Biol 213 Genetics: Friday, October 27, 2000
Lac Operon (Part II)

Most students find the *lac* operon complicated and confusing. If it weren't, it would not be able to perform all the functions that *E. coli* demands of it. The *lac* operon is useful to us as well, because it contains within it the most important themes of gene regulation. We are focusing our attention on the *lac* operon, rather than surveying many genes in a more superficial way, so that we can gain a deep understanding of some strategies of regulation that can be applied to a large number of genes. You will find that the concepts learned through study of the *lac* operon will appear again as we consider eukaryotic gene regulation and the control over multicellular development.

Outline

I. THE OPERATOR AND THE NATURE OF REPRESSION

II. THE MOLECULAR NATURE OF BINDING SITES: the promoter and the operator
pp.341-343

III. COMPLEMENTATION TESTS OF ELEMENTS OF *LAC* REGULATION

IV. POSITIVE REGULATION AND CRP, pp.404-405
- A. Symbolic representation of a cellular state: cAMP
- B. Model of positive regulation: CRP protein
- C. Experimental test of positive regulation of the *lac* operon through CRP and cAMP

V. WHAT CAN THE *LAC* OPERON TEACH US?
- A. Integration of Signals
- B. Types of regulation

I. The operator and the nature of repression

We left each other Wednesday with visions of the *lac* operon (e.g. Fig. 15-4) dancing in our heads, a picture that represents the product of the 20 years Jacques Monod had pursued the problem of how the expression of ß-galactosidase is regulated. That picture depicts a repressor protein as a critical element of the regulation, the element that mediates the response of *lacZ* to the presence or absence of lactose. However, by the end of the 1950's, when much of this picture was coming together, it was by no means clear how ß-galactosidase expression was regulated.

Monod had started off with a completely different notion. He figured that since lactose increases ß-galactosidase activity and *lacI* is required for proper regulation, lactose must act through the LacI product to induce expression (which is why he named the gene *lacI* -- *I* for induction). We know now that the protein encoded by *lacI* is a repressor, not an inducer and slowly Monod came to accept that possibility. But how to put the notion of a repressor to the test? Monod had the following test in mind. He

reasoned that if a repressor existed, it must repress something, and it should be possible to find a mutation in that something. Just as you can ruin the regulation of your front door by mutating either the lock or the key, so you should be able to ruin the regulation of the *lac* operon by mutating either the repressor or the object of repression.

Easily stated, not so easily achieved. Monod posed this problem to Francois Jacob, a coworker at Institut Pasteur in Paris: How do you find a hypothetical mutant defective in whatever the *lacI* "repressor" acts on (if *lacI* indeed encodes a repressor). Let's think this through with Jacob.

SQ1. Which scheme below is a good way to find a mutation in the object of repression of the hypothetical *lacI* repressor?

a. Mutagenize *E. coli*, screen mutants for the ability to make ß-galactosidase in the absence of inducer (i.e., no lactose or IPTG)

b. Mutagenize *E. coli*, select for mutants able to grow on PGal as the sole carbon source (recall that PGal is a substrate of ß-galactosidase but not an inducer)

c. Mutagenize a strain of *E. coli* that has two copies of *lacI*, select for mutants able to grow on PGal as the sole carbon source

Let's look at choice **a**. Recall that mutations of any specific type are rare. You'd be lucky if one mutagenized cell in a million carried a mutation in a specific gene. Even though student labor is rather cheap, consider how many student-hours would need to be spent going through millions of *E. coli* for a specific phenotype!

Let's look at choice **b**. In looking for very rare mutations, selection is much more practical than screens. In lab, you looked for rare recombinations between two mutations within *rII* by selecting for the ability to grow on *E. coli* K. It would have been horrible if you had infected *E. coli* B instead and had to screen thousands of plaques for one that was unusually large. Recall that PGal is a substrate for ß-galactosidase but does not bind to Lac repressor. Therefore, it cannot normally support growth: ß-galactosidase will not be produced because the repressor will remain bound to the operator, even in the presence of PGal. Demanding growth on PGal is therefore a way to select for rare mutants defective in repression of the *lac* operon. Unfortunately, the mutations found by this method generally mapped to *lacI*. The mutants simply lacked repressor, not the object of repression. A good idea, but it didn't work. We'll see why in a moment.

