substrates for the enzymes coded by *lacZYA*.
- In the absence of β-galactosides the *lac* operon is expressed only at a very low (basal) level.
- Addition of β-galactosides induces transcription of all three genes of the operon.
- Because the *lac* mRNA is extremely unstable, induction can be rapidly reversed.
- The same types of systems that allow substrates to induce operons coding for metabolic enzymes can be used to allow end-products to repress the operons that code for biosynthetic enzymes.

Bacteria need to respond swiftly to changes in their environment. Fluctuations in the supply of nutrients can occur at any time; survival depends on the ability to switch from metabolizing one substrate to another. Yet economy also is important, since a bacterium that indulges in energetically expensive ways to meet the demands of the environment is likely to be at a disadvantage. So a bacterium avoids synthesizing the enzymes of a pathway in the absence of the substrate; but is ready to produce the enzymes if the substrate should appear.

The synthesis of enzymes in response to the appearance of a specific substrate is

called **induction**. This type of regulation is widespread in bacteria, and occurs also in unicellular eukaryotes (such as yeasts). The lactose system of *E. coli* provides the paradigm for this sort of control mechanism.

When cells of *E. coli* are grown in the absence of a β-galactoside, there is no need for β-galactosidase, and they contain very few molecules of the enzyme – say, <5. When a suitable substrate is added, the enzyme activity appears very rapidly in the bacteria. Within 2-3 minutes some enzyme is present, and soon there are ~5000 molecules of enzyme per bacterium. (Under suitable conditions, β-galactosidase can account for 5-10% of the total soluble protein of the bacterium.) If the substrate is removed from the medium, the synthesis of enzyme stops as rapidly as it had originally started.

Figure 10.6 summarizes the essential features of induction. Control of transcription of the *lac* genes responds very rapidly to the inducer, as shown in the upper part of the figure. In the absence of inducer, the operon is transcribed at a very low **basal level**. Transcription is stimulated as soon as inducer is added; the amount of *lac* mRNA increases rapidly to an induced level that reflects a balance between synthesis and degradation of the mRNA.

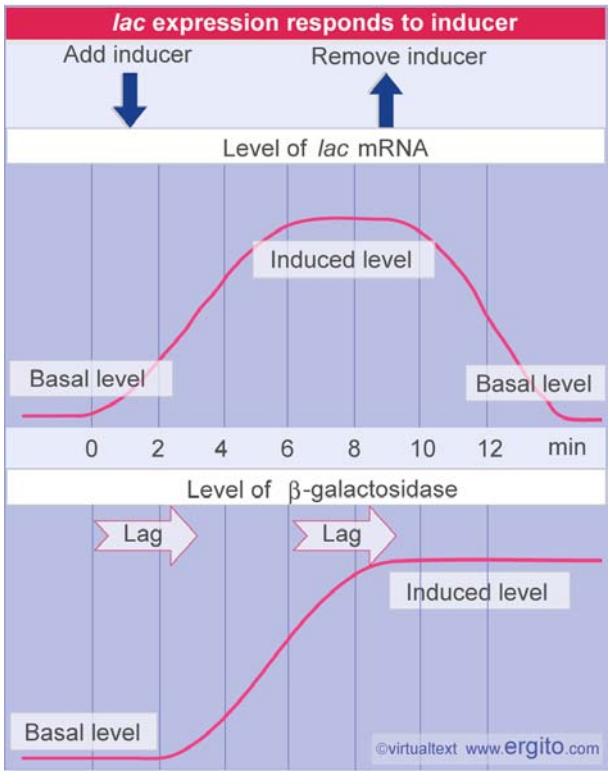

Figure 10.6 Addition of inducer results in rapid induction of *lac* mRNA, and is followed after a short lag by synthesis of the enzymes; removal of inducer is followed by rapid cessation of synthesis.

The *lac* mRNA is extremely unstable, and decays with a half-life of only ~3 minutes. This feature allows induction to be reversed rapidly. Transcription ceases as soon as the inducer is removed; and in a very short time all the *lac* mRNA has been destroyed, and the cellular content has returned to the basal level.

The production of protein is followed in the lower part of the figure. Translation of the *lac* mRNA produces β-galactosidase (and the products of the other *lac* genes). There is a short lag between the appearance of *lac* mRNA and appearance of the first completed enzyme molecules (it is ~2 min after rise of mRNA from basal level before protein begins to increase). There is a similar lag between reaching maximal induced levels of mRNA and protein. When inducer is removed, synthesis of enzyme ceases almost immediately (as the mRNA is degraded), but the β-galactosidase in the cell is more stable than the mRNA, so the enzyme activity remains at the induced level for longer.

This type of rapid response to changes in nutrient supply not only provides the ability to metabolize new substrates, but also is used to shut off endogenous synthesis of compounds that suddenly appear in the medium. For example, *E. coli* synthesizes the amino acid tryptophan through the action of the enzyme tryptophan synthetase. But if tryptophan is provided in the medium on which the bacteria are growing, the production of the enzyme is immediately halted. This effect is called **repression**. It allows the bacterium to avoid devoting its resources to unnecessary synthetic activities.

Induction and repression represent the same phenomenon. In one case the bacterium adjusts its ability to use a given substrate (such as lactose) for growth; in the other it adjusts its ability to synthesize a particular metabolic intermediate (such as an essential amino acid). The trigger for either type of adjustment is the small molecule that is the substrate for the enzyme, or the product of the enzyme activity, respectively. Small molecules that cause the production of enzymes able to metabolize them are called **inducers**. Those that prevent the production of enzymes able to synthesize them are called **corepressors**.

THE OPERON

3.10.6 Repressor is controlled by a small molecule inducer

Key Terms

Gratuitous inducers resemble authentic inducers of transcription but are not substrates for the induced enzymes.

Allosteric regulation describes the ability of a protein to change its conformation (and therefore activity) at one site as the result of binding a small molecule to a second site located elsewhere on the protein.

Coordinate regulation refers to the common control of a group of genes.

Key Concepts

- An inducer functions by converting the repressor protein into an inactive form.
- Repressor has two binding sites, one for the operator and another for the inducer.
- Repressor is inactivated by an allosteric interaction in which binding of inducer at its site changes the properties of the DNA-binding site.

The ability to act as inducer or corepressor is highly specific. Only the substrate/product or a closely related molecule can serve. *But the activity of the small molecule does not depend on its interaction with the target enzyme.* Some inducers resemble the natural inducers of the *lac* operon, but cannot be metabolized by the enzyme. The example *par excellence* is isopropylthiogalactoside (IPTG), one of several thiogalactosides with this property. Although it is not recognized by β-galactosidase, IPTG is a very efficient inducer of the *lac* genes.

Molecules that induce enzyme synthesis but are not metabolized are called **gratuitous inducers**. They are extremely useful because they remain in the cell in their original form. (A real inducer would be metabolized, interfering with study of the system.) The existence of gratuitous inducers reveals an important point. *The system must possess some component, distinct from the target enzyme, that recognizes the appropriate substrate; and its ability to recognize related potential substrates is different from that of the enzyme.*

The component that responds to the inducer is the repressor protein coded by *lacI*. The *lacZYA* structural genes are transcribed into a single mRNA from a promoter just upstream of *lacZ.* The state of the repressor determines whether this promoter is turned off or on:

- **Figure 10.7** shows that in the absence of an inducer, the genes are not transcribed, because repressor protein is in an active form that is bound to the operator.

- **Figure 10.8** shows that when an inducer is added, the repressor is converted into an inactive form that leaves the operator. Then transcription starts at the promoter and proceeds through the genes to a terminator located beyond the 3′ end of *lacA*.

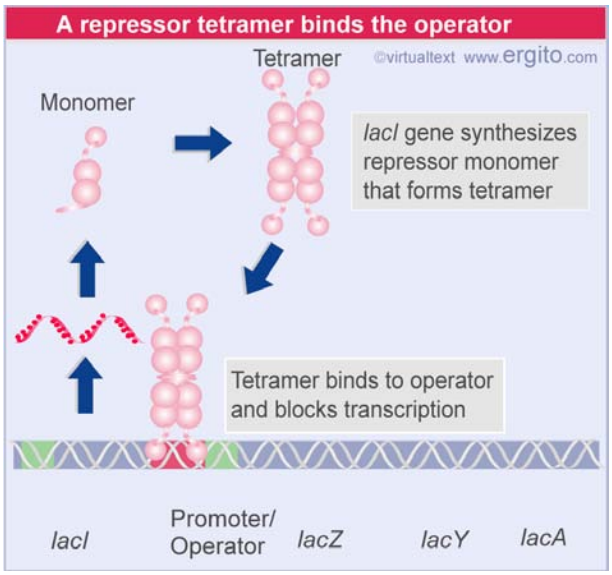

Figure 10.7 Repressor maintains the *lac* operon in the inactive condition by binding to the operator. The shape of the repressor is represented as a series of connected domains as revealed by its crystal structure (see later).
This is a static version of an interactive figure; see http://www.ergito.com/main.jsp?bcs=MBIO.3.10.6 to view properly.

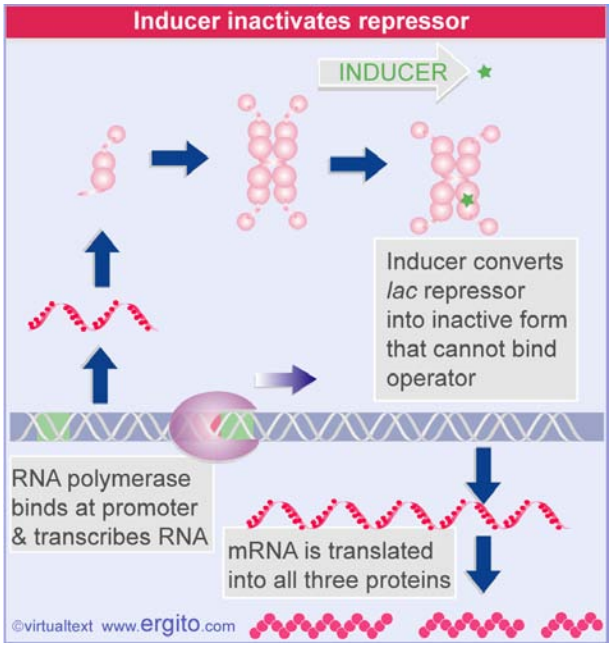

Figure 10.8 Addition of inducer converts repressor to an inactive form that cannot bind the operator. This allows RNA polymerase to initiate transcription.

