B M B 400
Part Four: Gene Regulation
Section I = Chapter 15
POSITIVE AND NEGATIVE CONTROL SHOWN BY THE *lac* OPERON OF *E. COLI*

A. **Definitions** and general comments

1. **Operons**

An **operon** is a cluster of coordinately regulated genes. It includes **structural genes** (generally encoding enzymes), **regulatory genes** (encoding, e.g. activators or repressors) and **regulatory sites** (such as promoters and operators).

2. **Negative versus positive control**

a. The type of control is defined by the response of the operon when no regulatory protein is present.

b. In the case of <u>negative control</u>, the genes in the operon are expressed unless they are switched off by a repressor protein. Thus the operon will be turned on constitutively (the genes will be expressed) when the repressor in inactivated.

c. In the case of <u>positive control</u>, the genes are expressed only when an active regulator protein, e.g. an activator, is present. Thus the operon will be turned off when the positive regulatory protein is absent or inactivated.

Table 4.1.1. Positive vs. negative control

	Regulatory protein is present	Example of regulatory protein	Mutate regulatory gene to lose function
Positive control	Operon ON	Activator	Operon OFF
Negative control	Operon OFF	Repressor	Operon ON

3.   Catabolic versus biosynthetic operons

a.   Catabolic pathways catalyze the breakdown of nutrients (the substrate for the pathway) to generate energy, or more precisely ATP, the energy currency of the cell. In the absence of the *substrate*, there is no reason for the catabolic enzymes to be present, and the operon encoding them is repressed. In the presence of the substrate, when the enzymes are needed, the operon is induced or de-repressed.

Table 4.1.2. Comparison of catabolic and biosynthetic operons

Operon encodes	Absence of	Effect	Presence of	Effect
catabolic enzymes	substrate	repressed	substrate	derepressed (induced)
biosynthetic enzymes	product	induced	product	repressed

For example, the *lac* operon encodes the enzymes needed for the uptake (lactose permease) and initial breakdown of lactose (the disaccharide β-D-galactosyl-1->4-D-glucose) into galactose and glucose (catalyzed by β-galactosidase). These monosaccharides are broken down to lactate (principally via glycolysis, producing ATP), and from lactate to CO_2 (via the citric acid cycle), producing NADH, which feeds into the electron-transport chain to produce more ATP (oxidative phosphorylation). This can provide the energy for the bacterial cell to live. However, the initial enzymes (lactose permease and β-galactosidase) are only needed, and only expressed, in the presence of lactose and in the absence of glucose. In the presence of the substrate lactose, the operon in turned on, and in its absence, the operon is turned off.

b.   Anabolic, or biosynthetic, pathways use energy in the form of ATP and reducing equivalents in the form of NAD(P)H to catalyze the synthesis of cellular components (the product) from simpler materials, e.g. synthesis of amino acids from small dicarboxylic acids (components of the the citric acid cycle). If the cell has plenty of the product already (in the presence of the *product*), the the enzymes catalyzing its synthesis are not needed, and the operon encoding them is repressed. In the absence of the product, when the cell needs to make more, the biosynthetic operon is induced.

E.g., the *trp* operon encodes the enzymes that catalyze the conversion of chorismic acid to tryptophan. When the cellular concentration of Trp (or Trp-tRNA^{trp}) is high, the operon is not expressed, but when the levels are low, the operon is expressed.

4. Inducible versus repressible operons

a. <u>Inducible</u> operons are turned on in reponse to a metabolite (a small molecule undergoing metabolism) that regulates the operon. E.g. the *lac* operon is induced in the presence of lactose (through the action of a metabolic by-product allolactose).

b. <u>Repressible</u> operons are switched off in reponse to a small regulatory molecule. E.g., the *trp* operon is repressed in the presence of tryptophan.

