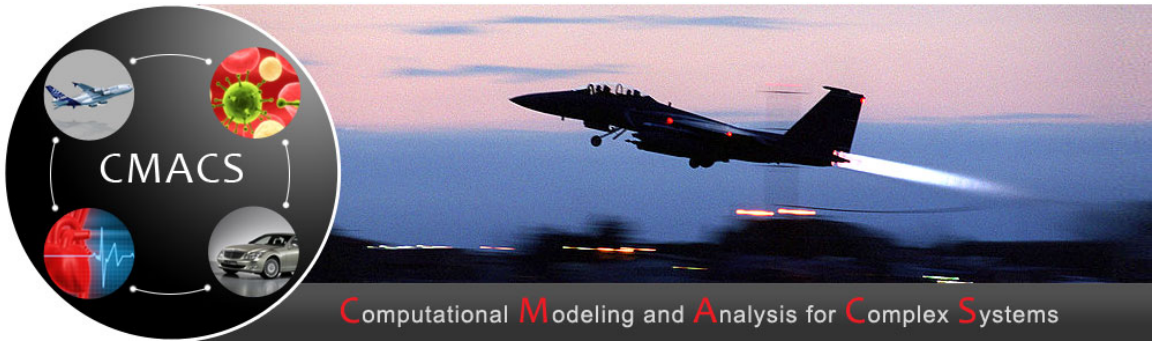


A Brief Introduction to Rule-Based Modeling

Jim Faeder

Department of Computational and Systems Biology

University of Pittsburgh School of Medicine

Workshop on Systems Biology and Formal Methods (SBFM'12)

March 29, 2012

Kimmel Center, NYU

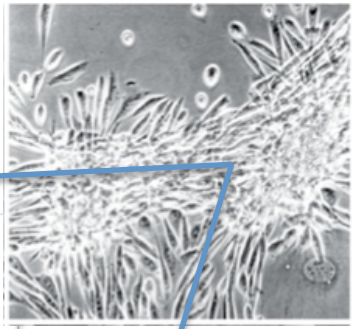
Thanks Patrick and NYU Staff!

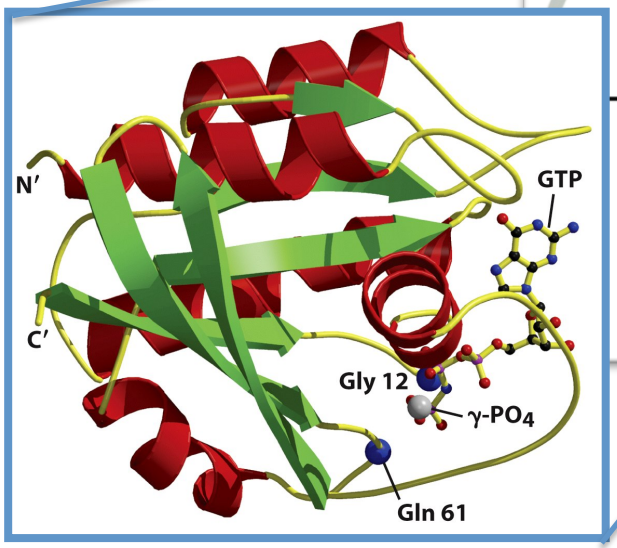
Outline

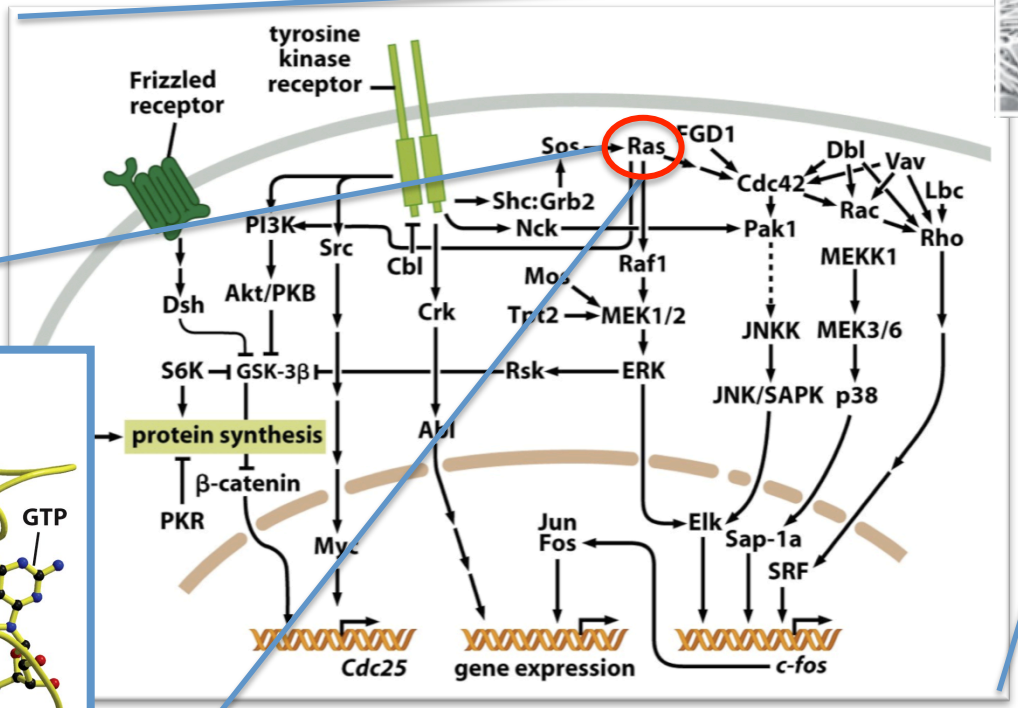
- Motivation
- Basics
- Origins
- Simulation
- Some examples
- Toward comprehensive models
- Community building

The need to model across scales in biology

Transformed







The need for a formal path to reasoning and understanding in biology

Yuri Lazebnik

Can a biologist fix a radio?—Or, what I learned while studying apoptosis

Figure 1. The radio that has been used in this study

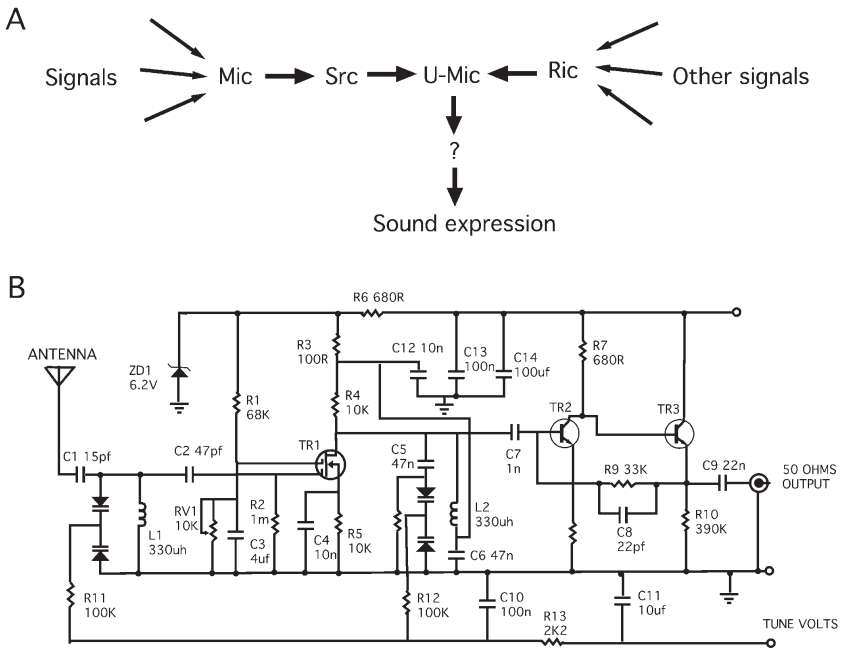
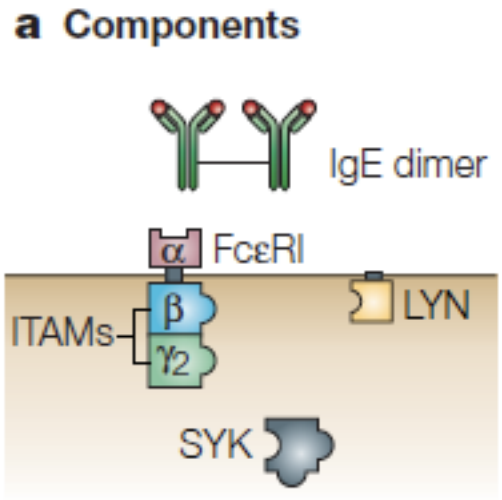


Figure 3. The tools used by biologists and engineers to describe processes of interest

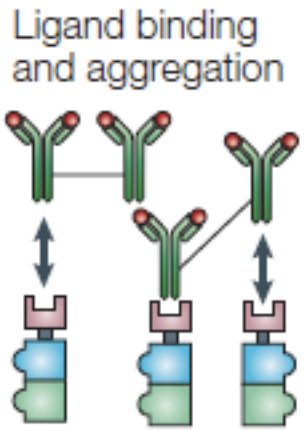
A: The biologist's view of a radio. See Figure 2 and text for description of the indicated components. **B:** The engineer's view of a radio. (Please note that the circuit diagram presented is not that of the radio used in the study. The diagram of the radio was lost, which, in part, explains why the radio remains broken.)

“Object-Oriented” Representation of Signaling Molecules



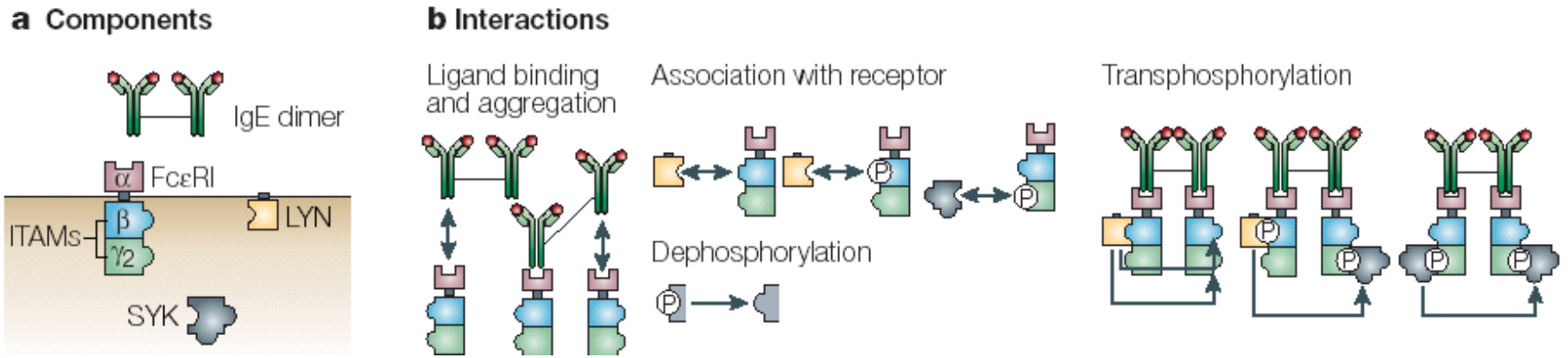
Formal representation

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




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

Composition of a Rule-Based Model



Molecules

```
begin molecules
Lig(l,l)
Lyn(U,SH2)
Syk(tSH2,l~U~P,a~U~P)
Rec(a,b~U~P,g~U~P)
end molecules
```

Reaction Rules

```
begin reaction_rules
# Ligand-receptor binding
1 Rec(a) + Lig(l,l) <-> Rec(a!1).Lig(l!1,l) kp1, km1
 Rec(a) + Lig(l,l) <-> Rec(a!1).Lig(l!1,l) kp1, km1

# Receptor-aggregation
2 Rec(a) + Lig(l,l!1) <-> Rec(a!2).Lig(l!2,l!1) kp2,km2

# Constitutive Lyn-receptor binding
3 Rec(b~Y) + Lyn(U,SH2) <-> Rec(b~Y!1).Lyn(U!1,SH2) kpL, kmL
…
```

Some Origins

- Bray and co-workers
- Finney and co-workers
- Goldstein and co-workers
- Meier-Schellersheim and Mack
- Regev and Shapiro
- Danos and co-workers

Some Origins

- Bray and co-workers (Lay, Morton-Firth, Le Novere, Shimizu, …)

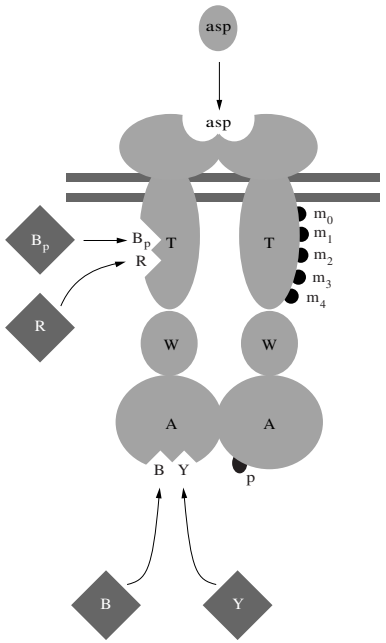
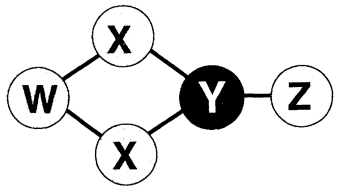


FIG. B2. Illustration of binding sites and methylation states for the chemotaxis receptor complex, which contains two molecules each of Tar, CheW and CheA. Tar dimers can be in one of five methylation states (representing the number of methyl groups), and bind aspartate, phosphorylated CheB and CheR; CheA dimers can be phosphorylated, and bind CheB and CheY; the complex is either in an inactive or active conformation (not shown). This diagram is not meant to represent the actual positions of binding sites on the complex.

