Modeling Cell Signaling

Bill Hlavacek

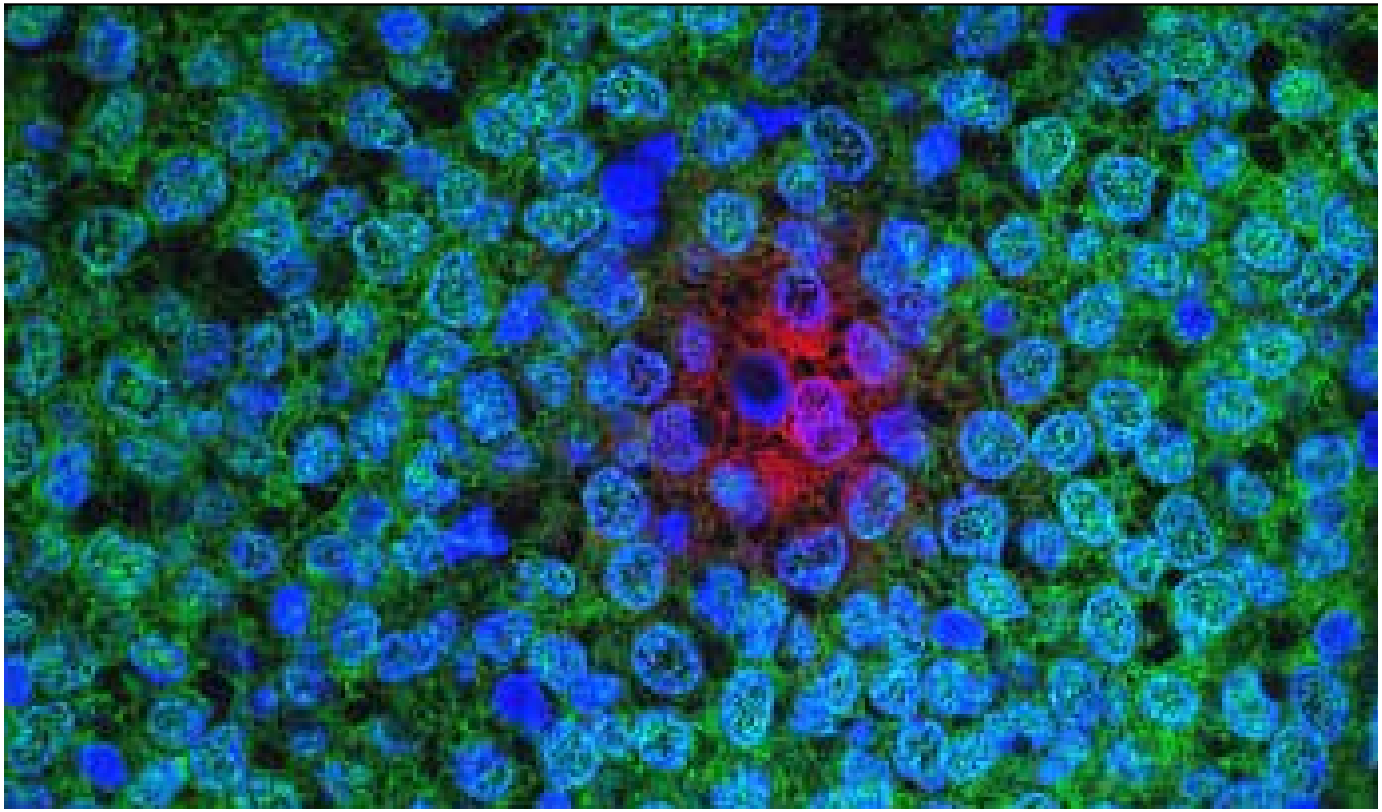
Theoretical Biology & Biophysics Group

Theoretical Division

Brian Munsky, William S. Hlavacek, and Lev S. Tsimring, editors

QUANTITATIVE BIOLOGY

Theory, Computational Methods, and Models

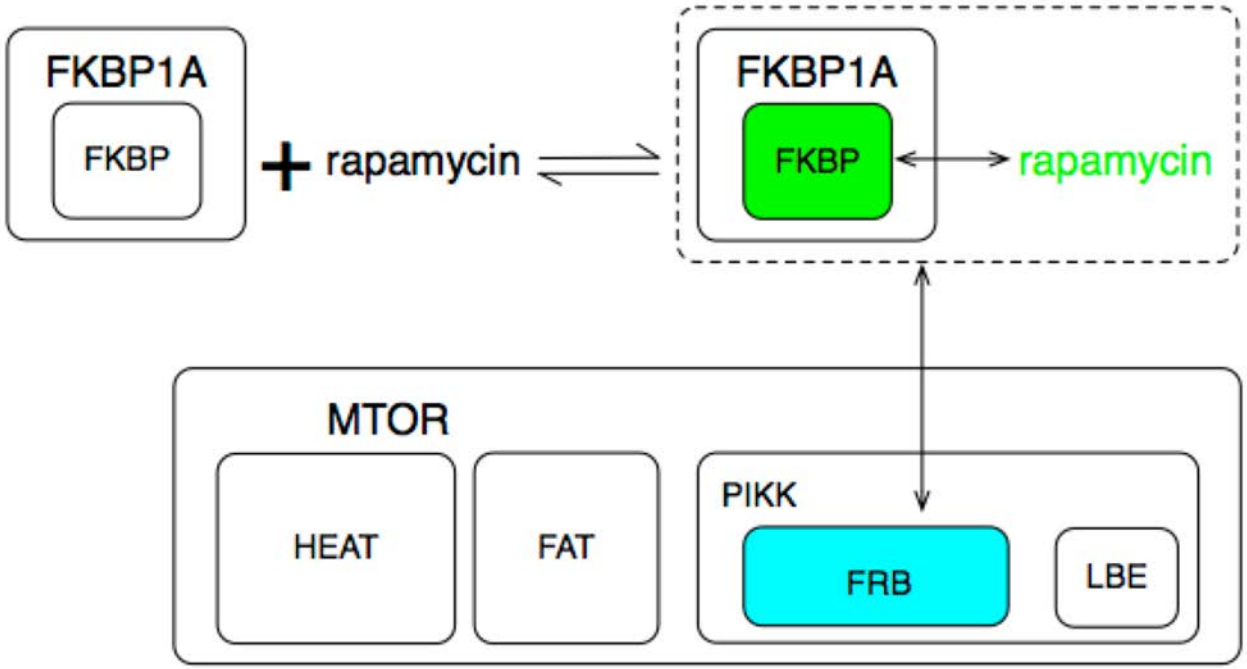


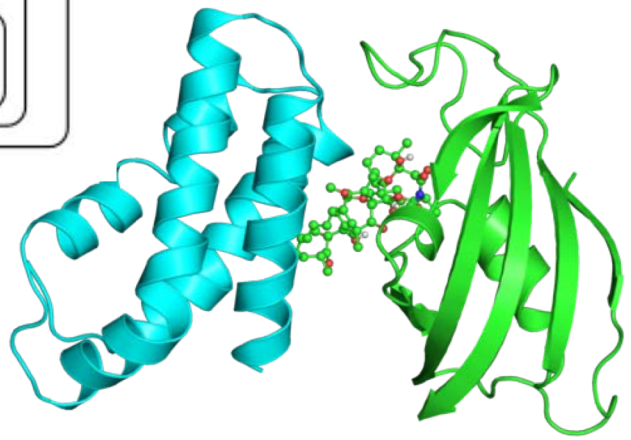
Value added by modeling of cellular regulatory systems

- **We can use models to organize and evaluate information**
  - To think with greater rigor and precision
  - To discover knowledge gaps
  - To identify key quantitative factors that affect system behavior
  - To summarize observations and preserve knowledge
- **We can analyze models to obtain insights and generate hypotheses**
  - To elucidate general design principles
  - To explain counterintuitive behavior
  - To enhance experimental efforts (e.g., through experimental design)
  - To guide interventions

Influences within the AMPK-MTORC1-ULK1 network

Illustration of details involved in tracking site dynamics

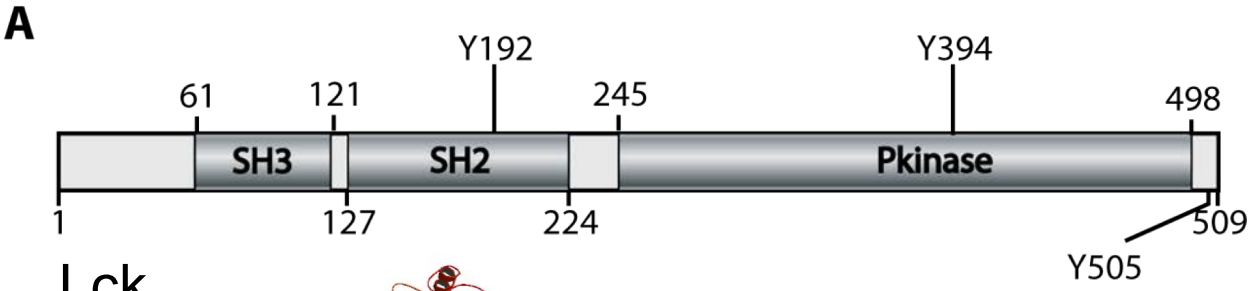




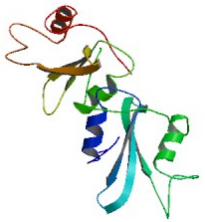
Outline

1. **Features of signaling proteins**
2. Combinatorial complexity: the key problem solved by rule-based modeling
3. Basic concepts of rule-based representation of biomolecular interactions
4. Simulation methods for rule-based models (indirect and direct)
5. Exercises (computer lab)

