“Gene Regulation in Prokaryotes”

Lectures 19-21
III. Gene Regulation

Questions

• How to make different proteins at different times?
• How to make different cells that do different things?
• How to make different structures from a single egg?
• And, how to make different species with similar sets of genes
The answers to each of these questions are all founded on the concept of “genetic switches”
Approaches

1. Genetics

   Find mutants for:
   a) Structural proteins that carry out functions
   b) Regulatory proteins that control activity of structural genes
   c) Regulatory elements that act in cis to control gene expression

2. Molecular Biology

   a) Identify and isolate genes
   b) Determine how they are regulated
   c) Determine how their protein products act.
Control of Gene Expression in Bacteria and their Viruses

I. Physiological Mechanisms
   1. The \textit{lac} operon (genetics)
   2. Promoters and repressors
   3. Other operons
   4. Simple life cycles
Physiology

- Cells should make the needed proteins at the right time
- Do not waste energy

**In Bacteria:**
- Often involves nutrient utilization pathway

**In Eukaryotes:**
- May involve the generation of specific proteins in specific types of cells

**The First Model:** Lactose utilization in *E. coli*:

\[
\text{lactose} \xrightarrow{\text{B-galactosidase}} \text{galactose + glucose} \quad (\text{a disaccharide}) \quad \text{H}_2\text{O} \quad \text{(energy)}
\]
Adaptation and Induction

The presence of substrate, lactose, caused the appearance of enzyme, $B$-galactosidase. Was this induction an "adaptation" of the enzyme to substrate just as the bacterium "adapts" to environment?

**Induction**

![Graph showing induction of lac mRNA, B-galactosidease (lacZ), and Permease (lacY) levels over time.](image)

- **lac mRNA**
- **B-galactosidease (lacZ)**
- **Permease (lacY)**

- Time points: lactose added at 0, removed at 5 and 10.
**B-galactosidase (and permease) are inducible**

1. In absence of lactose 1-2 molecules/cell  
   In presence of lactose 100,000 molecules/cell

2. Synthesized nearly simultaneously and only after lac mRNA becomes detectable.

3. Lactose (and analogs) is *inducer*

**How does a cell "KNOW" what to make?**

[The great Lysenko vs. Monod-Jacob divide]  
(Instructive) (Selective)

Monod, Jacob, Lwoff sought to explain induction *GENETICALLY.*
The *lac* operon of *E. coli*.
Key Approach: Mutants and Merozygotes

**Mutants** → To Define Key Genetic Elements

→ To Map Them in Relation to Each Other

**Types of Mutants:**

- *lac Z* - defective enzyme (*B*-galactosidase)
- *lac Y* - defective permease
  
  (trans) *lac I* - make enzyme continuously
  
  (cis) *lac 0c* - make enzyme continuously

"Constitutive"
Merozygotes (Partial Diploids)
- To test gene functions in cis/trans

**Note**
**Gene Order**

**NOTE:** All *lac* Genes Map Very Close Together on Chromosome
Mutations that abolished enzyme activity were *likely* to be **structural** (in the enzyme).

Mutations that resulted in altered inducibility (e.g. constitutive) are **regulatory** and of 2 types:

<table>
<thead>
<tr>
<th>$i^-$</th>
<th>$O^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>trans-acting</td>
<td>cis-dominant</td>
</tr>
</tbody>
</table>

$O^c$ - affects expression only on same chromosome or episome (DNA)

$i^-$ - is recessive to $i^+$, $i^+$ will act on other chromosomes

So, if a merozygote is: $i^+ / i^-$ it is **inducible**

$O^c/O^+$ it is **constitutive**

Using merozygotes, all sorts of cis, trans combinations could be generated and enzyme production measured.
The Operon Model

1. Structural genes: *lac* Z, Y, A (transport & metabolism)
   Regulatory elements: the *lac* I gene - repressor
   the *lac* O operator
   the *lac* P promoter
2. *lac* Z, Y, A in a single mRNA - polycistronic
3. *Promoter* is adjacent to operator (*lac* P − no mRNA)
4. *lac* I protein binds to operator - represses transcription
5. Inducers, e.g. Lactose, bind to and inactivate repressor
The lac Operon

