Gene regulation by operons

Matthew Macauley

Department of Mathematical Sciences
Clemson University
http://www.math.clemson.edu/~macaule/

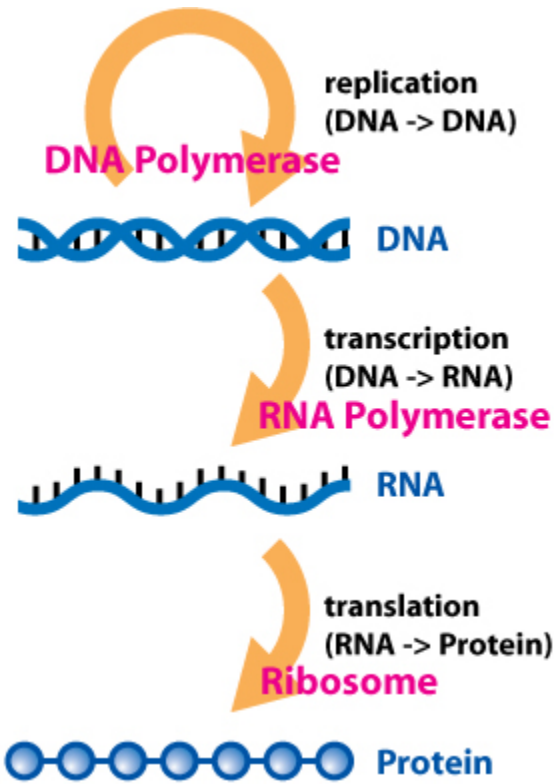
Math 4500, Spring 2025

Gene expression and the central dogma

Gene expression is a process that takes gene info and creates a functional gene product (e.g., a protein or enzyme).

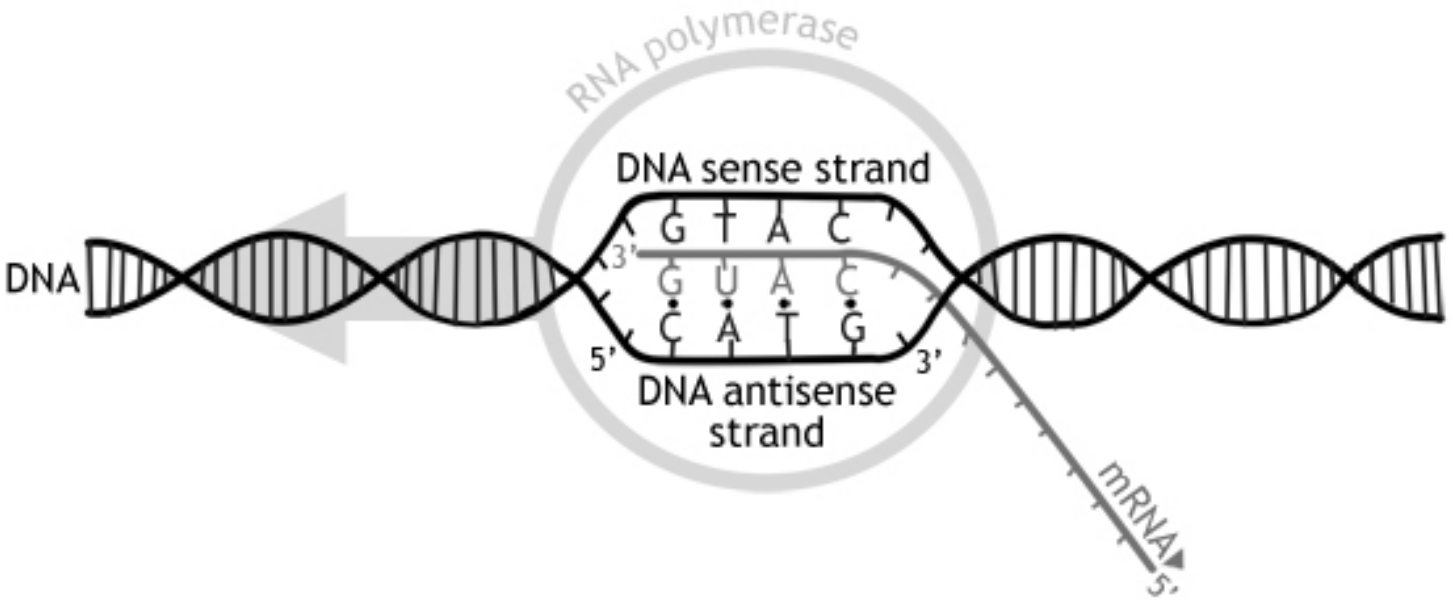
Gene expression is a 2-step process, called the central dogma:

