

The Operon

Within its tiny cell, the bacterium *E. coli* contains all the genetic information it needs to metabolize, grow, and reproduce. It can synthesize every organic molecule it needs from glucose and a number of inorganic ions.

Many of the genes in *E. coli* are expressed constitutively; that is, they are always turned "on". Others, however, are active only when their products are needed by the cell, so their expression must be regulated.

Two examples:

- If the amino acid tryptophan (Trp) is added to the culture, the bacteria soon stop producing the five enzymes previously needed to synthesize Trp from intermediates produced during the respiration of glucose. In this case, the presence of the products of enzyme action represses enzyme synthesis.
- Conversely, adding a new substrate to the culture medium may induce the formation of new enzymes capable of metabolizing that substrate. If we take a culture of *E. coli* that is feeding on glucose and transfer some of the cells to a medium contain lactose instead, a revealing sequence of events takes place.
 - At first the cells are quiescent: they do not metabolize the lactose, their other metabolic activities decline, and cell division ceases.
 - Soon, however, the culture begins growing rapidly again with the lactose being rapidly consumed. What has happened? During the quiescent interval, the cells began to produce three enzymes.

The three enzymes are

- a permease that transports lactose across the plasma membrane from the culture medium into the interior of the cell
- beta-galactosidase which converts lactose into the intermediate allolactose and then hydrolyzes this into glucose and galactose. Once in the presence of lactose, the quantity of beta-galactosidase in the cells rises from a tiny amount to almost 2% of the weight of the cell.
- a transacetylase whose function is still uncertain.

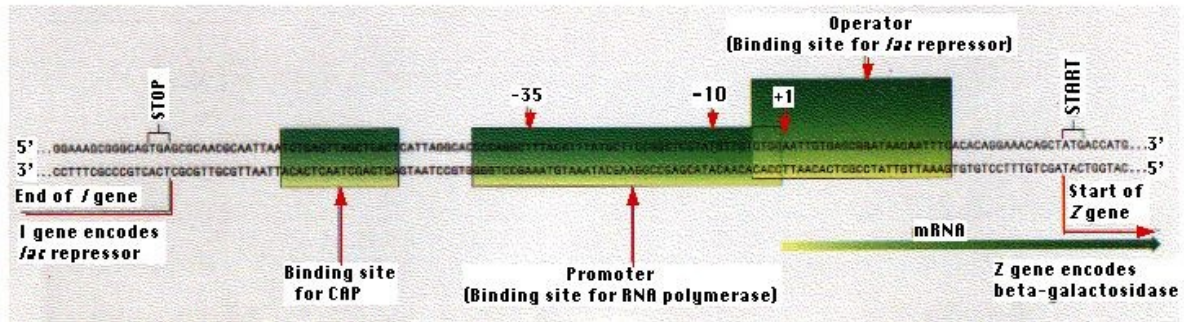
The *lac* operon

The capacity to respond to the presence of lactose was always there. The genes for the three induced enzymes are part of the genome of the cell. But until lactose was added to the culture medium, these genes were not expressed (β -galactosidase was expressed weakly — just enough to convert lactose into allolactose).

The most direct way to control the expression of a gene is to regulate its rate of transcription; that is, the rate at which RNA polymerase transcribes the gene into molecules of messenger RNA (mRNA).

Link to a general discussion of gene transcription.

Gene transcription begins at a particular nucleotide shown in the figure as "+1". RNA polymerase actually binds to a site "upstream" (i.e., on the 5' side) of this site and opens the double helix so that transcription of one strand can begin.

