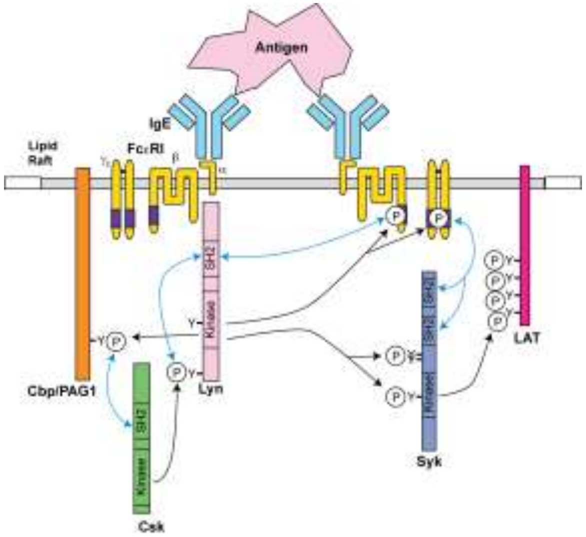
Navigating the Subway Map of the Cell



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

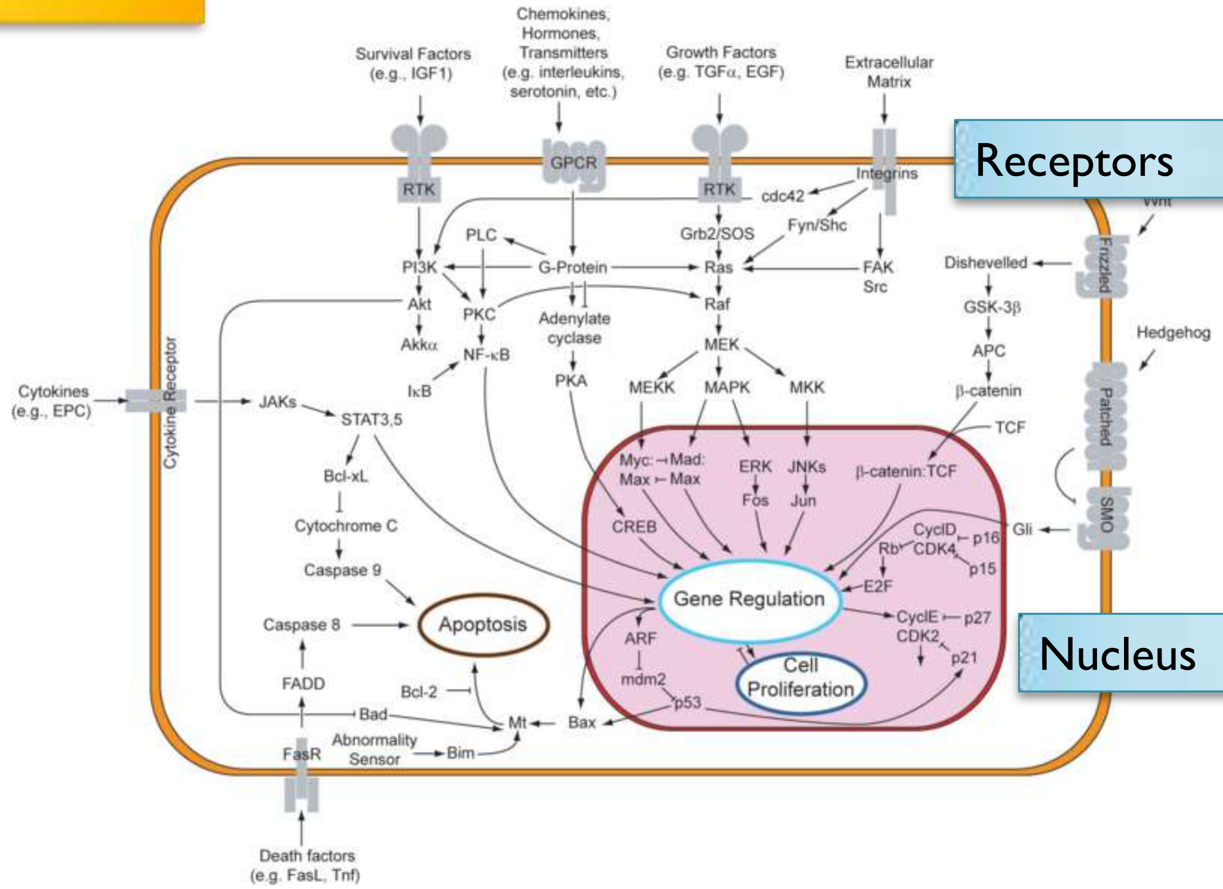
CMACS-NSF Meeting
March 4-5, 2010

faeder@pitt.edu
http://ccbb.pitt.edu/faeder

How Cells Process Information

Environment

Hormones, growth factors, etc.

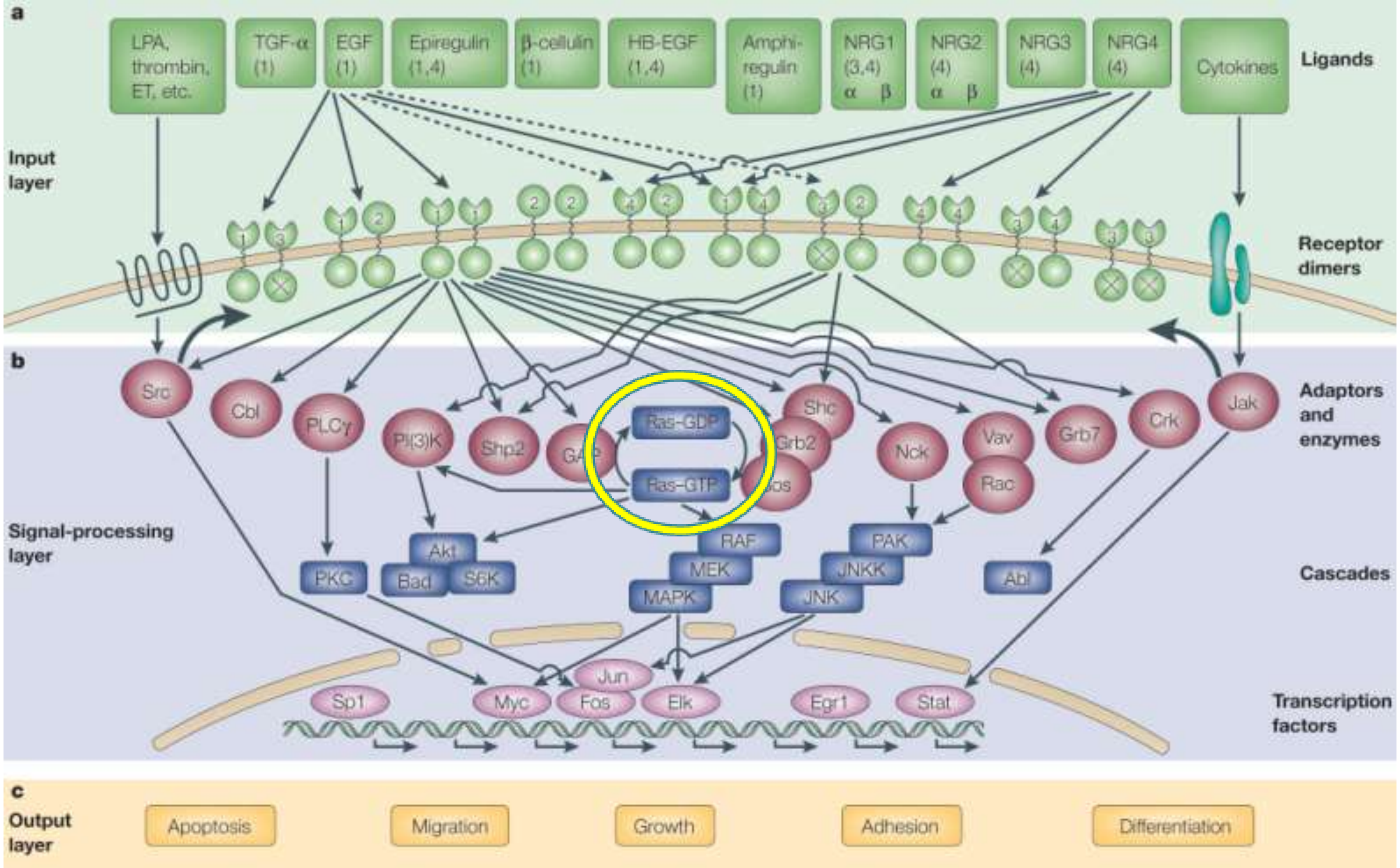


Receptors

Nucleus

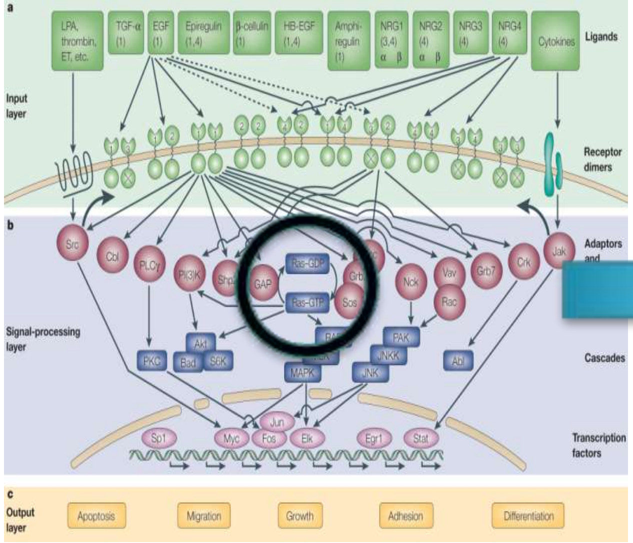
http://en.wikipedia.org/wiki/Cell_signaling

Architecture of a signaling network

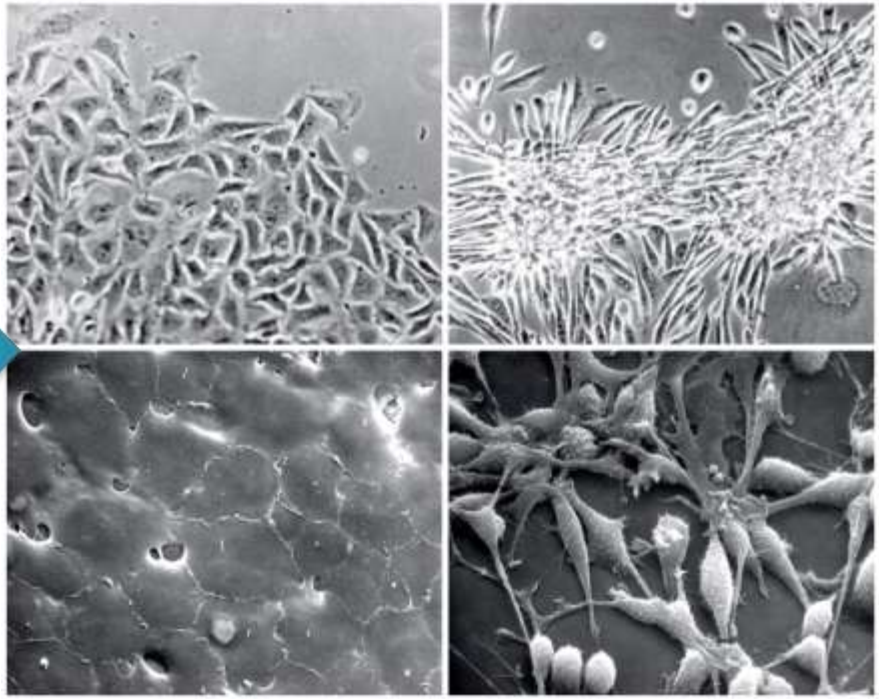


Yarden & Sliwkowski, *Nature Rev. Mol. Cell Biol.* **02**: 127-137 (2001).

Mutation of Ras Can Produce a Tumor Cell

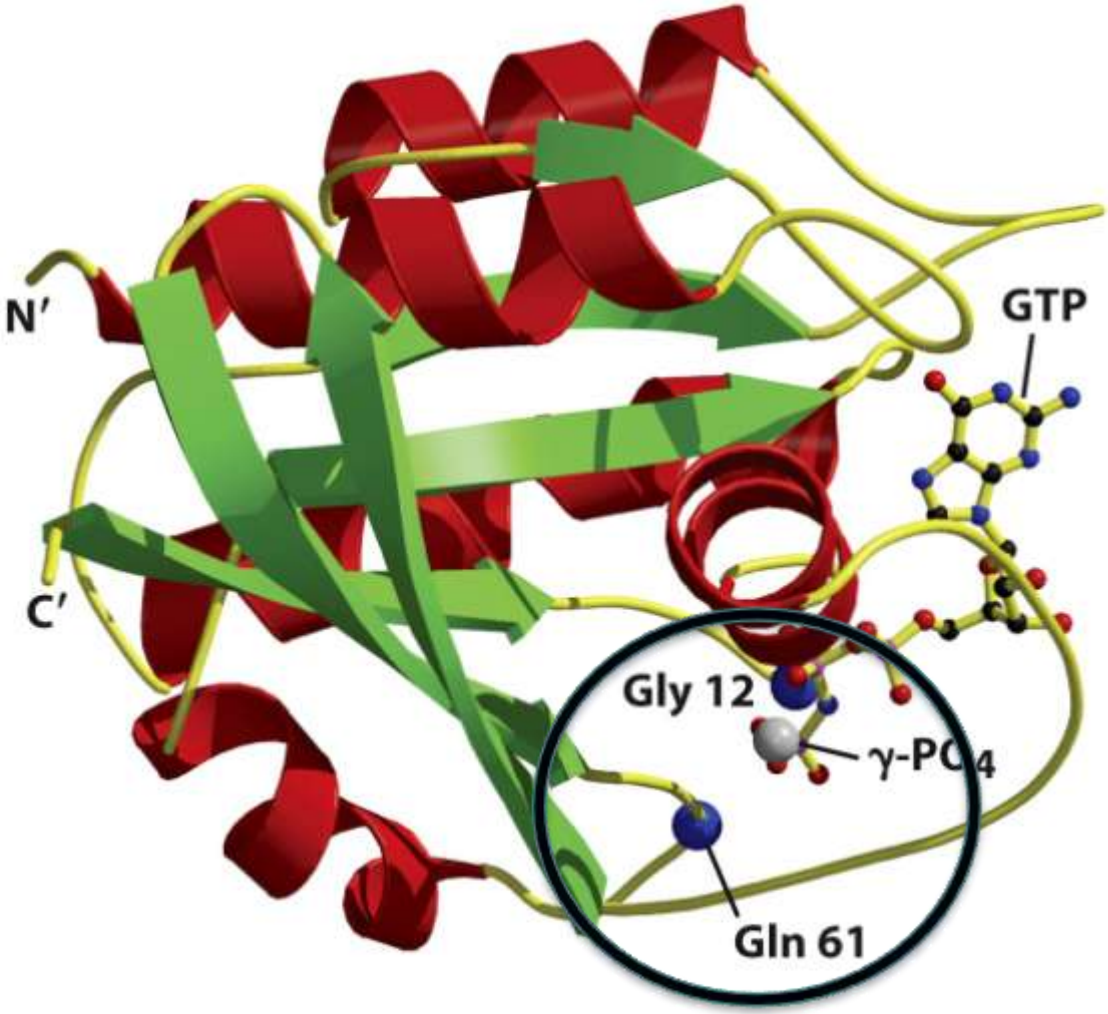


Normal **Transformed**



Ras mutations in cancer

Ras



>20% human tumors carry Ras point mutations.

