Rule-based Modeling of Signal Transduction

Jim Faeder

Theoretical Biology and Biophysics Group (T-10)
Los Alamos National Laboratory

q-bio Summer School
July 25, 2007

Los Alamos NATIONAL LABORATORY EST.1943

Center for Nonlinear Studies

NNSA

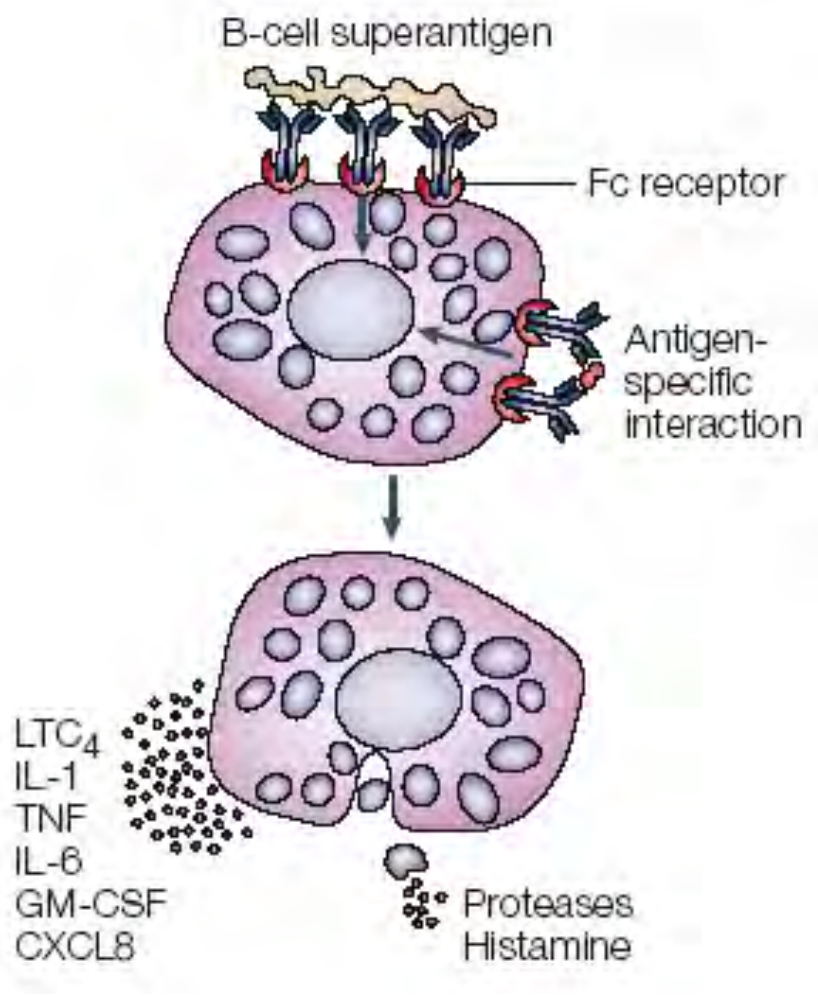
Cell signaling - *cellular information processing* - is critical to the survival of all organisms and plays a critical role in human health and disease.

Our goal is to develop predictive models of cell signaling to better understand and control these processes.

Cell signaling in allergic responses

Stimulus

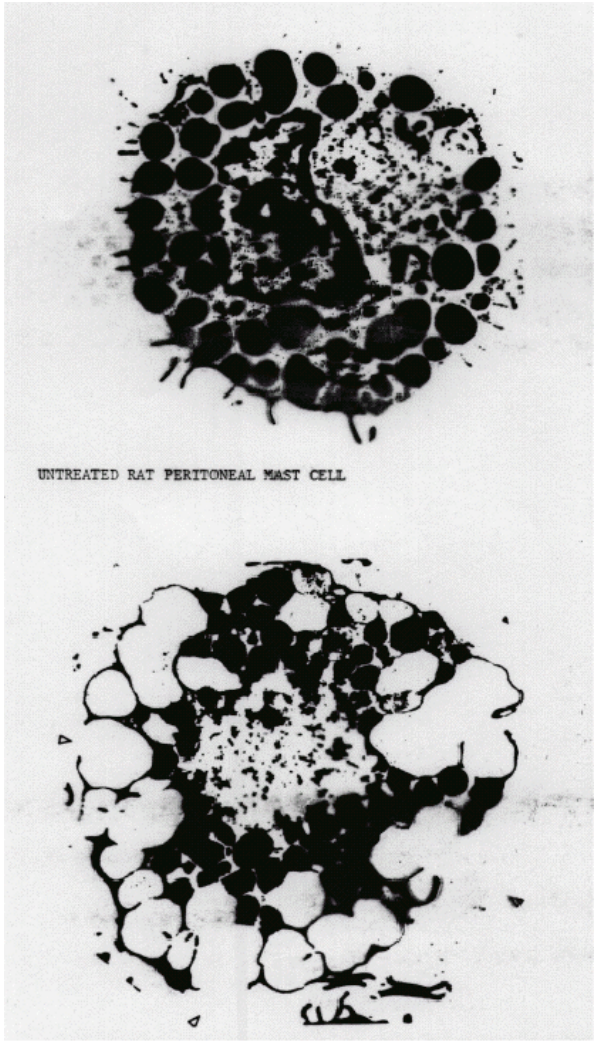
Response



Marshall, *Nat. Rev. Immunol*. (2004)

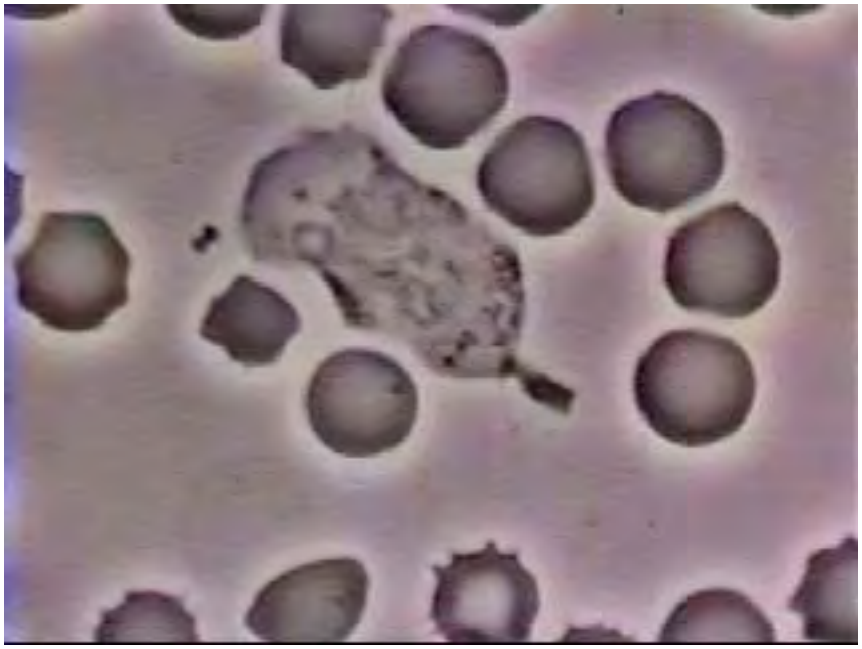
Mast cell “degranulation” is a critical component of many allergic responses

Mast cell at rest



3 min. after exposure to allergen

Movie: An immune cell in action



Original film from David Rogers (Vanderbuilt University)

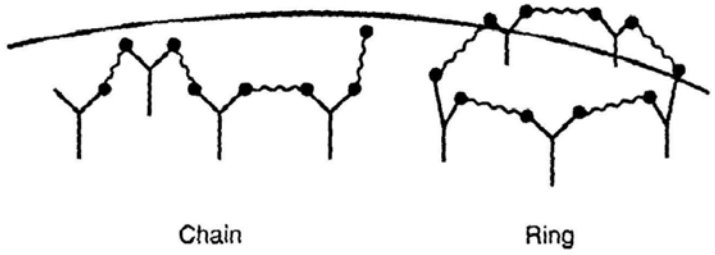
http://www.biochemweb.org/fenteany/research/cell_migration/neutrophil.html

Goals

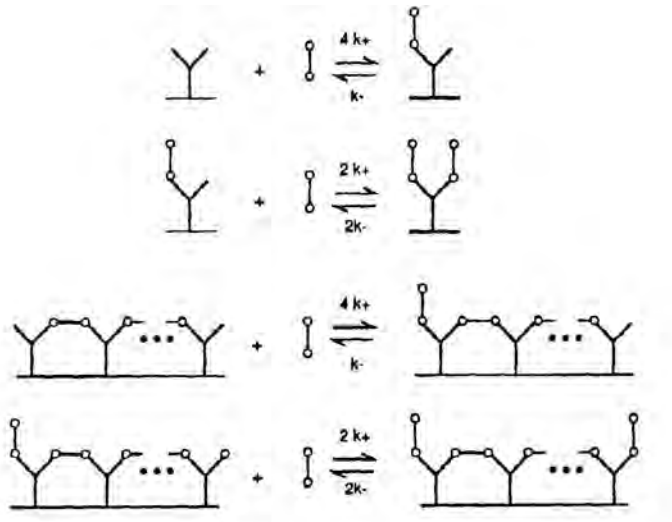
- Predictive understanding
  - Different stimulation conditions
  - Protein expression levels
  - Manipulation of protein modules
  - Site-specific inhibitors
- Mechanistic insights
  - Why do signal proteins contain so many diverse elements?
- Drug development
  - New targets
  - Combination therapies

Los Alamos approach to modeling: The past (70s-90s)

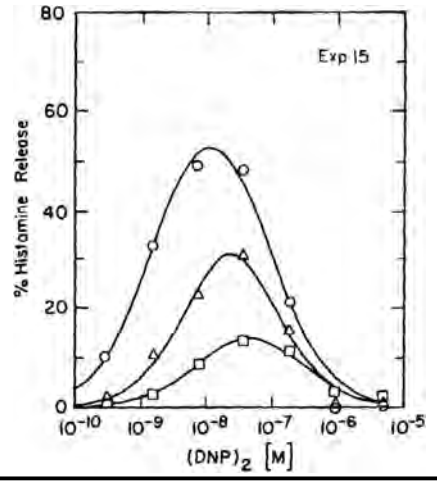
1. Multivalent binding process



2. Rules that define network

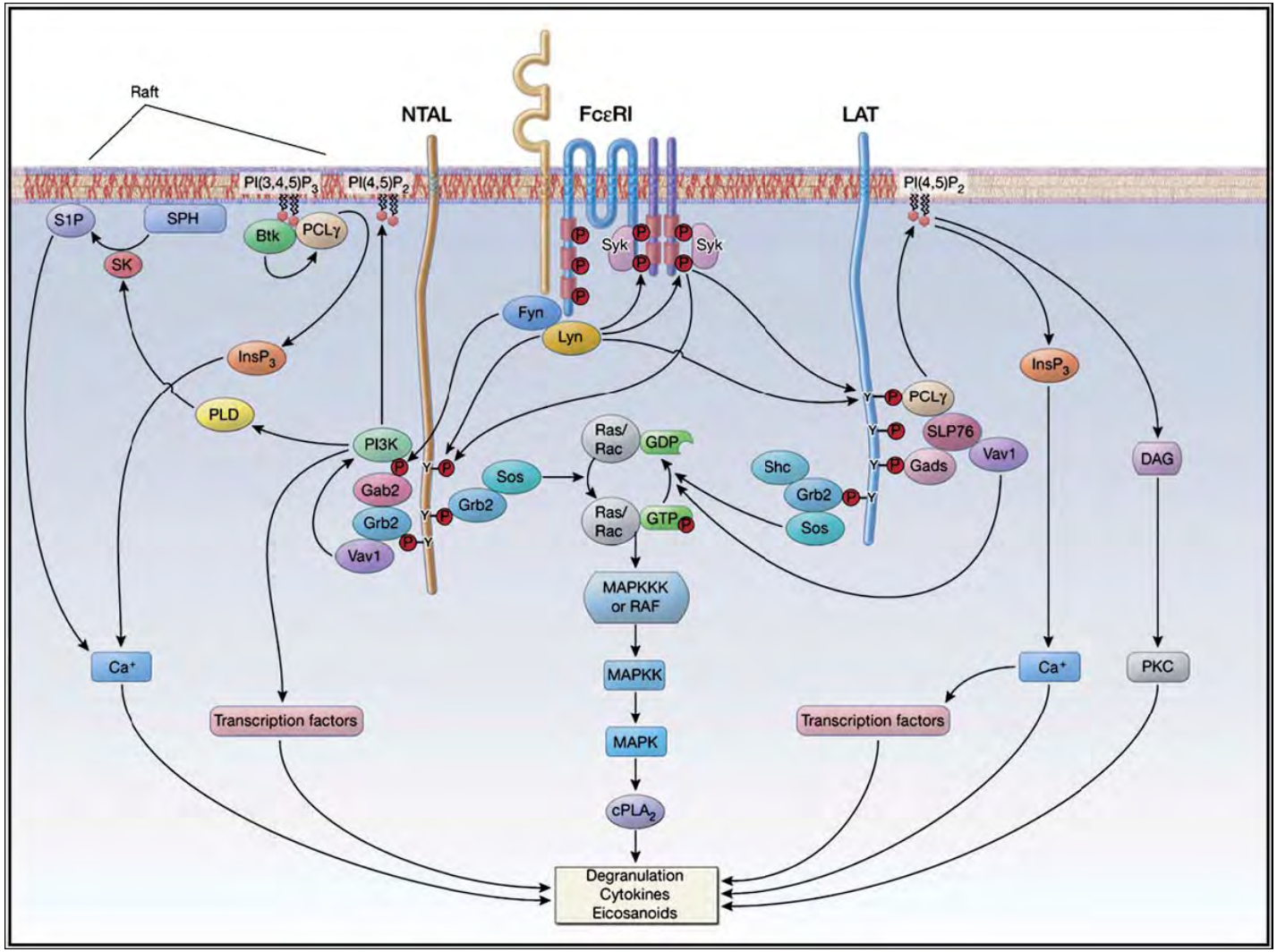


3. Quantitative predictions



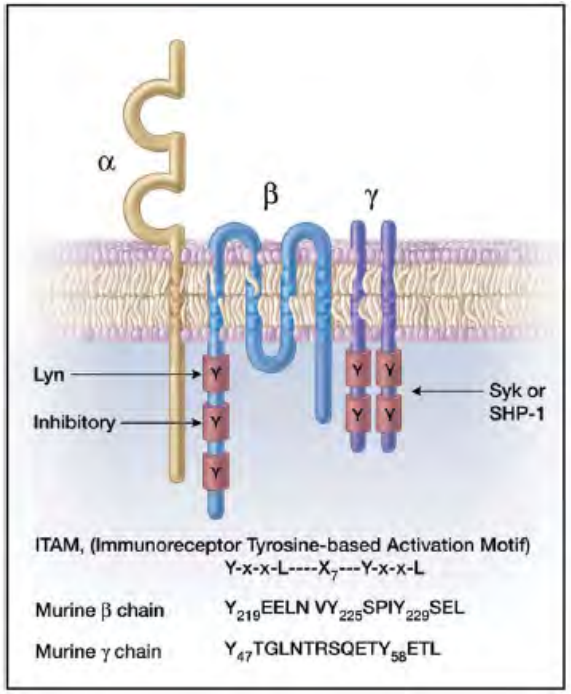
B. Goldstein, in *Theoretical Immunology, Part One*, Ed. A. S. Perelson

Toward models of intracellular signaling



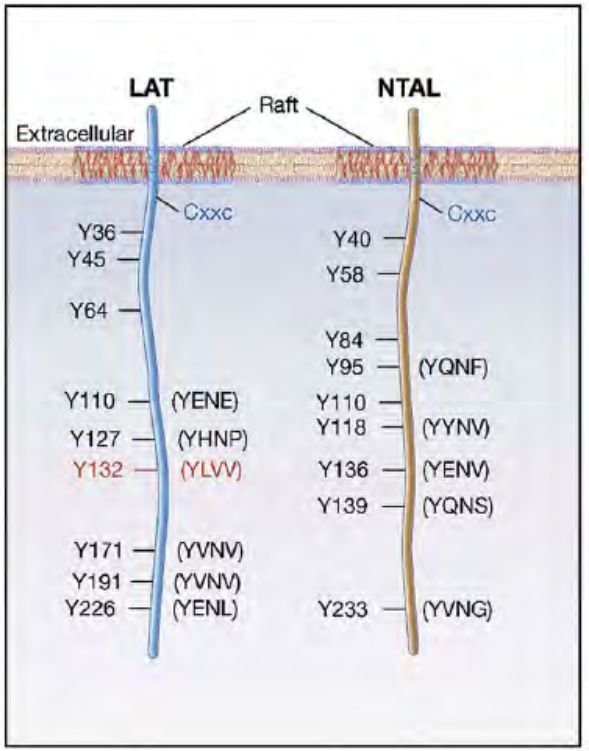
J. Rivera and A. Gilfillan, *J. Allergy Clin. Immunol.***117**, 1216 (2006).

Modularity of signaling proteins



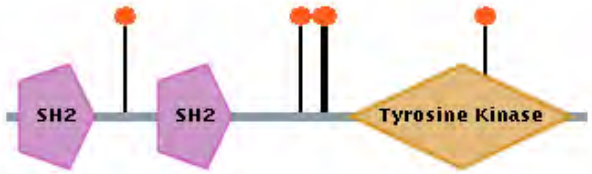
FcεRI

Lyn

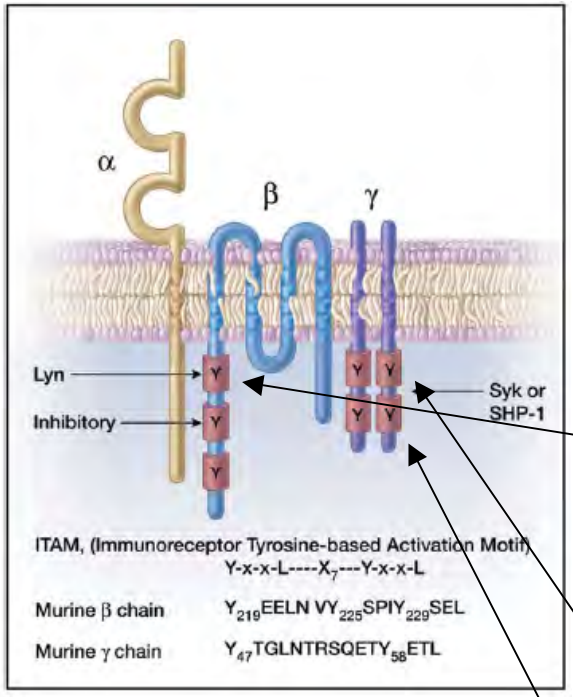


Transmembrane Adaptors

Syk

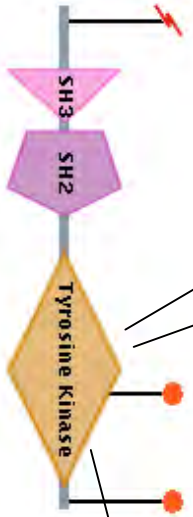


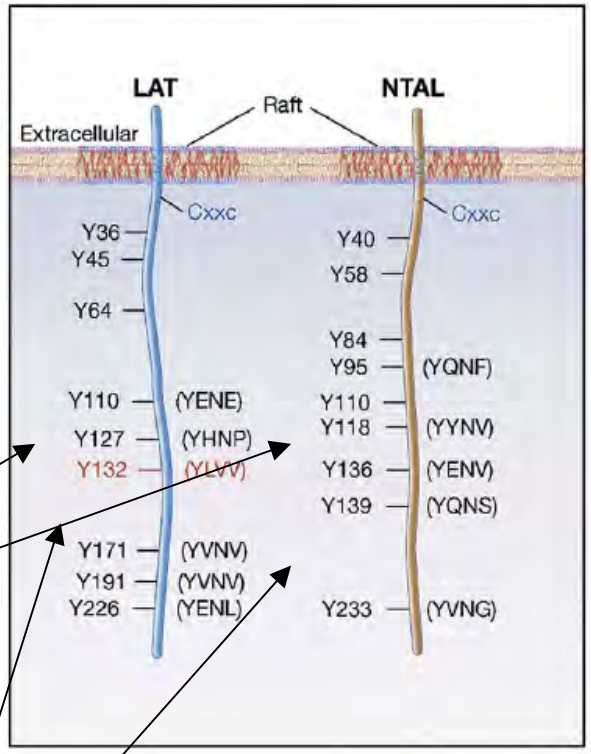
Signaling proteins contain domains and motifs that mediate interactions with other proteins



