Rule-based Modeling

STMC
SpatioTemporal Modeling Center

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. Methods for simulating a model
5. Suggested exercise

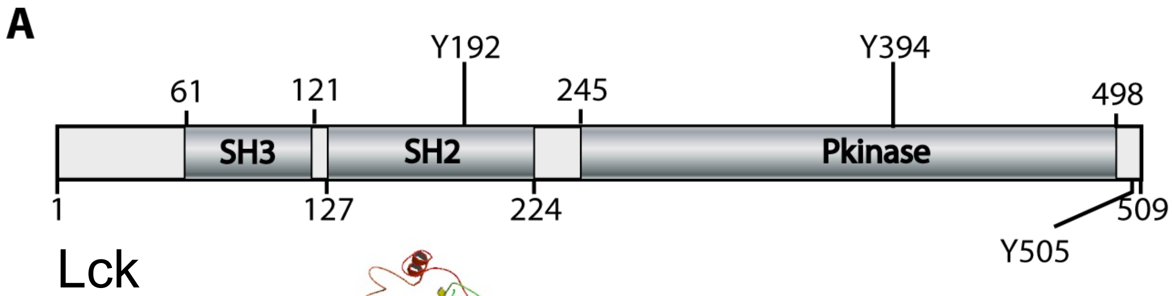
The need for predictive models of signal-transduction systems

- **These systems mediate cellular information processing and regulate cellular phenotypes**
- **They are complex**
- **Molecular changes that affect cell signaling cause/sustain disease (e.g., cancer)**
- **Numerous drugs that target signaling proteins are currently in clinical trials**
  - Spectacular successes (e.g., imatinib treatment of CML)
  - But results are disappointing for many patients
- **Many clinical trials are underway to test combinations of drugs (clinicaltrials.gov)**
  - There are too many combinations to consider all possibilities in trials

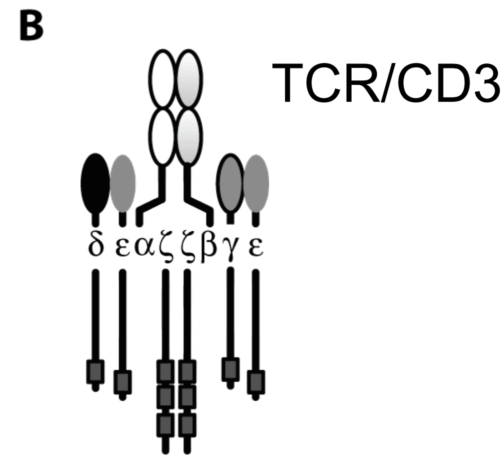
Value added by modeling

- **We can use models to organize information about a system with precision**
  - Introduces greater rigor and discipline
- **We can determine the logical consequences of a model specification**
  - Design principles can be elucidated (key for synthetic biology)
  - Certification (essential for personalized medicine)

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



Lck-SH2 (1bhh)

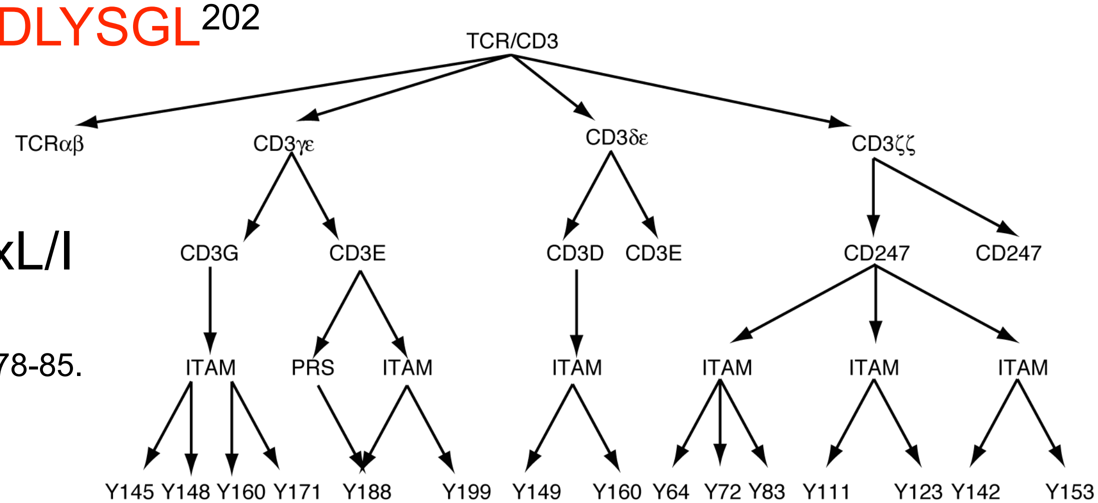


CD3E: ^{184}PNPDYEPIRKGQRDLYSGL^{202}

PRS: PxxDY

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

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



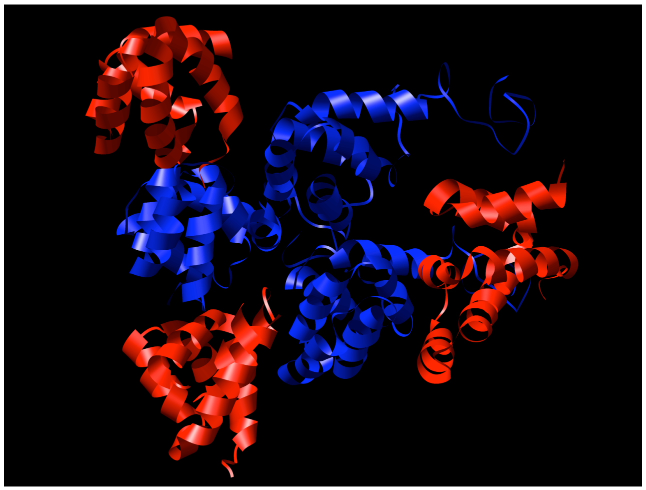
Los Alamos
NATIONAL LABORATORY
EST. 1943

There are many protein interaction domains

Los Alamos
NATIONAL LABORATORY
EST. 1943

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

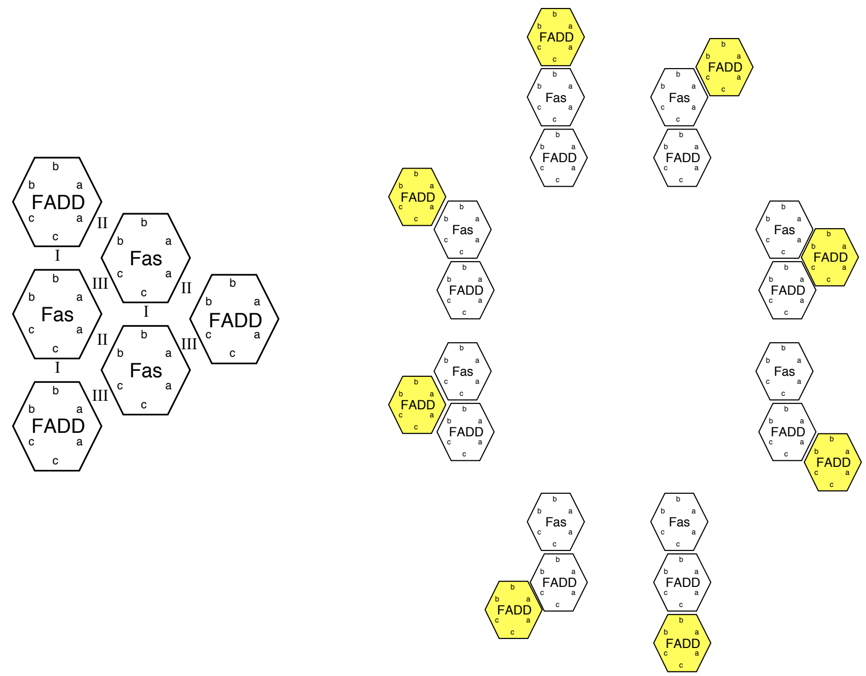
Some domains are multivalent and mediate oligomerization via domain-domain interactions



A hexamer of death domains

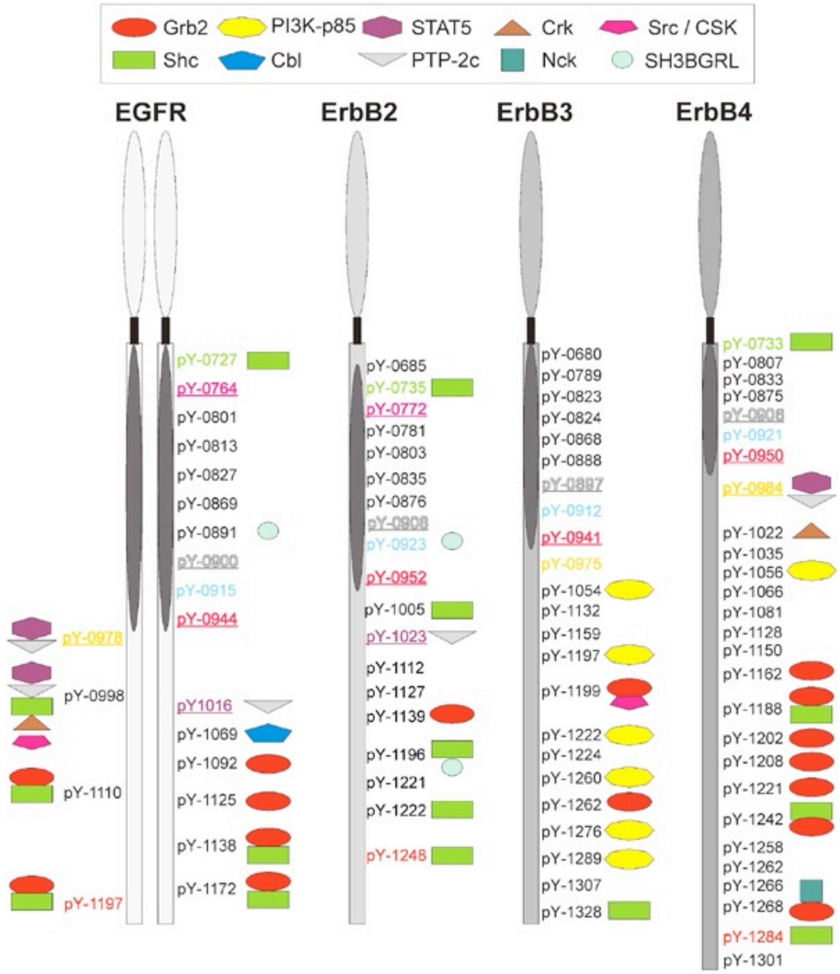
Weber and Vincenz (2001) *FEBS Lett*.

C.-T. Tung (Los Alamos)



There are many possible protein complexes!

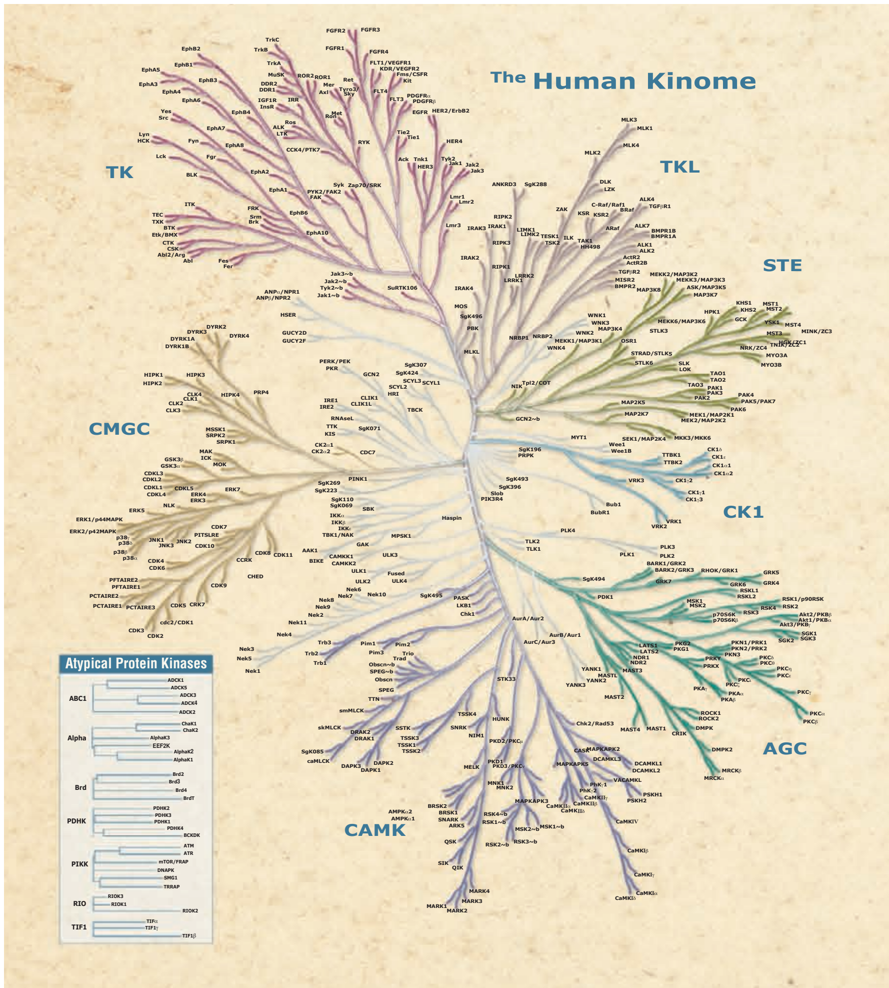
Domain-motif interactions are often controlled by post -translational modifications



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

There are many possible protein phosphoforms!

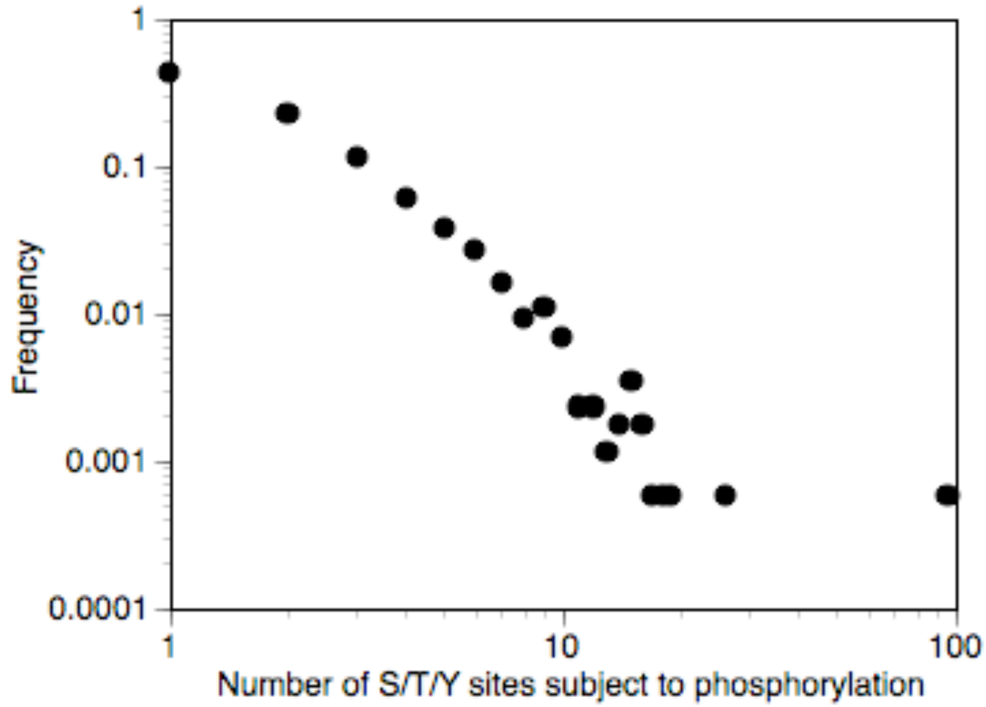
518 protein kinases (~2% of human genes)



There are phosphatases too!

