Control of Metabolic Processes

As described earlier, the metabolic processes occurring within living organisms (glycolysis, respiration, photophosphorylation, etc.) are dependent upon the enzymes present within cells, and these are determined by the genes present, or the information carried within DNA molecules. Whether or not a specific cell is using one or another metabolic process is determined by regulatory mechanisms functioning at various levels. In this section, two mechanisms for controlling metabolism will be described, one functioning at the enzyme level, and the other functioning at the gene level.

Feedback inhibition:

Enzymes are proteins (ribozymes are RNA molecules), so are composed of multiple amino acids connected together in long chains. They typically have tertiary structure (are 3-dimensional), and many are quaternary (are composed of multiple polypeptide chains). As described earlier, enzymes are specific in their action, and bind with their substrate or reactant molecules through regions on their surfaces called reactive sites or binding sites. Enzymes can also be inhibited, i.e., the catalytic activity of enzymes can be blocked through either competitive or allosteric inhibition.

Feedback inhibition is a regulatory mechanism involving the allosteric inhibition of one or more enzymes (usually the first or second) involved a common metabolic pathway. The inhibitor is usually the end-product of the pathway, so this mechanism can also be called end-product inhibition. In bacteria the biosynthesis of isoleucine (an amino acid) involves threonine (also an amino acid) as a substrate and a metabolic pathway with five steps catalyzed by five different enzymes. The enzymes involved are represented by the letters A-E, and the metabolic intermediates by the letters W-Z, in the diagram below.

```
Threonine  A (W)  B (X)  C (Y)  D (Z)  E  Isoleucine
       ↑  |  |  |  |  
       Allosteric inhibition of enzyme A
```

When isoleucine begins to accumulate within the cytoplasm, it acts as an allosteric inhibitor of the first enzyme in the pathway (enzyme A), and effectively shuts down isoleucine synthesis. Because the end-product of the pathway acts to reduce its own production, this is an example of negative feedback. Feedback inhibition or end-product inhibition occurs within both eukaryotic and prokaryotic cells and allows organisms to control metabolic processes relatively quickly (exerts rapid control). This mechanism is also reversible, because when the concentration of end-product is decreased, the first enzyme is no longer inhibited. Though used extensively, this mechanism is not efficient in terms of energy conservation, because when metabolic pathways are inhibited, enzymes are inactive. This means the cell had to expend considerable energy in the formation of mRNA molecules and polypeptides to establish a metabolic pathway no longer being used. A more efficient means of controlling metabolism can be exerted at the gene level.

In *Saccharomyces cerevisiae*, parts of the pathway used for isoleucine synthesis are also used for the synthesis of leucine and valine. The first enzyme in the pathway, threonine deaminase, is down regulated (inhibited) by both leucine and isoleucine, but upregulated by valine. The second enzyme in the pathway is down regulated by all three amino acids.
Genetic control:

In prokaryotic cells, the genes encoding enzymes involved in common metabolic pathways are often arranged together within specific regions of the chromosome called **operons**. An **operon** is a segment of DNA (a nucleotide sequence) containing a series of structural genes and the control elements regulating the transcription of those genes. The "control elements" typically include a **promoter site** or sequence and a region known as the **operator site** (an attenuator site may also be present, but will not be included here). Recall that promoter sites are nucleotide sequences recognized by the **sigma factors** of **RNA polymerase**. Promoters interacting with sigma factors determine where transcription will begin and in which direction it will proceed along DNA molecules. The operator site functions like an "on-off" switch, and is influenced by DNA-binding proteins called **repressors**. When a repressor protein is bound to the operator, transcription is blocked (repressed), but if the repressor is not bound, transcription is allowed to proceed.

There are many different metabolic pathways controlled by operons in different types of bacteria, but two such systems found within *E. coli* cells are commonly used as examples of genetic control mechanisms.

