Engineering Genetic Circuits

Chris J. Myers

Lecture 1: An Engineers Guide to Biology and Biochemistry
Biology has at least 50 more interesting years (1984).
DNA makes RNA, RNA makes protein, and proteins make us.
Chemical reactions
Macromolecules
Genomes
Cells and their structure
Genetic circuits
Viruses
Phage λ
Atoms are the basic building block for all matter.

About 98 percent of any living organism consists of: hydrogen (H), carbon (C), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S).

All material is created or destroyed via chemical reactions.

Chemical reactions combine atoms to form molecules and combine simpler molecules to form more complex ones.

Atoms form molecules via covalent, ionic, and hydrogen bonds.

Chemical reactions can also work in reverse.
Chemical Reaction Example

\[ 2H_2 + O_2 \overset{k}{\rightarrow} 2H_2O \]

- \( H_2, O_2, \) and \( H_2O \) are chemical (or molecular) species.
- Subscripts indicate \( H \) and \( O \) are present in \textit{dimer} form.
- The molecules \( H_2 \) and \( O_2 \) are known as the \textit{reactants}.
- The water molecule, \( H_2O \), is known as the \textit{product}.
- The 2’s indicate 2 \( H_2 \) molecules are used to produce 2 water molecules.
- These numbers are known as the \textit{stoichiometry} of the reaction.
- Since matter is conserved, atom counts on each side must equal.
- Some reactions in this course may not have this property.
The $k$ above the arrow is known as the rate constant. It indicates the probability or speed of this reaction. Used in many of the modeling techniques in this course. Often difficult to determine for bio-chemical reactions.
Rate of a chemical reaction is governed by rate constant and concentrations of reactants raised to power of stoichiometry.

This is known as the *law of mass action*.

The rate of water formation is:

\[
\frac{d[H_2O]}{dt} = 2k[H_2]^2[O_2]
\]

where [H$_2$O], [H$_2$], and [O$_2$] represent the concentration of water, hydrogen dimers, and oxygen dimers.

2 in front of $k$ is due to this reaction producing two water molecules.
Laws of Thermodynamics

- Chemical reactions must obey the laws of thermodynamics.
- First law is that energy can be neither created nor destroyed.
- Second law is entropy (disorder in the universe) must increase.
- These two laws can be combined into a single equation:

\[ \Delta H = \Delta G + T\Delta S \]

where \( \Delta H \) is change in bond energy, \( \Delta G \) is change in free energy, \( T \) is the absolute temperature, and \( \Delta S \) is change in entropy.
• $\Delta G$ is also known as the Gibb’s free energy after J. Willard Gibbs who introduced this concept in 1878.

• Consider a reversible reaction of the form:

$$2H_2 + O_2 \xrightleftharpoons{K_{eq}} 2H_2O$$

where $K_{eq} = k/k_{-1}$ is the equilibrium constant.

• The Gibb’s free energy for the forward reaction is:

$$\Delta G = \Delta G^\circ + RT \ln \{ ([H_2O]^2)/([H_2]^2[O_2]) \}$$

where $R = 1.987$ cal/mol is the gas constant and $T$ is the temperature.
The value of $\Delta G^\circ$ is related to $K_{eq}$:

$$\Delta G^\circ = -RT \ln K_{eq}$$

Combining equations results in:

$$\Delta G = RT \ln \frac{k_1[H_2O]^2}{k[H_2]^2[O_2]}$$

- When negative, forward reaction can occur spontaneously.
- When positive, reverse reaction can occur spontaneously.
- When zero, the reaction is in a steady state.
Hydrolysis of ATP

- How do chemical reactions with positive free energy occur?
- Free energies of chemical reactions are additive.
- Coupling with other reactions allows them to occur.
- *Hydrolysis* of ATP releases energy:

\[
ATP + H_2O \leftrightarrow HPO_4^{2-} + ADP.
\]

- These types of ATP reactions occur in all living organisms.
- ATP is the universal energy currency of living organisms.
Enzymes

- **Activation energy** barrier must be overcome.
- An *enzyme*, or catalyst, can accelerate a reaction without being consumed by the reaction.
- **Modifier** is a species that is not consumed by a reaction.
- Often enzyme amount much smaller than other reactants.
- Enzymes do not effect free energy of the reaction, but only help the reaction overcome its activation energy barrier.
Nearly 70 percent of all living organisms are made up of water.

Remainder largely *macromolecules* of 1000s of atoms.

There are four types:

- *Carbohydrates*
- *Lipids*
- *Nucleic acids*
- *Proteins*
Carbohydrates

- Made up of carbon and water \((C_n(H_2O)_m\) where \(m \approx n\)).
- Often called sugars.
- An example is glucose.
- Important source of chemical energy.
- Powers nearly all processes of a cell.
- Also part of the backbone for DNA and RNA.
Lipids

- Made up mostly of carbon and hydrogen atoms.
- Often have a *hydrophilic* (water-loving) part and a *hydrophobic* (water-fearing) part.
- Primary use is to form *membranes*.
- Membranes separate cells from one another and create compartments within cells as well as having other functions.
- Make good membranes because their hydrophobic parts attract to form *lipid bilayers* where exterior allows water, but interior repels water.
- This allows the lipid bilayers to form between areas containing water, but they do not allow water to easily pass through.
- Examples include fats, oils, and waxes.
Nucleic Acids

(Courtesy: National Human Genome Research Institute)
Deoxyribonucleic acid (DNA)

- Stores information within living organisms.
- Composed of base bound to sugar and phosphate molecule.
- Two forms: DNA (deoxyribonucleic acid) and RNA (ribonucleic acid).
- Sequence of nucleotides encode the instructions to construct proteins.
- Most organisms use DNA, but a few viruses use RNA.
- A DNA strand (chain) is made up of four chemical bases: adenine (A) and guanine (G), which are called purines, and cytosine (C) and thymine (T), referred to as pyrimidines.
- Each base has slightly different composition of O, C, N, H.
A strand of DNA is always synthesized in the 5’ to 3’ direction.
The so-called 5’ end terminates in a 5’ phosphate group (-PO4); the 3’ end terminates in a 3’ hydroxyl group (-OH).
DNA is a double-stranded with each strand running in opposite directions.
A-T and G-C base pairs are complementary.
Chemical makeup of this base pairing creates a force that twists the DNA into its coiled double helix structure.
DNA is readily copied since one strand of DNA can act as a template to direct the synthesis of a complementary strand.
A single-stranded chain of nucleotides with the same 5’ to 3’ direction.

Uses a different sugar and uracil replaces the thymine nucleotide.

All genes that code for proteins are first made into an RNA strand called a messenger RNA (mRNA).

mRNA carries the information encoded in DNA to the protein assembly machinery, or ribosome.

The ribosome complex uses mRNA as a template to synthesize the exact protein coded for by the gene.