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

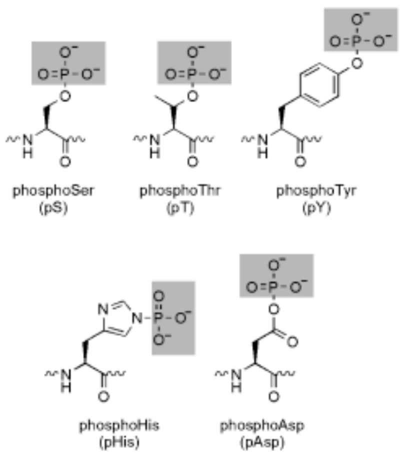
Signaling proteins typically contain multiple phosphorylation sites (S/T/Y)

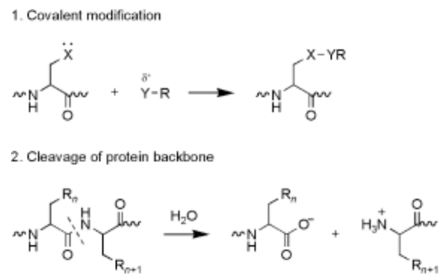


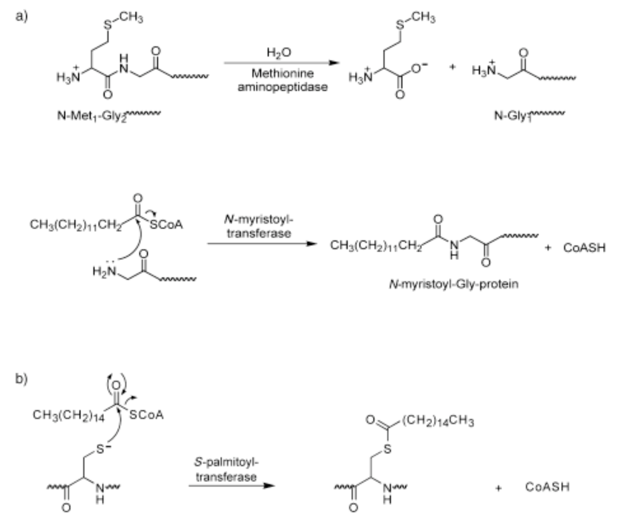
> 50% are phosphorylated at 2 or more sites

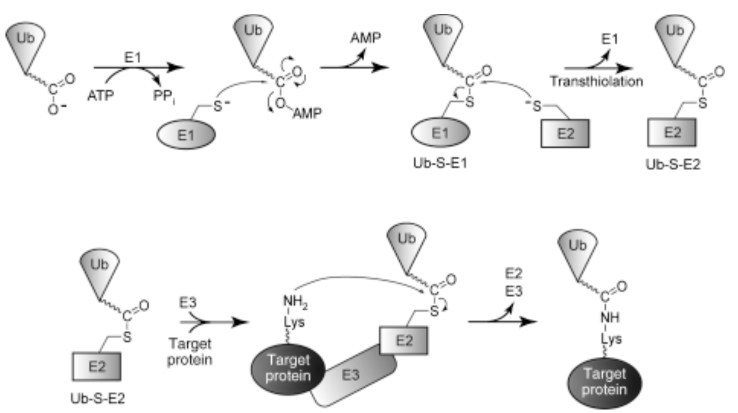
Phospho.ELM database v. 3.0 (http://phospho.elm.eu.org)

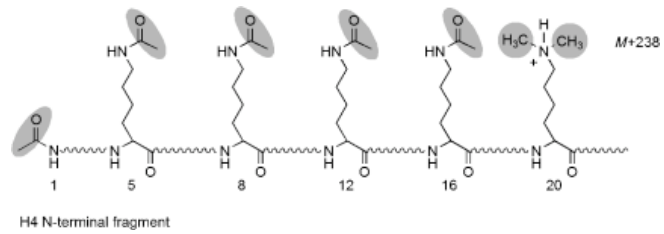
There are many different kinds of post-translational modifications of proteins







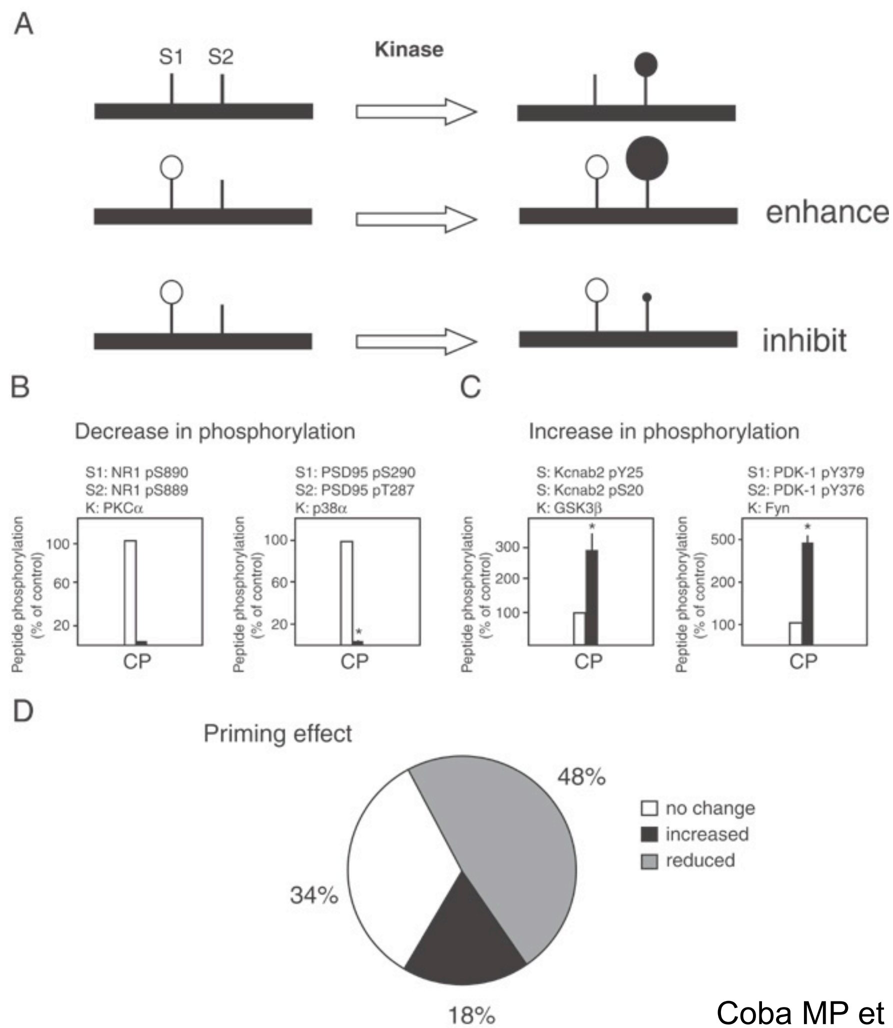




Los Alamos
NATIONAL LABORATORY
EST. 1943

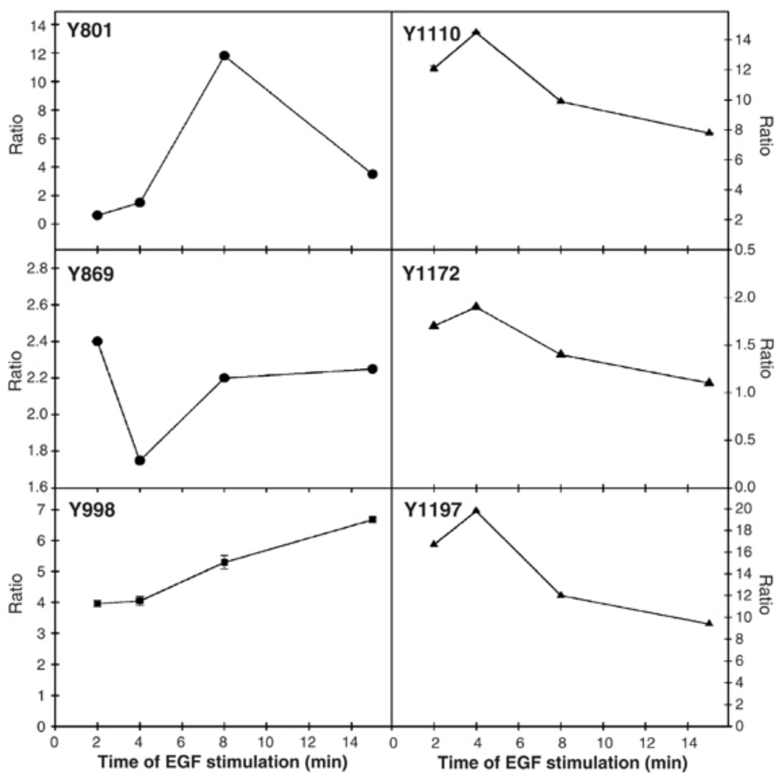
Walsh CT et al. (2005) Angew. Chem. Int. Ed. Engl. 44:7342-72.

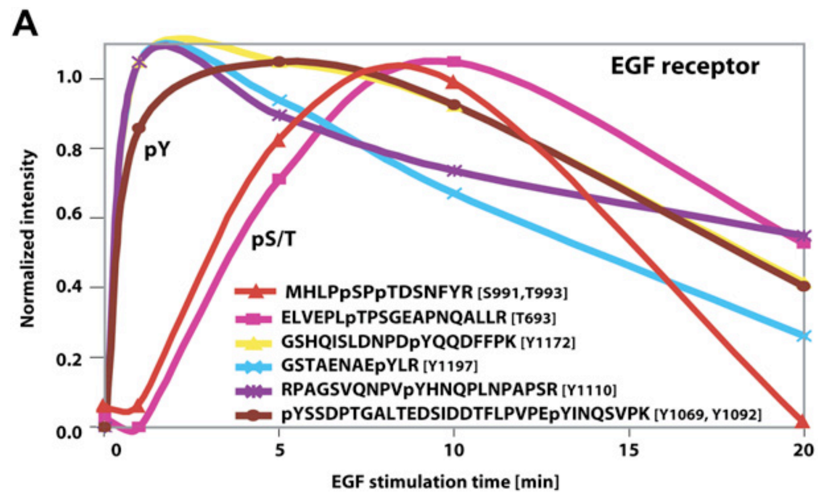
Priming – cooperative phosphorylation of neighboring kinase substrates is common



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.

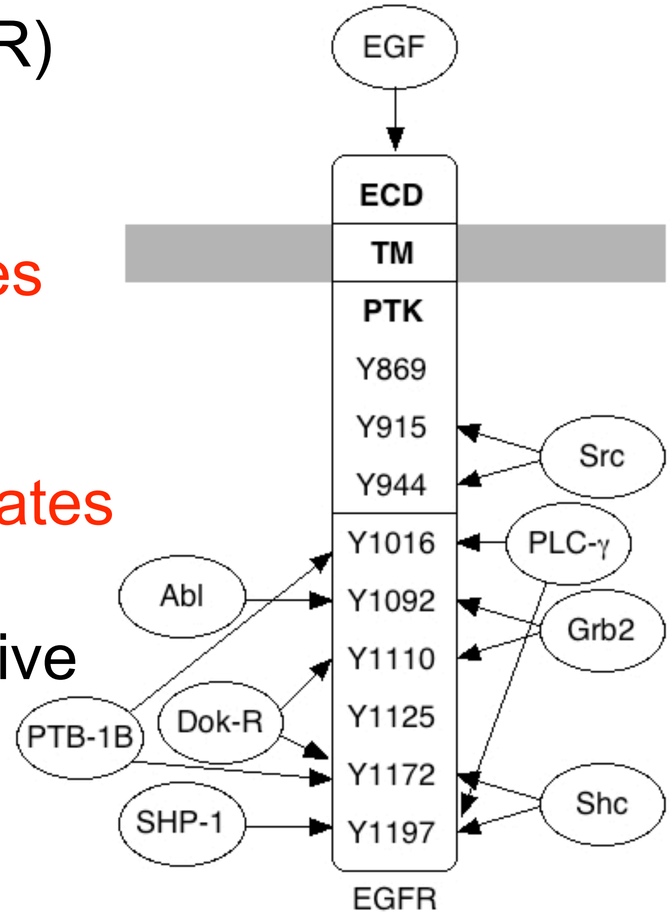
Combinatorial complexity – a serious problem for the conventional modeling approach

Epidermal growth factor receptor (EGFR)

9 sites => 2^9=512 phosphorylation states

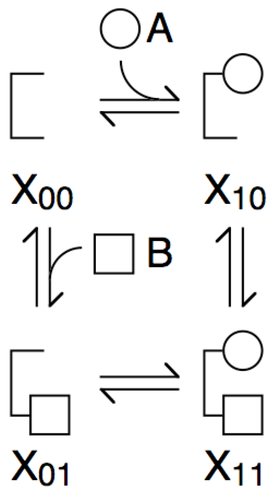
Each site has ≥ 1 binding partner
=> more than 3^9=19,683 total states

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



The textbook approach

Conventional representation of a biochemical reaction network

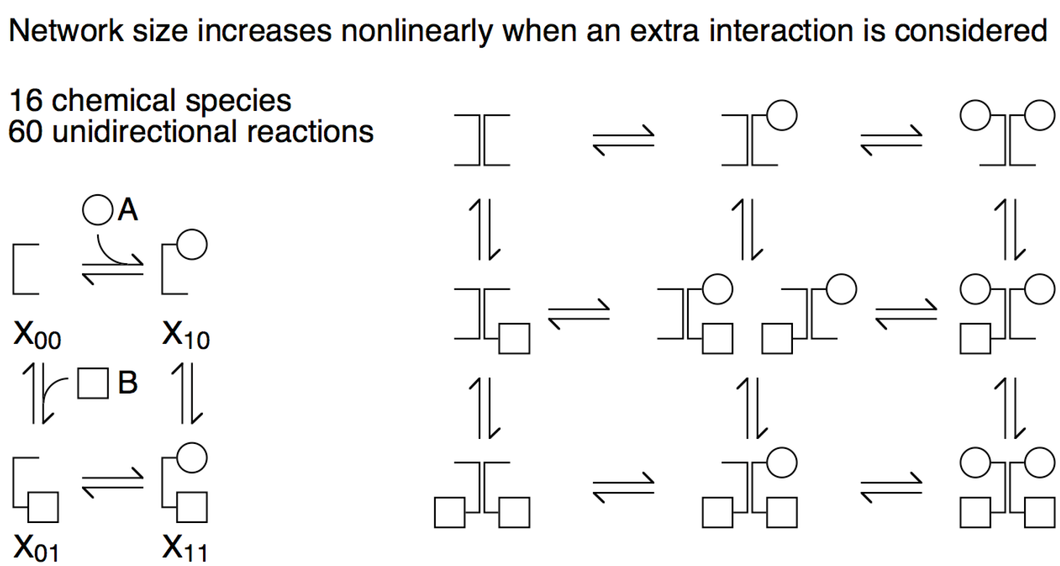


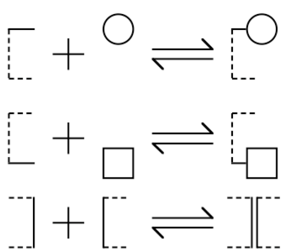
$$X_{00} + A \rightleftharpoons X_{10}$$
$$X_{00} + B \rightleftharpoons X_{01}$$
$$X_{10} + B \rightleftharpoons X_{11}$$
$$X_{01} + A \rightleftharpoons X_{11}$$

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

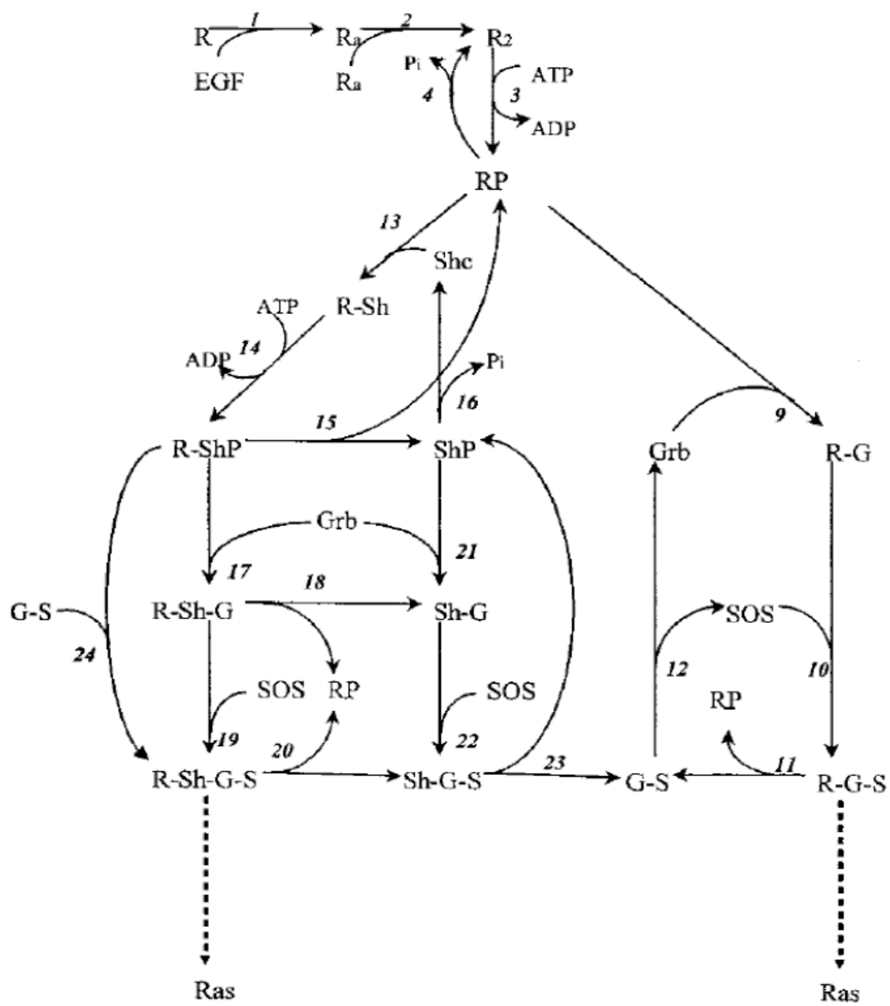




There are only three interactions. We can use a "rule" to model each of these interactions.

