Navigating the Subway Map of the Cell

Jim Faeder
Department of Computational Biology
University of Pittsburgh School of Medicine

CMACS-NSF Meeting
March 4-5, 2010

faeder@pitt.edu
http://ccbb.pitt.edu/faeder
How Cells Process Information

Environment

Hormones, growth factors, etc.

Receptors

Nucleus

Architecture of a signaling network

Mutation of Ras Can Produce a Tumor Cell
>20% human tumors carry Ras point mutations.

>90% in pancreatic cancer.
Modularity of Signaling Proteins

Figure 6.10a  The Biology of Cancer (© Garland Science 2007)
Modularity produces complex wiring
Complexity of Receptor Complexes

Figure 6.9  The Biology of Cancer (© Garland Science 2007)
The “curse” of complexity

Number of States

<table>
<thead>
<tr>
<th>Number of States</th>
<th>Monomers</th>
<th>Dimers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3,888</td>
<td>7,560,216</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Number of States

<table>
<thead>
<tr>
<th>Number of States</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>3,888</td>
<td>3</td>
<td>19,440</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7,560,216</td>
<td>4</td>
<td>188,966,520</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
AIM: Model the biochemical machinery by which cells process information (and respond to it).

**Representation**

BIONETGEN Language
kappa
etc.

**Simulation**

ODE, PDE
Stochastic Simulation Algorithm
Kinetic Monte Carlo
Brownian dynamics
Syk activation model

Key variables
• ligand properties
• protein expression levels
• multiple Lyn-FceRI interactions
• transphosphorylation

Mol. Immunol., 2002
J. Immunol., 2003
Defining Molecules

BIONETGEN Language

IgE(a,a)
FceRI(a,b~U~P,g2~U~P)
Lyn(U,SH2)
Syk(tSH2,1Y~U~P,aY~U~P)
Defining Interaction Rules

**BIO NET GEN Language**

\[
\text{IgE}(a,a) + \text{FceRI}(a) \leftrightarrow \text{IgE}(a,a^{11}).\text{FceRI}(a^{11}) \]

... binding and dissociation

\[
\text{Lyn}(U^{11}).\text{FceRI}(b^{11}).\text{FceRI}(b^{~U}) \rightarrow \text{Lyn}(U^{11}).\text{FceRI}(b^{11}).\text{FceRI}(b^{~P})
\]

component state change
Rule-based modeling protocol

- Objects and rules
- Reaction Network
- ODE Solver
- Stochastic Simulator (Gillespie)
- Output

```
http://bionetgen.org
```

“Normal Cell”
“Mutants”
BIONETGEN Editor - BiNGE

Yao Sun and Liz Marai, U. Pitt Computer Science
**BIONetGEN Editor - BiNGE**

![Image of the BIONetGEN Editor with a BGL file opened and showing the console output]

- **Save**
- **Save All**
- **Find**
- **Replace**
- **Contact Map**
- **Influence Map**
- **Check**
- **Run**
- **Par Scan**

**SimpleExampleExtended.bngl**

```plaintext
begin molecule types
begin seed species
EGF(R)   EGF0
EGFR(L,C,R1,Y1068-U) EGFRO
Grb2(SH2,SH3)  GRB20
Sos1(PxxP)  SOS10
Sos1(SxxS)  SOS11
egfr_simple.bngl
end parameters
```

**Console**

- Propagation took 1.10e-01 CPU seconds
- Final network file written to D:\BNGModels\SimpleExample\SimpleExampleExtended_ssa_end.net
- Program times: 0.17 CPU s 0.00 clock s
- Edge species became populated 0 times.
BIONETGEN Editor - BiNGE
Limits of the network generation approach

- Extending model to include Lyn regulation results in $>20,000$ species.
Kohn’s Wiring Diagram for the Cell

Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)
NF SIM
“Network-Free” Stochastic Simulator

- Generalization of rule-based kinetic Monte Carlo method of Yang et al.
- Particle-based method avoids combinatorial explosion
- Gillespie-based simulations capture stochastic effects

Sneddon, Faeder, and Emonet, in preparation.
Integration with BioNetGen
Subway Map of Cell Signaling

Hanahan and Weinberg, 2000
Rule-based Model of EGFR Signaling

Preliminary Model: 20 molecules / 532 rules / 496 parameters

Matt Creamer and Rich Posner
Stats

Model
- 20 Molecule Types
  - 4 Receptors
  - 3 Ligands
- 536 Parameters
- 547 Reaction Rules

Simulation
1500 sim sec
- ~10-18 million events
- ~ 1060 real sec
~ 6e-5 CPU seconds/event
(On a 2.4 GHz Intel Core2Duo on iMac with 4 GB RAM)
Model Validation

