Modeling Genetic & Biochemical Processes

Tomorrow's forecast: Mostly sunny with a chance of cell division

(0) Intro

(1) Bacteriophage T7 model

(2) Thoughts on modeling

(3) Yeast pheromone signal transduction model

(4) Other

Drew Endy
Complementing (?) Approaches

(1) **Top-down, existing data**
   e.g., analysis of mRNA expression array data leads to underlying genetic regulatory network architecture.

(2) **Bottom-up, existing data**
   e.g., bacteriophage λ genetic-switch model.

(3) **Bottom-up, driving experiments**
   e.g., bacteriophage T7 development, yeast pheromone mating response.

(4) **Bottom-up, designed/re-factored biological systems**
   e.g., bistable switch, repressillator, frankenphage.

(5) **Module/sub-module (whatever) analysis**
   e.g., “use-it-or-lose-it”, adaptation in chemotaxis
Biology has a good history of model building

For example: *lac operon* (see Benno Müller-Hill, 1996; Dienert (1900) Ann Inst Pasteur 14:139-189)
Quantitative Models?

Model

'time' = repressed

- mRNA
- Repressor binds to operator, preventing transcription of lac operon

Model output

Number of molecules per cell

'time' = induced

- mRNA
- Lac mRNA
- Inducer-repressor complex does not bind to operator

Graph:
- mRNALac
- LacZ
- IPTG
- LacI
- LacI-IPTG

Time [s] = seconds
Needed: Better description of process physics

Even when the system is “specified”, it is still very hard to represent a biological system realistically.

E. coli cytoplasm is ~200-320 mg/ml protein
GFP has an apparent diffusion coefficient 11-fold lower in E. coli cytoplasm than in water.

minC-GFP translocation

E-human 1.0?

Human

time < 1 week

nope

why?
(1) incomplete specification
(2) unknown transition rules
(3) uncertain physics
(4) unknown levels of resolution
(5) computational methods/power
(6) unknown function

time > 70 years
Possible now? Maybe bacteriophage T7 model...

What is phage T7?

Novagen, Inc

Spires & Brown, unpublished
T7 was isolated from standard anti-coll-phae mixture prepared by Dr. W.J. MacNeal... Demerec & Fano (1944).

Genome sequenced (100%) Dunn & Studier (1982).
39,937 bp linear dsDNA genome

56 genes thought to encode 60 proteins

Primary function assigned to ~33 genes

> 50 regulatory elements

“Protein” linkage map via yeast two-hybrid, Bartel et al. (1996).

~62nm diameter particle, short tail.

T7 RNA and DNA polymerases extremely well characterized.

And much more...
Innocent beginnings

Get to know: The characterized T7 genes
T7 model is a coupled series of smaller models

(1) DNA entry
(2) Transcription
(3) Translation
(4) Replication
(5) Prohead assembly
(6) DNA packaging
(7) Particle assembly
(8) Lysis
Given a T7 particle bound to host...

T7 DNA entry:
Example DNA entry experiment

(1) \( DpnI \) sites (cuts at GATC given M+)

restriction fragment time course

(2)

RNAP mediated entry

(3)

Series of models

Process

DNA entry

Transcription

Translation

DNA replication

Capsid assembly

DNA packaging

Particle assembly

[Lyis]
Model resolution = genetic elements

T7 sequence

T7 model genome
Series of models

**Process**
- DNA entry
- Transcription
- Translation
- DNA replication
- Capsid assembly
- DNA packaging
- Particle assembly
- [Lysis]

**Model**

Step 1: Assign RNA polymerase to each promoter
Step 2: Set RNA polymerase density for each mRNA
Step 3: Set translation level based on mRNA species
\[
\frac{d(mRNA_i)}{dt} = \left[ (k_{Ph}^i)(S_i) + (k_{Rd}^i)(mGPI) \right] - (k_{dm}^i)(mRNA_i) \quad \text{for } i = 1 \text{ to } N_m
\]

\[
\frac{d(gpi)}{dt} = \left[ (k_{T}^i)(R_d)(mGPI) \right] - (k_{ugp}^i)(gpi) + T 
\]

for \( i = 1 \) to \( N_g \)
Evaluation: protein synthesis
### Series of models

<table>
<thead>
<tr>
<th>Process</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA entry</td>
<td>In the simulation we assume that a replisome consists of a gp5:thioredoxin complex and six gp4A molecules. We assume that replisomes form instantly given stoichiometric amounts of these components. Furthermore, we assume thioredoxin is in excess at all times and all T7 DNA polymerase molecules (gp5) become processive after forming a 1:1 complex with thioredoxin. We also assume that up to two replication forks can form per complete T7 genome, given sufficient numbers of replisomes. This last assumption allows for an exponential increase in the total rate of DNA synthesis as replication proceeds. The roles of gp1, gp2.5, and gp3.5 are accounted for by incorporating &quot;hard&quot; switches such that DNA synthesis occurs only if the concentrations of gp2.5 and the gp1:gp3.5 complex are non-zero. Having met these constraints, the simulation treats replisome elongation as the rate-limiting step in DNA synthesis. This rate is set as a function of dNTP concentrations using Michaelis-Menten kinetics. The simulation ignores any intermediate steps between the formation of acid-soluble DNA fragments created from the phage-mediated digestion of the host genome and their eventual conversion to dNTPs. We assume that the digestion of the host genome proceeds at a constant rate if the concentrations of the T7 endo- (gp3) and exonucleases (gp6) are non-zero [L. You, Thesis Proposal, University of Wisconsin-Madison, Madison, WI (1999)].</td>
</tr>
<tr>
<td>Transcription</td>
<td></td>
</tr>
<tr>
<td>Translation</td>
<td></td>
</tr>
<tr>
<td>DNA replication</td>
<td></td>
</tr>
<tr>
<td>Capsid assembly</td>
<td></td>
</tr>
<tr>
<td>DNA packaging</td>
<td></td>
</tr>
<tr>
<td>Particle assembly</td>
<td></td>
</tr>
<tr>
<td>[Lysis]</td>
<td></td>
</tr>
</tbody>
</table>
Series of models

