Structure and function of genetic regulatory circuits in *Escherichia coli*

Michael Wall
Modeling Algorithms & Informatics Group
Computer & Computational Sciences Division
Los Alamos National Laboratory
mewall@lanl.gov
http://home.lanl.gov/mewall
Acknowledgments

• Genetic Regulatory Networks
  – Michael Blinov, LANL PD
  – William Hlavacek, LANL
  – Michael Savageau, UC/Davis
  – Uri Alon, Weizmann Institute

• Department of Energy
• National Institutes of Health

• Protein Function Inference (LANL)
  – William Bruno
  – Judith Cohn
  – Cliff Joslyn
  – James Howse
  – Dengming Ming, PD
  – Andreas Rechtsteiner, GRA
  – Luis Rocha
  – Charlie Strauss
  – Tom Terwilliger
  – Karin Verspoor
Gene Regulation

• Subject of countless studies, beginning with Jacob & Monod
  – Operon model
  – Repressor control
• Important means by which the cell controls levels of gene products
• Central to the cell’s ability to change its capabilities response to environmental and/or developmental signals
An E. coli cell is \( \sim 1\% \) mRNA

### TABLE 1 Composition of an average *E. coli* B/r cell\(^a\)

<table>
<thead>
<tr>
<th>Components</th>
<th>% Total dry wt(^b)</th>
<th>Amt ((10^{15})) per cell(^c)</th>
<th>Mol wt</th>
<th>Molecules per cell</th>
<th>No. of different kinds of molecules(^d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>55.0</td>
<td>156</td>
<td>4.0 (\times 10^4)</td>
<td>2,350,000</td>
<td>1,850</td>
</tr>
<tr>
<td>RNA</td>
<td>20.5</td>
<td>58</td>
<td>1.0 (\times 10^6)</td>
<td>18,700</td>
<td>1</td>
</tr>
<tr>
<td>23S rRNA</td>
<td>31.0</td>
<td>15.5</td>
<td>2.5 (\times 10^4)</td>
<td>198,000</td>
<td>60</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>5.0 (\times 10^5)</td>
<td>1.2</td>
<td>2.5 (\times 10^4)</td>
<td>138</td>
<td>600</td>
</tr>
<tr>
<td>tRNA</td>
<td>0.5</td>
<td>8.2</td>
<td>1.0 (\times 10^6)</td>
<td>138</td>
<td>600</td>
</tr>
<tr>
<td>mRNA</td>
<td>3.1</td>
<td>8.8</td>
<td>2.5 (\times 10^9)</td>
<td>2.1</td>
<td>1</td>
</tr>
<tr>
<td>DNA</td>
<td>3.1</td>
<td>8.8</td>
<td>2.5 (\times 10^9)</td>
<td>2.1</td>
<td>1</td>
</tr>
<tr>
<td>Lipid</td>
<td>9.1</td>
<td>25.9</td>
<td>705</td>
<td>22,000,000</td>
<td>1</td>
</tr>
<tr>
<td>Lipopolysaccharide</td>
<td>3.4</td>
<td>9.7</td>
<td>4.070</td>
<td>1,430,000</td>
<td>1</td>
</tr>
<tr>
<td>Peptidoglycan</td>
<td>2.5</td>
<td>7.1</td>
<td>(904)(\mu)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Glycogen</td>
<td>2.5</td>
<td>7.1</td>
<td>1.0 (\times 10^6)</td>
<td>4300</td>
<td>1</td>
</tr>
<tr>
<td>Polyamines</td>
<td>0.4</td>
<td>1.1</td>
<td>1.0 (\times 10^6)</td>
<td>4300</td>
<td>1</td>
</tr>
<tr>
<td>Putrescine</td>
<td>0.83</td>
<td>1.1</td>
<td>1.0 (\times 10^6)</td>
<td>4300</td>
<td>1</td>
</tr>
<tr>
<td>Spermidine</td>
<td>0.27</td>
<td>1.1</td>
<td>1.0 (\times 10^6)</td>
<td>4300</td>
<td>1</td>
</tr>
<tr>
<td>Metabolites, cofactors, ions</td>
<td>3.5</td>
<td>9.9</td>
<td>1.0 (\times 10^6)</td>
<td>4300</td>
<td>1</td>
</tr>
</tbody>
</table>

\(^a\)Calculated for an average cell in a population of *E. coli* B/r in balanced growth at 37°C in aerobic glucose minimal medium with a mass doubling time of 40 min. The cell is defined by dividing the total biomass, or the amount of any of its measured components, by the total number of cells in the population. This average cell, therefore, is approximately 44% through its division cycle (see reference 11 for the function describing the distribution of cell ages in an asynchronous, steady-state population) and, assuming that increase in cell mass is exponential, is approximately 33% larger than when it was born. This table is modified from data in Table 1 of reference 10.

