Rule-based Modeling

William S. Hlavacek

Theoretical Division

Los Alamos National Laboratory

Outline

1. **The motivation for rule-based modeling**
2. Basic concepts of rule-based modeling
3. An example model specification
4. Two methods for simulating a model
5. Suggested exercises

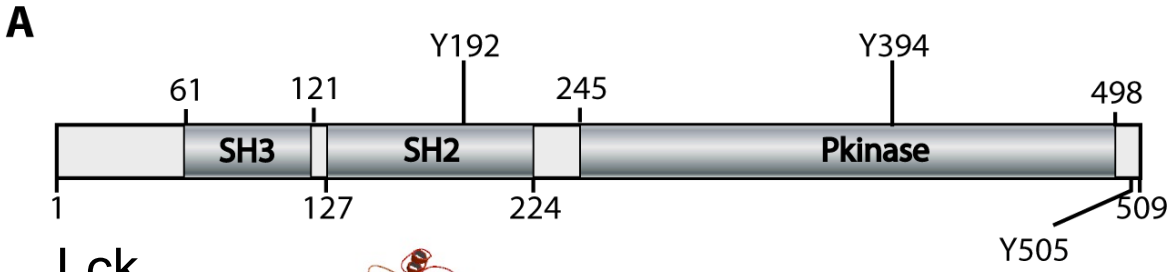
The need for predictive models of signal-transduction systems

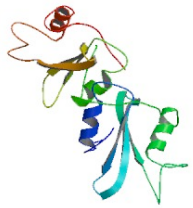
- **Molecular changes that affect cell signaling cause/sustain disease (cancer)**
- **Over 200 drugs that target signaling proteins are currently in clinical trials**
  - Spectacular success in some cases (Gleevec treatment of CML)
  - But results are largely disappointing for most patients
- **96 clinical trials are underway to test combinations of drugs (clinicaltrials.gov)**
  - There are too many combinations to consider all possibilities in trials

Value added by modeling

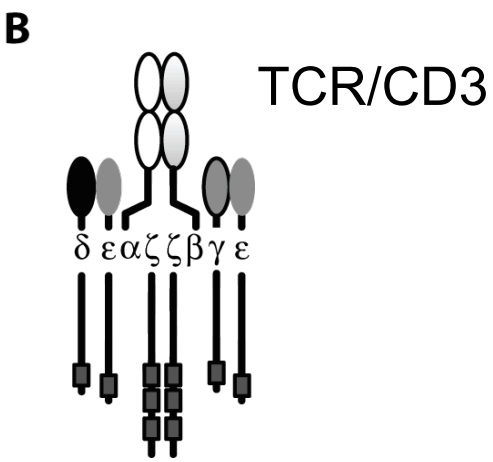
1. **We can use models to organize information about a system with precision**
2. **We can determine the logical consequences of a model specification**

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





Lck-SH2 (1bhh)

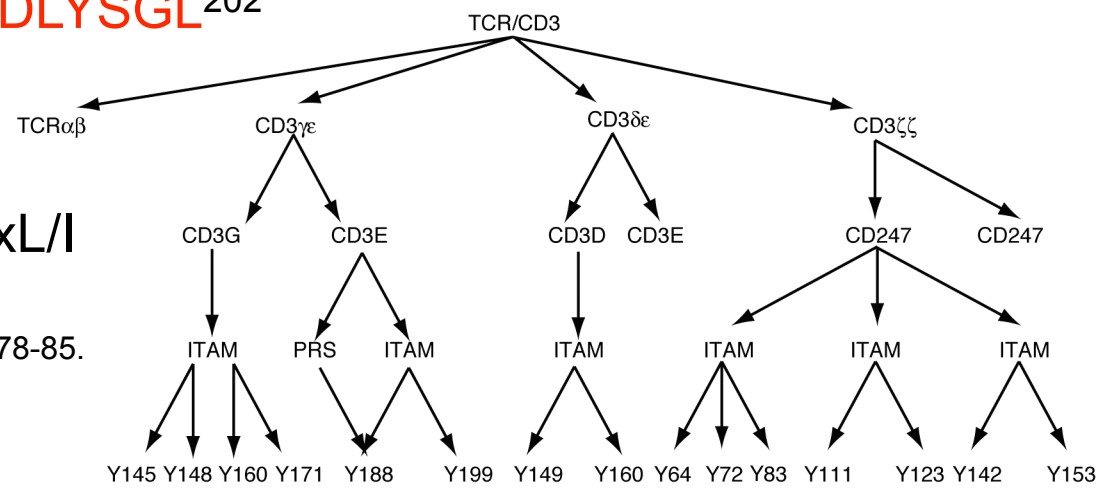


CD3E: ^{184}PNPDYEPIRKGQRDLYSGL^{202}

PRS: PxxDY

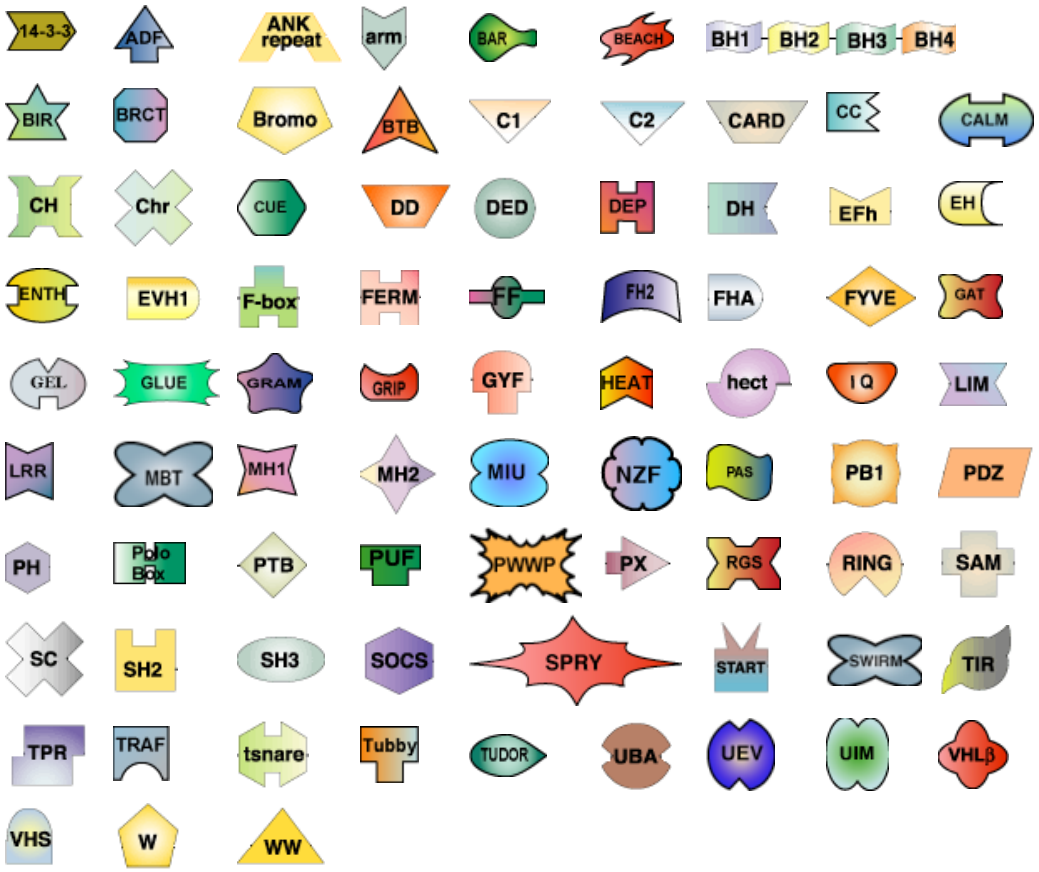
ITAM: YxxL/I(x_{6-8})YxxL/I

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



Los Alamos NATIONAL LABORATORY EST.1943

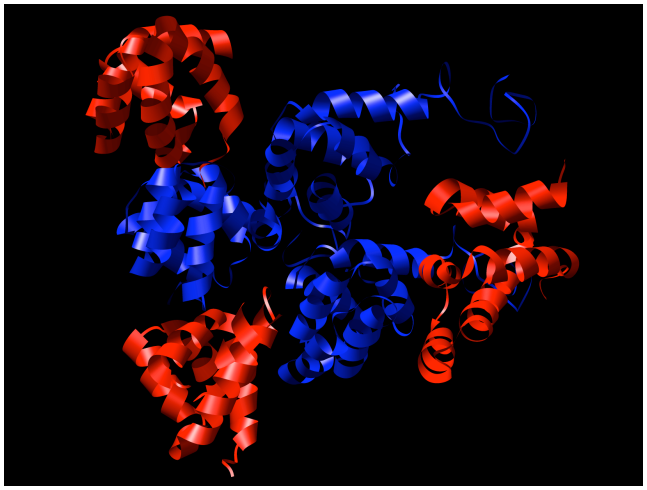
There are many protein interaction domains



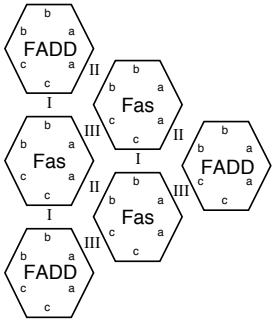
Los Alamos NATIONAL LABORATORY EST.1943

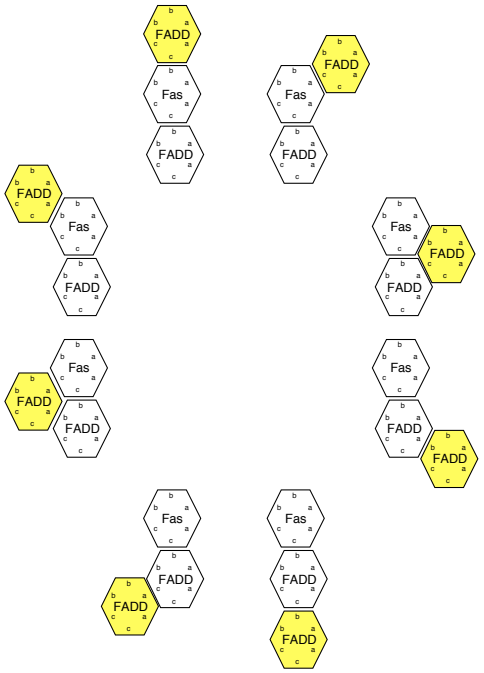
The Pawson Lab (http://pawsonlab.mshri.on.ca/)

Some domains are multivalent and mediate oligomerization via domain-domain interactions



A hexamer of death domains



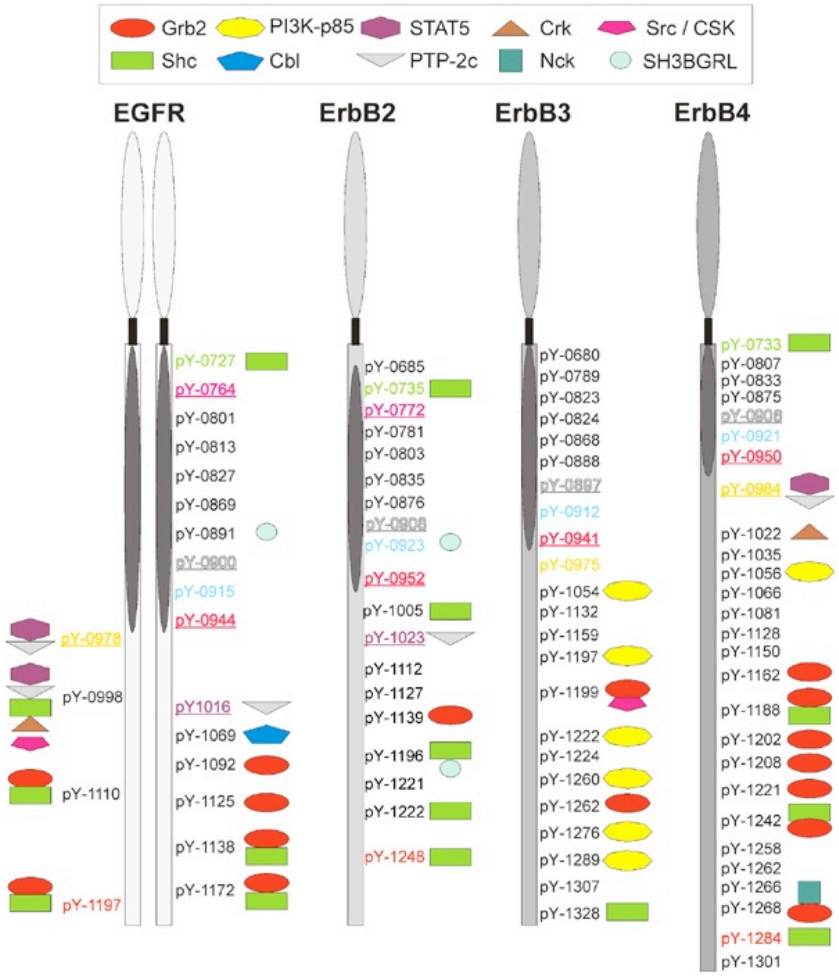


Weber and Vincenz (2001) *FEBS Lett.*

C.-T. Tung (Los Alamos)

There are many possible protein complexes!

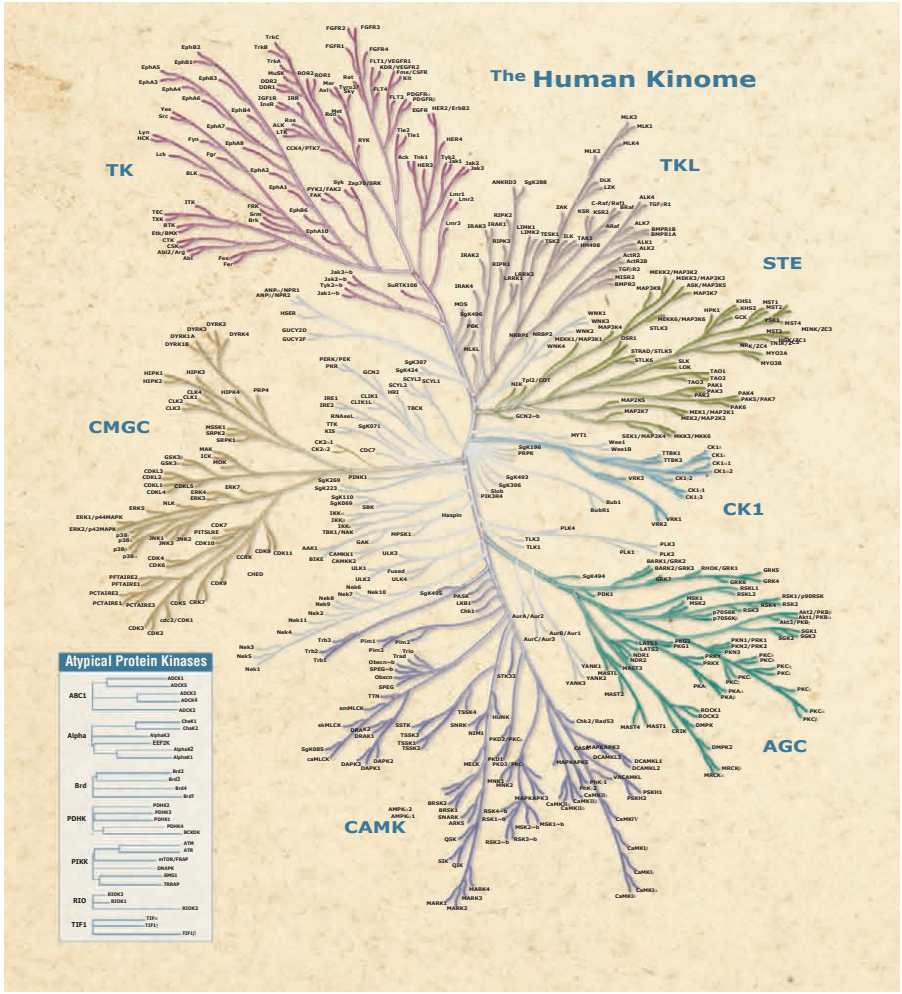
Domain-motif interactions are often controlled by post -translational modifications



There are many possible protein phosphoforms!