Let's look at choice **c**. Jacob hit upon a clever alternative, mutagenizing a strain with two copies of *lacI*.

SQ2. If there is a one in 10^6 chance that *lacI* is affected by mutation, what is the probability that two copies of *lacI* (each on a different molecule of DNA) are both affected?

This strategy therefore effectively eliminates the possibility of picking up *lacI* mutants. Unfortunately, *E. coli* comes with only one copy of *lacI*. Where can we pick up a second?

If you were at Institut Pasteur, you might have made yourself a hero and suggested pUR3! Recall that this plasmid from Lab 2 carries *lacI* as well as the entire *lac* operon. A normal *E. coli* strain carrying pUR3 would have two copies of *lacI*: the copy on pUR3 and the natural copy on the chromosome. Even though molecular cloning was a dozen years in the future, tricks had been developed to cause the cell to cut out regions of the chromosome and put them on indigenous plasmids. A plasmid (F'*lac*) was already available that carried the *lac* region. Now, for the first time, it was possible to find mutations affecting repression that were even more rare than those within *lacI*.

Now we can ask the big question: What does the repressor act on? Is the object ß-galactosidase itself? Maybe the repressor normally binds to the enzyme rendering it inactive. A mutant that can grow on PGal might then have a mutant ß-galactosidase that doesn't bind repressor and is always active. Or maybe the object of repression is an unknown protein that binds to the repressor and mediates its effects. How can we tell?

SQ3. You have available to you deletion mutants of *E. coli* covering different parts of the *lac* operon and *lacI* (just as you had several deletion mutants of T4 in the lab you just completed). How can you use these to assess whether mutants that cannot be repressed have a mutant ß-galactosidase that doesn't bind repressor?

Mapping is one of the basic tools of the geneticist, providing important clues to the function of genetic elements. Surprisingly, all the *E. coli*[F'*lac*] mutants able to use PGal turned out to have mutations that mapped to the same position: a tiny region right next to *lacZ*. This region Jacob and Monod termed the operator (*o*) and postulated that it is the DNA binding site for the Lac repressor (see Figs. 15-4 and 15-8 for pictures of the Lac repressor binding to the Lac operator). Mutant operators were called o^c (for *o*perator-*c*onstitutive, i.e. a defective operator that causes constitutive expression). We now know that mutants with defective operators are unable to bind the Lac repressor .

SQ4. Would the phenotype of a *lacI* mutant differ from the phenotype of a o^c mutant? If so, how? If not, why not? We're talking here of an otherwise normal *E. coli*, i.e. one that does not carry any extra copies of genes.

II. The molecular nature of binding sites: the promoter and the operator (pp.341-343)

Mutation identified another small region of importance, very close to the operator. Mutants in this region, called the promoter (*p*), had no expression of the enzymes encoded by the *lac* operon (i.e., ß-galactosidase or transacetylase) under any condition. It turns out that the promoter is the region that RNA polymerase binds to in order to initiate transcription of the operon. Fig 15-4 shows this interaction.

We thus know of two regions of DNA, the operator and the promoter, that are not genes but are nonetheless very important. Both regions are small compared to genes. Both regions are binding sites for protein.

Since the Lac repressor binds specifically to the operator and not to random DNA sequences, it follows that the base sequence of the operator is somehow sensed by

the repressor. Fig. 1 the base sequence of the Lac operator. It consists of about 20 basepairs of DNA that are protected by the binding of the Lac repressor. Mutations that change this sequence can diminish the binding of repressor and leave the *lac* operon constitutively expressed.

By the same reasoning, RNA polymerase must recognize the specific base sequence of the promoter. The sequence in the *lac* regulatory region that RNA polymerase binds to is shown in Fig. 1. There are of course thousands of genes in the *E. coli* genome that must be transcribed, and RNA must bind to promoters in front of each. The enzyme isn't very finicky, as illustrated by the list of promoter sequences in Fig. 13-3. There are two regions where RNA polymerase makes contact with the DNA: 6 bases positioned about 10 bases upstream from the transcription start site (called the -10 region) and 6 bases positioned about 35 bases upstream (called the -35 region). Mutations in either of these regions can prevent binding of RNA polymerase and thus prevent transcription.

