Rules for Modeling Signal-transduction Systems

William S. Hlavacek

Signal Transduction, Oct. 11th, 14:00-16:30

Systems Biology of Basic Biological Systems

ICSB 2006

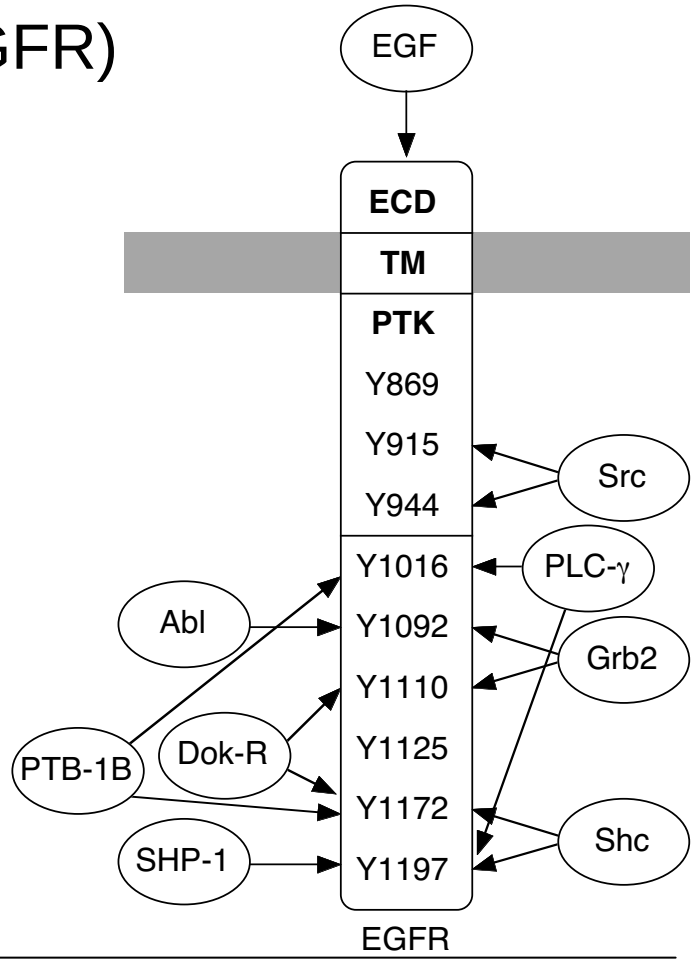
Los Alamos
NATIONAL LABORATORY
EST.1943

How do we predict the behavior of a system from knowledge of its parts and interactions?

- Combinatorial complexity
- Drawbacks of conventional modeling approaches
- Rule-based modeling - methods
  - Visual and computable representation of protein interactions with site-specific details
- An example of a rule-based model
- Future directions

Multiplicity of sites and binding partners gives rise to combinatorial complexity

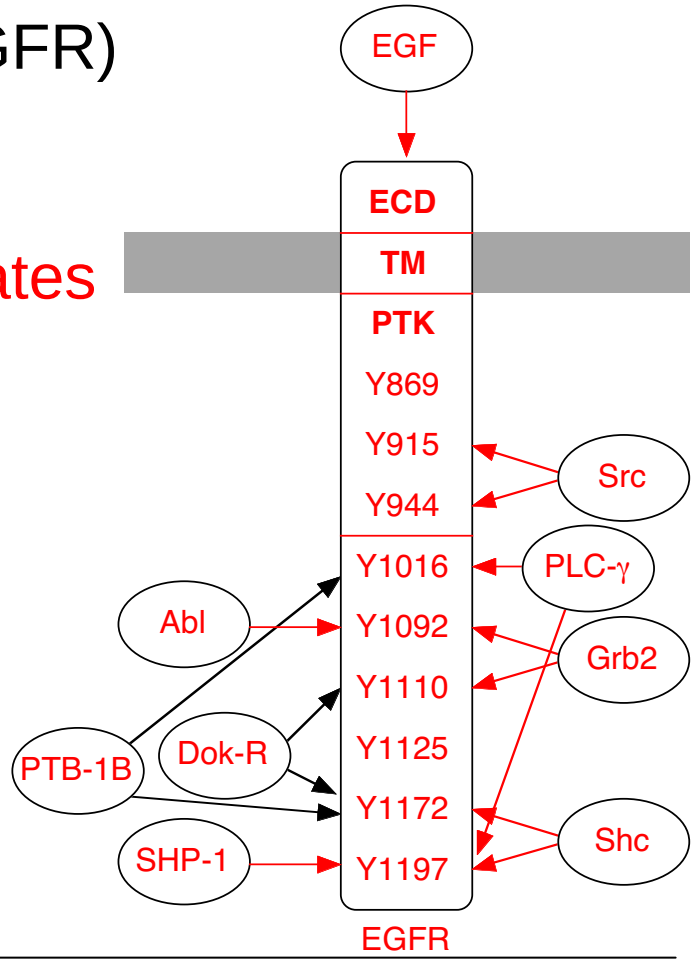
Epidermal growth factor receptor (EGFR)



Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

9 sites $\Rightarrow$ 2^9=512 phosphorylation states



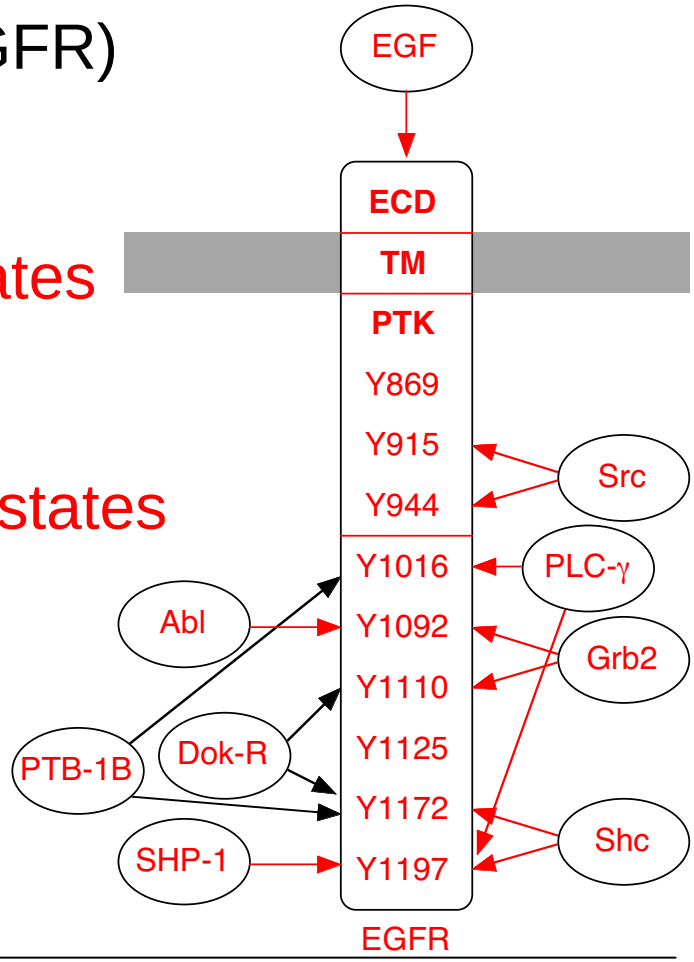
Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

9 sites $\Rightarrow$ 2^9=512 phosphorylation states

Each site has $\geq$ 1 binding partner

$\Rightarrow$ more than 3^9=19,683 total states



Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

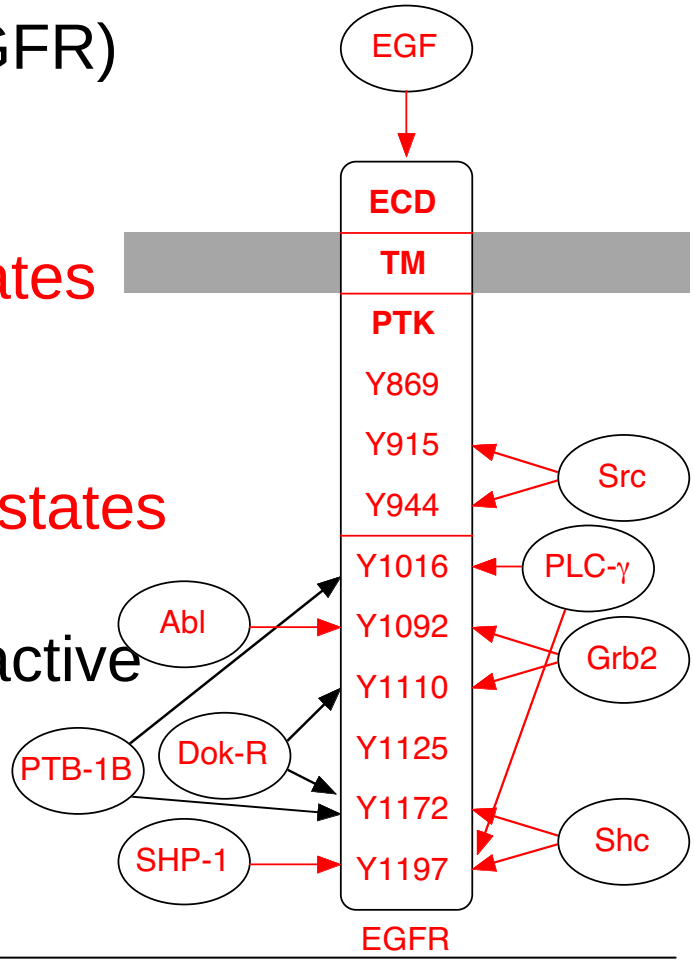
9 sites $\Rightarrow$ 2^9=512 phosphorylation states