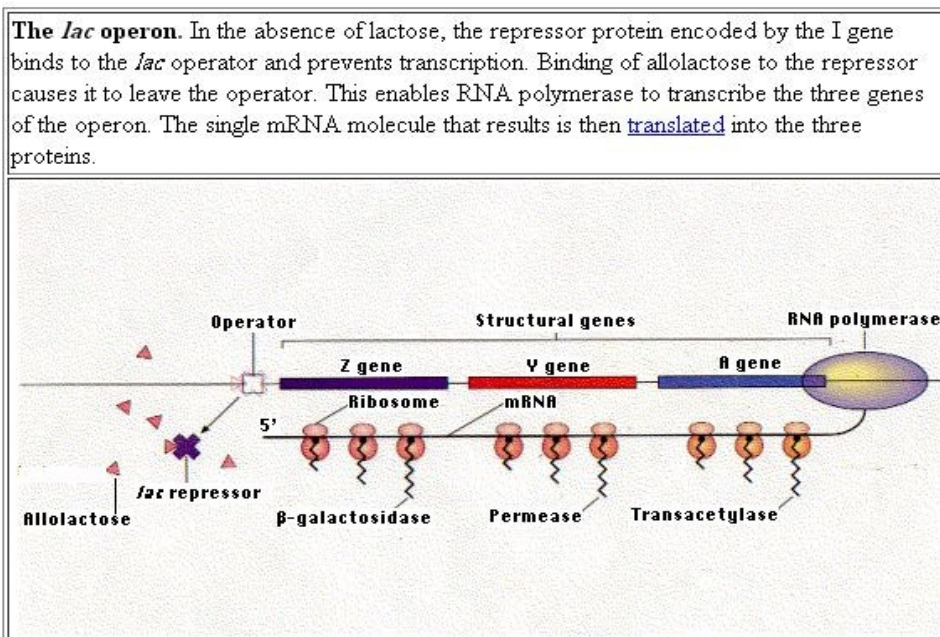


The binding site for RNA polymerase is called the promoter. In bacteria, two features of the promoter appear to be important:

- a sequence of TATAAT (or something similar) centered 10 nucleotides upstream of the +1 site and
- another sequence (TTGACA or something quite close to it) centered 35 nucleotides upstream.

The exact DNA sequence between the two regions does not seem to be important.

Each of the three enzymes synthesized in response to lactose is encoded by a separate gene. The three genes are arranged in tandem on the bacterial chromosome.



The *lac* repressor binds to a specific sequence of two dozen nucleotides called the operator. [Link to a discussion of how the DNA sequence of the operator site can be determined.] Most

of the operator is downstream of the promoter. When the repressor is bound to the operator, RNA polymerase is unable to proceed downstream with its task of gene transcription.

The *lac* repressor represents only a tiny fraction of the proteins in the *E. coli* cell. [Link to a discussion of how it can nonetheless be isolated and purified.](#)

The operon is the combination of the

- operator and
- the three protein-encoding genes associated with it.

The gene encoding the *lac* repressor is called the *I* gene. It happens to be located just upstream of the *lac* promoter. However, its precise location is probably not important because it achieves its effect by means of its protein product, which is free to diffuse throughout the cell. And, in fact, the genes for some repressors are not located close to the operators they control.

Although repressors are free to diffuse through the cell, how does — for example — the *lac* repressor find the single stretch of 24 base pairs of the operator out of the 4.6 million base pairs of DNA in the *E. coli* genome? It turns out the repressor is free to bind anywhere on the DNA using both

- hydrogen bonds and
- ionic (electrostatic) interactions between its positively-charged amino acids (Lys, Arg) and the negative charges on the deoxyribose-phosphate backbone of the DNA.

Once astride the DNA, the repressor can move along it until it encounters the operator sequence. Now an allosteric change in the tertiary structure of the protein allows the same amino acids to establish bonds — mostly hydrogen bonds and hydrophobic interactions — with particular bases in the operator sequence.

The *lac* repressor is made up of four identical polypeptides (thus a "homotetramer"). Part of the molecule has a site (or sites) that enable it to recognize and bind to the 24 base pairs of the *lac* operator. Another part of the repressor contains sites that bind to allolactose. When allolactose unites with the repressor, it causes a change in the shape of the molecule, so that it can no longer remain attached to the DNA sequence of the operator. Thus, when lactose is added to the culture medium,

- it causes the repressor to be released from the operator
- RNA polymerase can now begin transcribing the 3 genes of the operon into a single molecule of messenger RNA.