A



0	W + X = WX
1	X + Y = XY
2	Y + Z = YZ
3	W + XY = WXY
4	X + WX = WXX
5	X + XY = XXY
6	X + YZ = XYZ
7	Y + WX = WXY
8	Z + XY = XYZ
9	WX + XY = WXXY
10	WX + YZ = WXYZ
11	W + XXY = WXXY
12	W + XYZ = WXYZ
13	X + WXY = WXXY
14	X + XYZ = XXYZ
15	X + WXYZ = WXXYZ
16	Y + WXX = WXXY
17	Z + WXY = WXYZ
18	Z + XXY = XXYZ
19	Z + WXXY = WXXYZ
20	WX + XYZ = WXXYZ
21	YZ + WXX = WXXYZ
22	W + XXYZ = WXXYZ

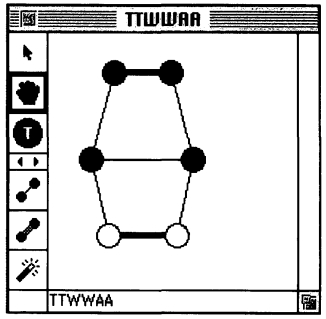


Fig. 2. Window displaying the structure of the Tar complex. The oligomer was assembled by clicking and dragging symbols representing the three protein species from the toolbar, and then linking them together by means of the two bonding tools, as described in the text.

OLIGO (1997)
StochSim (1997)

Some Origins

Andrew Finney, later joined by Le Novere and Shimizu (2001-2). Proposal for multistate extension to SBML, which also later included complexation

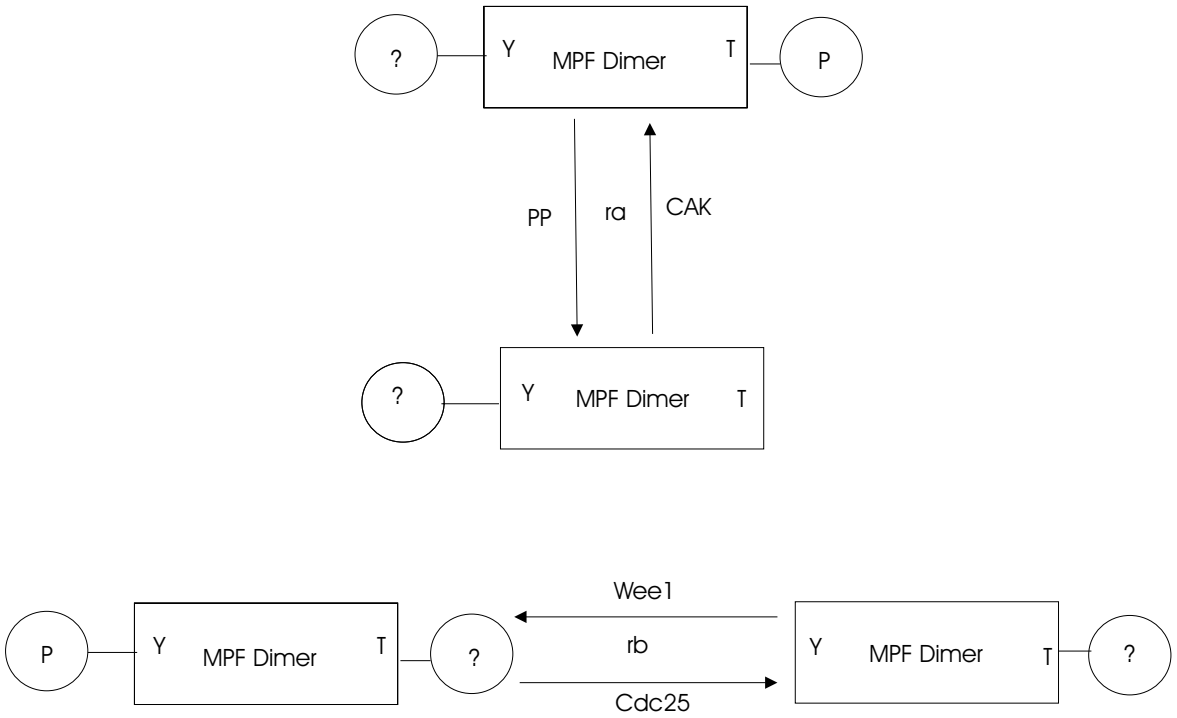
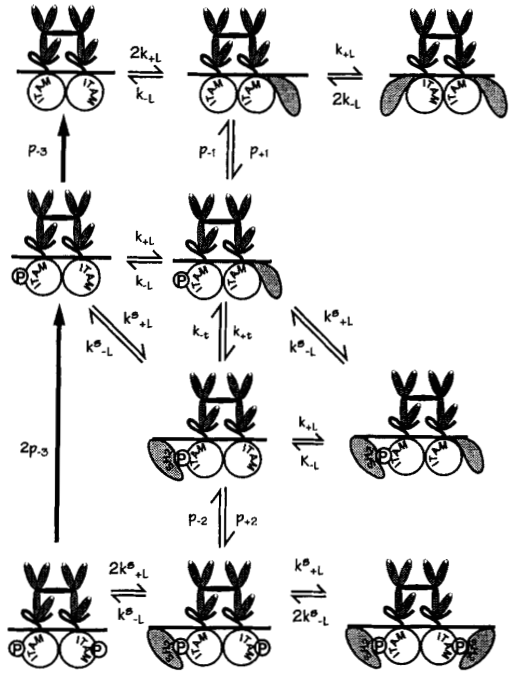


Figure 2: *A formulation of the Tyson model subset abstracting the dimer phosphorylation reactions*

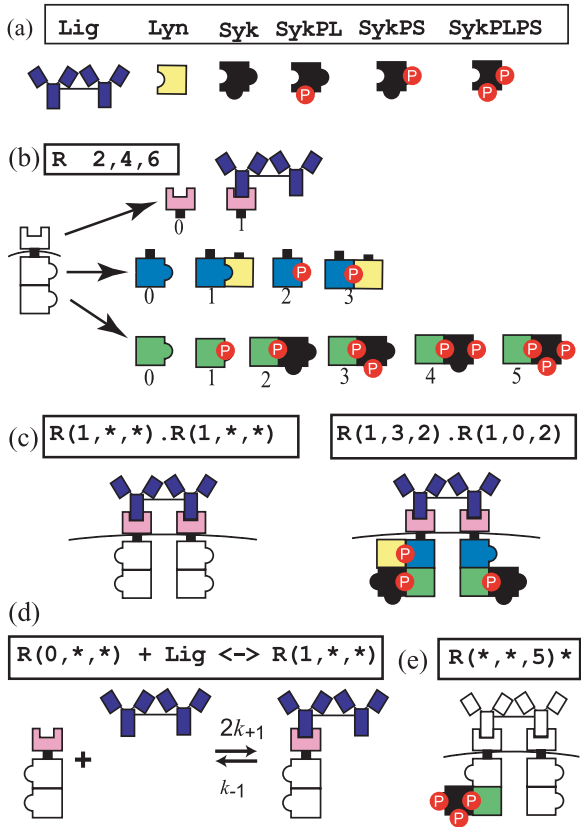
Some Origins

- Goldstein and co-workers (Blinov, Faeder, Hlavacek, Metzger, Redondo, Wofsy, …)



IGURE 3. Reactions of aggregated receptors in model 2. Shown for ggregates of two receptors are the 10 possible configurations that can ccur in model 2 and the reactions that lead to them.

Wofsy et al. (1997)



BioNetGen (2004)

Some Origins

- Meier-Schellersheim and Mack

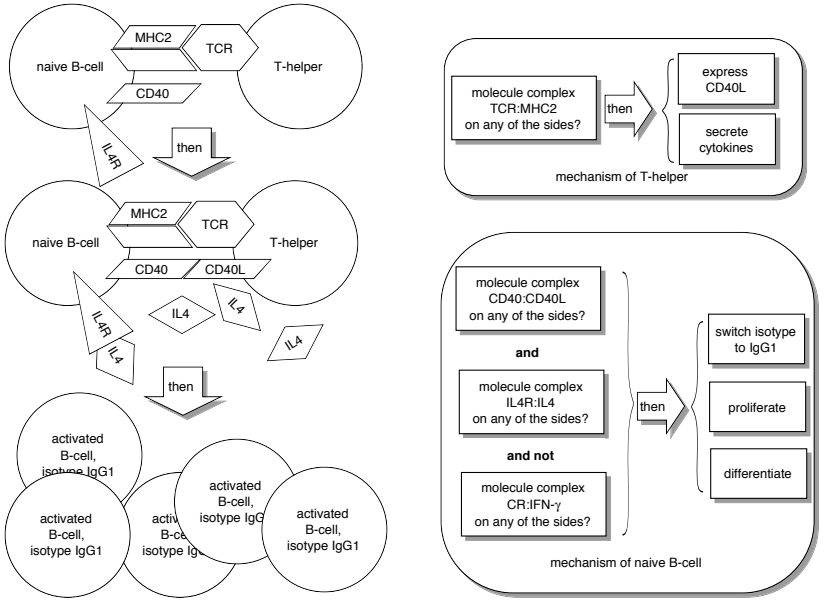
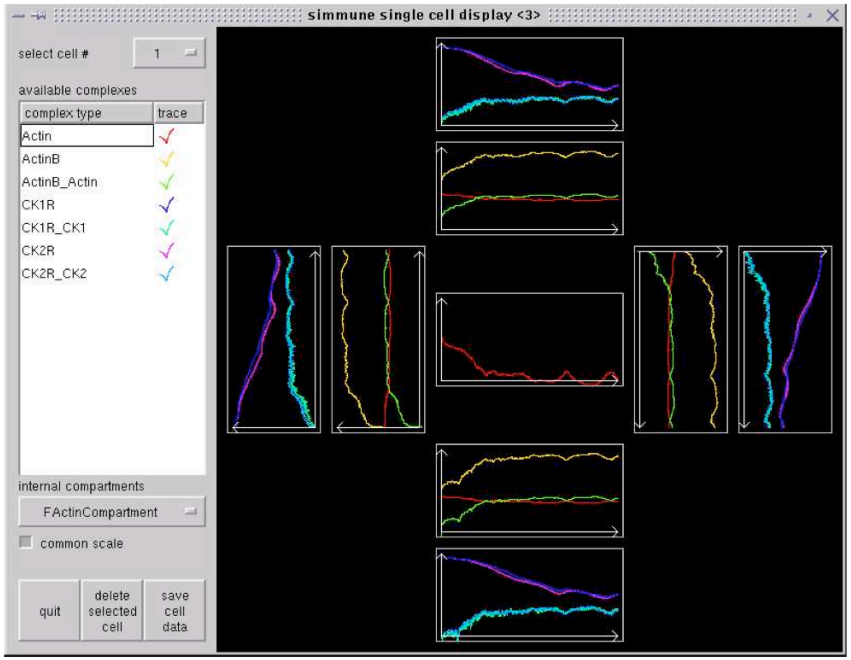


Figure 6.2: Mechanisms of B-cell activation by T-helper cells



Simmune (2001)

Some Origins

- Regev and Shapiro (2001)

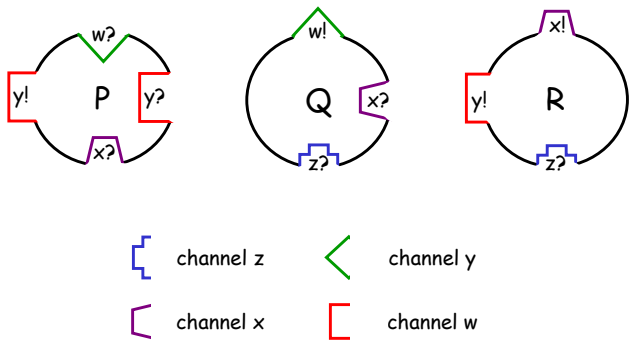


Fig. 2. π-calculus processes and channels: An intuitive view. Three processes, P, Q, R (ovals) with four communication channels (complementary shapes of protrusions and depressions).