Note that in this usage, the terms are defined by the reponse to a small molecule. Although lac is an inducible operon, we will see conditions under which it is repressed or induced (via derepression).
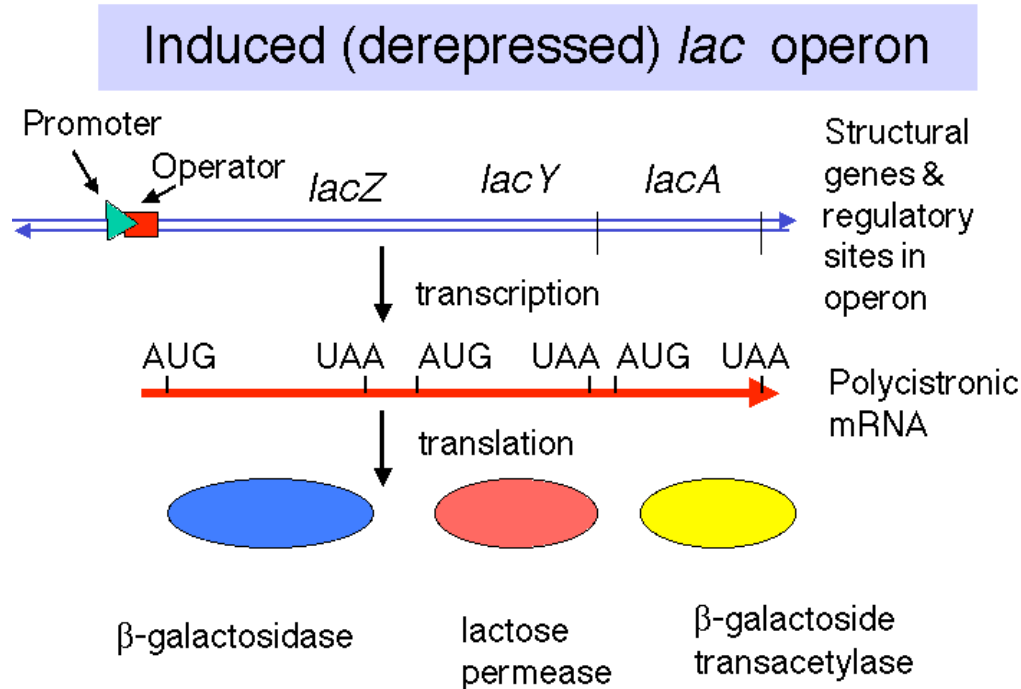
Table 4.1.3.

Inducible vs. repressible operons

Defined by response of operon to a metabolite (small molecule).

Type of operon	Presence of	Effect	Examples Metabolite	Operon
Inducible	metabolite	ON	lactose	*lac*
Repressible	metabolite	OFF	Trp	*trp*

B.   Map of the *E. coli lac* operon

Figure 4.1.1.

1. Promoters = *p* = binding sites for RNA polymerase from which it initiates transcription.

    There are separate promoters for the *lacI* gene and the *lacZYA* genes.

2. Operator = *o* = binding site for repressor; overlaps with the promoter for *lacZYA*.

3. Repressor encoded by *lacI* gene

4. Structural genes: *lacZYA*

    lacZ encodes β-galactosidase, which cleaves the disccharide lactose into galactose and glucose.
    lacY encodes the lactose permease, a membrane protein that faciltitates uptake of lactose.
   *lacA* encodes β-galactoside transacetylase; the function of this enzymes in catabolism of lactose is not understood (at least by me).

C. Negative control

The *lac* operon is under <u>both</u> negative and positive control. The mechanisms for these will be considered separately.

1. In negative control, the *lacZYA* genes are switched off by repressor when the inducer is absent (signalling an absence of lactose). <u>When the repressor tetramer is bound to *o*, *lacZYA* is not transcribed and hence not expressed.</u>

Figure 4.1.2. Repressed *lac* operon
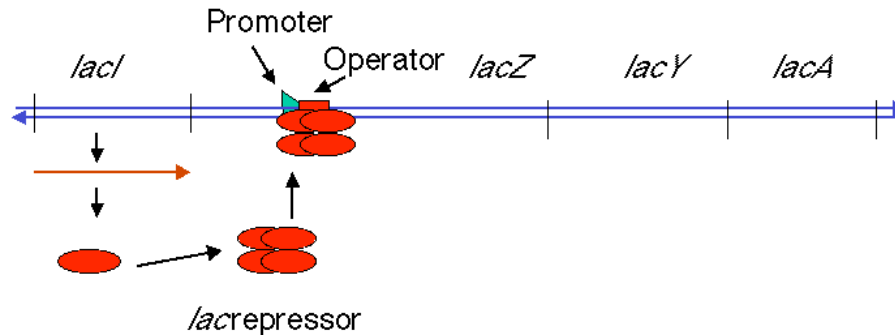
Repressor binds to the **operator** in the absence of the **inducer** (a metabolite of lactose), and blocks transcription of the *lac* operon.

2. When inducer is present (signalling the presence of lactose), it binds the repressor protein, thereby altering its conformation, decreasing its affinity for *o*, the operator. The dissociation of the repressor-inducer complex allows *lacZYA* to be transcribed and therefore expressed.