The closer a promoter is to the perfect -35 and -10 sequence, the better RNA polymerase binds to it, and the more it promotes transcription. The *lac* promoter has a -35 region TTTACA (differing by one base from the perfection defined in Fig. 13-3) and a -10 region of TATGTT (differing by two bases).

SQ5. A mutation (called *UV5*) of the *lac* promoter changes the -10 sequence from TATGTT to TATAAT. How would you expect this mutation to affect transcription of the *lac* operon?

SQ6. Suppose that the promoter region were flipped around. What would be the result? Would *lacZ* be expressed? Would *lacI*?

SQ7. What are the phenotypes of the following mutants (use + or -)?

	ß-galactosidase activity		
Genotype	**- lactose**	**+ lactose**	**Grows on lactose**
wild-type			
lacI$^-$			
lac o^c			
lac p$^-$			
lac p$^-$ *lac o*^{c}			

III. Complementation tests of elements of *lac* regulation

The experimental test of the ideas described above relied on complementation tests. Complementation requires that the genes being tested exist in multiple copy, so Jacob and Monod made use again of a plasmid (F'*lac*) carrying the *lac* region. Remember that wild-type genes can complement mutant genes because the former produce a protein not made by the latter. This is not the case with binding sites, like operators and

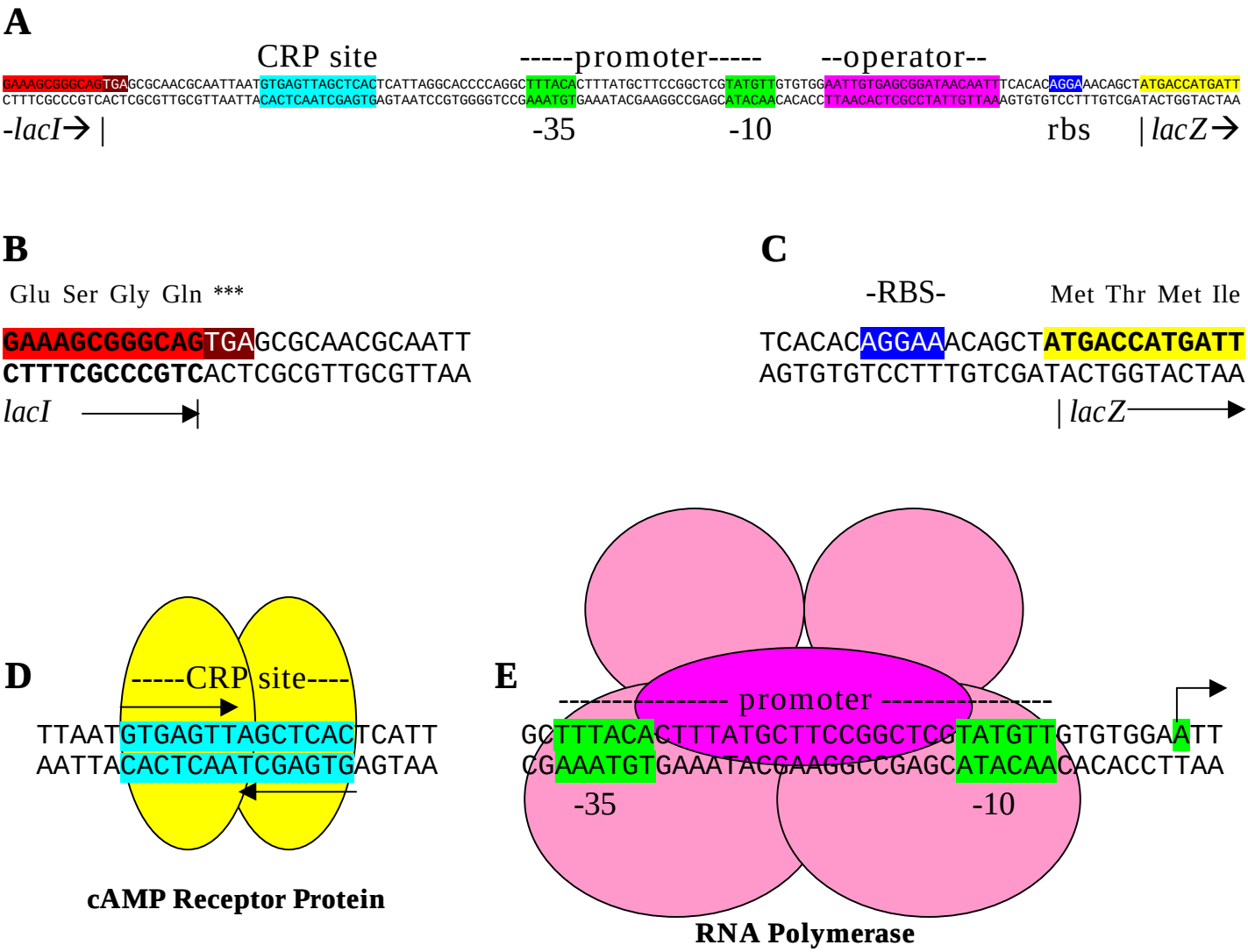

Fig. 1. Nucleotide sequence of the regulatory region of the Lac operon. Sites colored on both strands indicate DNA binding sites for protein. Sites colored on only one strand indicate features of interest on the transcribed RNA. Panel **A** shows the nucleotide sequence of the region between *lacI* and *lacZ*, containing