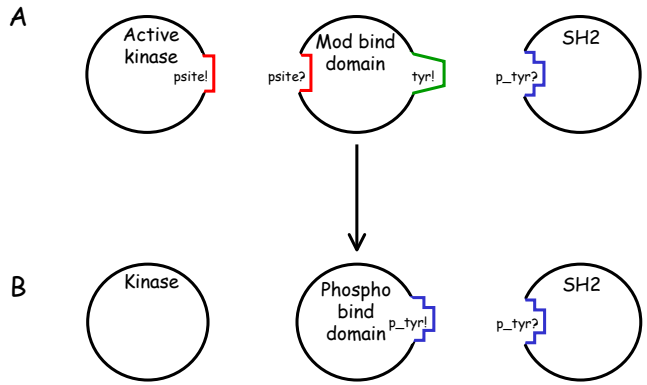
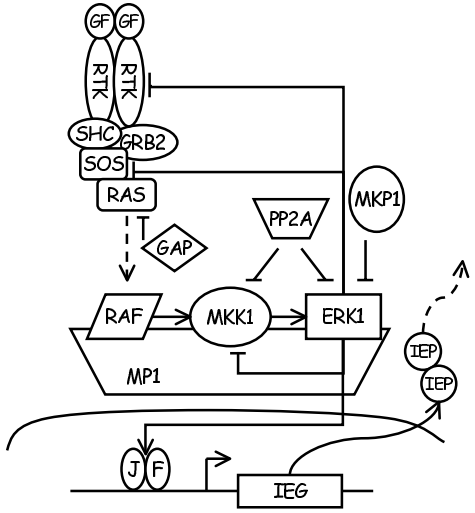


Fig. 5. Interaction and modification as communication and state change.



BioSPI

Some Origins

- Danos (Krivine, Feret, Fontana, Harmer, …)

V. Danos, C. Laneve / Theoretical Computer Science 325 (2004) 69–110

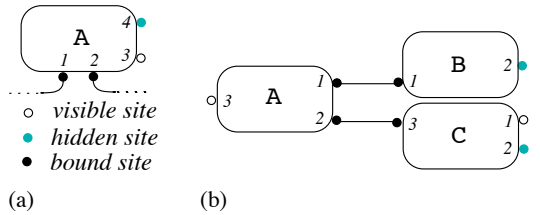


Fig. 1. A protein and a complex.

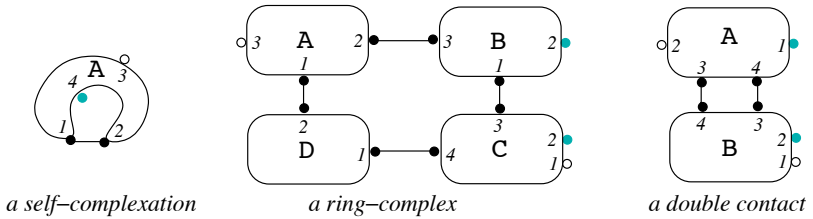


Fig. 2. Complexes.

Kappa Calculus (Danos and Laneve, 2003)

What are the formal expressive capabilities of biological systems?

"This seems a useful preliminary step towards computer-aided exploration and engineering of such systems, though for the moment *it is still unclear what kind of biologically relevant questions one would be prompted to ask if such tools were available*."

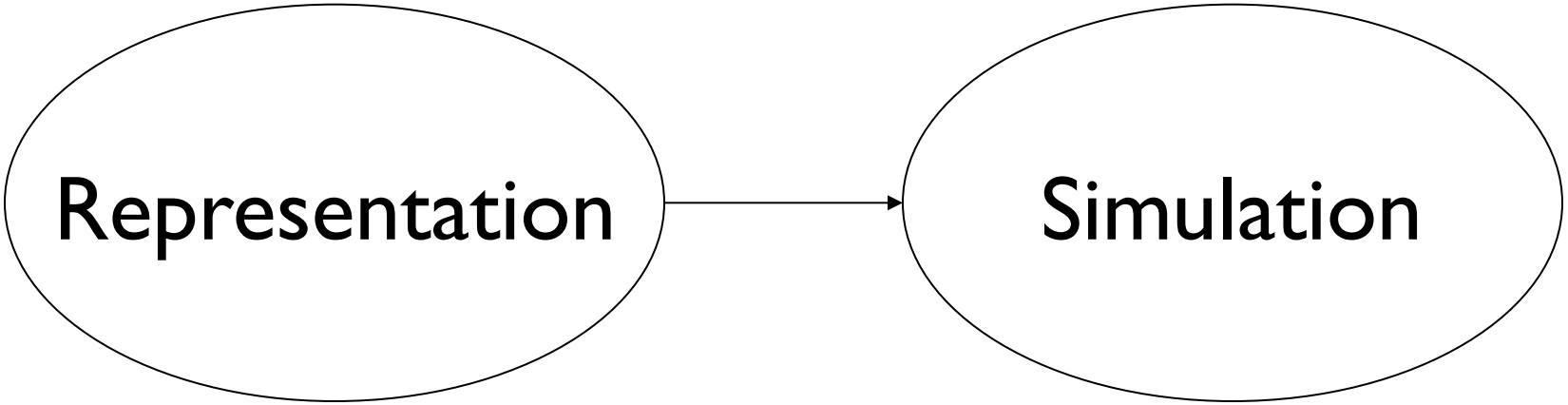
-Chiaverini and Danos (2003)

Some Rule-Based Modeling Tools

1. StochSim (1997)
2. BioSPI (2001)
3. Simmune (2001)
4. BioNetGen (2004)
5. Moleculizer (2005)
6. BioNetGen2 (2006)
7. Kappa Factory (2007)
8. ALC (2008)
9. BlenX (2008)
10. Virtual Cell / BioNetGen (2008)
11. Dynstoc (2009)
12. little b (2009)
13. SSC (2009)
14. GetBonNie (2009)
15. BioΨ (2010)
16. Meredys (2010)
17. RuleMonkey (2010)
18. Smoldyn 2.1 / libMoleculizer (2010)
19. SRsim (2010)
20. ANC (2010)
21. NFsim (2011)
22. RuleBender (2011)
23. ML-Rules (2011)
24. PySB (2011)
25. …

Modeling cell signaling

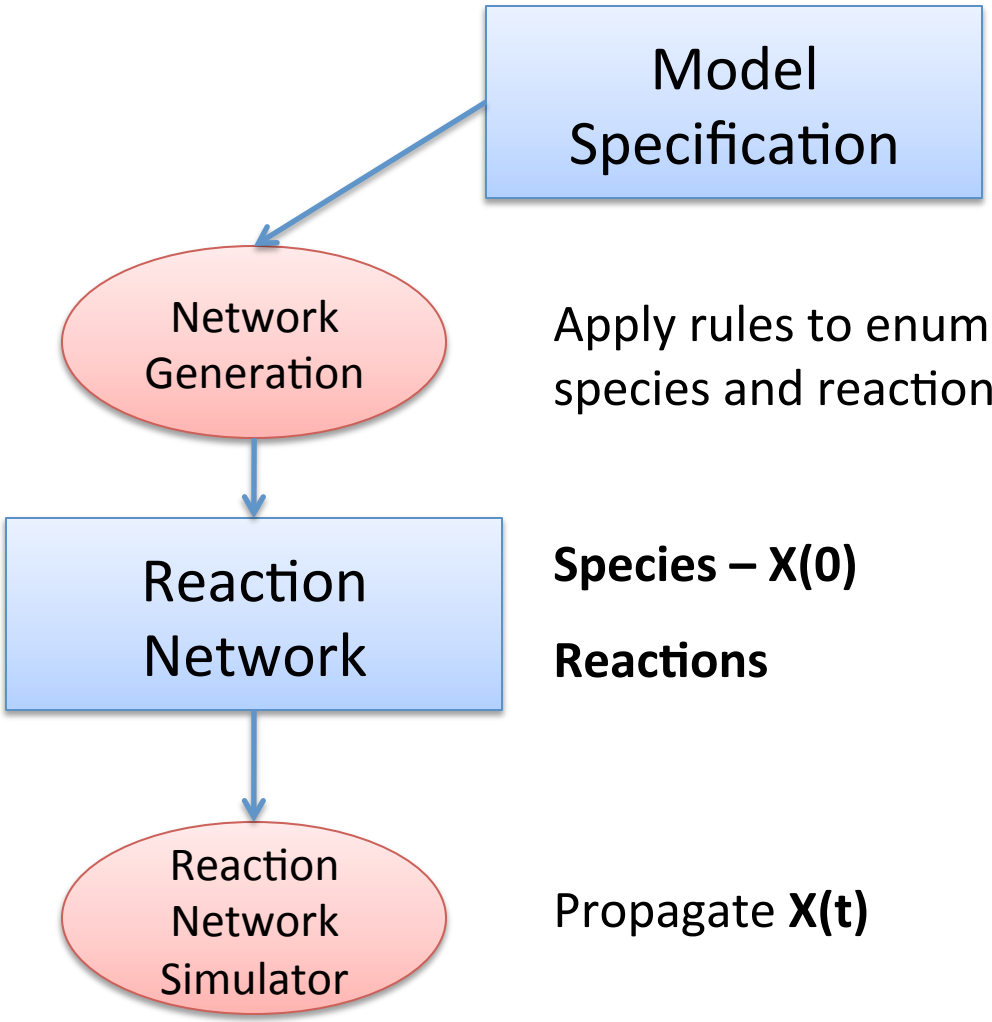
AIM: Model the biochemical machinery by which cells process information (and respond to it).



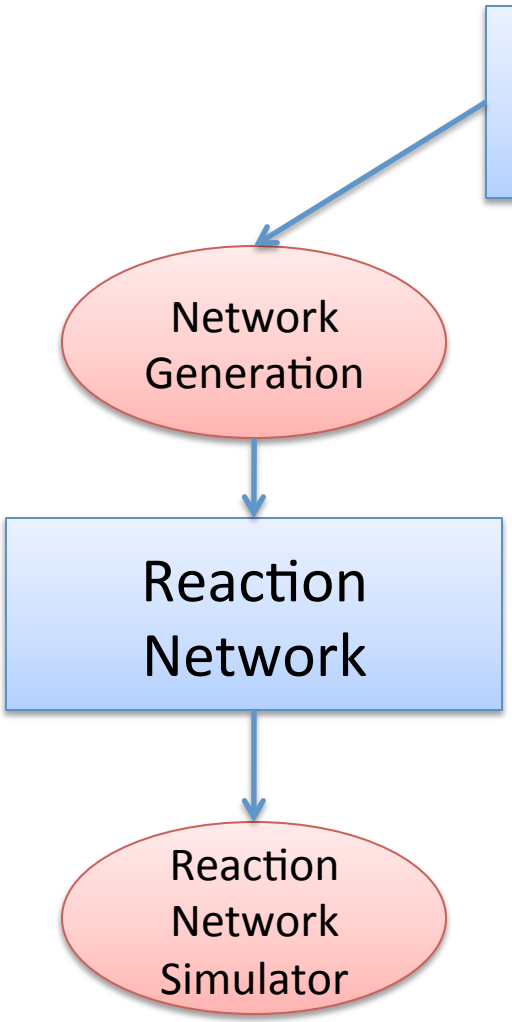
BIONETGEN Language
Kappa
…

ODE, PDE
Stochastic Simulation Algorithm
Kinetic Monte Carlo
Brownian dynamics

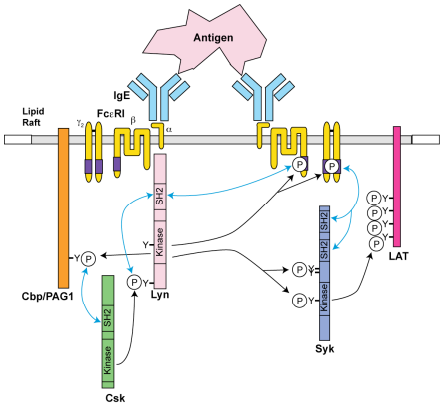
Network Generation Approach to Simulation