>90% in *pancreatic cancer.*

The Biology of Cancer (© Garland Science 2007)

Modularity of Signaling Proteins

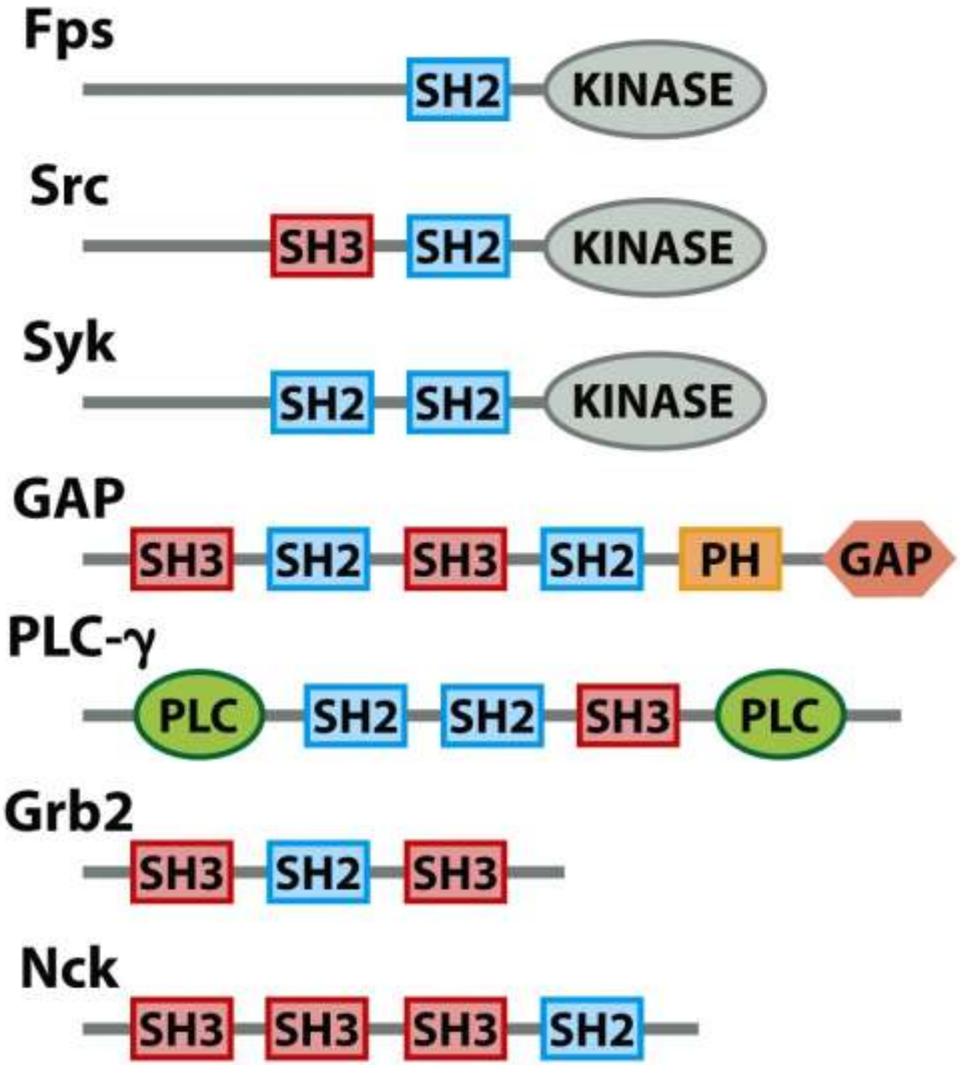
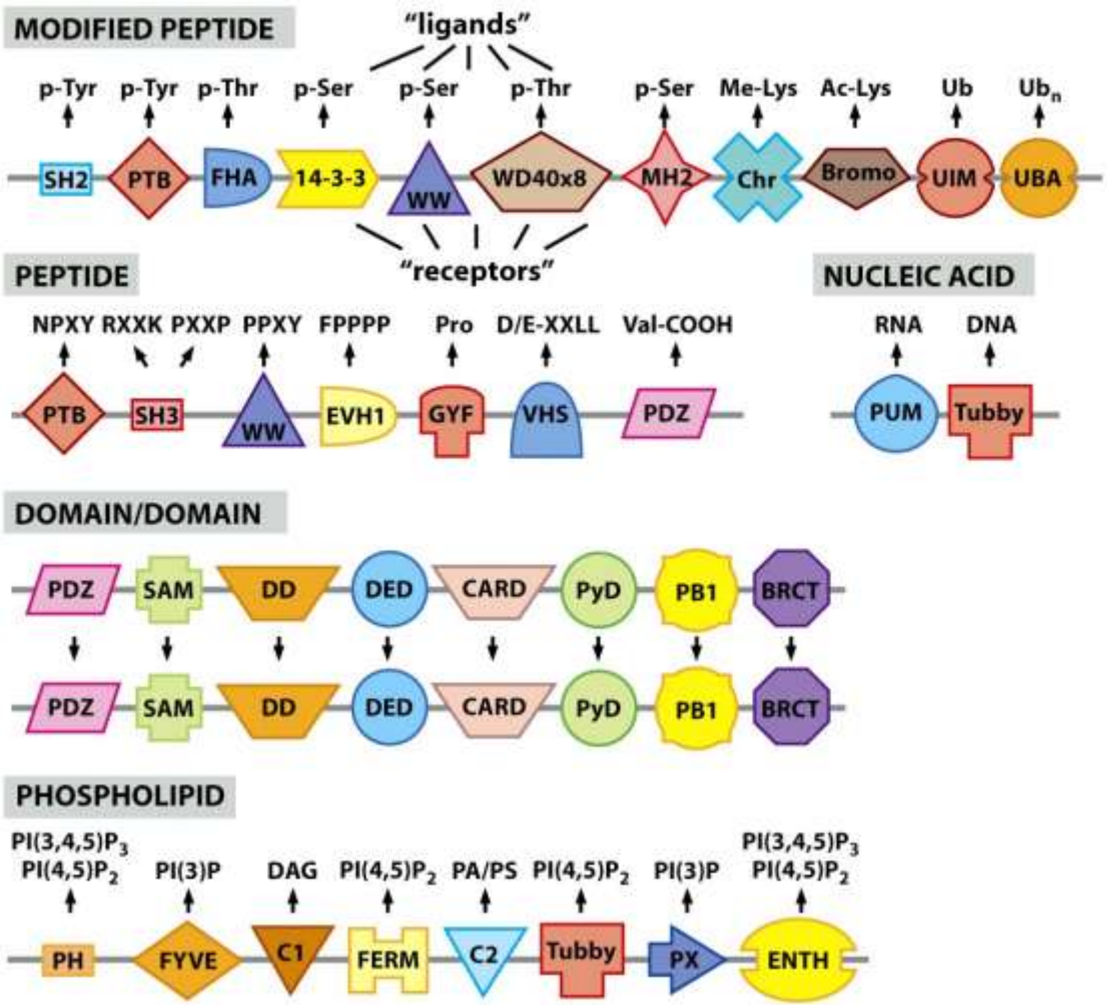


Figure 6.10a *The Biology of Cancer* (© Garland Science 2007)

Modularity produces complex wiring



Complexity of Receptor Complexes

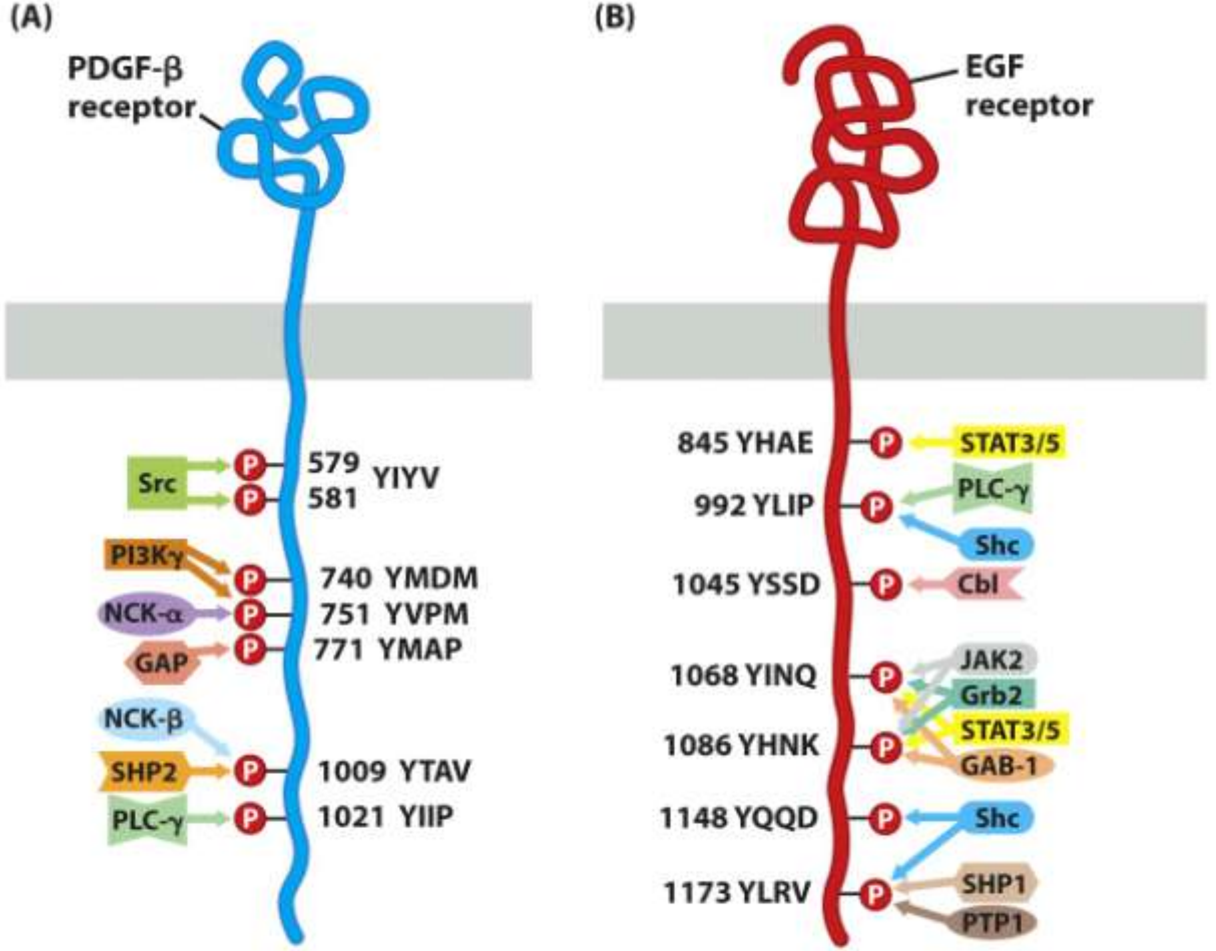
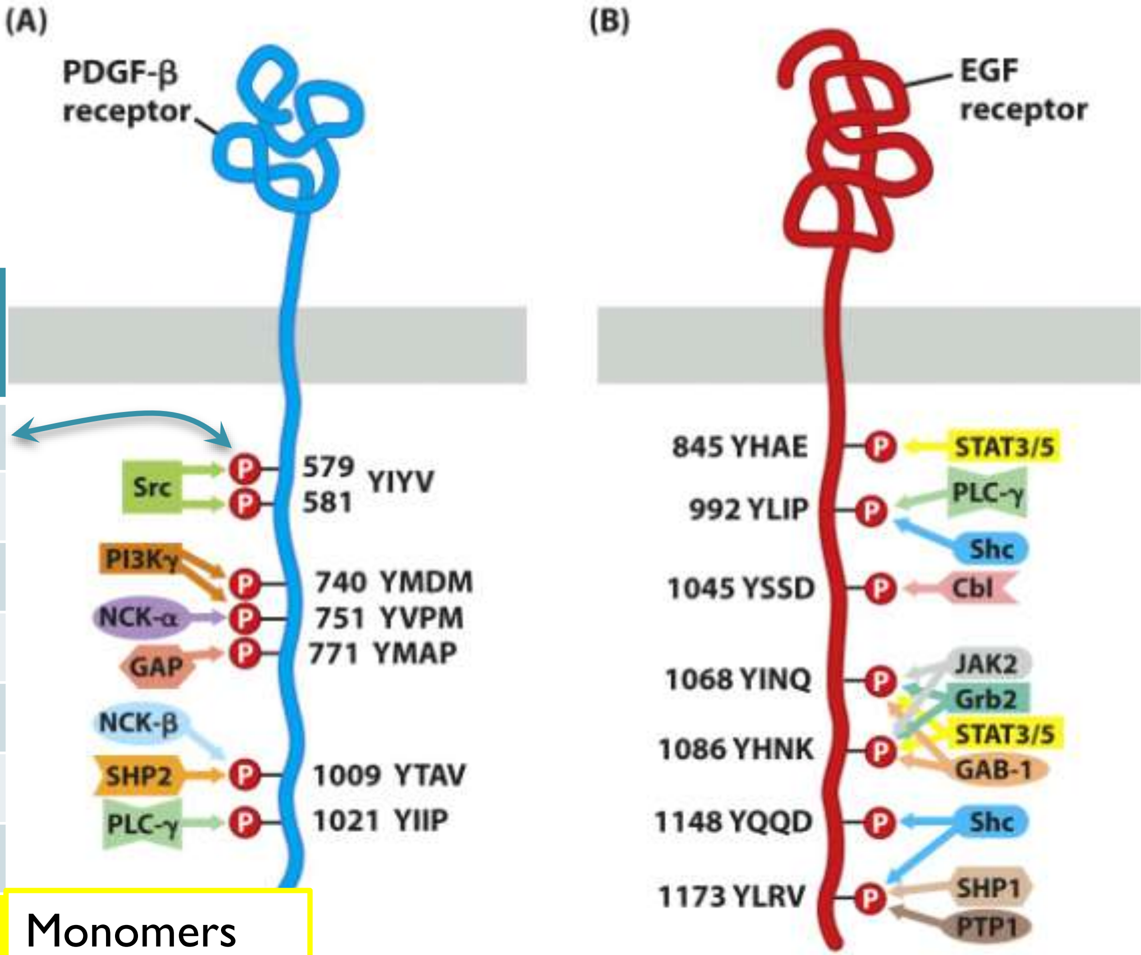


Figure 6.9 *The Biology of Cancer* (© Garland Science 2007)

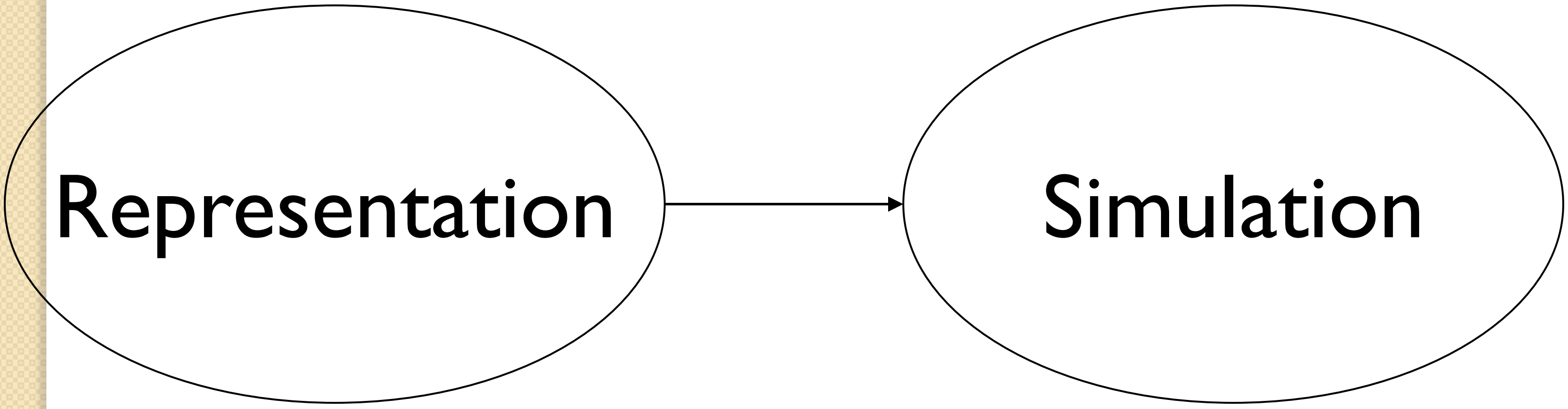
The “curse” of complexity



Number of States		Number of States
3		3
3		4
3		3
4		6
3		6
4		3
3		5
3,888	Monomers	19,440
7,560,216	Dimers	188,966,520

Modeling cell signaling

AIM: Model the biochemical machinery by which cells process information (and respond to it).



