Rule-based Modeling

STMC
SpatioTemporal Modeling Center

William S. Hlavacek

Theoretical Division

Los Alamos National Laboratory

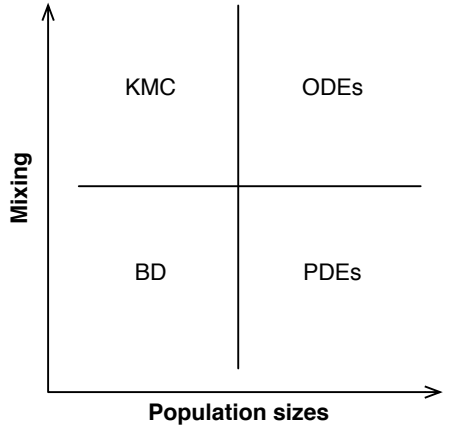
Los Alamos
NATIONAL LABORATORY
EST.1943

Center for
Nonlinear Studies

What is rule-based modeling?

- **A modeling approach based on the principles of chemical kinetics**
- **An approach distinguished by its means of model specification: local rules vs. equations**
- **An approach based on assumptions of modularity – the interaction represented by a rule depends on molecular context only as explicitly specified in the rule**
- **An approach useful for studying biomolecular site dynamics in cell signaling systems (e.g., changes in levels of phosphorylation at individual pTyr sites)**

There's an RBM equivalent for many traditional modeling approaches



Outline

1. **The motivation for modeling, and rule-based modeling in particular**
2. Basic concepts of rule-based modeling
3. Indirect and direct methods for simulating a model

The need for predictive models of cell signaling systems

- **These systems mediate cellular information processing and decision-making (regulation of cellular activities and fates)**
- **These systems 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**
  - 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
  - Targeted therapeutics can have counterintuitive effects (e.g., BRAFi)

Cell signaling behavior depends on quantitative properties

Sci. Signal., 16 October 2012
Vol. 5, Issue 246, p. pe46
[DOI: 10.1126/scisignal.2003354]

PERSPECTIVES

The Response of Cancers to BRAF Inhibition Underscores the Importance of Cancer Systems Biology

Edward C. Stites[*†]

Clinical Translational Research Division, The Translational Genomics Research Institute, Phoenix, AZ 85004, USA.

† Present address: Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Avenue, St. Louis, MO 63110, USA.

Abstract: The BRAF inhibitor vemurafenib has become an important treatment option for melanoma patients, the majority of whom have a BRAF(V600E) mutation driving their malignancy. However, this same agent does not generally benefit colon cancer patients who have the BRAF(V600E) mutation. Recent work suggests that BRAF(V600E) inhibition by vemurafenib results in decreased negative feedback to the epidermal growth factor receptor (EGFR) pathway and that the different clinical responses are due to differences in the amount of EGFR present in these two cancers. The experimental work that identified the feedback signaling was an elegant mix of functional genomic approaches and focused, hypothesis-driven cellular and molecular biology. The results of these studies suggest that combined treatment of BRAF(V600E)-driven colon cancers with both vemurafenib and EGFR inhibitors is worth clinical evaluation.

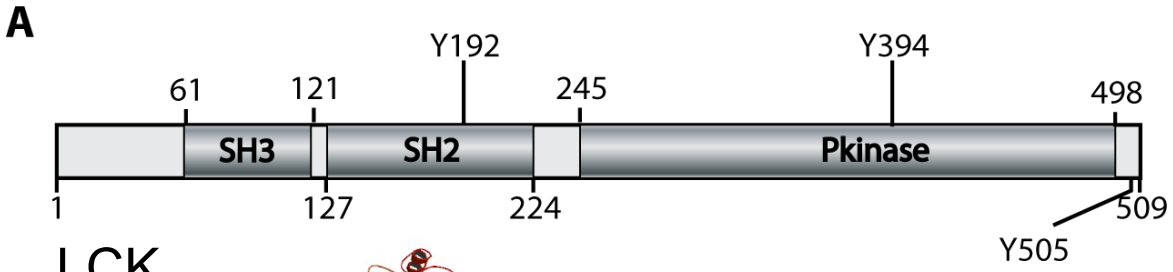
* Corresponding author. E-mail: estites@tgen.org

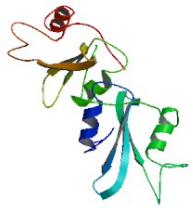
Citation: E. C. Stites, The Response of Cancers to BRAF Inhibition Underscores the Importance of Cancer Systems Biology. *Sci. Signal.* **5**, pe46 (2012).

What's the value added by modeling?

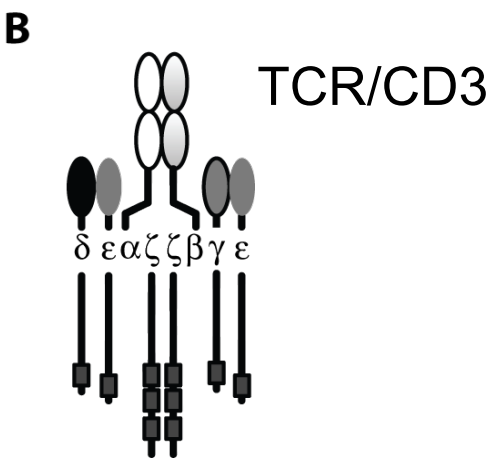
- **We can use models to organize information about a system with precision**
  - Introduces greater rigor and discipline
  - Discovery of knowledge gaps
- **We can determine the logical consequences of a model specification**
  - Design principles can be elucidated (key for intuitive understanding and perhaps synthetic biology)
  - Ranking of clinically available therapies for a given patient (key for personalized medicine)

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





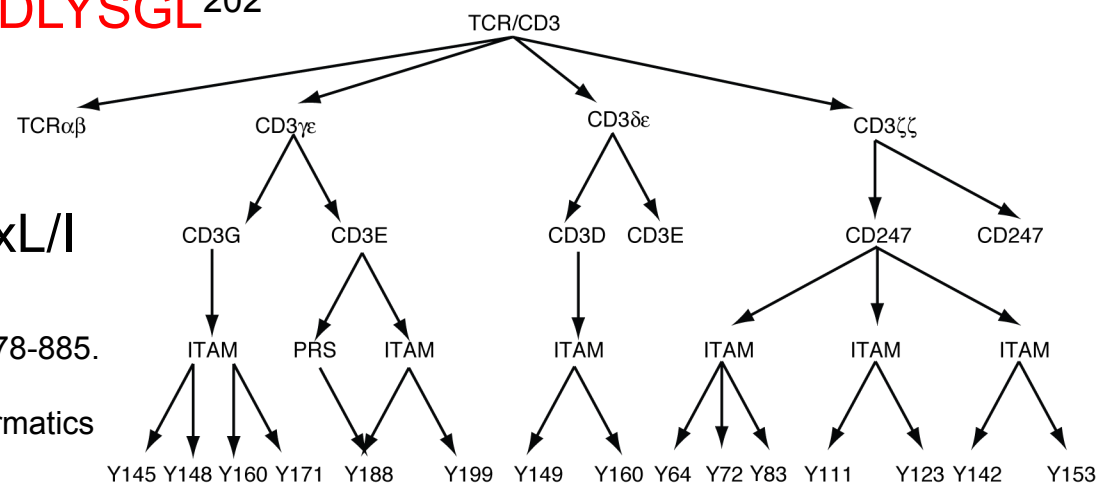
LCK-SH2 (1bhh)



CD3E: ^{184}PNPDYEPIRKGQRDLYSGL^{202}

PRS: PxxDY

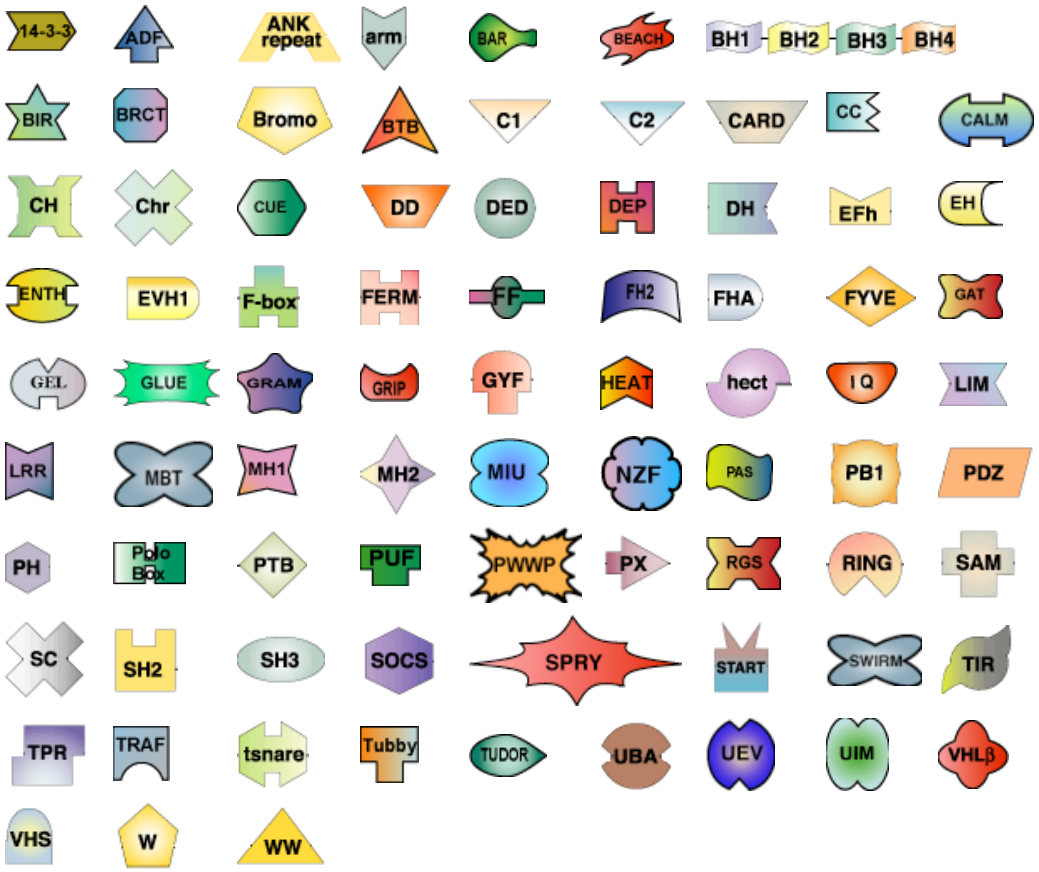
ITAM: YxxL/I(x_{6-8})YxxL/I



Kesti T, et al. (2007) J. Immunol. 179:878-885.

Lemons NW, et al. (2011) BMC Bioinformatics

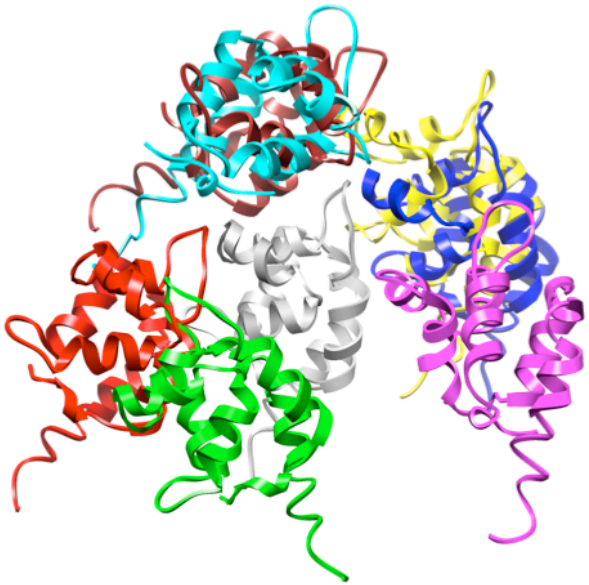
There are many protein interaction domains



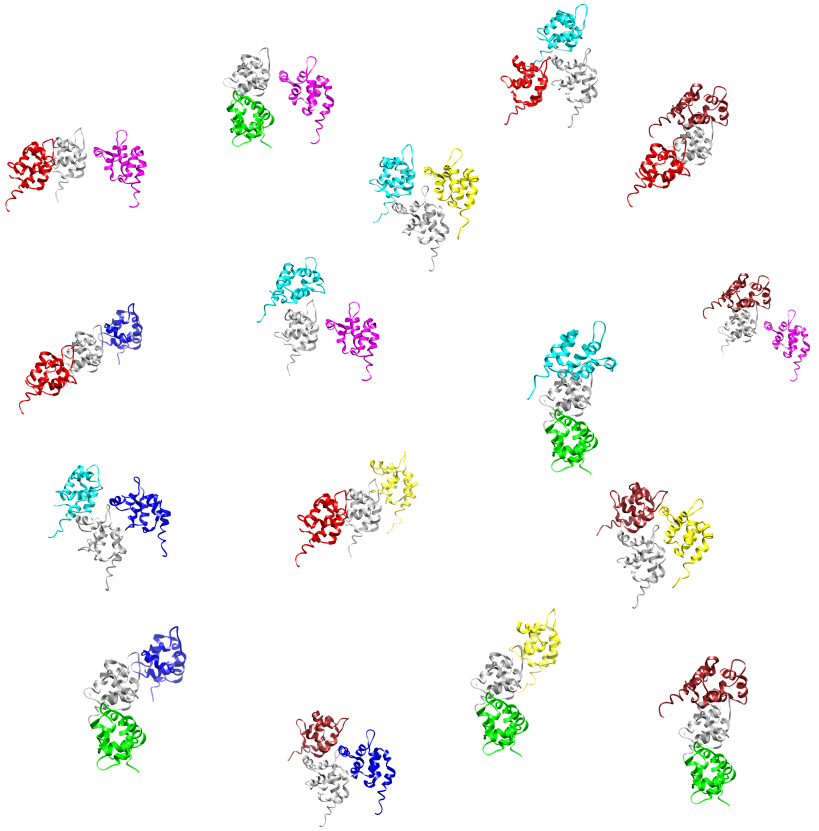
Los Alamos NATIONAL LABORATORY EST.1943

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

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



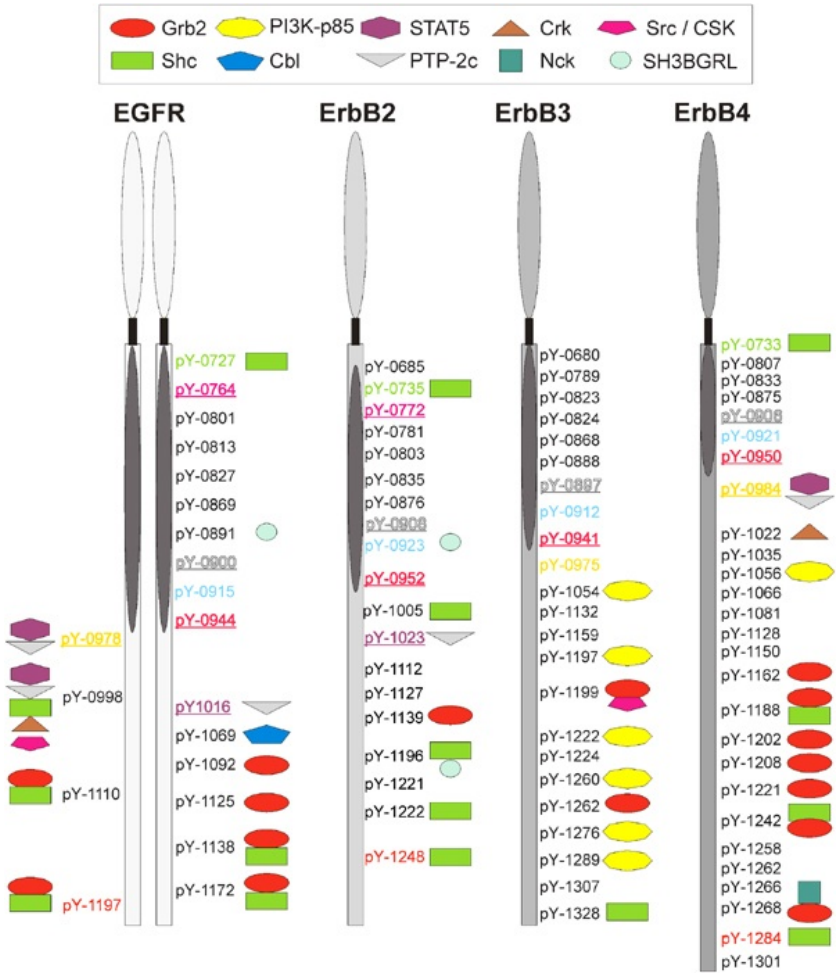
Gray: crd.dd.a; Magenta: crd.dd.b; Cyan: crd.dd.c
Yellow: crd.dd.d; Red: crd.dd.e; Blue: crd.dd.f;
Green: crd.dd.g; Brown: crd.dd.h



There are many possible trimers!