Each site has ≥ 1 binding partner

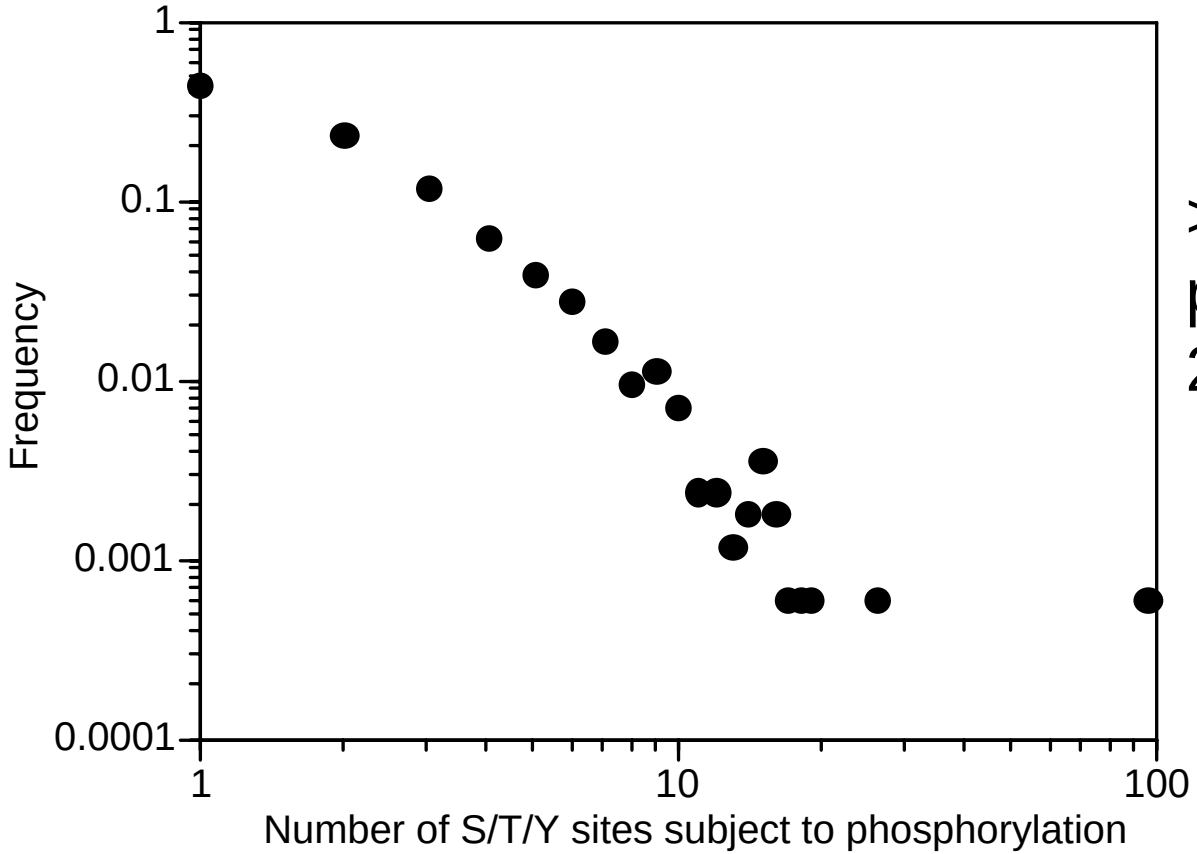
$\Rightarrow$ more than 3^9=19,683 total states

EGFR must form *dimers* to become active

$\Rightarrow$ more than 1.9×10^8 states



Signaling proteins typically contain multiple phosphorylation sites



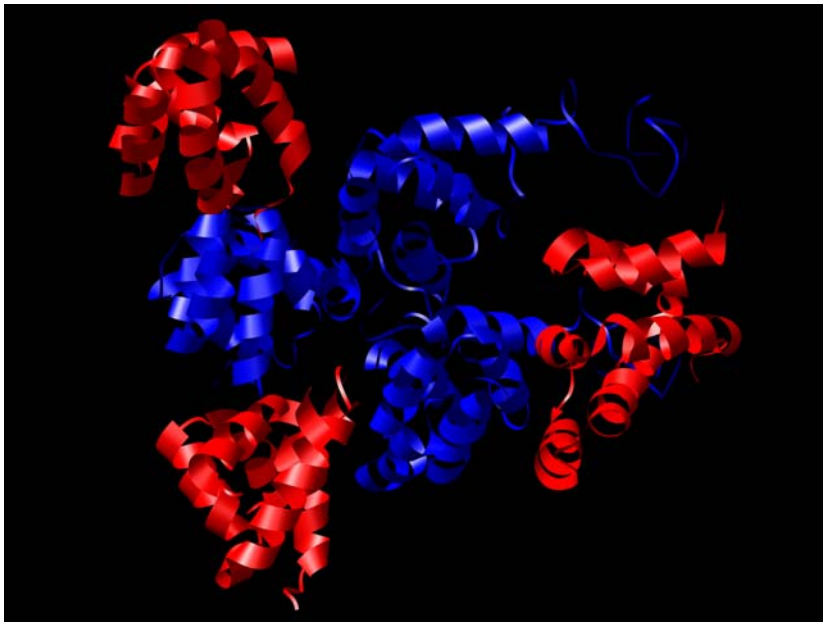
> 50% are phosphorylated at 2 or more sites

Phospho.ELM database v. 3.0 (http://phospho.elm.eu.org)

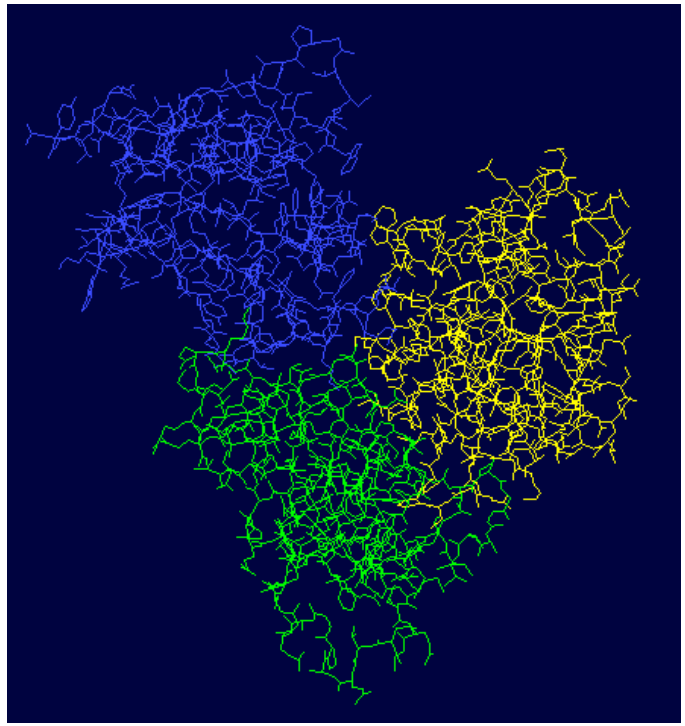
Oligomerization alone can generate many complexes

Complexes potentially involved in Toll-like receptor signaling

6 homodimers, 6 heterodimers, 20 homotrimers, 66 heterotrimers



A hexamer of death domains
Weber and Vincenz (2001) *FEBS Lett.*

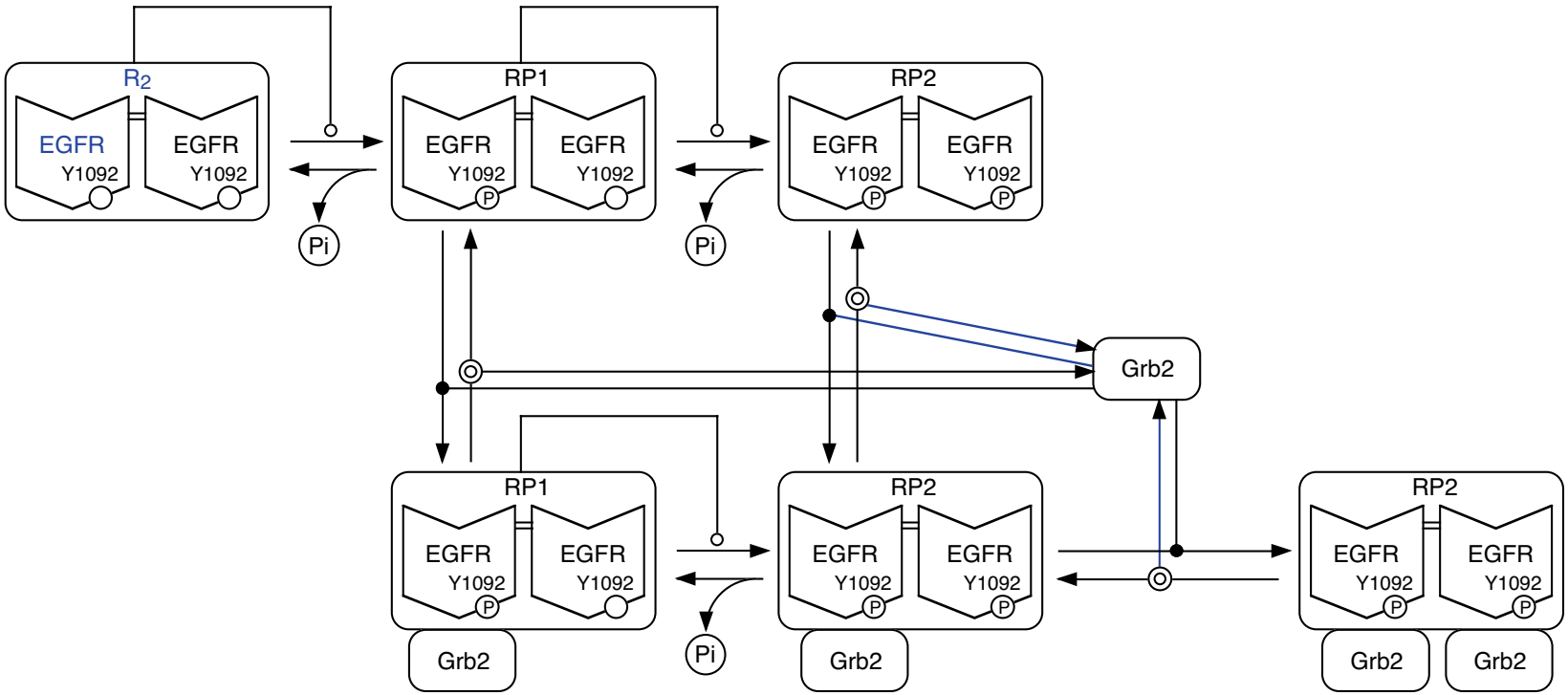


A trimer of TIR domains
C.-T. Tung (Los Alamos)