Network Generation Approach to Simulation

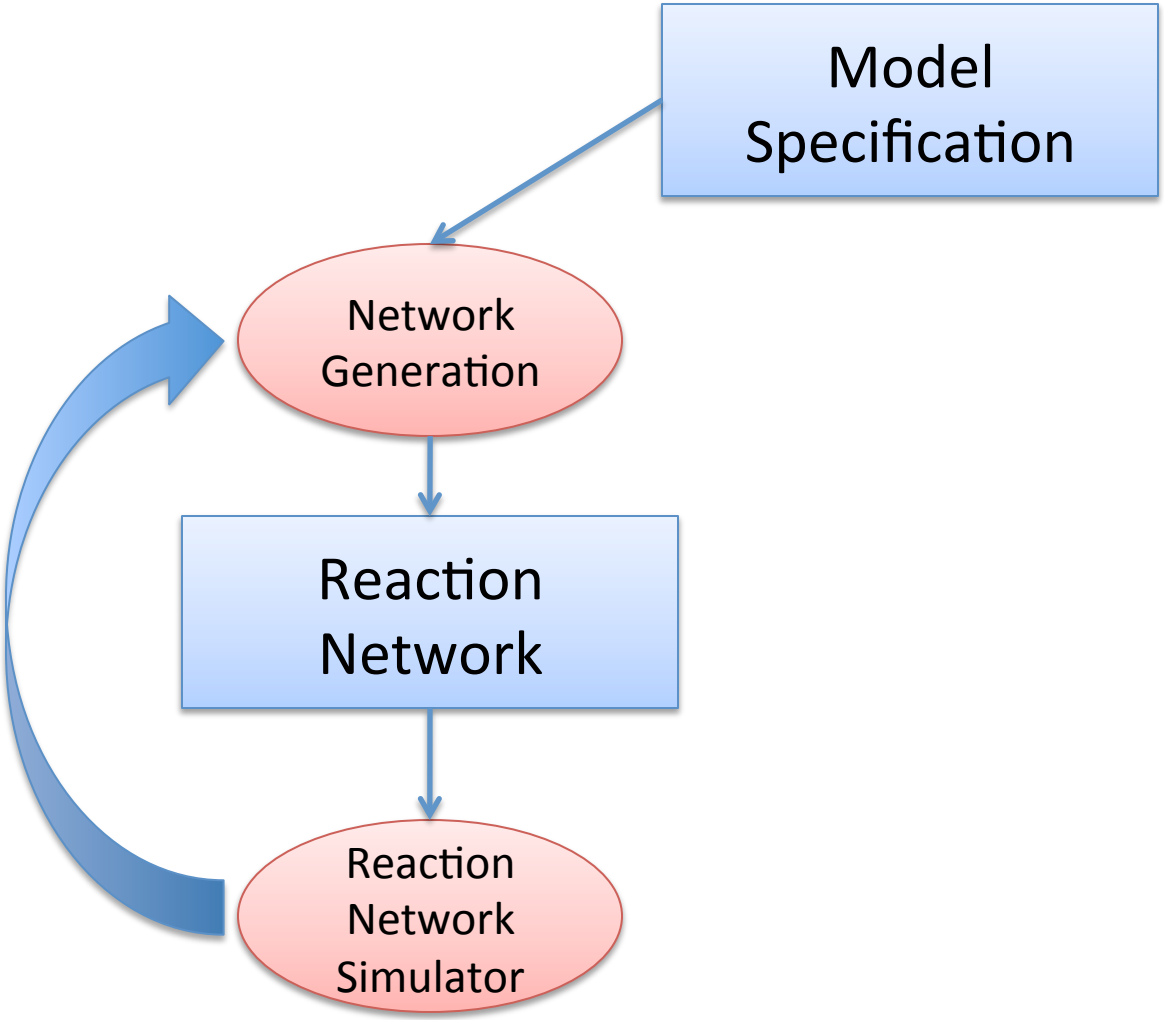


Problem: Network generation is not robust to combinatorial complexity. Almost any realistic description will lead to explosion.

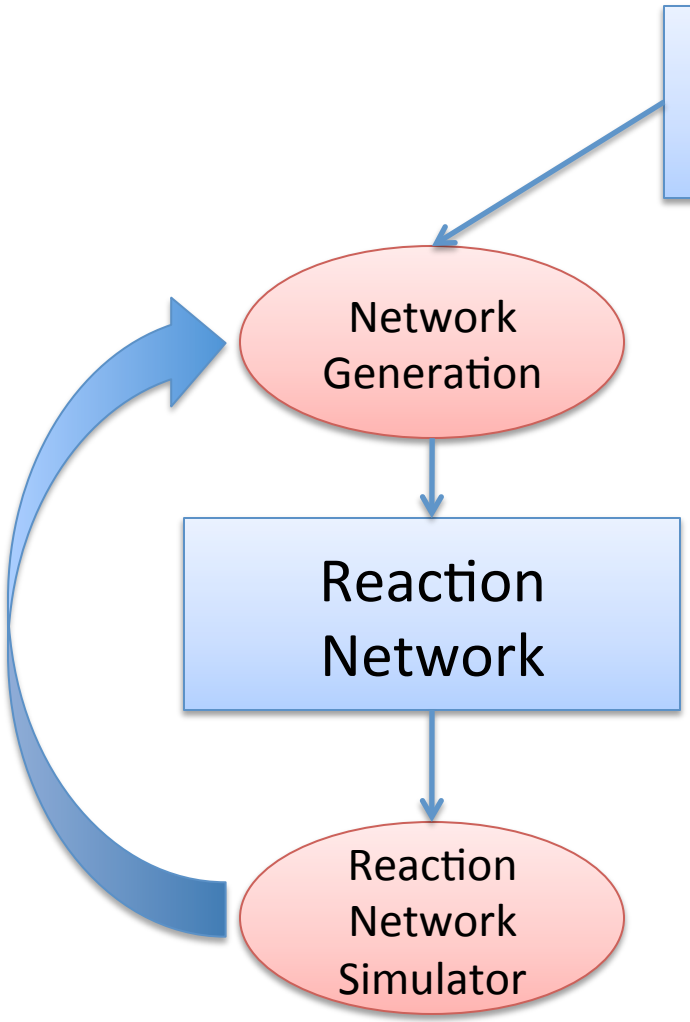


> 10^4 species

On-the-Fly Extends Range of Network Generation Approach



On-the-Fly Extends Range of Network Generation Approach



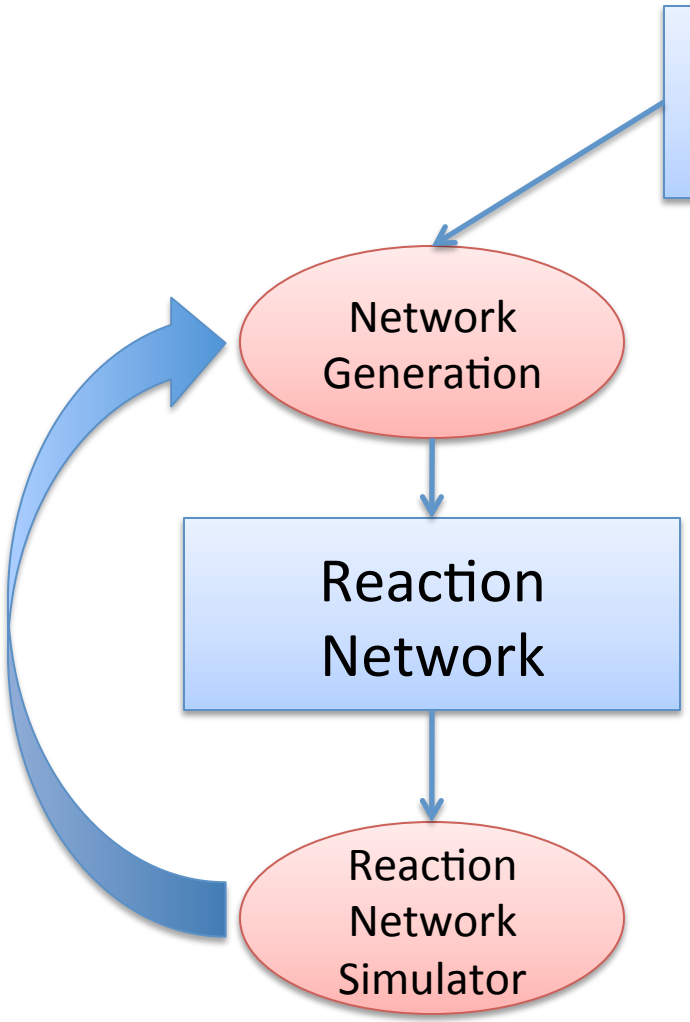
Lok and Brent (*Nat. Biotechnol.*,2005) – *Moleculizer*

Faeder et al. (*Complexity*, 2005) – *BioNetGen*

Meier-Schellersheim et al. (PLoS Comp. Biol., 2006; *Nature Methods*, 2012) – *Simmune*

Andrews et al. (PLoS Comp. Biol., 2010) – *Smoldyn 2.1*

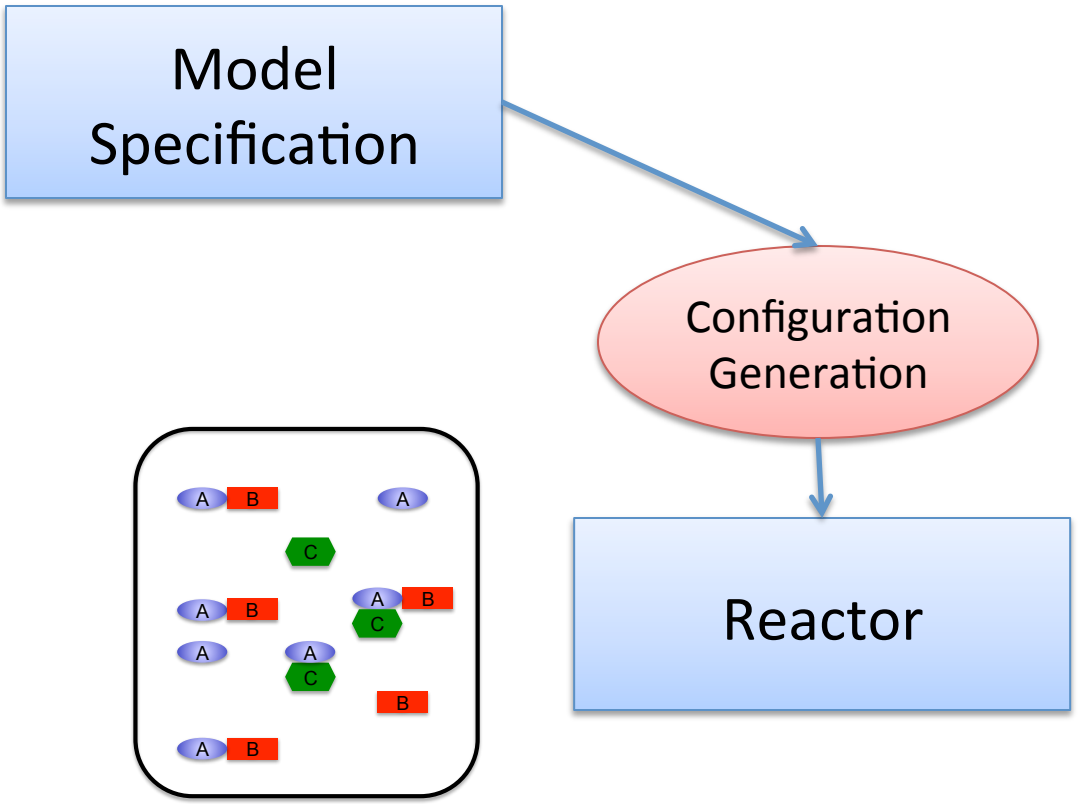
On-the-Fly Extends Range of Network Generation Approach



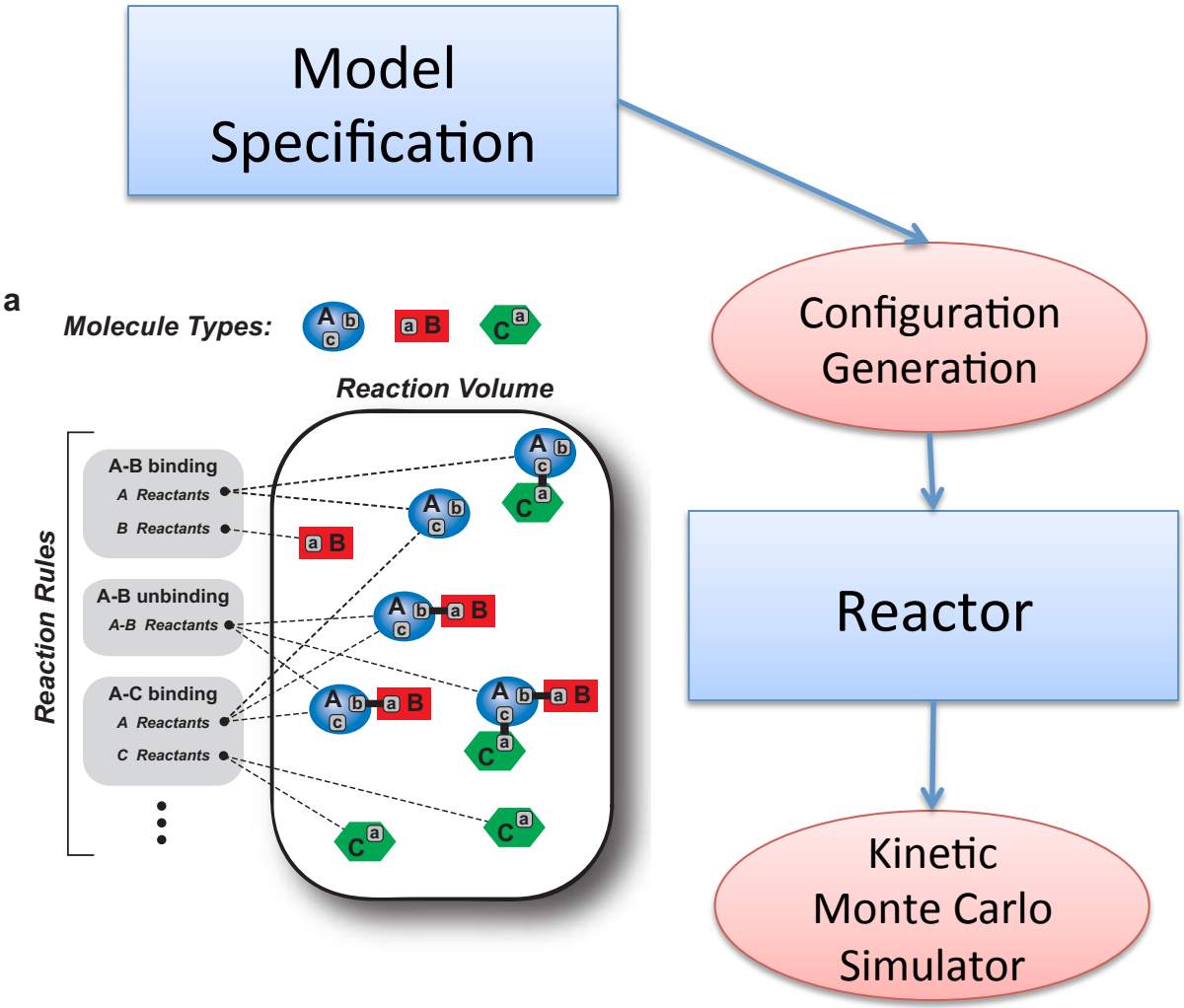
May break down when new species are generated frequently.