DNA also codes for ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), and small nuclear RNAs (snRNAs).
Proteins

- Basic building blocks of nearly all the machinery of a cell.
- Each cell contains thousands of different proteins.
- Long chains with as many as 20 kinds of amino acids.
- Genetic code carried by DNA specifies order and number of amino acids and, therefore, shape and function of the protein.
- Code from DNA is transferred to RNA through transcription.
- mRNA is translated by a ribosome into protein.
- mRNA decoded in blocks of three bases, or codons.
- Protein built one amino acid at a time, with order determined by the order of the codons in the mRNA.
<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU Phenylalanine</td>
<td>UCU Serine</td>
<td>UAU Tyrosine</td>
<td>UGU Cysteine</td>
</tr>
<tr>
<td></td>
<td>UUC Phenylalanine</td>
<td>UCC Serine</td>
<td>UAC Tyrosine</td>
<td>UGC Cysteine</td>
</tr>
<tr>
<td></td>
<td>UUA Leucine</td>
<td>UCA Serine</td>
<td>UAA Stop</td>
<td>UGA Stop</td>
</tr>
<tr>
<td></td>
<td>UUG Leucine</td>
<td>UCG Serine</td>
<td>UAG Stop</td>
<td>UGG Tryptophan</td>
</tr>
<tr>
<td>C</td>
<td>CUU Leucine</td>
<td>CCU Proline</td>
<td>CAU Histidine</td>
<td>CGU Arginine</td>
</tr>
<tr>
<td></td>
<td>CUC Leucine</td>
<td>CCC Proline</td>
<td>CAC Histidine</td>
<td>CGC Arginine</td>
</tr>
<tr>
<td></td>
<td>CUA Leucine</td>
<td>CCA Proline</td>
<td>CAA Glutamine</td>
<td>CGA Arginine</td>
</tr>
<tr>
<td></td>
<td>CUG Leucine</td>
<td>CCG Proline</td>
<td>CAG Glutamine</td>
<td>CGG Arginine</td>
</tr>
<tr>
<td>A</td>
<td>AUU Isoleucine</td>
<td>ACU Threonine</td>
<td>AAU Asparagine</td>
<td>AGU Serine</td>
</tr>
<tr>
<td></td>
<td>AUC Isoleucine</td>
<td>ACC Threonine</td>
<td>AAC Asparagine</td>
<td>AGC Serine</td>
</tr>
<tr>
<td></td>
<td>AUA Isoleucine</td>
<td>ACA Threonine</td>
<td>AAA Lysine</td>
<td>AGA Arginine</td>
</tr>
<tr>
<td></td>
<td>AUG Methionine</td>
<td>ACG Threonine</td>
<td>AAG Lysine</td>
<td>AGG Arginine</td>
</tr>
<tr>
<td>G</td>
<td>GUU Valine</td>
<td>GCU Alanine</td>
<td>GAU Aspartate</td>
<td>GGU Glycine</td>
</tr>
<tr>
<td></td>
<td>GUC Valine</td>
<td>GCC Alanine</td>
<td>GAC Aspartate</td>
<td>GGC Glycine</td>
</tr>
<tr>
<td></td>
<td>GUA Valine</td>
<td>GCA Alanine</td>
<td>GAA Glutamate</td>
<td>GGA Glycine</td>
</tr>
<tr>
<td></td>
<td>GUG Valine</td>
<td>GCG Alanine</td>
<td>GAG Glutamate</td>
<td>GGG Glycine</td>
</tr>
</tbody>
</table>
In 1961, Nirenberg and Matthaei correlated the first codon (UUU) with the amino acid phenylalanine.

A given amino acid can have more than one codon.

These redundant codons usually differ at the third position.

Serine is encoded by UCU, UCC, UCA, and/or UCG.

Redundancy is key to accommodating mutations that occur naturally as DNA is replicated and new cells are produced.

Some codons do not code for an amino acid at all but instruct the ribosome when to stop adding new amino acids.
Protein Structure

- Protein folds into a specific 3-dimensional configuration.
- Shape and position of the amino acids in this folded state determines the function of the protein.
- Understanding and predicting protein folding is an important area of research.

The structure of a protein is described in four levels.

- **Primary structure** - sequence of amino acids.
- **Secondary structure** - patterns formed by amino acids that are close (ex. α-helices and β-pleated sheets).
- **Ternary structure** - arrangement of far apart amino acids.
- **Quaternary structure** - arrangement of proteins that are composed of multiple amino acid chains.
Protein Structure

Primary protein structure
is sequence of a chain of amino acids

Amino Acids

Secondary protein structure
occurs when the sequence of amino acids
are linked by hydrogen bonds

Pleated sheet

Alpha helix

Tertiary protein structure
occurs when certain attractions are present
between alpha helices and pleated sheets

Pleated sheet

Alpha helix

Quaternary protein structure
is a protein consisting of more than one
amino acid chain.

(Courtesy: National Human Genome Research Institute)
What is a Genome?

- All of the 30 million types of organisms use the same basic materials and mechanisms to produce building blocks necessary for life.
- Information encoded in the DNA within its *genome* is used to produce RNA which produces proteins.
- A genome is divided into genes where each gene encodes the information necessary for constructing a protein.
- Some also control the production of proteins by other genes.
What are Genes?

- In 1909, Danish botanist Wilhelm Johanssen coined the word *gene* for the hereditary unit found on a chromosome.
- 50 years earlier, Gregor Mendel characterized hereditary units as *factors*—differences passed from parent to offspring.
- Mendel was an Austrian monk who experimented with his pea plants in the monastery gardens.
- Normally they self-fertilize, but he manipulated their parentage and thus their traits using a pair of clippers.
- Discovery went largely ignored for nearly 50 years until three researchers essentially duplicated his results.
Where are Genes?

- Until 1953, it was not known for certain that genes are made of DNA.
- In 1953, Watson and Crick, with support from x-ray data from Franklin and Wilkens, discovered the double helix structure of DNA.
- This discovery showed that DNA is composed of two strands composed of complementary bases.
- This base pairing idea shed light on how DNA could encode genetic information and be readily duplicated during cell division.
- Between 1953 and 1965, work by Crick and others showed how the DNA codes for amino acids and thus proteins.
In February 2001, two largely independent draft versions of the human genome were published in Nature.

Both estimated between 30,000 to 40,000 genes in the human genome (today’s estimate is between 20,000 and 25,000).

How do scientists estimate the number of genes in a genome?

- **Open reading frames**, a 100 bases without a stop codon;
- **Start codons** such as ATG;
- Specific sequences found at *splice junctions*; and
- **Gene regulatory sequences**.

When complete mRNA sequences known, software can align start and end sequences with the DNA sequence.
What is Contained in Our Genome?

- Sequences that code for proteins are called *structural genes*.
- *Regulatory sequences* are start/end of genes, sites for initiating replication/recombination, or sites to turn genes on/off.
- Over 98% of our genome has unknown function ("junk" DNA).
- "*Repetitive DNA*", short sequences repeated 100s of times, make up 40 to 45 percent of our genome.
- Although have no role in the coding of proteins, they are an excellent "*marker*" by which individuals can be identified.
- "*Pseudogenes*" are believed to be a remnant of a real gene that has suffered mutations and is no longer functional.
- Believed to carry a record of our evolutionary history.
Introns and Exons

- Genes make up about 1% of the total DNA in our genome.
- A eukaryotic gene is not found in a continuous stretch.
- The coding portions of a gene, called exons, are interrupted by intervening sequences, called introns.
- Both exons and introns are transcribed into mRNA, but before being transported to the ribosome, the mRNA transcript is edited.
- Removes introns, joins exons together, and adds unique features to end of transcript to make a “mature” mRNA.
- It is still unclear what all the functions of introns are, but may serve as the site for recombination.
About 40 percent of the expressed genome is *alternatively spliced* to produce multiple proteins from a single gene.

This process may have evolved to limit effects of mutations.

Genetic mutations occur randomly, and the effect of a small number of mutations on a single gene may be minimal.

However, an individual having many genes each with small changes could weaken the individual, and thus the species.