**Process**

- DNA entry
- Transcription
- Translation
- DNA replication
- Capsid assembly
- DNA packaging
- Particle assembly

**Model**

"Procapsid assembly is simulated with a 4.78-order nucleation-limited reaction derived from data for phage P22 (Prevelige et al., 1993). T7 and P22, both Podoviridae, have dissimilar genomes and growth cycles (Hausmann, 1988 in The Bacteriophages, R. Calendar, Ed., Plenum, NY), but their icosahedral capsids are each approximately 60 nm in diameter and are attached to short noncontractile tails (Ackermann & Berthiaume, 1995, Atlas of Virus Diagrams, CRC, NY). The P22 kinetic data for procapsid assembly is the most comprehensive for any phage and allows the development of a procapsid rate expression. This representation for the formation of procapsids includes the requirement that the major capsid protein concentration is above a nucleating concentration before assembly initiates. The consumption of procapsids as progeny are formed requires complete procapsids, T7 DNA, and enough of each structural protein to complete the phage. As procapsids and progeny phage particles are assembled the simulation accounts for the depletion of various T7 proteins. The simulation also accounts for the utilization of scaffold protein, gp9, during procapsid assembly. Unlike P22, which is known to recycle its scaffold protein, T7 does not (Roeder & Sadowski, 1977), despite the fact that gp9 does not remain in the final phage particle (Steven & Trus, 1983 in Electron Microscopy of Proteins: V, J.R.Harris & R.W. Horne, Eds., Academic Press, London). The simulation assumes that packaging of DNA into the procapsid is the rate limiting step for T7 progeny formation (given DNA, procapsids, and non-zero concentrations for all particle proteins and the DNA maturation proteins, gp18 and gp19)." adapted from D. Endy, Dissertation, Dartmouth College, Hanover NH (1998)"
Series of models

Process

DNA entry
Transcription
Translation
DNA replication
Capsid assembly
DNA packaging
Particle assembly
[Lysis]

Model

*not included*

Concentration dependent relationship between degradation of host membrane(s) and phage proteins is unknown at present.
Evaluation: intracellular progeny

Experimental data and simulation results from Endy et al., (1997)

*Simulated red line from T7v2 using $\eta_{dack} = 0.5$
All this trouble for what?

In T7, the position of a gene on the genome directly regulates the timing and level of gene expression.

If I have a model based on what is known, can I compute what happens when I move genes around?
Expectation: “sliding” gene 10A
Expected: “sliding” gene 1

Phage T7 per minute
Protein synthesis rates comparison
Protein synthesis rates comparison (cont.)

Observed

Computed

Faster

No change

Slower

Genome-wide protein synthesis rates

Fold change from wild-type protein synthesis rate

0.01 0.1 1 10
Acknowledge: The uncharacterized T7 genes

4.7

Query=
(135 letters)

Distribution of 2 Blast Hits on the Query Sequence

Mouse-over to show de fined and scores. Click to show alignments

Color Key for Alignment Scores

<table>
<thead>
<tr>
<th>Score Range</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;40</td>
<td>Red</td>
</tr>
<tr>
<td>40-50</td>
<td>Blue</td>
</tr>
<tr>
<td>50-80</td>
<td>Green</td>
</tr>
<tr>
<td>80-200</td>
<td>Purple</td>
</tr>
<tr>
<td>&gt;=200</td>
<td>Red</td>
</tr>
</tbody>
</table>

Score E

Sequences producing significant alignments:

<table>
<thead>
<tr>
<th>E-value</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2e-57</td>
<td>4.7</td>
</tr>
</tbody>
</table>

Sequence: gi|2737453|ref|NP_041981.1| gene 4.7 [Bacteriophage T7] gi|9759557|dbj|BAB11444.1| (AB010070) SMC-like protein [Arab...
Suspending disbelief: genome permutations

What is this system optimizing?

If I wanted this system to...

If I wanted a system that...
Arbitrary characterization: PCA of permuted genomes

Class 1 & 2: $\mu_{T7} < 5$ phage per minute (N=900)
Class 3: $\mu_{T7} > 5$ phage per minute (N=100)
T7 wildtype
First-order approximate description of resources
Computing consequences of intervention...
Radically different ways of choosing targets...
Partially inhibit component of negative feedback loop
Thinking about doing experiments “right”

T7 sequence near ϕ3.8

ϕ3.8

3.8 RBS

3.5 stop

mRNA start

R3.8 (cut site unknown)

Two approaches

(1) “Refactor” T7

(2) Build a new phage (ϕf) that can be modeled
Aside: technology infrastructure

Rob Carlson (http://www.molsci.org/~rcarlson/exp_biol_tech.html)
Modules defined by “function”, what is function?
At least as important as having complete description of components

Better description of environment
How do biological systems encode environmental information (time scales)?

Better description of constraints (from physics)

Biology is a medium for creation
Evolution is the default design algorithm
e.g., constraint of continued existence
Acknowledgements

Eric Lyons
Lynn Thomason
John Yin
Lingchong You
Ian Molineux
Roger Brent
[Sydney Brenner]
Rob Carlson
Alejandro Colman-Lerner
Larry Lok

http://virus.molsci.org/t7/
http://yeast.molsci.org/alpha/
Yeast pheromone signal transduction pathway

Traversing levels of resolution