The crucial features of the control circuit reside in the dual properties of the repressor: it can prevent transcription; and it can recognize the small-molecule inducer. The repressor has two binding sites, one for the operator and one for the inducer. When the inducer binds at its site, it changes the conformation of the protein in such a way as to influence the activity of the operator-binding site. The ability of one site in the protein to control the activity of another is called **allosteric** control (see *Molecular Biology Supplement 32.7 Allostery*).

Induction accomplishes a **coordinate regulation**: *all the genes are expressed (or not expressed) in unison.* The mRNA is translated sequentially from its 5′ end, which explains why induction always causes the appearance of β-galactosidase, β-galactoside permease, and β-galactoside transacetylase, in that order. Translation of a common mRNA explains why the relative amounts of the three enzymes always remain the same under varying conditions of induction.

Induction throws a switch that causes the genes to be transcribed. Inducers vary in their effectiveness, and other factors influence the absolute level of transcription or translation, but the relationship between the three genes is predetermined by their organization.

We notice a potential paradox in the constitution of the operon. The lactose operon contains the structural gene (*lacZ*) coding for the β-galactosidase activity needed to metabolize the sugar; it also includes the gene (*lacY*) that codes for the protein needed to transport the substrate into the cell. But if the operon is in a repressed state, how does the inducer enter the cell to start the process of induction?

Two features ensure that there is always a minimal amount of the protein present in

the cell, enough to start the process off. There is a basal level of expression of the operon: even when it is not induced, it is expressed at a residual level (0.1% of the induced level). And some inducer enters anyway via another uptake system.

THE OPERON

3.10.7 *cis*-acting constitutive mutations identify the operator

Key Terms

An **uninducible** mutant is one where the affected gene(s) cannot be expressed.

A **constitutive** process is one that occurs all the time, unchanged by any form of stimulus or external condition.

A ***cis*-dominant** site or mutation affects the properties only of its own molecule of DNA. *cis*-dominance is taken to indicate that a site does not code for a diffusible product. (A rare exception is that a protein is *cis*-dominant when it is constrained to act only on the DNA or RNA from which it was synthesized.)

Key Concepts

- Mutations in the operator cause constitutive expression of all three *lac* structural genes.
- They are *cis*-acting and affect only those genes on the contiguous stretch of DNA.

Mutations in the regulatory circuit may either abolish expression of the operon or cause unregulated expression. Mutants that cannot be expressed at all are called **uninducible**. The continued expression of a gene that does not respond to regulation is called **constitutive** gene expression, and mutants with this property are called constitutive mutants.

Components of the regulatory circuit of the operon can be identified by mutations that *affect the expression of all the structural genes and map outside them*. They fall into two classes. The promoter and the operator are identified as targets for the regulatory proteins (RNA polymerase and repressor, respectively) by *cis*-acting mutations. And the locus *lacI* is identified as the gene that codes for the repressor protein by mutations that eliminate the *trans*-acting product.

The operator was originally identified by constitutive mutations, denoted O^c, whose distinctive properties provided the first evidence for *an element that functions without being represented in a diffusible product.*

The structural genes contiguous with an O^c mutation are expressed constitutively because the mutation changes the operator so that the repressor no longer binds to it. So the repressor cannot prevent RNA polymerase from initiating transcription. The operon is transcribed constitutively, as illustrated in **Figure 10.9**.

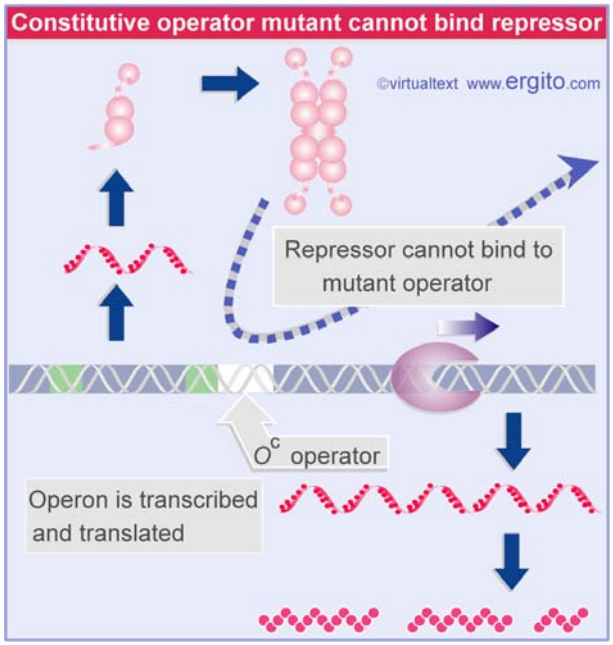

Figure 10.9 Operator mutations are constitutive because the operator is unable to bind repressor protein; this allows RNA polymerase to have unrestrained access to the promoter. The O^c mutations are *cis*-acting, because they affect only the contiguous set of structural genes.

The operator can control only the lac *genes that are adjacent to it.* If a second *lac* operon is introduced into the bacterium on an independent molecule of DNA, it has its own operator. Neither operator is influenced by the other. So if one operon has a wild-type operator, it will be repressed under the usual conditions, while a second operon with an O^c mutation will be expressed in its characteristic fashion.

These properties define the operator as a typical *cis*-acting site, whose function depends upon recognition of its DNA sequence by some *trans*-acting factor. The operator controls the adjacent genes irrespective of the presence in the cell of other alleles of the site. A mutation in such a site, for example, the O^c mutation, is formally described as ***cis*-dominant**.

A mutation in a *cis*-acting site cannot be assigned to a complementation group. (The ability to complement defines genes that are expressed as diffusible products.) When two *cis*-acting sites lie close together – for example, a promoter and an operator – we cannot classify the mutations by a complementation test. We are restricted to distinguishing them by their effects on the phenotype.

cis-dominance is a characteristic of any site that is *physically contiguous with the sequences it controls.* If a control site functions as part of a polycistronic mRNA, mutations in it will display *exactly the same pattern* of *cis*-dominance as they would if functioning in DNA. The critical feature is that the control site cannot be physically separated from the genes that it regulates. From the genetic point of view, it does not matter whether the site and genes are together on DNA or on RNA.

THE OPERON

3.10.8 *trans*-acting mutations identify the regulator gene

Key Concepts

- Mutations in the *lacI* gene are *trans*-acting and affect expression of all *lacZYA* clusters in the bacterium.
- Mutations that eliminate *lacI* function cause constitutive expression and are recessive.
- Mutations in the DNA-binding site of the repressor are constitutive because the repressor cannot bind the operator.
- Mutations in the inducer-binding site of the repressor prevent it from being inactivated and cause uninducibility.
- Mutations in the promoter are uninducible and *cis*-acting.

Constitutive transcription is also caused by mutations of the *lacI*$^-$ type, which are caused by loss of function (including deletions of the gene). When the repressor is inactive or absent, transcription can initiate at the promoter. **Figure 10.10** shows that the *lacI*$^-$ mutants express the structural genes all the time (constitutively), *irrespective of whether the inducer is present or absent*, because the repressor is inactive.

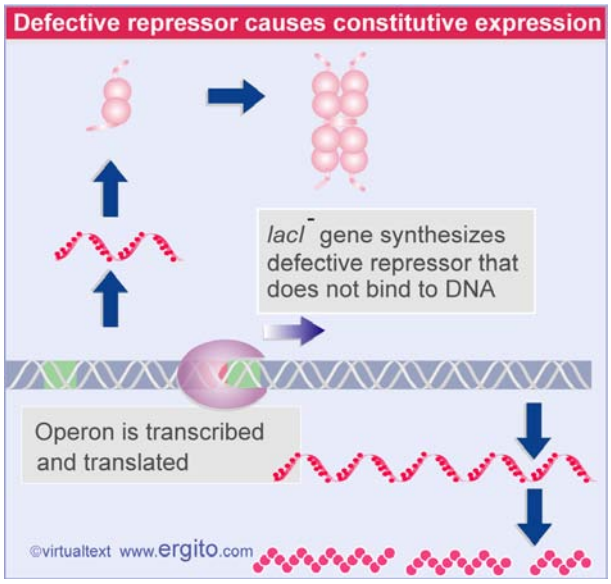

Figure 10.10 Mutations that inactivate the *lacI* gene cause the operon to be constitutively expressed, because the mutant repressor protein cannot bind to the operator.

The two types of constitutive mutations can be distinguished genetically. O^c mutants are *cis*-dominant, whereas $lacI^-$ mutants are recessive. This means that the introduction of a normal, $lacI^+$ gene restores control, irrespective of the presence of the defective $lacI^-$ gene.

Mutants of the operon that are uninducible fall into the same two types of genetic classes as the constitutive mutants:

- Promoter mutations are *cis*-acting. If they prevent RNA polymerase from binding at P_{lac}, they render the operon nonfunctional because it cannot be transcribed.
- Mutations that abolish the ability of repressor to bind the inducer are described as $lacI^s$. They are *trans*-acting. The repressor is "locked in" to the active form that recognizes the operator and prevents transcription. The addition of inducer has no effect because its binding site is absent, and therefore it is impossible to convert the repressor to the inactive form. The mutant repressor binds to all *lac* operators in the cell to prevent their transcription, and cannot be pried off, irrespective of the properties of any wild-type repressor protein that is present, so it is genetically dominant.

