DOWNLOADED BY
KAMI

Edited by Foxit Reader
Copyright(C) by Foxit Software Company,2005-2006
For Evaluation Only.

MICROBIAL GENETICS

Gene Regulation: The Operons

www.kamishin@yahoo.com

Pradeep Kumar Burma
Reader
Department of Genetics
University of Delhi South Campus
Benito Juarez Road
New Delhi-110021
E-mail: pburma@hotmail.com

05-May-2006 (revised 06-Oct-2006)

CONTENTS

Introduction
Genes can be constitutive, inducible or repressible
Inducible and repressible systems are built using positive and negative control mechanisms
What are operons
The Lactose Operon
    Induction of *lac* operon
    Mutants of *lac* operon
Catabolite repression
The Tryptophan Operon
The Arabinose Operon

Keywords

Gene regulation; Operon; Lactose operon; Tryptophan operon; Arabinose operon; Catabolite repression

Introduction

Genes govern the structure and function of a cell. All multicellular and complex organisms like humans, starts their development from a single celled zygote which divides to produce millions of cells. As this cell and its progeny divides, they maintain the same genomic content, yet groups of cells differentiate to perform different functions of the human body. This is achieved by switching 'ON' and 'OFF' different subsets of genes which are part of each individual cell. Thus, the phenotype (both structurally as well as functionally) achieved by a cell is dependent upon what genes have been expressed during its developmental history. Unicellular organisms like *E.coli* though simplistic in organization as compared to eukaryotes also respond to changes in their environment, as well as to their intracellular needs and thus need to regulate expression of their genes. For example, if *E.coli* cells find lactose in their environment to be used as a sole carbon/energy source, it would need to switch ON a set of genes to enable it to metabolize lactose. In another scenario, *E.coli* cells synthesize different amino acids, which accumulates in the cell. Once a certain level of a given amino acid is attained in the cell, it would be wasteful for *E.coli* to keep synthesizing it, so there is a need to switch OFF the genes which encodes the enzymes required for the synthesis of the said amino acid. This precise control of 'when', 'where' and 'what' genes are to be switched 'ON' and 'OFF' is called 'Gene Regulation'.

This paradigm of Gene Regulation finds its beginning in the seminal work of Francois Jacob, Jacques Monod and their colleagues. Their work at the Pasteur Institute at Paris led to the concept of 'Operon model of gene regulation'. Using utilization of lactose by *E.coli* as their model, their work laid the foundation for all the incisive work on gene control both in prokaryotes and eukaryotes. They along with Andre' Lwoff were awarded the Noble Prize in Physiology and Medicine in 1965.

Work of Jacob and Monod led to major conceptual changes as well as introduced novel ideas. Well before Jacob and Monod initiated their work, it was known that the enzyme β-galactosidase which could cleave lactose into glucose and galactose was only available in the cell in presence of its substrate (i.e. lactose). Scientists conceptualized that activation of β-galactosidase in the presence of lactose was a case of **'enzyme adaptation'** i.e. the inactive form of β-galactosidase already present in the cell was activated in the presence of lactose. Jacob and Monod demonstrated that this was not true. It was a case of **"enzyme induction**" – that is the enzyme was synthesized only in the presence of lactose. They further went on to show that the synthesis was controlled at the level of the gene. They were the first to introduce the concept of mRNA which could be involved as an intermediate molecule between the source of information, the gene (DNA) to the final product (the protein).

Gene regulation can take place at any stage in the pathway of information flow. The simplest depiction of this pathway in prokaryotes is shown in Fig. 1. There are five major levels at which a gene could be regulated in prokaryotes. Regulation could occur at the transcriptional level, i.e. whether a gene is switched ON or 'OFF' or at post-transcriptional level, which includes RNA processing to generate a mature RNA, the stability of mRNA and signals which could determine whether a mRNA is to be translated or not. The last step of regulation could be the post-translational modification of proteins before it performs its function. All points in the pathway of information flow thus act as hubs for controlling gene expression. In the present chapter, we will look at the examples of how gene regulation at the transcriptional level can be controlled.

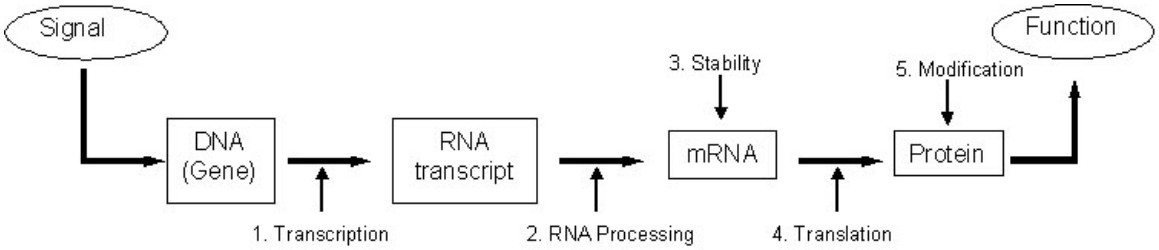


Fig. 1 The path of information flow: from signal to function

Genes can be constitutive, inducible or repressible

The *E. coli* genome encodes approximately 4000 and odd polypeptides whose concentration in the cell is not the same. While some proteins are present as few molecules per cell others are present in thousands.. This reflects the existence of some fundamental regulatory mechanisms that control the expression of genetic information. Further, some genes may be kept ON through out the life cycle of the cell, while others are turned ON for a small period in the cell's life.

In prokaryotes, gene that are involved in housekeeping functions like encoding rRNAs, tRNAs, ribosomal proteins etc are continuously kept 'ON' (transcribed) as these products are needed through out its life span. Such genes are called as **constitutive genes**.

The second class of genes are **inducible**. These genes are switched ON in response to some signal. For example, *E. coli* preferentially uses glucose as its energy source. However, it can use other carbohydrates like arabinose, galactose and lactose in the absence of glucose if these are present in the medium. These sugars have to be broken down – catabolised to enable the cell to utilize it. The enzymes that are needed for this catabolism is normally absent in the cell and are synthesized only when the concerned sugar is present. In other words, the presence of a carbohydrate like lactose in the medium signals *E.coli* to switch ON the genes which encode the enzymes needed for the utilization (breakdown) of lactose by the cell. Genes that code for protein(s) that are part of catabolic pathways are generally inducible. Inducible genes can be monitored by exposing *E. coli* cells to the appropriate signal e.g. lactose and analyzing the synthesis of the product (activity of enzymes involved in lactose utilization) over a period of time. A typical induction profile is presented in Fig. 2a. In this example, the activity of the enzyme involved in breakdown of lactose (β-galactosidase) is extremely low in the absence of lactose. On addition of lactose in the medium the level of the same enzyme increases significantly with time.

A third class of genes are those which are **r**epressible. Generally, genes encoding enzymes that are components of anabolic pathways (e.g. synthesis of tryptophan) are repressible in **nature**. These genes are generally transcriptionally ON but get switched OFF in the presence of a signal (high levels of tryptophan, in the above example). Taking the example of tryptophan synthesis, enzymes needed for synthesis of tryptophan are available in the cell till the time enough tryptophan accumulates. This accumulation of tryptophan then acts as a signal to shut down the gene activity involved in encoding enzymes for tryptophan synthesis. This can be experimentally mimicked by adding tryptophan in a culture of growing *E. coli* cells. In this case (Fig. 2b) activity of tryptophan synthatase (one of the enzymes involved in

tryptophan synthesis) will be repressed following addition of tryptophan. Thus, *E. coli* not only responds to the changes in the environment by controlling expression of its genes but also conserves its energy by curbing the wasteful expression of gene function.

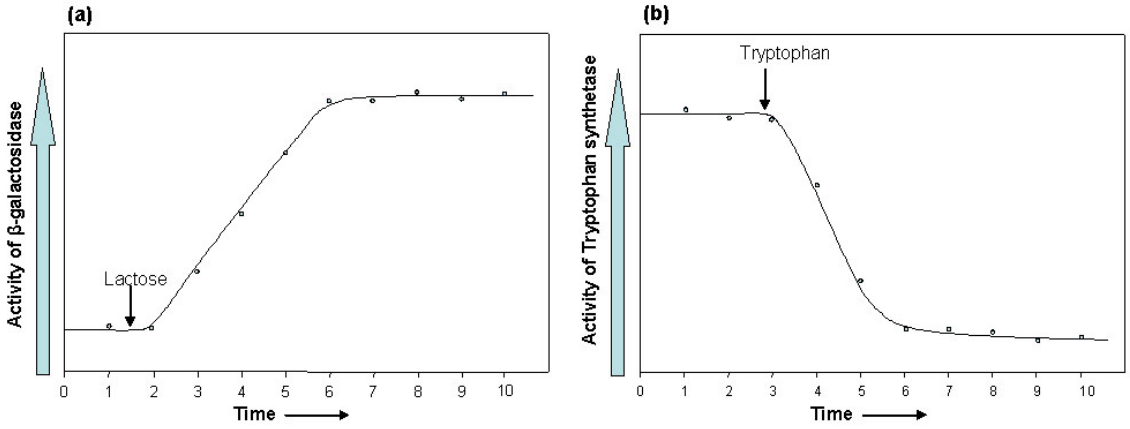


Fig. 2 Profile of activity of enzymes which are typically (a) induced or (b) repressed following perception of the signal, to a growing *E.coli* cell population

Inducible and repressible systems are built using positive and negative control mechanisms

Genes can be induced (switched ON) or repressed (switched OFF) by using either a positive control or a negative control mechanism. A positive control (Fig. 3a) envisages the need for a regulatory molecule which is required for expression of a gene. In other words, the gene will only be transcribed if this molecule is present and binds to the regulator binding site of the gene to be transcribed. In the absence of its binding, the gene will be OFF. On the other hand in a negative control (Fig. 3b) the binding of the regulatory molecule shuts down the expression of a gene while its lack of binding will switch the gene ON. These regulatory proteins influence the transcription of the gene which is carried out by RNA polymerases that binds to the promoter region. Thus, positive regulators induce transcription, while negative regulators repress transcription. It should be kept in mind that the concept of positive and negative regulator is different from inducibility or repressibility of gene expression. In fact, an inducible or a repressible system can be developed, using either a positive or a negative control mechanism. Of course the mechanisms used to achieve these states would be different (Fig. 4).