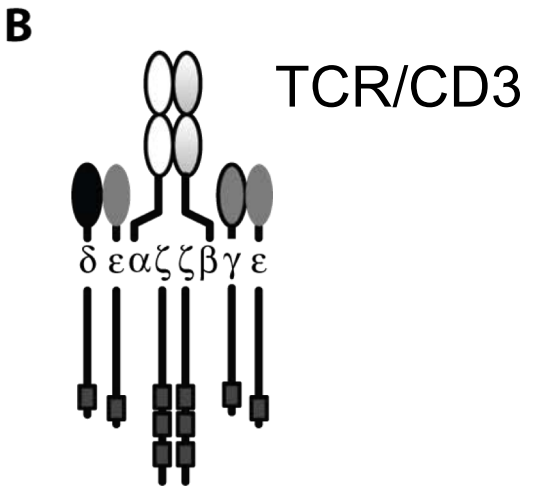
A signaling protein is typically composed of multiple components (subunits, domains, and/or linear motifs) that mediate interactions with other proteins



Lck



Lck-SH2 (1bhh)

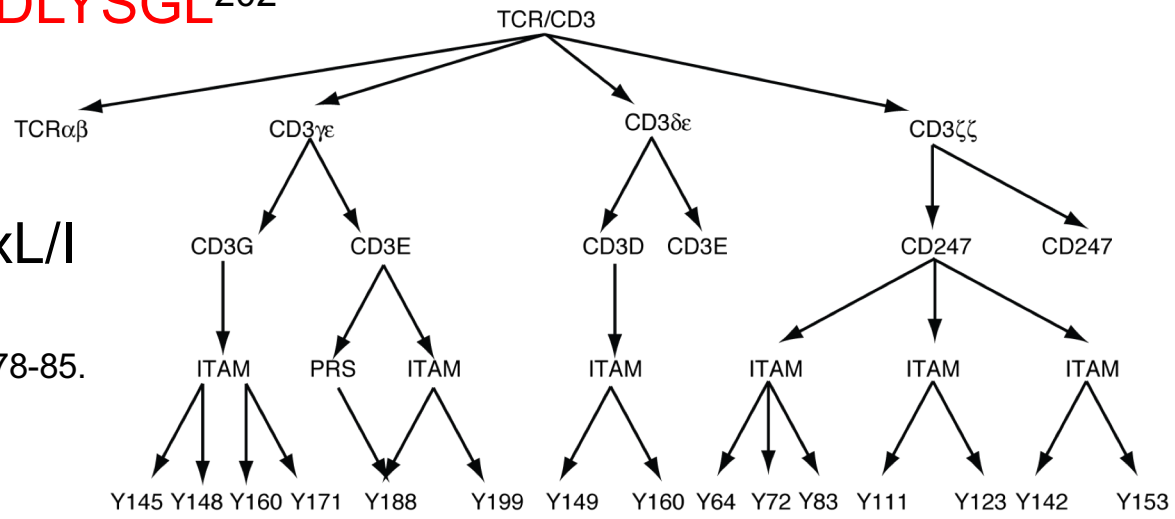


CD3E: ^{184}PNPDYEPIRKGQRDLYSGL^{202}

PRS: PxxDY

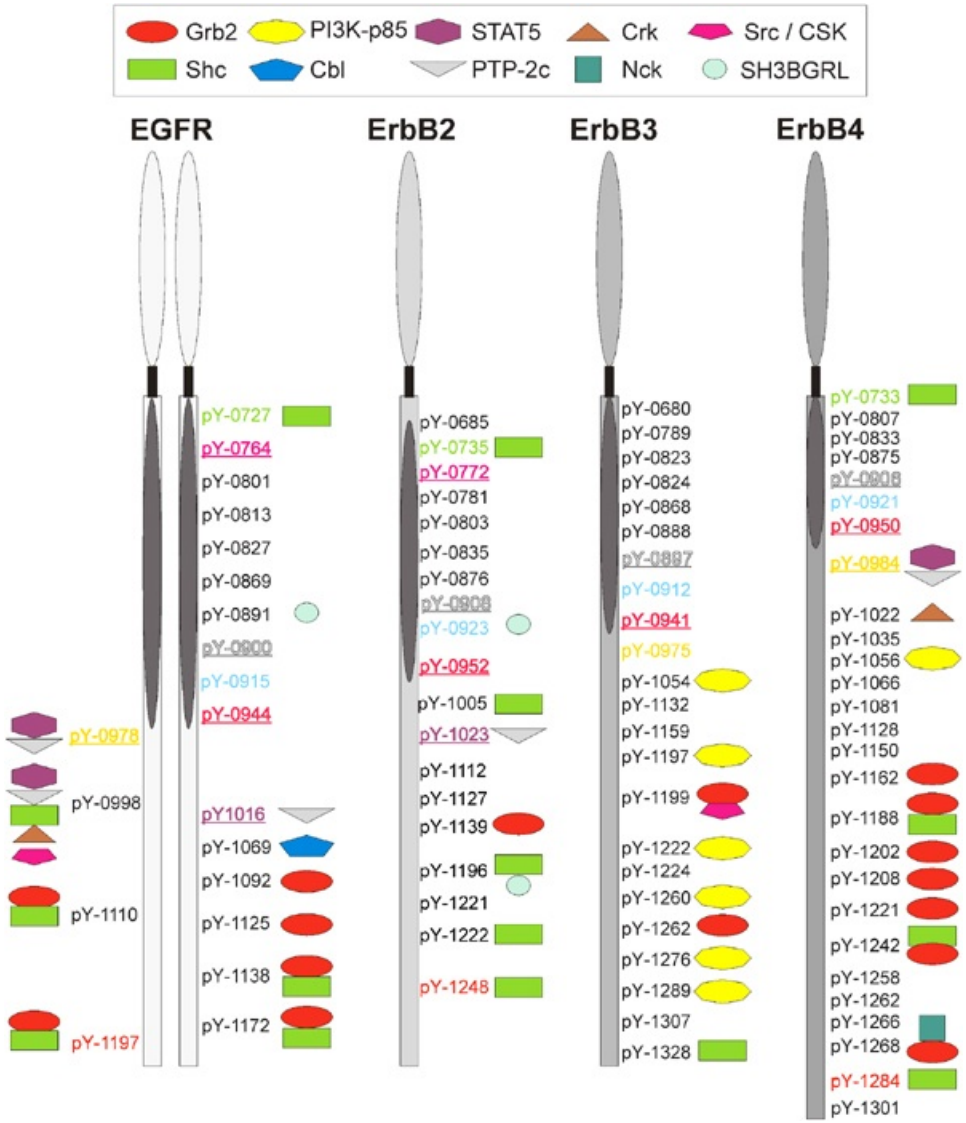
ITAM: YxxL/I(x_{6-8})YxxL/I

Kesti T et al. (2007) J. Immunol. 179:878-85.



Los Alamos NATIONAL LABORATORY EST. 1943

Domain-motif interactions are often controlled by post-translational modifications



Schulze WX et al. (2005)
Mol. Syst. Biol.

Los Alamos
NATIONAL LABORATORY
EST. 1943

Outline

1. Features of signaling proteins
2. **Combinatorial complexity: the key problem solved by rule-based modeling**
3. Basic concepts of rule-based representation of biomolecular interactions
4. Simulation methods for rule-based models (indirect and direct)
5. Exercises (computer lab)

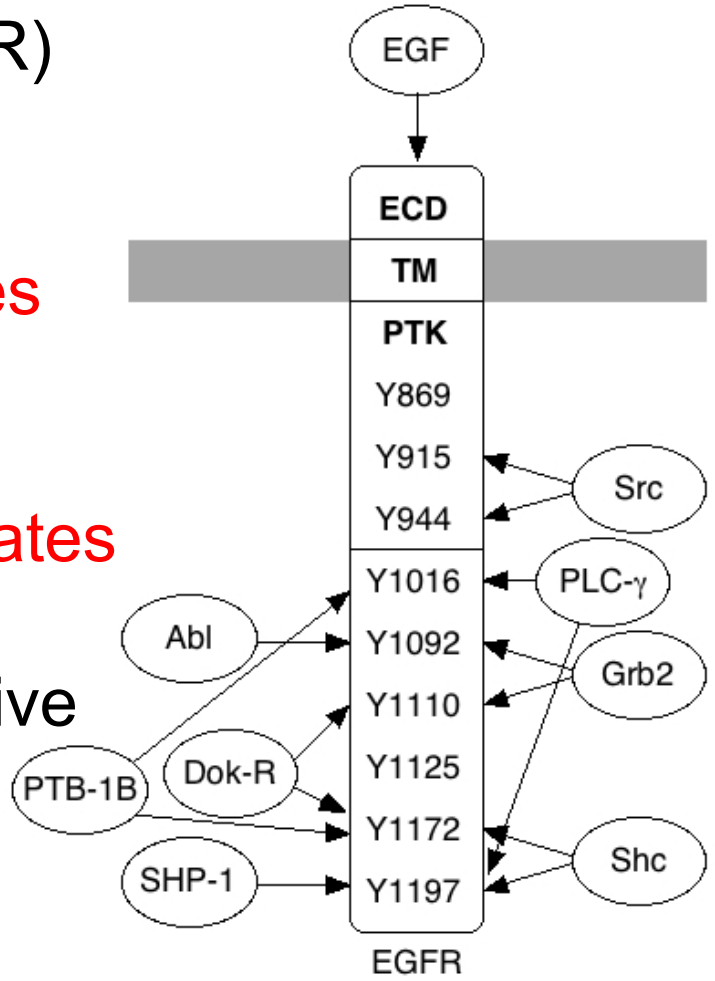
Complexity arises from post-translational modifications

