SEM1 CC2

Lactose and Tryptophan Operon

In bacteria, genes encoding the enzymes of a particular metabolic pathway are often grouped adjacent to one another in a cluster on the chromosome. Such clusters, together with the regulatory sequences that control their transcription, are called **operons.** This pattern of organization allows all of the genes in the group to be expressed in a coordinated fashion through transcription into a **single polycistronic mRNA** encoding all the enzymes of the metabolic pathway. A regulatory sequence lying adjacent to this unit of transcription determines whether it is transcribed. This sequence is termed the **operator**. The operator is located next to the promoter. Interaction of a **regulatory protein** with the operator controls transcription of the operon by governing the accessibility of RNA polymerase to the promoter. Although this is the paradigm for prokaryotic gene regulation, it must be emphasized that many regulated prokaryotic genes do not contain operators and are regulated in ways that do not involve protein:operator interactions.

Transcription of Operons Is Controlled by Induction and Repression

In prokaryotes, regulation is ultimately responsive to small molecules serving as signals of the nutritional or environmental conditions confronting the cell. Increased synthesis of enzymes in response to the presence of a particular substrate is termed **induction.** For example, lactose can serve as both carbon and energy source for E. *coli* . Metabolism of this substrate depends on hydrolysis into its component sugars, glucose and galactose, by the enzyme b-galactosidase. In the absence of lactose, E. *coli* cells contain very little b-galactosidase (less than 5 molecules per cell). However, lactose availability induces the synthesis of b-galactosidase by activating transcription of the *lac* operon. One of the genes in the *lac* operon, *lac*Z, is the structural gene for b-galactosidase. When its synthesis is fully induced, b-galactosidase can amount to almost 10% of the total soluble protein in E. *coli* . When lactose is removed from the culture, synthesis of b-galactosidase halts.The alternative to induction, namely decreased synthesis of enzymes in response to a specific metabolite, is termed **repression.** For example, the enzymes of tryptophan biosynthesis in E. *coli* are encoded in the trp operon. If sufficient Trp is available to the growing bacterial culture, the trp operon is not transcribed, so the Trp biosynthetic enzymes are not made; that is, their synthesis is repressed. Repression of the trp operon in the presence of Trp is an

eminently logical control mechanism: If the end product of the pathway is present, why waste cellular resources making unneeded enzymes?

Induction and repression are two faces of the same phenomenon. In induction, a substrate activates enzyme synthesis. Substrates capable of activating synthesis of the enzymes that metabolize them are called **co-inducers,** or often simply **inducers.** Some substrate analogs can induce enzyme synthesis even though the enzymes are incapable of metabolizing them. These analogs are called **gratuitous inducers.** A number of thiogalactosides, such as **IPTG** (isopropylthiogalactoside), are excellent gratuitous inducers of b-galactosidase activity in E. *coli* . In repression, a metabolite, typically an end product, depresses synthesis of its own biosynthetic enzymes. Such metabolites are called **co-repressors.**

lac: **The Paradigm of Operons**

In 1961, François Jacob and Jacques Monod proposed the **operon hypothesis** to account for the coordinate regulation of related metabolic enzymes. The operon was considered to be the unit of gene expression, consisting of two classes of genes: the structural genes for the enzymes, and regulatory elements or genes that controlled expression of the structural genes. The two kinds of genes could be distinguished by mutation. Mutations in a structural gene would abolish one particular enzymatic activity, but mutations in a regulatory gene would affect all of the different enzymes under its control. Mutations of both kinds were known in E. *coli* for lactose metabolism. Bacteria with mutations in either the *lacZ* gene or the *lacY* gene could no longer metabolize lactose—the *lacZ* mutants (*lacZ* strains) because b-galactosidase activity was absent, the *lacY* mutants because lactose was no longer transported into the cell. Lactose transport could still be induced in *lacZ* mutants, and *lacY* mutants displayed lactose-inducible b-galactosidase activity. Other mutations defined another gene, the *lac*I gene. *lacI* mutants were different because they both expressed b-galactosidase activity and immediately transported lactose, without prior exposure to an inducer. That is, a single mutation led to the expression of *lac*tose metabolic functions independently of inducer. Expression of genes independently of regulation is termed **constitutive expression.** Thus, *lac*I had the properties of a regulatory gene. The *lac* operon includes the regulatory gene *lac*I, its promoter p, and three structural genes, *lac*Z, *lac*Y, and *lac*A, with their own promoter p_{lac} and operator O.