Outline

- Combinatorial complexity
- **Drawbacks of conventional modeling approaches**
- Rule-based modeling - methods
- An example of a rule-based model
- Future directions

Draw a reaction scheme diagram and translate the diagram into equations



$$d[RP1]/dt = 2k_p[R_2] - k_d[RP1] - k_p[RP1] + 2k_d[RP2]$$
$$-k_{on}[Grb2][RP1] + k_{off}[RP1\text{-}Grb2] \ldots$$

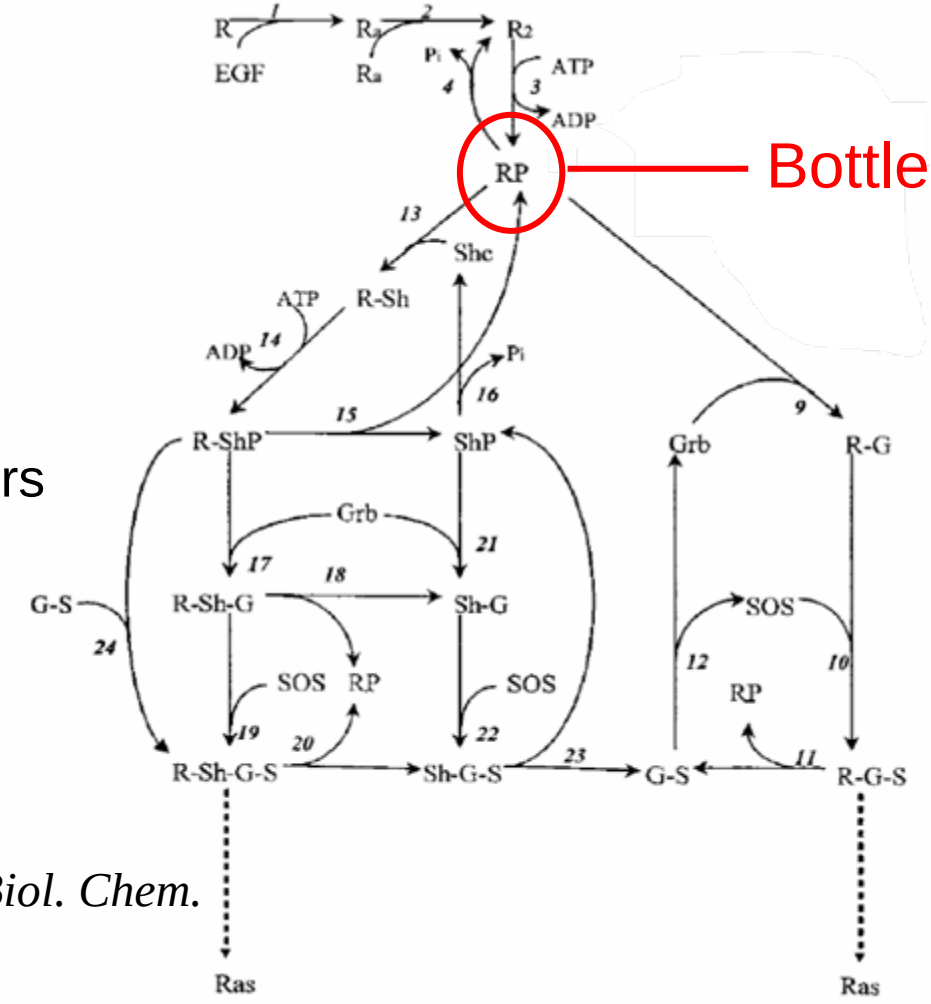
We need an equation for each chemical species and a term for each reaction…

- Sometimes it's easy to write the equations
  - 7 equations with 12 terms on the right-hand side for the scheme shown in the previous slide
- And sometimes it's difficult to write the equations
  - 356 equations with 3749 terms on the right-hand side for a model of signaling by the epidermal growth factor receptor [Blinov et al. (2006) *BioSystems*]

Assumptions are made to limit combinatorial complexity

Phosphorylation inhibits dimer breakup

No phosphorylated monomers



Kholodenko et al. (1999) *J. Biol. Chem.*

Los Alamos NATIONAL LABORATORY EST.1943

Outline

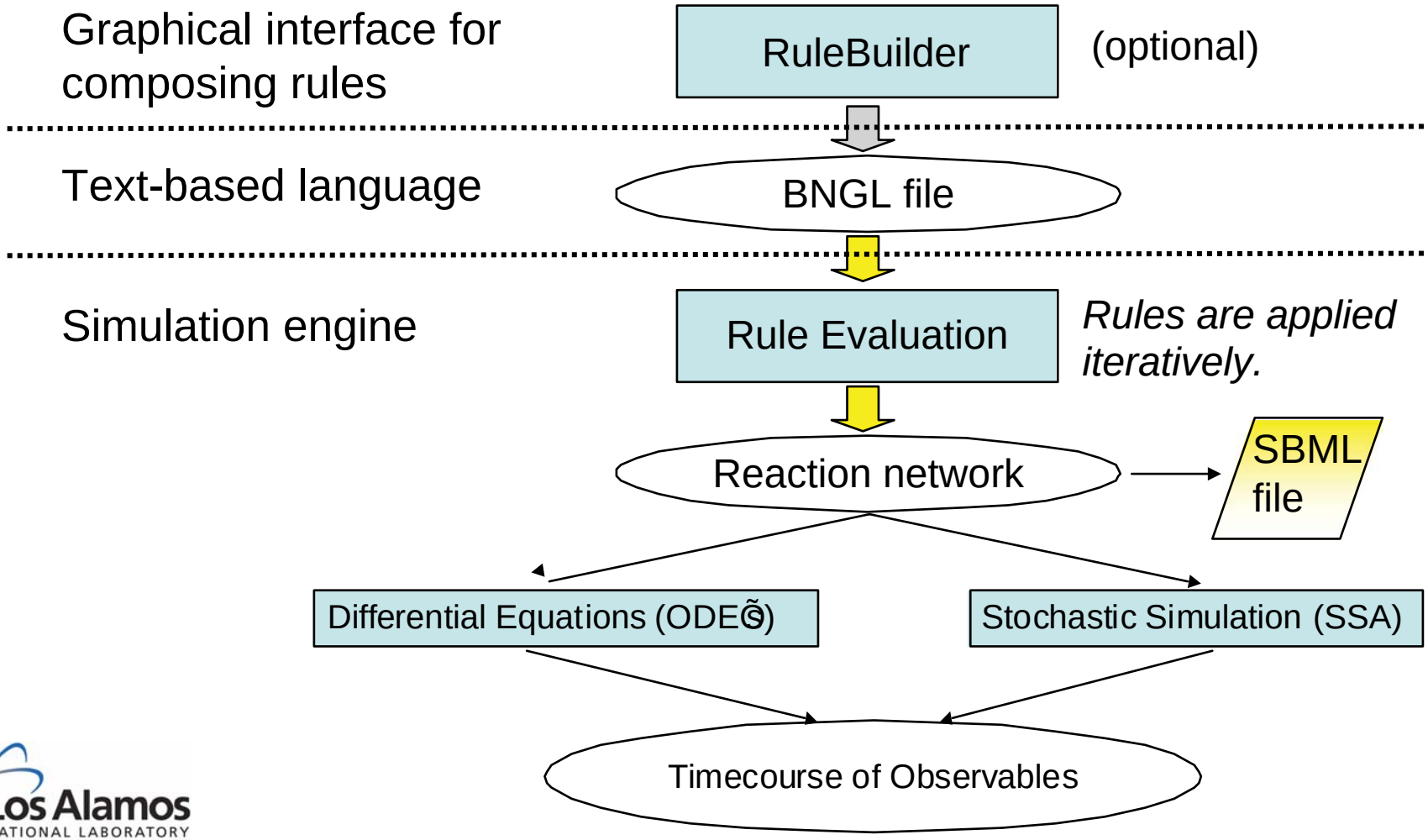
- Combinatorial complexity
- Drawbacks of conventional modeling approaches
- **Rule-based modeling - methods**
- An example of a rule-based model
- Future directions

Rule-based modeling in a nutshell

- Use graphs to represent molecules
  - Vertices correspond to molecular parts
  - Vertices of the same “color” comprise a molecule
  - Vertices are labeled - labels can be introduced for states (open, closed, phosphorylated, or not phosphorylated)
  - Edges represent bonds between molecular parts
- Use graph-rewriting rules to define interactions and their consequences
  - Involves pattern (subgraph) matching
  - Simple operations, adding or removing an edge or relabeling a vertex

Faeder et al. (2005) *Proc. ACM Symp. Appl. Computing*
Blinov et al. (2005) *Proc. BioCONCUR*