The two types of mutations in *lacI* can be used to identify the individual active sites in the repressor protein. The *DNA-binding site* recognizes the sequence of the operator. It is identified by constitutive point mutations that prevent repressor from binding to DNA to block RNA polymerase. The *inducer-binding site* is identified by point mutations that cause uninducibility, because inducer cannot bind to trigger the allosteric change in the DNA-binding site.

THE OPERON

3.10.9 Multimeric proteins have special genetic properties

Key Terms

Interallelic complementation (intragenic complementation) describes the change in the properties of a heteromultimeric protein brought about by the interaction of subunits coded by two different mutant alleles; the mixed protein may be more or less active than the protein consisting of subunits only of one or the other type.

Negative complementation occurs when interallelic complementation allows a mutant subunit to suppress the activity of a wild-type subunit in a multimeric protein.

A **dominant negative** mutation results in a mutant gene product that prevents the function of the wild-type gene product, causing loss or reduction of gene activity in cells containing both the mutant and wild-type alleles. The effect may result from the titration of another factor that interacts with the gene product or by an inhibiting interaction of the multimer.

Key Concepts

- Active repressor is a tetramer of identical subunits.
- When mutant and wild-type subunits are present, a single *lacI*$^{-d}$ mutant subunit can inactivate a tetramer whose other subunits are wild-type.
- *lacI*$^{-d}$ mutations occur in the DNA-binding site. Their effect is explained by the fact that repressor activity requires all DNA-binding sites in the tetramer to be active.

An important feature of the repressor is that it is multimeric. Repressor subunits associate at random in the cell to form the active protein tetramer. When two different alleles of the *lacI* gene are present, the subunits made by each can associate to form a heterotetramer, whose properties differ from those of either homotetramer. This type of interaction between subunits is a characteristic feature of multimeric proteins and is described as **interallelic complementation** (see *Molecular Biology Supplement 32.9 Complementation*).

Negative complementation occurs between some repressor mutants, as seen in the combination of *lacI*$^{-d}$ with *lacI*$^{+}$ genes. The *lacI*$^{-d}$ mutation alone results in the production of a repressor that cannot bind the operator, and is therefore constitutive like the *lacI*$^{-}$ alleles. Because the *lacI*$^{-}$ type of mutation inactivates the repressor, it is usually recessive to the wild type. However, the *–d* notation indicates that this variant of the negative type is dominant when paired with a wild-type allele. Such mutations are called **dominant negative** (see *Molecular Biology Supplement 32.9 Complementation*).

Figure 10.11 explains this phenomenon. The reason for the dominance is that the

lacI^{d} allele produces a "bad" subunit, which is not only itself unable to bind to operator DNA, but is also able as part of a tetramer to prevent any "good" subunits from binding. This demonstrates that the repressor tetramer as a whole, rather than the individual monomer, is needed to achieve repression. In fact, we may reverse the argument to say that, whenever a protein has a dominant negative form, this must mean it functions as part of a multimer. The production of dominant negative proteins has become an important technique in eukaryotic genetics.

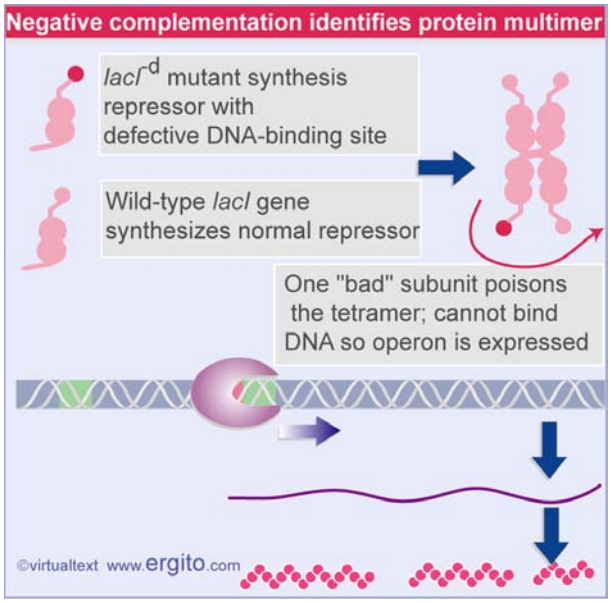

Figure 10.11 A *lacI*$^{-d}$ mutant gene makes a monomer that has a damaged DNA binding site (shown by the red circle). When it is present in the same cell as a wild-type gene, multimeric repressors are assembled at random from both types of subunits. It only requires one of the subunits of the multimer to be of the *lacI*$^{>-d}$ type to block repressor function. This explains the dominant negative behavior of the *lacI*$^{>-d}$ mutation.

Last updated on 7-25-2002

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.10.9

THE OPERON

3.10.10 Repressor protein binds to the operator

Key Terms

A **palindrome** is a DNA sequence that reads the same on each strand of DNA when the strand is read in the 5′ to 3′ direction. It consists of adjacent inverted repeats.

Key Concepts

- Repressor protein binds to the double-stranded DNA sequence of the operator.
- The operator is a palindromic sequence of 26 bp.
- Each inverted repeat of the operator binds to the DNA-binding site of one repressor subunit.

The repressor was isolated originally by purifying the component able to bind the gratuitous inducer IPTG (509; see *Great Experiments 2.1 Isolation of repressor*). (Because the amount of repressor in the cell is so small, in order to obtain enough material it was necessary to use a promoter up mutation to increase *lacI* transcription, and to place this *lacI* locus on a DNA molecule present in many copies per cell. This results in an overall overproduction of 100-1000-fold.)

The repressor binds to double-stranded DNA containing the sequence of the wild-type *lac* operator. The repressor does not bind DNA from an O^c mutant. The addition of IPTG releases the repressor from operator DNA *in vitro.* The *in vitro* reaction between repressor protein and operator DNA therefore displays the characteristics of control inferred *in vivo;* so it can be used to establish the basis for repression (510).

How does the repressor recognize the specific sequence of operator DNA? The operator has a feature common to many recognition sites for bacterial regulator proteins: it is a **palindrome**. The inverted repeats are highlighted in **Figure 10.12**. Each repeat can be regarded as a half-site of the operator.

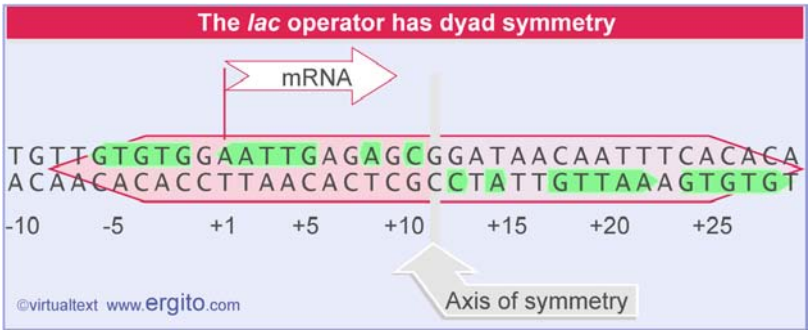

Figure 10.12 The *lac* operator has a symmetrical sequence. The sequence is numbered relative to the startpoint for transcription at +1. The pink arrows to left and right identify the two dyad repeats. The green blocks indicate the positions of identity.

We can use the same approaches to define the points that the repressor contacts in the operator that we used for analyzing the polymerase-promoter interaction (see *Molecular Biology 3.9.14 RNA polymerase binds to one face of DNA*). Deletions of material on either side define the end points of the region; constitutive point mutations identify individual base pairs that must be crucial. Experiments in which DNA bound to repressor is compared with unbound DNA for its susceptibility to methylation or UV crosslinking identify bases that are either protected or more susceptible when associated with the protein.

Figure 10.13 shows that the region of DNA protected from nucleases by bound repressor lies within the region of symmetry, comprising the 26 bp region from -5 to +21. The area identified by constitutive mutations is even smaller. Within a central region extending over the 13 bp from +5 to +17, there are eight sites at which single base-pair substitutions cause constitutivity. This emphasizes the same point made by the promoter mutations summarized earlier in **Figure 9.29**. *A small number of essential specific contacts within a larger region can be responsible for sequence-specific association of DNA with protein.*

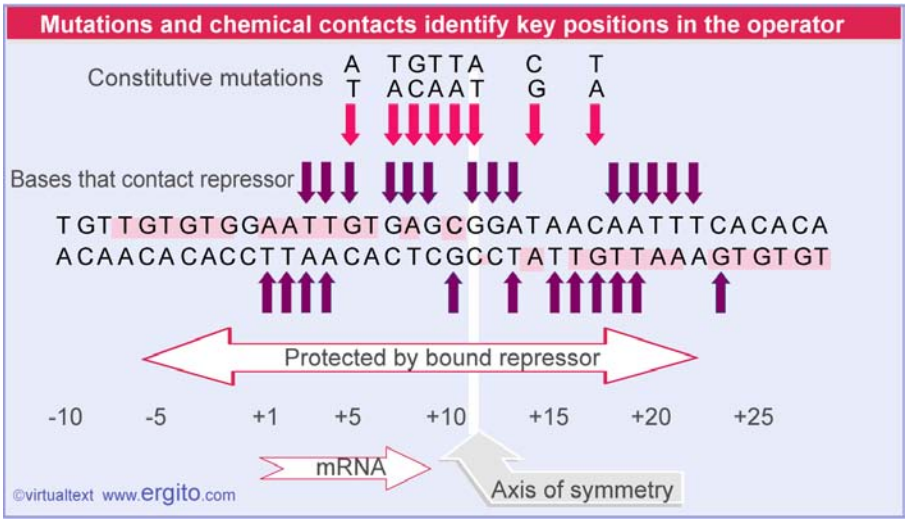

Figure 10.13 Bases that contact the repressor can be identified by chemical crosslinking or by experiments to see whether modification prevents binding. They identify positions on both strands of DNA extending from +1 to +23. Constitutive mutations occur at 8 positions in the operator between +5 and +17.