Model Steady State Response to EGF

John Sekar
Model Validation

Model Steady State Response to EGF

Basal activity is too high!
Stop # 2: TGF-β Pathway

Cell cycle model also under development
The Path Ahead

- Continue to build and analyze models of key pathways
- Systematic investigation of models using
  - Statistical and Bayesian Model Checking
  - Global parameter sensitivity analysis
  - Parameter estimation and synthesis
- Integration of pathway models
- Model reduction
  - Coarse-graining of detailed models (bottom up)
  - Comparison / Mapping to logical models (top down)
The Path Ahead

- Continue to build and analyze models of key pathways
- Systematic investigation of models using
  - Statistical and Bayesian Model Checking
  - Global parameter sensitivity analysis
  - Parameter estimation and synthesis
- Integration of pathway models
- Model reduction
  - Coarse-graining of detailed models (bottom up)
  - Comparison / Mapping to logical models (top down)

Can Abstract Interpretation provide powerful new approaches to this problem? Danos, Feret and colleagues
Boolean networks

- The **state of an element** in the signaling network can be described by a **Boolean variable**, expressing that it is:
  - Active or present (on or ‘1’)
  - Inactive or absent (off or ‘0’)

- **Boolean functions:**
  - Represent interactions between elements
  - The state of an element is calculated from states of other elements

- **Practical advantages**
  - No parameters – facilitates model development
  - Easy to understand – facilitates collaboration
Model development protocol

1. Experiments
2. Reading and Discussion
3. Model Formulation and Annotation
4. Model Testing

Outside Experts

bi-weekly

Product Model
Logical modeling - example

Biological network

Proteins: $p_1, p_2, p_3$
Protein states: $x_1, x_2, x_3$

Boolean network

$x_1(t+1) = x_2(t)$ or $x_3(t)$
$x_2(t+1) = \text{not } x_1(t)$ and $x_3(t)$
$x_3(t+1) = x_1(t)$ and $\text{not } x_3(t)$
Logical modeling - example

• $x_1x_2x_3$ — state vector

Boolean network

Logic circuit network

State transition table

<table>
<thead>
<tr>
<th>state</th>
<th>$x_1(t)x_2(t)x_3(t)$</th>
<th>$x_1(t+1)x_2(t+1)x_3(t+1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>000</td>
<td>000</td>
</tr>
<tr>
<td>$s_2$</td>
<td>001</td>
<td>110</td>
</tr>
<tr>
<td>$s_3$</td>
<td>010</td>
<td>100</td>
</tr>
<tr>
<td>$s_4$</td>
<td>011</td>
<td>110</td>
</tr>
<tr>
<td>$s_5$</td>
<td>100</td>
<td>001</td>
</tr>
<tr>
<td>$s_6$</td>
<td>101</td>
<td>100</td>
</tr>
<tr>
<td>$s_7$</td>
<td>110</td>
<td>101</td>
</tr>
<tr>
<td>$s_8$</td>
<td>111</td>
<td>100</td>
</tr>
</tbody>
</table>

$x_1(t+1) = x_2(t)$ or $x_3(t)$
$x_2(t+1) = \text{not } x_1(t) \text{ and } x_3(t)$
$x_3(t+1) = x_1(t)$ and $\text{not } x_3(t)$
Logical modeling - example

Boolean network

$x_1(t+1) = x_2(t) \text{ or } x_3(t)$

$x_2(t+1) = \text{not } x_1(t) \text{ and } x_3(t)$

$x_3(t+1) = x_1(t) \text{ and } \text{not } x_3(t)$

Logic circuit network

State transition diagram
A sequence of connected states forms a **trajectory** of the system.

The number of states and the number of trajectories in the state space are **finite**.

All initial states of a trajectory will eventually reach a **steady state** or a **state cycle**.
Synchronous vs. asynchronous updates

Synchronous updates:
- \( x_1(t+1) = x_2(t) \) or \( x_3(t) \)
- \( x_2(t+1) = \text{not } x_1(t) \) and \( x_3(t) \)
- \( x_3(t+1) = x_1(t) \) and \( \text{not } x_3(t) \)

Asynchronous updates:

State transition table:

<table>
<thead>
<tr>
<th>State</th>
<th>( x_1, x_2, x_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>000</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>001</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>010</td>
</tr>
<tr>
<td>( s_4 )</td>
<td>011</td>
</tr>
<tr>
<td>( s_5 )</td>
<td>100</td>
</tr>
<tr>
<td>( s_6 )</td>
<td>101</td>
</tr>
<tr>
<td>( s_7 )</td>
<td>110</td>
</tr>
<tr>
<td>( s_8 )</td>
<td>111</td>
</tr>
</tbody>
</table>
Regulatory graph for mammalian cell cycle network