C.-S. Tung

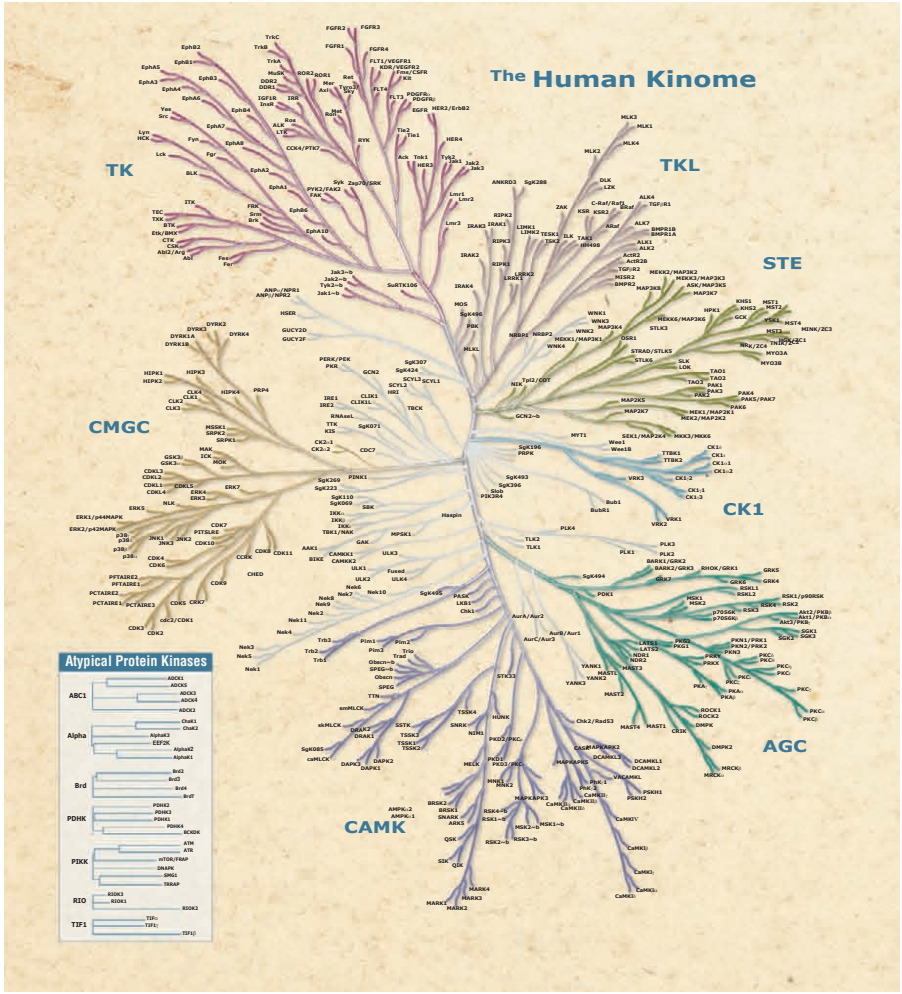
Domain-motif interactions are often controlled by post-translational modifications



There are many possible protein phosphoforms!

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

518 protein kinases (~2% of human genes)

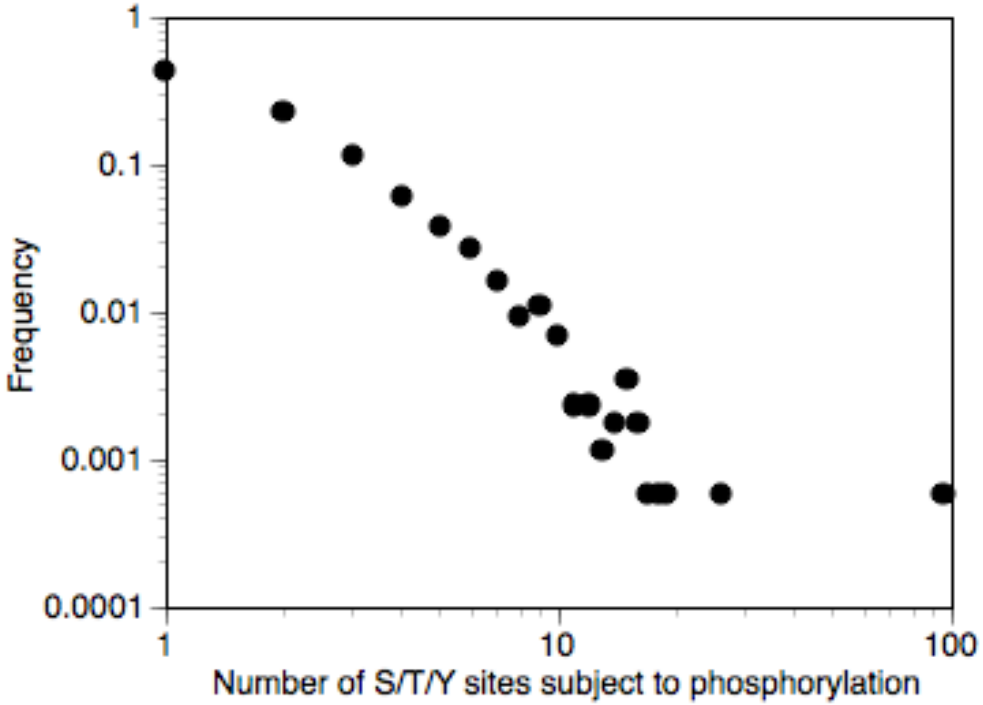


There are phosphatases too!

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)

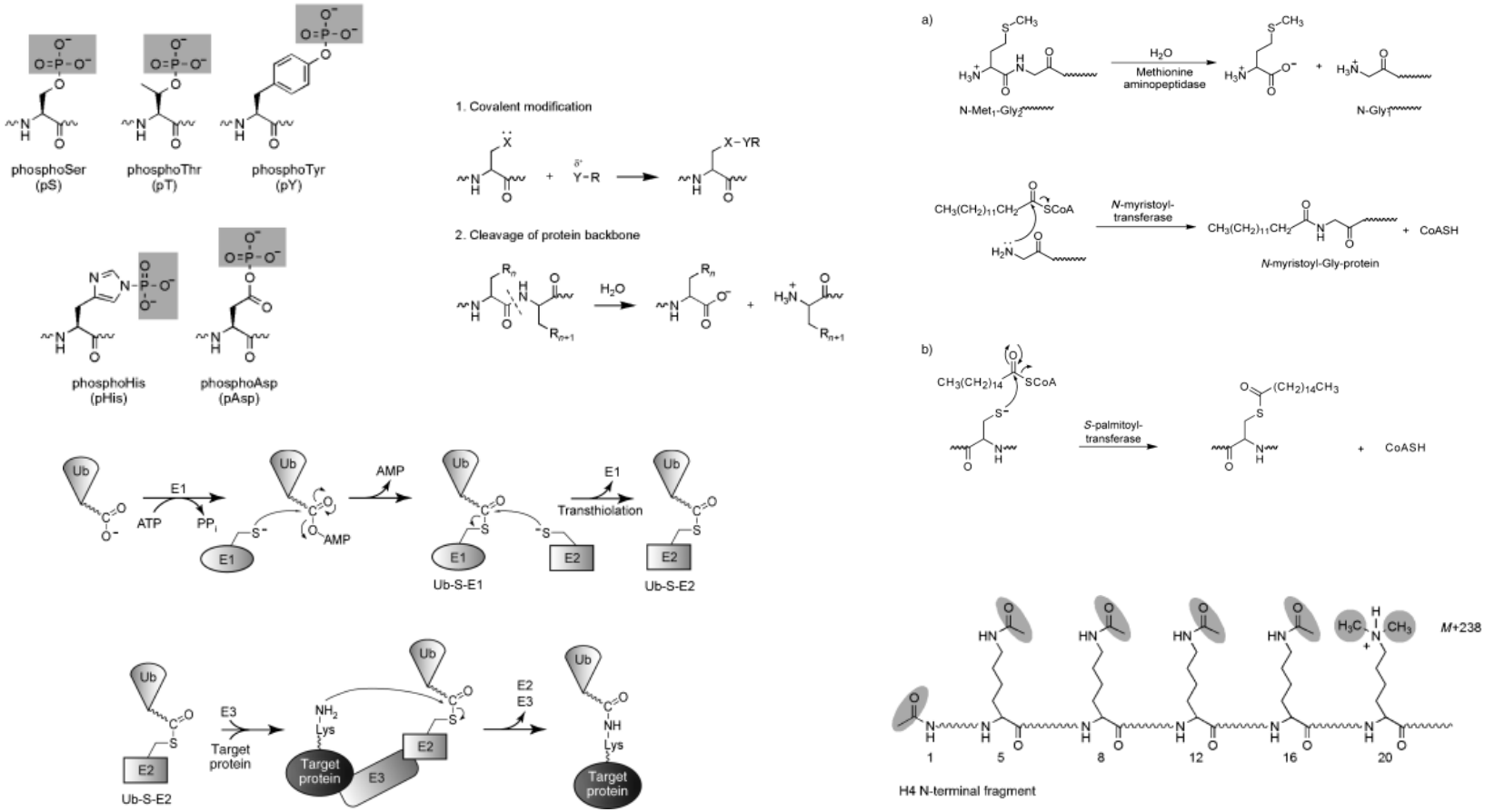


> 50% are phosphorylated at 2 or more sites

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

Hlavacek et al. (2006) Science's STKE

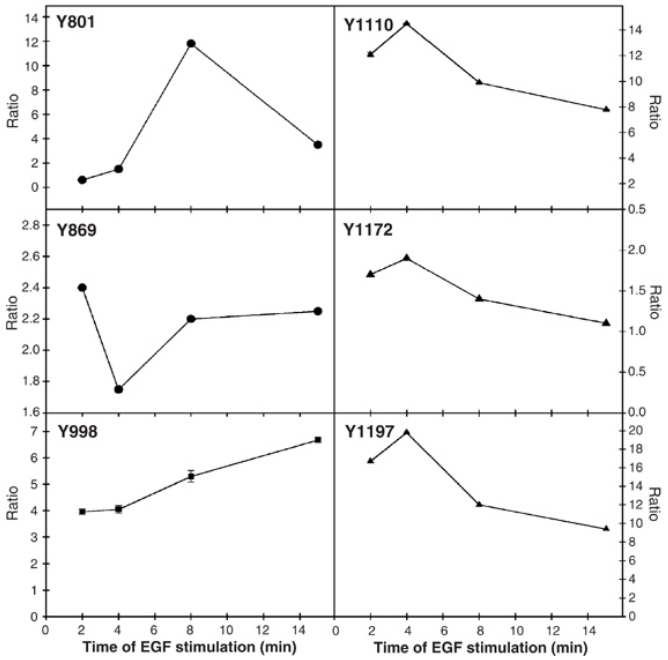
There are many different kinds of post-translational modifications of proteins



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

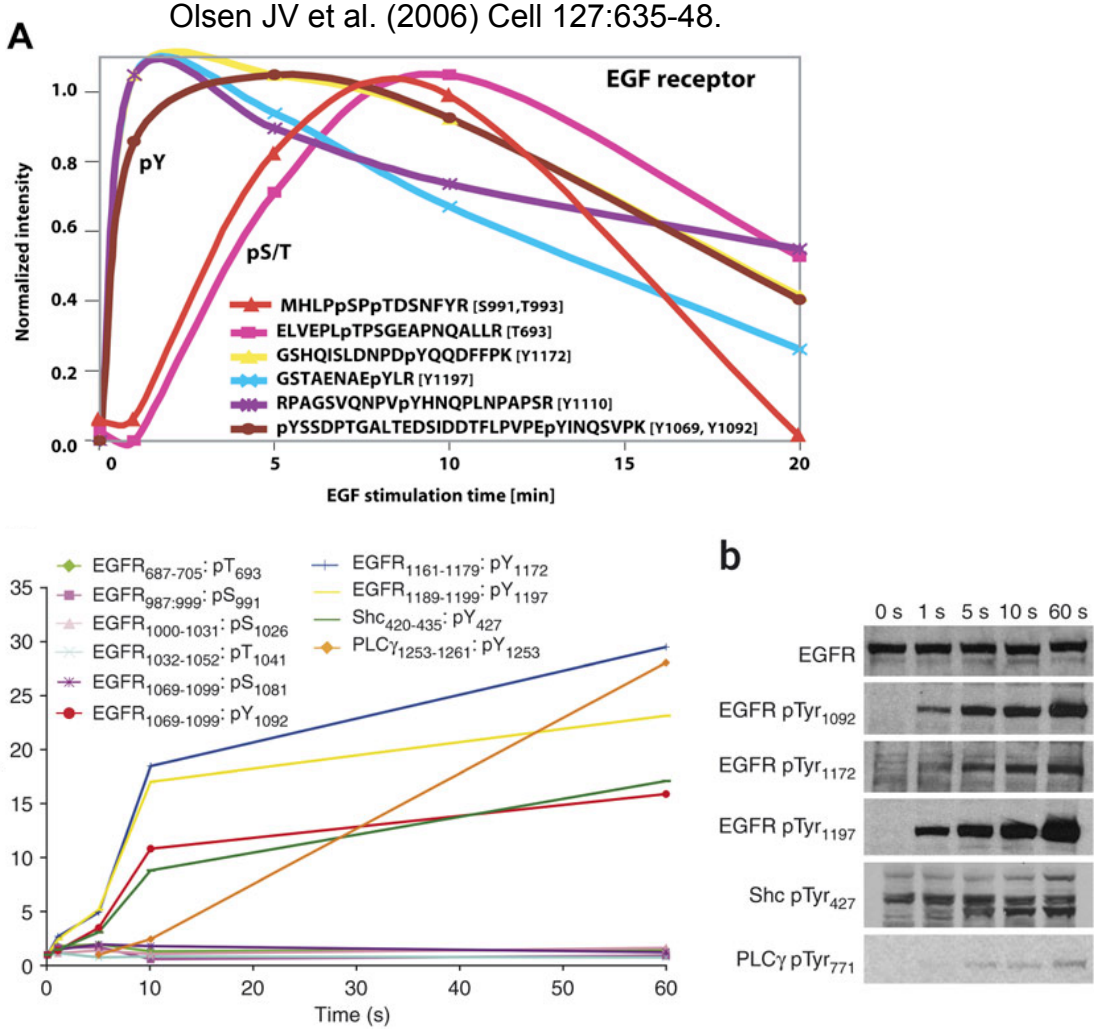
Distinct time courses of phosphorylation for different amino acid residues within the same protein

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



Dengjel J et al. (2007) Nat Biotechnol

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



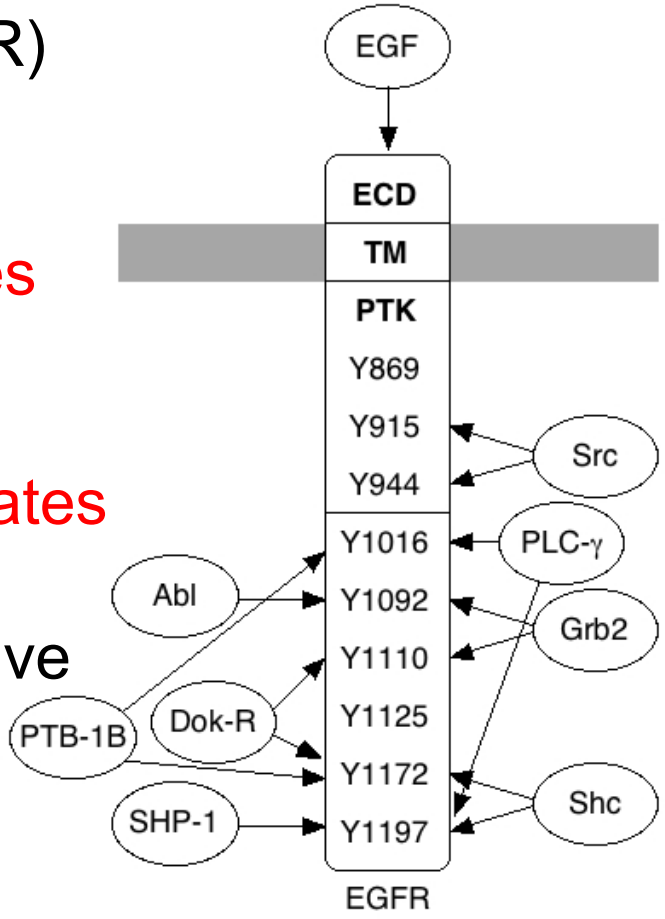
Combinatorial complexity

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



CC is a serious problem for the textbook approach: why

A

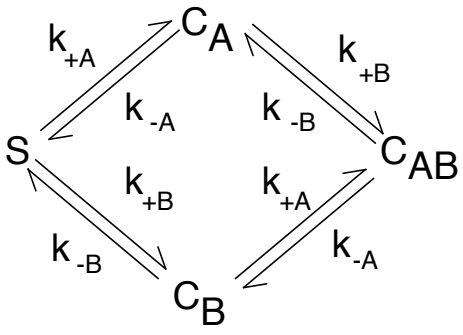
$$A + S \underset{k_{-A}}{\overset{k_{+A}}{\rightleftharpoons}} C_A$$

$$B + S \underset{k_{-B}}{\overset{k_{+B}}{\rightleftharpoons}} C_B$$

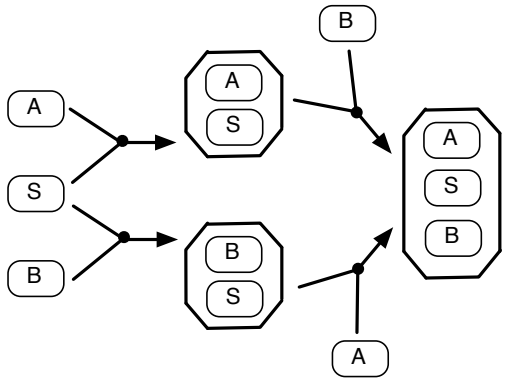
$$A + C_B \underset{k_{-A}}{\overset{k_{+A}}{\rightleftharpoons}} C_{AB}$$

