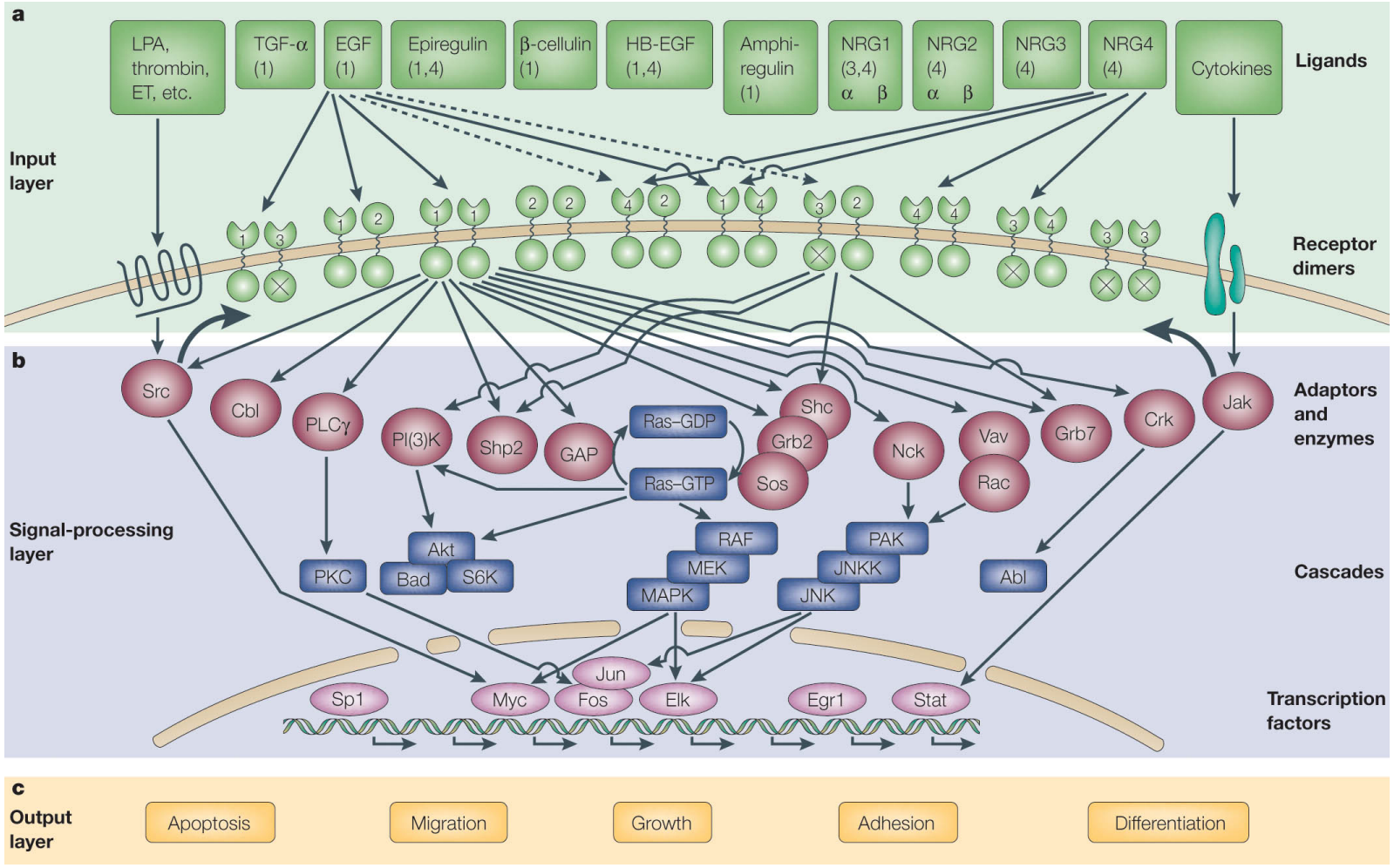
Rule-based modeling

NIMBioS tutorial, April 8-10th, 2013

Michael Blinov

- Vcell BioNetGen version: http://vcell.org/bionetgen
- Stand-alone BioNetGen version: http://bionetgen.org
- References used in the slides: http://www.ccam.uchc.edu/mblinov/Blinov_publications.html

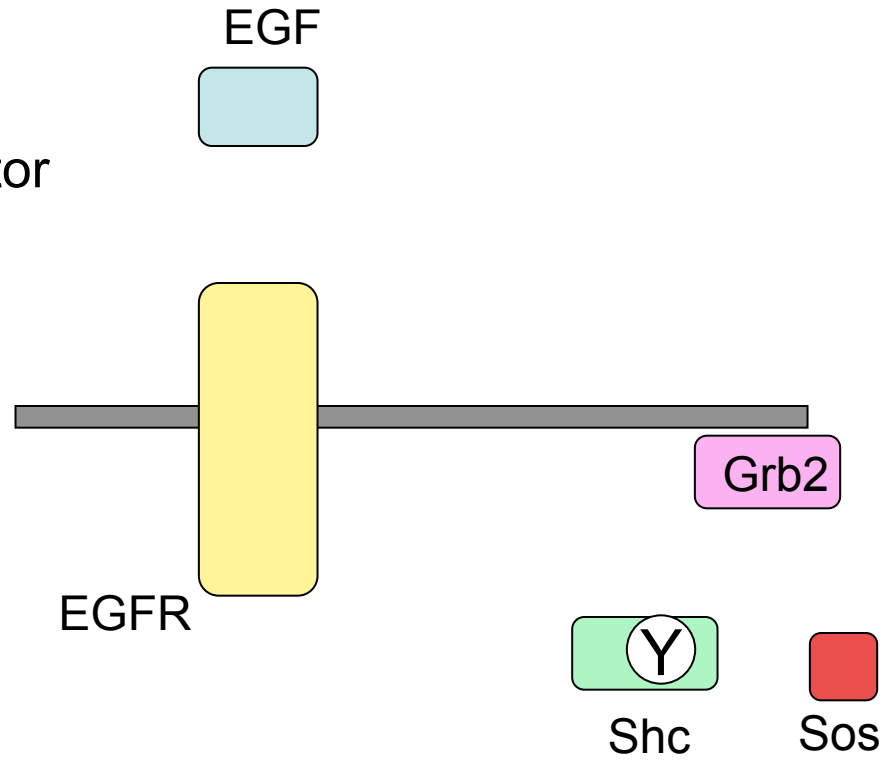
The first step in modeling: a cartoon – identifying a networks of proteins and other molecules that are involved in signaling



Yarden & Sliwkowski, *Nature Rev. Mol. Cell Biol.* **02**: 127-137 (2001).

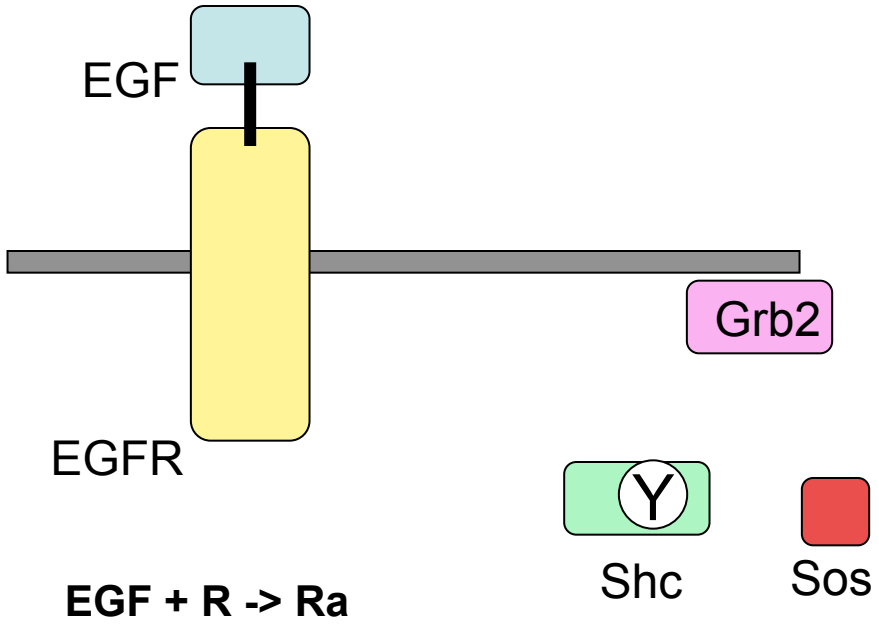
Early events in EGFR signaling

EGF = epidermal growth factor
EGFR = epidermal growth factor receptor
Grb2, Shc = Adapter proteins
Sos = GEF, Ras activator