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

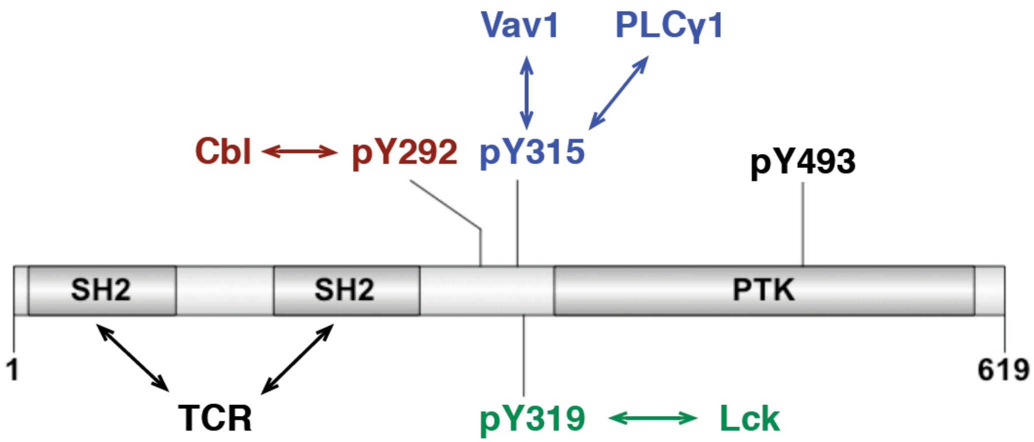


A reaction scheme incorporated in many published models of EGFR signaling

Rule-based modeling solves the problem of combinatorial complexity

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

- ZAP-70



Outline

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

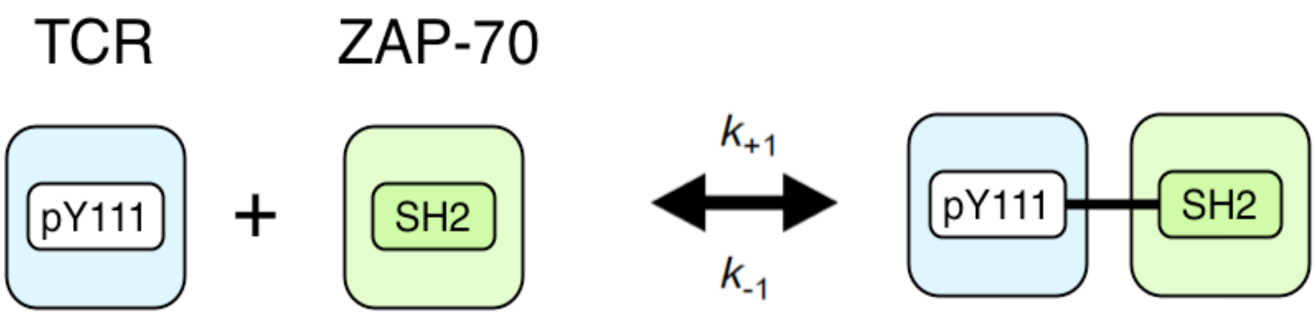
Rule-based modeling: basic concepts

Graphs represent molecules, their component parts, and “internal states”

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

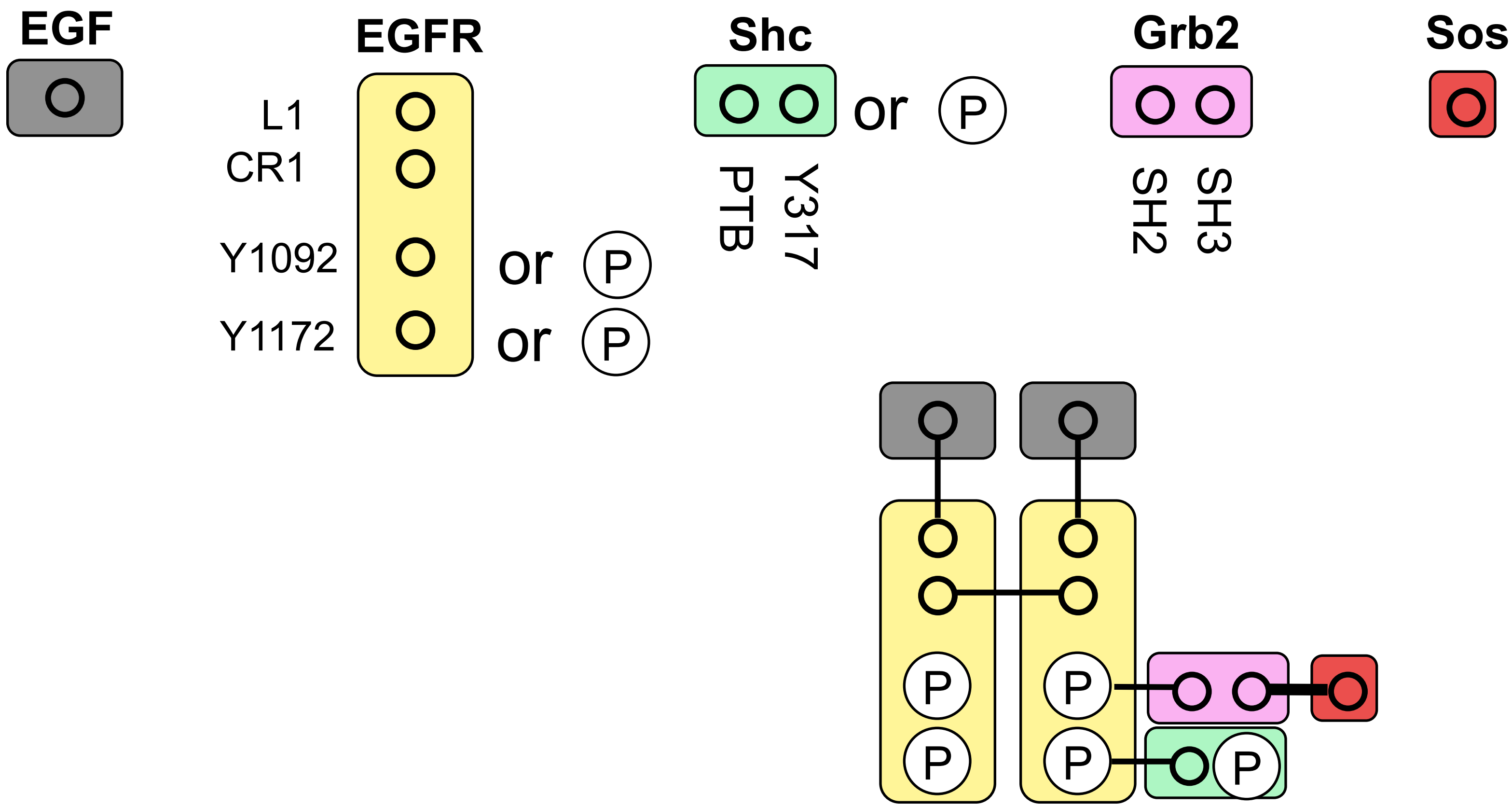
Graph-rewriting rules represent molecular interactions

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



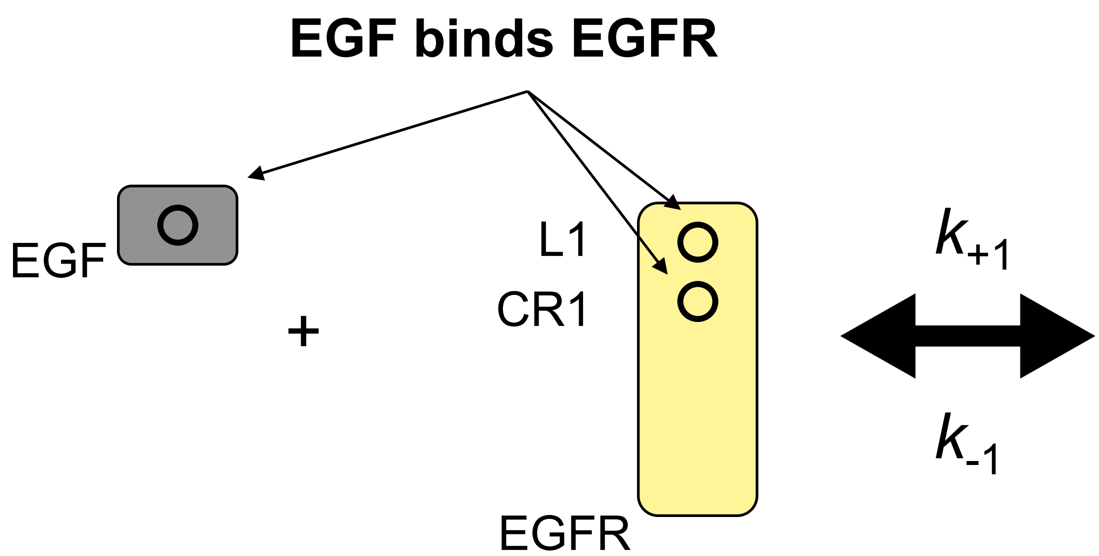

```
TCR(Y111~p)+ZAP70(SH2)<->TCR(Y111~p!1).ZAP70(SH2!1)
```

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

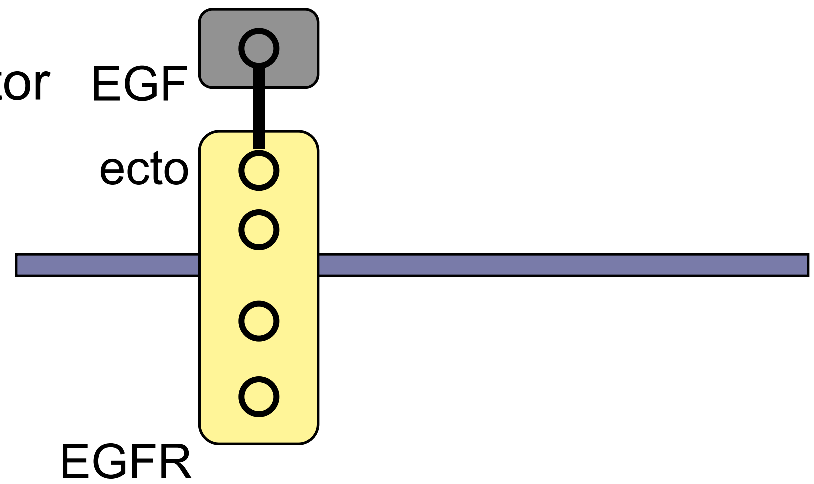
1. The motivation for rule-based modeling
2. Basic concepts of rule-based modeling
3. **An example model specification**
4. Methods for simulating a model
5. Suggested exercise

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

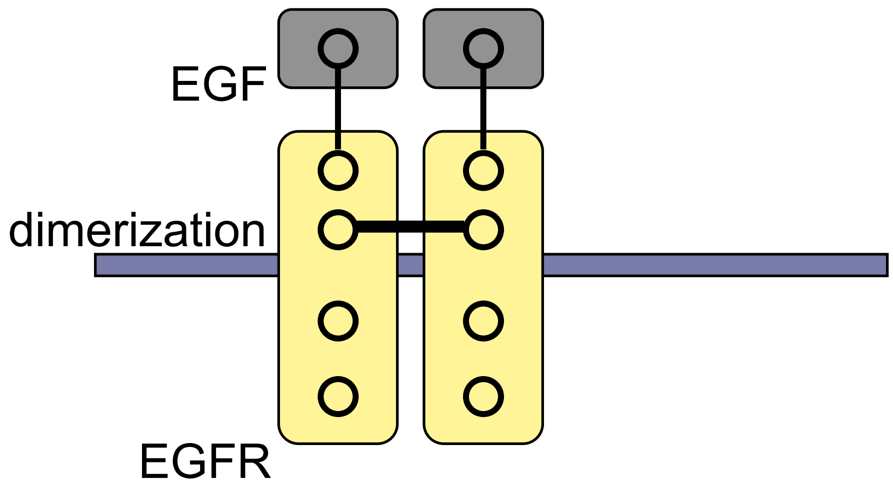


Blinov ML et al. (2006) *BioSystems*

Early events in EGFR signaling

1. EGF binds EGFR

2. EGFR dimerizes

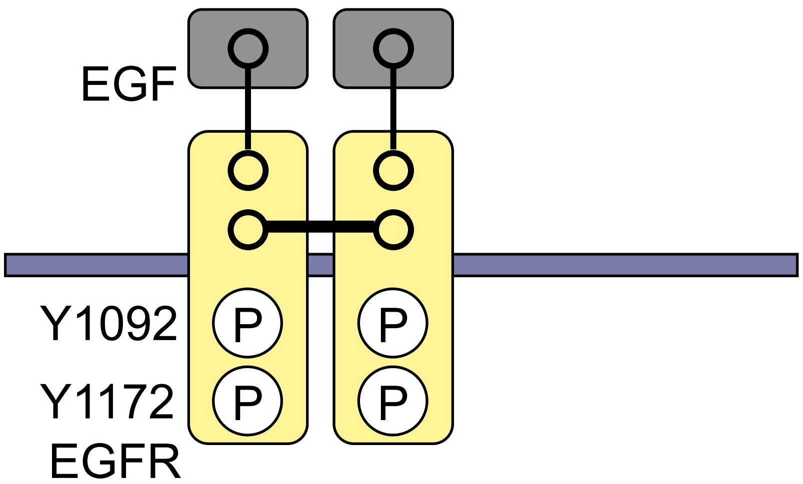


Early events in EGFR signaling

1. EGF binds EGFR

2. EGFR dimerizes

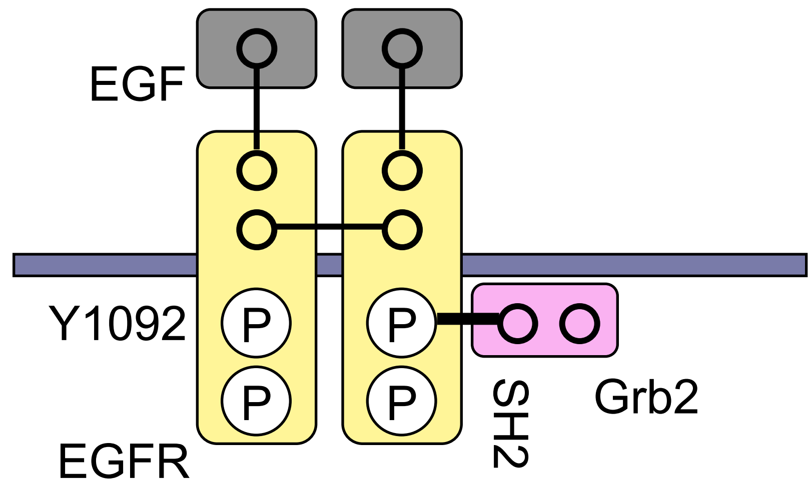
3. EGFR transphosphorylates a copy of itself



Early events in EGFR signaling

Grb2 pathway

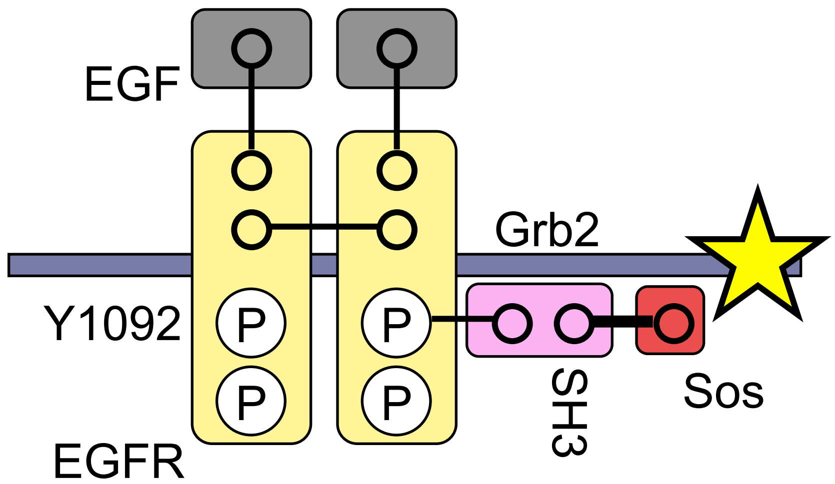
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Grb2 binds phospho-EGFR**



Early events in EGFR signaling

Grb2 pathway

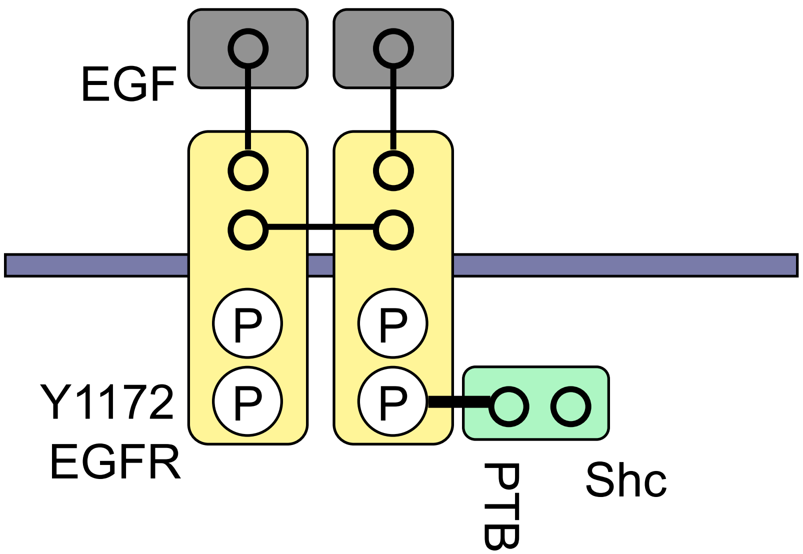
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Grb2 binds phospho-EGFR
5. **Sos binds Grb2 (Activation Path 1)**



Early events in EGFR signaling

Shc pathway

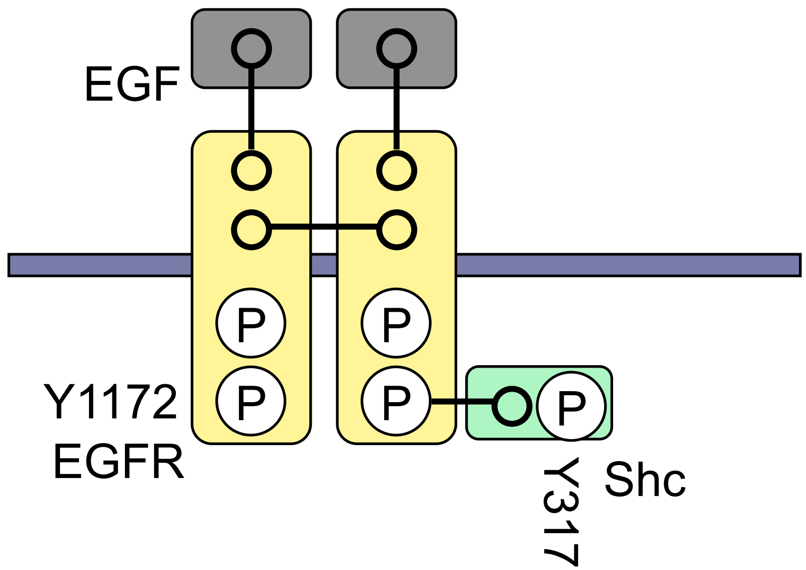
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Shc binds phospho-EGFR**



Early events in EGFR signaling

Shc pathway

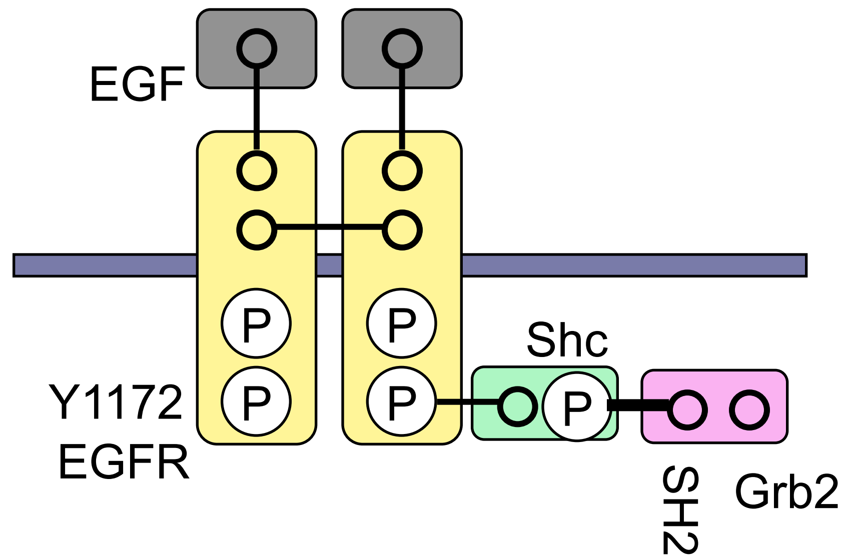
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. **EGFR transphosphorylates Shc**