Figure 4.1.3. Induction of the *lac* operon by derepression.
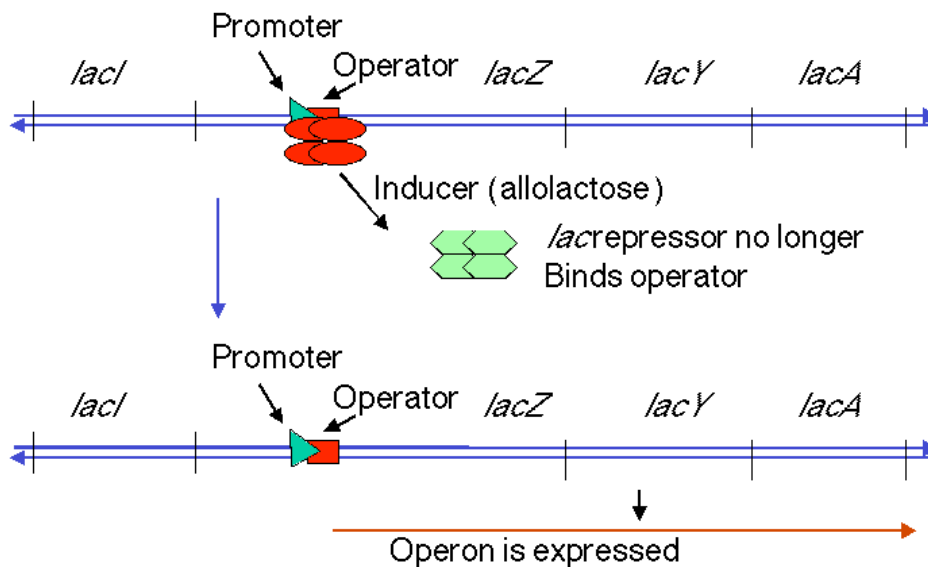
D. Inducers

1. The <u>natural inducer</u> (or antirepressor), is <u>allolactose</u>, an analog of lactose. It is made as a metabolic by-product of the reaction catalyzed by β-galactosidase. Usually this enzyme catalyzes the cleavage of lactose to galactose + glucose, but occasionally it will catalyze an isomerization to form allolactose, in which the galacose is linked to C6 of glucose instead of C4.

2. A <u>gratuitous inducer</u> will induce the operon but not be metabolized by the encoded enzymes; hence the induction is maintained for a longer time. One of the most common ones used in the laboratory is a synthetic analog of lactose called isopropylthiogalactoside (IPTG). In this compound the β-galactosidic linkage is to a thiol, which is not an efficient substrate for β-galactosidase.

E. Regulatory mutants

<u>Regulatory mutations</u> affect the <u>amount</u> of all the enzymes encoded by an operon, whereas mutations in a structural gene affects only the <u>activity</u> of the encoded (single) polypeptide.

1. <u>Repressor mutants</u>

    a. Wild-type strains ($lacI^+$) are <u>inducible</u>.

    b. Most strains with a defective repressor ($lacI^-$) are <u>constitutive</u>, i.e. they make the enzymes encoded by the *lac* operon even in the absence of the inducer.

    c. Strains with repressor that is not able to interact with the inducer ($lacI^S$) are <u>noninducible</u>. Since the inducer cannot bind, the repressor stays on the operator and prevents expression of the operon even in the presence of inducer.

d. Deductions based on phenotypes of mutants

Table 4.1.4. Phenotypes of repressor mutants

Genotype	β-galactosidase		transacetylase		Conclusion
	-IPTG	+IPTG	-IPTG	+IPTG	
$I^+Z^+A^+$	<0.1	100	<1	100	Inducible
$I^+Z^-A^+$	<0.1	<0.1	<1	100	*lacZ* encodes β-galactosidase
$I^-Z^+A^+$	100	100	90	90	Constitutive
$I^+Z^-A^+$ /F' $I^-Z^+A^+$	<0.1	100	<1	200	$I^+ > I^-$ in *trans*
$I^sZ^+A^+$	<0.1	<1	<1	<1	Noninducible
$I^sZ^+A^+$ /F' $I^+Z^+A^+$	<0.1	1	<1	1	$I^s > I^+$ in *trans*

(1) The wild-type operon is inducible by IPTG.

(2) A mutation in *lacZ* affects only β-galactosidase, not the transacetylase (or other products of the operon), showing that *lacZ* is a structural gene.

(3) A mutation in *lacI* affects both enzymes, hence *lacI* is a regulatory gene. Both are expressed in the absence of the inducer, hence the operon is constitutively expressed (the strain shows a constitutive phenotype).

(4) In a merodiploid strain, in which one copy of the *lac* operon is on the chromosome and another copy is on an F' factor, one can test for dominance of one allele over another. The wild-type *lacI+* is dominant over *lacI⁻* in *trans.* In a situation where the only functional *lacZ* gene is on the same chromosome as *lacI⁻*, the functional *lacI* still causes repression in the absence of inducer.

(5) The *lacIˢ* allele is noninducible.

(6) In a merodiploid, the *lacIˢ* allele is dominant over wild-type in *trans*.

e. The fact that the product of the *lacI* gene is *trans*-acting means that it is a diffusible molecule that can be encoded on one chromosome but act on another, such as the F' chromosome in example (d) above. In fact the product of the *lacI* gene is a repressor protein.

2. Operator mutants

   a. Defects in the operator lead to constitutive expression of the operon, hence one can isolate <u>operator constitutive mutations</u>, abbreviated o^c. The wild-type o^+ is inducible.

   b. Mutations in the operator are *cis*-acting; they only affect the expression of structural genes on the same chromosome.

