Gene Regulation -- The Lac Operon

While the mechanisms of transcription and translation are still fresh in our minds, let's take a look at another very significant problem: specifically, how is gene expression regulated? We know that some genes are active only at certain times during development and that different genes are expressed in different tissues. We also know that almost every cell of a higher organism must have essentially the same set of genes. This was proven first with plants where a single differentiated phloem cell of a carrot (and now many cell types from many other plants) could be induced in culture to regenerate a whole new carrot plant. More recently, cloning of whole animals has demonstrated the same thing; differentiation into specific cell and tissue types does discard un-needed genes. Starting with Dolly, a sheep created when a nucleus from a mammary gland cell was injected into an enucleated egg, several species of mammals have been cloned. Examples include a bull Second Chance
www.reporternews.com/1999/texas/clone0902.html and though not yet successful, the "Missyplicity" project www.missyplicity.com/ to clone a pet dog, both of which are projects of the TAMU Vet. School. Regeneration of genetically identical copies, (at least for genes in the nucleus of the donor) clearly shows that all the genes are still present in mature cells, even if they are not transcribed in all tissues.

Although it is easy to demonstrate differential gene activity in higher organisms, it is much simpler to study the molecular events in gene regulation in simpler organisms. The main effort of this lecture is to explore an extremely well described system of gene regulation in the bacterium E. coli,. The system provides a foundation for models that explain regulation in more complicated systems.

The model we will examine is called the Lac Operon, and it has to do with the ability of E. coli to utilize the sugar lactose. Lactose is a 12 Carbon sugar made of 2 simpler 6 Carbon sugars, glucose and galactose. As you likely know, glucose is a very efficient carbon source; it can enter directly into the metabolic paths that provide both energy and substrates for making more complex compounds. If lactose is provided as the carbon source, it must first be broken down into the two component sugars before it can be used.
The enzyme for breaking down lactose in *E. coli* is called β-galactosidase. The following observations demonstrated that the gene that codes for β-galactosidase in *E. coli* is regulated:

*E. coli* grown in glucose as the sole carbon source have about 3 copies of the enzyme β-galactosidase/cell.

*E. coli* grown in lactose as the sole carbon source have about 3,000 copies of the enzyme β-galactosidase/cell.

Thus, there are a thousand times more copies of the enzyme when it is needed than when it is not.

The system of regulation seen here is called "induction" since synthesis of the enzyme is "turned on" only when needed. Induction typically is used to regulate "breakdown" pathways as opposed to "synthetic" pathways.

Two Frenchmen, Francios Jacob and Jaque Monod won a Nobel prize for their work in describing how the lac-operon functions. They used a genetic approach to address the problem, by identifying mutants that did not have normal regulation of β-galactosidase. We will first look at the model they derived, and then see how the behavior of mutants led to the model.

The lac-operon is actually a series of adjacent genes and regulatory elements in one small part of the *E. coli* circular chromosome.

**Lac-Operon components**

![Lac-Operon components diagram]

Definitions: P strands for **promoter**; it is the site where RNA polymerase attaches in order to transcribe mRNA.

Although all promoters have the same function and share similar sequences that are recognized by RNA polymerase, they differ enough so that some are very strong (leading to high levels of transcription) and others are weak (rarely transcribed). Thus, one level of regulating gene expression comes as a consequence of the strength of the promoter at the beginning of the gene.
The \textit{I} gene is called a \textbf{regulator gene}; it is transcribed to make a mRNA which is translated to a \textbf{repressor protein}. There is a termination signal at the end of the \textit{I} gene.

\textit{O} stands for \textbf{Operator}; it is a short sequence of bases that acts like a switch that can be recognized by repressor protein.

\textit{Z}, \textit{Y} and \textit{A} are all "structural genes (genes that code for polypeptides)

\textit{Z} codes for \textit{β}-galactosidase; \textit{Y} codes for lactose permease, a protein that functions to actively bring lactose from outside to cell to the inside, even against a concentration gradient. \textit{A} codes for transacetylase, an enzyme that is also needed to breakdown many sugars related to lactose.

One long mRNA is made for the \textit{Z}, \textit{Y} and \textit{A} genes; this is the basis for the system being called an operon. All 3 genes that code for enzymes needed to use \textit{β}-galactoside molecules as a source of carbon and energy are adjacent and are coordinately turned on or off by regulating transcription. Operons are only found in prokaryotes; in eukaryotes, each structural gene has its own promoter and regulatory elements.

Let's first examine how the lac-operon functions when only glucose is present; that is when we expect it to be turned off:

\textbf{Stepwise:} 1. The Promoter for the \textit{I} gene is always "on", but is very weak, so it is transcribed only rarely. A gene that is not regulated,
other than by the strength of its promoter, is said to be "constitutive".

2. The I mRNA is translated into the repressor protein. A typical cell will have only about 10 copies of this protein.

3. In the absence of lactose, the repressor protein binds to the Operator, preventing transcription form the second promoter. Almost no ZYA mRNA is made.