The symmetry of the DNA sequence reflects the symmetry in the protein. Each of the identical subunits in a repressor tetramer has a DNA-binding site. Two of these sites contact the operator in such a way that each inverted repeat of the operator makes the same pattern of contacts with a repressor monomer. This is shown by symmetry in the contacts that repressor makes with the operator (the pattern between +1 and +6 is identical with that between +21 and +16) and by matching constitutive mutations in each inverted repeat. (However, the operator is not perfectly symmetrical; the left side binds more strongly than the right side to the repressor. A stronger operator would be created by a perfect inverted duplication of the left side.)

Last updated on 7-29-2002

References

509. Gilbert, W. and Muller-Hill, B. (1966). *Isolation of the lac repressor*. Proc. Natl. Acad. Sci. USA 56, 1891-1898.

510. Gilbert, W. and Muller-Hill, B. (1967). *The lac operator is DNA*. Proc. Natl. Acad. Sci. USA 58, 2415-2421.

THE OPERON

3.10.11 Binding of inducer releases repressor from the operator

Key Concepts

- Inducer binding causes a change in repressor conformation that reduces its affinity for DNA and releases it from the operator.

Various inducers cause characteristic reductions in the affinity of the repressor for the operator *in vitro*. These changes correlate with the effectiveness of the inducers *in vivo*. This suggests that induction results from a reduction in the attraction between operator and repressor. So when inducer enters the cell, it binds to free repressors and in effect prevents them from finding their operators. But consider a repressor tetramer that is already bound tightly to the operator. How does inducer cause this repressor to be released?

Two models for repressor action are illustrated in **Figure 10.14**:

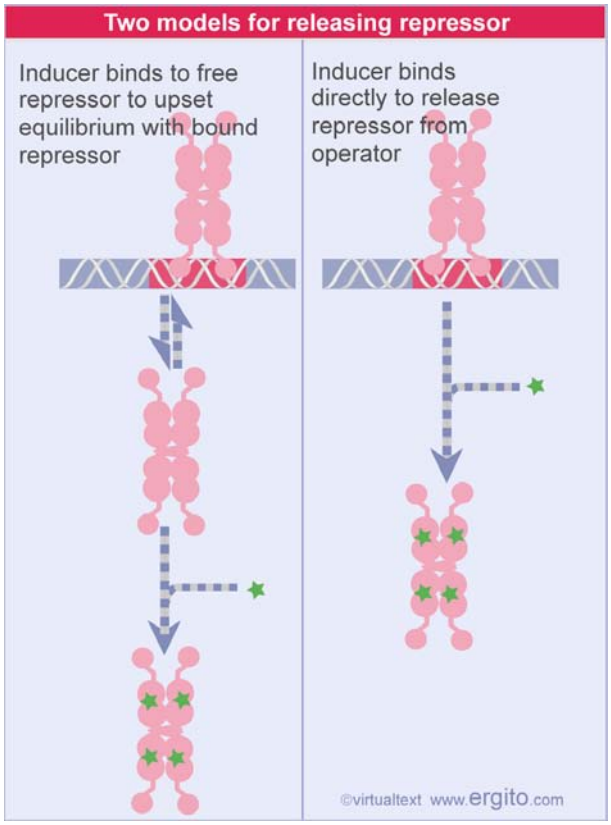

Figure 10.14 Does the inducer bind to free repressor to upset an equilibrium (left) or directly to repressor bound at the operator (right)?

- The equilibrium model (left) calls for repressor bound to DNA to be in rapid equilibrium with free repressor. Inducer would bind to the free form of repressor, and thus unbalance the equilibrium by preventing reassociation with DNA.

- But the rate of dissociation of the repressor from the operator is much too slow to be compatible with this model (the half-life *in vitro* in the absence of inducer is >15 min). This means that instead the *inducer must bind directly to repressor protein complexed with the operator.* As indicated in the model on the right, inducer binding must produce a change in the repressor that makes it release the operator. Indeed, addition of IPTG causes an immediate destabilization of the repressor-operator complex *in vitro*.

Binding of the repressor-IPTG complex to the operator can be studied by using greater concentrations of the protein in the methylation protection/enhancement assay. The large amount compensates for the low affinity of the repressor-IPTG complex for the operator. The complex makes exactly the same pattern of contacts with DNA as the free repressor. An analogous result is obtained with mutant repressors whose affinity for operator DNA is increased; they too make the same pattern of contacts.

Overall, a range of repressor variants whose affinities for the operator span seven orders of magnitude all make the same contacts with DNA. *Changes in the affinity of the repressor for DNA must therefore occur by influencing the general conformation of the protein in binding DNA, not by making or breaking one or a few individual bonds.*

THE OPERON

3.10.12 The repressor monomer has several domains

Key Terms

The **headpiece** is the DNA-binding domain of the *lac* repressor.

Key Concepts

- A single repressor subunit can be divided into the N-terminal DNA-binding domain, a hinge, and the core of the protein.
- The DNA-binding domain contains two short α-helical regions that bind the major groove of DNA.
- The inducer-binding site and the regions responsible for multimerization are located in the core.

The repressor has several domains. The DNA-binding domain occupies residues 1-59. It is known as the **headpiece**. It can be cleaved from the remainder of the monomer, which is known as the *core*, by trypsin. The crystal structure illustrated in **Figure 10.15** offers a more detailed account of these regions (512; 513).

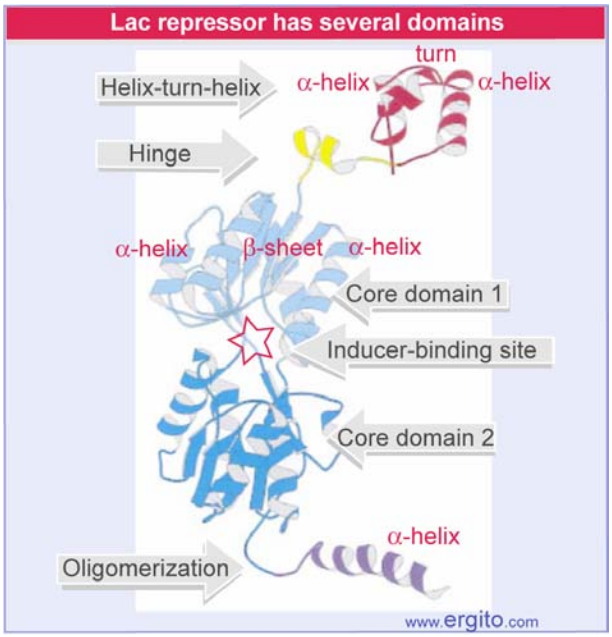

Figure 10.15 The structure of a monomer of Lac repressor identifies several independent domains. Photograph kindly provided by Mitchell Lewis.

The N-terminus of the monomer consists of two α-helices separated by a turn. This

is a common DNA-binding motif, known as the HTH (helix-turn-helix); the two α-helices fit into the major groove of DNA, where they make contacts with specific bases (see *Molecular Biology 3.12.12 Repressor uses a helix-turn-helix motif to bind DNA*). This region is connected by a *hinge* to the main body of the protein. In the DNA-binding form of repressor, the hinge forms a small α-helix (as shown in the figure); but when the repressor is not bound to DNA, this region is disordered. The HTH and hinge together correspond to the headpiece.

The bulk of the core consists of two regions with similar structures (core domains 1 and 2). Each has a six-stranded parallel β-sheet sandwiched between two α-helices on either side. The inducer binds in a cleft between the two regions .

At the C-terminus, there is an α-helix that contains two leucine heptad repeats. This is the oligomerization domain. The oligomerization helices of four monomers associate to maintain the tetrameric structure.

Last updated on 7-29-2002

References

512. Friedman, A. M., Fischmann, T. O., and Steitz, T. A. (1995). *Crystal structure of lac repressor core tetramer and its implications for DNA looping*. Science 268, 1721-1727.

513. Lewis, M. et al. (1996). *Crystal structure of the lactose operon repressor and its complexes with DNA and inducer*. Science 271, 1247-1254.

THE OPERON

3.10.13 Repressor is a tetramer made of two dimers

Key Concepts

- Monomers form a dimer by making contacts between core domain 2 and between the oligomerization helices.
- Dimers form a tetramer by interactions between the oligomerization helices.

Figure 10.16 shows the structure of the tetrameric core (using a different modeling system from **Figure 10.15**). It consists in effect of two dimers. The body of the dimer contains a loose interface between the N-terminal regions of the core monomers, a cleft at which inducer binds, and a hydrophobic core (top). The C-terminal regions of each monomer protrude as parallel helices. (The headpiece would join on to the N-terminal regions at the top.) Together the dimers interact to form a tetramer (center) that is held together by a C-terminal bundle of four helices.