\(^b\)Relative amounts of the major components based on information in references 3, 12, and 16 and on unpublished experiments of the author (F.C.N.; see text). In some cases, data from strains other than B/r, from growth conditions other than the reference one, or from both had to be used (see references concerning glycogen [5], polyamine [9], and lipid [16]).

\(^c\)Based on measurements of the total dry mass and the number of cells measured in portions of a reference culture (unpublished observations). The wet weight is calculated from the assumption that 70% of *E. coli* protoplasm is water. The total dry weight per cell is 2.8 \(\times 10^{-12}\) g; the water content (assuming that 70% of the cell is water [3]) is 6.7 \(\times 10^{-13}\) g; the total weight of one cell is 9.5 \(\times 10^{-13}\) g.

\(^d\)Based on the following information sources: protein, examination of two-dimensional polycrylamide gels (unpublished observations); stable RNA, chapter 88; mRNA, assuming three genes per average transcriptional unit; lipid, an indeterminant number of species because of the variety of fatty acids associated with the four major types of phospholipids exclusive of lipopolysaccharide (i.e., 76% phosphatidylethanolamine, 20% phosphatidylycerol, and small amounts of cardiolipin and an unidentified species) (1, 12); and metabolites, cofactors, and ions, roughly estimated as described in Table 3 of reference 10.
Understanding Gene Regulation

- There are subsystems that adjust gene expression in response to specific environmental signals
  - Genetic regulatory circuits
- The overall genetic regulatory response depends on the detailed structure of a gene circuit
  - Gene-circuit design
- Gene-circuit performance may depend on gene-circuit design
  - Mathematically controlled comparison
- Does evolution select for designs that optimize performance?
  - Design principles
Activator and Repressor Control

\[ X \xrightarrow{+/-} Y \]

+ indicates positive mode of control, or activator control
- Indicates negative mode of control, or repressor control

Are there natural preferences for mode of control?

Demand Theory
Demand Theory

Regulator $\rightarrow$ Enzyme

Demand Theory:

If enzyme is in high demand, expect positive control

If enzyme is in low demand, expect negative control

Savageau. Genetics 1998

Example of a design principle
Demand Theory

High Demand

Regulator → + Enzyme

Loss of enzyme
High Penalty

Regulator → - Enzyme

Overexpression
Low penalty

Any observed preference should be for positive control

Savageau. Genetics 1998
Genetic Regulatory Circuit

- Genetic regulatory interactions may be modulated by small-molecule signals
- The genes and gene products involved in the genetic regulatory response to a signal comprise a genetic regulatory circuit
- To understand the genetic regulatory response, it is necessary to know signal interactions as well as genetic regulatory interactions

1 April 2004
Michael Wall / mewall@lanl.gov
Elementary Gene Circuits

- Elementary gene circuits involve just one transcription factor (TF)
- The genetic regulatory interactions are like those of a single-input module
  - Effector genes are coordinately regulated
- The TF may regulate its own expression
- TF activity is modulated by a signal
Types of Elementary Gene Circuit

- Activator vs. Repressor
- Function types and properties
- Logarithmic gain
- Coupling type

**Activator vs. Repressor**
(a) Inducible
- $\log_{[\text{Capacity}]}$
- Basal Expression

(b) Repressible
- Expression in the Absence of Signal
- $\log_{[\text{Capacity}]}$

**Log**
- $\log_{[\text{Enzyme}]}$
- $\log_{[\text{Regulator}]}$
- $\log_{[\text{Inducer}]}$
- $\log_{[\text{Co-Repressor}]}$
Functions of Elementary Gene Circuits

• Catabolic gene circuit
  – E.g. lac
  – Turn on catabolic enzymes when a key metabolite is present
  – Inducible

• Biosynthetic gene circuit
  – E.g. trp
  – Turn on anabolic enzymes when a key metabolite is scarce
  – Repressible

1 April 2004 Michael Wall / mewall@lanl.gov
The Search for Design Principles

• How can we understand the functional consequences of selecting alternative designs?

