

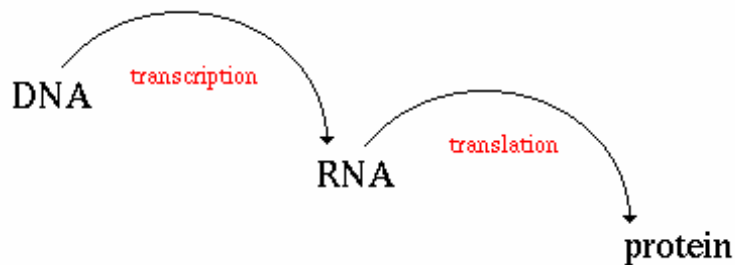
Fundamentals of Gene Regulation and the Arabinose Operon

Abstract: Every living organism has the capability to turn on and off genes at different times and under certain conditions. This occurs continuously in our own bodies without us even being aware. This process of gene regulation is similar in all organisms, increasing in complexity as you climb up the evolutionary ladder. Much of the regulation occurs at the transcriptional level, during the synthesis of messenger RNA (mRNA). In the bacteria, *E. coli*, the genes that degrade the sugar arabinose (the arabinose operon), will serve as a model system for learning about gene regulation.

Introduction:

Each cell that makes up our various organs and tissues contains the exact same complement of DNA - known as our personal genome. In order for our cells to become a skin cell vs. a hair cell, certain genes must be turned on or kept off. This process of switching the expression of genes on and off is known as gene regulation. Some genes, such as the genes for ribosomal proteins, must be continually expressed in all cells to maintain protein production. However, many genes are active only in specific cells or tissues. For example, the gene opsin, which encodes for a protein that helps color reception, is only expressed in the retina cells of the eye.

The Central Dogma Simplified -



Regulation at the level of transcription:

Genes that are regulated are usually controlled during transcription, when messenger RNA (mRNA) is being synthesized from the DNA. There are specific DNA sequences, called regulatory signals or sequences that are often positioned near the structural genes. These short regulatory sequences serve as binding sites for regulatory proteins, called transcription factors. Interaction of these regulatory proteins with the DNA can either block or allow transcription. The proteins that block transcription are called repressors, whereas those that allow transcription are called activators. Therefore, genes expressed only at certain times or only under specific environmental conditions contain regulatory sequences that interact with proteins that are present or active under specific conditions.

The arabinose operon of *E. coli*:

Let's explore how the bacteria, *E. coli*, regulates the genes involved in the breakdown of the sugar arabinose. Arabinose, a five carbon sugar, serves as energy and carbon source for the bacteria. *E. coli* encodes three genes (*araB*, *araA*, and *araD*) that need to be expressed for the transport and breakdown of arabinose. These genes are only expressed when the sugar arabinose is present in the environment. These three *ara* genes are clustered together and make up the arabinose operon (Figure 1).

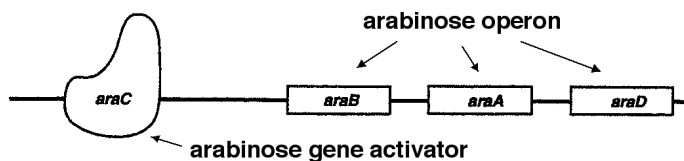


Figure1: Arabinose operon consists of 3 genes, *araB*, *araA* and *araD*, regulated by the arabinose gene activator, *araC*. (Bio-Rad)

The *ara* regulatory sequences precede the 5' end of the *araB* gene. This is called "upstream" from the *ara* coding sequences. The promoter region is the site that binds the RNA polymerase enzyme, which is always required for the transcription of genes.

Just upstream from the promoter lies the binding site for the arabinose gene activator, *araC*. This activator protein is bound to the regulatory sequence and changes shape when the sugar, arabinose, binds to it. This conformational (shape) change allows RNA polymerase to bind at the promoter sequence and transcription is initiated for the arabinose degradation genes, *araB*, *araA*, and *araD* (Figure 2).

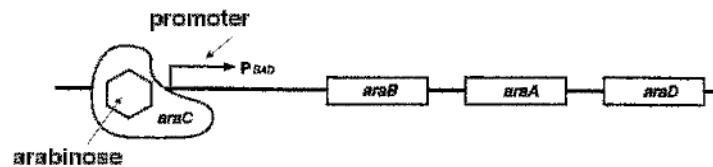


Figure 2: When arabinose is present, it binds to *araC*. Binding causes a change in *araC* conformation revealing the promoter sequence, P_{BAD}. RNA polymerase can now bind and initiate transcription of the 3 genes in the arabinose operon. (Bio-Rad)

So when arabinose is in the environment of *E. coli*, the sugar attaches to the DNA bound activator protein, AraC. The AraC-arabinose complex undergoes a conformational change so that RNA polymerase can now bind to the promoter region. The three genes, *araB*, *araA*, and *araD* are transcribed and translated to produce the three enzymes required for the degradation of arabinose. As the arabinose is used up and broken down, AraC has no arabinose to bind to it, AraC returns to its original shape, RNA polymerase binding does not occur and transcription is then turned off (Figure 3).