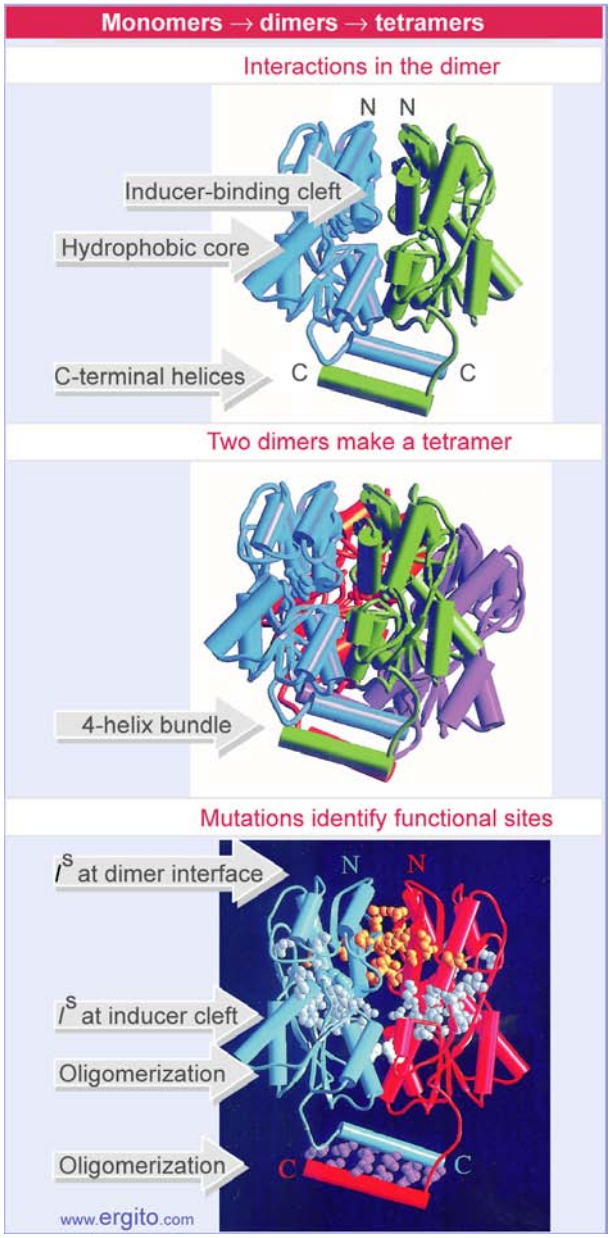

Figure 10.16 The crystal structure of the core region of Lac repressor identifies the interactions between monomers in the tetramer. Each monomer is identified by a different color. Mutations are colored as: dimer interface - yellow; inducer-binding blue; oligomerization - white and purple. Photographs kindly provided by Alan Friedman.

Sites of mutations are shown by beads on the structure at the bottom. *lacI* ^{s} mutations make the repressor unresponsive to the inducer, so that the operon is uninducible. They map in two groups: gray shows those in the inducer-binding cleft, and yellow shows those that affect the dimer interface. The first group abolish the inducer binding site; the second group prevent the effects of inducer binding from being transmitted to the DNA-binding site. *lacI* $^{-d}$ mutations that affect oligomerization map in two groups. White shows mutations in core domain 2 that prevent dimer

formation. Purple shows those in the oligomerization helix that prevent tetramer formation from dimers.

From these data we can derive the schematic of **Figure 10.17**, which shows how the monomers are organized into the tetramer. Two monomers form a dimer by means of contacts at core domain 2 and in the oligomerization helix. The dimer has two DNA-binding domains at one end of the structure, and the oligomerization helices at the other end. Two dimers then form a tetramer by interactions at the oligomerization interface.

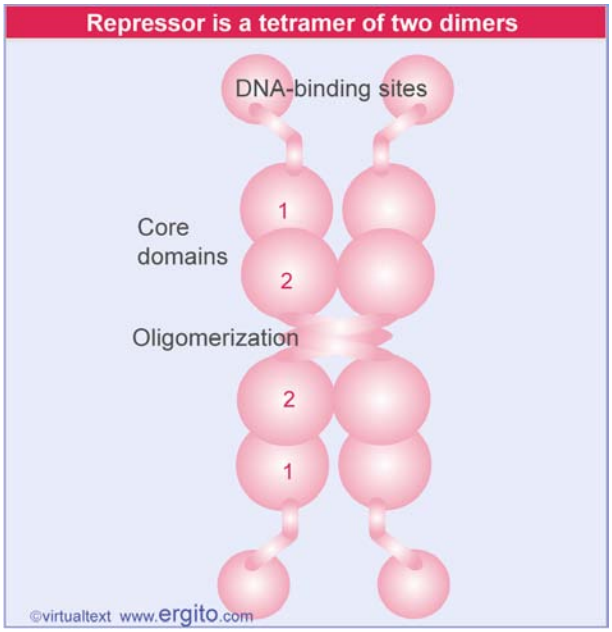

Figure 10.17 The repressor tetramer consists of two dimers. Dimers are held together by contacts involving core domain 2 as well as by the oligomerization helix. The dimers are linked into the tetramer by the oligomerization interface.

Last updated on January 26, 2004

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.10.13

THE OPERON

3.10.14 DNA-binding is regulated by an allosteric change in conformation

Key Concepts

- The DNA-binding domain of a monomer inserts into the major groove of DNA.
- Active repressor has a conformation in which the two DNA-binding domains of a dimer can insert into successive turns of the double helix.
- Inducer binding changes the conformation so that the two DNA-binding sites are not in the right geometry to make simultaneous contacts.

Early work suggested a model in which the headpiece is relatively independent of the core. It can bind to operator DNA by making the same pattern of contacts with a half-site as intact repressor. However, its affinity for DNA is many orders of magnitude less than that of intact repressor. The reason for the difference is that the dimeric form of intact repressor allows two headpieces to contact the operator simultaneously, each binding to one half-site. **Figure 10.18** shows that the two DNA-binding domains in a dimeric unit contact DNA by inserting into successive turns of the major groove. This enormously increases affinity for the operator.

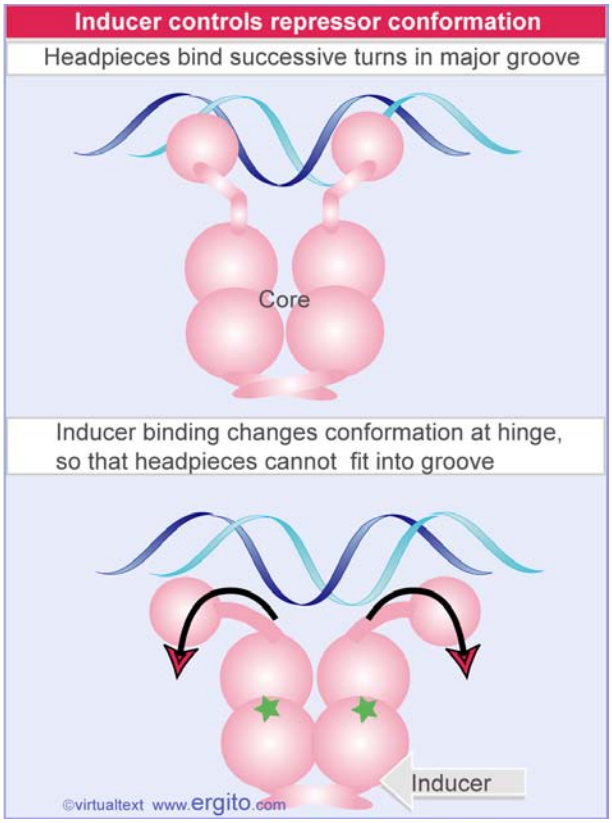

Figure 10.18 Inducer changes the structure of the core so that the headpieces of a repressor dimer are no longer in an orientation that permits binding to DNA.

Binding of inducer causes an immediate conformational change in the repressor protein. Binding of two molecules of inducer to the repressor tetramer is adequate to release repression. Binding of inducer changes the orientation of the headpieces relative to the core, with the result that the two headpieces in a dimer can no longer bind DNA simultaneously. This eliminates the advantage of the multimeric repressor, and reduces the affinity for the operator.

Last updated on 7-29-2002

This content is available online at http://www.ergito.com/main.jsp?bcs=MBIO.3.10.14

THE OPERON

3.10.15 Mutant phenotypes correlate with the domain structure

Key Concepts

- Different types of mutations occur in different domains of the repressor subunit.

Mutations in the Lac repressor identified the existence of different domains even before the structure was known (2568; 2569). We can now explain the nature of the mutations more fully by reference to the structure, as summarized in**Figure 10.19** (for review see 2567).

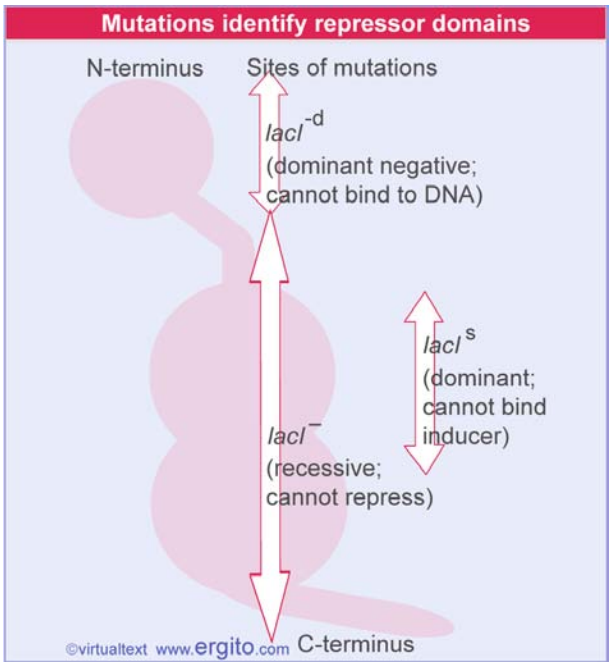

Figure 10.19 The locations of three type of mutations in lactose repressor are mapped on the domain structure of the protein. Recessive *lacI*$^{-}$ mutants that cannot repress can map anywhere in the protein. Dominant negative *lacI*$^{-d}$ mutants that cannot repress map to the DNA-binding domain. Dominant *lacI*s mutants that cannot induce because they do not bind inducer map to core domain 1.

Recessive mutations of the *lacI*$^{-}$ type can occur anywhere in the bulk of the protein. Basically any mutation that inactivates the protein will have this phenotype. The more detailed mapping of mutations on to the crystal structure in **Figure 10.16** identifies specific impairments for some of these mutations, for example, those that affect oligomerization.