some of sites important in the regulation of the Lac operon. Panel **B** and **C** show the end of *lacI* and the beginning of *lacZ*, respectively. The ribosomal binding site (RBS) preceding *lacZ* is highlighted. Panel **D** shows cAMP Receptor Protein (CRP) binding to its binding site. CRP is a dimeric protein, each subunit recognizing 5'-GTGAGTT-3' (shown by arrows). Panel **E** shows RNA polymerase binding to the Lac promoter at two sites: approximately 10 and 35 nucleotides upstream from the start of base at which transcription begins (shown by an arrow pointing in the direction of transcription). Panel **F** shows the Lac repressor binding to the operator. The repressor is a dimeric protein, each subunit recognizing 5'-AATTGT-3' (shown by arrows).

promoters. Binding sites produce nothing, so a wild-type operator cannot complement a mutant operator.

Figures 2A and 2B illustrate the difference between genes, which can complement, and binding sites, which cannot, a key distinction in Jacob and Monod's experiments. In panel **A,** the Lac repressor, encoded by *lacI*, binds to the operator (*o*) of the *lac* operon on the chromosome and the operator on F'*lac* and prevents transcription. Mutant *lacI*$^-$ does not produce repressor, but that's all right, because there's plenty of repressor in the cell already from the wild-type *lacI*. Notice that the figure does not specify which molecule of DNA is the chromosome and which is the plasmid. It doesn't make any difference. A repressor from either DNA represses all operators in the cell. The phenotype is therefore wild type: expression of ß-galactosidase dependent upon the presence of inducer.