Source: Faure et al., Bioinformatics, 2006.
## Logical rules

<table>
<thead>
<tr>
<th>Product</th>
<th>Logical rules leading to an activity of the product</th>
<th>Justification/References</th>
</tr>
</thead>
<tbody>
<tr>
<td>CycD</td>
<td>CycD</td>
<td>CycD is an input, considered as constant.</td>
</tr>
<tr>
<td>Rb</td>
<td>$(\neg CycD \land CycE \land CycA \land CycB)$ $\lor$ $(p27 \land CycD \land CycB)$</td>
<td>Rb is expressed in the absence of the cyclins, which inhibit it by phosphorylation (Novak and Tyson, 2004; Taya, 1997); it can be expressed in the presence of CycE or CycA if their inhibitory activity is blocked by p27 (Coqueret, 2003).</td>
</tr>
<tr>
<td>E2F</td>
<td>$(Rb \land CycA \land CycB)$ $\lor$ $(p27 \land \neg Rb \land CycB)$</td>
<td>E2F is active in the absence of Rb, that blocks E2F self-transcriptional activation (Helin, 1998), and in the absence of CycA and CycB, that inhibit E2F (Novak and Tyson, 2004); CycA may be present, if its inhibitory activity is blocked by p27 (Coqueret, 2003).</td>
</tr>
<tr>
<td>CycE</td>
<td>$(E2F \land \neg Rb)$</td>
<td>CycE activity requires the presence of E2F and the absence of Rb (Helin, 1998).</td>
</tr>
<tr>
<td>CycA</td>
<td>$(E2F \land \neg Rb \land Cdc20 \land (Cdh1 \land Ubc))$ $\lor$ $(CycA \land \neg Rb \land Cdc20 \land (Cdh1 \land Ubc))$</td>
<td>The transcription of CycA is activated by E2F in the absence of Rb, which blocks this activation (Helin, 1998), in the absence of Cdc20, as well as of the pair formed by Cdh1 and UbcH10, which both lead to the degradation of CycA (Harper et al., 2002; Rape and Kirschner, 2004); CycA is stable in the absence of its inhibitors Rb, Cdc20, and of the pair Cdh1 and UbcH10.</td>
</tr>
<tr>
<td>p27</td>
<td>$(\neg CycD \land CycE \land CycA \land CycB)$ $\lor$ $(p27 \land (CycE \land CycA) \land CycB \land \neg CycD)$</td>
<td>p27 is active in the absence of the cyclins; when p27 is already present, it blocks the action of CycE or CycA (but not both of them) by sequestration (Coqueret, 2003).</td>
</tr>
<tr>
<td>Cdc20</td>
<td>CycB</td>
<td>CycB indirectly activates Cdc20 (Harper et al., 2002).</td>
</tr>
<tr>
<td>Cdh1</td>
<td>$(CycA1 \land \neg CycB)$ $\lor$ $(Cdc20 \land p27 \land CycB)$</td>
<td>The activity of Cdh1 requires the absence of CycB and CycA, which inhibit it by phosphorylation (Harper et al., 2002); Cdc20 further activates Cdh1. (Novak and Tyson, 2004); p27 allows the presence of CycA, by blocking its activity.</td>
</tr>
<tr>
<td>UbcH10</td>
<td>$(\neg Cdh1) \lor (Cdh1 \land Ubc \land (Cdc20 \lor CycA \lor CycB))$</td>
<td>UbcH10 is active in the absence of Cdh1; this UbcH10 activity can be maintained in the presence of Cdh1 when at least one of its other targets is present (CycA, Cdc20, or CycB) (Rape and Kirschner, 2004).</td>
</tr>
<tr>
<td>CycB</td>
<td>$(Cdc20 \land \neg Cdh1)$</td>
<td>CycB is active in the absence of both Cdc20 and Cdh1, which target CycB for destruction (Harper et al., 2002).</td>
</tr>
</tbody>
</table>

*Source: Faure et al., Bioinformatics, 2006.*
Updating approaches

Source: Faure et al., Bioinformatics, 2006.
Vision

- Logical models of subway map components
- Begin with cell cycle models and link with other regulatory pathways connected to receptor signaling
- Basis for both simulation and formal analysis
- Complement to reaction network models being developed with TGEN collaborators
Collaborators

**FaederLab**
Natasa Miskov-Zivanov
John Sekar
Leonard Harris
Justin Hogg
Jintao Liu

**TGen**
Rich Posner
Matthew Creamer
Josh Colvin
Daniel Von Hoff

**CMU**
Ed Clarke
Haijun Gong
Paolo Zuliani
Anvesh Komuravelli
Chris Langmead
Sumit Jha

**Yale**
Thierry Emonet
Michael Sneddon

**CMU**
Nancy Griffeths

**Lehmann**

---

http://bionetgen.org
Thank You!

Photo by John Sekar