An inducible system using a negative control can be achieved by the presence of a negative regulator (repressor) which blocks transcription of the gene. The presence of the inducer does not allow the repressor to bind to the regulatory site. This in turn allows the RNA polymerase to transcribe the gene (switched 'ON'; Fig. 4a). An inducible system can also be built up using a positive control. In this case, the binding of the positive regulator (lets call it an **activator**) is needed to enable the RNA polymerase to promote transcription. However, this activator is in an inactive form in the absence of an inducer and thus cannot bind to the regulatory region and consequently the gene is OFF. The presence of the inducer activates the activator which can now bind to the regulatory site, and promotes RNA polymerase to transcribe the gene. The gene is switched ON (Fig. 4a).

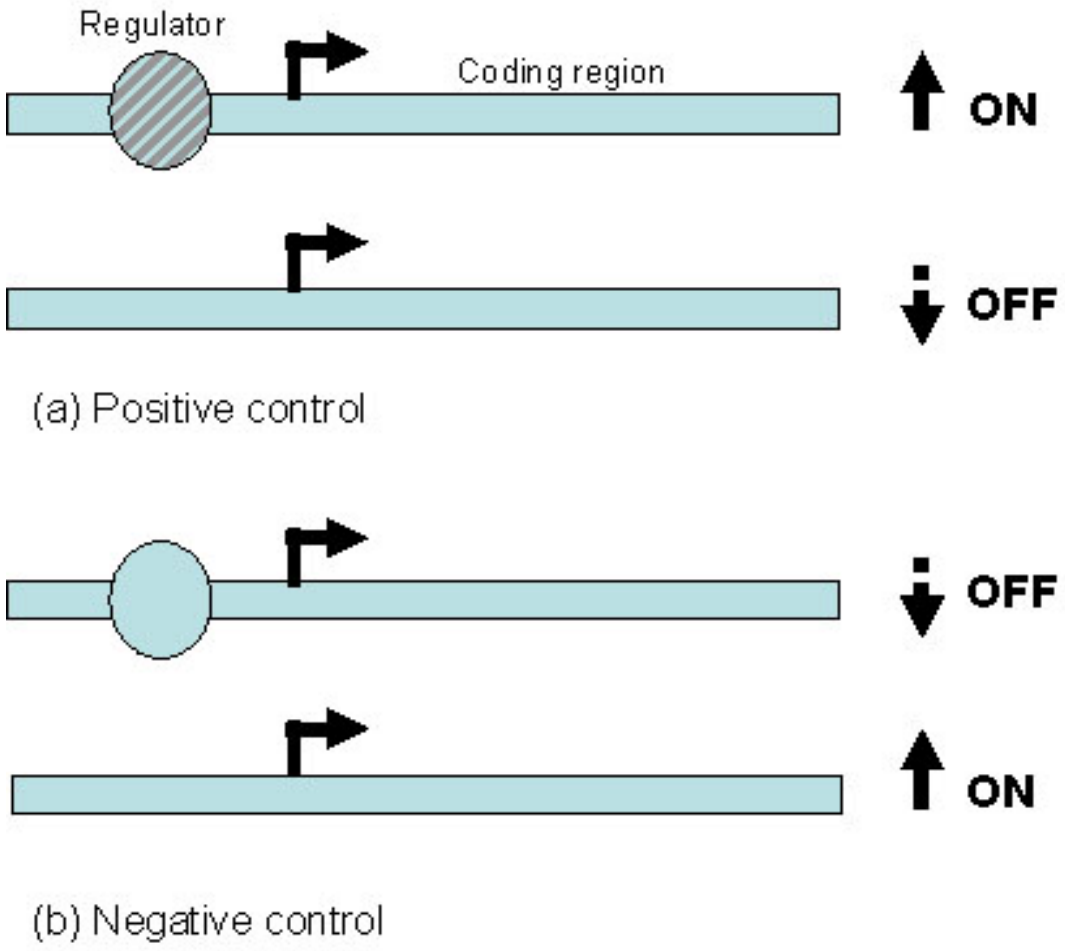
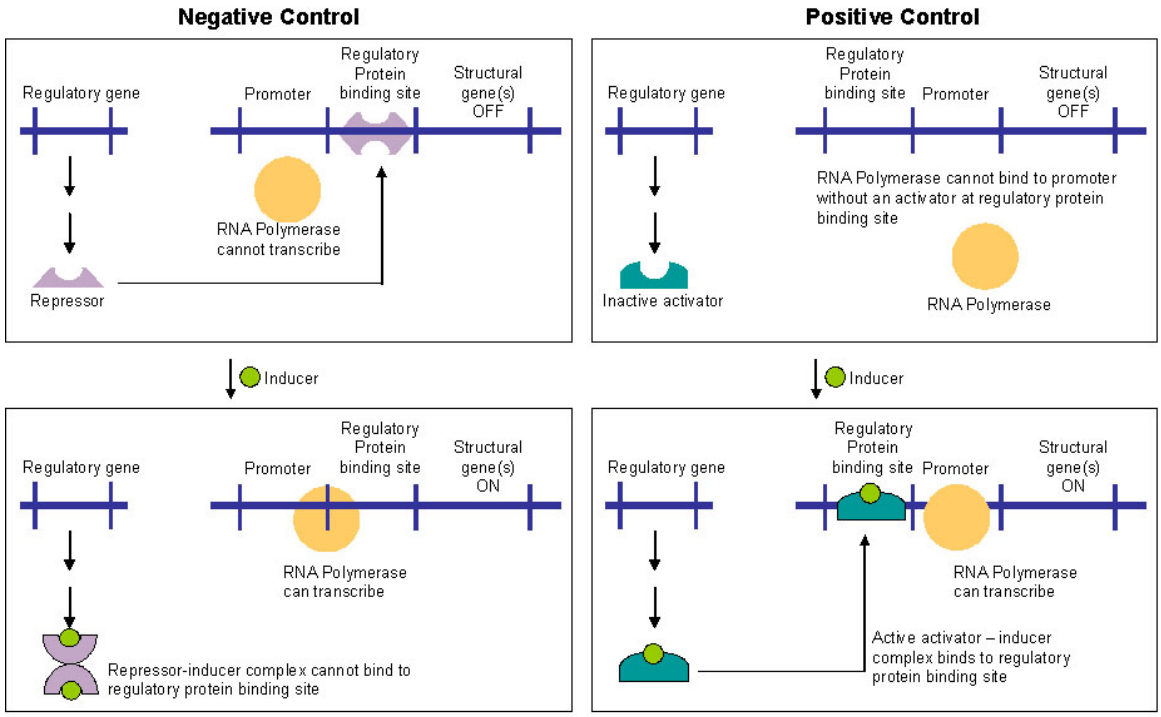