- Oligomerization
- # species >> # molecules

Agent-Based Approach Avoids State-Space Explosion

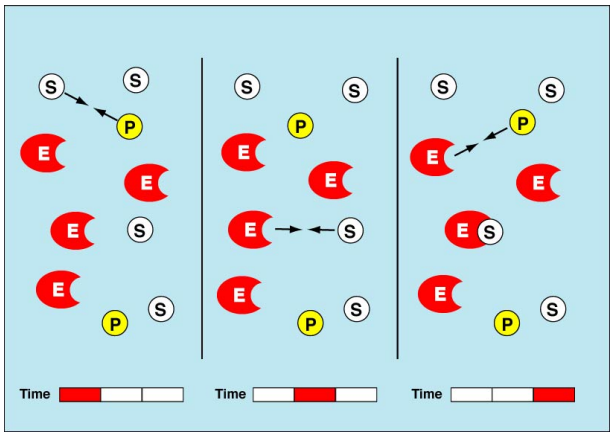


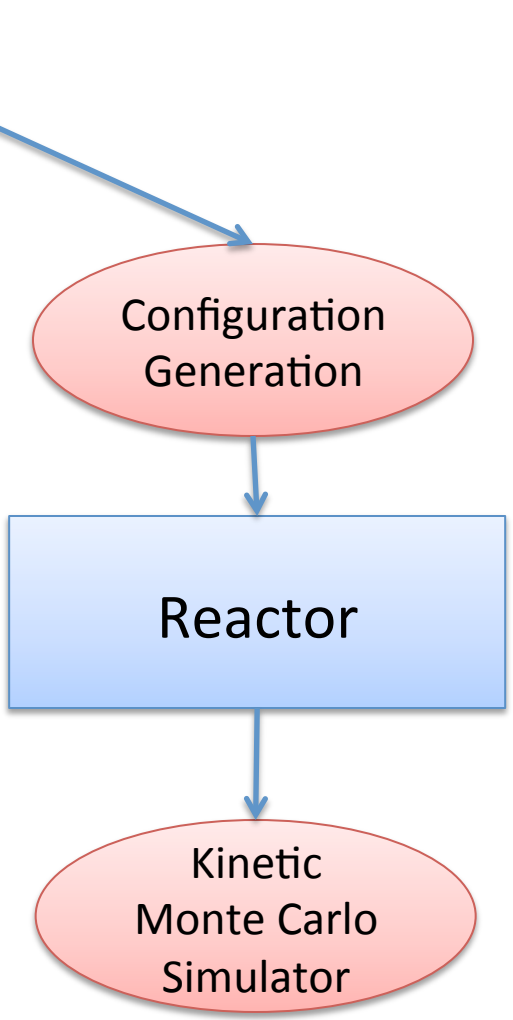
Agent-Based Approach Avoids State-Space Explosion



Agent-Based Approach Avoids State-Space Explosion

StochSim (1997) was first agent-based simulator for biochemical systems.



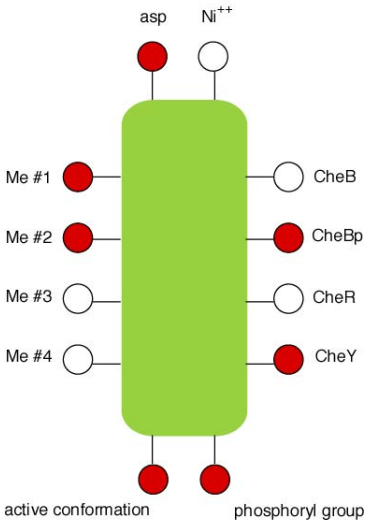


Agent-Based Approach Avoids State-Space Explosion

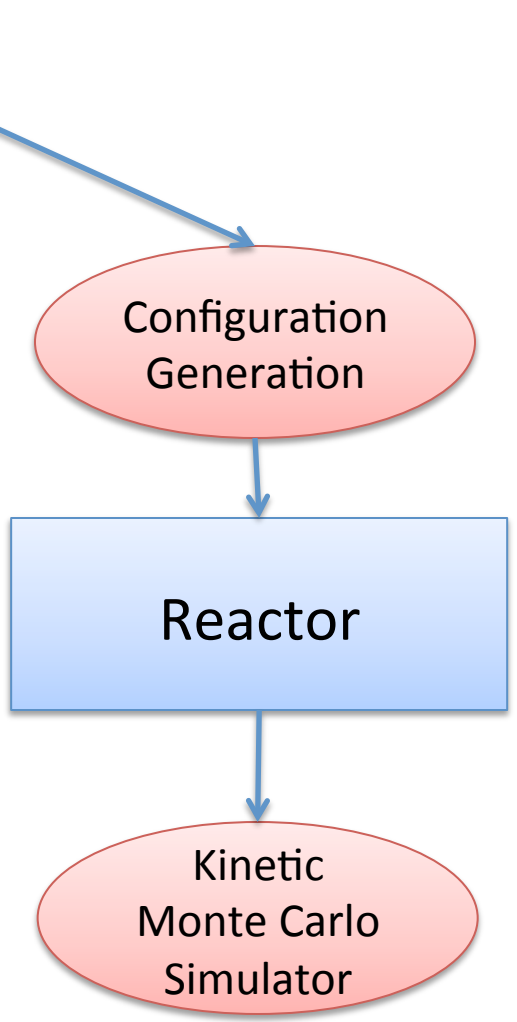
StochSim (1997) was first agent-based simulator for biochemical systems.

Two main drawbacks.

1. Limited connectivity of molecules.
2. Slow simulation algorithm based on random selection of reactants ➔Rejection of most reaction attempts



StochSim Agent

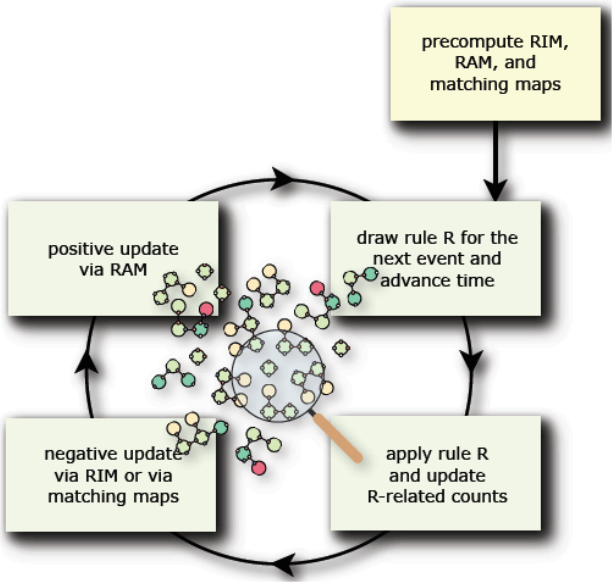


Scalable simulation of cellular signaling networks

Vincent Danos[1,4⋆], Jérôme Feret[3], Walter Fontana[1,2], and Jean Krivine[5]

APLAS '07 (invited paper)

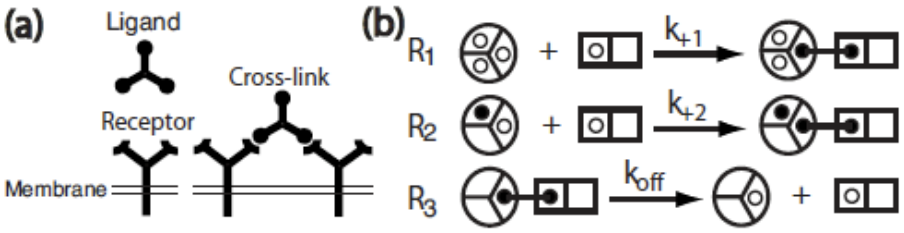
Based on Gillespie stochastic simulation algorithm for propagating CTMC's.



PHYSICAL REVIEW E **78**, 031910 (2008)

Kinetic Monte Carlo method for rule-based modeling of biochemical networks

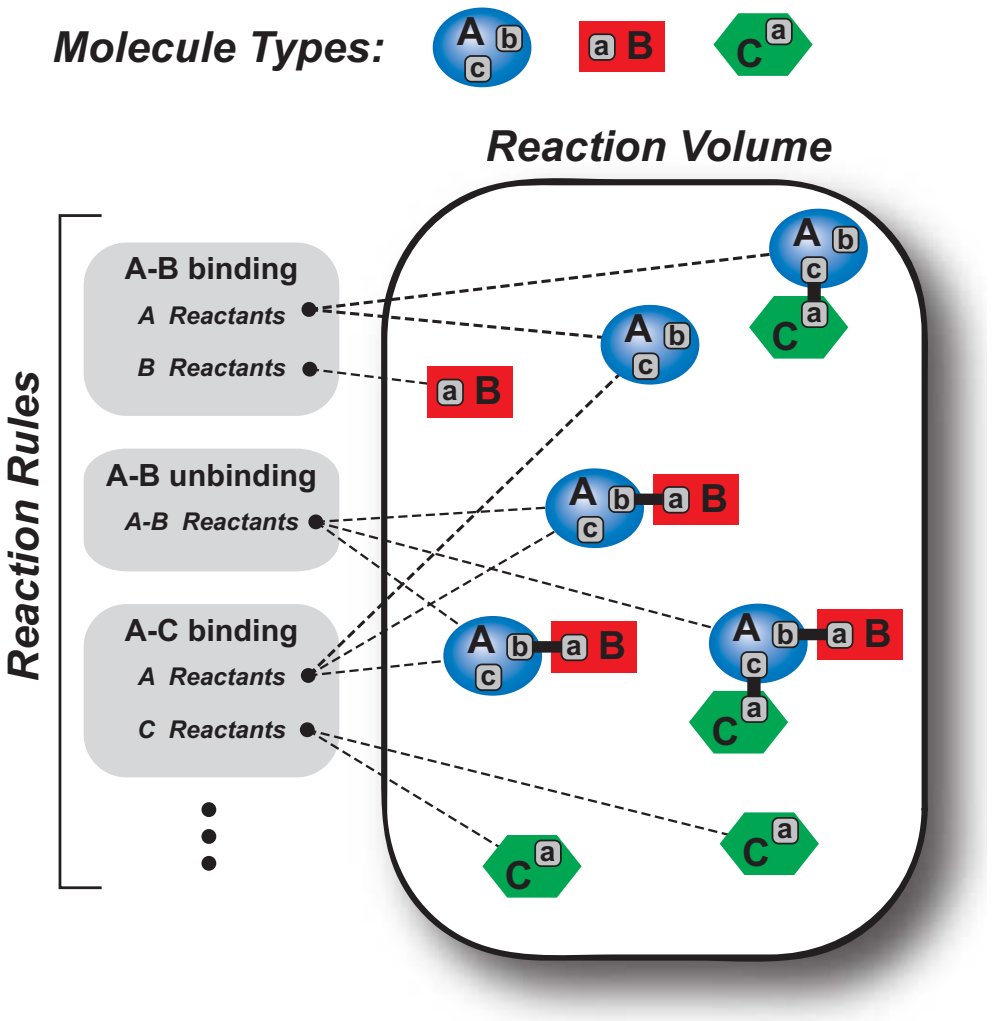
Jin Yang,[1,*] Michael I. Monine,[2] James R. Faeder,[3,†] and William S. Hlavacek[2,‡]