1. **transcription** of genes: messenger RNA synthesis
2. **translation** of genes: protein synthesis



- DNA consists of bases adenine (A), cytosine (C), guanine (G), and thymine (T).
- RNA consists of bases A, C, G, and uracil (U).
- Proteins are long chains of amino acids.
- Gene expression is used by all known life forms.

Step 1: Transcription (mRNA synthesis)



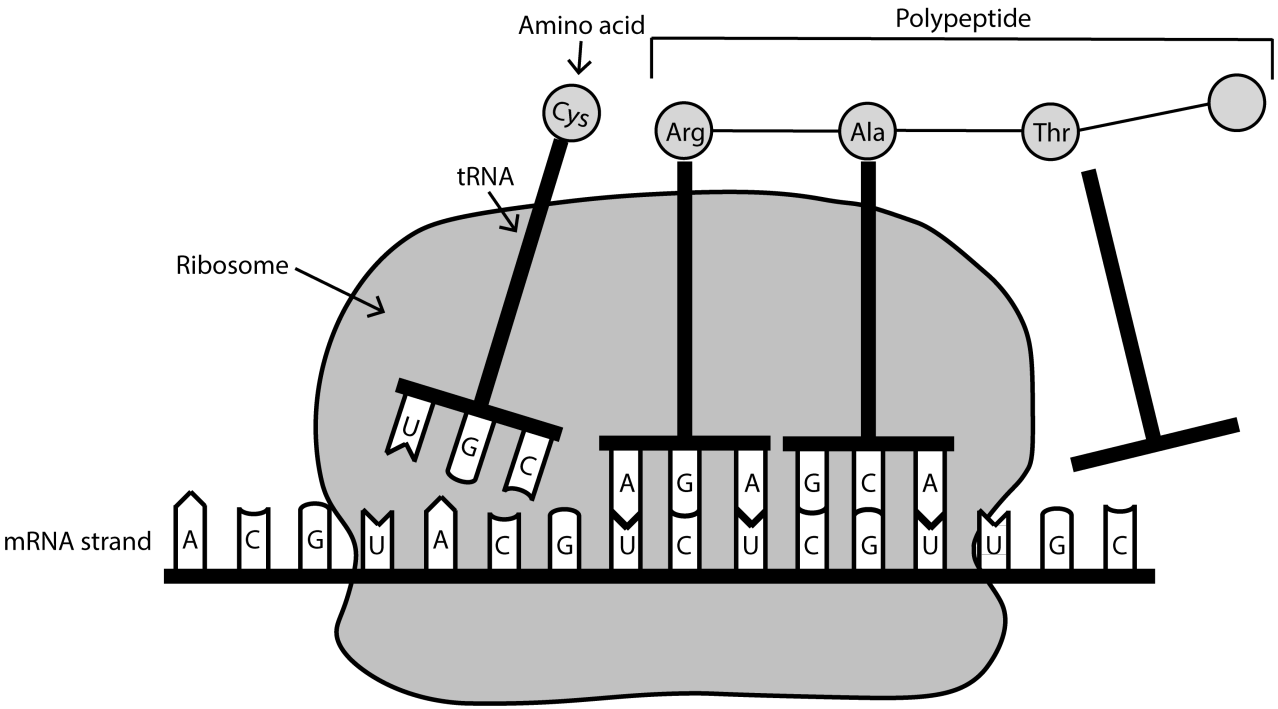
Transcription in eukaryotes occurs inside the cell nucleus.