Hardly does transcription begin, before ribosomes attach to the growing mRNA molecule and move down it to translate the message into the three proteins. You can see why punctuation codons — UAA, UAG, or UGA — are needed to terminate translation between the portions of the mRNA coding for each of the three enzymes.

This mechanism is characteristic of bacteria, but differs in several respects from that found in eukaryotes:

- Genes in eukaryotes are not linked in operons (except for nematodes like *C. elegans* and tunicates like *Ciona intestinalis*).
- Primary transcripts in eukaryotes contain the transcript of only a single gene (with the above exceptions).

C. elegans differs from most eukaryotes in having a substantial fraction (15–20%) of its genes grouped in operons containing from 2 to 8 genes each. Like bacteria, all the genes in an operon are transcribed from a single promoter producing a single primary transcript (pre-mRNA). Some of the genes in these operons appear — as in bacteria — to be involved in the same biochemical function, but this may not be the case for most. *C. elegans* operons also differ from those in bacteria in that each pre-mRNA is processed into a separate mRNA for each gene rather than being translated as a unit.

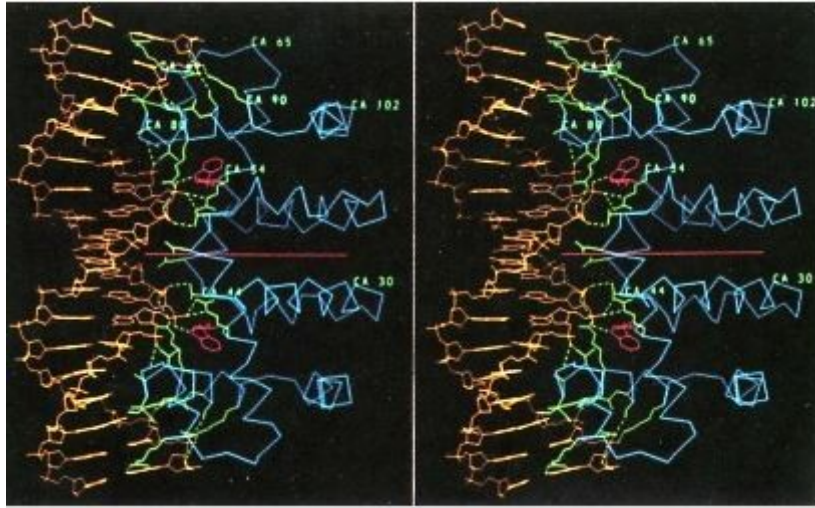
- Transcription and translation are not physically linked in eukaryotes as they are in bacteria; transcription occurs in the nucleus while translation occurs in the cytosol (with a few exceptions).

[Link to a discussion of gene regulation in eukaryotes.](#)

Corepressors

As mentioned above, the synthesis of tryptophan from precursors available in the cell requires 5 enzymes. The genes encoding these are clustered together in a single operon with its own promoter and operator. In this case, however, the presence of tryptophan in the cell shuts down the operon. When Trp is present, it binds to a site on the Trp repressor and enables the Trp repressor to bind to the operator. When Trp is not present, the repressor leaves its operator, and transcription of the 5 enzyme-encoding genes begins.

Stereo view of the tryptophan repressor (right side of each panel) bound to its operator DNA (left side). The repressor is a homodimer of two identical polypeptides (on either side of the horizontal red line). Binding to DNA occurs only when a molecule of tryptophan (red rings) is bound to each monomer of the repressor. You may be able to fuse the images by holding a stiff paper or cardboard between the views so that your left eye sees only the left-hand image, your right eye the right-hand image. Persevere; the results are worth it!. (Image courtesy of P. B. Sigler.)



The usefulness to the cell of this control mechanism is clear. The presence in the cell of an essential metabolite, in this case tryptophan, turns off its own manufacture and thus stops unneeded protein synthesis.

As its name suggests, repressors are negative control mechanisms, shutting down operons

- in the absence of a substrate (lactose in our example) or
- the presence of an essential metabolite (tryptophan is our example).