Fig. 3 Positive and Negative Control

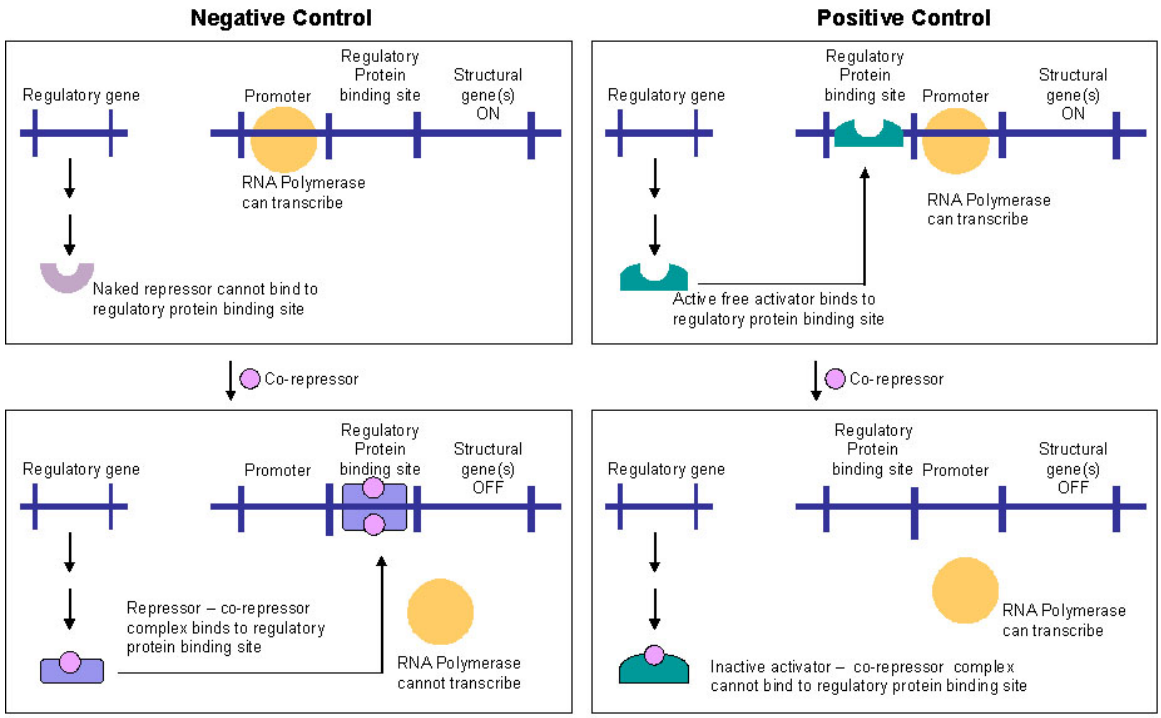
Similarly, repressible systems can be built using negative and positive regulators. A repressible system using a negative control envisages the presence of a negative regulator (repressor) which is normally inactive. This inactive repressor is unable to bind to the regulatory site and block gene expression. In the presence of the signaling molecule (called as a **co-repressor**), the repressor is activated to bind to its regulatory site inhibiting transcription of the gene. The gene is thus switched OFF (Fig. 4b). A repressible system using a positive control mechanism needs the presence of an activator which can positively regulate transcription of a gene, in the absence of the co-repressor, so that the gene is ON. However, in the presence of a co-repressor, the activator is inactivated and thus cannot promote transcription of the gene (switching it OFF, Fig. 4b).

These regulatory mechanisms are built using the following components that were identified by detailed genetic and biochemical analyses: (i) structural genes: genes that encode the proteins needed for a specific function (e.g., catabolism of lactose), (ii) regulatory genes: those that encode the regulatory protein (i.e., the repressors or activators) and (iii) the signal:

which will either induce (called as inducer, e.g. lactose) or repress (called as co-repressor, e.g. tryptophan).



(a) Inducible system



(b) Repressible system

Fig. 4 Organization of (a) an inducible and (b) repressible expression system using negative and positive control mechanism

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

We will now look into specific examples of how these mechanisms are operational in *E. coli* to regulate its gene expression taking examples of three different OPERONS viz., lac (lactose) trp (tryptophan) and ara (arabinose).

What are operons?

For realizing a function, all genes involved in a given pathway of anabolism or catabolism have to be switched ON or OFF simultaneously. In order to co-ordinate this regulation, a set of structural genes form one transcriptional unit which is under the control of a single promoter and other regulatory sequences (e.g. the binding site for the regulatory protein). Such a functional unit is referred to as an operon. Thus, all the structural genes of an operon can be switched ON or OFF in a coordinated fashion and a polycistronic mRNA is transcribed. A polycistronic mRNA is one which carries information for more than one protein. Following transcription, the different proteins encoded by this mRNA are translated individually. Operons are thus defined as functionally coordinated group of genes producing polycistronic transcripts (Fig. 5). Although operons were thought to occur only in the prokaryotes, these are also found in higher organisms like the nematode, Caenorhabditis. Many genes of C. elegans seem to be coordinately regulated and transcribed into polycistronic RNA that is processed into mRNA by trans-splicing.

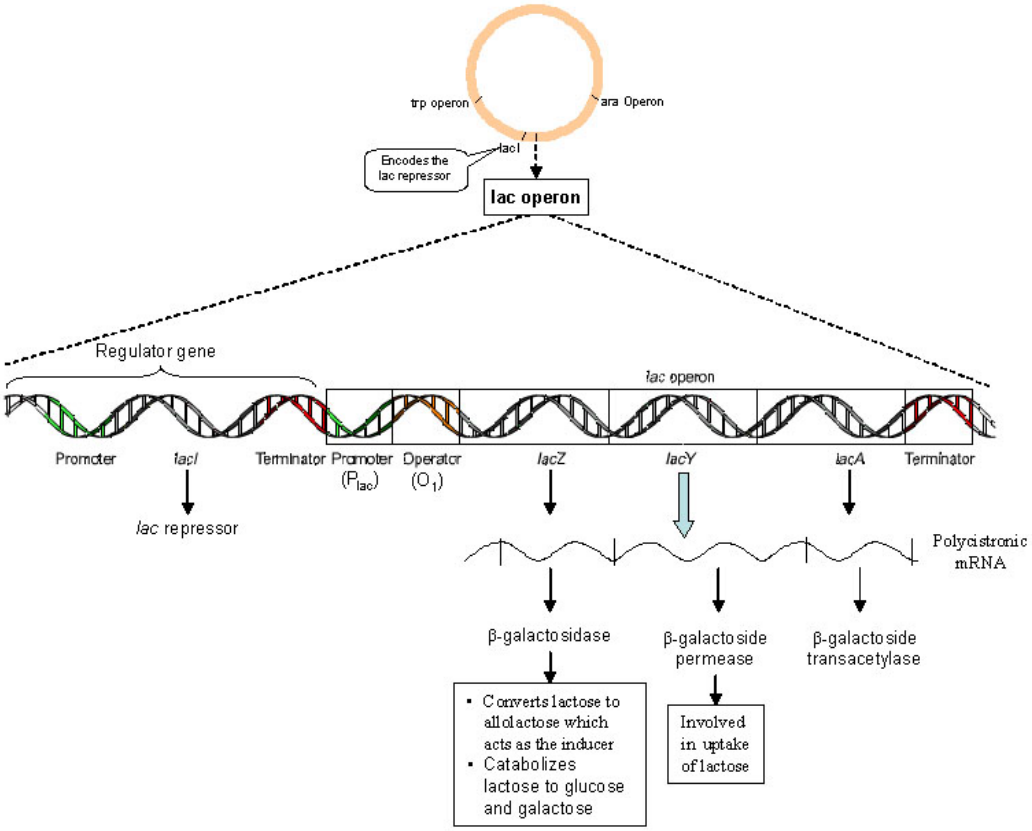


Fig. 5(a) Structural organization of the lac operon. The lacI gene encoding the lac repressor (the regulatory protein) is not part of the operon

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

Coordinate control of a set of genes which are not contiguous (that is not present as a single transcriptional unit) can also be carried out by regulating their expression using the same regulatory protein. These set of genes constitute what is called as regulons.

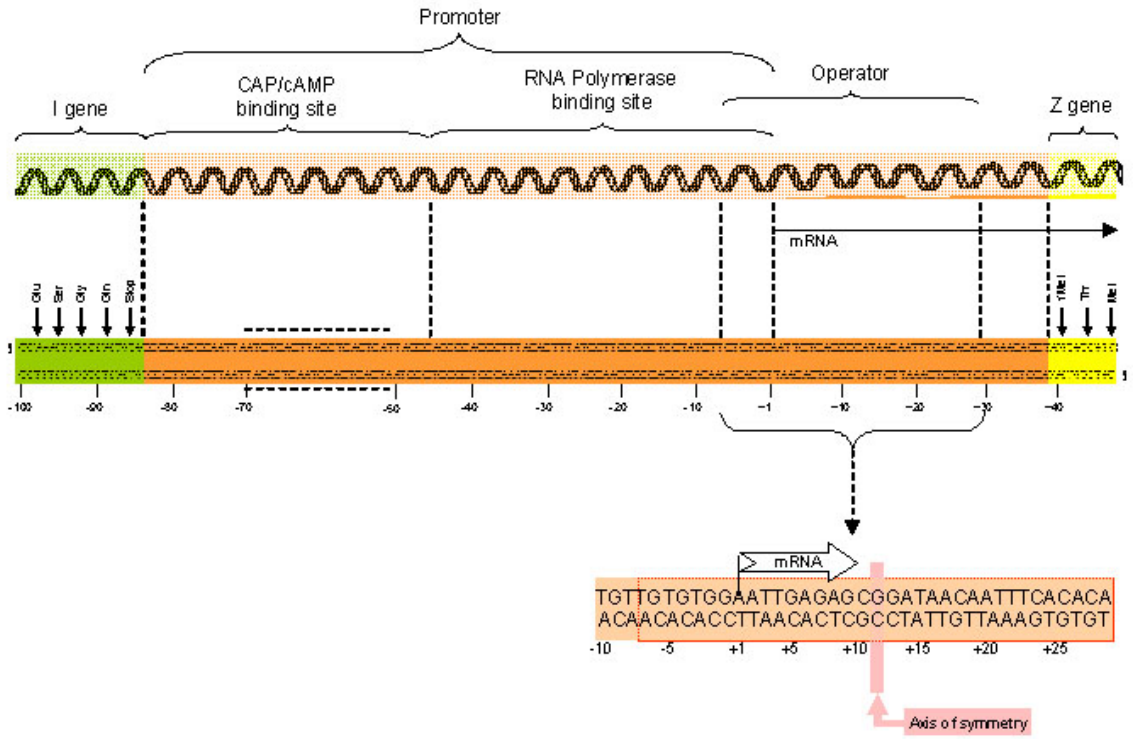


Fig. 5(b) The organization of the DNA region which regulates the transcription of the lac operon