$$B + C_A \underset{k_{-B}}{\overset{k_{+B}}{\rightleftharpoons}} C_{AB}$$

B



C



D

$$\frac{dx_1}{dt} = k_{-A}x_4 + k_{-B}x_5 - k_{+A}x_1x_2 - k_{+B}x_1x_3$$

$$\frac{dx_2}{dt} = k_{-A}x_4 - k_{+A}x_1x_2$$

$$\frac{dx_3}{dt} = k_{-B}x_5 - k_{+B}x_1x_3$$

$$\frac{dx_4}{dt} = k_{+A}x_1x_2 + k_{-B}x_6 - k_{-A}x_4 - k_{+B}x_4x_3$$

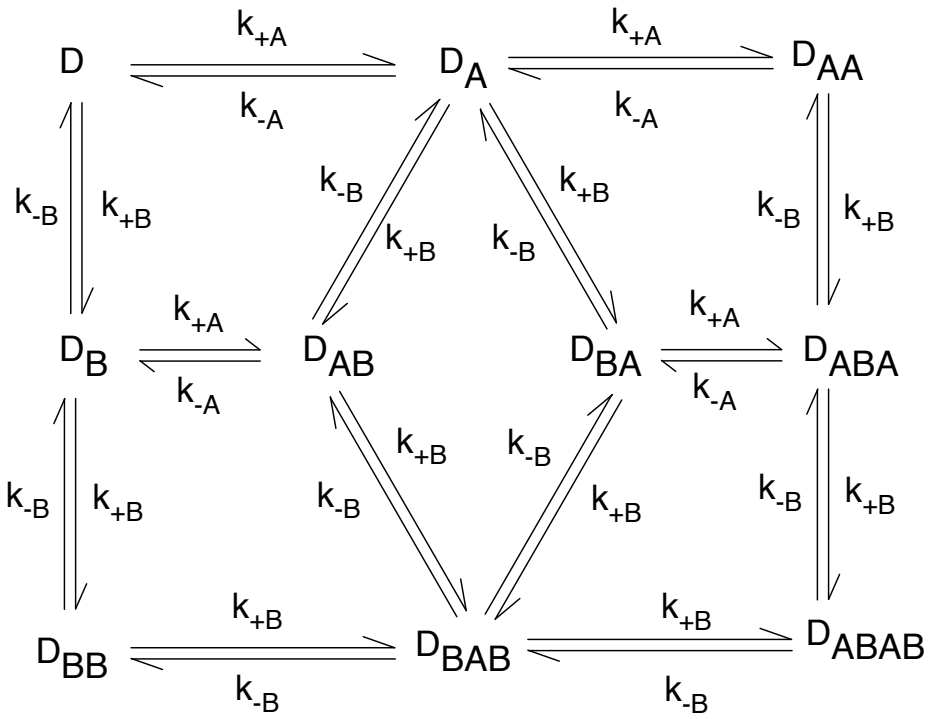
$$\frac{dx_5}{dt} = k_{+B}x_1x_3 + k_{-A}x_6 - k_{-B}x_5 - k_{+A}x_5x_2$$

$$\frac{dx_6}{dt} = k_{+B}x_4x_3 + k_{+A}x_5x_2 - k_{-B}x_6 - k_{-A}x_6$$

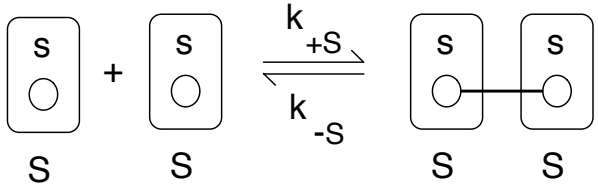
Chylek et al. (2013)

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

A



C



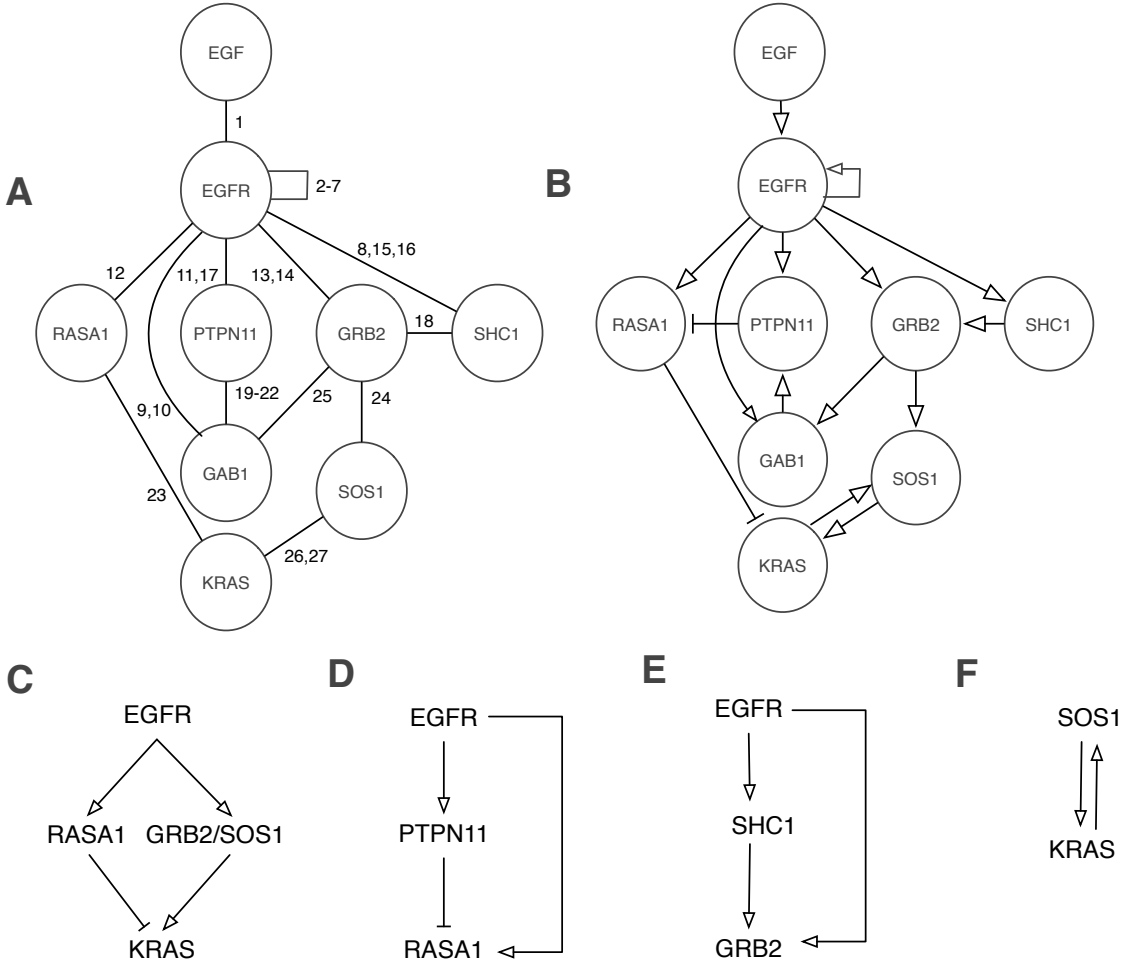
B

$$\frac{dx_1}{dt} = k_{-A}x_4 + k_{-B}x_5 - k_{+A}x_1x_2 - k_{+B}x_1x_3 - k_{+S}x_1x_1 - k_{+S}x_1x_4 - k_{+S}x_1x_5 + 2k_{-S}x_6 - k_{+S}x_1x_7 + k_{-S}x_8 + k_{-S}x_8 + k_{-S}x_{13}$$

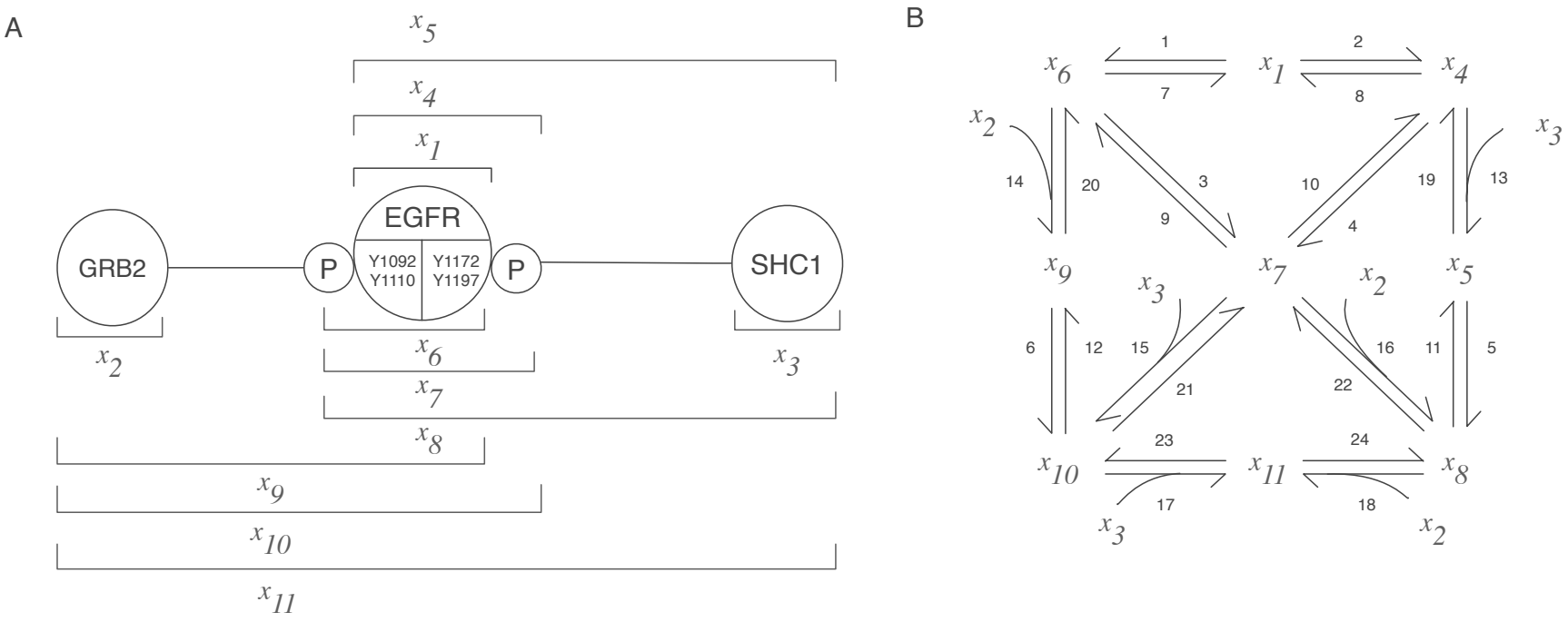
Los Alamos NATIONAL LABORATORY EST.1943

Chylek et al. (2013)

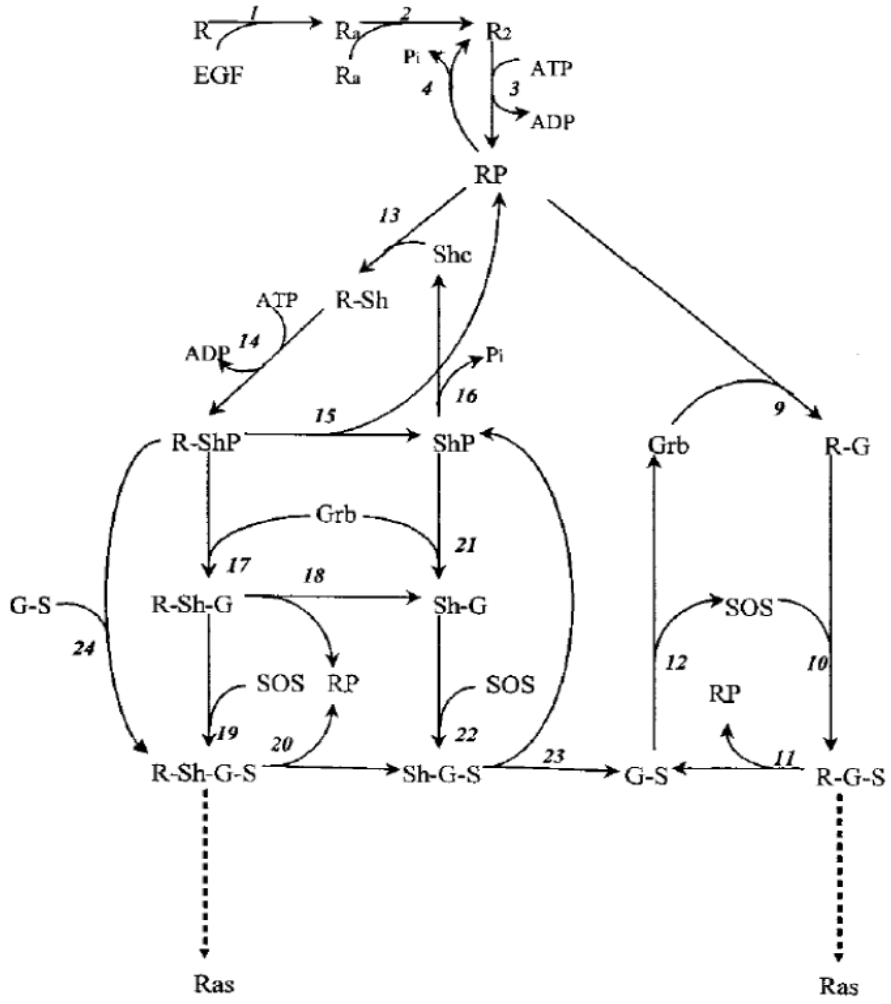
A typical representation of a cell signaling network



The edges of a typical network graph hide considerable complexity



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 numerous published models for EGFR signaling

Kholodenko et al. (1999)

The problem of combinatorial complexity

- **Inside a Chemical Plant**
  - Large numbers of molecules…
  - …of a few types
  - Traditional state variables (concentrations) can be measured.
  - Traditional modeling of chemical kinetics, developed before the existence of molecules was widely accepted (starting in 1865), works.
- **Inside a Cell**
  - Possibly small numbers of molecules…
  - …of myriad possible types (more than a googol, easily)
  - Traditional state variables CANNOT be measured – no data.
  - What do we know? Most studies are focused on elucidating the “rules” of interactions, meaning the parts of biomolecules responsible for interactions and the contextual constraints on interactions. Can we use this information more directly?

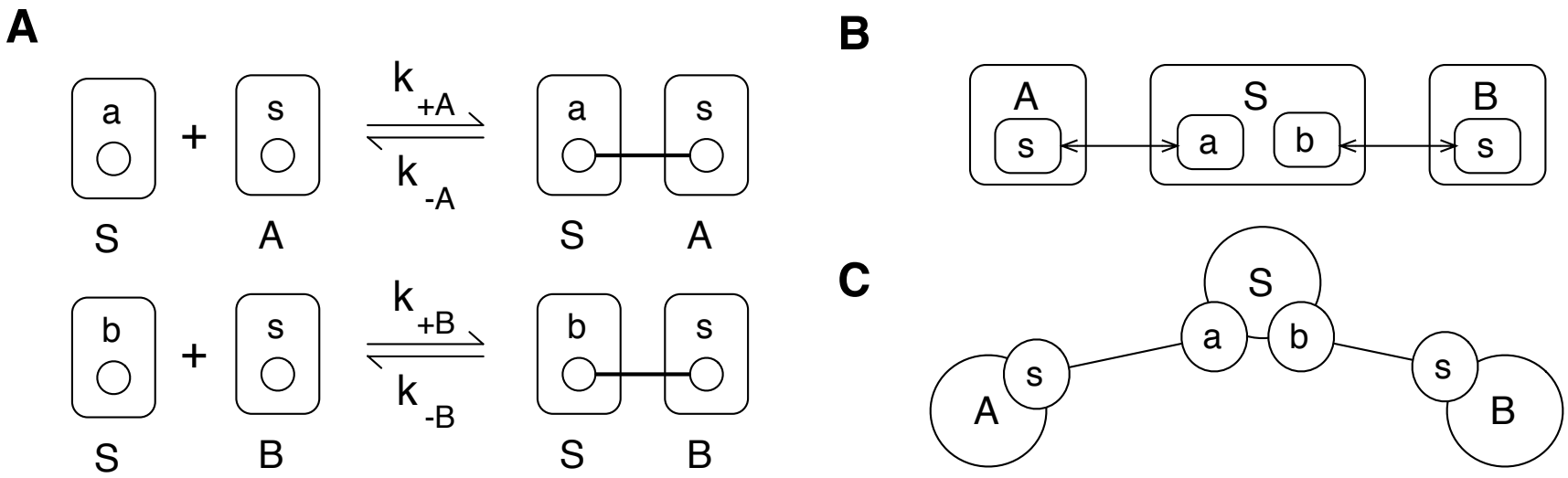
Outline

1. The motivation for modeling, and rule-based modeling in particular
2. **Basic concepts of rule-based modeling**
3. Indirect and direct methods for simulating a model

Rule-based modeling: basic concepts

Use graphs represent molecules and their component parts and "internal states"

Formalize interactions as graph-rewriting rules



A “language” is used to encode graphs and and graph-rewriting rules in plain text

Listing 1

```
begin molecule types
S(s,a,b)
A(s)
B(s)
end molecule types
begin seed species
S(s,a,b) S_init
A(s) A_init
B(s) B_init
end seed species
begin reaction rules
S(a) + A(s) <-> S(a!1).A(s!1) kpa,kma
S(b) + B(s) <-> S(b!1).B(s!1) kpb,kmb
S(s) + S(s) <-> S(s!1).S(s!1) kps,kms
end reaction rule
```