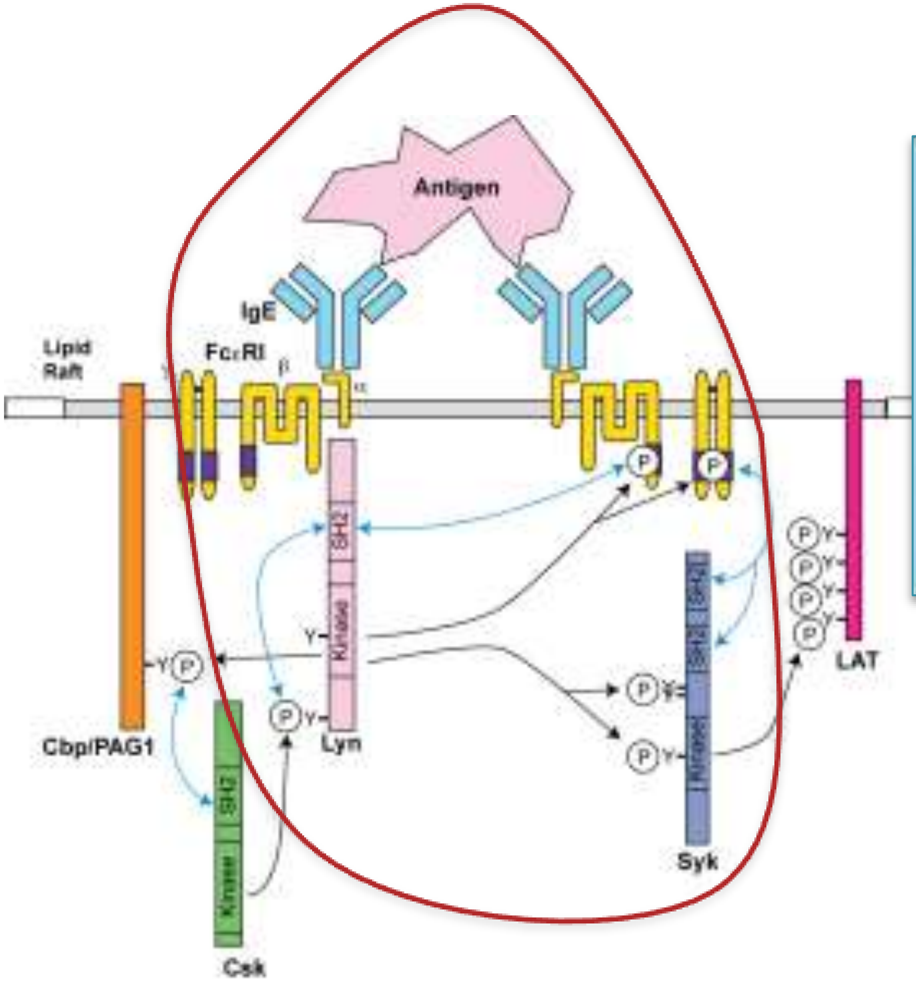
Representation

Simulation

BioNetGen Language
kappa
etc.

ODE, PDE
Stochastic Simulation Algorithm
Kinetic Monte Carlo
Brownian dynamics

Syk activation model

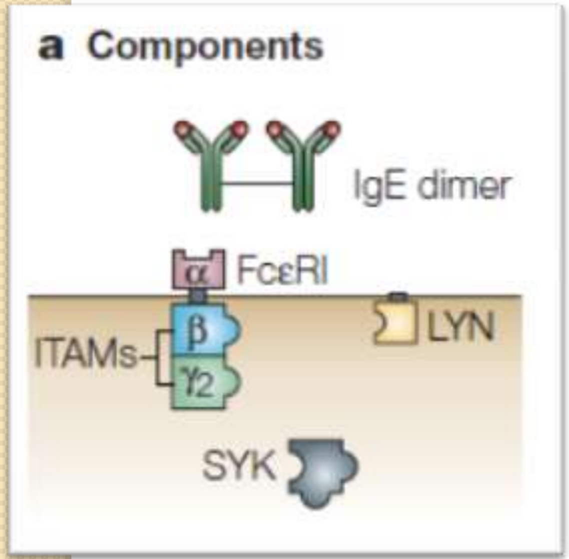


Key variables

- ligand properties
- protein expression levels
- multiple Lyn-FceRI interactions
- transphosphorylation

Mol. Immunol.,2002
J. Immunol., 2003

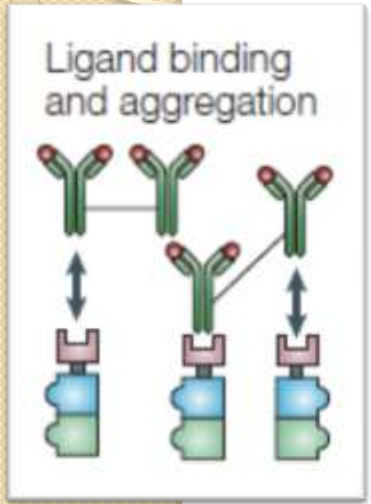
Defining Molecules



BioNetGen Language

```
IgE(a,a)
FceRI(a,b~U~P,g2~U~P)
Lyn(U,SH2)
Syk(tSH2,lY~U~P,aY~U~P)
```

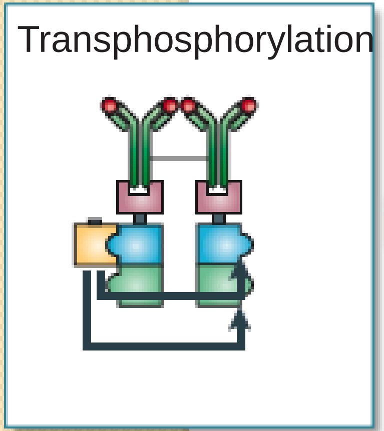
Defining Interaction Rules



BioNetGen Language

```
IgE(a,a)+ FceRI(a)<-> IgE(a,a!1).FceRI(a!1)
…
```

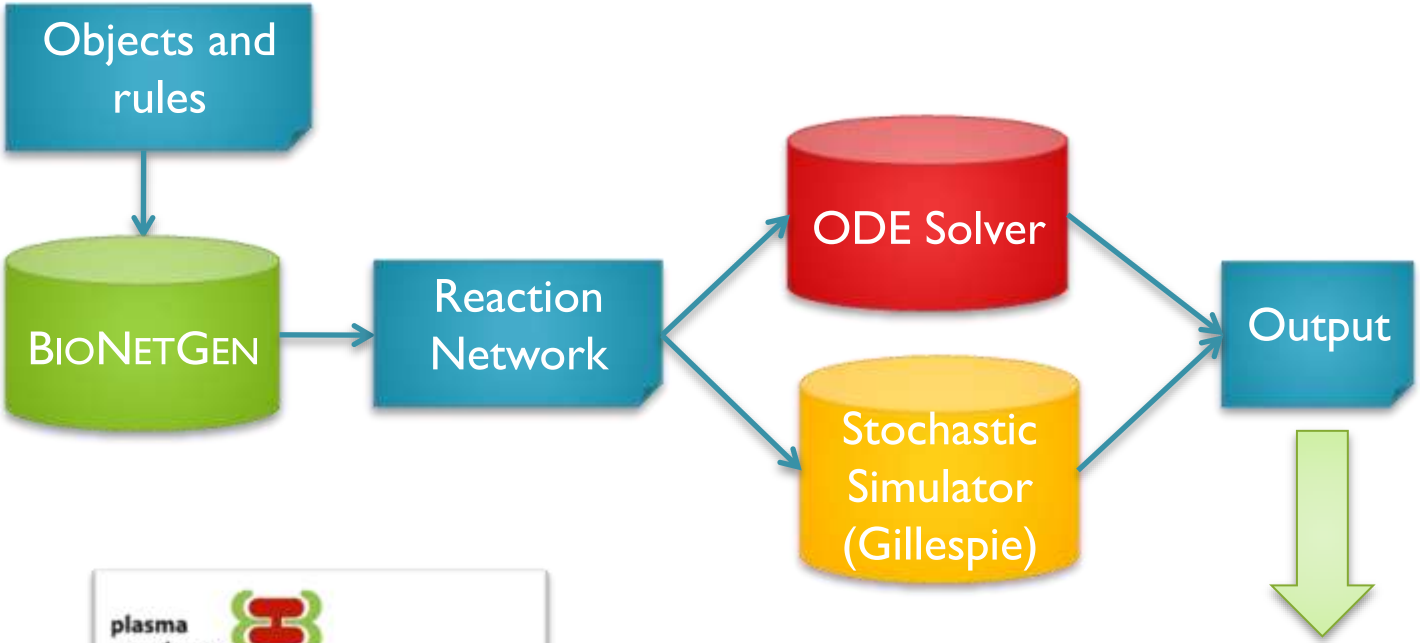
binding and dissociation



```
 Lyn(U!1).FceRI(b!1).FceRI(b~U)-> \
 Lyn(U!1).FceRI(b!1).FceRI(b~P)
```

component state change

Rule-based modeling protocol



Objects and rules

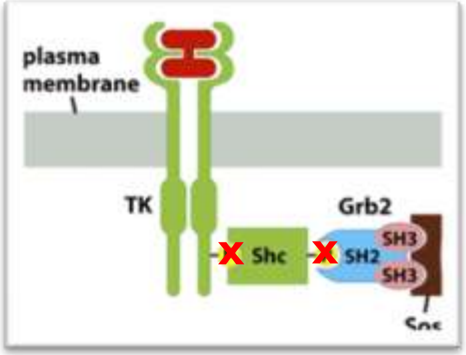
BioNetGen

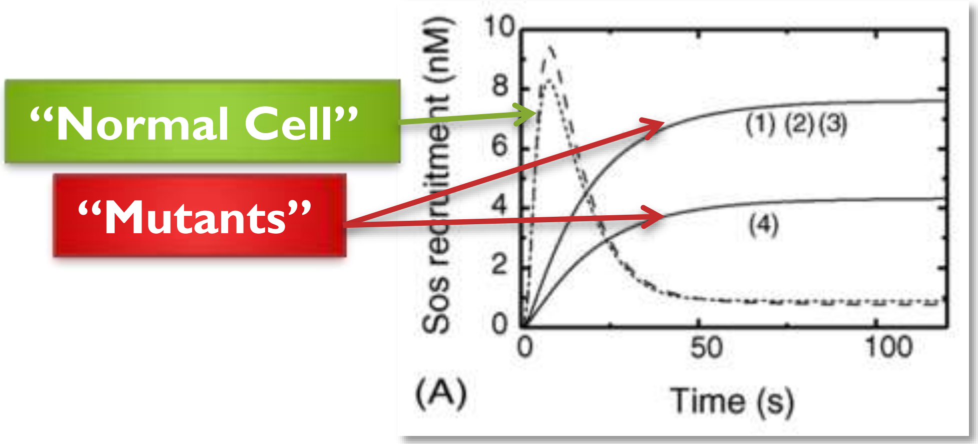
Reaction Network

ODE Solver

Stochastic Simulator (Gillespie)