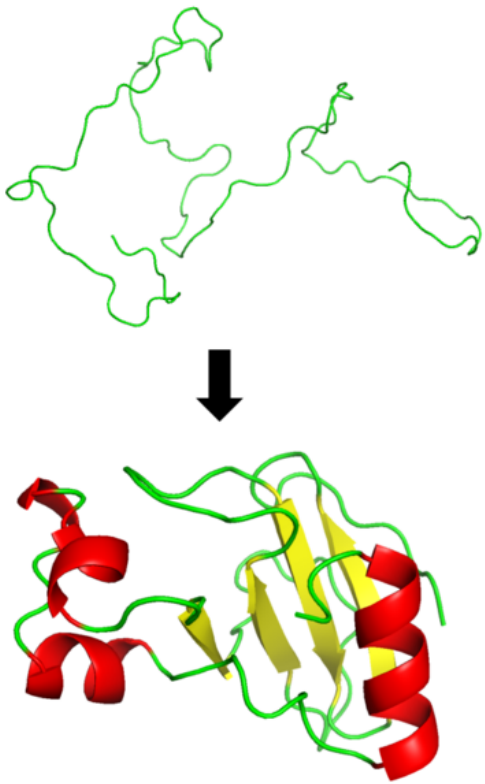
RNA polymerase uses a helicase enzyme to bind to DNA, "unzipping" it to read it.

DNA is copied into messenger RNA.

Segments of RNA not needed for protein coding are removed.

Step 2: Translation (protein synthesis)





During translation, mRNA is read by ribosomes.

Each triple of RNA bases codes for an amino acid.

The result is a protein: a long chain of amino acids.

Proteins fold into a 3D shape which determines function.

Gene expression and the central dogma

The expression level is the rate at which a gene is being expressed.

Housekeeping genes are continuously expressed; they are essential for basic life processes.

Regulated genes are expressed only under certain outside factors (e.g., environmental, physiological, etc.). Expression is controlled by the cell.

It is easiest and most efficient to control gene regulation by affecting transcription.

One way to block transcription is for a repressor protein to bind to the DNA or RNA.

Goal

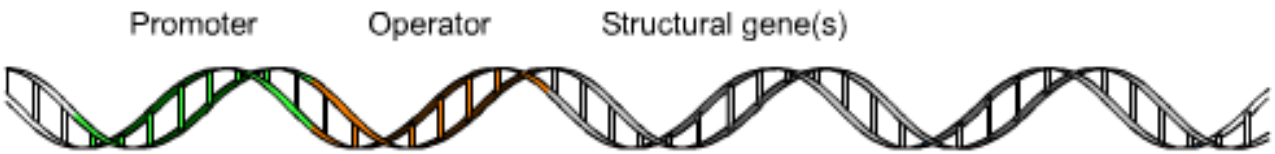
Understand the complex cell behaviors of gene regulation—how the cell turns on/off certain genes depending on the organism's requirements.

Regulation by operons

An operon is a region of DNA that contains a cluster of genes that are transcribed together.

An operon has three regions:

1. **Promoter**: Where RNA polymerase binds and transcription is initiated.
2. **Operator**: Where the repressor protein binds.
3. **Strucural genes** to be transcribed.



DNA replication and gene expression were all studied in prokaryotes before they were studied in eukaryotic cells.

Operons are more common in prokaryotes than eukaryotes.

They are also used by viruses, such as bacteriophages. The well-studied T7 phage has two operons.