BNGL

Listing 2

```
%agent: S(s,a,b)
%agent: A(s)
%agent: B(s)
%init: 1e5 * S(s,a,b)
%init: 1e5 * A(s)
%init: 1e5 * B(s)
S(a),A(s) ->S(a!1),A(s!1) @0.1
S(a!1),A(s!1) ->S(a),A(s) @0.1
S(b),B(s) ->S(B!1),B(s!1) @0.1
S(B!1),B(s!1) ->S(b),B(s) @0.1
S(S),S(s) ->S(s!1),S(s!1) @0.1
S(s!1),S(s!1) ->S(S),S(s) @0.1
```

Kappa (almost the same as BNGL)

PySB (Lopez et al., 2013): an embedded language

A

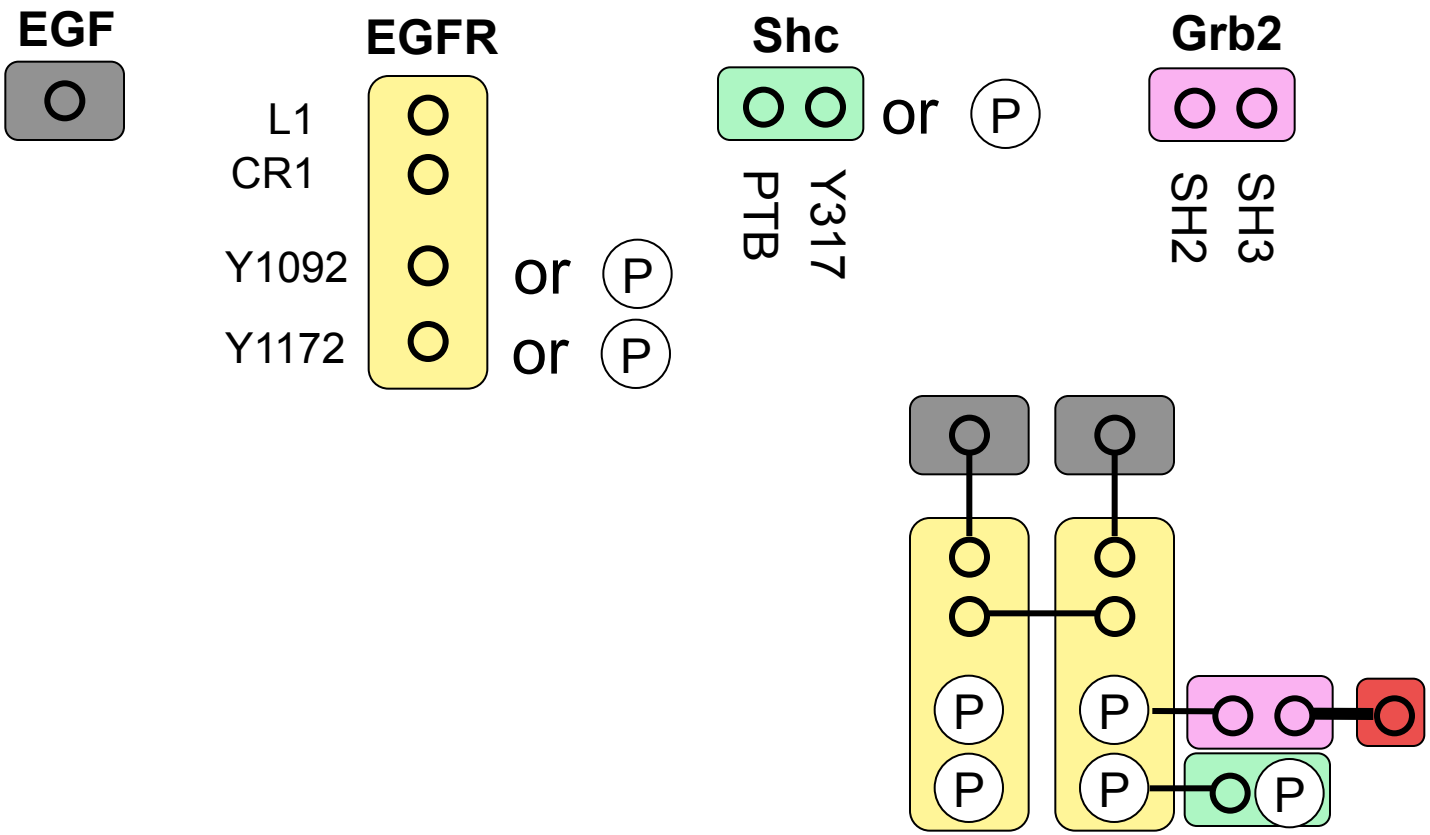
```
from pysb import *

Model()

# Declare molecule types
Monomer('IGF1', ['ds', 'hs'])
tyrosines = ['Y973', 'Y980', 'Y1161', 'Y1166', 'Y1280', 'Y1281', 'Y1346']
Monomer('IGF1R', ['S1', 'S2']+tyrosines, {i:['O', 'P'] for i in tyrosines})

# Create a dictionary of rate constants
kp = {}
for tyr in tyrosines:
    kp[tyr]=Parameter('kp'+tyr, 1.0) # all rate constants are set to 1

# Define phosphorylation rules for IGF1R tyrosines
for tyr in tyrosines:
    Rule("phos1"+tyr, IGF1(ds=1, hs=2) % IGF1R(S1=1) % IGF1R({'S2':2, tyr:'O'}) >> \
         IGF1(ds=1, hs=2) % IGF1R(S1=1) % IGF1R({'S2':2, tyr:'P'}), kp[tyr])
    Rule("phos2"+tyr, IGF1(ds=1, hs=2) % IGF1R({'S1':1, tyr:'O'}) % IGF1R(S2=2) >> \
         IGF1(ds=1, hs=2) % IGF1R({'S1':1, tyr:'P'}) % IGF1R(S2=2), kp[tyr])
```

B

```
# Rule 1
IGF1(ds!1,hs!2).IGF1R(S1!1).IGF1R(S2!2,Y973~0)->\
IGF1(ds!1,hs!2).IGF1R(S1!1).IGF1R(S2!2,Y973~P) kpY973
# Rule 2
IGF1(ds!1,hs!2).IGF1R(S1!1,Y973~0).IGF1R(S2!2)->\
IGF1(ds!1,hs!2).IGF1R(S1!1,Y973~P).IGF1R(S2!2) kpY973
...
# Rule 13
IGF1(ds!1,hs!2).IGF1R(S1!1).IGF1R(S2!2,Y1346~0)->\
IGF1(ds!1,hs!2).IGF1R(S1!1).IGF1R(S2!2,Y1346~P) kpY1346
# Rule 14
IGF1(ds!1,hs!2).IGF1R(S1!1,Y1346~0).IGF1R(S2!2)->\
IGF1(ds!1,hs!2).IGF1R(S1!1,Y1346~P).IGF1R(S2!2) kpY1346
```

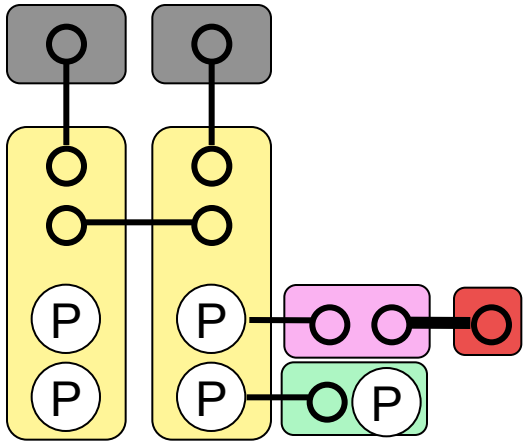
Code example from Chylek et al.

Representing molecules



Blinov ML et al. (2006) *BioSystems*

Representing complexes: connected sets of graphs for molecules

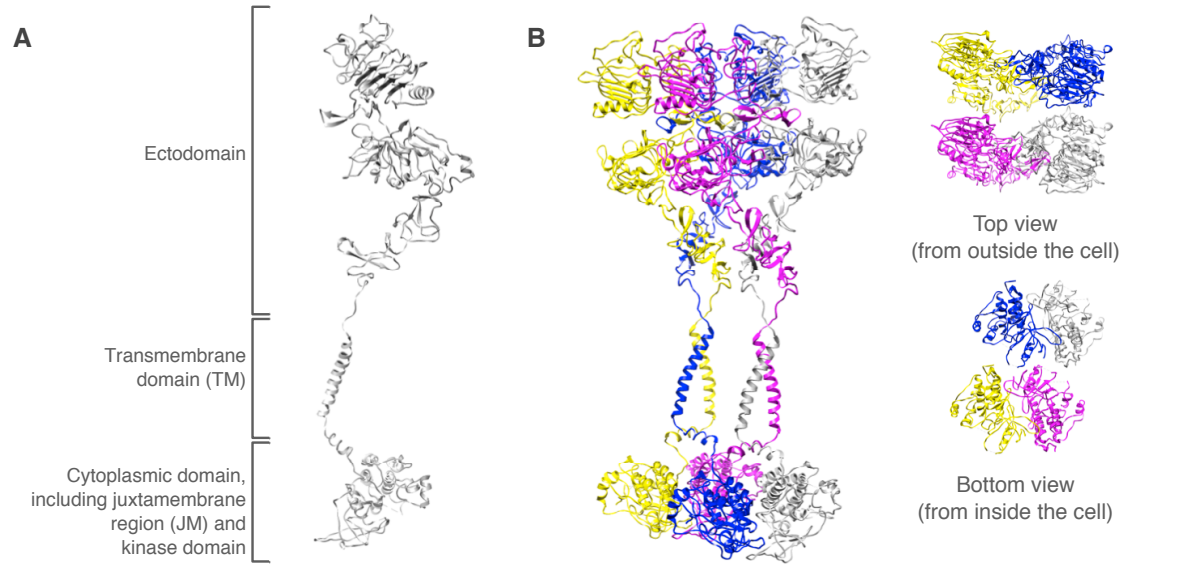


No need to introduce a unique name (e.g., X_{123} or ShP-RP-G-Sos) for each chemical species, as in conventional modeling

Edges represent bonds between components

Bonds may be intra- or intermolecular

The granularity of molecule representation: intermediate between atomic and traditional state variable representations



D `EGF(rec)`

E `EGFR(lig,back,cd~c~o,Y~u~p)`

F

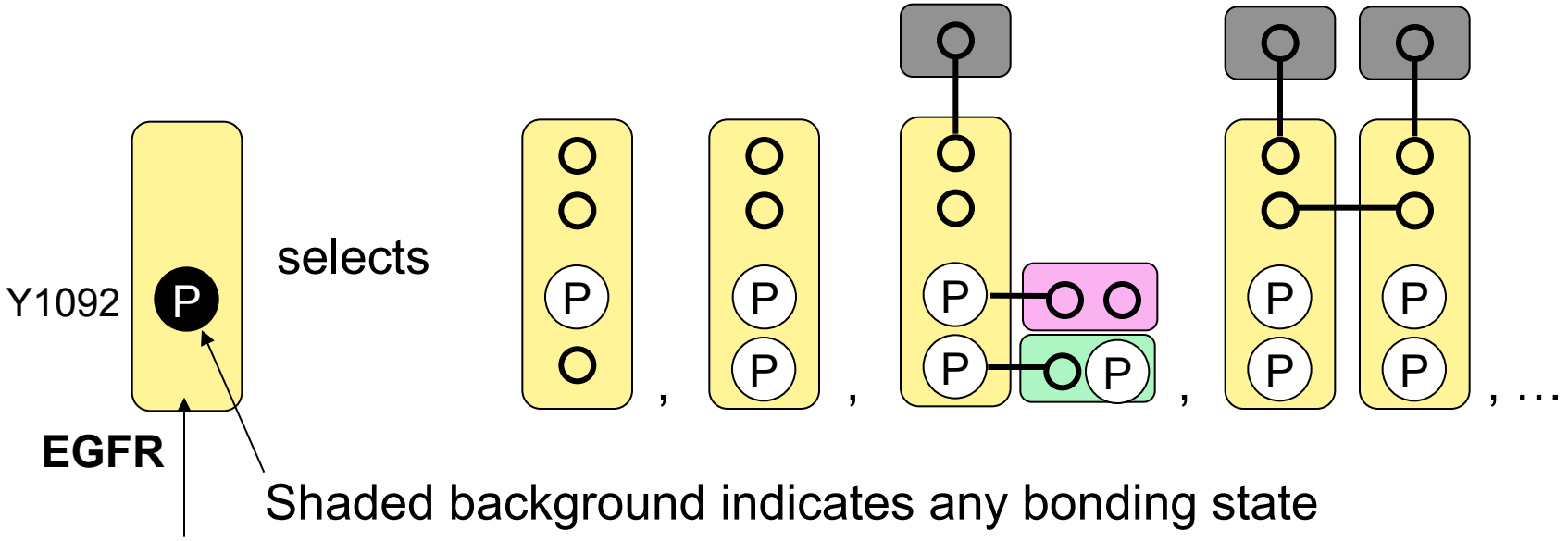
```
EGF(rec!5).EGFR(lig!5,back!1,cd~o!4,Y~p).\
EGF(rec!6).EGFR(lig!6,back!1,cd~o!3,Y~p).\
EGF(rec!7).EGFR(lig!7,back!2,cd~o!3,Y~p).\
EGF(rec!8).EGFR(lig!8,back!2,cd~o!4,Y~p)
```

Kozer et al. (2013) Mol BioSyst

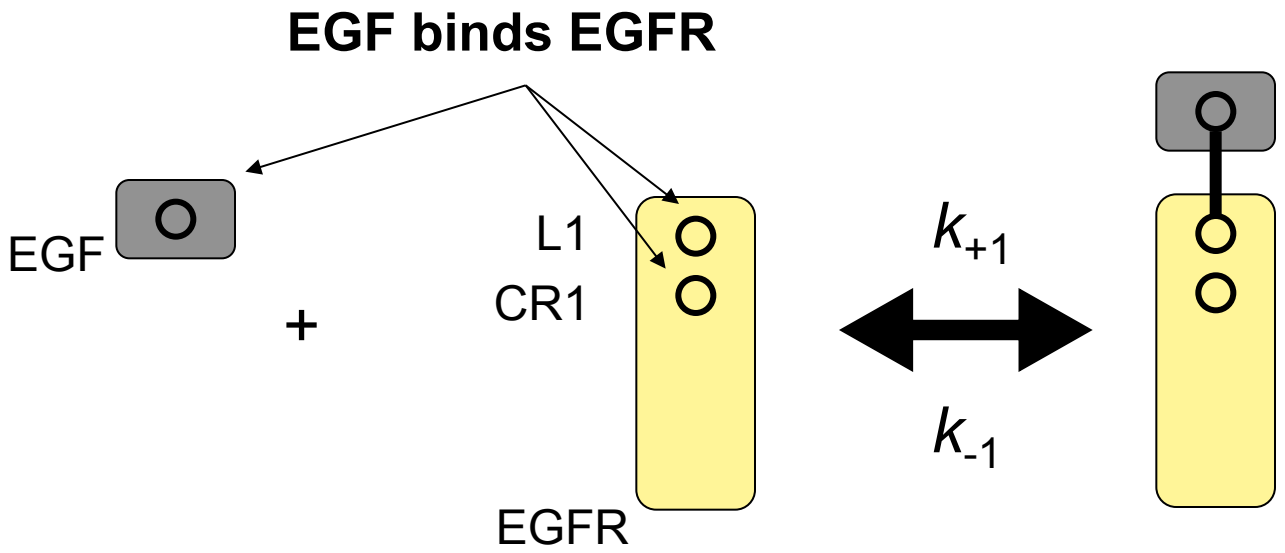
L.A. Chylek & C.-S. Tung

Representing interactions: patterns (subgraphs) define sets of chemical species with common features

A pattern that matches EGFR phosphorylated at Y1092



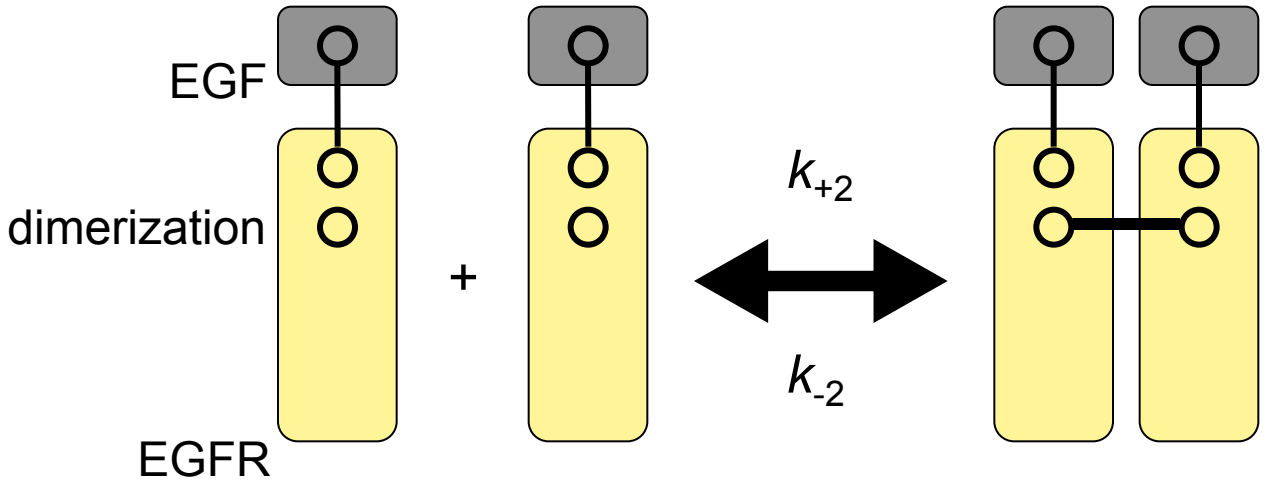
Representing interactions: rules are composed of patterns




```
begin reaction rules

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

end reaction rules
```