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

518 protein kinases (~2% of human genes)

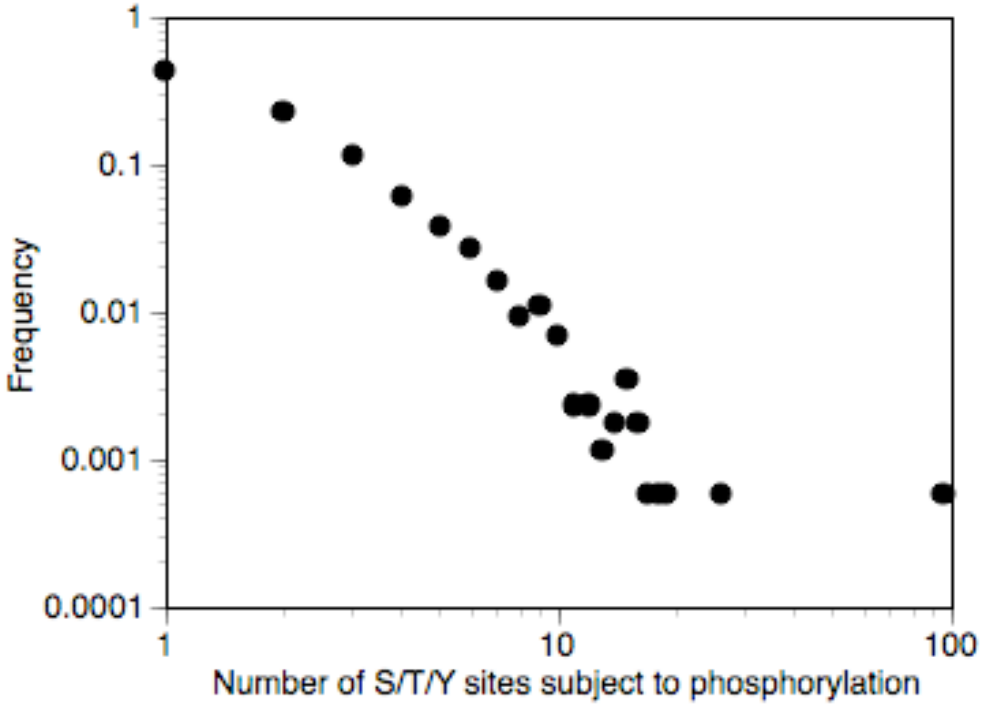


There are phosphatases too!

Manning G et al. (2002)
Science 298:1912-34.

Los Alamos NATIONAL LABORATORY EST.1943

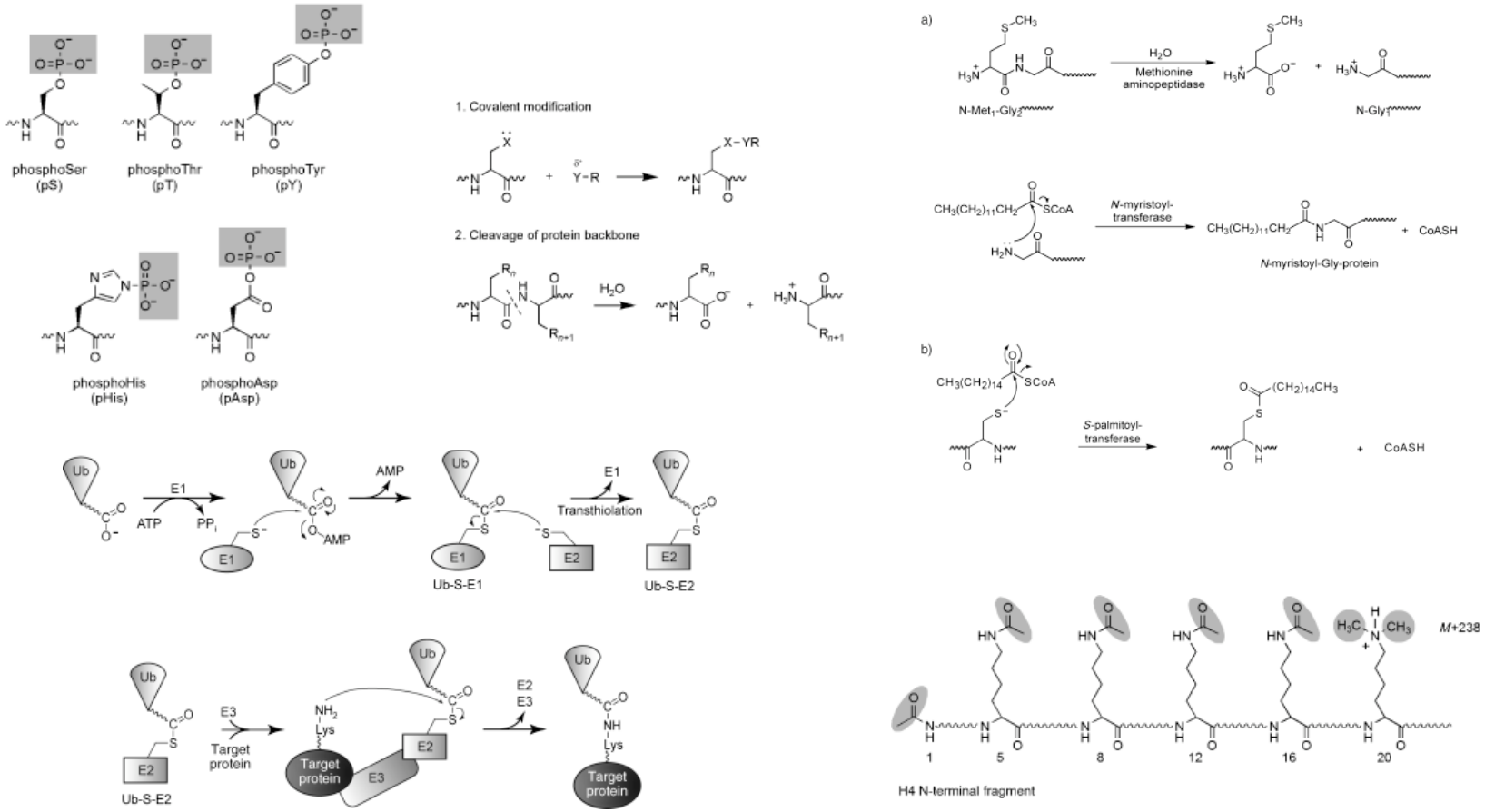
Signaling proteins typically contain multiple phosphorylation sites (S/T/Y)



> 50% are phosphorylated at 2 or more sites

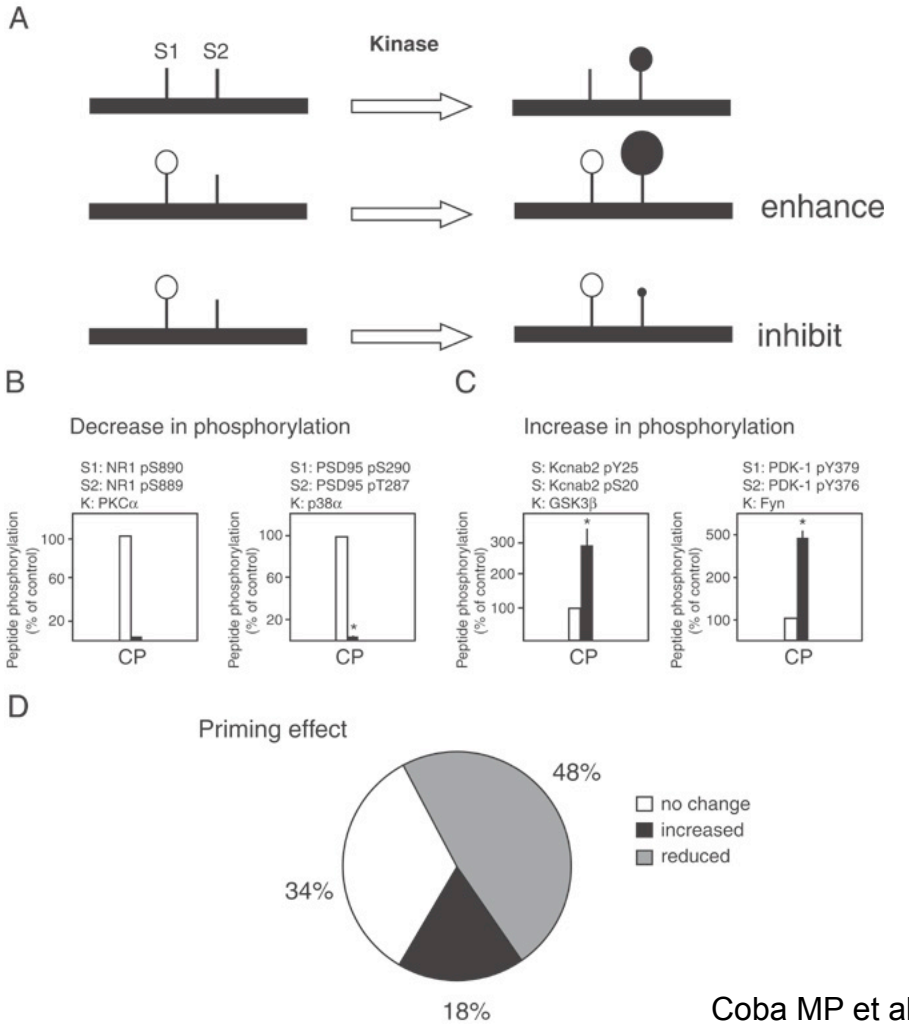
Phospho.ELM database v. 3.0 (http://phospho.elm.eu.org)

There are many different kinds of post-translational modifications of proteins



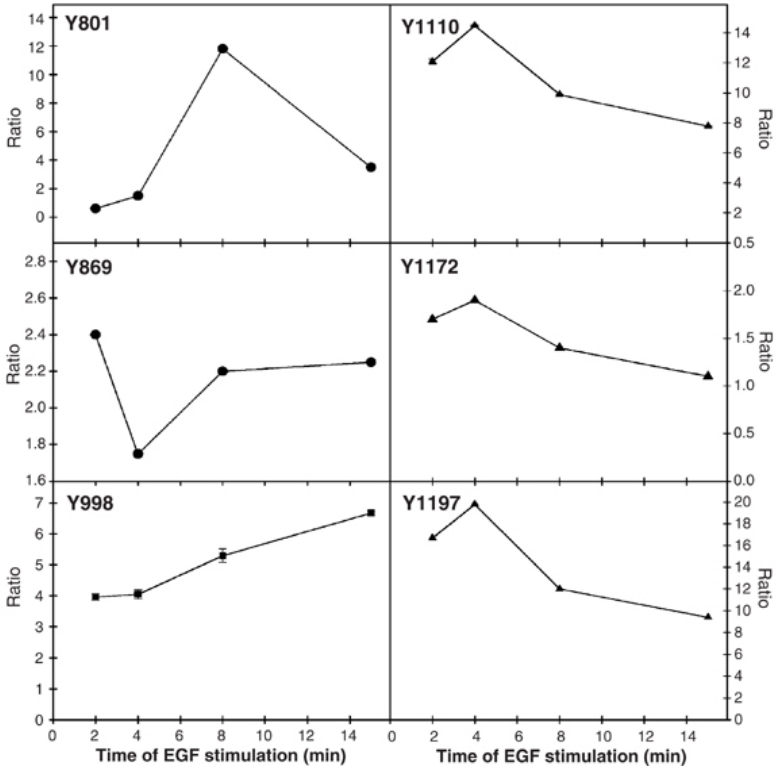
Walsh CT et al. (2005) Angew. Chem. Int. Ed. Engl. 44:7342-72.

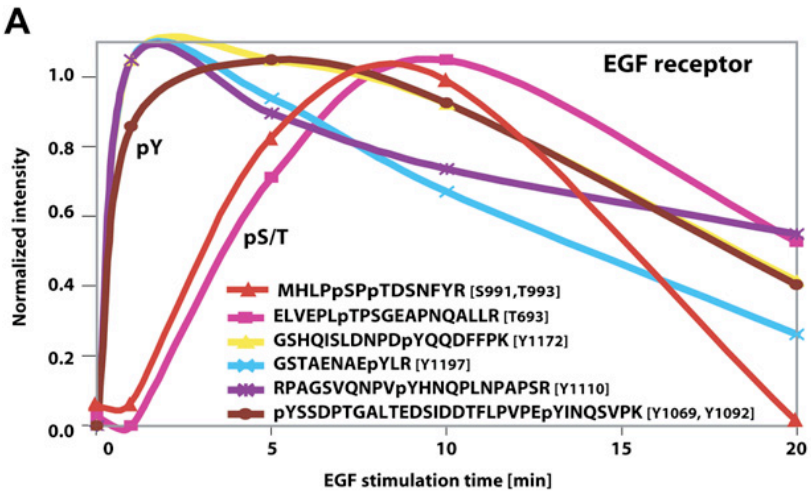
Priming – cooperative phosphorylation of neighboring kinase substrates is common



Coba MP et al. (2009) Sci. Signal.

Distinct time courses of phosphorylation for different amino acid residues within the same protein





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

Olsen JV et al. (2006) Cell 127:635-48.

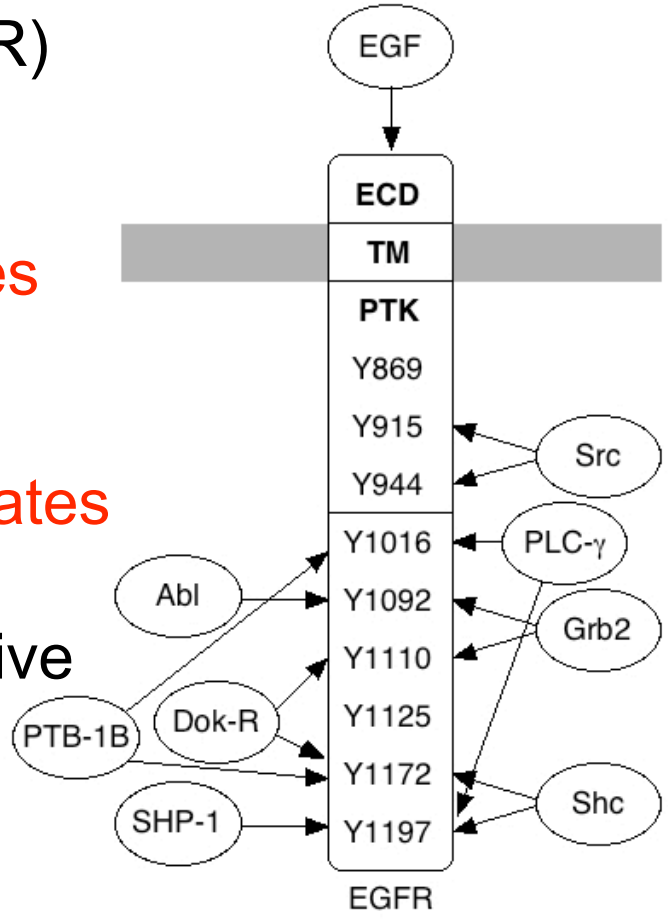
Combinatorial complexity – a serious problem for the conventional modeling approach

Epidermal growth factor receptor (EGFR)

9 sites => 2^9=512 phosphorylation states

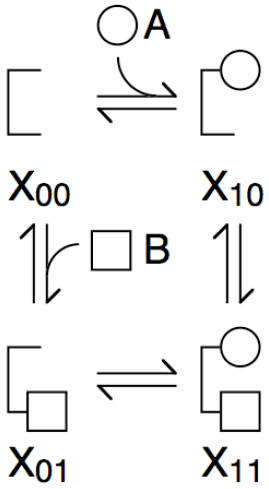
Each site has ≥ 1 binding partner
=> more than 3^9=19,683 total states

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



The textbook approach

Conventional representation of a biochemical reaction network



$$X_{00} + A \rightleftharpoons X_{10}$$
$$X_{00} + B \rightleftharpoons X_{01}$$
$$X_{10} + B \rightleftharpoons X_{11}$$
$$X_{01} + A \rightleftharpoons X_{11}$$

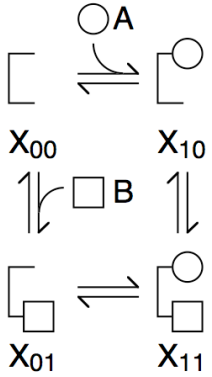
6 chemical species
8 unidirectional reactions

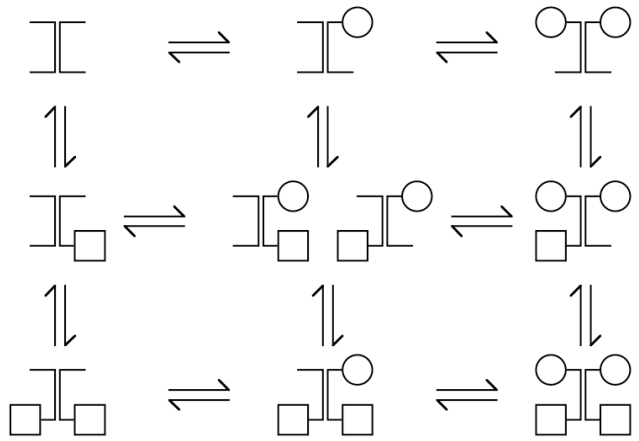
$$d[X_{00}]/dt = -k_{+1}[A][X_{00}] + k_{-1}[X_{10}] - k_{+2}[B][X_{00}] + k_{-2}[X_{01}]$$

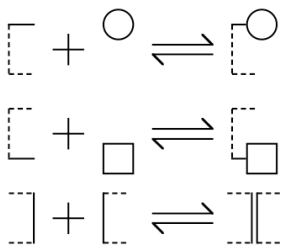
Network (model) size tends to grow nonlinearly (exponentially) with the number of molecular interactions in a system when molecules are structured

