

The Regulation of Transcription ("Gene Regulation") in Prokaryotes¹

Introduction: In this section we will investigate a couple of examples of how the production of proteins is turned on or off. Obviously, this is very important. For instance, it would hardly make sense for a cell to synthesize proteins that it doesn't need.

Give two reasons why this would be wasteful for the cell.

Moreover, we know that in animals and plants, cells all contain the same DNA and therefore the same genetic information, yet they are **differentiated** one from another structurally and functionally. We have learned that this **differentiation** is the result of the presence of:

- Differences in types of proteins (whether they be structural, enzymatic, or whatever in function) found in different cells. This does not mean that each cell type has entirely different proteins present. Most proteins are the same in different cell types; it is differences in a relatively few types that result in differentiation.
- Differences in the quantity of these proteins in different cells

Give several examples of proteins or pathways that would be expected to be found in all cells. Use proteins and pathways we have already considered in this course.

These differences are largely achieved by a process that is termed **gene regulation** or **gene expression**. What this really amounts to is regulating the rate of transcription of various structural genes. Thus, **regulation of transcription** is pretty much a synonym for gene regulation. Cells that are differentiated with respect to each other are different, not because of genetic differences, but instead because the rates of transcription of some genes are different in different cells.

Why should the rate of the transcription result in structural/functional differences in cells?

When we discuss the "rate of transcription" what we really mean is everything from a rate of "zero" where the genes are never transcribed (*i.e.*, they are turned off) to those that might be transcribed more or less continuously -- at least for a short period of time. Most examples fall somewhere in between.

¹ © 2006 by K.N. Prestwich, Dept. Biology, College of the Holy Cross, Worcester, MA 01610

Cells cannot only can differ with respect to each other -- they can also differ from themselves over time. This will be the case with the two examples we will study in these notes. Both of our examples will have to do with the control of the transcription of enzymes needed for metabolic processes. In one case, the enzymes have to do with obtaining and metabolizing a disaccharide called lactose (this is a common sugar in milk, it is composed of the hexoses glucose and galactose). The other has to do with synthesis of the non-essential amino acid tryptophan.

What is a non-essential amino acid? An essential amino acid? Are non-essential amino acids not required in proteins?

In the organism we are going to study, the common intestinal bacteria *E. coli*, lactose is not frequently available. So, the bacterium does not want to make proteins required for its metabolism when it probably won't need them. However, if lactose appears, the bacterium gains an advantage if it is able to rapidly produce the needed enzymes. On the other hand, if the lactose disappears, the *E. coli* needs to be able to stop the production of these same enzymes. In cases like this, we term the enzymes that appear in response to some specific environmental event but that are lacking at all other times **inducible enzymes**.

What is the specific environmental event in this case?

By contrast, other enzymes are often needed most of the time. Normally the genes that code for these enzymes are constantly being transcribed and the resulting mRNA translated at a certain rate. This allows for the replacement of proteins are constantly being broken down.

Why are proteins constantly being broken down? What is the advantage to this?

Nevertheless, there may be times when these enzymes are not needed. For instance, *E. coli* normally makes much of the tryptophan it needs for various proteins it must synthesize. However, if the *E. coli* finds itself in an environment that contains tryptophan, there is no need to synthesize more and no reason to keep the tryptophan synthesis enzymes on hand. So, we need a mechanism to stop protein the synthesis of tryptophan enzymes. Enzymes that are normally present but whose production can be decreased or stopped, such as those associated with tryptophan production, are called **repressible enzymes**.