BioNetGen2: Software for graphical rule-based modeling

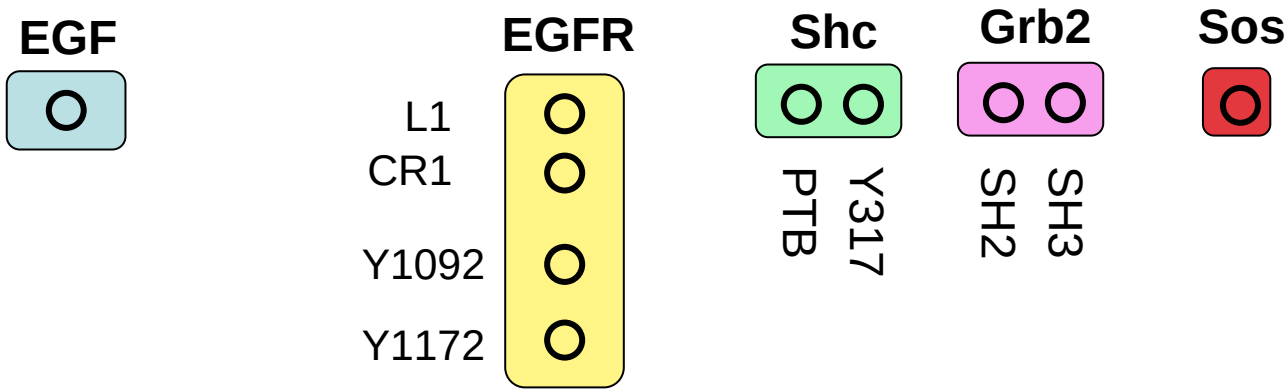


Other software for rule-based modeling

- StochSim
- Moleculizer
- Simmune

Five proteins involved in EGFR signaling represented as graphs

Molecule templates

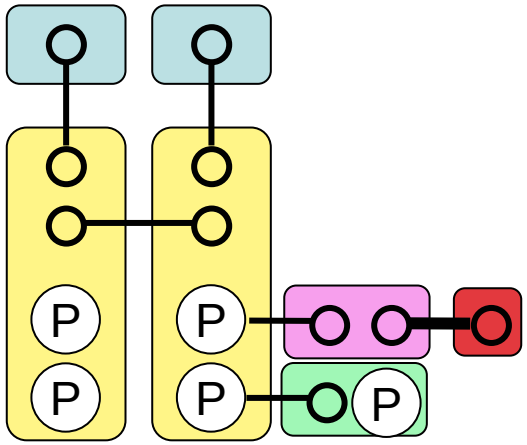


Vertices correspond to components of proteins

Components may have attributes: ○ or Ⓟ

Complexes are represented as connected instances of molecule graphs

An EGFR dimer

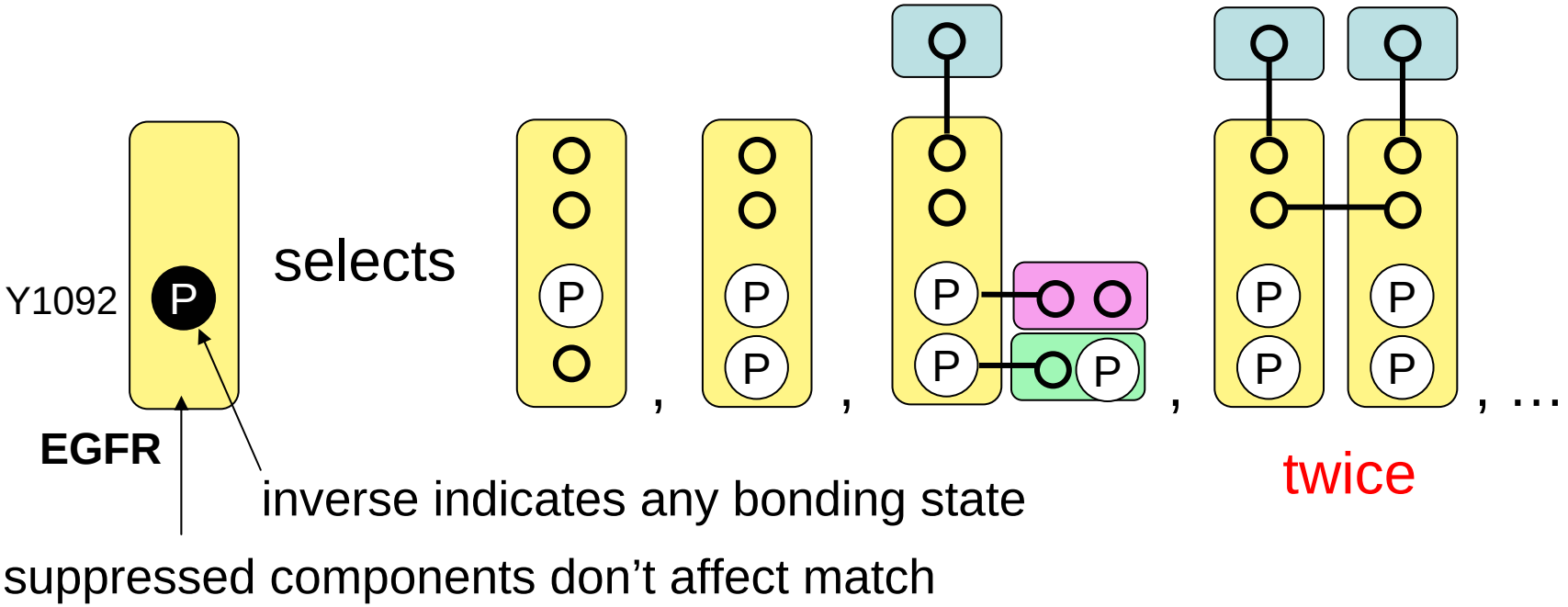


Edges represent bonds between components

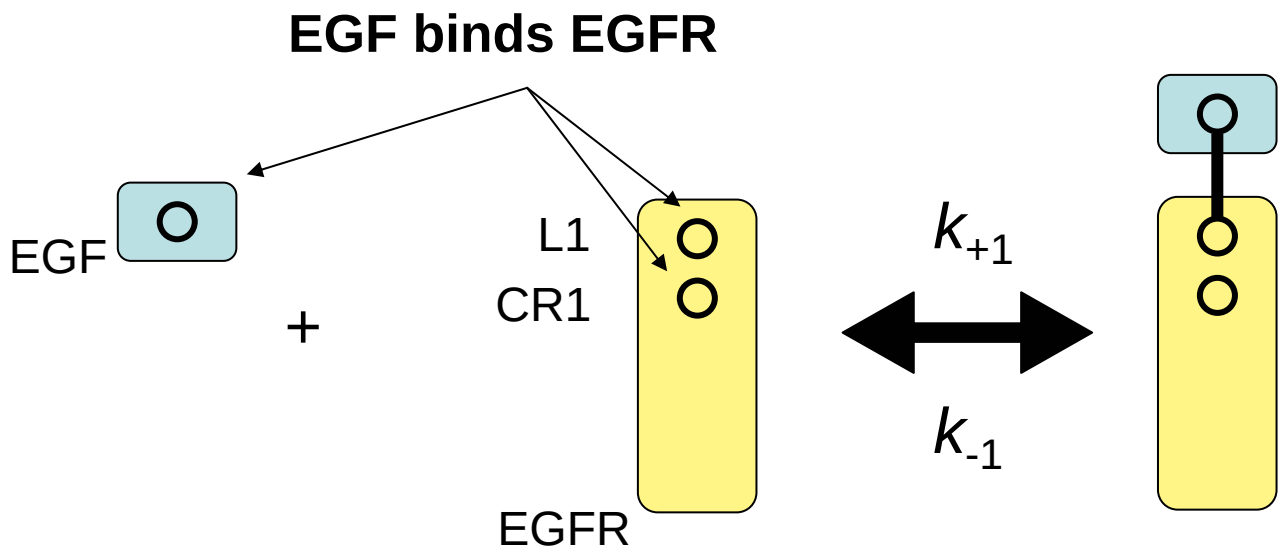
Bonds may be intra- or intermolecular

Sets of chemical species with common features are represented by patterns (subgraphs)

A pattern that selects EGFR phosphorylated at Y1092.



Interactions are represented as graph-rewriting rules, which are comprised of patterns



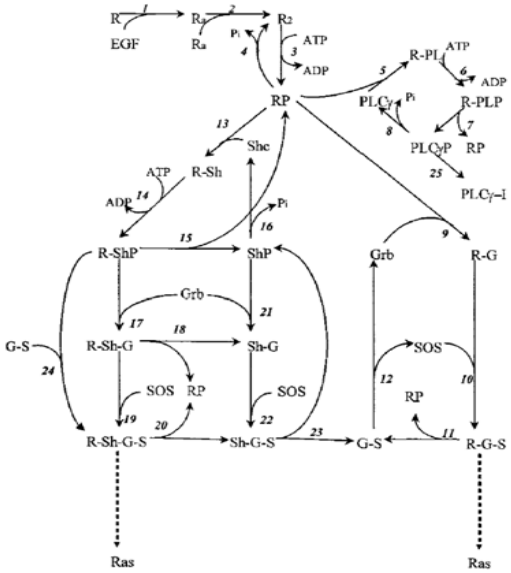
Patterns select reactants and specify graph transformation

- **Addition of bond between EGF and EGFR**

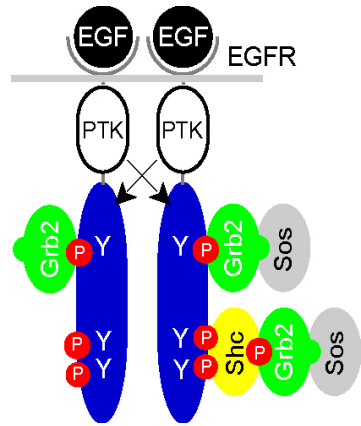
Rule-based version of the Kholodenko model

- 5 molecule types
- 23 reaction rules
- No new rate parameters (!)

18 species
34 reactions



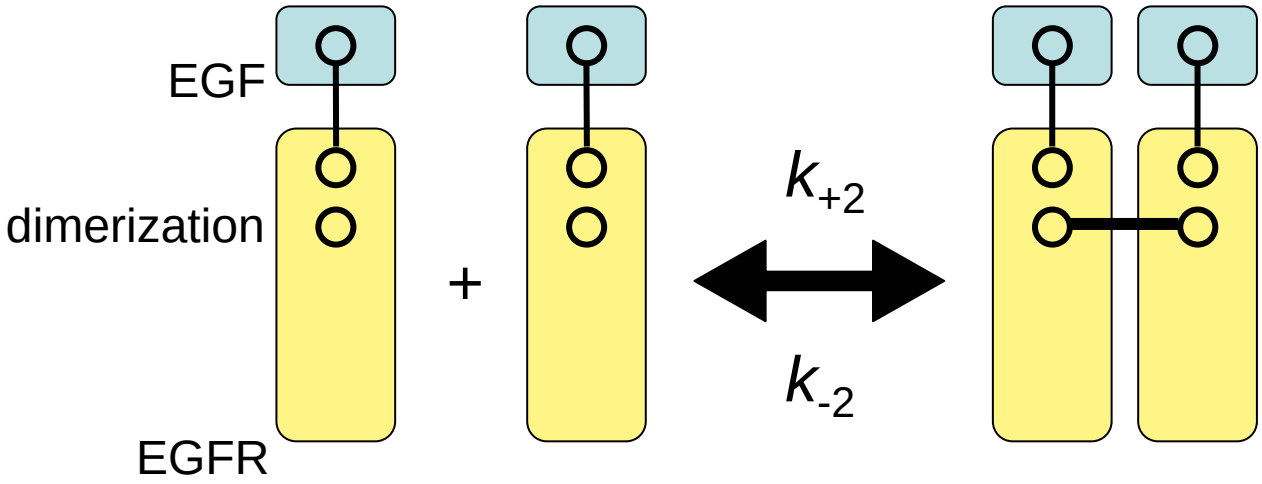
356 species
3749 reactions



Blinov et al. (2006) *BioSystems*

Dimerization rule eliminates previous assumption restricting breakup of receptors