The structural genes of the *lac* operon are controlled by **negative regulation.** That is, they are transcribed to give an mRNA unless turned off by the *lac*I gene product. This gene product is the *lac* repressor, a tetrameric protein. The *lac* repressor has two kinds of binding sites—one for inducer and

another for DNA. In the absence of inducer, *lac* repressor blocks *lac* gene expression. It accomplishes repression by binding to the operator DNA site upstream from the *lac* structural genes. Despite the presence of *lac* repressor, RNA polymerase can still initiate transcription at the p_{lac} promoter, but *lac* repressor blocks elongation of transcription, so initiation is aborted. In *lac*I mutants, the *lac* repressor is absent or defective in binding to operator DNA, *lac* gene transcription is not blocked, and the *lac* operon is constitutively expressed in these mutants. Note that *lac*I is normally expressed constitutively from its promoter, so that *lac* repressor protein is always available to fill its regulatory role. About 10 molecules of *lac* repressor are present in an E. *coli* cell. Derepression of the *lac* operon occurs when appropriate b-galactosides occupy the inducer site on *lac* repressor, causing a conformational change in the protein that lowers the repressor's affinity for operator DNA. As a tetramer, *lac* repressor has four inducer binding sites and its response to inducer shows cooperative allosteric effects. Thus, as a consequence of the "inducer"-induced conformational change, the inducer:*lac* repressor complex dissociates from the DNA, and RNA polymerase transcribes the structural genes. Induction reverses rapidly, *lac* mRNA has a half-life of only 3 minutes, and once the inducer is used up through metabolism by the enzymes, free *lac* repressor re-associates with the operator DNA, transcription of the operon is halted, and the residual *lac* mRNA decays.

The *lac* Operator

The *lac* operator is a palindromic DNA sequence. **Palindromes,** or "inverted repeats", provide a twofold, or dyad, symmetry, a structural feature common at sites in DNA where proteins specifically bind. While the operator consists of 35 bp, 26 of which are protected from nuclease digestion when *lac* repressor is bound, a central core defined by 13 bp (from +5 to +17) is involved in specific contacts with *lac* repressor. Mutations at eight sites in this restricted region lead to constitutive expression of the *lac* operon because repressor can no longer bind. These mutants are so-called O^c, or **operator-constitutive, mutants.** Note that the distribution of O^c mutants is not symmetrical about the axis of symmetry. Further, certain O^c mutations, as in $G:C^nA:T$ changes at positions 7 or 9, actually render the palindrome more perfect. The distribution of O^c mutants indicates that repressor contacts with the left half of the palindrome may be more crucial than those with the right half. The operator and promoter (p_{lac}) sites overlap: *lac* repressor protects a region roughly covering nucleotides -5 to +21 from nuclease digestion, whereas RNA polymerase binding and nuclease protection defines p_{lac} as falling within the -45 to +18 region.

Interactions of *lac* Repressor with DNA

Limited digestion of *lac* repressor with trypsin removes an N-terminal, 5'-residue fragment from each subunit, leaving a "core" tetramer that is no longer capable of binding to operator DNA. IPTG binding by the "core" tetramer is unaffected. The N-terminal, 5'-residue fragment retains DNA-binding ability. Thus, the protein is composed of an N-terminal, DNA-binding domain, with the rest of the protein functioning in inducer binding and tetramer formation. In the absence of inducer, intact *lac* repressor nonspecifically binds to duplex DNA with an association constant, K_A, of 2×10^6 M^{-1}, and to the *lac* operator DNA sequence with much higher affinity, $K_A = 2 \times 10^{13}$ M^{-1}. Thus, *lac* repressor binds 10^7 times better to *lac* operator DNA than to any random DNA sequence. IPTG binds to *lac* repressor with an association constant of about 10^6 M^{-1}. The IPTG:*lac* repressor complex binds to operator DNA with an association constant, $K_A = 2 \times 10^{10}$ M^{-1}. Although this affinity is high, it is 3 orders of magnitude less than the affinity of inducer-free repressor for *lac* operator. There is no difference in the affinity of free *lac* repressor and *lac* repressor with IPTG bound for nonoperator DNA. The *lac* repressor apparently acts by binding to DNA and sliding along it, testing sequences in a one-dimensional search until it finds the *lac* operator. The *lac* repressor then binds there with high affinity until inducer causes this affinity to drop by 3 orders of magnitude.