If single mutation affects several alternate transcripts, it is likely that the individual will not survive.
What is a Cell?

- The structural and functional unit of all living organisms.
- Some organisms, such as bacteria, are *unicellular*.
- Other organisms, such as humans, are *multicellular*.
- Humans have an estimated 100,000,000,000,000 cells!
- Each cell can take in nutrients, convert these into energy, carry out specialized functions, and reproduce as necessary.
- Each cell stores its own set of instructions in its genome for carrying out each of these activities.
Life arose on earth about 3.5 billion years ago.
The first types of cells were prokaryotic cells.
They are unicellular organisms that lack a nuclear membrane.
They do not develop or differentiate into multicellular forms.
Bacteria are the best known and most studied form.
Some bacteria grow in masses, but each cell is independent.
They are capable of inhabiting almost every place on the earth.
They lack intracellular organelles and structures.
Most functions of organelles are taken over by the plasma membrane.
Prokaryotic Features

(Courtesy: National Human Genome Research Institute)
Eukaryotic Organisms

- Eukaryotes appear in the fossil record about 1.5 billion years ago.
- They include fungi, mammals, birds, fish, invertebrates, mushrooms, plants, and complex single-celled organisms.
- Eukaryotic cells are about 10 times the size of a prokaryote and can be as much as 1000 times greater in volume.
- Use same genetic code and metabolic processes as prokaryotes, but higher level organizational complexity permits multicellular organisms.
- Have membrane-bounded compartments called organelles.
- Most important is the nucleus that houses the cell’s DNA.
Eukaryotic Features

(Courtesy: National Human Genome Research Institute)
The Plasma Membrane (A Cell’s Protective Coat)

- Outer lining of a eukaryotic cell.
- Separates and protects a cell from its environment.
- Made mostly of lipids, proteins, and carbohydrates.
- Embedded within are a variety of molecules that act as channels and pumps, moving molecules into and out of the cell.
- In prokaryotes, usually referred to as the cell membrane.
The Cytoskeleton (A Cell’s Scaffold)

- Acts to organize and maintain the cell’s shape.
- Anchors organelles in place.
- Helps during *endocytosis*, the uptake of external materials.
- Moves parts of the cell in processes of growth and motility.
- Involves many proteins each controlling a cell’s structure by directing, bundling, and aligning filaments.
The large fluid-filled space inside the cell.
In prokaryotes, this space is relatively free of compartments.
In eukaryotes, “soup” within which organelles reside.
Home of the cytoskeleton.
Contains dissolved nutrients, helps break down waste products, and moves material around the cell.
Contains many salts and is an excellent conductor of electricity, creating the perfect environment for the mechanics of the cell.
The function of the cytoplasm, and the organelles which reside in it, are critical for a cell’s survival.
Organelles

- They are a set of “little organs” that are adapted and/or specialized for carrying out one or more vital functions.
- Organelles are found only in eukaryotes and are always surrounded by a protective membrane.
The Nucleus: A Cell’s Center

- A spheroid membrane-bound region that contains genetic information in long strands of DNA called *chromosomes*.
- Most conspicuous organelle found in a eukaryotic cell.
- Separated from the cytoplasm by *nuclear envelope* which isolates and protects DNA from molecules that could damage its structure or interfere with its processing.
- Where almost all DNA replication and RNA synthesis occurs.
- During processing, DNA is transcribed into mRNA.
- mRNA is transported out of the nucleus, where it is translated into a specific protein molecule.
- In prokaryotes, DNA processing takes place in the cytoplasm.
The Ribosome: The Protein Production Machine

- Found in both prokaryotes and eukaryotes.
- It is a large complex composed of RNAs and proteins.
- They process genetic instructions carried by mRNA.
- Translation is the process of converting a mRNA’s genetic code into the exact sequence of amino acids that make up a protein.
- Protein synthesis is extremely important, so there are a large number of ribosomes (100s or 1000s) in a cell.
- They float freely in the cytoplasm or sometimes bind to another organelle called the endoplasmic reticulum.
The endoplasmic reticulum (ER) is the transport network for molecules targeted for modifications and specific destinations.

The ER has two forms: the rough ER and the smooth ER.

The rough ER has ribosomes adhering to its outer surface.

Translation of the mRNA for proteins that either stay in the ER or are exported out of the cell occurs at these ribosomes.

The smooth ER serves as the recipient for those proteins synthesized in the rough ER.

Proteins to be exported are passed to the Golgi apparatus for further processing, packaging, and transport to other locations.
Mitochondria and Chloroplasts

- **Mitochondria** are self-replicating organelles that occur in various numbers, shapes, and sizes.
- Have two functionally distinct membrane systems:
  - The outer membrane, which surrounds the organelle; and
  - The inner membrane, which has folds called *cristae* that project inwards to increase its surface area.
- Plays a critical role in generating energy in the eukaryotic cell.
- **Chloroplasts** are similar but are found only in plants.
- Both surrounded by double membrane and involved in energy metabolism.
- Chloroplasts convert sun’s light energy into ATP using *photosynthesis*.
Plants and animals have genome within mitochondria and chloroplasts. Mitochondrial DNA is only inherited from our mother. Independent aerobic function may have evolved from bacteria living inside other organisms in a symbiotic relationship. These organisms evolved to become incorporated into the cell. Many diseases caused by mutations in mitochondrial DNA. Mitochondrial Theory of Aging suggests accumulation of mutations in mitochondria drives aging process.
Lysosomes and Peroxisomes

- Often referred to as the garbage disposal system of a cell.
- Spherical, bound by membrane, and rich in digestive enzymes.
- Peroxisomes often resemble a lysosome, but are self replicating, whereas lysosomes are formed in the Golgi complex.
- Lysosomes contain three dozen enzymes for degrading proteins, nucleic acids, and certain sugars called polysaccharides.
- These enzymes work best at low pH, reducing risk that they will digest their own cell if they escape from the lysosome.
- The cell could not house such destructive enzymes if they were not contained in a membrane-bound system.
- Lysosome can digest foreign bacteria that invade a cell.
Lysosomes and Peroxisomes (cont)

- They also recycle receptor proteins and other membrane components and degrade worn out organelles.
- They can even help repair damage to the plasma membrane by serving as a membrane patch, sealing the wound.
- Peroxisomes rid body of toxic substances, such as hydrogen peroxide, and contain enzymes for oxygen utilization.
- High numbers of peroxisomes can be found in the liver.
- All enzymes in them are imported from cytosol and have a special sequence, called a PTS or peroxisomal targeting signal.
- They also have membrane proteins for importing proteins into their interiors and to replicate.
Genetic Circuits

- Genes encoded in DNA used as templates to synthesize mRNA through the process of transcription.
- Genes include *coding sequences* and *regulatory sequences*.
- Regulatory sequences can bind to other proteins which in turn either activate or repress transcription.
- Transcription is also regulated through *post-transcriptional modifications*, *DNA folding*, and other feedback mechanisms.
- This regulatory network increases an organism’s complexity.
- Behavior analogous to electrical circuits in which multiple inputs are processed to determine multiple outputs.
- Therefore, these regulatory networks known as *genetic circuits*. 
Transcription