Network-Free Stochastic Simulator
NFsim

Michael Sneddon
Thierry Emonet
Yale University

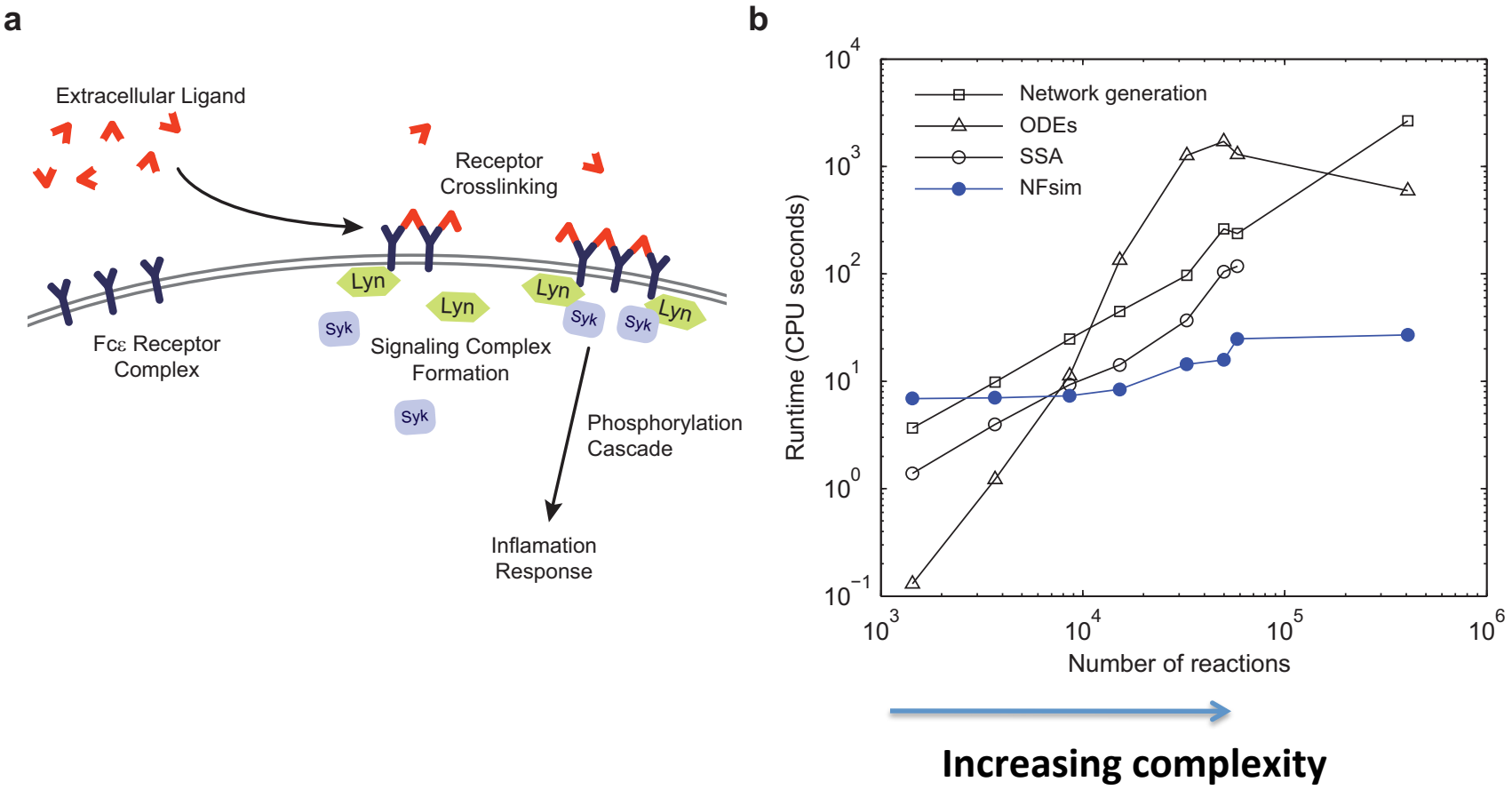
- Generalizes and extends Yang et al. algorithm
- Simulate any BioNetGen model
- Generalized rate laws / Coarse-grained reactions
- Highly efficient implementation
- Freely-available and open source.



Sneddon et al. (2011) *Nat. Methods*, **8**, 177

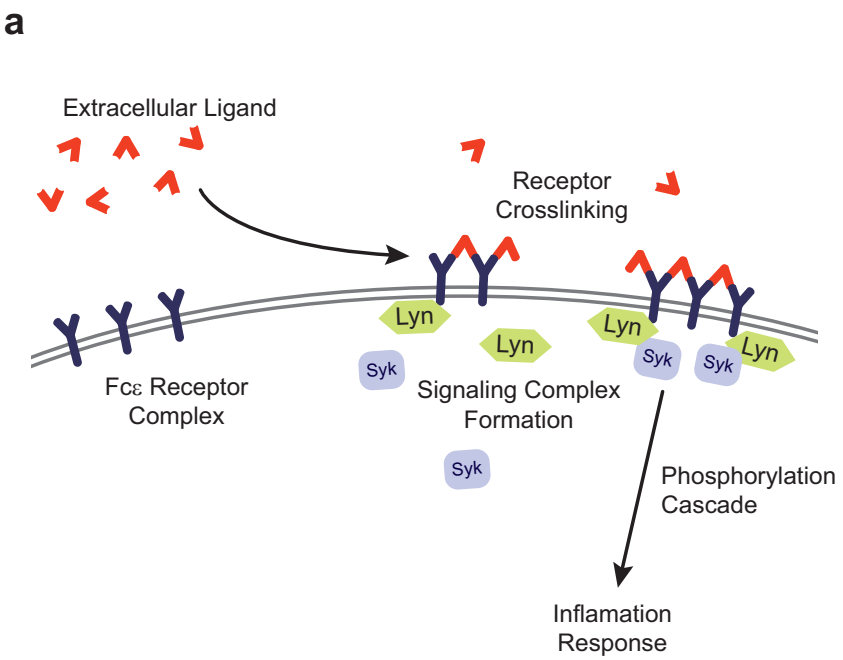
http://emonet.biology.yale.edu/nfsim/

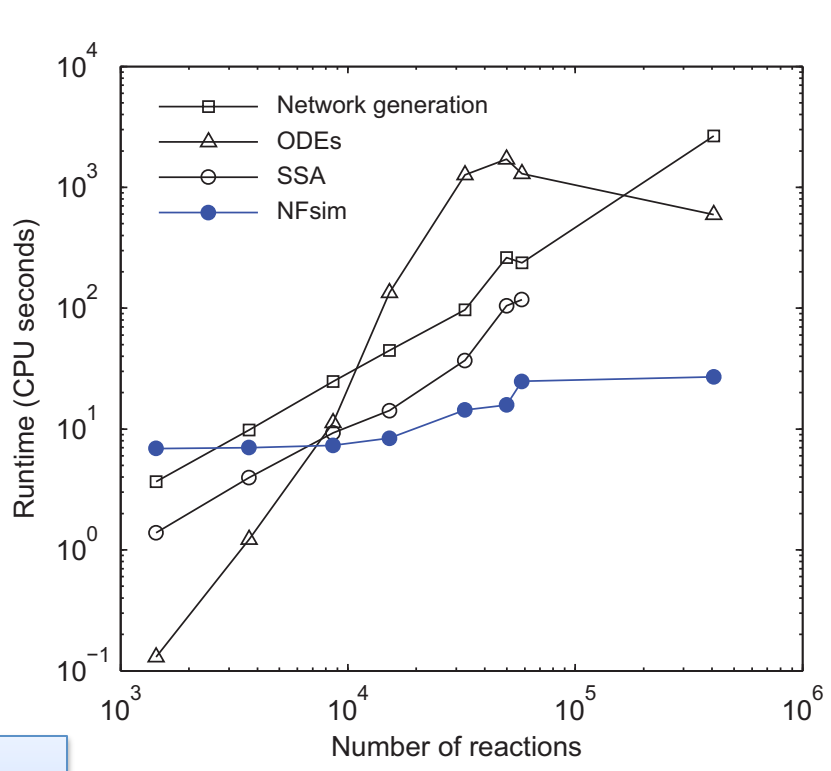
FcεRI signaling models



Sneddon et al. (2011) *Nat. Methods*, **8**, 177

FcεRI signaling models

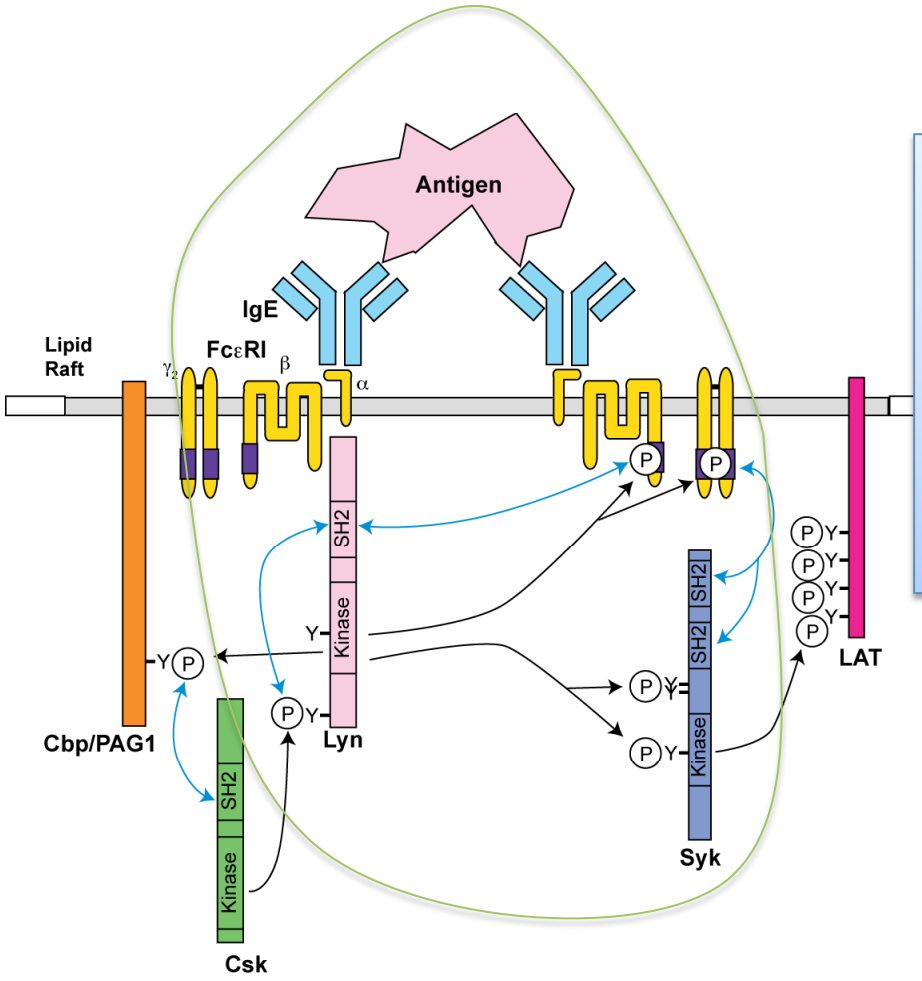




NFSIM can simulate models of greatly increased complexity with manageable increase in cost.