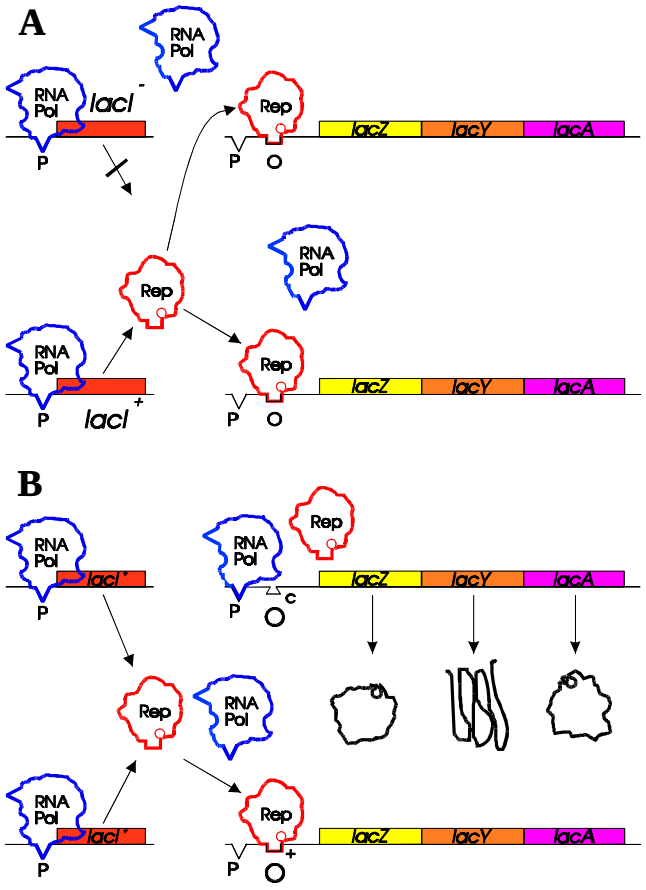

Fig. 2. Example of Jacob and Monod's complementation test. Panel **A** illustrates the phenotype of *E. coli* (*lacI*$^+$) into which F'*lac* (*lacI*$^-$) is introduced. The repressor encoded by *lacI*$^+$ acts on both operators, repressing transcription. Panel **B** illustrates the phenotype of *E. coli* (*O*$^+$) into which F'*lac* (*O*^{C}) is introduced. The defective operator cannot interact with repressor.

Fig. 2B presents a seemingly similar situation with very different results. Here, both molecules make wild type repressor from *lacI*$^+$, but one molecule has a normal operator while the other one has *o*^{c}, an mutant operator that cannot bind repressor. This time the repressor works on only <u>one</u> of the two operons. The genes downstream from *o*^{c} are expressed whether or not inducer is present.

We say that a gene that makes a diffusible regulatory protein, acts in <u>trans</u> (distant): its product acts not only on the DNA molecule that made it but on <u>any</u> DNA molecule in the cell with the appropriate binding site. In contrast, the binding site acts in <u>cis</u> (close): its effects are felt only by the DNA to which it is attached.

SQ8. Just to see if we understand these terms, does a mutation in a start codon act in cis or in trans?

SQ9. Would a mutation in a promoter act in cis or in trans?

Table 1. Expression of Lac operon in *E. coli* with or without F'*lac*

	Genotype[a]	ß-galactosidase activity[b] - IPTG	ß-galactosidase activity[b] + IPTG	Permease activity[b] - IPTG	Permease activity[b] +IPTG
1	$I^+ Z^+ Y^+$	~	+	~	+
2	$I^- Z^+ Y^+$	+	+	+	+
3	$I^+ Z^- Y^+$ / F': $I^- Z^+ Y^+$	~	+	~	+
4	$O^c Z^+ Y^+$	+	+	+	+
5	$O^+ Z^- Y^-$ / F': $O^c Z^+ Y^+$	+	+	+	+
6	$O^+ Z^+ Y^+$ / F': $O^c Z^- Y^+$	-	+	+	+
7	$O^+ Z^- Y^+$ / F': $O^c Z^+ Y^-$	+	+	-	+
8	$I^S O^+ Z^+ Y^+$ / F': $I^+ O^c Z^+ Y^+$	+	+	+	+

[a] *I, O, Z,* and *Y* refer to the genotypes of *lacI, lac* operator, *lacZ*, and *lacY*, respectively. I^S represents the *lacI*^{S} allele, which encodes a Lac repressor unable to bind inducer. The first genotype shown is that of the *E. coli* chromosome. The second (if present) is that of F'*lac*. Any genotype not shown is wild type.

[b] ß-galactosidase activity and Lac permease activity in the presence or absence of IPTG is reported either as high (+) or low (~).

Now we can grasp the experiment performed by Jacob and Monod, some results of which are shown in Table 1. They brought different versions of F'*lac* that had different genotypes for *lac* genes into different strains of *E. coli* that had its own *lac* genotype. It's important to realize that when the genotype for a gene is not given, it is wild type. Each line with a genotype including F'*lac* reports the results of a complementation test, which taken together tells us a lot about how LacI and the Lac operator work.