Positive Control of the *lac* Operon by CAP (Catabolite Activator Protein)

Transcription by RNA polymerase from some promoters proceeds with low efficiency unless assisted by an accessory protein that acts as a positive regulator. One such protein is **CAP,** or **catabolite activator protein.** Its name derives from the phenomenon of catabolite repression in E. *coli*. Catabolite repression is a global control that coordinates gene expression with the total physiological state of the cell: As long as glucose is available, E. *coli* catabolizes it in preference to any other energy source, such as lactose or galactose. Catabolite repression ensures that the operons necessary for metabolism of these alternative energy sources, that is, the *lac* and *gal* operons, remain repressed until the supply of glucose is exhausted. Catabolite repression overrides the influence of any inducers that might be present.

Catabolite repression is mediated by cAMP levels, which in turn are regulated by glucose. Transport of glucose into the cell is accompanied by deactivation of **E. *coli* adenylyl cyclase,** leading to lower cAMP levels. The action of CAP as a positive regulator is cAMP-dependent. cAMP binding to CAP enhances its DNA-binding affinity. CAP, also referred to as **CRP** (for **cAMP receptor protein**), is a dimer of identical 210-residue (22.5-kD) polypeptides. The N-terminal domains bind cAMP; the C-terminal domains constitute the DNA-binding site. Two molecules of cAMP are bound per dimer. The CAP-$(cAMP)_2$ complex binds to specific target sites near the promoters of operons. Its presence assists closed

promoter complex formation by RNA polymerase holoenzyme. For example, CAP binding at the -72 to -52 region of *lac* DNA promotes formation of an RNA polymerase holoenzyme:p_{lac} DNA closed promoter complex. Analysis of the structure of the CAP:DNA complex reveals that the DNA is bent more than 90° about the center of dyad symmetry. This bend may be related to the ability of CAP to assist in transcription initiation.

Positive Vs Negative Control

Negative- and positive-control systems are fundamentally different. Genes under negative control are transcribed *unless* they are turned off by the presence of a repressor protein. Often, transcription activation is essentially *anti-inhibition;* that is, the reversal of negative control. In contrast, genes under positive control are expressed *only if* an active regulator protein is present. The *lac* operon illustrates these differences. The action of *lac* repressor is negative. It binds to operator DNA and blocks transcription; expression of the operon only occurs when this negative control is lifted through release of the repressor. In contrast, regulation of the *lac* operon by CAP is positive: Transcription of the operon by RNA polymerase is stimulated by CAP's action as a positive regulator. Operons can also be classified as **inducible** or **repressible,** or both, depending on how they respond to the small molecules that mediate their expression. Repressible operons are expressed only in the absence of their co-repressors. Inducible operons are transcribed only in the presence of small-molecule co-inducers

Trp operon

Trp operon is an operon - a group of genes that are used, or transcribed, together - that codes for the components for production of tryptophan. The Trp operon is present in many bacteria, but was first characterized in *E.coli*. It is regulated so that when tryptophan is present in the environment, it is not used. It was an important experimental system for learning about gene regulation, and is commonly used to teach gene regulation.

Discovered in 1953 by Jacques Monod and colleagues, the trp operon in *E. coli* was the first repressible operon to be discovered. While the lac operon can be activated by a chemical (allolactose), the tryptophan (Trp) operon is inhibited by a chemical (tryptophan). This operon contains five structural genes: trp E, trp D, trp C, trp B, and trp A, which encodes tryptophan synthetase. It also contains a promoter which binds to RNA polymerase and an operator which blocks transcription when bound to the protein synthesized by the repressor gene (trp R) that binds to the operator. In the lac operon, allolactose binds to the repressor protein, allowing gene transcription, while in the trp operon,

tryptophan binds to the repressor protein effectively blocking gene transcription. In both situations, repression is that of RNA polymerase transcribing the genes in the operon. Also unlike the lac operon, the trp operon contains a leader peptide and an attenuator sequence which allows for graded regulation.