(Courtesy: National Human Genome Research Institute)
Transcription

- Initiated at *promoter site* by *RNA polymerase* (RNAP).
- Promoter is a unidirectional sequence found on one strand which instructs RNAP where to start and in which direction.
- RNAP unwinds double helix at that point and begins synthesis of an mRNA complementary to one of the strands of DNA.
- This strand is called the *antisense* or *template* strand, whereas the other strand is referred to as the *sense* or *coding* strand.
- Synthesis proceeds in a unidirectional manner.
- Terminates when polymerase stumbles upon a stop signal.
- In eukaryotes, not fully understood, but prokaryotes have short region of G’s and C’s that folds in on itself causing polymerase to trip and release the *nascent*, or newly formed, mRNA.
Regulation of Transcription

- Ability of RNAP to bind to promoter site can be either enhanced or precluded by *transcription factors*.
- They recognize portions of the DNA sequence near the promoter region known as *operator sites*.
- Those that help RNAP bind are *activators* and those that block RNAP from binding are *repressors*.
- These sequences can be *cis-acting* (affecting adjacent genes), or *trans-acting* (affecting distant genes).
Genetic Circuit Example

DNA

\[ P_{RE} \]

\[ O_E \quad O_R \]

\[ P_R \]
Genetic Circuit Example

DNA

$P_{RE}$

$O_E$

$O_R$

$P_R$

cl

cll

Genes

Chris J. Myers (Lecture 1: Engineers Guide)
Genetic Circuit Example
Genetic Circuit Example
Genetic Circuit Example

DNA

$P_{RE}$

OE

OR

RNAP

cl

cll

Promoters

Genes

Transcription

Operator Sites

Promoters

Genes

RNAP
Genetic Circuit Example

DNA

Promoters

Genes

cl

cll

\( P_{RE} \)

\( O_E \)

\( O_R \)

\( P_R \)

mRNA
Genetic Circuit Example

DNA

- Operator Sites
- Promoters
- Genes

ClI Protein

- Translation
- mRNA

- Transcription

- Activation

CI Protein

- mRNA

- Translation

ClI Protein

- Transcription

- Activation
Genetic Circuit Example

DNA

cl

Operator Sites

Promoters

Genes

CI Protein

ClII Protein

Translation

mRNA

Activation

Transcription

Pr

Pr

Pp

OE OR
Genetic Circuit Example

- Dimerization
- Repression
- Translation
- mRNA

- CI Dimer
- CI Protein
- CII Protein
- DNA
- Operator Sites
- Promoters
- Genes
- cl
- cII

- Activation
- Transcription
- mRNA

- **Promoters**
- **Genes**
- **Operator Sites**

- **CI Protein**

- **CI Dimer**

- **DNA**

- **mRNA**

- **P\text{\textsubscript{RE}}**
- **P\text{\textsubscript{R}}**

- **O\text{\textsubscript{E}}**
- **O\text{\textsubscript{R}}**
Genetic Circuit Example

- **Dimerization**
- **Repression**
- **Activation**
- **Degradation**

**DNA**

**Operator Sites**

**Promoters**

**Genes**

- **cl**
- **cII**

- **P_{RE}**
- **P_{R}**

- **CI Dimer**
- **CI Protein**
- **CII Protein**
Genetic Circuit Example

CI Dimer
Repression

DNA

cl

Promoters
Genes

Operator Sites

P_{RE} O_{E} O_{R} P_{R}
Genetic Circuit Example

- **RNAP**: RNA polymerase
- **CI Dimer**: Repression
- **Activation**: CI Protein
- **DNA**
- **Operator Sites**: $O_E$, $O_R$
- **Promoters**: $P_{RE}$, $P_R$
- **Genes**: cl, cII
- **mRNA**: Translation
- **cI cII**: Gene expression

Diagram showing the interaction between DNA, promoters, genes, and proteins in a genetic circuit.
Methylation

- Transcription also regulated by variations in DNA structure.
- One chemical modification of DNA, called *methylation*, involves the addition of a *methyl group* (-CH3).
- Methylation frequently occurs at cytosine residues preceded by guanine bases, often in vicinity of promoter sequences.
- Inhibits transcription by attracting a protein that binds to methylated DNA, interfering with polymerase binding.
- The methylation status of DNA often correlates with its functional activity.
Translation

(Courtesy: National Human Genome Research Institute)
Translation

- **Ribosome** has two subunits.
- **Small subunit** finds mRNA to begin translating.
- **Large subunit** has two sites for amino acids to bind.
- **A site** accepts transfer RNA (tRNA) bearing an amino acid.
- **P site** binds the tRNA to the growing chain.
- Each tRNA has a specific **acceptor site** that binds a particular triplet of nucleotides, called a **codon**,
- Also has an **anti-codon site** that binds a sequence of three unpaired nucleotides, the **anti-codon**, which binds to the codon.
- Also has specific **charger protein** that only binds to a specific tRNA and attaches correct amino acid to the acceptor site.
Start signal is the codon ATG that codes for methionine.

A tRNA charged with methionine binds to the start signal.

Large subunit binds to the mRNA and the small subunit, and so begins elongation, the formation of the polypeptide chain.

After the first charged tRNA appears in the A site, the ribosome shifts so that the tRNA is now in the P site.

New tRNAs, corresponding to codons of the mRNA, enter the A site, and a bond is formed between the two amino acids.

The first tRNA is now released, and the ribosome shifts again so that a tRNA carrying two amino acids is now in the P site.
This continues until the ribosome reaches a *stop codon*.

Ribosome breaks apart releasing the mRNA and new protein.

Note that a protein will often undergo further modification, called *post-translational modification*.

*Translational regulation* occurs through the binding of *repressor proteins* to a sequence found on an RNA molecule.

Translational control plays a significant role in the process of embryonic development and cell differentiation.
Where Do Viruses Fit?

- They are not classified as cells and therefore are neither unicellular nor multicellular organisms.
- They are not “living” because they lack a metabolic system and are dependent on host cells they infect to reproduce.
- They have genomes of either DNA or RNA which are either double-stranded or single-stranded.
- Their genomes code for both proteins to package its genetic material and those needed to reproduce.
Because viruses are acellular, they must utilize the machinery and metabolism of a host cell to reproduce.

For this reason, known as *obligate intracellular parasites*.

Before entering host, it is a *virion*—package of genetic material.

They can be passed from host to host either through direct contact or through a vector, or carrier.

*Bacteriophages* attach to the cell wall surface, make a small hole, and inject their DNA into the cell.

Others (such as HIV) enter the host via *endocytosis*, the process in which a cell takes in material from its environment.

After entering the cell, its genetic material takes over cell and forces it to produce new viruses.
Types of Viruses

(Courtesy: National Center for Biotechnology Information)
The form of genetic material contained in the *viral capsid* determines the exact replication process.

- If it has DNA, it is replicated by the host along with its own DNA.
- If it has RNA, it is copied using **RNA replicase** making a template to produce 100s of duplicates of the original RNA.
- **Retroviruses** use **reverse transcriptase** to synthesize a complementary strand of DNA which is then replicated using the host cell machinery.
Steps Associated with Viral Reproduction

1. **Attachment**, sometimes called *absorption*: The virus attaches to receptors on the host cell wall.

2. **Penetration**: Viral genome moves through plasma membrane into host, capsid of a *phage* remains outside.

3. **Replication**: Virus induces host to synthesize the necessary components for its replication.

4. **Assembly**: The newly synthesized viral components are assembled into new viruses.

5. **Release**: Assembled viruses are released from the cell and can now infect other cells, and the process begins again.
When the virus takes over the cell, it directs host to manufacture the proteins necessary for virus reproduction.

The host produces three kinds of proteins:

- *Early proteins*, enzymes used in nucleic acid replication;
- *Late proteins*, proteins used to construct the virus coat; and
- *Lytic proteins*, enzymes used to break open the cell for exit.