What Cells Use Gene Regulation Processes? Any kind of cell can vary over time with respect to proteins present within it. In bacteria (prokaryotes), this type of regulation is the main way cells within a species differ. After all, there is no true cellular differentiation in unicellular organisms such as these. We will use bacterial examples because they are relatively easy to understand and there is a certain beauty about the systems they use in regulation. However, there are

major differences between pro- and eukaryotic cells that you should be aware of with regard to both transcription and translation. Study the next box carefully:

Some Important Differences Between Prokaryotic and Eukaryotic Transcription and Translation

We will investigate regulation in prokaryotes since it is far simpler than what happens in eukaryotes. However, keep in mind the following differences:

- Prokaryotes do not have introns and no processing is done to the mRNA molecule
- Prokaryotes often produce one mRNA molecule (1 transcript) that contains the information from several structural genes. We say that such a transcript is **polycistronic** where the "cistron" refers to the information for one protein. By contrast, eukaryotic transcripts contain only the information needed to synthesize one kind of protein.
- Since there is no nucleus in prokaryotes, translation can start immediately as the mRNA is being formed.
- A given mRNA molecule is often simultaneously translated by several ribosomes.

In most ways these details are not important for understanding gene regulation. However, some knowledge of them is quite useful to avoid confusion.

Moreover, we will also see that there are major differences in the way that regulation is achieved in most prokaryotes compared to eukaryotes. We will take these differences up at the end of these notes.

Setting the Stage: Remember from the last class that:

- Messenger RNA molecules can be translated few or many times but no mRNA molecule lasts forever. This is especially true in prokaryotes where typical life spans of mRNA molecules are measured in minutes (by contrast, in animals mRNA molecules often persist for several days). Messenger RNA molecules are hydrolyzed by nucleases. The resulting parts (NMPs) are "recharged" with high-energy phosphates and reused in the making of more RNA or used as a energy sources such as GTP or ATP (remember -- there is no difference between ATP used to drive active transport and ATP used in RNA synthesis).
- Proteins do not last forever. Proteases hydrolyze them at various rates and the amino acids are usually re-used in synthesis of new protein molecules. Thus, a given amino acid may, over its time in a cell, be a member of a number of different proteins.

Study Questions:

With all of this RNA and protein degradation going on, what must also occur at the same rate just to keep the cell looking and functioning the same over time?

What are the possible advantages of all of this "re-sculpting" of the cell?

What happens to someone who is in excellent aerobic shape and who then quits working out with respect to their capacity for aerobic metabolism? What happens on the molecular level?

Construct a "water tank" model of enzyme activity where the tank's water level is analogous to the amount of a particular enzyme. Water flowing in represents the addition of new enzyme and water flowing out represents the loss of enzyme molecules. What cellular processes specifically go with the inputs and outputs to the tank?

Organization of Genetic Information

We have learned earlier that genes are of two general types: structural and regulatory. Let's consider the regulatory ones in a bit more detail. We can consider them as belonging to two general groups:

- Regulatory sequences -- sections of a DNA molecule that code for nothing but which serve as attachment points for specific proteins and
- Regulatory structural genes – genes that code for proteins whose only function is in the regulation of gene expression.

Regulatory Sequences contain information that is used to control transcription. Most commonly, they serve as places where proteins bind to DNA and thereby start transcription or stop it. Today we will consider two types of bacterial regulatory sequences:

- **The promoter**: a sequence **recognized by RNA polymerase** as a place to attach. After it attaches, RNA polymerase will begin to read the DNA molecule in a 3 to 5' direction. It will continue this movement unless it is physically blocked by a protein attached to the DNA (see below) or until it encounters a specific **termination sequence** that causes the RNA polymerase to drop off the DNA molecule.
- **The operator**: a sequence immediately downstream (in the 5' direction) from the promoter and just before the structural genes. This sequence is recognized by a specific type of protein called a **repressor protein**. Notice that when the repressor protein is bound to the operator, RNA polymerase cannot transcribe the structural genes that follow -- it blocks the 3' to 5' movement of the polymerase. Thus, operator sites represent regions of "negative control" over gene expression.
- **Positive Regulation Sites**: These are **upstream (in the 3' direction) from the promoter**. When various signals bind to these sites, the resulting molecular complex makes it easier for RNA polymerase to bind to the promoter region. Thus, we say they are regions that exercise "positive control" over gene expression.