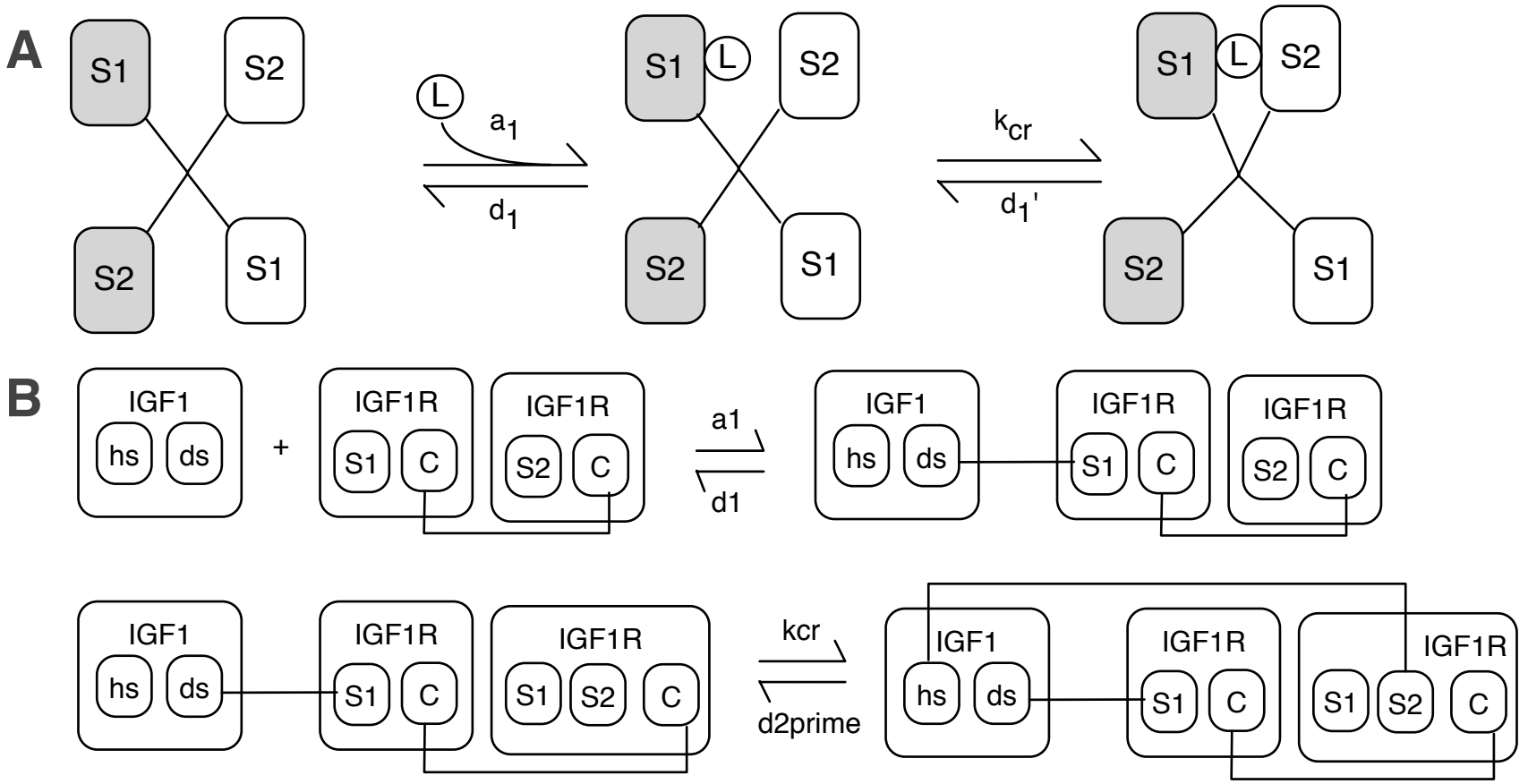
Dimerization rule

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.

Another example of how to represent interactions (part 1)



Kiselyov et al. (2009) Mol Syst Biol; Chylek et al. (submitted)

Another example of how to represent interactions (part 2)

C

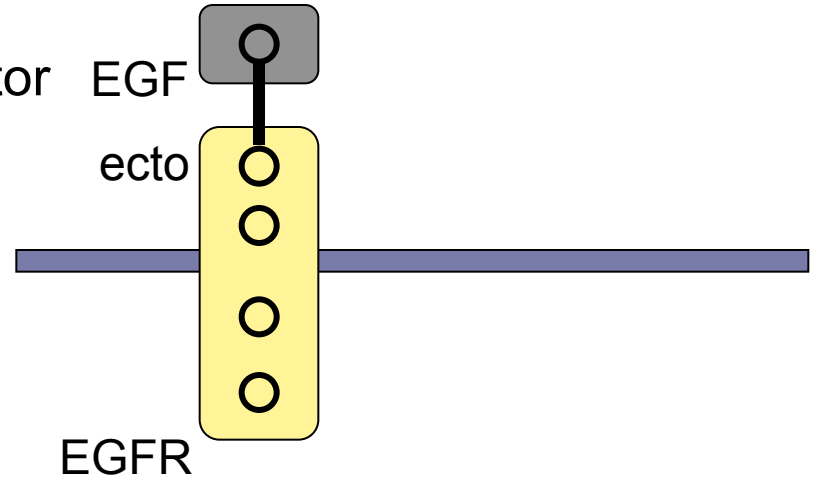
```
begin molecule types
IGF1(ds,hs)
IGF1R(S1,S2,C) # monomer
end molecule types

begin seed species
IGF1R(S1,S2,C!0).IGF1R(S1,S2,C!0) IGF1R_total # dimer
IGF1(ds,hs) IGF1_total
end  seed species

begin observables
Molecules Crosslink IGF1R(S1!+).IGF1R(S2!+)
end observables

begin reaction rules
# Rule 1
IGF1(ds,hs)+IGF1R(S1,C!0).IGF1R(S2,C!0)<->\
IGF1(ds!1,hs).IGF1R(S1!1,C!0).IGF1R(S2,C!0) a1,d1
# Rule 2
IGF1(ds,hs)+IGF1R(S2,C!0).IGF1R(S1,C!0)<->\
IGF1(ds,hs!1).IGF1R(S2!1,C!0).IGF1R(S1,C!0) a2,d2
# Rule 3
IGF1R(S1!1,C!0).IGF1R(S1,S2,C!0).IGF1(ds!1,hs)<->\
IGF1R(S1!1,C!0).IGF1R(S1,S2!2,C!0).IGF1(ds!1,hs!2)
kcr,d2prime
# Rule 4
IGF1R(S1!1,S2,C!0).IGF1R(S1!+,S2,C!0).IGF1(ds!1,hs)<->\
IGF1R(S1!1,S2,C!0).IGF1R(S1!+,S2!2,C!0).IGF1(ds!1,hs!2)
kcr,d2prime
# Rule 5
%x:IGF1R(S2!1,C!0).IGF1R(S1,C!0).IGF1(ds,hs!1)<->\
%x:IGF1R(S2!1,C!0).IGF1R(S1!2,C!0).IGF1(ds!2,hs!1)\
   if(Crosslink(x)>0,0,kcr),d1prime
end reaction rules
```

Early events in EGFR signaling – the rules of a model tell a story (and a story can be translated into a set of rules)

EGF = epidermal growth factor

EGFR = epidermal growth factor receptor

1. EGF binds EGFR

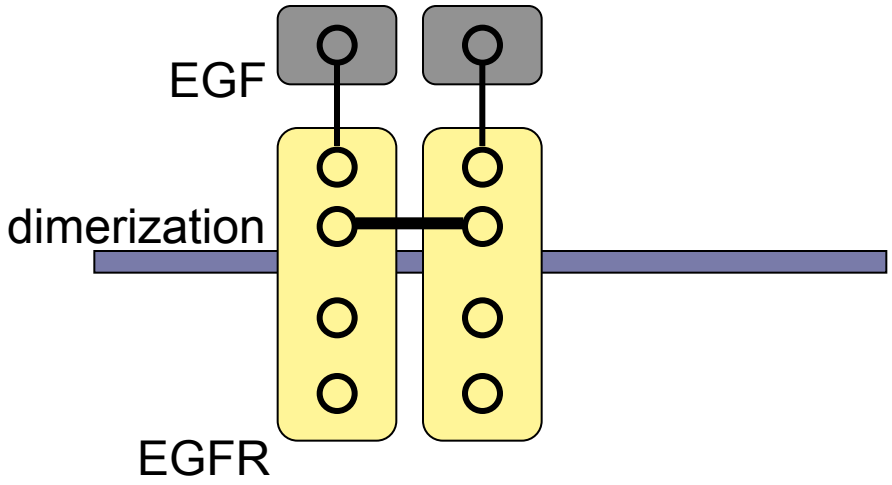


Los Alamos NATIONAL LABORATORY EST.1943

Blinov ML et al. (2006) *BioSystems*; Danos et al. (2007)

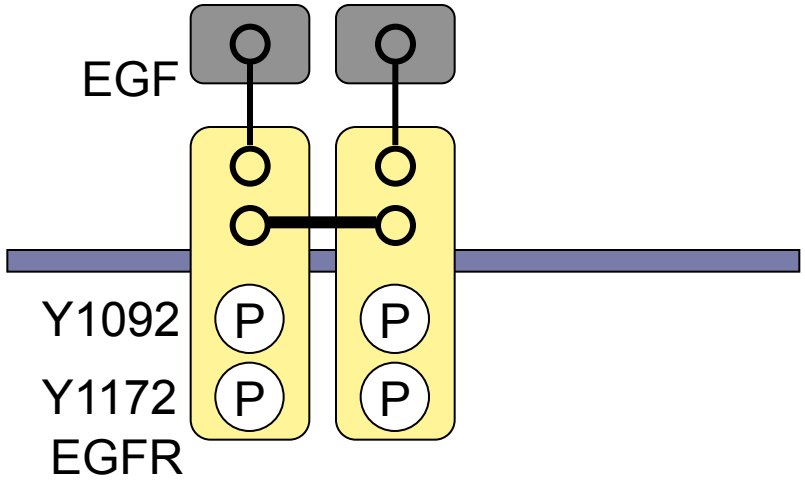
Early events in EGFR signaling

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



Early events in EGFR signaling

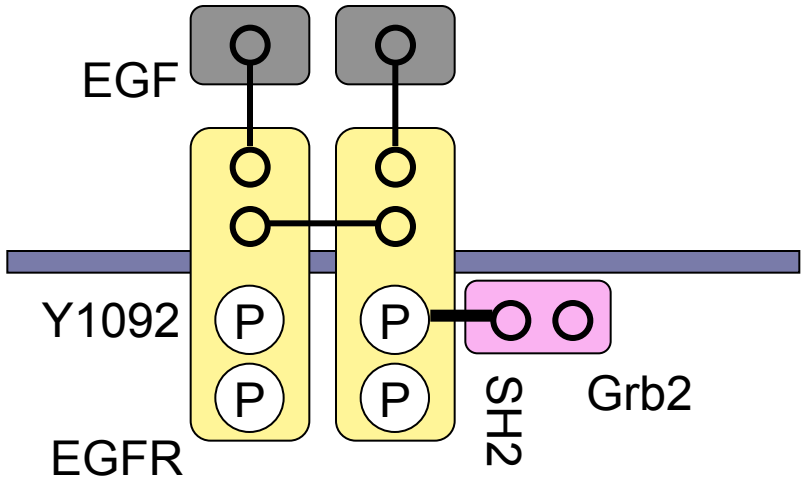
1. EGF binds EGFR
2. EGFR dimerizes
3. **EGFR transphosphorylates a copy of itself**



Early events in EGFR signaling

Grb2 pathway

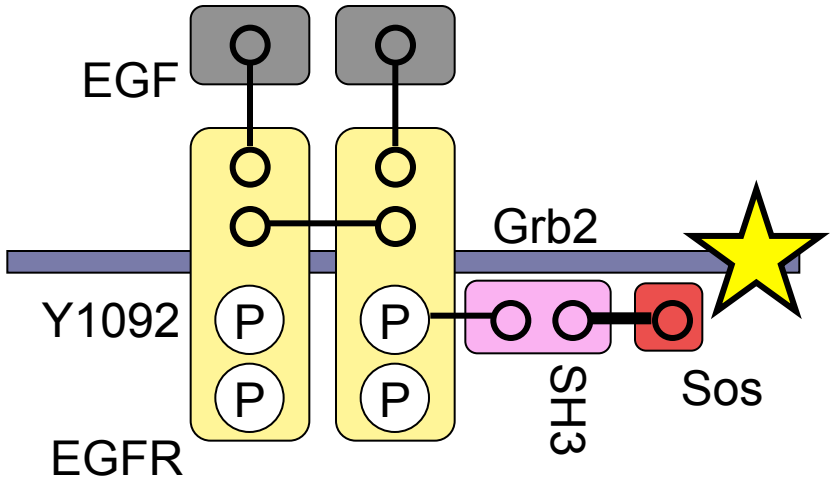
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Grb2 binds phospho-EGFR**



Early events in EGFR signaling

Grb2 pathway

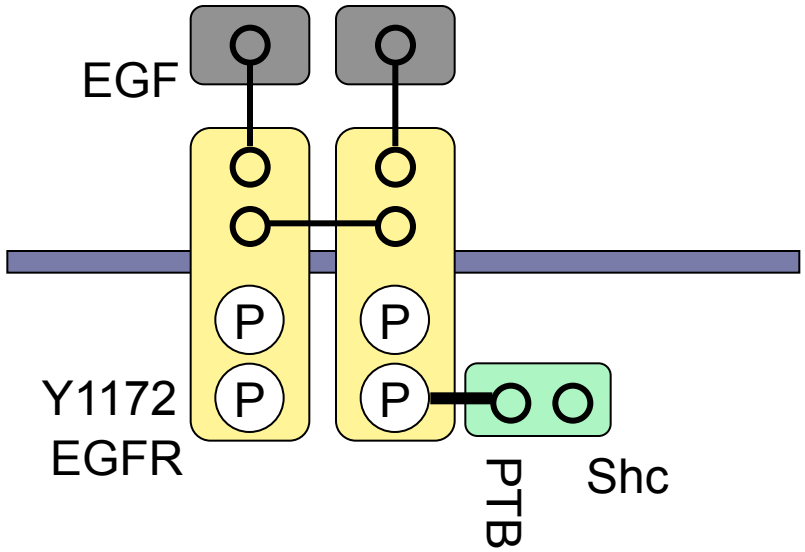
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Grb2 binds phospho-EGFR
5. **Sos binds Grb2 (Activation Path 1)**



Early events in EGFR signaling

Shc pathway

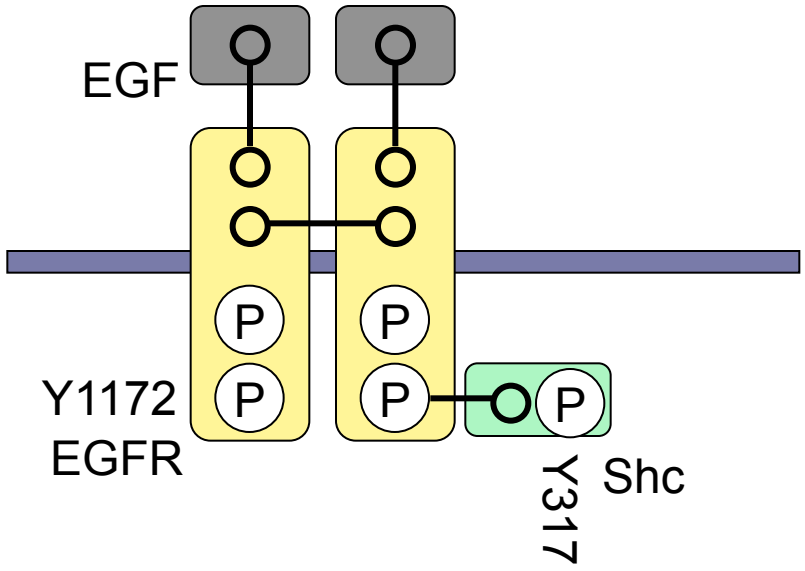
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Shc binds phospho-EGFR**