FcεRI

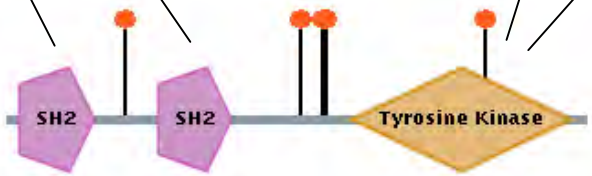
Lyn





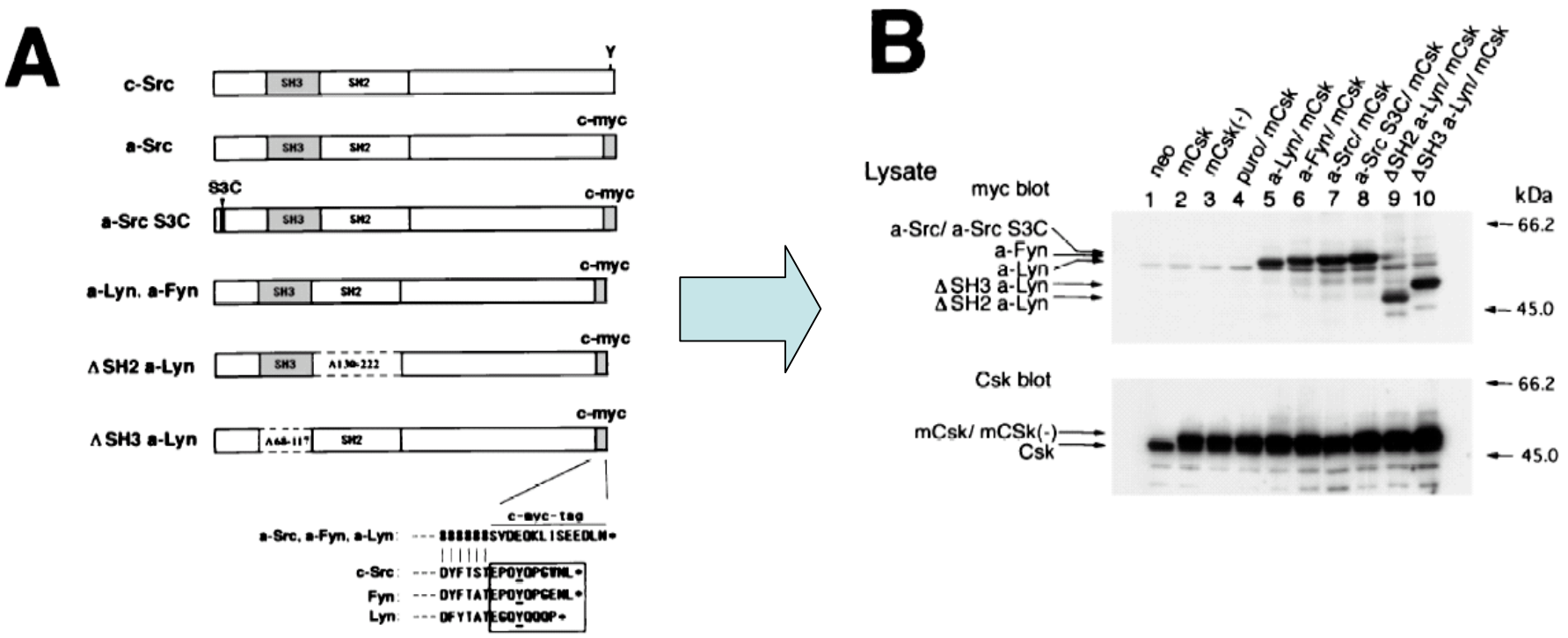
Transmembrane Adaptors

Syk



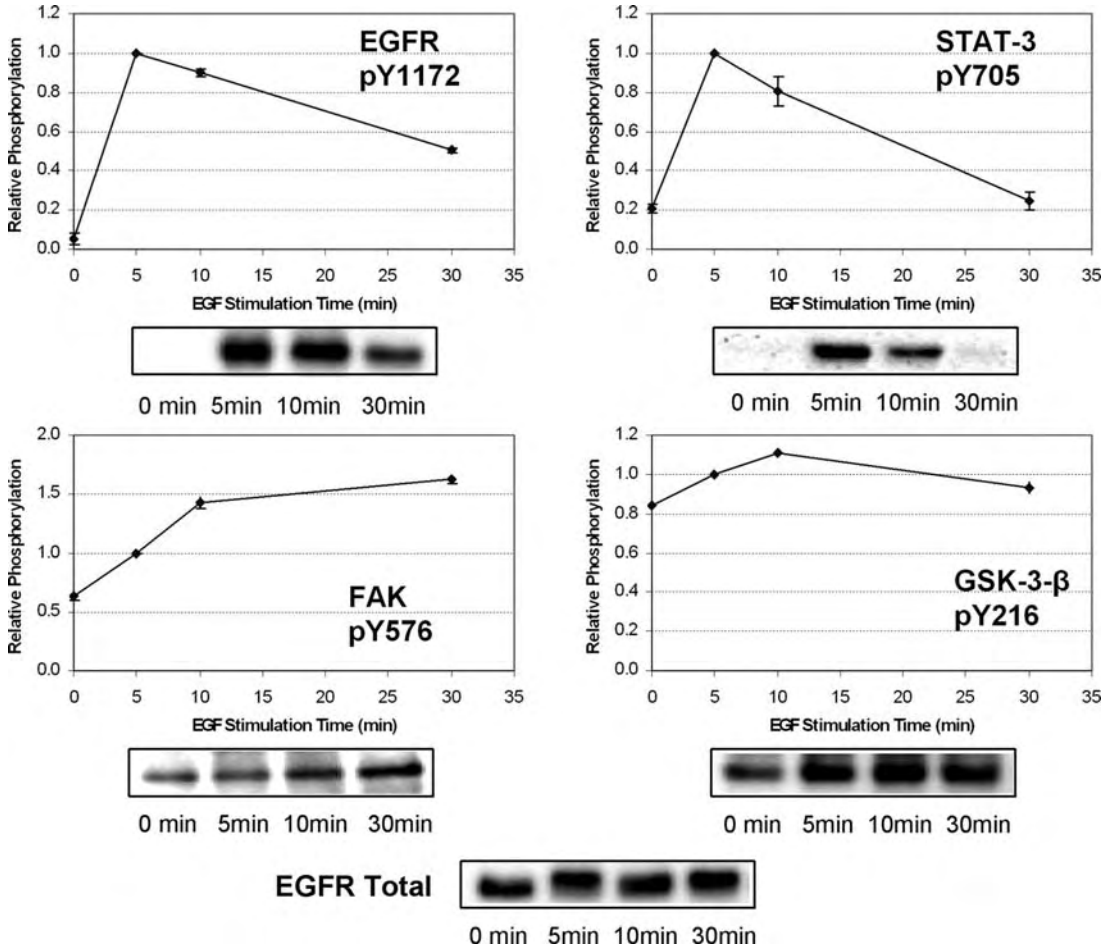
Experiments probe the function of protein modules

These modules may mediate both protein-protein and protein-lipid interactions



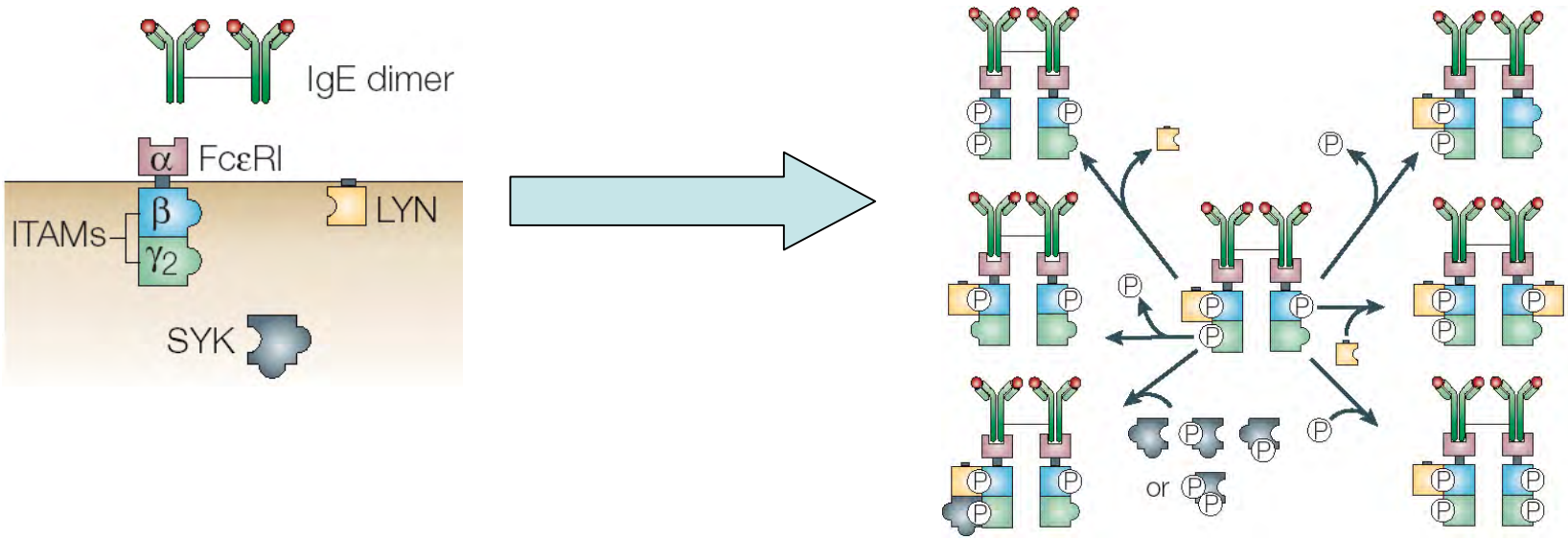
Honda *et al.*, Mol. Cell. Biol. (2000), **20**, 1759.

Experiments probe the kinetics of multiple phosphorylation sites



Zhang et al., *Mol. Cell. Proteomics* **4**, 1240 (2005).

Early IgE receptor signaling exhibits combinatorial complexity



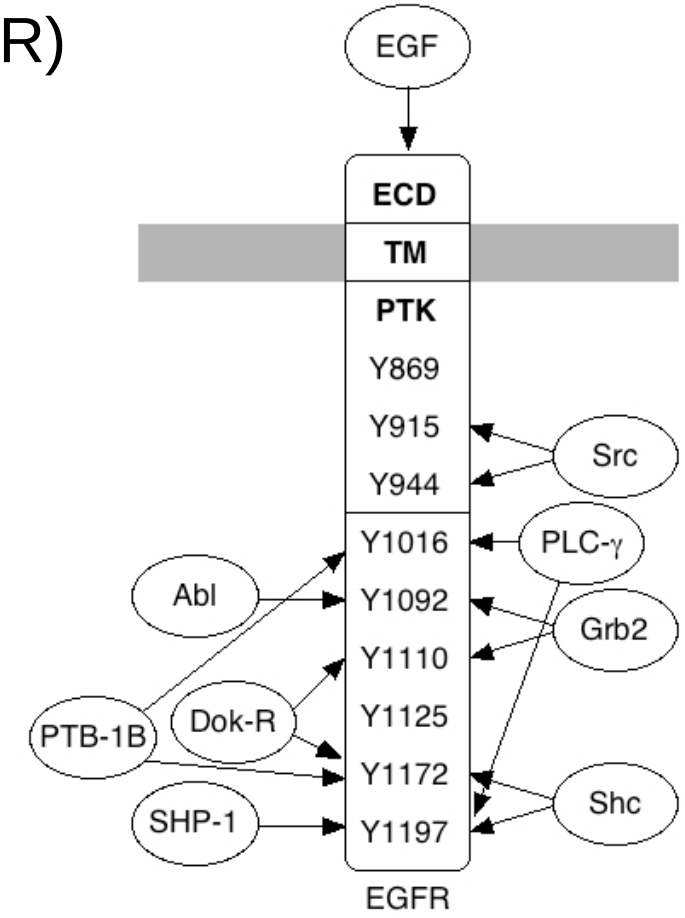
354 species / 3680 reactions

Combinatorial complexity = small number of components and interactions gives rise to a large network of species and reactions

Goldstein et al. *Mol. Immunol.* (2002) ; Faeder et al. *J. Immunol.* (2003)

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)

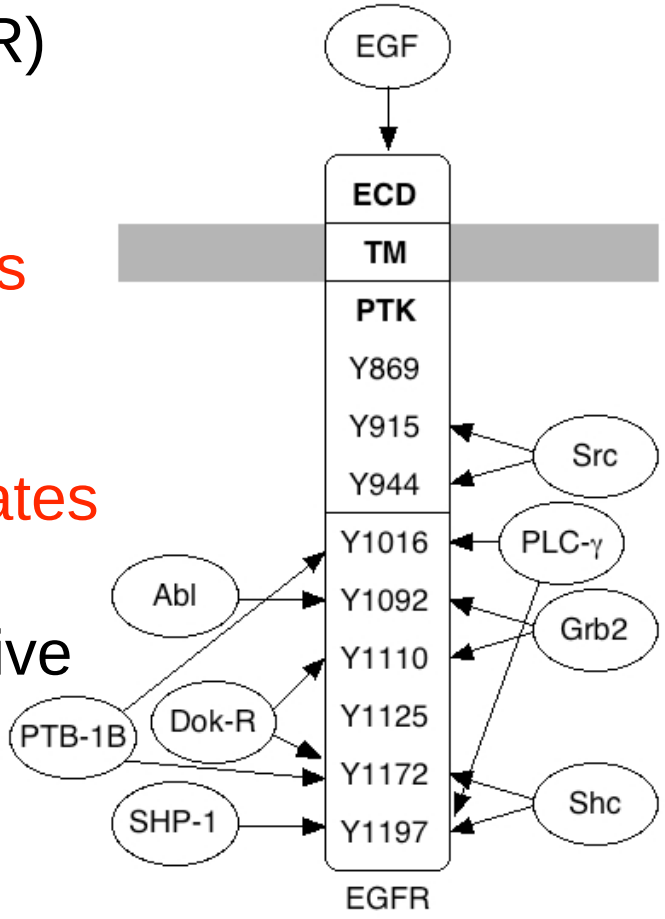
9 sites $\Rightarrow$ 2^9=512 phosphorylation states

Each site has ≥ 1 binding partner

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

EGFR must form *dimers* to become active

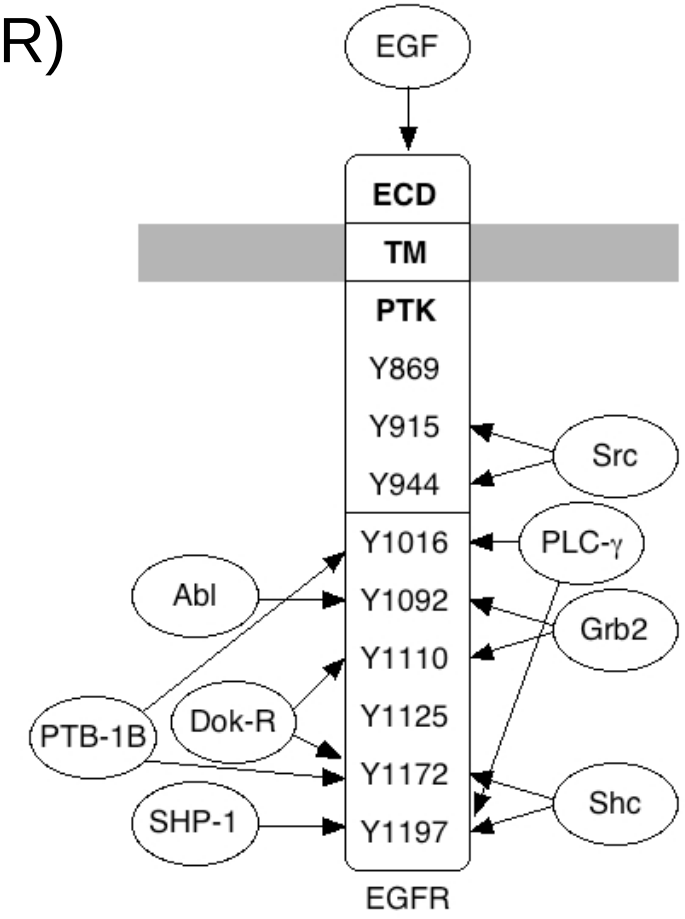
$\Rightarrow$ more than 1.9×10^8 states



Multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

…but the number of interactions is relatively small.



Summary

What functional role do protein domains and motifs play in signaling?

Combinatorial complexity

- *Modularity of protein structure*
- *Multivalent interactions*

BioNetGen language provides explicit representation of molecules and interactions

Molecules are *structured objects* (hierarchical graphs)

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

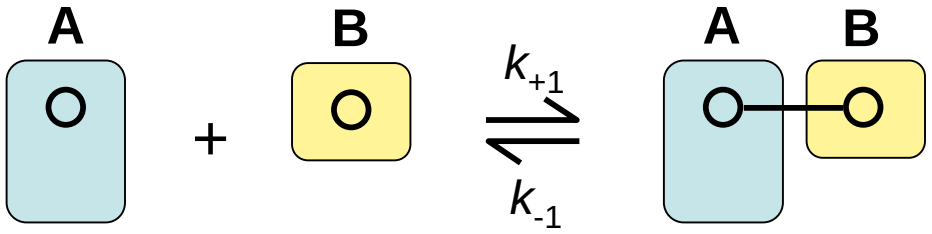
Faeder et al., *Proc. ACM Symp. Appl. Computing* (2005)

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`

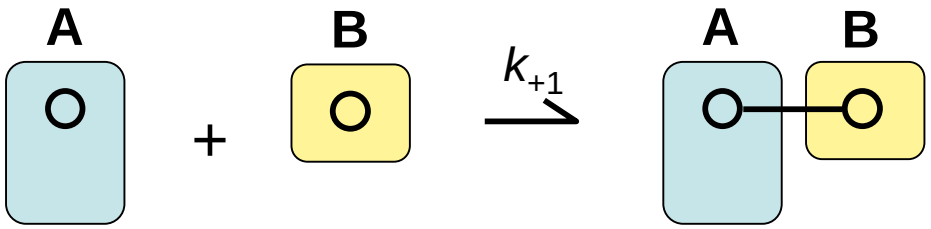
a bond between two components

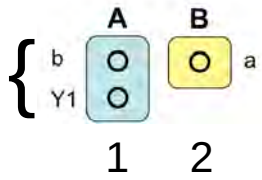
Faeder et al., *Proc. ACM Symp. Appl. Computing* (2005)

Rules generate events

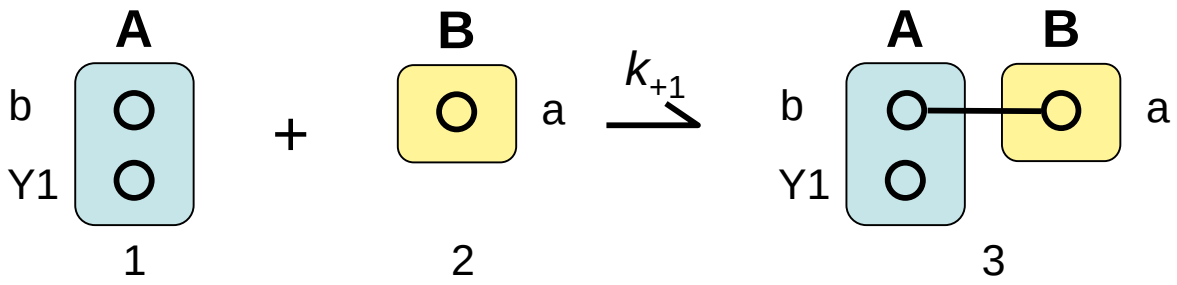
Example of reaction generation:

Rule1



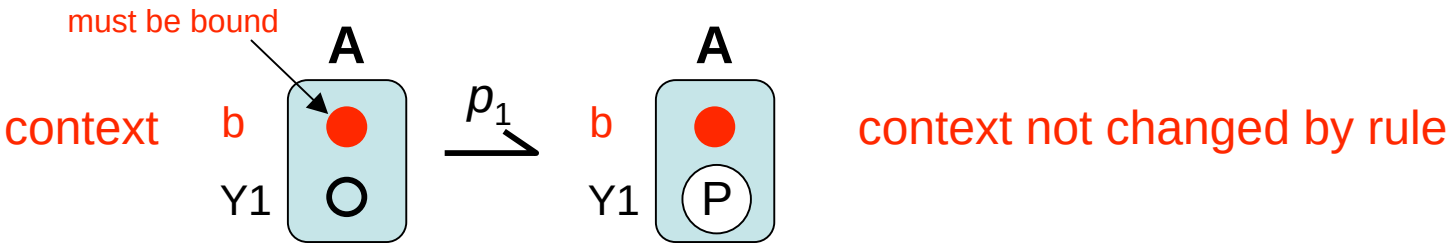
Rule1 applied to {  } generates

Reaction1

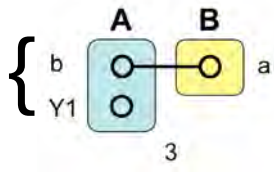


Rules may specify contextual requirements

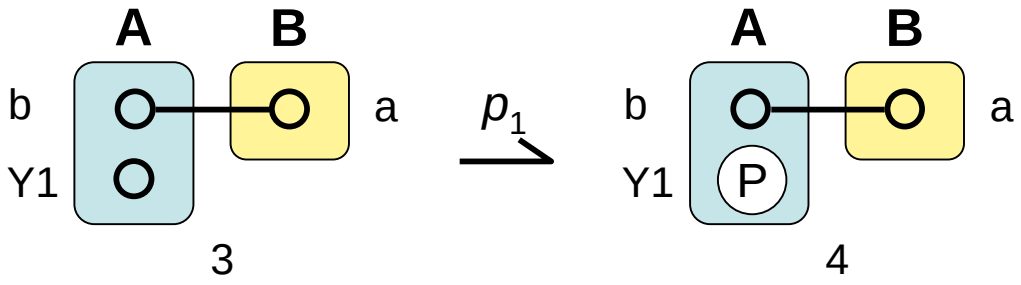
Rule2



BNGL: `A(b!+,Y1~U) <-> A(b!+,Y1~P) p1`

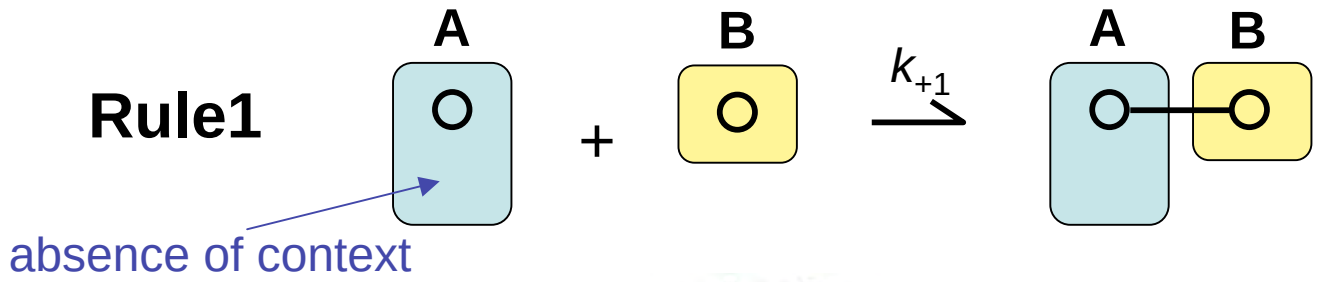
Rule2 applied to {  } generates

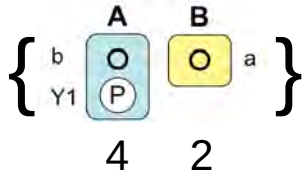
Reaction2

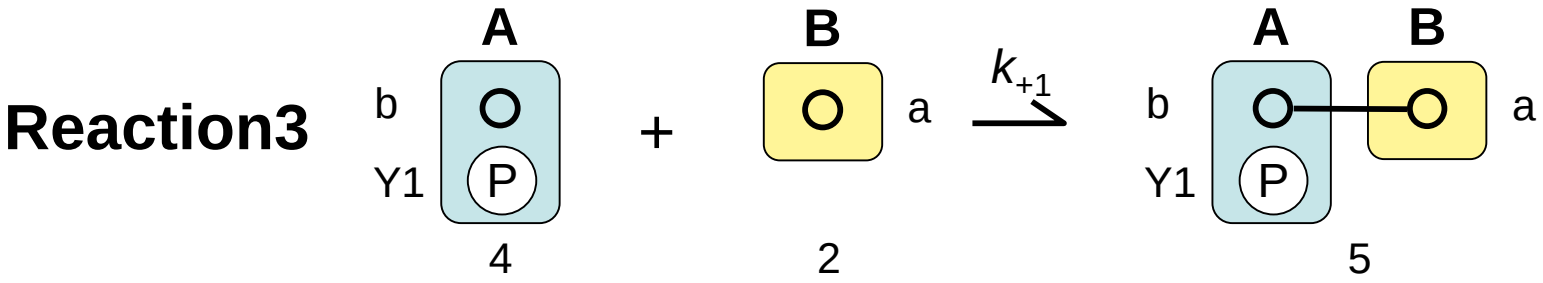


Rules may generate multiple events

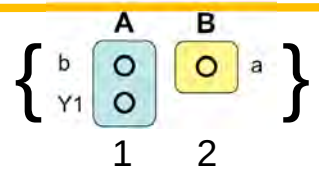
Second example of reaction generation:

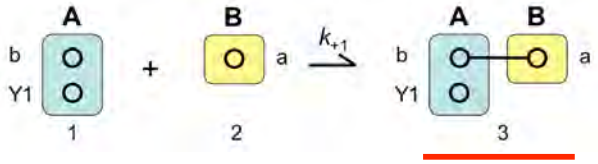


Rule1 applied to  generates

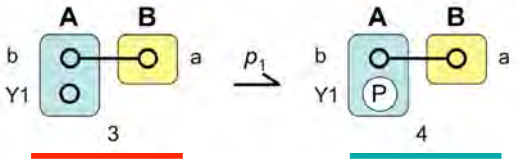


Iterative application of rules generates standard mass action reaction network

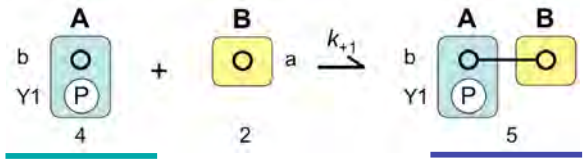
Initial species: 

Iteration 1: Rule1
Rule2 → Reaction1: 

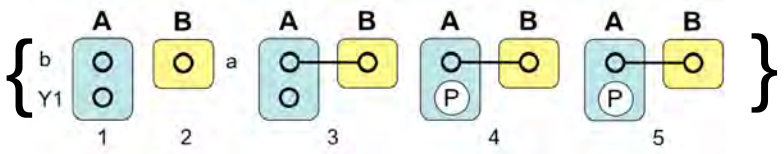
(3 species, 1 reaction)

Iteration 2: Rule1
Rule2 → Reaction2: 

(4 species, 2 reactions)

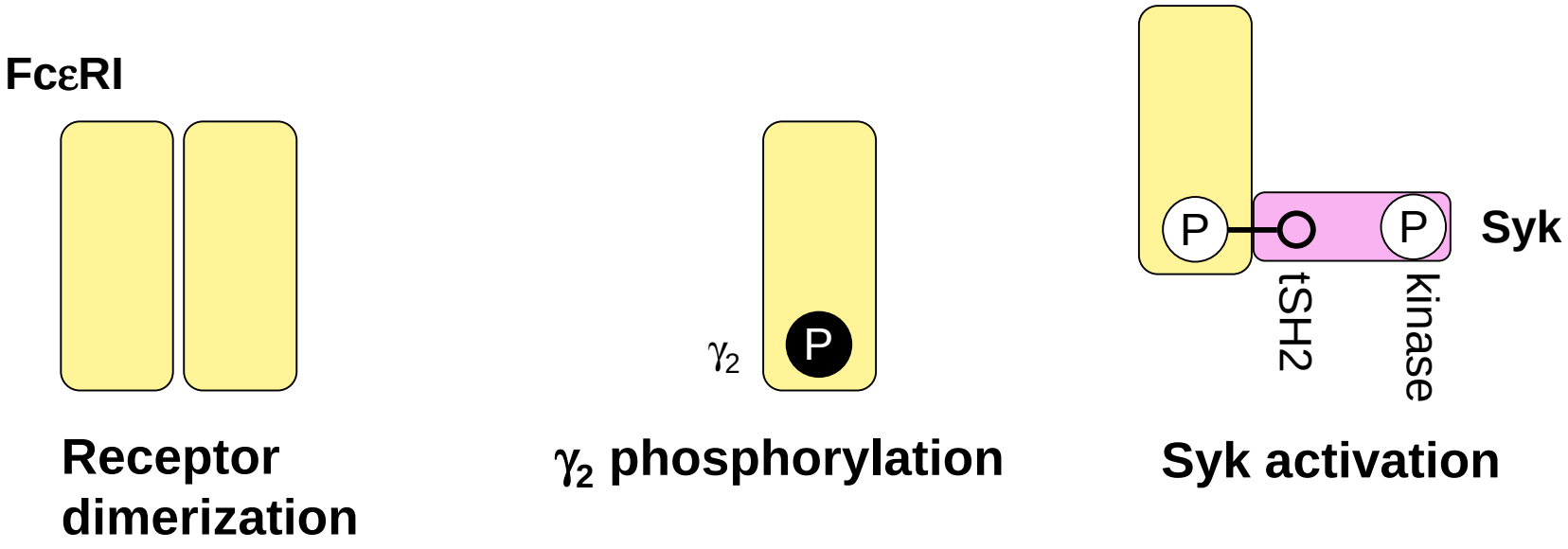
Iteration 3: Rule1
Rule2 → Reaction3: 

(5 species, 3 reactions)

Final species: 

Observables use patterns to define model outputs

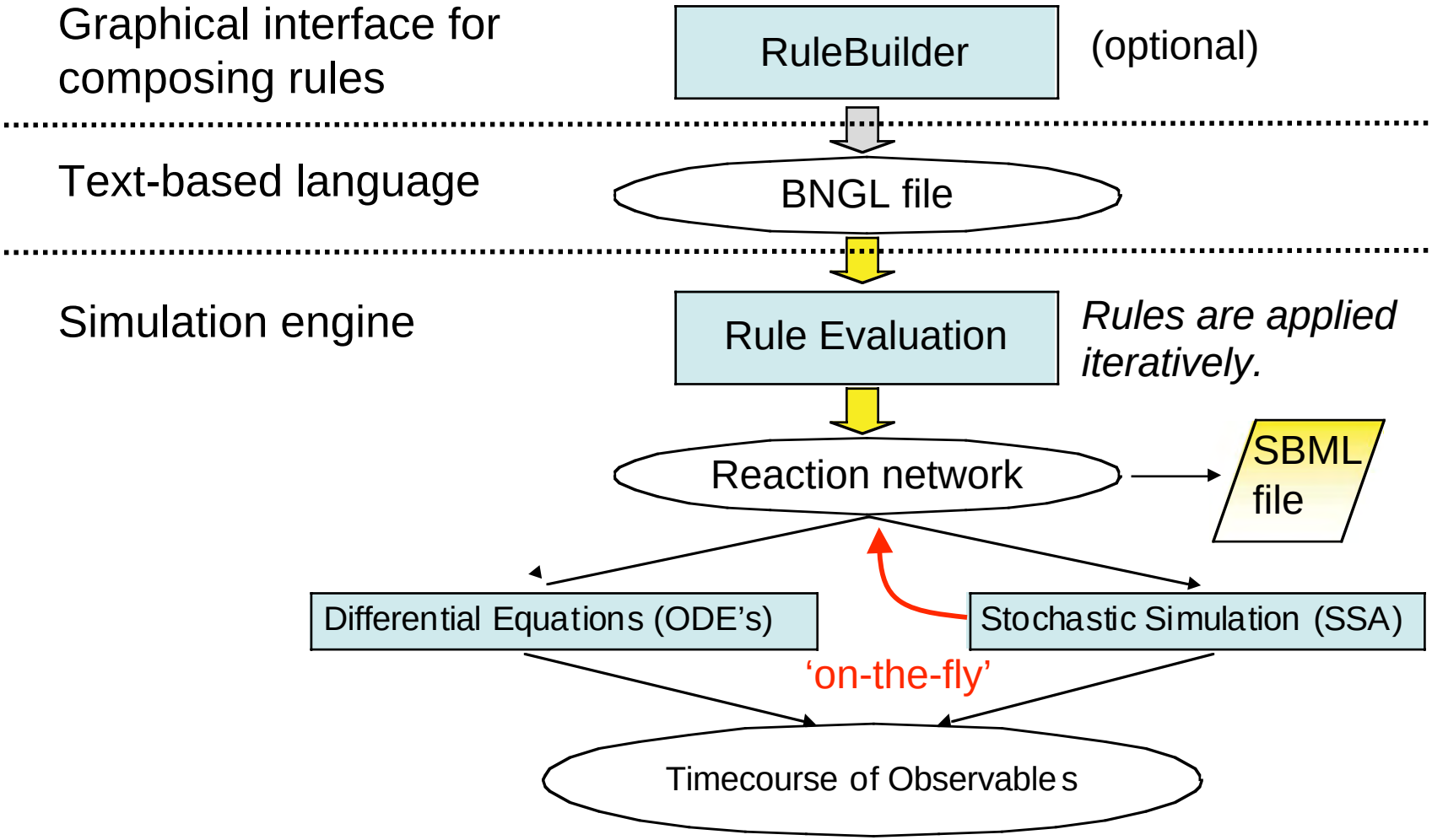
- Microscopic species generated by applying rules to molecules are difficult to observe directly.
- *Observables* define quantities that can be measured in experiments.



Elements of BNG Model

- **parameters**
  - defined anywhere
  - math expressions provide annotation
- **seed species**
  - Any molecule with non-zero initial concentration
- **reaction rules**
- **observables**
  - define model outputs
- **actions**
  - network generation
  - simulation
  - output
  - change parameters

BioNetGen2: Software for graphical rule-based modeling



http://bionetgen.lanl.gov

Advantages of BNGL

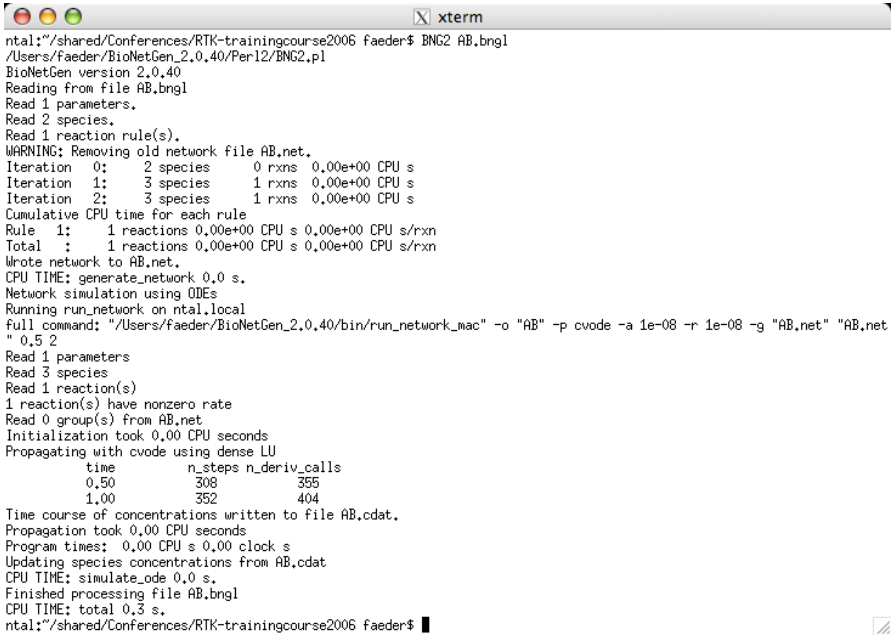
- Precise and flexible modeling language
- *Human readable*
  - Rules can be embedded in wikis, databases, applications, and papers (Ty Thomson)
- *Machine readable*
  - Forms basis for SBML L3 proposal (Blinov)
  - Interoperability (Vcell, Dynstoc, Kappa Factory, …)
  - Molecule and rule definition could be automated using databases of protein-protein interactions as a source

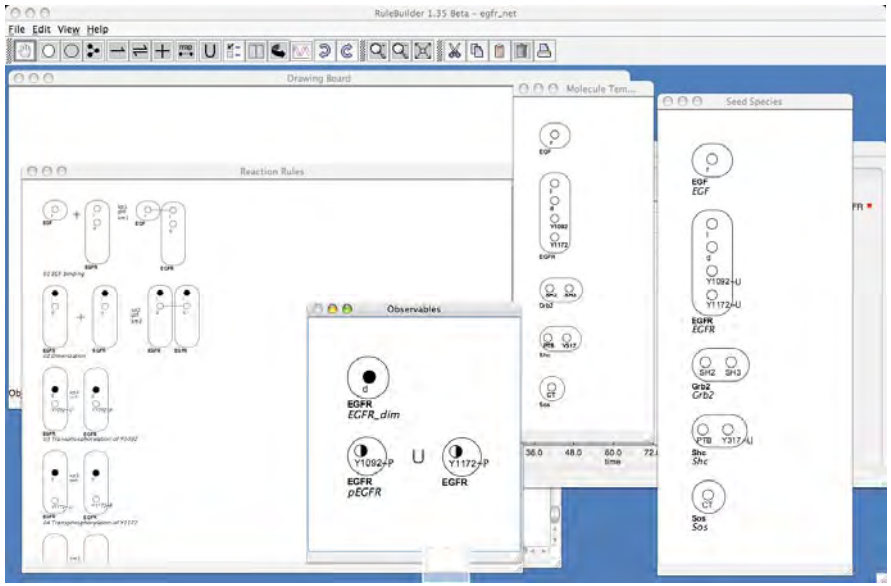
Hlavacek et al. (2006) *Sci. STKE*, **2006**, re6.

Two interfaces to BNG

Terminal interface
(text-based input)

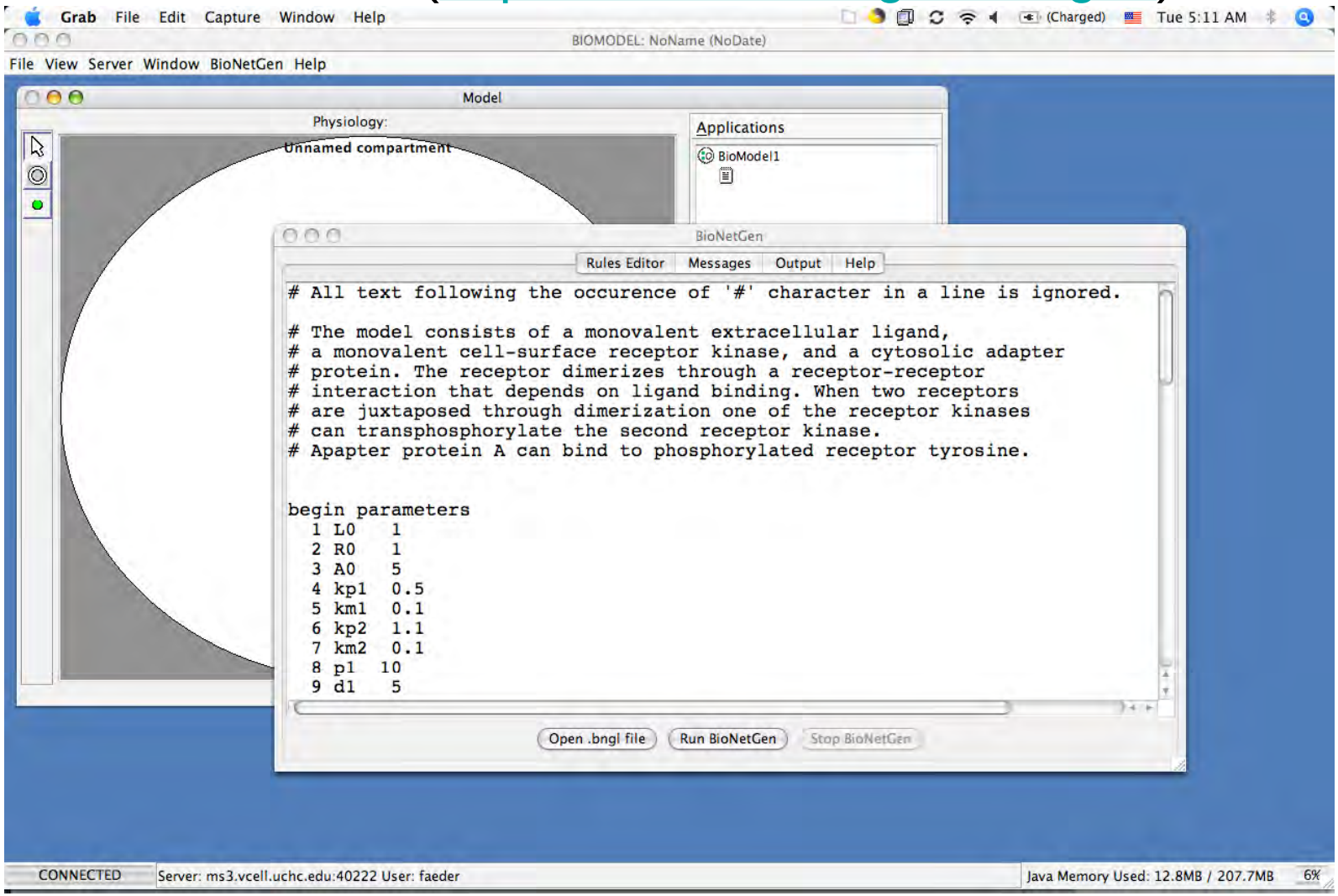
RuleBuilder GUI





A Third Way - Virtual Cell Interface

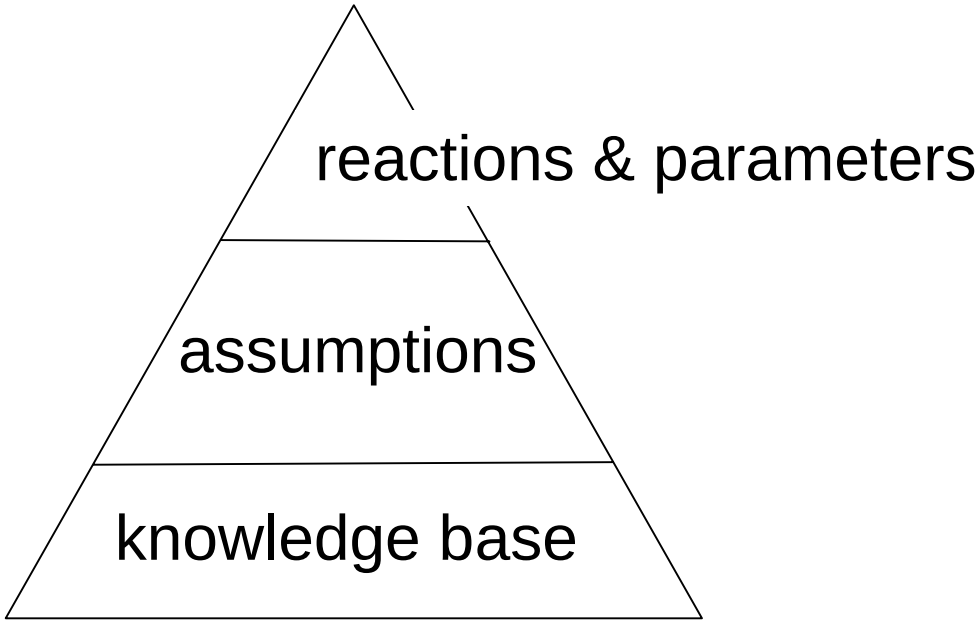
BioNetGen@Vcell (http://www.vcell.org/bionetgen)