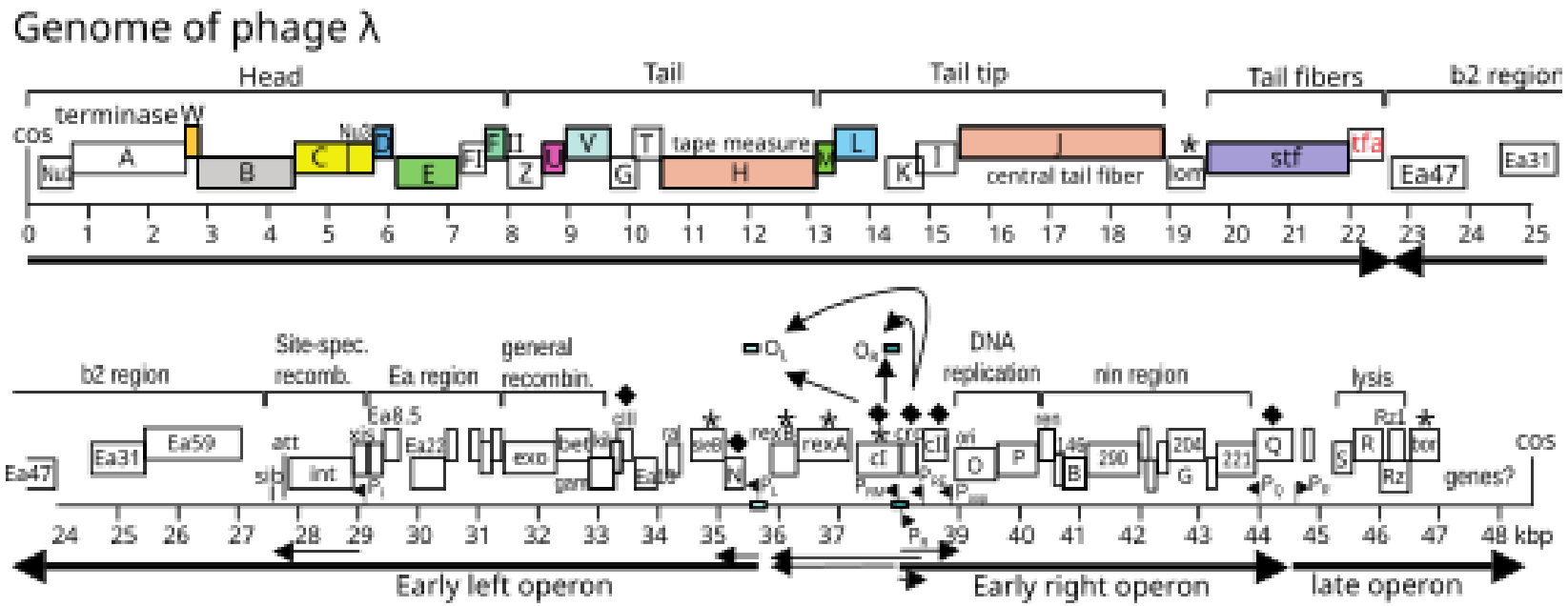
Regulation by operons

The genome of the T7 phage virus was sequenced in 1983.

It has a linear genome of nearly 50,000 base pairs.

It has two operons:

- One codes for gene products
- One contols *lysis*—cell death by bursting.



Gene regulation in eukaryotes

From B. Alberts, A. Johnson, J. Lewis, et al. *Molecular Biology of the Cell.* 7th edition (2022). W. W. Norton.

The regulation of transcription in eucaryotes differs in three important ways from that typically found in bacteria.

- First, eucaryotes make use of gene regulatory proteins that can act even when they are bound to DNA thousands of nucleotide pairs away from the promoter that they influence, which means that a single promoter can be controlled by an almost unlimited number of regulatory sequences scattered along the DNA.
- Second, as we saw in the last chapter, eucaryotic RNA polymerase II, which transcribes all protein-coding genes, cannot initiate transcription on its own. It requires a set of proteins called *general transcription factors*, which must be assembled at the promoter before transcription can begin. (The term "general" refers to the fact that these proteins assemble on all promoters transcribed by RNA polymerase II; in this they differ from gene regulatory proteins, which act only at particular genes.) This assembly process provides, in principle, multiple steps at which the rate of transcription initiation can be speeded up or slowed down in response to regulatory signals, and many eucaryotic gene regulatory proteins influence these steps.
- Third, the packaging of eucaryotic DNA into chromatin provides opportunities for regulation not available to bacteria.