However, some gene transcription in *E. coli* is under positive control.

Positive Control of Transcription: CAP

Absence of the lac repressor is essential but not sufficient for effective transcription of the lac operon. The activity of RNA polymerase also depends on the presence of another DNA-binding protein called catabolite activator protein or CAP. Like the lac repressor, CAP has two types of binding sites:

- One binds the nucleotide cyclic AMP; the other
- binds a sequence of 16 base pairs upstream of the promoter (back up to graphic)

However, CAP can bind to DNA only when cAMP is bound to CAP. so when cAMP levels in the cell are low, CAP fails to bind DNA and thus RNA polymerase cannot begin its work, even in the absence of the repressor.

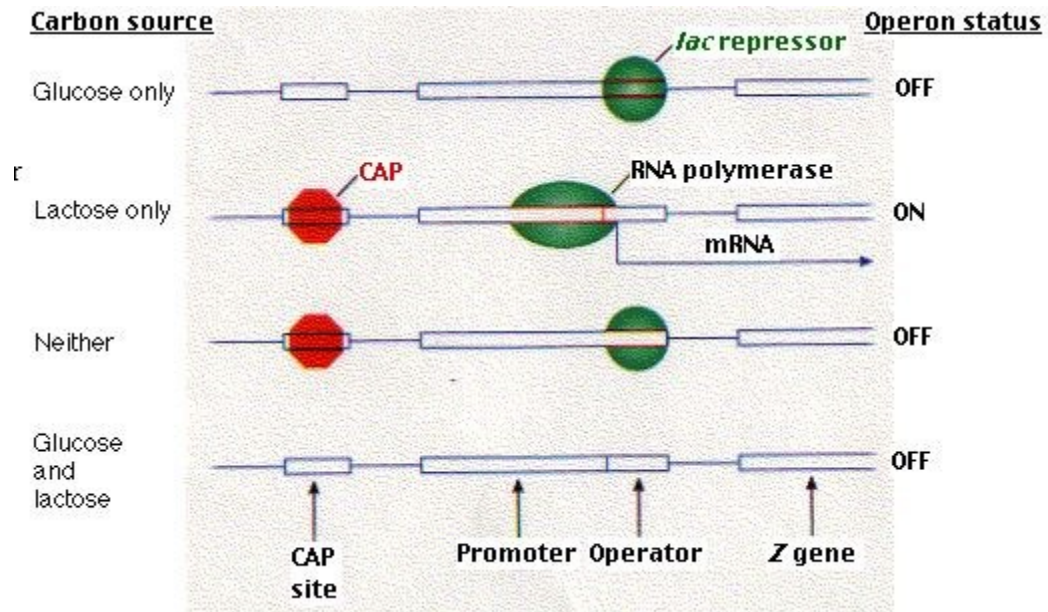
So the *lac* operon is under both negative (the repressor) and positive (CAP) control.

Why?

It turns out that it is not simply a matter of belt and suspenders. This dual system enables the cell to make choices. What, for example, should the cell do when fed both glucose and lactose? Presented with such a choice, *E. coli* (for reasons about which we can only speculate) chooses glucose. It makes its choice by using the interplay between these two control devices.

- Although the presence of lactose removes the repressor,
- the presence of glucose lowers the level of cAMP in the cell and thus removes CAP.

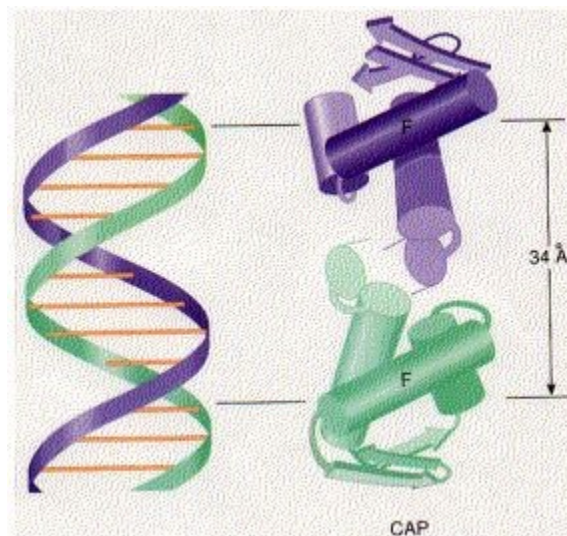
Without CAP, binding of RNA polymerase is inhibited even though there is no repressor to interfere with it if it could bind. The molecular basis for its choices is shown in the