It is an example of negative regulation of gene expression. Within the operon's regulatory sequence, the operator is blocked by the repressor protein in the presence of tryptophan (thereby preventing transcription) and is liberated in tryptophan's absence (thereby allowing transcription). The process of attenuation complements this regulatory action.

Repression

This is a negative repressive feedback mechanism. The repressor for the trp operon is produced upstream by the trpR gene, which is continually expressed at a low level. It creates monomers, which associate into tetramers. These tetramers are inactive and "floating" around within the cell. When tryptophan is present, these Tetramers bind to the tryptophan repressor tetramers causing a change in conformation, which allows the repressor to bind the operator. This prevents RNA polymerase from binding to and transcribing the operon, so tryptophan is not produced from its precursor. When tryptophan is not present, the repressor is in its native conformation and cannot bind the operator region, so transcription is not inhibited by the repressor.

Attenuation

In addition to repression, the *trp* operon is controlled by **transcription attenuation.** Unlike the mechanisms discussed thus far, attenuation regulates transcription after it has begun. Charles Yanofsky, the discoverer of this phenomenon, has defined attenuation as any regulatory mechanism that manipulates transcription termination or transcription pausing to regulate gene transcription downstream. In prokaryotes, transcription and translation are coupled, and the translating ribosome is affected by the formation and persistence of pause and termination structures in the mRNA. Attenuation occurs under normal conditions but is blocked when levels of specific **charged tRNAs** (aminoacyl-tRNAs) are lowered on account of amino acid limitation. In many operons encoding enzymes of amino acid biosynthesis, a transcribed 150- to 300-bp leader region is positioned between the promoter and the first major structural gene. These regions encode a short leader peptide containing **multiple codons** for the pertinent amino acid. For example, the leader peptide of the *leu* operon has four Leu codons, the *trp* operon has two tandem Trp codons, and so forth. Translation of these codons

depends on an adequate supply of the relevant aminoacyl-tRNA, which in turn rests on the availability of the amino acid. When Trp is scarce, the entire *trp* operon from *trp*L to *trp*A is transcribed to give a polycistronic mRNA. But, as [Trp] increases, more and more of the trp transcripts consist of only a 140-nucleotide fragment corresponding to the 5'-end of *trp*L. Trp availability is causing premature termination of *trp* transcription, that is, transcription attenuation. The secondary structure of the 160-bp leader region transcript is the principal control element in transcription attenuation. This RNA segment includes the coding region for the 14-residue leader peptide. Three critical base-paired hairpins can form in this RNA: the **1:2 pause** structure, the **3:4 terminator,** and the **2:3 antiterminator.** Obviously, the 1:2 pause, 3:4 terminator, and the 2:3 antiterminator represent mutually exclusive alternatives. A significant feature of this coding region is the tandem UGG Trp codons.

Transcription by RNA polymerase begins and progresses until position 92 is reached, whereupon the 1:2 hairpin is formed, causing RNA polymerase to pause in its elongation cycle. While RNA polymerase is paused, a ribosome begins to translate the leader region of the transcript. Translation by the ribosome releases the paused RNA polymerase and transcription continues, with RNA polymerase and the ribosome moving in unison. As long as Trp is plentiful enough that Trp-$tRNA^{Trp}$ is not limiting, the ribosome is not delayed at the two Trp codons and follows closely behind RNA polymerase, translating the message soon after it is transcribed. The presence of the ribosome atop segment 2 blocks formation of the 2:3 antiterminator hairpin, allowing the alternative 3:4 terminator hairpin to form. Stable hairpin structures followed by a run of Us are features typical of *rho*-independent transcription termination signals, so the RNA polymerase perceives this hairpin as a transcription stop signal and transcription is terminated at this point. On the other hand, a paucity of Trp and hence Trp-$tRNA^{Trp}$ causes the ribosome to stall on segment 1. This leaves segment 2 free to pair with segment 3 and to form the 2:3 antiterminator hairpin in the transcript. Because this hairpin precludes formation of the 3:4 terminator, termination is prevented and the entire operon is transcribed. Thus, transcription attenuation is determined by the availability of charged $tRNA^{Trp}$ and its transitory influence over the formation of alternative secondary structures in the mRNA.

