Modeling and Analysis of Gene Regulatory Networks

Hamid Bolouri

http://www.its.caltech.edu/~hbolouri/

Quantitative Systems Immunology Summer Course,

San Antonio, July 2010
What is a model?

Why model?
What is a model?

Pattern matching

Red sky at night,
Sailor's delight;
Red sky at morning,
Sailor's warning.

Probabilistic

\[ P(\text{high}|S) = \frac{P(S|\text{high})P(\text{high})}{P(S|\text{low})P(\text{low}) + P(S|\text{high})P(\text{high})} \]

Regression based

Logic

<table>
<thead>
<tr>
<th>Node</th>
<th>Boolean Regulatory Rule</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO</td>
<td>( \overline{\text{NO}}^* = \text{NIA12} \text{ and } \overline{\text{NOS}} )</td>
</tr>
</tbody>
</table>
Declarative/data-driven/implicit modeling

e.g. http://cytoscape.org/

Guilt by association:

e.g. to discover

• Novel cluster members
  • co-expression
  • Gene Ontology
  • highly connected

• Global context for focused data
  • e.g. pathway interactions

• Patterns of activity across
  • time
  • cell conditions/types

• Recurring network motifs

• Evolutionary-conserved motifs
DNA-centered GRNs

http://www.biotapestry.org/

View from the Genome

View from All Nuclei

View from T1 Nuclei

View from T2 Nuclei
Why make models of network dynamics?

- Complex concentration
  - Time
  - Total concentration of the green protein
... surprising behavior
Only mechanistically correct models extrapolate reliably

Red : measured [complex]
Black : trend-line of [complex]
Green : [complex] = [scaffold]
Building mechanistic GRN models from data
A protocol for unraveling gene regulatory networks

Stefan C Materna¹ & Paola Oliveri¹,²

NATURE PROTOCOLS | VOL.3 NO.12 | 2008
ChIP-Seq Overview

1. Cross-link proteins to DNA

2. Fragment chromatin to <500bp

3. Immunoprecipitate antibody bound DNA fragments

4. Reverse cross-links and sequence fragment ends

5. Map sequence reads to genome

6. Identify genomic regions with enriched number of mapped reads

Barbara Wold’s lab, Science, 2007, 16(5830):1497-502
Kharchenko, Tolstorukov and Park

Torres, Metta, Ottenwälder & Schlötterer, Genome Research, 2007, 18:000
Estimating the average DNA fragment length from the strand-specific tag shift

Sarkar, Gentleman, Lawrence, Zhang, Yao

Kharchenko, Tolstorukov and Park
These 2 peaks are apart by only ~150bp.

If I had assumed a larger average DNA fragment size, these peaks would overlap!
Not all large peaks are true binding sites
The benefit of having negative controls
Ly9 expression is repressed in the bone marrow
A typical pipeline:

(Hes1 ChIPseq in HSCs)

> 500 lines of R

Uses 12 R “packages”

Includes > 10 major user-defined parameters (e.g. # read mismatches, Baye’s PP, ncRNAs?)

> 200 interim results files
The Better Way to Crunch data

- Use privately, or share data & R scripts
- Access the power of R through point & click menus
- Run your analysis on the Amazon cloud
  - free for small jobs, or
  - buy virtually unlimited resources directly from Amazon

CRdata.org

Everyday researchers lose valuable time waiting for computations on their desktops. CRdata.org frees your desktop for other work while our servers analyze your data in the cloud.

You can use CRdata as a private resource, share your data and R scripts with selected others, or with the whole community.

To ensure you have access to the latest resources, all R and Bioconductor libraries are updated nightly on CRdata.