Operons in *E. coli*

Escherichia coli is a bacterium in the gut of mammals and birds. Its genome has been sequenced and its physiology is well-understood.

E. coli has 4307 mRNA-encoding genes controlled by

- 788 operons ("polycistronic mRNAs")
- 1781 monocistronic mRNAs.

The lactose (*lac*) operon controls the transport and metabolism of lactose in *E. coli*.

The *lac* operon was discovered by Francois Jacob and Jacques Monod, which earned them the 1965 Nobel Prize.

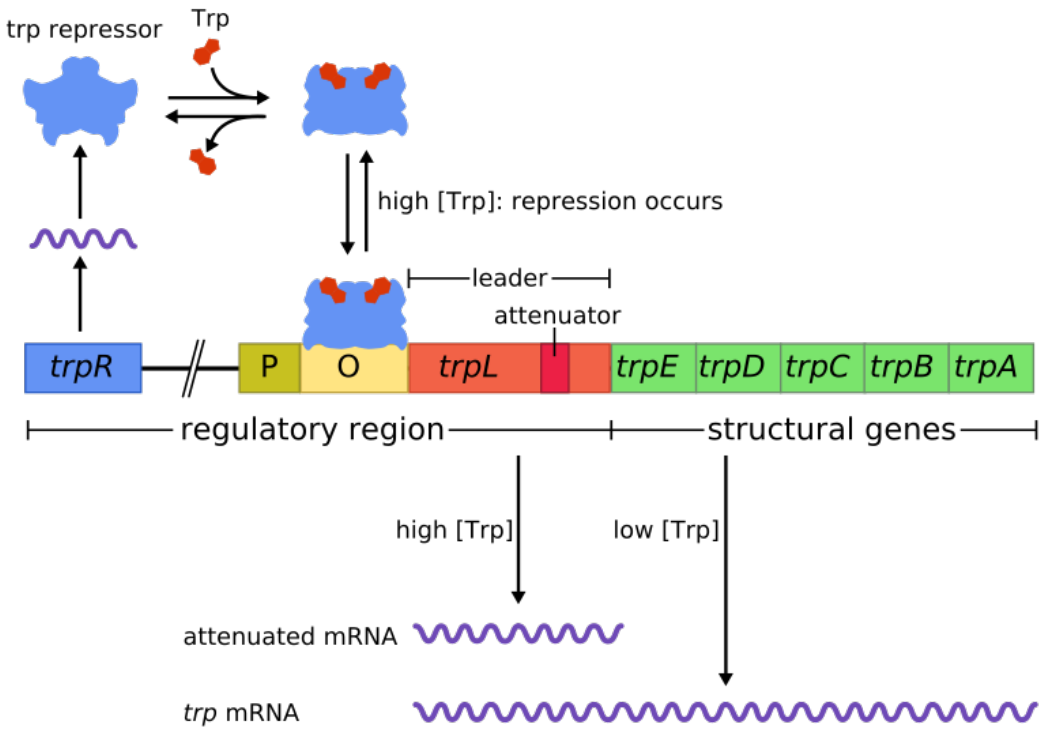
It was the first operon discovered and is the most widely studied mechanism of gene regulation.

It is used as a "test system" for models of gene regulation.