All of these regulatory sequences are named according to the structural genes they control or sometimes for the compound they bind. Thus, today we will talk about a **lac** (for lactose) **promoter** and **lac operator** and a **trp** (for tryptophan) **operator** and **trp promoter**. In a moment we will add a bit of detail about these two types of sequences.

Regulatory structural genes, or, unfortunately, "regulatory genes", are sequences that code for proteins. I prefer to think of these as no different from any other structural genes except that in their case, the protein products have a regulatory role. For the purposes of our discussion today, we will look at two types of regulatory genes:

- **Genes for repressor proteins.** These genes code for proteins that can attach to the operator sequences mentioned above. For our examples today we will consider two such genes -- one for a repressor specific for lac operator (see previous material in these notes) called the **lac repressor** and the other specific for the trp operator called the **trp repressor**. These are two rather different proteins -- don't think of them as being the same -- they just perform similar functions.
- **The gene for RNA polymerase** -- some would not count this one as a regulatory gene even though its product is a central player. This just goes to show how artificial these classifications are!

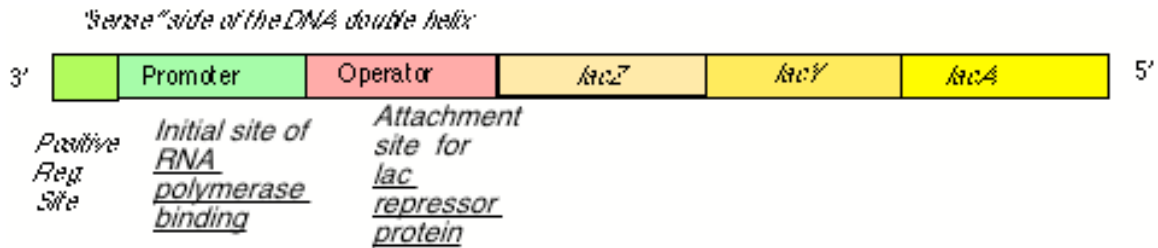
The relationships between regulatory and structural genes

Structural genes with related functions and the regulatory sequences that help to control their expression (control their transcription) are grouped together. This is especially evident in prokaryotes. These functional groups of regulatory sequences and related structural genes are called **operons**. To make this concrete, let's look at the first operon to be described -- the **lac operon**. First described in 1961 by the great François Jacob and Jacques Monod, the operon consists of three regulatory sequences and three structural genes. The structural genes and the functions of their products are:

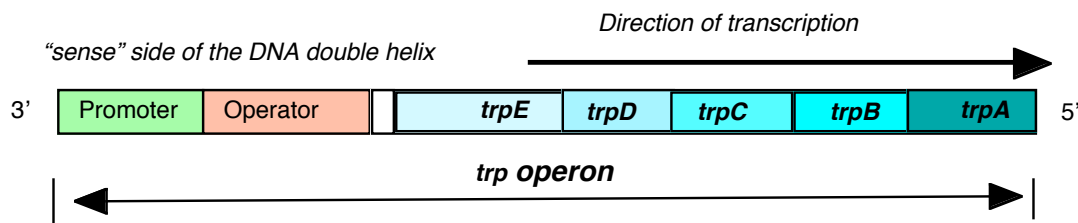
- *lacZ* -- codes for β -galactosidase, a protein that split lactose into glucose and galactose
- *lacY* -- permease -- a protein that increases the permeability of the plasma membrane to lactose
- *lacA* -- transacetylase -- a protein with unknown function in lactose metabolism or related processes.

DO NOT learn the names of these proteins or structural genes unless you want to; just understand the fact that two of them are very important in lactose metabolism and will be induced together.

The regulatory sequences are located immediately upstream (3' side) from these structural genes. So, the whole **lac operon** looks like this:



For comparisons sake, let's also see how the ***trp* operon** is set up. There are 5 structural genes in this case, each coding for a different enzyme used in the biochemical pathway that synthesizes tryptophan. These genes are named *trp A* through *trp E*. As with the *lac* operon there are also promoter and operator regulatory sequences immediately upstream from the structural genes:



Note: please don't get the idea that the regulatory sequences are as long as the structural genes. for they are not!. Also don't get the idea that the *trp* operon has the same total DNA length as they *lac* operon-- the *trp* operon is far longer. These are schematic drawings.