The AlphaWiki Yeast Pheromone Response Model

Ty Thomson & Drew Endy (MIT)

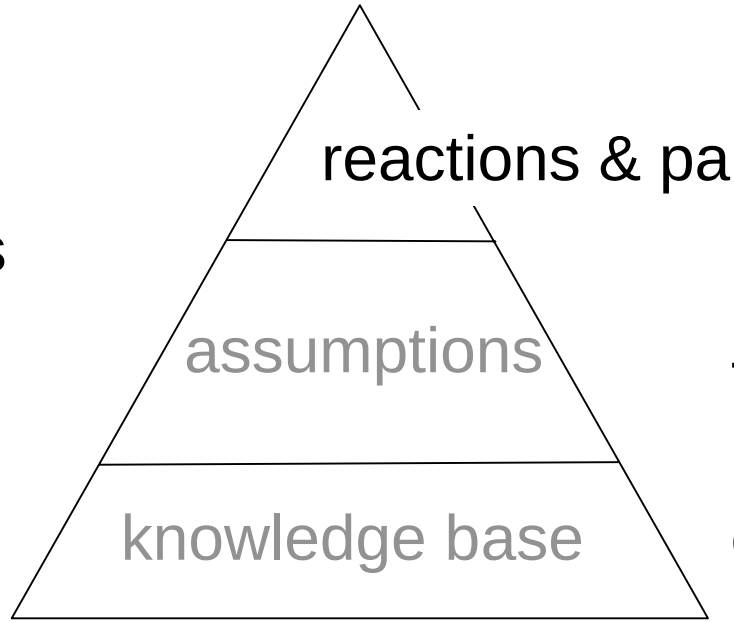
Biological models



The AlphaWiki Yeast Pheromone Response Model

Ty Thomson & Drew Endy (MIT)

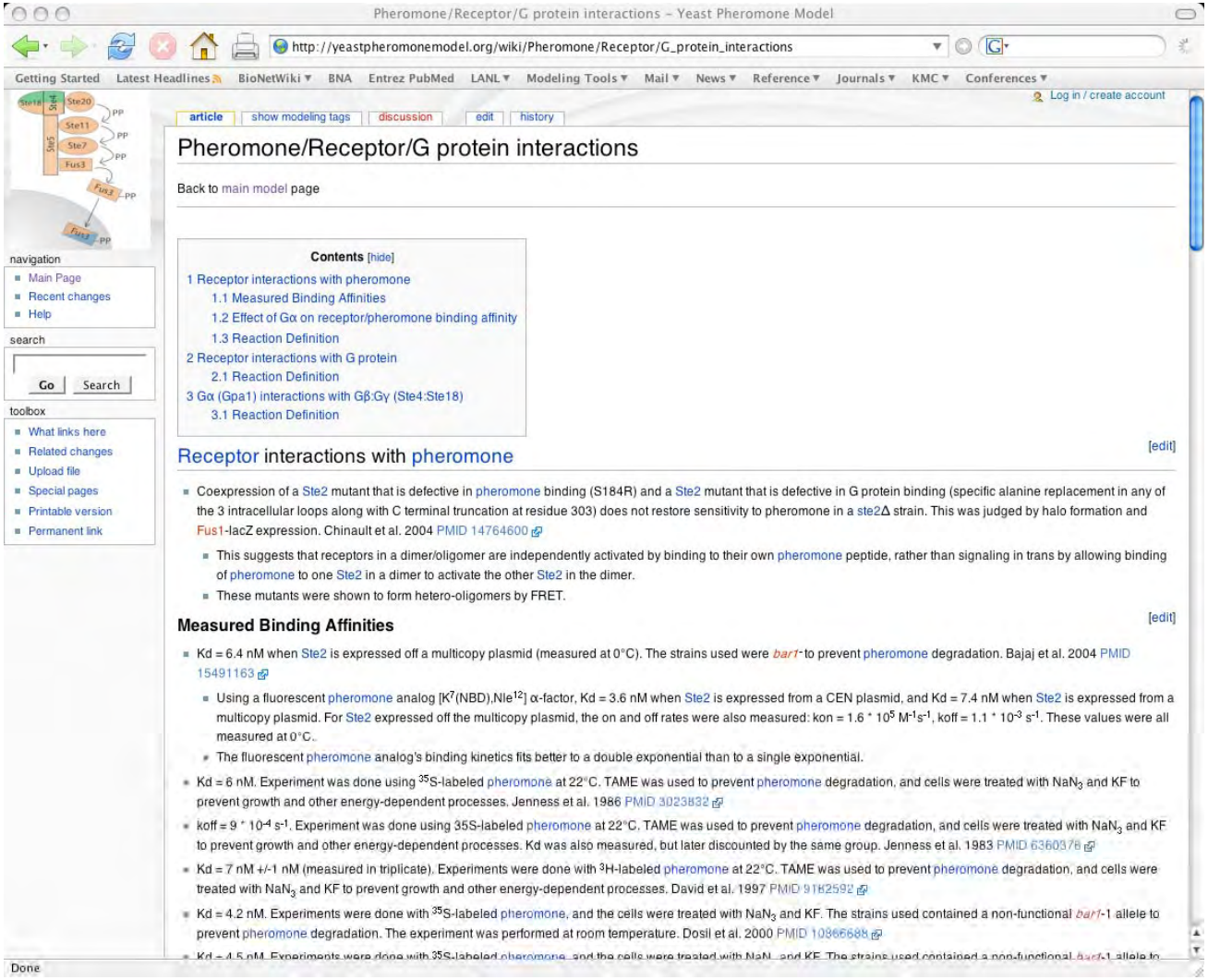
Biological models



These are usually left out of the literature

The AlphaWiki Yeast Pheromone Response Model

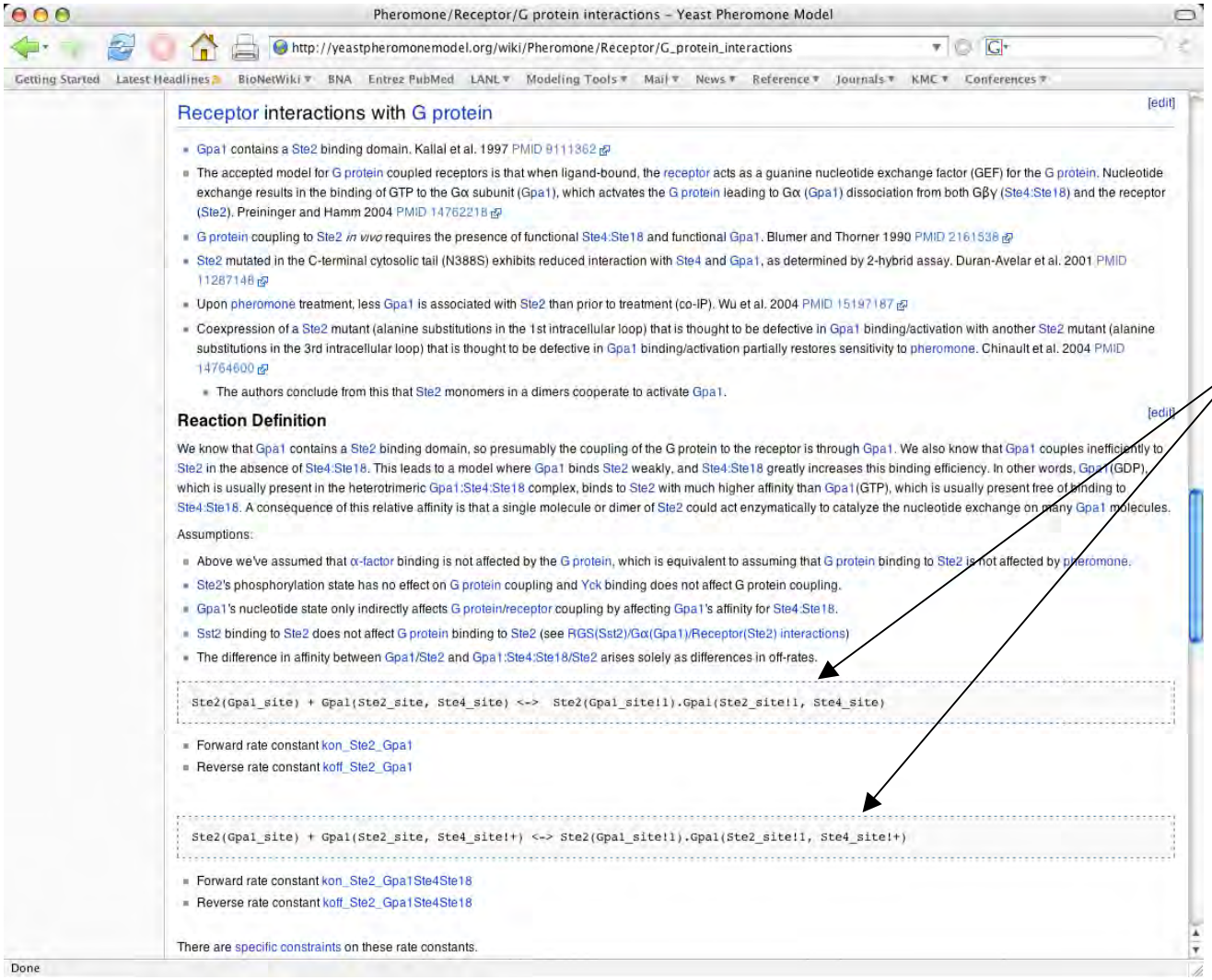
Ty Thomson & Drew Endy (MIT)



Structured wiki may offer a solution

The AlphaWiki Yeast Pheromone Response Model

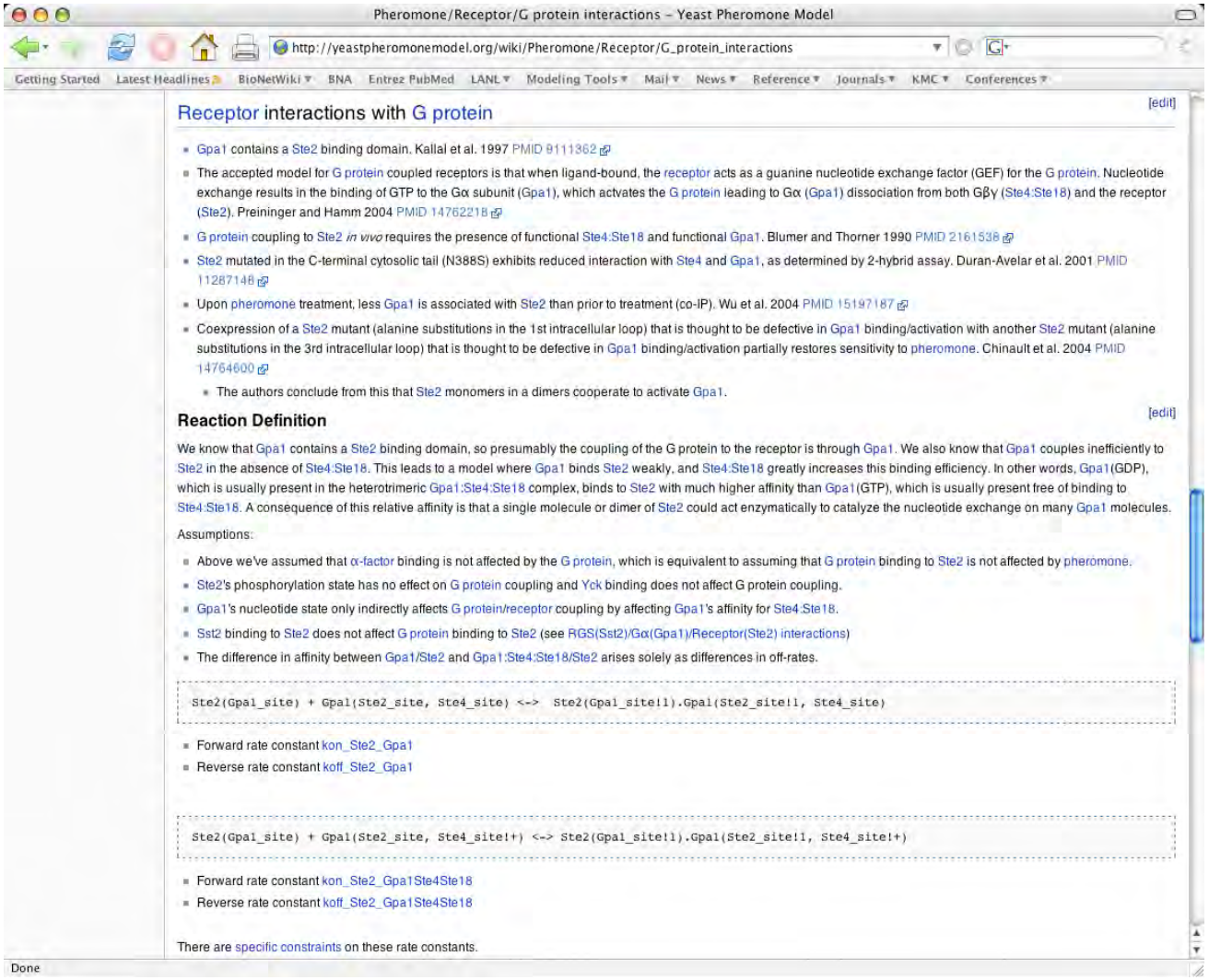
Ty Thomson & Drew Endy (MIT)



BNG rules are used for precise reaction definitions

The AlphaWiki Yeast Pheromone Response Model

Ty Thomson & Drew Endy (MIT)



Model is automatically generated from wiki

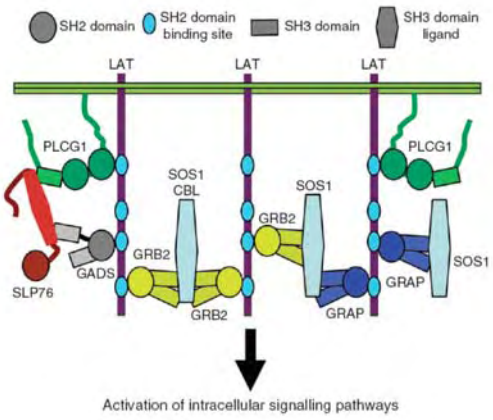
Systems Modeled

- IgE Receptor (FcεRI)
  - Faeder et al. *J. Immunol.* (2003)
  - Goldstein et al. *Nat. Rev. Immunol.* (2004)
- Growth Factor Receptors
  - Blinov et al. *Biosyst.* (2006) [EGFR]
  - Barua et al. *Biophys. J.* (2006) [Shp2]
- TLR4, TCR, IFNγ, TNF-α, TGF-β, …
- Carbon Fate Maps
  - Mu et al., submitted.

Key insights

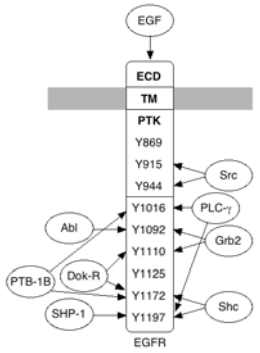
- RBM's are straightforward to construct and do not require more parameters
  - New predictions

- Important role of multivalent interactions
  - Complex formation can produce ligand specificity (kinetic proofreading)
  - Intuition often fails
  - Oligomerization may be a common feature of biological signaling

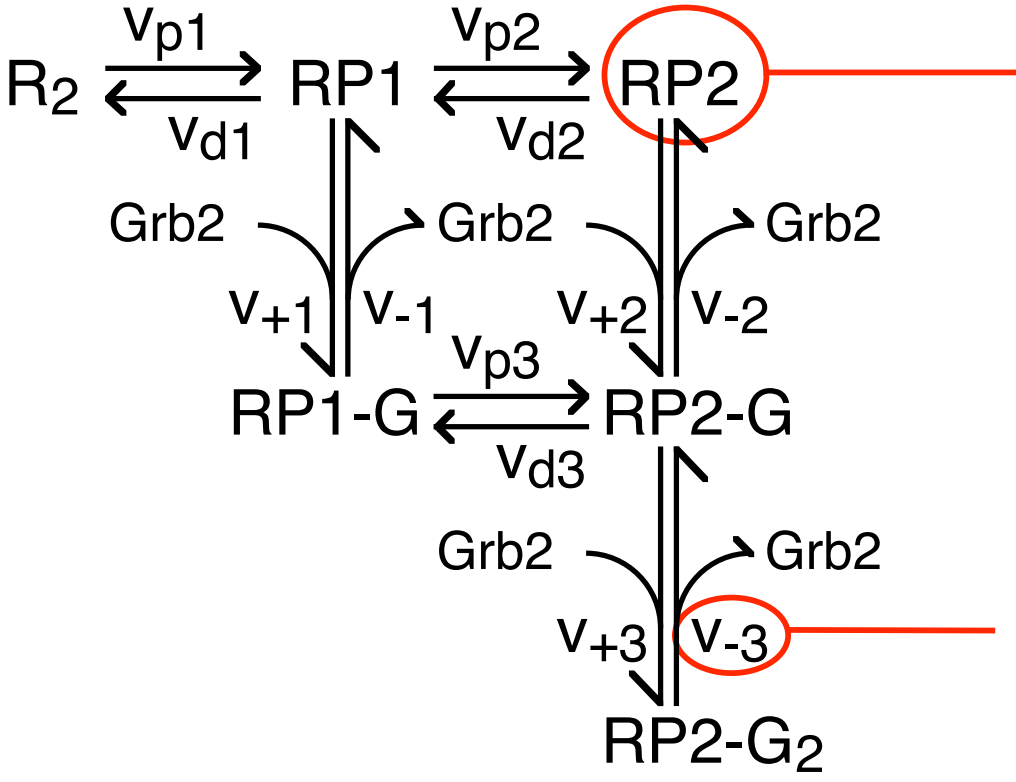


Ambarish Nag

- Concurrency in biological information processing
  - Scaffolds can activate multiple pathways independently
  - Strong potential for interaction among pathways (largely unexplored)