Network size increases nonlinearly when an extra interaction is considered

16 chemical species
60 unidirectional reactions



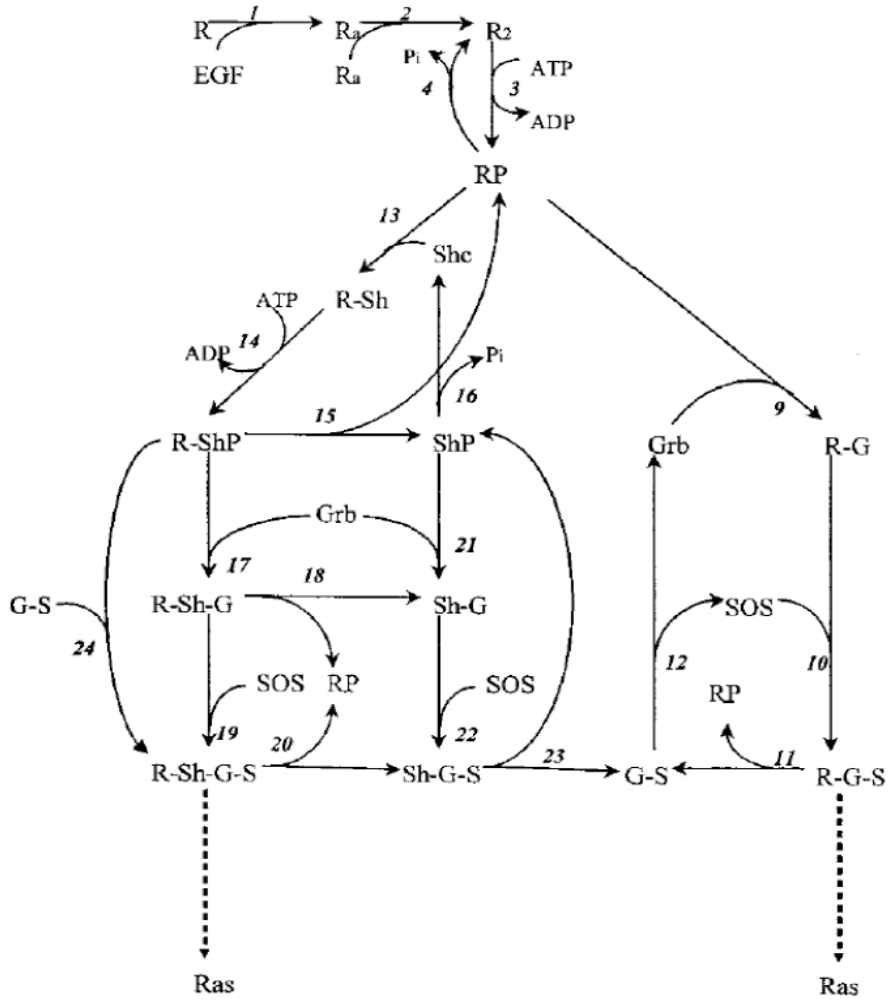




There are only three interactions. We can use a "rule" to model each of these interactions.

Science's STKE re6 (2006)

If you can write the model by hand, it may look like a mechanistic model, but it's probably just a complicated fitting function



A reaction scheme incorporated in many published models of EGFR signaling

Los Alamos NATIONAL LABORATORY EST.1943

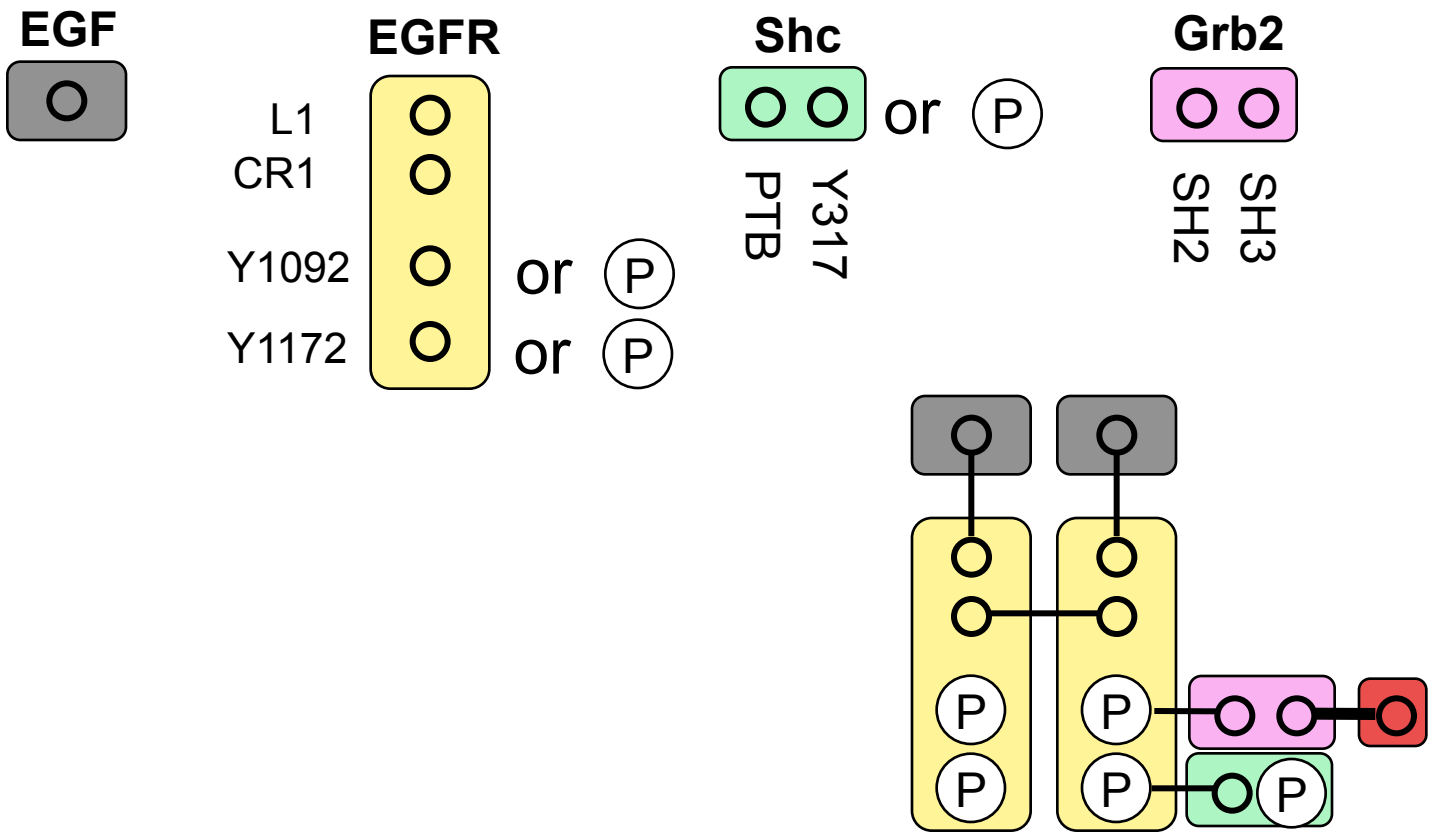
The problem of combinatorial complexity necessitates a new modeling approach

- **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 is designed to deal with this situation

Outline

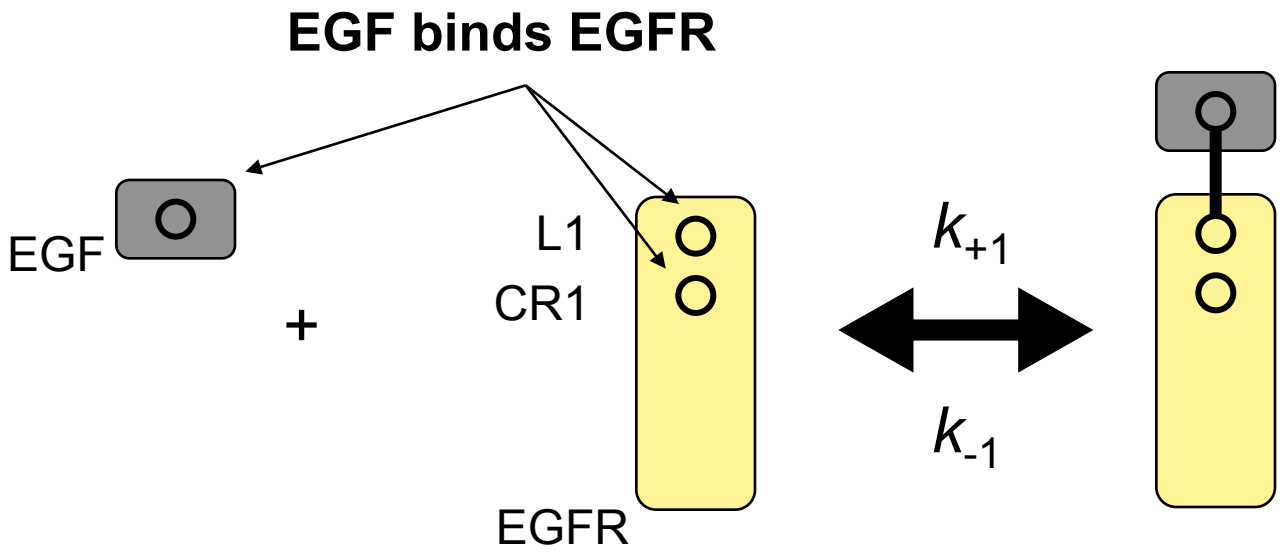
1. The motivation for rule-based modeling
2. **Basic concepts of rule-based modeling**
3. An example model specification
4. A survey of methods for simulating a model
5. Suggested exercises

Structured objects are naturally represented by graphs, so we use graphs to represent molecules and molecular complexes in signal-transduction systems



Blinov ML et al. (2006) *BioSystems*

Use graph-rewriting rules to represent interactions




```
begin reaction rules

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

end reaction rules
```

Outline

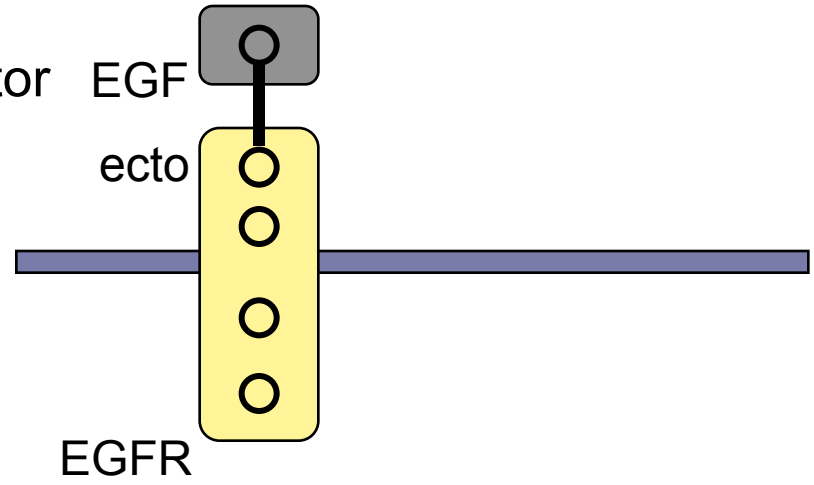
1. The motivation for rule-based modeling
2. Basic concepts of rule-based modeling
3. **An example model specification**
4. Two methods for simulating a model
5. Suggested exercises

Early events in EGFR signaling, illustrated with the same (sub)graphs used to specify a rule-based model for these events

EGF = epidermal growth factor

EGFR = epidermal growth factor receptor

1. EGF binds EGFR

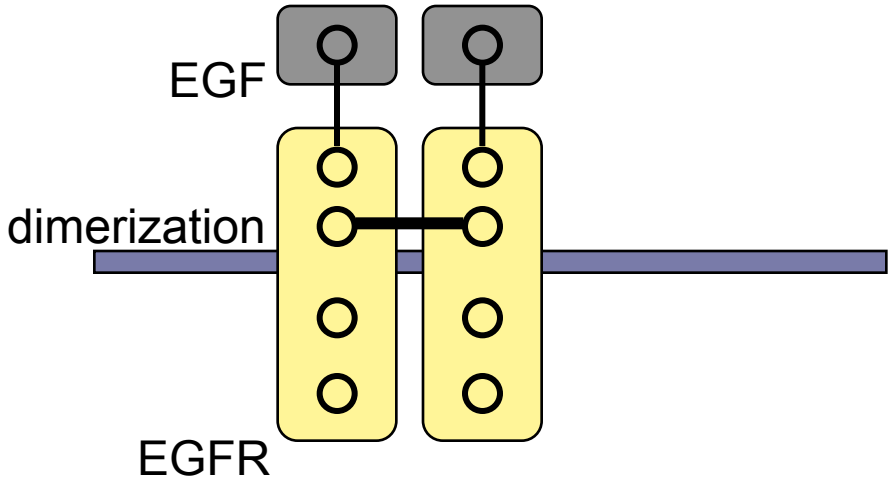


Los Alamos NATIONAL LABORATORY EST.1943

Blinov ML et al. (2006) *BioSystems*

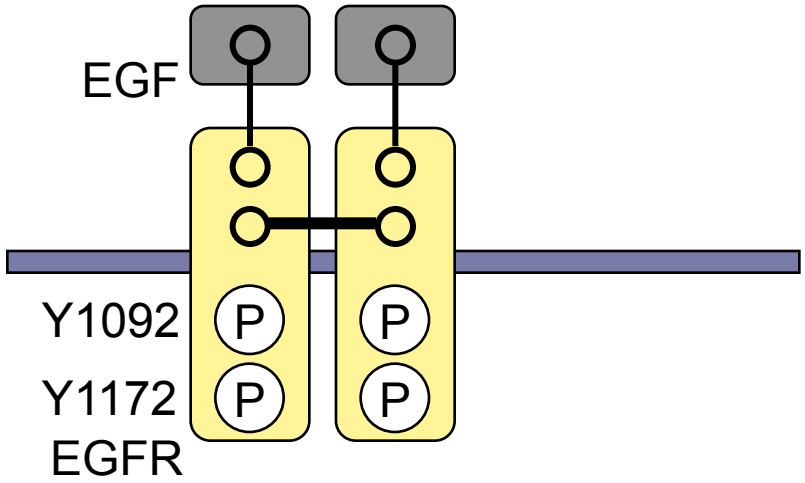
Early events in EGFR signaling

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



Early events in EGFR signaling

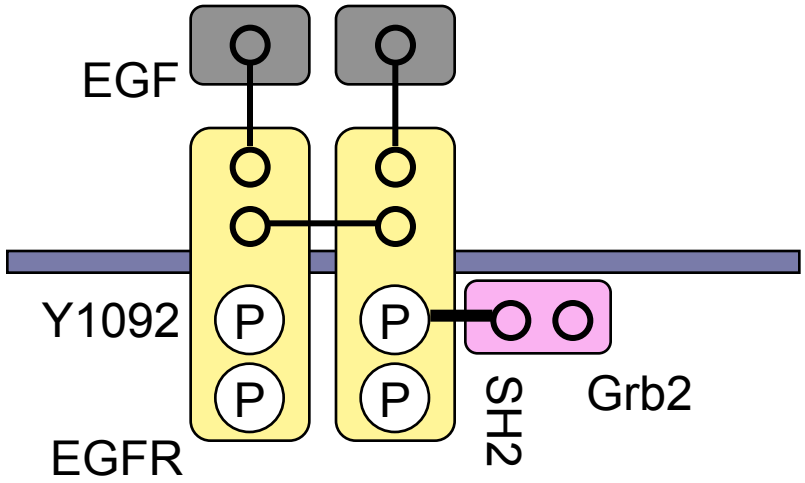
1. EGF binds EGFR
2. EGFR dimerizes
3. **EGFR transphosphorylates a copy of itself**



Early events in EGFR signaling

Grb2 pathway

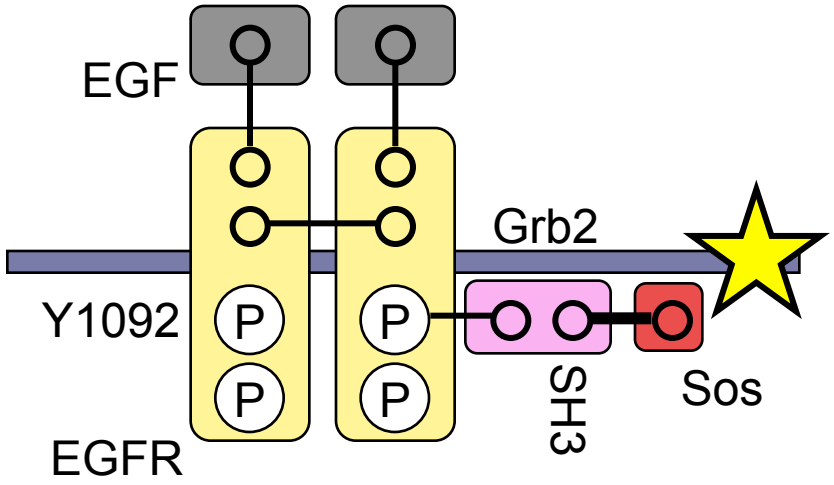
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Grb2 binds phospho-EGFR**