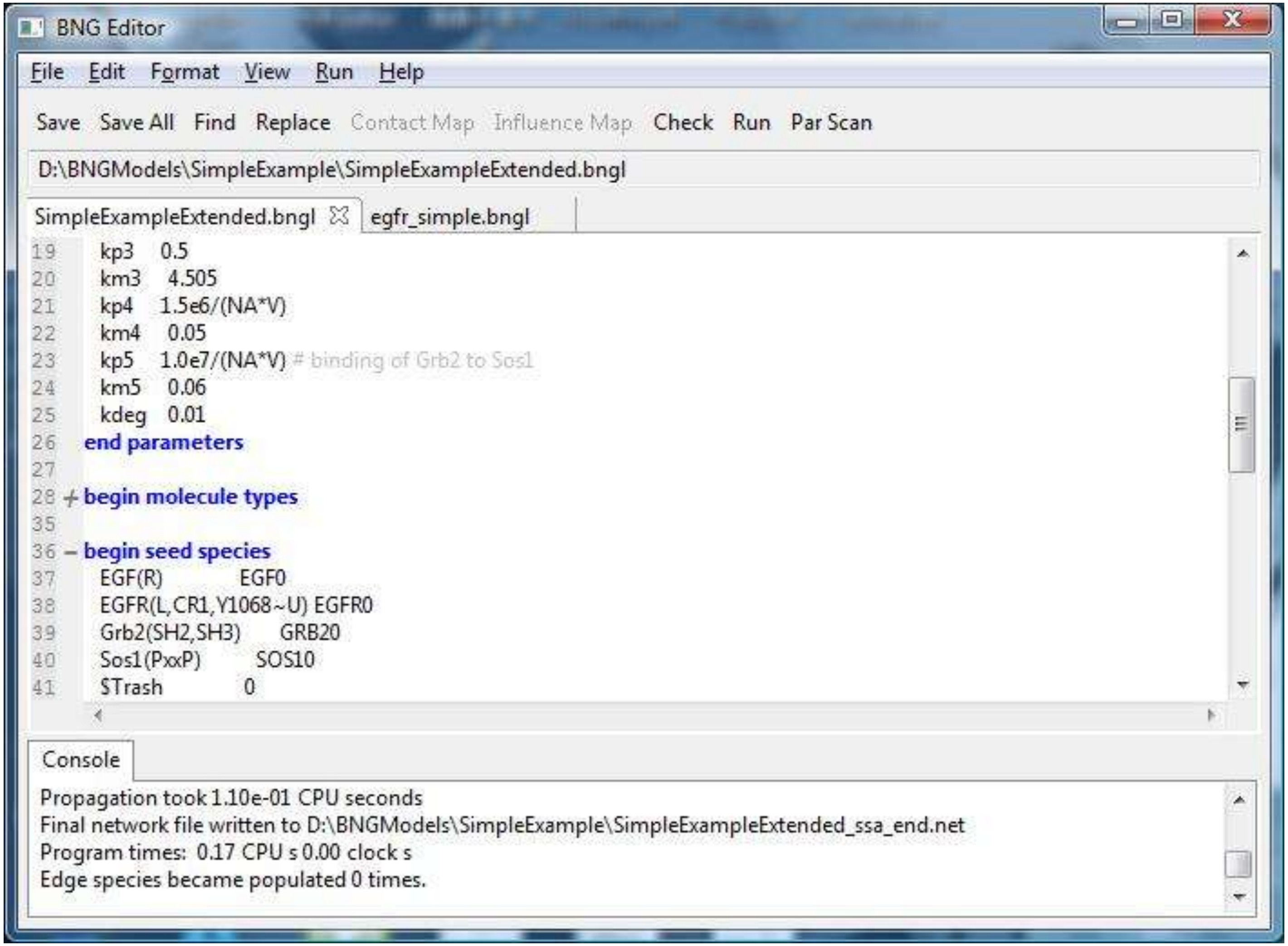
Output





http://bionetgen.org

BIONETGEN Editor - BiNGE

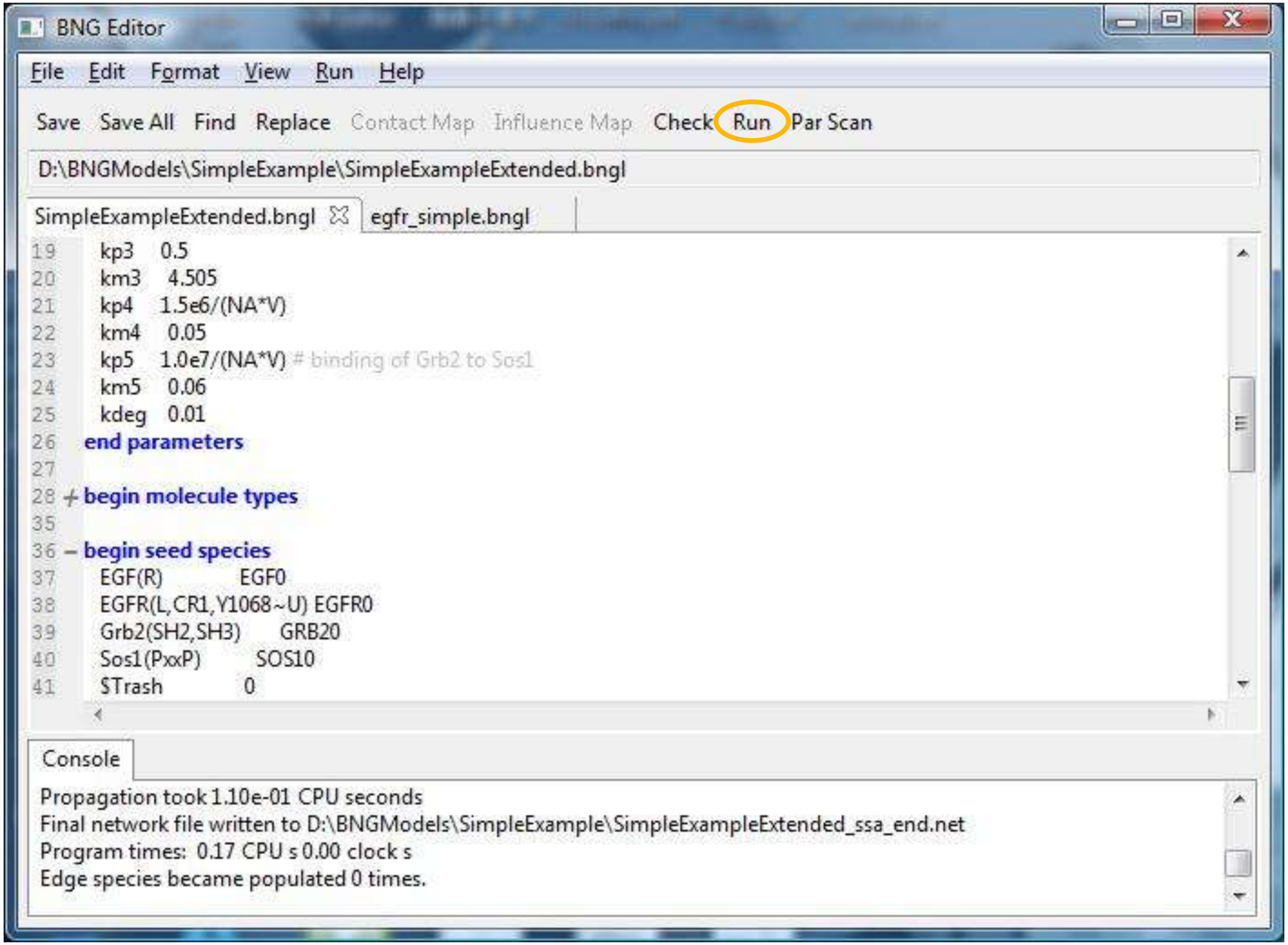


```
    kp3   0.5
    km3    4.505
    kp4   1.5e6/(NA*V)
    km4    0.05
    kp5   1.0e7/(NA*V) # binding of Grb2 to Sos1
    km5    0.06
    kdeg   0.01
end parameters

begin molecule types

begin seed species
    EGF(R)            EGF0
    EGFR(L,CR1,Y1068~U) EGFR0
    Grb2(SH2,SH3)       GRB20
    Sos1(PxxP)        SOS10
    $Trash            0
```

```
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.
```

Yao Sun and Liz Marai, U. Pitt Computer Science

BIONETGEN Editor - BiNGE

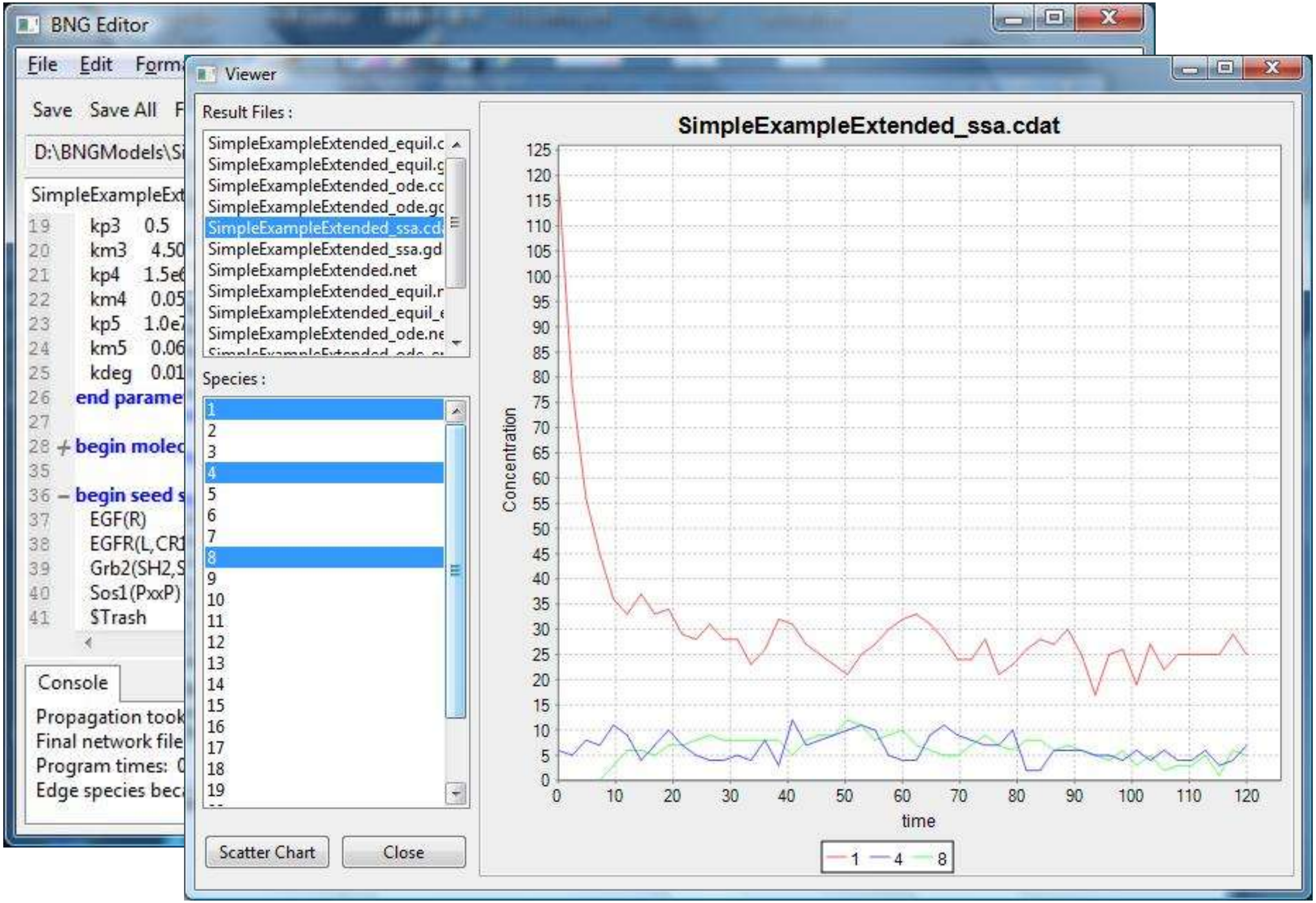


D:\BNGModels\SimpleExample\SimpleExampleExtended.bngl

```
  kp3   0.5
  km3   4.505
  kp4   1.5e6/(NA*V)
  km4   0.05
  kp5   1.0e7/(NA*V) # binding of Grb2 to Sos1
  km5   0.06
  kdeg  0.01
end parameters

begin molecule types

begin seed species
  EGF(R)             EGF0
  EGFR(L,CR1,Y1068~U) EGFR0
  Grb2(SH2,SH3)      GRB20
  Sos1(PxxP)         SOS10
  $Trash             0
```

```
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



```
D:\BNGModels\Si
SimpleExampleExt
kp3   0.5
km3   4.50
kp4   1.5e6
km4   0.05
kp5   1.0e7
km5   0.06
kdeg  0.01
end parame
+ begin molec
- begin seed s
  EGF(R)
  EGFR(L,CR1
  Grb2(SH2,S
  Sos1(PxxP)
  $Trash
```

Console