No lactose present

Operon

Structural genes

RNA polymerase

DNA

mRNA

Polypeptide Folding

Repressor protein

Figure 10-6a
Introduction to Genetic Analysis, Ninth Edition
© 2008 W.H. Freeman and Company
The lac Operon

Lactose present

DNA
mRNA
Polypeptide Folding
Repressor protein

RNA polymerase

Operon
Structural genes

Medium
Lactose

β-Galactosidase
Permease
Transacetylase

Figure 10-6b
Introduction to Genetic Analysis, Ninth Edition
© 2008 W.H. Freeman and Company
**Operons**

**General Theme:**

A metabolite controls the expression of a battery of genes that have evolved to utilize it (CATABOLISM)

**Other Examples:**

- trp - tryptophan biosynthesis
- ara - arabinose utilization
- his - histidine biosynthesis

**Think About:**

Genetic logic of *negative* or *positive* control

i. e. repression & activation
Operator mutations act in *cis*

*O*⁺/O⁻ heterozygote

![Diagram showing the action of operator mutations in cis](image)

- Repressor cannot bind to altered operator
- Expression blocked
- Expression even in absence of inducer

*Figure 10-8*

*Introduction to Genetic Analysis, Ninth Edition*

© 2008 W.H. Freeman and Company
Repressor is trans-acting
Repressors that cannot bind lactose are “super-repressors”
Negative Control of the *lac* Operon by the *lacI* Repressor

*lac* repressor is allosteric: it has two different conformations

1. In presence of inducer, it does not bind DNA
2. In absence of inducer, it binds strongly to *lac* operator DNA

*lac I* mutations:

- *lac I* - no repressor function - constitutive
- *lac Is* - superrepressor - not inducible

What about *lac O*c?

Operator mutations that won't be recognized by repressor protein.

Sequence of operator matters

Repressor is sequence - specific
The DNA Base Sequence of the Operator
Positive Control of the \textit{lac} Operon: CAP and Catabolite Repression

Glucose indirectly inhibits \textit{lac} expression

\begin{align*}
\uparrow & \text{glucose} \quad \downarrow \text{\textit{lac}} \\
\text{If glucose is high} & \quad \text{cAMP is low} \\
\text{low} & \quad \text{cAMP is high}
\end{align*}

Remember lactose $\xrightarrow{\text{lacZ}}$ galactose + glucose

If lactose & glucose are present - no \textit{lacZ} is made until glucose is depleted. How?

High cAMP is necessary for activation of \textit{lac} operon.

cAMP is bound by CAP (catabolite activator protein)

\textit{cAMP-CAP} binds to distal part of promoter and facilitates transcription
Glucose levels control the lac operon

(a) Glucose levels regulate cAMP levels

High glucose

ATP → No cAMP

Low glucose

ATP → cAMP

(b) cAMP–CAP complex activates transcription

Complex binds to promoter

Figure 10-13
Introduction to Genetic Analysis, Ninth Edition
© 2008 W.H. Freeman and Company
CAP, RNA POLYMERASE AND REPRESSOR BINDING SITES

E. coli chromosome

CAP site

RNA polymerase interaction site

E. coli DNA

Repessor

Stop codon

Promoter

CAP contact region

mRNA

Operator

fMet Thr Met

ATGCTTCCGCTCGTATGTGTGTGGTTGAAATTTGAGCGGATAACAAATTTTACACAGGAAACAGCTATGAGCATG
TACGAGGCCGAGCATAACAACACACCTTAACACTCGCCTTTGTAAAGTGTTGGCTCTTTGCGATACTGGTAC 5'
The "Logic" of Operon Control and Resource Utilization

- Glucose is normal energy and carbon source.
- Cell has "back-up" system to use lactose (lac).
- Even if lactose is present, it won't bother to make lacZ if glucose is present.
- Even if no glucose is present, operon isn't unduced until lactose appears.
- Lactose inhibits inhibition of lac expression
- Damn, ain't that beautiful?
Some other logical operons

*arabinose utilization:*

*araC* protein has dual role, both positive and negative regulator
+ arabinose, *araC*-ara complex promotes *ara* expression
- arabinose, *araC* keeps operon shut off