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

The Lactose Operon (the classic example of an inducible operon)

A set of structural genes (constituting the lac operon) which are involved in the utilization of lactose are found to be coordinately switched ON (induced) when *E. coli* cells find lactose in the medium. This induction is however, dependent on the absence of glucose in the medium, i.e., in the presence of glucose the operon is not switched on even if lactose is added. Thus, the regulation of the *lac* operon involves (i) induction by lactose and (ii) repression by glucose (as we will see later this has been termed as catabolite repression or glucose effect). First let us see the organization of the *lac* operon. The operon has three structural genes : *lacZ* encoding the enzyme β-galactosidase; *lacY*, β-galactoside permease; and *lacA*, β-galactoside acetyl transferase (Fig. 5). The structural genes are under the control of a promoter P_{lac} which contains the biding site for RNA polymerase as well as for the regulatory proteins : the repressor (called as lac repressor) and the activator (cAMP-Catabolite Activator Protein complex). The binding site for the repressor is the operator, while that for the activator is called as the CAP binding site. The repressor protein is encoded by the regulatory gene, *lacI*, which is not a part of the *lac* operon. It happens to lie adjacent to the *lac* operon. In other examples like the *trp* operon, the regulatory genes are at different location than the operon itself.

Induction of lac operon

In *E. coli* cells growing in a medium devoid of lactose, the *lac* operon is in a repressed state (switched OFF). Addition of the disaccharide lactose to the medium derepresses or induces the operon leading to the transcription of the polycistronic mRNA encoding the three proteins LacZ, LacY and LacA. Lac operon is thus an inducible operon. The repressed state is maintained by the binding of the lac repressor (encoded by *lacI* gene) to the operator (negative regulation). This binding blocks the RNA polymerase from transcribing the structural genes. On addition of lactose in the medium, the lactose is brought into the cell with the help of the protein β-galactoside permease (product of *lacY*) and converted to allolactose (Fig. 6) by the enzyme β-galactosidase (product of *lacZ*). The allolactose then binds to the repressor protein which in turn prevents it from binding to the operator. Once the operator is free of the bound repressor, RNA polymerase is able to transcribe the lac operon. The operon is thus induced. In the next section we will look at some of the questions on induction.

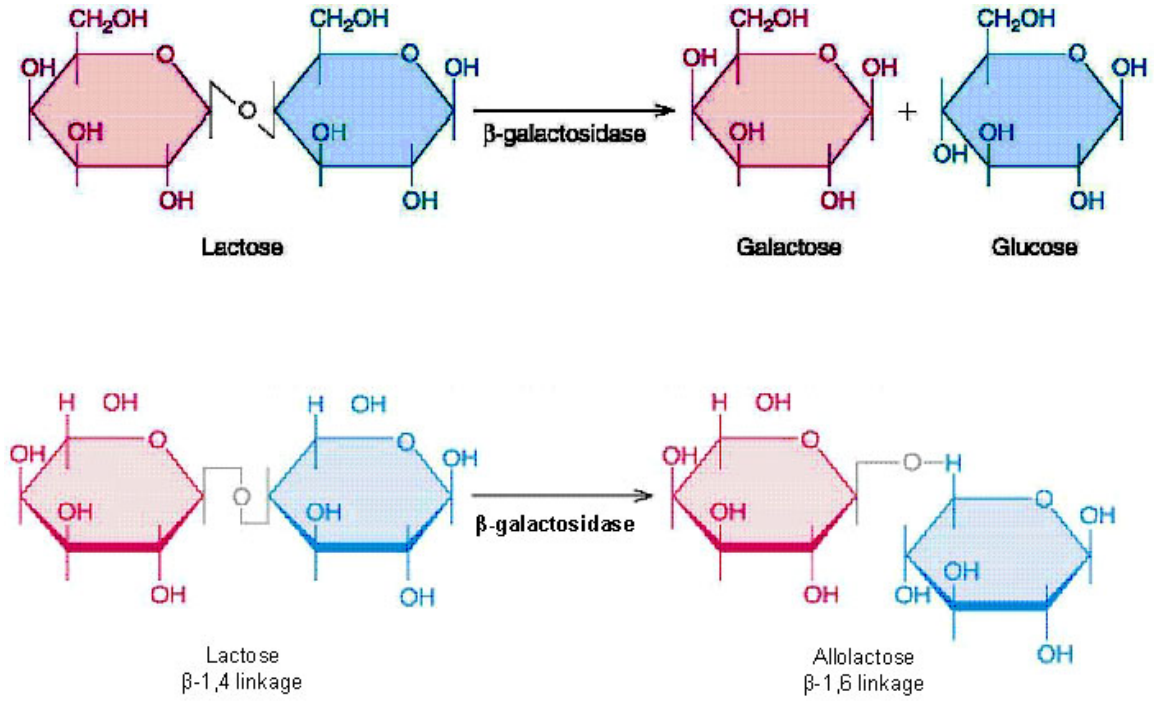


Fig. 6 Enzymatic actions of β-galactosidase

How is the lactose brought into the cell in the first place, if the operon was OFF in the absence of lactose?

Although we refer to the operon being repressed or OFF in the absence of lactose, in reality the operon is transcribed even in the absence of lactose but at an extremely low level (basal level). This happens as from time to time the repressor bound at the operator falls off on its own and the RNA polymerase is able to transcribe the genes. Thus, the cells always have few molecules of β-galactoside permease which is used to bring in the first few molecules of lactose which then leads to induction. It is thus more appropriate to say that an operon is repressed rather than OFF, as repressed means lowering down but not completely shutting it OFF.

How does the repressor function?

The repressor is an allosteric protein, that is its binding to one particular molecule (the inducer allolactose in this case) changes its conformation which in turn changes its ability to bind to the second particular entity/site (the operator). The repressor thus binds to the operator in the absence of the inducer. The Lac repressor binds as a tetramer and the protein would thus have sites for inducer binding, DNA binding as well as for oligomerization. The operator is a short stretch of DNA sequence lying just downstream to the RNA polymerase binding site. It is a palindrome, that is it is an inverted repeat with an axis of symmetry (Fig 5b). The repressor binds to the operator with a certain affinity such that transcription by RNA polymerase is blocked for most of the time. When allolactose is bound to the repressor, the ability of the repressor to bind to the operator is greatly reduced. Since no covalent bonds are involved in the repressor-operator binding, the repressor falls off the operator and the RNA polymerase is able to transcribe through the operon. Initially, one operator sequence (also called as O_1) just downstream to the promoter was identified. Later two more operator sequences called O_2 (within the *lacZ* gene) and O_3 (found further upstream to the *lac* promoter) were identified. Although all the three operators play a role in the regulation of *lac* operon O_1 has the maximum influence.

Which is the true inducer of lac operon ?

Although the presence of lactose induces the *lac* operon, allolactose is the true inducer as its binding to the repressor molecule de-represses the operon. Allolactose is derived from lactose as described earlier, and lactose here may be called as the signal molecule. Apart from allolactose several other molecules have been found to induce the *lac* operon. One of the most popular inducers used for inducing the *lac* operon in experiments has been the isopropyl β-D-thiogalactoside (IPTG). In fact in most of the work by Jacob and Monod, IPTG was used as the inducer. IPTG is, however, not a substrate for β-galactosidase. Such inducers are called as gratuitous inducers.

Mutants of lac operon

One of the ways of studying different components of any regulatory network is to generate mutants and analyze their behavior. This is also called as genetic analysis. The operon concept developed by Jacob and Monod was largely based on experiments with various mutants of the *lac* operon. We will now discuss some of the different mutant phenotypes observed for *lac* operon in light of what they tell us regarding regulation.

***E.coli* strain that expresses *lac* operon constitutively:** Such mutants also called as constitutive mutants express the genes of the *lac* operon even in the absence of the inducer (allolactose or IPTG). This could be possible if one of the components of the negative regulation (repression) is either lost or mutated. The two components involved in this are the repressor and its binding site on the DNA. Thus mutations in the *lacI* gene or the operator can lead to constitutive expression of the *lac* operon. The former is called as *lacI⁻* mutants while the latter are known as o^c (operator constitutive) mutants. In the first case, a functional repressor protein is not formed due to mutation in the *lacI* gene. In the second case the repressor is not able to bind to the operator due to mutations at this site.

The two types of constitutive mutants were distinguished by (i) mapping the site of mutation by standard mapping techniques, like recombination mapping and (ii) the use of **merodiploids** or **merozygotes**. Merodiploids (partial diploids) can be created by sexduction

in *E. coli*. In this case, one copy of the operon is present on the chromosome while the other copy is present on the F' factor. In this scenario, a constitutive operator (o^c) only affects the operon it is physically a part of (that is it controls those structural genes to which it is in ***cis***). On the other hand, a *lacI* gene mutation is recessive to its wild type regulator gene. Hence, the operon will be induced irrespective of which operon (chromosomal or F' factor) the mutation is on, as long as a wild-type copy of *lacI* is available. Results from a set of haploids and merodiploids are presented in Table 1.