      (1) The merodiploid $I^+o^cZ^+/I^+o^+Z^-$ [this is an abbreviation for *lacI*$^+o^c$*lacZ*$^+$/*lacI*$^+o^+$*lacZ*$^-$] expresses β-galactosidase constitutively. Thus o^c is dominant to o^+ when o^c is in *cis* to *lacZ+*.

      (2) The merodiploid $I^+o^cZ^-/I^+o^+Z^+$ is inducible for β-galactosidase expression. Thus o^+ is dominant to o^c when o^+ is in *cis* to *lacZ*$^+$.

      (3) The allele of o that is in *cis* to the active reporter gene (i.e., on the same chromosome as *lacZ*$^+$ in this case) is the one whose phenotype is seen. Thus the operator is *cis*-acting, and this property is referred to as *cis*-dominance. As in most cases of *cis*-regulatory sequences, these are sites on DNA that are required for regulation. In this case the operator is a binding site for the *trans*-acting repressor protein.

F. Interactions between operator and repressor

1. **Sequence of operator**

    a. The operator <u>overlaps</u> the start the site of transcription and the promoter.

    b. It has a <u>dyad symmetry</u> centered at +11. Such a dyad symmetry is commonly found within binding sites for symmetrical proteins (the repressor is a homotetramer).

    c. The sequence of DNA that consititutes the operator was defined by the position of o^C mutations, as well as the nucleotides protected from reaction with, e.g. DMS, upon binding of the repressor.

Figure 4.1.4.
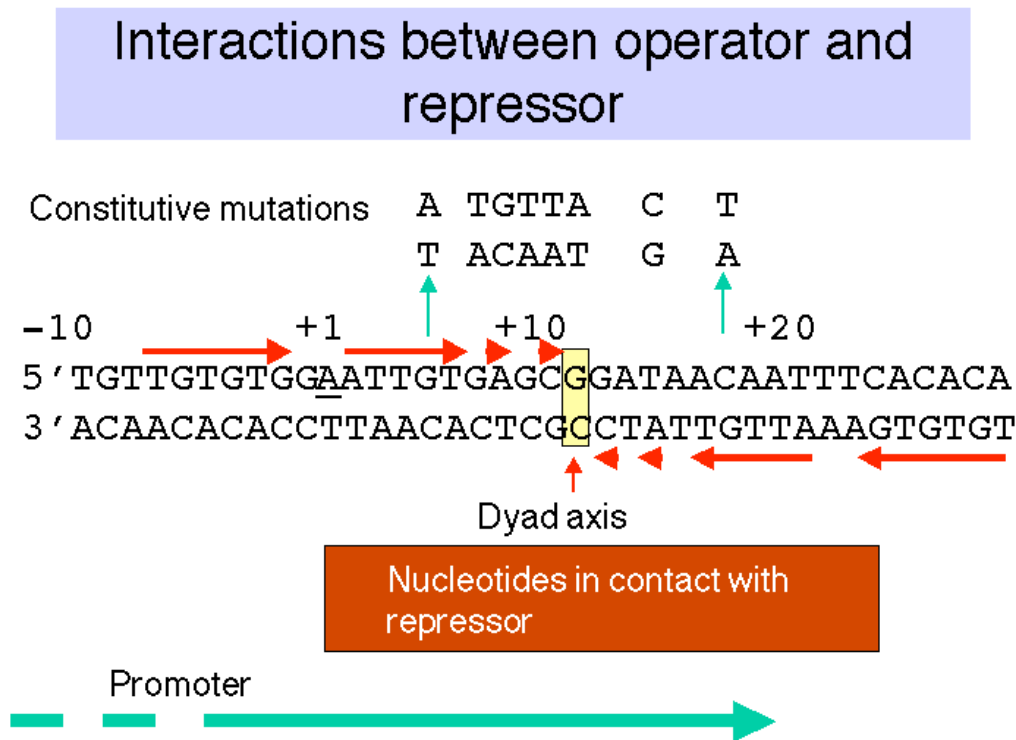
2. **Repressor**

    a. Purification

        (1) Increase the amount of repressor in the starting material by <u>over-expression.</u>

        A wild-type cell has only about 10 molecules of the repressor tetramer. Isolation and purification of the protein was greatly aided by use of mutant strain with up-promoter mutations for *lacI*, so that many more copies of the protein were present in each cell. This general strategy of over-producing the protein is widely used in purification schemes. Now

the gene for the protein is cloned in an expression vector, so that the host (bacteria in this case) makes a large amount of the protein - often a substantial fraction of the total bacterial protein.

(2) Assays for repressor

[1] Binding of radiolabeled IPTG (gratuitous inducer) to repressor

[2] Binding of radiolabeled operator DNA sequence to repressor. This can be monitored by the ability of the protein-DNA complex to bind to nitrocellulose (whereas a radiolabeled mutant operator DNA fragement, o^c, plus repressor will not bind). Electrophoretic mobility shift assays would be used now in many cases.