```
Propagation took
Final network file
Program times: 0
Edge species beca
```

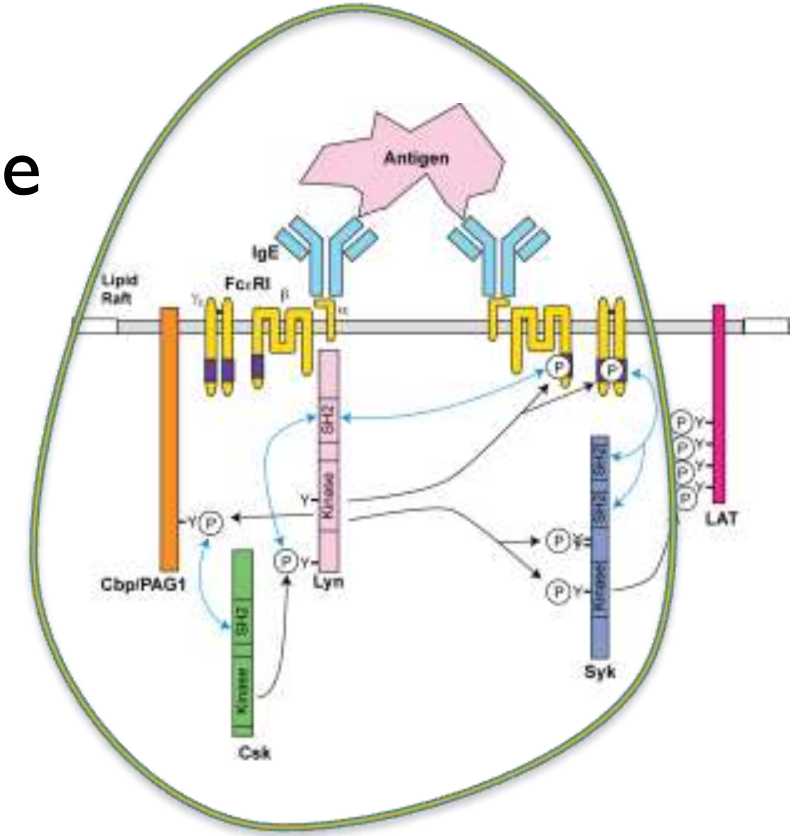
Result Files :

- SimpleExampleExtended_equil.c
- SimpleExampleExtended_equil.g
- SimpleExampleExtended_ode.cc
- SimpleExampleExtended_ode.gc
- SimpleExampleExtended_ssa.cd
- SimpleExampleExtended_ssa.gd
- SimpleExampleExtended.net
- SimpleExampleExtended_equil.r
- SimpleExampleExtended_equil_e
- SimpleExampleExtended_ode.ne

Species : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

Limits of the network generation approach

- Extending model to include Lyn regulation results in >20,000 species.



Kohn's Wiring Diagram for the Cell

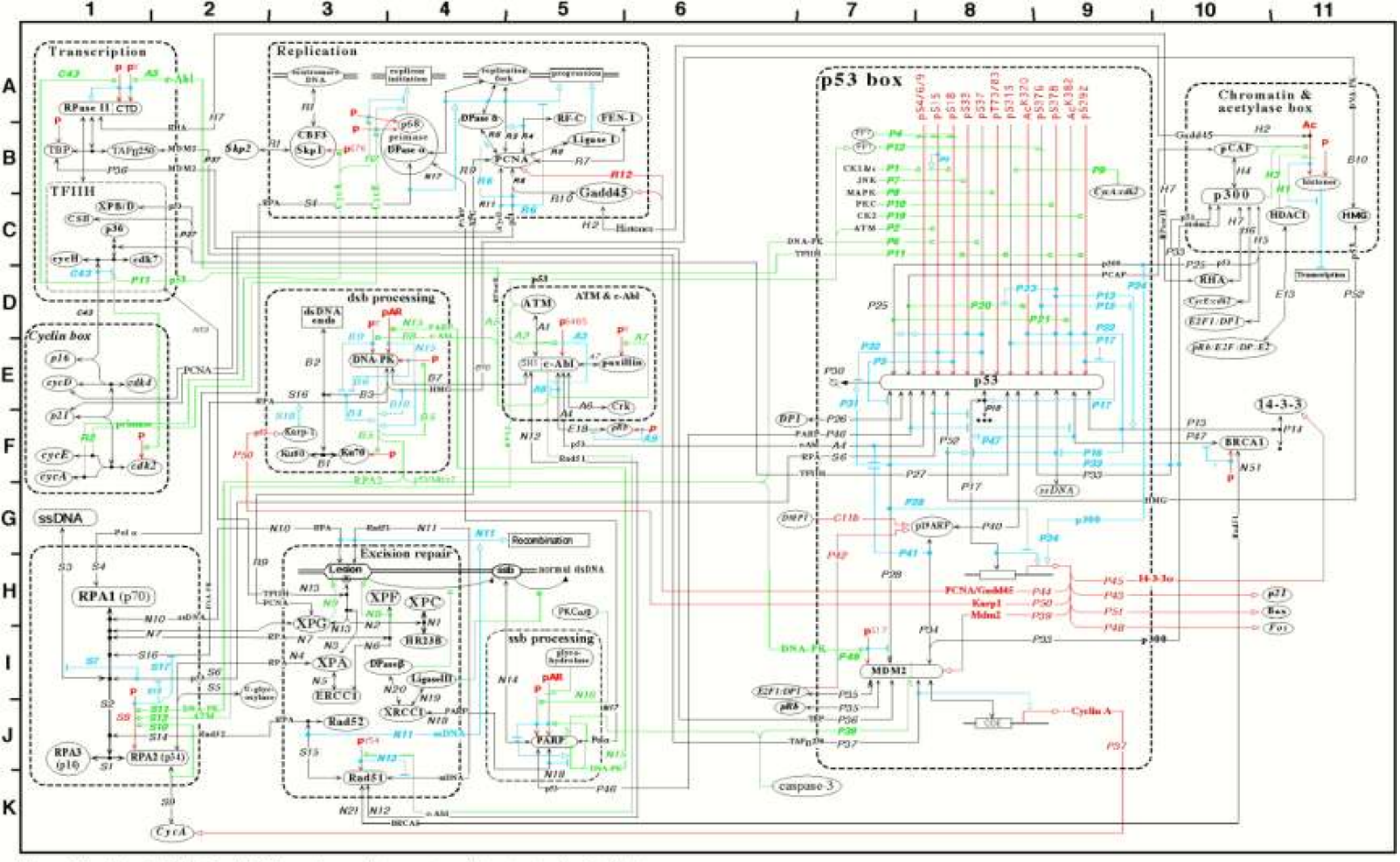


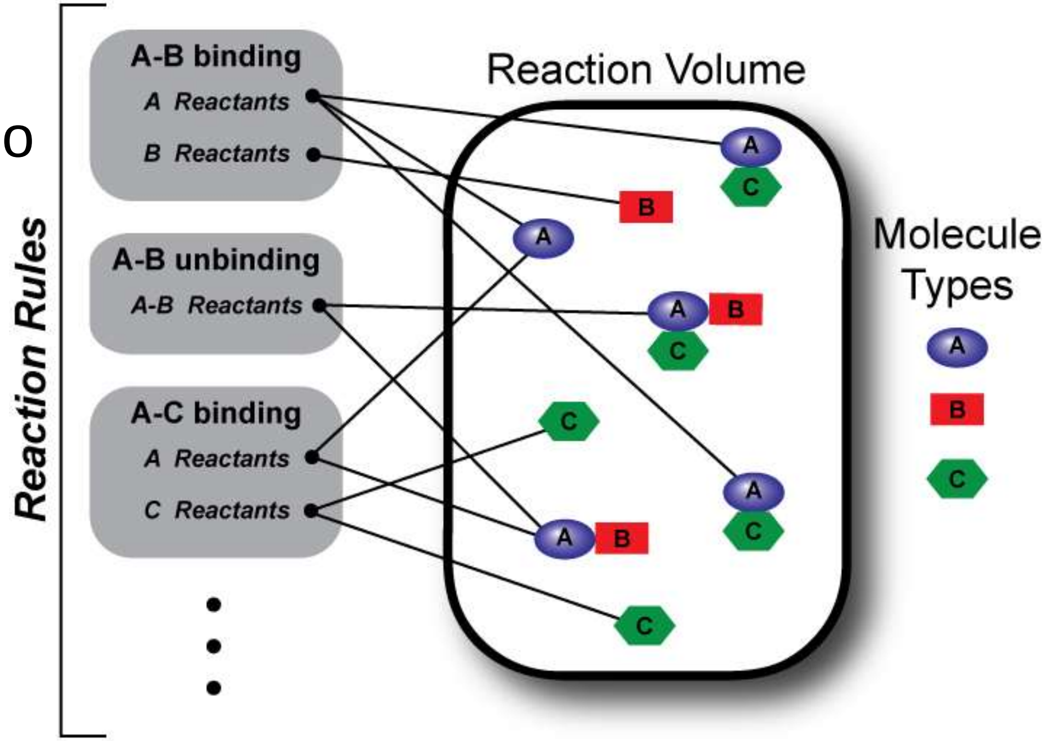
Figure 6B: The p53-Mdm2 and DNA repair regulatory network (version 2p - May 19, 1999)

Kohn, Molecular Biology of the Cell 1999

NFSIM

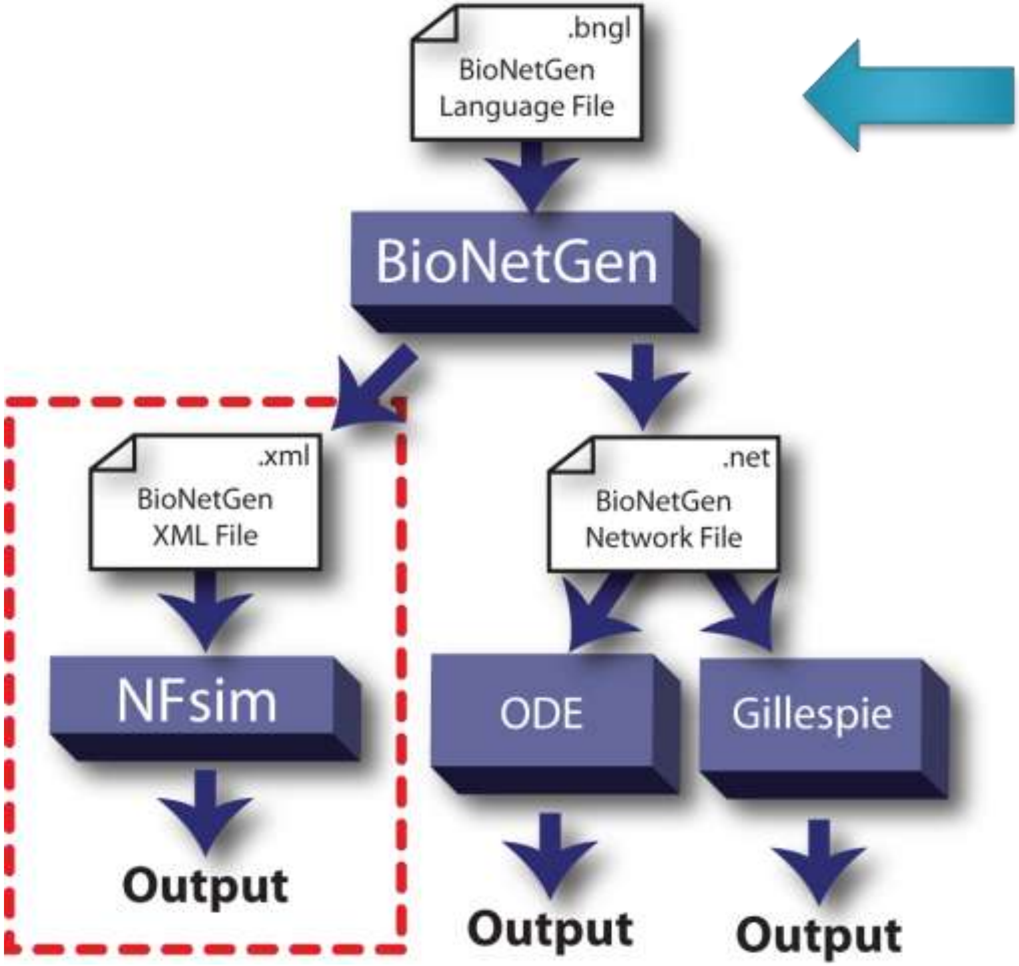
"Network-Free" Stochastic Simulator

- 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



.bngl BioNetGen Language File

BioNetGen

.xml BioNetGen XML File

NFsim

Output

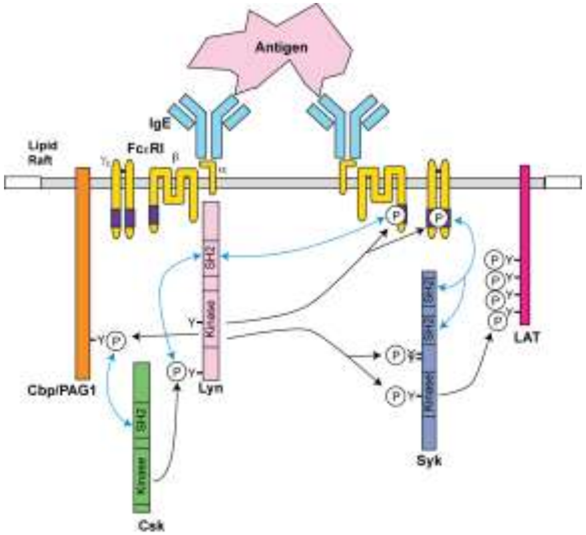
.net BioNetGen Network File

ODE

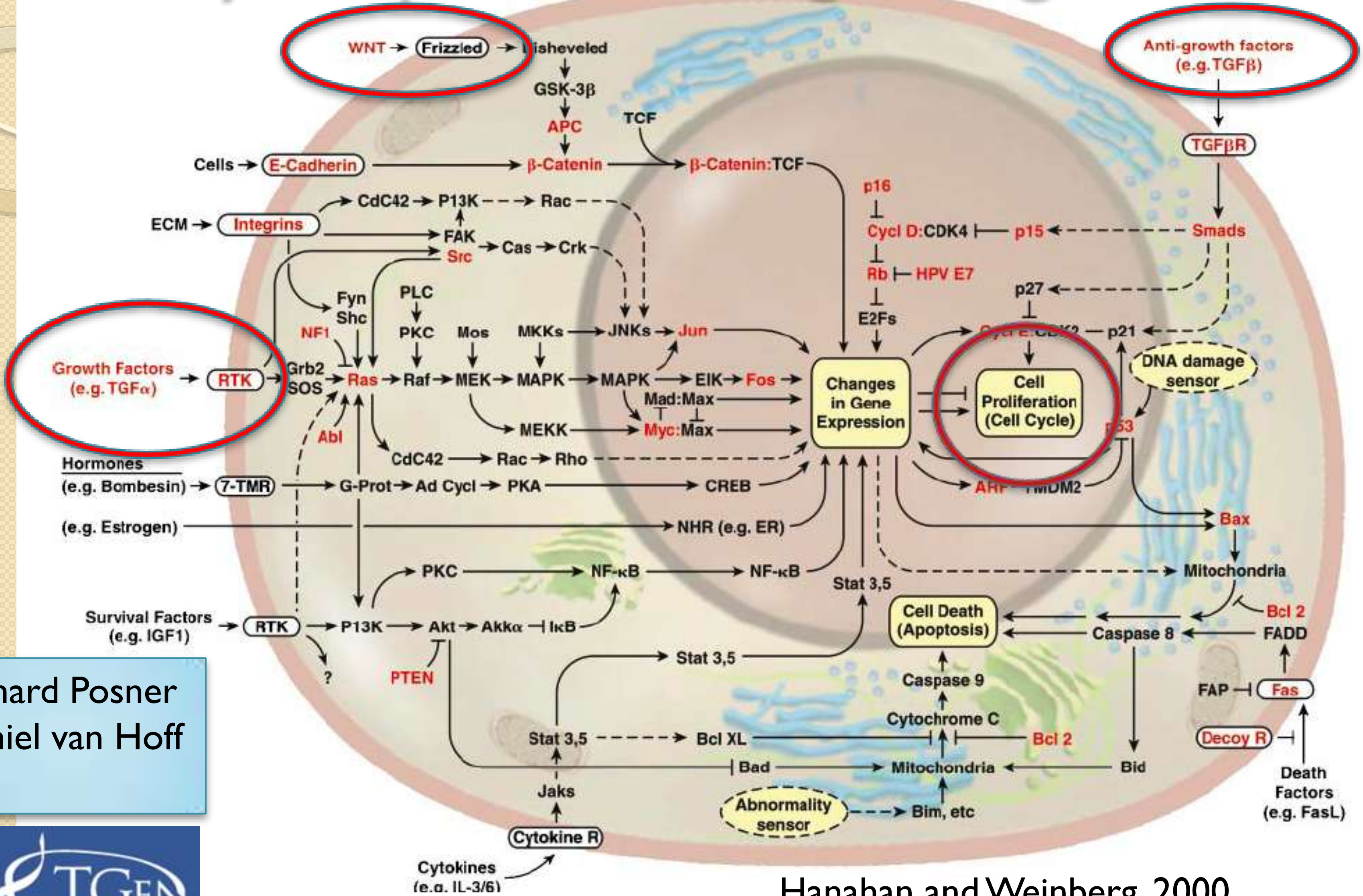
Gillespie

Output

Output



Subway Map of Cell Signaling



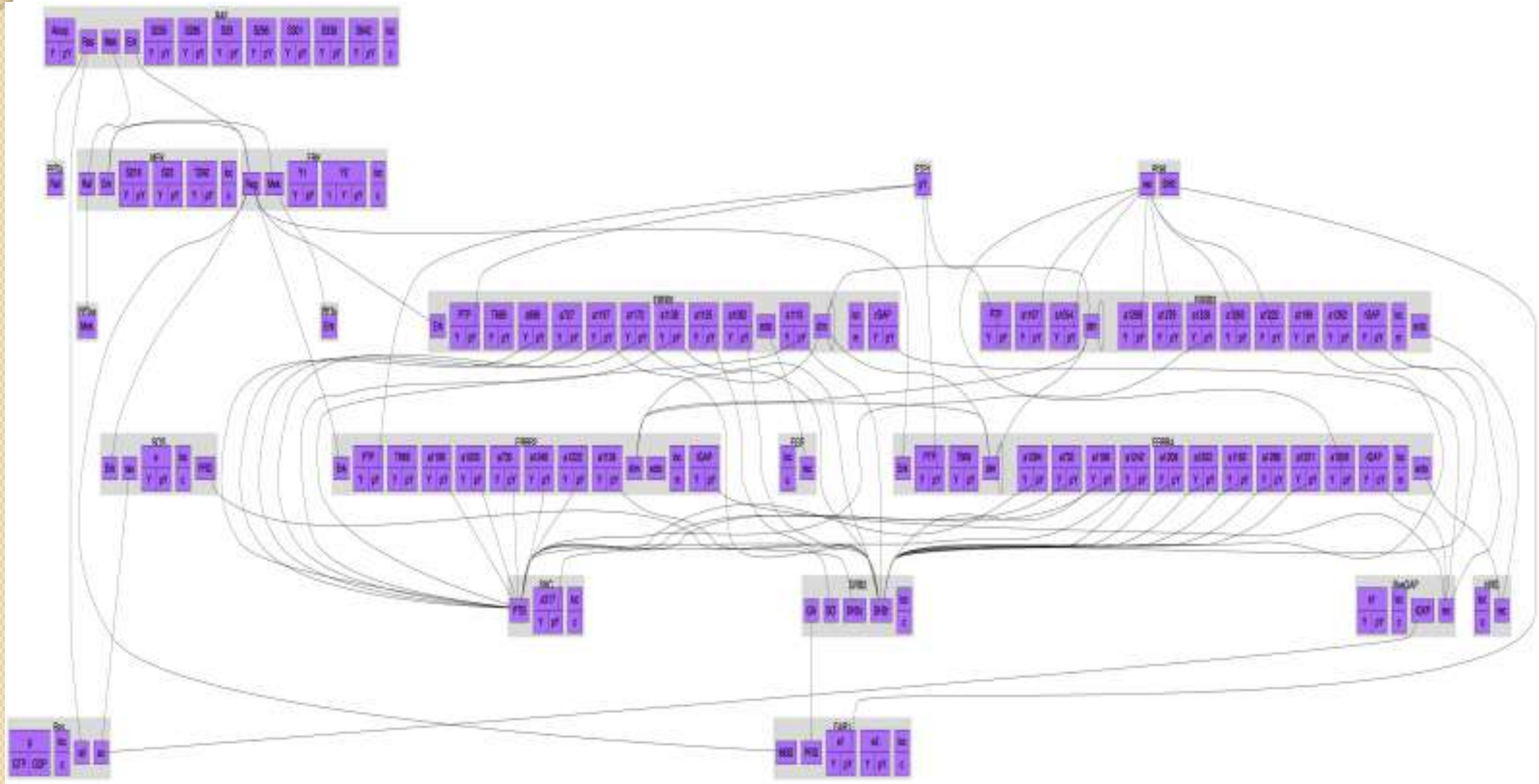
Hanahan and Weinberg, 2000

Richard Posner
Daniel van Hoff
…

TGEN

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

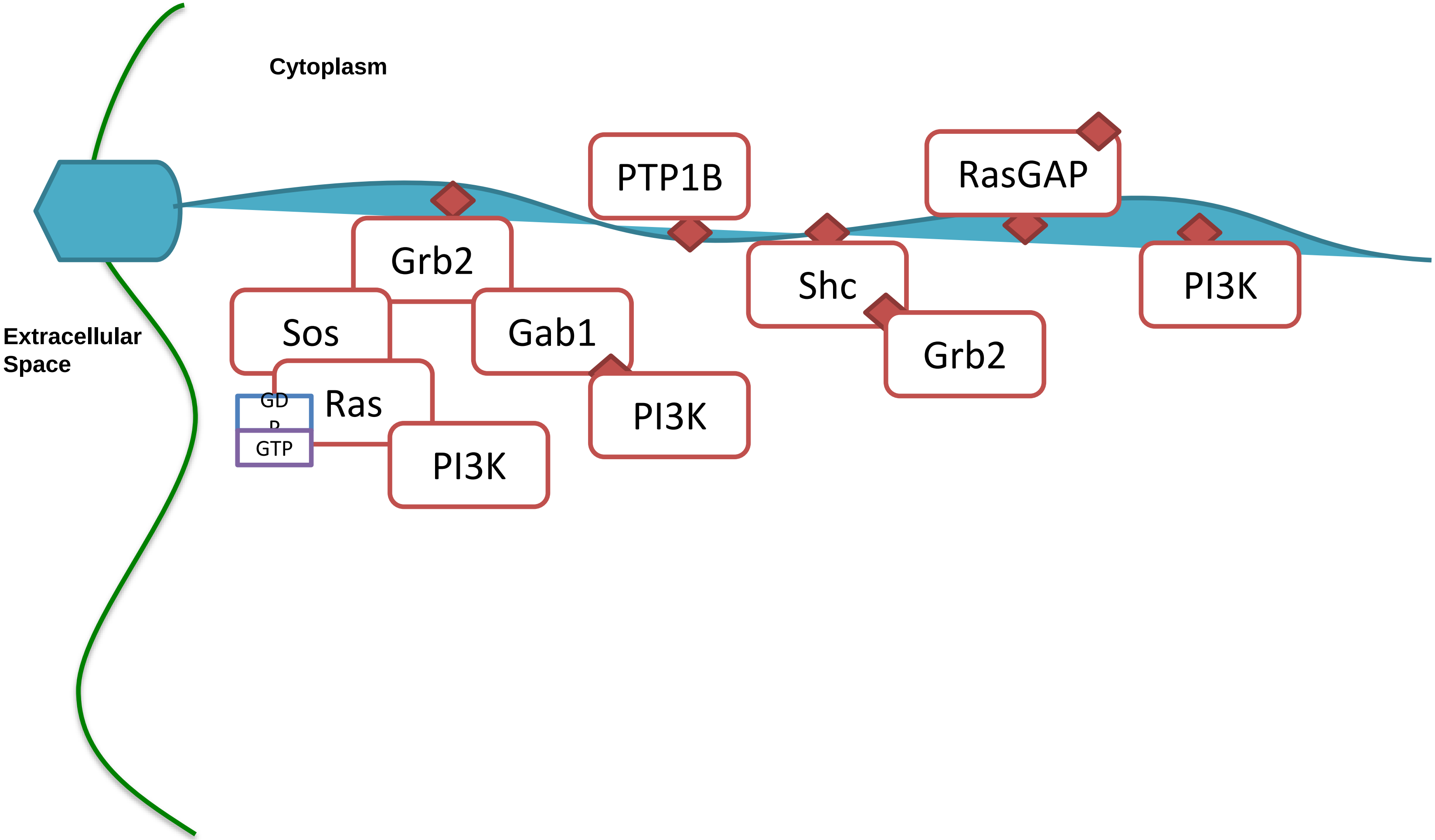
1500 sim sec

- ~10-18 million events
- ~ 1060 real sec

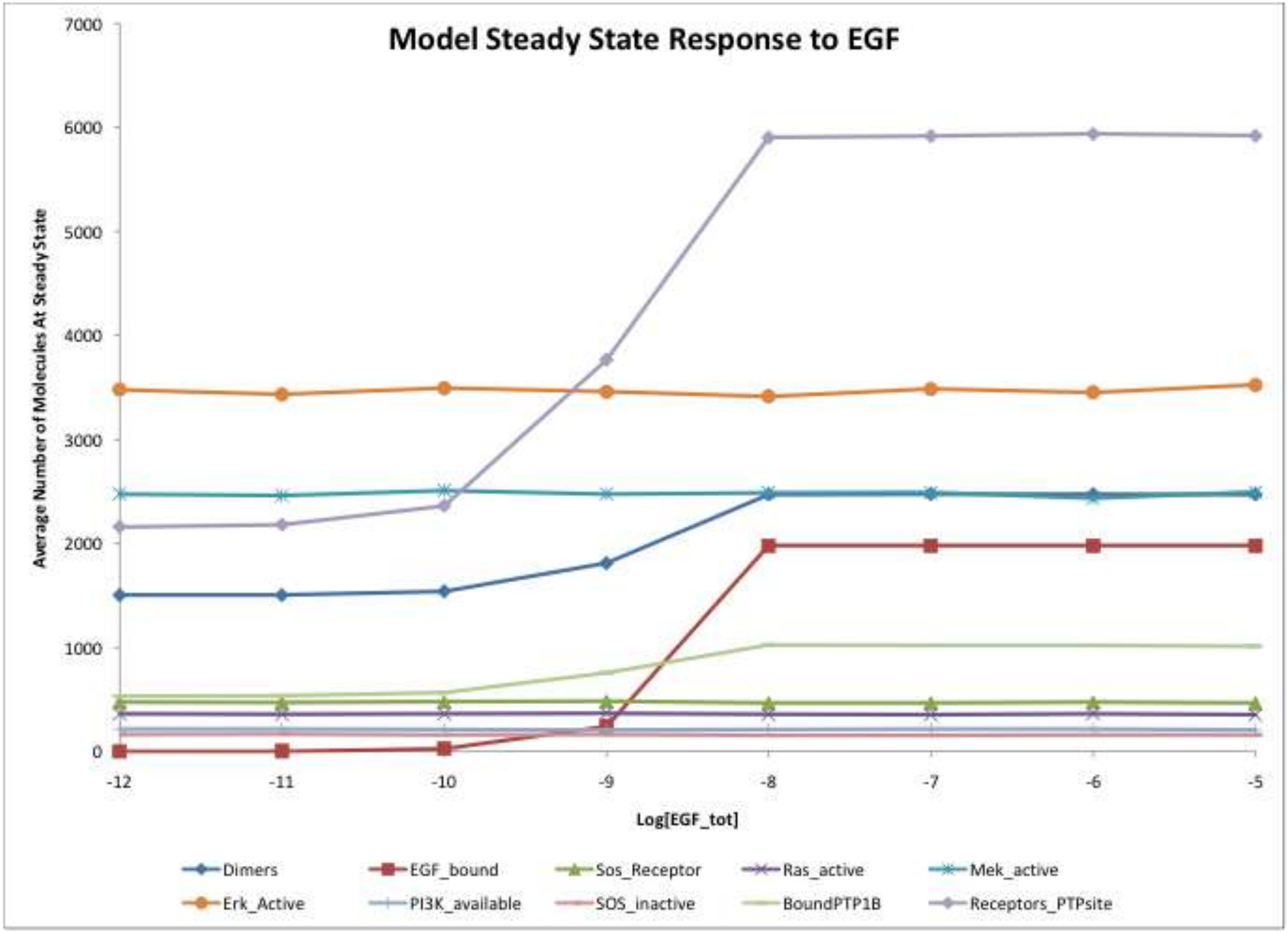
~ 6e-5 CPU seconds/event

(On a 2.4 GHz Intel Core2Duo on iMac with 4 GB RAM)

Visual Annotation of the Model

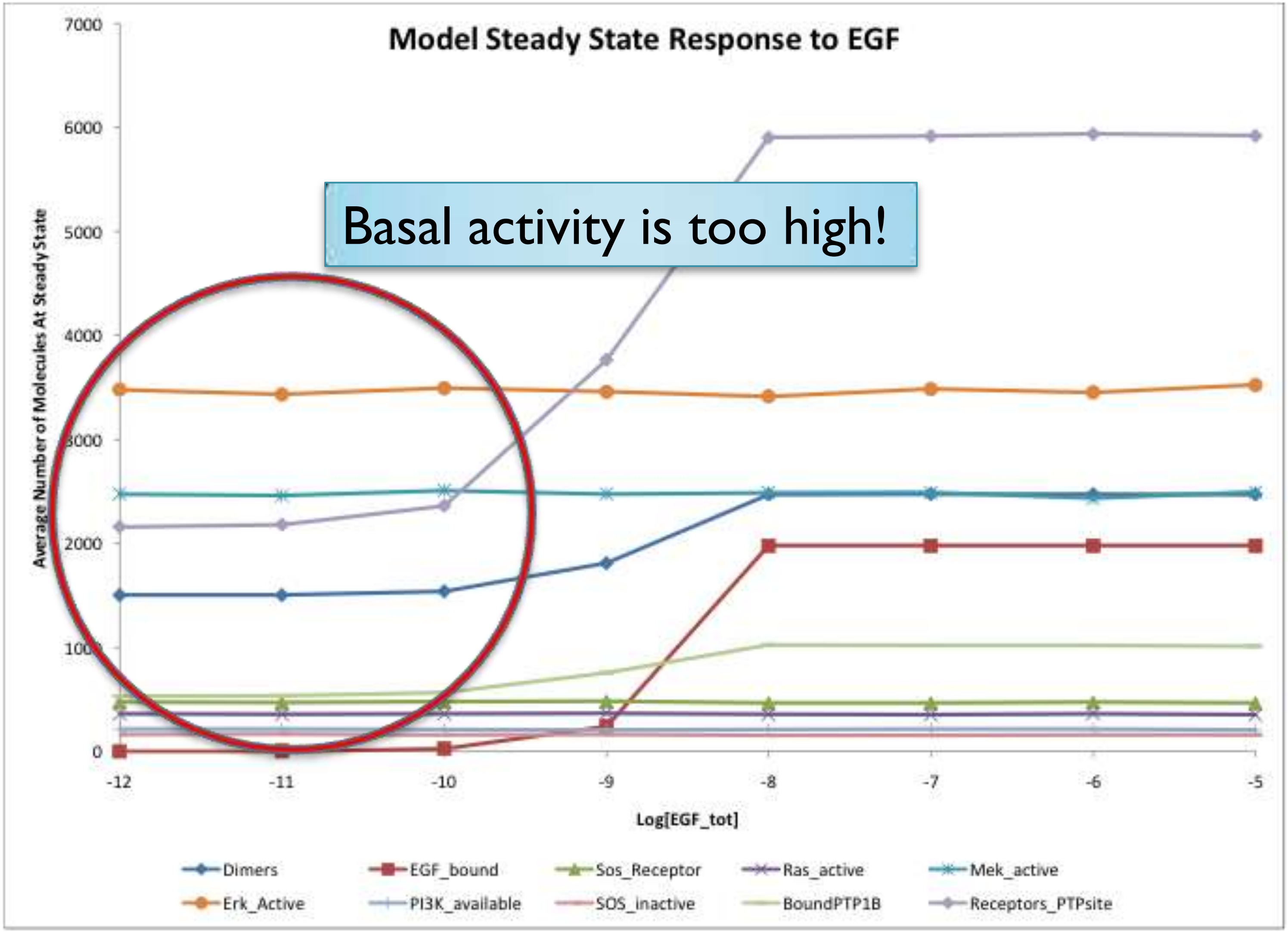


Model Validation



John Sekar

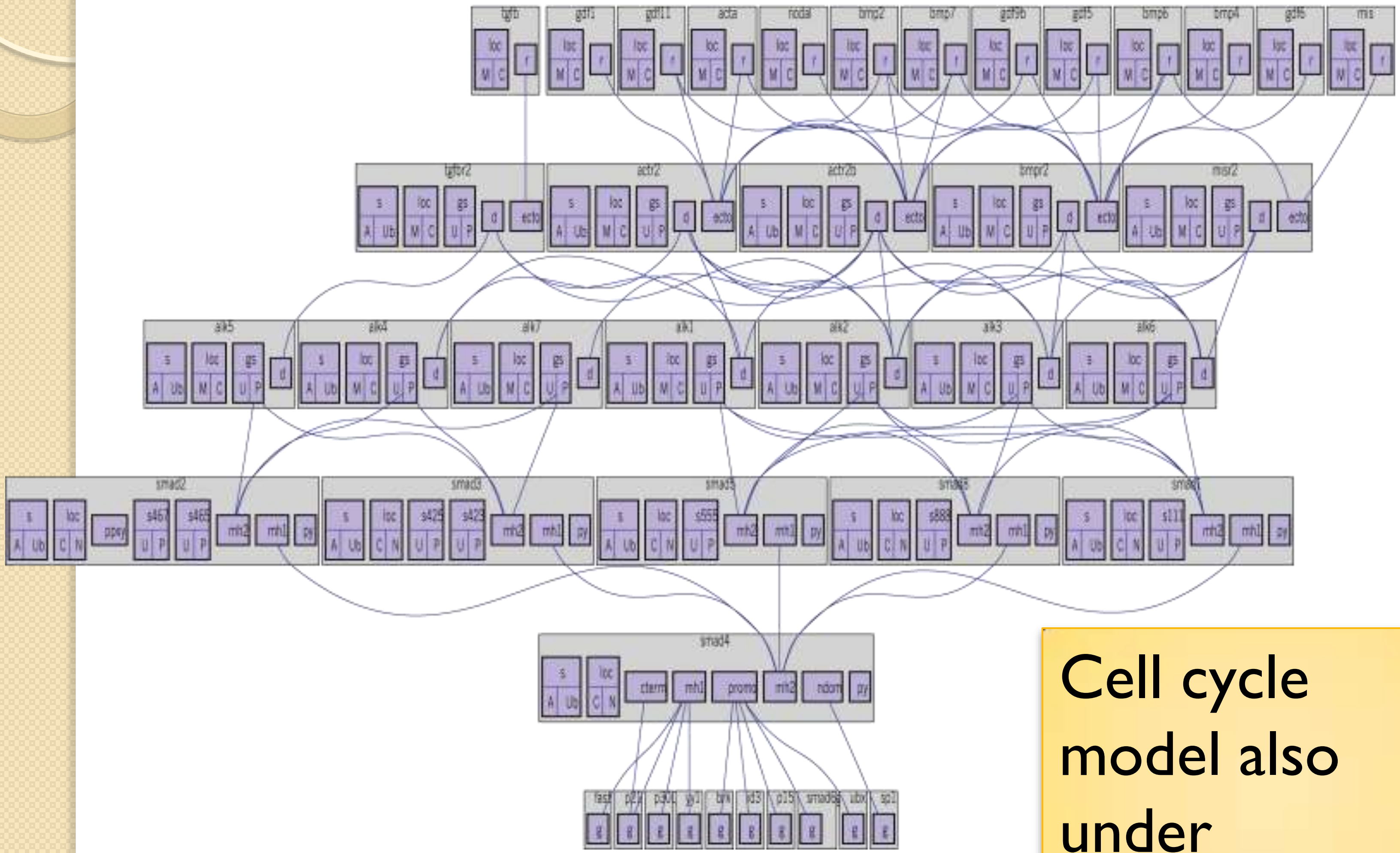