Transcription Is Regulated by a Diversity of Mechanisms

A surprising variety of control mechanisms operate in transcriptional regulation, as we have just seen. Several organizing principles materialize. First, DNA:protein interactions are a central feature in transcriptional control, and the DNA sites where regulatory proteins bind commonly display at least partial dyad symmetry or inverted repeats. Further, DNA-binding proteins themselves are generally

even-numbered oligomers (for example, dimers, tetramers) that have an innate twofold rotational symmetry. Second, protein:protein interactions are an essential component of transcriptional activation. We see this latter feature in the activation of RNA polymerase by CAP-$(cAMP)_2$ Third, the regulator proteins receive cues that signal the status of the environment (for example, Trp, lactose, cAMP) and act to communicate this information to the genome, typically via the medium of conformational changes and DNA:protein interactions.

Transcriptional Activators work through Protein: Protein Contacts with RNA Polymerase

Although transcriptional control is governed by a variety of mechanisms, an underlying principle of transcriptional activation has emerged. Transcriptional activation can take place when a **transcriptional activator** protein (such as CAP-$(cAMP)_2$) bound to DNA makes protein:protein contacts with RNA polymerase, and the degree of transcriptional activation is proportional to the strength of the protein:protein interaction. Generally speaking, a nucleotide sequence that provides a binding site for a DNA-binding protein can serve as an **activator site** if the DNA-binding protein bound there can interact with promoter-bound RNA polymerase. DNA-binding proteins that activate transcription thus have a **DNA-binding domain** and an **activation domain** capable of interacting with RNA polymerase. Such activation domains activate transcription through protein:protein interactions with either the a, b, b', or s subunits of RNA polymerase. Further, if the DNA-bound transcriptional activator makes contacts with two different components of RNA polymerase, a synergistic effect takes place such that transcription is markedly elevated. Thus, transcriptional activation at specific genes relies on the presence of one or more activator sites where one or more transcriptional activator proteins can bind and make contacts with RNA polymerase bound at the promoter of the gene. Indeed, transcriptional activators may facilitate the recruitment and binding of RNA polymerase to the promoter. This general principle applies to transcriptional activation in both prokaryotic and eukaryotic cells

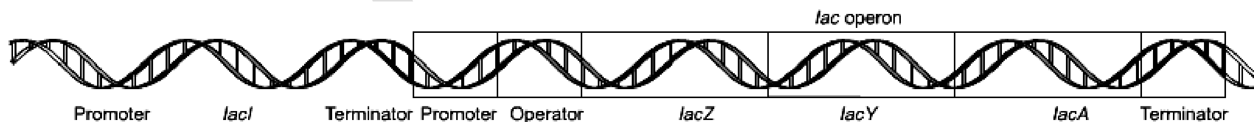

Fig- Overview of lac operon

		p	lacI	p_{lac}, O	lacZ	lacY	lacA
DNA							
bp			1080	82	3069	1251	609
mRNA							
Polypeptide	Amino acids		360		1023	417	203
	kD		38.6		116.4	46.5	22.7
Protein	Structure		Tetramer		Tetramer	Membrane protein	Dimer
	kD		154.4		465	46.5	45.4
Function			Repressor		β-Galactosidase	Permease	Trans-acetylase

Fig- The *lac* operon. The operon consists of two transcription units. In one unit, there are three structural genes, *lac*Z, *lac*Y, and *lac*A, under control of the promoter, p_{lac}, and the operator O. In the other unit, there is a regulator gene, *lac*I, with its own promoter, p_{lacI}. *lac*I encodes a 360-residue, 38.6-kD polypeptide that forms a tetrameric *lac* repressor protein. *lac*Z encodes b-galactosidase, a tetrameric enzyme of 116-kD subunits. *lac*Y is the b-galactoside permease structural gene, a 46.5-kD integral membrane protein active in b-galactoside transport into the cell. The remaining structural gene encodes a 22.7-kD polypeptide that forms a dimer displaying thiogalactoside transacetylase activity in vitro, transferring an acetyl group from acetyl-CoA to the C-6 OH of thiogalactosides, but the metabolic role of this protein in vivo remains uncertain. *lac*A mutants show no identifiable metabolic deficiency. Perhaps the *lac*A protein acts to detoxify toxic analogs of lactose through acetylation.