[3] This ability of particular sequences to bind with high affinity to the desired protein is frequently exploited to rapidly isolate the protein. The binding site can be synthesized as duplex oligonucleotides. These are ligated together to form multimers, which are then attached to a solid substrate in a column. The desired DNA-binding protein can then be isolated by <u>affinity chromatography</u>, using the binding site in DNA as the affinity ligand.

b. The isolated, functional repressor is a <u>tetramer</u>; each of the four monomers is the product of the *lacI* gene (i.e. it is a homotetramer).

c. The **DNA-binding domain** of the *lac* repressor folds into a **helix-turn-helix** domain. We will examine this structural domain in more in Chapter III. It is one of the most common DNA-binding domains in prokaryotes, and a similar structural domain (the homeodomain) is found in some eukaryotic transcriptional regulators.

3. **Contact points between repressor and operator**

a. Investigation of the contact points between repressor and the operator utiblized the same techniques that we discussed previously for mapping the binding site of RNA polymerase on the promoter, e.g. electrophoretic mobility shift assays (does the DNA fragment bind?), DNase footprints (where does the protein bind?) and methylation interference assays (methylation of which purines will prevent binding?). Alternative schemes will allow one to identify sites at which methylation is either prevented or enhanced by the binding of the repressor. These techniques provide a biochemical defintion of the operator = binding site for repressor.

b. The key contact points (see Figure 4.1.4.):

(1) are within the dyad symmetry.

(2) coincide (in many cases) with nucleotides that when mutated lead to constitutive expression. Note that the latter is a genetic definition of the operator, and it coincides with the biochemically-defined operator.

(3) tend to be distributed symmetrically around the dyad axis (+11).

(4) are largely on one face of the DNA double helix.

c. The partial overlap between the operator and the promoter initially suggested a model of steric interference to explain the mechanism of repression. As long a repressor was bound to the operator, the polymerase could not bind to the promoter. But, as will be explored in the next chapter, this is *not* the case. RNA polymerase *can* bind to the *lac* promoter even when repressor is boudn to the *lac* operator. However, the polymerase *cannot initiate* transcription when juxtaposed to the repressor.

4. **Conformational shift in repressor when inducer binds**

a. The repressor has two different domains, one that binds to DNA ("headpiece" containing the helix-turn-helix domain) and another that binds to the inducer (and other subunits) (called the "core). These are connected by a "hinge" region.

b. These structural domains can be distinguished by the phenotypes of mutations that occur in them.

lacI^{-d} prevents binding to DNA, leads to constitutive expression.

lacI^{S} prevents binding of inducer, leads to a noninducible phenotype.

c. Binding of inducer to the "core" causes an allosteric shift in the repressor so that the "headpiece" is no longer able to form a high affinity complex with the DNA, and the repressor can dissociate (go to one of the many competing nonspecific sites).

G. Positive control: "catabolite repression"

1. <u>Catabolite repression</u>

   a. Even bacteria can be picky about what they eat. Glucose is the preferred source of carbon for *E. coli*; the bacterium will consume the available glucose before utilizing alternative carbon sources, such as lactose or amino acids.

   b. Glucose leads to repression of expression of *lac* and some other catabolic operons. This phenomenon is called <u>catabolite repression</u>.

2. <u>Two components</u> are needed for this form of regulation

   a. <u>cAMP</u>

      [1] In the presence of glucose, the [cAMP] inside the cell decreases from 10^{-4} M to 10^{-7} M. A high [cAMP] will relieve catabolite repression.

      [2] cAMP synthesis is catalyzed by adenylate cyclase (product of the *cya* gene)

      $$ATP \rightarrow cAMP + PP_i$$

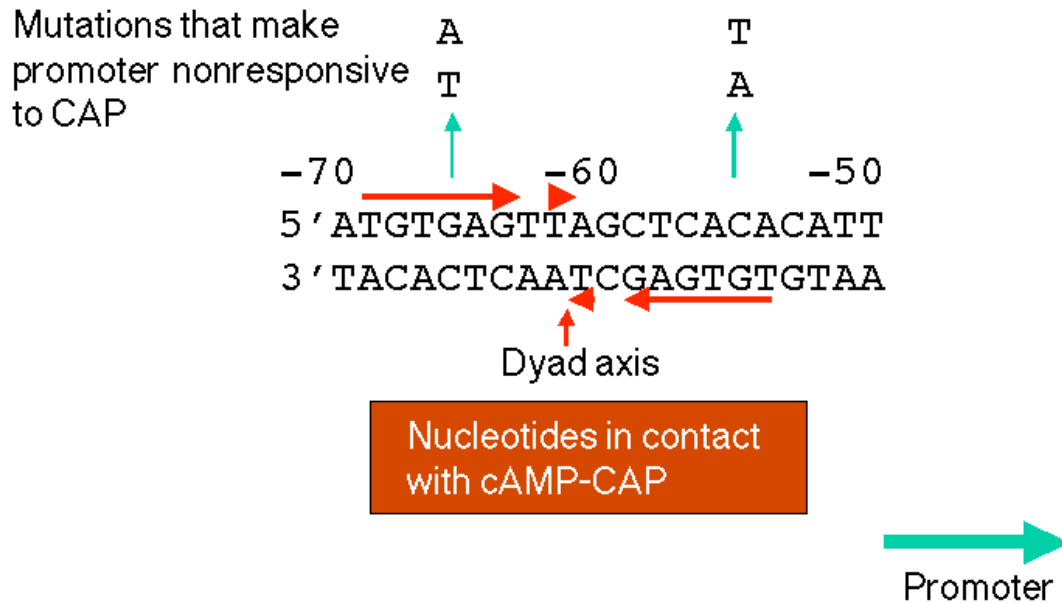   b. <u>Catabolite Activator Protein = CAP</u>