Early events in EGFR signaling

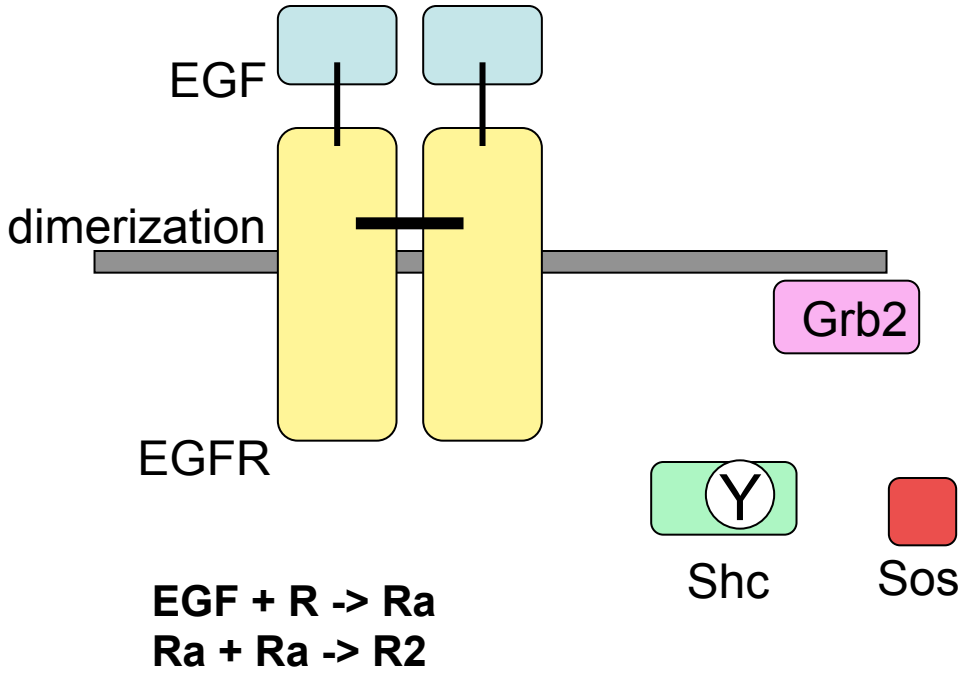
1. EGF binds EGFR



EGF + R -> Ra

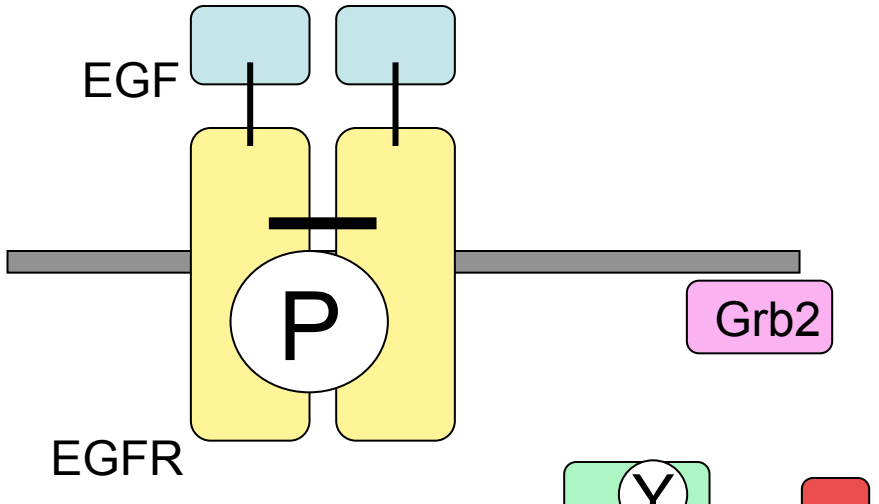
Early events in EGFR signaling

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



Early events in EGFR signaling

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

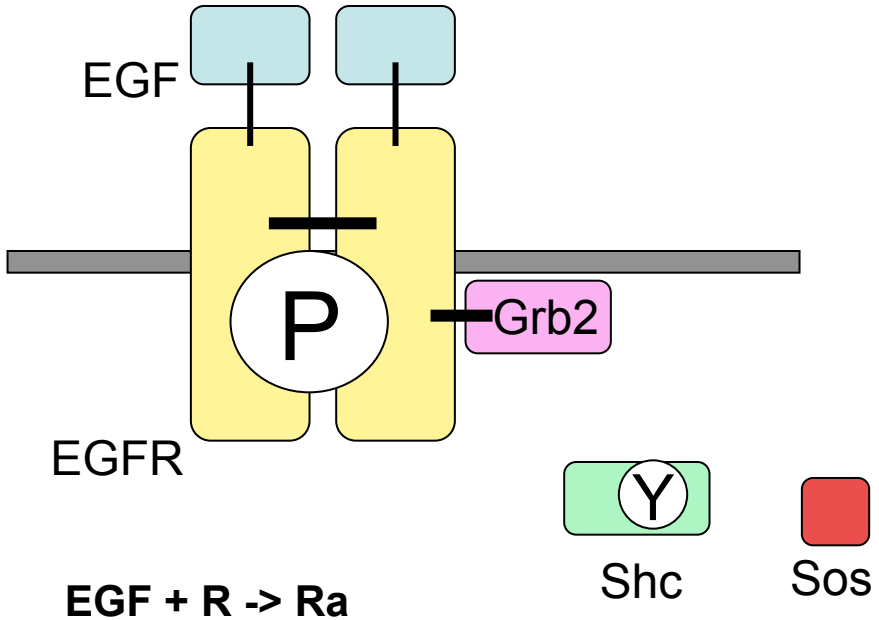


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp

Early events in EGFR signaling

Grb2 pathway

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

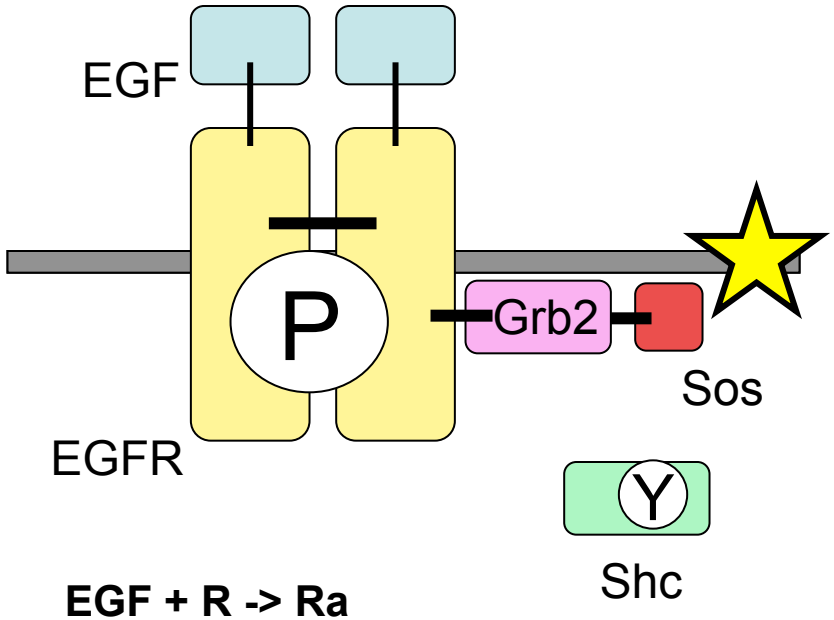


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G

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)**

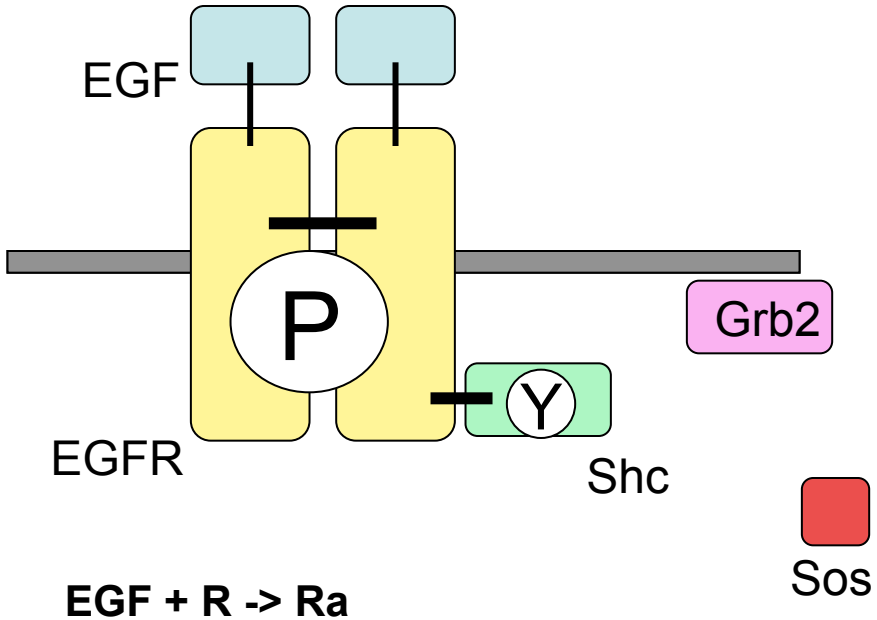


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G
R-G + Sos -> R-G-S

Early events in EGFR signaling

Shc pathway

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

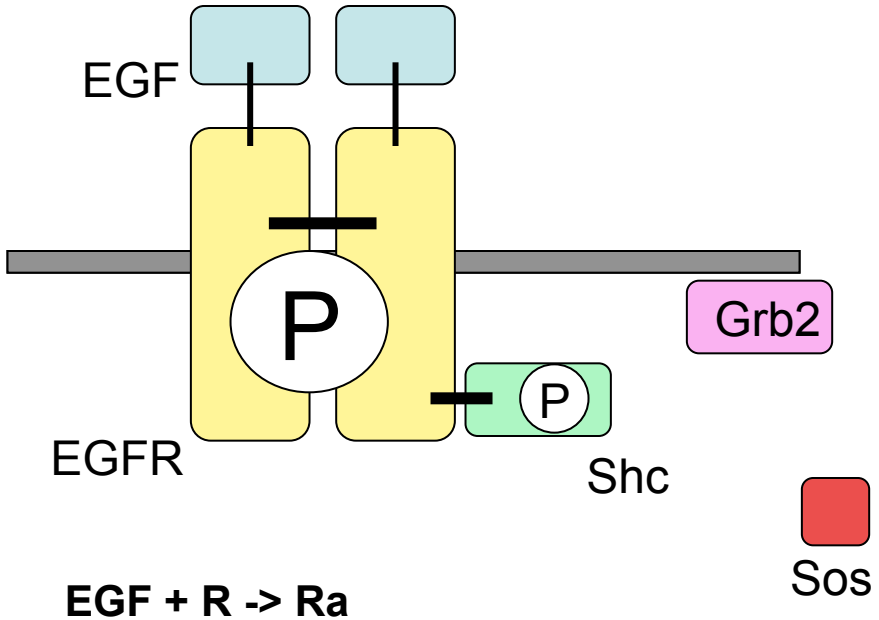


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G
R-G + Sos -> R-G-S
Rp + Shc -> R-Sh