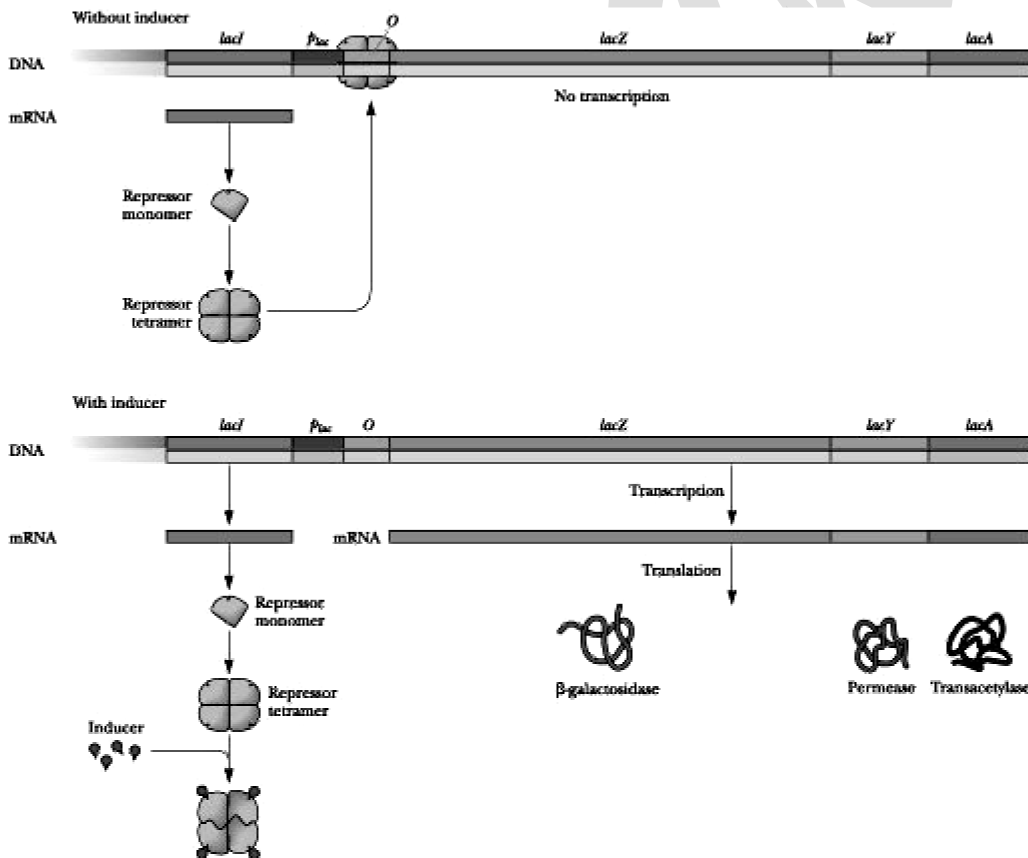

Fig- The mode of action of *lac* repressor

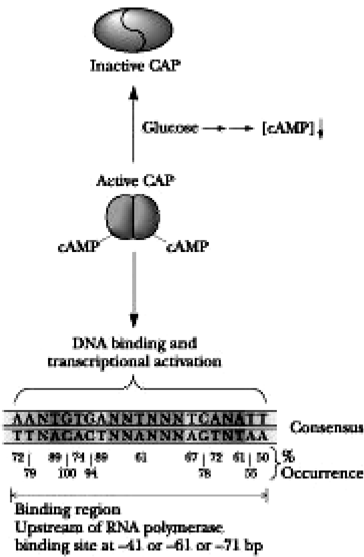

Fig- The mechanism of catabolite repression and CAP action. Glucose instigates catabolite repression by lowering cAMP levels. cAMP is necessary for CAP binding near promoters of operons whose gene products are involved in the metabolism of alternative energy sources such as lactose, galactose, and arabinose. The binding sites for the CAP-$(cAMP)_2$ complex are consensus DNA sequences containing the conserved pentamer TGTGA and a less well conserved inverted repeat, TCANA (where N is any nucleotide).