      [1] Product of the *cap* gene, also called *crp* (cAMP receptor protein).

      [2] Is a dimer

      [3] Binds cAMP, and then the cAMP-CAP complex binds to DNA at specific sites

3. Binding site for cAMP-CAP

   a. In the *lac* operon, the binding site is a region of about 20 bp located just upstream from the promoter, from -52 to -72.

   b. The pentamer TGTGA is an essential element in recognition. For the *lac* operon, the binding site is a dyad with that sequence in both sides of the dyad.

   c. Contact points betwen cAMP-CAP and the DNA are close to or coincident with mutations that render the *lac* promoter no longer responsive to cAMP-CAP.

   d. cAMP-CAP binds on one face of the helix.

Figure 4.1.5. Binding site for cAMP-CAP

Mutations that make
promoter nonresponsive
to CAP

A   T
T   A

-70   -60   -50

5'ATGTGAGTTAGCTCACACATT
3'TACACTCAATCGAGTGTGTAA

Dyad axis

Nucleotides in contact
with cAMP-CAP

Promoter

4. Binding of cAMP-CAP to its site will enhance efficiency of transcription initiation at promoter

    a. The *lac* promoter is not a particularly strong promoter. The sequence at -10, TATGTT, does not match the consensus (TATAAT) at two positions.

    b. In the presence of cAMP-CAP, the RNA polymerase will initiate transcription more efficiently.

    c. The *lac* UV5 promoter is an up-promoter mutation in which the -10 region matches the consensus. The *lac* operon driven by the UV5 promoter will achieve high level induction without cAMP-CAP, but the wild-type promoter requires cAMP-CAP for high level induction.

Figure 4.1.6. Regulatory region of *lac* operon, including CAP binding site
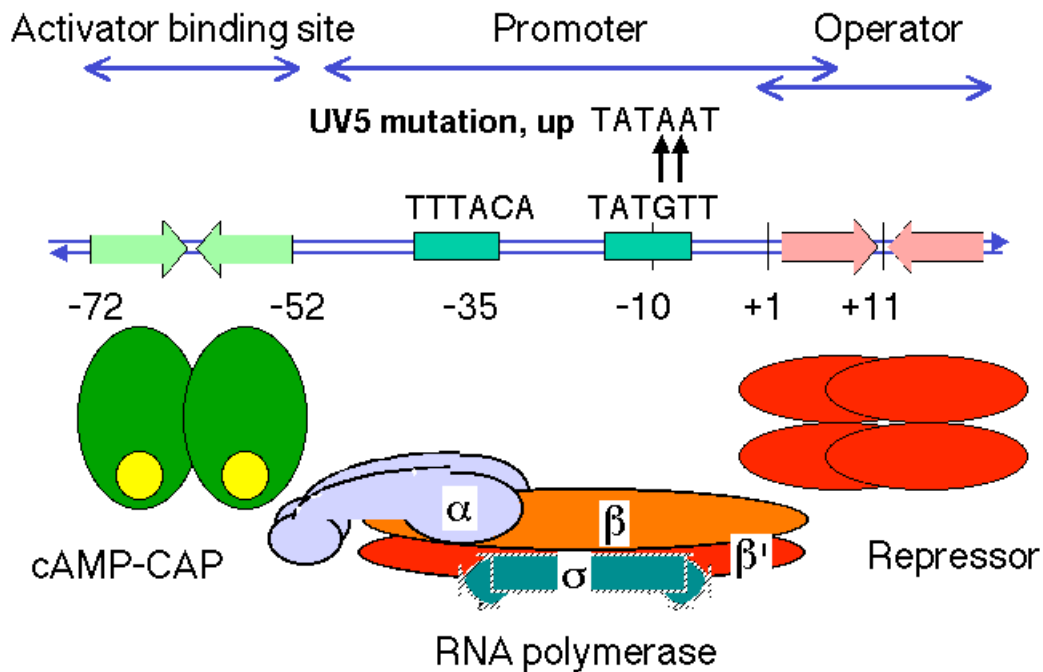
5. **Mode of action of cAMP-CAP**

    a. Direct positive interaction with RNA polymerase. The C-terminus of the α subunit is required for RNA polymerase to be activated by cAMP-CAP. This will be explored in more detail in Chapter 16.

    b. cAMP-CAP bends the DNA about 90°.