Table 1: β-galactosidase activity observed in wild type, mutant and mero-diploid strains of *E. coli* . The galactosidase activity (normally expressed in Miller's Unit) observed in wild type *E. coli* following induction has been taken as 100 in these examples

Genotype	Galactosidase activity	
	Non-induced	Induced
$i^+ o^+ z^+ y^+$	10	100
$i^- o^+ z^+ y^+$	100	100
$i^+ o^c z^+ y^+$	100	100
$i^+ o^+ z^+ y^- / i^- o^+ z^- y^+$	10	100
$i^+ o^c z^+ y^+ / i^+ o^+ z^- y^+$	100	100
$i^+ o^+ z^+ y^+ / i^+ o^c z^- y^+$	10	100

***E.coli* mutants which are super-repressible:** Such strains remain uninduced even in the presence of the inducer. In reality a higher concentration of inducer is required to induce these strains as compared to that needed for the wild type. Logically this can happen if (i) there is an overproduction of the repressor protein, (ii) the repressor has lost the ability to bind the inducer, (iii) the repressor can bind to the inducer but has lost its allosteric property i.e. the inducer bound repressor can still bind to the operator or (iv) the mutated repressor binds to the operator more tightly. Mutants for two of these classes are known. I^s mutants code for a repressor which has lost its allosteric property, while I^q mutants overproduce the Lac repressor than normal due to a mutation in the promoter of the *lacI* gene which up regulates its activity.

Lac operon needs positive inputs for induction

Though the inability of the *lac* repressor to bind to the operator in the presence of the inducer (allolactose) removes the negative control yet the genes of the operon cannot be transcribed as the binding of a positive regulatory molecule is required. This regulatory protein is the CAP (Catabolite Activator Protein) which when bound by small effector molecule called cyclic AMP (cAMP) binds to the CAP binding site upstream to the *lac* promoter (Fig 5b). Binding of this protein positively regulates the transcription from the *lac* promoter.

The *lac* operon (an inducible operon) is thus under both negative and positive regulation. For the operon to be switched ON the repressor should be inactivated as well as the CAP-cAMP complex should bind to the promoter for its activation.

Catabolite repression

When *E .coli* cells are grown in presence of both glucose and lactose, the growth curve observed is as presented in Fig. 7. After the initial exponential growth, the cells reach the stationary phase after which the second phase of growth is initiated. Such a growth profile is called as a diauxie curve. If one was to assay for β-galactosidase activity during the entire growth period very low activity would be observed in the first phase of growth, while in the second phase there would be an increasing activity (Fig 7). This is due to the fact that in the first phase of growth only glucose is utilized. Once glucose is depleted from the medium the catabolism of lactose is initiated and this supports the second growth phase. Thus, the presence of glucose prevents the induction of the *lac* operon (as well as other operons controlling enzymes involved in carbohydrate catabolism). This phenomenon is called as catabolite repression or the glucose effect. Such a mechanism has evolved to ensure that as long as glucose is available, *E. coli* preferentially uses it, rather than wasting its energy in utilizing less efficient energy resources.

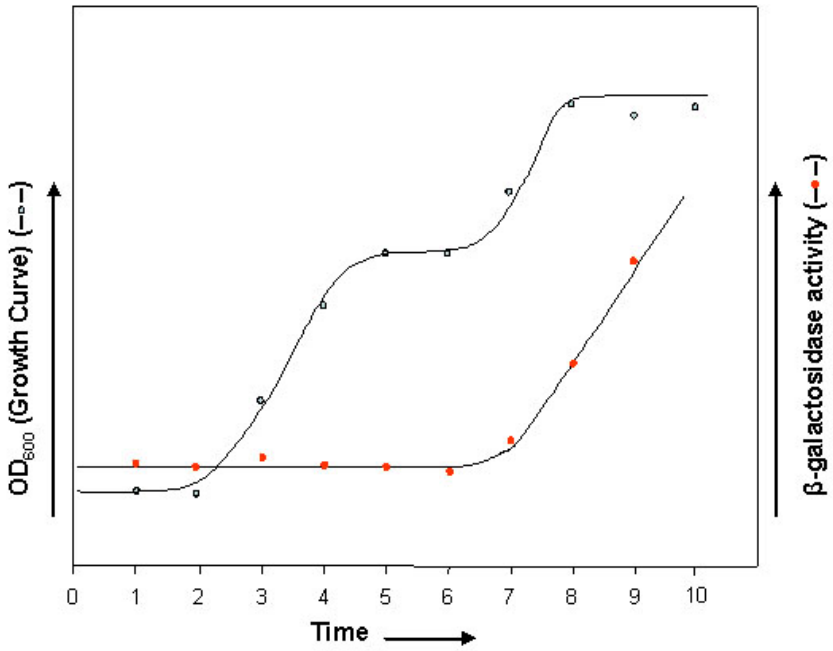


Fig. 7 Growth profile and β-galactosidase activity exhibited by *E.coli* cells when grown in a medium having glucose and lactose as carbon source

Initial studies on unraveling the mechanism of catabolite repression suggested that the presence of glucose somehow down regulates the levels of cAMP in the cell. The model proposed that in the presence of glucose in the medium, the internal concentration of cAMP was reduced. The low concentrations of cAMP will inhibit the formation of the cAMP-CRP complex which in turn would affect the positive regulatory pathway of the lac operon, thus not supporting the switching ON of the operon. In the absence of glucose, the cAMP concentration is appreciably high in the cell thus making the activation complex with CRP. This would lead to switching ON of the operon in the presence of alternative carbon source, such as lactose and when Lac repressor is also inactivated. Though thought to be the mechanism responsible for glucose-lactose diauxie in *E. coli*, later evidences suggested that the glucose effect is brought about by an entirely different mechanism named as inducer

exclusion. Under this model which has been experimentally supported, the intake of glucose into the cell by the phosphoenol-pyruvate dependent phopshotransferase system (PTS) decreases the level of phosphorylation of one of its components, the enzyme IIA^{Glc}. The resulting dephophorylated form then prevents the uptake of the inducer by binding to the Lac permease (Fig. 8). It should be noted that CRP-cAMP complex is still required for the positive regulation but does not play a role in catabolite repression. However in other examples of catabolite repression, the cAMP pathway is still thought to be the mechanism.

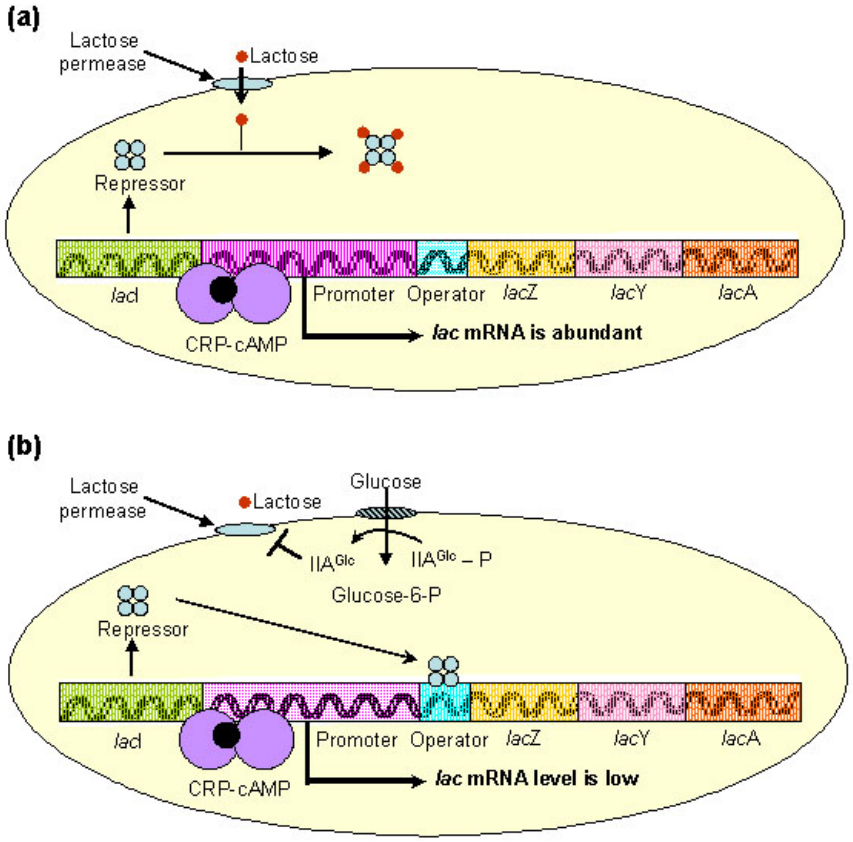


Fig. 8 Mechanism of catabolite repression. Lactose is taken into the cell by lactose permease in the (a) absence of glucose. Its activity is however inhibited in the (b) presence of glucose in the medium

(Source: Inada et al. Genes to Cells 1 (1996) : p. 293-301)

In summary, the lac operon is an example of an inducible operon demonstrating dual regulation, both negative and positive. The negative regulation is brought about by the repressor protein while the CRP-cAMP complex acts as the positive regulator. The operon is induced when the activity of the repressor is abolished in the presence of the inducer. The second control of this operon is the catabolite repression. In the presence of glucose, lactose fails to induce the operon as the uptake of glucose inhibits the uptake of lactose by the Lac permease protein.