When only lactose is present the model works as follows:

Stepwise: 1. The Promoter for the I gene occasionally is bound by an RNA polymerase to initiate transcription.

2. The I mRNA is translated into the repressor protein; a typical cell will have only about 10 copies of this protein.

3. Lactose (actually one stereo-isomer called allolactose) binds to the repressor very efficiently and converts the repressor into an inactive state, where it can't bind the Operator. The process is reversed when all the lactose is digested, so the system again will turn off.

4. When the very strong Promoter for making Z-Y-A mRNA is not blocked, many copies of the mRNA are made. The small amount of lactose that diffuses in is able to initiate induction.
of transcription of the Z-Y-A mRNA. Even as the message is being made, translation begins and the 3 proteins are made.

5. Translation begins at the 5' end of the mRNA and makes β-galactosidase from the Z gene. There is a stop codon, followed immediately by another AUG start, so many, but not all, ribosomes read on through and make permease from the Y gene. The same process allows some A gene product to also be made.

This whole system of regulating gene expression by regulating transcription is especially effective because messages are the most fragile part of the system. The average half-life of a mRNA in *E. coli* is about 1.8 minutes. Messages in eukaryotic tissues have a longer "life expectancy", but on the average are still much more fragile than proteins. It should be pointed out though that even proteins "turn-over", that is they wear out or become nonfunctional and are subject to degradation.

(continued on the next page)
Mutations that define the Lac-Operon model:

Jacob and Monod found mutants that did not show the normal regulation, that is, synthesis of Z, Y, and A proteins only when lactose was present.

Two kinds of mutants gave constitutive (continuous) synthesis of β-galactosidase, permease and TAase, no matter what carbon source was present

a) Mutants in the I gene (i^- or i^c) mutants all were mapped (we will take up mapping later in the course) to a similar location. Special tricks were used to make cells that were partial diploids with two copies of the lac-operon. When the copies created heterozygous cells

(I^+, I^-) normal regulation was observed. This indicated that the I gene codes for something that can move and interact with the operators of both copies of the lac-operon present in these cells.

b) Mutants in the promoter that change the base sequence so that it is no longer recognized by the repressor protein are also constitutive for Z, Y and A expression. In this case however, the mutation is dominant in partial diploids:
Since the operator in the upper strand cannot be bound by active repressor, β-galactosidase and TAase will always be made. In this cell, permease will show normal regulation since it is only made by the lower copy of the lac-operon. These results told Jacob and Monod that the Operator regulated transcription only of gene on the same DNA molecule.

From this simple model it is easy to see how a cascade of events could lead to complex regulatory schemes; the product of one reaction could induce the next pathway etc. until development is accomplished.

Analogous systems exist in bacteria for regulation of biosynthetic pathways. For example, if we look at the pathway for the synthesis of the amino acid histidine, several levels of regulation can be seen. Since *E. coli* spend part of their life cycle in a colon, they may well have plenty of histidine available from digested proteins. On the other hand, if no histidine is available they can make it "from scratch". The pathway for histidine biosynthesis starts with ATP+ PRPP and after 11 enzyme catalyzed reaction histidine is made. Given that a bacterium that wastes the resources and energy used to make histidine when it is not needed
would not compete well in nature, it is not surprising that several levels of regulation has evolved. One very common mechanism for biosynthetic (anabolic) pathways that is used for fine-tuning the level of the product is called "end-product inhibition". In end-product inhibition, the product of the pathway binds to the first enzyme unique to the pathway to inactivate it, thus slowing the rate of synthesis. As is the case with lactose interaction with the lac-repressor, histidine does not bind to the active site of the enzyme but to a different (allosteric site) which causes the protein to assume an inactive configuration. The reaction is also reversible, so that when the endproduct concentration becomes low, the enzyme will regain its active conformation.

However, if there is really an excess of histidine, it would also be costly to make 11 unneeded enzymes only to shut off the first enzyme. In this case, a system much like the lac-operon takes over. In his case, since the function is to turn off genes when a compound is present, the model is an example of repression.

\[
\begin{array}{ccccccccccccc}
P & \text{Reg} & \text{P} & \text{O} & \text{E1} & \text{E2} & \text{E3} & \text{E4} & \text{E5} & \text{E6} & \text{E7} & \text{E8} & \text{E9} & \text{E10} & \text{E11} \\
\end{array}
\]

In the case of the histidine operon, the regulator gene product is called a co-repressor. By itself, it cannot bind to the Operator unless it is first bound to histidine. Thus, a relatively subtle difference accounts for regulation in the inducible lac-operon and the repressible his-operon.

Many other variations on the same theme are known. For example, if both glucose and lactose are present cells use up the glucose before turning on the lac-operon. When energy begins to become limiting, a signal molecule (cAMP) builds up, binds to a CAP protein, which in turn binds to a site between I and P₂ of the lac operon, which opens the promoter for RNA polymerase binding.