Natural and synthetic gene regulatory networks

Friday, November 17, 2017
BME/CHE/PHY 558, Physical & Quantitative Biology
Rutgers University: Chemical Thermodynamics
Lecturer: Gábor Balázsi
Challenge: The larger picture

How does gene expression depend on: promoter kinetics, mRNA, protein synthesis/decay?

Write down the relevant reactions!
Reactions: The larger picture

How does gene expression depend on: promoter kinetics, mRNA, protein synthesis/decay?

**Promoter:**

\[ D \xleftarrow{k_{ON}} B \xrightarrow{k_{OFF}} \]

*D*=empty DNA  
*B*=Polymerase-bound DNA

**mRNA, M:**

\[ B \xrightarrow{k_M} M + B \]

\[ M \xrightarrow{g_M} \emptyset \]

**Protein, P:**

\[ M \xrightarrow{k_P} P + M \]

\[ P \xrightarrow{g_P} \emptyset \]
Models of gene expression

Models of gene expression focus on: promoter, mRNA, protein.

Promoter:

\[
\begin{align*}
D & \overset{k_{ON}}{\underset{k_{OFF}}{\rightleftharpoons}} B \\
\frac{dD}{dt} &= -k_{ON}D + k_{OFF}B \\
D + B &= \text{const} = 1
\end{align*}
\]

mRNA:

\[
\begin{align*}
B & \overset{k_M}{\rightarrow} M + B \\
\frac{dM}{dt} &= k_M B - g_M M
\end{align*}
\]

Protein:

\[
\begin{align*}
M & \overset{k_P}{\rightarrow} P + M \\
\frac{dP}{dt} &= k_P M - g_P P
\end{align*}
\]
Gene expression reflects promoter states

Promoter: \[ k_{ON} (1 - B) = k_{OFF} B \]

mRNA: \[ M = \frac{k_M}{g_M} B \]

Protein: \[ P = \frac{k_P}{g_P} M \]
\[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} B \]
\[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} \frac{k_{ON}}{k_{OFF} + k_{ON}} \]
Definition of Networks (Graphs)

- A network (graph) consists of interconnected objects

- Components of a network:
  - Nodes (vertices)
  - Links (edges)
Gene-regulatory networks consist of regulatory interactions between genes.

- Proteins called transcription factors can bind to promoter regions.
- Activator transcription factors can enhance the synthesis of other proteins.
- Repressor transcription factors inhibit gene activity.
Measuring gene expression

Intracellular proteins are invisible. How to measure their quantity?

- The GFP protein shines green.
- Green cells contain more of the original protein as well.
Model for gene repression

We focus on: repressor ($R$), repressor-bound promoter ($B$), unbound ($D$), mRNA ($M$), protein ($P$).

![Diagram of gene repression model with equations and states]

Promoter: $D + R \xleftarrow{k_{ON}} B \xrightarrow{k_{OFF}} D$

$$\frac{dD}{dt} = -k_{ON} DR + k_{OFF} B$$

$$D + B = \text{const} = 1$$

mRNA: $D \xrightarrow{k_M} M + D$

$$\frac{dM}{dt} = k_M D - g_M M$$

Protein: $M \xrightarrow{k_P} P + M$

$$\frac{dP}{dt} = k_P M - g_P P$$
Gene repression

Promoter: \( k_{ON} (1 - B) R = k_{OFF} B \) \[ D = 1 - B = \frac{K}{K + R} \] \[ K = \frac{k_{OFF}}{k_{ON}} \]

mRNA: \[ M = \frac{k_M}{g_M} D \]

Protein: \[ P = \frac{k_P k_M}{g_P g_M} \frac{K}{K + R} \]
Model of gene activation

We focus on: activator (A), promoter bound (B), unbound (D), mRNA (M), protein (P).