**Tryptophan biosynthesis – Control involving a repressible operon:**

The operon controlling **tryptophan biosynthesis** in *E. coli* is commonly used as an example of a **repressible operon**, i.e., one in which transcription is usually occurring, but can be repressed or "turned off". **Tryptophan** is an amino acid synthesized from glutamine and chorismic acid by means of a metabolic pathway involving five enzymes, as diagrammed below.

\[
\begin{align*}
\text{Chorismic acid} & \quad + \\
\text{Glutamine} & \quad \to \quad (W) \quad (X) \quad (Y) \quad (Z) \quad \to \quad \text{Tryptophan}
\end{align*}
\]

In this pathway, enzymes are designated by the letters "A-E", and metabolic intermediates formed within the pathway are designated by the letters "W-Z". The genes encoding the enzymes above are arranged together within the **tryptophan biosynthesis operon**. This operon includes a promoter site, an operator site and an attenuator site, but only the first two control elements will be described here.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Operator</th>
<th>Attenuator</th>
<th>Gene A</th>
<th>Gene B</th>
<th>Gene C</th>
<th>Gene D</th>
<th>Gene E</th>
</tr>
</thead>
</table>

The **promoter site** is the nucleotide sequence **sigma factor** binds with to start transcription. The **operator site** is the nucleotide sequence the **repressor protein** binds with to block or **repress transcription**. In this case, the gene encoding the repressor protein is not part of the operon, but is located in a different region of the chromosome. The repressor protein associated with the tryptophan biosynthesis operon is **inactive alone**, so cannot bind with the operator site. Under most circumstances this operon is "on", i.e., transcription is allowed to proceed. While the operon is active, transcription results in the formation of mRNA molecules that travel to ribosomes and are used in the production of enzymes A-E. The biosynthesis is then allowed to proceed and tryptophan levels within the cell increase.

Tryptophan serves as a **corepressor**, i.e., a compound that can bind with the inactive repressor protein and activate it (much like coenzymes and cofactors activate enzymes). The tryptophan-repressor complex can then bind with the operator site and block transcription of the genes present within the operon.
This mechanism is very efficient in terms of energy conservation because when tryptophan concentrations become high, the cell will not just inhibit the enzymes involved in tryptophan synthesis, it will stop making the m-RNA molecules necessary to form the polypeptides needed to run the tryptophan biosynthesis pathway.

**Lactose utilization – Control involving an inducible operon:**

Lactose utilization in *E. coli* is controlled by an **inducible operon**, i.e., one in which transcription is usually repressed or "off", but can be **induced** or "turned on". This operon (often referred to as the **lac operon**) includes three structural genes, a promoter and an operator. A regulatory gene (*lacI*) near the lac operon promoter site encodes a repressor protein that is **active alone**. The repressor protein is **constitutive**, so is always being made; consequently, under most circumstances the structural genes within the operon are not being transcribed. Note – When *E. coli* cells are living within the intestines of adult cows, there is no lactose available, so making enzymes to catabolize it would be a waste of energy. The *E. coli* cells conserve energy by repressing the transcription of the **lac** operon genes.

<table>
<thead>
<tr>
<th>Promoter</th>
<th>lacI</th>
<th>Operator</th>
<th>lacZ</th>
<th>lacY</th>
<th>lacA</th>
</tr>
</thead>
</table>

In this diagram, the *lacI* gene is the regulatory gene encoding the repressor protein, promoter sites are nucleotide sequences where sigma factors bind to begin transcription, the operator site is where the active repressor protein binds to block (repress) transcription and the structural genes *lacZ*, *lacY* and *lacA* encode enzymes associated with lactose utilization. Since the repressor is constitutive, and active alone, transcription of the three structural genes is usually being repressed, but not entirely. Why not?