A standard reaction scheme



Species: One for every possible modification state of every complex

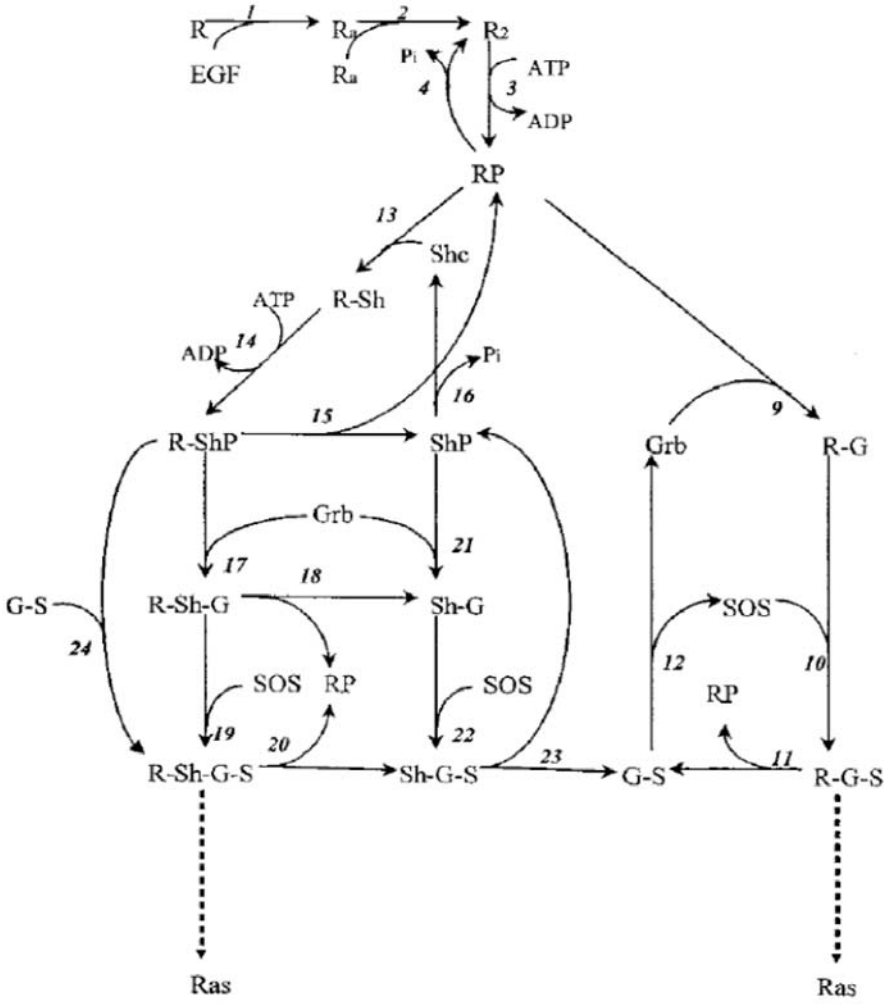
Reactions: One for every transition among species

Mass action kinetics gives rise to a set of ODEs, one for each species

A conventional model for EGFR signaling

The Kholodenko model*

Avoids combinatorial complexity by assuming that certain reaction events must occur in a particular order



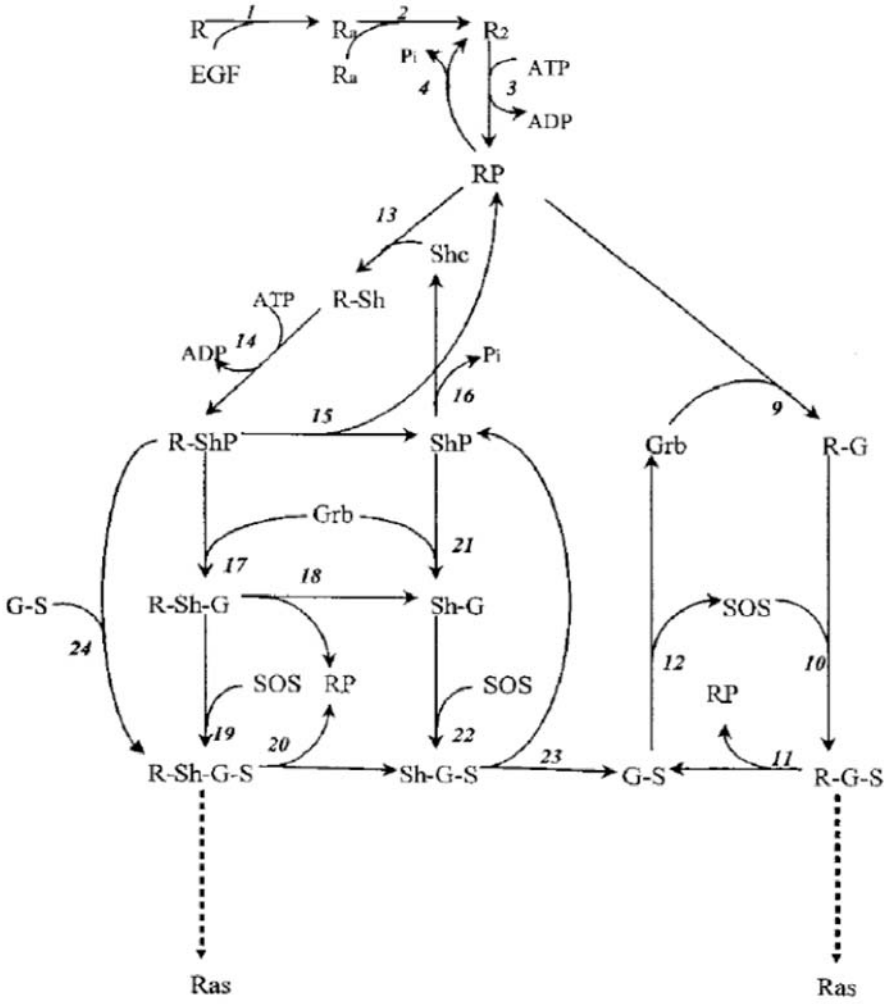
J. Biol. Chem. **274**, 30169 (1999)

A conventional model for EGFR signaling

The Kholodenko model*

5 components

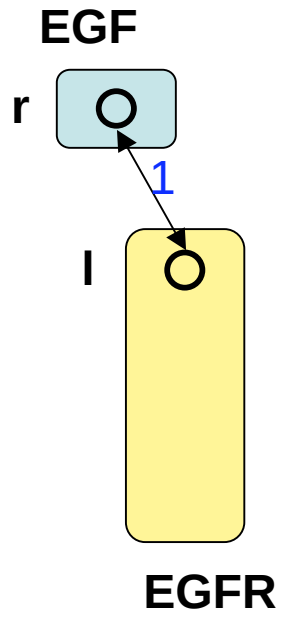
18 species
34 reactions



Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```



Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```

```
EGFR(l) + EGF(r) <-> EGFR(l!1).EGF(r!1)
```

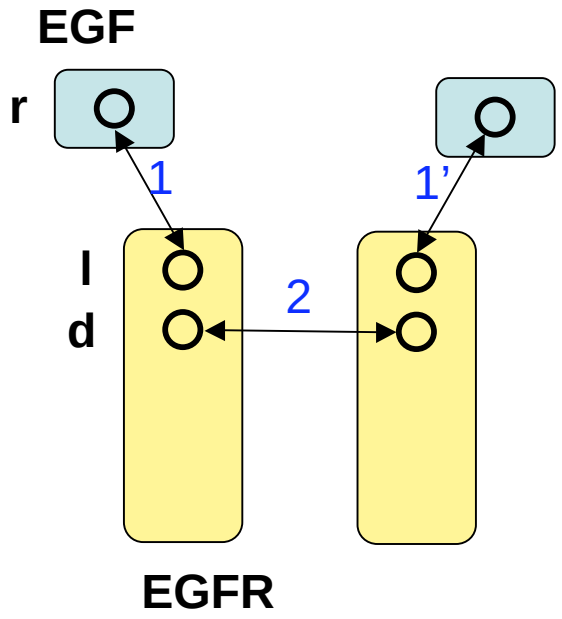
Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```

2. EGFR dimerization

```
Ra + Ra <-> R2
```

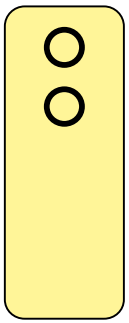


Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```

```
EGFR(l,d) + EGF(r) <-> EGFR(l!1,d).EGF(r!1)
```



additional context because representation is flat

2. EGFR dimerization

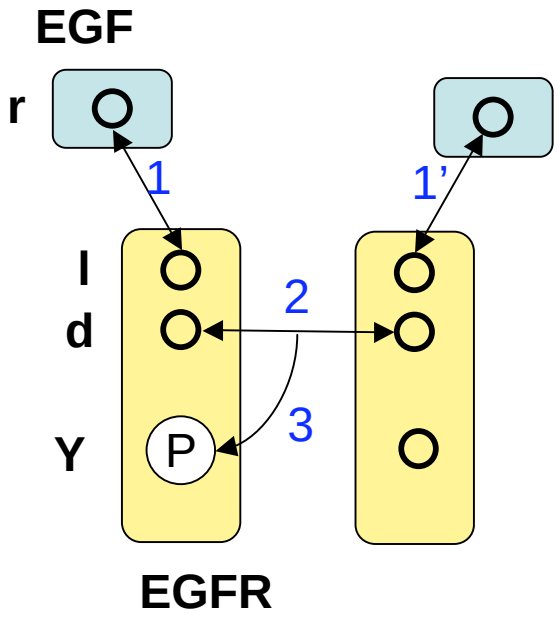
```
Ra + Ra <-> R2
```

```
EGFR(l!1,d).EGF(r!1) + EGFR(l!2,d).EGF(r!2) <->
        EGFR(l!1,d!3).EGF(r!1).EGFR(l!2,d!3).EGF(r!2)
```

Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```



2. EGFR dimerization

```
Ra + Ra <-> R2
```

3. EGFR autophosphorylation

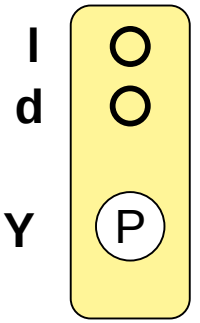
```
R2 <-> RP
```

Dissecting the reaction scheme

1. EGF binding to EGFR

```
R + EGF <-> Ra
```

```
EGFR(l,d,Y~U) + EGF(r) <-> EGFR(l!1,d,Y~U).EGF(r!1)
```



2. EGFR dimerization

```
Ra + Ra <-> R2
```

```
EGFR(l!1,d,Y~U).EGF(r!1) + EGFR(l!2,d,Y~U).EGF(r!2) <->
    EGFR(l!1,d!3,Y~U).EGF(r!1).EGFR(l!2,d!3,Y~U).EGF(r!2)
```

3. EGFR autophosphorylation

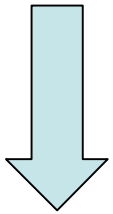
```
R2 <-> RP
```

```
EGFR(d!+,Y~U) <-> EGFR(d!+,Y~P)
```

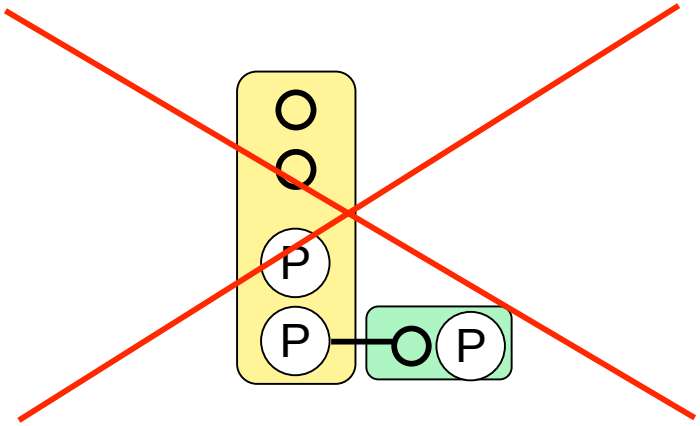
Assumptions accumulate!

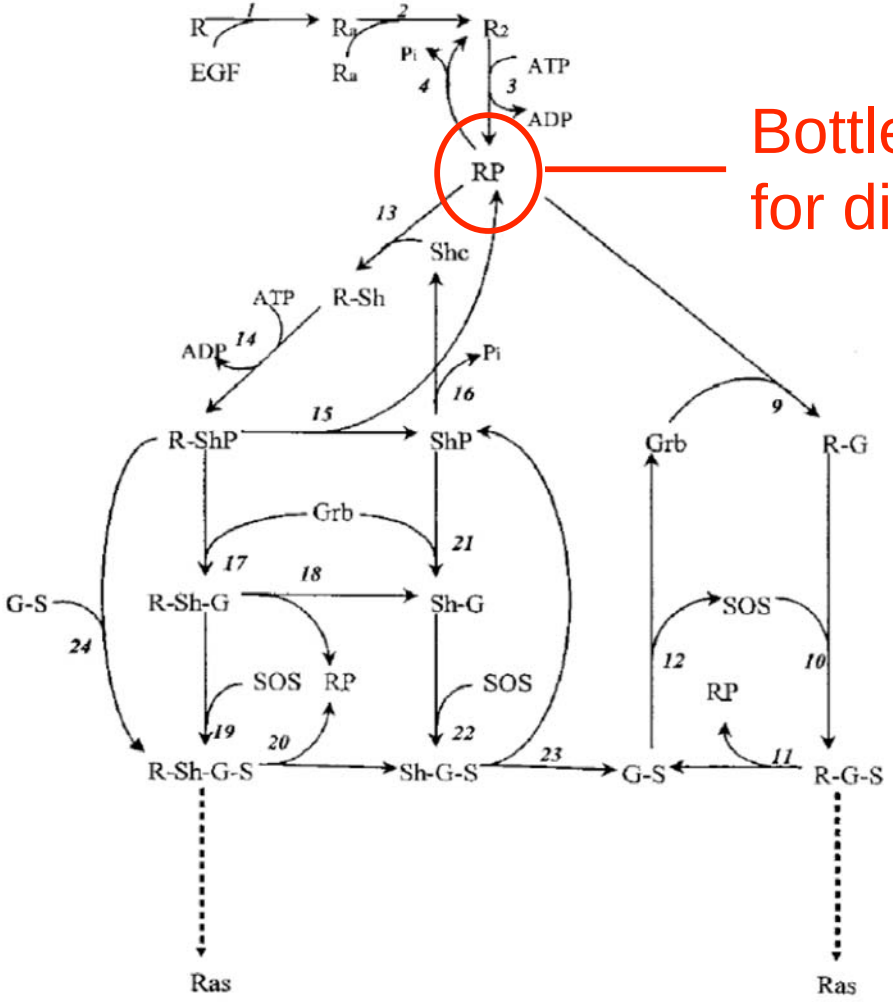
Effect of assuming receptor activation is sequential