Early events in EGFR signaling

Shc pathway

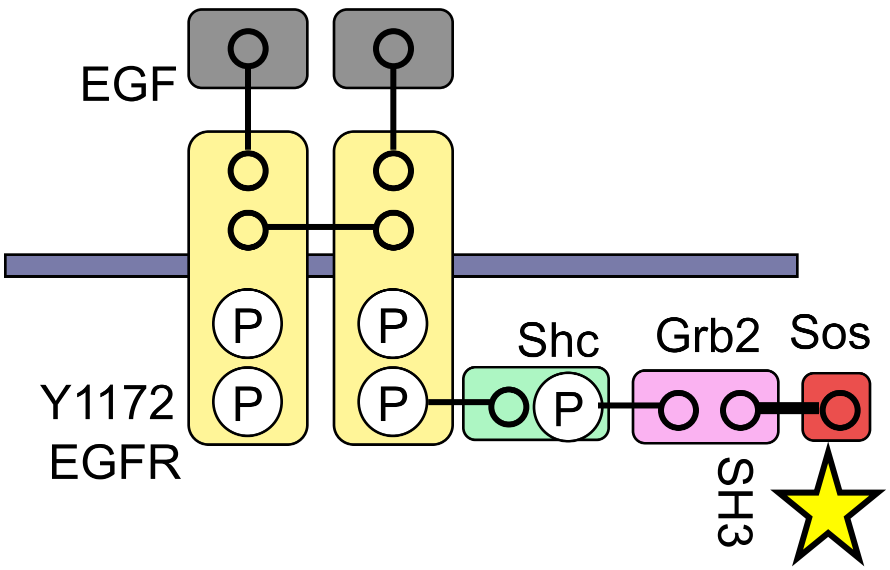
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. **Grb2 binds phospho-Shc**



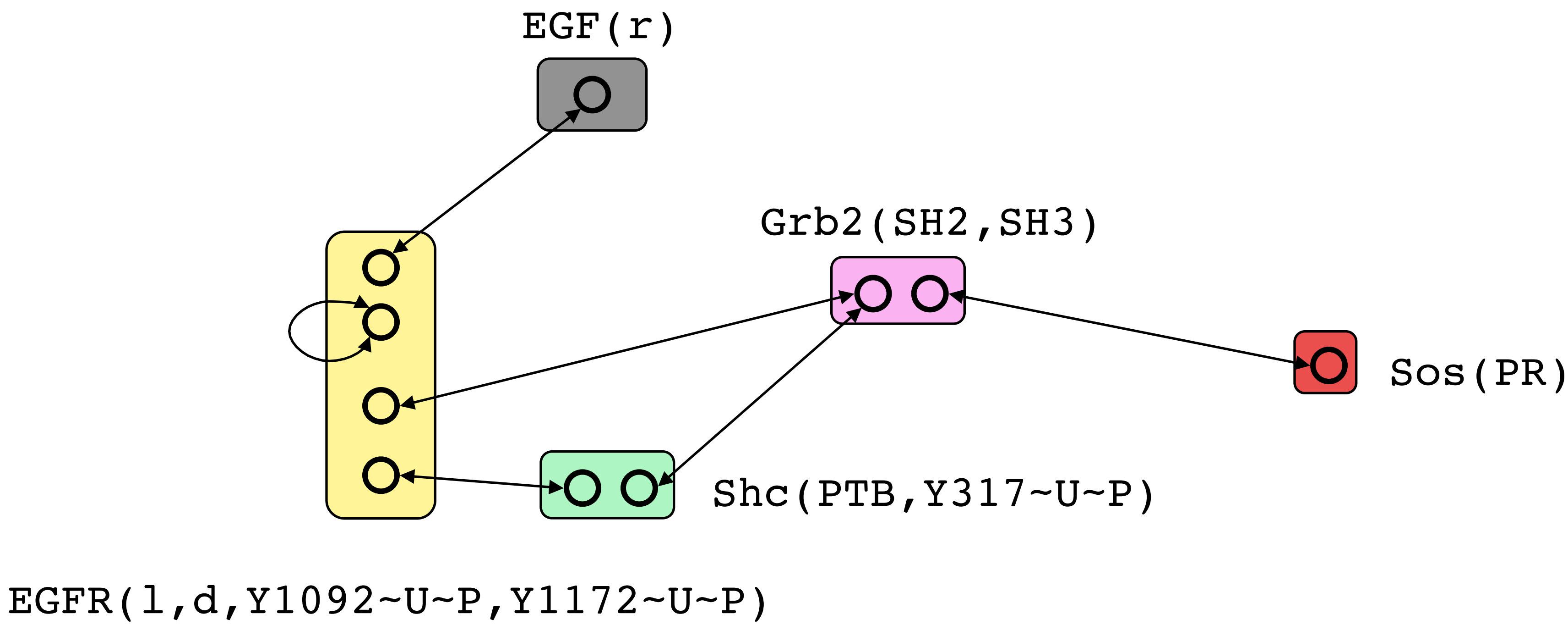
Early events in EGFR signaling

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)**



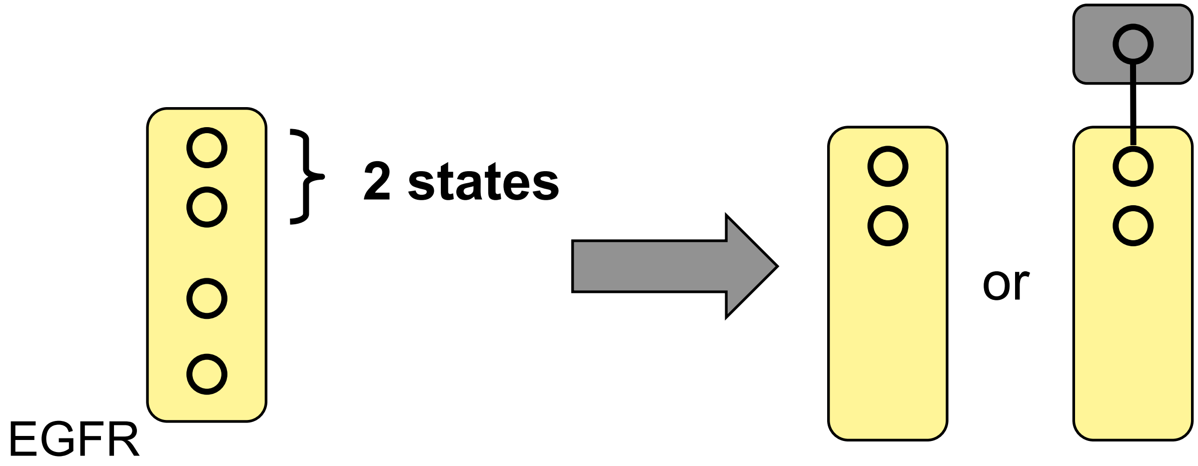
Summary of molecules and their interactions in a simple model of early events in EGFR signaling



Blinov et al. (2006)

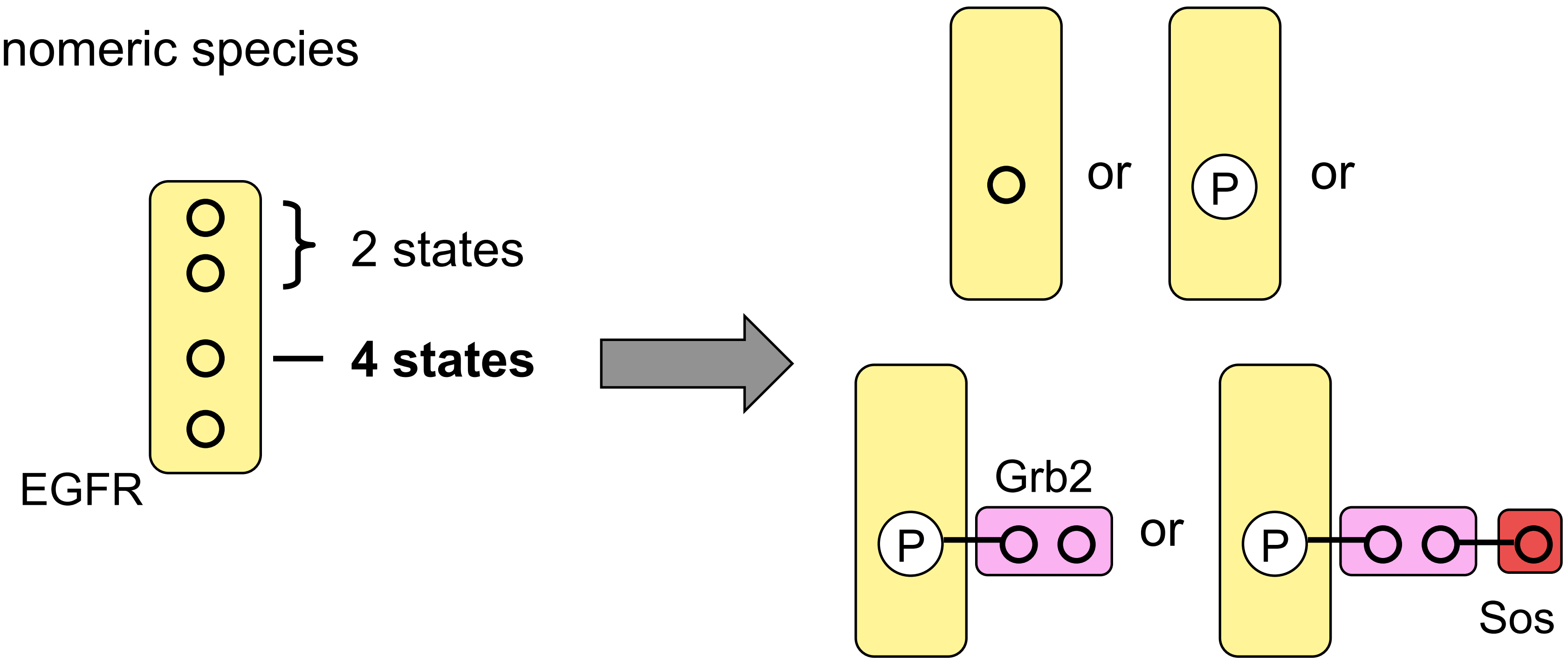
Combinatorial complexity of early events

Monomeric species



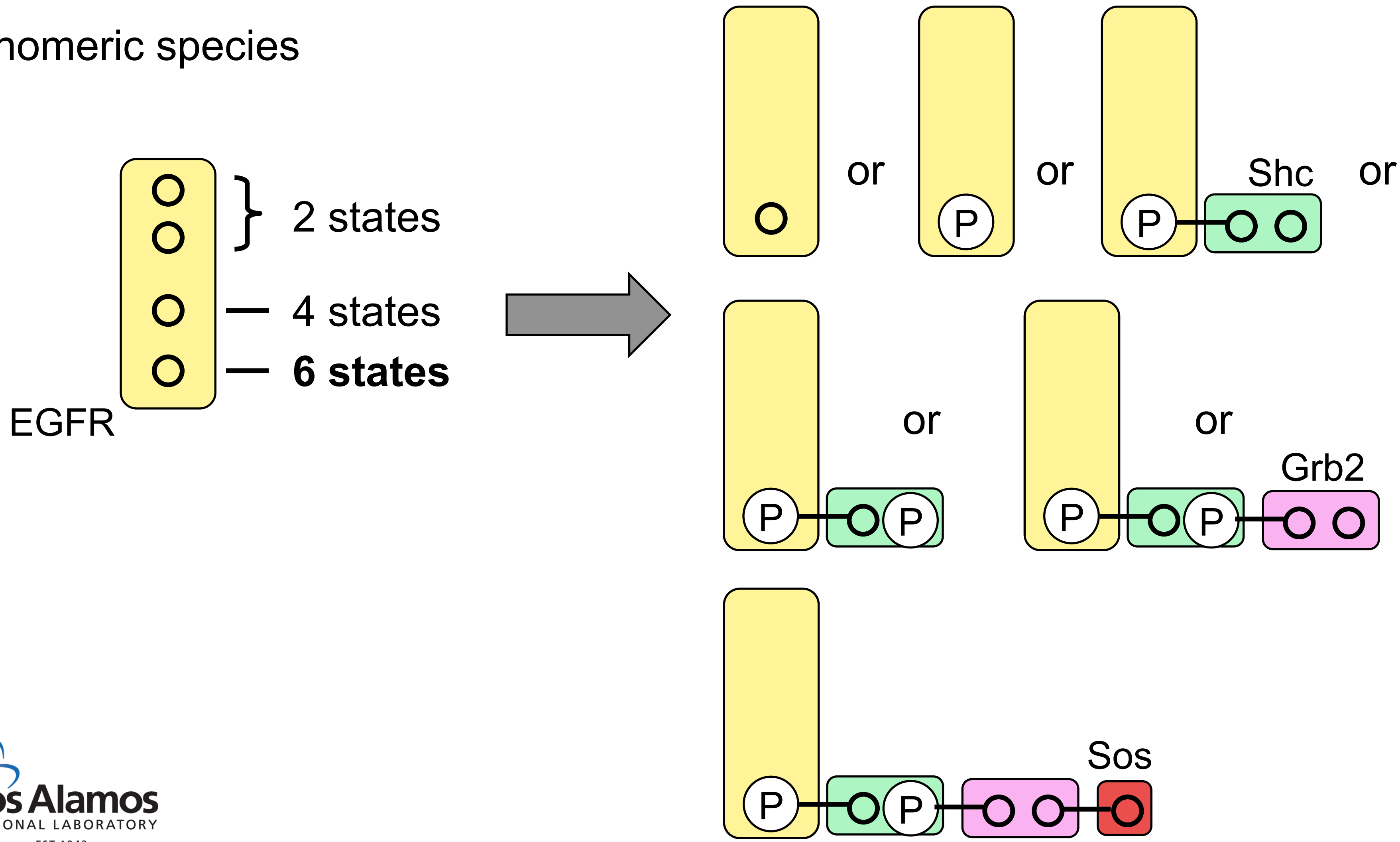
Combinatorial complexity of early events

Monomeric species



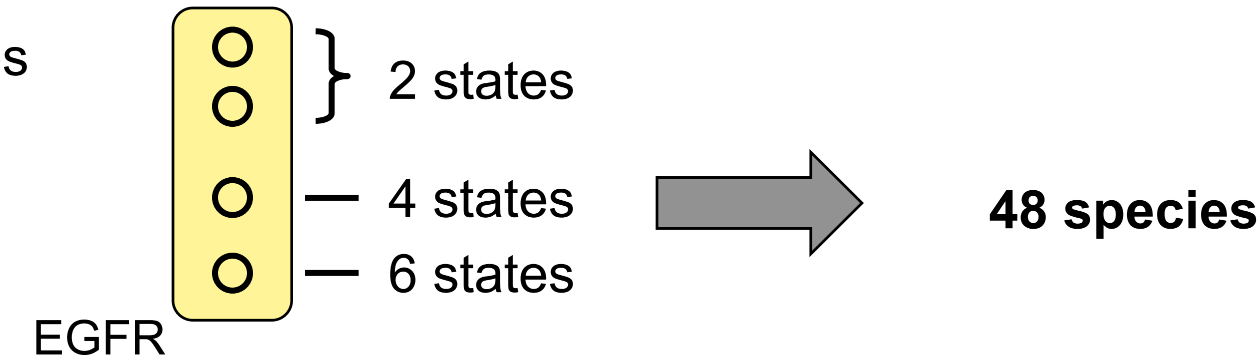
Combinatorial complexity of early events

Monomeric species



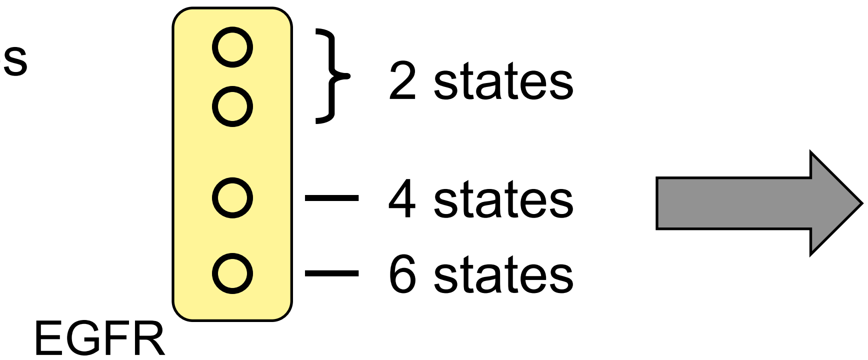
Combinatorial complexity of early events

Monomeric species



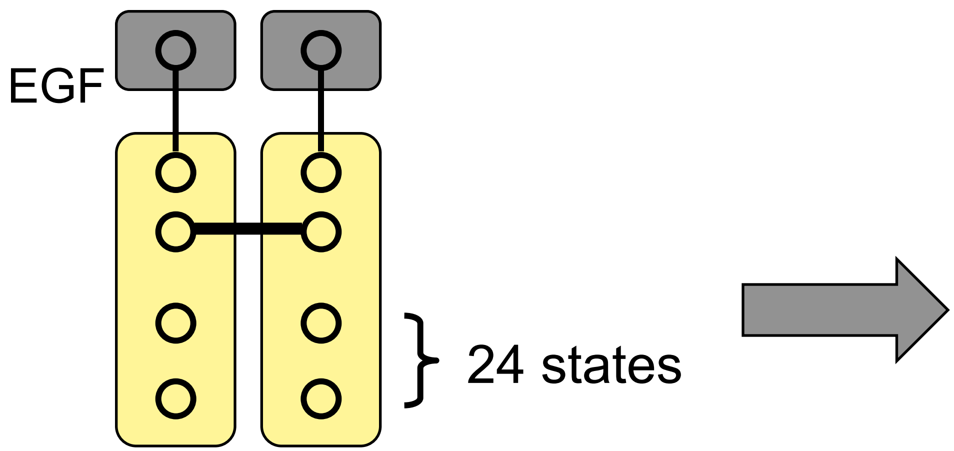
Combinatorial complexity of early events