Viral product is assembled spontaneously, that is, the parts are made separately by the host and are joined together by chance.

Self-assembly often aided by *molecular chaperones*, or proteins made by the host that help the capsid parts come together.
New viruses leave the cell either by *exocytosis* or by *lysis*.

Animal viruses instruct host’s endoplasmic reticulum to make *glycoproteins* which collect in clumps along the cell membrane. Virus then discharged at these exit sites.

Bacteriophages must break open, or *lyse*, the cell to exit.

They have a gene that codes for an enzyme called *lysozyme*.

This breaks down cell wall, causing the cell to swell and burst.

New viruses released into the environment, killing the host.
In 1953, Lwoff et al. discovered that a strain of *E. Coli* when exposed to UV light *lyse* spewing forth *λ* viruses.

Some of the newly infected *E. Coli* would soon lyse while others grow and divide normally until exposed to UV light.

In other words, some cells follow a *lysis* pathway while other followed a *lysogeny* pathway.

The decision between the lysis and the lysogeny developmental pathway is made by a fairly simple genetic circuit.
Phage λ

(Courtesy of Maria Schnos and Ross Inman, Institute for Molecular Virology, University of Wisconsin, Madison)
E. coli bacterial cell

Host chromosome

Phage λ

Attachment

Penetration

Lysogeny

Lysis

Replication

Assembly

Release

Cell division

Lysogeny Pathway

Induction event

Lysis Pathway

Phage λ Developmental Pathways
The $O_R$ Operator

\[ P_{RM} \quad cl \quad O_{R3} \quad O_{R2} \quad O_{R1} \quad P_R \quad cro \]
CI, the $\lambda$ Repressor

The figure illustrates the interaction between CI, CI$_2$, $cI$, cro, and three operators ($O_{R1}$, $O_{R2}$, $O_{R3}$). The repressor CI binds to the operator sites, modulating gene expression. The binding of CI to the operator sites leads to repression of the adjacent promoters ($P_{RM}$ and $P_R$).
\(\lambda\)'s Cro Molecule

\[ P_{RM} \quad cI \quad O_{R3} \quad O_{R2} \quad O_{R1} \quad cro \]

Cro\textsubscript{2} and Cro

\[ P_{R} \]
Cl$_2$ Bound to $O_{R1}$ Turns Off $P_R$
Cl$_2$ Bound to $O_{R2}$ Turns On $P_{RM}$
Nothing in biology is clear-cut.

Without Cl_2 bound to OR2, RNAP can still bind to P_{RM} and initiate transcription of CI at a reduced *basal rate*.

With Cl_2 bound to OR2, transcription occurs at enhanced *activated rate*.
Cl$_2$ Bound to $O_{R3}$ Turns Off $P_{RM}$
$P_R$ Active When $O_R$ Sites Are Empty
Cro_2 Bound to OR3 Turns off P_{RM}
Cro₂ Bound to \( O_{R1} \) or \( O_{R2} \) Turns off \( P_R \)
Low Concentrations of Cl$_2$
Cooperativity Aids Cl₂ Binding to OR2

![Diagram showing cooperativity in binding of Cl₂ to OR2](image)

- **PRM**: Promoter for OR3
- **OR3**: Operator site for OR3
- **OR2**: Operator site for OR2
- **OR1**: Operator site for OR1
- **cro**: Operator site for cro
High Concentrations of \( \text{Cl}_2 \)
Low Concentrations of Cro₂

Diagram showing genetic regulatory elements with promoters and operators.
Moderate Concentrations of Cro$_2$
High Concentrations of Cro$_2$
Cooperativity of Cl$_2$ Binding

The image shows a genetic regulatory circuit with promoters $P_{RM}$ and $P_R$, and various operators $OR_1$, $OR_2$, $OR_3$, and the genes $cl$ and $cro$. The circuit illustrates how cooperativity affects the binding of regulatory proteins to these elements, potentially influencing gene expression.
Cooperativity of CI₂ Binding

\[ P_{RM} \quad cI \quad O_{R3} \quad O_{R2} \quad O_{R1} \quad cro \quad P_R \]
Cooperativity of CI<sub>2</sub> Binding

The diagram illustrates the binding of RNAP (RNA polymerase) to various regulatory elements, including cl, OR3, OR2, OR1, and cro. The cooperativity of CI<sub>2</sub> binding is shown by the interaction between RNAP and the different promotors (PRM and PR) along the DNA sequence.
Cooperativity of CI$_2$ Binding
At low to moderate concentrations of Cl and Cro, there are three common configurations:

- No molecules bound to $O_R$, Cro produced at full rate and Cl produced at low basal rate.
- Cl$_2$ bound to $O_R$1 and $O_R$2, Cro production repressed, and Cl activated.
- Cro$_2$ bound to $O_R$3, Cl cannot be produced, Cro is produced.

Feedback of the products as transcription factors coupled with affinities makes $O_R$ behave as a bistable switch.

- In *lysis* state, Cro produced locking out production of Cl.
- In *lysogeny* state, Cl produced locking out production of Cro.
In lysogeny state, cell is immune to further infection.

cro genes on DNA inserted by further infections are shut off by Cl$_2$ molecules from first infection.

Once cell commits to lysogeny, it becomes very stable and does not easily change over to the lysis pathway.
Induction

UV Light

Activated RecA

RecA

$P_{RM}$

$P_R$

$cl$  $O_{R3}$  $O_{R2}$  $O_{R1}$  $cro$
Cl$_2$ and Cro$_2$ bind to operator sites that are 17 base pairs long.

How do these proteins locate these sequences from amongst the millions within the bacteria?

Observing from midpoint, a strand on one side is nearly symmetric with complimentary strand on other side.
Near Symmetry in the Operator Sequences

<table>
<thead>
<tr>
<th>Op</th>
<th>Operator sequences</th>
<th>Operator half sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>$O_{R1}$</td>
<td>$TATCA_CCG_cCAGAGGTAT$</td>
<td>$TATCA_CCG_c$</td>
</tr>
<tr>
<td></td>
<td>$ATAAGTGCGG_TCTCCTCAT$</td>
<td>$TATCA_CCG_c$</td>
</tr>
<tr>
<td>$O_{R2}$</td>
<td>$TAACACCG_tGGTGGTGG$</td>
<td>$TAACACCG_t$</td>
</tr>
<tr>
<td></td>
<td>$ATAATGGCGaCCGACCACAC$</td>
<td>$CAAACACGCG$</td>
</tr>
<tr>
<td>$O_{R3}$</td>
<td>$TATCA_CCG_cAAGGGAATA$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td></td>
<td>$ATAAGTGCGG_TCTCCTCAT$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td>$O_{L1}$</td>
<td>$TATCA_CCG_cCAGTGGTA$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td></td>
<td>$ATAATGGCGG_TCTCCTCAT$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td>$O_{L2}$</td>
<td>$TATCTCTCTGGC_GGTGGTGG$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td></td>
<td>$ATAAGAGACcGCCACACAC$</td>
<td>$CAAACACGCG$</td>
</tr>
<tr>
<td>$O_{L3}$</td>
<td>$TATCA_CCG_cAGATGGTT$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td></td>
<td>$ATAATGGCGG_TCTCCTCAT$</td>
<td>$TAACACCG_c$</td>
</tr>
<tr>
<td>Con.</td>
<td></td>
<td>$T_9 \ A_{12} \ T_6 \ C_{12} \ A_9 \ C_{11} \ C_7 \ G_9 \ C_5$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$C_2 \ C_3 \ T_2 \ T_1 \ T_4 \ T_2 \ T_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$A_1 \ A_3 \ C_1 \ G_1 \ C_1$</td>
</tr>
</tbody>
</table>
The consensus sequence is as follows:

\[
\text{TATCACCGcCGGTGATA}
\]
\[
\text{ATAGTGGCgGCCACTAT}
\]

Many entries are highly preserved.