Figure 4.1.7. DNA (top helical structure) is bent by the CAP dimer .
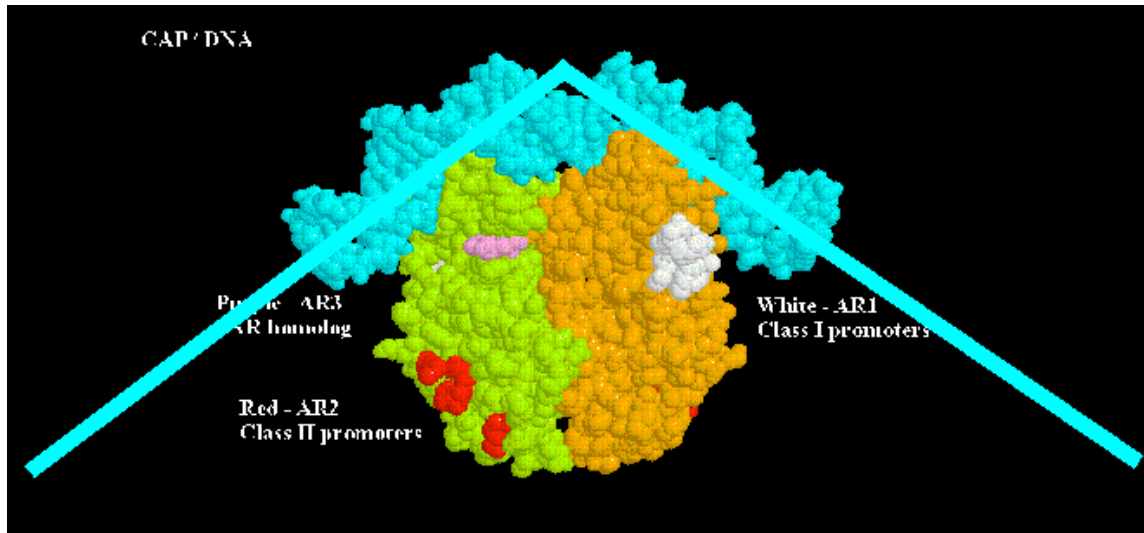
I. Some generalities:

1. Repressors, activators and polymerases interact primarily with one face of the DNA double helix.

2. Regulatory proteins, such as activators and repressors, are frequently symmetrical and bind symmetrical sequences in DNA.

3. RNA polymerases are not symmetrical, and the promoters to which they bind also are asymmetrical. This confers directionality on transcription.

Questions for Chapter 15. Positive and Negative Transcriptional Control

15.1   Amber mutations are one class of nonsense mutations. They lead to premature termination of translation by alternation of an amino acid-encoding codon to a UAG terminator, e.g. CAG (Gln) may be changed to UAG (stop; amber). The phenotype of such amber mutants can be suppressed by amber-suppressor genes, which are mutant tRNA genes that encode tRNAs that recognize UAG codons and allow insertion of an amino acid during translation. Which genes or loci in the *lac* operon can give rise to amber-suppressible mutations?

15.2   (POB) Negative regulation.
In the *lac* operon, describe the probable effect on *lacZ* gene expression of:
a) Mutations in the *lac* operator
b) Mutations in the *lacI* gene
c) Mutations in the promoter

15.3   Consider a <u>negatively</u> controlled operon with two structural genes (<u>A</u> and <u>B</u>, for enzymes A and B) an operator gene (<u>0</u>) and a regulatory gene (<u>R</u>). In the wild-type haploid strain grown in the absence of inducer, the enzyme activities of A and B are both 1 unit. In the presence of an inducer, the enzyme activities of A and B are both 100 units. For parts a-d, choose the answer that best describes the enzyme activities in the designated strains.

		Uninduced		Induced	
		Enz A	Enz B	Enz A	Enz B
a)	$R^+0^CA^+B^+$				
		a)	1	1	100	100
		b)	1	100	100	1
		c)	50	50	100	100
b)	$R^-0^+A^+B^-$				
		a)	1	1	100	100
		b)	100	100	100	100
		c)	100	0	100	0
c)	$R^+0^CA^+B^+/R^+0^+A^+B^+$				
		a)	2	2	200	200
		b)	51	51	200	200
		c)	200	2	2	200
d)	$R^-0^+A^+B^+/R^+0^+A^+B^+$				
		a)	2	2	200	200
		b)	2	101	2	101
		c)	200	200	200	200

15.4   (POB) Positive regulation.
A new RNA polymerase activity is discovered in crude extracts of cells derived from an exotic fungus. The RNA polymerase initiates transcription only from a single, highly specialized promoter. As the polymerase is purified, its activity is observed to decline. The purified enzyme is completely inactive unless crude extract is added to the reaction mixture. Suggest an explanation for these observations.