CRdata is free & open source
Basics of kinetic modeling with Ordinary Differential Equations
Using Ordinary Differential Equations (ODEs) to model dynamics

distance traveled = current position – starting position
  = speed x (time running)

speed = [(position at time t2) – (position at time t1)] / (t2 – t1)

speed = \frac{d(position)}{d(time)} = \frac{position_2 - position_1}{time_2 - time_1} \quad \text{as} \quad (time_2 - time_1) \to 0

acceleration = \frac{d(speed)}{d(time)} = \frac{d^2(position)}{d(time)^2}
Basics of Gene Expression Kinetics
A simple 2-step ODE model of transcription and translation

\[
\frac{d[mRNA]}{dt} = k_t \cdot Y - k_{dm} \cdot mRNA
\]

\(k_t\) is the maximal rate of transcription

\[
\frac{d[P]}{dt} = k_s \cdot mRNA - k_{dp} \cdot P
\]

\(k_s\) is the protein synthesis rate/mRNA concentration unit

At steady-state:

\[P_{ss} = \frac{k_s}{k_{dp}} \cdot \frac{k_t}{k_{dm}} \cdot Y\]

\[P_{ss} \propto \text{Fractional Saturation of TF complex on its DNA binding site}\]
Transcriptional activation by one factor

Fractional DNA Occupancy $Y_A = \frac{[A : DNA]}{[DNA] + [A : DNA]} = \frac{K_A . A . DNA}{DNA + K_A . A . DNA}$

$Y_A = \frac{K_A . A}{1 + K_A . A}$ \hspace{1cm} or equivalently: \hspace{1cm} $Y_A = \frac{A}{K_{DA} + A}$

$\frac{d(mRNA)}{dt} = k_t . Y_A - k_{dm} . mRNA$

$\frac{d(P)}{dt} = k_s . mRNA - k_{dp} . P$
Transcriptional repression

\[ Y_R = \frac{K_{AR} \cdot R}{1 + K_{AR} \cdot R} \]  
(fraction of DNA occupied by R)

\[ \frac{d(mRNA)}{dt} = k_t \cdot (1 - Y_R) - k_d \cdot mRNA \]

\[ \frac{d(mRNA)}{dt} = k_t \cdot \left( \frac{1}{1 + K_{AR} \cdot R} \right) - k_d \cdot mRNA \]
DNA occupancy by 2 factors

“independence of the path”: \[ K_A \cdot K_{AB} = K_B \cdot K_{BA} = K_q \cdot K_A \cdot K_B \]

Where \( K_q = \text{cooperativity factor} \)

occupancy = \( \frac{\text{steady state level of activating states}}{\text{sum of the steady state levels of all DNA configurations}} \)

\[
\begin{align*}
\frac{d(DA)}{dt} &= k_a \cdot A.D - k_{-a} \cdot DA \\
\frac{d(DAB)}{dt} &= k_{ab} \cdot B.DA - k_{-ab} \cdot DAB \\
DA &= \frac{k_a}{k_{-a}} \cdot A.D = K_A \cdot A.D \\
DAB &= \frac{k_{ab}}{k_{-ab}} \cdot B.DA = K_{AB} \cdot B.DA = K_{AB} \cdot B \cdot K_A \cdot A.D = K_q \cdot K_A \cdot K_B \cdot A.B.D
\end{align*}
\]
occupancy = \frac{\text{steady state level of activating states}}{\text{sum of the steady state levels of all DNA configurations}}

For 2 activators

\[
\]

For highly cooperative factors

\[
Y \approx \frac{K_A[A].K_B[B].K_q}{K_A[A].K_B[B].K_q + 1}
\]

For 1 TF multimer

\[
Y \approx \frac{(K_A.A)^N}{(K_A.A)^N + 1}
\]

or

\[
Y \approx \frac{A^N}{A^N + K_{DA}N}
\]

For 1 repressor, 1 activator

\[
\]
GRN building blocks
A Recurrent Network Involving the Transcription Factors PU.1 and Gfi1 Orchestrates Innate and Adaptive Immune Cell Fates


Multilineage Transcriptional Priming and Determination of Alternate Hematopoietic Cell Fates

Laslo et al, Cell 2006, 126:755-766

http://www.BioTapestry.org
http://www.its.caltech.edu/~tcellgrn/TCellMap.html
Topological motifs as network building blocks

Ecoli

- feedforward loop
- single input module (SIM)
- dense overlapping regulons (DOR)

Uri Alon & colleagues

york

Rick Young & colleagues

BMC Bioinformatics

Research article

Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network

Radu Dobrin¹, Qasim K Beg¹, Albert-László Barabási² and Zoltán N Oltvai*¹
Gene regulatory building blocks