Differences exist that cause the differences in affinity for Cl$_2$ and Cro$_2$ for the different operators.

Notice that the first half of the operator sites $O_{R1}$ and $O_{R3}$ agree perfectly with the consensus sequence while second half has several differences.
Amino Acid-Base Pair Interactions

**CI$_2$**

- gln
- gly
- gly
- leu
- asn
- ser
- ala
- phe

**Cro$_2$**

- gln
- ala
- asp
- ala
- his
- ser
- lys
- ile
- ile

**OR$_1$**

- T
- A
- C
- C
- T
- C
- T
- G

- A
- T
- G
- G
- A
- G
- A
- C

**OR$_2$**

- C
- A
- A
- C
- A
- C
- G
- C
- G
- T
- T
- G
- T
- G
- C
- G
- T

**OR$_3$**

- T
- A
- T
- C
- C
- C
- T
- T

- A
- T
- A
- G
- G
- G
- A
- A
- C
\[ \lambda \text{ Promoters} \]

<table>
<thead>
<tr>
<th></th>
<th>-35</th>
<th>-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>T T G A C A</td>
<td>T A T A A T</td>
</tr>
<tr>
<td>( \lambda P_{RM} )</td>
<td>T A G A T A</td>
<td>T A G A T T</td>
</tr>
<tr>
<td>( \lambda P_R )</td>
<td>T T G A C T</td>
<td>G A T A A T</td>
</tr>
</tbody>
</table>
The \( \lambda \) Genome

(Bacteriophage \( \lambda \) Genome)

(Courtesy of Richard Wheeler)
Patterns of Gene Expression

- Late Lysogeny
- Late Lysis
- Early
- Inactive CII
- Active CII
- Very Early
Very Early Events
The Action of N RNAP
The Action of N

- "PNUT_L\(\text{RNAP}\)NUT_L\(N\)TL1\(\text{cll}\)xis\(\text{int}\)"
The Action of N

\[ P_L \rightarrow N \rightarrow T_{L1} \rightarrow clll \rightarrow xis \rightarrow int \]
The Action of N

[Diagram showing a biological process]
The Action of N

The diagram illustrates the action of N in the context of genetic circuits. It shows the interaction between RNAP (RNA polymerase), NUTL, N, T_{L1}, cIII, xis, and int.
The Action of N

\[ P_L \rightarrow NUT_L \rightarrow N \rightarrow T_{L1} \rightarrow cIII \rightarrow xis \rightarrow int \rightarrow RNAP \]
Early Events

Early
Retroregulation of Int

RNase3

sib  int  xis  cIII N
Retroregulation of Int

int xis cIII N
Retroregulation of Int

xis  cIII  N
Lysis/Lysogeny Decision

- Key to lysis/lysogeny decision is the protein CII.
- CII activates $P_{RE}$ which jump-starts production of CI.
- After jump-start, positive feedback in $P_{RM}$ increases CI and locks out Cro production resulting in lysogeny decision.
- Activity of CII is determined by environmental factors.
- Bacterial proteases attack and destroy CII.
- Growth in a rich medium activates these proteases whereas starvation has the opposite effect.
- Thus, λ tends to lysogenize starved cells.
- CIII protein protects CII from degradation promoting lysogeny.
- Production of CII and CIII enhanced by anti-terminator, N.
- Higher multiplicity of infection means more N, cII, and cIII gene copies leading to higher probability of lysogeny.
High protease levels and low gene counts
⇒ CII produced slowly and degrades rapidly
⇒ little CI is synthesized
⇒ Q and Cro are synthesized
⇒ lysis decision

Low protease levels and high gene counts
⇒ more N, CII, and CIII are produced
⇒ CI and Int are made from $P_{RE}$ and $P_I$
⇒ Int integrates phage chromosome and CI turns off all genes except CI
⇒ lysogeny decision
Late Lytic Events

Late Lysis
Late Lysogeny
Integration and Induction

attP

attB

int xis cIII N cI sib
Create a species for RNAP as well as for each promoter and protein.
Create degradation reactions for each protein.
Create open complex formation reactions for each promoter.
Create dimerization reactions, if needed.
Create repression reactions for each repressor.
Create activation reactions for each activator.
Degradation Reactions

\[
\text{CI} \xrightarrow{r} \text{Cl} \quad \text{CII} \xrightarrow{r} \text{CII}
\]

\[
k_d [\text{CI}] \quad k_d [\text{CII}]
\]

\[
\text{Cl} \xrightarrow{k_d} (\quad) \\
\text{CII} \xrightarrow{k_d} (\quad)
\]

\[
\text{Constant Value} = \frac{k_d}{0.0075 \text{ sec}^{-1}}
\]
Open Complex Formation Reactions

\[ P_{RE} + \text{RNAP} \xleftrightarrow{K_{o1}} S_1 \]
\[ P_R + \text{RNAP} \xleftrightarrow{K_{o2}} S_2 \]
\[ S_1 \xrightarrow{k_b} S_1 + np \text{ CI} \]
\[ S_2 \xrightarrow{k_o} S_2 + np \text{ CII} \]

*Constant* | *Value*
---|---
\[ \text{RNAP}_0 \] | 30 nM
\[ K_{o1} \] | 0.01 M\(^{-1}\)
\[ K_{o2} \] | 0.69422 M\(^{-1}\)
\[ k_b \] | 0.00004 sec\(^{-1}\)
\[ k_o \] | 0.014 sec\(^{-1}\)
\[ np \] | 10
Dimerization Reactions

\[
\begin{align*}
2\text{CI} & \underset{\text{2, r}}{\overset{\text{K}_d}{\rightarrow}} \text{Cl}_2 \\
K_d[\text{Cl}]^2 - [\text{Cl}_2] & \overset{p}{\underset{\text{K}_d}{\rightarrow}} \text{Cl}_2 \\
\text{Constant Value} & = \frac{K_d}{0.1 M^{-1}}
\end{align*}
\]
Repression Reactions

\[ PR \rightarrow K_r [P_R][CI_2]^{nc} - [S_3] \]

\[ P_R + nc CI_2 \quad \xrightarrow{K_r} \quad S3 \]

**Constant**

| \( K_r \) | 0.2165 M\(^{-nc}\) |
| \( nc \) | 1 |

Chris J. Myers (Lecture 1: Engineers Guide)
Activation Reactions

\[
P_{RE} + na \text{ CII} + \text{RNAP} \xleftarrow{K_a} S4 \\
S4 \xrightarrow{k_a} S4 + np \text{ Cl}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(K_a)</td>
<td>0.00161 (M^{-(na+1)})</td>
</tr>
<tr>
<td>(k_a)</td>
<td>0.015 sec(^{-1})</td>
</tr>
<tr>
<td>(na)</td>
<td>1</td>
</tr>
<tr>
<td>(np)</td>
<td>10</td>
</tr>
</tbody>
</table>
Complete Reaction-Based Model

- $P_{RE}$
- RNAP
- $P_R$
- $S_3$
- $S_1$
- $S_2$
- $S_4$
- $CI$
- $CI_2$
- $CII$
- $r_2$
- $r_3$
- $r_4$
- $r_5$
- $r_6$
- $r_7$
- $r_8$
- $r_9$
- $r_{10}$
- $r_1$

Reactions:
- $na,r$
- $np,p$
- $m$
- $nc,r$
- $2,r$

Connections:
- $r$
- $p$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
- $r$
Why Study Phage λ?