The Tryptophan Operon (a repressible operon with an additional transcription attenuation control)

The tryptophan operon also known as *trp* (pronounced as 'trip") operon is involved in responding to the cells' need to produce the amino acid, tryptophan. The operon consists of a

set of five structural genes : *trp E* through *trp A* (Fig. 9) which encode for the three key enzymes needed for the synthesis of tryptophan from chorismic acid. The three enzymes are; Anthranilate synthetase (made up of two subunits), Indole glycerolphopshate synthetase, and Trptophan synthetase (made up of α and β subunits). The regulatory region of the *trp* operon consists of a major promoter (P_1) which drives the transcription of the five structural genes as a polycistronic RNA. Overlapping the P_1 promoter is the operator region (O) to which binds the repressor protein. An additional promoter called the P_2 promoter is found at the end of the *trpD* genes which can transcribe the genes *trp C, B* and *A*. We will not consider the regulatory control at the P_2 promoter in the present chapter. In addition to the promoters and operator there is a third control region called *trpL*. The *trpL* which is the first region to be transcribed under P_1 promoter encodes a 162 nucleotide long mRNA leader sequence. Although this region is translated, the small protein encoded by *trpL* has no function in tryptophan synthesis. This region provides a second control mechanism called attenuation which we will discuss a little later.

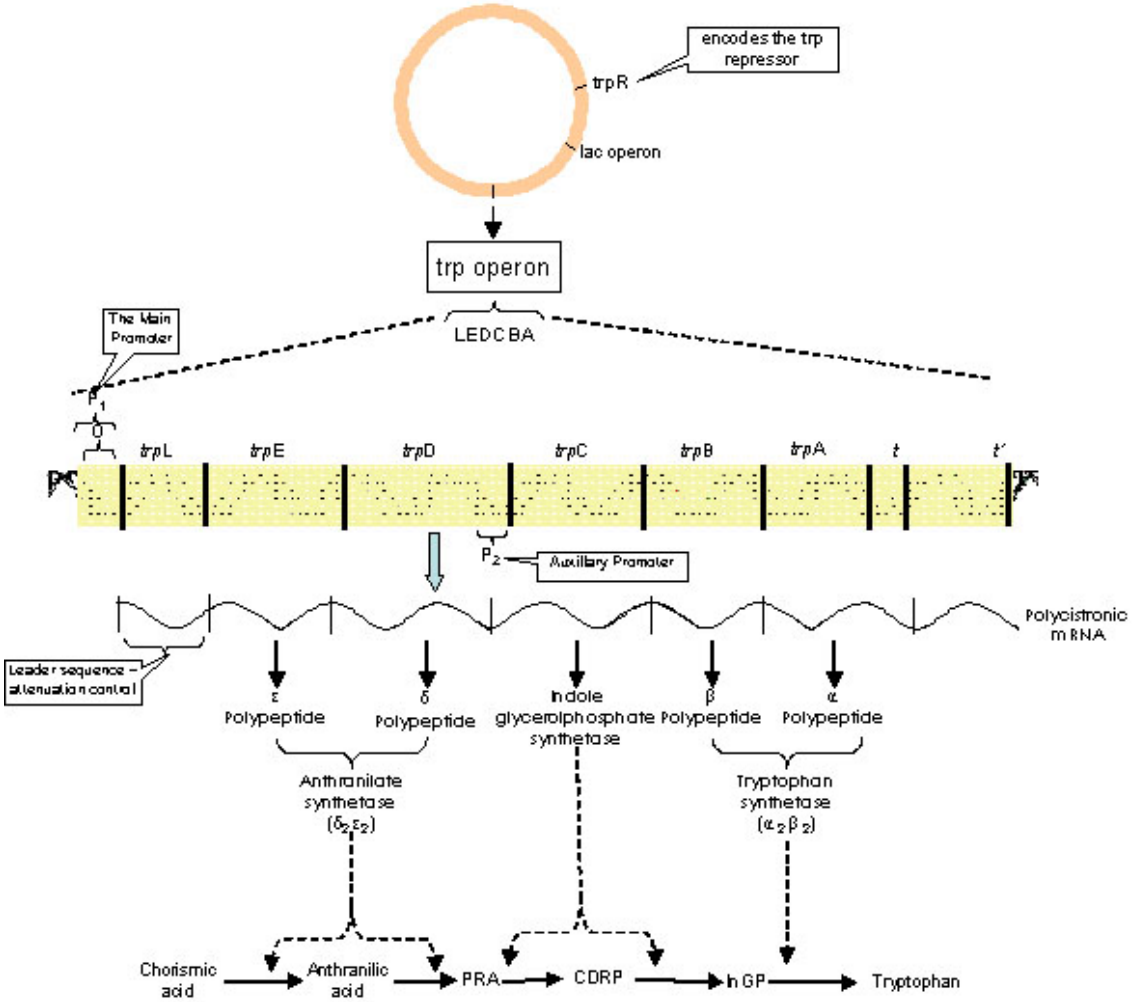


Fig. 9 Structural organization of the trp operon.

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

The regulatory circuit operating at the *trp* operon is designed such that if the cells have enough tryptophan (either from synthesis within the cell or by supplementation of tryptophan in the growth medium) the operon is switched 'OFF'. This is logical, as the cell already has

enough of tryptophan it would not waste its energy in synthesizing the enzymes needed for tryptophan biosynthesis. As the tryptophan is used up by the cell and the internal level falls below a threshold, the operon is switched 'ON' to reinitiate the process of tryptophan synthesis. Therefore, in the presence of low concentration of tryptophan, the operon is 'ON', while a high concentration of the same switches it 'OFF'. The operon is thus repressible as the presence of the signal (high levels of tryptophan) shuts the operon. (Compare it to the lac operon where the presence of the signal activates the operon).

Control by the repressor: One of the control points of the *trp* operon is the repressor protein which is encoded by the *trpR*. The product of the *trpR* gene is an apo-repressor which functions as a repressor only when bound by tryptophan (called as co-repressor). This functional repressor moiety binds to the operator to down regulate the operon. In the absence of enough tryptophan in the cell the apo-repressor is nonfunctional and thus cannot repress the operon. Like the Lac repressor, the Trp repressor also regulates via a negative control mechanism (all repressors are negative regulators) but is activated in the presence of the signal (tryptophan). The Lac repressor, on the other hand, is inactivated in the presence of the signal (allolactose). Taking analogy from *lac* operon, mutations in the *trp* repressor gene or its operator should make the operon constitutive, i.e. the operon should be equally active in the absence or presence of tryptophan.

Attenuation control: In reality, mutations in the trp repressor/operator system only relieves about 70% of the repression. This is so, because in addition to the negative regulatory scheme described above, the *trp* operon also employs another control mechanism called attenuation. Attenuation is a mechanism of transcriptional control that involves premature transcription termination. Normally the transcription of the *trp* operon is terminated at the rho-dependent terminator, t' located about 330bp beyond the end of *trpA*. However, under conditions when there is excess of tryptophan within the cells transcription can be prematurely terminated in the *trpL* region even before the RNA polymerase transcribes any of the structural genes of the operon. When the *trpL* region is transcribed, the RNA formed has the potential to base pair in short stretches due to the presence of inverted repeats as shown in Fig. 10. Region 1 can base pair with region 2, region 2 can base pair with region 3 and region 3 can base pair with region 4. This region of the transcript (encoding the *trp* leader) can thus take up two different secondary structures (Fig. 11). Under the conditions when region 1 pairs with 2, region 3 pairs up with 4 (as 2 is not available to pair with 3), the structure formed by the pairing of the regions 3 and 4 is that of a rho-independent terminator (a stem loop structure followed by a run of Us). This terminator blocks further transcription. The terminator structure is formed as long as region 2 is not available for base pairing with region3. Under conditions when the region 1 is prevented from base pairing with region 2, region2 pairs with region 3, the rho-independent terminator is not formed, allowing transcription to proceed beyond into the structural genes of the *trp* operon.

The level of tryptophan in the cell dictates whether the terminator would be formed or not. This is controlled by the presence of an open reading frame in the *trpL* region which encodes for a 14 amino acid long peptide. The translatable region overlaps within the region 1 of the *trpL* transcript (Fig 10). The important feature of this ORF is the presence of two codons encoding the amino acid tryptophan. Following transcription initiation from the *trp* promoter, the *trpL* region is transcribed, and subsequently, translation of the ORF is initiated (transcription and translation are coupled in the case of prokaryotes). By the time the RNA polymerase reaches region 2 of the *trpL* the ribosome is in a position to translate the two trp codons. The time taken to incorporate the two trp amino acids decides whether attenuation would take place or not. Under conditions when the tryptophan levels are low in the cell the

ribosome is stalled in region 1 (Fig 12a) as it has to wait for a longer time to incorporate a tryptophan molecule. While the ribosome stalls, the RNA polymerase transcribes the rest of the *trpL* region. As the region 3 is transcribed it immediately base pairs with region 2 (which has already been transcribed). Thus, region 3 is not available for pairing with region 4 which prevents the formation of the termination signal. The RNA polymerase thus moves ahead and transcribes the structural genes of the *trp* operon. However, under conditions of high levels of tryptophan in the cell the ribosome does not stall in the region 1 (Fig 12b) but moves ahead encompassing the proximal portion of region 2. While RNA polymerase continues transcribing regions 3 and 4, region 3 cannot base pair with region 2 as the ribosome masks this region. Region 3 thus pairs with region 4 leading to the formation of rho-independent terminator signal. This results in the termination of transcription or what is called attenuation, before the structural genes *trpE* through *trpA* can be transcribed. The attenuation system is thus able to sense the intracellular concentration of trptophan and modulate the regulation accordingly.