Epidermal growth factor receptor (EGFR)

9 sites => 2^9=512 phosphorylation states

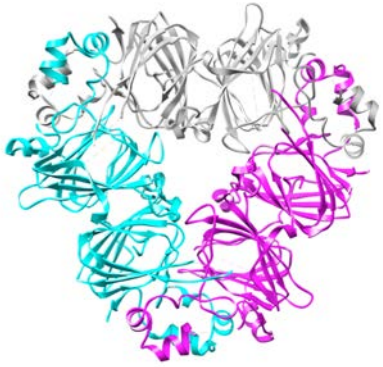
Each site has ≥ 1 binding partner
=> more than 3^9=19,683 total states

EGFR must form *dimers* to become active
=> more than 1.9×10^8 states

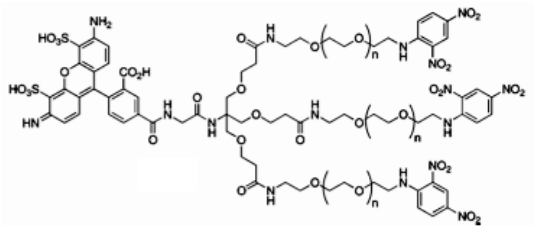


Complexity arises from oligomerization/aggregation

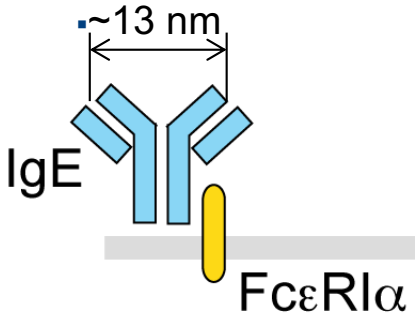
Mahajan et al. (2014) *ACS Chem Biol* **9**: 1508-1519.



DF3

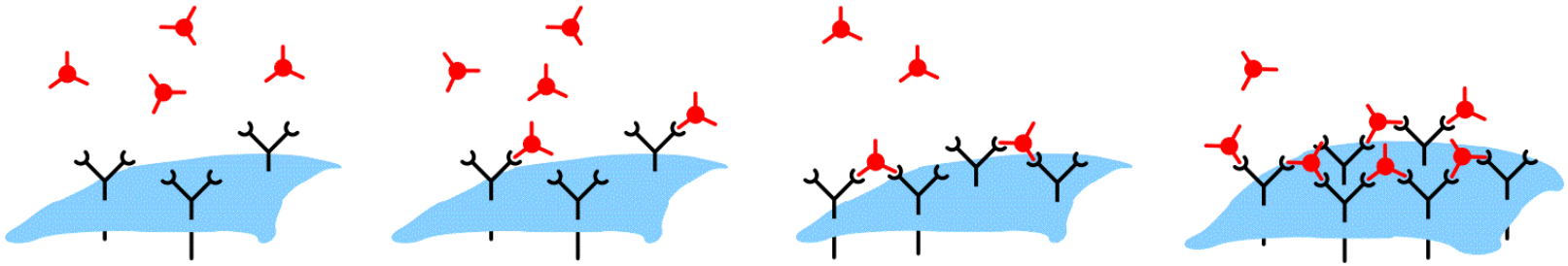


Compound 6a



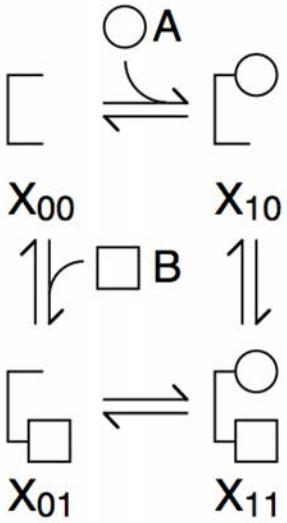
Ara h 1 (major peanut allergen), PDB 3S7E

Posner et al. (2007) *Org Lett* **9**: 3551



The textbook approach

Conventional representation of a biochemical reaction network



$X_{00} + A \rightleftharpoons X_{10}$

$X_{00} + B \rightleftharpoons X_{01}$

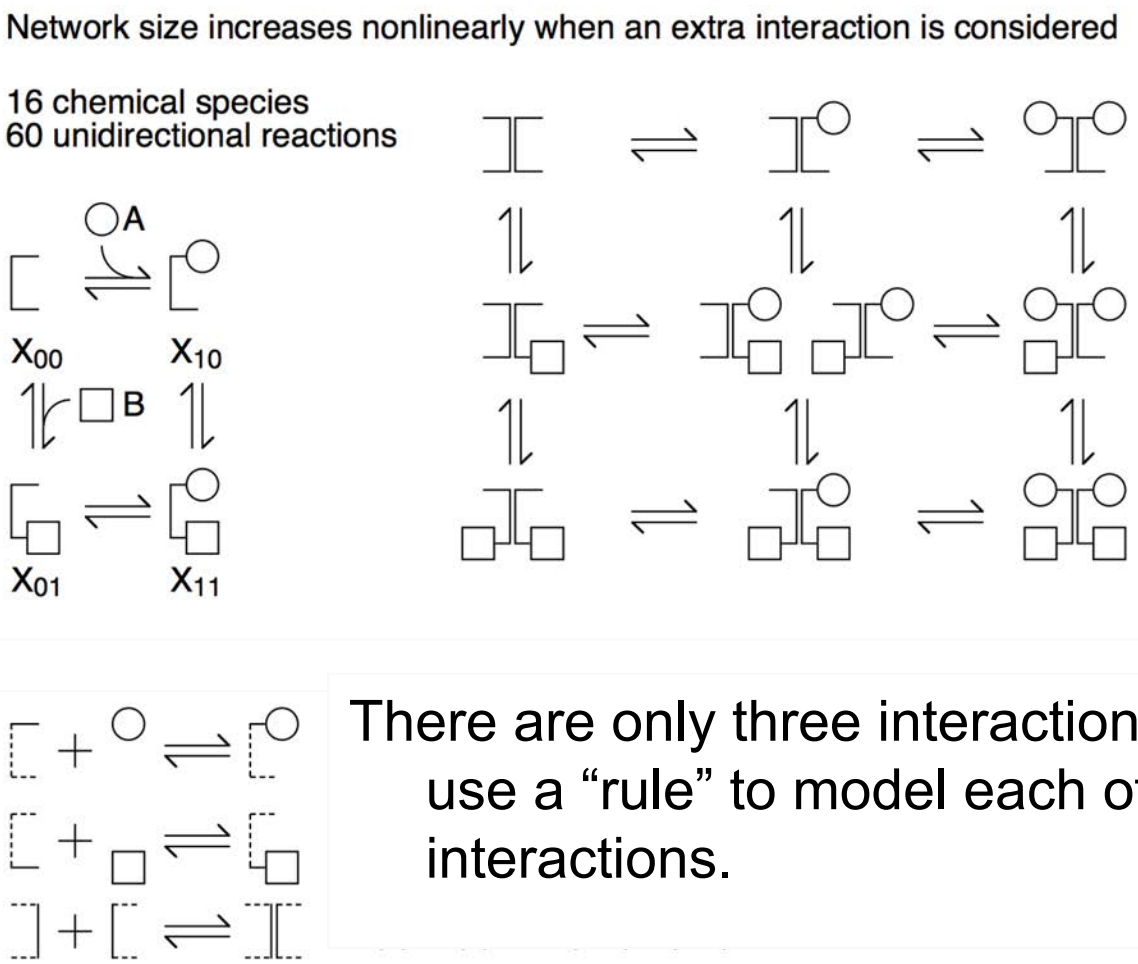
$X_{10} + B \rightleftharpoons X_{11}$

$X_{01} + A \rightleftharpoons X_{11}$

6 chemical species
8 unidirectional reactions

$$d[X_{00}]/dt = -k_{+1}[A][X_{00}] + k_{-1}[X_{10}] - k_{+2}[B][X_{00}] + k_{-2}[X_{01}]$$

Network (model) size tends to grow nonlinearly (exponentially) with the number of molecular interactions



There are only three interactions. We can use a "rule" to model each of these interactions.