Increasing complexity

Syk activation model

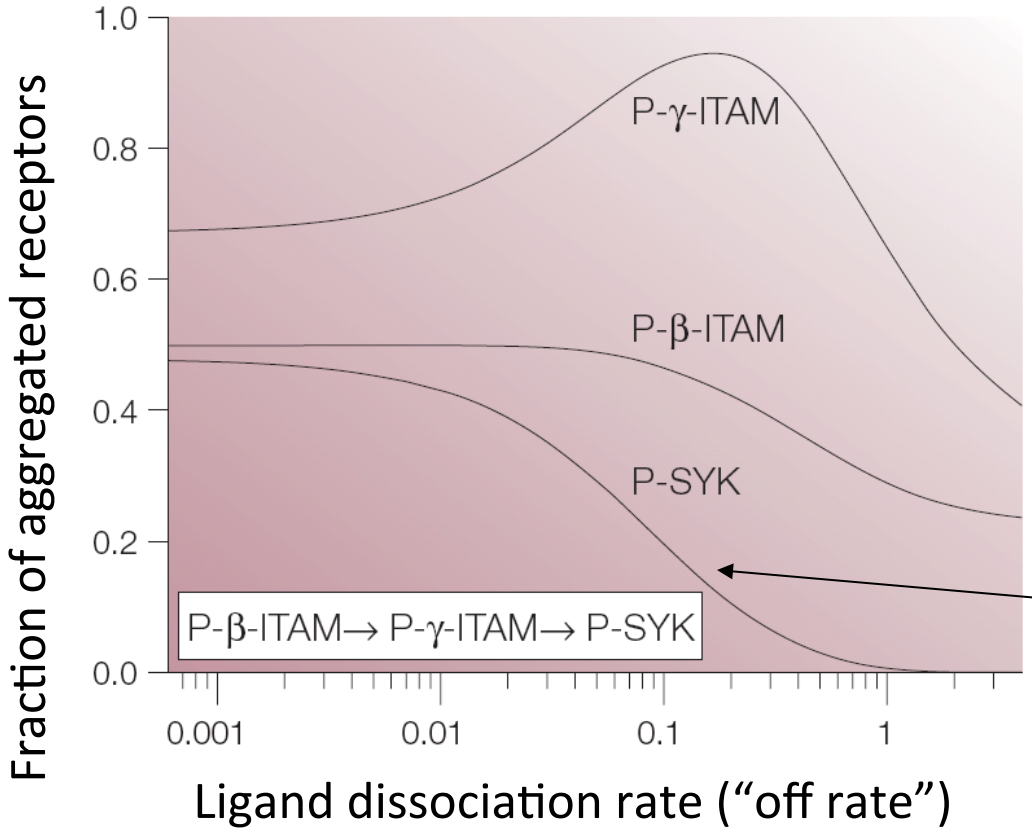


Key variables

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

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

Kinetic proofreading of Syk activation but not receptor phosphorylation

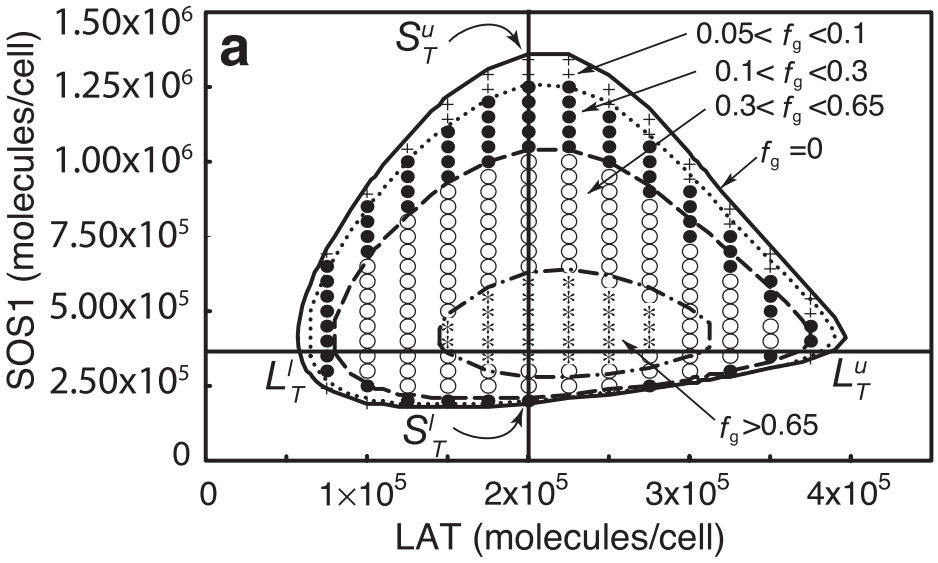


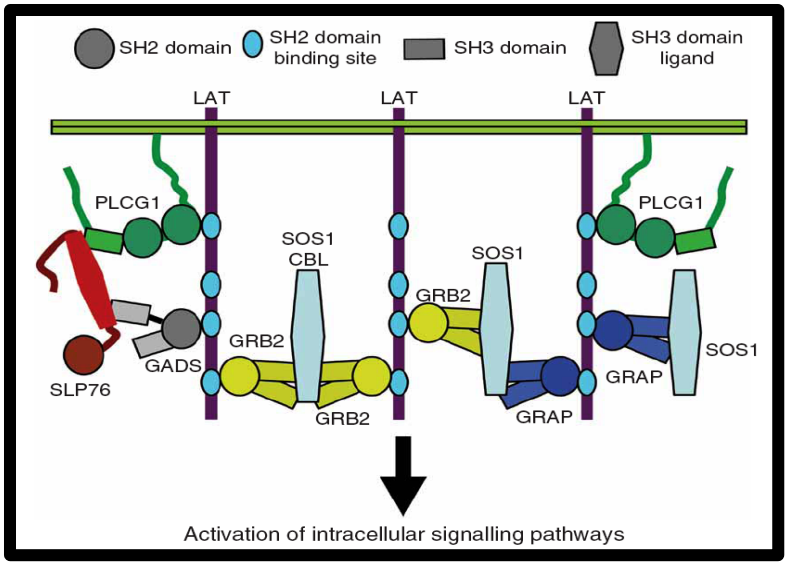
Rapid fall in efficiency of Syk phosphorylation

Goldstein *et al.* (2004) *Nat. Rev. Immunol*. **4**, 445-456.

Extension of the Syk activation model

- LAT may form large oligomers under physiological conditions.

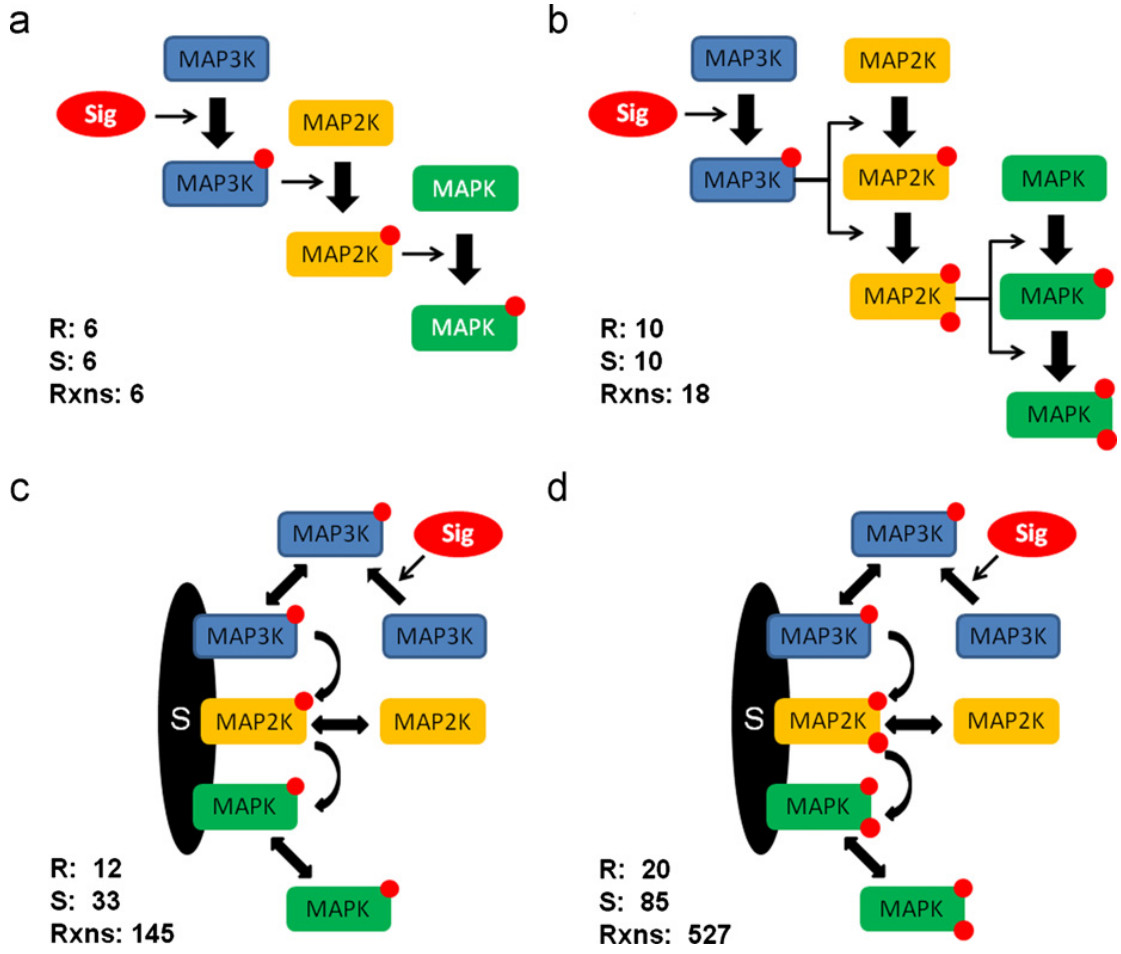




Houtman et al., *Nat. Struct. Mol. Biol.* (2006)

Nag et al., *Biophys. J.* (2009)

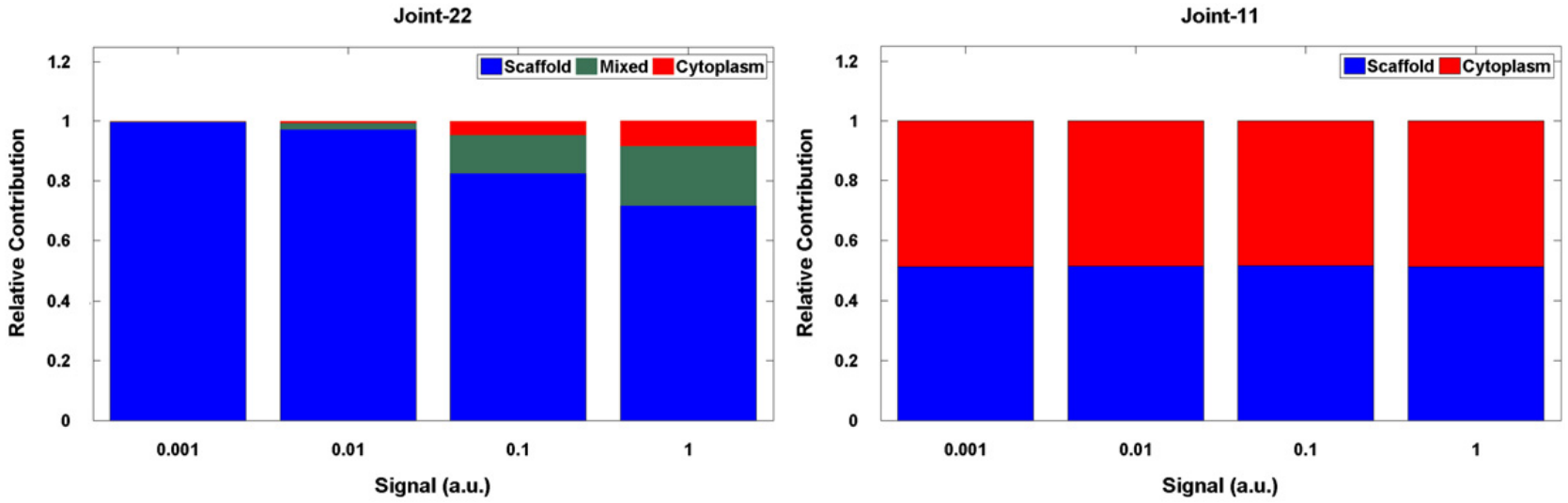
Interplay of double-phosphorylation and scaffolding



Kocieniewski, Faeder, and Lipniacki, *J. Theor. Biol.* **295**, 116-124 (2012).

Interplay of double-phosphorylation and scaffolding

Key finding: Requirement for double phosphorylation directs signal through scaffolds preferentially over cytoplasm



Kocieniewski, Faeder, and Lipniacki, *J. Theor. Biol.* **295**, 116-124 (2012).

Toward comprehensive models

- RBM offers the potential for high-resolution models of large-scale networks
  - Is the development of such models feasible?
  - Will such models be useful?

Future of rule-based modeling

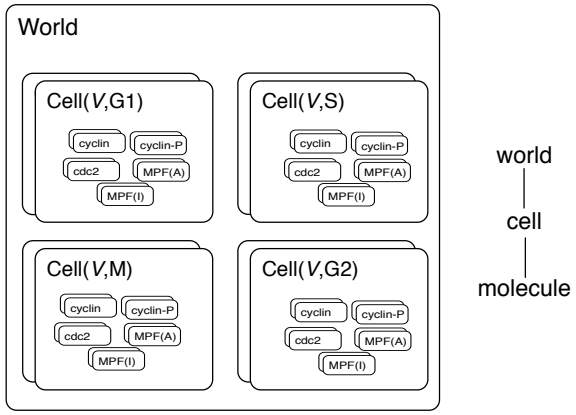
- Despite many successes, rule-based modeling remains a niche in systems biology.
- Many recognize that it may be necessary for some systems or in the future, but continue to use or promote standard methods based on chemical reaction networks.
- These approaches are powerful, but they do not seem to appeal to most biologists.
- Can rule-based modeling have a broader appeal?