The structural genes within the **lac** operon encode enzymes; two of these, *lacY* and *lacZ* encode enzymes directly involved in lactose utilization, i.e., **β-galactoside permease** or **lactose permease** (an enzyme allowing lactose to enter the cell) and **β-galactosidase** (an enzyme that breaks lactose into glucose and galactose), respectively. The third gene, *lacA*, encodes the enzyme **thiogalactoside transacetylase**. This enzyme catalyzes a chemical reaction converting lactose into **allolactose**. Allolactose is significant, because it serves as the **inducer** for the lac operon. When allolactose is abundant, it binds with the repressor protein and **inactivates** it, i.e., changes its configuration so it can no longer bind with the operator site. With the repressor removed, transcription is allowed to proceed.

At this point, one might well ask the following question. If the enzymes encoded by the genes of the **lac** operon are not being made, lactose cannot enter the cell and allolactose cannot be formed; therefore, how can this operon ever be induced? Apparently the operon is "off" like a leaky faucet, some transcription occurs even when the operon is being repressed. This is because although the repressor protein binds tightly to the operator site, it only **interferes** with RNA polymerase, and doesn't completely block transcription.

Arabinose utilization in *E. coli* is controlled by an operon that involves a regulatory protein (*AraC* protein) that binds to two sections of the operator site causing DNA loop formation. Since RNA polymerase cannot get past this loop, the binding of the *AraC* protein represses transcription of the arabinose utilization genes. In this operon, arabinose serves as an inducer because it can bind with the *AraC* protein and prevent it binding the operator site.

Although inducers (e.g., allolactose) partially control the transcription of **inducible genes**, some **inducible operons** can also be regulated by a mechanism called **catabolite repression**. This mechanism involves **cyclic-AMP** as a regulatory molecule.
Catabolite repression:

Catabolite repression is a mechanism allowing bacteria such as *E. coli* to utilize constitutive enzymes in favor of inducible ones. This mechanism adds a second layer of control and improves the efficiency of inducible operons, e.g., the “leaky” systems as described above.

A catabolite is any substance a cell can catabolize (break down) to release energy. Glucose is a common catabolite, and as described earlier can be broken down by various metabolic pathways. Since *E. coli* cells are facultatively anaerobic, they can use either fermentation or respiratory pathways to catabolize glucose. In either case, when glucose is available, ATP is being made, and energy is available to drive cellular processes. The flow of energy through *E. coli* cells is vastly simplified in the diagram presented below.

In this diagram, the covalent bonds within glucose molecules are being broken through catabolic processes (recall fermentation and cellular respiration), and the energy released is being used to convert $\text{AMP} + \text{P}_\text{~P}$ into ATP. Since ATP is the energy currency of the cell, it is constantly being broken down and the energy released is used to drive cellular processes, e.g., active transport, kinesin motion, synthesis reactions, etc. The catabolism of ATP yields $\text{AMP} + \text{P}_\text{~P}$. This diagram is not accurate, because as was explained earlier, ATP is formed from $\text{ADP} + \text{P}_\text{i}$ (adenosine diphosphate and inorganic phosphate); however, the diagram is accurate to this extent: When ATP levels increase, c-AMP levels decrease. When associated with glucose catabolism, this occurs because the transport of glucose into cells inhibits the activity of adenylate cyclase, the enzyme responsible for converting ATP into cyclic-AMP.

In its cyclic form, 3'-5'-cyclic adenosine monophosphate (cyclic-AMP), AMP serves as a regulatory molecule, i.e., as a "second messenger" or a molecule involved in signal transduction. Cyclic-AMP can bind with a protein called catabolite activating protein (CAP) also called cyclic-AMP receptor protein (CRP), to form a complex that can interact with DNA. The cyclic-AMP-CAP complex binds to DNA at a site very near the promoter site on the lac operon and makes it easier for the sigma factor of RNA polymerase to bind, i.e., it enhances the promotor site (makes it more attractive to sigma factor). When cyclic-AMP levels are high, the c-AMP-CAP complex is bound to DNA and the structural genes of the lac operon are transcribed (transcription is increased), assuming lactose is available and some of it has been converted into allolactose. When the lac operon promoter site is not "enhanced", sigma factor is only weakly attracted to it.