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**

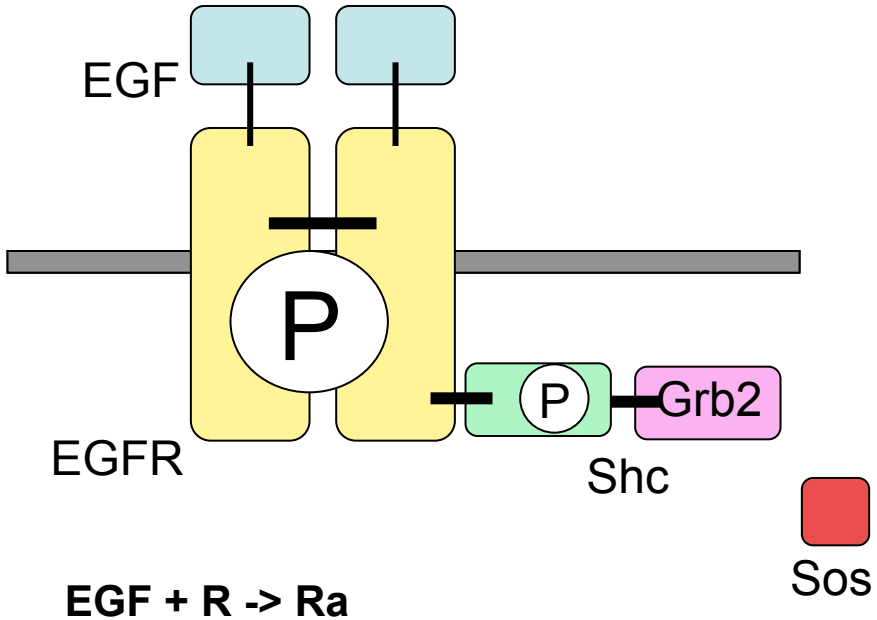


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G
R-G + Sos -> R-G-S
Rp + Shc -> R-Sh
R-Sh -> R-ShP

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**

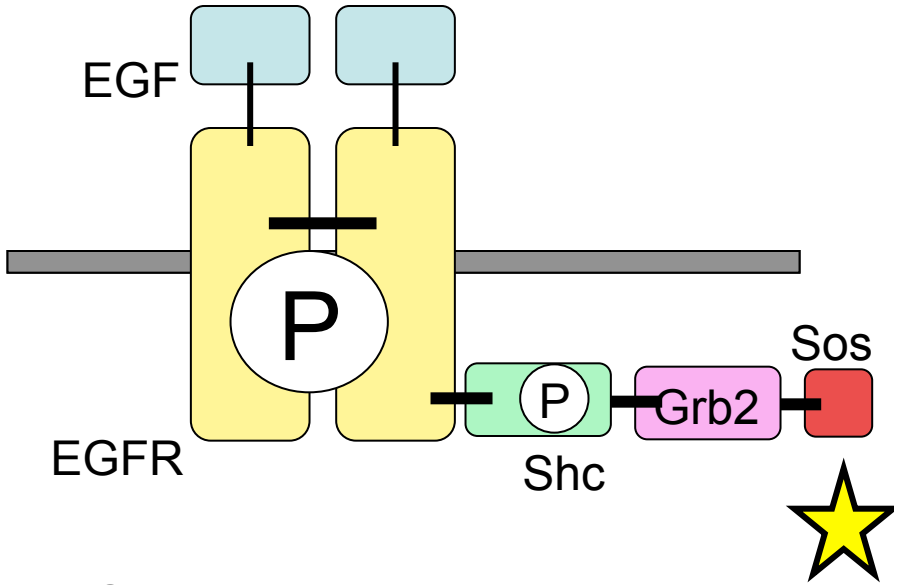


EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G
R-G + Sos -> R-G-S
Rp + Shc -> R-Sh
R-Sh -> R-ShP
R-ShP + Grb2 -> R-Sh-G

Early events in EGFR signaling

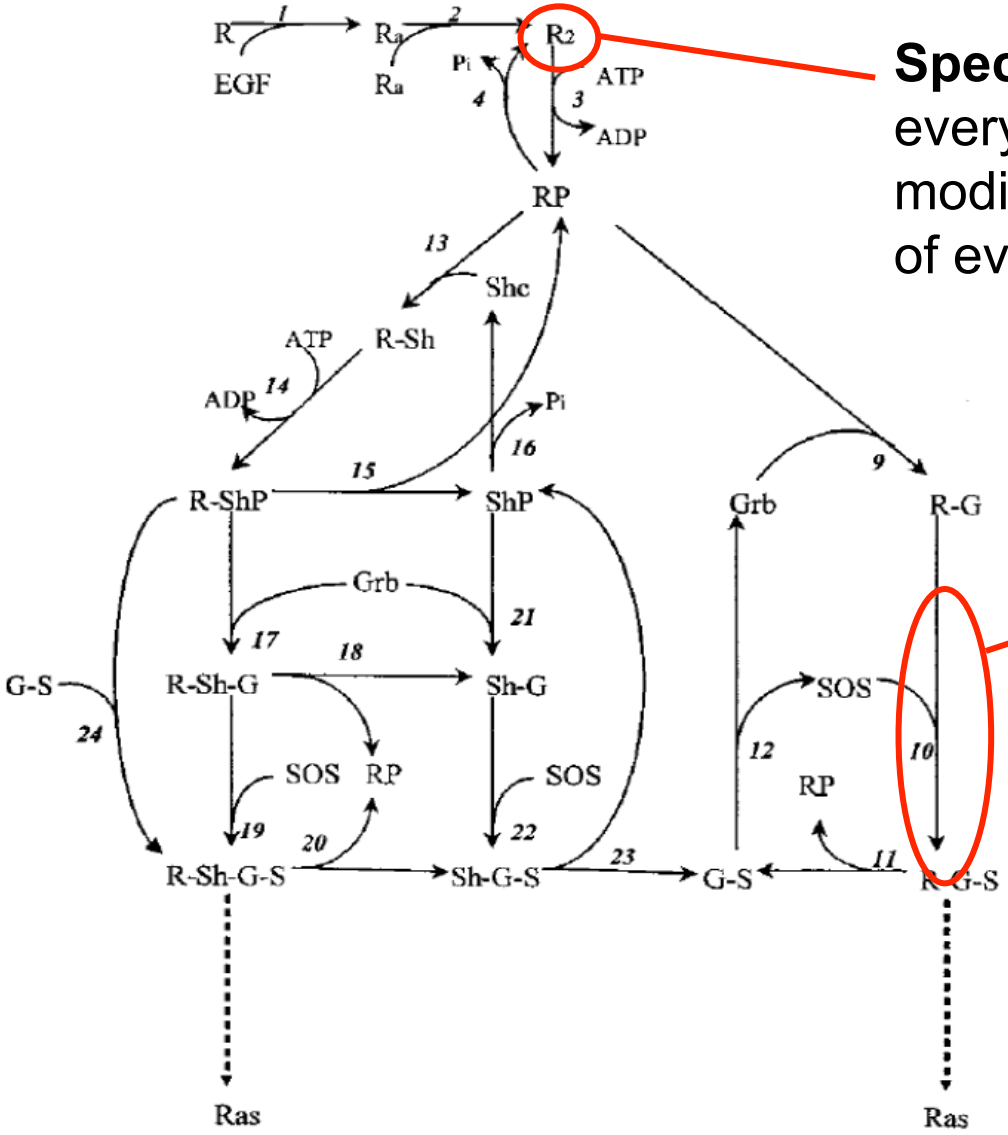
Shc pathway

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)**



EGF + R -> Ra
Ra + Ra -> R2
R2 -> Rp
Rp + Grb2 -> R-G
R-G + Sos -> R-G-S
Rp + Shc -> R-Sh
R-Sh -> R-ShP
R-ShP + Grb2 -> R-Sh-G
R-Sh-G + Sos -> R-Sh-G-S

The next step: write down reaction network



Species: One for every possible modification state of every complex

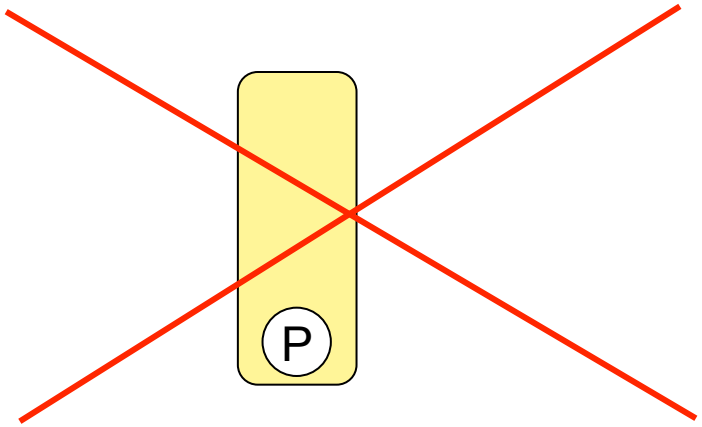
Reactions: One for every transition among species

