Rule-based Modeling

Bill Hlavacek
Theoretical Division
Los Alamos National Laboratory

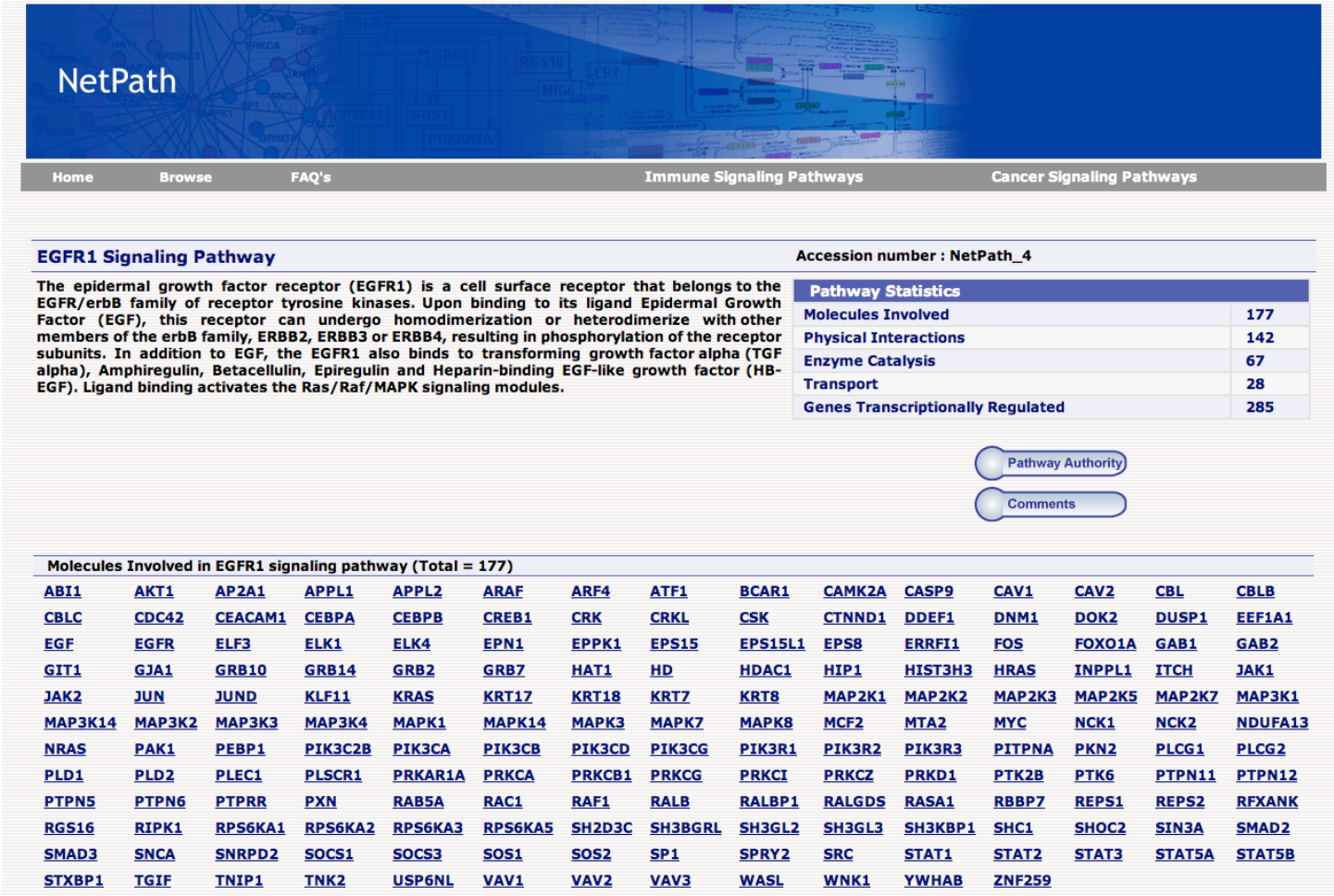
Los Alamos NATIONAL LABORATORY EST.1943

NNSA

Who is this?

http://bionetgen.org

Cellular regulatory systems are complex



NetPath

Home | Browse | FAQ's | Immune Signaling Pathways | Cancer Signaling Pathways

EGFR1 Signaling Pathway

Accession number : NetPath_4

The epidermal growth factor receptor (EGFR1) is a cell surface receptor that belongs to the EGFR/erbB family of receptor tyrosine kinases. Upon binding to its ligand Epidermal Growth Factor (EGF), this receptor can undergo homodimerization or heterodimerize with other members of the erbB family, ERBB2, ERBB3 or ERBB4, resulting in phosphorylation of the receptor subunits. In addition to EGF, the EGFR1 also binds to transforming growth factor alpha (TGF alpha), Amphiregulin, Betacellulin, Epiregulin and Heparin-binding EGF-like growth factor (HB-EGF). Ligand binding activates the Ras/Raf/MAPK signaling modules.

Pathway Statistics	
Molecules Involved	177
Physical Interactions	142
Enzyme Catalysis	67
Transport	28
Genes Transcriptionally Regulated	285

Pathway Authority

Comments

Molecules Involved in EGFR1 signaling pathway (Total = 177)

ABI1	AKT1	AP2A1	APPL1	APPL2	ARAF	ARF4	ATF1	BCAR1	CAMK2A	CASP9	CAV1	CAV2	CBL	CBLB
CBLC	CDC42	CEACAM1	CEBPA	CEBPB	CREB1	CRK	CRKL	CSK	CTNND1	DDEF1	DNM1	DOK2	DUSP1	EEF1A1
EGF	EGFR	ELF3	ELK1	ELK4	EPN1	EPPK1	EPS15	EPS15L1	EPS8	ERRFI1	FOS	FOXO1A	GAB1	GAB2
GIT1	GJA1	GRB10	GRB14	GRB2	GRB7	HAT1	HD	HDAC1	HIP1	HIST3H3	HRAS	INPPL1	ITCH	JAK1
JAK2	JUN	JUND	KLF11	KRAS	KRT17	KRT18	KRT7	KRT8	MAP2K1	MAP2K2	MAP2K3	MAP2K5	MAP2K7	MAP3K1
MAP3K14	MAP3K2	MAP3K3	MAP3K4	MAPK1	MAPK14	MAPK3	MAPK7	MAPK8	MCF2	MTA2	MYC	NCK1	NCK2	NDUFA13
NRAS	PAK1	PEBP1	PIK3C2B	PIK3CA	PIK3CB	PIK3CD	PIK3CG	PIK3R1	PIK3R2	PIK3R3	PITPNA	PKN2	PLCG1	PLCG2
PLD1	PLD2	PLEC1	PLSCR1	PRKAR1A	PRKCA	PRKCB1	PRKCG	PRKCI	PRKCZ	PRKD1	PTK2B	PTK6	PTPN11	PTPN12
PTPN5	PTPN6	PTPRR	PXN	RAB5A	RAC1	RAF1	RALB	RALBP1	RALGDS	RASA1	RBBP7	REPS1	REPS2	RFXANK
RGS16	RIPK1	RPS6KA1	RPS6KA2	RPS6KA3	RPS6KA5	SH2D3C	SH3BGRL	SH3GL2	SH3GL3	SH3KBP1	SHC1	SHOC2	SIN3A	SMAD2
SMAD3	SNCA	SNRPD2	SOCS1	SOCS3	SOS1	SOS2	SP1	SPRY2	SRC	STAT1	STAT2	STAT3	STAT5A	STAT5B
STXBP1	TGIF	TNIP1	TNK2	USP6NL	VAV1	VAV2	VAV3	WASL	WNK1	YWHAB	ZNF259			

Akhilesh Pandey (Johns Hopkins)

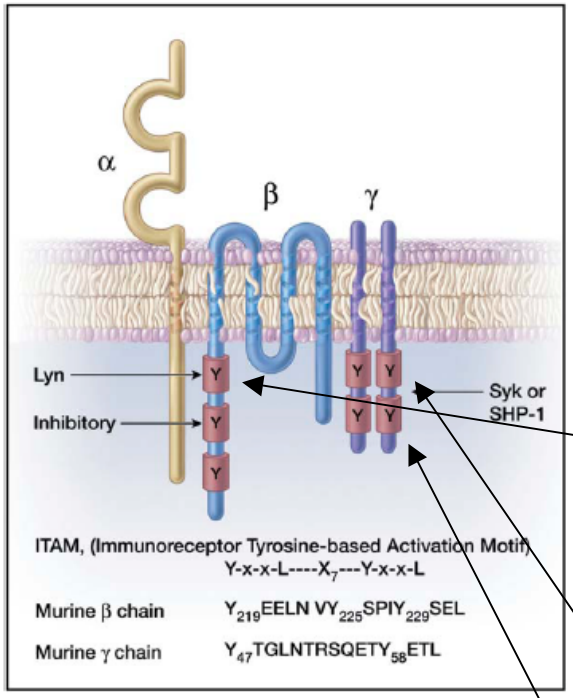
Value added by modeling

1. We can use models to organize information about a system with precision
2. We can determine the logical consequences of a model specification

Outline

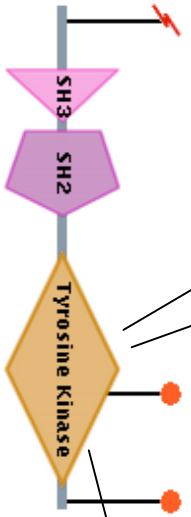
1. **Combinatorial complexity**
2. The conventional approach to modeling
3. The rule-based approach to modeling
4. Tools
5. New simulation methods

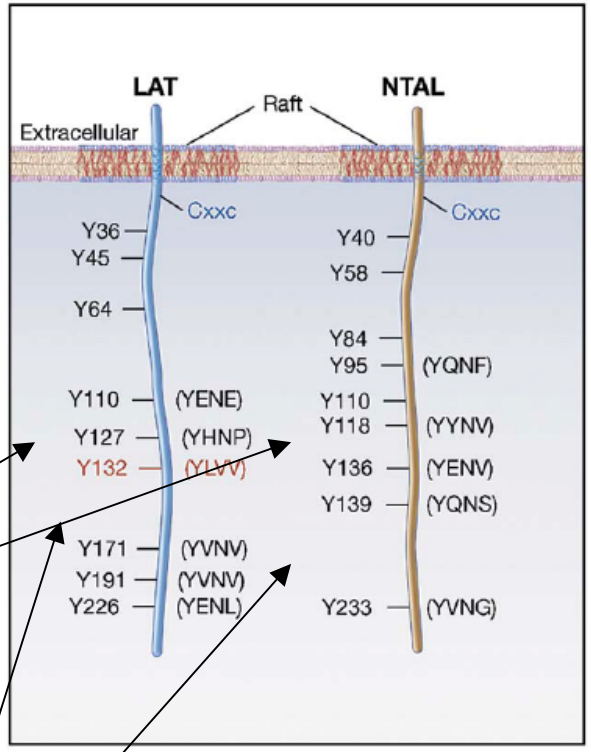
Signaling proteins contain domains and motifs that mediate interactions with other proteins



FcεRI

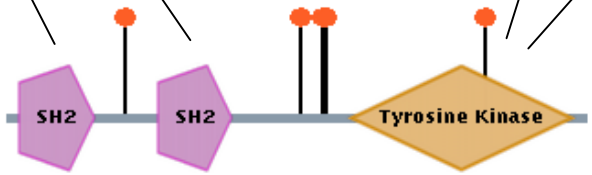
Lyn





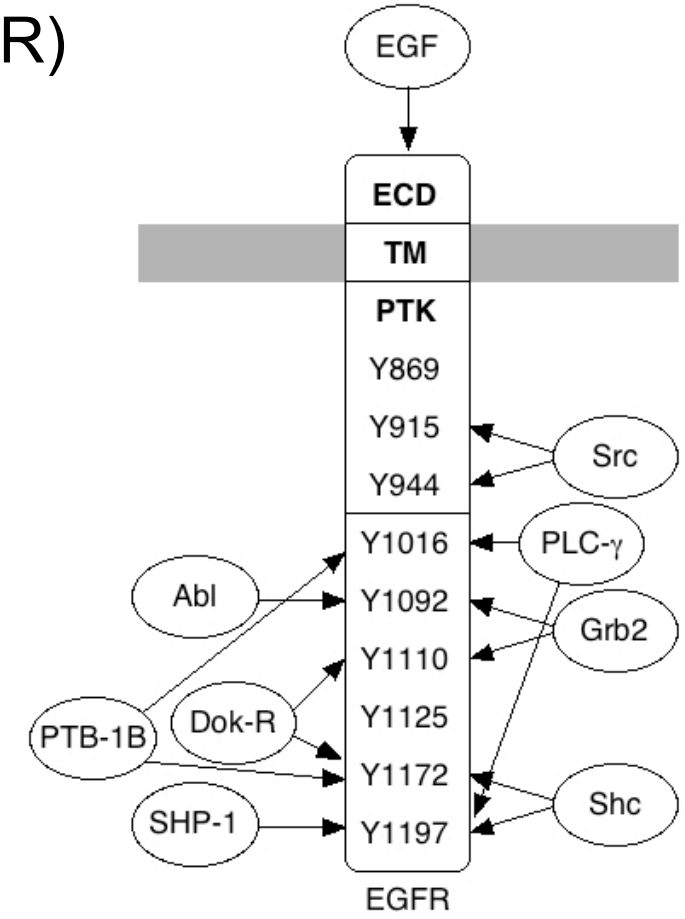
Transmembrane Adaptors

Syk



Multiplicity of sites and binding partners gives rise to combinatorial complexity

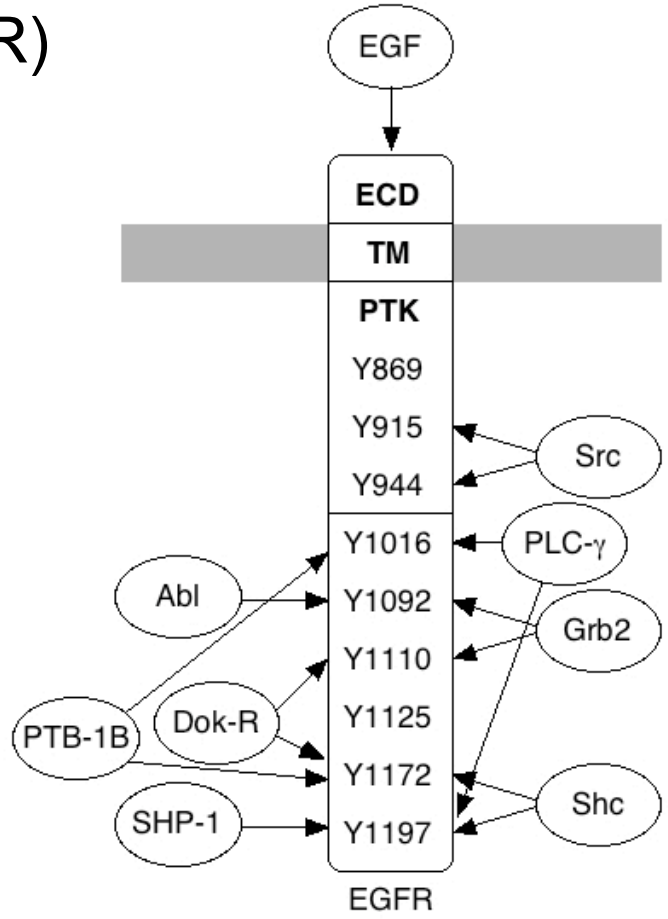
Epidermal growth factor receptor (EGFR)



Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

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



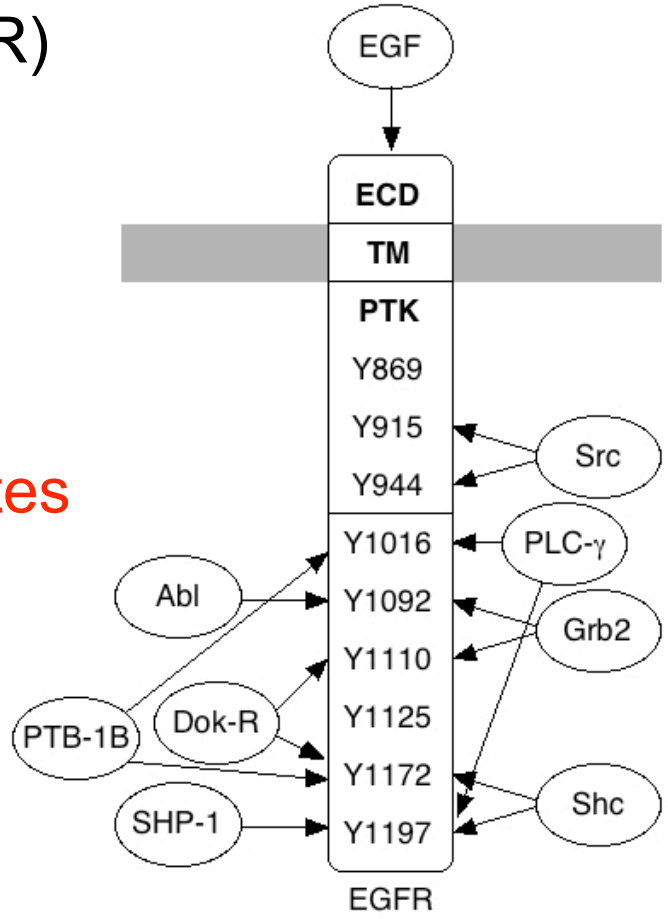
Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

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

Each site has ≥ 1 binding partner

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



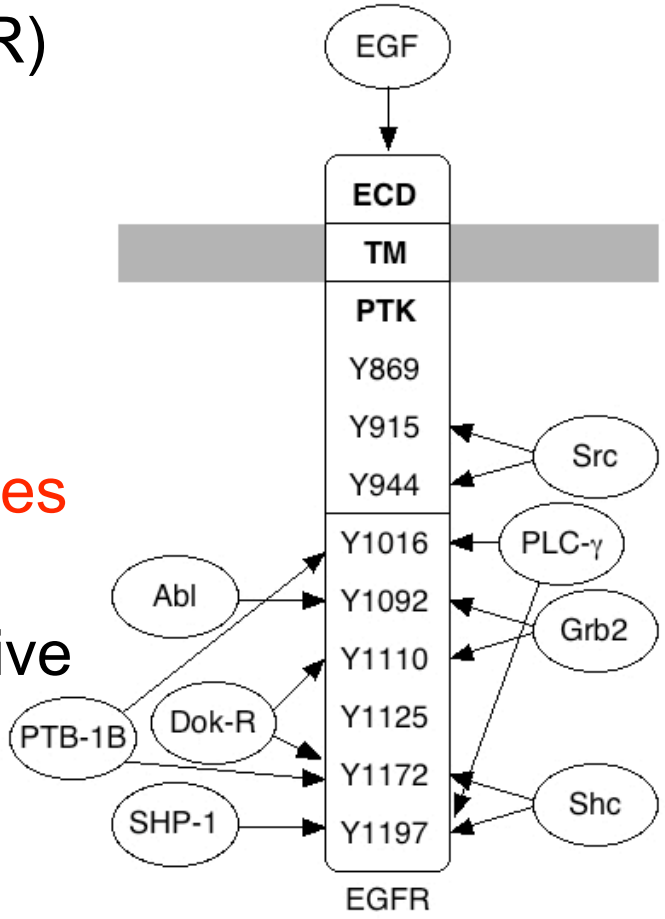
Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

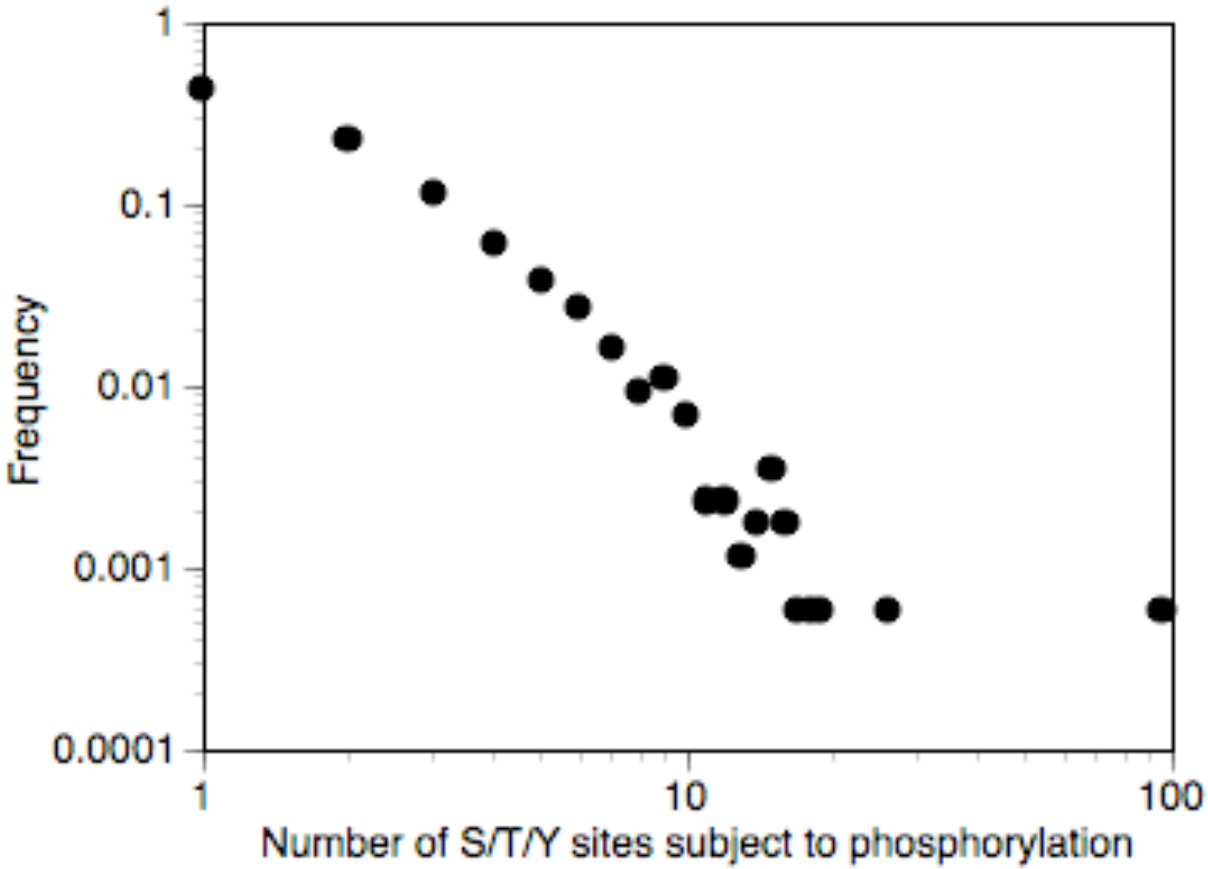
9 sites $\Rightarrow 2^9 = 512$ phosphorylation states

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

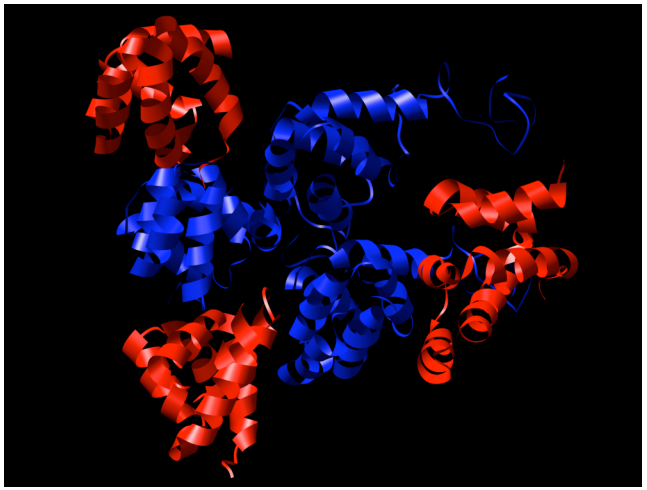


> 50% are phosphorylated at 2 or more sites

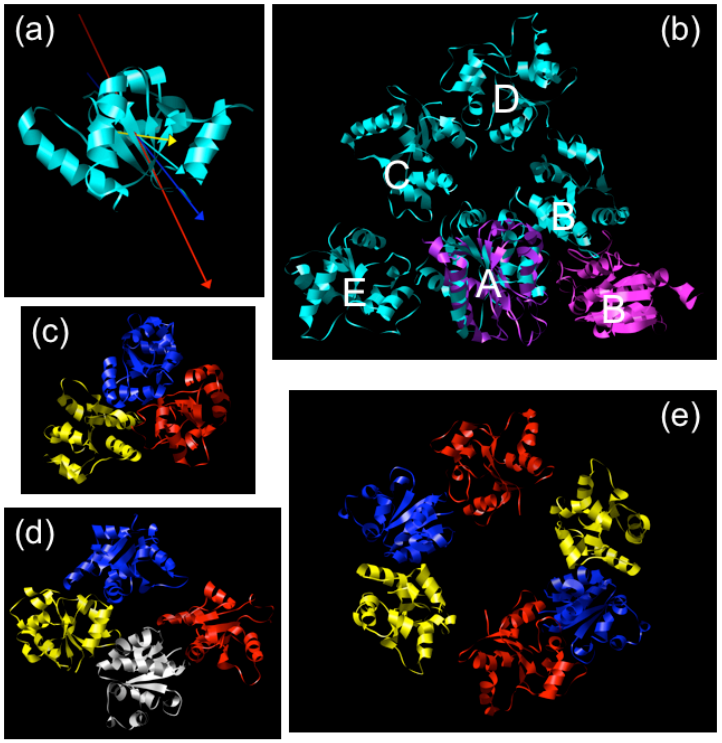
Source: Phospho.ELM database v. 3.0 (http://phospho.elm.eu.org)

Oligomerization alone can generate many complexes

Complexes potentially involved in Toll-like receptor signaling



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



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

The problem of combinatorial complexity necessitates a new modeling approach

- Inside a Chemical Plant
  - Large numbers of molecules…
  - …of a few types
  - Conventional modeling works fine
- Inside a Cell
  - Small numbers of molecules…
  - …of many possible types
  - Rule-based modeling addresses this situation

The need for predictive models of large scale with site-specific details

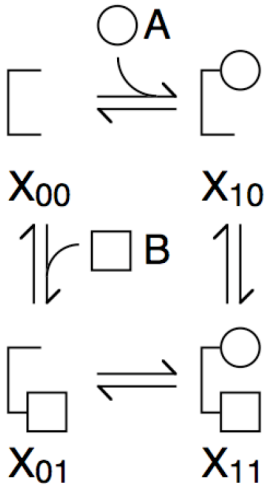
- **Molecular changes that affect cell signaling cause disease (cancer)**
- **Over 200 drugs that target malfunctioning signaling proteins are currently in clinical trials**
  - One spectacular success (Gleevec)
  - But results are largely disappointing for most patients
- **96 clinical trials are underway to test combinations of drugs (clinicaltrials.gov)**
  - There are too many combinations to consider all of them in trials

Outline

1. The biochemistry of cell signaling and combinatorial complexity
2. **The conventional approach to modeling**
3. The rule-based approach to modeling
4. Tools
5. New simulation methods

Models can be specified in different ways

Conventional representation of a biochemical reaction network



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

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

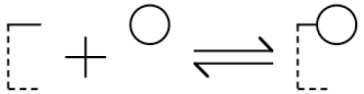
$X_{10} + B \rightleftharpoons X_{11}$

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

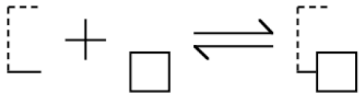
6 chemical species
8 unidirectional reactions

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

Rule-based representation



X(a)+A(x)<->X(a!1).A(x!1)

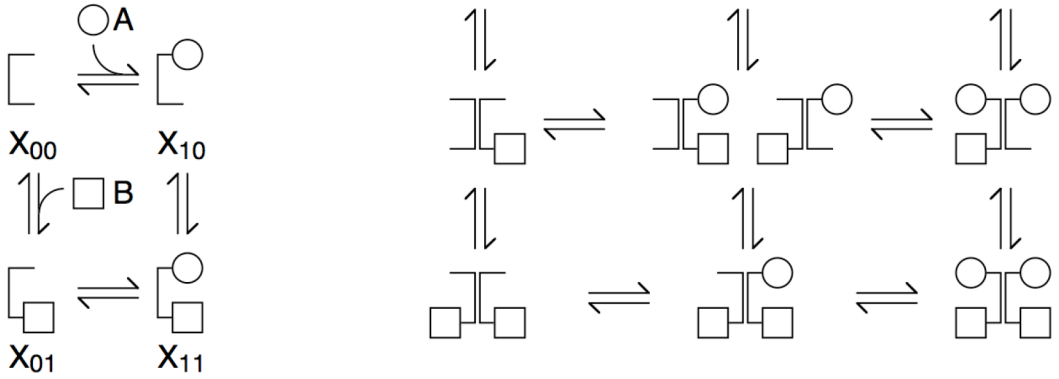


X(b)+B(x)<->X(b!1).B(x!1)

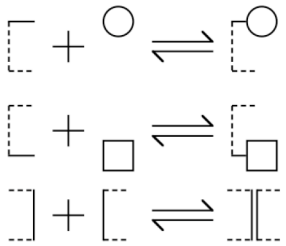
Rules representing molecular interactions allow for compact model specifications

Network size increases nonlinearly when an extra interaction is considered

16 chemical species
60 unidirectional reactions



The number of rules scales linearly with the number of molecular interactions in a system



X(a)+A(x)<->X(a!1).A(x!1)

X(b)+B(x)<->X(b!1).B(x!1)

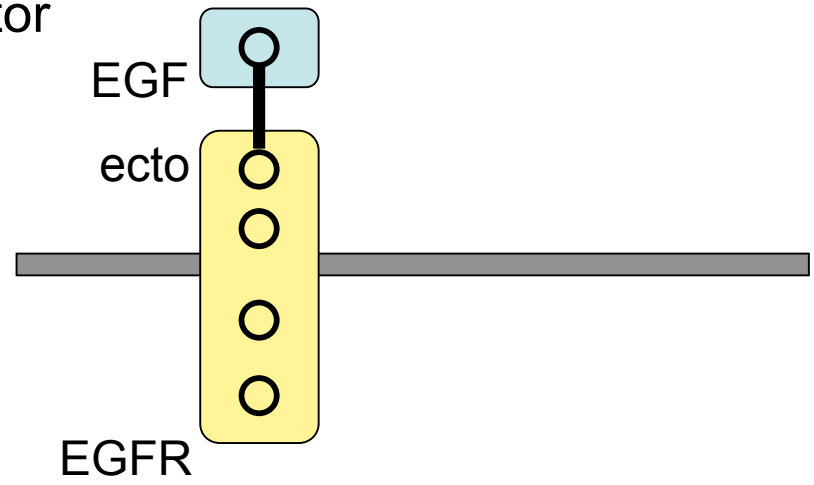
X(x)+X(x)<->X(x!1).X(x!1)

Science's STKE re6 (2006)

Early events in EGFR signaling - we'll consider these events to illustrate modeling approaches

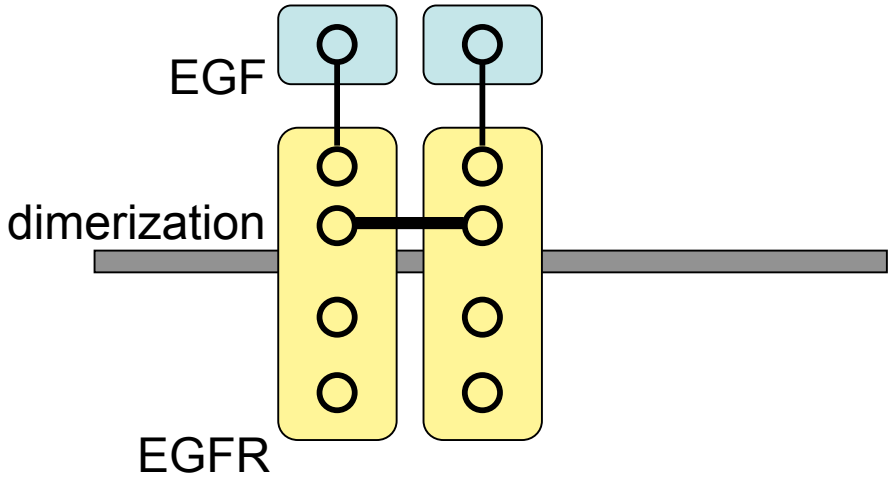
EGF = epidermal growth factor
EGFR = epidermal growth factor receptor