Operation of the Repressor Proteins: The first thing to remember about regulatory proteins, called **repressors** in the *trp* and *lac* operons, is that they are always present in the cell. We will see shortly that if these proteins are absent, as they are in some mutant strains of *E. coli*, then the regulation schemes won't work. Think of these repressor proteins as being **coded for by structural genes that are transcribed and translated at a low, constant rate**. Very little of these proteins is needed in a cell. Unlike a typical enzyme where a great many copies of the protein are needed to catalyze a reaction at an appreciable rate, cells will only require a very small number of repressors of each type.

The second thing we need to know about repressor proteins is that they have **binding sites for**

- **Their specific operator AND for**
- **A specific "signal molecule".**

The exact details of how the system will work depends on the system. So, let's look at the *lac* repressor protein and operon first.

Negative Control by the *lac* Repressor:

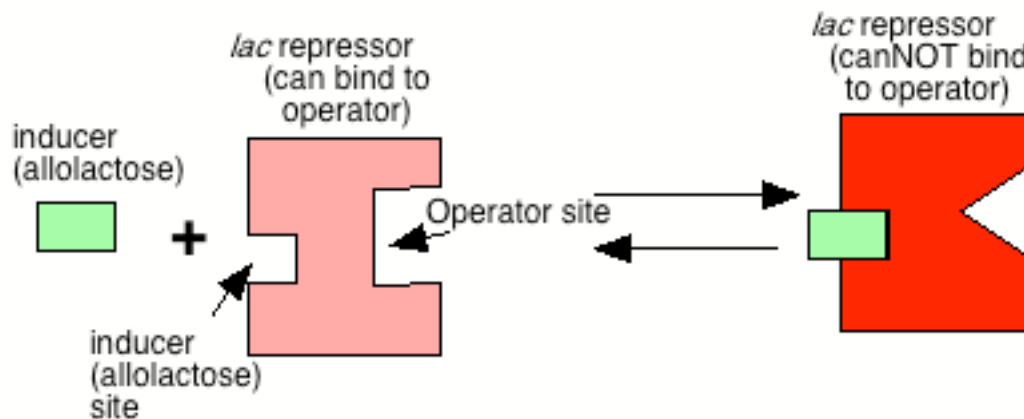
The *lac* repressor protein **can bind EITHER** to:

- A signal substance called **ALLOLACTOSE** (termed the **INDUCER**) or to
- **The operator sequence** but
- **Not both at the same time.**
- Allolactose, the inducer, is only present in cells when lactose is present.

So, if there is lactose present, allolactose is also present and it binds with the *lac* repressor proteins. **This causes an allosteric change in the molecule that prevents it from binding to the operator.**

On the other hand, if allolactose is not present (for instance, after all the lactose in a cell has been used up or when there has been none for a long time), the *lac* repressor protein does not have this inducer attached. As a result the *lac* repressor assumes a shape that allows it to bind to the operator and it does.

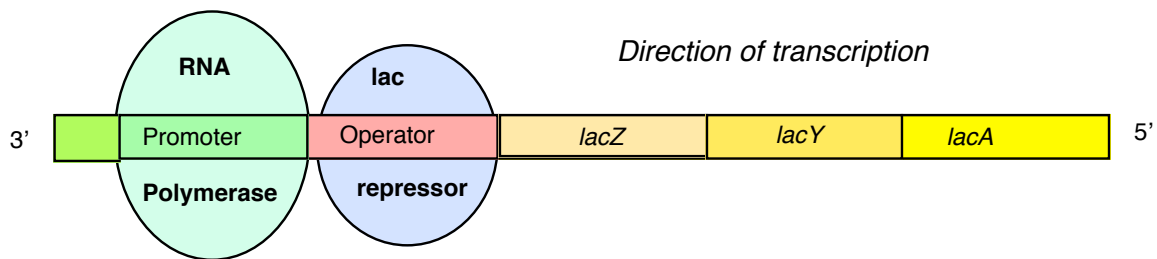
What kind of binding is there between the *lac* repressor and the inducer? The operator?
What happens if the *lac* repressor is bound to the operator and the allolactose concentration increases?
Where is the operator relative to the promoter (upstream or downstream)?
What happens to the *lac* repressor if it is bound to the inducer and the cell uses up all the free (unbound) inducer via catabolic metabolism?