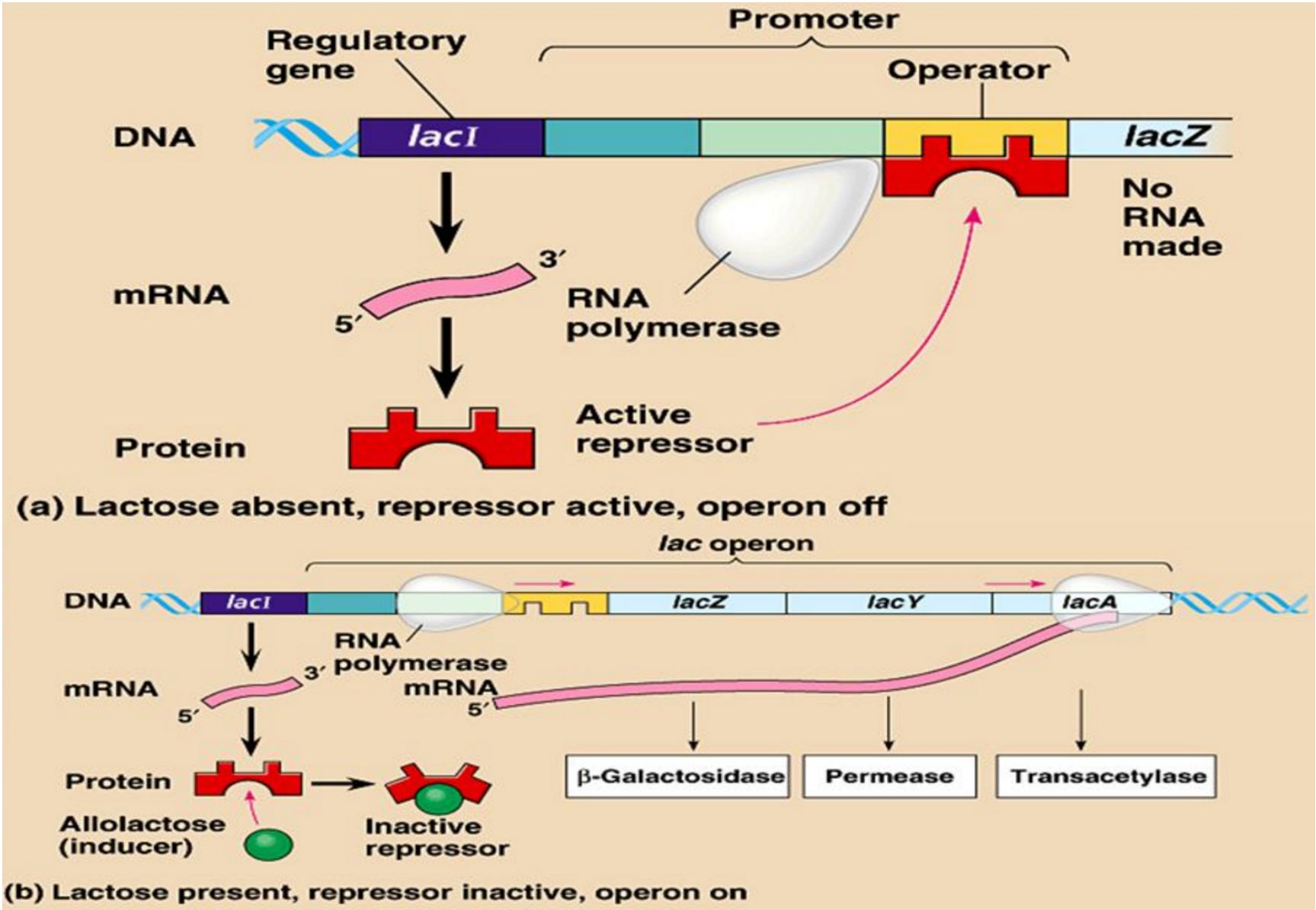
Operons in *E. coli*

The *lac* operon is inducible: it is OFF by default.

In contract, the tryptophan (*trp*) operon is repressible: it is ON by default.