1. EGF binds EGFR



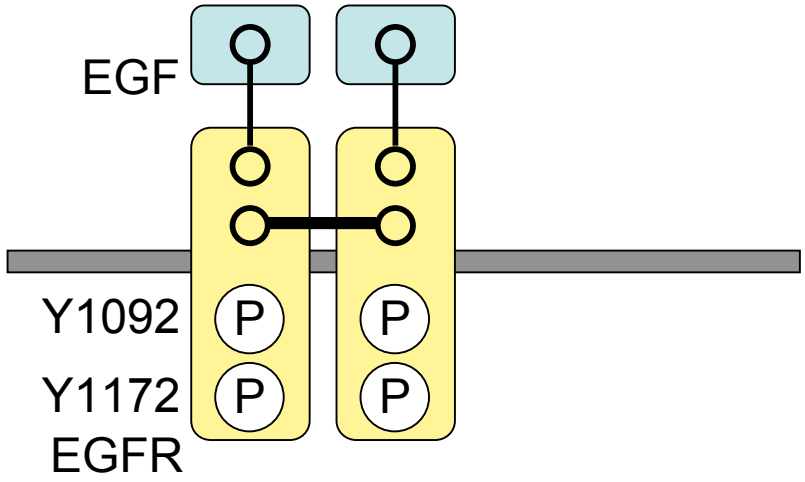
Early events in EGFR signaling

1. EGF binds EGFR
2. **EGFR dimerizes**



Early events in EGFR signaling

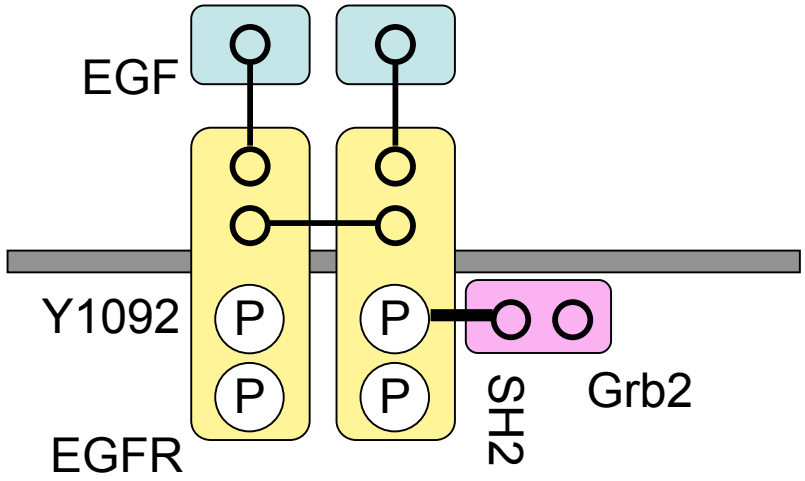
1. EGF binds EGFR
2. EGFR dimerizes
3. **EGFR transphosphorylates itself**



Early events in EGFR signaling

Grb2 pathway

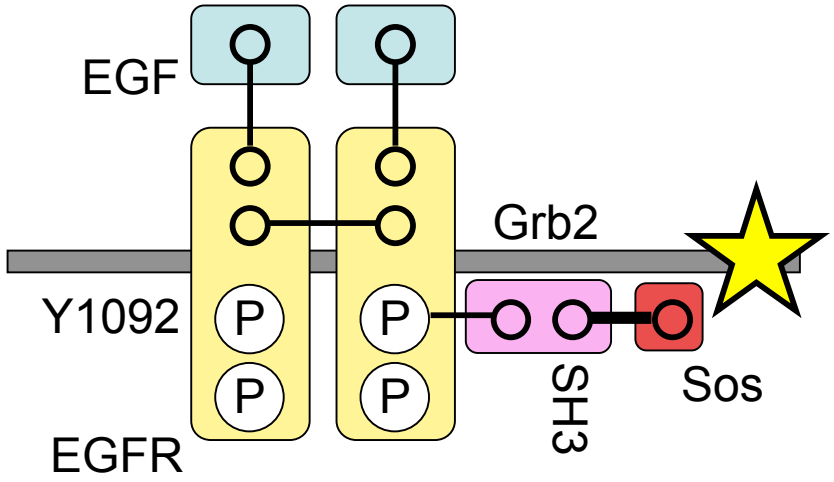
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. **Grb2 binds phospho-EGFR**



Early events in EGFR signaling

Grb2 pathway

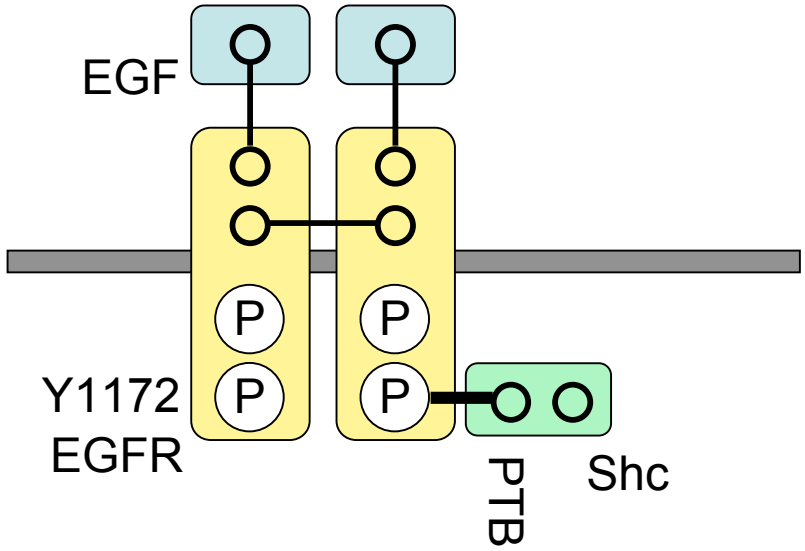
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. Grb2 binds phospho-EGFR
5. **Sos binds Grb2 (Activation Path 1)**



Early events in EGFR signaling

Shc pathway

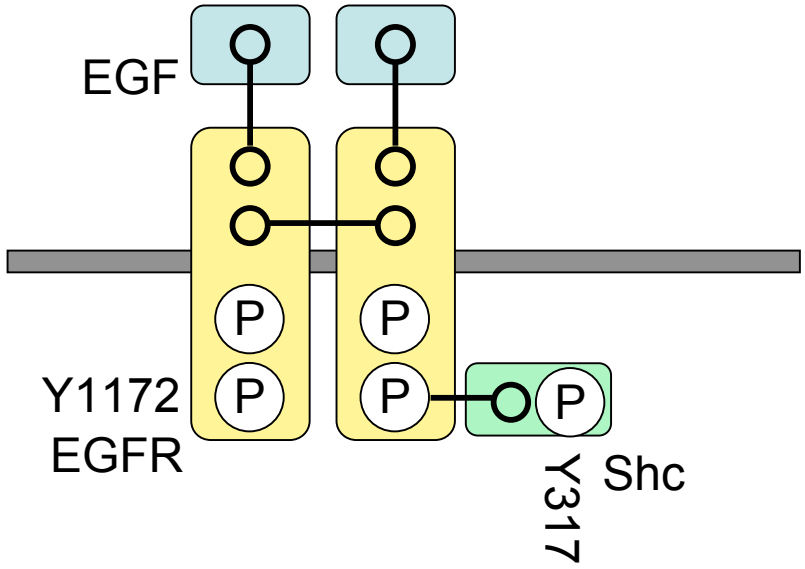
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. **Shc binds phospho-EGFR**



Early events in EGFR signaling

Shc pathway

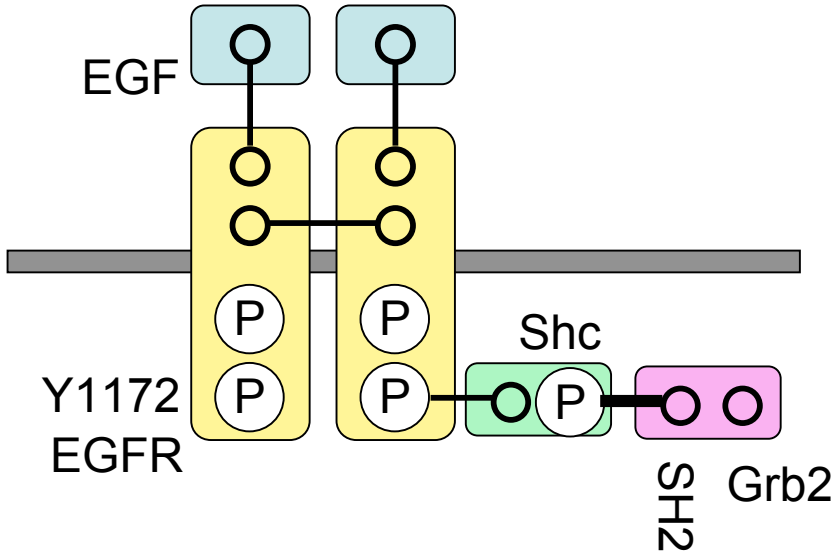
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. Shc binds phospho-EGFR
5. **EGFR transphosphorylates Shc**



Early events in EGFR signaling

Shc pathway

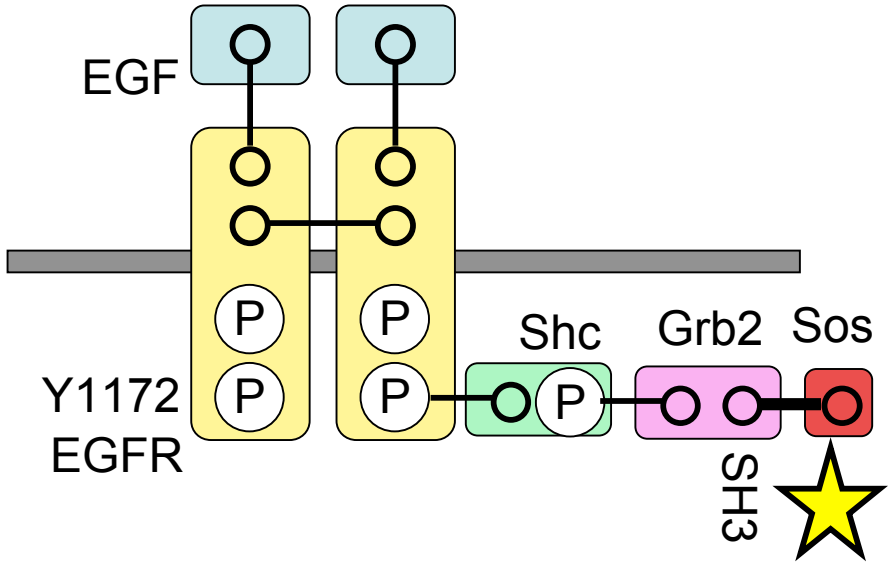
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. **Grb2 binds phospho-Shc**



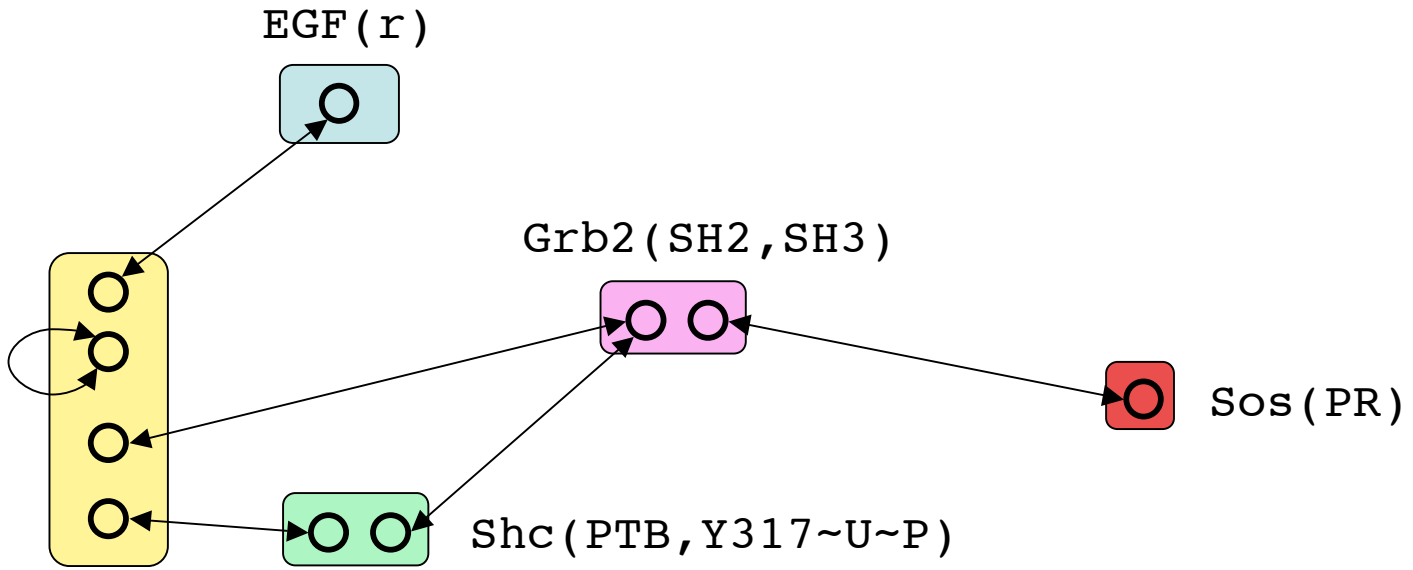
Early events in EGFR signaling

Shc pathway

1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates itself
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. Grb2 binds phospho-Shc
7. **Sos binds Grb2 (Activation Path 2)**



Representation of molecules in a simple model of early events in EGFR signaling

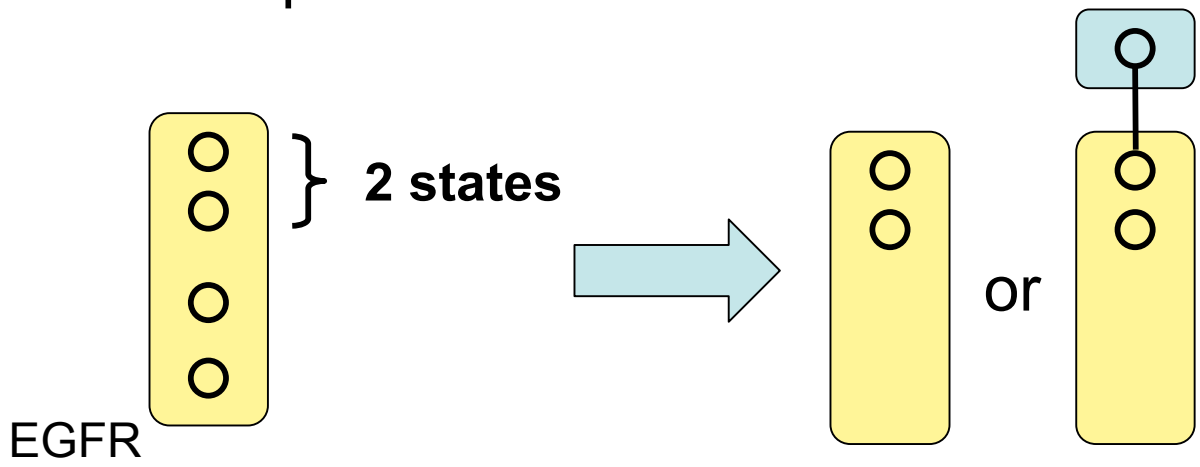


EGFR(l,d,Y1092~U~P,Y1172~U~P)

Blinov et al. (2006)