Early events in EGFR signaling

Grb2 pathway

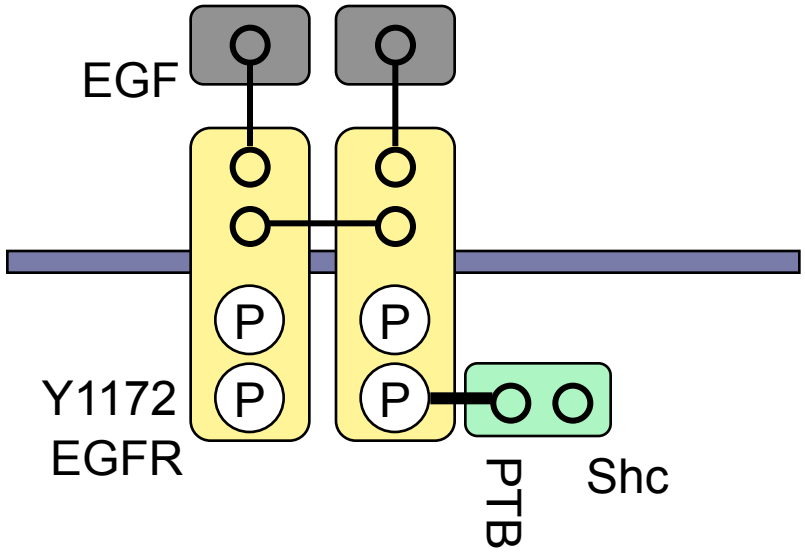
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Grb2 binds phospho-EGFR
5. **Sos binds Grb2 (Activation Path 1)**



Early events in EGFR signaling

Shc pathway

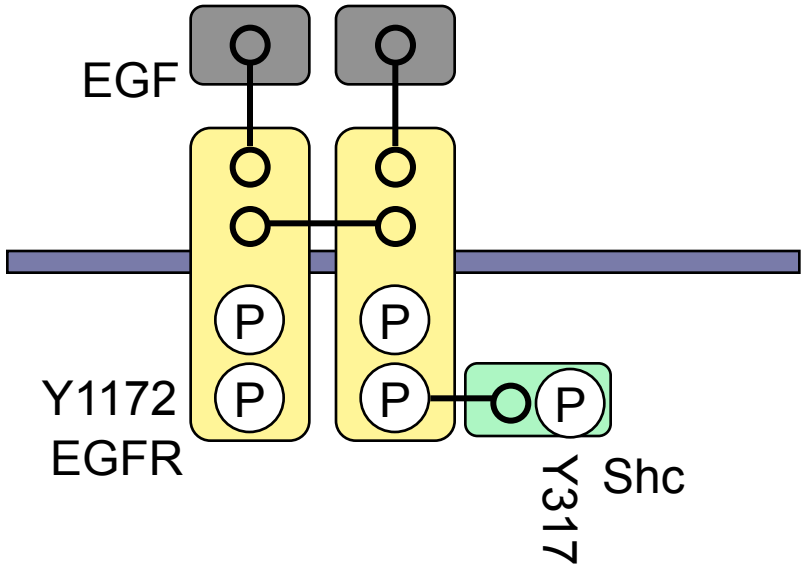
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Shc binds phospho-EGFR**



Early events in EGFR signaling

Shc pathway

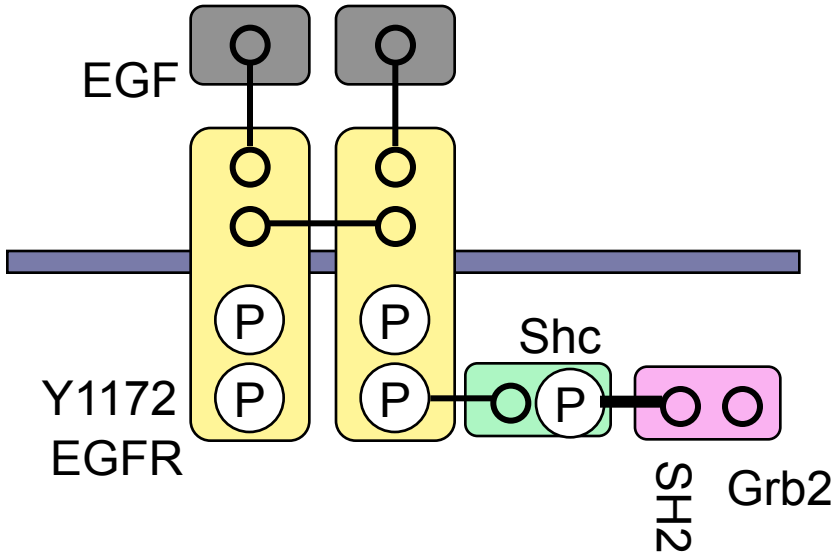
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. **EGFR transphosphorylates Shc**



Early events in EGFR signaling

Shc pathway

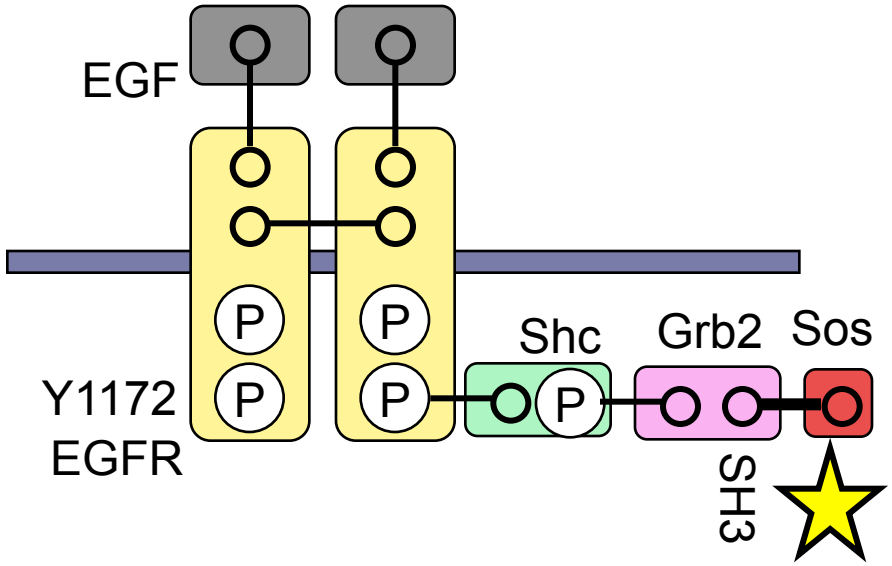
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. **Grb2 binds phospho-Shc**



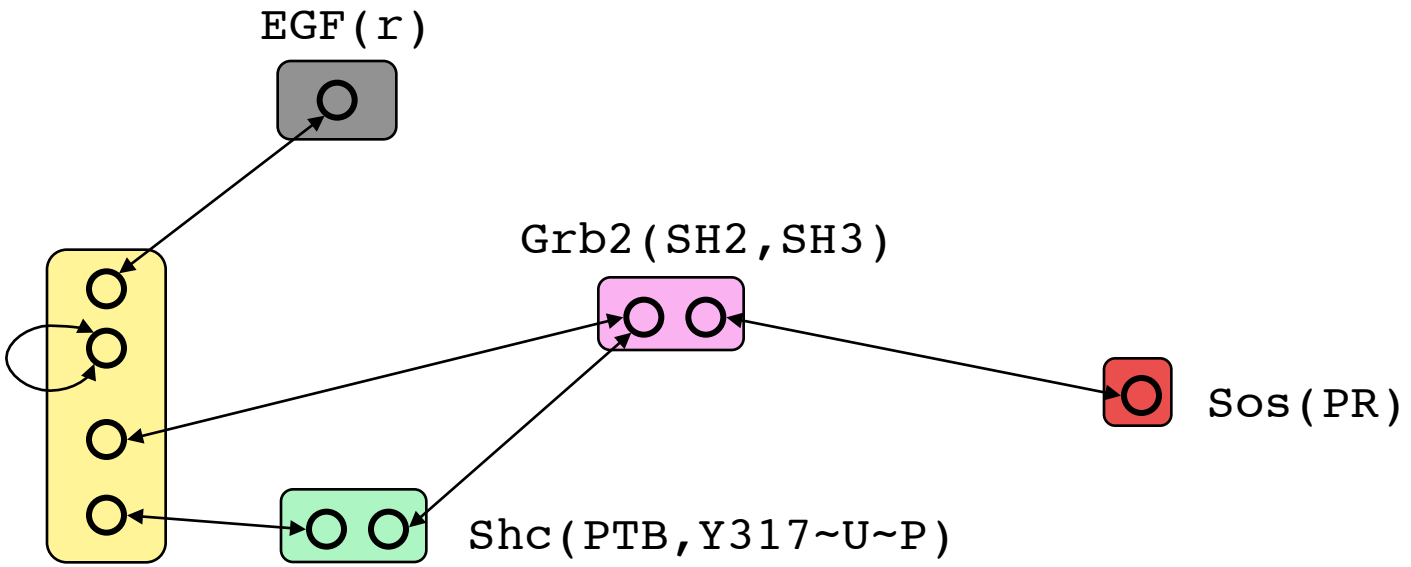
Early events in EGFR signaling

Shc pathway

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



Summary of molecules and their interactions in a simple model of early events in EGFR signaling

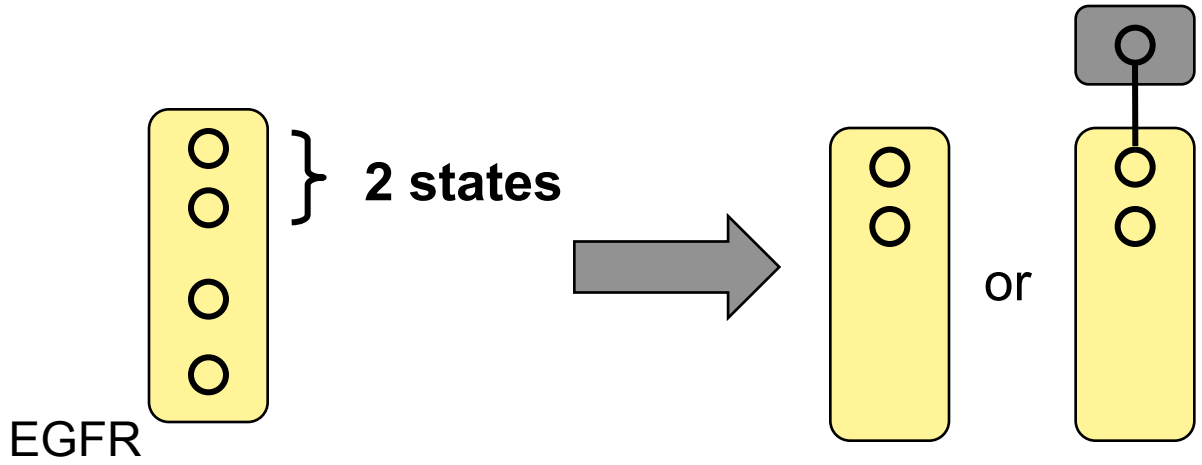


EGFR(l,d,Y1092~U~P,Y1172~U~P)

Blinov et al. (2006)

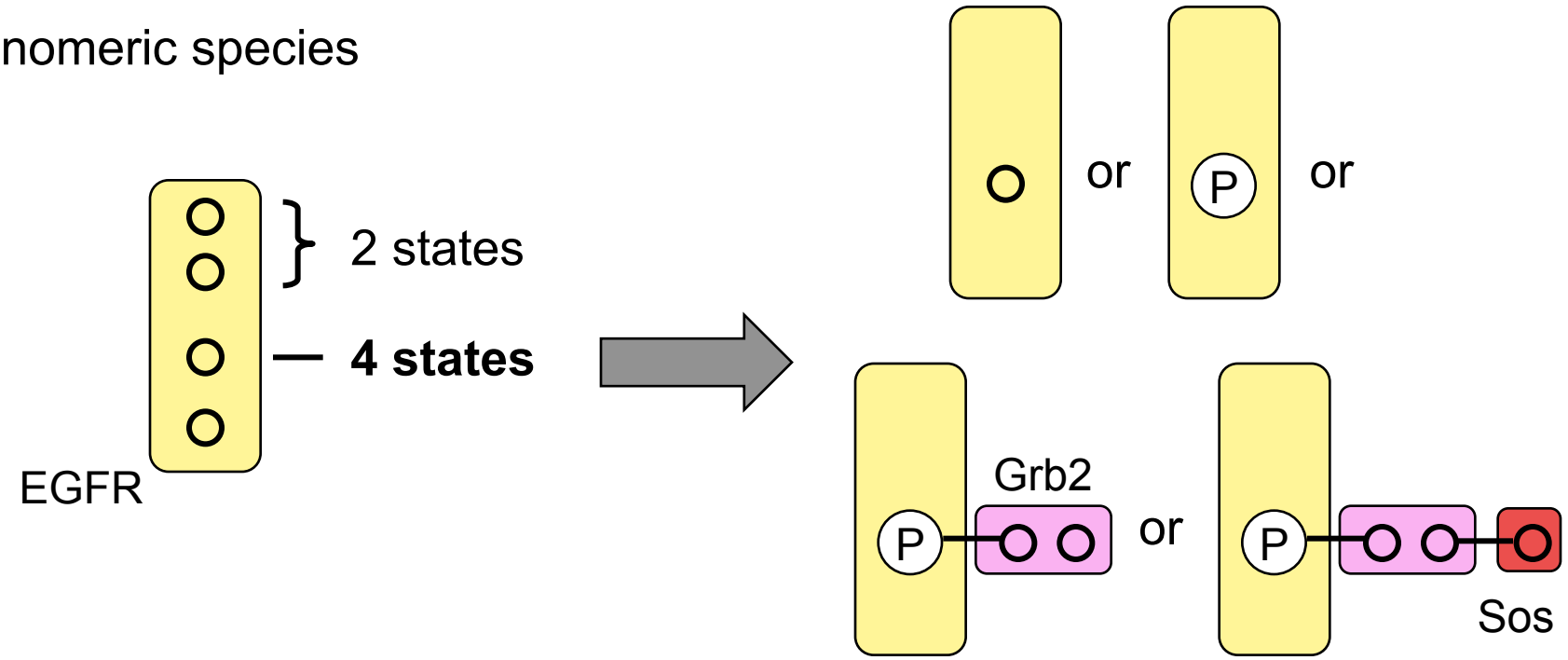
Combinatorial complexity of early events