Monomeric species



48 species

Dimeric species



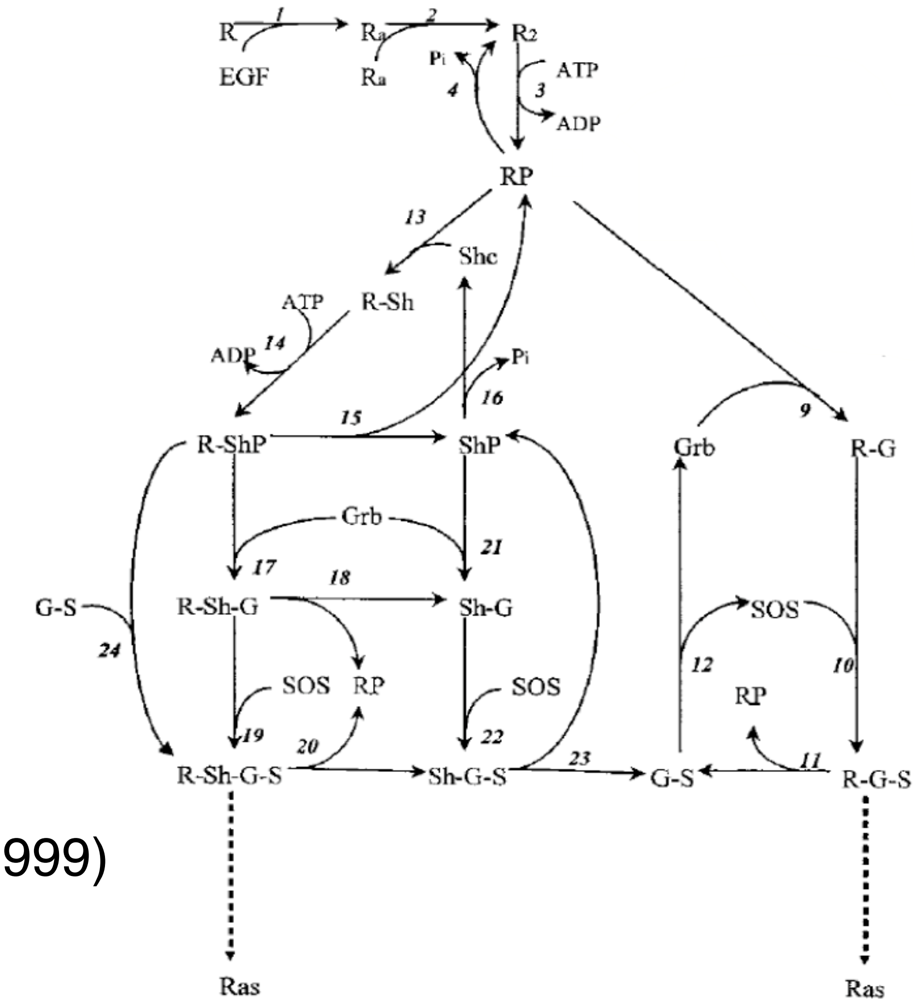
$N(N+1)/2$ = **300 species**

A conventional model for EGFR signaling

The Kholodenko model*

5 proteins

18 species
34 reactions



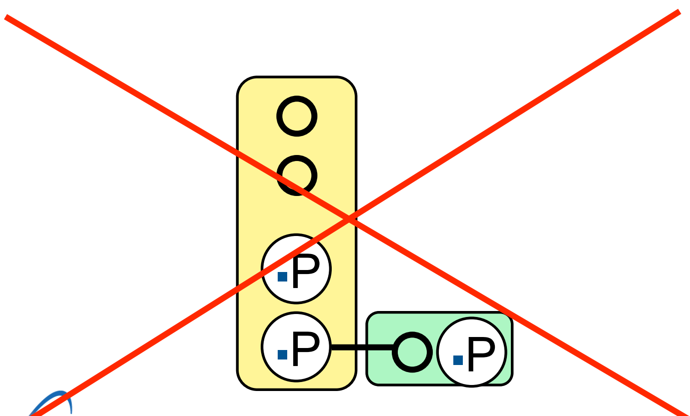
*_J. Biol. Chem._ **274**, 30169 (1999)

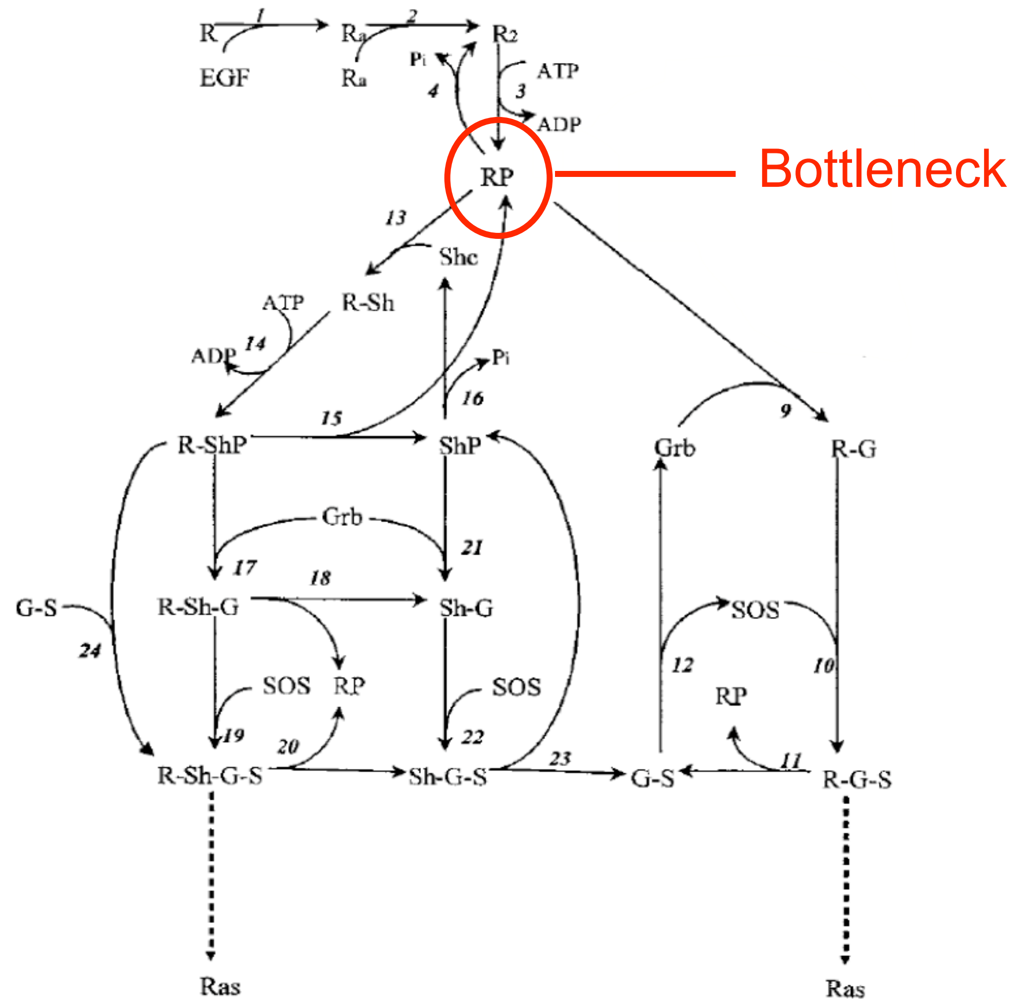
Los Alamos
NATIONAL LABORATORY
EST.1943

Assumptions made to limit combinatorial complexity

1. Phosphorylation inhibits dimer breakup

No modified monomers

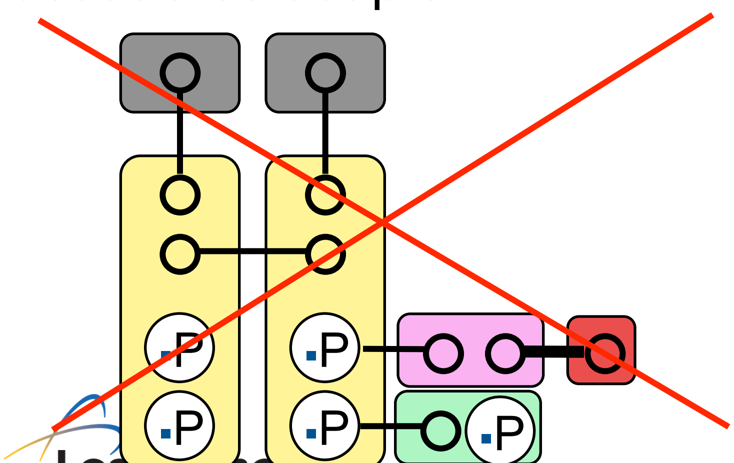


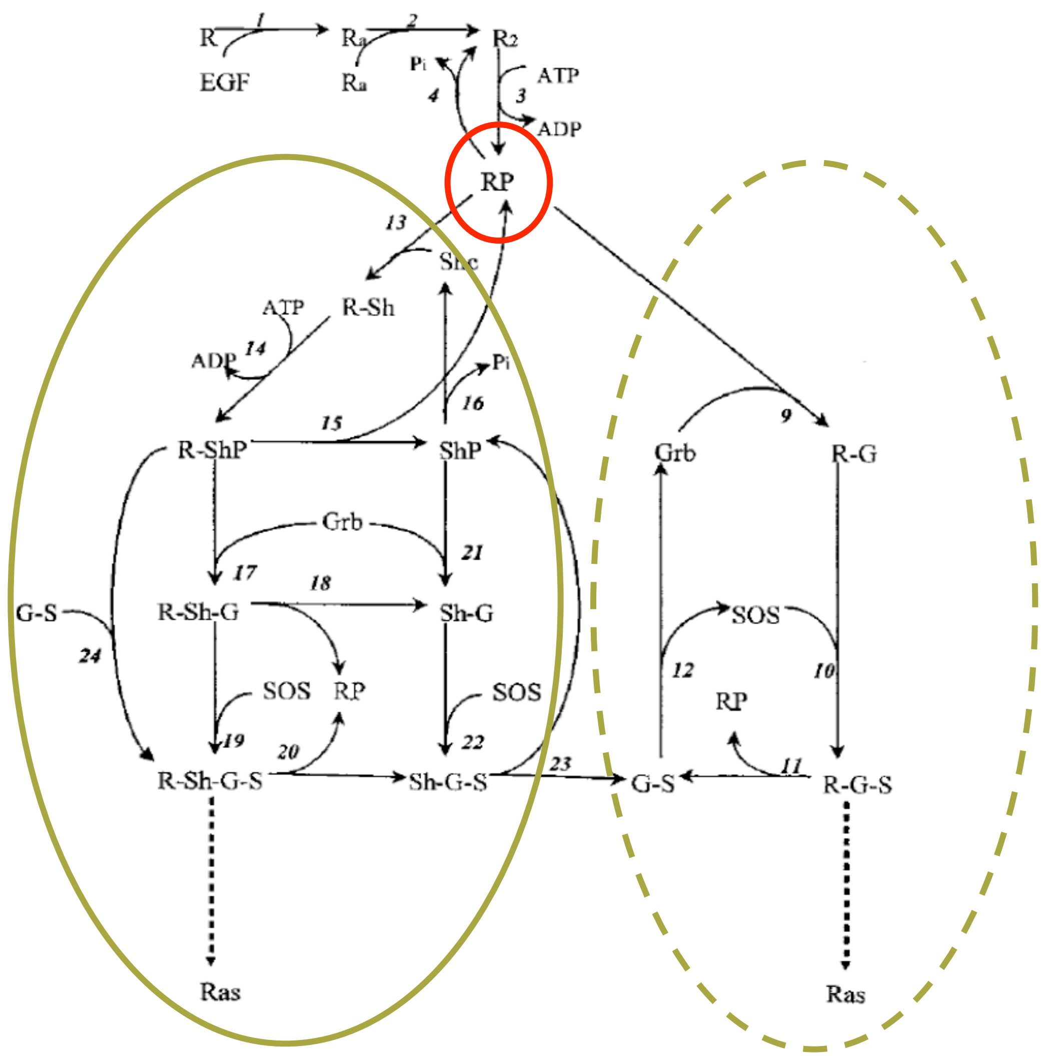


Assumptions made to limit combinatorial complexity

2. Adaptor binding is competitive

No dimers with more than one associated adapter





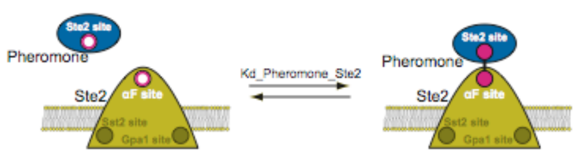
Reminders

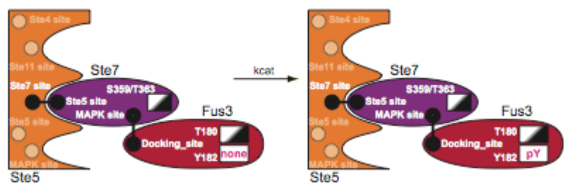
Graphs represent molecules, their component parts, and states

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



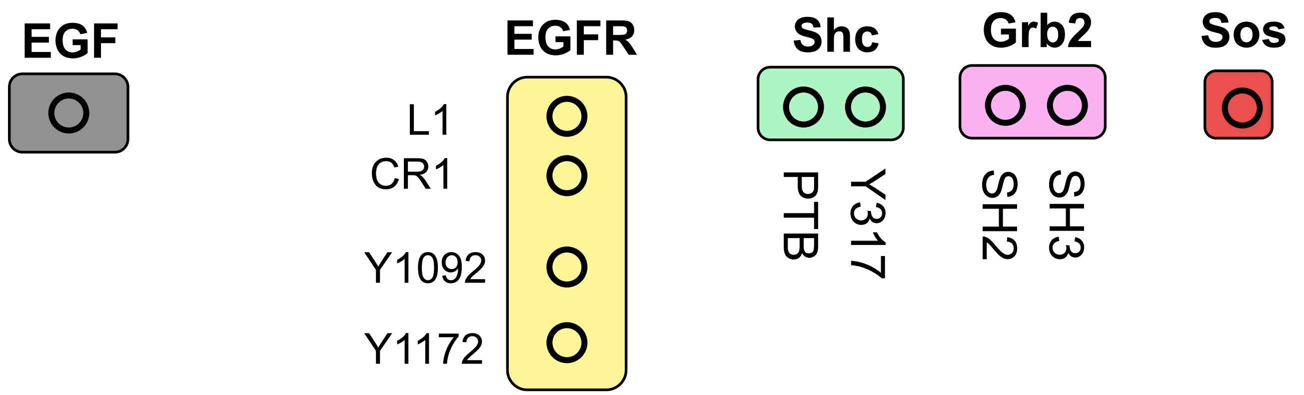


Ste7	Fus3	kcat
S359/T363	T180	
pS	none	kcat_Ste5Ste7***pS***Fus3_pY
pS	pT	kcat_Ste5Ste7***pS***Fus3***pT***_pY
pSpT	none	kcat_Ste5Ste7***pSpT***Fus3_pY
pSpT	pT	kcat_Ste5Ste7***pSpT***Fus3***pT***_pY

Ty Thomson (MIT) - yeastpheromonemodel.org

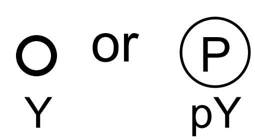
Molecules are modeled as graphs

Molecules

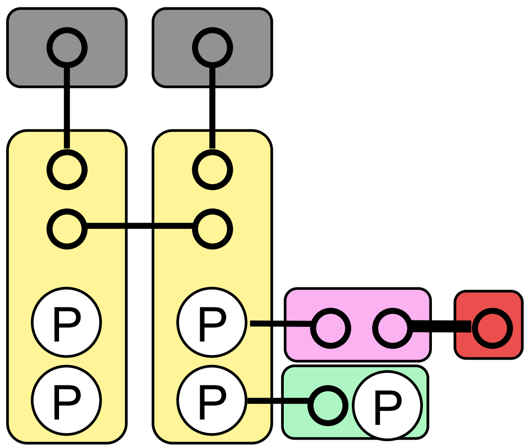


Nodes represent components of proteins

Y components may have labels:



Molecular complexes are simply connected molecules



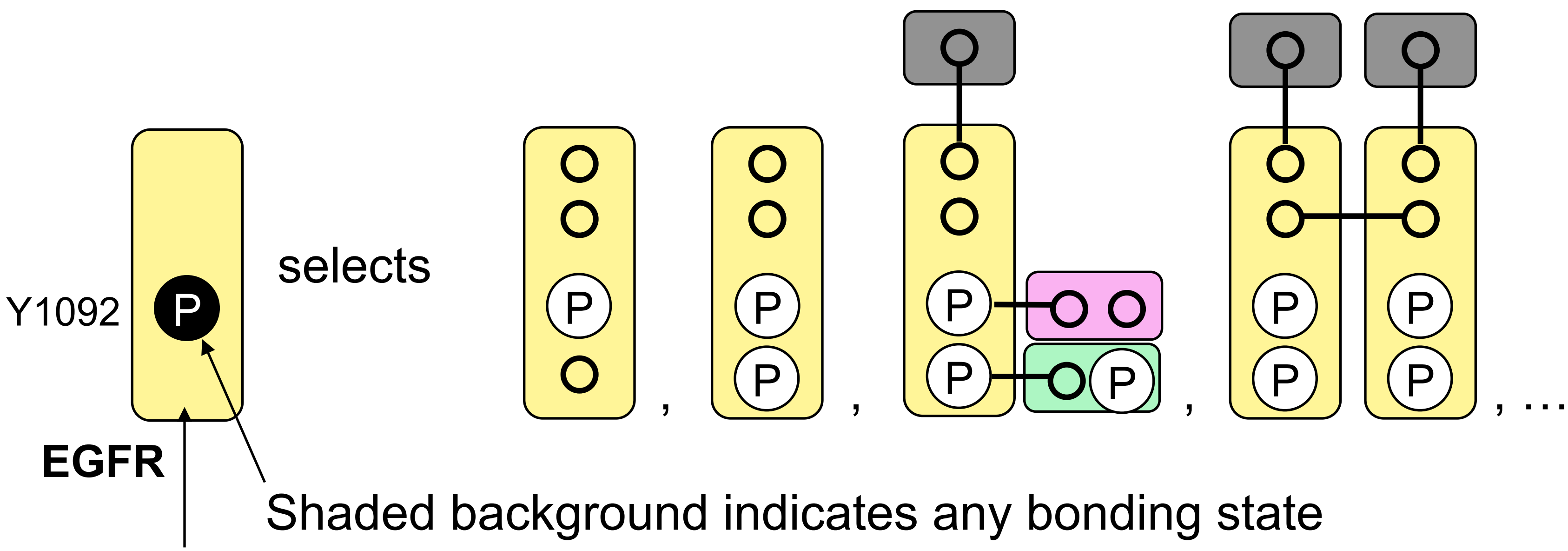
No need to introduce a unique name (e.g., X_{123} or ShP-RP-G-Sos) for each chemical species, as in conventional modeling