The special class of dominant-negative *lacI*^{d} mutations lie in the DNA-binding site

of the repressor subunit (see *Molecular Biology 3.10.9 Multimeric proteins have special genetic properties*). This explains their ability to prevent mixed tetramers from binding to the operator; a reduction in the number of binding sites reduces the specific affinity for the operator. The role of the N-terminal region in specifically binding DNA is shown also by its location as the site of occurrence of "tight binding" mutations. These increase the affinity of the repressor for the operator, sometimes so much that it cannot be released by inducer. They are rare.

Uninducible *lacI*^{s} mutations map in a region of the core domain 1 extending from the inducer-binding site to the hinge. One group lies in amino acids that contact the inducer, and these mutations function by preventing binding of inducer. The remaining mutations lie at sites that must be involved in transmitting the allosteric change in conformation to the hinge when inducer binds.

Last updated on 7-29-2002

Reviews

2567. Pace, H. C., Kercher, M. A., Lu, P., Markiewicz, P., Miller, J. H., Chang, G., and Lewis, M. (1997). *Lac repressor genetic map in real space.* Trends Biochem. Sci. 22, 334-339.

References

2568. Markiewicz, P., Kleina, L. G., Cruz, C., Ehret, S., and Miller, J. H. (1994). *Genetic studies of the lac repressor. XIV. Analysis of 4000 altered E. coli lac repressors reveals essential and non-essential residues, as well as spacers which do not require a specific sequence.* J. Mol. Biol. 240, 421-433.

2569. Suckow, J., Markiewicz, P., Kleina, L. G., Miller, J., Kisters-Woike, B., and Muller-Hill, B. (1996). *Genetic studies of the Lac repressor. XV: 4000 single amino acid substitutions and analysis of the resulting phenotypes on the basis of the protein structure.* J. Mol. Biol. 261, 509-523.

THE OPERON

3.10.16 Repressor binds to three operators and interacts with RNA polymerase

Key Concepts

- Each dimer in a repressor tetramer can bind an operator, so that the tetramer can bind two operators simultaneously.
- Full repression requires the repressor to bind to an additional operator downstream or upstream as well as to the operator at *lacZ*.
- Binding of repressor at the operator stimulates binding of RNA polymerase at the promoter.

The allosteric transition that results from binding of inducer occurs in the repressor dimer. So why is a tetramer required to establish full repression?

Each dimer can bind an operator sequence. This enables the intact repressor to bind to two operator sites simultaneously. In fact, there are two further operator sites in the initial region of the *lac* operon. The original operator, *O1*, is located just at the start of the *lacZ* gene. It has the strongest affinity for repressor. Weaker operator sequences (sometimes called pseudo-operators) are located on either side; *O2* is 410 bp downstream of the startpoint, and *O3* is 83 bp upstream of it.

Figure 10.20 shows what happens when a DNA-binding protein can bind simultaneously to two separated sites on DNA. The DNA between the two sites forms a loop from a base where the protein has bound the two sites. The length of the loop depends on the distance between the two binding sites. When Lac repressor binds simultaneously to *O1* and to one of the other operators, it causes the DNA between them to form a rather short loop, significantly constraining the DNA structure A scale model for binding of tetrameric repressor to two operators is shown in **Figure 10.21** (511).

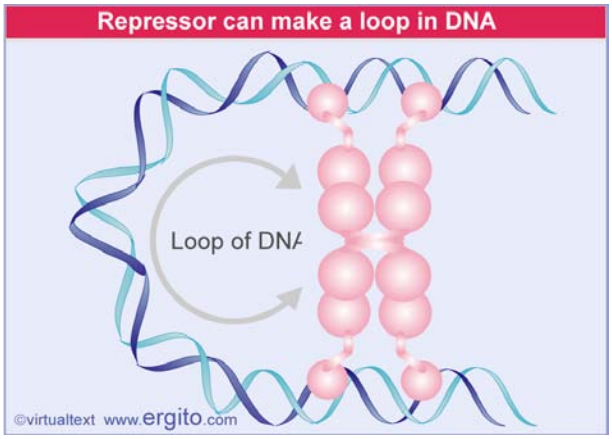

Figure 10.20 If both dimers in a repressor tetramer bind to DNA, the DNA between the two binding sites is held in a loop.

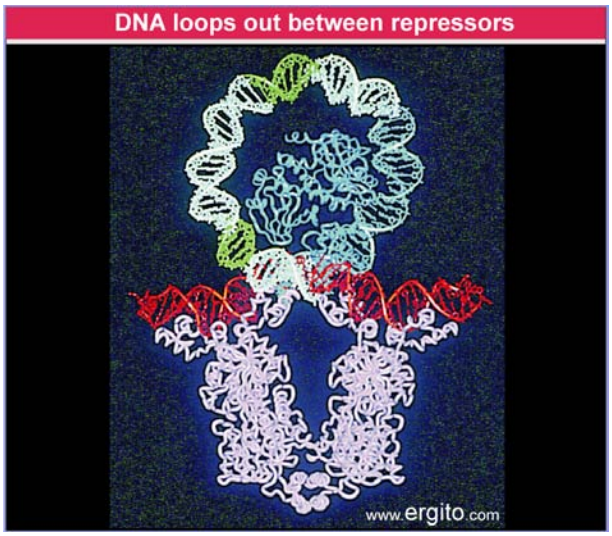

Figure 10.21 When a repressor tetramer binds to two operators, the stretch of DNA between them is forced into a tight loop. (The blue structure in the center of the looped DNA represents CAP, another regulator protein that binds in this region). Photograph kindly provided by Mitchell Lewis.

Binding at the additional operators affects the level of repression. Elimination of either the downstream operator (*O2*) or the upstream operator (*O3*) reduces the efficiency of repression by 2-4×. However, if *both O2* and *O3* are eliminated, repression is reduced 100×. *This suggests that the ability of the repressor to bind to one of the two other operators as well as to O1 is important for establishing repression.* We do not know how and why this simultaneous binding increases repression.

We know most about the direct effects of binding of repressor to the operator (*O1*). It was originally thought that repressor binding would occlude RNA polymerase from binding to the promoter. However, we now know that the two proteins may be bound to DNA simultaneously, and *the binding of repressor actually enhances the binding of RNA polymerase!* But the bound enzyme is prevented from initiating transcription.

The equilibrium constant for RNA polymerase binding alone to the *lac* promoter is 1.9×10^7 M^{-1}. The presence of repressor increases this constant by two orders of magnitude to 2.5×10^9 M^{-1}. In terms of the range of values for the equilibrium constant K_B given in **Figure 9.20**, repressor protein effectively converts the formation of closed complex by RNA polymerase at the *lac* promoter from a weak to a strong interaction.

What does this mean for induction of the operon? The higher value for K_B means that, when occupied by repressor, the promoter is 100 times more likely to be bound by an RNA polymerase. And by allowing RNA polymerase to be bound at the same time as repressor, it becomes possible for transcription to begin immediately upon induction, instead of waiting for an RNA polymerase to be captured.

The repressor in effect causes RNA polymerase to be stored at the promoter. The complex of RNA polymerase·repressor·DNA is blocked at the closed stage. When inducer is added, the repressor is released, and the closed complex is converted to an open complex that initiates transcription. The overall effect of repressor has been to speed up the induction process.

Does this model apply to other systems? The interaction between RNA polymerase, repressor, and the promoter/operator region is distinct in each system, because the operator does not always overlap with the same region of the promoter (see **Figure 10.26**). For example, in phage lambda, the operator lies in the upstream region of the promoter, and binding of repressor occludes the binding of RNA polymerase (see *Molecular Biology 3.12 Phage strategies*). So a bound repressor does not interact with RNA polymerase in the same way in all systems.

Last updated on 7-29-2002

References

511. Oehler, S. (1990). *The three operators of the lac operon cooperate in repression*. EMBO J. 9, 973-979.

THE OPERON

3.10.17 Repressor is always bound to DNA

Key Concepts

- Proteins that have a high affinity for a specific DNA sequence also have a low affinity for other DNA sequences.
- Every base pair in the bacterial genome is the start of a low-affinity binding-site for repressor.
- The large number of low-affinity sites ensures that all repressor protein is bound to DNA.
- Repressor binds to the operator by moving from a low-affinity site rather than by equilibrating from solution.

Probably all proteins that have a high affinity for a specific sequence also possess a low affinity for any (random) DNA sequence. A large number of low-affinity sites will compete just as well for a repressor tetramer as a small number of high-affinity sites. There is only one high-affinity site in the *E. coli* genome: the operator. The remainder of the DNA provides low-affinity binding sites. Every base pair in the genome starts a new low-affinity site. (Just moving one base pair along the genome, out of phase with the operator itself, creates a low-affinity site!) So there are 4.2×10^6 low-affinity sites.

The large number of low-affinity sites means that, even in the absence of a specific binding site, all or virtually all repressor is bound to DNA; none is free in solution. **Figure 10.22** shows how the equation of describing the equilibrium between free repressor and DNA-bound repressor can be rearranged to give the proportion of free repressor.

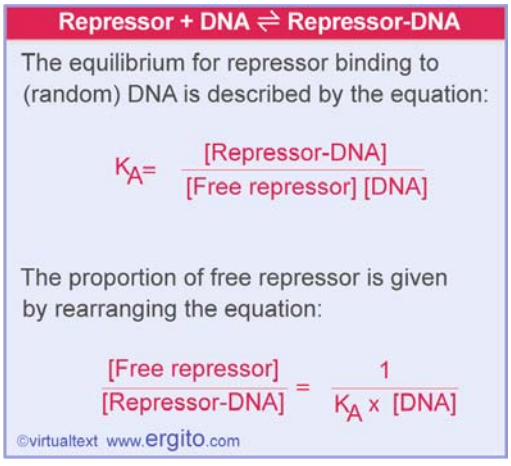

Figure 10.22 Repressor binding to random sites is governed by an equilibrium equation.