EGFR dimerizes (600 reactions)



Dimers form and break up independent of phosphorylation of cytoplasmic domains

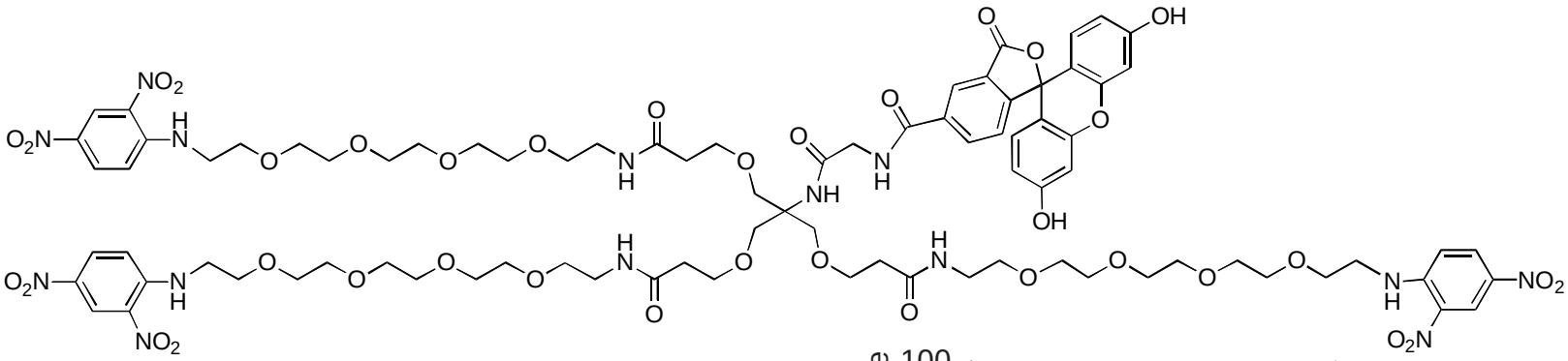
Outline

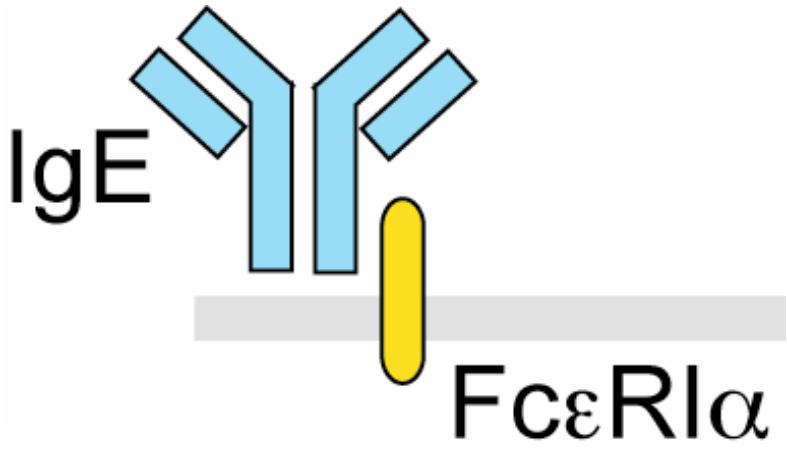
- Combinatorial complexity
- Drawbacks of conventional modeling approaches
- Rule-based modeling - methods
- **An example of a rule-based model**
- Future directions

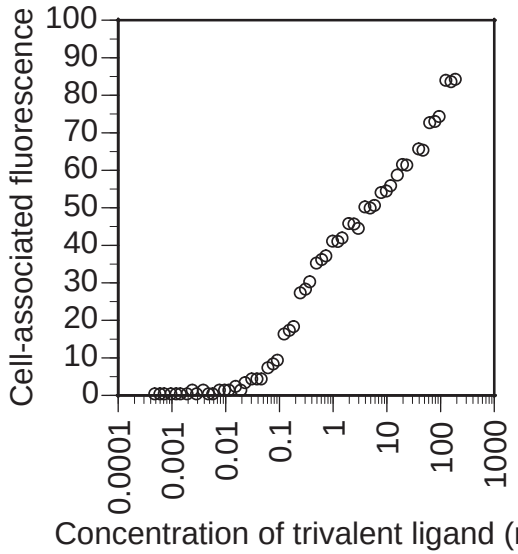
Creating a model

- Define graphs (molecules)
- Define graph-rewriting rules (interactions)
- The rest is automatic - we'll look at how that works

The system: interaction of a trivalent ligand with a bivalent cell-surface receptor



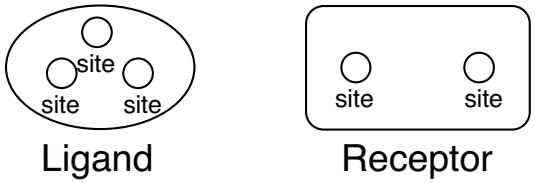




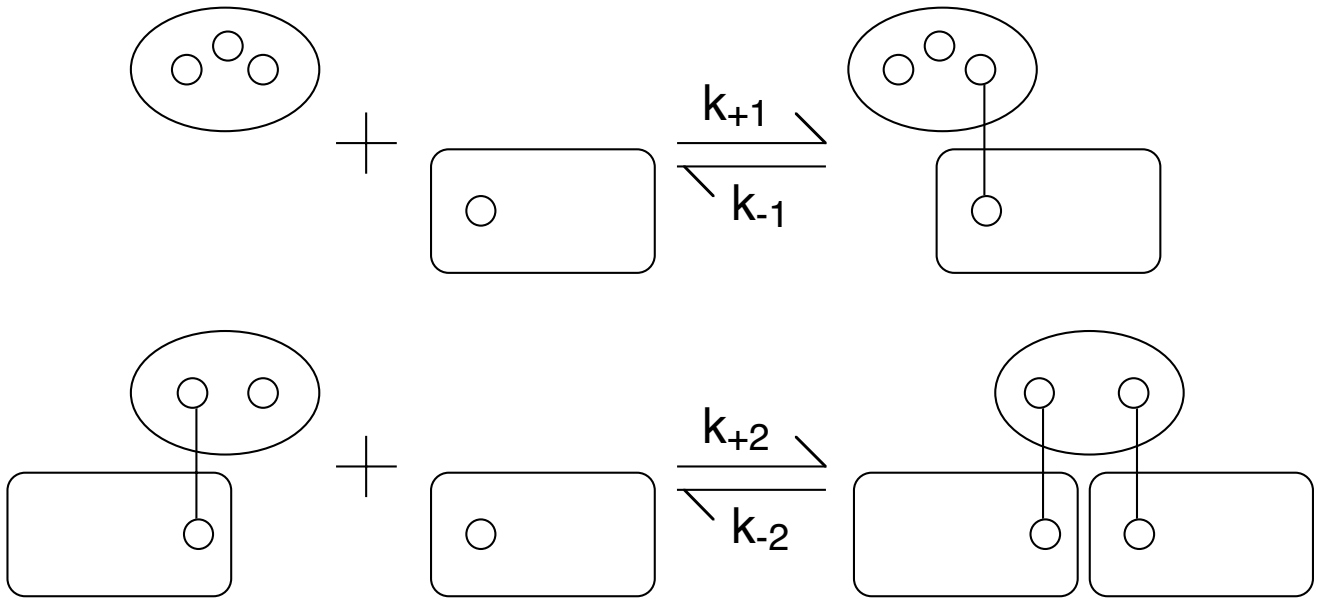
R.G. Posner (TGen)

A model specification

Molecules



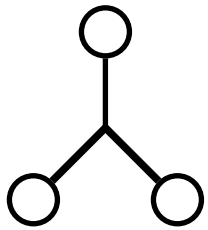
Interactions (reaction rules)



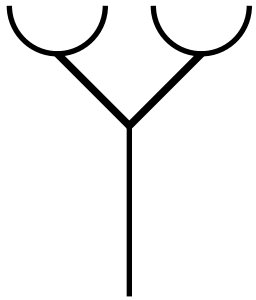
Rule processing

- Define seed species
- Determine if a pattern in a rule matches any species
- If so, apply the transformation defined in the rule
- Iteratively apply rules to new product species