1. Phosphorylation inhibits dimer breakup



No modified monomers





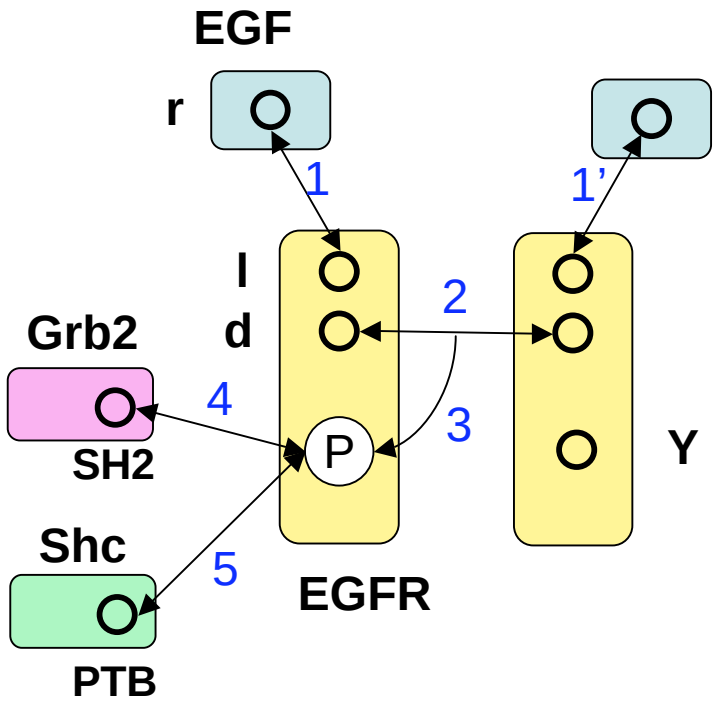
Adaptor protein binding

4. Grb2 binding to pEGFR

```
Grb2 + RP <-> RP-Grb2
```

5. Shc binding to pEGFR

```
Shc + RP <-> RP-Shc
```



Binding is assumed to be competitive

- either 4 or 5 may occur but not both
- only 1 adaptor per EGFR dimer

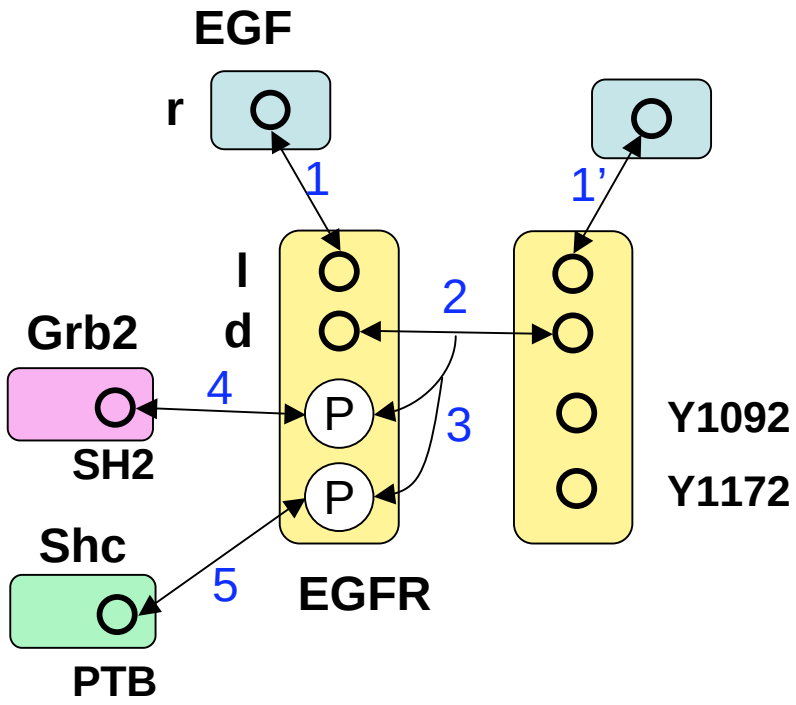
Splitting the adaptor binding site

4. Grb2 binding to pEGFR

```
Grb2 + RP <-> RP-Grb2
```

5. Shc binding to pEGFR

```
Shc + RP <-> RP-Shc
```



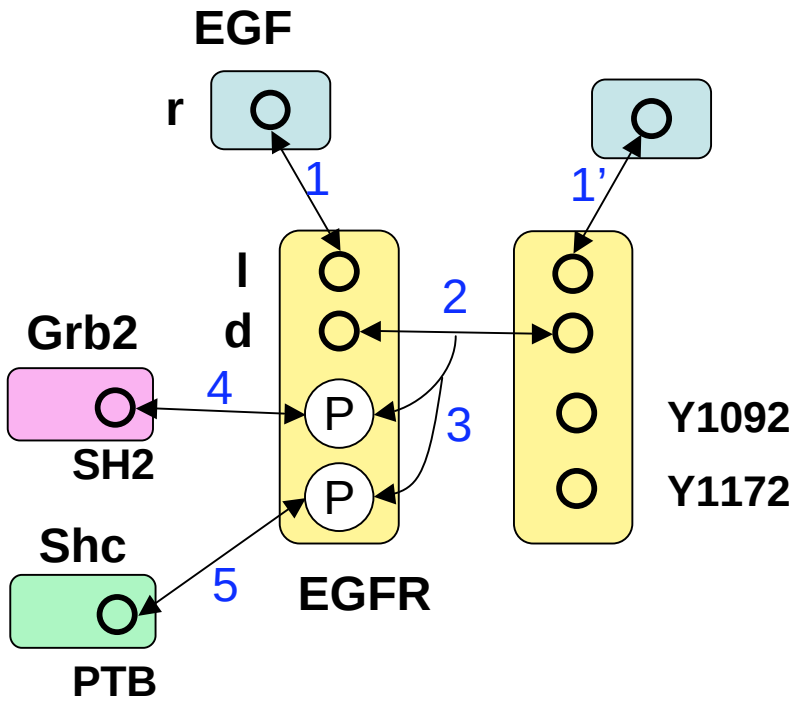
Splitting the adaptor binding site

4. Grb2 binding to pEGFR

```
Grb2 + RP <-> RP-Grb2
```

5. Shc binding to pEGFR

```
Shc + RP <-> RP-Shc
```

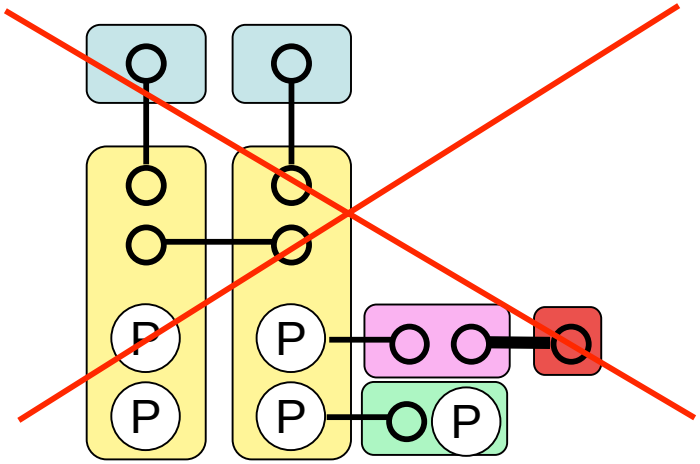


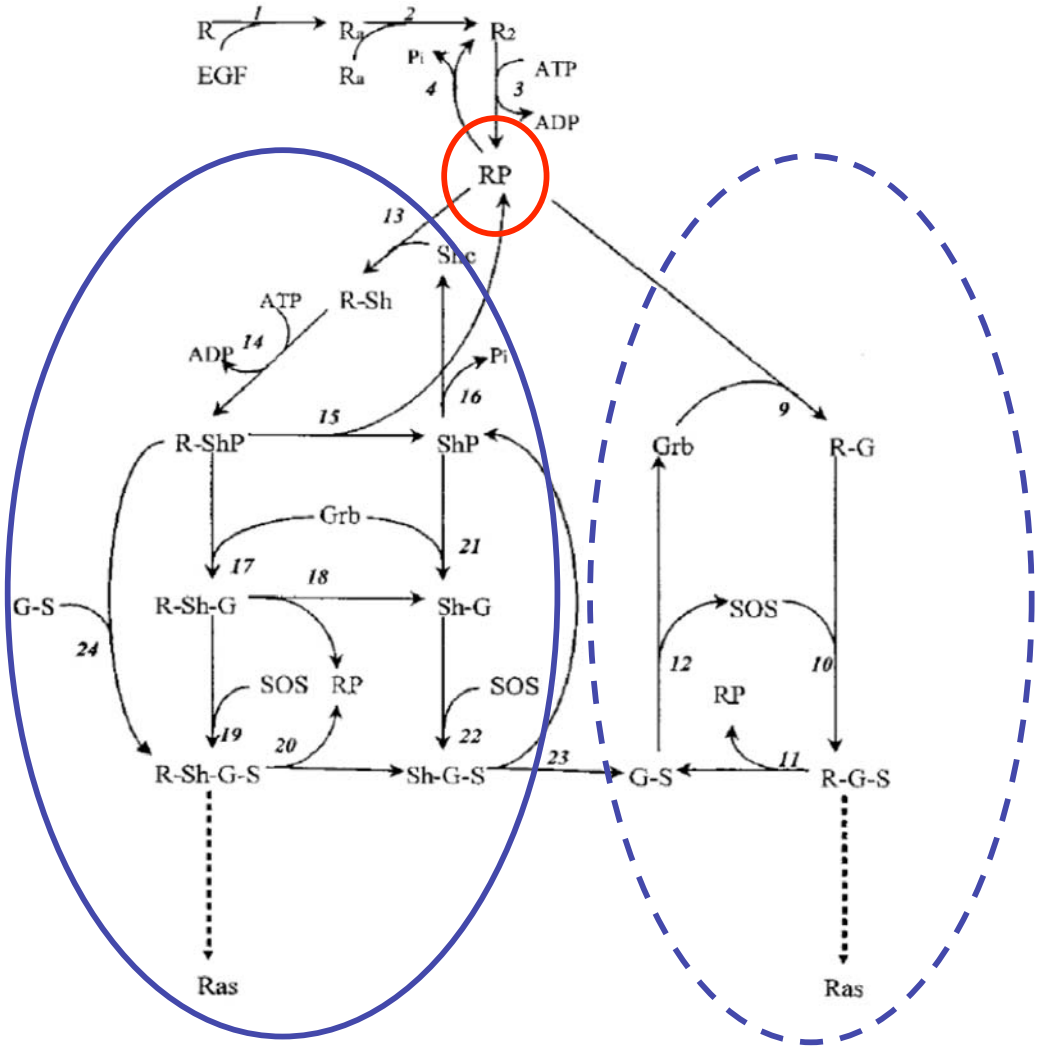

```
4. EGFR(d!+,Y1092~P) + Grb2(SH2,SH3) <->
        EGFR(d!+,Y1092~P!1).Grb2(SH2!1,SH3)

5. EGFR(d!+,Y1172~P) + Shc(PTB,Y317~U) <->
        EGFR(d!+,Y1172~P!1).Shc(PTB!1,Y317~U)
```

Effect of assuming adaptor binding is competitive

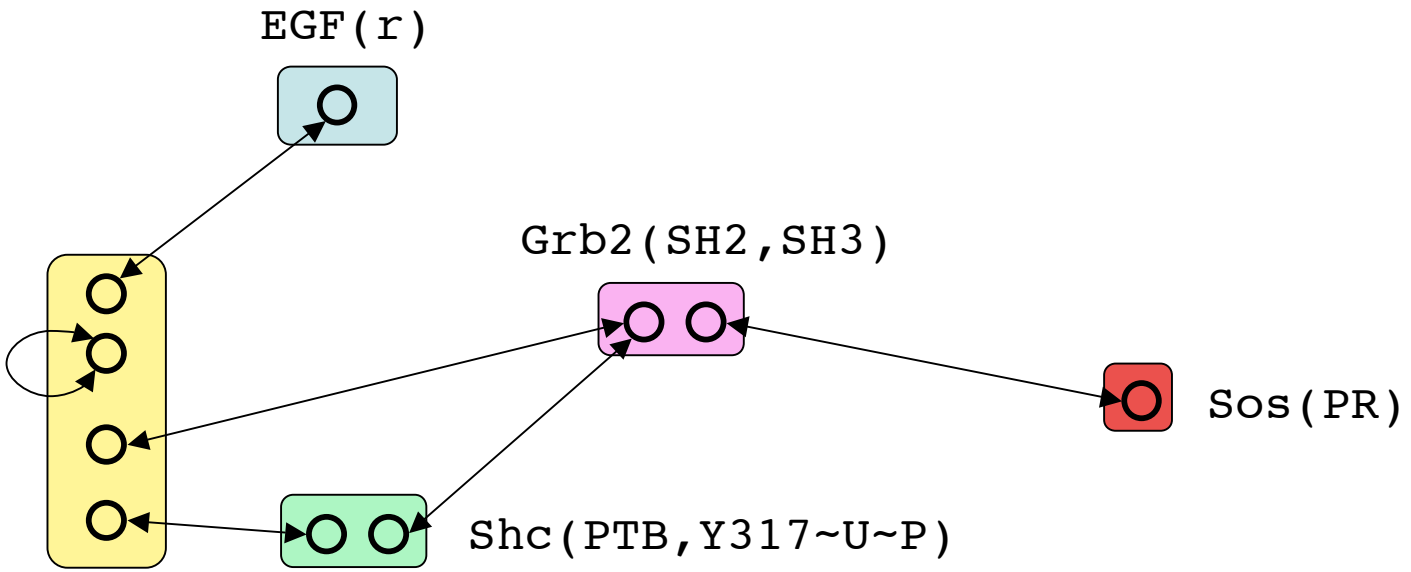
2. Adaptor binding is competitive

No dimers with more than one site modified





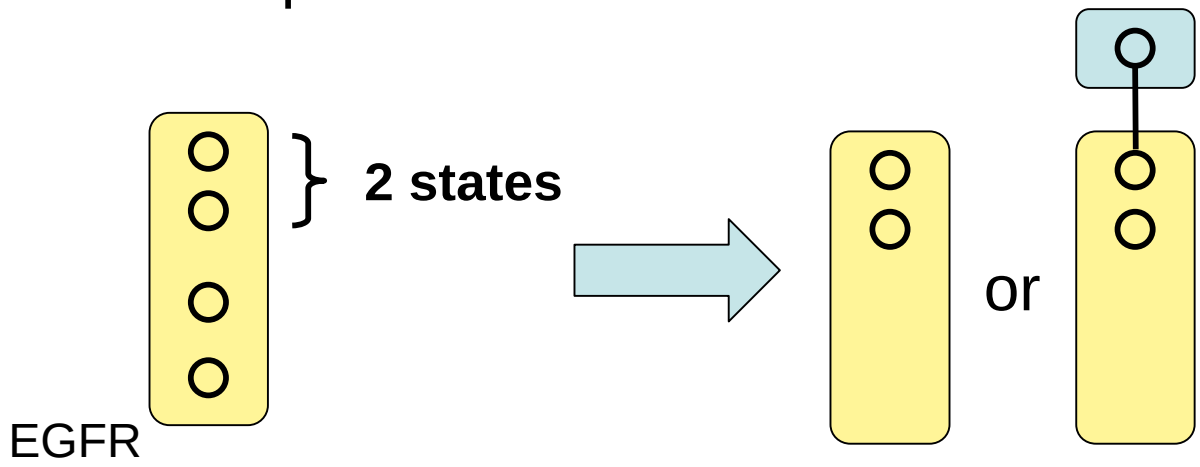
Molecules, components, and Interactions of the Kholodenko Model



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

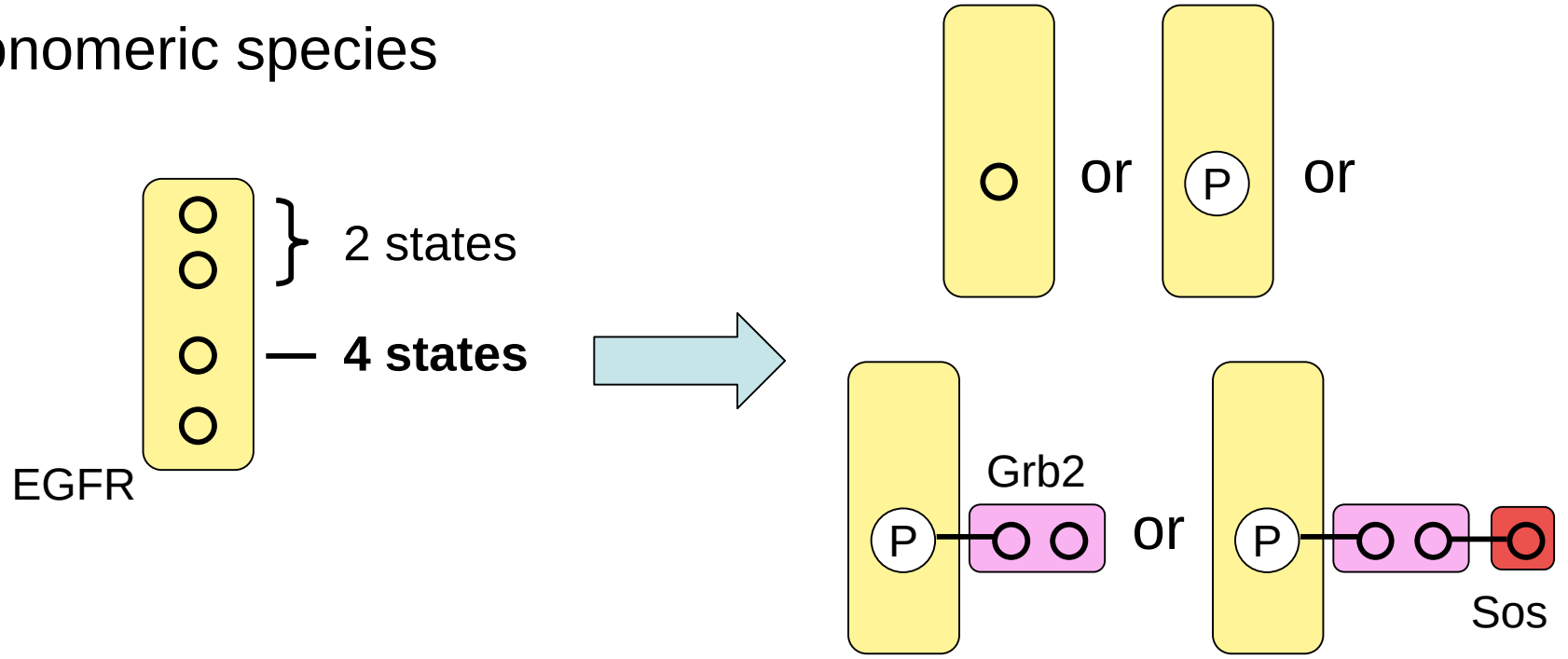
Combinatorial complexity of early events

Monomeric species

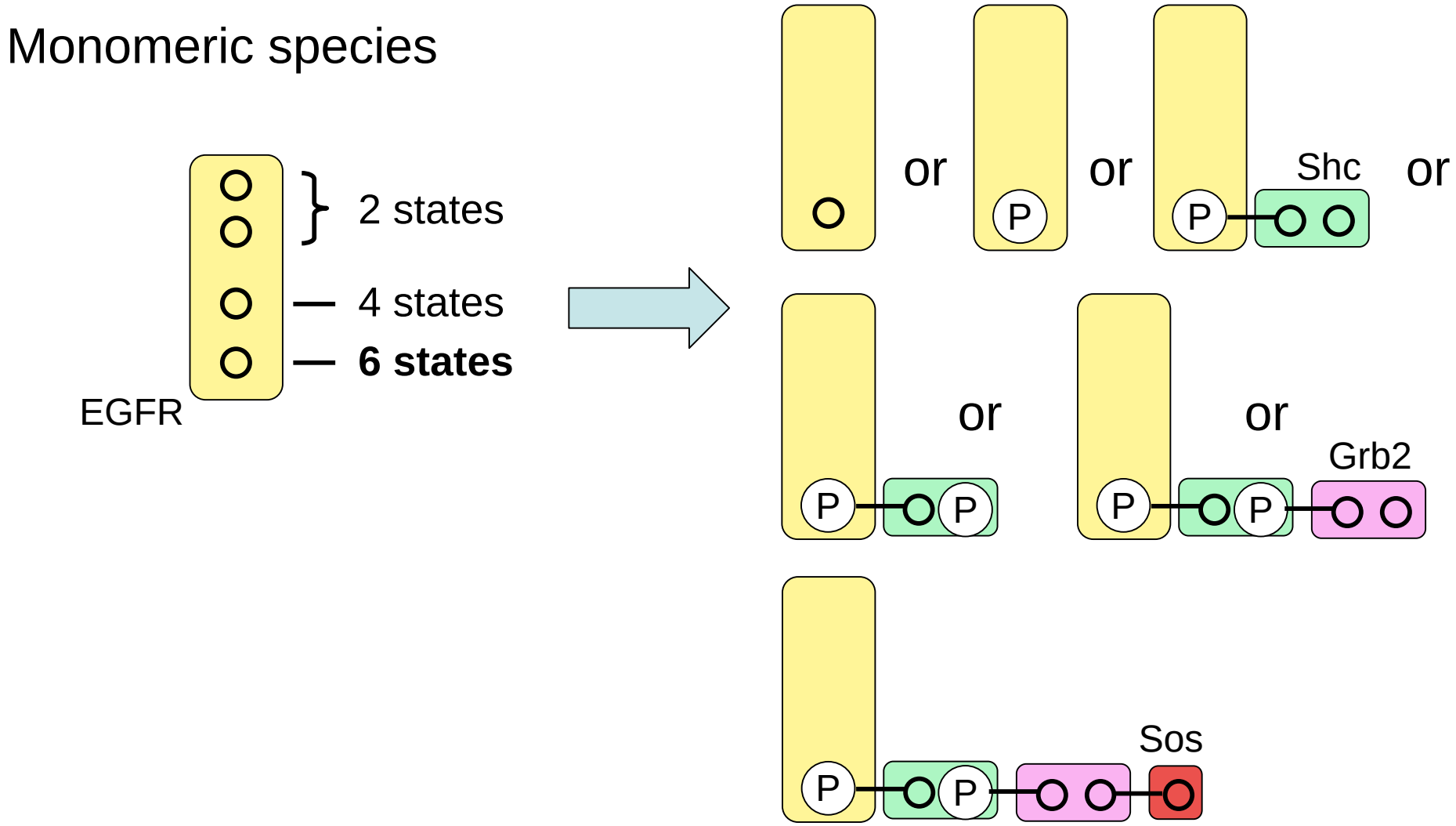


Combinatorial complexity of early events

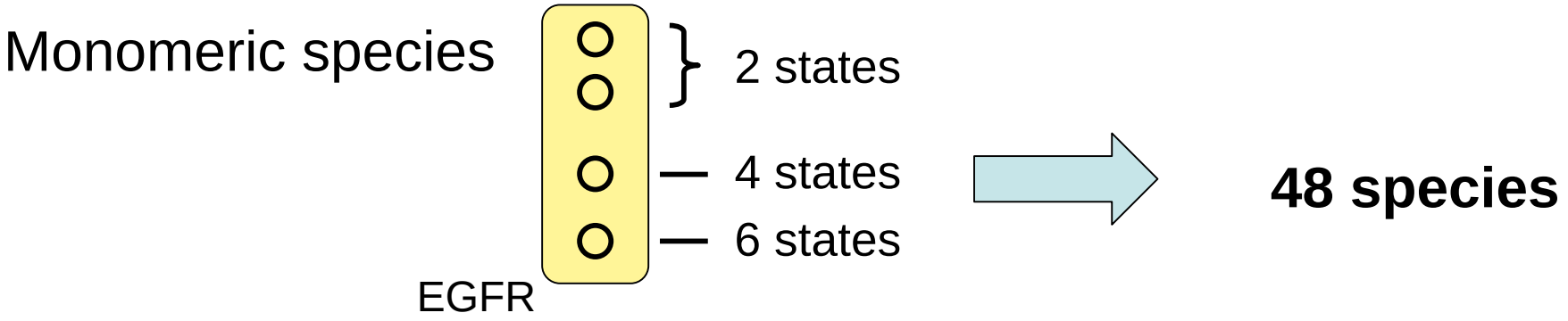
Monomeric species



Combinatorial complexity of early events

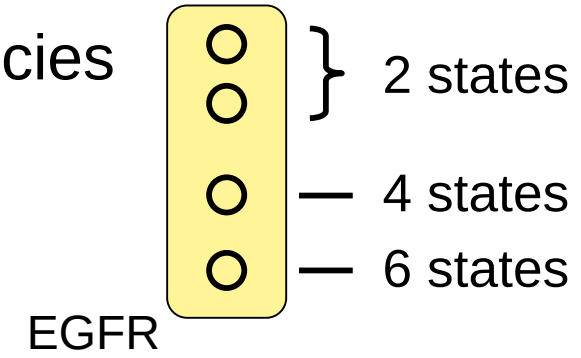


Combinatorial complexity of early events



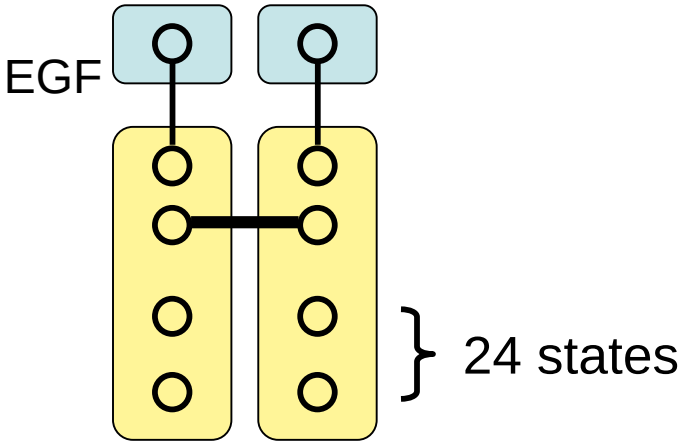
Combinatorial complexity of early events