Early events in EGFR signaling

Shc pathway

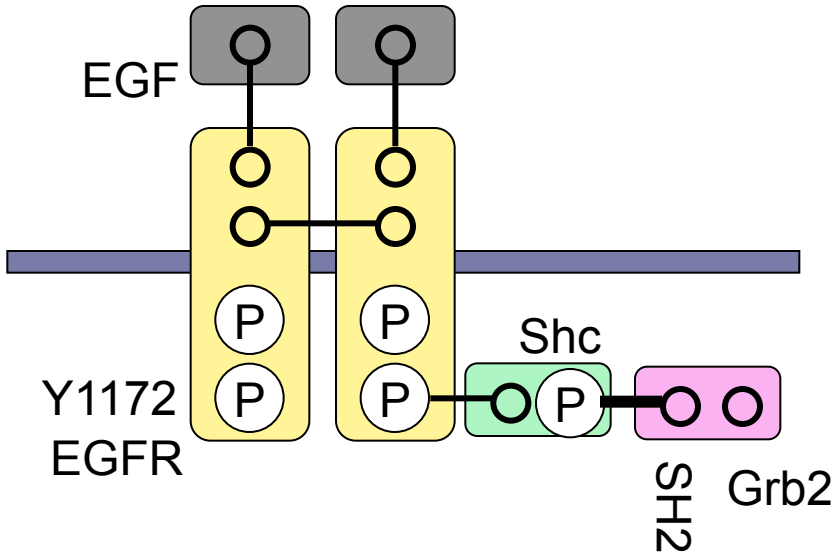
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. **EGFR transphosphorylates Shc**



Early events in EGFR signaling

Shc pathway

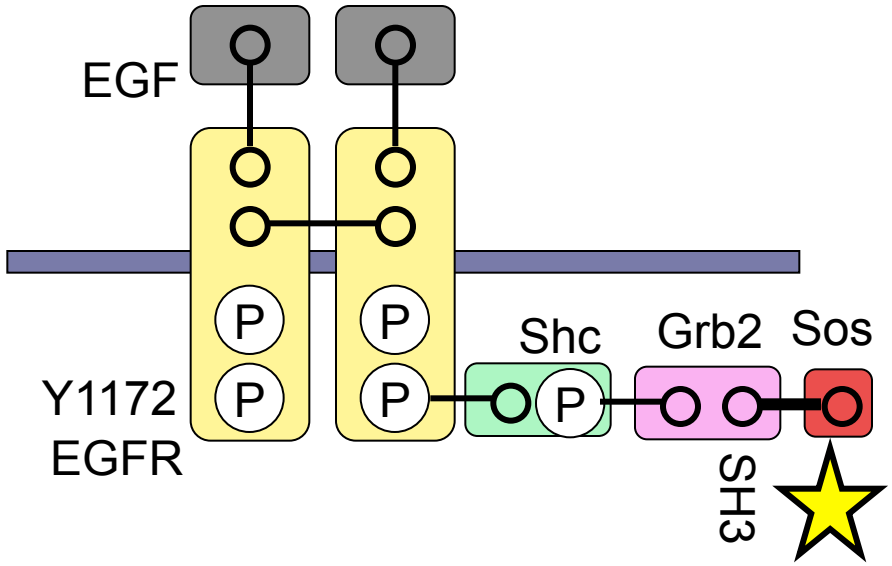
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. **Grb2 binds phospho-Shc**



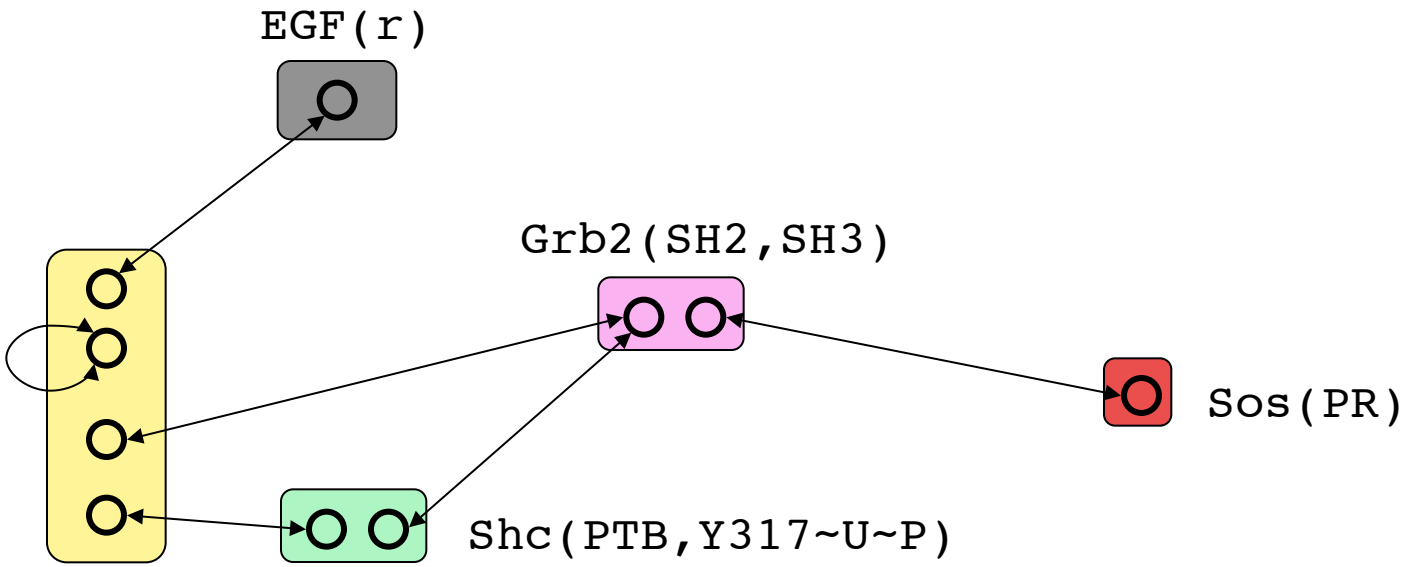
Early events in EGFR signaling

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



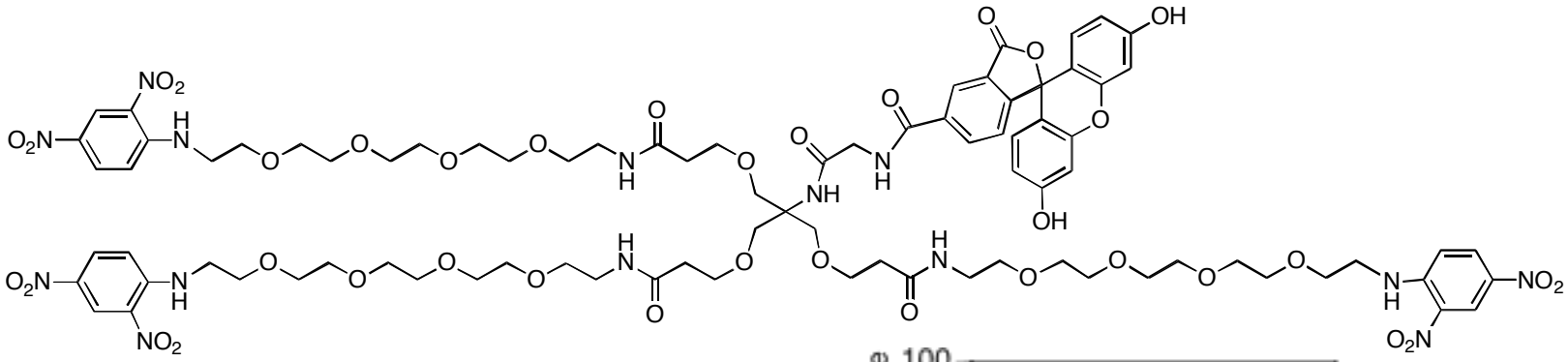
EGFR(l,d,Y1092~U~P,Y1172~U~P)

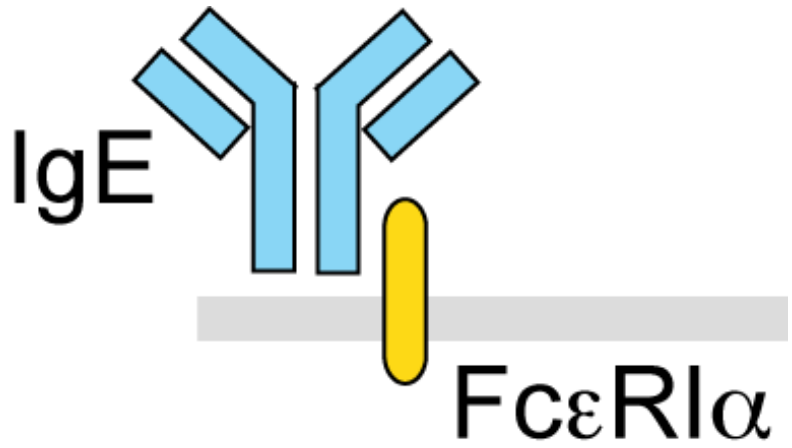
Blinov et al. (2006)

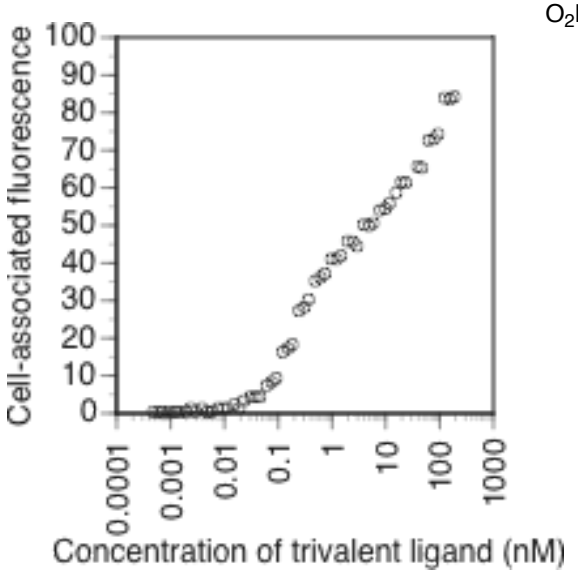
Outline

1. The motivation for modeling, and rule-based modeling in particular
2. Basic concepts of rule-based modeling
3. **Indirect and direct methods for simulating a model**

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



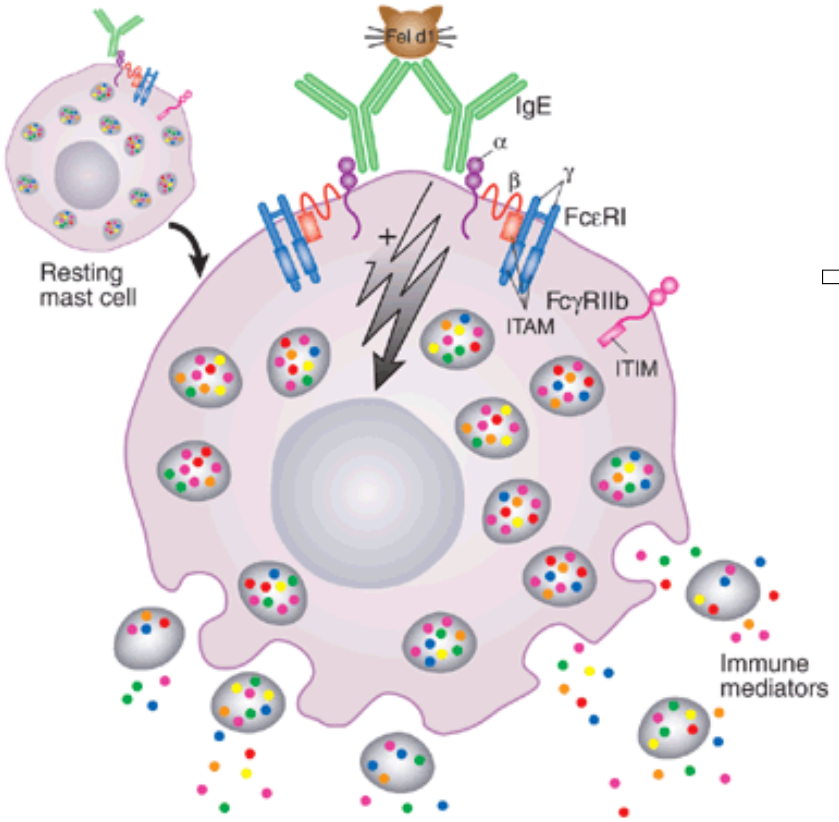


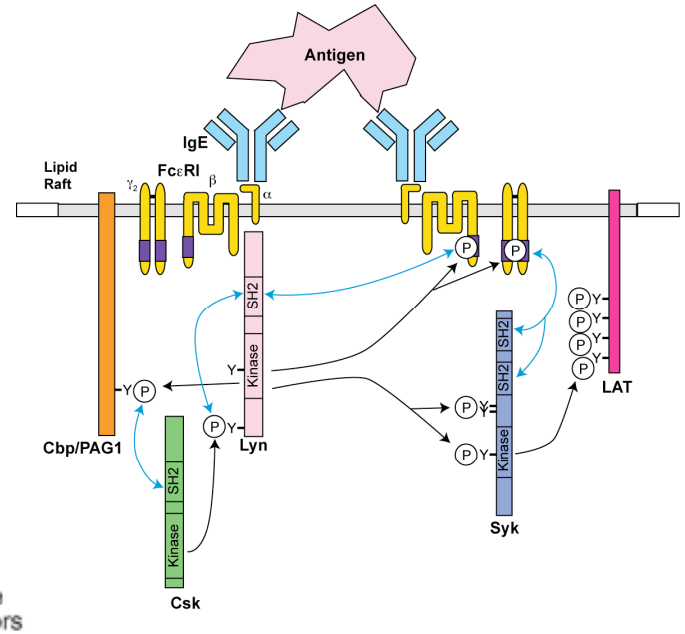


Los Alamos NATIONAL LABORATORY EST.1943

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

Signaling by FcεRI begins with ligand-induced receptor clustering

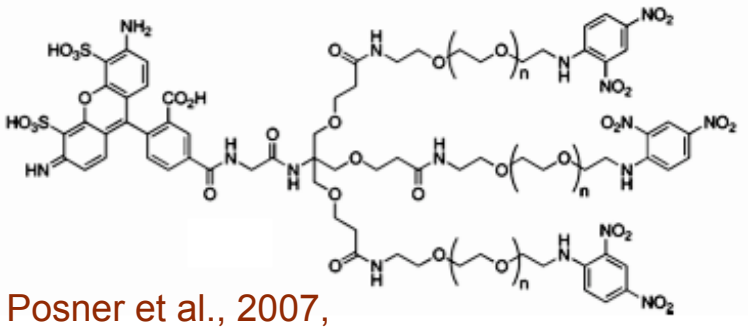




FcεRI-dependent mast cell activation:
Degranulation (release of histamine, etc.)
Cytokine/chemokine release (TNF-α, IL-6, IL-8, etc.)
Arachidonic acid metabolite release (LTC_4, PGD_2, etc.)
↓
Anaphylaxis, allergy (atopic asthma), etc.

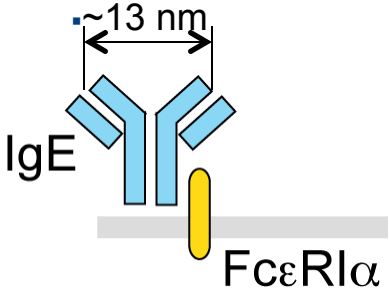
Los Alamos NATIONAL LABORATORY EST.1943

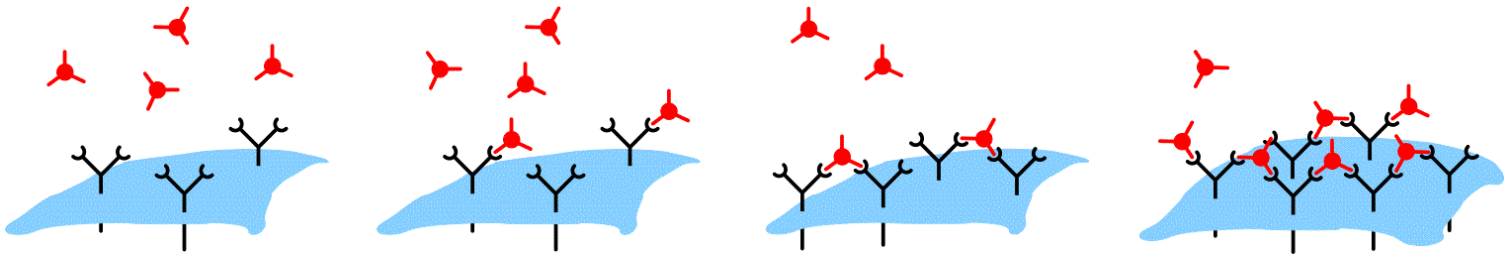
Trivalent ligands



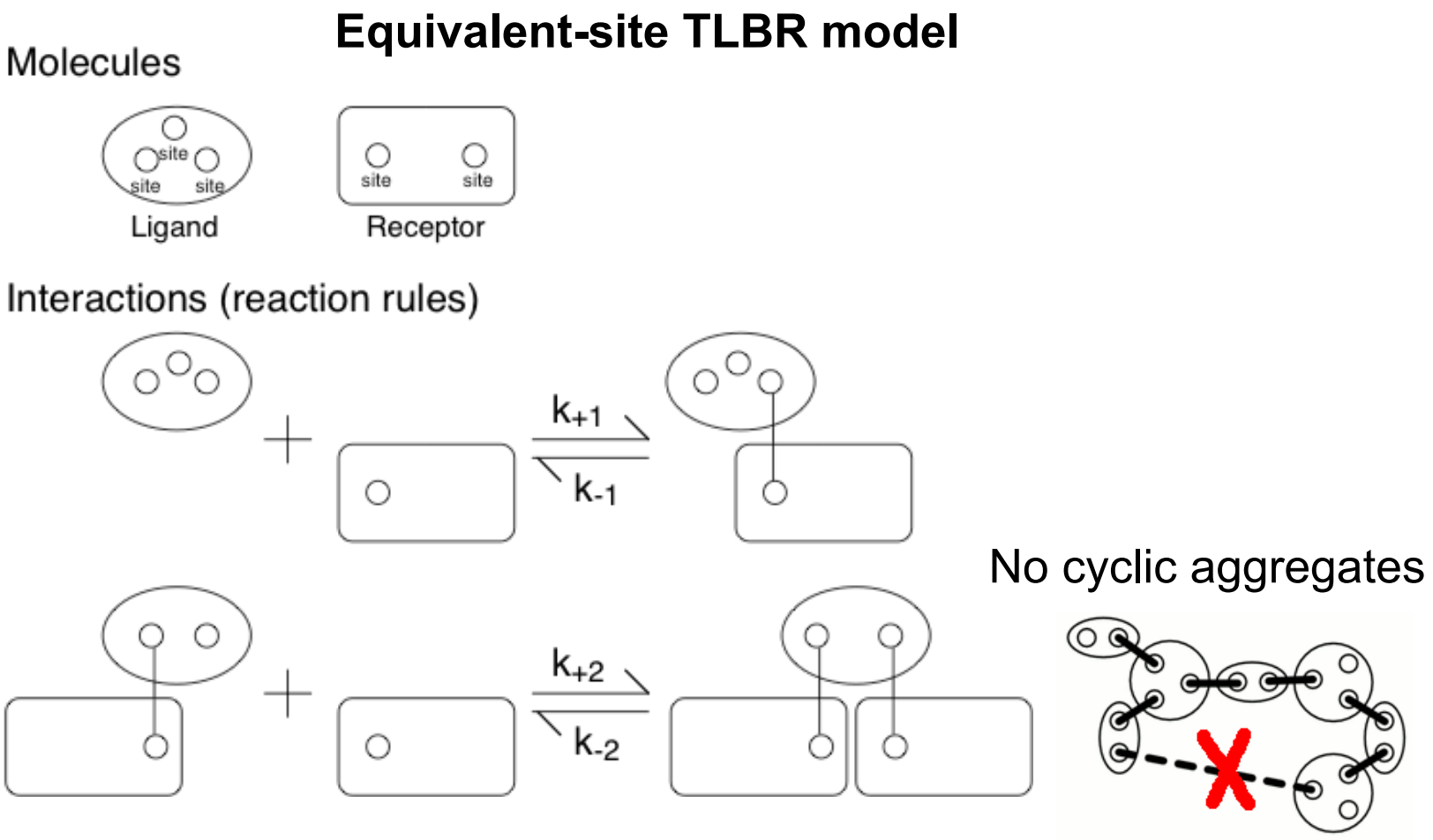
Posner et al., 2007,
Org. Lett, **9**:3551