Monomeric species

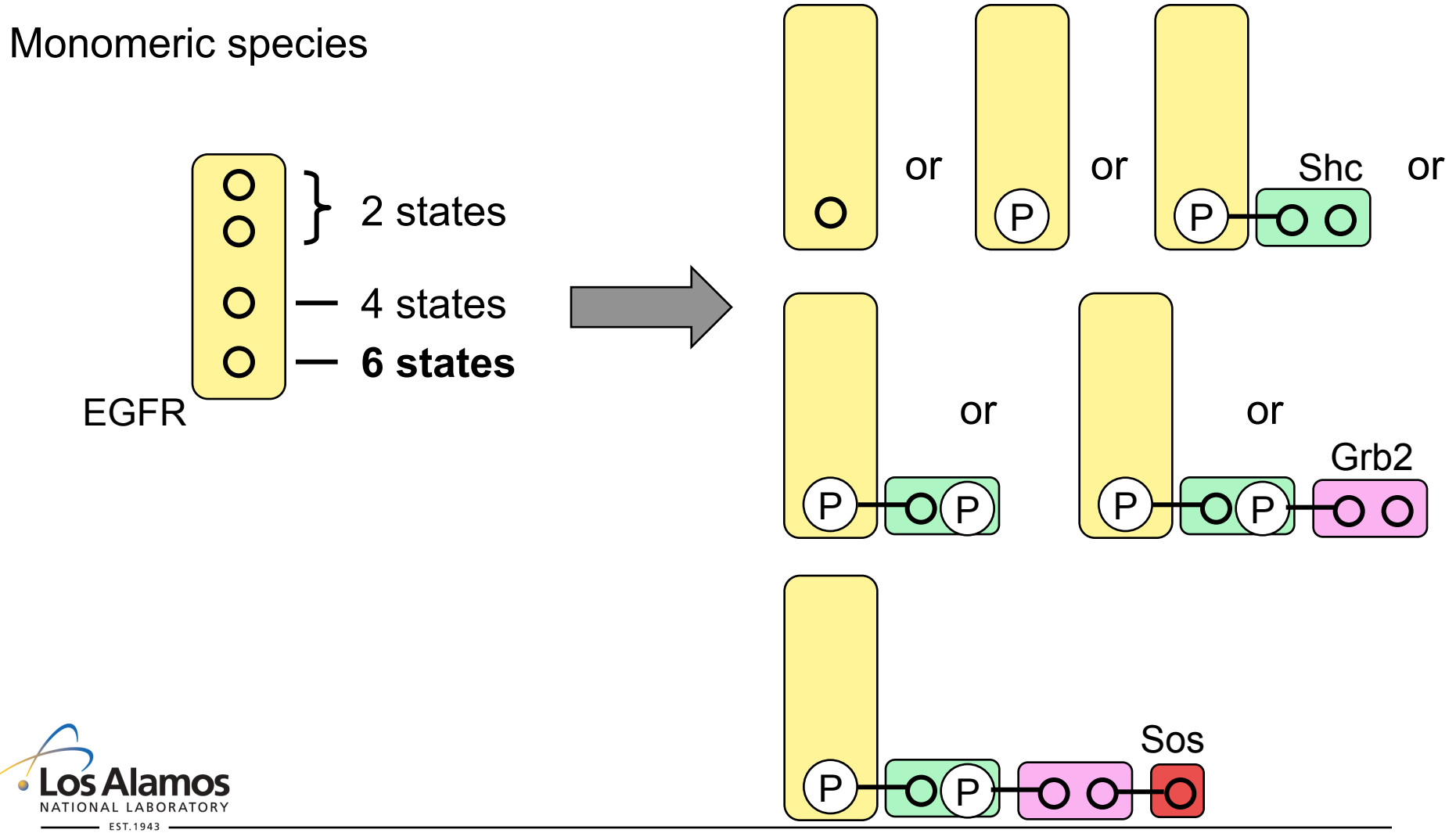


Combinatorial complexity of early events

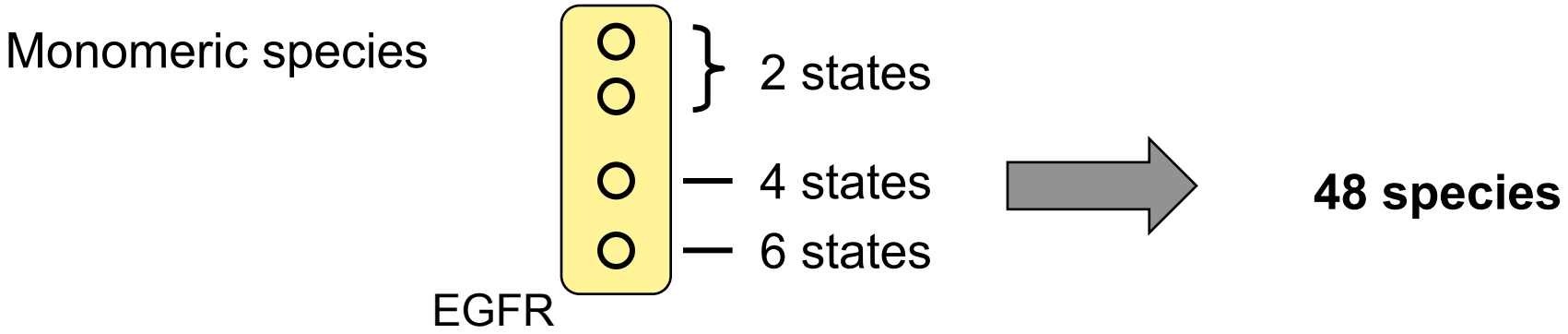
Monomeric species



Combinatorial complexity of early events

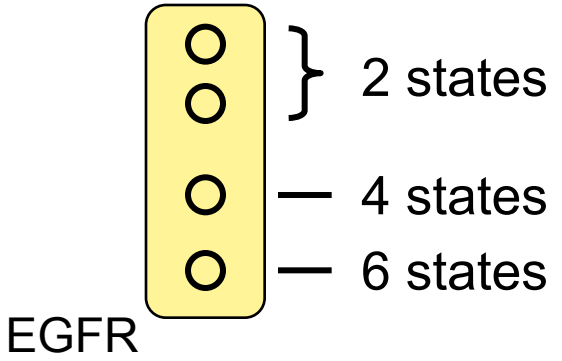


Combinatorial complexity of early events



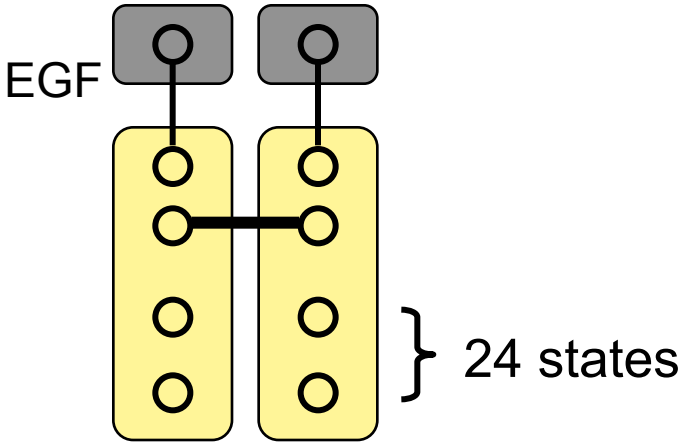
Combinatorial complexity of early events

Monomeric species



48 species

Dimeric species



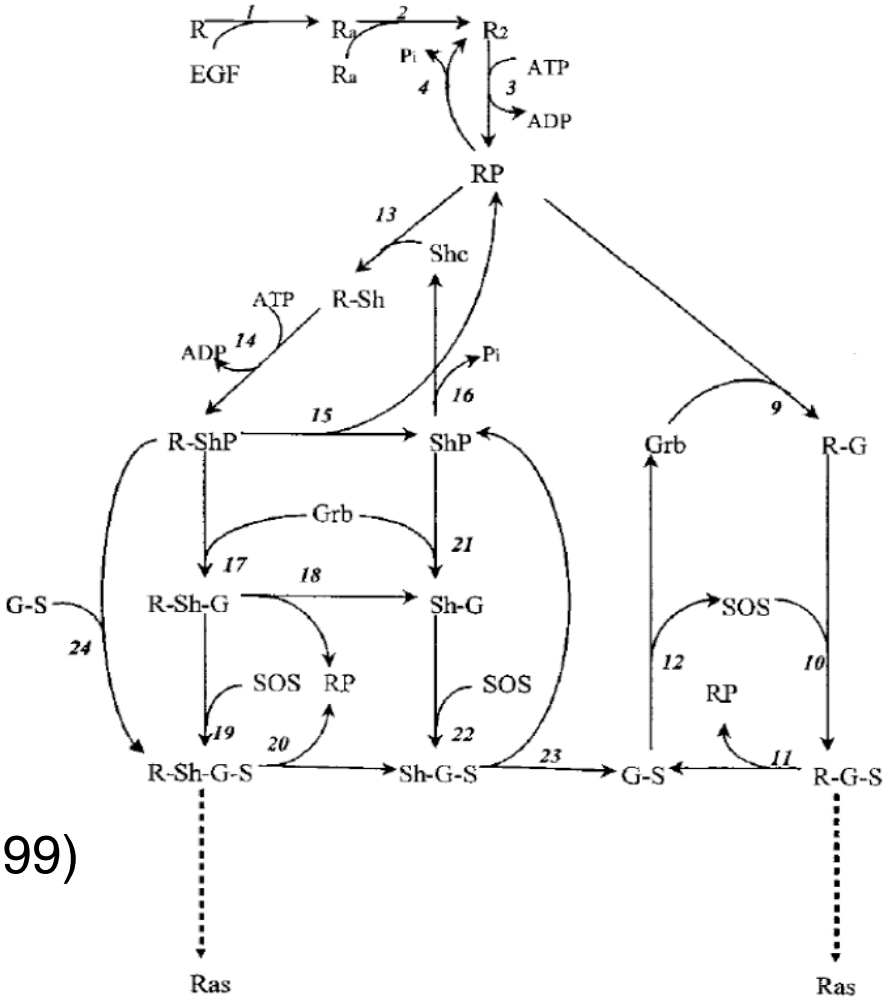
$N(N+1)/2$ = **300 species**

A conventional model for EGFR signaling

The Kholodenko model*

5 proteins

18 species
34 reactions



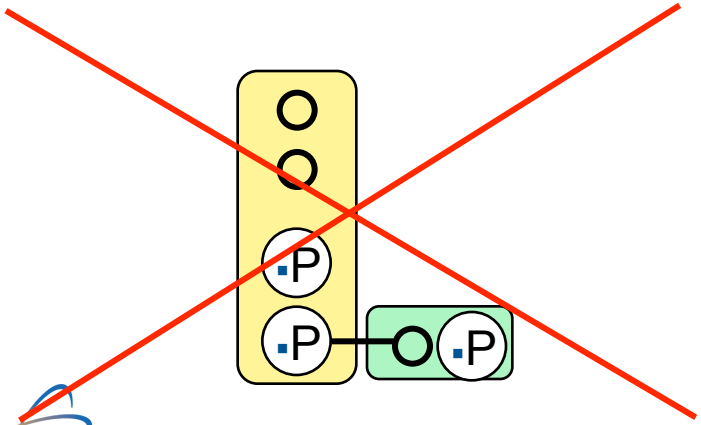
J. Biol. Chem. **274**, 30169 (1999)

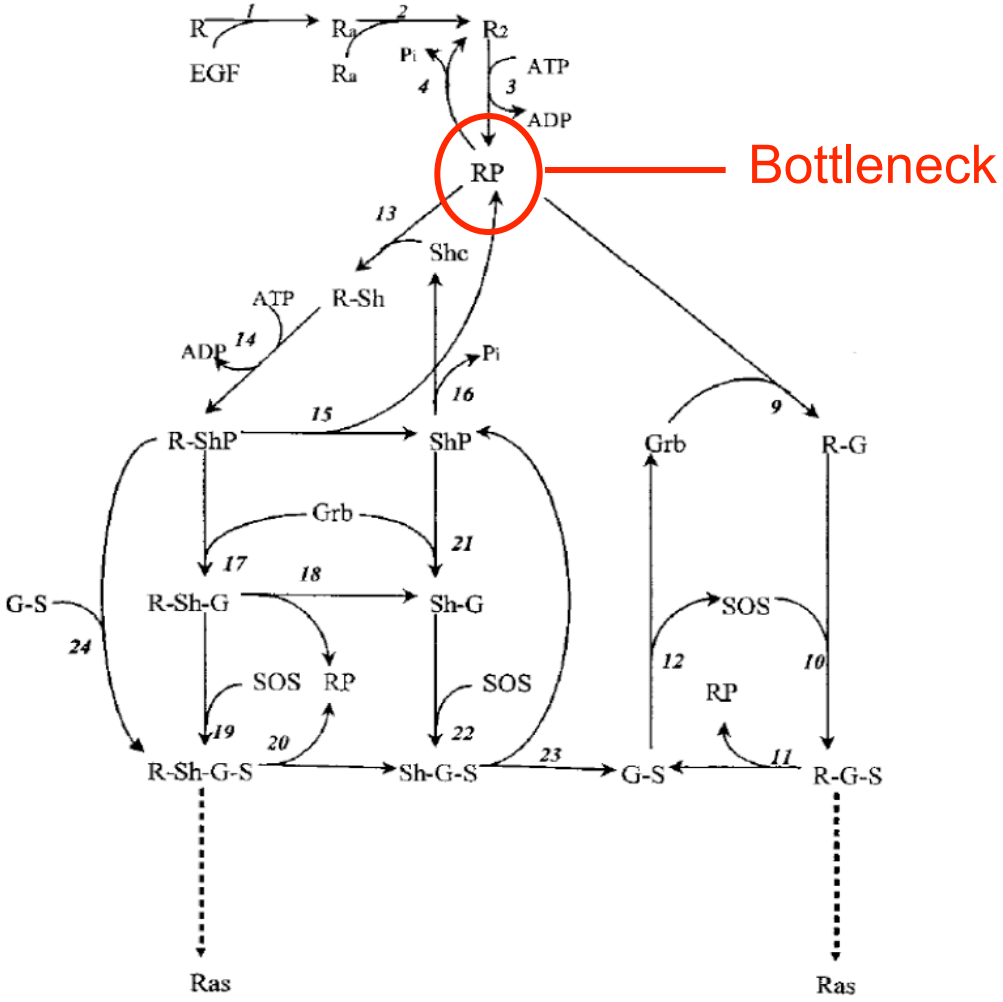
Los Alamos NATIONAL LABORATORY EST.1943

Assumptions made to limit combinatorial complexity

1. Phosphorylation inhibits dimer breakup

No modified monomers



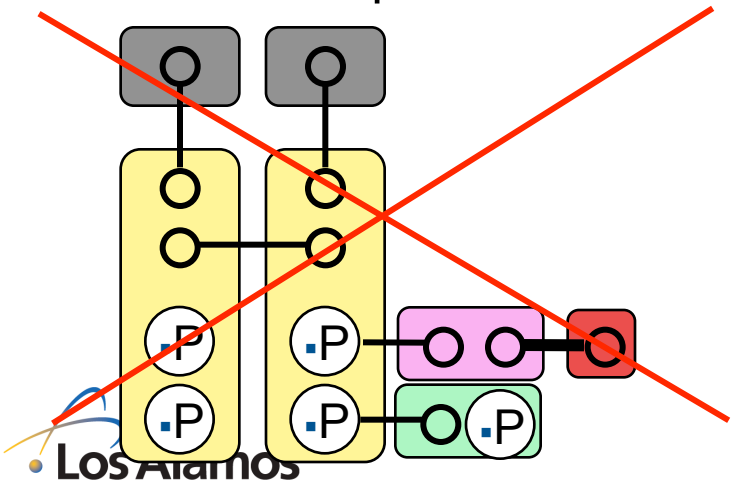


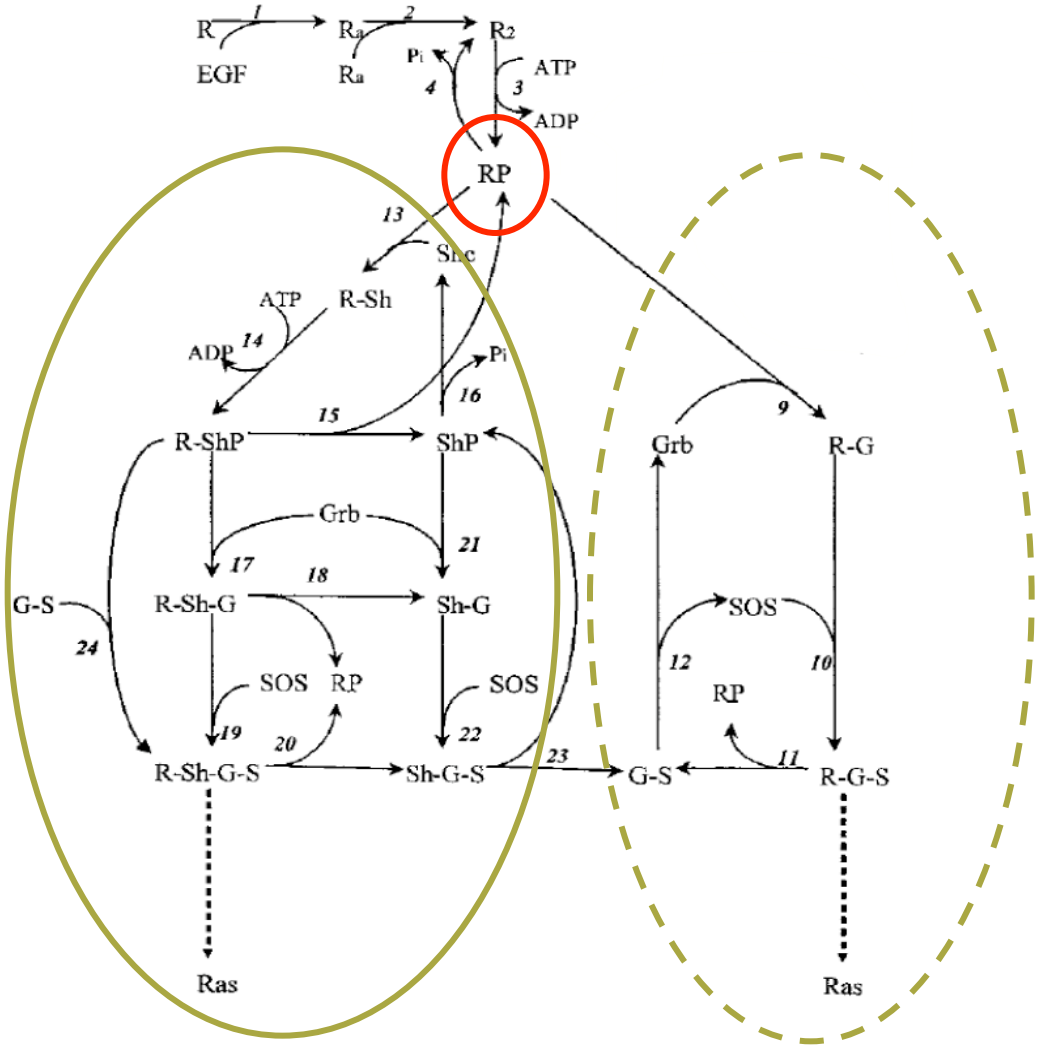
Los Alamos NATIONAL LABORATORY EST.1943

Assumptions made to limit combinatorial complexity

2. Adaptor binding is competitive

No dimers with more than one associated adapter





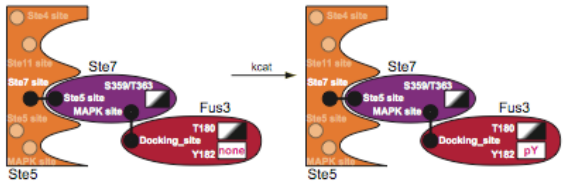
Reminders

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

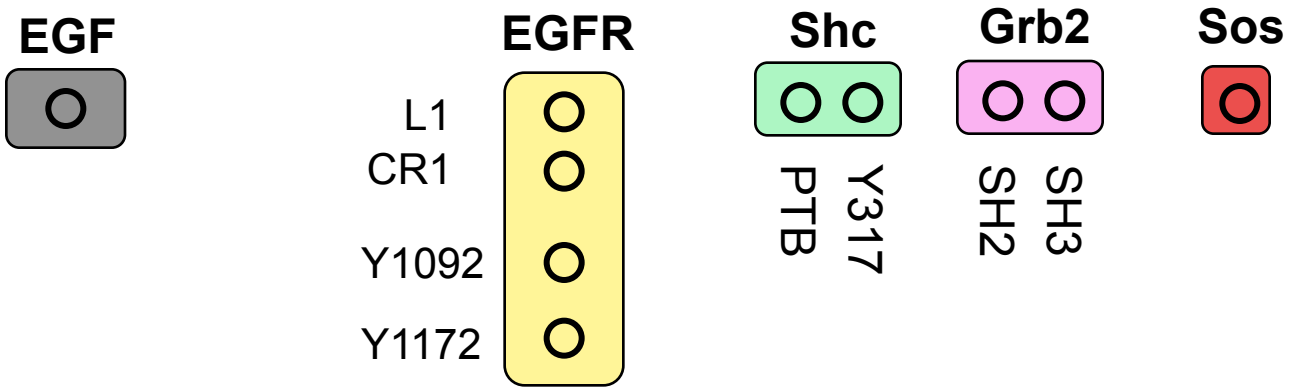


Ste7	Fus3	kcat
S359/T363	T180	
pS	none	kcat_Ste5Ste7*pS*Fus3_pY
pS	pT	kcat_Ste5Ste7*pS*Fus3*pT*_pY
pSpT	none	kcat_Ste5Ste7*pSpT*Fus3_pY
pSpT	pT	kcat_Ste5Ste7*pSpT*Fus3*pT*_pY