Science's STKE re6 (2006)

Rule-based modeling solves the problem of combinatorial complexity

- **Inside a Chemical Plant**
  - Large numbers of molecules…
  - …of a few types
  - Conventional modeling works fine (a good idea since Harcourt and Esson, 1865)
- **Inside a Cell**
  - Possibly small numbers of molecules…
  - …of many possible types
  - Rule-based modeling is designed to deal with this situation (new)

Outline

1. Features of signaling proteins
2. Combinatorial complexity: the key problem solved by rule-based modeling
3. **Basic concepts of rule-based representation of biomolecular interactions**
4. Simulation methods for rule-based models (indirect and direct)
5. Exercises (computer lab)

Rule-based modeling: basic concepts

Graphs represent molecules/complexes, their component parts, and "internal states"

collections of same-colored vertices represent "molecule types"

vertices represent "sites"

vertex labels represent "states"

edges represent bonds

connnected molecule types represent complexes

Graph-rewriting rules represent molecular interactions

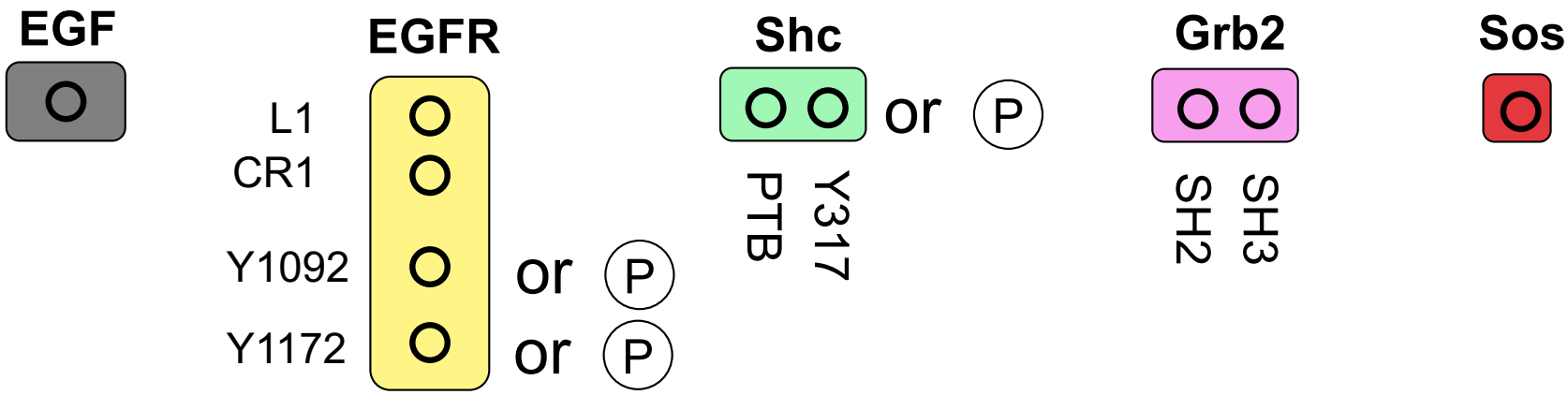
addition of an edge to represent bonding

removal of an edge to represent dissociation

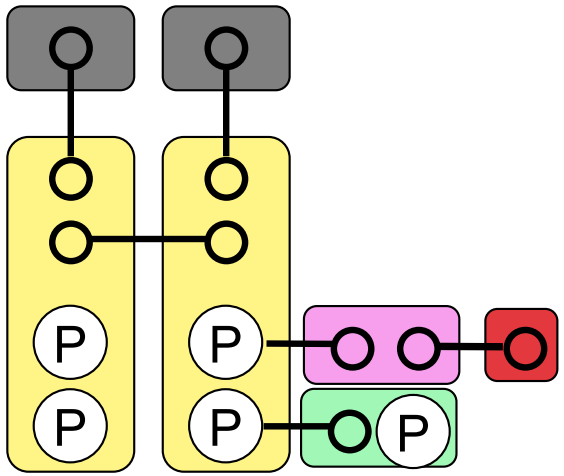
change of a vertex label to represent change of state

(e.g., change of conformation, location, or PTM status)

The site graphs of a model for EGFR signaling



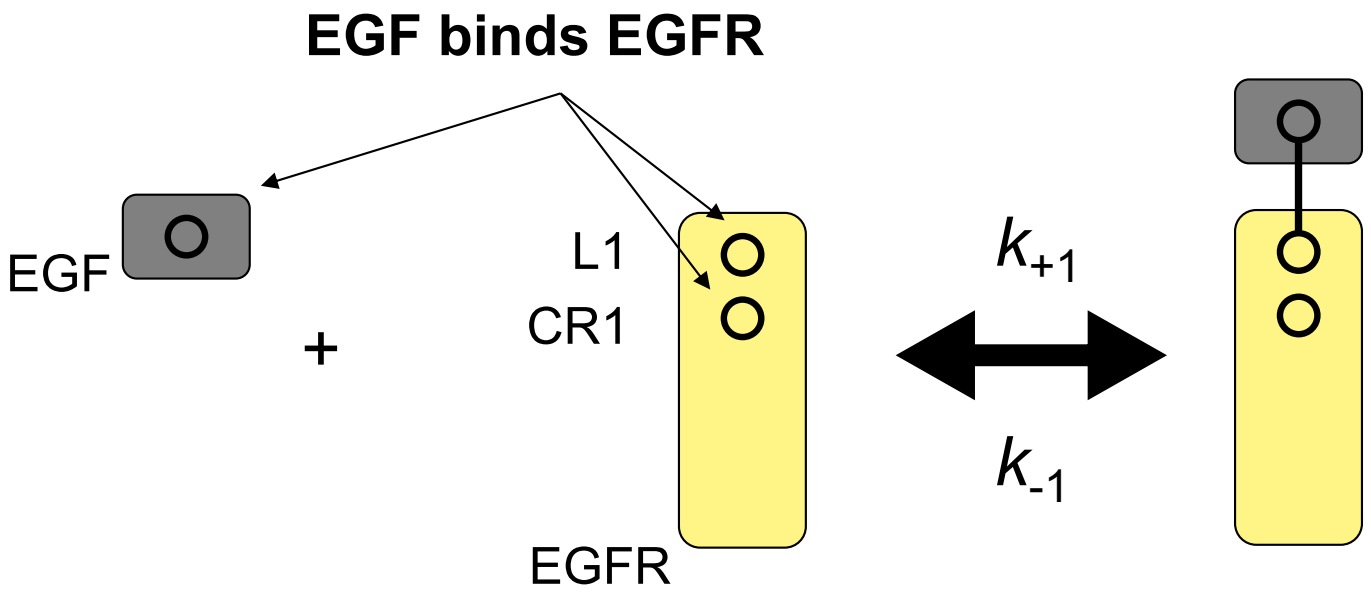
No need to introduce a unique name (e.g., X_{123} or ShP-RP-G-Sos) for each chemical species, as in conventional modeling



Los Alamos NATIONAL LABORATORY EST. 1943

Blinov ML et al. (2006) *BioSystems*

A rule for EGF-EGFR binding



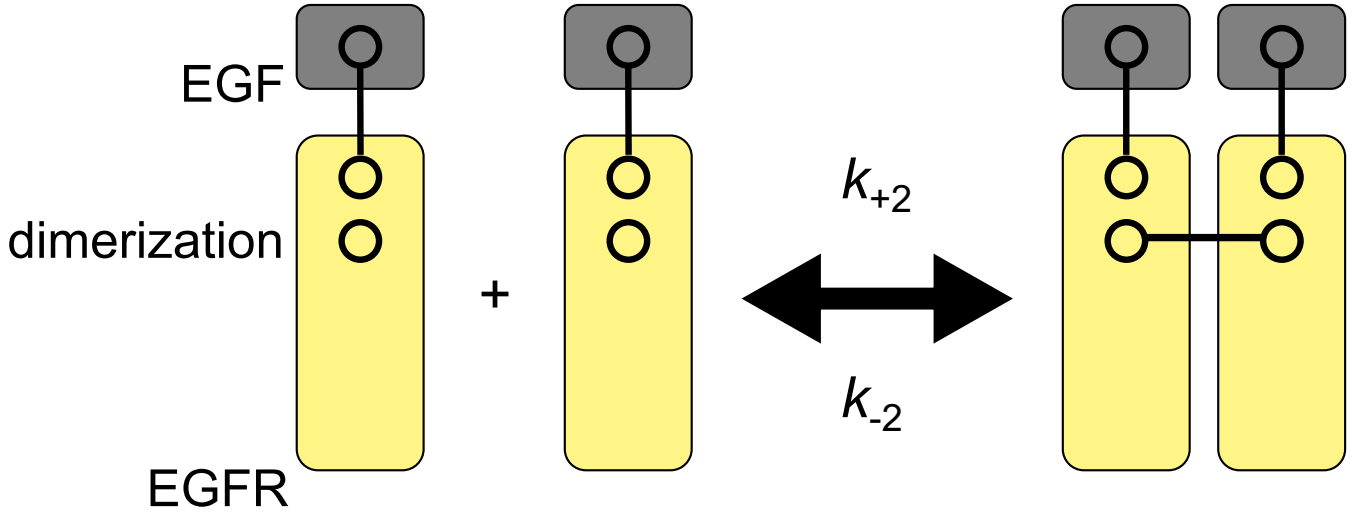

```
begin reaction rules

 EGF(R)+EGFR(L1,CR1)<->EGF(R!1).EGFR(L1!1,CR1)

end reaction rules
```

A rule for ligand-dependent EGFR dimerization

EGFR dimerizes (600 reactions are implied by this one rule)



No free lunch: According to this rule, dimers form and break up with the same fundamental rate constants regardless of the states of cytoplasmic domains, which is a simplification.