- Compound 6a

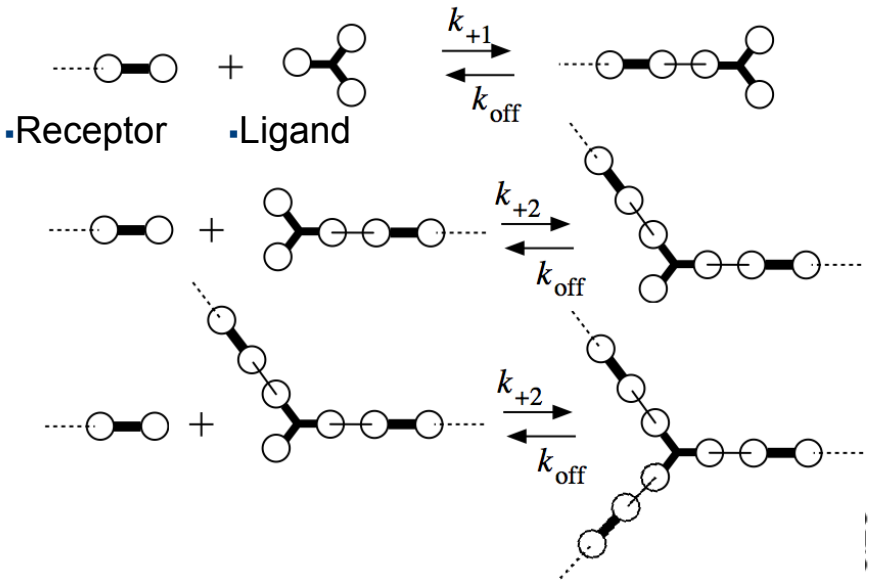




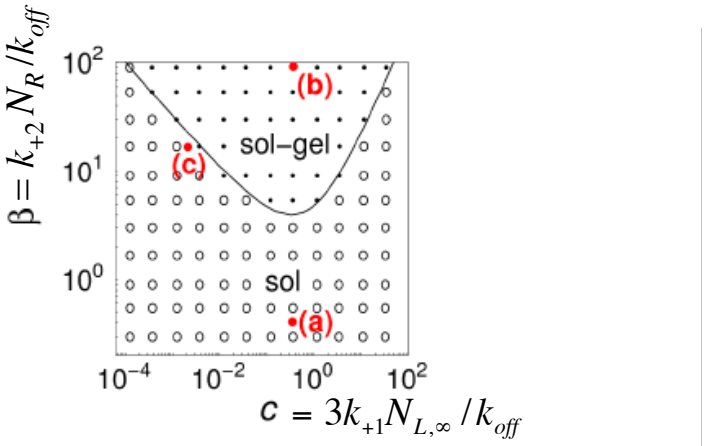
Rule-based model specification corresponding to equilibrium model of Goldstein and Perelson (1984)

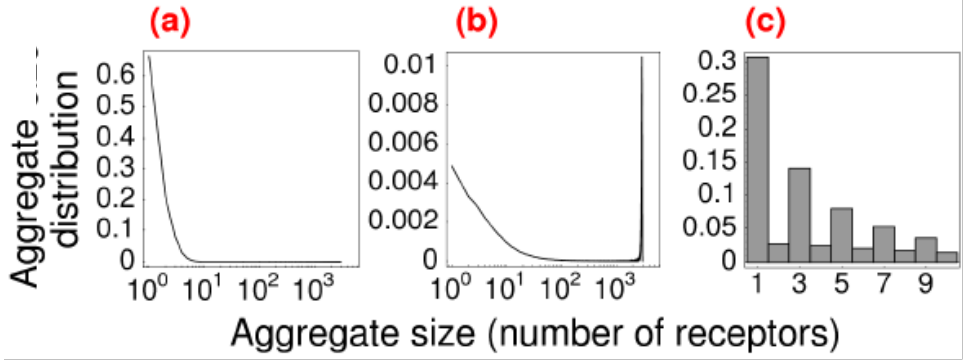


Goldstein-Perelson and TLBR models



Equilibrium properties:



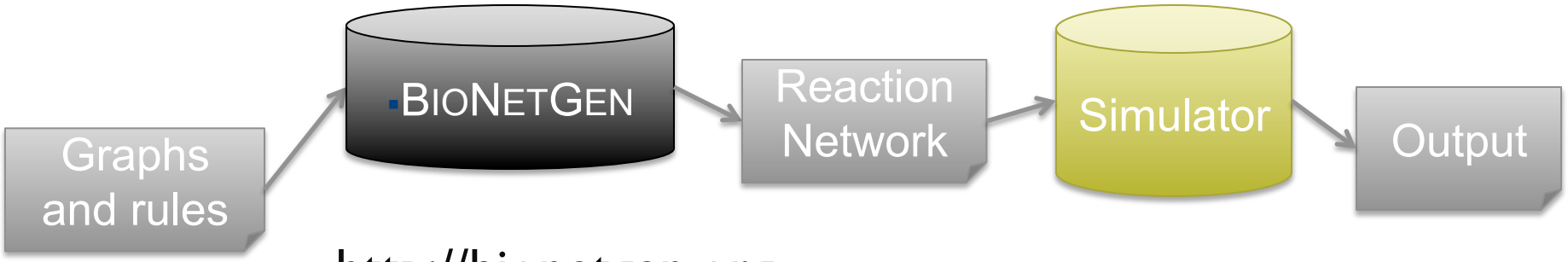


Goldstein and Perelson (1984) *Biophys. J.*, 45:1109

Yang et al. (2008) *Phys. Rev. E,* 78:31910

Protocol for "generate-first" simulation (an indirect method)

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

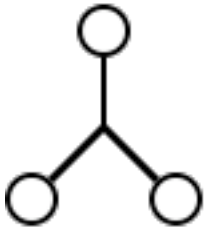


http://bionetgen.org

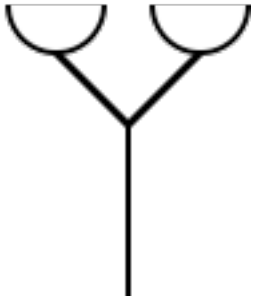
Los Alamos NATIONAL LABORATORY EST.1943

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

“Generate-first” method starts with seed species

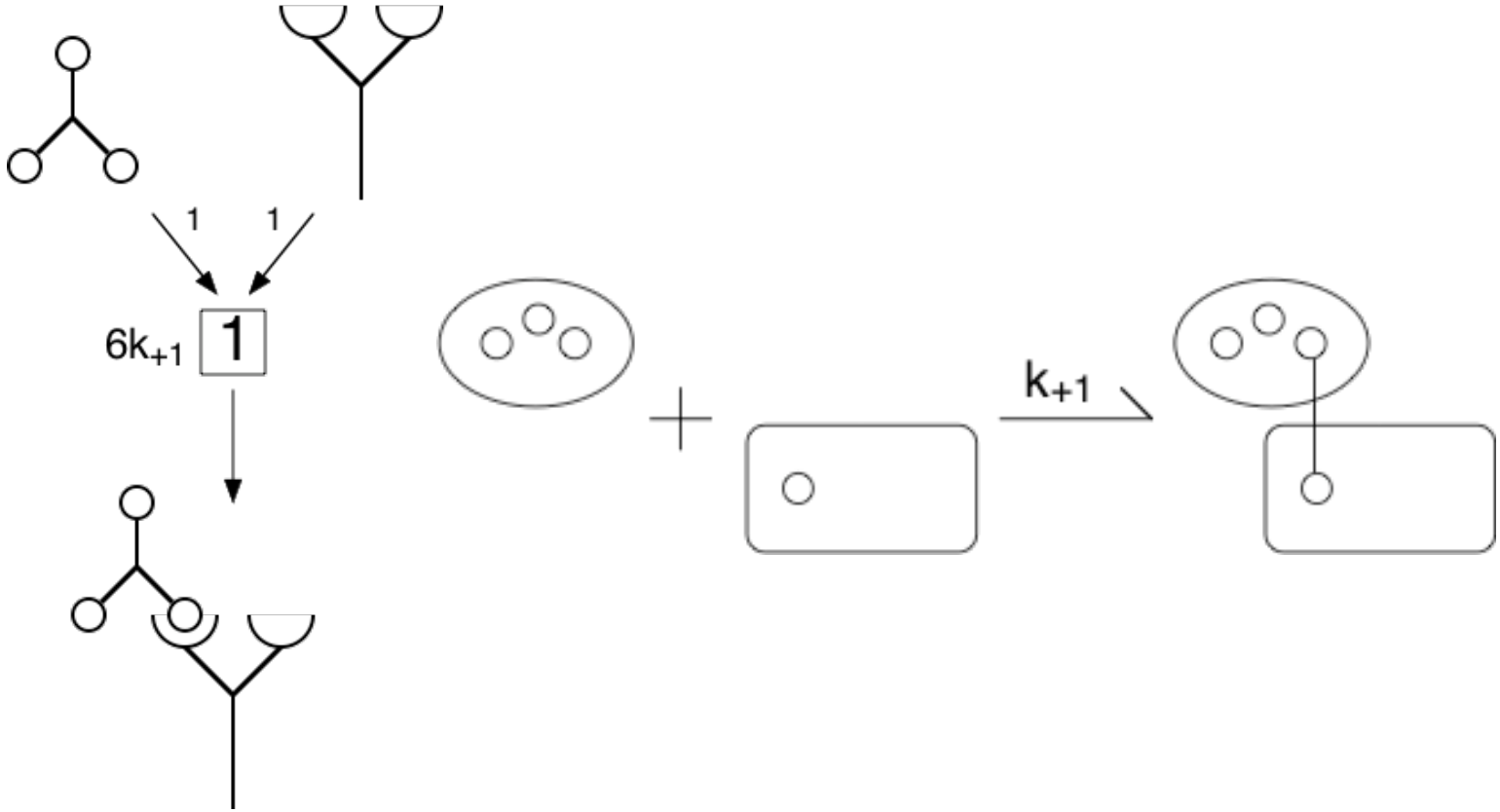


Ligand

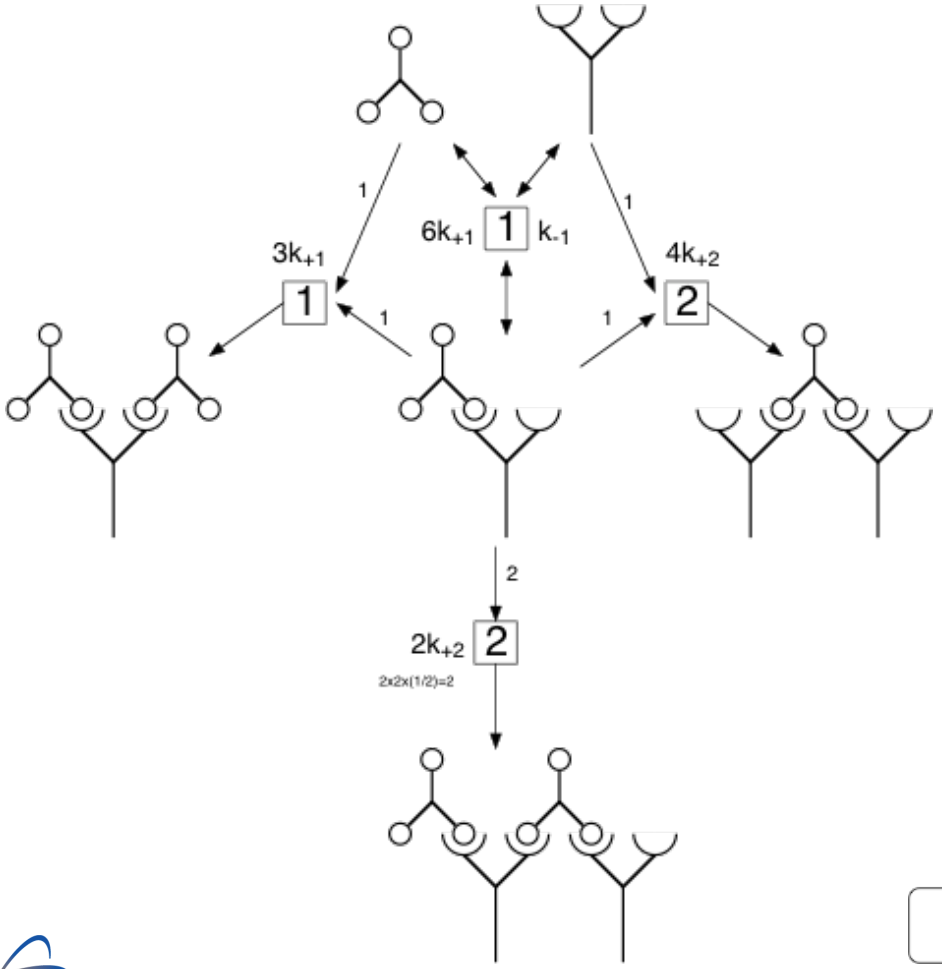


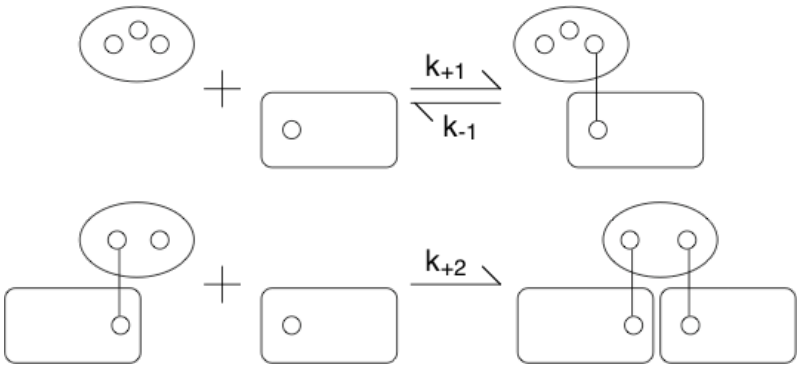
Receptor

After first round of rule application

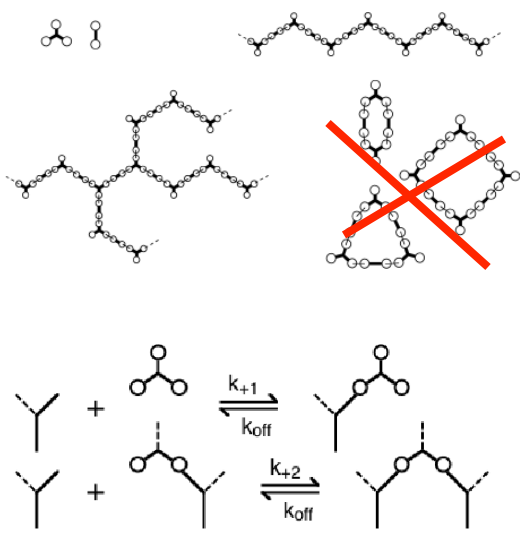


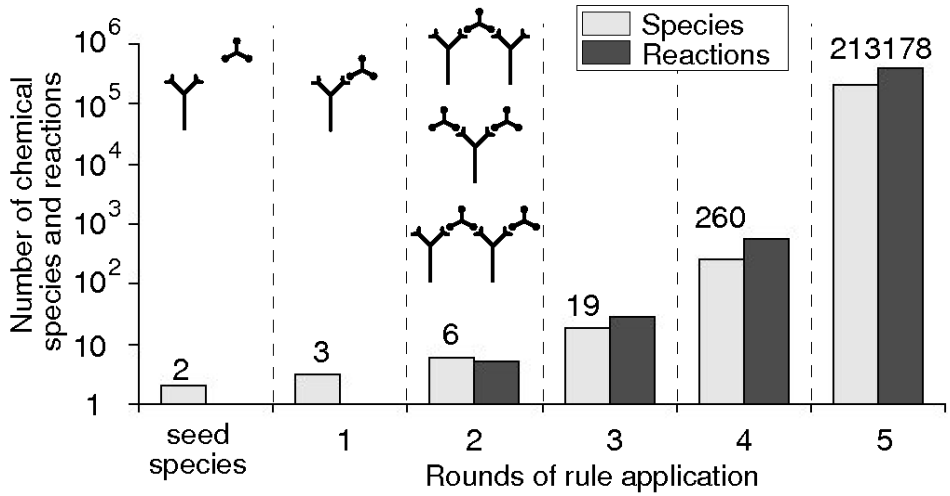
After the second round of rule application