Latch

 PERFECT ADAPTATION

Toggle Switch

Mutual Exclusion

Single-pulse / adaptation
Rapid response
Fixed expression level
Oscillator

Bolouri & Davidson, BioEssays, 2002
Longabaugh & Bolouri, Current Genomics, 2006
Examples of positive feedback

Auto-regulation

Mutual repression
Steady states in feedback networks

conceptual model:

- stable steady state 1
- stable steady state 2
- unstable steady state 1

At steady states, production rate = clearance rate

- stable steady states
- unstable steady state

\[
\frac{dG}{dt} = k_t \cdot \frac{G^N}{K^N + G^N} - k_d \cdot G
\]

rate of production > rate of clearance \(\rightarrow\) G will increase over time

rate of clearance > rate of production \(\rightarrow\) G will decrease over time
Two ways of providing input:

1. **Independent activating input**

   ![Diagram](image1)

   Add a second occupancy term:

   \[
   \frac{dG}{dt} = k_{t1} \cdot \frac{G^N}{K_G + G^N} + k_{t2} \cdot \frac{i^n}{K_i^n + i^n} - k_d \cdot G
   \]

2. **Input is the same protein as feedback**

   ![Diagram](image2)

   Add to G in the occupancy term:

   \[
   \frac{dG}{dt} = k_t \cdot \frac{(G+i)^N}{K^N + (G+i)^N} - k_d \cdot G
   \]

---

198x141 gene G
67x67 Two ways of providing input:
270x298 iK
260x280 i.k
286x193 GK
260x174 G.kdt
260x88 dG
274x336 dnn
292x274 i
253x284 n
274x243 n2NN
293x167 G
274x187 N
274x137 t1
274x137 -
286x287 +
270x222 +
286x182 +
270x115 =
529x100 rate of clearance
501x73 rate of production
198x469 gene G
104x93 independent activating input
124x103 adding a second occupancy term:
229x126 add to G in the occupancy term:
272x669 .Gk
288x615 i)GK
262x596 i)(G.kdt
262x485 dG
276x661 dNN
255x605 N
276x605 -
288x604 +
262x585 +=
288x584 (.add to G in the occupancy term:
360x624 rate of clearance
415x621 rate of production
Auto-regulation: hysteresis & bistable lock-on switches

\[ \text{mRNA}' = k_t \cdot \left[ \frac{(\text{in} + P)^N}{K_{Diss}^N + (\text{in} + P)^N} \right] - k_{dm} \cdot \text{mRNA} \]

\[ P' = k_s \cdot \text{mRNA} - k_{dp} \cdot P \]

At steady state:

\[ \text{mRNA} = \frac{k_t}{k_{dm}} \cdot \left[ \frac{(\text{in} + P)^N}{K_{Diss}^N + (\text{in} + P)^N} \right] \]

\[ P = \frac{k_s}{k_{dp}} \cdot \text{mRNA} \rightarrow \text{mRNA} = \frac{k_{dp}}{k_s} \cdot P \]
Prediction and measurement of an autoregulatory genetic module

Farren J. Isaacs*, Jeff Hasty†, Charles R. Cantor*, and J. J. Collins**

7714–7719 | PNAS | June 24, 2003 | vol. 100 | no. 13

Hysteresis in a synthetic mammalian gene network

Beat P. Kramer and Martin Fussenegger†

PNAS | July 5, 2005 | vol. 102 | no. 27 | 9517–9522
The Community Effect: robust Wnt signaling in early sea urchin development

Ligand sharing – each cell receives $\frac{1}{4}$ of its ligands auto-catalytically, $\frac{3}{4}$ from its neighbors

Bolouri & Davidson, BioEssays, 2002
& Developmental Biology, 2010.
Mutual repression

\[
\frac{dG_1}{dt} = k_{t1}\left(\frac{K^N_{Diss1}}{K^N_{Diss1} + G_2^N}\right) - k_{d1}G_1
\]

\[
\frac{dG_2}{dt} = k_{t2}\left(\frac{K^N_{Diss2}}{K^N_{Diss2} + G_1^N}\right) - k_{d2}G_2
\]