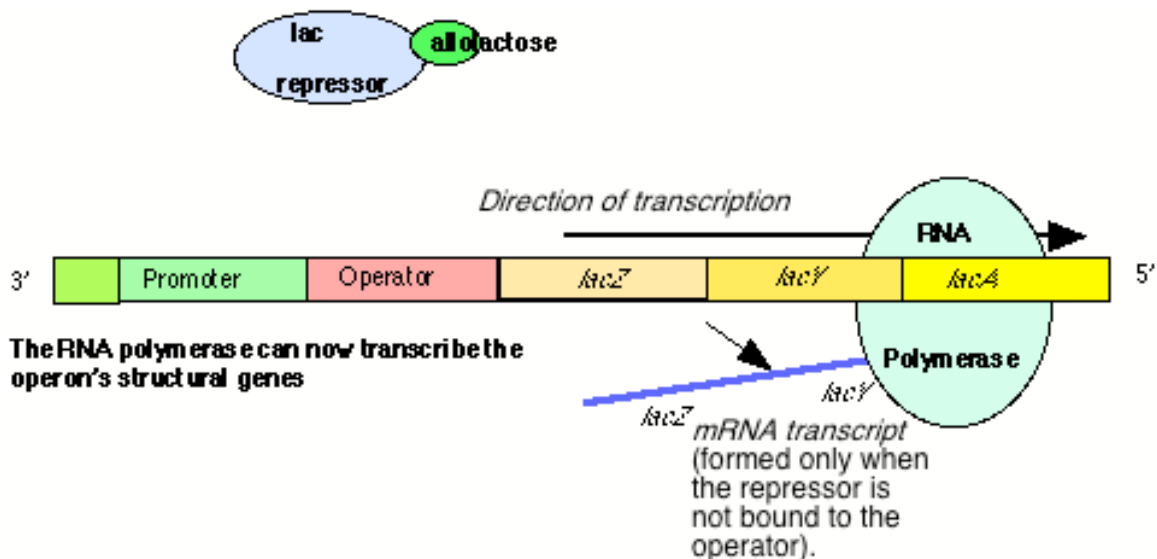
Now let's discuss the promoter and then we'll be ready to put all of this together. **RNA polymerase can bind to the promoter. Once it does, it will begin reading the DNA in the typical 3 to 5' direction until it encounters a roadblock or stop sign.** The secret of operon regulation will be the presence and absence of roadblocks.

Negative Control at the *lac* Operon:

Recall that normally there is no lactose and therefore no allolactose present in *E. coli*. As a result, the small number of *lac* repressor proteins present in the cell are in a conformation that allows any one of them to bind to the *lac* operator. They are only loosely bound to the *lac* operator, but whenever one comes off, another will soon replace it. For all intents, the *lac* operator is constantly occupied by the *lac* repressor. Now, what if DNA polymerase binds to the promoter and starts to move in the 5' direction? Notice that almost immediately it runs into the *lac* repressor. This prevents any transcription from occurring:



On the other hand, if lactose is present, the inducer allolactose will also be present. The *lac* repressor proteins will bind the inducer, undergo an allosteric change, and be unable to bind to the operator. The next time a molecule of RNA polymerase comes along and attaches to the promoter, it will be able to move along and transcribe the operon – there is no roadblock at the *lac* operator:



Look back at this process and explain why the compound allolactose is an **inducer**. What will be the effect of this process on the ability of the cell to pick up and metabolize lactose?