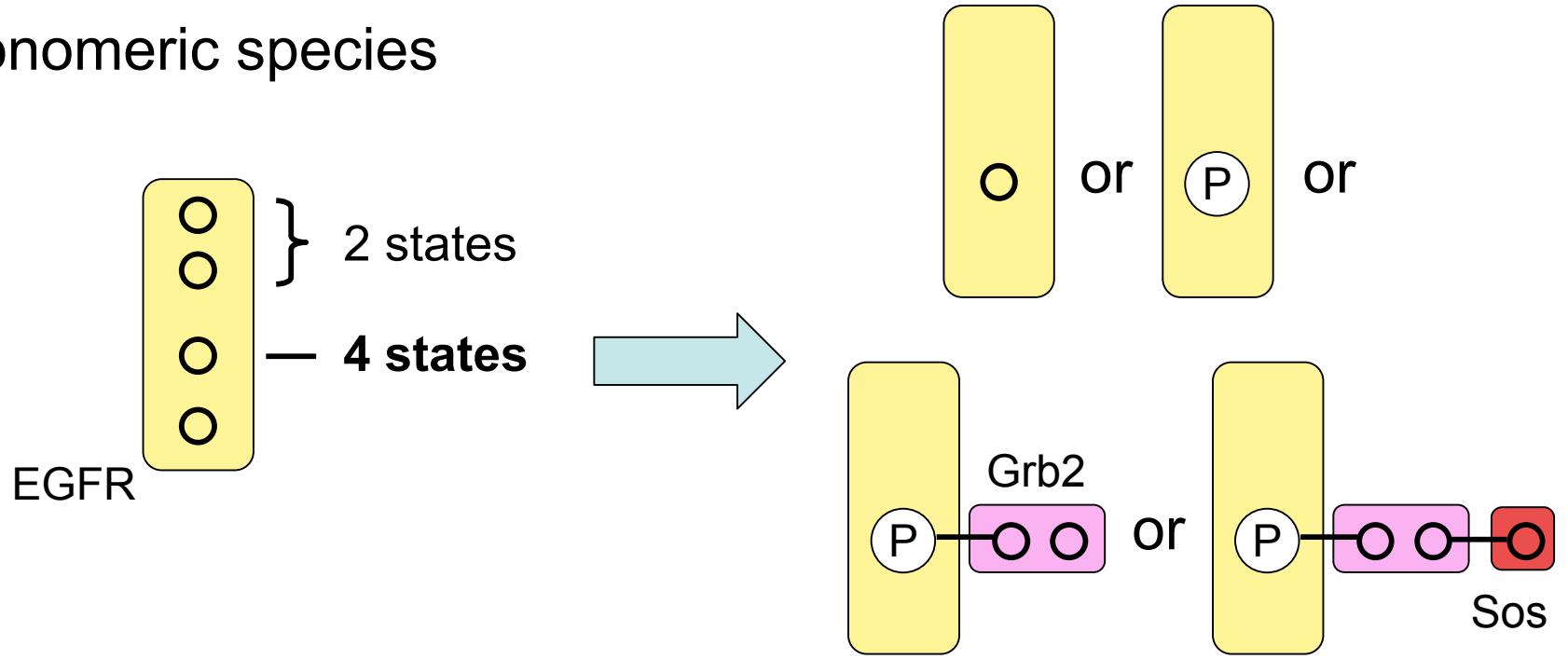
Combinatorial complexity of early events

Monomeric species

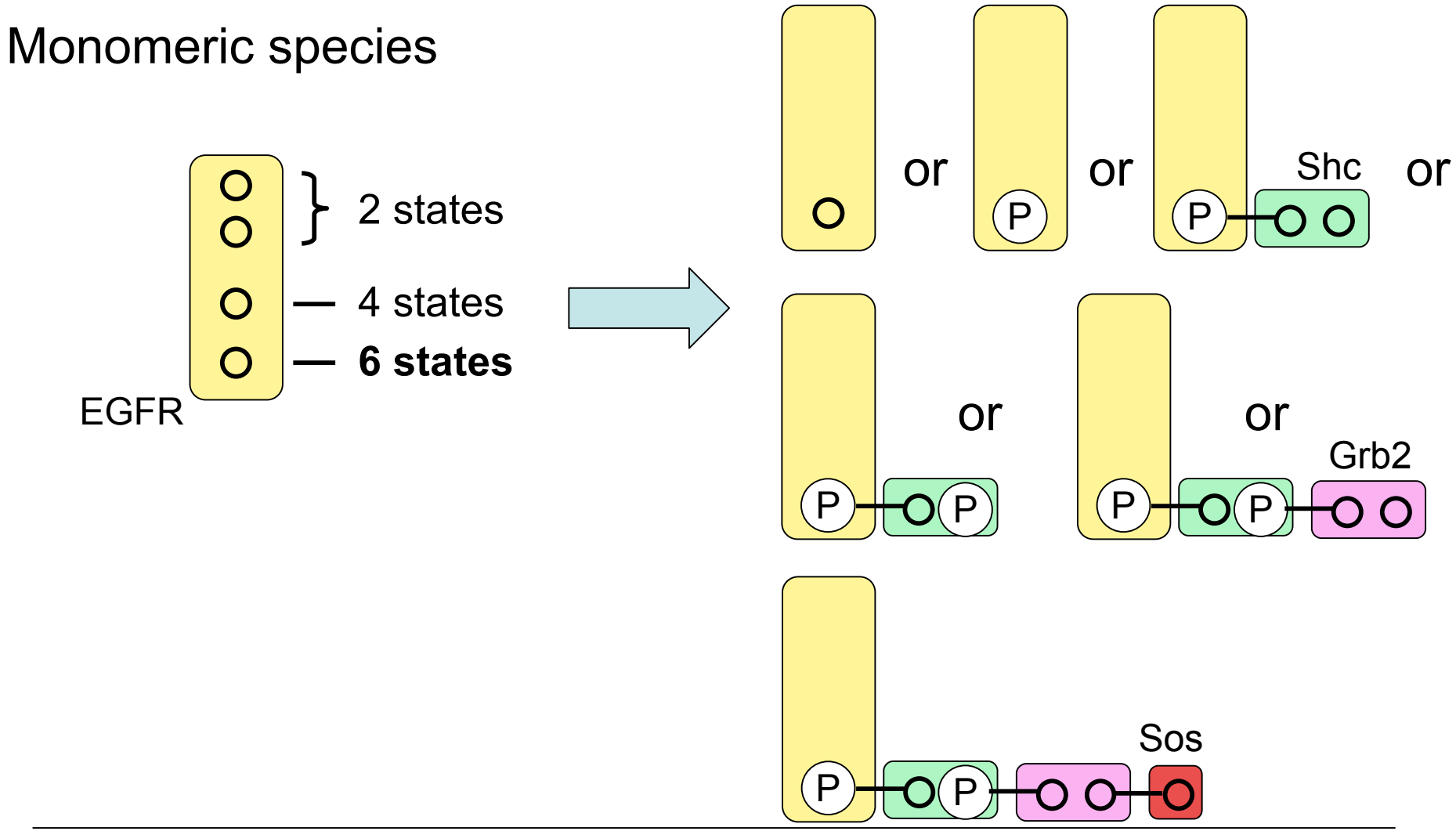


Combinatorial complexity of early events

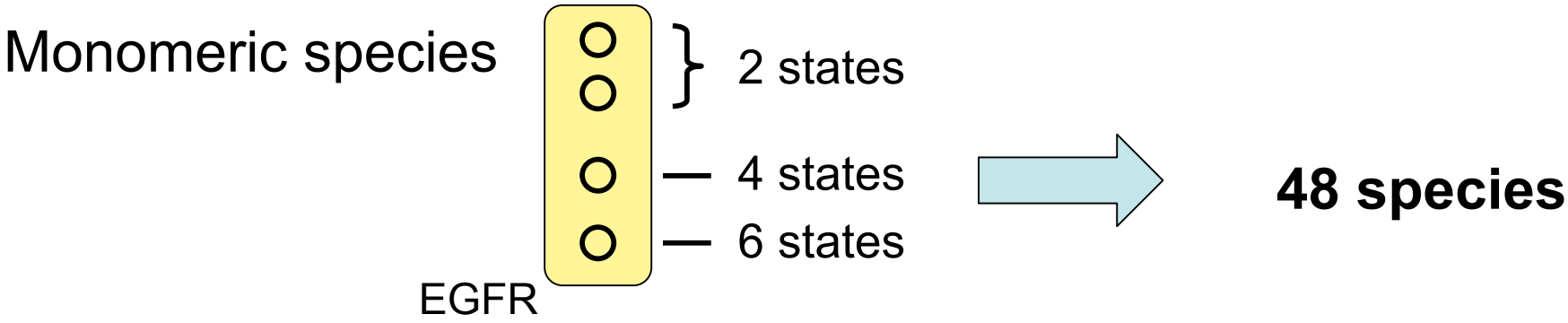
Monomeric species



Combinatorial complexity of early events

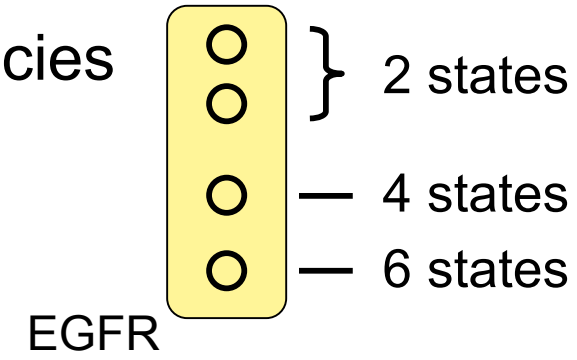


Combinatorial complexity of early events



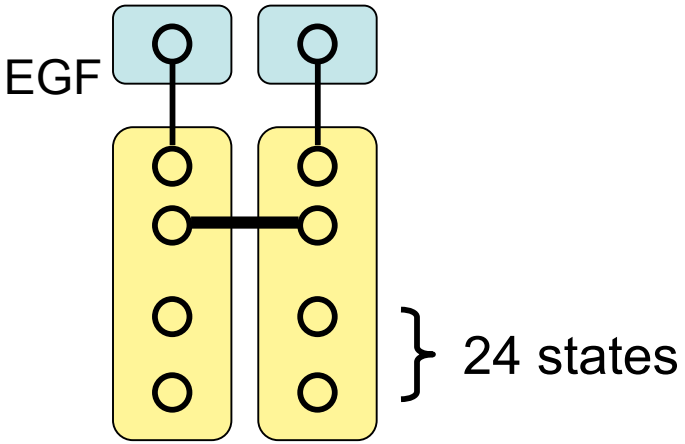
Combinatorial complexity of early events

Monomeric species



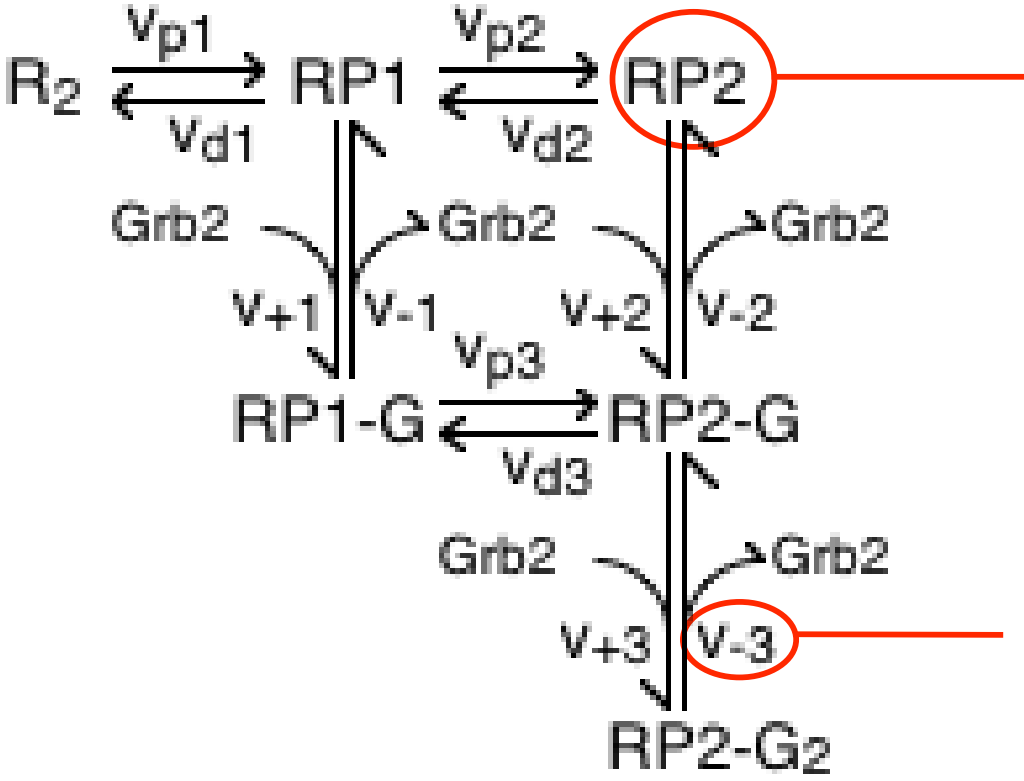
48 species

Dimeric species



$N \times (N+1)/2$ = **300 species**

A reaction-scheme diagram



Species: One for every possible modification state of every complex

Reactions: One for every transition among species

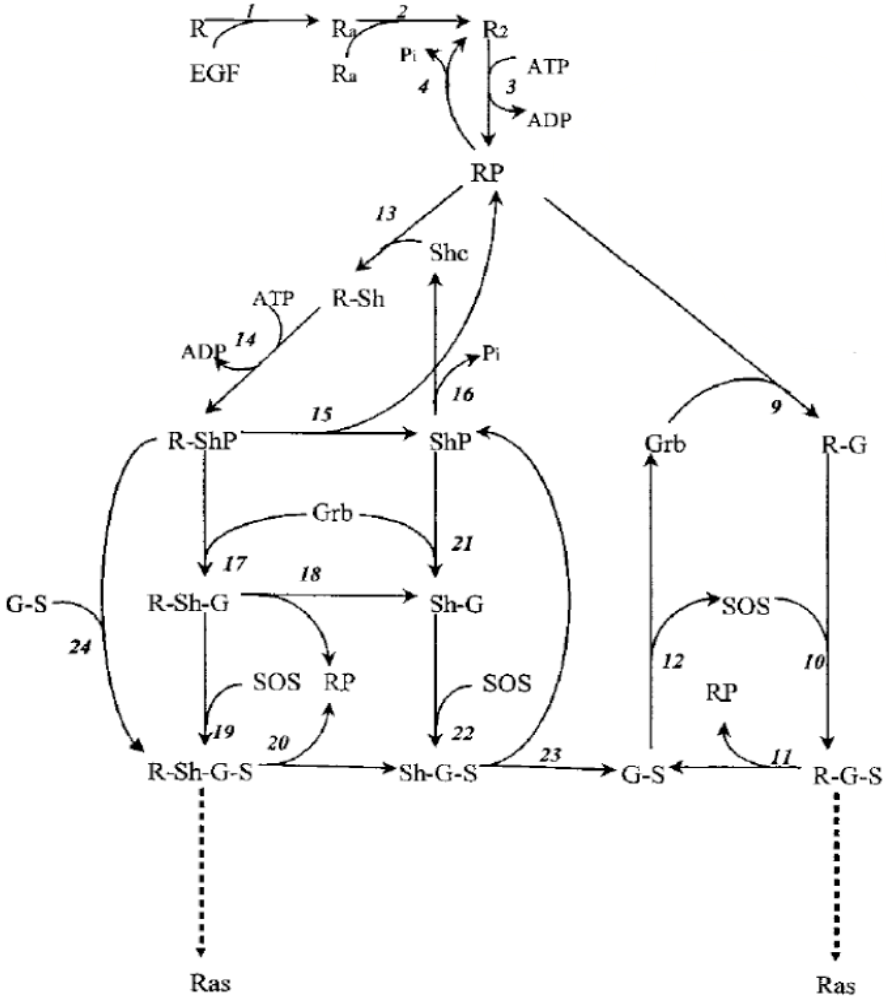
This scheme can be translated to obtain a set of ODEs, one for each species