Ty Thomson (MIT) - yeastpheromonemodel.org

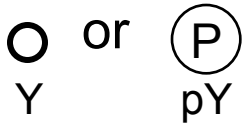
Molecules are modeled as graphs

Molecules

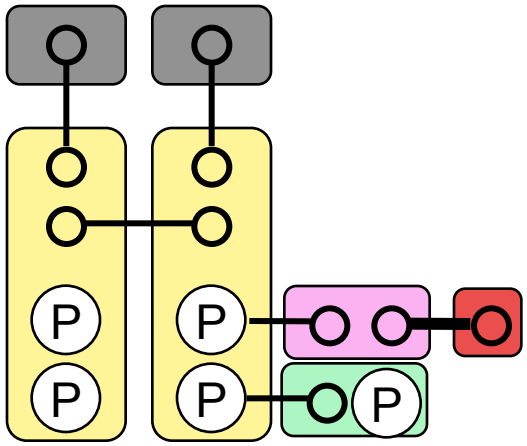


Nodes represent components of proteins

Y components may have labels:



Molecular complexes are simply connected molecules



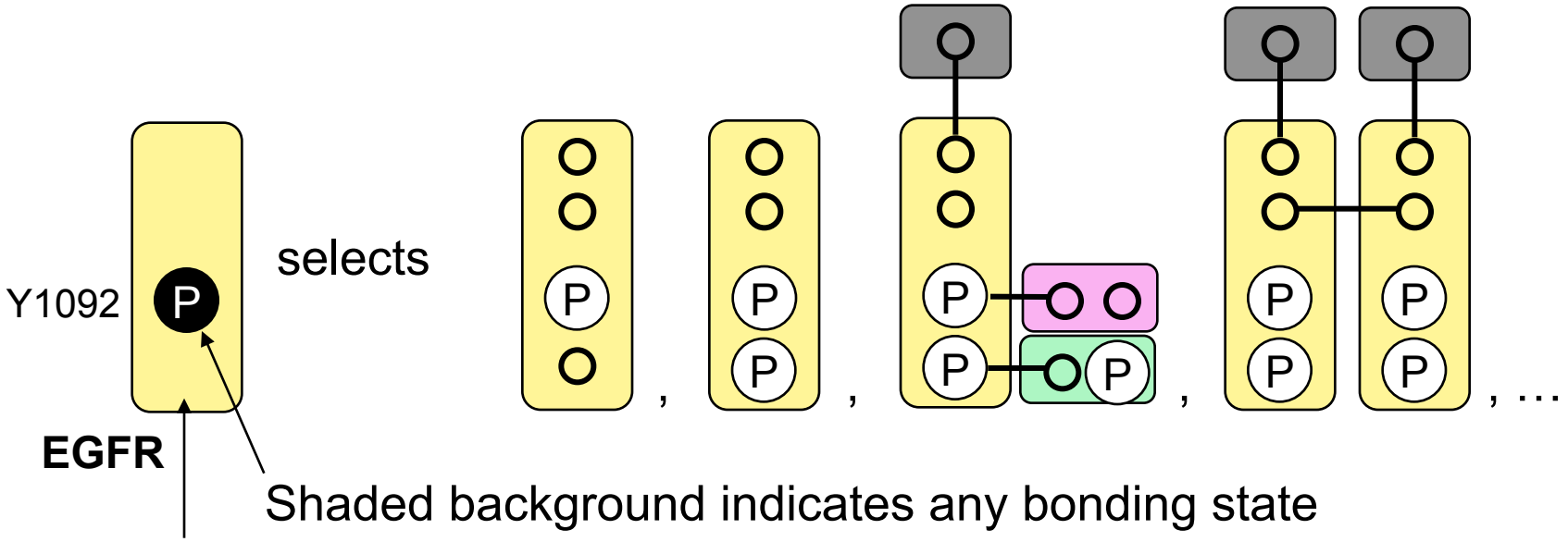
No need to introduce a unique name (e.g., X_{123} or ShP-RP-G-Sos) for each chemical species, as in conventional modeling

Edges represent bonds between components

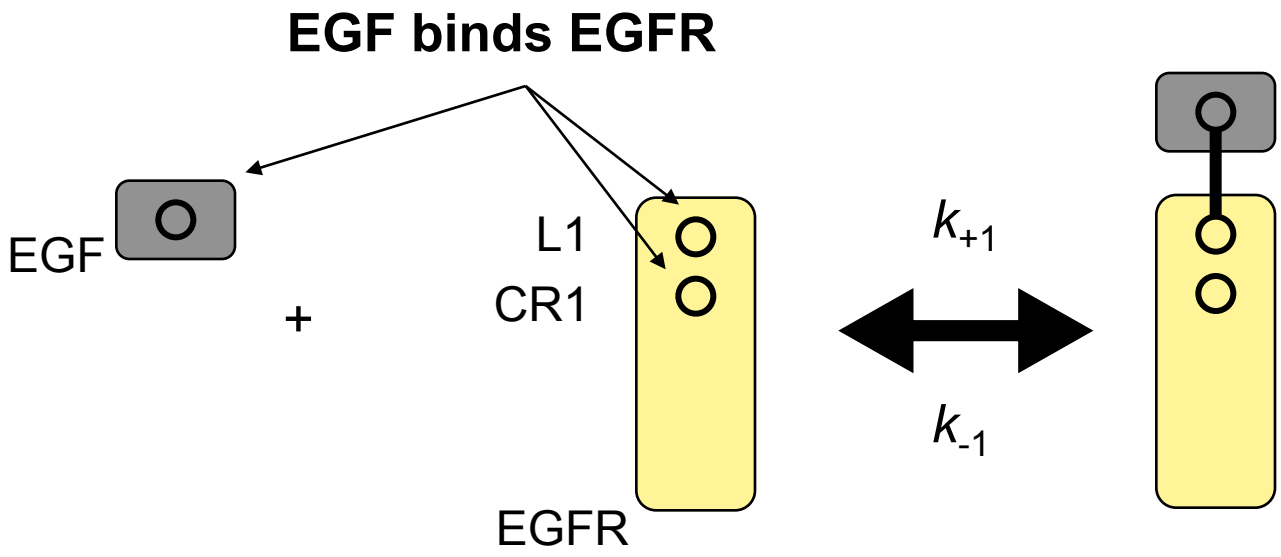
Bonds may be intra- or intermolecular

Patterns (subgraphs) define sets of chemical species with common features

A pattern that matches EGFR phosphorylated at Y1092



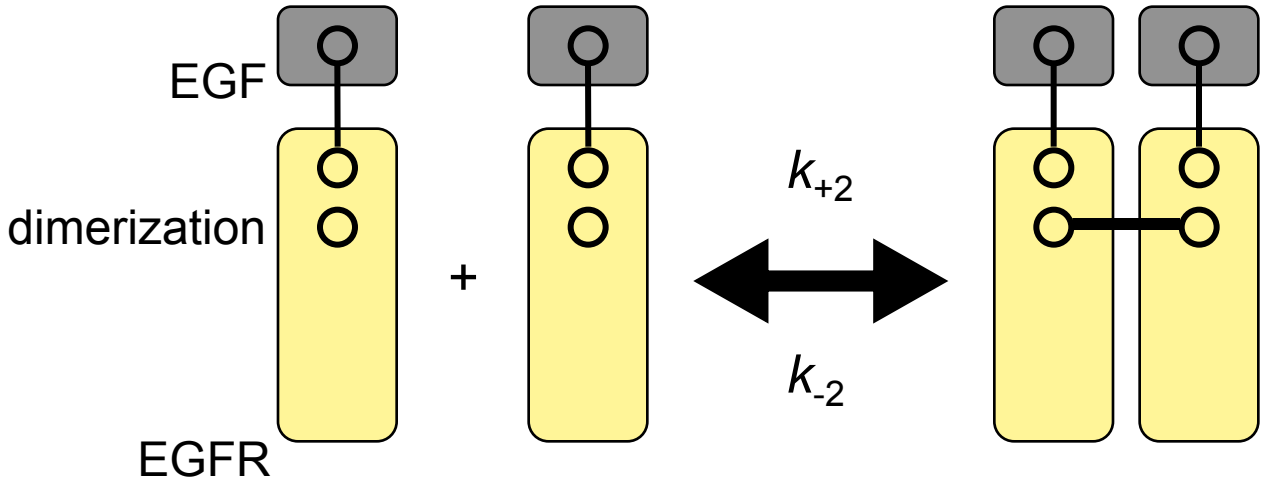
A reaction rule, composed of patterns, defines a class of reactions



Patterns select reactants (by matching graphs representing chemical species) and specify a transformation of the graphs representing reactants - **Addition of bond between EGF and EGFR in this case**

Dimerization rule eliminates previous assumption restricting breakup of receptors

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



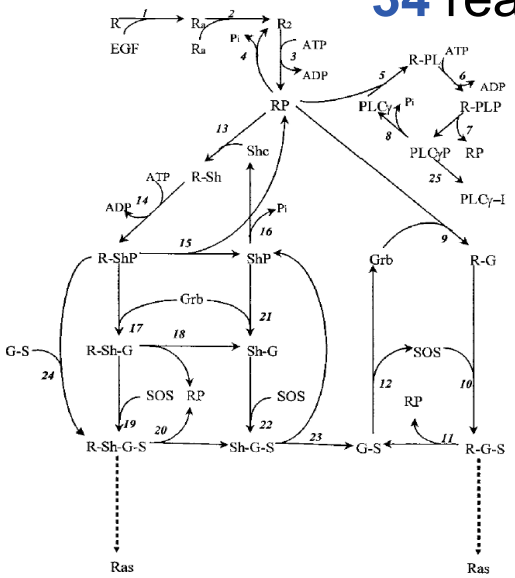
No free lunch: According to this rule, dimers form and break up with the same fundamental rate constants regardless of the states of cytoplasmic domains, which is an idealization.

Rule-based version of the Kholodenko model

- **5 molecule types**
- **23 reaction rules**
- **No new rate parameters! – Q: How? A: a rule provides a coarse-grained description of the reactions implied by the rule. All these reactions are parameterized by the same fundamental rate constant(s).**

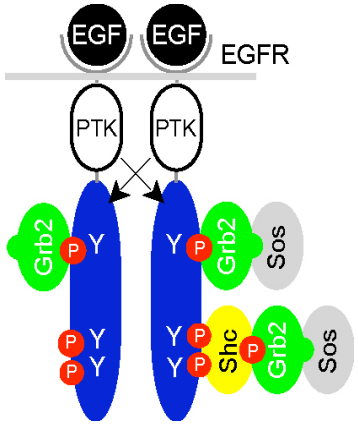
18 species

34 reactions



356 species

3749 reactions

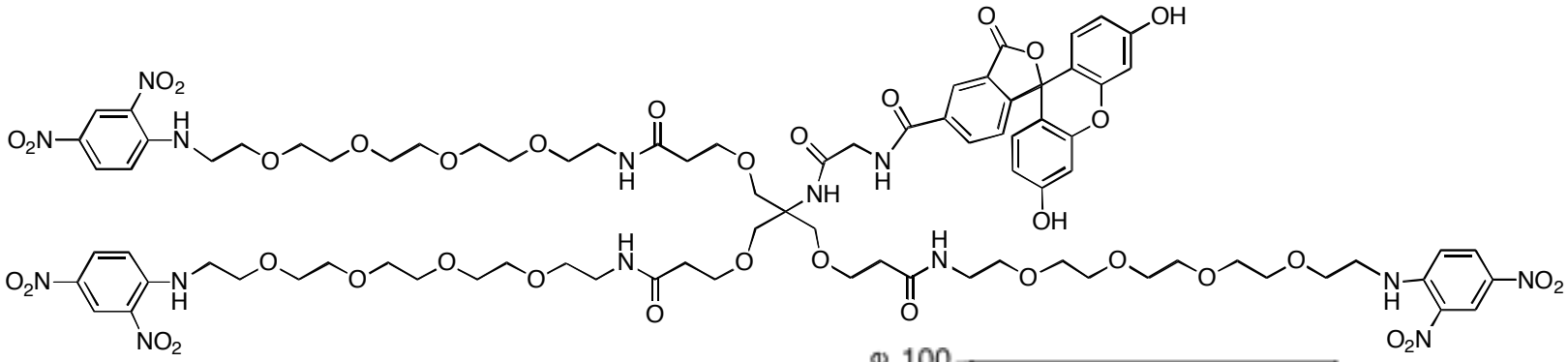


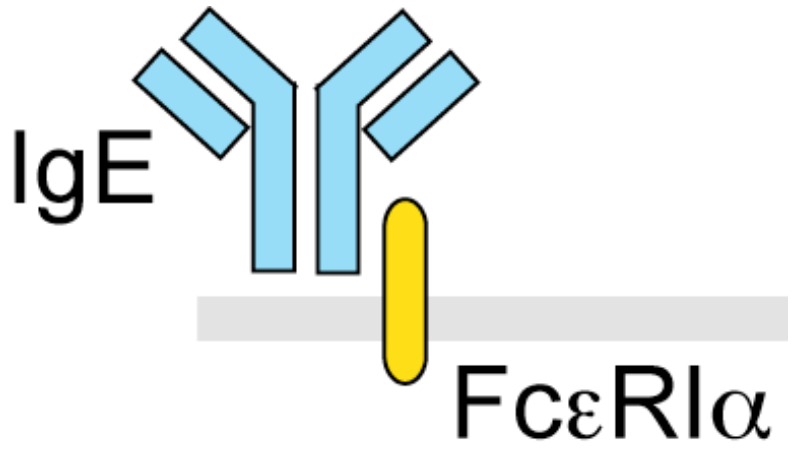
Blinov et al. *Biosystems* **83**, 136 (2006).