Requirements for modeling tools

- Descriptive
  - Complex biochemistry
  - **Spatial organization**
- Accessible
  - Visual interfaces for biologists that present common concepts and hide mathematical details
  - Programming interfaces for developers and advanced modelers
- Scalable
  - Visualizations that do not become unwieldy for large models
  - Simulation methods that can handle combinatorial complexity
    - Model reduction
    - Accelerated stochastic simulations
  - Tools for model analysis – relating models to data / experiments
    - parameter estimation
    - uncertainty estimation
    - structure identification

Future Challenges

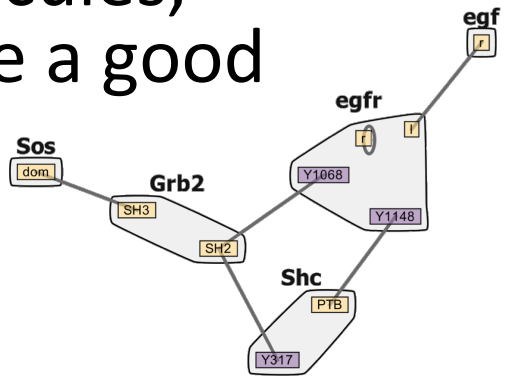
- Improving the efficiency of network-free simulation
  - memory (Hogg)
  - leaping – many events at one time
- Spatial dynamics
  - cell compartments
  - PDEs (Vcell,Simmune)
  - subvolumes (SSC)
  - particles (Meredys)
  - molecular scale (SRSim)
  - multicellular dynamics (ML-Rules, Simmune)
- Parameter estimation
  - Need for uncertainty estimates in model predictions
  - Standard methods apply
- Structure determination / refinement
  - Identification of missing interactions
  - Methods adapted to rules



Maus et al., BMC *Syst Biol* (2011)

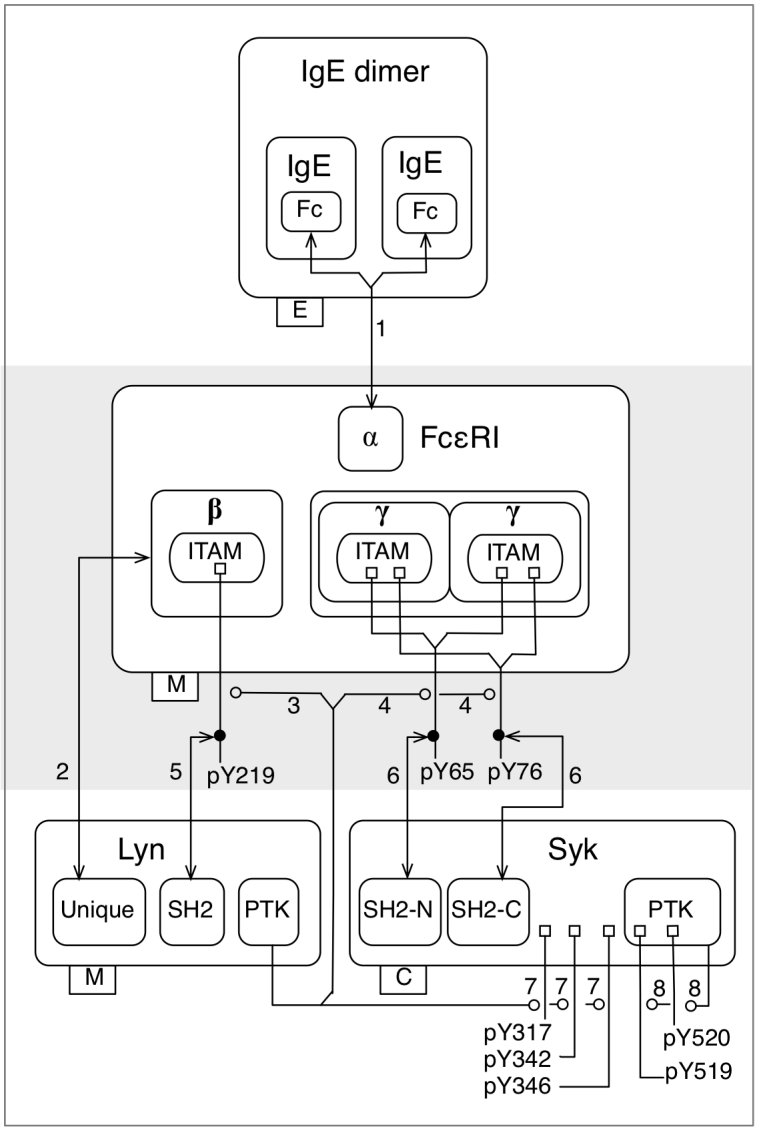
Interfaces

- For biologists the interfaces need to be visual – a network representation is natural
- Contact maps, which show molecules, components and interactions are a good start.



- Extended contact maps (Chylek et al.) show more detail and are scalable….

Extended contact maps



Reveals the big picture:

1) **a box for each molecule in a model**
2) **an arrow for each interaction in a model**
3) **a flag for each PTM in a model.**

Do not yet provide an executable representation. The translation step is a barrier.

Chylek et al. *Mol. Biosyst.* (2011).

Interfaces II

- At the other end, “power modelers” need open frameworks to build new capabilities, such as higher level organization of models (e.g. PySB, Sekar poster).
- Open source is not enough – open architecture, e.g., well-designed and documented API’s can facilitate and greatly increase the productivity of such efforts.
  - Current situation is worrisome.

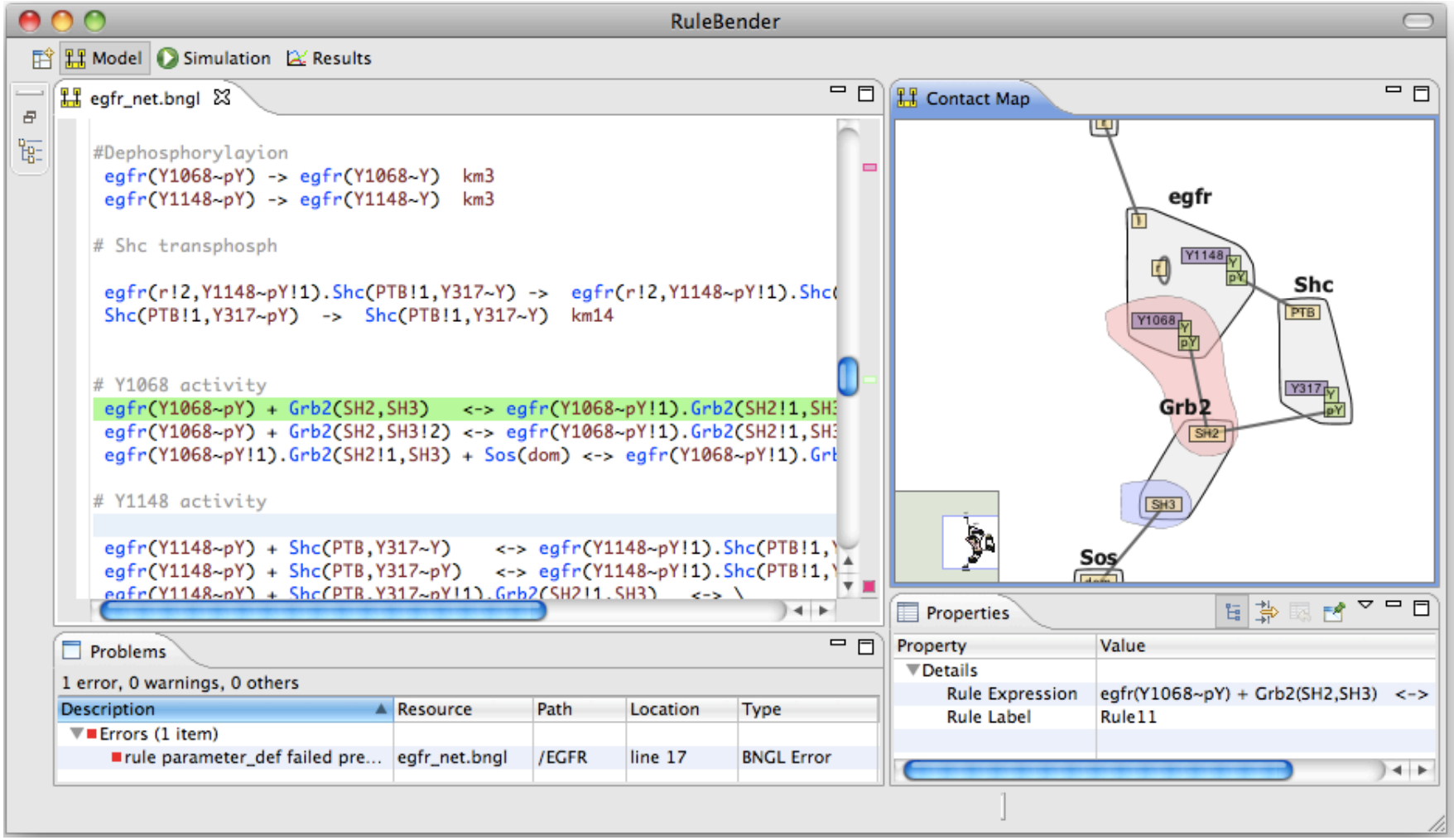
An IDE may have the potential to address both needs

- Integrates programming and graphical interfaces
- Eclipse-based IDE's have been developed for both Kappa and BNG
  - RuleStudio
  - RuleBender
- Views can be customized to support different workflows and levels of visual or text-based interaction.

RuleBender

Built in Eclipse RCP

http://rulebender.org



Xu et al. *Bioinformatics* (2011); Smith et al. *BioVis12 (Best Paper)*; *BMC Bioinformatics* (2012)

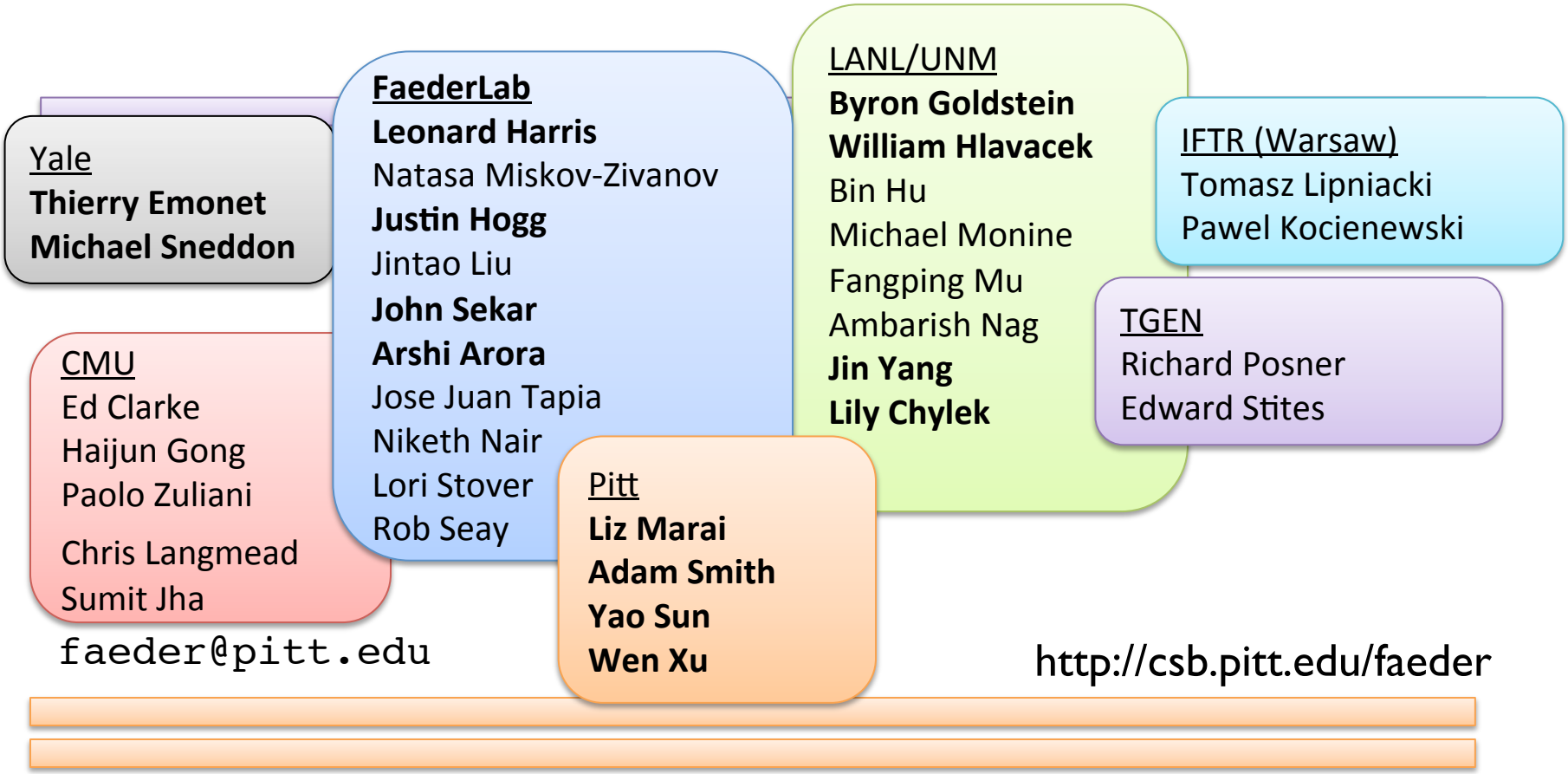
CMACS Pancreatic Cancer Challenge Project

- Integrate 12 core signaling pathways that have been identified from cancer genome project as playing a key role in the disease.
  - Mechanistic explanation for observed mutations
  - Testbed for potential therapeutic mechanisms
- What tools and organization are needed to make such an effort successful?
  - BioNetGen modeling language with some basic annotation standards – protein naming conventions
  - Wiki and model repository

Community Building

- **Language interoperability**
  - SBML L3 Multi – accepted but not implemented
  - BNG <-> Kappa translators
- **Rule-based modeling libraries**
  - Can we pool our efforts to develop a common repository for models? Ongoing?
  - What standards are needed for model annotation to make models interoperable?
  - Simulation benchmarks
- **Common visual interfaces**
  - Development of an IDE with support for multiple rule-based modeling tools? Is RuleBender appealing for that?
  - Can ECM's be a standard for diagramming rule-based models and can it be integrated with SBGN effort?

Acknowledgments



http://bionetgen.org
http://emonet.biology.yale.edu/nfsim
http://rulebender.org

Funding: NIH (NIGMS), NSF (EMT), NSF (Expeditions in Computing), CTSI