SQ10. Look at line 6 of Table 1. What does "$O^+ Z^+ Y^+$ / F': $O^c Z^- Y^+$" mean? Is ß-galactosidase expression in this strain be dependent upon the presence of inducer? Why (not)?

IV. Positive regulation and CRP (pp.404-405)

IV.A. Symbolic representation of a cellular state: cAMP

We've come a long way from Monod's original experiment demonstrating diauxic growth in *E. coli*. Let's see if we can now explain his results.

SQ11. Given the nature of the operator and the promoter and the ability of the Lac repressor to regulate the expression of the Lac operon, how can you explain the diauxic growth of *E. coli* grown on glucose + lactose?

All the wonderful molecular machinery we've discussed tells us a lot about how *E. coli* ensures that ß-galactosidase and related protein are made only when lactose is present, but it doesn't tell us anything about how *E. coli* prevents the synthesis of these protein in the presence of the preferred carbon source, glucose! Something beyond what we've considered must allow *E. coli* to sense the level of glucose and act on it. Part of the mechanism concerns the molecule cyclic AMP (cAMP) (see Fig. 3).

Cyclic AMP is formed from ATP by joining the innermost phosphate to a second hydroxyl group. What is cAMP good for? The remarkable thing is that it has no real function. Make no mistake: it is important, but unlike ATP and AMP it has no direct chemical function. Rather, it acts as a pure symbol.

A pure symbol... let me explain. At an intersection, a red light symbolizes that we should stop. The light, of course, has no power to stop cars. It works only because we agree to its meaning, and so all cars stop at red lights, except those of a few mutants. All organisms use molecules as symbols. For example, we use adrenaline to symbolize the need for heightened metabolism in the face of stress. cAMP is another commonly used symbol. The molecule is not food, it doesn't do anything, but it serves as a signal that the cell is hungry for sugar. Any gene or enzyme that ought to turn on or off in response to that information can recognize the message of cAMP and react accordingly.

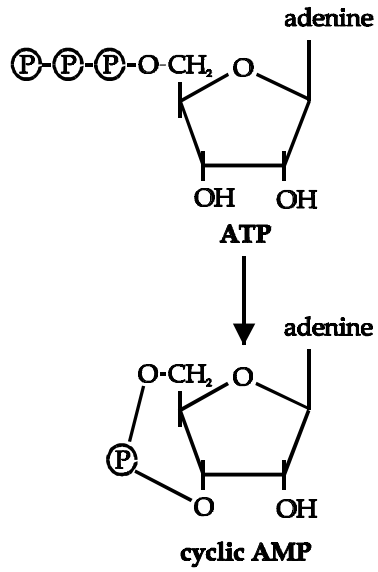

Fig. 3. Formation of cyclic AMP (cAMP). cAMP is formed from a common metabolite (ATP) and is easily recognized by protein due to its shape and negative charge (from the phosphate).

IV.B. Model of positive regulation: CRP protein

We can see red lights, but how can *E. coli* see cAMP? Just as there is the Lac repressor to sense lactose, there is a protein to bind to cAMP and act on its message. That protein, cAMP receptor protein (CRP), binds to cAMP and produces the response to sugar starvation. Fig. 15-6 illustrates how CRP protein complexed with cAMP binds close to the *lac* promoter and increases the affinity of the promoter for RNA polymerase and consequently the transcription of the *lac* operon.

SQ12. For the scheme of glucose control suggested above to work, the level of cAMP must change in response to the level of sugar. Should it go up or down in response to glucose? In response to lactose?

The binding site for CRP is another short DNA sequence (Fig. 1B). We've now discussed three protein binding sites: that for CRP (CRP site), for RNA polymerase (the promoter), and for the repressor (the operator). If you take away the protein binding sites from between *lacI* and *lacZ*, there's not much left. That's most of what bacterial DNA is: genes with protein binding sites preceding them.