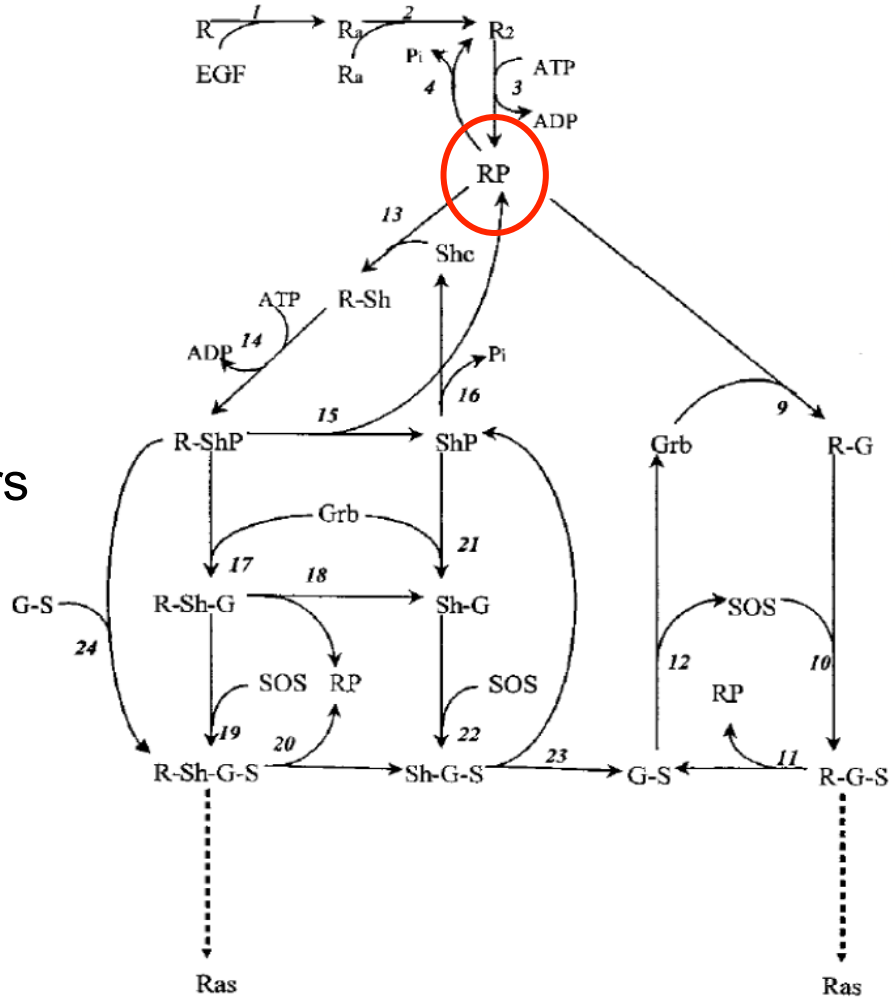
Kholodenko et al., J. Biol. Chem. **274**, 30169 (1999)

Assumptions made:

Phosphorylation inhibits dimer breakup

No phosphorylated monomers

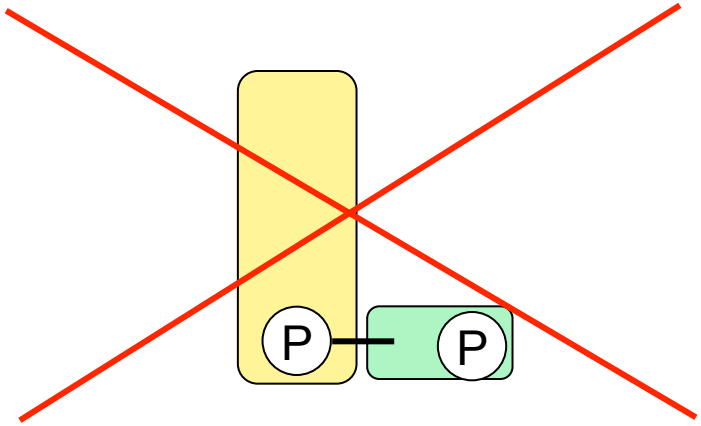


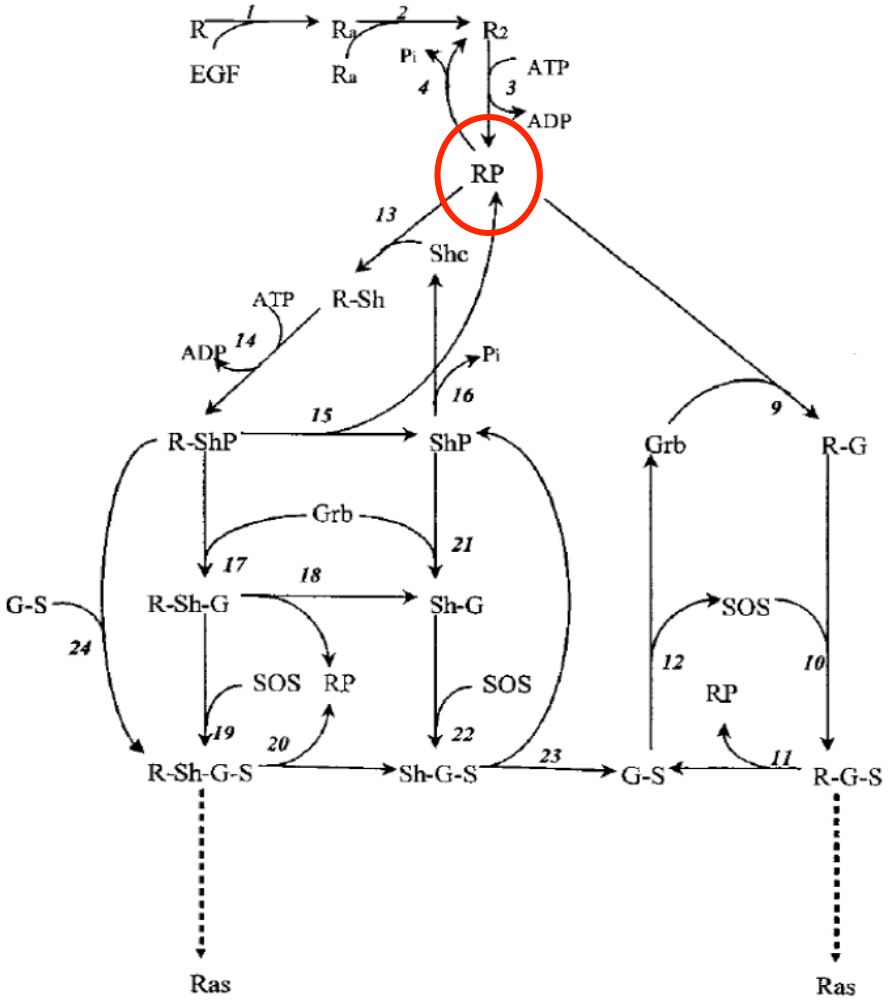


Assumptions made:

Phosphorylation inhibits dimer breakup

No monomeric complexes

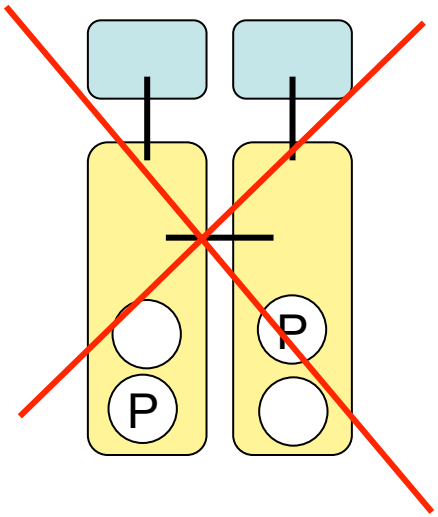


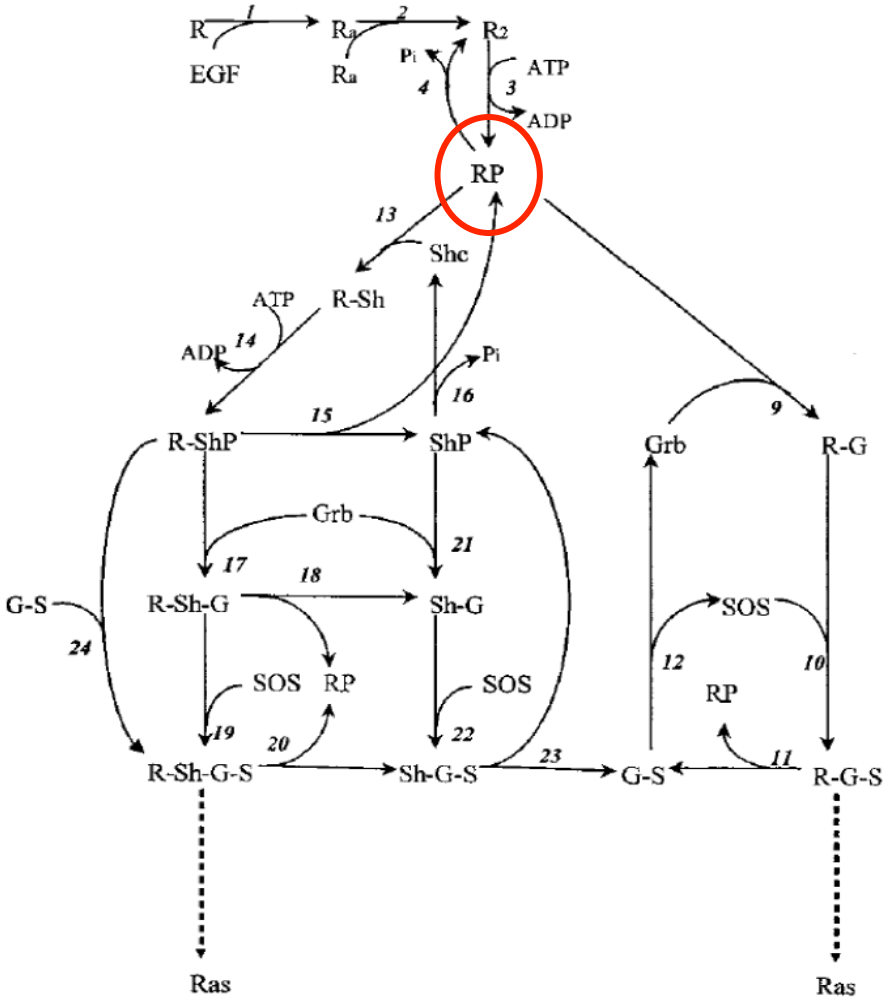


Assumptions made:

Phosphorylation is simultaneous

Same phosphorylation timecourses for all residues

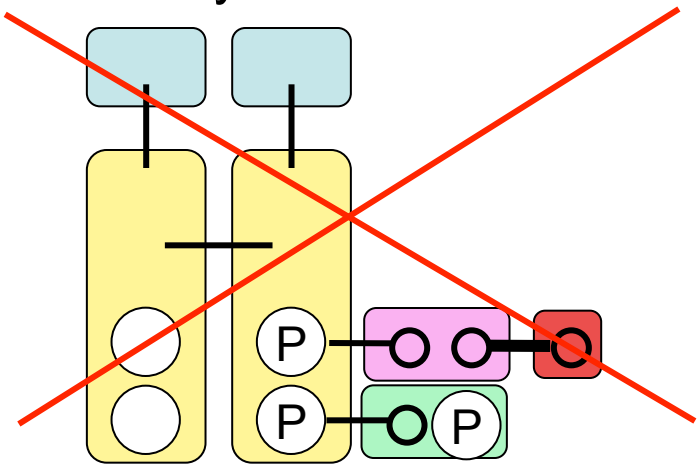


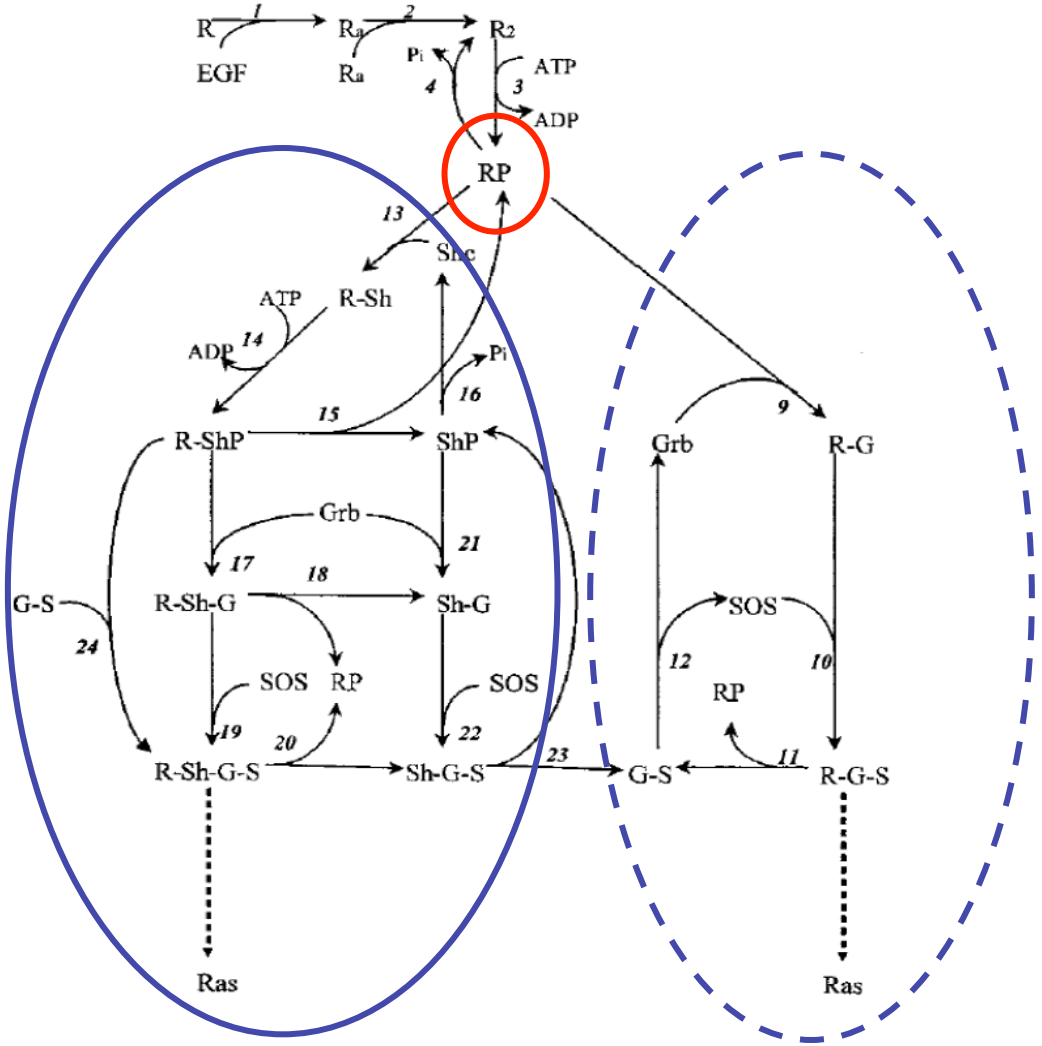


Assumptions made:

Adaptor binding is competitive

Only one adapter ptotein can bind at any time



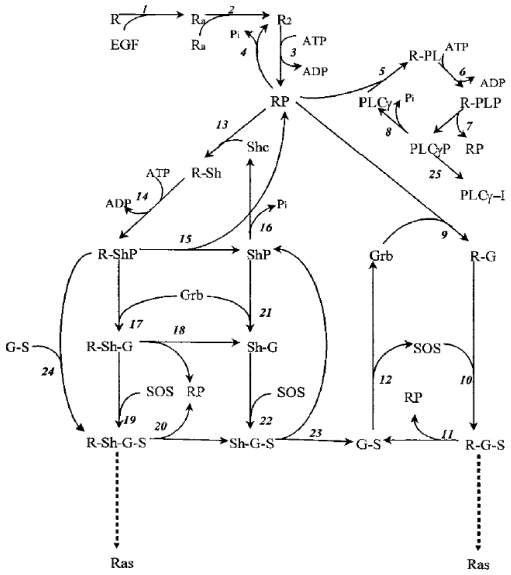


Blinov et al., BioSystems 2006

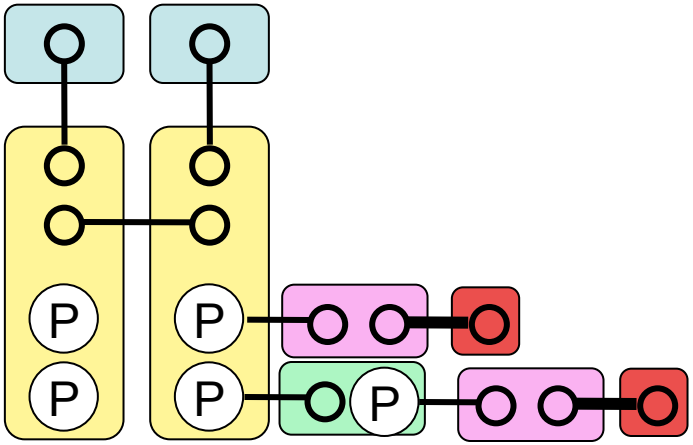
Rule-based version of the Kholodenko model

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

18 species
34 reactions



356 species
3749 reactions



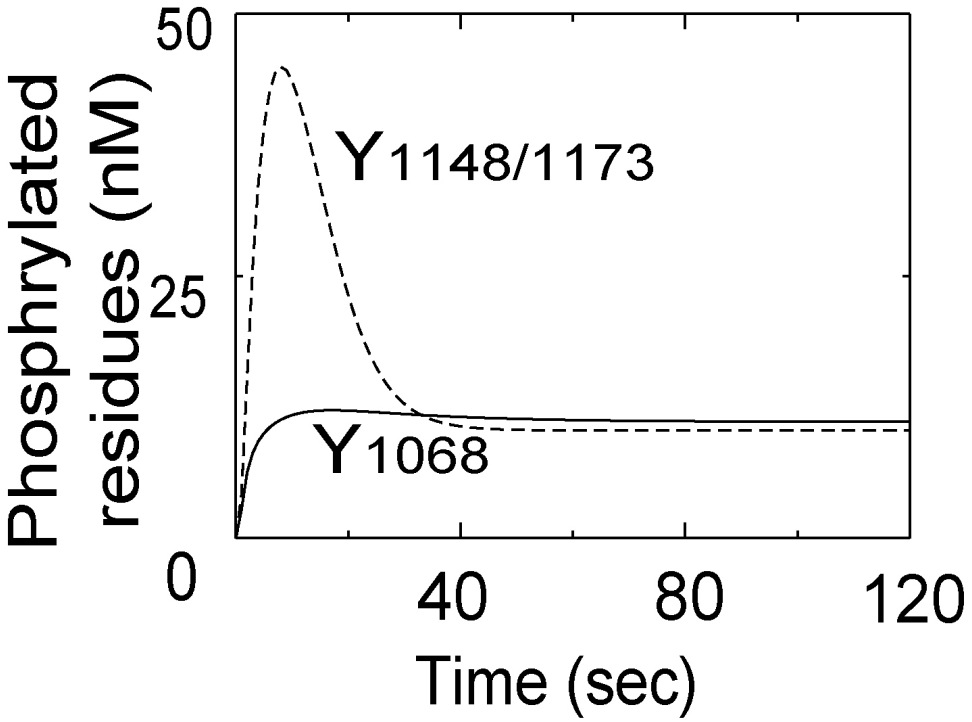
Blinov et al. *Biosystems* **83**, 136 (2006).