Seed species

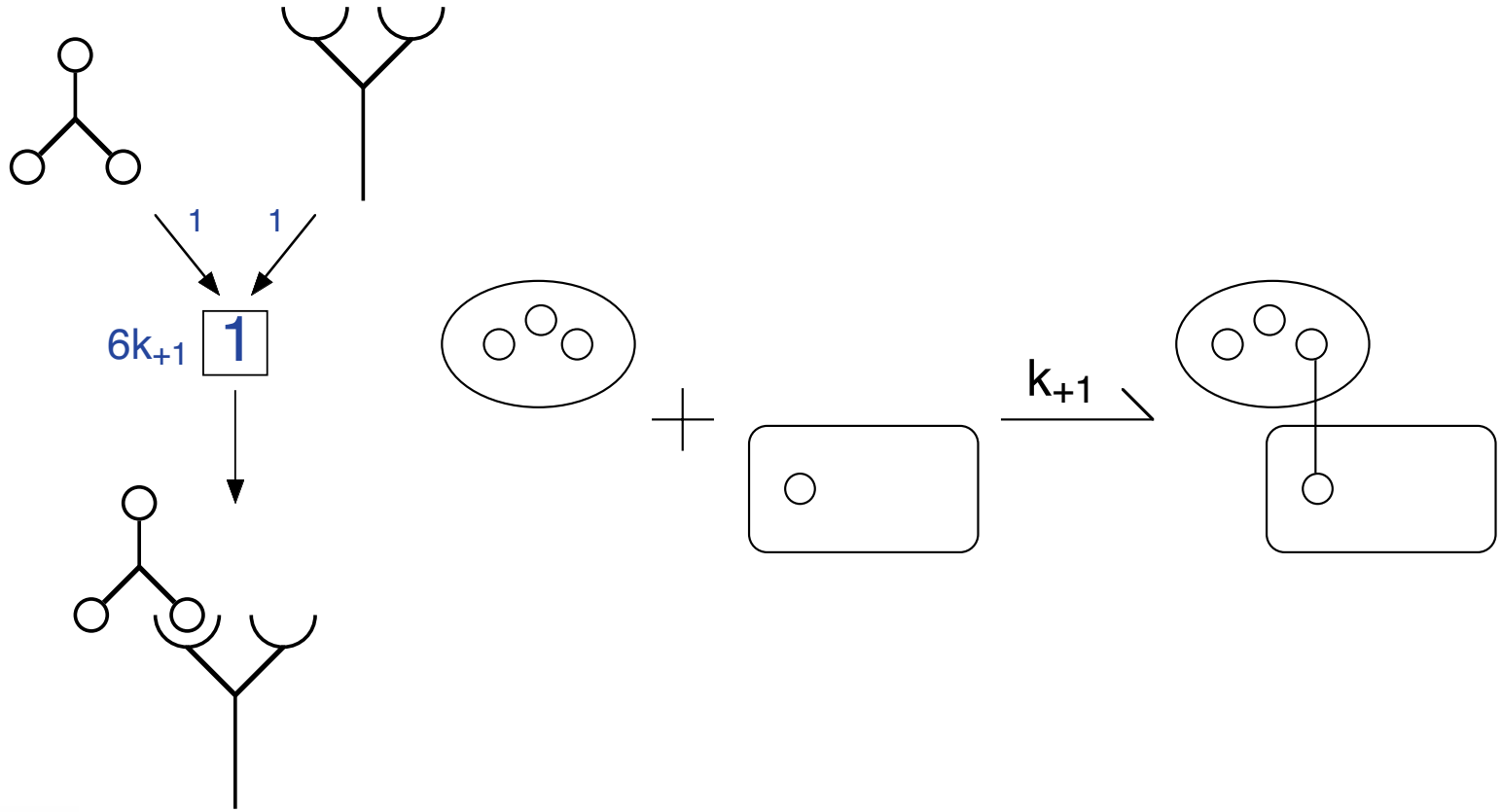


Ligand

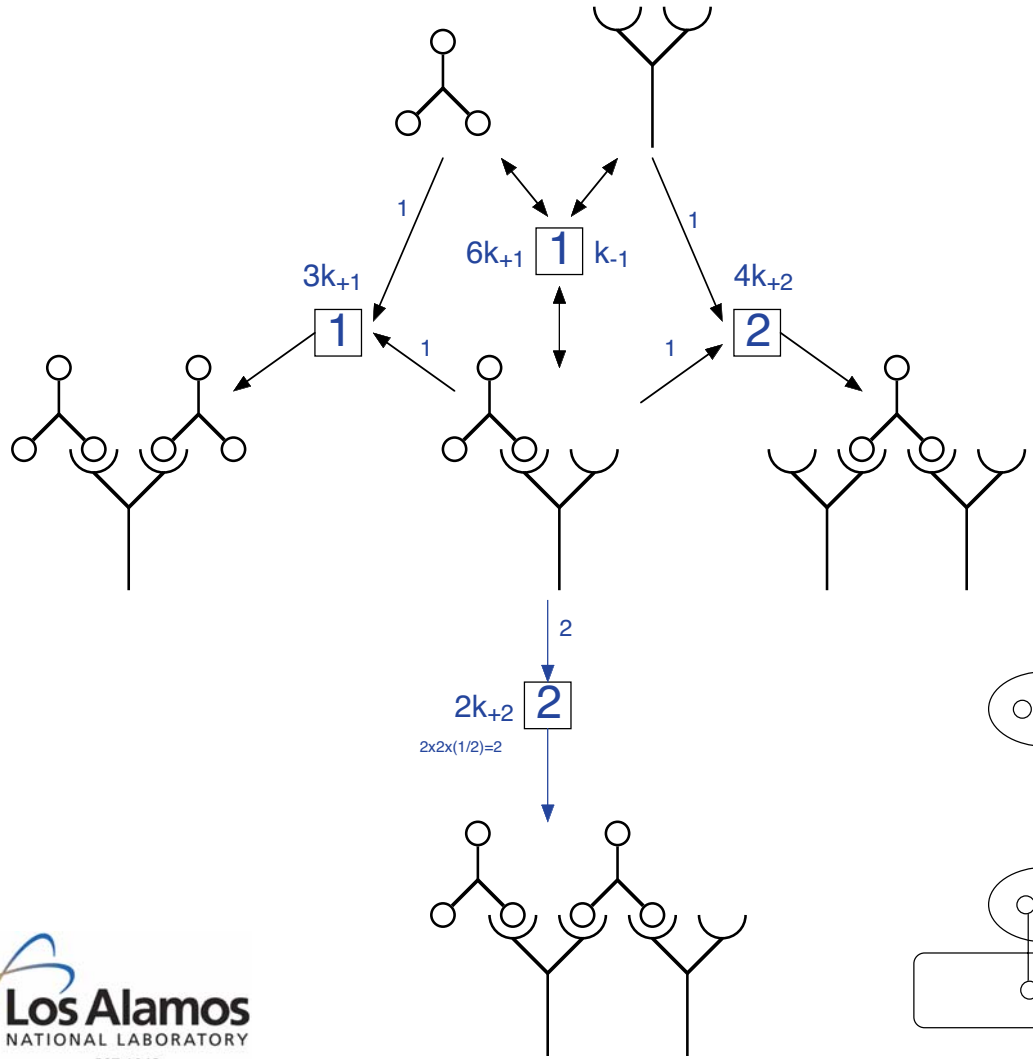


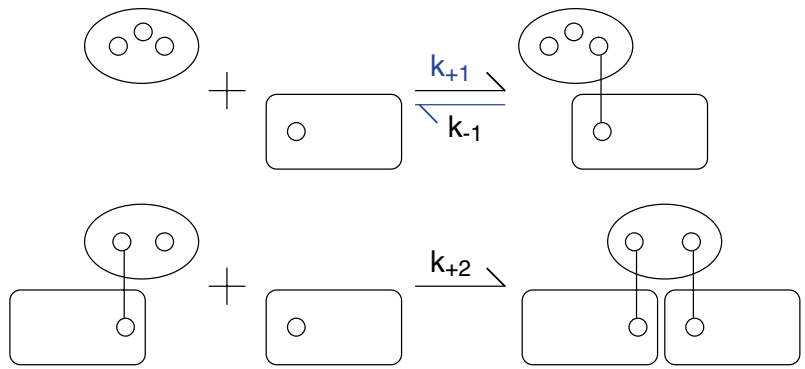
Receptor

After first round of rule application

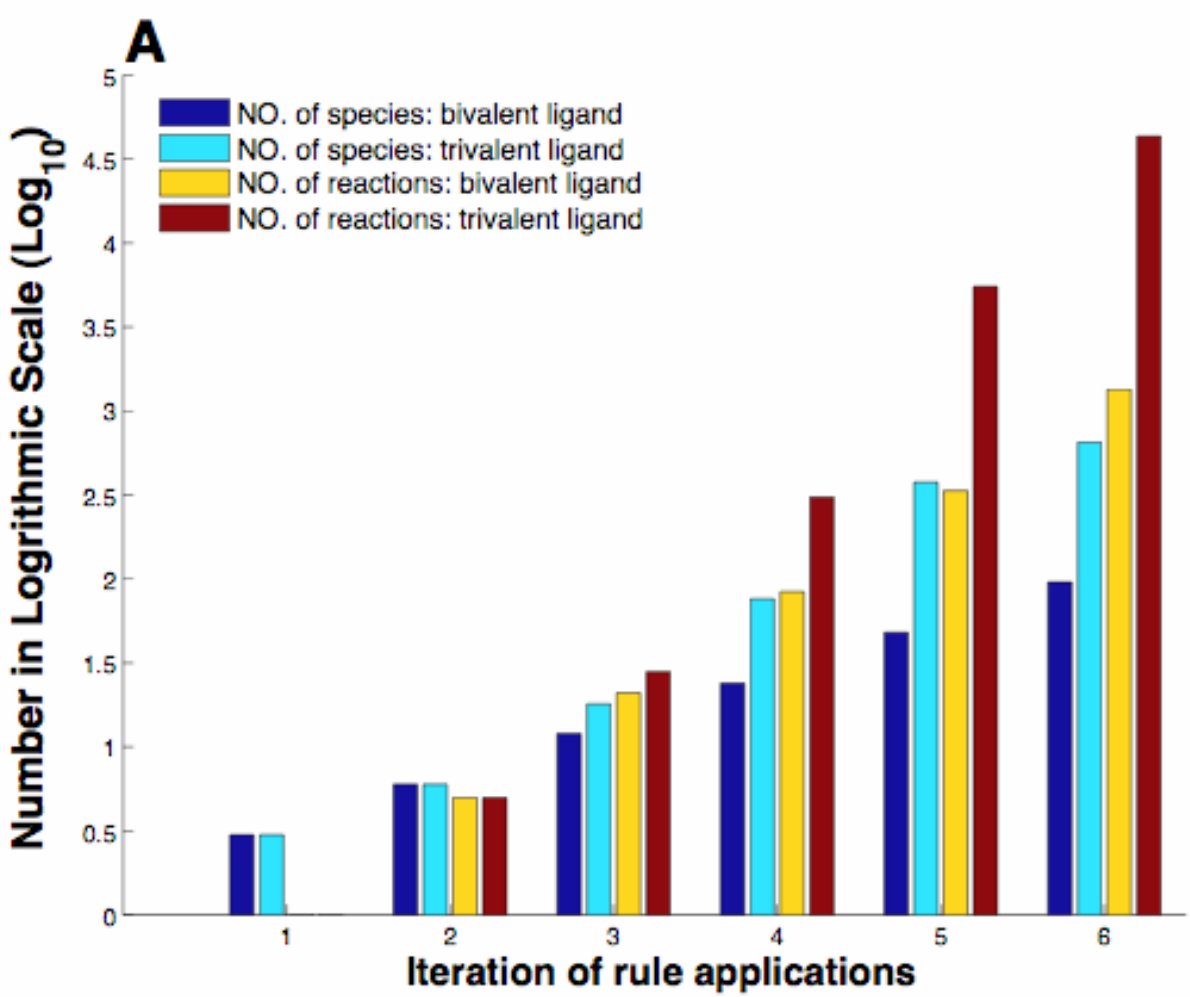


After the second round of rule application





Two rules generate a vast number of chemical species and reactions

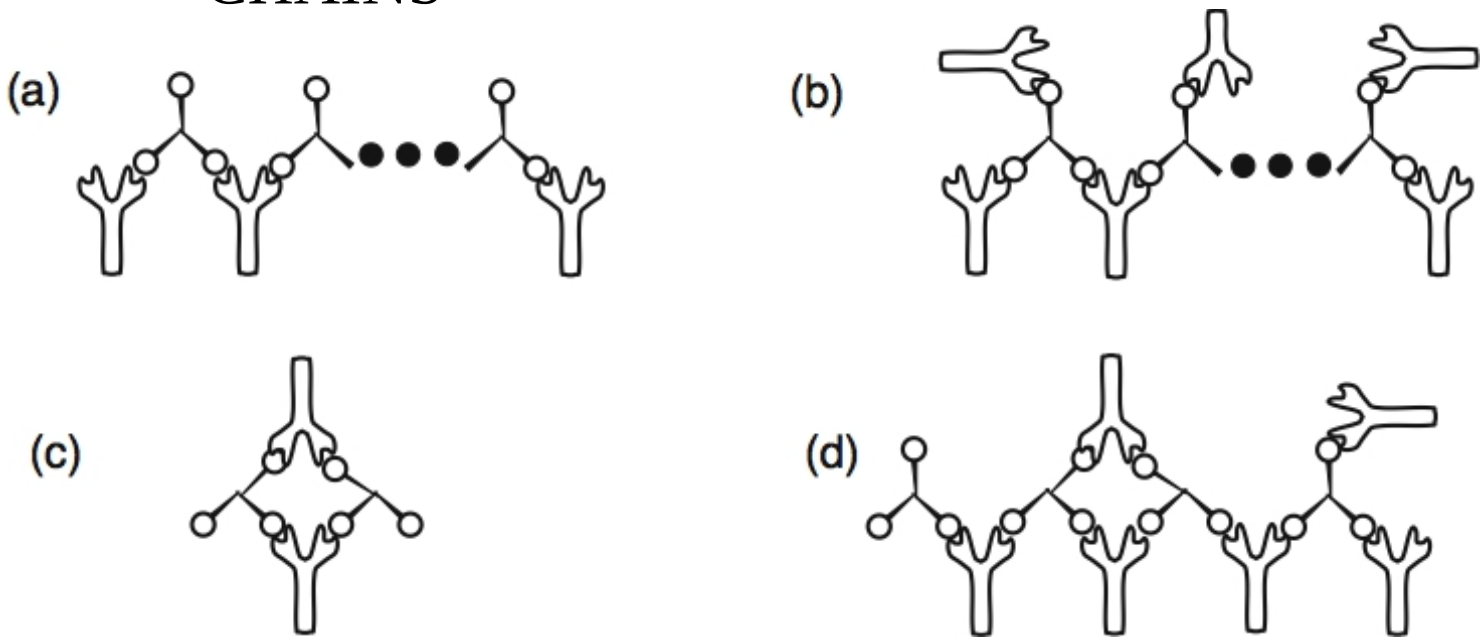


