Regulation of gene expression in prokaryotes and eukaryotes

Gene structure in prokaryotes. Bacterial operons: lac, ara, trp, gal. Gene structure in eukaryotes. Regulation of transcription: transcription factors.

1. Define the terms: operon, cistron, promoter.
2. Explain the functioning and regulation of the following operons: lac, ara, trp, gal.
3. Explain positive and negative controls of operons.
4. Differentiate between constitutive and inducible promoters.
5. Explain the mechanism of transcription regulation in eukaryotes.
6. Describe the structure of the promoter: TATA-box, GC-box.
7. Explain the functions of enhancers and silencers.
8. Describe the role of transcription factors and activators in the regulation of transcription
9. Describe the structure and significance of DNA-binding domains and transcription activation domains.
10. Compare translation regulation in pro- and eukaryotes

Genes determine everything about us, from the outward physical traits we can see to the behind the scenes structures inside our cells that allow them to carry out all of our body functions. Genes are transcribed into copies of themselves called RNA, which is then translated to protein. **Structural genes** are genes that code for proteins in the body needed for structure or function. Unlike structural genes, **regulatory genes** code for protein products that control other genes, instead of making structures of their own. Regulatory genes code for proteins that act like switches, turning other genes on or off. These genes are essential to controlling cell function, and without them, cells can grow out of control, causing diseases, like cancer, in the body.

Gene expression is the phenotypic manifestation of genes by the processes of transcription and translation. Gene expression via transcription and translation is a fundamental principle of molecular biology that is often referred to as the central dogma of molecular biology. Gene expression is the process by which the information encoded in a gene is used to direct the assembly of a protein molecule. The cell reads the sequence of the gene in groups of three bases. Each group of three bases (codon) corresponds to one of 20 different amino acids used to build the protein.

Gene expression in humans is complex and highly regulated. Regulation occurs at many points during the transcription and translation processes and involves epigenomic compounds, which are chemical compounds and proteins that can attach to DNA and influence gene expression.

The number of genes in an organism’s genome (the entire set of chromosomes) varies significantly between species. For example, whereas the genome of the bacterium *Escherichia coli* O157:H7 houses precisely 5,416 genes. *Arabidopsis thaliana*—the first plant for which a complete genomic sequence was recovered—has roughly 25,500 genes; its genome is one of the smallest known to plants. Among extant independently replicating organisms, the bacterium *Mycoplasma genitalium* has the fewest number of genes, just 517. In the human genome, there are a little less than 20,000 genes. In some cells, many genes are active--say, 10,000--and the other 10,000 would be inactive. In other kinds of cells, maybe the other 10,000 would be active and the first 10,000 would be inactive.

And so, gene regulation is the process by which the cell determines which genes will be active and which genes will not be active. And gene regulation is at the bottom of what makes a cell decide to become a red blood cell, or a neuron, or a hepatocyte in the liver, or a muscle cell. So different gene regulation will give you a different program of genes and different genes expressed.

There are several different kinds of gene regulation. Some genes, called housekeeping genes, are expressed in almost every cell. And these require a regulatory network or machinery that keeps them on in almost every cell, so these are the enzymes that help make DNA, and perform glycolysis, and burn sugar, and things like that. There are other genes that are called

tissue-specific genes. These are genes that, say, would only be expressed in a red blood cell or a neuron. Very often, these genes have transcription factors, which are proteins that bind to DNA, near these genes. And those transcription factors actually help the RNA machinery get there and transcribe that gene in those cells, and those tissues, transcription factors, rather, are expressed specifically in those tissues. There are also factors expressed in those tissues that will be suppressors that can turn a gene off. And then there are genes that are regulated during development. Sometimes they're expressed in fetal life and then turned off in adults, and sometimes it's vice versa. So there are very complex different ways that genes are regulated.

Gene regulation is the process of turning genes on or off. Gene regulation can occur at any point of the transcription-translation process but most often occurs at the transcription level.

Proteins that can be activated by other cells and signals from the environment are called transcription factors. Transcription factors bind to regulatory regions of the gene and increase or decrease the level of transcription. Other mechanisms of gene regulation include regulating the processing of RNA, the stability of mRNA and the rate of translation.

Turning the correct genes on and off is an essential component to maintaining a cell's functionality.