Edges represent bonds between components

Bonds may be intra- or intermolecular

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

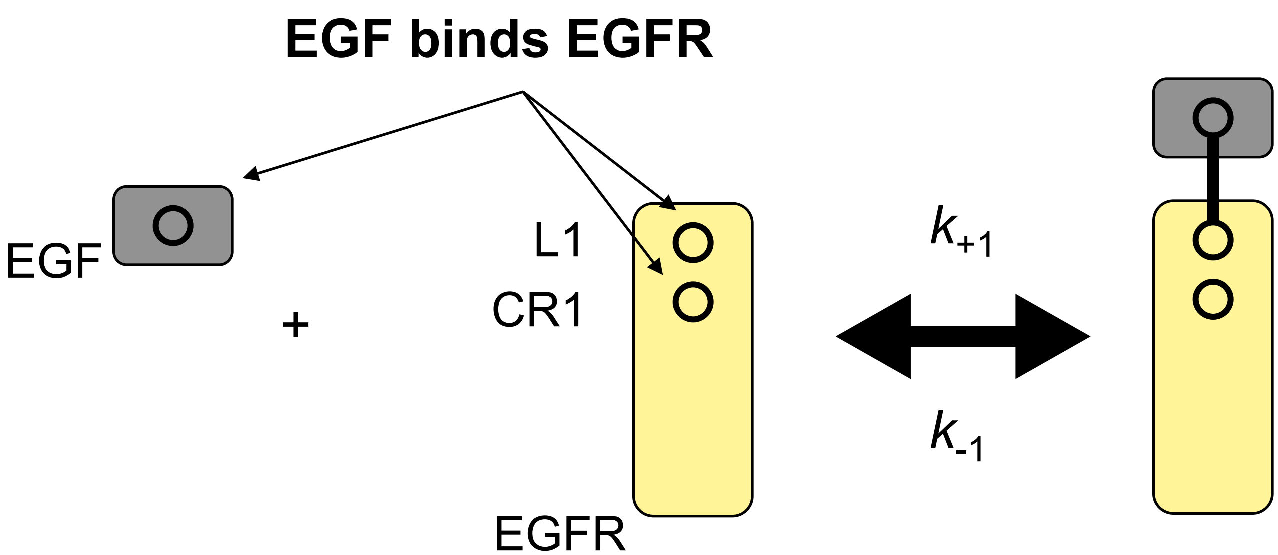
A pattern that matches EGFR phosphorylated at Y1092



Shaded background indicates any bonding state

Suppressed components don't affect match

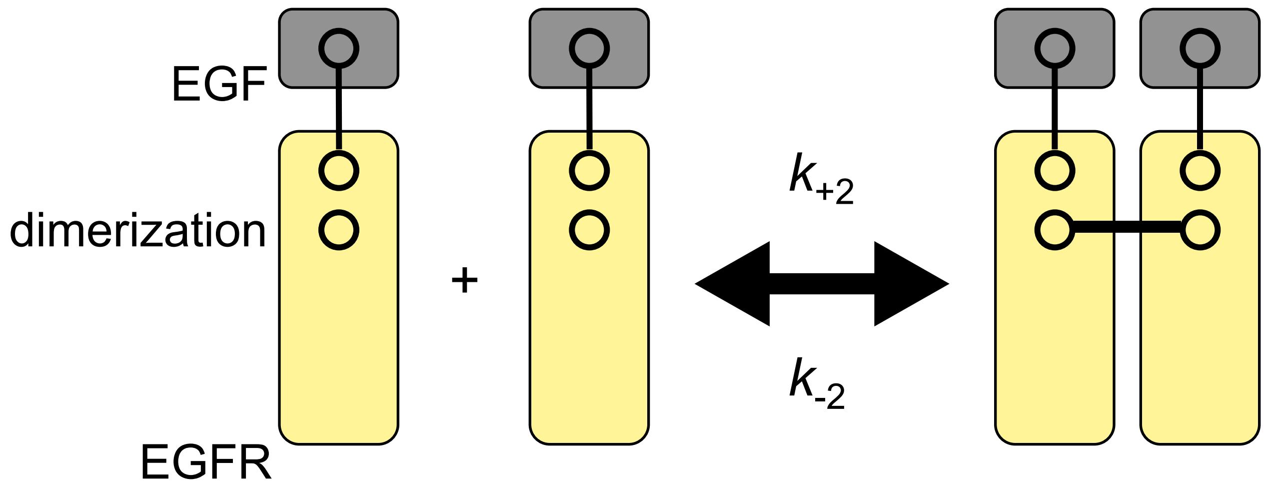
A reaction rule, composed of patterns, defines a class of reactions



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)



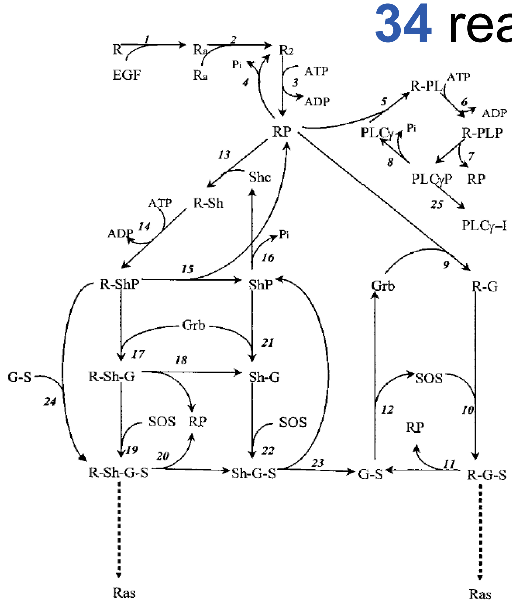
No free lunch: According to this rule, dimers form and break up with the same fundamental rate constants regardless of the states of cytoplasmic domains, which is an idealization.

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).**

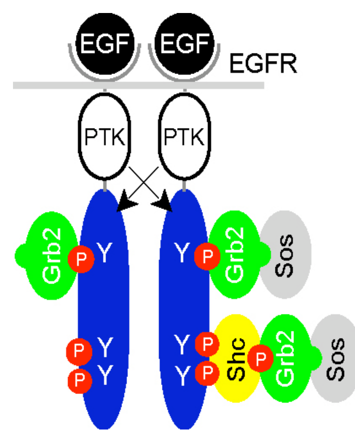
18 species

34 reactions



356 species

3749 reactions

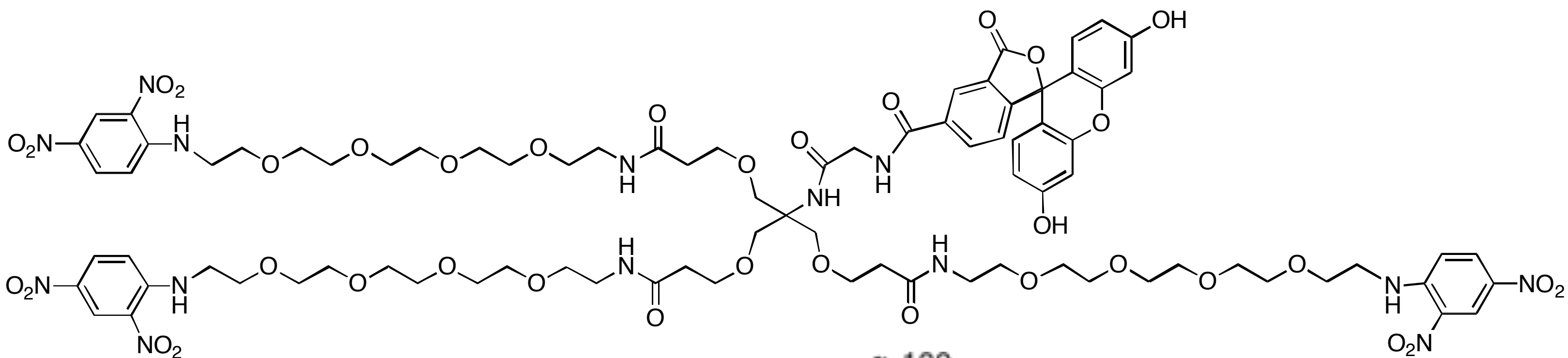


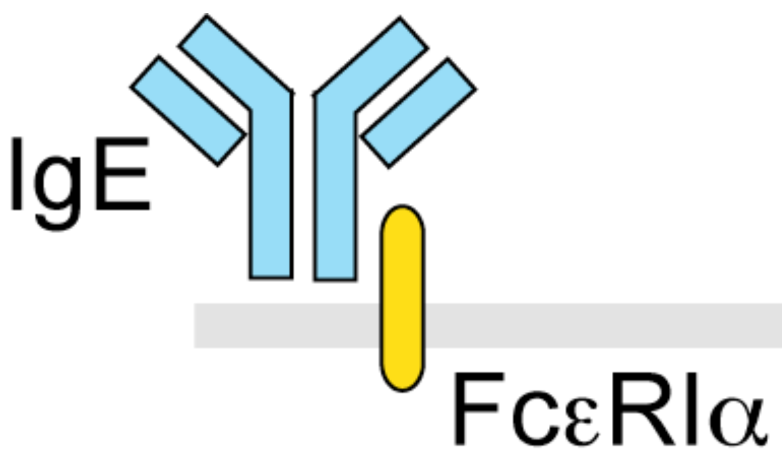
Blinov et al. *Biosystems* **83**, 136 (2006).

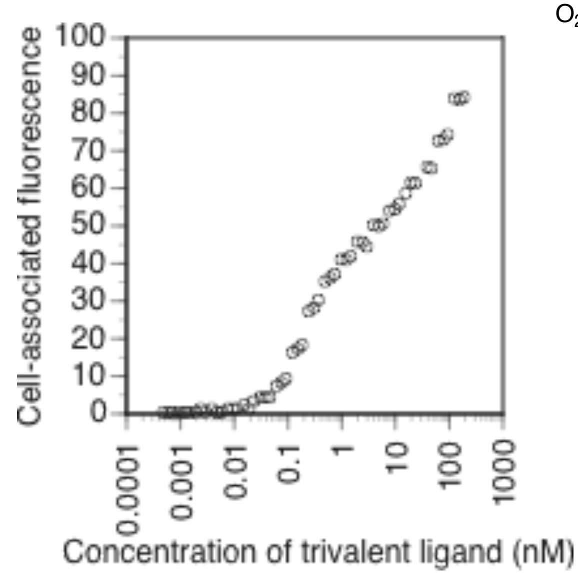
Outline

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

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

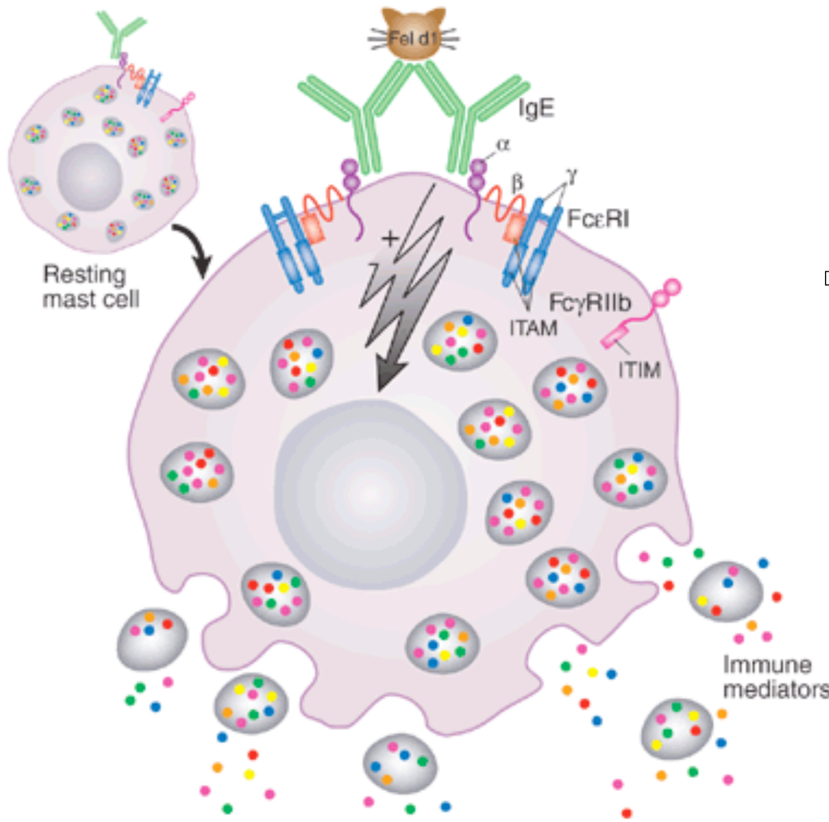




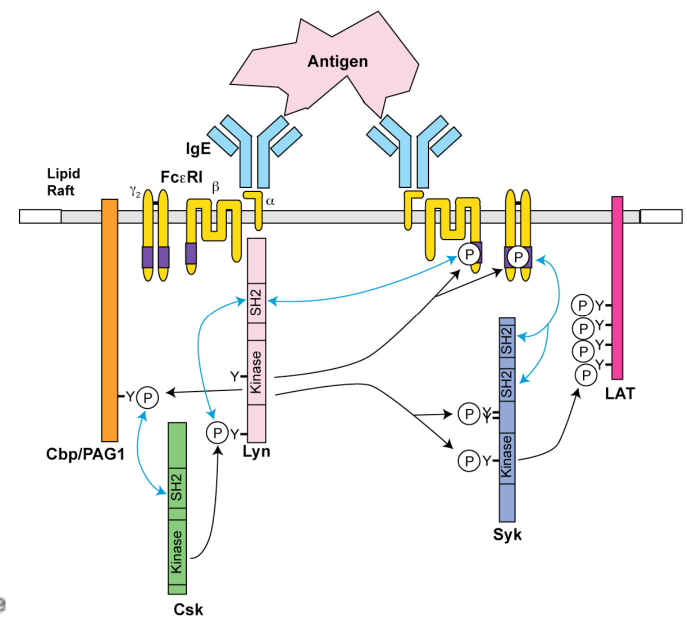


R.G. Posner (TGen) and P.B. Savage (BYU)

Signaling by FcεRI begins with ligand-induced receptor clustering

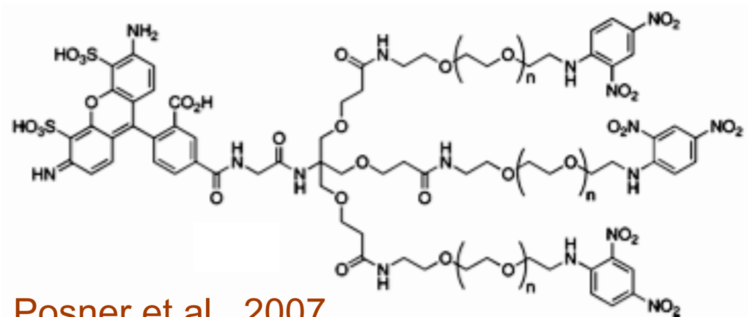


FcεRI-dependent mast cell activation:
Degranulation (release of histamine, etc.)
Cytokine/chemokine release (TNF-α, IL-6, IL-8, etc.)
Arachidonic acid metabolite release (LTC_4, PGD_2, etc.)
↓
Anaphylaxis, allergy (atopic asthma), etc.