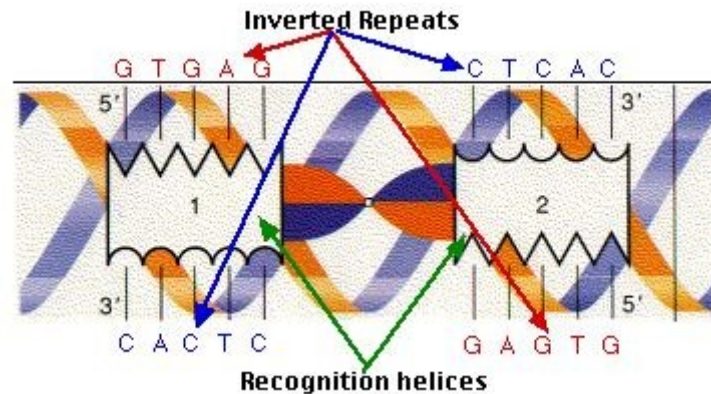
graphic.

CAP consists of two identical polypeptides (hence it is a homodimer). Toward the C-terminal, each has two regions of alpha helix with a sharp bend between them. The longer of these is called the recognition helix because it is responsible for recognizing and binding to a particular sequence of bases in DNA.

The graphic shows a model of CAP. The two monomers are identical. Each monomer recognizes a sequence of nucleotides in DNA by means of the region of alpha helix labeled F. Note that the two recognition helices are spaced 34Å apart, which is the distance that it takes the DNA molecule (on the left) to make precisely one complete turn.



The recognition helices of each polypeptide of CAP are, of course, identical. But their orientation in the dimer is such that the sequence of bases they recognize must run in the opposite direction for each recognition helix to bind properly. This arrangement of two identical sequences of base pairs running in opposite directions is called an inverted repeat.



The strategy illustrated by CAP and its binding site has turned out to be used widely. As more and more DNA-regulating proteins have been discovered, many turn out to share the traits we find in CAP:

- They usually contain two subunits. Therefore, they are dimers.
- They recognize and bind to DNA sequences with inverted repeats.
- In bacteria, recognition and binding to a particular sequence of DNA is accomplished by a segment of alpha helix. Hence these proteins are often described as helix-turn-helix proteins. The Trp repressor shown above is a member of this group.

Riboswitches

Protein repressors and corepressors are not the only way in which bacteria control gene transcription. It turns out that the regulation of the level of certain metabolites can also be controlled by riboswitches. A riboswitch is section of the 5'-untranslated region (5'-UTR) in a molecule of messenger RNA (mRNA) which has a specific binding site for the metabolite (or a close relative).

Some of the metabolites that bind to riboswitches:

- the purines adenine and guanine
- the amino acids glycine and lysine
- flavin mononucleotide (the prosthetic group of NADH dehydrogenase)
- S-adenosyl methionine (that donates methyl groups to many molecules, including
 - DNA
 - the cap at the 5' end of messenger RNA [Link]

In each case, the riboswitch regulates transcription of genes involved in the metabolism of that molecule. The metabolite binds to the growing mRNA and induces an allosteric change that

- for some genes causes further synthesis of the mRNA to terminate before forming a functional product and
- for other genes, enhances completion of synthesis of the mRNA.
- In both cases, one result is to control the level of that metabolite (a kind of feedback inhibition).

Some riboswitches control mRNA translation rather than its transcription. [Link]

It has been suggested that these regulatory mechanisms, which do not involve any protein, are a relict from an "RNA world".

Reference

http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/L/LacOperon.html#The_lac_operon