• Define quantitative performance criteria, and analyze the dependence of performance on system design

• Design principles may then be developed to predict how performance may be optimized
Practical Impact of Design Principles

**Engineered Gene Circuits**

- **Bioremediation**
  - Active biological containment (ABC) circuit (Ramos 2001)
    - $\text{Pm::asd} / (\text{Pm::lacI, xylS, lac::gef})$ system in a $\Delta\text{asd strain}$ of *P. putida*

- **Biosensors**
  - Toxin-induced fluorescent circuit (Bechor 2002)
    - $\text{fabA::lux}$ fusion plasmid in *E. coli*

- **Gene therapy**
  - Cancer-specific viral circuit (Ramachandra 2001)
    - Engineered virus that is repressed at normal p53 levels

- **Metabolic engineering**
  - Environment-sensitive metabolic circuit (Farmer & Liao 2000)
    - ACP-induced lycoprene production in *E. coli* using a modified Ntr regulon
Elementary Gene Circuit Model

Dashes = inducible-catabolic gene circuit  
Dot-dashes = repressible-biosynthetic gene circuit

1 April 2004  Michael Wall / mewall@lanl.gov
Coupled ODE Models of Gene Regulation

Transcription rate

\[ \frac{dX}{dt} = V^{(+)} - V^{(-)} \]

First-order decay/dilution

\[ V^{(-)} = B \ X \]

Hill-function regulation model

\[ V^{(+)} = A \left(1 + R^n\right)^{-1} \quad n < 0 \text{ activation} \]
\[ n > 0 \text{ repression} \]
Model of Activation

\[ x^{**2}(1+x^{**2})^{**(-1)} \]
\[ x^{**4}(1+x^{**4})^{**(-1)} \]
\[ x^{**8}(1+x^{**8})^{**(-1)} \]
Model of Repression

\[(1+x^2)^{-1}\]
\[(1+x^4)^{-1}\]
\[(1+x^8)^{-1}\]

1 April 2004
Michael Wall / mewall@lanl.gov
Power-Law Approximation

\[ \frac{dX_i}{dt} = \alpha_i \prod_j X^{g_{ij}} - \beta_i \prod_j X^{h_{ij}} \]

1 April 2004

Michael Wall / mewall@lanl.gov
Elementary Gene Circuits
Mathematical Model

\[
\frac{dX_1}{dt} = \alpha_1 X_6^{g_{16}} X_3^{g_{13}} X_5^{g_{15}} - \beta_1 X_1^{h_{11}}
\]

\[
\frac{dX_2}{dt} = \alpha_2 X_7^{g_{27}} X_1^{g_{21}} - \beta_2 X_2^{h_{22}}
\]

\[
\frac{dX_3}{dt} = \alpha_3 X_8^{g_{38}} X_9^{g_{39}} X_2^{g_{32}} X_3^{g_{33}} - \beta_3 X_2^{h_{32}} X_3^{h_{33}}
\]

\[
\frac{dX_4}{dt} = \alpha_4 X_6^{g_{46}} X_3^{g_{43}} X_5^{g_{45}} - \beta_4 X_4^{h_{44}}
\]

\[
\frac{dX_5}{dt} = \alpha_5 X_7^{g_{57}} X_4^{g_{54}} - \beta_5 X_5^{h_{55}}
\]

X_1 = Enzyme mRNA
X_2 = Enzyme
X_3 = Intracellular signal
X_4 = TF mRNA
X_5 = TF
X_6 = nucleic acid pool
X_7 = amino acid pool
X_8 = Signal precursor
X_9 = Extracellular signal

1 April 2004
Michael Wall / mewall@lanl.gov
Performance Criteria

• Stability
  – Ability of a unique, steady-state enzyme level to be determined by signal level, even when model parameter values may vary
  – Linearize equations about the steady-state solution, check for negative eigenvalues
  – Measure the distance in parameter space between the stable system and the nearest unstable system

• Robustness
  – Ability of the steady-state enzyme level to be maintained when model parameter values vary
  – Parameter sensitivity analysis

• Responsiveness
  – Ability of the system to equilibrate quickly after a change in signal
Mathematically Controlled Comparison

• Compare performance criteria for alternative systems that carry out the same function (e.g. repressible system)

• Internal Equivalence
  – Maintain identical parameter values for all processes but those being considered as alternative designs
    • All processes but transcription of regulator and enzyme are equivalent

• External Equivalence
  – Ensure that systems being compared have the same overall biological function
    • Steady-state effector gain is the same for alternative systems

1 April 2004 Michael Wall / mewall@lanl.gov
Prediction

Negative TF self-regulation Increases Performance

- Regulator +/− Enzyme

• Increased stability, robustness
• Increased responsiveness

Savageau. Nature 1974
Hlavacek & Savageau. JMB 1996
Wall et al. JMB 2003
Experimental Confirmations