Gene regulation is how a cell controls which genes, out of the many genes in its genome, are "turned on" (expressed). Thanks to gene regulation, each cell type in your body has a different set of active genes – despite the fact that almost all the cells of your body contain the exact same DNA. These different patterns of gene expression cause your various cell types to have different sets of proteins, making each cell type uniquely specialized to do its job.

Differences in gene regulation makes the different cell types in a multicellular organism (such as yourself) unique in structure and function. Gene regulation can also help us explain some of the differences in form and function between different species with relatively similar gene sequences.

For instance, humans and chimpanzees have genomes that are about 98%, percent identical at the DNA level. The protein-coding sequences of some genes are different between humans and chimpanzees, contributing to the differences between the species. However, researchers also think that changes in gene regulation play a major role in making humans and chimps different from one another. For instance, some DNA regions that are present in the chimpanzee genome but missing in the human genome contain known gene-regulatory sequences that control when, where, or how strongly a gene is expressed.

Experiments have shown that many of the genes within the cells of organisms are inactive much or even all of the time. Thus, at any time, in both eukaryotes and prokaryotes, it seems that a gene can be switched on or off. The regulation of genes between eukaryotes and prokaryotes differs in important ways.

Gene expression can be regulated at any step: from transcriptional initiation, to RNA processing, to post-translational modification of the protein. The regulation of lactose metabolism genes in *E. coli* (*lac* operon) was the first such mechanism to be described in 1961.

Regulation of gene expression in prokaryotes

We tend to think of bacteria as simple. But even the simplest bacterium has a complex task when it comes to gene regulation! The bacteria in your gut or between your teeth have genomes that contain thousands of different genes. Most of these genes encode proteins, each with its own role in a process such as fuel metabolism, maintenance of cell structure, and defense against viruses.

Some of these proteins are needed routinely, while others are needed only under certain circumstances. Thus, cells don't express all the genes in their genome all the time. You can think of the genome as being like a cookbook with many different recipes in it. The cell will only use the recipes (express the genes) that fit its current needs.

There are various forms of **gene regulation**, that is, mechanisms for controlling which genes get expressed and at what levels. However, a lot of gene regulation occurs at the level of transcription.

Bacteria have specific regulatory molecules that control whether a particular gene will be transcribed into mRNA. Often, these molecules act by binding to DNA near the gene and helping or blocking the transcription enzyme, RNA polymerase. Let's take a closer look at how genes are regulated in bacteria.

In bacteria, related genes are often found in a cluster on the chromosome, where they are transcribed from one **promoter** (RNA polymerase binding site) as a single unit. Such a cluster of genes under control of a single promoter is known as an **operon**. Operons are common in bacteria, but they are rare in eukaryotes such as humans.

A bacterial cell with has a circular bacterial chromosome A small segment of the chromosome is an operon. The DNA of the operon contains three genes, Gene 1, Gene 2, and Gene 3, which are found in a row in the DNA. They are under control of a single promoter (site where RNA polymerase binds) and they are transcribed together to make a single mRNA that has contains sequences coding for all three genes. When the mRNA is translated, the three different coding sequences of the mRNA are read separately, making three different proteins (Protein 1, Protein 2, and Protein 3). The operon does not consist of just the three genes. Instead, it also includes the promoter and other regulatory sequences that regulate expression of the genes.

In general, an operon will contain genes that function in the same process. For instance, a well-studied operon called the *lac* operon contains genes that encode proteins involved in uptake and metabolism of a particular sugar, lactose. Operons allow the cell to efficiently express sets of genes whose products are needed at the same time.

Anatomy of an operon

Operons aren't just made up of the coding sequences of genes. Instead, they also contain **regulatory DNA sequences** that control transcription of the operon. Typically, these sequences are binding sites for **regulatory proteins**, which control how much the operon is transcribed. The promoter, or site where RNA polymerase binds, is one example of a regulatory DNA sequence.

The promoter is found in the DNA of the operon, upstream of (before) the genes. When the RNA polymerase binds to the promoter, it transcribes the operon and makes some mRNAs.

Most operons have other regulatory DNA sequences in addition to the promoter. These sequences are binding sites for regulatory proteins that turn expression of the operon "up" or "down."

- Some regulatory proteins are **repressors** that bind to pieces of DNA called **operators**. When bound to its operator, a repressor reduces transcription (e.g., by blocking RNA polymerase from moving forward on the DNA).