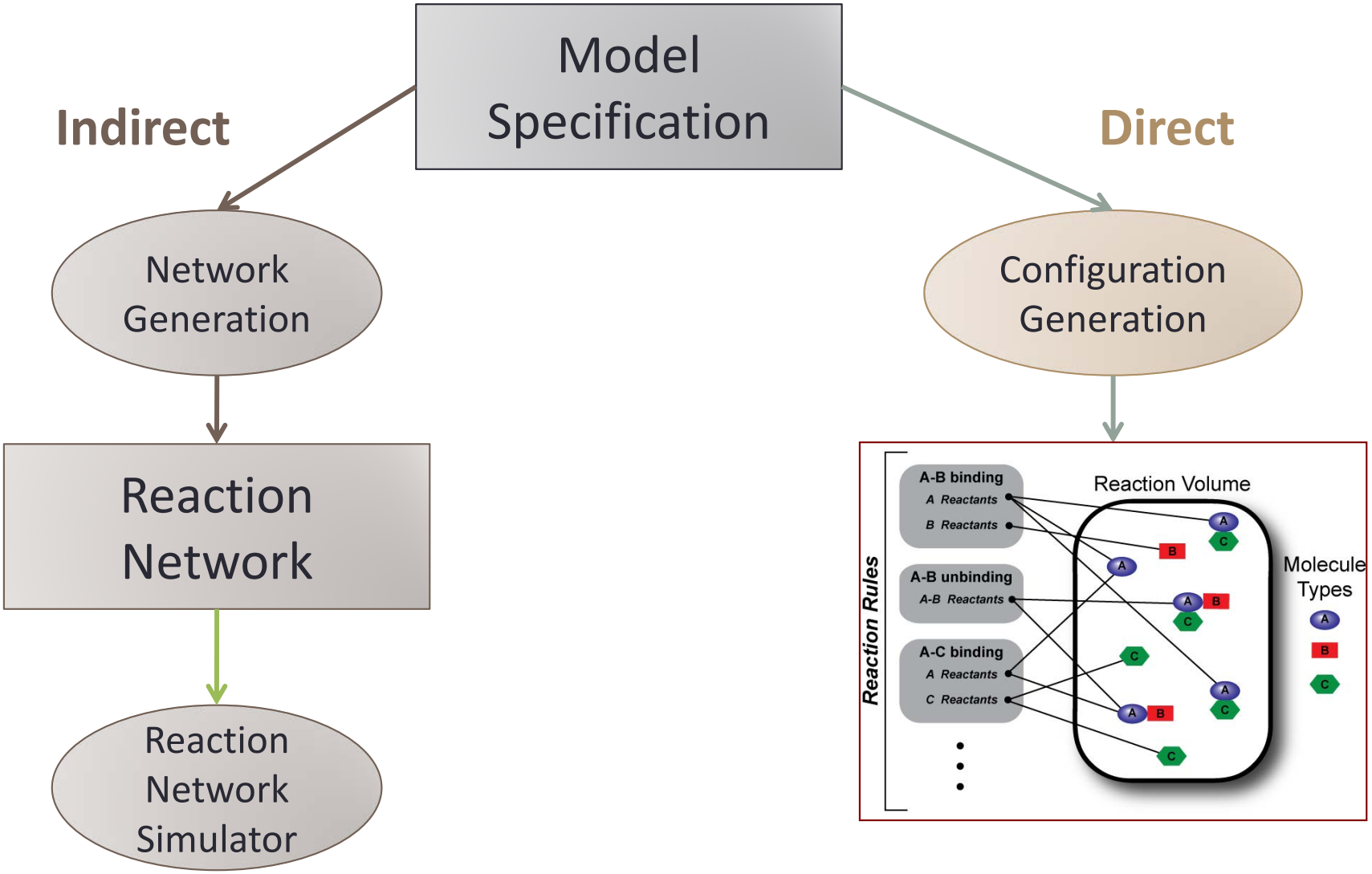
Outline

1. Features of signaling proteins
2. Combinatorial complexity: the key problem solved by rule-based modeling
3. Basic concepts of rule-based representation of biomolecular interactions
4. **Simulation methods for rule-based models (indirect and direct)**
5. Exercises (computer lab)

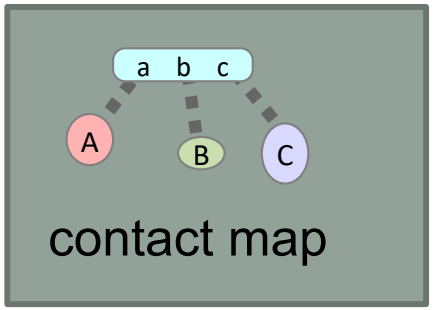
Many different traditional simulation techniques are compatible with RBM

1. Ordinary differential equations (ODEs) - **BioNetGen**
   - One equation per chemical species in the reaction network
   - Each reaction contributes a negative term to a reactant's equation and a positive term to a product's equation
2. Markov chains – BioNetGen + **NFsim**
   - Gillespie's method or stochastic simulation algorithm (SSA) or KMC
   - Each trajectory represents one sample from probability space of the chemical master equation (CME)
3. Partial differential equations (PDEs) - **VCell**
   - Species concentrations are resolved in space
4. Particle-based stochastic spatial simulations – **Smoldyn** + **MCell**
5. Force field- or potential-based calculations with excluded volume and orientation constraints (molecular dynamics) - **SRSim**

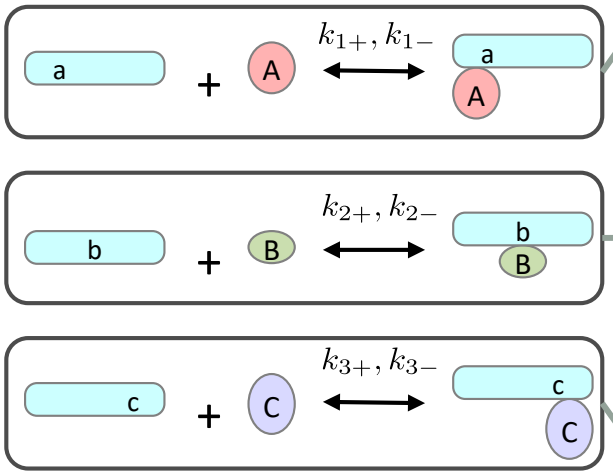
Two types of methods for simulating RBMs



Indirect Methods – Network Generation



reaction rules



seed species

S,A,B,C

S	+	A	$\overset{k_{1+},k_{1-}}{\rightleftharpoons}$	SA	1
SB	+	A	$\overset{k_{1+},k_{1-}}{\rightleftharpoons}$	SAB	2
SC	+	A	$\overset{k_{1+},k_{1-}}{\rightleftharpoons}$	SAC	3
SBC	+	A	$\overset{k_{1+},k_{1-}}{\rightleftharpoons}$	SABC	4
S	+	B	$\overset{k_{2+},k_{2-}}{\rightleftharpoons}$	SB	1
SA	+	B	$\overset{k_{2+},k_{2-}}{\rightleftharpoons}$	SAB	2
SC	+	B	$\overset{k_{2+},k_{2-}}{\rightleftharpoons}$	SBC	3
SAC	+	B	$\overset{k_{2+},k_{2-}}{\rightleftharpoons}$	SABC	4
S	+	C	$\overset{k_{3+},k_{3-}}{\rightleftharpoons}$	SC	1
SA	+	C	$\overset{k_{3+},k_{3-}}{\rightleftharpoons}$	SAC	2
SB	+	C	$\overset{k_{3+},k_{3-}}{\rightleftharpoons}$	SBC	3
SAB	+	C	$\overset{k_{3+},k_{3-}}{\rightleftharpoons}$	SABC	4