SQ13. Consider the possible order of the elements shown in Fig. 1A. Which would not make sense?

a. Lac promoter – *lacZ* – *lacY* – *lacA* – *Lac* operator

b. Ribosome binding site – promoter – *lacZ*

c. *lacZ* – *lacZ* start codon – Lac promoter

IV.C. Experimental test of positive regulation of the *lac* operon through CRP and cAMP

The model shown in Fig. 15-6 provides a satisfying explanation of the diauxic effect found by Monod. The model rests on several experimental tests of its assumptions:

1. CRP binds upstream from the promoter, dependent upon the presence of cAMP
   Binding studies demonstrate this to be true
2. CRP protein is required for good expression of the *lac* operon
   Mutation of crr (encoding CRP) causes near loss of lac expression
3. CRP binding is required for good expression of the *lac* operon
   The L8 mutation, which destroys the CRP binding site, causes near loss of lac expression (see Fig. 4B).
4. The concentration of cAMP is responsive to the level of glucose
   This has been shown to be true

Curiously, one important assumption has not been met:

5. Unlike glucose, lactose should not increase the level of cAMP
   The level of cAMP in lactose-grown cells is the same as in glucose-grown cells.

This result poses a serious challenge to the model, since the model relies on low cAMP levels in the presence of glucose and high levels in the presence of lactose to explain the diauxic effect.

SQ14. Draw a graph showing the growth of *E. coli* on a 1:1 mixture of glucose and lactose and superimpose on it the predicted levels of ß-galactosidase and of cAMP.

Recently, Kimata et al (1997) [Proc Natl Acad Sci USA 94:12914-12919] put the model to an additional test. They used a mutant *E. coli* that carried the same *L8* mutation described above (point 3) that also had a mutant *lac* promoter, altered so that it binds RNA polymerase even in the absence of CRP protein. This promoter mutation is called *UV5* (see Fig. 4C).

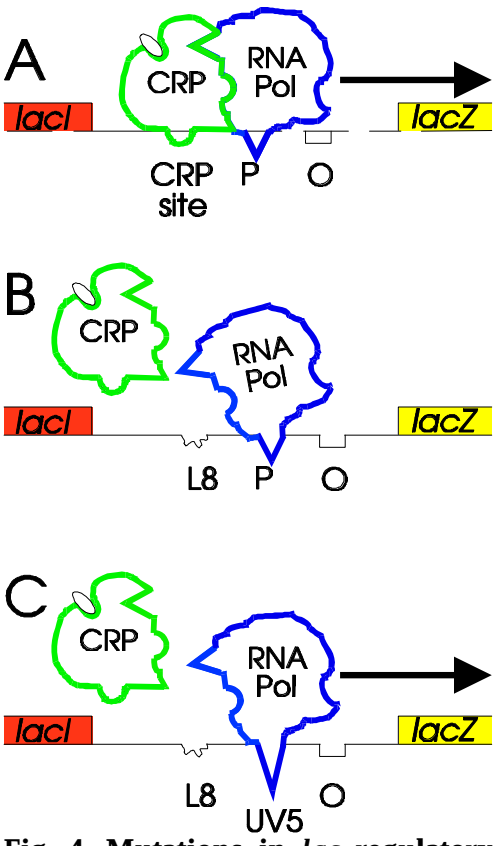

Fig. 4. Mutations in *lac* regulatory region. Only part of the *lac* operon is shown. (**A**) Wildtype regulatory region. The binding of CRP to its site stabilizes the binding of RNA polymerase to the promoter, resulting in a high level of transcription. (**B**) The *L8* mutation prevents CRP from binding. Poor binding of RNA polymerase to the promoter leads to a low level of transcription. (**C**) The *UV5* mutation strengthens the promoter so that RNA polymerase can bind stably without the help of CRP, leading to a high level of transcription.

SQ15. Justify the following predictions for the *L8 UV5* double mutant.

a. In the absence of inducer, ß-galactosidase activity should be low
b. In the presence of IPTG, ß-galactosidase activity should be high
c. In the presence of IPTG and glucose, ß-galactosidase should remain high
d. The strain grown in glucose + lactose should not exhibit the diauxic effect

One result from Kimata et al's experiments is shown in Fig. 5. The experiment monitored both growth of *E. coli* on glucose + lactose -- the usual diauxic condition -- and the appearance of ß-galactosidase during growth. Compare this figure with that of Fig. 1 of Wednesday's notes.