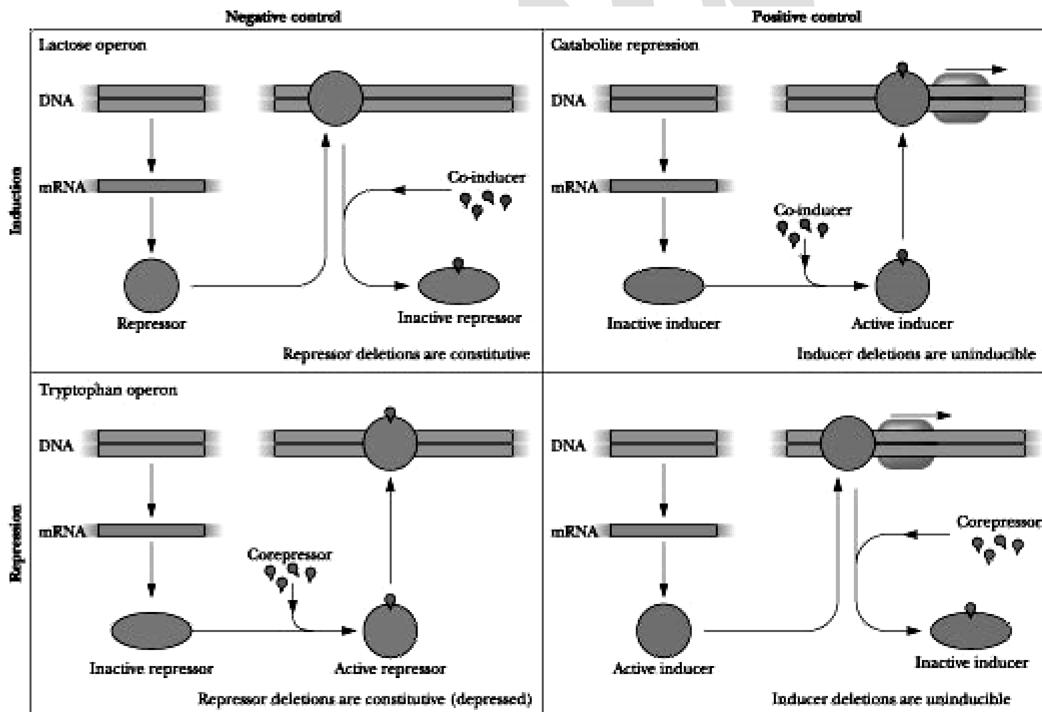

Fig- Control circuits governing the expression of genes. These circuits can be negative or positive, inducible or repressible