reactions

S, A, B, C,	1
SA, SB, SC,	2
SAB, SAC, SBC,	3
SABC	4

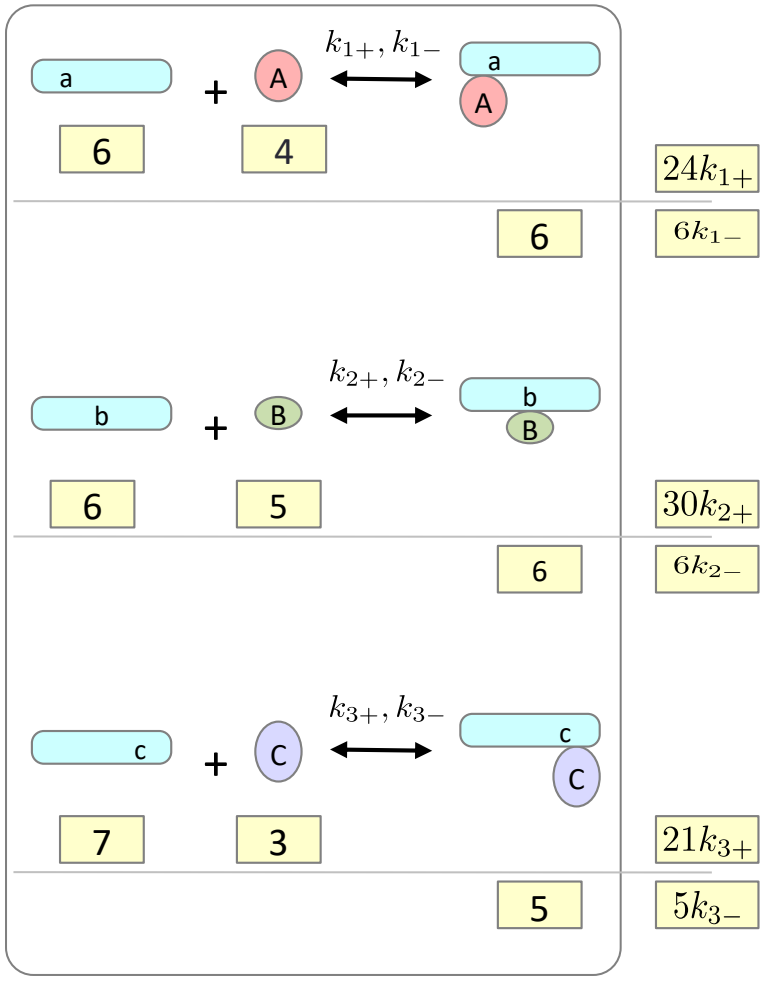
species

4 molecule types
3 rules

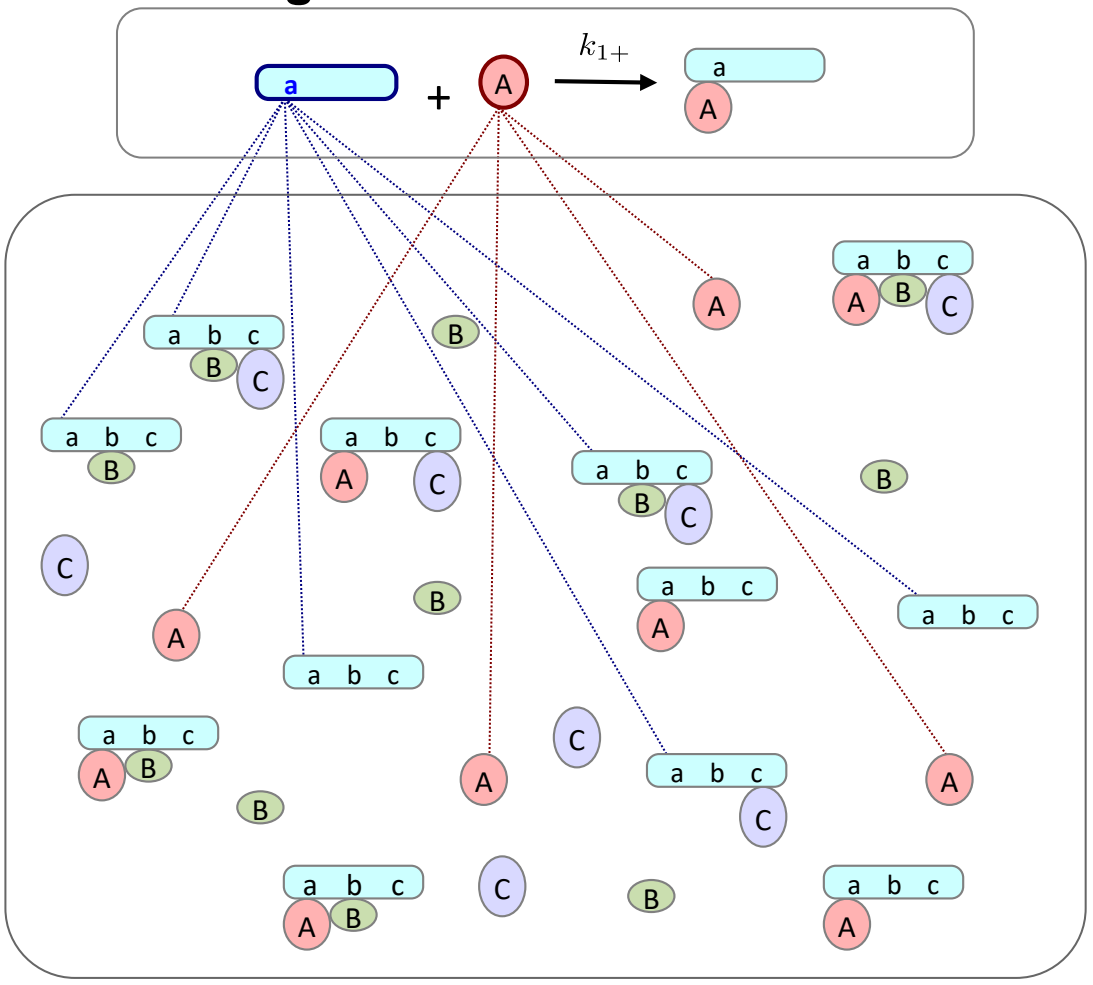
11 species
12 reactions

Direct Methods – rules generate reaction events and system configurations

reaction rules



event generation

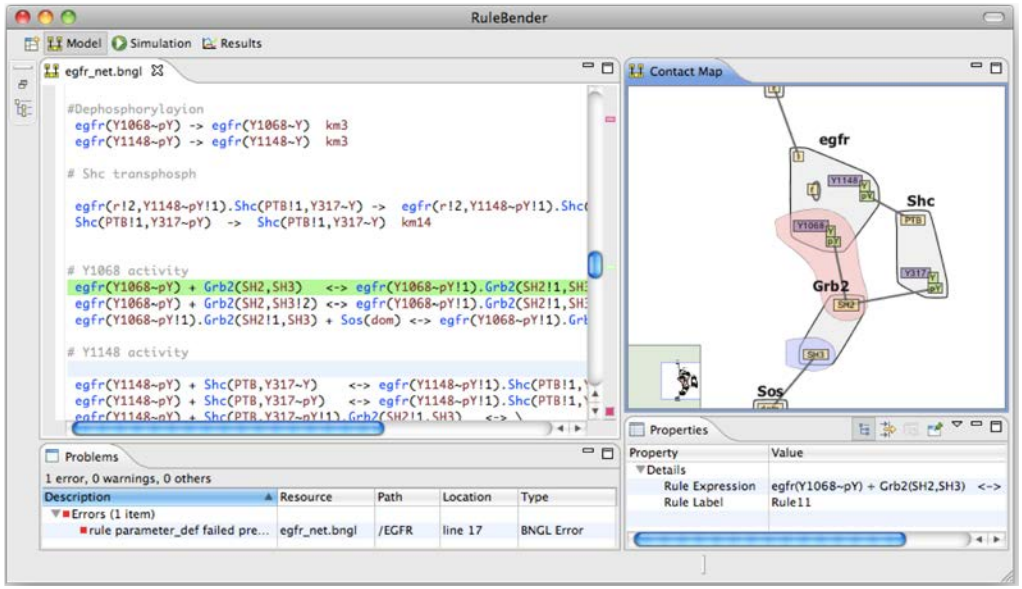


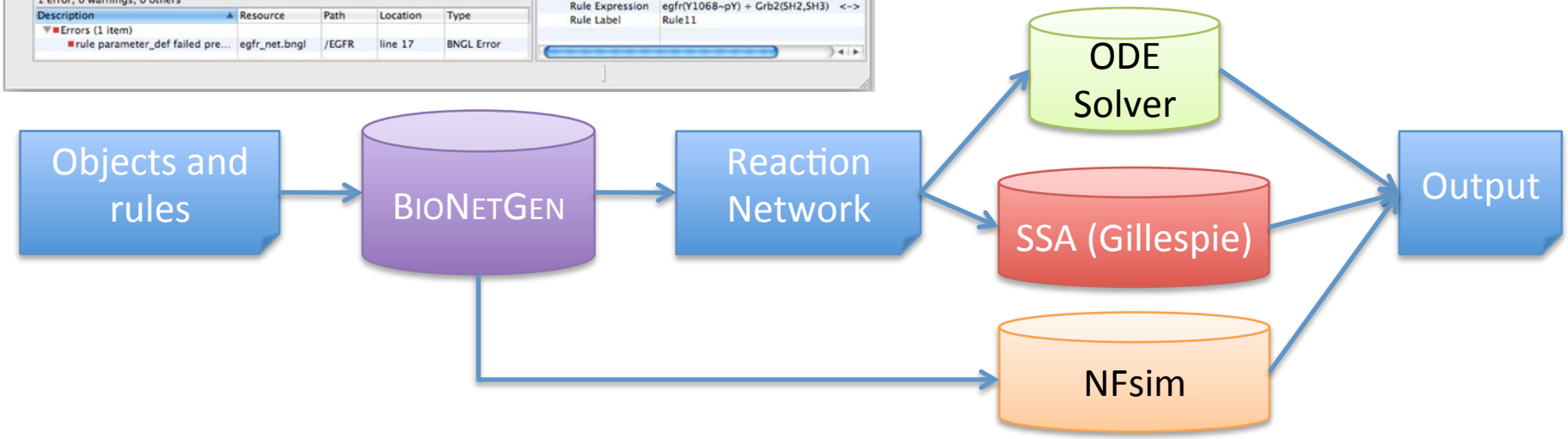
system configuration

total propensity $24k_{1+} + 6k_{1-} + 30k_{2+} + 6k_{2-} + 21k_{3+} + 5k_{3-}$

RuleBender/BioNetGen/NFsim

RuleBender – integrated development environment (IDE)