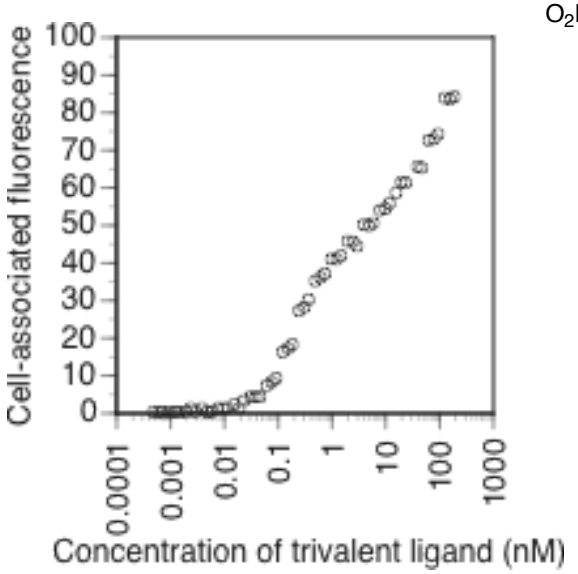
Outline

1. The motivation for rule-based modeling
2. Basic concepts of rule-based modeling
3. An example model specification
4. **Two methods for simulating a model**
5. Suggested exercises

Consider interaction of a trivalent ligand with a bivalent cell -surface receptor



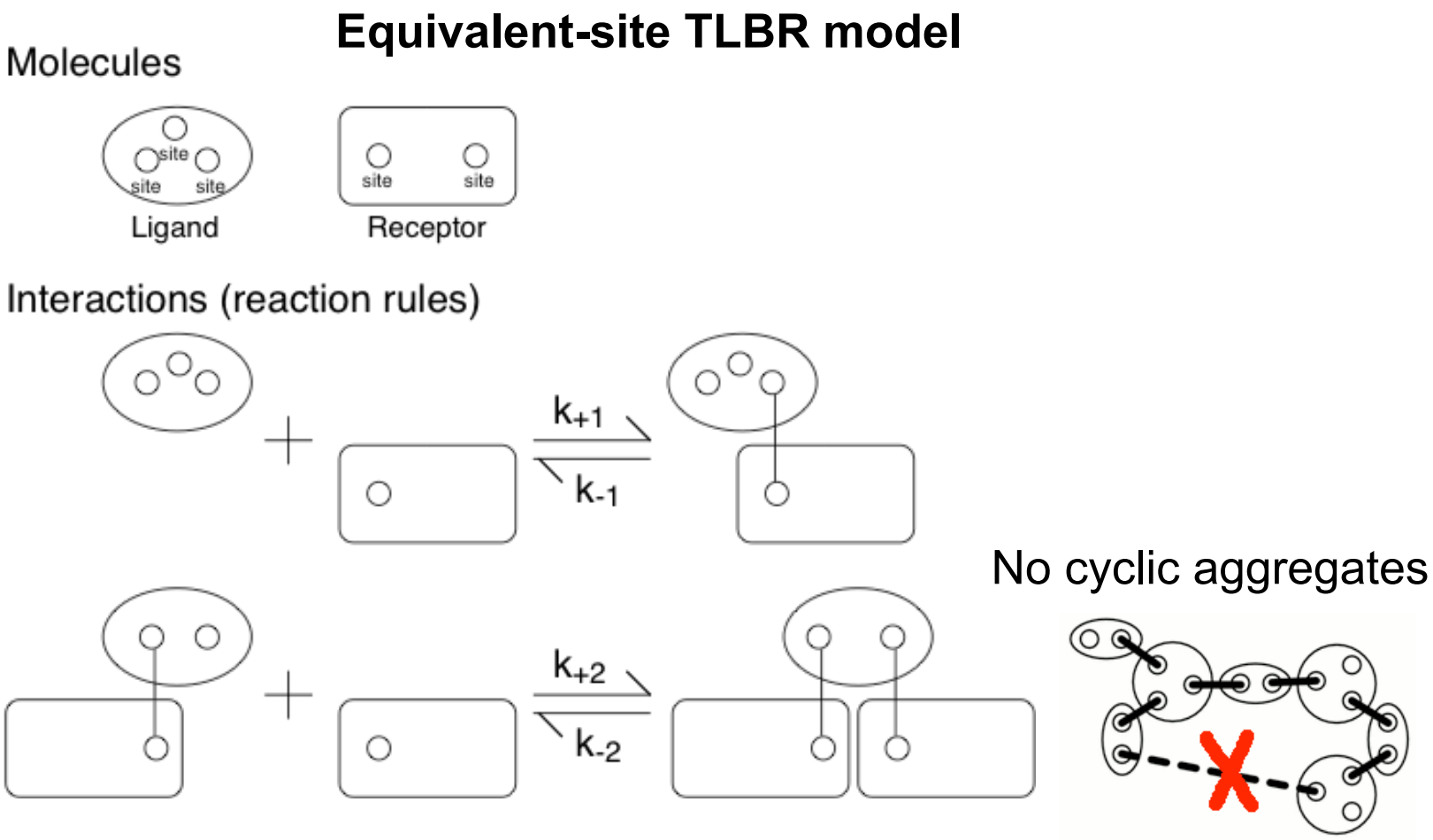




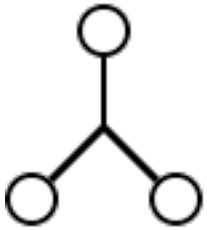
Los Alamos NATIONAL LABORATORY EST.1943

R.G. Posner (TGen) and P.B. Savage (BYU)

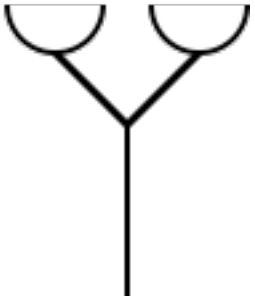
Rule-based model specification corresponding to equilibrium model of Goldstein and Perelson (1984)



Seed species

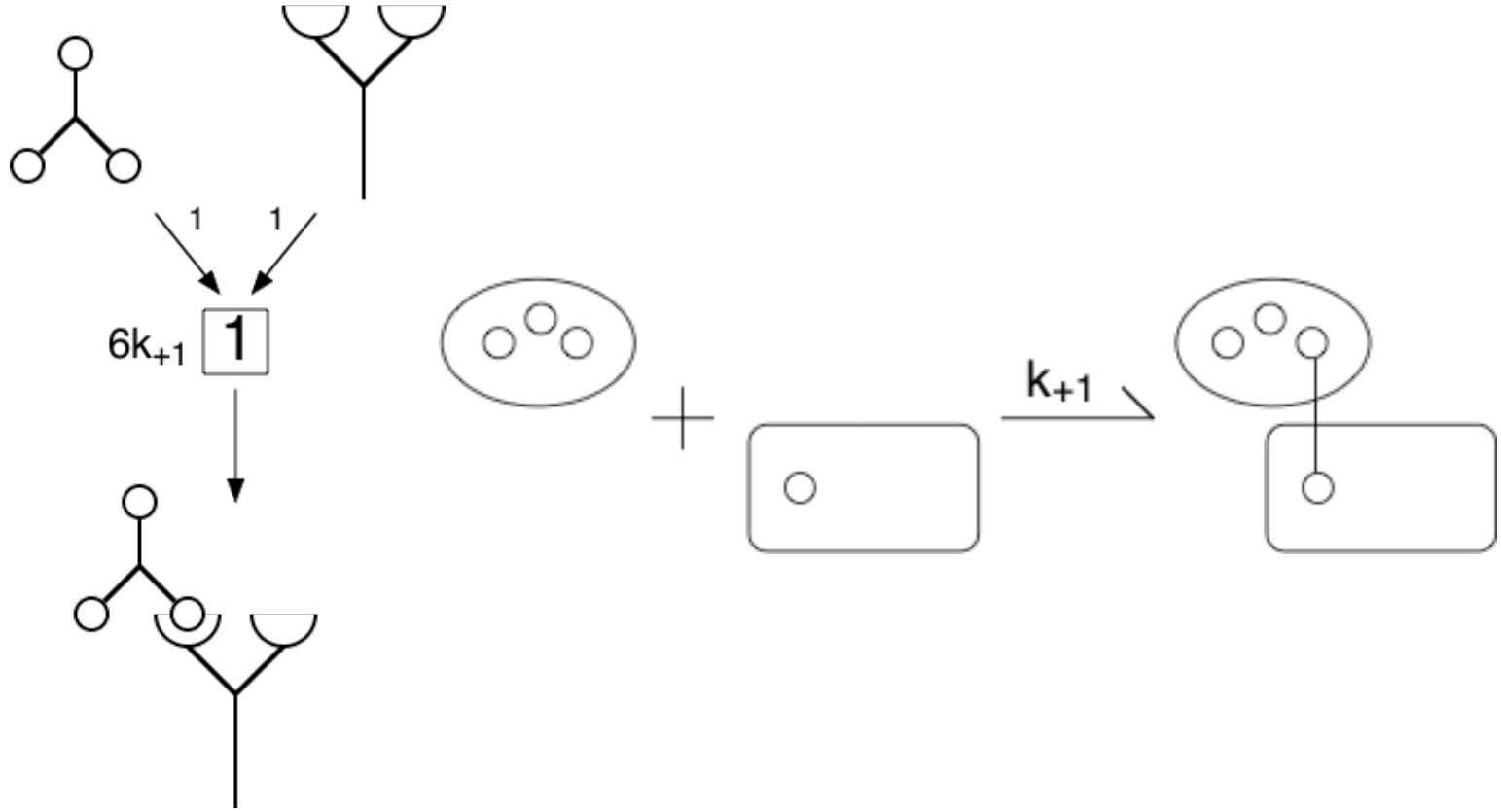


Ligand

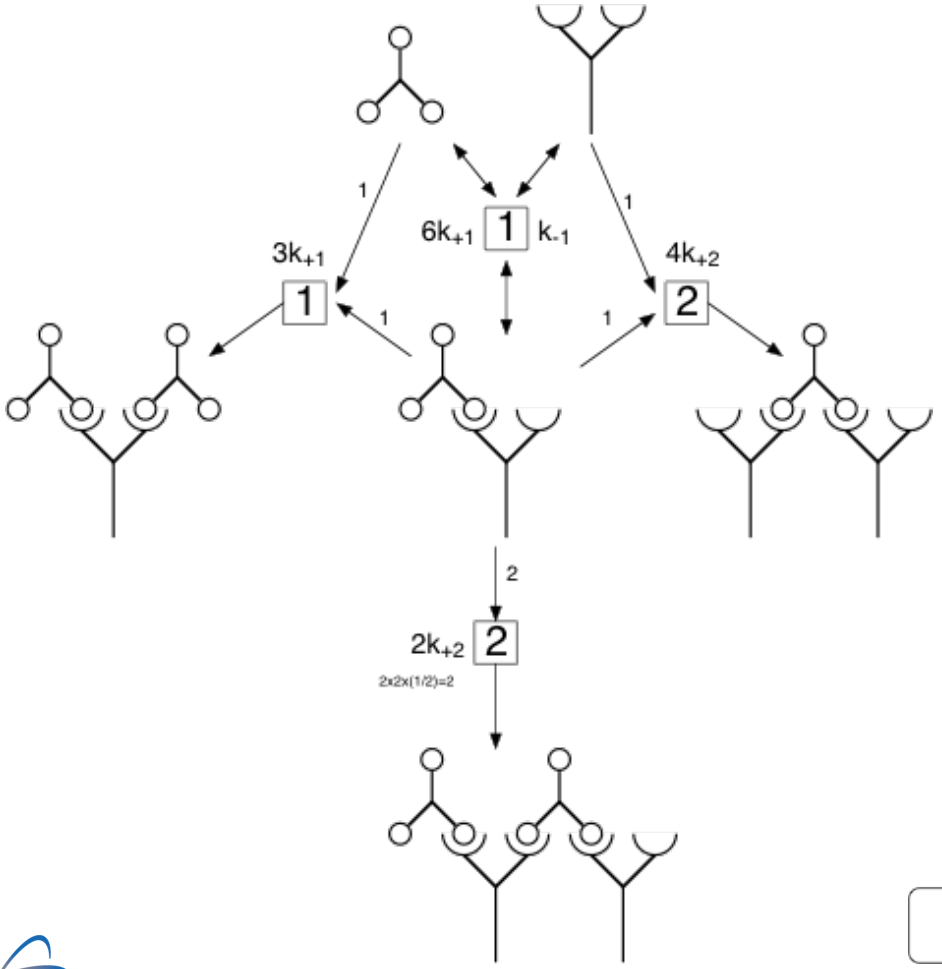


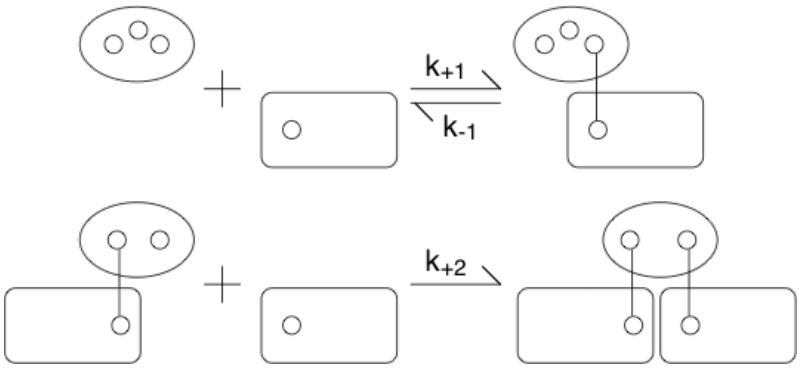
Receptor

After first round of rule application

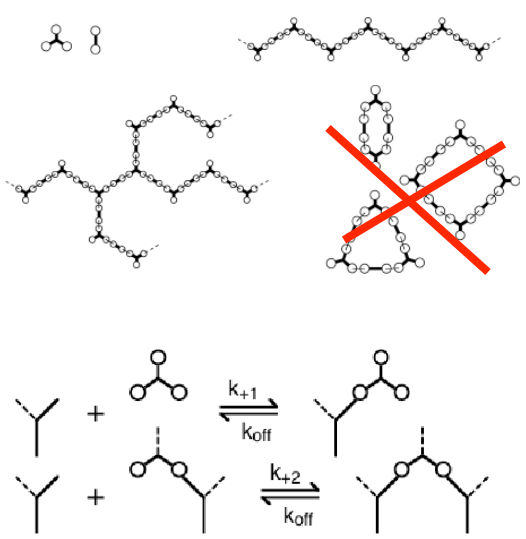


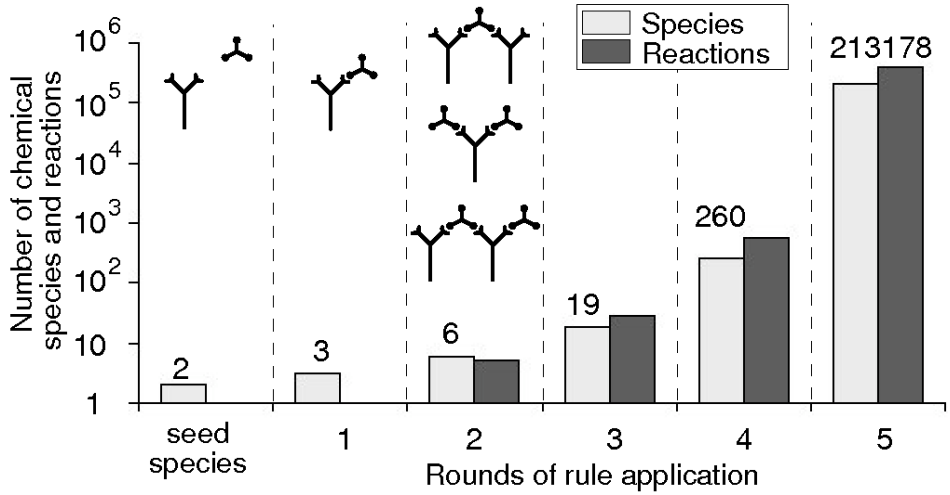
After the second round of rule application

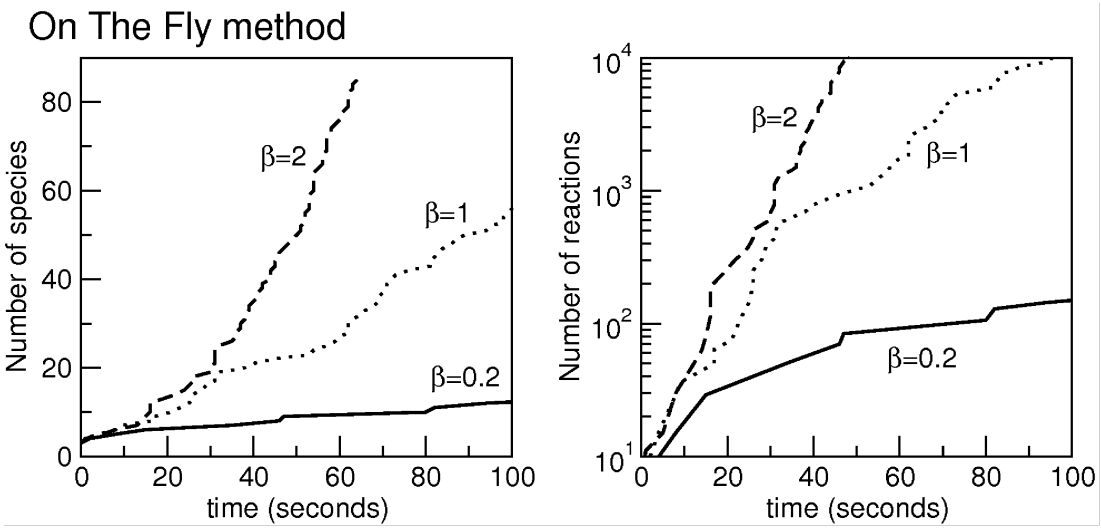




Rule-derived network can be too large to simulate using conventional population-based methods