A conventional model for EGFR signaling

The Kholodenko model*

5 proteins

18 species
34 reactions

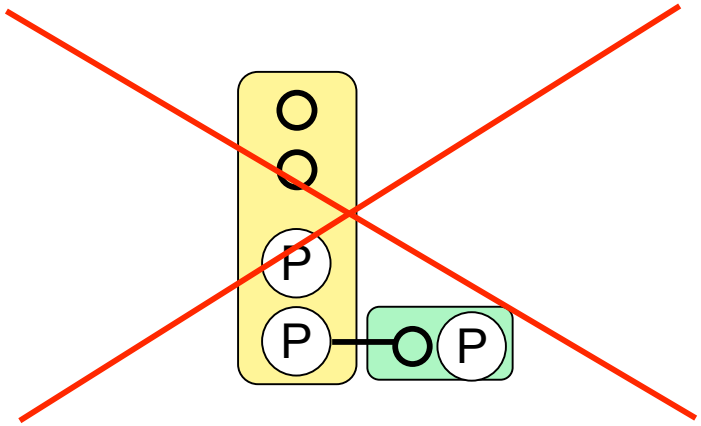


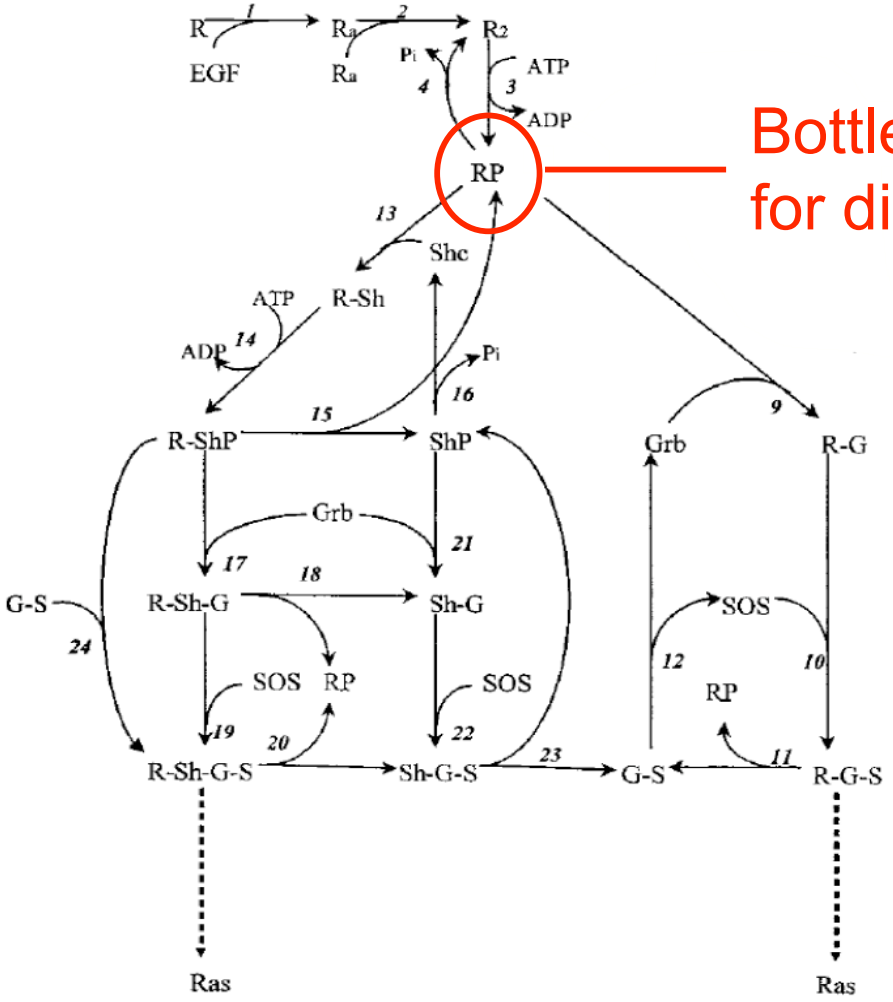
J. Biol. Chem. **274**, 30169 (1999)

Assumptions made to limit combinatorial complexity

1. Phosphorylation inhibits dimer breakup

No modified monomers



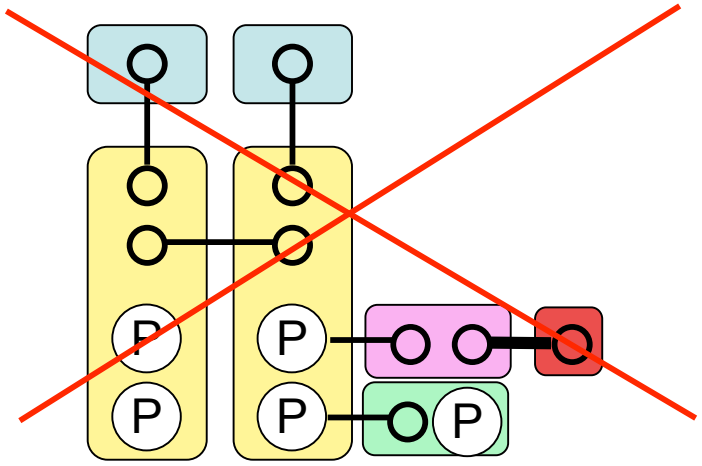


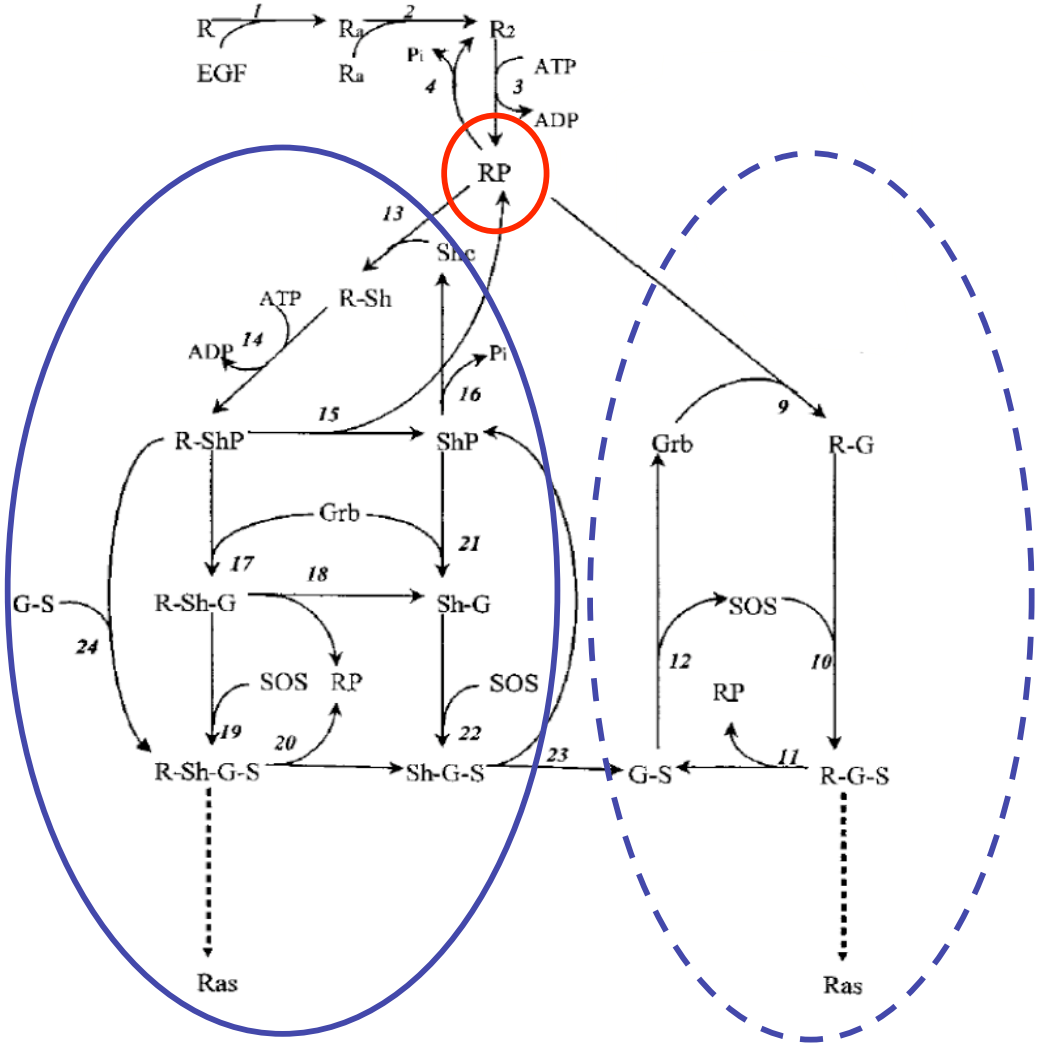
Bottleneck for dimers

Assumptions made to limit combinatorial complexity

2. Adaptor binding is competitive

No dimers with more than one associated adapter





Outline

1. The biochemistry of cell signaling and combinatorial complexity
2. The conventional approach to modeling
3. **The rule-based approach to modeling**
4. Tools
5. New simulation methods

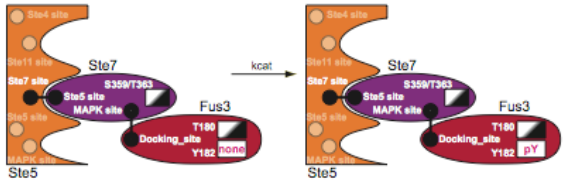
Rules operate on structured objects (graphs)

Graphs represent molecules, their component parts, and states

A (graph-rewriting) rule specifies the addition or removal of an edge to represent binding or unbinding, or the change of a state label to represent, for example, post-translational modification of a protein at a particular site

A model specification is readily visualized and compositional

Molecules, components, and states can be directly linked to annotation in databases

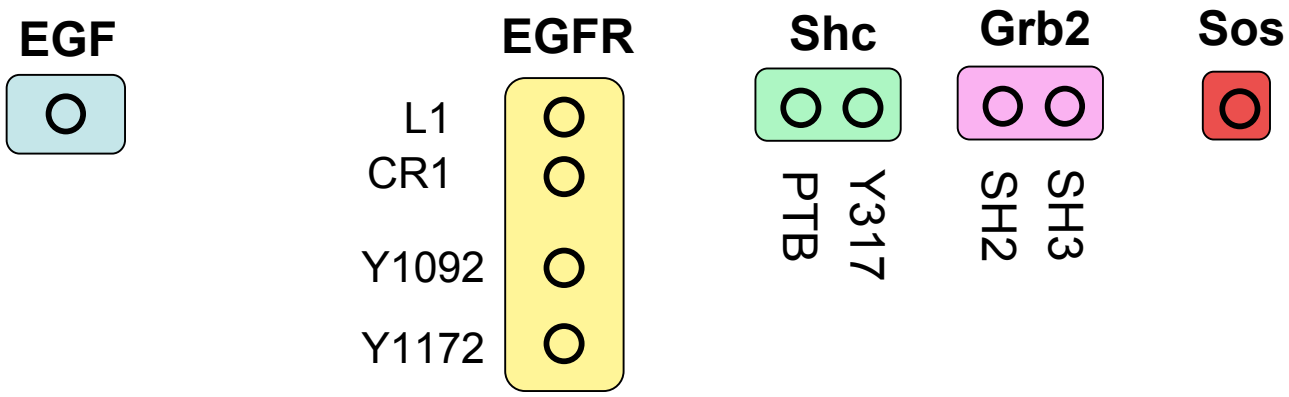


Ste7	Fus3	kcat
S359/T363	T180	
pS	none	kcat_Ste5Ste7*pS*Fus3_pY
pS	pT	kcat_Ste5Ste7*pS*Fus3*pT*_pY
pSpT	none	kcat_Ste5Ste7*pSpT*Fus3_pY
pSpT	pT	kcat_Ste5Ste7*pSpT*Fus3*pT*_pY

Ty Thomson (MIT) - yeastpheromonemodel.org

Proteins in a model are introduced with molecule templates

Molecule templates

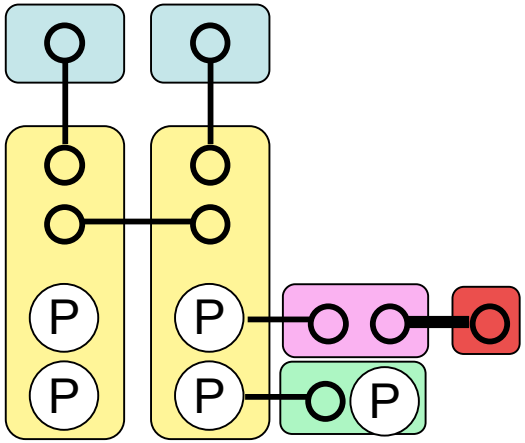


Nodes represent components of proteins

Components may have attributes: O or (P)

Complexes are connected instances of molecule templates

An EGFR dimer

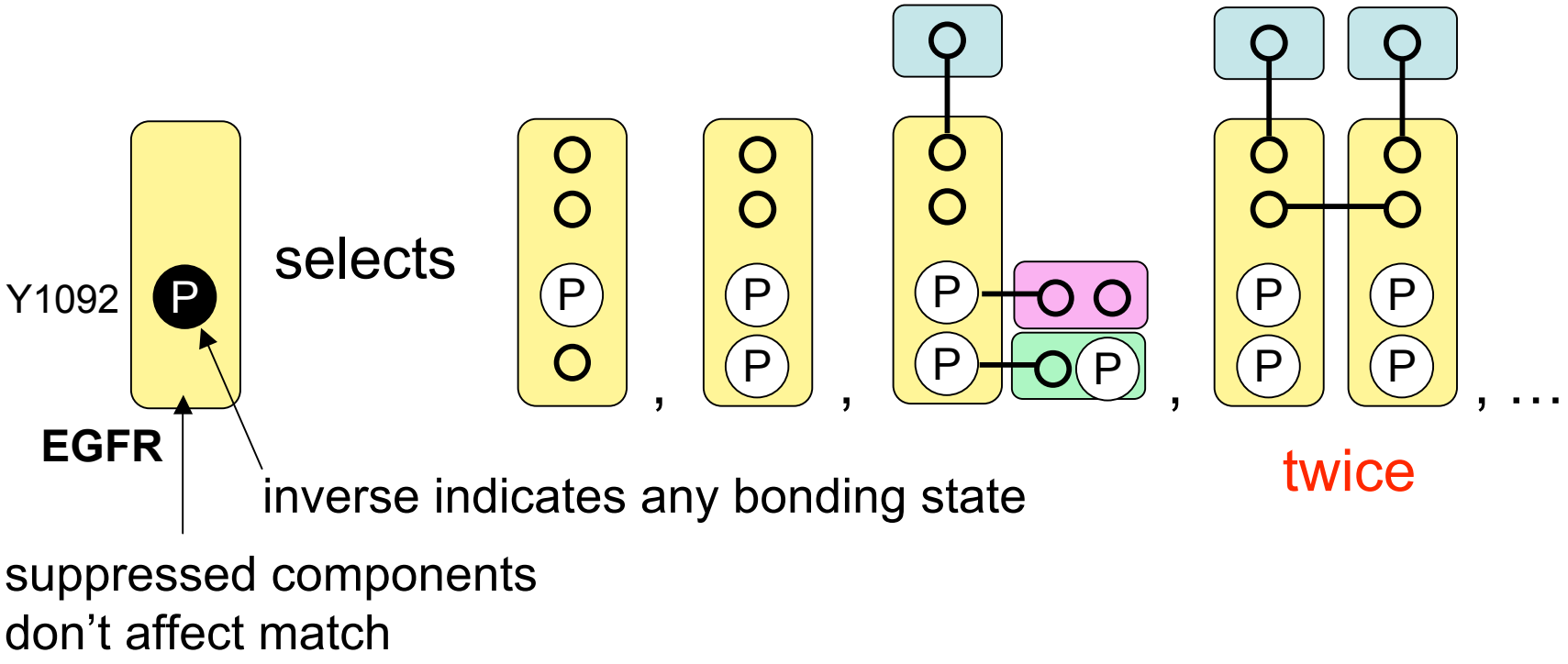


Edges represent bonds between components

Bonds may be internal or external

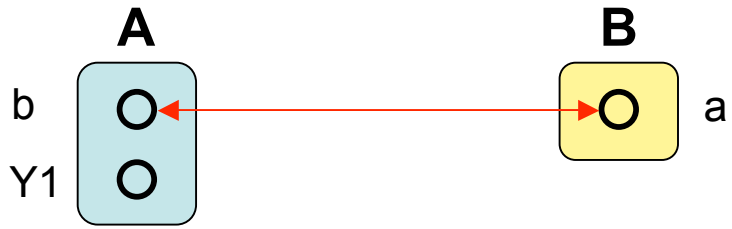
Patterns select sets of chemical species with common features

Pattern that selects EGFR phosphorylated at Y1092.