Applying the parameters for the *lac* system, we find that:

- The nonspecific equilibrium binding constant is $K_A = 2 \times 10^6\ M^{-1}$.
- The concentration of nonspecific binding sites is 4×10^6 in a bacterial volume of 10^{-15} liter, which corresponds to [DNA] = 7×10^{-3} M (a very high concentration).

Substituting these values gives: Free / Bound repressor = 10^{-4}.

So all but 0.01% of repressor is bound to (random) DNA. Since there are ~10 molecules of repressor per cell, this is tantamount to saying that there is no free repressor protein. This has an important implication for the interaction of repressor with the operator: it means that we are concerned with the *partitioning* of the repressor on DNA, in which the single high-affinity site of the operator *competes* with the large number of low-affinity sites.

In this competition, the absolute values of the association constants for operator and random DNA are not important; what is important is the ratio of K_{sp} (the constant for binding a specific site) to K_{nsp} (the constant for binding any random DNA sequence), that is, the specificity.

THE OPERON

3.10.18 The operator competes with low-affinity sites to bind repressor

Key Concepts

- In the absence of inducer, the operator has an affinity for repressor that is $10^7\times$ that of a low affinity site.
- The level of 10 repressor tetramers per cell ensures that the operator is bound by repressor 96% of the time.
- Induction reduces the affinity for the operator to $10^4\times$ that of low-affinity sites, so that only 3% of operators are bound.
- Induction causes repressor to move from the operator to a low-affinity site by direct displacement.
- These parameters could be changed by a reduction in the effective concentration of DNA *in vivo*.

We can define the parameters that influence the ability of a regulator protein to saturate its target site by comparing the equilibrium equations for specific and nonspecific binding. As might be expected intuitively, the important parameters are:

- The size of the genome dilutes the ability of a protein to bind specific target sites.
- The specificity of the protein counters the effect of the mass of DNA.
- The amount of protein that is required increases with the total amount of DNA in the genome and decreases with the specificity.
- The amount of protein also must be in reasonable excess of the total number of specific target sites, so we expect regulators with many targets to be found in greater quantities than regulators with fewer targets.

Figure 10.23 compares the equilibrium constants for *lac* repressor/operator binding with repressor/general DNA binding. From these constants, we can deduce how repressor is partitioned between the operator and the rest of DNA, and what happens to the repressor when inducer causes it to dissociate from the operator.

Repressor specifically binds operator DNA		
DNA	Repressor	Repressor + inducer
Operator	2×10^{13}	2×10^{10}
Other DNA	2×10^{6}	2×10^{6}
Specificity	10^{7}	10^{4}
Operators bound	96%	3%
Operon is:	repressed	induced

Figure 10.23 Lac repressor binds strongly and specifically to its operator, but is released by inducer. All equilibrium constants are in M^{-1}.

Repressor binds ~10^7 times better to operator DNA than to any random DNA sequence of the same length. So the operator comprises a single high-affinity site that will compete for the repressor 10^7 better than any low-affinity (random) site. How does this ensure that the repressor can maintain effective control of the operon?

Using the specificity, we can calculate the distribution between random sites and the operator, and can express this in terms of occupancy of the operator. If there are 10 molecules of *lac* repressor per cell with a specificity for the operator of 10^7, the operator will be bound by repressor 96% of the time. The role of specificity explains two features of the *lac* repressor-operator interaction:

- When inducer binds to the repressor, the affinity for the operator is reduced by ~10^3-fold. The affinity for general DNA sequences remains unaltered. So the specificity is now only 10^4, which is insufficient to capture the repressor against competition from the excess of 4.2×10^6 low-affinity sites. Only 3% of operators would be bound under these conditions.
- Mutations that reduce the affinity of the operator for the repressor by as little as 20-30× have sufficient effect to be constitutive. Within the genome, the mutant operators can be overwhelmed by the preponderance of random sites. The occupancy of the operator is reduced to ~50% if the repressor's specificity is reduced just 10× .

The consequence of these affinities is that in an uninduced cell, one tetramer of repressor usually is bound to the operator. All or almost all of the remaining tetramers are bound at random to other regions of DNA, as illustrated in **Figure 10.24**. There are likely to be very few or no repressor tetramers free within the cell.

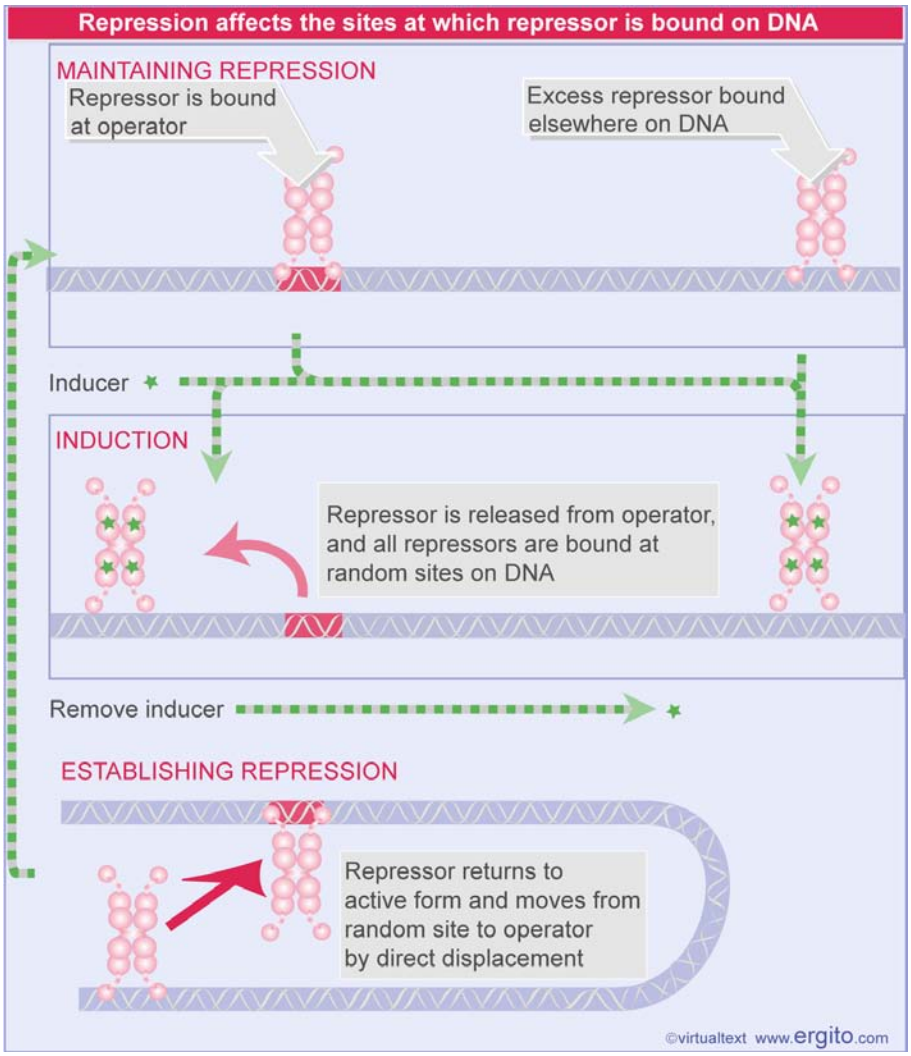

Figure 10.24 Virtually all the repressor in the cell is bound to DNA.

The addition of inducer abolishes the ability of repressor to bind specifically at the operator. Those repressors bound at the operator are released, and bind to random (low-affinity) sites. So in an induced cell, the repressor tetramers are "stored" on random DNA sites. In a noninduced cell, a tetramer is bound at the operator, while the remaining repressor molecules are bound to nonspecific sites. *The effect of induction is therefore to change the distribution of repressor on DNA, rather than to generate free repressor.*

When inducer is removed, repressor recovers its ability to bind specifically to the operator, and does so very rapidly. This must involve its movement from a nonspecific "storage" site on DNA. What mechanism is used for this rapid movement? The ability to bind to the operator very rapidly is not consistent with the time that would be required for multiple cycles of dissociation and reassociation with nonspecific sites on DNA. The discrepancy excludes random-hit mechanisms for finding the operator, suggesting that the repressor can move directly from a random site on DNA to the operator. This is the same issue that we encountered previously with the ability of RNA polymerase to find its promoters (see **Figure 9.22** and

Figure 9.23). The same solution is likely: movement could be accomplished by direct displacement from site to site (as indicated in **Figure 10.24**). A displacement reaction might be aided by the presence of more binding sites per tetramer (four) than are actually needed to contact DNA at any one time (two).

The parameters involved in finding a high-affinity operator in the face of competition from many low-affinity sites pose a dilemma for repressor. Under conditions of repression, there must be high specificity for the operator. But under conditions of induction, this specificity must be relieved. Suppose, for example, that there were 1000 molecules of repressor per cell. Then only 0.04% of operators would be free under conditions of repression. But upon induction only 40% of operators would become free. We therefore see an inverse correlation between the ability to achieve complete repression and the ability to relieve repression effectively. We assume that the number of repressors synthesized *in vivo* has been subject to selective forces that balance these demands.

The difference in expression of the lactose operon between its induced and repressed states *in vivo* is actually $10^3\times$. In other words, even when inducer is absent, there is a basal level of expression of ~0.1% of the induced level. This would be reduced if there were more repressor protein present, increased if there were less. So it could be impossible to establish tight repression if there were fewer repressors than the 10 found per cell; and it might become difficult to induce the operon if there were too many (514).

It is possible to introduce the *lac* operator-repressor system into the mouse. When the *lac* operator is connected to a tyrosinase reporter gene, the enzyme is induced by the addition of IPTG (2240). This means that the repressor is finding its target in a genome 10^3 times larger than that of *E. coli*. Induction occurs at approximately the same concentration of IPTG as in bacteria. However, we do not know the concentration of Lac repressor and how effectively the target is induced.