If we would include protein domain, we would be able to

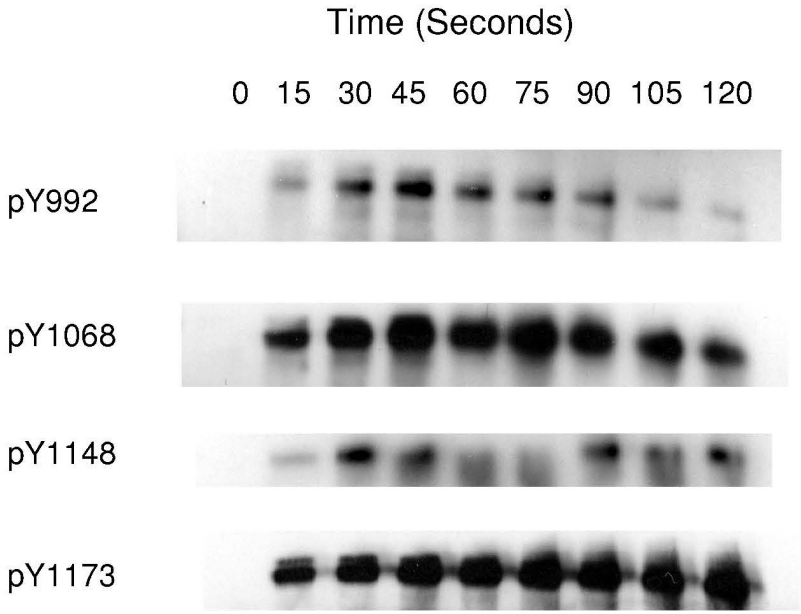
Blinov et al. *Biosystems* **83**, 136 (2006).

New testable predictions

Different dynamics for phoshorylation of different tyrosine residues.

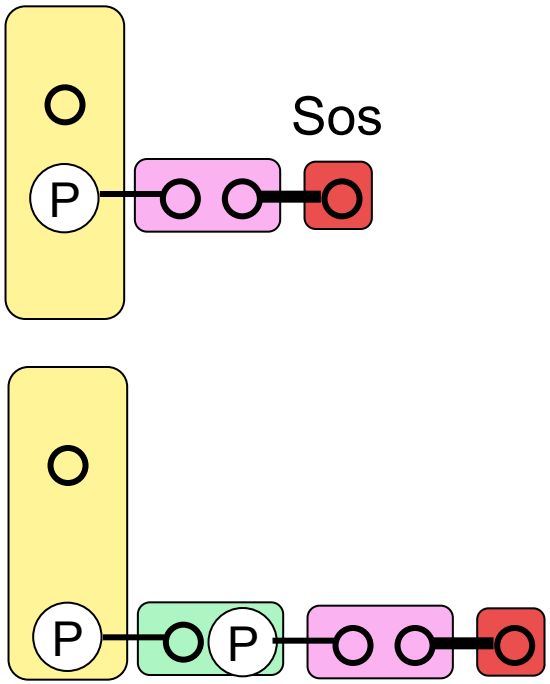


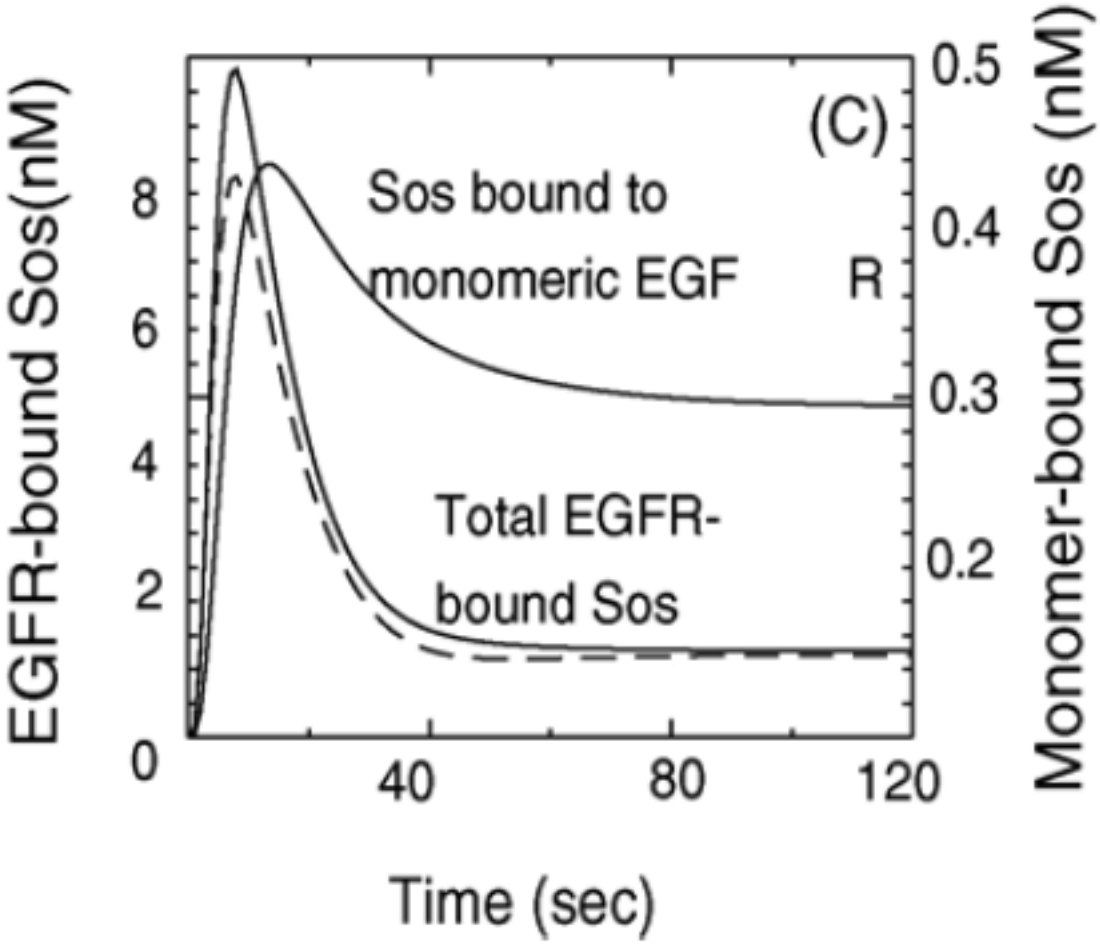
Edward Stites and Kodi Ravichandran (preliminary data, 2004



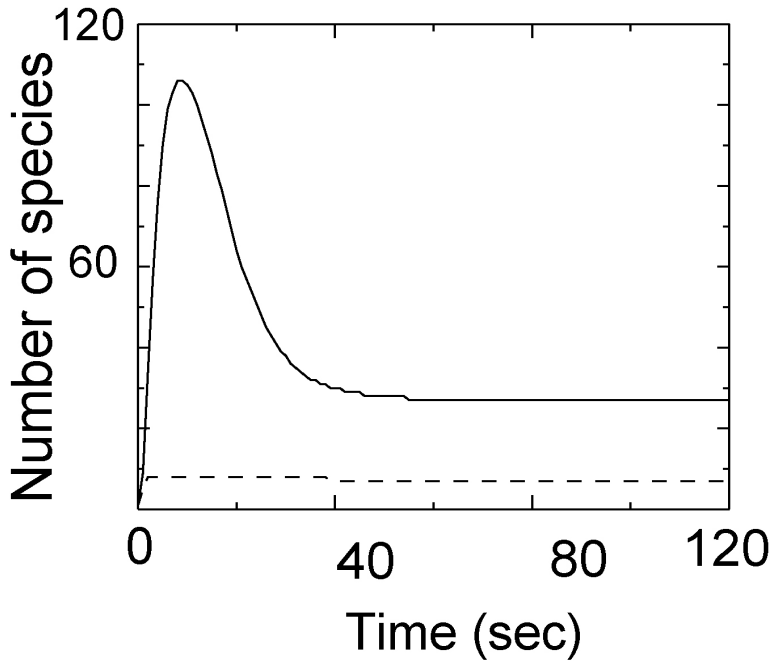
Also predicts monomers make substantial contribution to steady state Sos activation

36% of active Sos associates with EGFR monomers

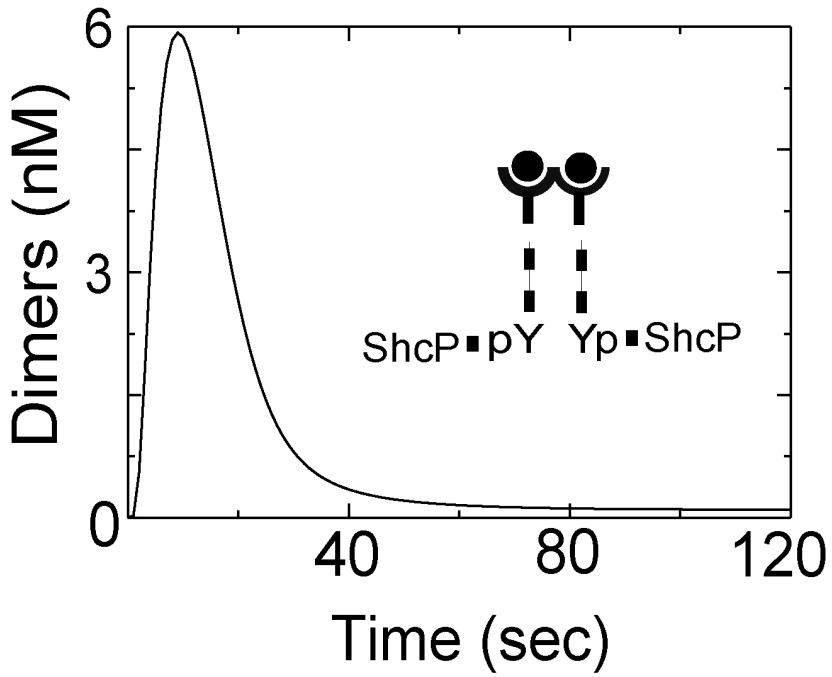




Much larger number of distinct chemical species participates in signaling at short times than at steady state