Monomeric species



48 species

Dimeric species

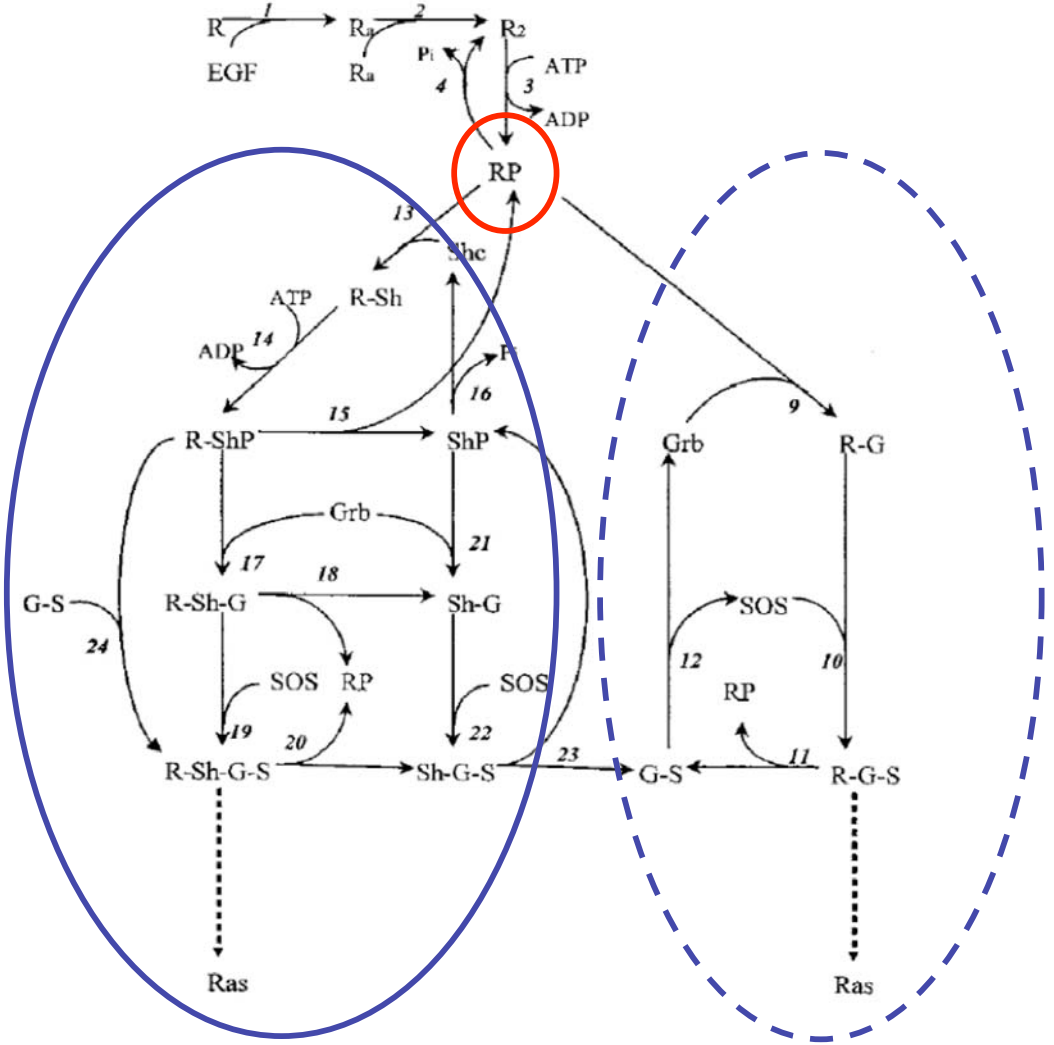


$N \times (N+1)/2$ = **300 species**

Assumptions made to limit combinatorial complexity

1. Phosphorylation inhibits dimer breakup
2. Adaptor binding is competitive

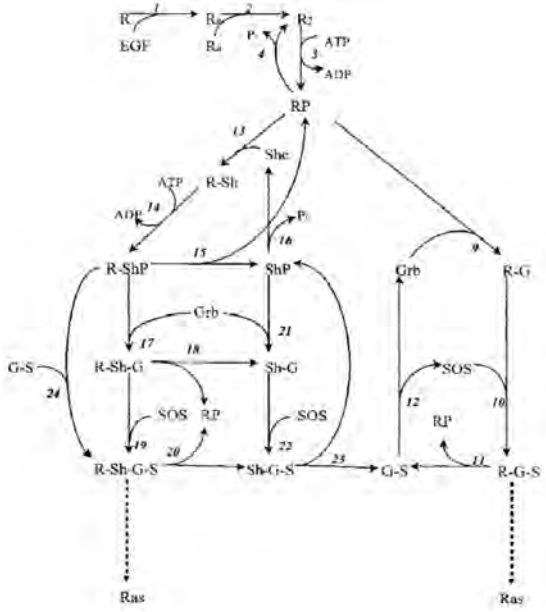
Experimental evidence contradicts both assumptions.



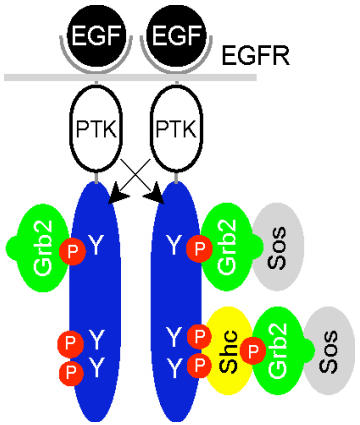
Rule-based version of the Kholodenko model

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

18 species
34 reactions



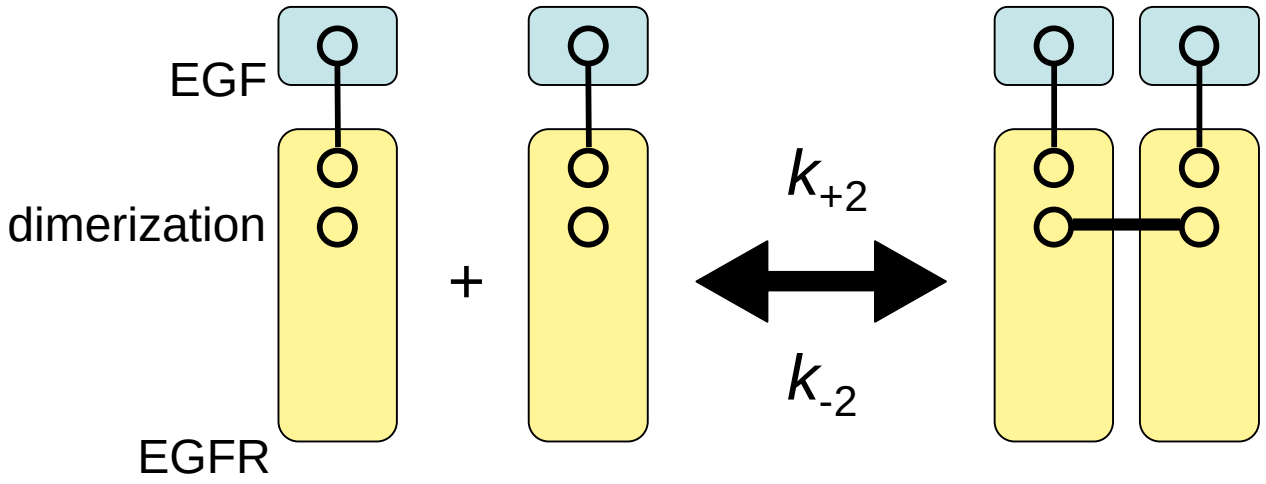
356 species
3749 reactions



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

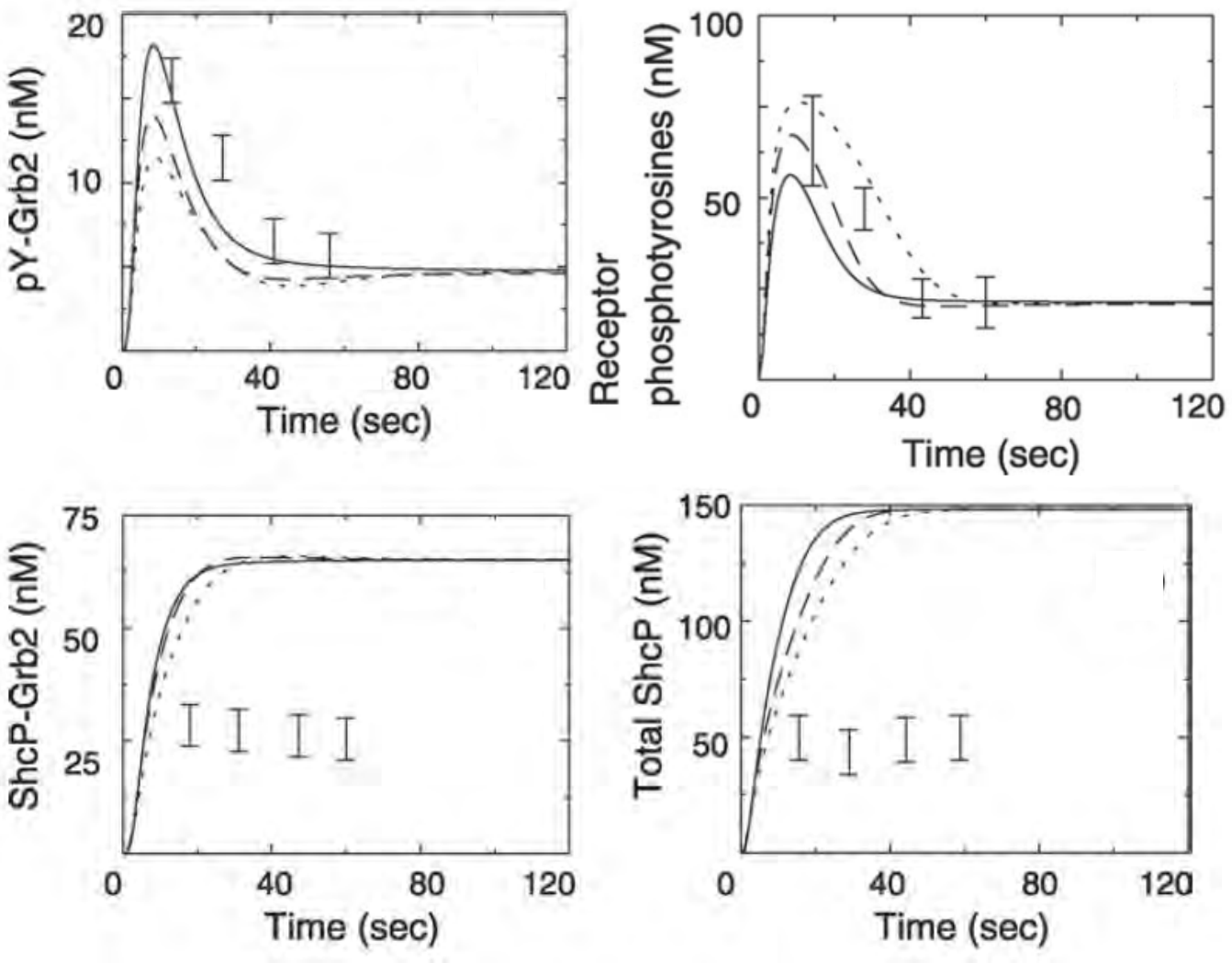
Dimerization rule eliminates previous assumption

EGFR dimerizes (600 reactions)

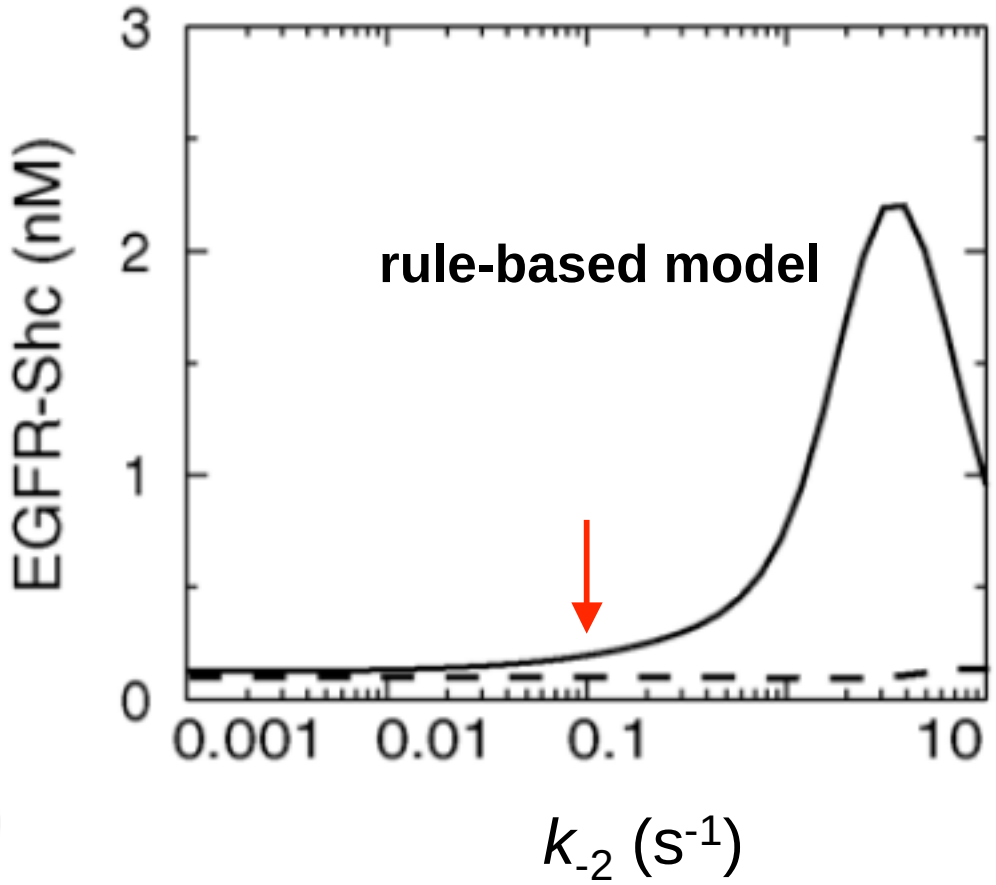


Dimers form and break up independent of phosphorylation of cytoplasmic domains

Two models predict similar overall binding and phosphorylation kinetics

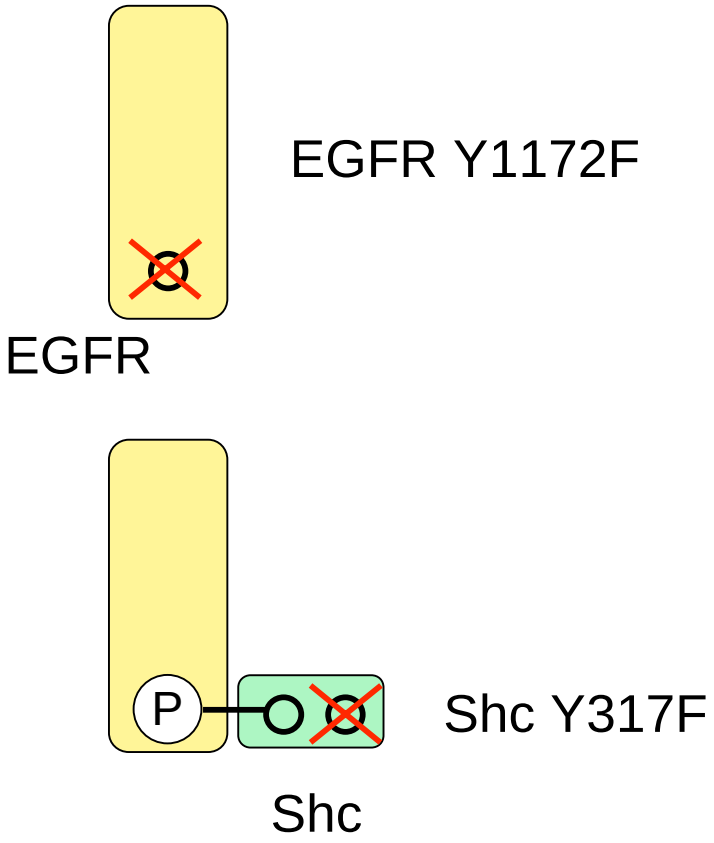


Strong differences when dimer dissociation rate is varied

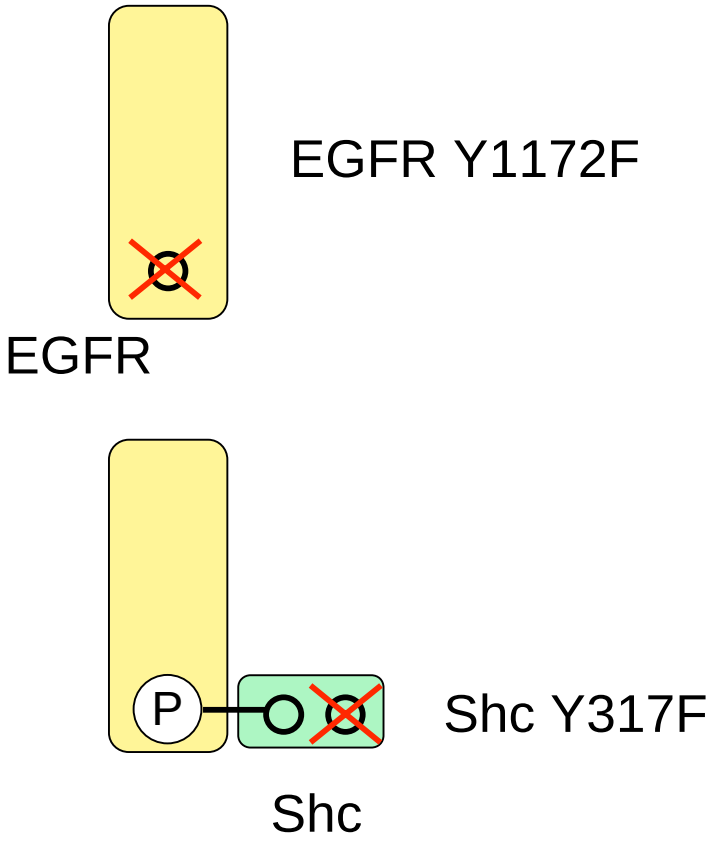


Kholodenko model

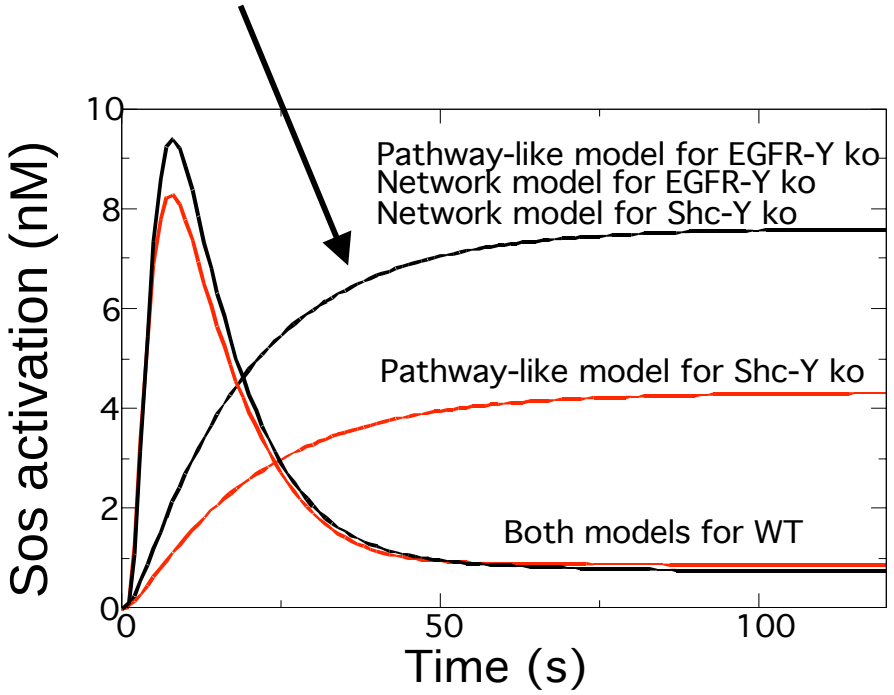
Results for two different knockouts of the Shc pathway



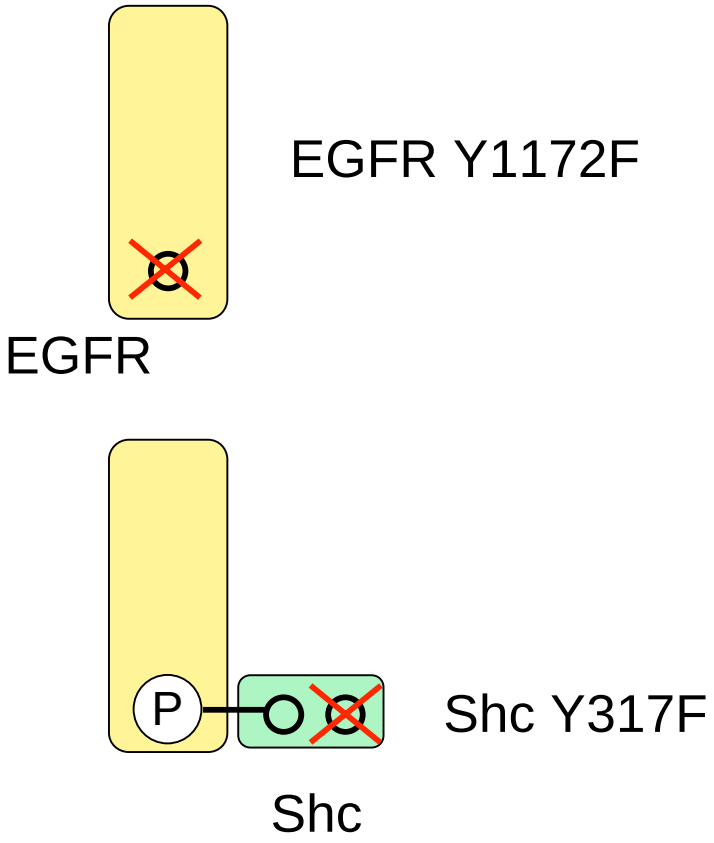
Results for two different knockouts of the Shc pathway