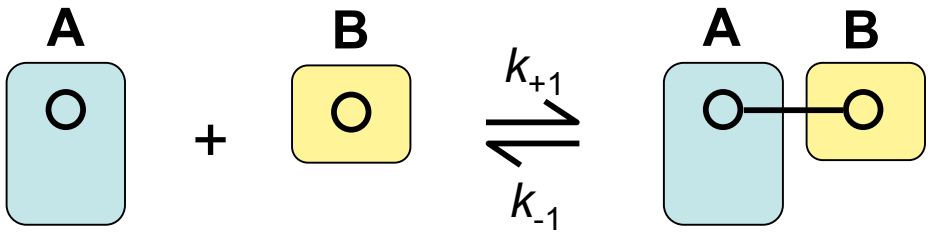
BioNetGen language provides explicit representation of molecules and interactions

Molecules are *structured objects* (hierarchical graphs)



BNGL: `A(b,Y1)` `B(a)`

Rules define interactions (graph rewriting rules)



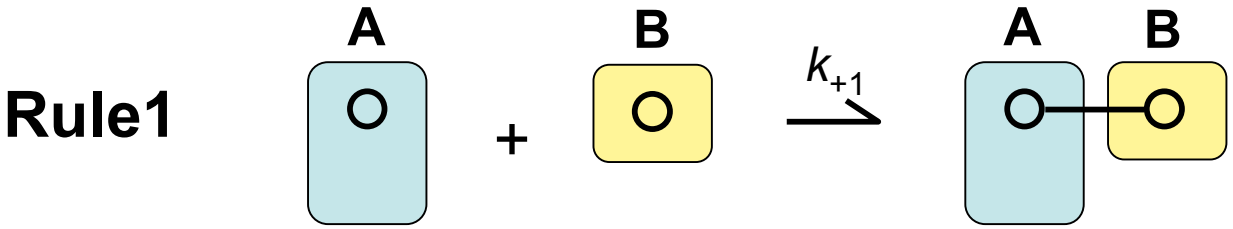
BNGL: `A(b) + B(a) <-> A(b!1).B(a!1) kp1,km1`

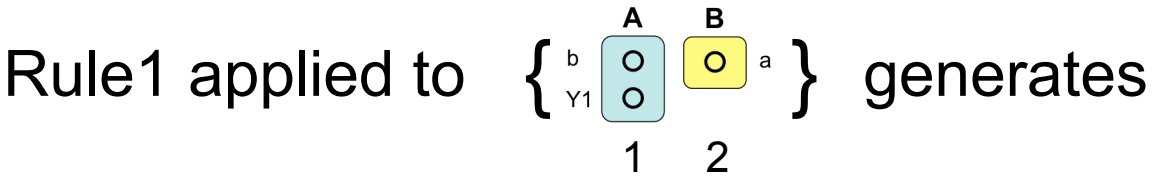
a bond between two components

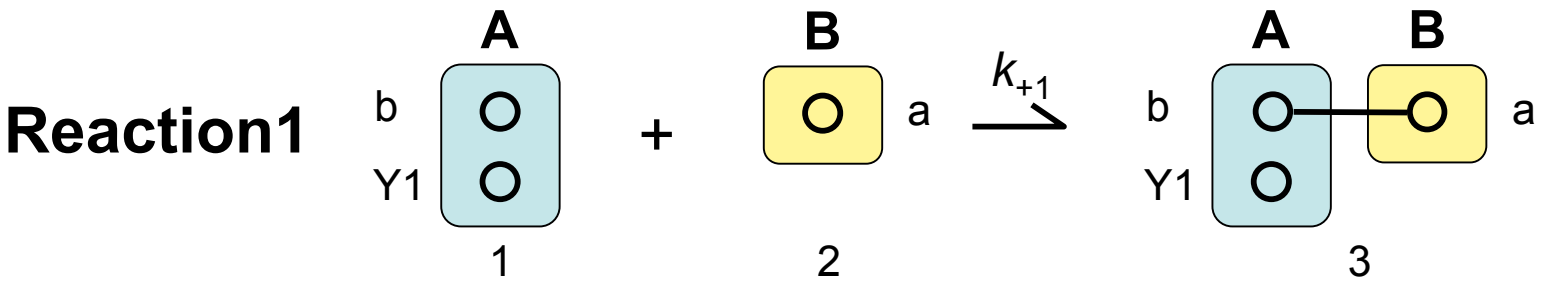
Faeder et al., *Proc. ACM Symp. Appl. Computing* (2005)

Rules generate events

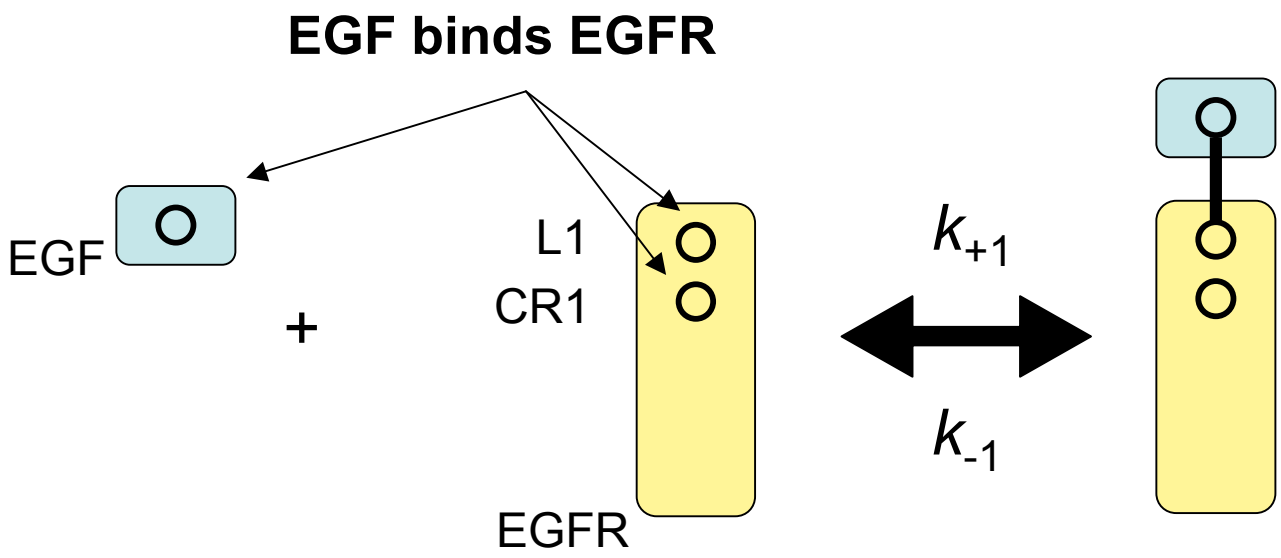
Example of reaction generation:







Reaction rules, composed of patterns, generalize reactions



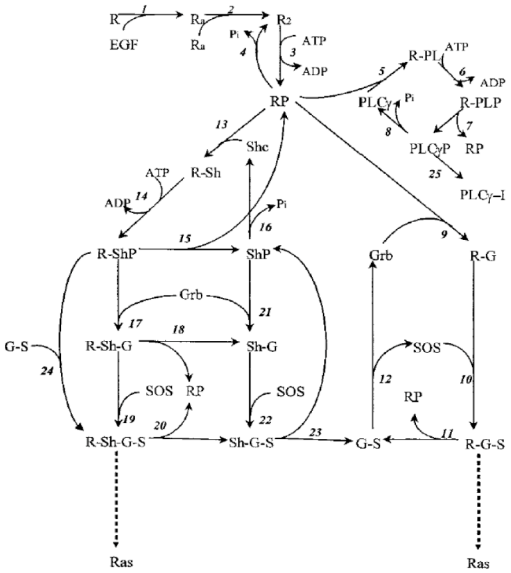
Patterns select reactants and specify graph transformation

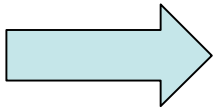
- **Addition of bond between EGF and EGFR**

Rule-based version of the Kholodenko model

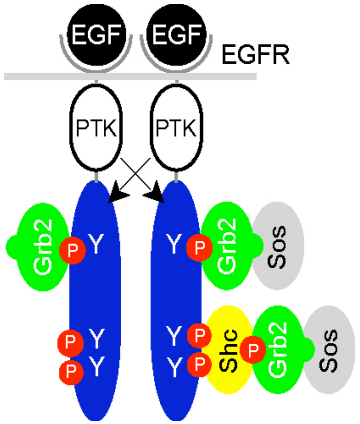
- 5 molecule types
- 23 reaction rules
- No new rate parameters (!)

18 species
34 reactions





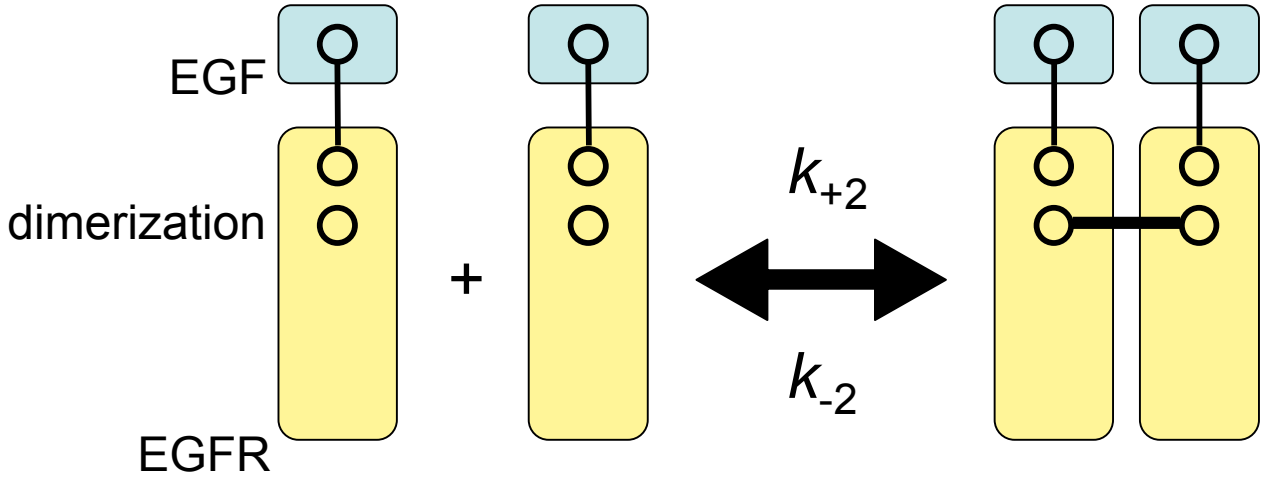
356 species
3749 reactions



Blinov et al. *Biosystems* **83**, 136 (2006).

Dimerization rule eliminates previous assumption restricting breakup of receptors

EGFR dimerizes (600 reactions)

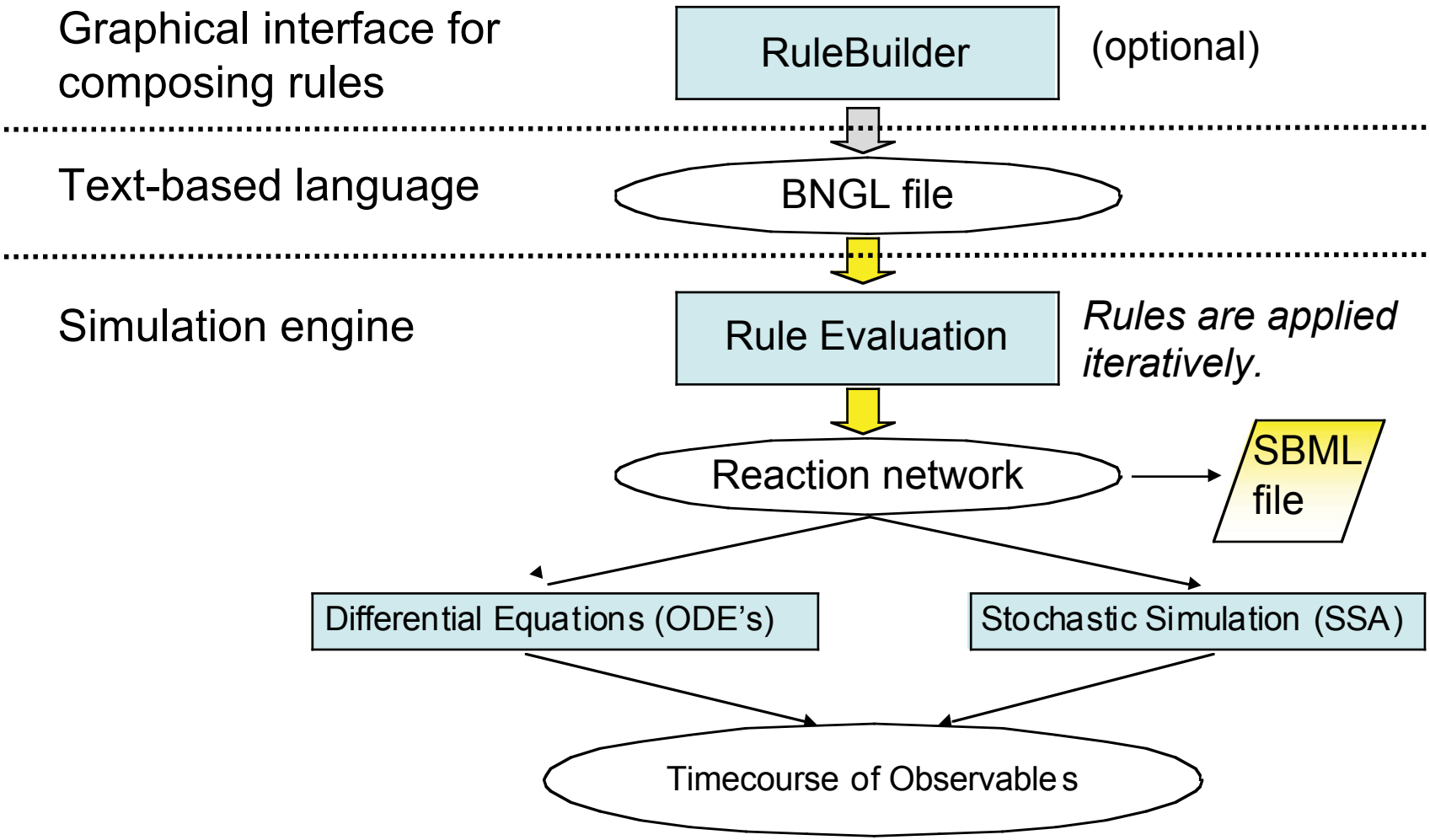


Dimers form and break up independent of phosphorylation of cytoplasmic domains

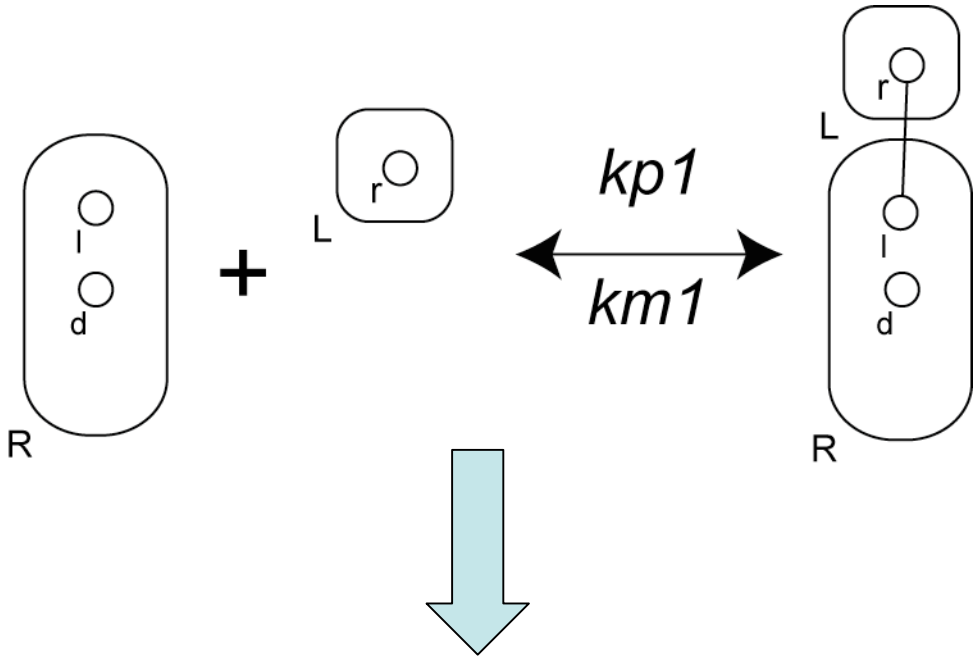
Outline

1. The biochemistry of cell signaling and combinatorial complexity
2. The conventional approach to modeling
3. The rule-based approach to modeling
4. **Tools**
5. New simulation methods