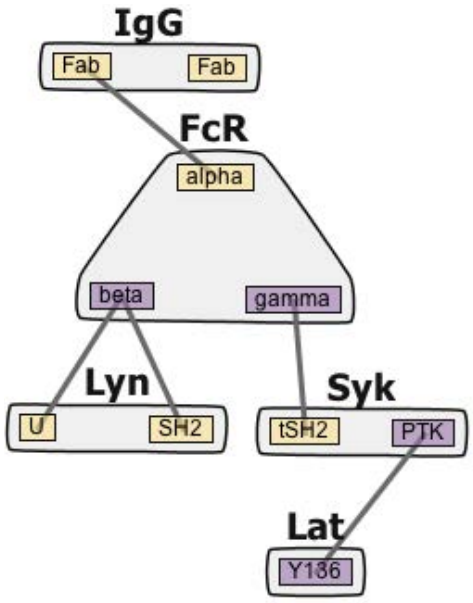




http://bionetgen.org/index.php/Download

Why use rule-based modeling techniques?

- ***Concise and precise* representation of biochemical knowledge**
  - Rules provide a convenient language for representing biomolecular interactions
  - Intricate molecular mechanisms can be captured easily in rule-based models
- ***Flexible* with respect to simulation method**
  - Deterministic / Stochastic
  - Well-mixed / Compartmental / Spatial
- **Model elements are *modular* and *reusable***
  - Rule libraries (Chylek et al., 2014) *Frontiers in Immunology*
- **Compact and automatic *visualization***
  - Contact map and beyond
- **Easy *annotation***
  - Model elements can be directly mapped to database entries



Contact map

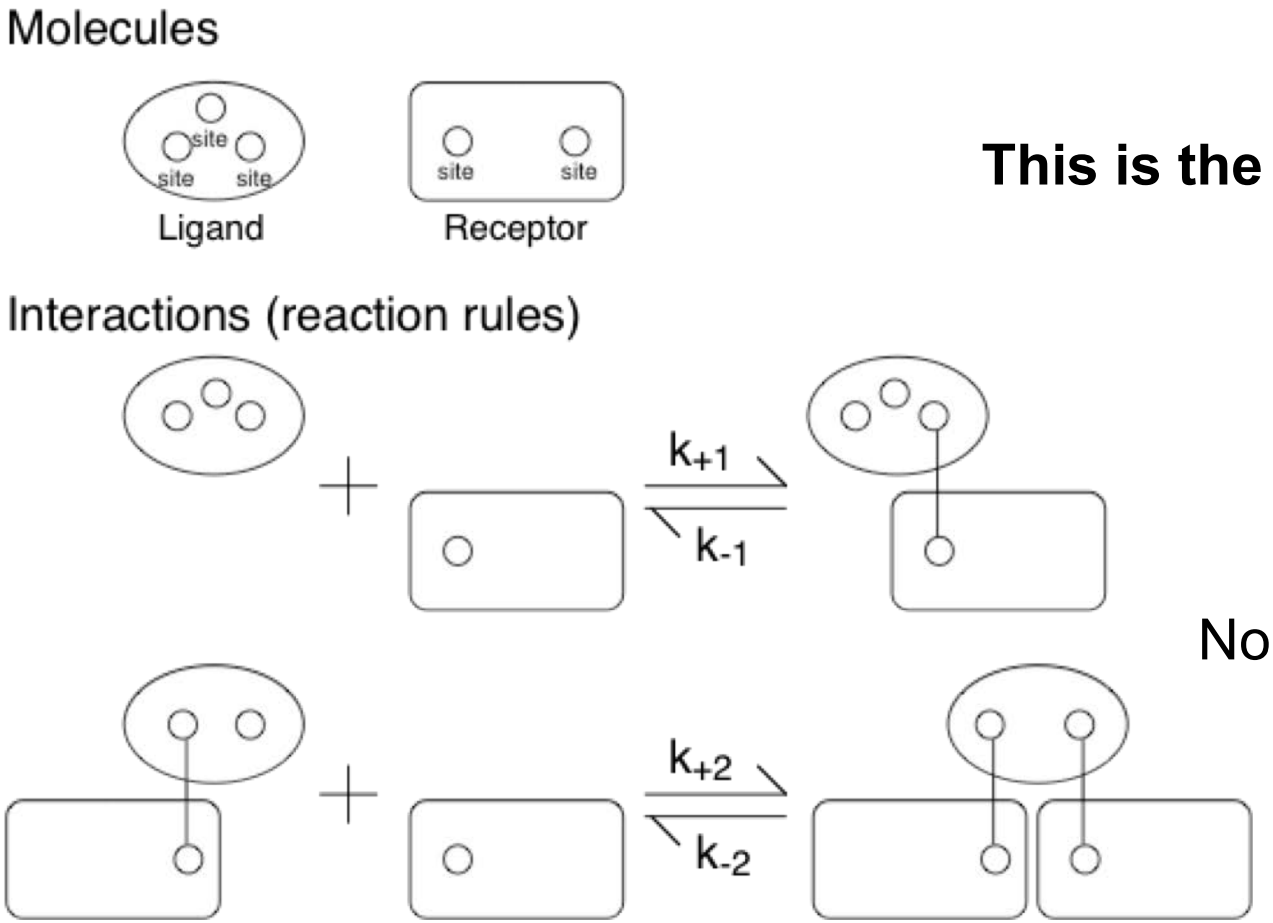
Outline

1. Features of signaling proteins
2. Combinatorial complexity: the key problem solved by rule-based modeling
3. Basic concepts of rule-based representation of biomolecular interactions
4. Simulation methods for rule-based models (indirect and direct)
5. **Exercises (computer lab)**

During the afternoon computer lab (**6/11, Mon**), we will build a simple rule-based model using RuleBender and look at several example models presented in this tutorial/review: Chylek et al. (2015) *Phys Biol* **12**: 045007.

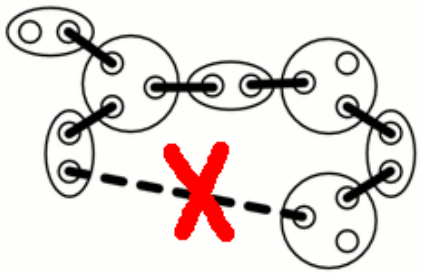
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4526164/

A rule-based model corresponding to the equilibrium continuum model of Goldstein and Perelson (1984)



This is the "TLBR model."