Los Alamos
NATIONAL LABORATORY
EST.1943

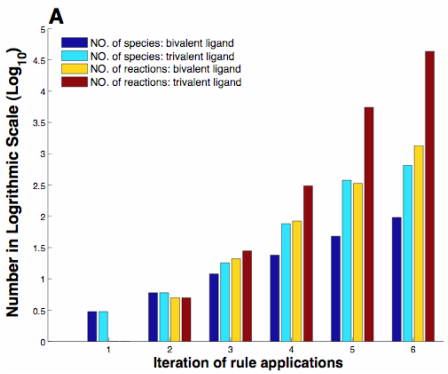
Diverse ligand-receptor complexes are possible

CHAINS

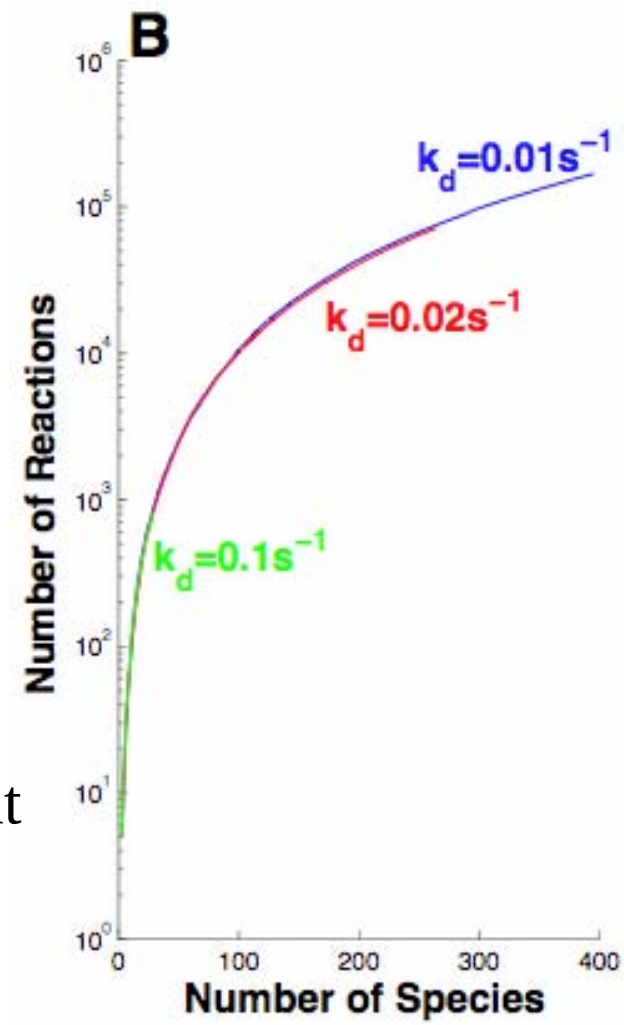
TREES

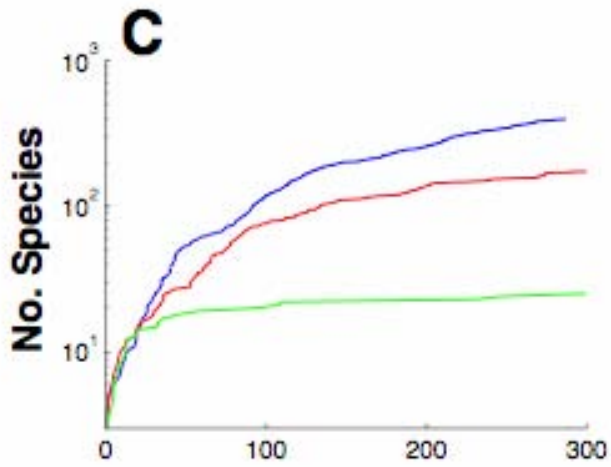


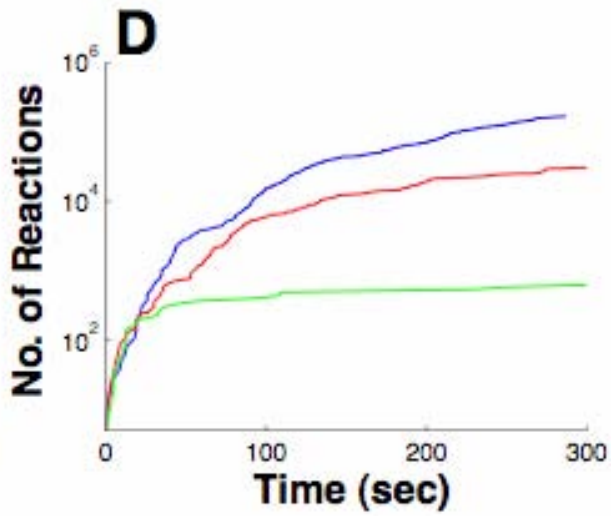
The network is large even when we consider only the populated portion



On-the-fly method of Lok and Brent (2005)