BioNetGen2: Software for graphical rule-based modeling

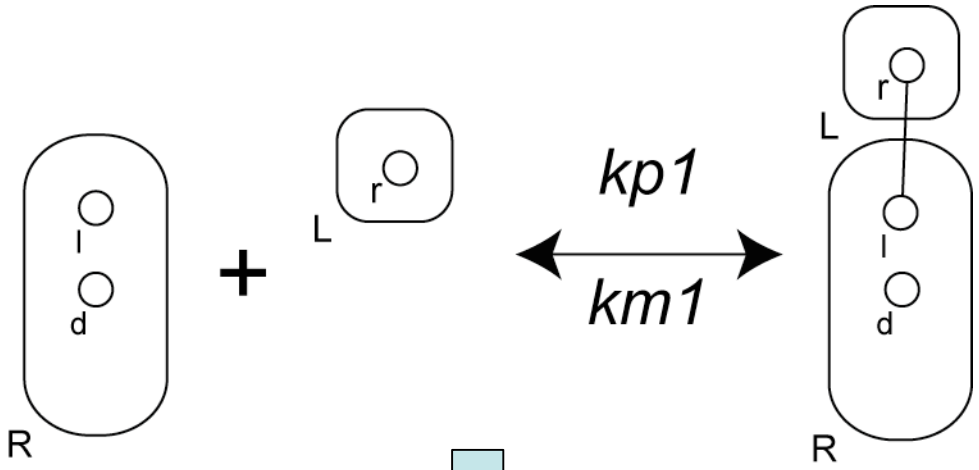


BNGL: A textual language for graphical rules




```
L(r) + R(l,d) <-> L(r!1).R(l!1,d) kp1, km1
```

BNGL: A textual language for graphical rules

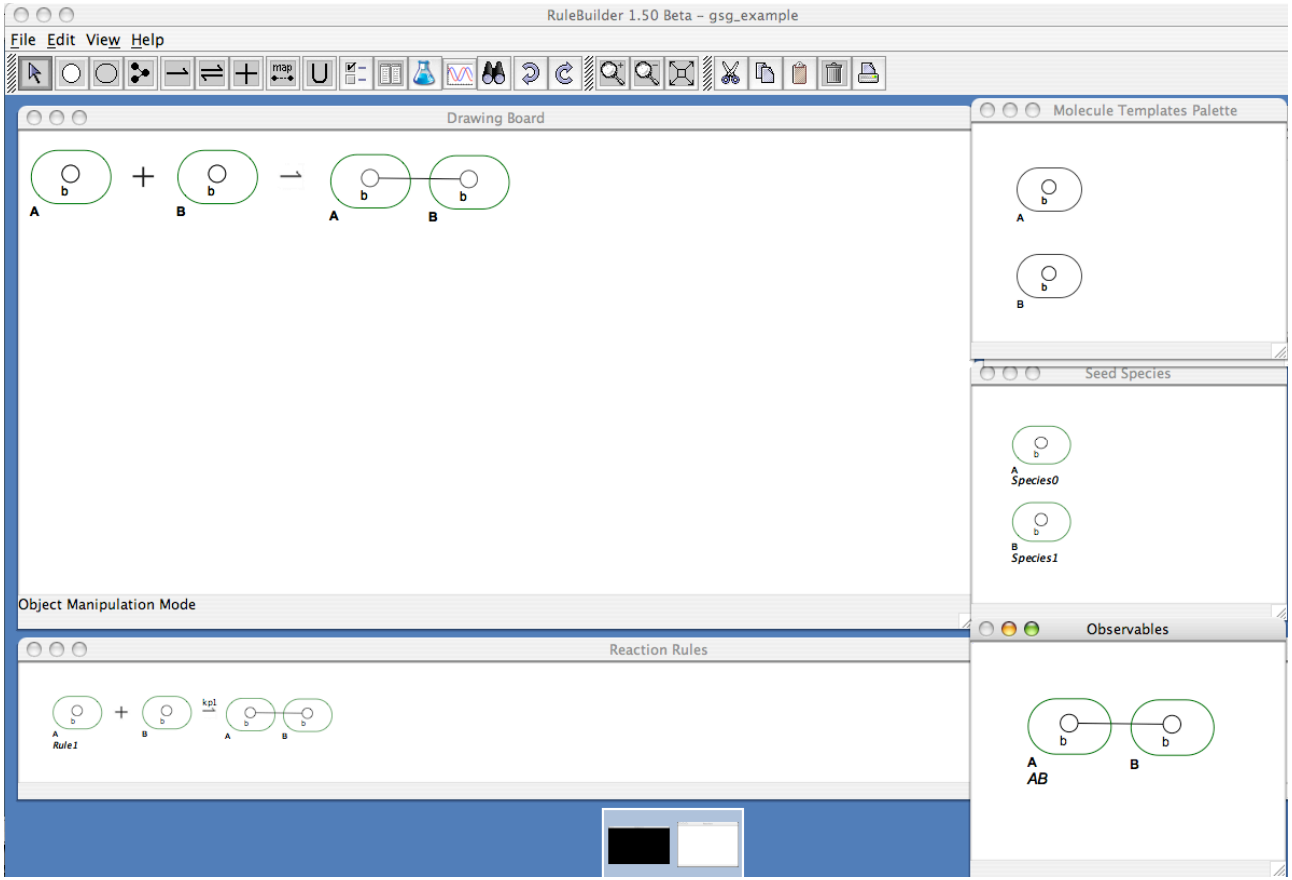


reactant patterns **product pattern** **rate law(s)**

```
L(r) + R(l,d) <-> L(r!1).R(l!1,d) kp1, km1
```

molecule **components (unbound)** **a bond**

Graphical Interface to BioNetGen

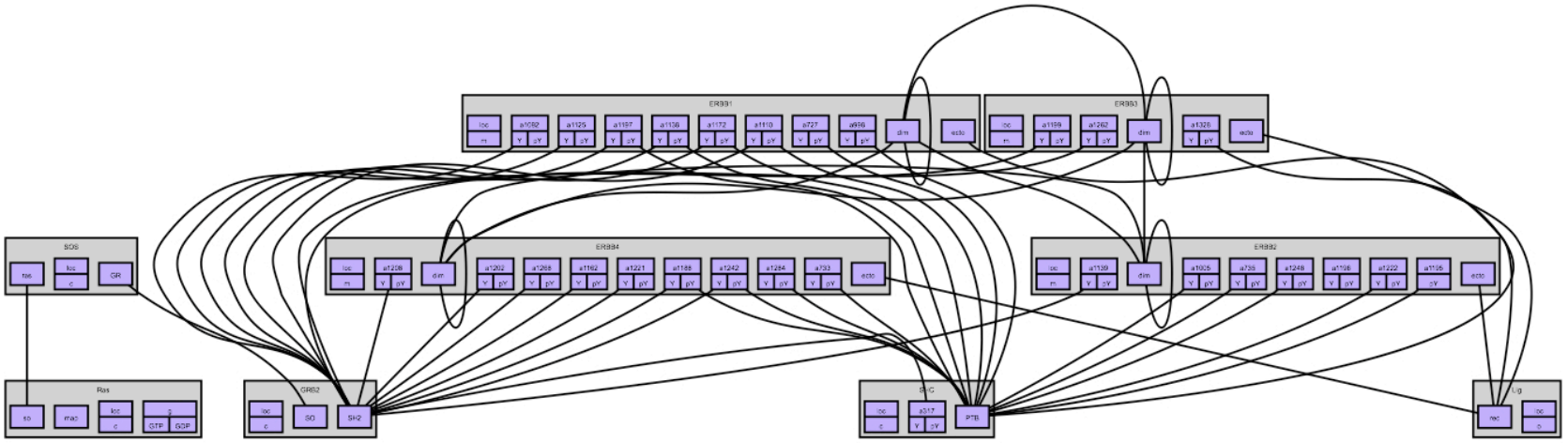


Greatly simplifies construction, visualization, and simulation of complex models

We can take advantage of collective intelligence to build large-scale models

One model currently under construction incorporates approximately 20 proteins involved in EGFR signaling
(EGF, HRG, EGFR, ErbB2, ErbB3, ErbB4, Shc, Grb2, Sos1, Gab1, PI3K, Akt, Ras, Raf, MEK, ERK)

And approximately 1,000 annotated rules capturing the site-specific details of protein-protein interactions



Outline

1. The biochemistry of cell signaling and combinatorial complexity
2. The conventional approach to modeling
3. The rule-based approach to modeling
4. Tools
5. **New simulation methods**

Rule-based models can be difficult to simulate

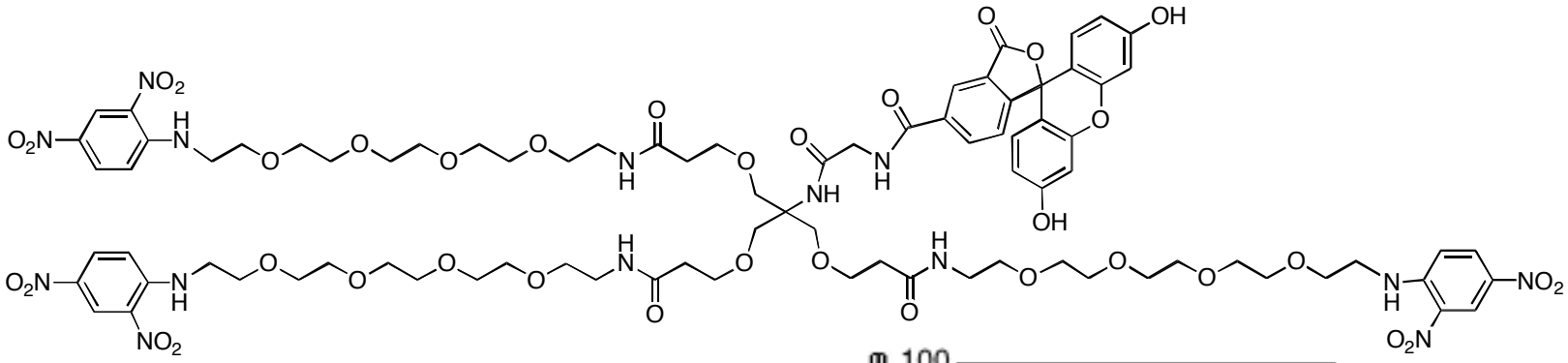
Rule-based models may encompass a large or even an unbounded number of species

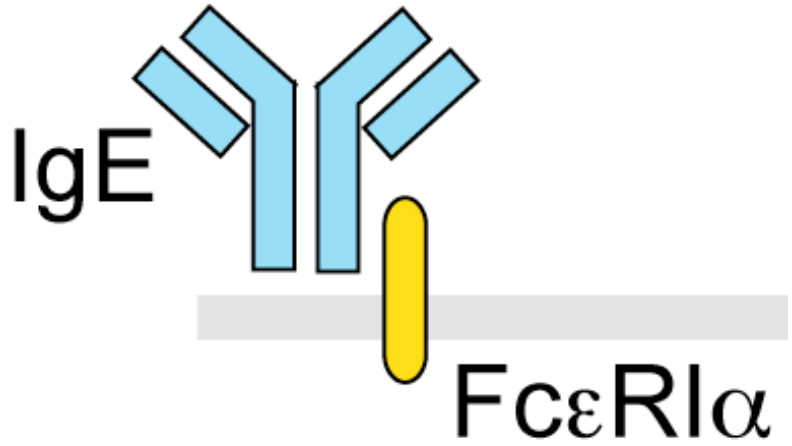
Computational costs for standard simulation methods increase with number of species and reactions in a model

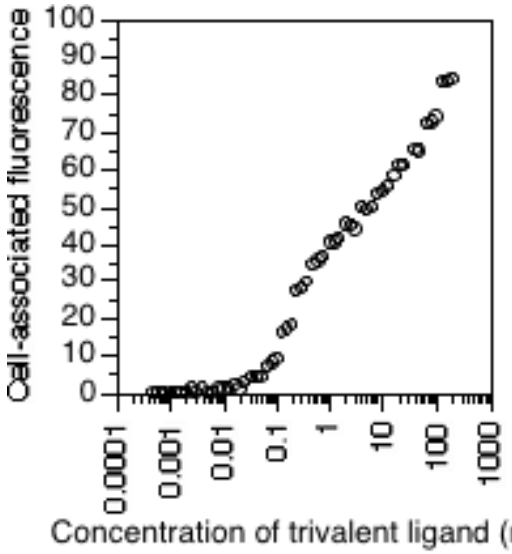
Parameter estimation and data fitting require running model simulations for a large number of parameter sets

We need a simulation method that is independent of the size of the reaction network implied by rules

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





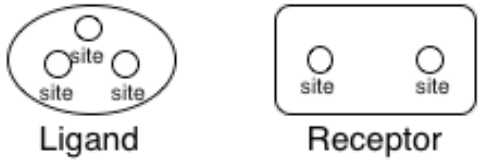


R.G. Posner (TGen)

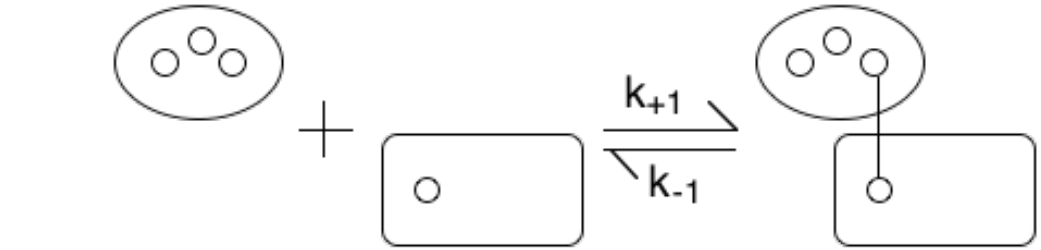
Rule-based model specification corresponding to equilibrium model of Goldstein and Perelson (1984)

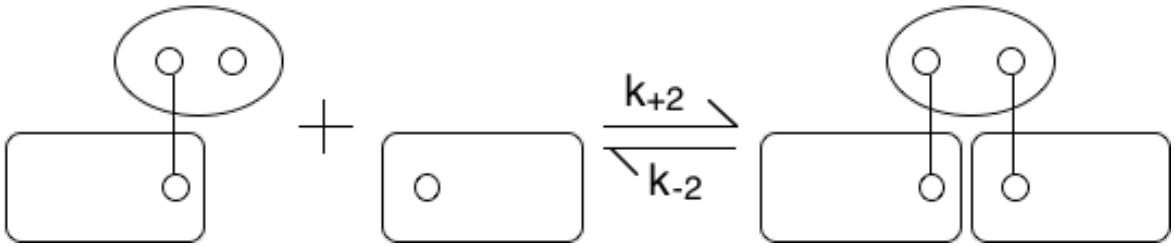
Equivalent-site model

Molecules

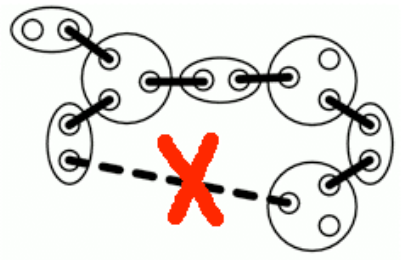


Interactions (reaction rules)