*araC* works through alternative binding sites
The dual role of AraC

(a) Activation

CAP + cyclic AMP

AraC protein + arabinose

Active transcription

mRNA

(b) Repression

AraC protein
Some Other Logical Operons

Metabolic Pathways

**General Strategy:**
Concentration of product controls operon expression

e.g. tryptophan biosynthesis in *E. coli*

---

*Figure 10-21*
*Introduction to Genetic Analysis, Ninth Edition*
© 2008 W.H. Freeman and Company
Tryptophan Operon

1. trpE is first gene in operon
2. trpE mRNA has long leader (untranslated 5' region)
3. region of mRNA works as **attenuator**:
   - in presence of tryptophan, transcription is halted about 140 bases into mRNA
   - in absence of tryptophan, transcription continues

Mechanism is complex, but logical

- leader sequence contains short 11 amino acid peptide with two trp residues
- when trp is abundant, trp-tRNA can be used to translate leader mRNA which terminates transcription
- when trp is limiting, translation stalls and transcription is permitted

his and phe operons use similar mechanisms of attenuation
The Leader and Attenuator of the *trp* Operon
The \textit{trp} Leader
Leader sequences

(a) \textit{trp} operon

Met - Lys - Ala - Ile - Phe - Val - Leu - Lys - Gly - Trp - Trp - Arg - Thr - Ser - Stop

\begin{align*}
5' & \text{ AUG - AAA - GCA - AUU - UUC - GUA - CUG - AAA - GGU - UGG - UGG - CGC - ACU - UCC - UGA } \\
3' & \text{ } \\
\end{align*}

(b) \textit{phe} operon

Met - Lys - His - Ile - Pro - Phe - Phe - Phe - Ala - Phe - Phe - Phe - Thr - Phe - Pro - Stop

\begin{align*}
5' & \text{ AUG - AAA - CAC - AUA - CCG - UUU - UUU - UUC - GCA - UUC - UUU - UUU - ACC - UUC - CCC - UGA } \\
3' & \text{ } \\
\end{align*}

(c) \textit{his} operon

Met - Thr - Arg - Val - Gln - Phe - Lys - His - His - His - His - His - His - Pro - Asp

\begin{align*}
5' & \text{ AUG - ACA - CGC - GUU - CAA - UUU - AAA - CAC - CAC - CAU - CAU - CAU - CAU - CCU - GAC } \\
3' & \text{ } \\
\end{align*}

Figure 10-24

\textit{Introduction to Genetic Analysis, Ninth Edition}

© 2008 W. H. Freeman and Company
Jacob & Monod (lac fame, trp suggestion) also suggested operon-type control of temperate phages which have two states:

- **lysogenic:** phage genome is **inactive**
- **lytic:** phage genome is **active**
Bacteriophage $\lambda$ life cycle

1. Lytic pathway
   - Many viral chromosomes
   - Viral assembly
   - Cell lysis

2. Lysogenic pathway
   - Recombination and integration
   - $\lambda$ prophage
   - Lysogenic growth
   - 3. Prophage induction
   - Cell lysis

Figure 10-25
*Introduction to Genetic Analysis, Ninth Edition*
© 2008 W.H. Freeman and Company
Approach Regulation of Life Cycle

Genetically

Once again, mutations were isolated which allowed for the deduction of the logic of control.

e.g. Isolate phages that are unable to lysogenize bacteria

- they form clear plaques c mutants

   c is analogous to I or O mutants in lac

   cI mutant makes a defective repressor

Also found that genes with related functions in biology of phage are clustered.
Phage λ Genome is Organized for Coordinate Control of gene expression
The choice between lytic and lysogenic cycles is regulated by a genetic switch.

The binding to and occupancy of two repressors on two DNA operators determines the state of the switch - shown is key operator.
**λ Phage Logic**

λ repressor (cI) binds 3 sites at each operator. $O_R$ overlaps Promoter of cI gene itself (repressor regulates its own synthesis)

If high repressor, the third $O_R$ site is occupied and further repressor synthesis is blocked.