15.5 Consider a hypothetical regulatory scheme in which citrulline induces the production of urea cycle enzymes. Four genes (citA, citB, citC, citD) affecting the activity or regulation of the enzymes were analyzed by assaying the wild-type and mutant strains for argininosuccinate lyase activity and arginase activity in the absence (-cit) or presence (+cit) of citrulline. In the following table, wild-type alleles of the genes are indicated by a + under the letter of the cit gene and mutant alleles are indicated by a - under the letter. The activities of the enzymes are given in units such that 1 = the uninduced wild-type activity, 100 = the induced activity of a wild-type gene, and 0 = no measurable activity. In the diploid analysis, one copy of each operon is present in each cell.

Strain number	genes								lyase activity		arginase act.	
Haploid:	A	B	C	D					- cit	+ cit	- cit	+ cit
1	+	+	+	+					1	100	1	100
2	-	+	+	+					100	100	100	100
3	+	-	+	+					0	0	1	100
4	+	+	-	+					100	100	100	100
5	+	+	+	-					1	100	0	0
Diploid:	A	B	C	D /	A	B	C	D				
6	+	+	+	- /	+	-	+	+	1	100	1	100
7	-	+	+	+ /	+	-	+	+	1	100	2	200
8	+	+	-	+ /	+	-	+	-	100	100	100	100
9	+	-	-	+ /	+	+	+	-	1	100	100	100

Use the data in the table to answer the following questions.

a) What is the phenotype of the following strains with respect to lyase and arginase activity? A single word will suffice for each phenotype.

	Lyase activity	Arginase activity
Strain 2	_____	_____
Strain 3	_____	_____
Strain 4	_____	_____
Strain 5	_____	_____
Strain 6	_____	_____

b) What can you conclude about the roles of *citB* and *citD* in the activity or regulation of the urea cycle in this organism? Brief answers will suffice.

c) What is the relationship (recessive or dominant) between wild-type and mutant alleles of *citA* and *citC*? Be as precise as possible in your answer.

d) What can you conclude about the roles of *citA* and *citC* in the activity or regulation of the urea cycle in this organism? Brief answers will suffice.

15.6 Consider a hypothetical operon responsible for synthesis of the porphyrin ring (the heterocyclic ring that is a precursor to heme, cytochromes and chlorophyll). Four genes or loci, *porA, porB, porC,* and *porD* that affect the activity or regulation of the biosynthetic enzymes were studied in a series of haploid and diploid strains. In the following table, wild-type alleles of the genes or loci are indicated by a + under the letter of the *por* gene or locus and mutant alleles are indicated by a — under the letter. The activities of two enzymes involved in porphyrin biosynthesis, δ-aminolevulinic acid synthetase and δ-aminolevulinic acid dehydrase (referred to in the table as ALA synthetase and ALA dehydrase), were assayed in the presence or absence of heme (one product of the pathway). The units of enzyme activity are 100 = non-repressed activity of the wild-type enzyme, 1 = repressed activity of the wild-type enzyme (in the presence of heme), and 0 = no measurable activity. In the diploid analysis, one copy of each operon is present in each cell.

Strain number	*por*				ALA synthetase		ALA dehyd.	
					- *heme*	+*heme*	- *heme*	+ *heme*
Haploid:	A	B	C	D				
1	+	+	+	+	100	1	100	1
2	-	+	+	+	100	100	100	100
3	+	-	+	+	0	0	100	1
4	+	+	-	+	100	1	0	0
5	+	+	+	-	100	100	100	100

Diploid:	A B C D / A B C D								ALA synth -heme	+heme	ALA dehyd -heme	+heme
6	+ - + + / + + - +								100	1	100	1
7	- + + + / + + - +								200	101	100	100
8	+ + + - / + + - +								200	2	100	1
9	- + - + / + - + -								100	100	100	1

Use the data in the table to answer the following questions.

a) Describe the phenotype of the following the strains with respect to ALA synthetase and ALA dehydrase activities. A single word will suffice for each phenotype.

    ALA synthetase   ALA dehydrase

Strain 2 _____   _____

Strain 3 _____   _____

Strain 4 _____   _____

Strain 5 _____   _____

Strain 6 _____   _____

b) What is the relationship (dominant or recessive) between wild-type and mutant alleles of the four genes, and which strain demonstrates this? Please answer in a sentence with the syntax in this example: "Strain 20 is repressible, which shows that mutant *grk*1 is dominant to wild-type."

    porA Strain ___ is _____, which shows that _____

porA is _____.

    porB   Strain ___ is _____, which shows that _____

porB is _____.

    porC   Strain ___ is _____, which shows that _____

porC is _____.

    porD   Strain ___ is _____, which shows that _____

porD is _____.

    c)   What is the role of each of the genes in activity or regulation of porphyrin biosynthesis? Brief phrases will suffice.

    d)   Is this operon under positive or negative control?