SQ.16. Is diauxic growth apparent in the conditions of this experiment?

SQ.17. Is the appearance of ß-galacto-sidase repressed by glucose in this experiment?

SQ.18. Describe how this experiment supports or refutes the model of the *lac* operon given in the text.

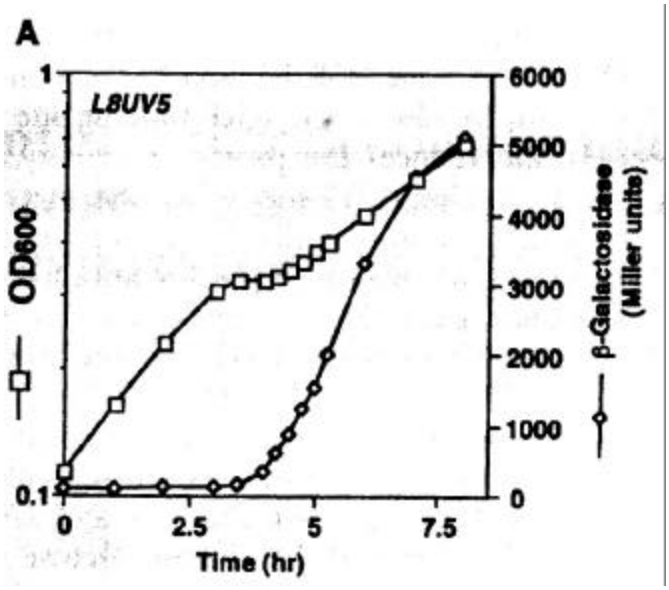

Fig. 5. Diauxic growth of *L8 UV5* double mutant. Growth (squares) and ß-galactosidase activity (diamonds) of an *E. coli* strain carrying the *L8 UV5* mutations in the control region of the *lac* operon was grown in a medium containing both glucose and lactose. Taken from Kimata et al (1997) Proc Natl Acad Sci USA 94:12914-12919.

V. What can the *lac* operon teach us?

V.A. Integration of Signals

E. coli must integrate different kinds of information from the environment to make informed metabolic decisions. For example, the *lac* operon should respond to both the level of lactose and the level of glucose. It is important that all genes whose products work together be expressed together.

SQ19. How are bacterial genes of related function arranged so that they are turned on or off at the same time?

SQ20. Control of what process regulates expression of the *lac* operon?

a. Replication of the gene
b. Transcription of the gene
c. Degradation of mRNA
d. Translation of mRNA
e. Degradation of protein

All of these control points are used by some genes, but regulation of the initiation of transcription, as exemplified by the *lac* operon, is the most common strategy in both prokaryotes and eukaryotes.

SQ21. What protein is the focal point that ultimately integrates information about the level of lactose and glucose?

V.B. Types of regulation

Regulation of gene expression is effected by the binding of protein to DNA regions that influence transcription. The protein often are affected by small molecules that directly or indirectly provide information about the environment (Table 2).

Table 2: Summary of characteristics of regulatory protein acting on *lac* operon

Protein	Target	Acts on	Influenced by
lac repressor	operator	*lac* operon	lactose
CRP	CRP site	several operons	cAMP (sugar)
RNA polymerase	promoter	hundreds of operons	CRP and other protein

Ironically, the effects of lactose, which acts to increase expression of the *lac* operon, are mediated by a protein (Lac repressor) whose function it is to decrease transcription. Likewise, the effects of glucose, which acts to decrease expression of the *lac* operon, are mediated by a protein (CRP) whose function it is to increase transcription.

The generality is that only protein can read DNA, so the effects of all environmental stimuli must be mediated by some protein that binds near the genes that ought to respond appropriately.

SQ22. You want to modify *E. coli* so that it will respond to plutonium by turning on the *lac* operon and appearing red on MacConkey plates, allowing it to serve as a biosensor for plutonium contamination in the environment. How do you need to engineer the bacterium to make this happen?