- Bacteria and their phages multiply quickly making it easier to analyze gene regulation with them than higher organisms.
- Phage λ has been the subject of study for over 50 years now.
- It is one of, if not the best, understood genetic circuit.
- Excellent illustration of a circuit that analyzes its environment makes decision between two competing pathways.
- Similarities with bacteria that must respond to stress and circuits involved in development and cell differentiation.
- Genes from phage λ are used in synthetic biology where DNA is produced to perform particular functions.
- Phage λ is an excellent testbed for trying new ideas.
- Virtually every modeling method has been applied to phage λ.
- This course also uses it as a running example throughout.
Tozeren/Byers - “New Biology for Eng. and Comp. Scientists”.
Gonick/Wheelis - “The Cartoon Guide to Genetics”.
King/Stansfield - “A Dictionary of Genetics”.
Berg/Tymoczko/Stryer - “Biochemistry”.
Watson et al. - “Molecular Biology of the Gene”.
Alberts et al. - “Molecular Biology of the Cell”.

Phage \( \lambda \) first isolated by Esther Lederberg (1951).

Excellent history in paper by Gottesman and Weisberg (2004).

Discovery of UV induction by Lwoff and Gutmann (1950).

Discovery of the genetic switch by Lwoff, Jacob, and Monod (1961).

Mark Ptashne - “A Genetic Switch”.

Model inspired by Arkin et al.’s phage \( \lambda \) model (1998).
Variety of ways to model a system using chemical reactions.

This model includes:

- Five genes: \( cl, \ cro, \ N, \ cI{\text{I}}, \) and \( cIII. \)
- Four promoters: \( P_{RM}, P_{R}, P_{RE}, \) and \( P_{L}. \)

Model somewhat simplified (see appendix at end of Chapter 2).
Phage λ Decision Circuit

![Diagram of the Phage λ Decision Circuit]

- **cl**
  - **P$_{RM}$**
  - **CI**
  - **CI$_2$**
  - **Cro$_2$**
  - **Cro**

- **cro**
  - **P$_R$**
  - **T$_{R1}$ 50%**
  - **P$_{RE}$**
  - **NUT$_R$**
  - **O$_{R3}$**
  - **O$_{R2}$**
  - **O$_{R1}$**

- **OL1**
  - **OL2**
  - **P$_L$**
  - **CI$_2$**
  - **Cro$_2$**

- **n**
  - **T$_{L1}$ 80%**
  - **NUT$_L$**

- **OE1**
  - **OE2**
  - **CII**

- **clII**
  - **deg**
  - **CIII**

- **clIII**
  - **P$_R$**
  - **T$_L$ 80%**

- **O$_{L1}$**
  - **O$_{L2}$**

Chris J. Myers (Lecture 1: Engineers Guide) Engineering Genetic Circuits 130 / 139
Model of the Promoter $P_{RE}$

\[
\begin{align*}
P_{RE} + \text{RNAP} & \quad \rightleftharpoons \quad K_{PRE2} \quad P_{RE} \cdot \text{RNAP} \\
P_{RE} + \text{CII} & \quad \rightleftharpoons \quad K_{PRE3} \quad P_{RE} \cdot \text{CII} \\
P_{RE} + \text{CII} + \text{RNAP} & \quad \rightleftharpoons \quad K_{PRE4} \quad P_{RE} \cdot \text{CII} \cdot \text{RNAP} \\
P_{RE} \cdot \text{RNAP} & \quad \rightarrow \quad k_{PREb} \quad P_{RE} \cdot \text{RNAP} + n\text{Cl} \\
P_{RE} \cdot \text{CII} \cdot \text{RNAP} & \quad \rightarrow \quad k_{PRE} \quad P_{RE} \cdot \text{CII} \cdot \text{RNAP} + n\text{Cl}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{PRE2}$</td>
<td>0.01 $M^{-1}$</td>
</tr>
<tr>
<td>$K_{PRE3}$</td>
<td>0.00726 $M^{-1}$</td>
</tr>
<tr>
<td>$K_{PRE4}$</td>
<td>0.00161 $M^{-1}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{PREb}$</td>
<td>0.00004 sec$^{-1}$</td>
</tr>
<tr>
<td>$k_{PRE}$</td>
<td>0.015 sec$^{-1}$</td>
</tr>
<tr>
<td>$n$</td>
<td>10</td>
</tr>
</tbody>
</table>
Model of the $O_R$ Operator

- Modeling method just described requires $(n - 1)$ chemical reactions where $n$ is number of potential configurations of transcription factors and RNAP bound to the operator and promoter sites.
- Also requires $m$ reactions for configurations leading to transcription.
- $O_R$ operator has 40 possible configurations with 13 leading to transcription.
  - $O_R3$ has four states (empty, $Cl_2$, $Cro_2$, and RNAP).
  - $O_R2$ and $O_R1$ also can bind all transcription factors but bind to RNAP jointly (i.e., 10 possibilities).
Simplified Model of the $O_R$ Operator (1 Site Occupied)

$$O_R + \text{Cl}_2 \overset{K_{OR2}}{\rightleftharpoons} O_R \cdot \text{Cl}_2$$

$$O_R + \text{Cro}_2 \overset{K_{OR5}}{\rightleftharpoons} O_R \cdot \text{Cro}_2$$

$$O_R + \text{RNAP} \overset{K_{OR8}}{\rightleftharpoons} O_R3 \cdot \text{RNAP}$$

$$O_R + \text{RNAP} \overset{K_{OR9}}{\rightleftharpoons} O_R12 \cdot \text{RNAP}$$

$$O_R3 \cdot \text{RNAP} \overset{k_{PRMb}}{\rightarrow} O_R3 \cdot \text{RNAP} + n\text{Cl}$$

$$O_R12 \cdot \text{RNAP} \overset{k_{PR}}{\rightarrow} O_R12 \cdot \text{RNAP} + n\text{Cro}$$

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{OR2}$</td>
<td>0.2165 $M^{-1}$</td>
<td>$k_{PRMb}$</td>
<td>0.001 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR5}$</td>
<td>0.449 $M^{-1}$</td>
<td>$k_{PRM}$</td>
<td>0.011 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR8}$</td>
<td>0.1362 $M^{-1}$</td>
<td>$k_{PR}$</td>
<td>0.014 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR9}$</td>
<td>0.69422 $M^{-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simplified Model of the $O_R$ Operator (2 Sites Occupied)