Model Validation



Basal activity is too high!

John Sekar

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

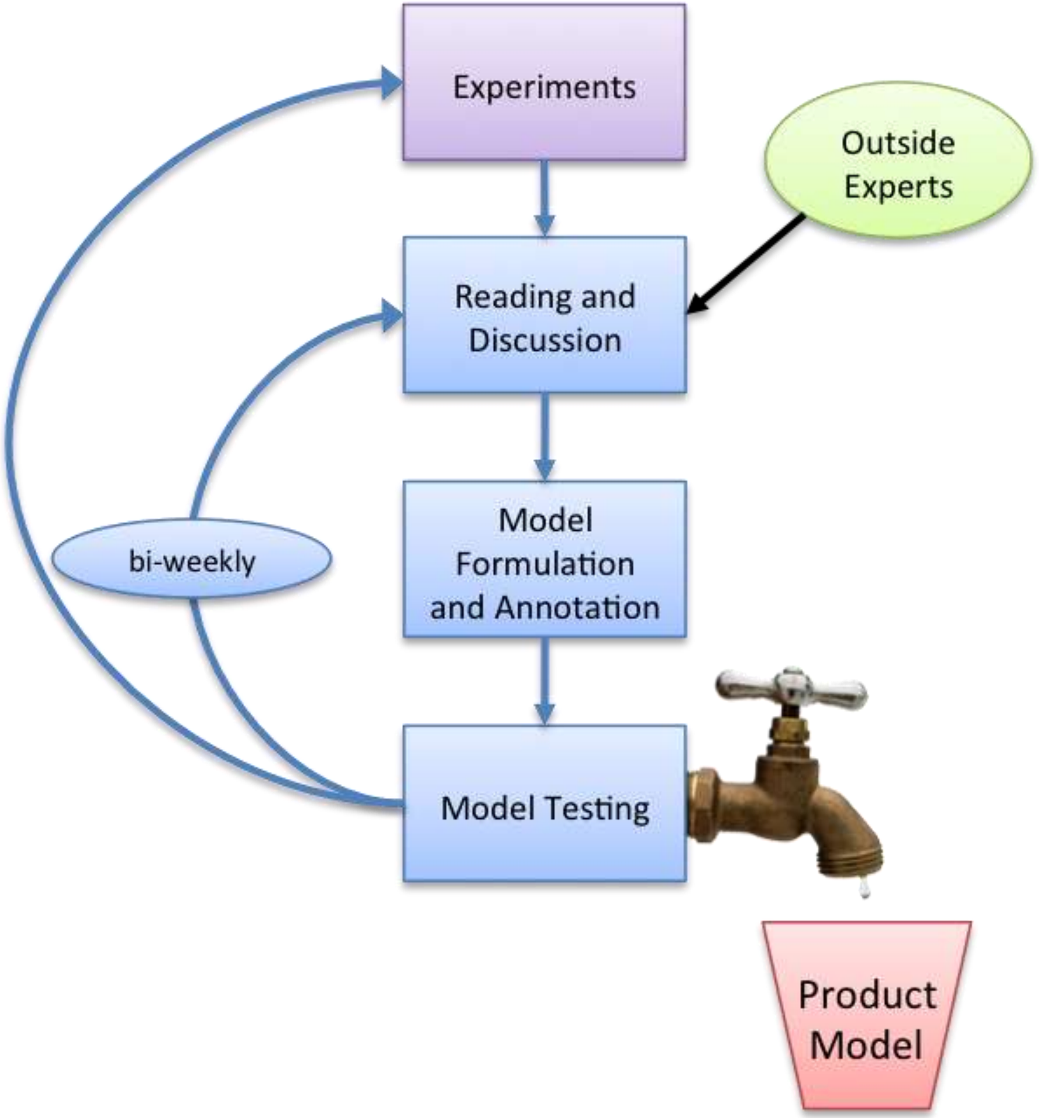
- Continue to build and analyze models of key pathways
- Systematic investigation of models using
  - Statistical and Bayesian Model Checking
  - Global parameter sensitivity analysis
  - Parameter estimation and synthesis
- Integration of pathway models
- Model reduction
  - Coarse-graining of detailed 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

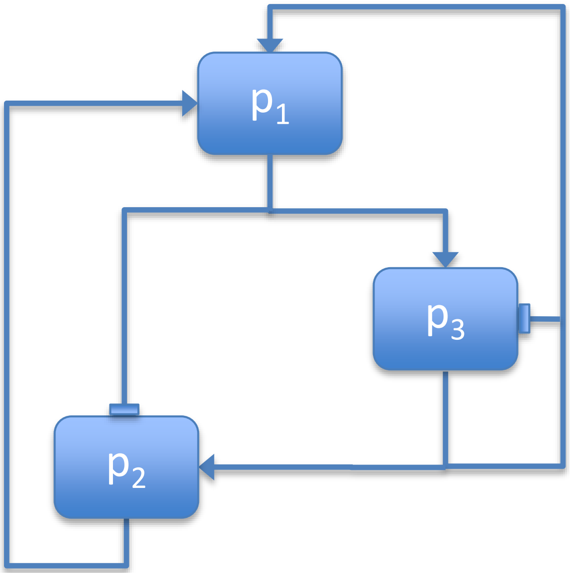
- The **state of an element** in the signaling network can be described by a **Boolean variable**, expressing that it is:
  - Active or present (on or '1')
  - Inactive or absent (off or '0')
- Boolean functions:
  - Represent interactions between elements
  - The state of an element is calculated from states of other elements
- Practical advantages
  - No parameters – facilitates model development
  - Easy to understand – facilitates collaboration

Model development protocol



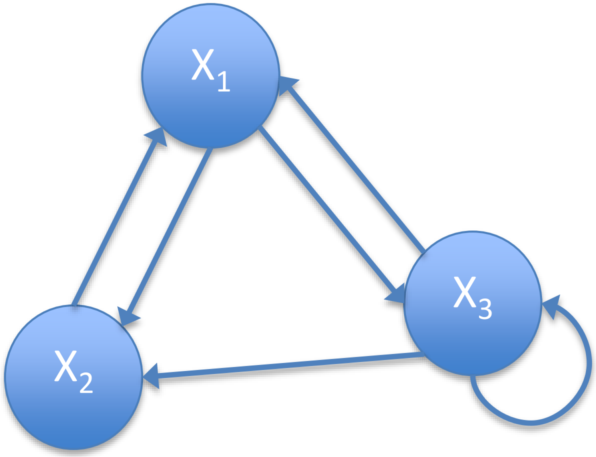
Logical modeling - example

Biological network



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

Boolean network

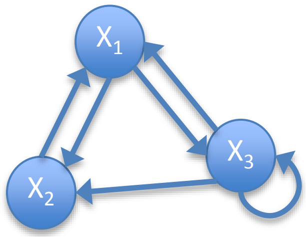


$x_1(t+1)$ = $x_2(t)$ or $x_3(t)$
$x_2(t+1)$ = not $x_1(t)$ and $x_3(t)$
$x_3(t+1)$ = x_1 (t) and not $x_3(t)$

Logical modeling - example

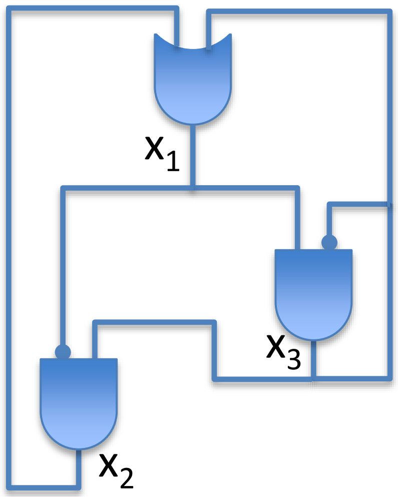
- $x_1x_2x_3$ – state vector

Boolean network



$x_1(t+1)$ = $x_2(t)$ or $x_3(t)$
$x_2(t+1)$ = not $x_1(t)$ and $x_3(t)$
$x_3(t+1)$ = x_1 (t) and not $x_3(t)$

Logic circuit network

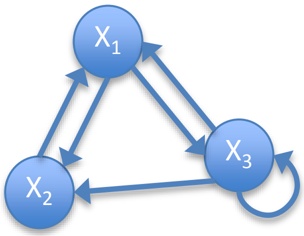


State transition table

state	$x_1(t)x_2(t)x_3(t)$	$x_1(t+1)x_2(t+1)x_3(t+1)$
s_1	000	000
s_2	001	110
s_3	010	100
s_4	011	110
s_5	100	001
s_6	101	100
s_7	110	101