**Promoter:**

\[ D + A \xrightleftharpoons[k_{OFF}]^{k_{ON}} B \]

\[ \frac{dD}{dt} = -k_{ON}DA + k_{OFF}B \]

\[ D + B = \text{const} = 1 \]

**mRNA:**

\[ B \xrightarrow{k_M} M + B \]

\[ \frac{dM}{dt} = k_M B - g_M M \]

**Protein:**

\[ M \xrightarrow{k_P} P + M \]

\[ \frac{dP}{dt} = k_P M - g_P P \]
Gene activation

**Promoter:**

\[ k_{ON} (1 - B) A = k_{OFF} B \]

\[ B = \frac{A}{K + A} \]

\[ K = \frac{k_{OFF}}{k_{ON}} \]

**mRNA:**

\[ M = \frac{k_M}{g_M} B \]

**Protein:**

\[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} \frac{A}{K + A} \]
DNA states: Two activator binding sites

Use binding polynomials: \[ QD = D \left(1 + K_1A + K_2A + K_1K_2A^2\right) \]

Probability of empty promoter:
\[ D = \frac{1}{1 + K_1A + K_2A + K_2K_1A^2} \]

Promoter bound by 1 activator:
\[ B_1 = \frac{K_1A + K_2A}{1 + K_1A + K_2A + K_2K_1A^2} \]

Promoter bound by 2 activators:
\[ B_{12} = \frac{K_2K_1A^2}{1 + K_1A + K_2A + K_2K_1A^2} \]
Protein: two activator binding sites

Protein synthesis, empty promoter:

\[ Da_0 = \frac{a_0}{1 + K_1A + K_2A + K_2K_1A^2} \]

Protein, 1 activator-bound promoter:

\[ B_1a_1 = \frac{(K_1 + K_2)AAa_1}{1 + K_1A + K_2A + K_2K_1A^2} \]

Protein, promoter + 2 activators bound:

\[ B_2a_2 = \frac{K_2K_1A^2a_2}{1 + K_1A + K_2A + K_2K_1A^2} \]

Total protein synthesis:

\[ P = \frac{a_0 + (K_1 + K_2)AAa_1 + K_2K_1A^2a_2}{1 + K_1A + K_2A + K_2K_1A^2} \]
Complete activator cooperativity

Promoter: \[ k_{ON} (1 - B) A^n = k_{OFF} B \]

mRNA: \[ M = \frac{k_M}{g_M} B \]

Protein: \[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} \frac{A^n}{K^n + A^n} \]

\[ K = \sqrt[n]{\frac{k_{OFF}}{k_{ON}}} \]

\[ n=5 \]

\[ 0 \quad 10^{-5} \quad 10^0 \quad 10^5 \]
Comparison: Single activator

Promoter: \[ k_{ON}(1 - B)A = k_{OFF}B \]

mRNA: \[ M = \frac{k_M}{g_M} B \]

Protein: \[ P = \frac{k_P}{g_P} \frac{k_M}{g_M} B \]

\[ B = \frac{A}{K + A} \]

\[ K = \frac{k_{OFF}}{k_{ON}} \]
Activation + Repression: the lac operon

The bacterium really likes glucose. It wants nothing else when glucose is present. However, when there is no glucose, it has to eat something... Such as lactose.

<table>
<thead>
<tr>
<th>Glucose</th>
<th>Lactose</th>
<th>Eat lactose?</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>NO</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>NO</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>NO</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>YES</td>
</tr>
</tbody>
</table>

How can the cell decide? By regulating the lac operon (3 lactose-eating genes: *lacZ, lacY, lacA*).
# Cellular logic: the lac operon

<table>
<thead>
<tr>
<th>Glucose</th>
<th>Lactose</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Diagram:**

1. **Glucose +, Lactose -**
   - Repressor active
   - Very little lac mRNA

2. **Glucose +, Lactose +**
   - Inducer-repressor inactive
   - Very little lac mRNA

3. **Glucose -, Lactose +**
   - Inducer-repressor active
   - Abundant lac mRNA
Real gene regulatory networks are large and complex
Other examples of large, complex networks

Internet

Protein-protein interaction network

Social network

Neuronal network
Biological network motifs:
some subgraphs occur more frequently than expected