The Arabinose Operon

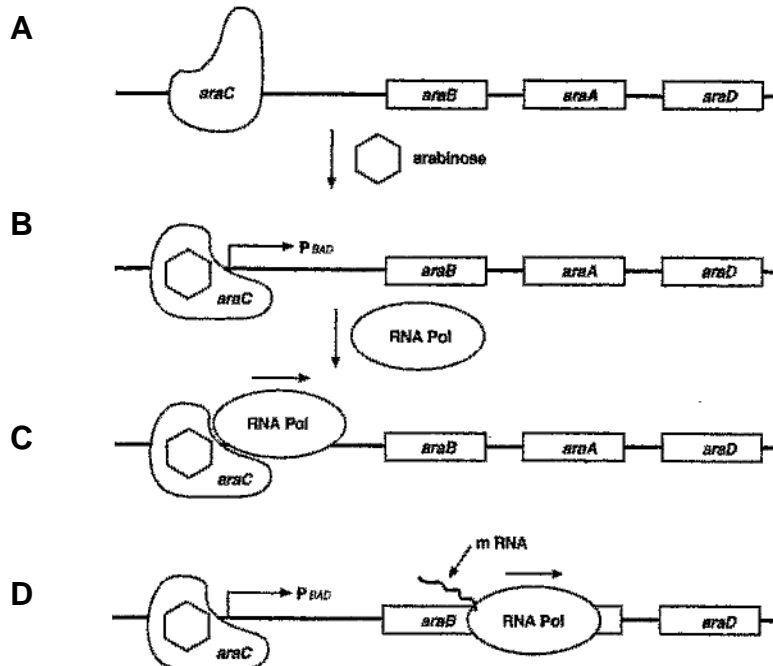


Figure 3:
A. AraC is blocking the binding of RNA polymerase for transcription of the gene responsible for degradation of arabinose.

B. Arabinose binds to *araC* causing a conformational change that reveals the promoter P_{BAD}.

C. RNA polymerase binds to P_{BAD}.

D. Transcription of the arabinose operon is achieved. (Bio-Rad)

Regulation of the GFP using the arabinose promoter:

A plasmid, pGLO, used in expressing the green fluorescent protein (GFP), has been engineered to incorporate various parts of the arabinose operon regulatory regions (Figure 4). Both the promoter (PBAD) and the *araC* gene are present. But the *araB*, *araA*, and *araD* region has been replaced by the single gene for GFP. This places the GFP gene under the regulation of the arabinose promoter and regulatory regions. When bacteria harboring the pGLO plasmid are grown in the presence of arabinose, the AraC protein will bind with arabinose, promoting the binding of RNA polymerase, and transcription of GFP will occur. The bacterial cells that express GFP in this manner will fluoresce a brilliant green color when exposed to UV light. When arabinose is used up, or removed from the growth medium, AraC will not assist the binding of RNA polymerase and GFP will not be transcribed. These cells will remain white, the natural (wild type) phenotype, and not be fluorescent when exposed to UV light.



Figure 4: Plasmid map of pGLO showing the insertion of GFP gene in place of the arabinose operon.

References:

<http://www.accessexcellence.org/AB/GG/central.html>

Access Excellence site for gene regulation and the classic lactose operon example.

Figures on arabinose operon and PBAD-GFP from Bio-Rad's *Secrets of the Rainforest* manual.

DNA

The material inside the nucleus of the cells that carries genetic information. The abbreviation, DNA, stands for deoxyribonucleic acid.

enzyme

A protein that facilitates a specific chemical reaction.

gene

A unit of inheritance; a small section of the DNA strand. Genes contain the code for a specific product, typically a protein such as an enzyme.

gene cloning (DNA cloning)

A lab technique which uses DNA manipulation procedures to produce a recombinant DNA molecule and then to make multiple copies of it by inserting it into the genome of a host microorganism which is then grown in culture.

gene expression

The process by which the genetic information is decoded into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

gene regulation

The process by which genes are turned on and off. This process can occur by a variety of mechanisms.

genetic engineering (gene manipulation, genetic manipulation)