s_8	111	100

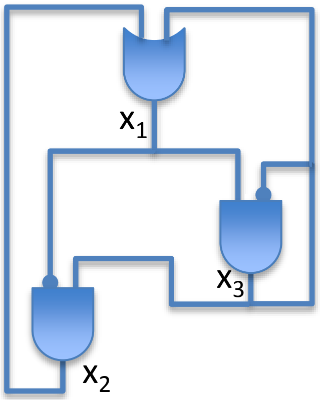
Logical modeling - example

Boolean network

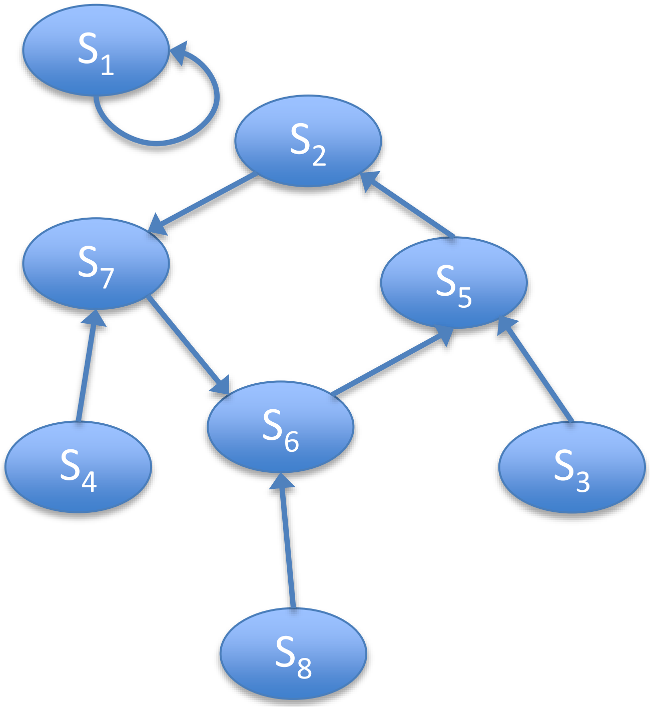


$x_1(t+1) = x_2(t)$ or $x_3(t)$
$x_2(t+1) =$ not $x_1(t)$ and $x_3(t)$
$x_3(t+1) = x_1(t)$ and not $x_3(t)$

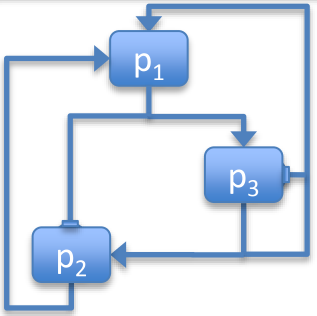
Logic circuit network



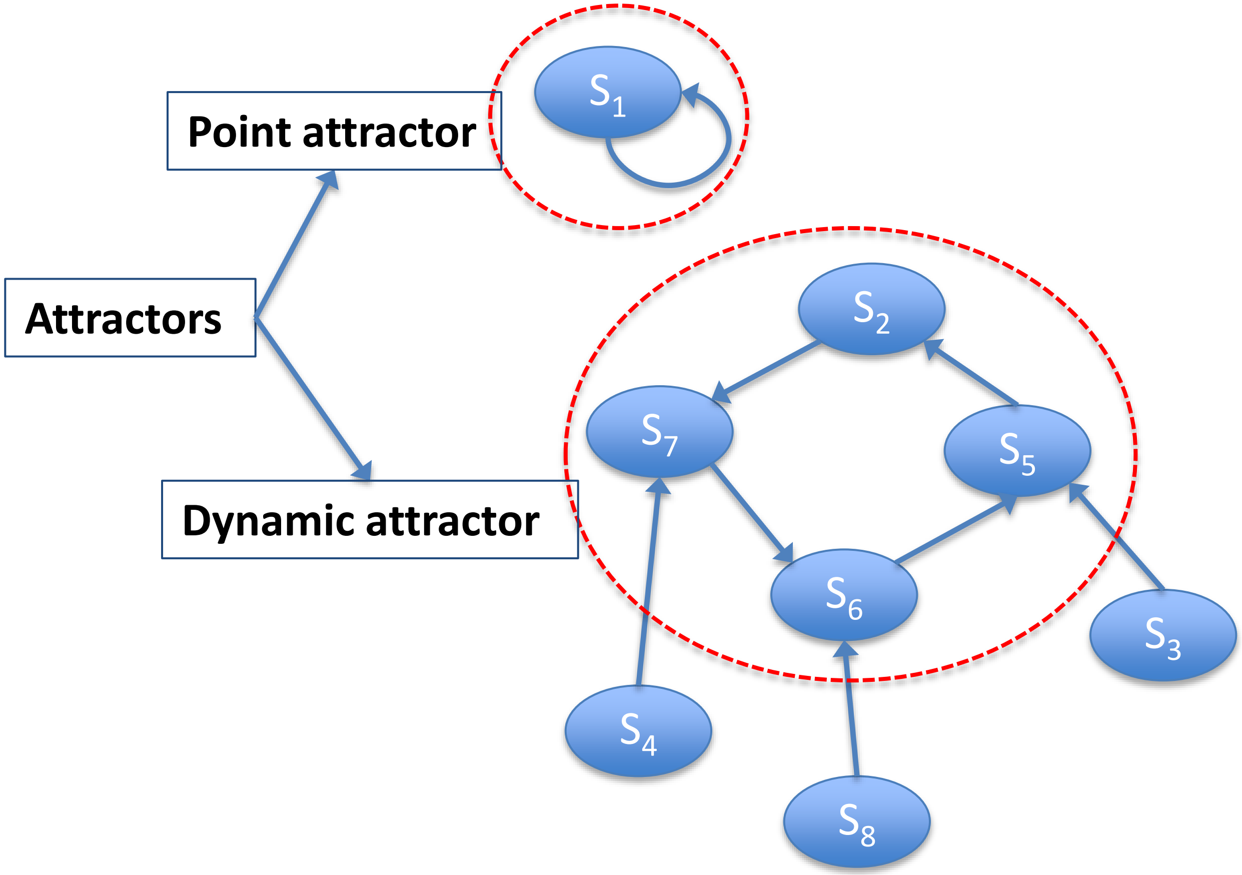
State transition diagram



Logical modeling - example



state	$x_1x_2x_3$
s_1	000
s_2	001
s_3	010
s_4	011
s_5	100
s_6	101
s_7	110
s_8	111


Point attractor

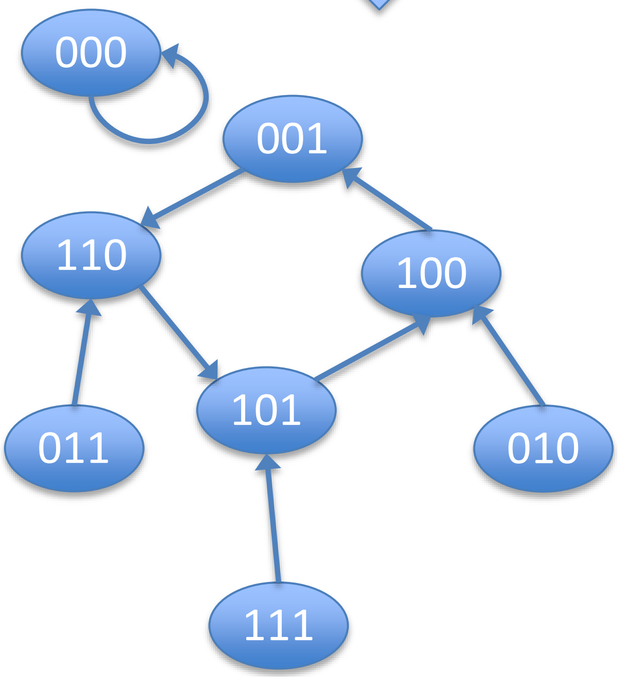
Attractors

Dynamic attractor

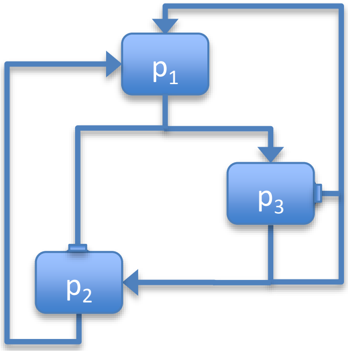
- A sequence of connected states forms a **trajectory** of the system
- The number of states and the number of trajectories in the state space are **finite**
- All initial states of a trajectory will eventually reach a **steady state** or a **state cycle**

Synchronous vs. asynchronous updates

Synchronous updates

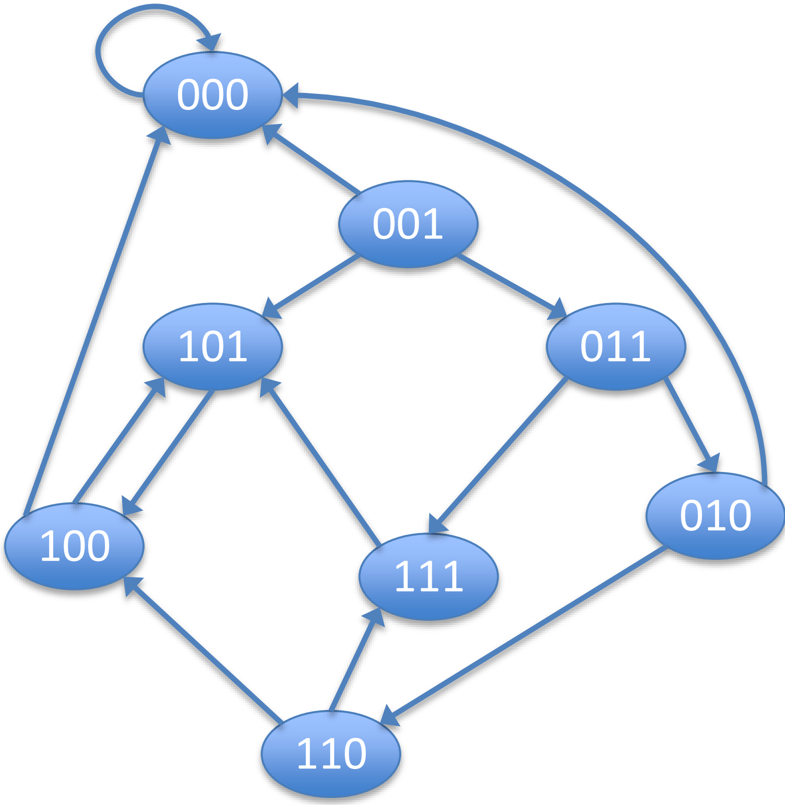


$x_1(t+1) = x_2(t)$ or $x_3(t)$
$x_2(t+1) =$ not $x_1(t)$ and $x_3(t)$
$x_3(t+1) = x_1(t)$ and not $x_3(t)$

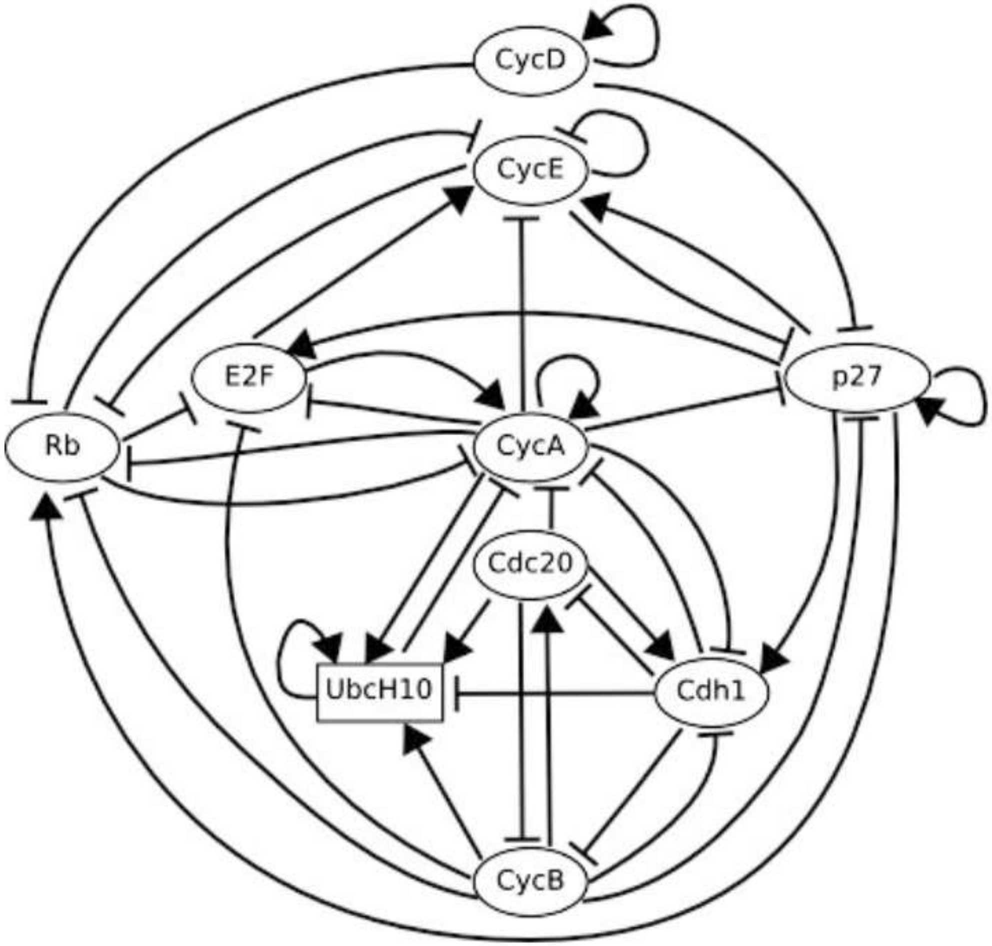


state	$x_1x_2x_3$
s_1	000
s_2	001
s_3	010
s_4	011
s_5	100
s_6	101
s_7	110
s_8	111

Asynchronous updates

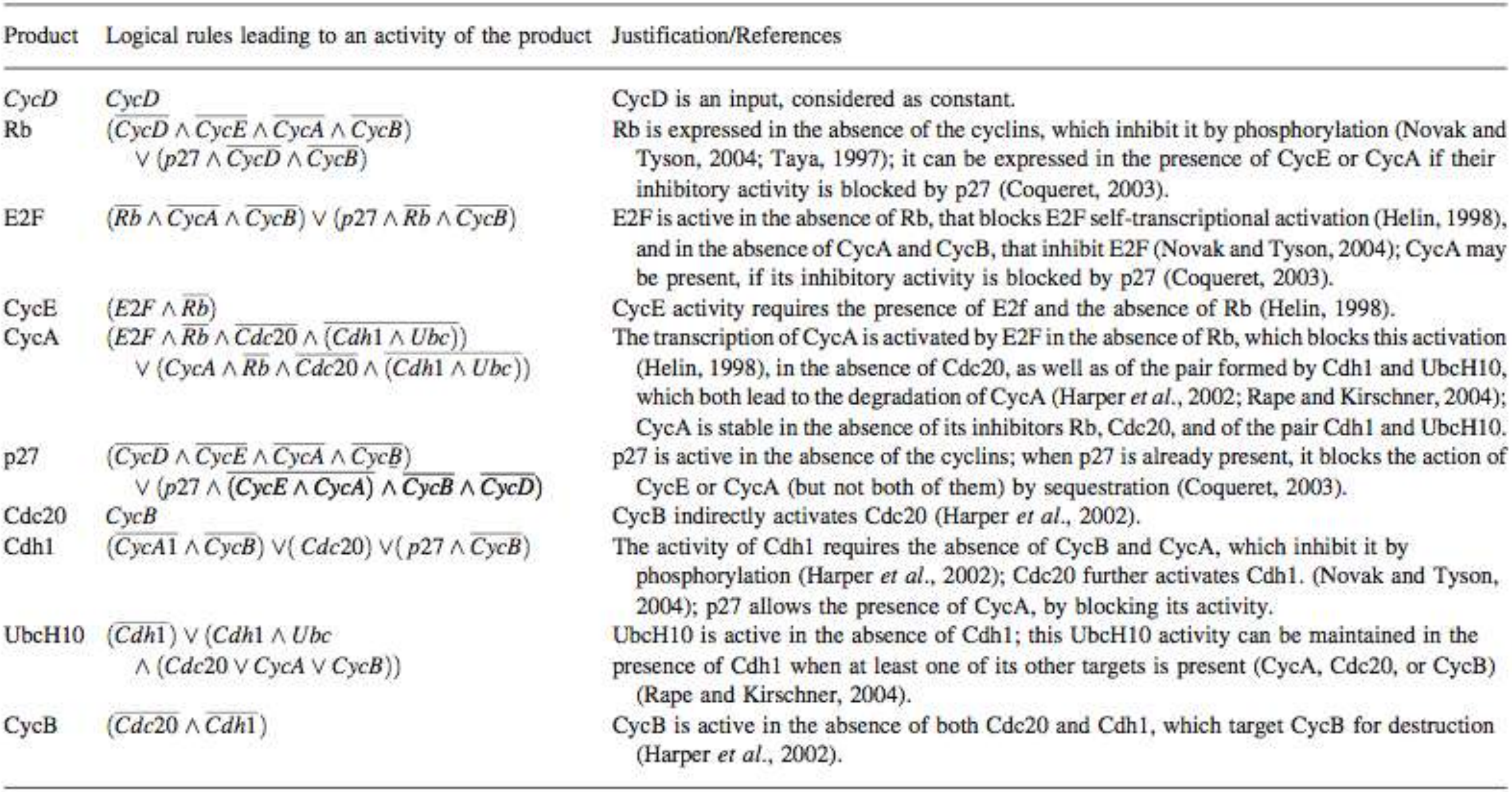


Regulatory graph for mammalian cell cycle network



Source: Faure et al., Bioinformatics, 2006.

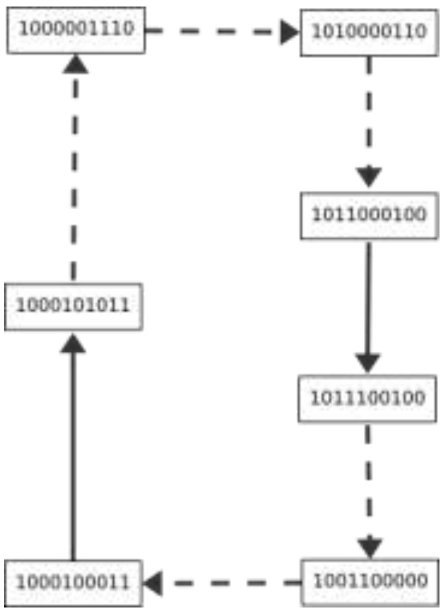
Logical rules



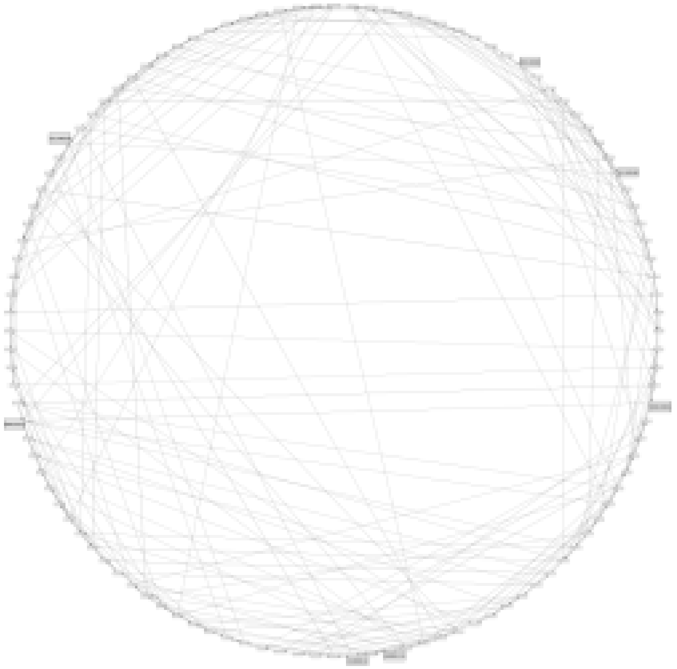
Product	Logical rules leading to an activity of the product	Justification/References
CycD	$CycD$	CycD is an input, considered as constant.
Rb	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB})$ $\vee (p27 \wedge \overline{CycD} \wedge \overline{CycB})$	Rb is expressed 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).