A repressor protein binds to a site called on the operator. In this case (and many other cases), the operator is a region of DNA that overlaps with or lies just downstream of the RNA polymerase binding site (promoter). That is, it is in between the promoter and the genes of the operon. When the repressor binds to the operator, it prevents RNA polymerase from binding to the promoter and/or transcribing the operon. When the repressor is bound to the operator, no transcription occurs and no mRNA is made.

- Some regulatory proteins are **activators**. When an activator is bound to its DNA binding site, it increases transcription of the operon (e.g., by helping RNA polymerase bind to the promoter).

The activator protein binds to a specific sequence of DNA, in this case immediately upstream of (before) the promoter where RNA polymerase binds. When the activator binds, it helps the polymerase attach to the promoter (makes promoter binding more energetically favorable). This causes the RNA polymerase to bind firmly to the promoter and transcribe the

genes of the operon much more frequently, leading to the production of many molecules of mRNA.

Like any other protein produced in an organism, they are encoded by genes in the bacterium's genome. The genes that encode regulatory proteins are sometimes called **regulatory genes**.

Many regulatory proteins can themselves be turned "on" or "off" by specific small molecules. The small molecule binds to the protein, changing its shape and altering its ability to bind DNA. For instance, an activator may only become active (able to bind DNA) when it's attached to a certain small molecule.

When the small molecule is absent, the activator is "off" - it takes on a shape that makes it unable to bind DNA. When the small molecule that activates the activator is added, it binds to the activator and changes its shape. This shape change makes the activator able to bind its target DNA sequence and activate transcription.

Operons may be inducible or repressible

Some operons are usually "off," but can be turned "on" by a small molecule. The molecule is called an **inducer**, and the operon is said to be **inducible**.

- For example, the *lac* operon is an inducible operon that encodes enzymes for metabolism of the sugar lactose. It turns on only when the sugar lactose is present (and other, preferred sugars are absent). The inducer in this case is allolactose, a modified form of lactose.

Other operons are usually "on," but can be turned "off" by a small molecule. The molecule is called a **corepressor**, and the operon is said to be **repressible**.

- For example, the *trp* operon is a repressible operon that encodes enzymes for synthesis of the amino acid tryptophan. This operon is expressed by default, but can be repressed when high levels of the amino acid tryptophan are present. The corepressor in this case is tryptophan.

These examples illustrate an important point: that gene regulation allows bacteria to respond to changes in their environment by altering gene expression (and thus, changing the set of proteins present in the cell).

Some genes and operons are expressed all the time

Many genes play specialized roles and are expressed only under certain conditions, as described above. However, there are also genes whose products are constantly needed by the cell to maintain essential functions. These **housekeeping genes** are constantly expressed under normal growth conditions ("constitutively active"). Housekeeping genes have promoters and other regulatory DNA sequences that ensure constant expression.

E. coli bacteria can break down lactose, but it's not their favorite fuel. If glucose is around, they would much rather use that. Glucose requires fewer steps and less energy to break down than lactose. However, if lactose is the only sugar available, the *E. coli* will go right ahead and use it as an energy source.

To use lactose, the bacteria must express the *lac* operon genes, which encode key enzymes for lactose uptake and metabolism. To be as efficient as possible, *E. coli* should express the *lac* operon only when two conditions are met:

- Lactose is available, **and**
- Glucose is not available

How are levels of lactose and glucose detected, and how do changes in levels affect *lac* operon transcription? Two regulatory proteins are involved:

- One, the *lac* repressor, acts as a lactose sensor.
- The other, catabolite activator protein (CAP), acts as a glucose sensor.

These proteins bind to the DNA of the *lac* operon and regulate its transcription based on lactose and glucose levels. Let's take a look at how this works.

Structure of the *lac* operon

The *lac* operon contains three genes: *lacZ*, *lacY*, and *lacA*. These genes are transcribed as a single mRNA, under control of one promoter.

Genes in the *lac* operon specify proteins that help the cell utilize lactose. *lacZ* encodes an enzyme that splits lactose into monosaccharides (single-unit sugars) that can be fed into glycolysis. Similarly, *lacY* encodes a membrane-embedded transporter that helps bring lactose into the cell.

In addition to the three genes, the *lac* operon also contains a number of regulatory DNA sequences. These are regions of DNA to which particular regulatory proteins can bind, controlling transcription of the operon.