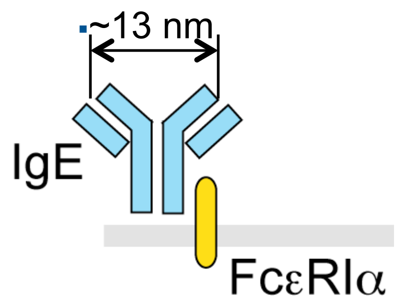
Los Alamos
NATIONAL LABORATORY
EST.1943

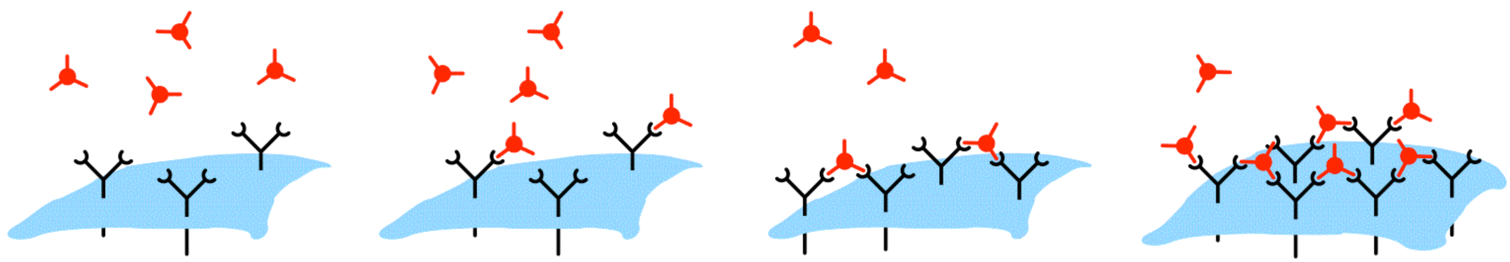
Trivalent ligands



Posner et al., 2007, *Org. Lett*, **9**:3551

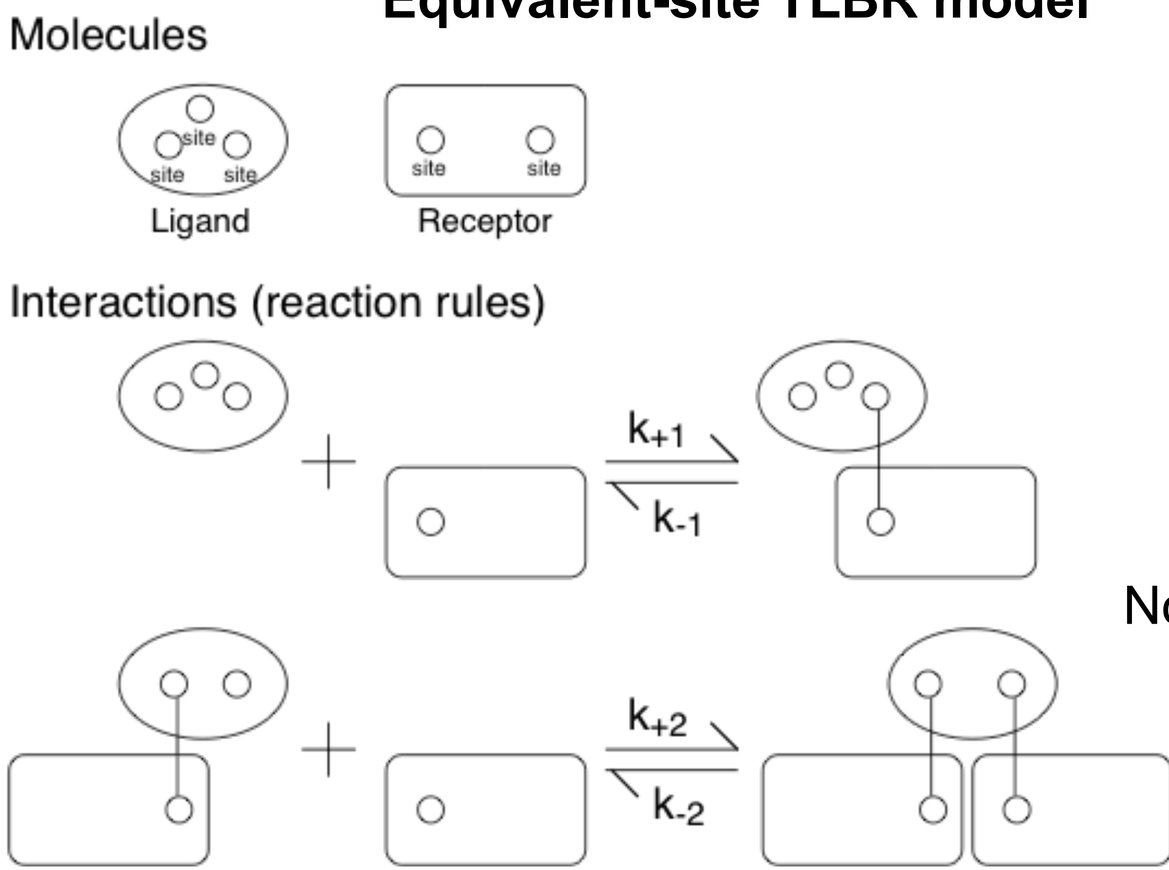
- Compound 6a



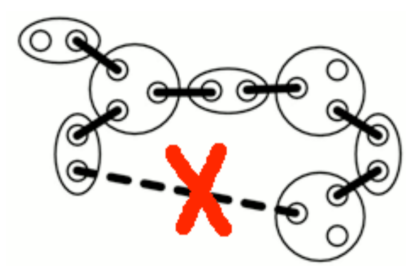


Rule-based model specification corresponding to equilibrium model of Goldstein and Perelson (1984)

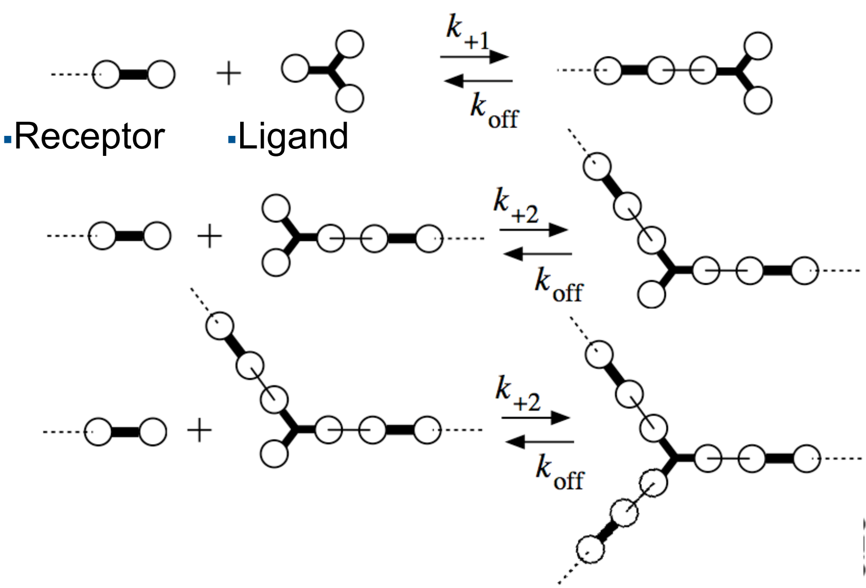
Equivalent-site TLBR model



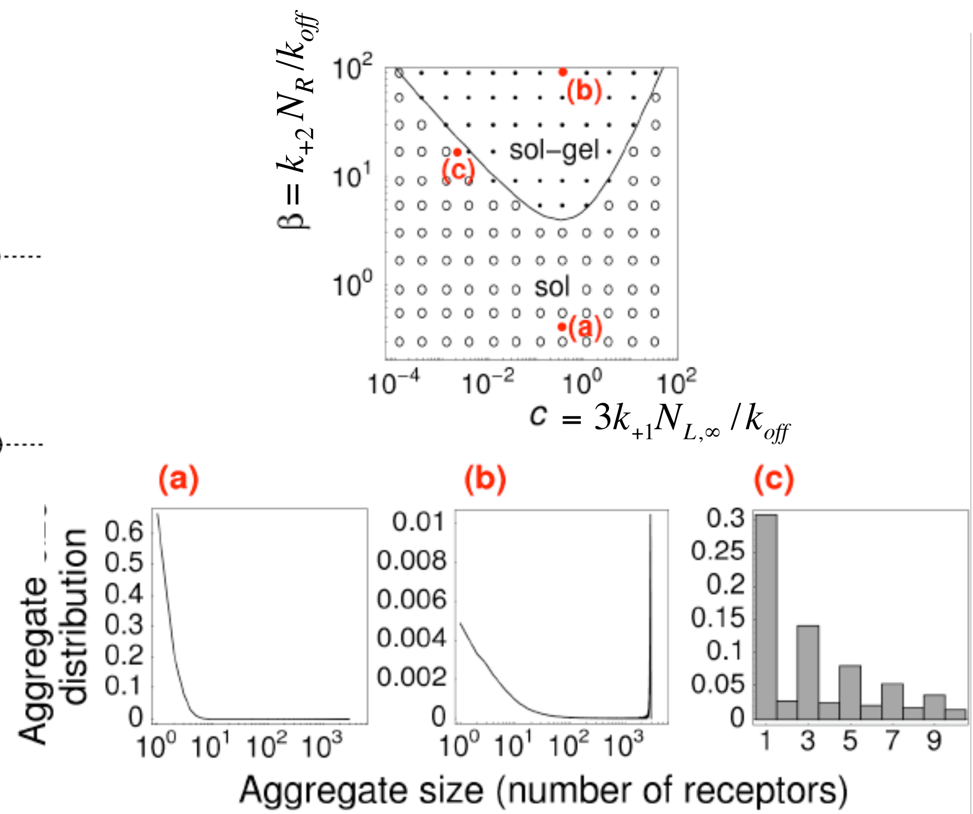
No cyclic aggregates



Goldstein-Perelson and TLBR models



- Equilibrium properties:

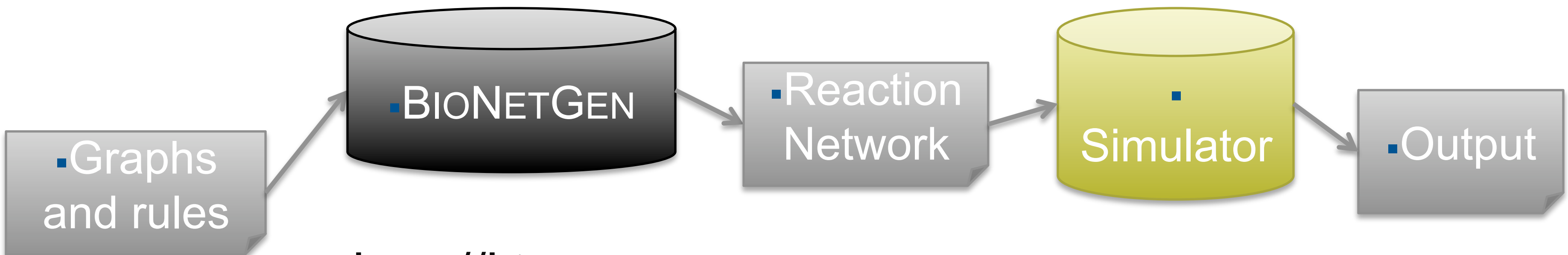


- Goldstein and Perelson (1984) *Biophys. J.*, 45:1109
- Yang et al. (2008) *Phys. Rev. E,* 78:31910

Protocol for “generate-first” simulation

- 1. Define molecules as *graphs* and interactions as *graph-rewriting rules*.
- 2. Specify **concentrations** and **rate constants**

3. Generate the implied reaction network and then simulate the network dynamics using conventional methods

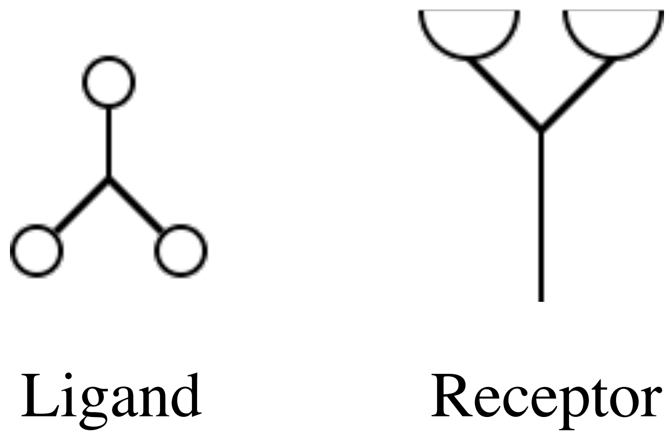


- http://bionetgen.org

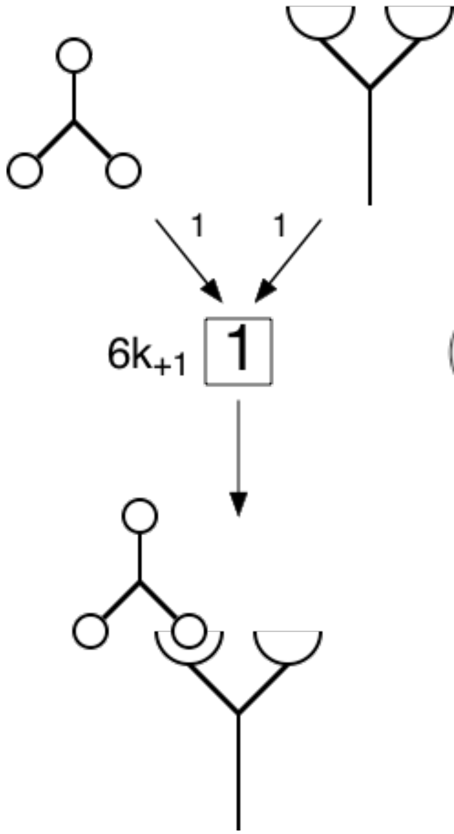
Los Alamos
NATIONAL LABORATORY
EST. 1943

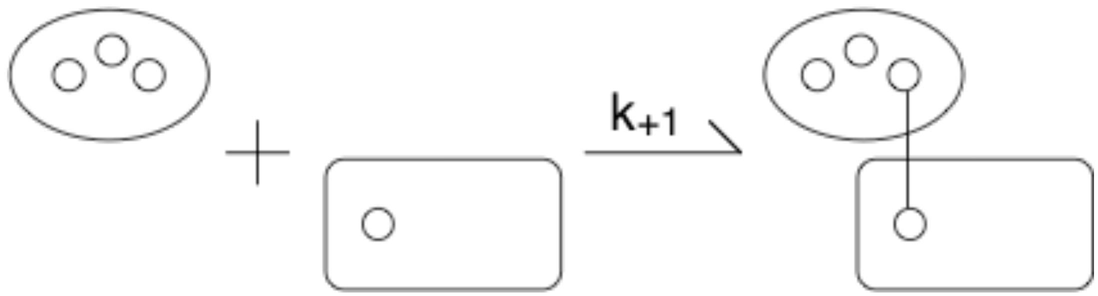
- Faeder, Blinov, and Hlavacek, *Methods Mol. Biol.* (2009)

“Generate-first” method starts with seed species

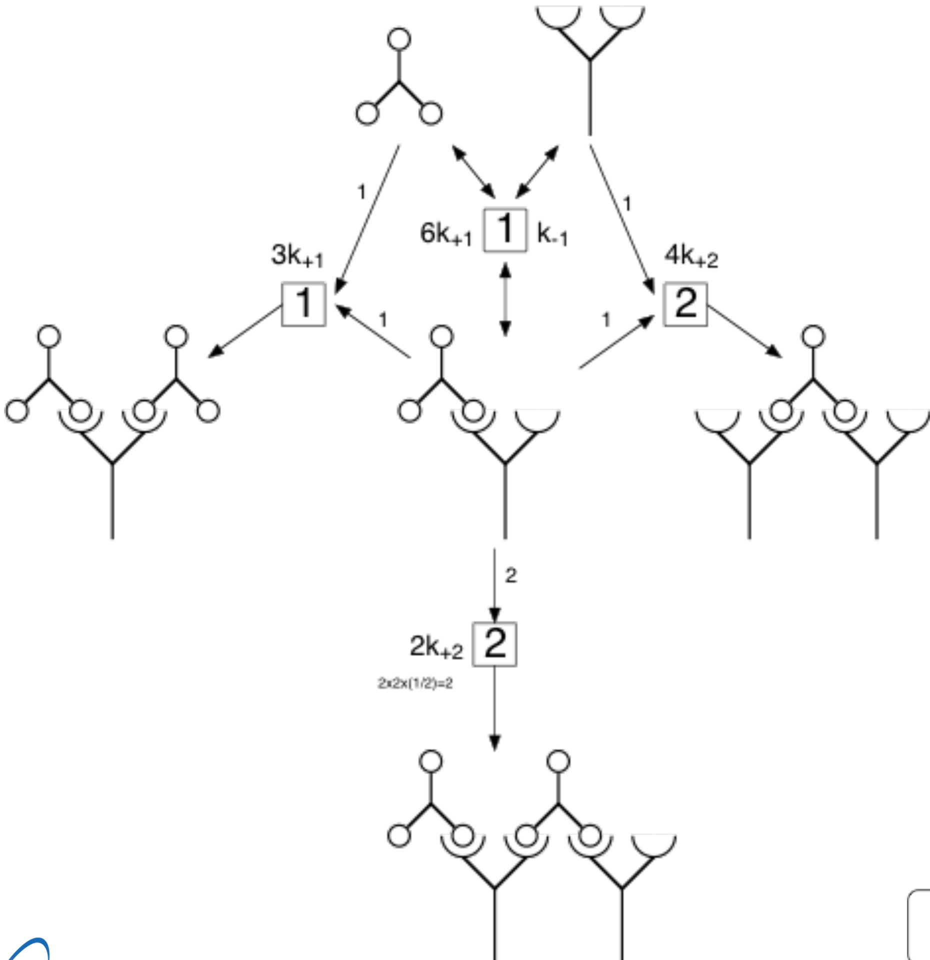


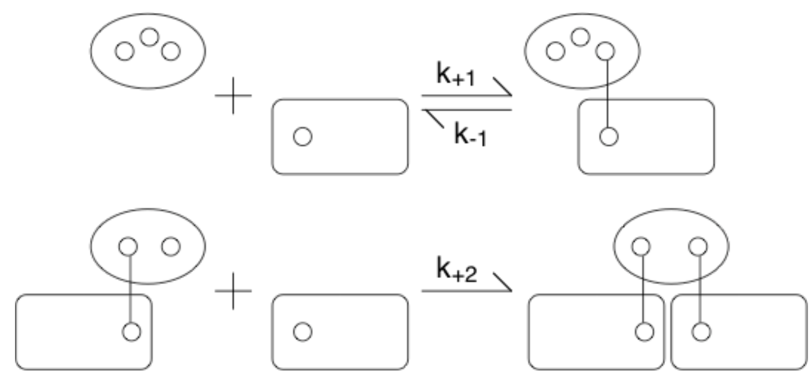
After first round of rule application





After the second round of rule application



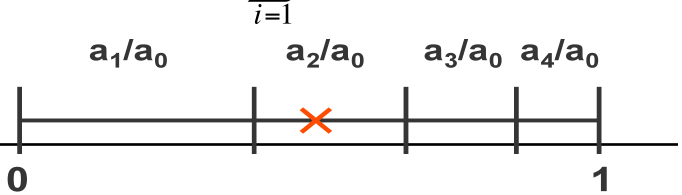


Gillespie method: generate-first or on-the-fly simulation

- 0
  - Iterative application of rules to generate network (species, reactions)
  - Set **x**(0)
  - Calculate **a**(0), $a_0(0) = \sum_i a_i(0)$