**Autogenous Regulators Increase Performance**

**Theoretical Prediction**

- Flux to product $V_3$

**Experimental Confirmation**

- Normalized free monomer concentration

Wall, Hlavacek & Savageau. JMB 2003

Rosenfeld, Elowitz & Alon. JMB 2002
Robustness: Little et al. EMBO J 1999

1 April 2004 Michael Wall / mewall@lanl.gov
Responsiveness of Alternative Coupling Types

Dimensionless time $\tau = F_2 t$

Normalized enzyme concentration $u_2$

1 April 2004 Michael Wall / mewall@lanl.gov
# Prediction of Coupling Types for Optimal Performance

<table>
<thead>
<tr>
<th>Effector TU</th>
<th>Low Gain Gain</th>
<th>Intermediate Gain</th>
<th>High Gain Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inducible (+)</td>
<td>Inverse</td>
<td>Uncoupled</td>
<td>Direct</td>
</tr>
<tr>
<td>Inducible (-)</td>
<td>Direct</td>
<td>Uncoupled</td>
<td>Inverse</td>
</tr>
<tr>
<td>Repressible (+)</td>
<td>Inverse</td>
<td>Direct</td>
<td>Direct</td>
</tr>
<tr>
<td>Repressible (-)</td>
<td>Direct</td>
<td>Uncoupled</td>
<td>Inverse</td>
</tr>
</tbody>
</table>

1 April 2004

Michael Wall / mewall@lanl.gov
Does Natural Selection Depend on Performance Criteria?

- In general we do not know which features of an organism confer selective advantage.
- It is important to consider ecological context, as was done for demand theory.
- Still, we may formulate a hypothesis that stability, robustness and responsiveness confer selective advantage.
- We may then test this hypothesis by seeing whether natural systems exhibit our predicted preferred designs...
EcoTFs
A database of E. coli TFs and signals
http://ecotfs.lanl.gov

<table>
<thead>
<tr>
<th>TF</th>
<th>Regulator TU(s)</th>
<th>Effector TU(s)</th>
<th>Coupling</th>
<th>Signal(s)</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Betl</td>
<td>betlBA (−)</td>
<td>betlBA (−)</td>
<td>D</td>
<td>choline (Ind)</td>
<td>Lamark T et al. (1996)</td>
</tr>
<tr>
<td></td>
<td>betT (−)</td>
<td></td>
<td></td>
<td></td>
<td>Rokenes TP et al. (1996)</td>
</tr>
</tbody>
</table>

Test theoretical predictions
Develop models
Not just elementary gene circuits
## Summary of Systems in EcoTFs

<table>
<thead>
<tr>
<th></th>
<th>(-) at Regulator TU</th>
<th>(+) at Regulator TU</th>
<th>TF does not control Regulator expression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
<td>U</td>
<td>D</td>
</tr>
<tr>
<td><strong>Inducible (+)</strong></td>
<td>4(^b)</td>
<td>3(^c)</td>
<td>4(^d)</td>
</tr>
<tr>
<td><strong>Inducible (-)</strong></td>
<td>0</td>
<td>0</td>
<td>9(^g)</td>
</tr>
<tr>
<td><strong>Repressible (+)</strong></td>
<td>0</td>
<td>3(^i)</td>
<td>0</td>
</tr>
<tr>
<td><strong>Repressible (-)</strong></td>
<td>0</td>
<td>1(^k)</td>
<td>9(^l)</td>
</tr>
</tbody>
</table>

\(^b\)(AraC, IlvY, MetR, SoxS); \(^c\)(CynR, SoxR, TorR); \(^d\)(CysB, DsdC, MelR, RhaS); \(^e\)(CpxR, IdnR, MarA, RhaR, XylR); \(^f\)(MalT, MhpR, Rob, XapR); \(^g\)(BetI, CytR, EmR, GalS, MarR, NagC, PdhR, PutA, UxuR); \(^h\)(GalR, GlpR, LacI, RbsR); \(^i\)(AsnC, GcvA, PspF); \(^j\)(FadR, FruR); \(^k\)(TyrR); \(^l\)(ArgR, DnaA, Fur, H-NS, IscR, MazEF, MetJ, PurR, TrpR); \(^m\)(ModE).

1 April 2004 Michael Wall / mewall@lanl.gov
Patterns in EcoTFs System Types

• Negative Self-Regulation
  – Preference observed in natural systems (33/49)

• Positive Self-regulation
  – Other functions?
    • Discontinuous switches
  – Other performance criteria?

• Cannot test coupling type predictions -- yet

• Inverse coupling is not found among repressible systems

1 April 2004 Michael Wall / mewall@lanl.gov
Design Principles of Binary Gene Circuits

• Gene circuits with two TFs
• Noise rejection as a performance criterion
  – Complementary to responsiveness

Simple binary
Simple cascade
Feed-forward cascade

1 April 2004 Michael Wall / mewall@lanl.gov