The DNA of the *lac* operon contains (in order from left to right): CAP binding site, promoter (RNA polymerase binding site), operator (which overlaps with promoter), *lacZ* gene, *lacY* gene, and *lacA* gene. The activator protein CAP, when bound to a molecule called cAMP , binds to the CAP binding site and promotes RNA polymerase binding to the promoter. The *lac* repressor protein binds to the operator and blocks RNA polymerase from binding to the promoter and transcribing the operon.

_ The **promoter** is the binding site for RNA polymerase, the enzyme that performs transcription.

- The **operator** is a negative regulatory site bound by the *lac* repressor protein. The operator overlaps with the promoter, and when the *lac* repressor is bound, RNA polymerase cannot bind to the promoter and start transcription.
- The **CAP binding site** is a positive regulatory site that is bound by catabolite activator protein (CAP). When CAP is bound to this site, it promotes transcription by helping RNA polymerase bind to the promoter.

Let's take a closer look at the *lac* repressor and CAP and their roles in regulation of the *lac* operon.

The *lac* repressor

The *lac* repressor is a protein that represses (inhibits) transcription of the *lac* operon. It does this by binding to the operator, which partially overlaps with the promoter. When bound, the *lac* repressor gets in RNA polymerase's way and keeps it from transcribing the operon.

When lactose is not available, the *lac* repressor binds tightly to the operator, preventing transcription by RNA polymerase. However, when lactose is present, the *lac* repressor loses its ability to bind DNA. It floats off the operator, clearing the way for RNA polymerase to transcribe the operon.

Upper panel: No lactose. When lactose is absent, the *lac* repressor binds tightly to the operator. It gets in RNA polymerase's way, preventing transcription.

Lower panel: With lactose. Allolactose (rearranged lactose) binds to the *lac* repressor and makes it let go of the operator. RNA polymerase can now transcribe the operon.

This change in the *lac* repressor is caused by the small molecule **allolactose**, an isomer (rearranged version) of lactose. When lactose is available, some molecules will be converted to allolactose inside the cell. Allolactose binds to the *lac* repressor and makes it change shape so it can no longer bind DNA.

Allolactose is an example of an **inducer**, a small molecule that triggers expression of a gene or operon. The *lac* operon is considered an **inducible operon** because it is usually turned off (repressed), but can be turned on in the presence of the inducer allolactose.

Catabolite activator protein (CAP)

When lactose is present, the *lac* repressor loses its DNA-binding ability. This clears the way for RNA polymerase to bind to the promoter and transcribe the *lac* operon. That sounds like the end of the story, right?

Well...not quite. As it turns out, RNA polymerase alone does not bind very well to the *lac* operon promoter. It might make a few transcripts, but it won't do much more unless it gets extra help from **catabolite activator protein** (**CAP**). CAP binds to a region of DNA just before the *lac* operon promoter and helps RNA polymerase attach to the promoter, driving high levels of transcription.

When glucose levels are low, cAMP is produced. The cAMP attaches to CAP, allowing it to bind DNA. CAP helps RNA polymerase bind to the promoter, resulting in high levels of

When glucose levels are high, no cAMP is made. CAP cannot bind DNA without cAMP, so transcription occurs only at a low level.

CAP isn't always active (able to bind DNA). Instead, it's regulated by a small molecule called **cyclic AMP** (**cAMP**). cAMP is a "hunger signal" made by *E. coli* when glucose levels are low. cAMP binds to CAP, changing its shape and making it able to bind DNA and promote transcription. Without cAMP, CAP cannot bind DNA and is inactive.

CAP is only active when glucose levels are low (cAMP levels are high). Thus, the *lac* operon can only be transcribed at high levels when glucose is absent. This strategy ensures that bacteria only turn on the *lac* operon and start using lactose after they have used up all of the preferred energy source (glucose).

So, when does the *lac* operon <u>really</u> turn on?

The *lac* operon will be expressed at high levels if two conditions are met:

- *Glucose must be unavailable:* When glucose is unavailable, cAMP binds to CAP, making CAP able to bind DNA. Bound CAP helps RNA polymerase attach to the *lac* operon promoter.
- *Lactose must be available*: If lactose is available, the *lac* repressor will be released from the operator (by binding of allolactose). This allows RNA polymerase to move forward on the DNA and transcribe the operon.

These two events in combination – the binding of the activator and the release of the repressor – allow RNA polymerase to bind strongly to the promoter and give it a clear path for transcription. They lead to strong transcription of the *lac* operon and production of enzymes needed for lactose utilization.

Literature:

1. Alberts et al., pp. 369-392;
2. Alberts et al., pp. 413-429.