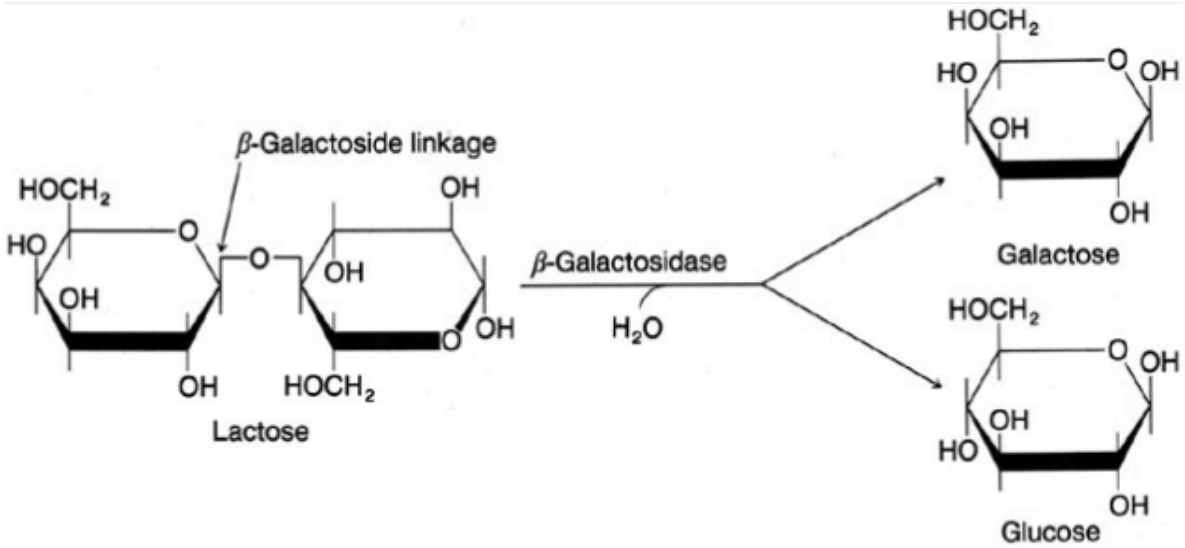
The *lac* operon in *E. coli*



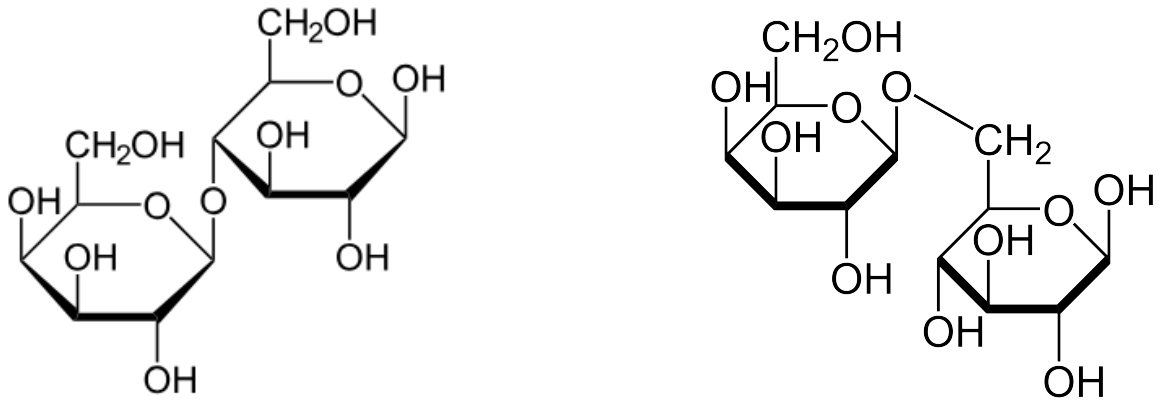
The *lacZ* gene and β-galactosidase

The first structural gene codes for the β-galactosidase enzyme, which has two functions:

1. Cleaves lactose ($C_{12}H_{22}O_{11}$) into glucose ($C_6H_{12}O_6$) and galactose ($C_6H_{12}O_6$).

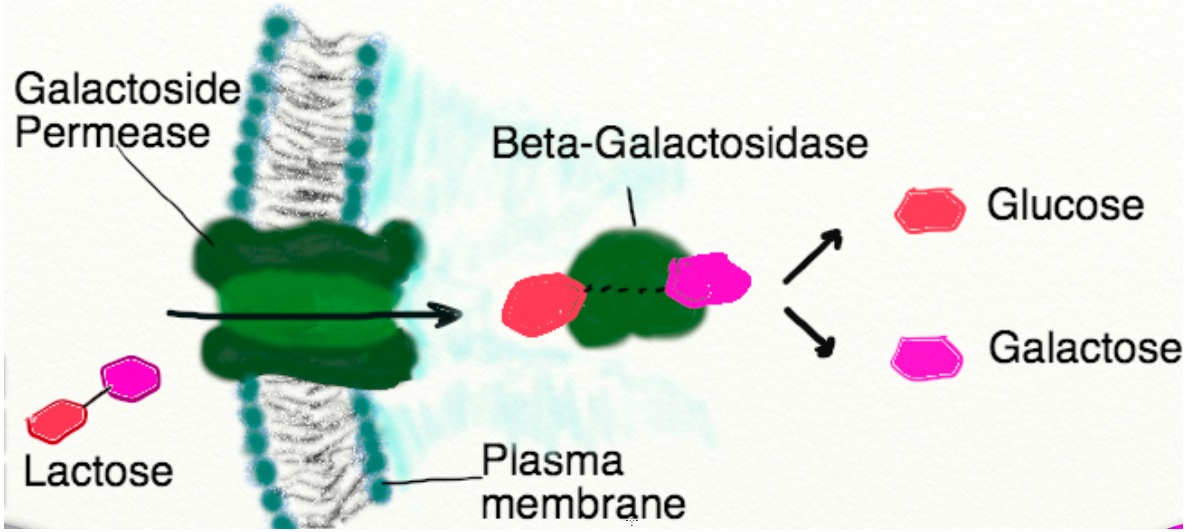


2. Catalyzes the isomerization of lactose to allolactose.



The *lacY* gene and *lac*-permease

The second structural gene codes for *lac* permease (LacY), a transporter protein.



The third structural gene codes for β-galactoside transacetylase (LacA).

It is unclear how this is involved in this gene regulatory network.

Summary: with lactose, but no glucose

Lactose is brought into the cell by the *lac* permease transporter protein.

β-galactosidase cleaves lactose into glucose and galactose.

β-galactosidase also converts lactose into allolactose.

Allolactose binds to the *lac* repressor protein, preventing it from binding to the operator region of the operon.

Transcription begins: mRNA encoding the *lac* genes is produced.

Lac proteins are synthesized, and more lactose is brough into the cell. The operon is ON.

Eventually, all lactose is used up, so there will be no more allolactose.

The *lac* repressor can now bind to the operator, so mRNA transcription stops. The operon has turned itself OFF.

Catabolite repression: with lactose and glucose

Glucose is the preferred carbon source of *E. coli*.

Thus, if glucose and lactose are both present, then the *lac* operon should be OFF.

This is done by a cellular mechanism called catabolite repression.

The lac operon promoter region has two binding sites, for:

- RNA polymerase (this "unzips" and reads the DNA)
- the CAP-cAMP complex (Analogy: "the key that starts the engine.")

This is a complex of two molecules: catabolite activator protein (CAP), and the cyclic AMP receptor protein (cAMP, or crp).

Intracellular glucose causes cAMP concentration to decrease.

When cAMP levels get too low, so do CAP-cAMP complex levels.

Without the CAP-cAMP complex, the promoter is inactivated, and the lac operon is OFF.

Modeling the *lac* operon

Models of molecular networks have variables that represent concentration levels of key gene products and substrates.

Any reasonable model of the *lac* operon should be able to capture the following:

- No lactose: operon is OFF
- Lactose and glucose: operon is OFF
- Lactose and no glucose: operon is ON.

Molecular concentrations are highly nonlinear, and they depend on complex biochemical reactions.

There are also other features such as time delays, dilution, degradation, and bistability that modeling frameworks should be able to handle.

We will see two very different approaches to modeling the lac operon:

- Delay differential equations (quantitative continuous framework).
- Boolean models (qualitative discrete framework).

Both have their pros and cons, which we will discuss.