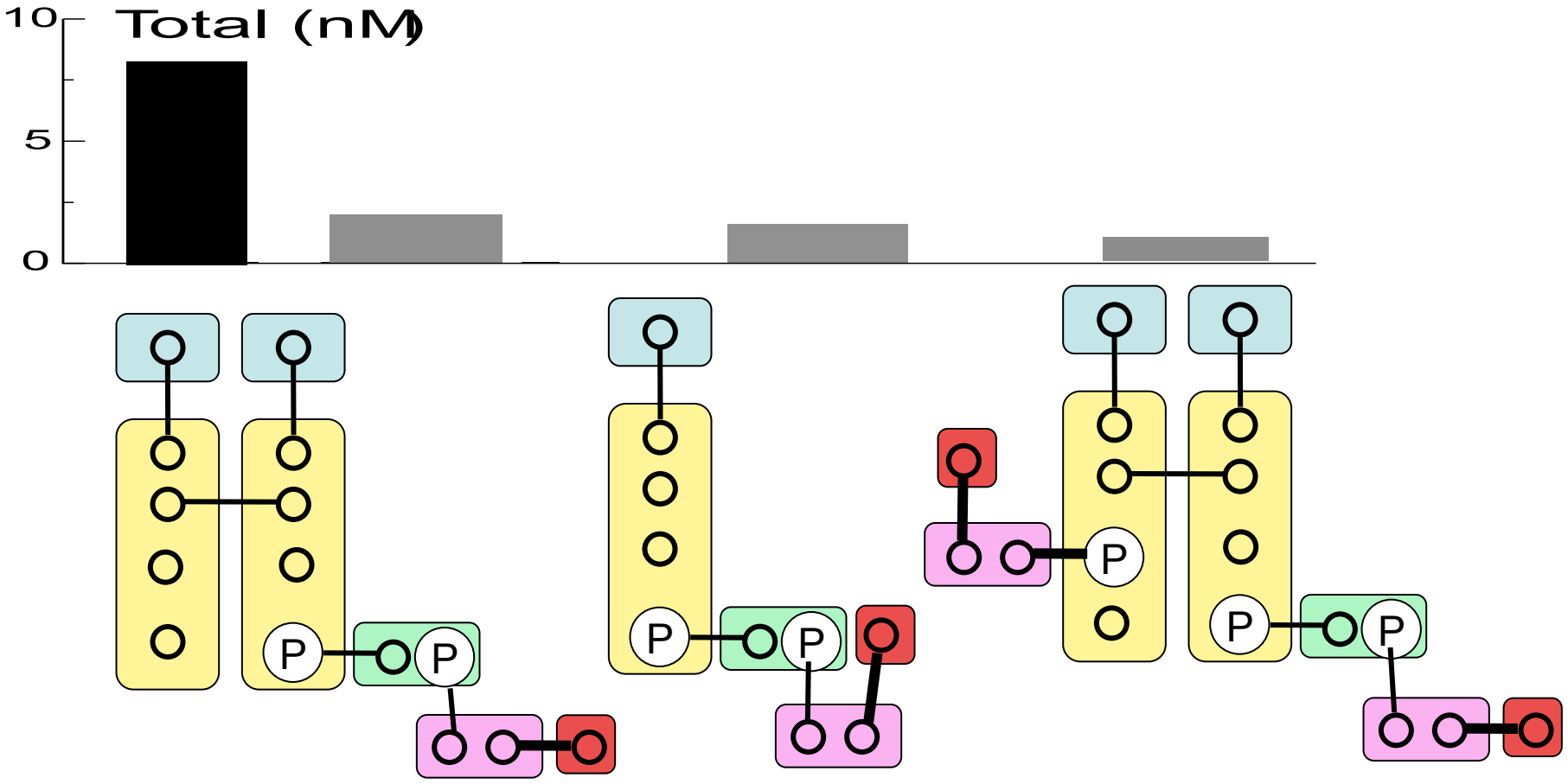
Significant amount of dimers have multiple bound proteins at short times



7% of dimers form complexes with two ShcP
30% of ShcP at transient is in complexes with one more ShcP

Dominant molecular complexes

Few chemical species are predicted to account for almost all recruited Sos at steady state.

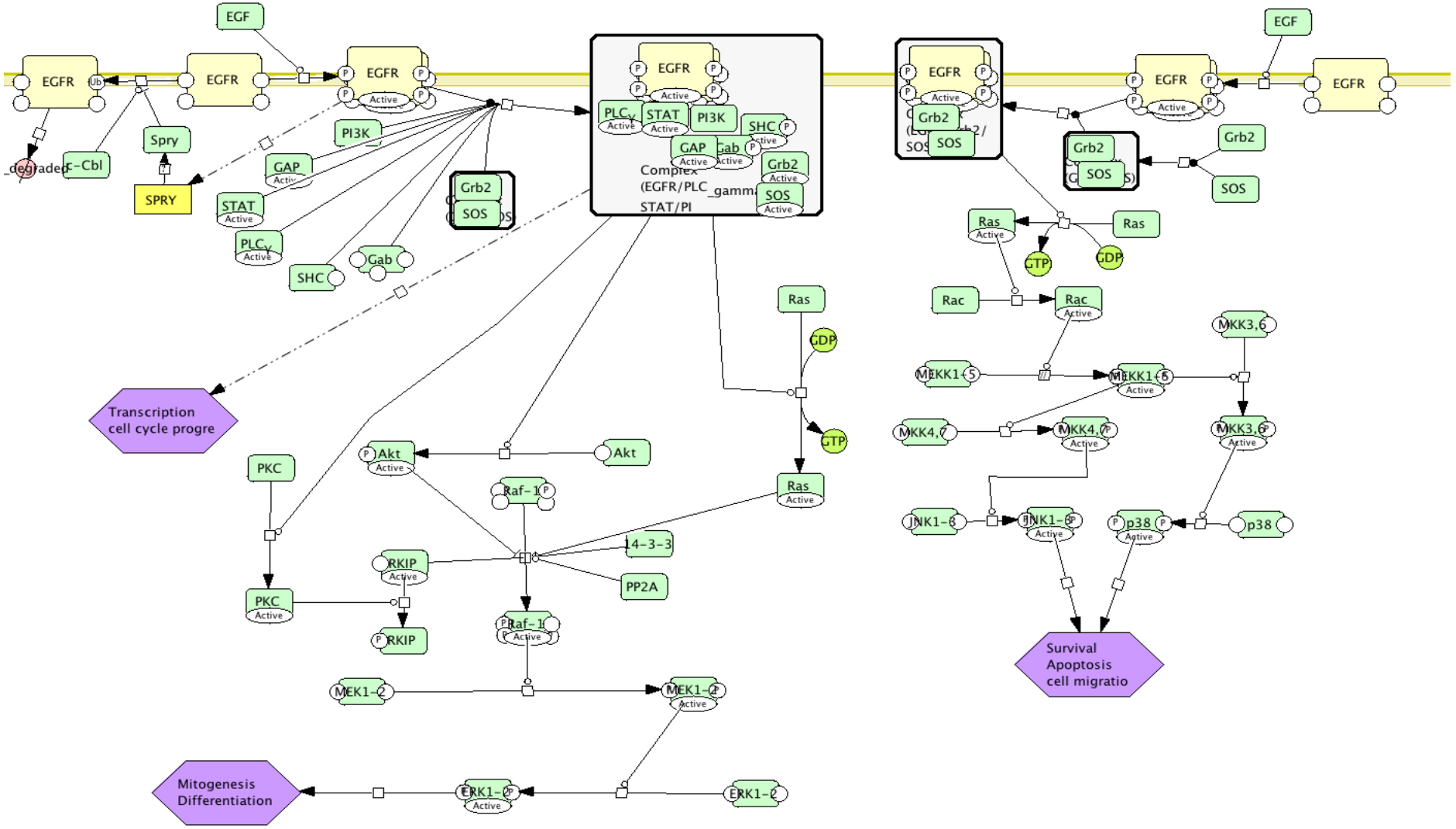


Our problem is: complexity

Hlavacek et al. Biotechnology Bioengineering (2003)

PANTHER (Protein ANalysis THrough Evolutionary Relationships)