In order to extrapolate *in vivo* from the affinity of a DNA-protein interaction *in vitro*, we need to know the effective concentration of DNA *in vivo*. The "effective concentration" differs from the mass/volume because of several factors. The effective concentration is increased, for example, by molecular crowding, which occurs when polyvalent cations neutralize ~90% of the charges on DNA, and the nucleic acid collapses into condensed structures. The major force that decreases the effective concentration is the inaccessibility of DNA that results from occlusion or sequestration by DNA-binding proteins.

One way to determine the effective concentration is to compare the rate of a reaction *in vitro* and *in vivo* that depends on DNA concentration. This has been done using intermolecular recombination between two DNA molecules. To provide a control, the same reaction is followed as an intramolecular recombination, that is, the two recombining sites are presented on the same DNA molecule. We assume that concentration is the same *in vivo* and *in vitro* for the *intramolecular* reaction, and therefore any difference in the ratio of intermolecular/ intramolecular recombination rates can be attributed to a change in the effective concentration *in vivo*. The results of such a comparison suggest that the effective concentration of DNA is reduced >10-fold *in vivo* (516).

This could affect the rates of reactions that depend on DNA concentration, including DNA recombination, and protein-DNA binding. It emphasizes the problem encountered by all DNA-binding proteins in finding their targets with sufficient speed, and reinforces the conclusion that diffusion is not adequate (see **Figure 9.22**).

Last updated on 12-20-2001

References

514. Lin, S.-y. and Riggs, A. D. (1975). *The general affinity of lac repressor for E. coli DNA: implications for gene regulation in prokaryotes and eukaryotes*. Cell 4, 107-111.

516. Hildebrandt, E. R. et al. (1995). *Comparison of recombination in vitro and in E. coli cells: measure of the effective concentration of DNA in vivo*. Cell 81, 331-340.

2240. Cronin, C. A., Gluba, W., and Scrable, H. (2001). *The lac operator-repressor system is functional in the mouse.* Genes Dev. 15, 1506-1517.

THE OPERON

3.10.19 Repression can occur at multiple loci

Key Terms

Autogenous control describes the action of a gene product that either inhibits (negative autogenous control) or activates (positive autogenous control) expression of the gene coding for it.

Key Concepts

- A repressor will act on all loci that have a copy of its target operator sequence.

The *lac* repressor acts only on the operator of the *lacZYA* cluster. However, some repressors control dispersed structural genes by binding at more than one operator. An example is the *trp* repressor, which controls three unlinked sets of genes:

- An operator at the cluster of structural genes *trpEDBCA* controls coordinate synthesis of the enzymes that synthesize tryptophan from chorismic acid.
- An operator at another locus controls the *aroH* gene, which codes for one of the three enzymes that catalyze the initial reaction in the common pathway of aromatic amino acid biosynthesis.
- The *trpR* regulator gene is repressed by its own product, the *trp* repressor. So the repressor protein acts to reduce its own synthesis. This circuit is an example of **autogenous** control. Such circuits are quite common in regulatory genes, and may be either negative or positive (see *Molecular Biology 3.11.11 r-protein synthesis is controlled by autogenous regulation* and *Molecular Biology 3.12.9 Repressor maintains an autogenous circuit*).

A related 21 bp operator sequence is present at each of the three loci at which the *trp* repressor acts. The conservation of sequence is indicated in **Figure 10.25**. Each operator contains appreciable (but not identical) dyad symmetry. The features conserved at all three operators include the important points of contact for *trp* repressor. This explains how one repressor protein acts on several loci: *each locus has a copy of a specific DNA-binding sequence recognized by the repressor* (just as each promoter shares consensus sequences with other promoters).

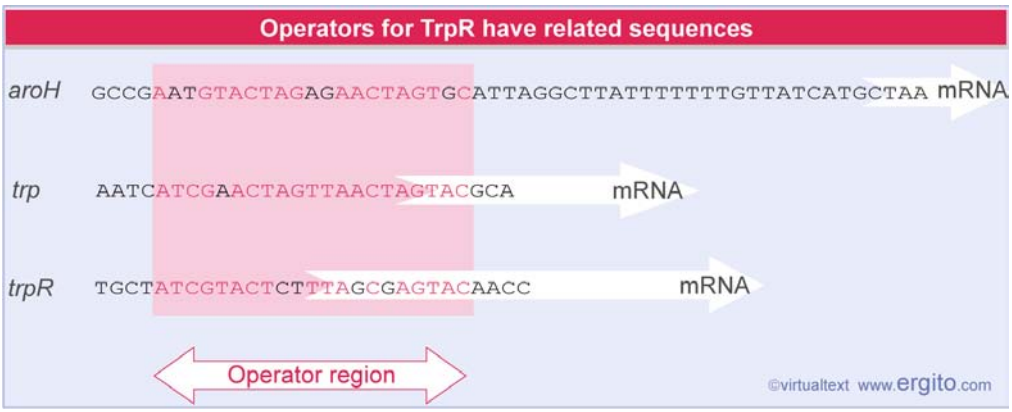

Figure 10.25 The *trp* repressor recognizes operators at three loci. Conserved bases are shown in red. The location of the startpoint and mRNA varies, as indicated by the white arrows.

Figure 10.26 summarizes the variety of relationships between operators and promoters. A notable feature of the dispersed operators recognized by TrpR is their presence at different locations within the promoter in each locus. In *trpR* the operator lies between positions -12 and +9, while in the *trp* operon it occupies positions -23 to -3, but in the *aroH* locus it lies farther upstream, between -49 and -29. In other cases, the operator lies downstream from the promoter (as in *lac*), or apparently just upstream of the promoter (as in *gal*, where the nature of the repressive effect is not quite clear). The ability of the repressors to act at operators whose positions are different in each target promoter suggests that there could be differences in the exact mode of repression, the common feature being that RNA polymerase is prevented from initiating transcription at the promoter.

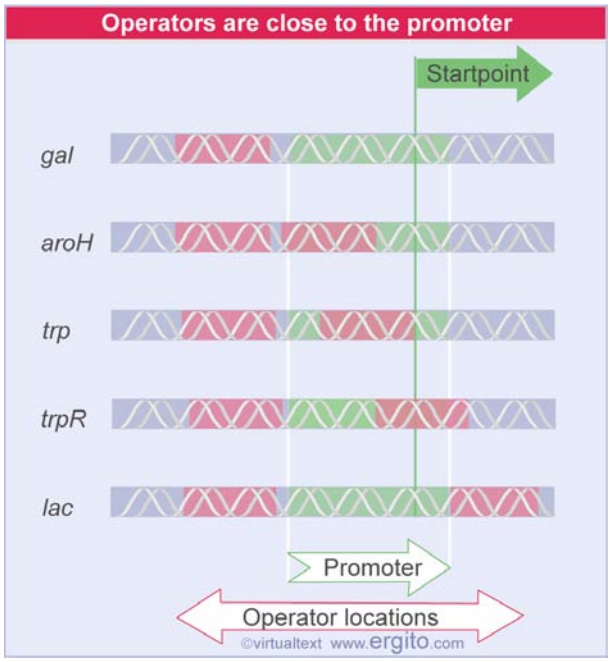

Figure 10.26 Operators may lie at various positions relative to the promoter.

THE OPERON

3.10.20 Summary

Transcription is regulated by the interaction between *trans*-acting factors and *cis*-acting sites. A *trans*-acting factor is the product of a regulator gene. It is usually protein but can be RNA. Because it diffuses in the cell, it can act on any appropriate target gene. A *cis*-acting site in DNA (or RNA) is a sequence that functions by being recognized *in situ*. It has no coding function and can regulate only those sequences that are physically contiguous with it. Bacterial genes coding for proteins whose functions are related, such as successive enzymes in a pathway, may be organized in a cluster that is transcribed into a polycistronic mRNA from a single promoter. Control of this promoter regulates expression of the entire pathway. The unit of regulation, containing structural genes and *cis*-acting elements, is called the operon.

Initiation of transcription is regulated by interactions that occur in the vicinity of the promoter. The ability of RNA polymerase to initiate at the promoter is prevented or activated by other proteins. Genes that are active unless they are turned off are said to be under negative control. Genes that are active only when specifically turned on are said to be under positive control. The type of control can be determined by the dominance relationships between wild type and mutants that are constitutive/derepressed (permanently on) or uninducible/super-repressed (permanently off).

A repressor protein prevents RNA polymerase either from binding to the promoter or from activating transcription. The repressor binds to a target sequence, the operator, that usually is located around or upstream of the startpoint. Operator sequences are short and often are palindromic. The repressor is often a homomultimer whose symmetry reflects that of its target.

The ability of the repressor protein to bind to its operator is regulated by a small molecule. An inducer prevents a repressor from binding; a corepressor activates it. Binding of the inducer or corepressor to its site produces a change in the structure of the DNA-binding site of the repressor. This allosteric reaction occurs in both free repressor proteins and directly in repressor proteins already bound to DNA.

The lactose pathway operates by induction, when an inducer β-galactoside prevents the repressor from binding its operator; transcription and translation of the *lacZ* gene then produce β-galactosidase, the enzyme that metabolizes β-galactosides. The tryptophan pathway operates by repression; the corepressor (tryptophan) activates the repressor protein, so that it binds to the operator and prevents expression of the genes that code for the enzymes that biosynthesize tryptophan. A repressor can control multiple targets that have copies of an operator consensus sequence.

A protein with a high affinity for a particular target sequence in DNA has a lower affinity for all DNA. The ratio defines the specificity of the protein. Because there are many more nonspecific sites (any DNA sequence) than specific target sites in a genome, a DNA-binding protein such as a repressor or RNA polymerase is "stored" on DNA; probably none or very little is free. The specificity for the target sequence must be great enough to counterbalance the excess of nonspecific sites over specific

sites. The balance for bacterial proteins is adjusted so that the amount of protein and its specificity allow specific recognition of the target in "on" conditions, but allow almost complete release of the target in "off" conditions.