The manipulation of an organism's genetic endowment by introducing or eliminating specific genes through modern molecular biology techniques. A broad definition of genetic engineering also includes selective breeding and other means of artificial selection. See recombinant DNA technologies.

green fluorescent protein (GFP)

A protein found in jellyfish that fluoresces, or emits a green visible light when excited by UV light with a wavelength of 395 nanometers. It can function as a biological marker when co-expressed with other proteins. The structure of the protein is cylindrical with the glowing component, an amino acid complex called a *fluorophore*, in the middle of it.

light

Consists of electromagnetic waves of specific frequencies and wavelengths. Light is often drawn as a ray. The ray shows the path taken by the light waves, the direction in which the energy is being carried.

bioluminescence

A complex chemical reaction that occurs within a living organism where the end product, energy, is released in the form of light instead of heat. Organisms which have this ability include glowworms, fireflies, jellyfish, fungi, and some deep-sea fish.

chemiluminescence

A general term for the production of light when the excitation energy comes from a chemical reaction. Bioluminescence is a type of chemiluminescence that occurs in a living organism.

electromagnetic waves or radiation

Are made up of oscillating electric and magnetic fields that can travel through many substances. Particles that make up the waves are called photons. They consist of many frequencies. At the high frequencies are the gamma rays and X rays, followed by ultraviolet, then the visible light. At the lower frequencies are the infrared radiation waves and microwaves. Within the lower frequencies are the radio waves.

fluorescence

A phenomenon shown by certain substances when they are hit by ultraviolet radiation. The substance absorbs high frequency wavelengths and emits it at a lower frequency light. This emission stops as soon as the high frequency radiation is removed. For example, GFP absorbs the higher frequency blue light emitted by aequorin, undergoes a chemical reaction, and emits the lower wavelength green light.

phosphorescence

Occurs when certain substances are hit by high frequency (short wavelength) electromagnetic waves, and waves of longer wavelength are emitted and released for a longer period of time, continuing long after the

original hit by the high frequency wavelengths. It is similar to fluorescence, except that the excited product is more stable so that the time until the energy is released is much longer. This results in a glow after the light has been removed. This is the basis behind glow-in-the-dark stickers.

ultraviolet radiation (UV)

Electromagnetic radiation produced by the sun and or produced when an electrical current passes through ionized gas between two electrodes. It consists of wavelengths between 200 and 400 nanometers. Exposure to excessive amounts of UV radiation damages DNA and can cause health problems such as skin cancer and cataracts in the eyes.

visible light

Electromagnetic waves or radiation that are produced by the sun and detectable by our eyes. It consists of wavelengths between 400 and 750 nanometers. Colors depend on the wavelength lengths; a short wavelength (the 400 nm side) looks blue and a long wavelength (the 750 nm side) looks red.

molecular biology

The study of the biochemical and molecular processes within cells. Emphasis on the processes of replication, transcription, and translation.

molecular genetics

The study of the flow and regulation of genetic information between DNA, RNA, and protein molecules.

operon system**operon**

A controllable unit of transcription consisting of a number of structural genes transcribed together. Contains at least two distinct regions: the **operator** and the **promoter**. For an example, see the lac operon page.

i gene (lac I operon)

A gene which is a part of the **lac operon**. It codes for a repressor protein, which prevents the transcription of genes that code for the enzymes that bring lactose into the bacterial cell and uses it as a carbon source. The repressor protein prevents the unnecessary production of these enzymes when no lactose is present.

promoter

A site on DNA to which RNA polymerase will bind and initiate transcription.

operator

Site of repressor binding on a DNA molecule; part of an operon.

recombinant DNA

Recombinant DNA is a fragment of DNA incorporated artificially into the DNA molecule of a suitable vector so that it can replicate itself many times. This way a large quantity of the DNA in question can be obtained. The DNA is usually one that contains genes of interest, such as interferon, insulin, or an oncogene. The DNA may also be intended to fix mutated genes causing diseases, such as hemophilia or sickle cell anemia. The vector could be plasmids, bacteriophages, and cosmids (packaged plasmid DNA into a phage particle).

recombinant DNA technologies

Procedures used to join together DNA segments in a cell-free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Vectors**cloning vector**

A DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self-replication. Vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

expression vector

A cloning vector that contains the necessary regulatory sequences to allow transcription and translation of a cloned gene or genes. A vector which is able to replicate and transcribe cloned DNA.

host-vector system

A combination of a bacterial host cell (i.e. a specific strain) and a virus vector (i.e. a particular bacteriophage strain) which work well together for DNA cloning.