Reducing the number of unknown parameters – a technique to simplify model exploration

re-write as:

\[
\frac{dG_1}{dt} = k_{t1}\left(\frac{1}{1 + \frac{G_2^N}{K^N_{Diss1}}}\right) - k_{d1}G_1
\]

\[
\frac{dG_2}{dt} = k_{t2}\left(\frac{1}{1 + \frac{G_1^N}{K^N_{Diss2}}}\right) - k_{d2}G_2
\]

If we measure \( G \) in units of \( K_{Diss} \), then

\[
\frac{dG_1}{dt} = k_{t1}\left(\frac{1}{1 + G_2^N}\right) - k_{d1}G_1
\]

\[
\frac{dG_2}{dt} = k_{t2}\left(\frac{1}{1 + G_1^N}\right) - k_{d2}G_2
\]
Mutual repression

\[ \frac{dG_1}{dt} = k_{t1} \left( \frac{1}{1 + G_2^N} \right) - k_{d1} \cdot G_1 \]

\[ \frac{dG_2}{dt} = k_{t2} \left( \frac{1}{1 + G_1^N} \right) - k_{d2} \cdot G_2 \]

Joshua L. Cherry and Frederick R. Adler

*J. theor. Biol.* (2000) 203, 117-133
Controlling the state of a mutual repression switch with 2 independent activating inputs

\[ k_1, k_2 = 5 \]

- \[ x_1 = \text{off}, x_2 = \text{on} \]
  - Inputs = 0
- \[ x_1 = \text{on}, x_2 = \text{off} \]
  - Inputs = 0
- \[ x_1 = \text{off}, x_2 = \text{on} \]
  - Input1 = 0, Input2 = 0.5
- \[ x_1 = \text{high}, x_2 = \text{low} \]
  - Input1 = 0.75, Input2 = 0.5
Controlling the state of a mutual repression switch with 2 independent activating inputs

$k_1, k_2 = 10$

$x_1 = 0.5 + 10/(1 + x_2^2)$

$x_2 = 0 + 10/(1 + x_1^2)$

$x_1 = 0 + 10/(1 + x_2^2)$

$x_2 = 0 + 10/(1 + x_1^2)$

$x_1 = 0.25 + 10/(1 + x_2^2)$

$x_2 = 0.5 + 10/(1 + x_1^2)$

Only steady state when $i_1 = 1, i_2 = 0.5$
Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells
Rekhtman et al, Genes & Dev, 1999 13:1398-1411

Erythroblasts

Myceloblasts

Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1

Hoogenkamp et al., "Molecular and Cellular Biology, Nov. 2007, p. 7425–7438"

-14kb conserved Sequence (URE)

-20kb

Promoter

PU.1

SFFV

C/EBP

PU.1/Elf-1/FLi1

Runx1 Runx1 Sp1

-14kb enhancer

Myeloid cells

B cells

T cells

Bold ChIP

Light EMSA


http://genomequebec.mcgill.ca/PReMod

Blanchette lab, McGill
Example PU.1 – Gata1 mutual repression models

Chickarmane V, Enver T, Peterson C (2009)

Bokes P, King JR, Loose M.

Huang et al, Developmental Biology 305 (2007) 695–713
developmental trajectory for neutrophils?

developmental trajectory for macrophages

Ikaros

PU.1

GFi

Ids

Egr

E2A

NAB2

Macrophage genes

B cell genes

The Better Way to Crunch data

- Use privately, or share data & R scripts
- Access the power of R through point & click menus
- Run your analysis on the Amazon cloud
  - free for small jobs, or
  - buy virtually unlimited resources directly from Amazon

Proceed as a guest user  Take a Tour

CRdata.org

Everyday researchers lose valuable time waiting for computations on their desktops. CRdata.org frees your desktop for other work while our servers analyze your data in the cloud.

You can use CRdata as a private resource, share your data and R scripts with selected others, or with the whole community.

To ensure you have access to the latest resources, all R and Bioconductor libraries are updated nightly on CRdata.