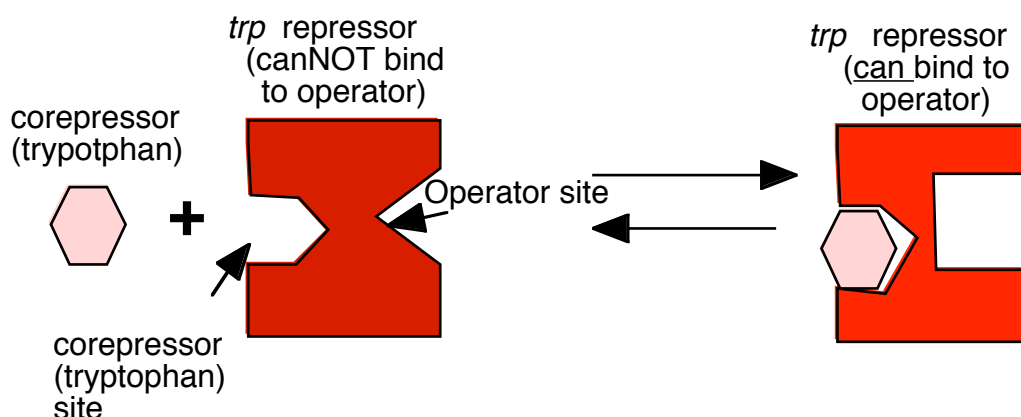
If lactose there is only a limited amount of lactose, what will eventually happen to all of the lactose in the cell? What happens to transcription at this point? Explain. What will be the fate of the mRNA and enzymes, over time? Explain why this process is useful to the cell.

Negative Control at the *trp* Operon:

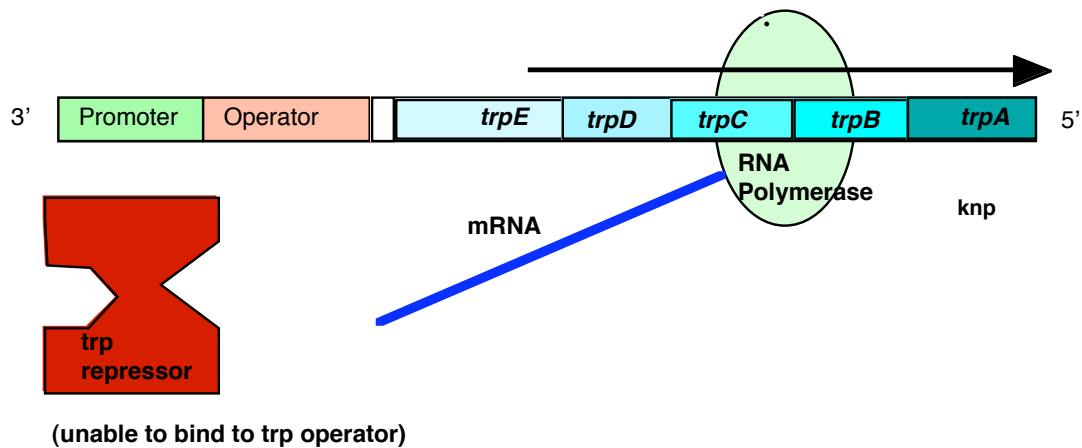
The *trp* repressor protein works in what might be described as the opposite manner to that of the *lac* operon. The signal molecule in this case is tryptophan itself. The signal is called a **co-repressor** because its presence at high concentrations will prevent transcription of the *trp* operon. Compare this with the role of allolactose in the *lac* operon -- when allolactose was added, it caused transcription to begin. By contrast, high levels of tryptophan will shut off transcription. Tryptophan is called a **co-repressor** because it works with the *trp* repressor protein to do this – high levels of tryptophan without the *trp* repressor protein will not shut off expression of the *trp* operon.

Here's how the *trp* operon works.