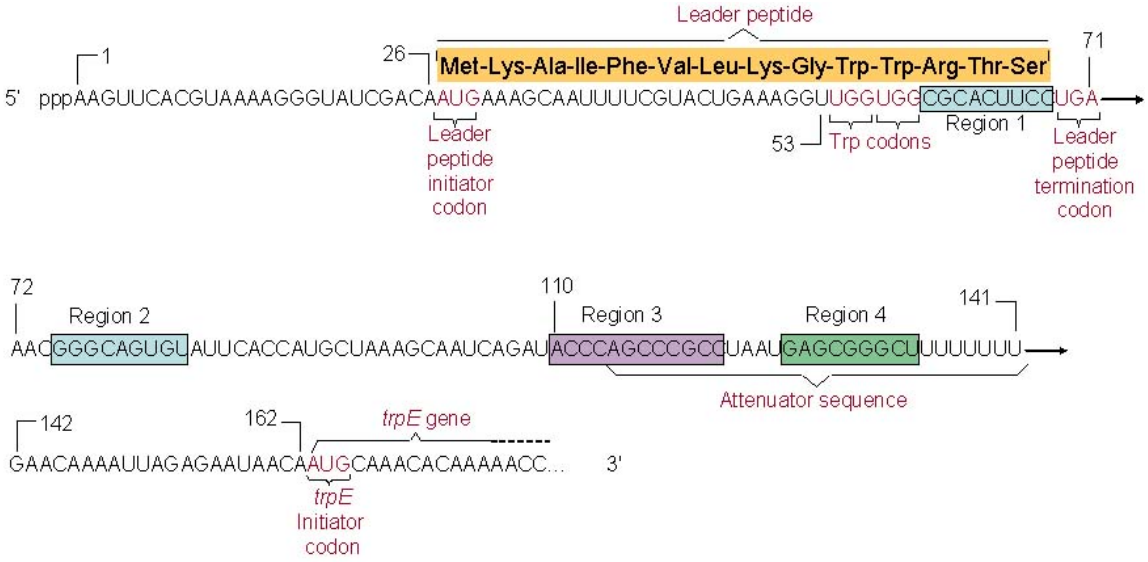
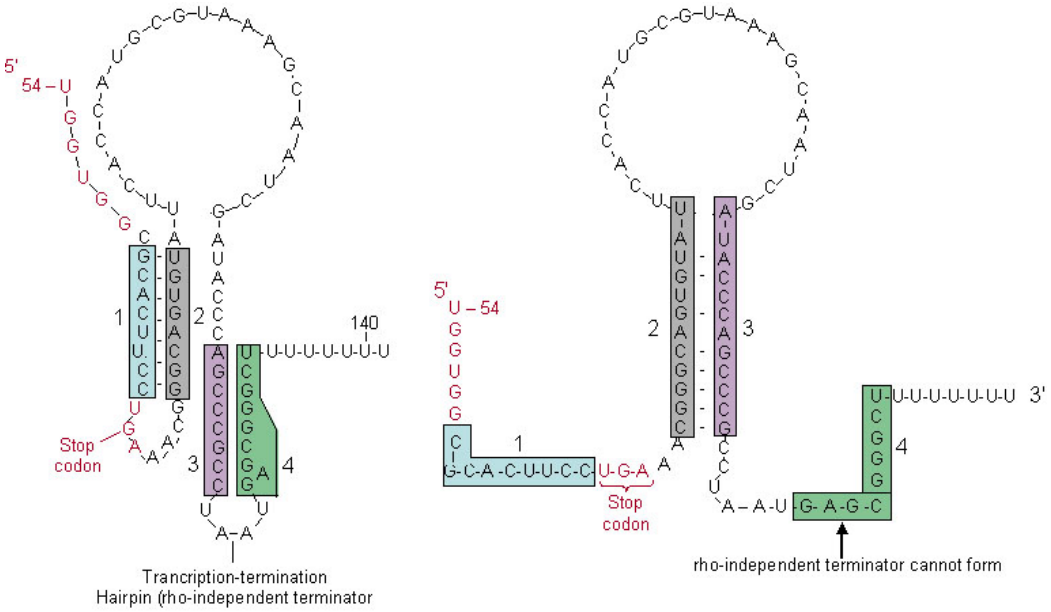


Fig. 10 Regulatory components of the trpL region

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

In summary, the trp operon is a repressible operon having two independent mechanisms to shut off transcription of the operon when the intracellular tryptophan levels are high. Repression is partly brought about by the negative regulator i.e. the trp repressor which binds to the operator only when bound to its co-repressor, tryptophan. The second, is a control via attenuation which uses the phenomenon of coupled transcription and translation to create a transcription terminator signal (in the presence of high tryptophan levels) before the structural genes of the trp operon can be transcribed. Attenuation, therefore, works by sensing intracellular concentration of trptophan. Many biosynthetic operons are regulated by attenuation and interestingly in each case studied, the leader region carries codons specific for that amino acid.



Case 1. Regions 1 and 2 base-paired, Regions 3 and 4 base-paired

Case 2. Regions 2 and 3 base-paired

Fig. 11 Alternate secondary structures that can be formed by the *trpL* transcript

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

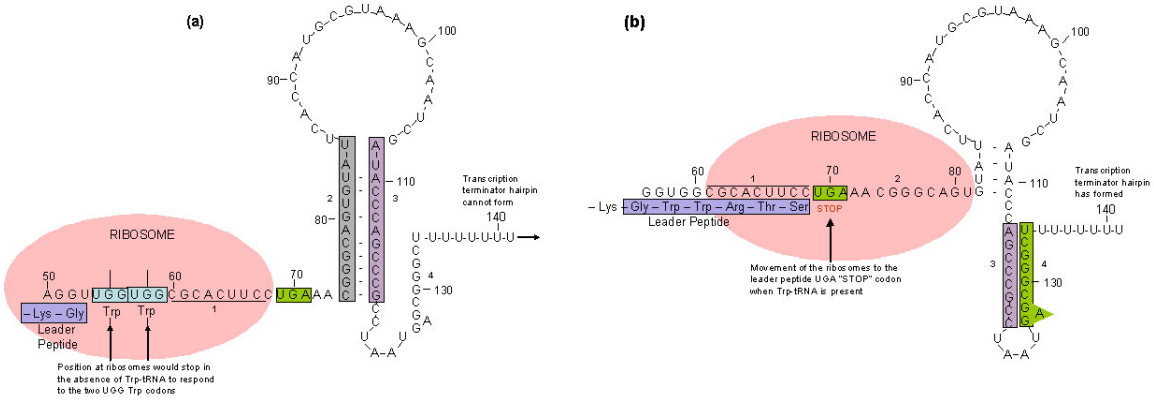


Fig. 12 Schematic representation of the movement of the ribosome (a) in the absence of tryptophan , (b) in its presence leading to attenuation control of the trp operon

(Source: Snustad and Simmons. Principles of Genetics. 2nd edition.)

The Arabinose Operon (control by DNA looping; single regulatory protein-dual function)

E. coli cells can also grow on L- arabinose as a carbon and energy source. The arabinose is brought into the cell through the help of transport system involving several gene products. The proteins involved in arabinose uptake are encoded by genes which are not part of the arabinose operon (Fig 13). The genes involved are *araE*, *araF*, *araG* and *araH*. While *araF*, *G* and *H* constitute an independent operon under the control of the promoter P_{FGH}, *araE* is an independent gene with its promoter P_E. The enzymes, Ribulokinase, Isomerase and

Epimerase needed for catabolism of arabinose are encoded by *araB, araA and araD* genes, respectively which constitutes the arabinose (*ara*) operon. The P_{BAD} promoter transcribes the poly-cistronic message of the operon. Transcription from this promoter is regulated by the AraC protein which is encoded by the *araC* gene lying upstream to the *ara* operon.

The AraC protein has a dual function. In the presence of arabinose it acts as a positive regulator to stimulate transcription from the P_{BAD} promoter while in its absence it helps in negatively regulating this promoter. It also functions as a negative regulator for it's own gene promoter that is P_C. The AraC protein exerts its function by binding to different sites upstream to these promoters: I (made up of two half repeats I_1 and I_2), O_1 (O_{1L} and O_{1R}) and O_2 (Fig 14).