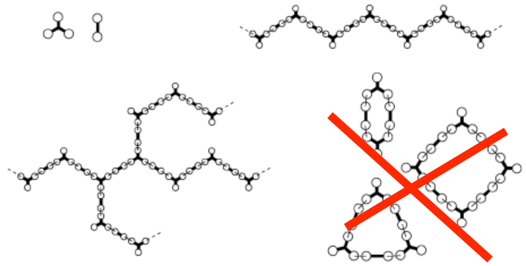
- 1
  - Select next event time
    $\tau = -\ln \rho_1 / a_0(t)$

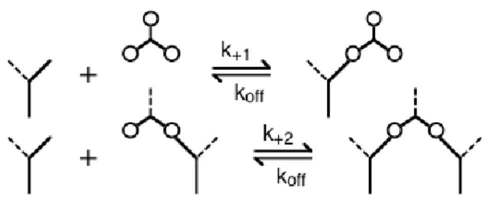
- 2
  - Select next reaction, r
    $\min r \text{ s.t. } \sum_{i=1}^{r} a_i(t) \geq \rho_2 a_0(t)$

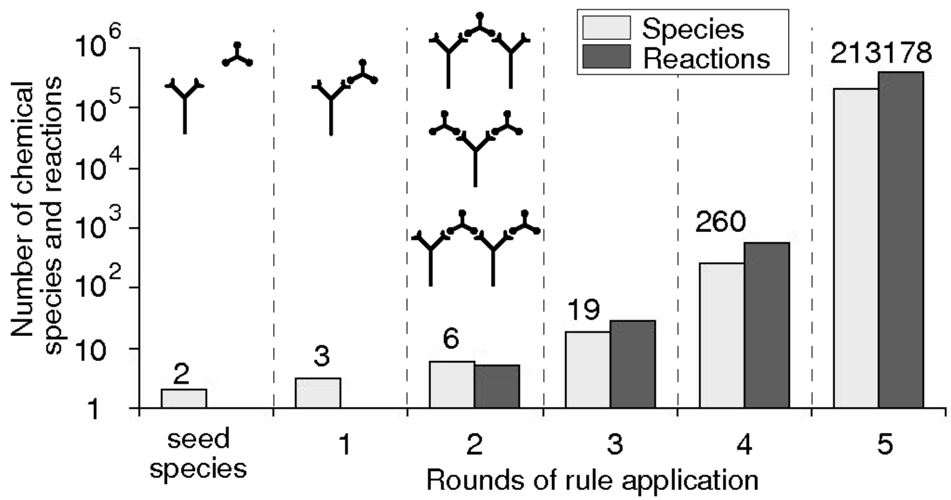


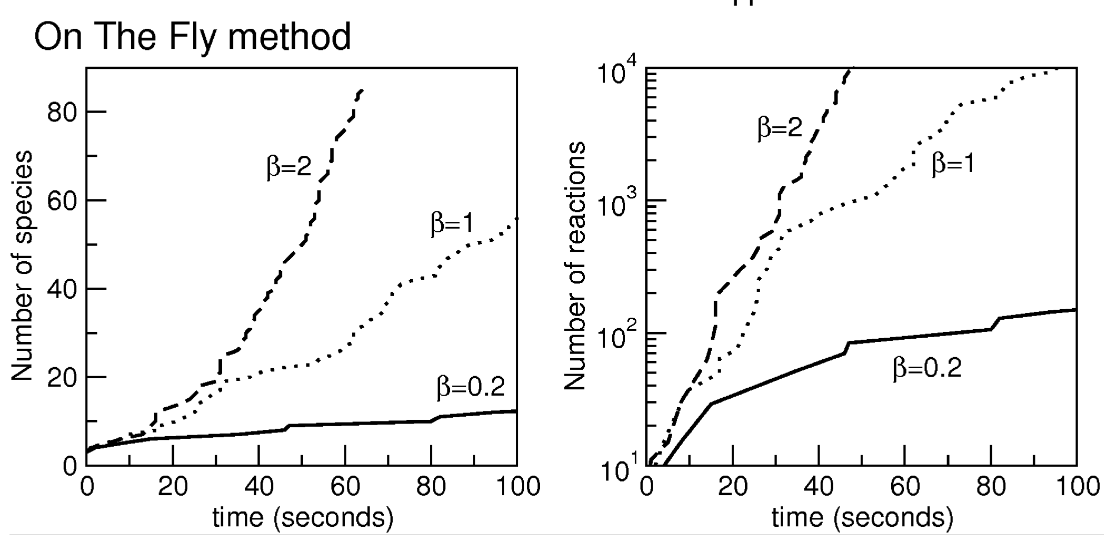
- 3
  - Update **x**, **a**(t), a_0, t
    Update only $a_i(t),\ i \in \mathrm{dep}(r)$*
    $\mathbf{x}(t+\tau) = \mathbf{x}(t) + \mathbf{S}_r$
  - *rxn q depends on rxn r iff. a reactant of rxn q is a reactant or product of rxn r.

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

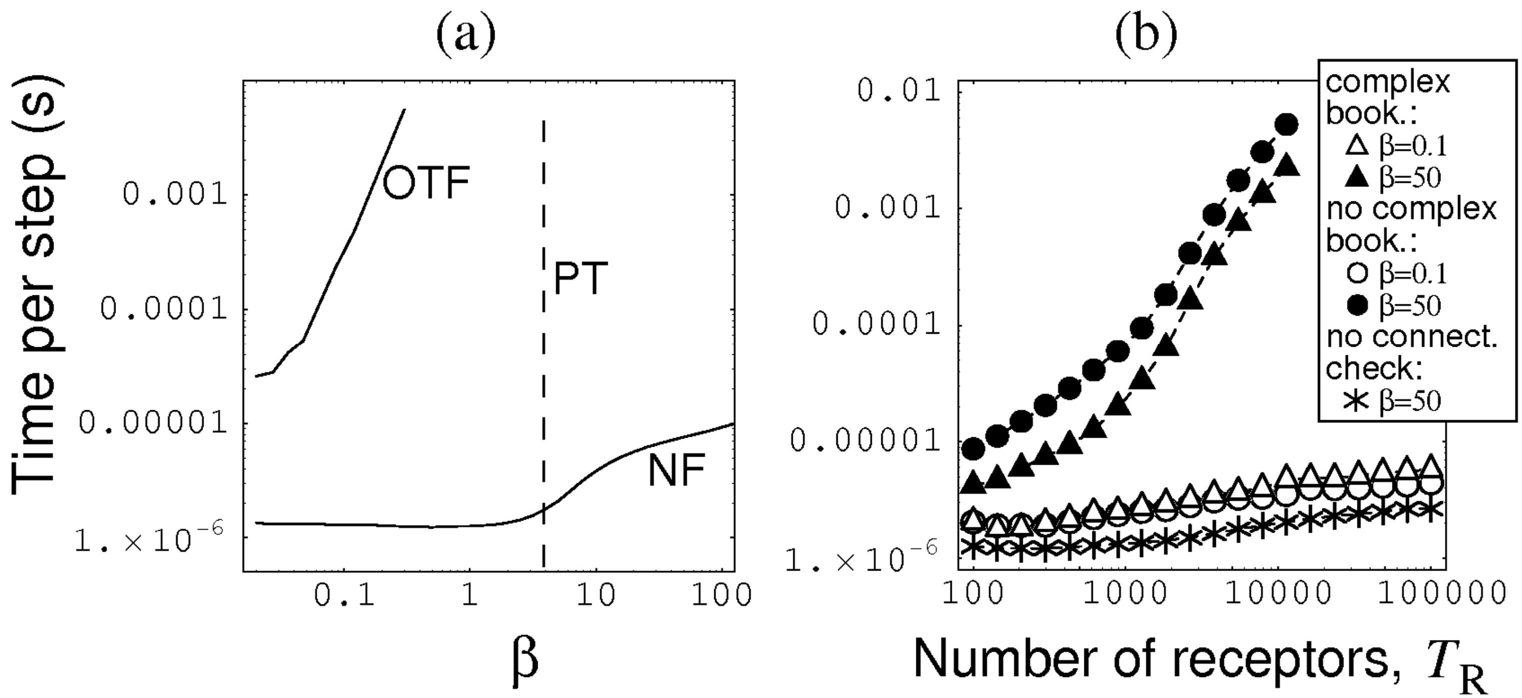








Performance of on-the-fly (OTF) simulation method



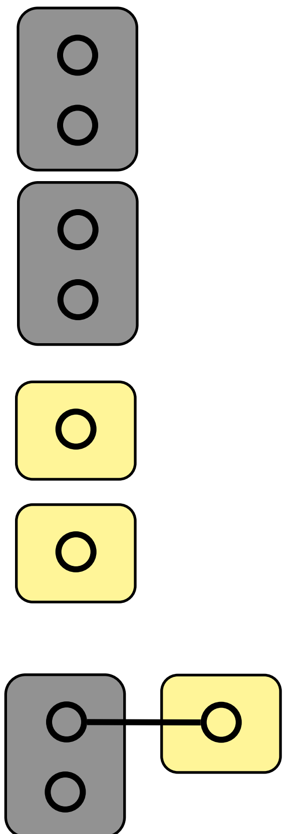
Yang et al. (2008) Phys. Rev. E

Network-free simulation

- **Agents/particles in simulation "box"**

Network-free simulation

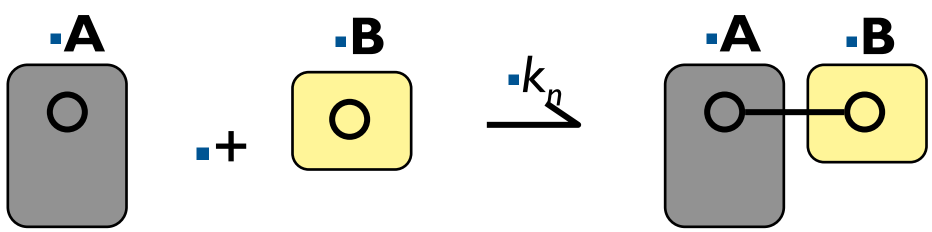
- **<u>Agents/particles</u>**



- ...

- **Rules are event generators**

- ...



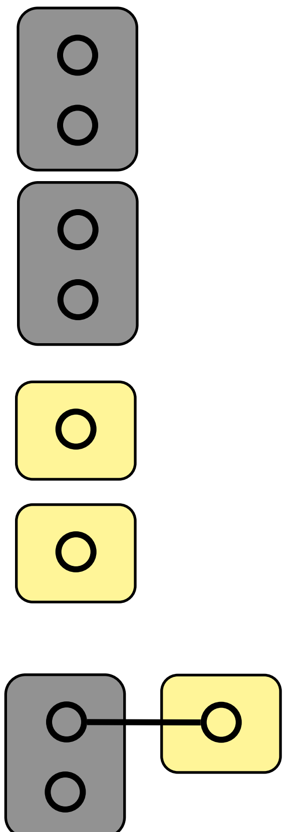
- **Rule *n***

- Cumulative rate = $a_n = k_n [A][B]$

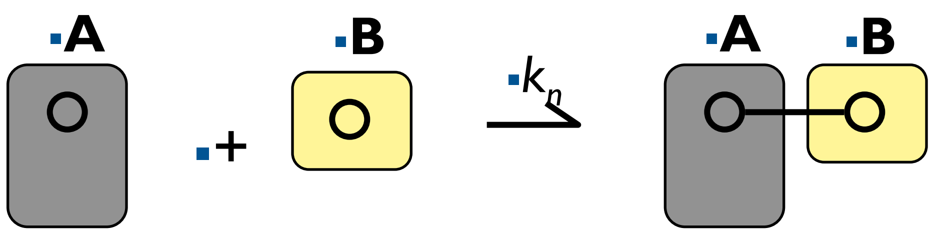
- ...

Network-free simulation

<u>Agents/particles</u>



Event *n* is chosen to fire using Gillespie algorithm



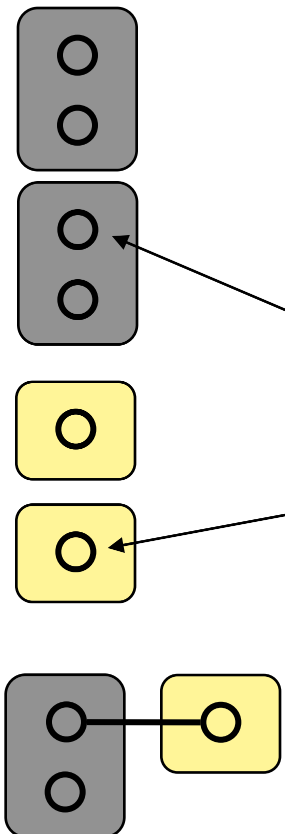
Rule *n*

Cumulative rate = a_n

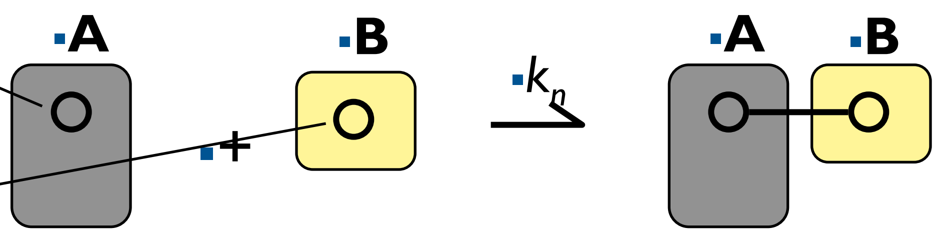
Los Alamos
NATIONAL LABORATORY
EST. 1943

Network-free simulation

Agents/particles



Reactant molecules chosen randomly

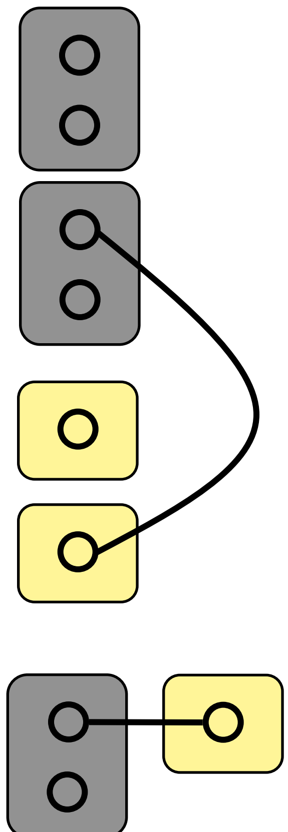


Rule *n*

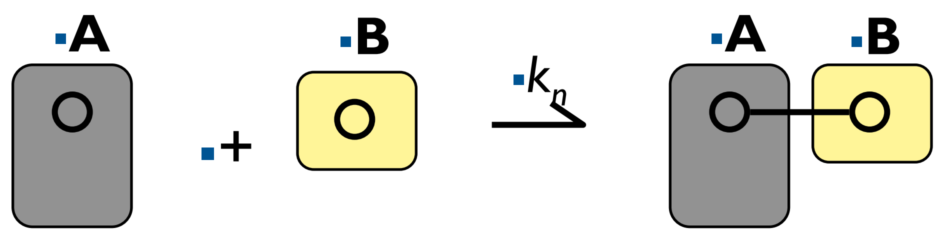
Cumulative rate = a_n

Network-free simulation

- **<u>Agents/particles</u>**



- **Rule transformation is applied**



- **Rule *n***

- Cumulative rate = a_n

Kinetic Monte Carlo method for "network-free" simulation of rule-based models

1. Instantiate molecules with components and states.

2. Determine cumulative rate for each mth reaction type,

$$r_m = k_m \prod_n^{n_m} N_n$$

3. Select next reaction time,

$$\Delta t = -\ln(z_1)/r_{tot}$$

4. Select next reaction type using the following condition:

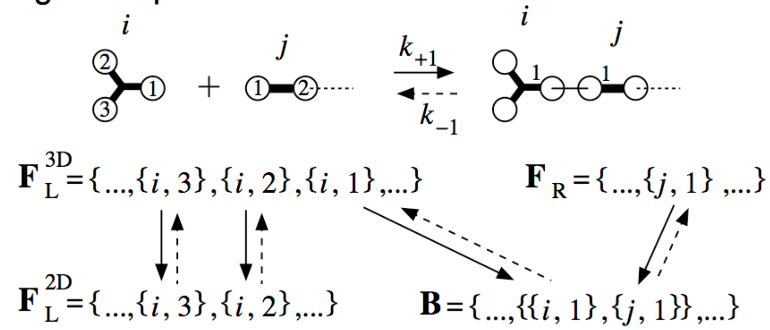
$$\sum_{j=1}^{J-1} r_j < z_2 r_{tot} \le \sum_{j=1}^{J} r_j$$

5. Select reactant molecules and **check context**.

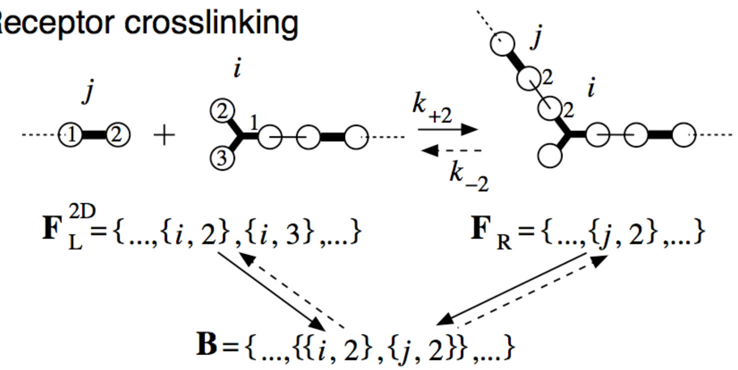
6. **Update lists**. Iterate.

List updates:

Ligand capture



Receptor crosslinking



Yang et al. (2008) *Phys. Rev. E,* 78:031910

Danos et al. (2007) *Lect. Notes Comp. Sci.*

Los Alamos NATIONAL LABORATORY EST. 1943

Conclusions

- Mechanistic models of cell signaling systems can be formulated via the rule-based modeling approach, simulated and used, for example, to provide a mechanistic interpretation of temporal phosphoproteomic data (not shown)
- Comprehensive models of cell signaling systems (on the way) should serve as launching pads for investigating a wide array of issues related to development of predictive models for cell signaling systems
  - What is required for model validation?
  - What are the best strategies for certification (e.g., model-guided experimental design)?
  - Can we quantify and track how consistent a model is with available knowledge?

Outline

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