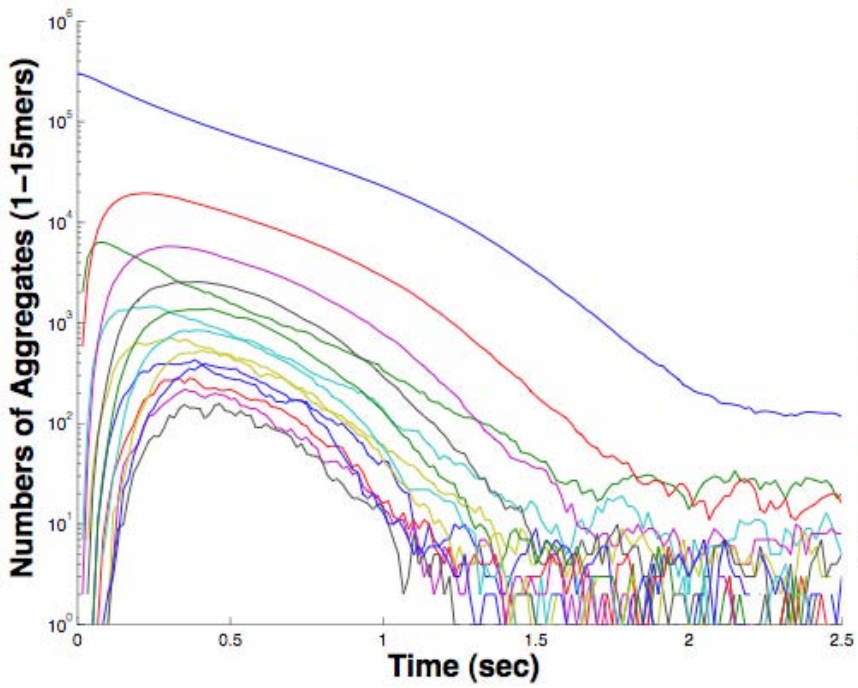
Los Alamos NATIONAL LABORATORY EST.1943

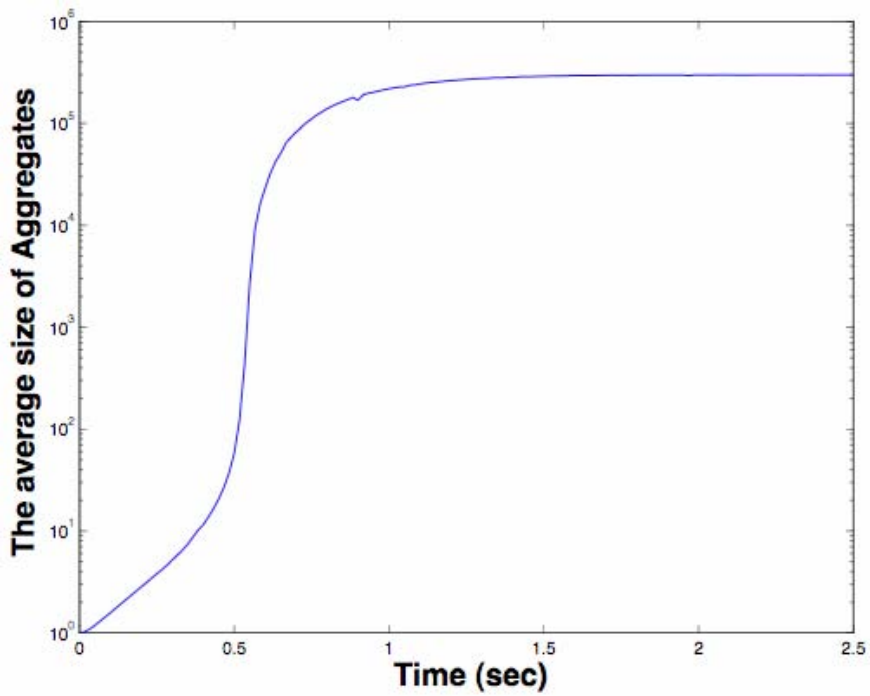
Rule-based Gillespie

- Set t=0, determine cumulative rate for each reaction **type**, set stop criteria
- Select next reaction time and reaction type (rather than an individual reaction) using direct method
- Select reactants using one or two additional random numbers and fire reaction
- Update reaction type rates
- Iterate

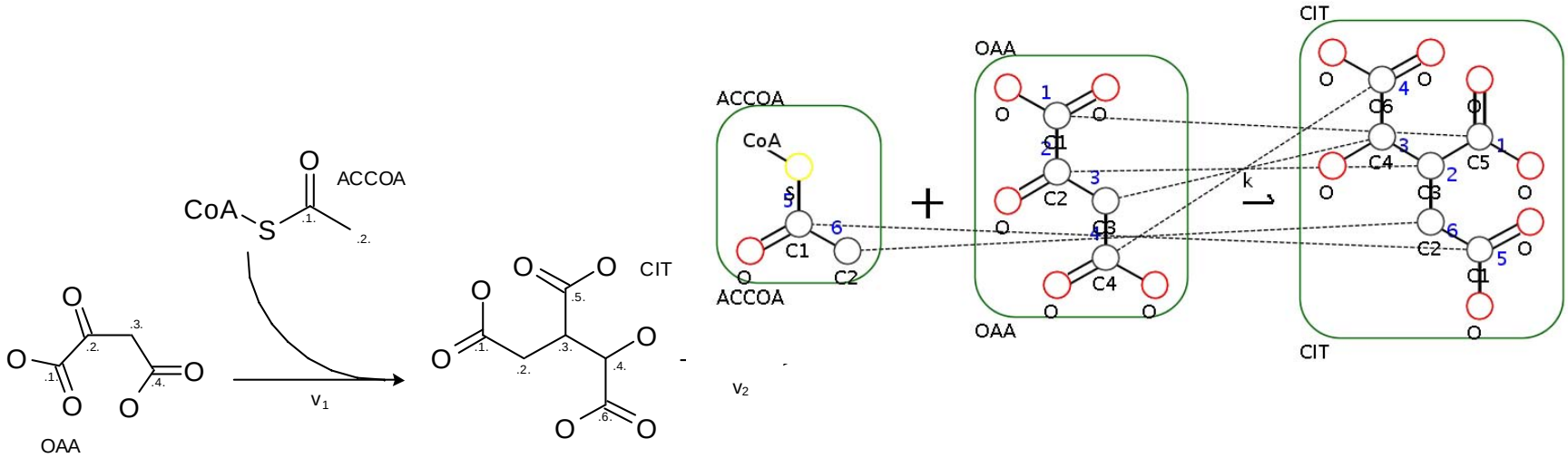
Kinetics of sol-gel phase transition

Four reaction types vs. many more reactions





Other applications of rule-based modeling



$$V_{CIT}\frac{d(CIT[100000])}{dt} = OAA[1000] * ACCOA[00] * v_1 - CIT[100000] * v_2$$

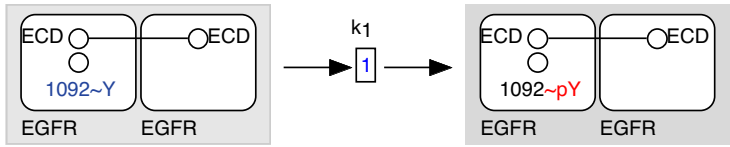
A simple metabolic network of 23 reactions involving
21 metabolites can involve up to 622 isotopomer fractions

Recently, we encoded reactions with carbon-fate maps in the form of BioNetGen rules for 5,153 metabolic reactions catalogued in the KEGG database

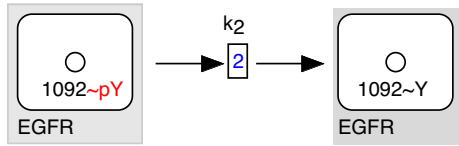
Outline

- Combinatorial complexity
- Drawbacks of conventional modeling approaches
- Rule-based modeling - methods
- An example of a rule-based model
- **Future directions**

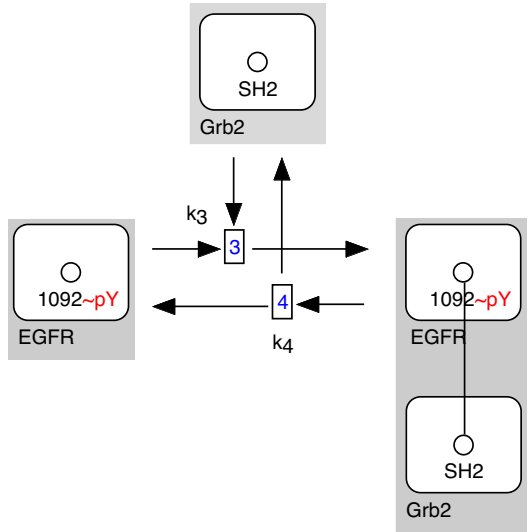
We can use rules to visualize site-specific details of protein interactions



EGFG(ECD!1,aa1092~Y).EGFR(ECD!1) -> EGFR(ECD!1,aa1092~pY).EGFR(ECD!1) k1



EGFR(aa1092~pY) -> EGFR(aa1092~Y) k2



EGFR(aa1092~pY) + Grb2(SH2) <-> EGFR(aa1092~pY!1).Grb2(SH2!1) k3,k4

Other visualization styles are possible and may be useful as well

At the same time, we can compute with rules

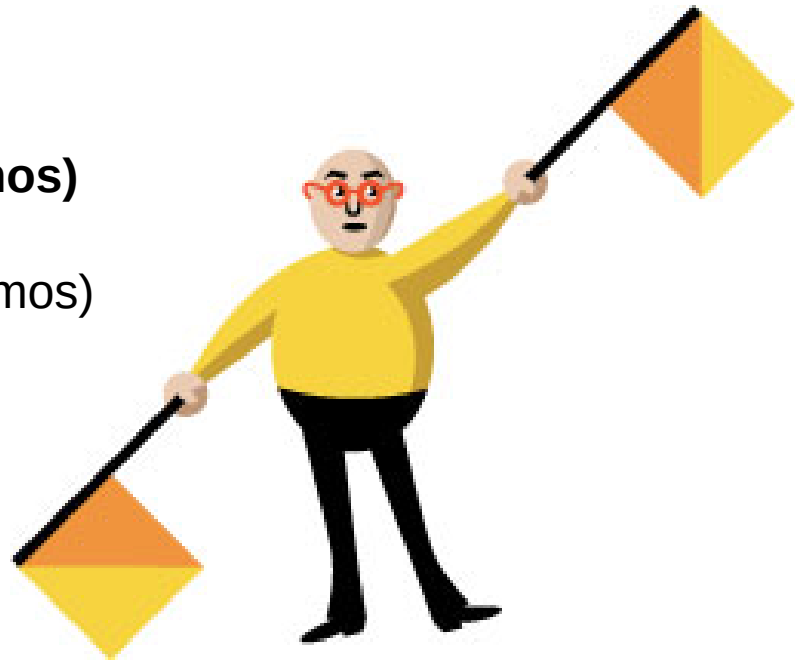
Is it time for high-throughput, extensible, community-driven modeling?

- Can we translate available information about proteins and protein interactions into model elements?
- Can we build models incrementally?
- How to establish conventions for annotating rule-based models?
- How best to collaborate on development of large-scale models?

cellsignaling.lanl.gov

Jin Yang (postdoc)
Fangping Mu (postdoc)

Michael L. Blinov (UConn)
James R. Faeder (Los Alamos)
Richard G. Posner (TGen)
Chang-Shung Tung (Los Alamos)



Funding
NIH
DOE

Los Alamos NATIONAL LABORATORY EST.1943

Los Alamos NATIONAL LABORATORY EST.1943