**Upon Infection:**

- host RNApol initiates at $O_L \rightarrow N, O_R \rightarrow cro$
- $N$ allows transcription to read through to genes in replication, recombination, lysogeny lysis
- Cro represses repressor $\rightarrow$ Lytic growth

$Cro \rightarrow lytic$

$cI \rightarrow lysogeny$
The $cI$ Control Regions for $\lambda$

- Repressor promotes lysogeny
- $cro$ site
- $cI$ site
- Repressor represses $cI$ lysis

<table>
<thead>
<tr>
<th>N gene</th>
<th>$O_L1$</th>
<th>$O_L2$</th>
<th>$O_L3$</th>
<th>$cI$ gene</th>
<th>$O_R3$</th>
<th>$O_R2$</th>
<th>$O_R1$</th>
<th>$cro$ gene</th>
</tr>
</thead>
</table>

- mRNA 5'
- 3' mRNA 5'
- Protein $cI$
- $cro$
The affinity of each repressor for operator sites is determined by amino acid side chains. This is the basis of specificity of all protein-DNA interactions.
But what controls the switch decision?

$cII$ - a positive activator of $cI$

$cII$ - $cI$ and Bacterial proteases

inactive

active

Q

(lysis)

(lysogeny)

$cIII$ is a protective protein for $cII$.
If $cIII$ is absent, $cII$ is inactive and phage only grow lytically.
Overview of Phage Genetic Switch

Upon Infection,

Host RNA polymerase transcribes *cro* and *N*

Then,

*N* allows for transcription via anti-termination of *cIII* and recombination genes (left) and *cII*, *O*, *P*, and *Q* (right)

Then,

*cIII*, protected by *cIII*, turns on *cI* and *int* (a gene required for integration)

Then, if

proteases are high,

*CII* is inactivated, *cro* represses *cI*, and *Q* acts to allow transcription of the late head and tail genes

If proteases are low,

*cII* is active, *cI* transcription proceeds at a high rate, *int* protein integrates the phage chromosome and *cI* shuts off everybody but itself.

**Key Feature**

Only a few regulatory proteins acting through a few sites control all genes in a "cascade" mechanism.
Multioperon Repression

Multicomponent systems can be subject to same repressor.

   e.g. the SOS repair pathway (UV damage)

   *lexA* represses *recA*, *uvrA*, *uvrB*, *umuC*, and regulates its own synthesis

w/irradiation

   *recA* protease inactivates *lexA* and all the *lexA* repressed genes are now on.

note:

   *recA* also cleaves *cI*, which is the basis for lysogen induction by UV-light, i.e. it inactivates the *λ* repressor that maintains lysogenic state.
Patterns of gene expression. The genes shown in red are on at each of the indicated stages of growth. Genes of related function are turned on and off together. These coordinately regulated genes lie in contiguous blocks except at the late stages of lysogenic growth where only two genes, $cI$ and $int$, are active.
λ Gene Expression

The patterns of gene expression during λ growth are summarized in the previous slide.

The first two stages of development - very early and early - occur before the decision to lyse or lysogenize is made. We can summarize the pattern of gene expression at each stage as follows:

**Very early**
Only genes *N* and *cro* are on.

**Early**
The list of active genes is extended to include the recombination genes, and the DNA replication genes.

**Late**
Here the pathway splits.

- If the phage chromosome is growing lytically, the various early genes are off and the heads, tails, and lysis genes are on. New phage particles are formed and released when the cell lyses.

- If the phage is lysogenizing, only two genes are on - *cl* and *int*. The *int* gene product, located in the recombination region, integrates the phage chromosome into the host chromosome. Finally, in the lysogen, only the repressor gene, *cl* is active.
One more strategy for coordinate control of genes - alternative sigma factors

The spore-forming bacteria *Bacillus subtilis* utilizes several sigma factors in its life cycle.
One more strategy for coordinate control of genes - alternative sigma factors

Each sigma factor recognizes different promoter sequences and controls different sets of genes

![Diagram showing promoter sequences and sigma factors]

**σE-regulated promoters**
- \( ybaN \): TCGGTTATATTCAATTGT-CCATGCTCATAAGAT...
- \( ydcC \): GTCTGCATATTAGGGAAACCCCACTCATATATT...
- \( ydcA \): TACGTACTATTTAAATGG-TTTTTCTCATAAACG...

**σF-regulated promoters**
- \( yrrR \): ACTGGTTTAGCAGGAAACACCTCTGCCCCAATG...
- \( ytfT \): CCGGTTTATTTTTTTTAGGAATTGGCGATAATG...
- \( yuiC \): TTGTGAATAGCTCTCTCACCCTGGGAACAATG...