No cyclic aggregates

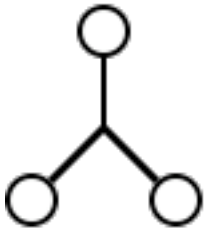


Generate-first method of simulation

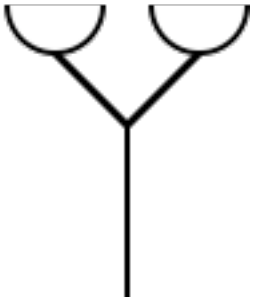
1. Define seed species
2. Determine if a pattern in a rule matches any species

   If so, apply the transformation defined in the rule
3. Iteratively apply rules to new product species
4. Simulate using conventional methods once network has been generated

Seed species

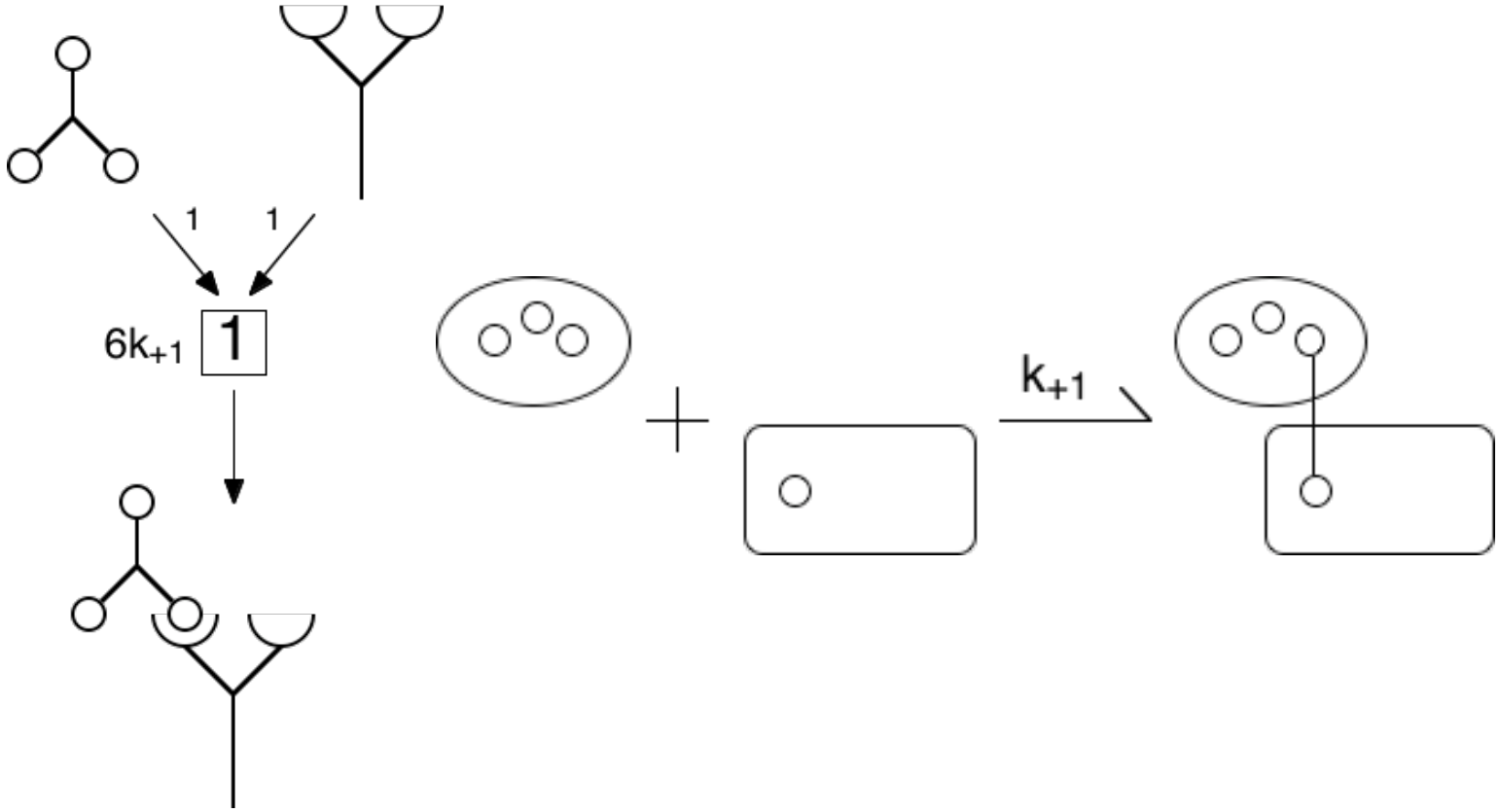


Ligand

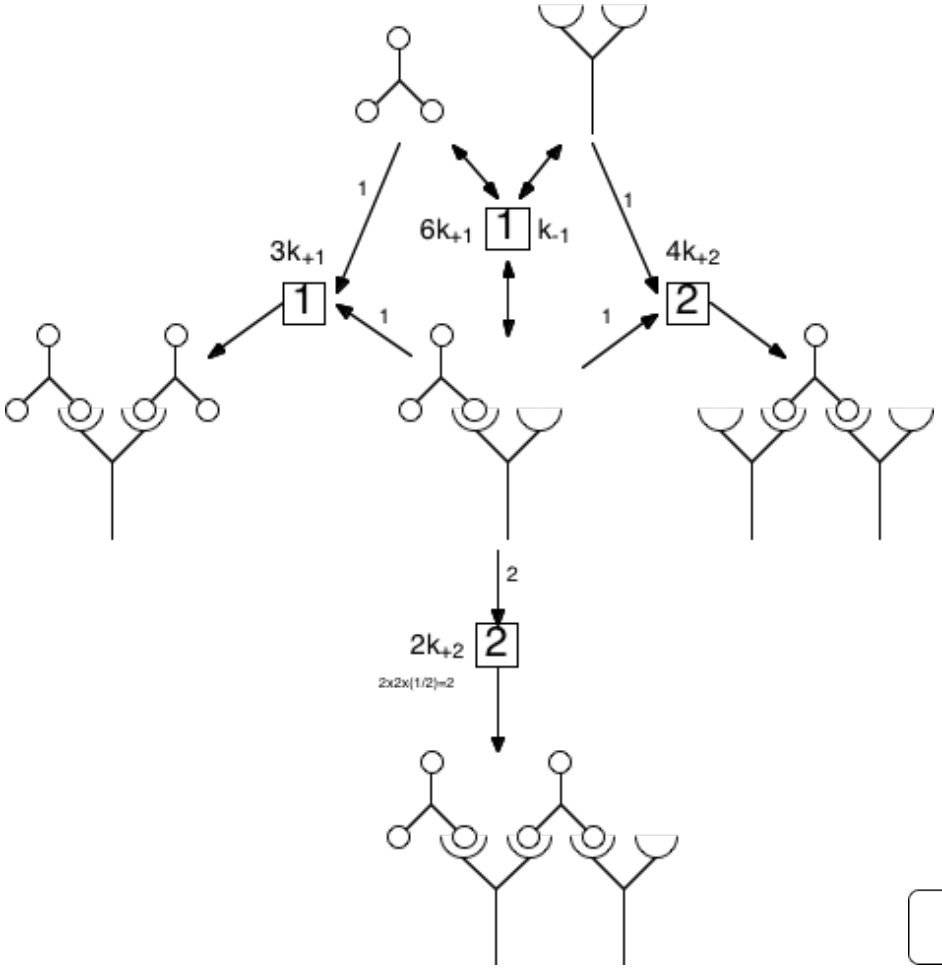


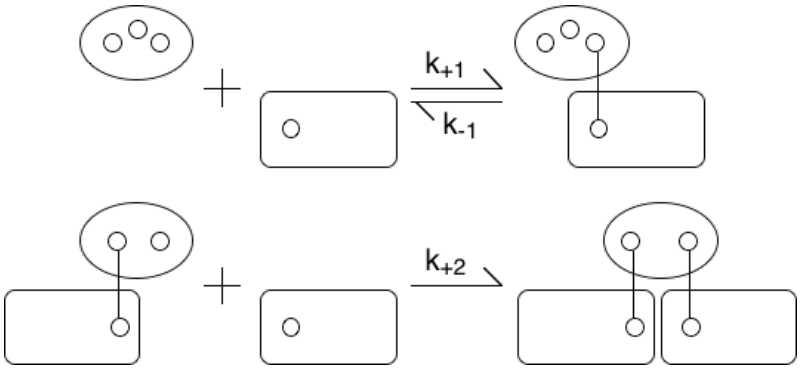
Receptor

After first round of rule application

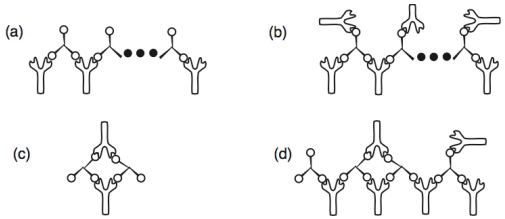


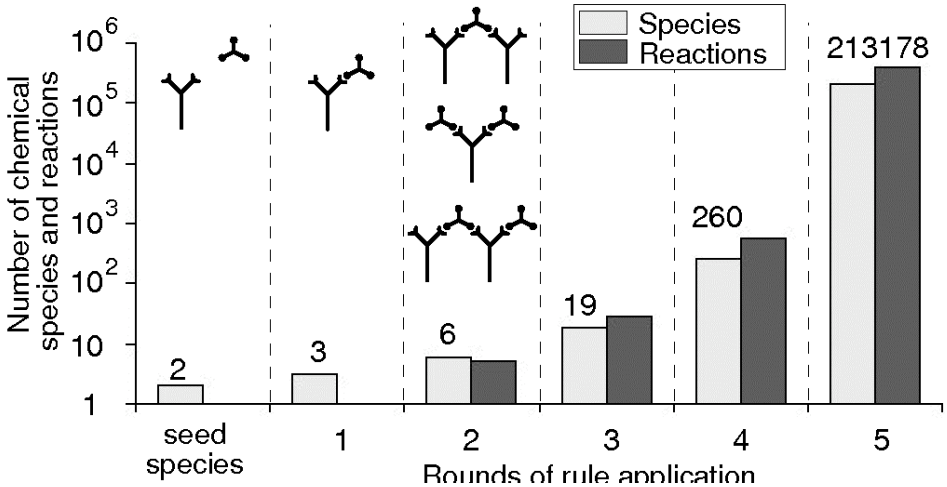
After the second round of rule application



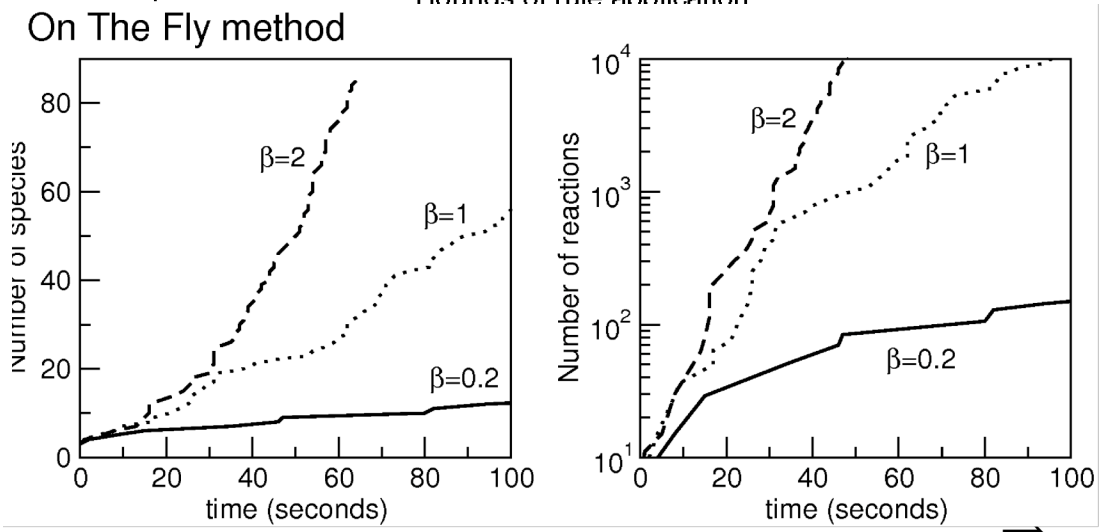


Rule-derived network is too large to simulate using conventional population-based methods





Two rules generate a vast number of chemical species and reactions



⇒

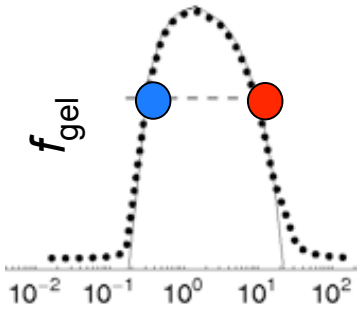
Rule-based KMC Method

(Particle-based version of Gillespie's Direct Method with rules)

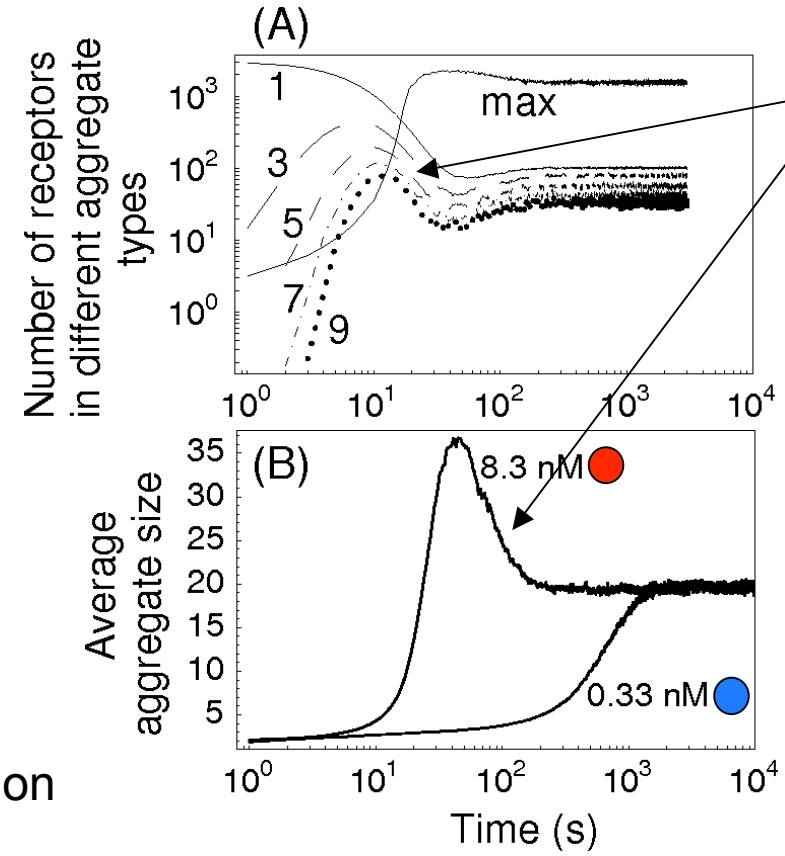
1. Instantiate molecules with components and states.
2. Determine cumulative rate for each reaction type, a_m
3. Select next reaction time, $\Delta t = -\ln\left(r_1\right) / a_{\text{tot}}$
4. Select next reaction type using $\sum_{j=1}^{J-1} a_j < r_2 a_{\text{tot}} \leq \sum_{j=1}^{J} a_j$
5. Select reactant molecules and components.
6. Update reaction type rates. Iterate.

Kinetics of aggregate formation

Equilibrium



Ligand concentration



At high ligand concentration, small aggregates form transiently

Performance