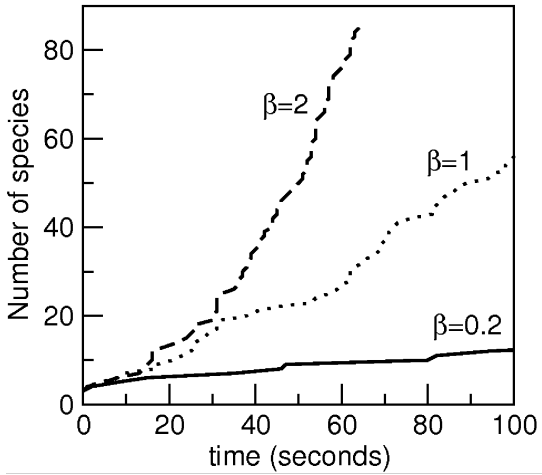


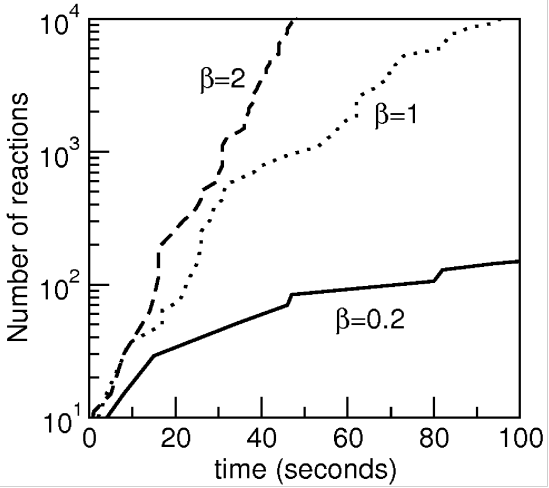
Rule-derived network can be too large to simulate using an indirect method





On The Fly method





Direct methods (particle-based KMC algorithms)

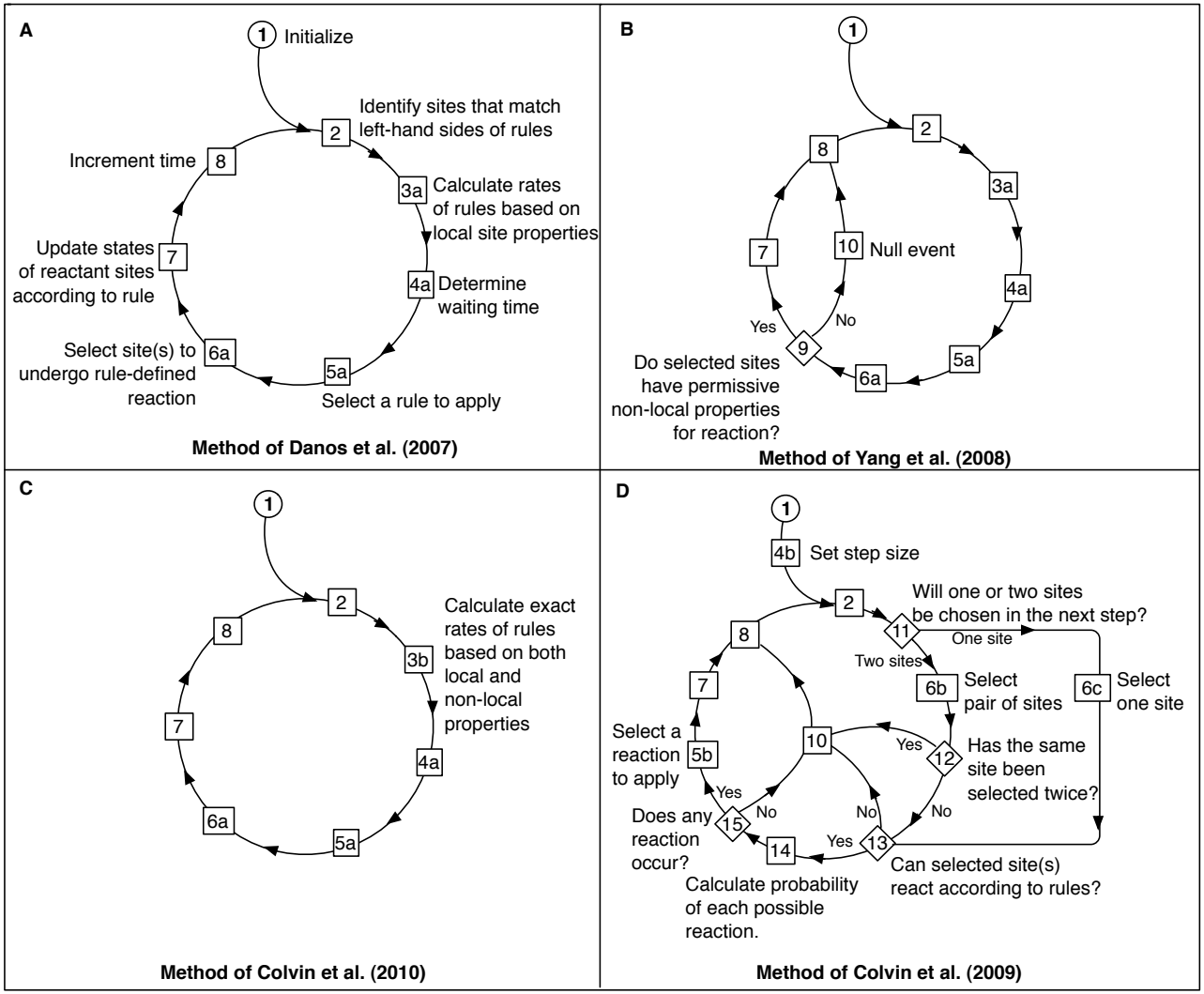
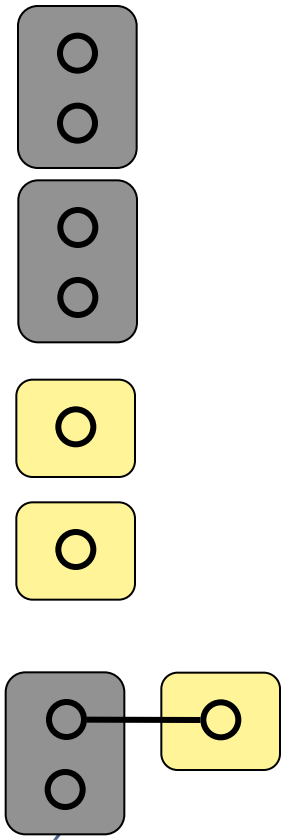


Diagram from Chylek et al. (2013)

Network-free (direct) simulation

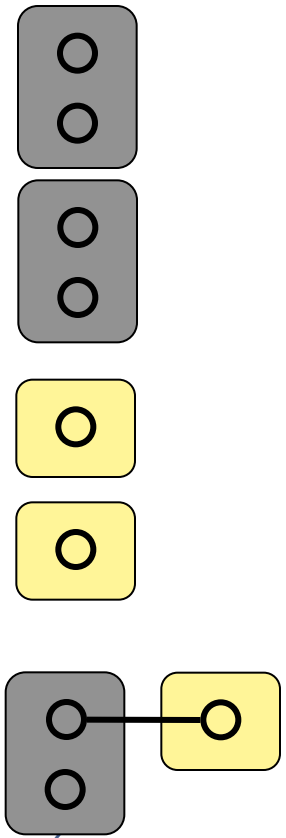
List of molecules and sites (particles) in a simulation "box"



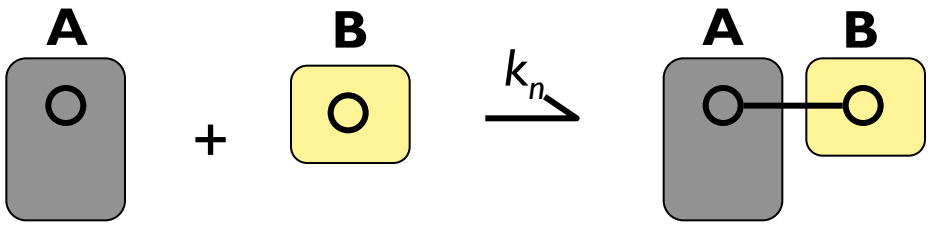
Los Alamos
NATIONAL LABORATORY
EST.1943

Network-free (direct) simulation

Particles



Rules are used as event generators – a rate is calculated for each rule



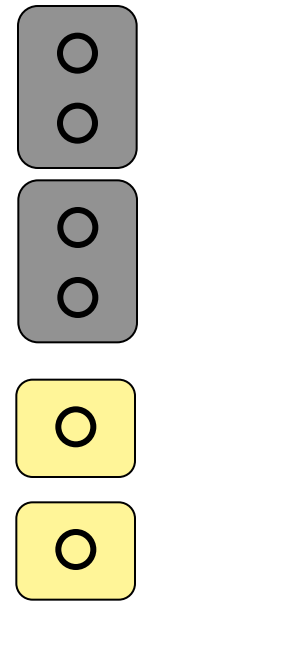
Rule *n*

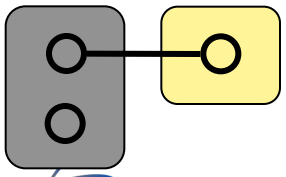
Rule rate = $a_n = k_n\,[A][B]$

Los Alamos NATIONAL LABORATORY EST.1943

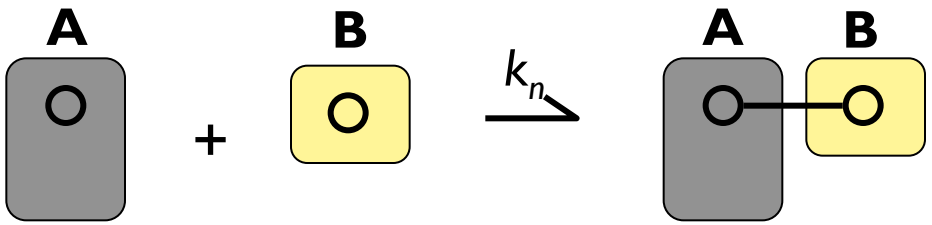
Network-free (direct) simulation

Particles





Let's assume that Reaction *n* is chosen to fire (using a procedure similar to that of the Gillespie algorithm)

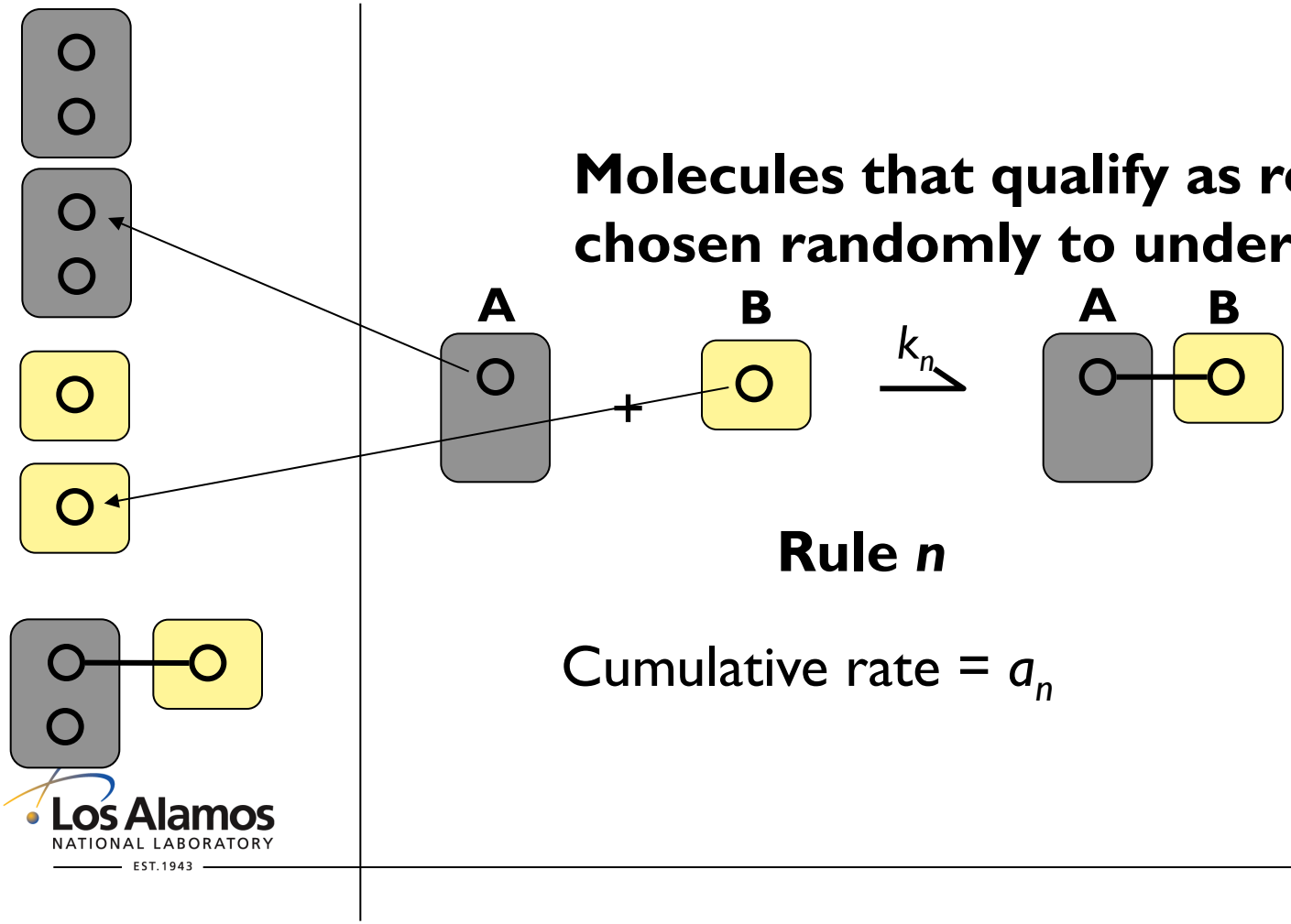


Rule *n*

Cumulative rate = a_n

Network-free (direct) simulation

Particles



Molecules that qualify as reactants are chosen randomly to undergo reaction

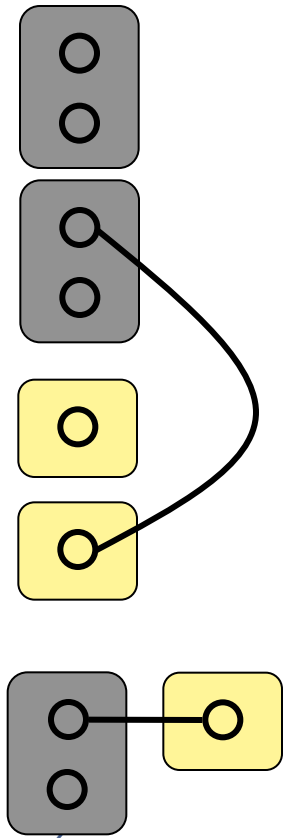
Rule *n*

Cumulative rate = a_n

Los Alamos
NATIONAL LABORATORY
EST. 1943

Network-free (direct) simulation

Particles



The transformation of the rule is applied, which generates a new system state

Los Alamos
NATIONAL LABORATORY
EST. 1943

Conclusions

- To capture the physicochemical dynamics of a cell signaling system, a coarse modeling approach (vs. an approach like MD) is required, one based on the principles of chemical kinetics, because relevant time scales are s to min.
- Traditional approaches for modeling chemical kinetics are difficult to apply if one is interested in complex mechanisms (anything except idealized small circuits) – why is this surprising?
  - To formulate an ODE model, one must enumerate the chemical species that are populated – astronomical numbers of chemical species can potentially be populated (so not all are important) but there is no data available to guide identification of the important species
- Descriptions of protein interactions in the biological literature can be viewed as descriptions of "local rules," which can be formalized to specify models for biomolecular site dynamics.
- Rule-based models can be simulated using indirect or direct methods
  - Indirect = derive equiv. model in trad. form + apply std. simulation algorithm
  - Direct = use rules as event generators (requires unconventional definition of system state)