The AraC protein functions as a homodimer. The monomer is made up of a DNA-binding domain as well as a dimerization domain (Fig 14a). The binding site for arabinose also lies within the dimerization domain. The two domains are connected by a flexible linker region. The AraC dimer can take up different conformations depending on whether arabinose is bound to it or not. In the absence of arabinose, the linker region is stretched out in such a manner that the two DNA binding domains of the dimer can reach out to the two far off half sites, O_2 and I_1 (Fig 14b). This binding is further facilitated by the looping out of the intervening 210 bp DNA region. However, when arabinose is bound to the AraC protein the outreach ability of the dimer is curtailed. Under these conditions, the DNA binding domains can bind only to the closely spaced direct half repeats I_1 and I_2 (Fig 14c). This binding releases the DNA loop and the negative control exerted by it . Further, the AraC protein now works as a positive regulator by interacting with the RNA polymerase and helping it to transcribe the genes of the *ara* operon.

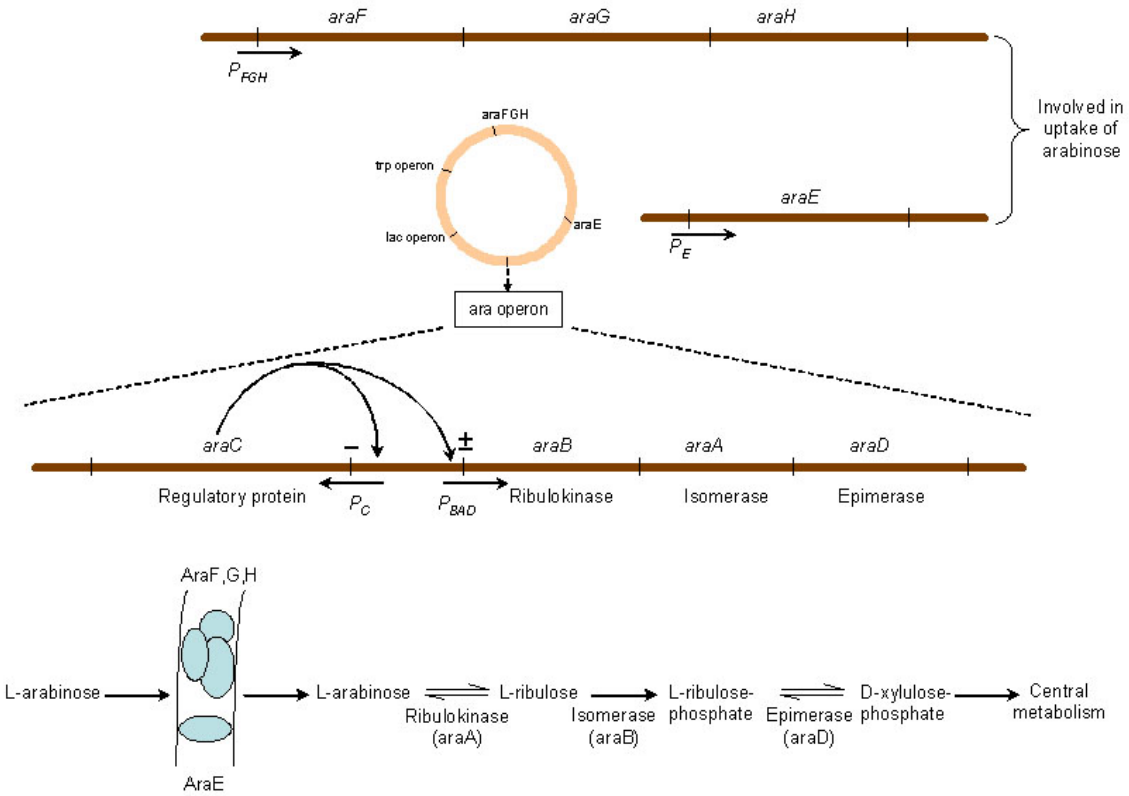


Fig. 13 Structural organization of Ara operon

(Source: Schleif . Trends in Genetics (2000), 16(12): p. 559-564)

The structural genes of the *ara* operon *B*, *A* and *D* are transcribed from the P_{BAD} promoter which needs two positive inputs : the binding of the AraC protein to the I_2 half site, as well as the binding of a CRP (cAMP receptor protein) upstream to the promoter. The AraC protein thus acts like a positive regulator in supporting transcription from the P_{BAD} promoter.

The dual conformation of AraC protein helps regulate the activity of the *ara* operon. In the absence of arabinose, the Ara C protein binds to O_2 and $I_{1.}$, also resulting in the formation of the DNA loop. The DNA loop formed in the absence of arbinose negatively regulates the *ara* operon by sterically hindering the access of RNA polymerase to the P_{BAD} promoter as well as the binding of the CRP protein to its regulatory site on the DNA molecule. Further, the binding of the AraC protein to the 0_2 half site prevents its binding to the I_2 which is essential for transcriptional activation from P_{BAD} promoter. The AraC protein also negatively regulates it own promoter $P_{c.}$. When high levels of AraC protein is attained in the cell, it binds to the half sites O_{1L} and O_{1R} which leads to the turning OFF of the *araC* gene.

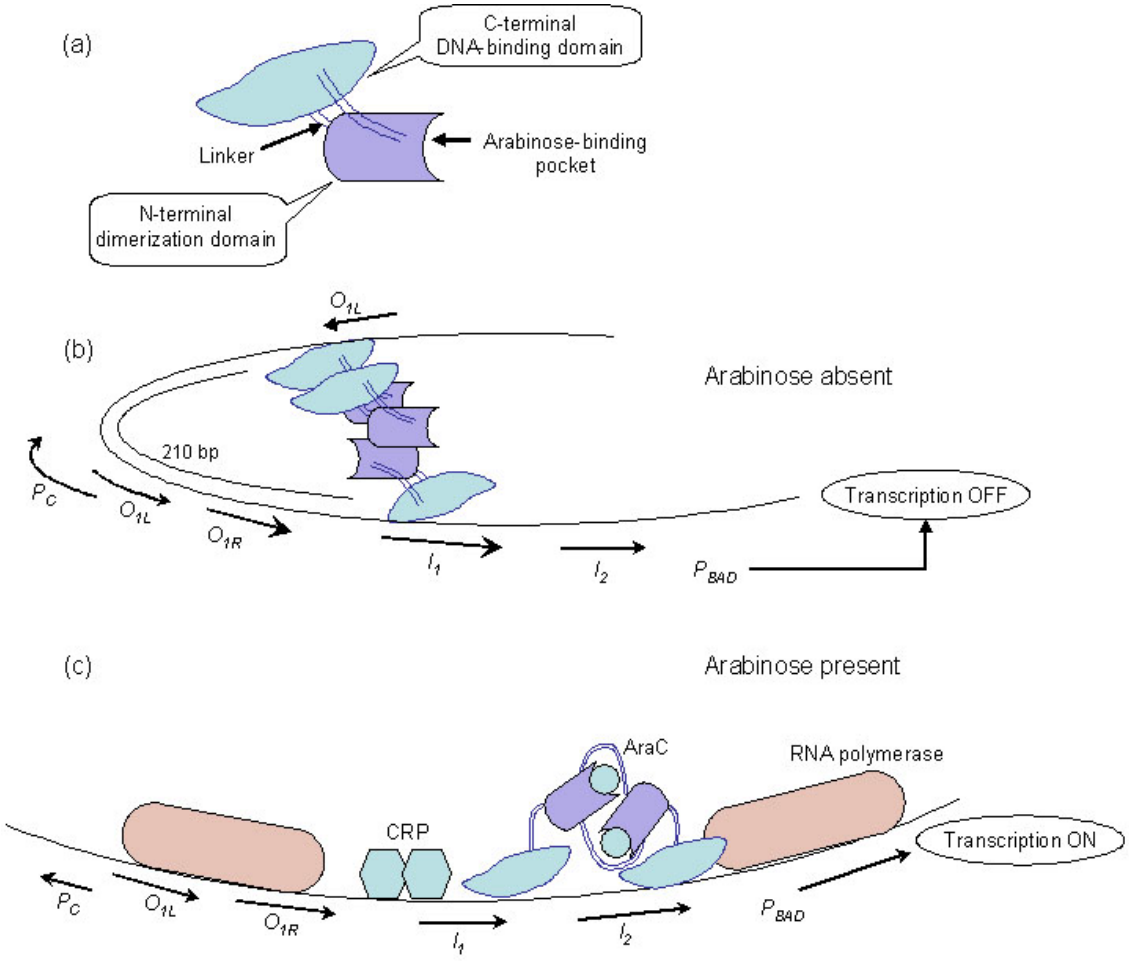


Fig. 14 (a) Functional domains of the AraC protein; (b) Binding of AraC protein of arabinose leading to DNA looping and (c) AraC functioning as a positive regulation in presence of arabinose

(Source: Schleif . Trends in Genetics (2000), 16(12): p. 559-564)

In summary, the arabinose operon is an example of an inducible operon which uses a positive regulator, the AraC protein, to turn ON the operon in the presence of arabinose. In the absence of arabinose, the appropriate interaction of the positive regulator is inhibited and thus the operon is turned OFF. The operon is further kept repressed by the formation of a DNA loop

Suggested Reading

1. Snustad and Simmons. Principles of Genetics. 2nd edition.
2. Inada et al. Genes to Cells 1 (1996) : p. 293-301
3. Schleif . Trends in Genetics (2000), 16(12): p. 559-564