\[
\begin{align*}
O_R + 2\text{Cl}_2 & \quad \Longleftrightarrow \quad K_{OR10} \quad O_R \cdot 2\text{Cl}_2 \\
O_R + \text{Cl}_2 + \text{Cro}_2 & \quad \Longleftrightarrow \quad K_{OR17} \quad O_R \cdot \text{Cl}_2 \cdot \text{Cro}_2 \\
O_R + 2\text{RNAP} & \quad \Longleftrightarrow \quad K_{OR16} \quad O_R \cdot 2\text{RNAP} \\
O_R + \text{Cro}_2 + \text{RNAP} & \quad \Longleftrightarrow \quad K_{OR26} \quad O_R \cdot \text{Cro}_2 \cdot \text{RNAP} \\
O_R \cdot 2\text{RNAP} & \quad \xrightarrow{k_{PRM_b}} \quad O_R \cdot 2\text{RNAP} + n\text{Cl} \\
O_R \cdot 2\text{RNAP} & \quad \xrightarrow{k_{PR}} \quad O_R \cdot 2\text{RNAP} + n\text{Cro} \\
O_R \cdot \text{Cro}_2 \cdot \text{RNAP} & \quad \xrightarrow{k_{PR}} \quad O_R \cdot \text{Cro}_2 \cdot \text{RNAP} + n\text{Cro}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{OR10}$</td>
<td>0.06568 $M^{-1}$</td>
<td>$k_{PRM_b}$</td>
<td>0.001 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR16}$</td>
<td>0.09455 $M^{-1}$</td>
<td>$k_{PRM}$</td>
<td>0.011 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR17}$</td>
<td>0.1779 $M^{-1}$</td>
<td>$k_{PR}$</td>
<td>0.014 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR26}$</td>
<td>0.25123 $M^{-1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Simplified Model of the $O_R$ Operator (3 Sites Occupied)

\[
\begin{align*}
O_R + 2\text{Cl}_2 + \text{Cro}_2 & \quad \xrightleftharpoons{K_{OR31}} \quad O_R \cdot 2\text{Cl}_2 \cdot \text{Cro}_2 \\
O_R + \text{RNAP} + 2\text{Cl}_2 & \quad \xrightleftharpoons{K_{OR37}} \quad O_R \cdot \text{RNAP} \cdot 2\text{Cl}_2 \\
O_R \cdot \text{RNAP} \cdot 2\text{Cl}_2 & \quad \xrightarrow{k_{PRM}} \quad O_R \cdot \text{RNAP} \cdot 2\text{Cl}_2 + n\text{Cl}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{OR31}$</td>
<td>0.02133 $M^{-1}$</td>
<td>$k_{PRM}$</td>
<td>0.011 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{OR37}$</td>
<td>0.0079 $M^{-1}$</td>
<td>$k_{PRMb}$</td>
<td>0.001 sec$^{-1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_{PR}$</td>
<td>0.014 sec$^{-1}$</td>
</tr>
</tbody>
</table>
CI and Cro Dimerazation and Degradation

\[
\begin{align*}
2\text{Cl} & \xrightleftharpoons{K_2} \text{Cl}_2 \\
2\text{Cro} & \xrightleftharpoons{K_5} \text{Cro}_2 \\
\text{Cl} & \xrightarrow{k_1} () \\
\text{Cro} & \xrightarrow{k_4} () \\
\end{align*}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(k_1)</td>
<td>0.0007 sec(^{-1})</td>
</tr>
<tr>
<td>(K_2)</td>
<td>0.1 M(^{-1})</td>
</tr>
<tr>
<td>(k_4)</td>
<td>0.0025 sec(^{-1})</td>
</tr>
<tr>
<td>(K_5)</td>
<td>0.1 M(^{-1})</td>
</tr>
</tbody>
</table>
Production from promoter $P_L$ and $N$ Degradation

$$P_L + \text{Cro}_2 \quad \overset{K_{PL2}}{\longleftrightarrow} \quad P_L \cdot \text{Cro}_2$$

$$P_L + \text{Cl}_2 \quad \overset{K_{PL4}}{\longleftrightarrow} \quad P_L \cdot \text{Cl}_2$$

$$P_L + 2\text{Cro}_2 \quad \overset{K_{PL7}}{\longleftrightarrow} \quad P_L \cdot 2\text{Cro}_2$$

$$P_L + \text{Cl}_2 + \text{Cro}_2 \quad \overset{K_{PL8}}{\longleftrightarrow} \quad P_L \cdot \text{Cl}_2 \cdot \text{Cro}_2$$

$$P_L + 2\text{Cl}_2 \quad \overset{K_{PL10}}{\longleftrightarrow} \quad P_L \cdot 2\text{Cl}_2$$

$$P_L + \text{RNAP} \quad \overset{K_{PL6}}{\longleftrightarrow} \quad P_L \cdot \text{RNAP}$$

$$P_L \cdot \text{RNAP} \quad \overset{k_{PL}}{\rightarrow} \quad P_L \cdot \text{RNAP} + nN$$

$$N \quad \overset{k_7}{\rightarrow} \quad ()$$

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{PL2}$</td>
<td>0.4132 M$^{-1}$</td>
<td>$K_{PL8}$</td>
<td>0.014 M$^{-1}$</td>
</tr>
<tr>
<td>$K_{PL4}$</td>
<td>0.2025 M$^{-1}$</td>
<td>$K_{PL10}$</td>
<td>0.058 M$^{-1}$</td>
</tr>
<tr>
<td>$K_{PL6}$</td>
<td>0.6942 M$^{-1}$</td>
<td>$k_{PL}$</td>
<td>0.011 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{PL7}$</td>
<td>0.0158 M$^{-1}$</td>
<td>$k_7$</td>
<td>0.00231 sec$^{-1}$</td>
</tr>
</tbody>
</table>
CIII and CII Production

\[
\begin{align*}
\text{NUT}_L + N & \quad \xrightleftharpoons{K_{\text{NUT}}} \quad \text{NUT}_L \cdot N \\
P_L \cdot \text{RNAP} + \text{NUT}_L & \quad \xrightarrow{0.2 \times k_{PL}} \quad P_L \cdot \text{RNAP} + \text{NUT}_L + n\text{CIII} \\
\text{NUT}_R + N & \quad \xrightleftharpoons{k_{PL}} \quad \text{NUT}_R \cdot N \\
O_{R12} \cdot \text{RNAP} + \text{NUT}_R & \quad \xrightarrow{0.5 \times k_{PR}} \quad O_{R12} \cdot \text{RNAP} + \text{NUT}_R + n\text{CII} \\
O_{R12} \cdot \text{RNAP} + \text{NUT}_R \cdot N & \quad \xrightarrow{k_{PR}} \quad O_{R12} \cdot \text{RNAP} + \text{NUT}_R \cdot N + n\text{CII}
\end{align*}
\]

Note that \( K_{\text{NUT}} \) is 0.2 \( M^{-1} \).
Model for CII and CIII Degradation

\[
\begin{align*}
\text{CII} + \text{P1} & \quad \underset{K_8}{\overset{k_{10}}{\rightleftharpoons}} \quad \text{CII} \cdot \text{P1} \\
\text{CII} \cdot \text{P1} & \quad \xrightarrow{k_{11}} \quad \text{P1} \\
\text{CIII} + \text{P1} & \quad \underset{K_{11}}{\overset{k_{13}}{\rightleftharpoons}} \quad \text{CIII} \cdot \text{P1} \\
\text{CIII} \cdot \text{P1} & \quad \xrightarrow{k_{13}} \quad \text{P1}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Constant</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_8$</td>
<td>1.0 $M^{-1}$</td>
</tr>
<tr>
<td>$k_{10}$</td>
<td>0.002 sec$^{-1}$</td>
</tr>
<tr>
<td>$K_{11}$</td>
<td>10.0 $M^{-1}$</td>
</tr>
<tr>
<td>$k_{13}$</td>
<td>0.0001 sec$^{-1}$</td>
</tr>
<tr>
<td>P1</td>
<td>35nM</td>
</tr>
</tbody>
</table>