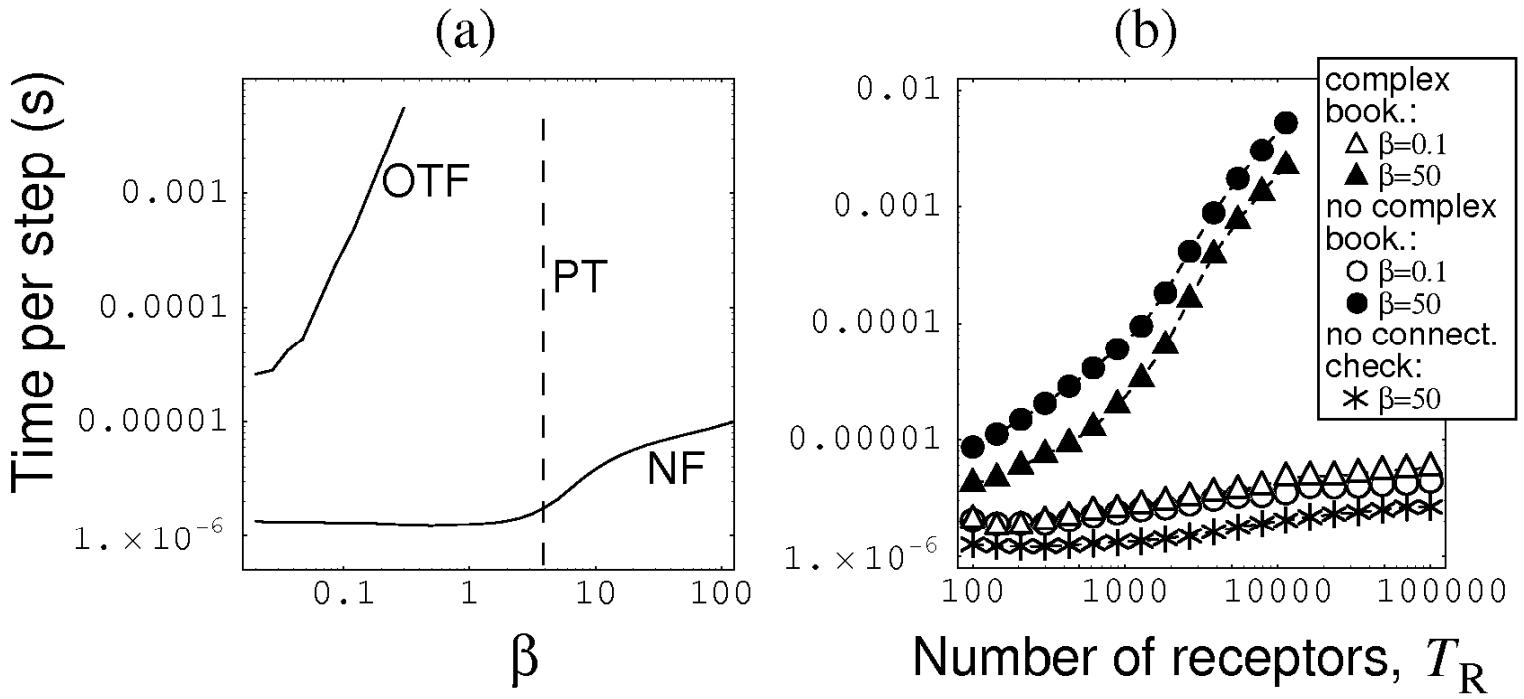






Los Alamos NATIONAL LABORATORY EST.1943

Performance of on-the-fly (OTF) simulation method

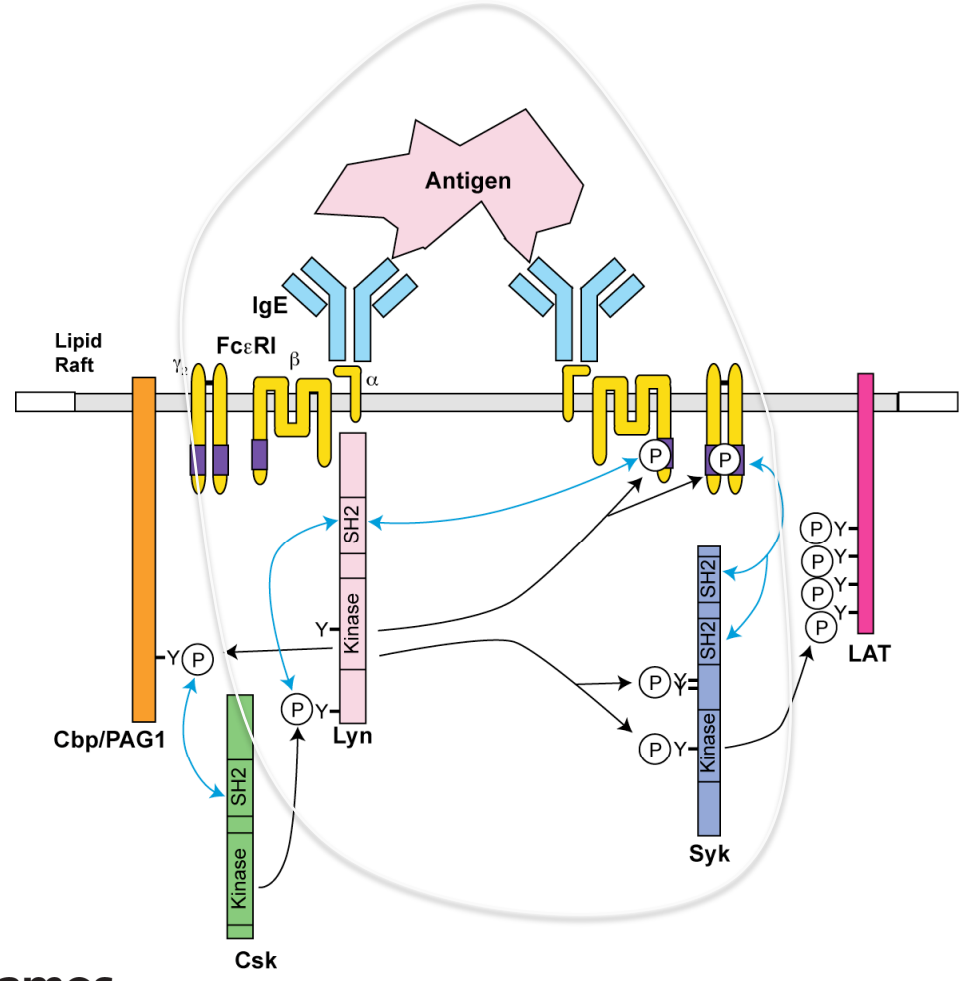


Yang et al. (2008) Phys. Rev. E

Outline

1. The motivation for rule-based modeling
2. Basic concepts of rule-based modeling
3. An example model specification
4. Two methods for simulating a model
5. **Suggested exercises**

Model for early events in FcεRI signaling leading to Syk activation

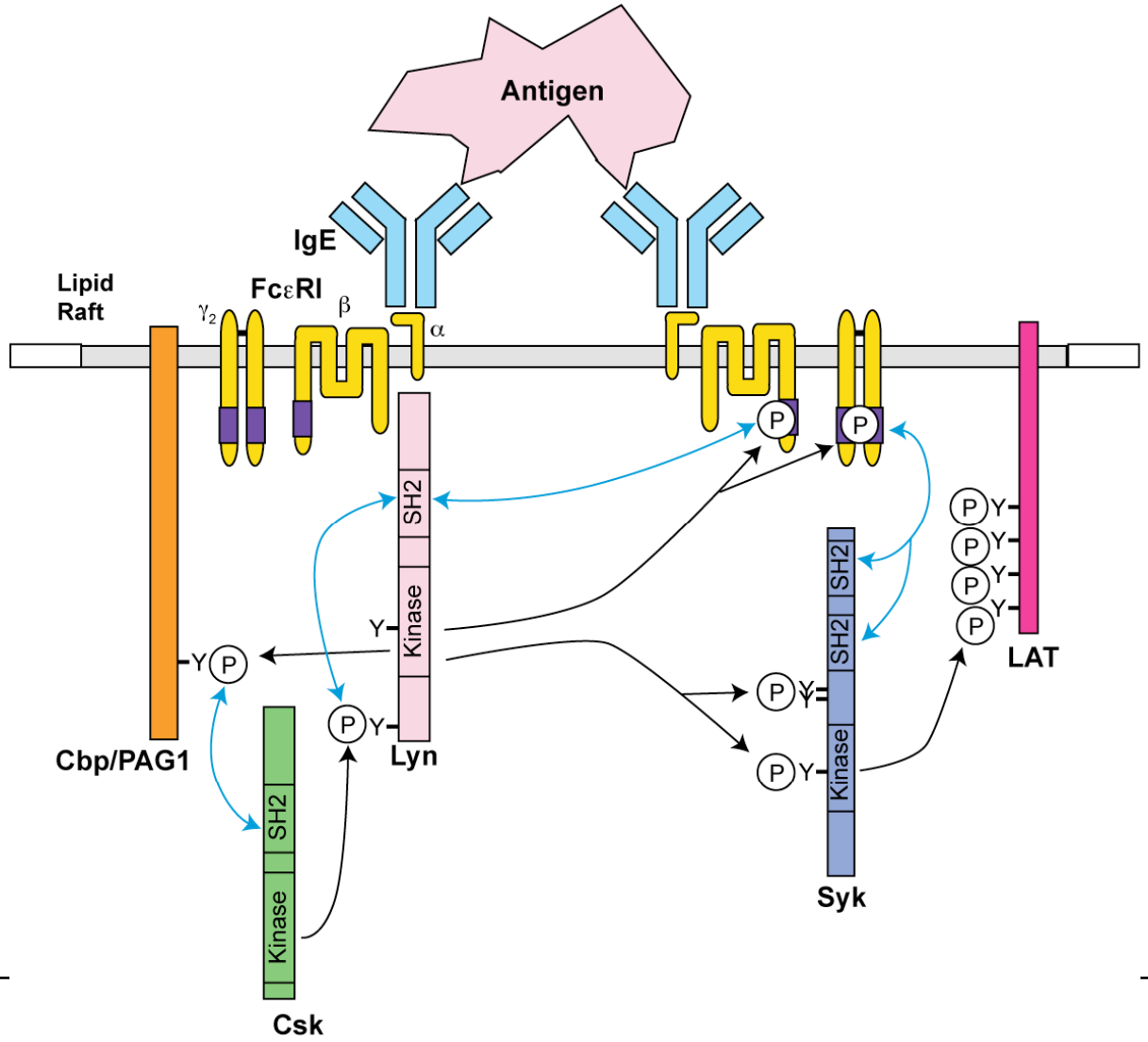


Mol. Immunol., 2002

J. Immunol., 2003

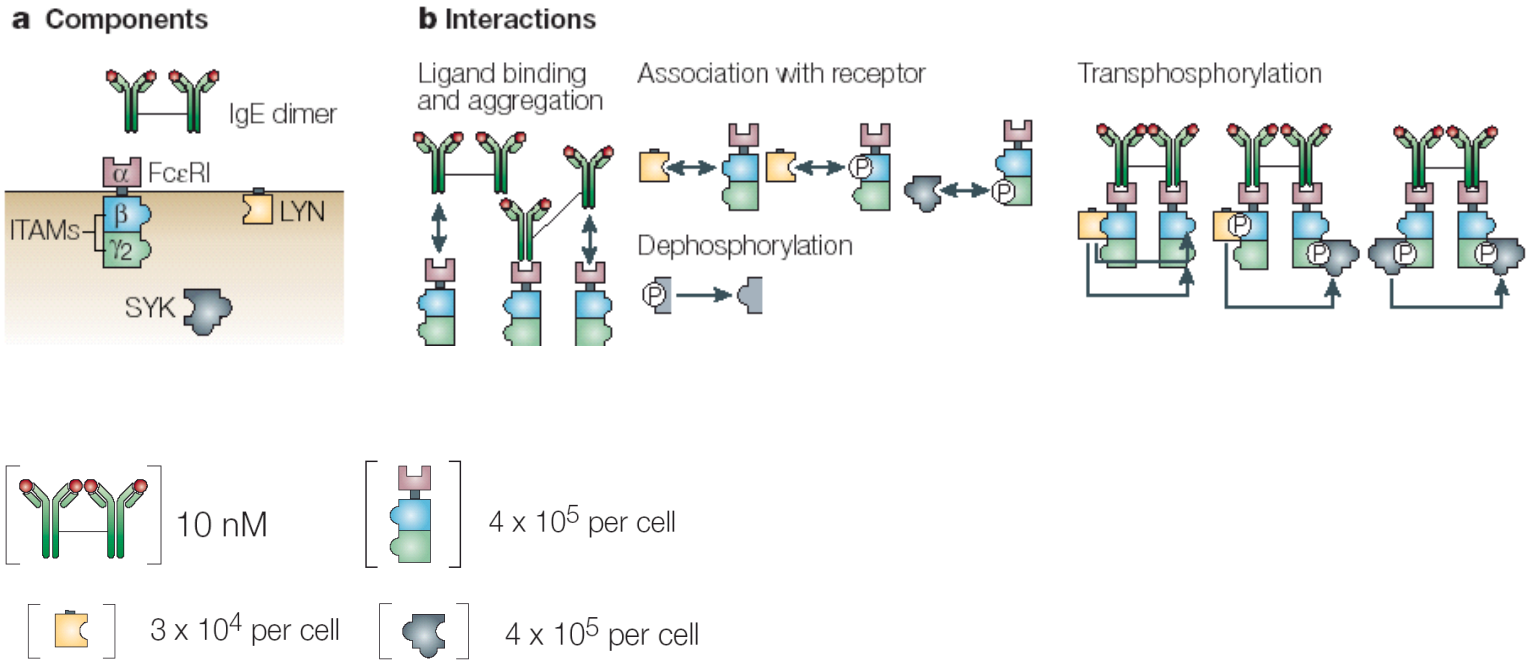
Los Alamos NATIONAL LABORATORY EST.1943

#1: What's the effect of Lyn regulation on Syk activation?



Los Alamos NATIONAL LABORATORY EST.1943

#2: What if the substrates tethered to a kinase saturate that kinase?

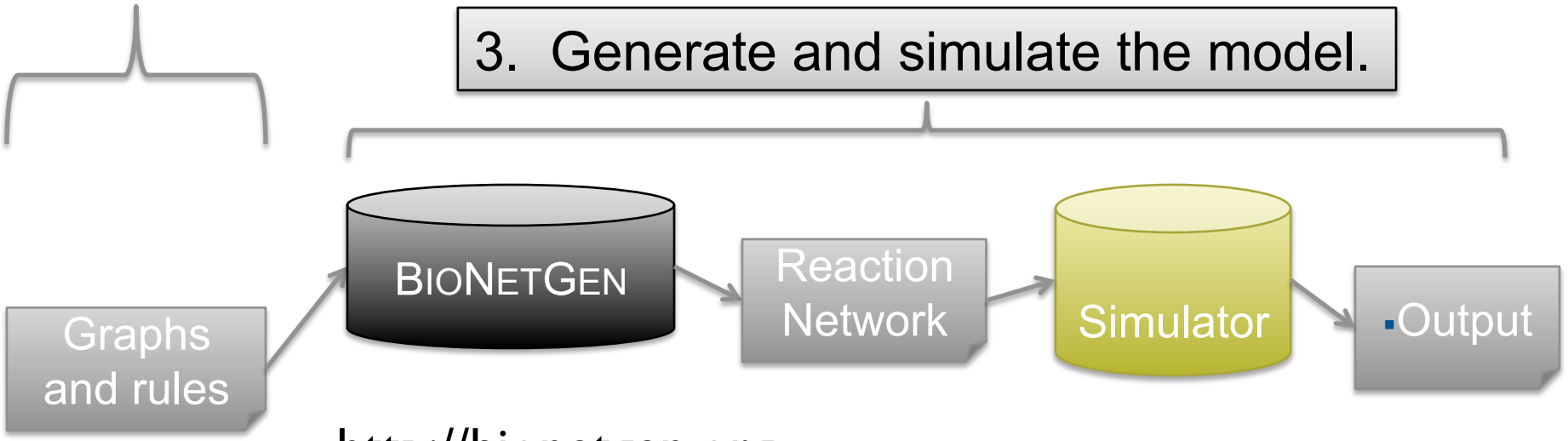


Rule-based modeling protocol

1. Define molecules as *graphs* and interactions as *rules*.

2. Determine **concentrations** and **rate constants**

3. Generate and simulate the model.



http://bionetgen.org

Los Alamos NATIONAL LABORATORY EST.1943

Faeder, Blinov, and Hlavacek, *Methods Mol. Biol.* (2009)

Summary

- Combinatorial complexity poses a barrier to specifying and simulating models of biological systems
- A rule-based modeling approach allows one to account for site-specific details of molecular interactions in a model
- Rule-base models can be simulated in a number of ways – only the two simplest approaches were mentioned in this talk