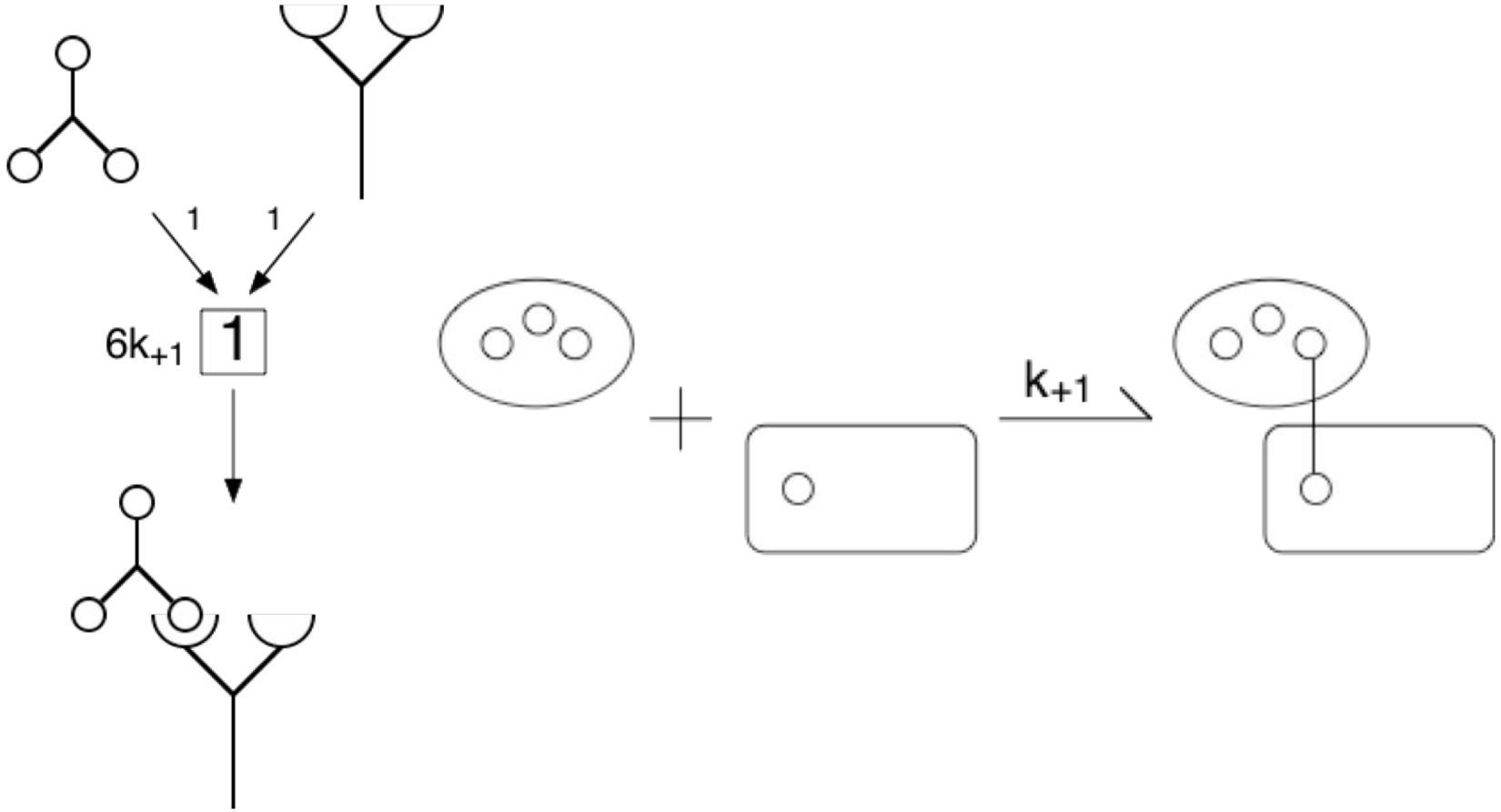
No cyclic aggregates



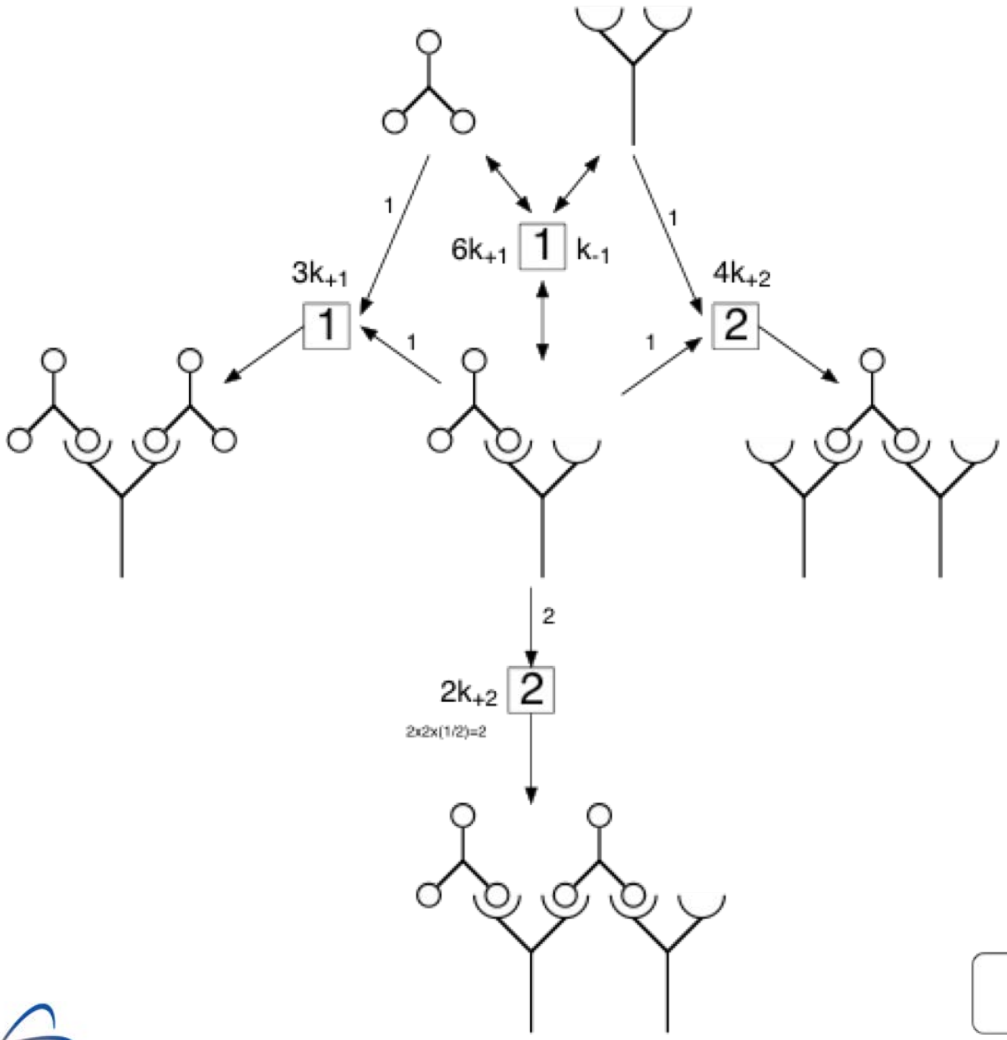
“Generate-first” method starts with seed species

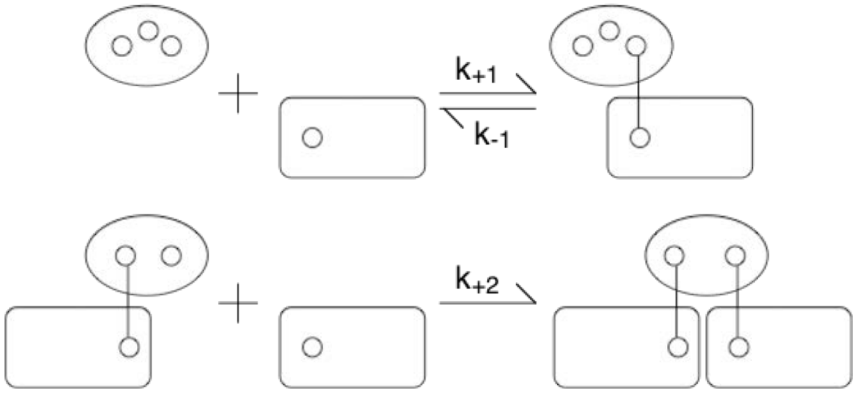
Ligand Receptor

After first round of rule application

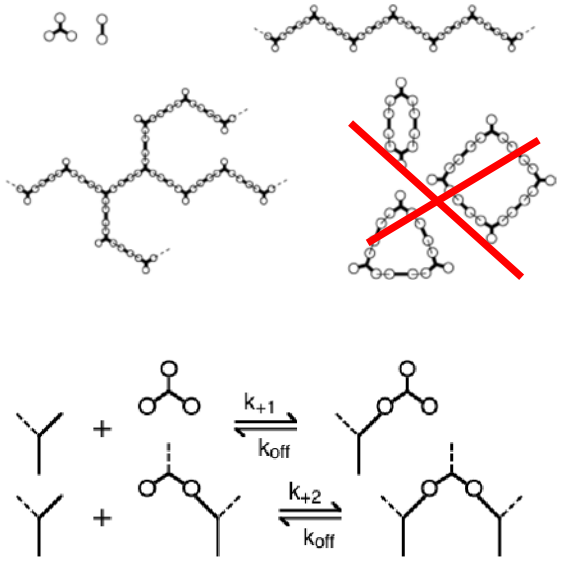


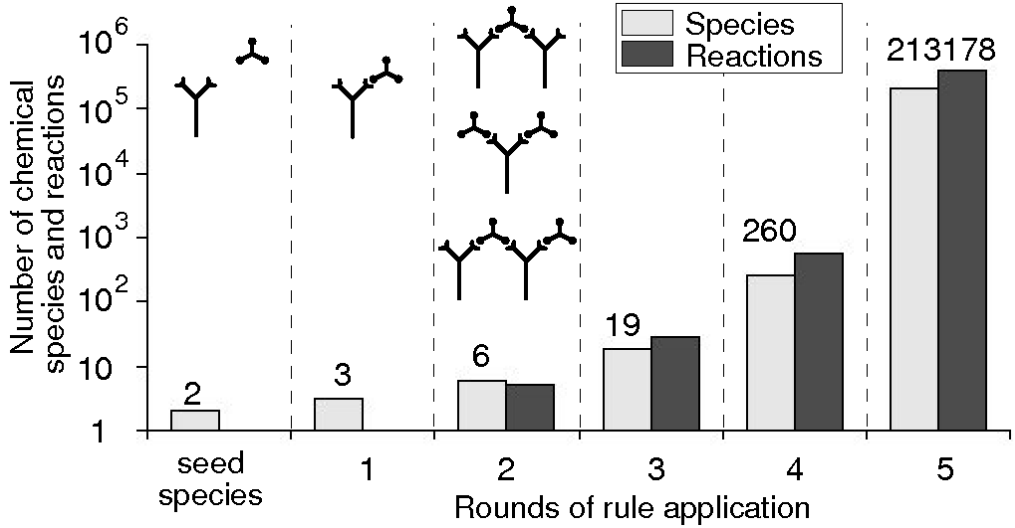
After the second round of rule application





Rule-derived network can be too large to simulate using conventional population-based methods





On The Fly method