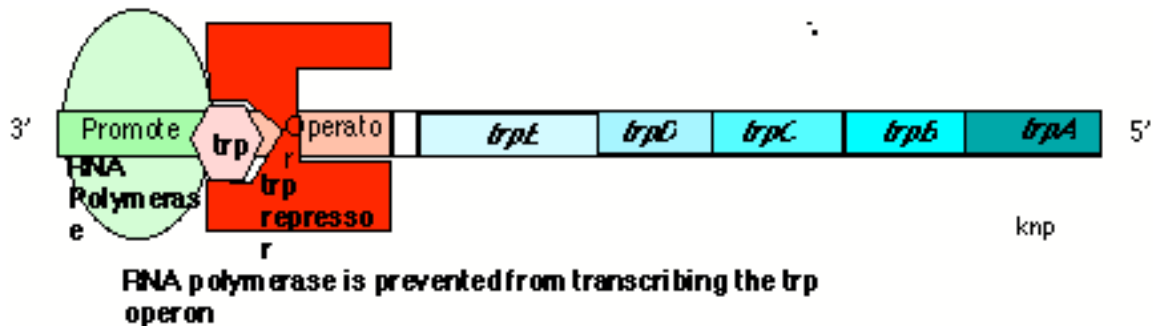
- Assume that the repressor protein (*trp* repressor) has binding sites for both the *trp* operator sequence and the co-repressor (tryptophan).
- If there is too little signal (note that there will always be some tryptophan present in the cell since tryptophan is needed in protein synthesis), the *trp* repressor does not bind the co-repressor.
- The result is that the *trp* repressor protein **cannot bind to the *trp* operon**.
- On the other hand, **if there is TOO MUCH SIGNAL** (too much tryptophan), such as would occur when the *E. coli* was absorbing tryptophan from the environment, **tryptophan binds to the *trp* repressor protein and changes its shape such that it can bind to the *trp* operon**. This blocks transcription:



And so, when low levels of tryptophan are present, the *trp* repressor does not bind to the operator and therefore RNA polymerase can transcribe the *trp* operon:



and if tryptophan is present *at sufficiently high concentrations*, transcription is stopped (the *trp* operon is repressed):



Be sure you understand both the similarities and differences between these two methods of control. Be sure you know why tryptophan is called a co-repressor and why the trp repressor protein only binds to it when it is in high concentrations.

Positive Regulation

The two previous examples are termed negative regulation. What is common to both of them is that the binding of a protein to the operator **stops** transcription and protein production. In negative regulation, the binding site that blocks RNA polymerase (the operator) is always located on the 5' (downstream) side of the promoter (so it can block the movement). On the other hand, in positive regulation, the regulatory sequence(s) is/are located on the 3' (upstream) end of the promoter. Thus they do not block the movement of RNA polymerase. On the other hand, they form complexes that enhance the ability of RNA polymerase to attach to the promoter.

How should you think about this? Imagine that the promoter/RNA polymerase "match" is not perfect. Thus, RNA polymerase will bind to the promoter but not at a high rate. With positive regulation, the attachment of

"inducer" or "activator" molecules to the DNA upstream from the promoter helps the RNA polymerase to attach and increases the transcription rate.

We will briefly consider how this works in eukaryotes in class.

Study Questions:

1. There is only one type of RNA polymerase. Yet it can bind to any promoter. What does that tell you about the nucleotide sequence in promoters?
2. The rate at which RNA polymerase can attach and bind to different promoters (and therefore the rate at which operons can be transcribed) is different in different cases. What could cause this?
3. What would be the effect of a mutation that prevented caused a non-functional version of the *lac* repressor (or no *lac* repressor at all) to be produced? What about for the *trp* operon?
4. In the case of the *trp* operon, is its regulation part of a negative or positive feedback process (see metabolic control notes) or neither? Explain.

From the reading:

5. The two regulatory processes we looked at in the notes are both called **negative regulation**. What is positive regulation? Simply define it; you need not describe the process that is in the book.
6. What are DNA methylation and histone methylation used to accomplish?
7. What are enhancers in eukaryotic regulation? Are control sequences immediately connected (right next to) the regulated structural gene in eukaryotes?