<table>
<thead>
<tr>
<th>Gene regulation (transcription)</th>
<th><img src="image" alt="Forward feed loop (FFL)" /></th>
<th>Feed-forward loop (FFL)</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Bi-fan" /></td>
<td><img src="image" alt="Bi-fan" /></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Network</th>
<th>Nodes</th>
<th>Edges</th>
<th>$N_{\text{real}}$</th>
<th>$N_{\text{rand}} \pm \text{SD}$</th>
<th>$Z$ score</th>
<th>$N_{\text{real}}$</th>
<th>$N_{\text{rand}} \pm \text{SD}$</th>
<th>$Z$ score</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. coli</em></td>
<td>424</td>
<td>519</td>
<td>40</td>
<td>$7 \pm 3$</td>
<td>10</td>
<td>203</td>
<td>$47 \pm 12$</td>
<td>13</td>
</tr>
<tr>
<td><em>S. cerevisiae</em></td>
<td>685</td>
<td>1,052</td>
<td>70</td>
<td>$11 \pm 4$</td>
<td>14</td>
<td>1812</td>
<td>$300 \pm 40$</td>
<td>41</td>
</tr>
</tbody>
</table>


Pillowcase, Transylvania: **22 tulips**

*Escherichia coli* (Ishihara, 2005)
67 genes and 102 regulations in 42 FFLs
Feedback regulation: a network motif

Core regulatory network of *Escherichia coli*
Feedback loops: When a gene controls itself

Types of feedback:

<table>
<thead>
<tr>
<th></th>
<th>Direct feedback (autoregulation)</th>
<th>Indirect feedback</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive (+)</td>
<td><img src="image1" alt="Positive feedback diagram" /></td>
<td><img src="image2" alt="Positive indirect feedback diagram" /></td>
</tr>
<tr>
<td>Negative (−)</td>
<td><img src="image3" alt="Negative feedback diagram" /></td>
<td><img src="image4" alt="Negative indirect feedback diagram" /></td>
</tr>
</tbody>
</table>

Short feedback (autoregulation and 2-gene feedback) are common motifs in transcriptional networks (~20% of yeast TFs have feedback regulation).

Examples:

- **STE12**: Yeast mating
- **ROX1**: Yeast hypoxia
- **SOX2** and **NANOG**: Human embryonic stem cells
Synthetic Biology:

**Building biological systems for predefined purposes**

making oscillators, switches, gates, biofuels, medicine

---

**Electronic components:**
(transistors, resistors, capacitors, etc.)

- Standard
- Well-characterized
- Reliable
- Low noise
- No replication

---

**Biological components:**
(genes, promoters, RNA-s, etc.)

- Diverse
- Uncharacterized
- Mutate & evolve
- Noisy
- Replication
Cells as dynamical systems: Steady states (equilibria)

Equilibrium: when time derivative = 0

\[
\frac{dP}{dt} = k - \gamma P = 0
\]

Equilibrium: when \( \text{synthesis} = \text{degradation} \)

\[
P_{eq} = \frac{k}{\gamma}
\]
Overdamped systems: Steady states (equilibria)

Analogy: parachute equation of motion

\[ \frac{dv}{dt} = g - \frac{\zeta}{m} v \]

Friction force:

\[ F_f = \zeta v \]

Equilibrium: when time derivative = 0

\[ \frac{dP}{dt} = k - \gamma P = 0 \]

\[ P_{eq} = \frac{k}{\gamma} \]
Gene 1 = lacI
Promoter 1

Gene 2 = tetR
Promoter 2

TetR

LacI

IPTG

ATc

Gene 2 ON

Gene 1 ON

Toggle switch: dynamics

Equations (ODE system)

Protein 1 = $u$; Protein 2 = $v$

\[
\frac{du}{dt} = \frac{\alpha_1}{1 + v^\beta} - u
\]

\[
\frac{dv}{dt} = \frac{\alpha_2}{1 + u^\gamma} - v
\]

NR: A regulatory cascade

\[
\frac{dx}{dt} = a - bxy
\]

- **synth.**
- **degr.**

Potential

Repressor, x

Rate of change

Synthesis
Degradation

Higher ATc
NR gene circuit: dose-response

[Diagram showing the NR gene circuit with tetR, Atc, and GFP, and graphs for fluorescence mean and noise (CV, %) vs. [ATc] (ng/ml)].
NR: gene expression at intermediate induction
The “Linearizer”

TetR represses its own promoter

- Negative Feedback
- Identical promoters

\[ \frac{dx}{dt} \approx aF(x) - bxy \]

Higher ATc

\[ F(x) = \frac{K^n}{K^n + x^n} \]
Comparison: NR and NF mean and CV

Nevozhay et al., *PNAS* **106**:5123, 2009
NF: gene expression at intermediate induction