So, if *E. coli* cells are placed into a TSI slant, which of the three sugars present will they use first, and why? Although the lac operon is "leaky" in terms of the control exerted upon it by the repressor protein, very little transcription of the lac operon structural genes will occur as long as glucose is available to the *E. coli* cells. The enzymes involved in glucose catabolism are constitutive, and not under control of a repressor. The *E. coli* will use glucose first, and only when glucose is no longer available, will they use lactose.

New findings (1997 – today) show that prokaryotic cells can also exert control over DNA entering their cells using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and a group of CRISPR associated (Cas) proteins (DNA binding proteins) that form complexes around CRISPR RNA molecules (crRNA), and then “target” DNA regions complementary to the “guide” crRNA.
CRISPR/Cas

The CRISPR/Cas system is now recognized as an acquired or adaptive immune system within prokaryotic cells. The Clustered Regularly Interspaced Short Palindromic Repeats characteristic of the CRISPR/Cas system are not truly palindromes (regions of DNA with sequences reading the same forward and back on complementary strands), but are nucleotide sequences 24 – 48 bases long with recognizable repeat patterns. The spacer regions located between the repeats are sections of DNA of similar size taken from foreign agents (viruses or plasmids) that have previously enter the cells. Around 90% of the archaea genomes currently sequenced contain CRISPR/Cas systems with large, apparently stable collections of spacer regions. Around 40% of bacterial genomes carry CRISPR/Cas systems, but the spacer collections are smaller and apparently change over time, i.e., represent collections of DNA from viruses or plasmids only recently encountered. The CRISPR associated proteins (Cas) are encoded by genes often clustered near the CRISPR repeat/spacer arrays. There are multiple different types of proteins involved, and much interest in determining their specific functions. The primary functions of the Cas proteins are acquisition of foreign DNA sections, and destruction of foreign DNA (with the same nucleotide sequences) if they enter the cell a second or subsequent time.

There are currently two classes of CRISPR/Cas systems, with class I type using a cluster of proteins to degrade foreign nucleic acids, and class II using a single large protein for the same purpose. Within these broad categories there are multiple variations (systems) involving different specific proteins. Bacteria often carry more than one CRISPR/Cas system, so they are compatible, and may share components. Their sporadic distribution suggests they are readily acquired through horizontal gene transfer. Proteins identified as Cas1 and Cas2 are encoded by genes common to all recognized CRISPR/Cas systems of bacteria, and are required for acquisition of spacer DNA sequences (those cut from viruses or plasmids). A large protein called Cas9 can also cut DNA to add spacers, and is involved in cutting DNA targeted by guide crRNA molecules.

CRISPR Relates to Genetic Regulation in Humans

The CRISPR associated protein Cas9 has been found to have DNA cutting ability and will bind with specific regions of DNA targeted by crRNA. In type II CRISPR/Cas systems, Cas9 is the only protein needed to cut host DNA for spacer insertion, and is the protein responsible for gene silencing, i.e., will cut both strands of a DNA duplex. In eukaryotic cells, DNA repair mechanisms often insert or remove one or more bases (causing addition or deletion mutations) so the gene cut cannot be expressed (transcribed and translated) correctly. The Cas9 protein binds with crRNA, but also requires a separate trans-activating crRNA (transRNA) for activation. Genetic manipulation has been used to create a combination crRNA/transRNA molecule that the Cas9 protein will bind with and use to target specific locations on human DNA molecules.

Multiple variations of Cas9 have been developed, some that cut both DNA strands and some that cut only one. They can be used in combinations that allow eukaryotic DNA to be mutated, cut and spliced, “silenced”, or “tagged” for recognition (by binding fluorescent markers). CRISPR/Cas9 targeting and regulation of transcription can be used to turn genes “on” or “off” with precision. The CRISPR/Cas9 system can also be used to target a specific region of DNA and then recruit other proteins to that location with many potential applications possible.