E2F	$(\overline{Rb} \wedge \overline{CycA} \wedge \overline{CycB}) \vee (p27 \wedge \overline{Rb} \wedge \overline{CycB})$	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).
CycE	$(E2F \wedge \overline{Rb})$	CycE activity requires the presence of E2f and the absence of Rb (Helin, 1998).
CycA	$(E2F \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)})$ $\vee (CycA \wedge \overline{Rb} \wedge \overline{Cdc20} \wedge \overline{(Cdh1 \wedge Ubc)})$	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.
p27	$(\overline{CycD} \wedge \overline{CycE} \wedge \overline{CycA} \wedge \overline{CycB})$ $\vee (p27 \wedge \overline{(CycE \wedge CycA)} \wedge \overline{CycB} \wedge \overline{CycD})$	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).
Cdc20	$CycB$	CycB indirectly activates Cdc20 (Harper *et al.*, 2002).
Cdh1	$(\overline{CycA1} \wedge \overline{CycB}) \vee (Cdc20) \vee (p27 \wedge \overline{CycB})$	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.
UbcH10	$(\overline{Cdh1}) \vee (Cdh1 \wedge Ubc$ $\wedge (Cdc20 \vee CycA \vee CycB))$	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).
CycB	$(\overline{Cdc20} \wedge \overline{Cdh1})$	CycB is active in the absence of both Cdc20 and Cdh1, which target CycB for destruction (Harper *et al.*, 2002).

Source: Faure et al., Bioinformatics, 2006.

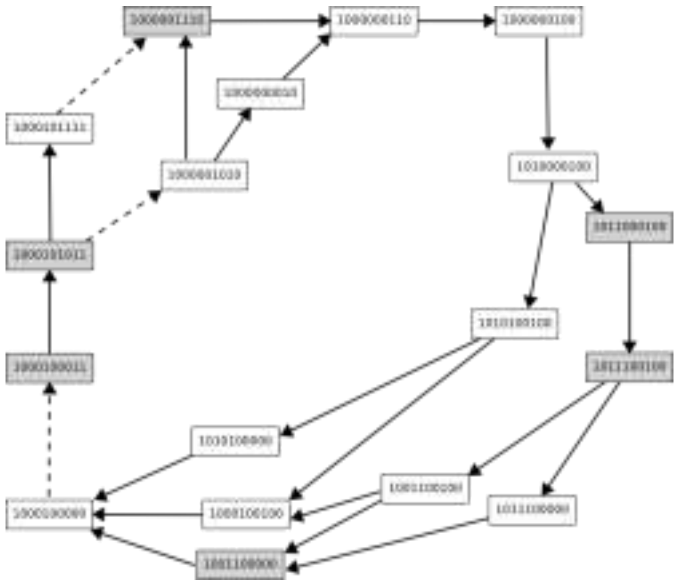
Updating approaches



synchronous



asynchronous



mixed

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

$$ NSF-EMT

Yale
Thierry Emonet
Michael Sneddon

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

$$ NSF-Expeditions in Computing

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

TGen
Rich Posner
Matthew Creamer
Josh Colvin
Daniel Von Hoff

Lehmann
Nancy Griffeths

http://bionetgen.org

Thank You!

Photo by John Sekar