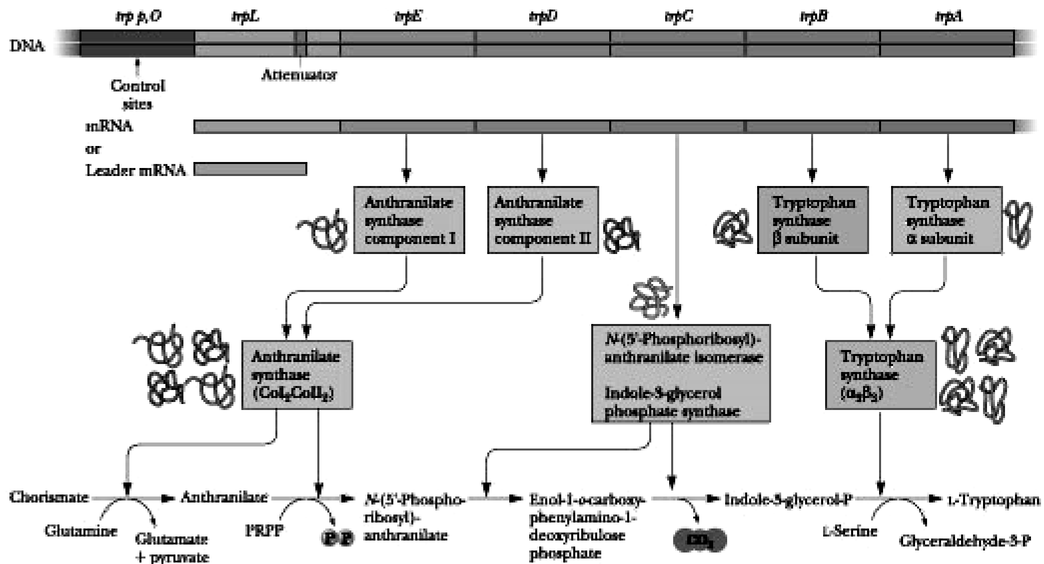

Fig- The *trp* operon of E. *coli*

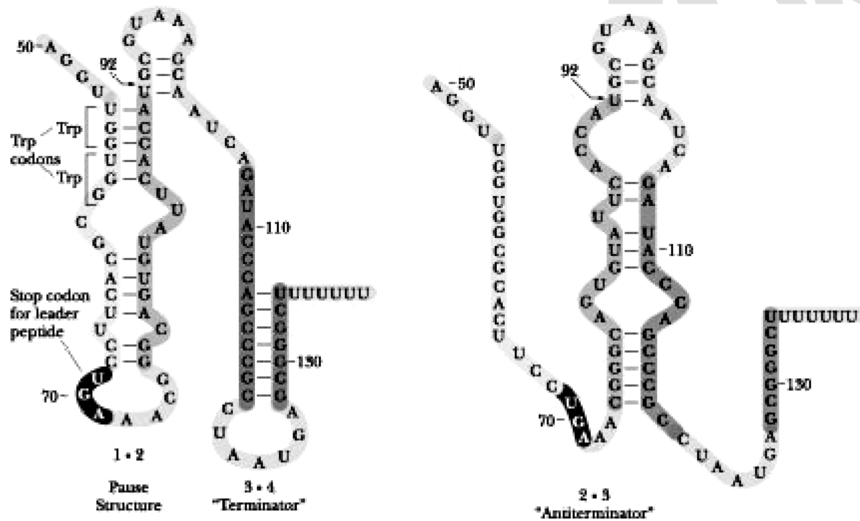

Fig- Alternative secondary structures for the leader region (*trp*L mRNA) of the *trp* operon transcript

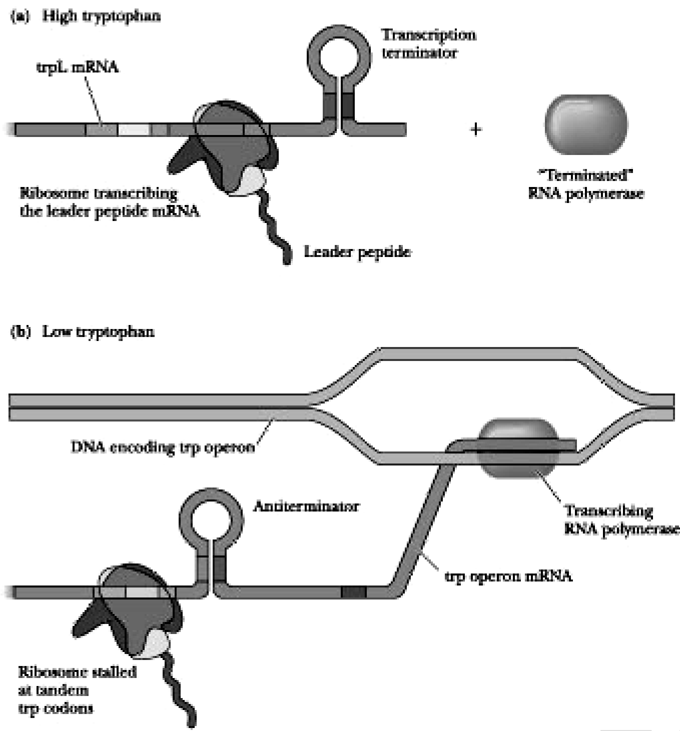

Fig- The mechanism of attenuation in the *trp* operon.

pinaki