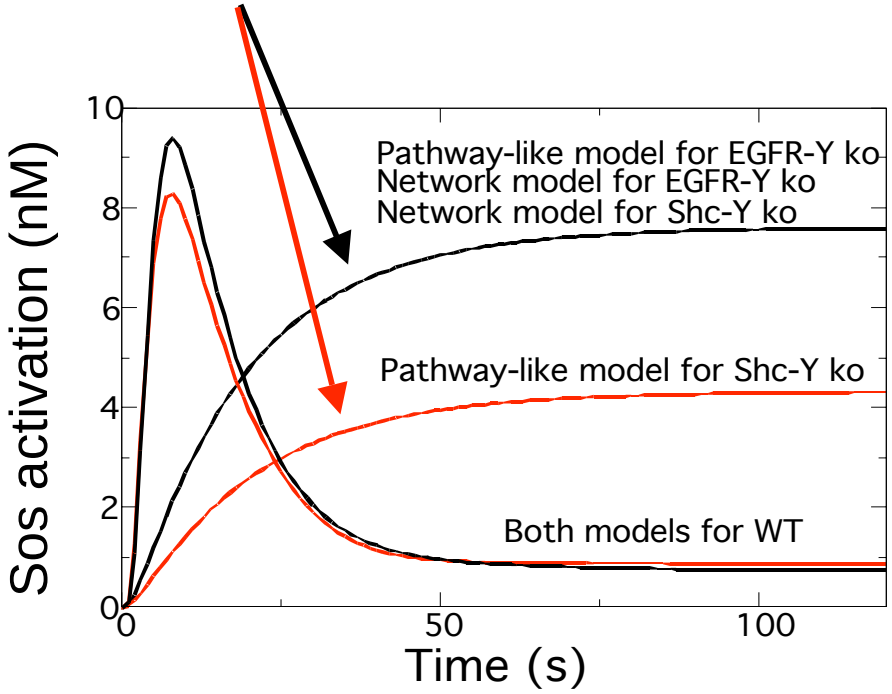
Rule-based model predicts same behavior for both knockouts



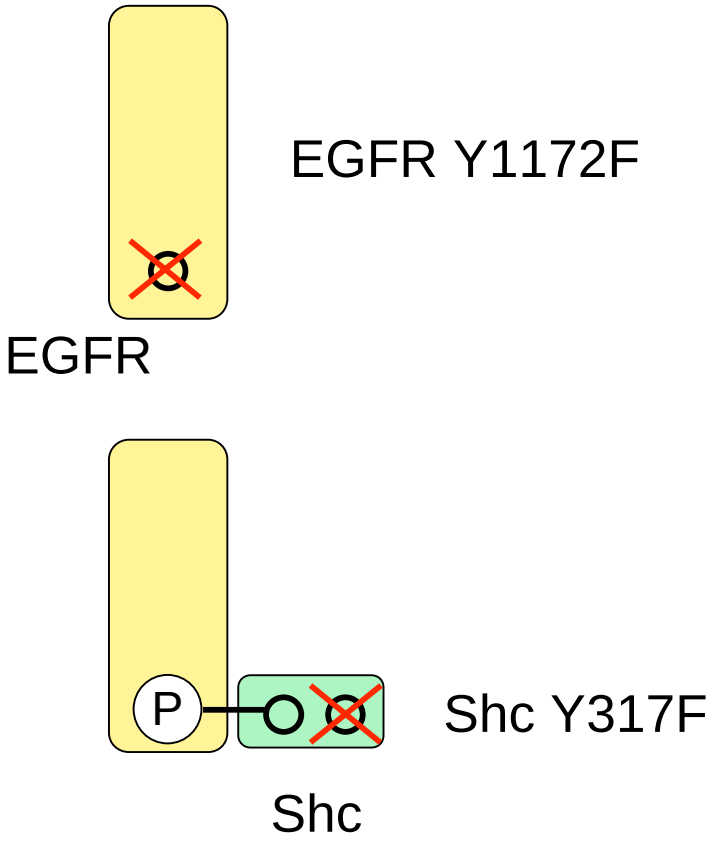
Results for two different knockouts of the Shc pathway



Kholodenko model predicts lower activation for Shc Y317F



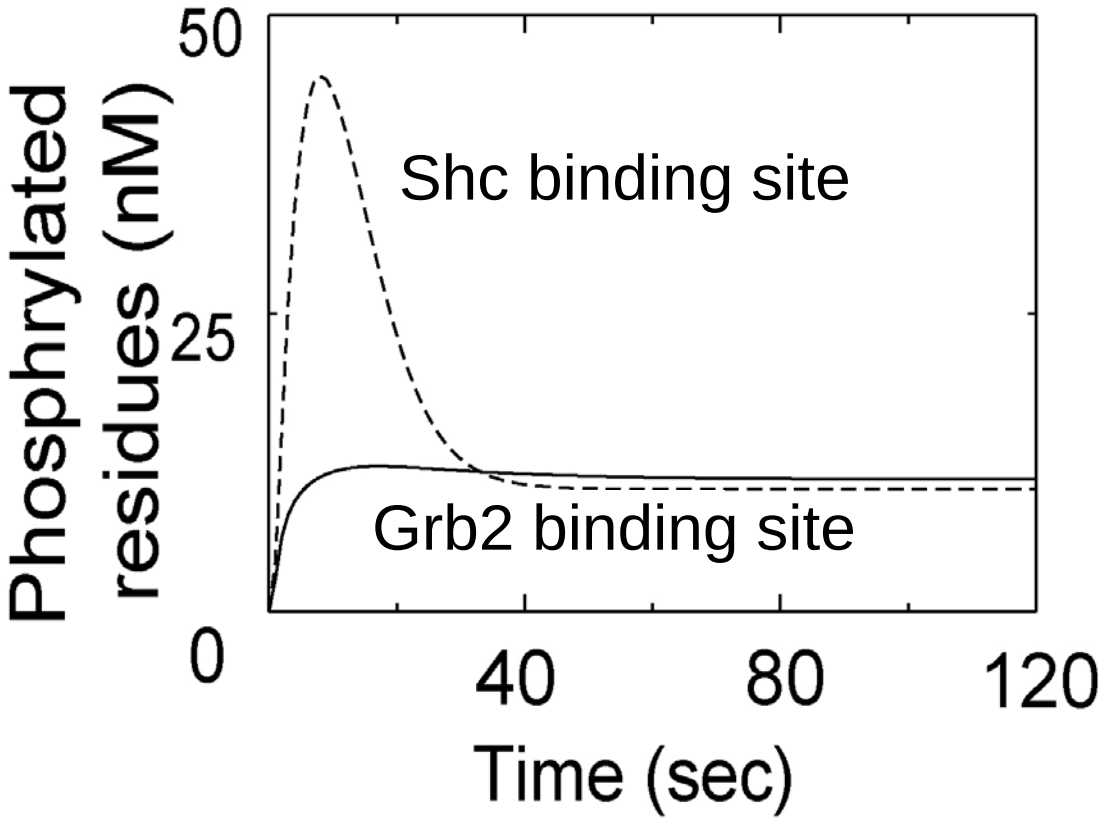
Results for two different knockouts of the Shc pathway



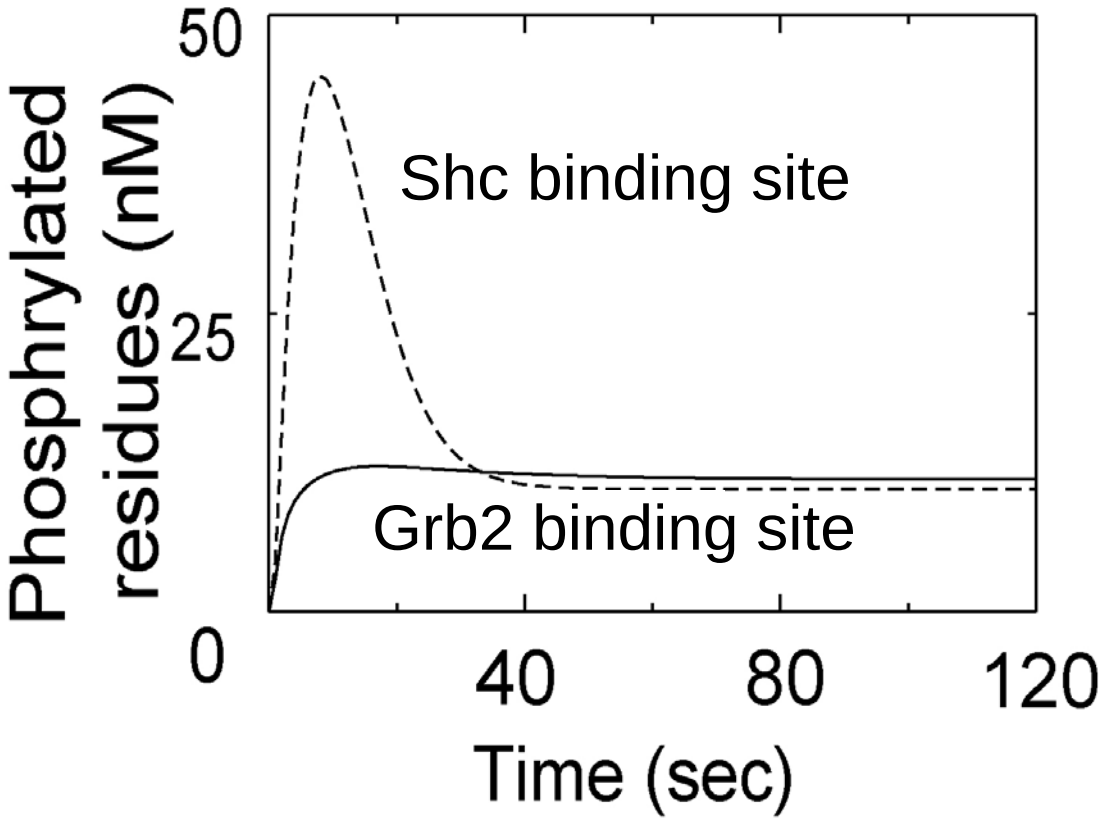
Kholodenko model predicts lower activation for Shc Y317F

… because mutant Shc blocks binding of Grb2 (competitive binding)

Rule-based model predicts distinct kinetics for two phosphorylation sites

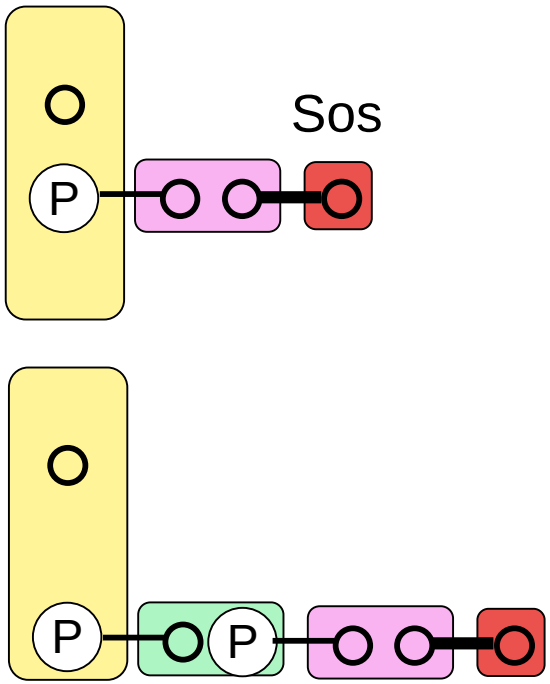


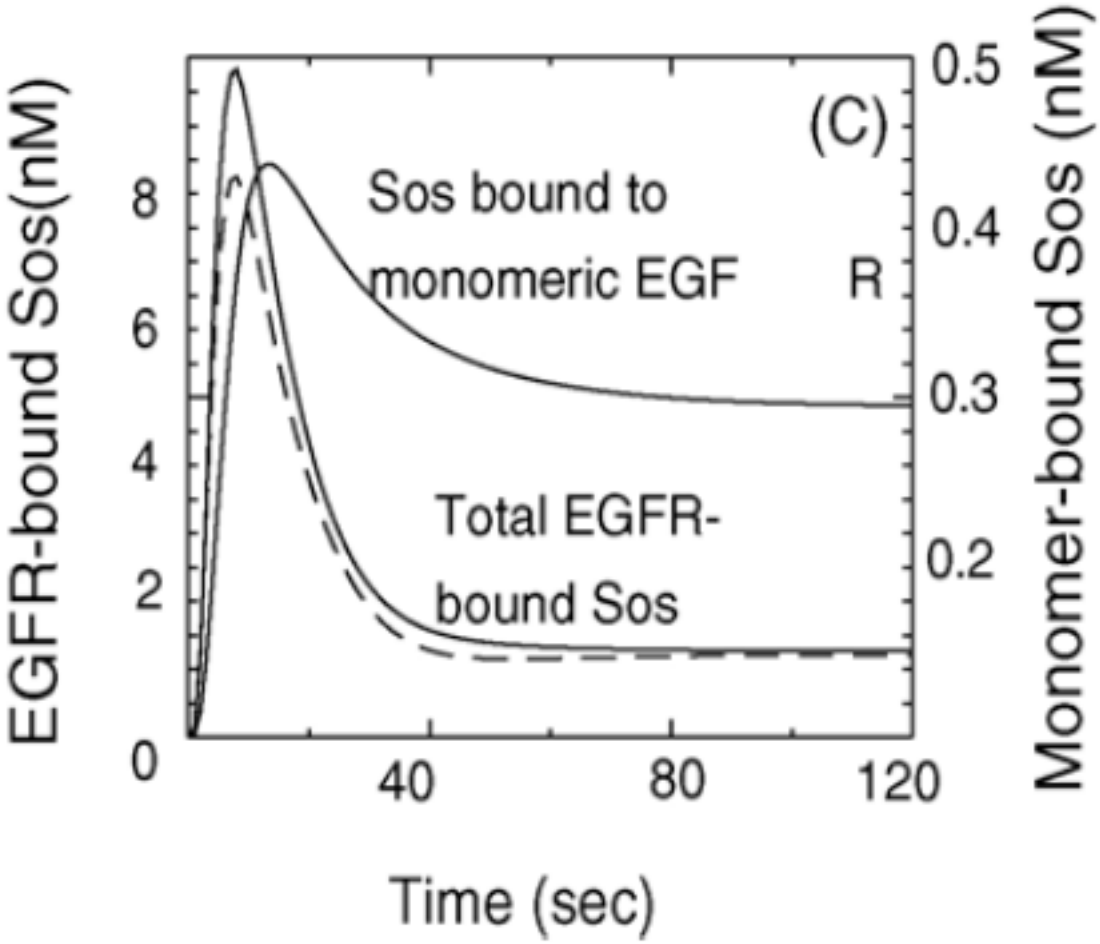
Rule-based model predicts distinct kinetics for two phosphorylation sites



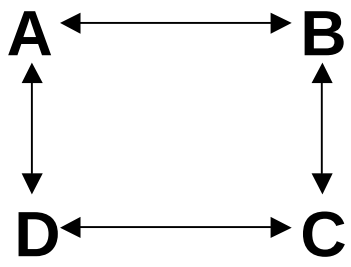
Also predicts monomers make substantial contribution to steady state Sos activation

36% of active Sos associates with EGFR monomers





Principle of detailed balance: Making sure that models obey laws of thermodynamics



Around any loop in the reaction network, the total free energy change (ΔG) must equal 0.

$$\Delta G = \Delta G_{AB} + \Delta G_{BC} - \Delta G_{DC} - \Delta G_{AD} = 0$$

$$\Rightarrow -RT(\ln K_{AB} + \ln K_{BC} - \ln K_{DC} - \ln K_{AD}) = 0$$

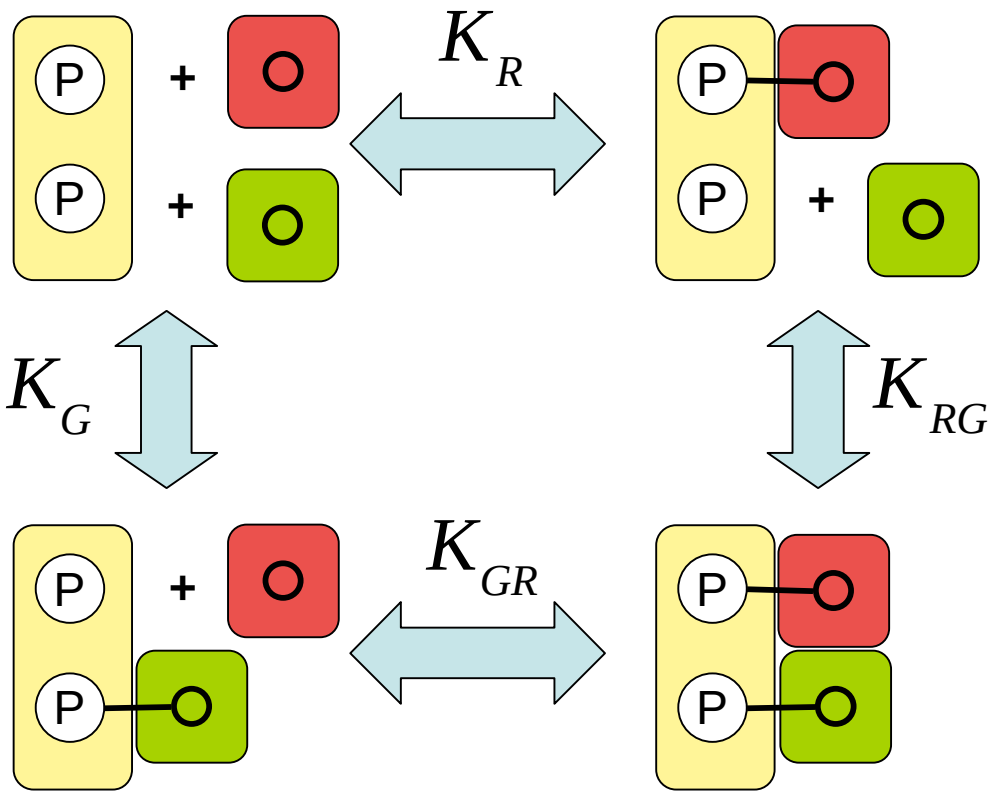
$$\Rightarrow K_{AB}K_{BC} / K_{DC}K_{AD} = 1$$

Kholodenko model has 5 such constraints, but some subsequent models have not enforced these.

See reference list on the q-bio wiki (Lecture 2, Bibilography and Links).

Worked example: cooperative binding to a scaffold

Xmas chile scaffold (XCeS) protein



$$K_R K_{RG} = K_G K_{GR}$$

"The enchilada is just as hot no matter which chile you eat first."

…but where's the SMOKING GUN?

Question is often raised: "Does the data available justify this complicated approach?"

We can argue with the question, but we are still looking for the definitive application where RBM is absolutely required and provides novel insight.

q-bio Model Inspection Program (aka Project 3)

“Looking for (Models of Mass Deception” (MMD)

Suspicious assumptions to look for (and test)

- *Sequential activation*
  - Particularly analyses whose results depend on such assumptions
- *Exclusive (one-at-a-time) interactions* or limits on the stoichiometry of complexes
- *Violations of principle of detailed balance*
  - Check model of Schoeberl et al. (*Nat. Biotechnol.*, 2002)

cellsignaling.lanl.gov

Michael Blinov
Jin Yang
Ambarish Nag
Michael Monine
Fangping Mu

Matthew Fricke
Leigh Fanning
Nathan Lemons
James Cavenaugh
Jeremy Kozdon
Paul Loriaux
Michelle Costa
Agate Ponder-Sutton
Michael Saelim
Jordan Atlas

Funding
NIH (NIGMS)
DOE
LANL-LDRD

William Hlavacek
Byron Goldstein*

http://bionetgen.lanl.gov

Collaborators
Ilona Reischl (NIH/FDA)
Henry Metzger (NIH)
Chikako Torigoe (Cornell)
Janet Oliver (UNM)
David Holowka(Cornell)
Barbara Baird (Cornell)

Dipak Barua (NC State)
Jason Haugh (NC State)

Josh Colvin (TGen)
Rich Posner (TGen)

Ty Thomson (MIT)

Nikolay Borisov
Boris Kholodenko