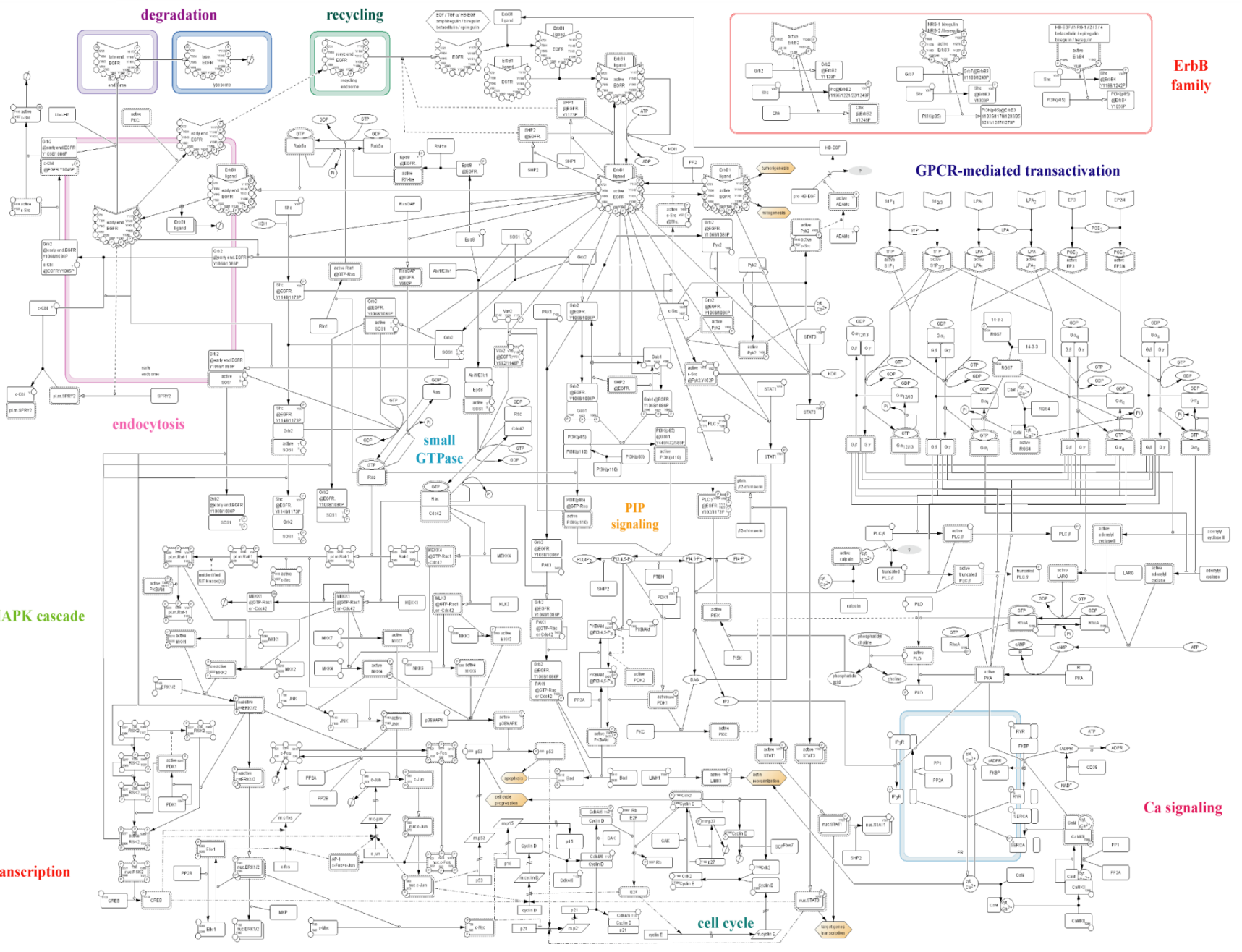
- http://www.pantherdb.org/
- SBGN



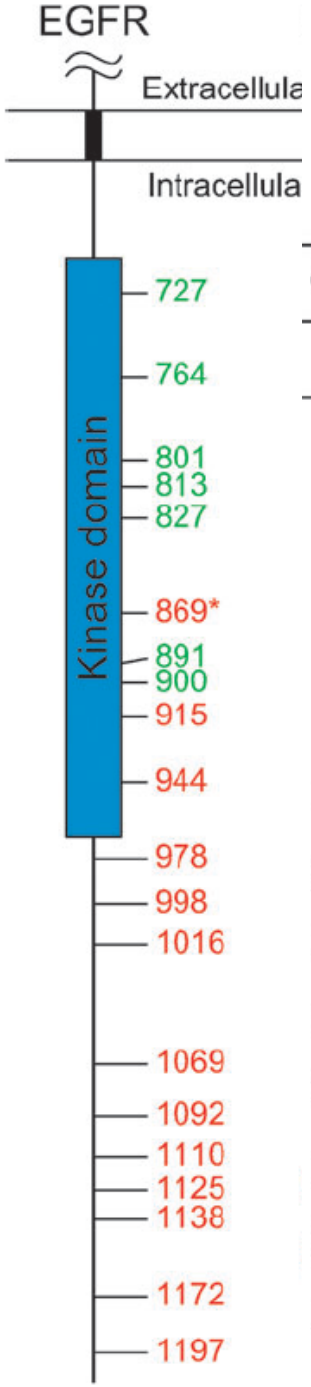
Epidermal Growth Factor Receptor Pathway Map

(EGFR0504V2)

Kanae Oda (1,2), Yukiko Matsuoka (3), Hiroaki Kitano (1,2,4)
(1) The Systems Biology Institute, (2) Department of Fundamental Science and Technology, Keio University,
(3) ERATO-SORST Kitano Symbiotic Systems Project, Japan Science and Technology Agency, (4) Sony Computer Science Laboratories, Inc.

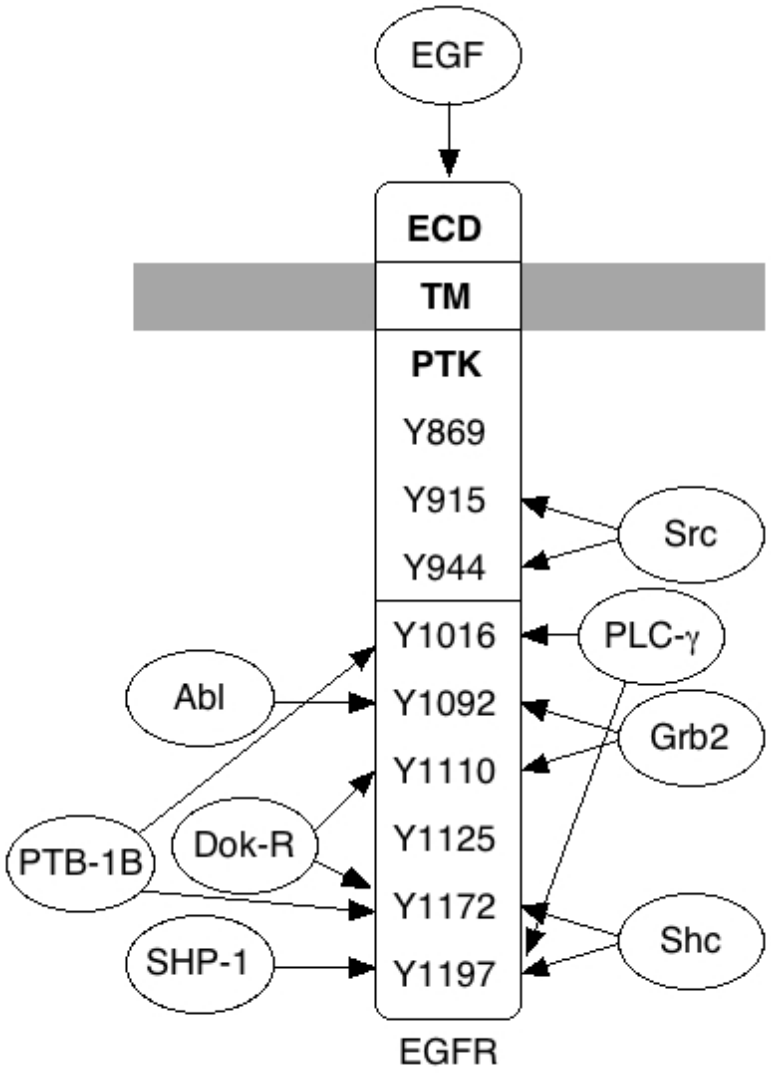


Domain-domain interactions



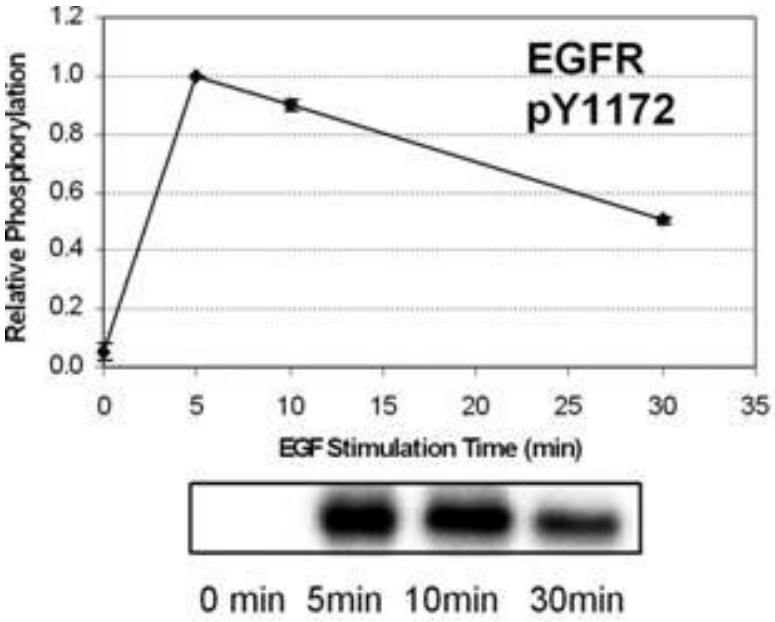
EGFR

Site	Peptide sequence
869	KLLGAEEKEpYHAEGGKVD
915	ELMTFGSKPpYDGIPASED
944[a]	QPPICTIDVpYMIMVKCWD
978	SKMARDPQRpYLVIQGDED
998	LPSPTDSNFpYRALMDEED
1016	MDDVVDADEpYLIPQQGFD
1069	IKEDSFLQRpYSSDPTGAD
1092	DDTFLPVPEpYINQSVPKD
1110	PAGSVQNPVpYHNQPLNPD
1125	NPAPSRDPHpYQDPHSTAD
1138	HSTAVGNPEpYLNTVQPTD
1172	HQISLDNPDpYQQDFFPKD
1197	KGSTAENAEpYLRVAPQSD



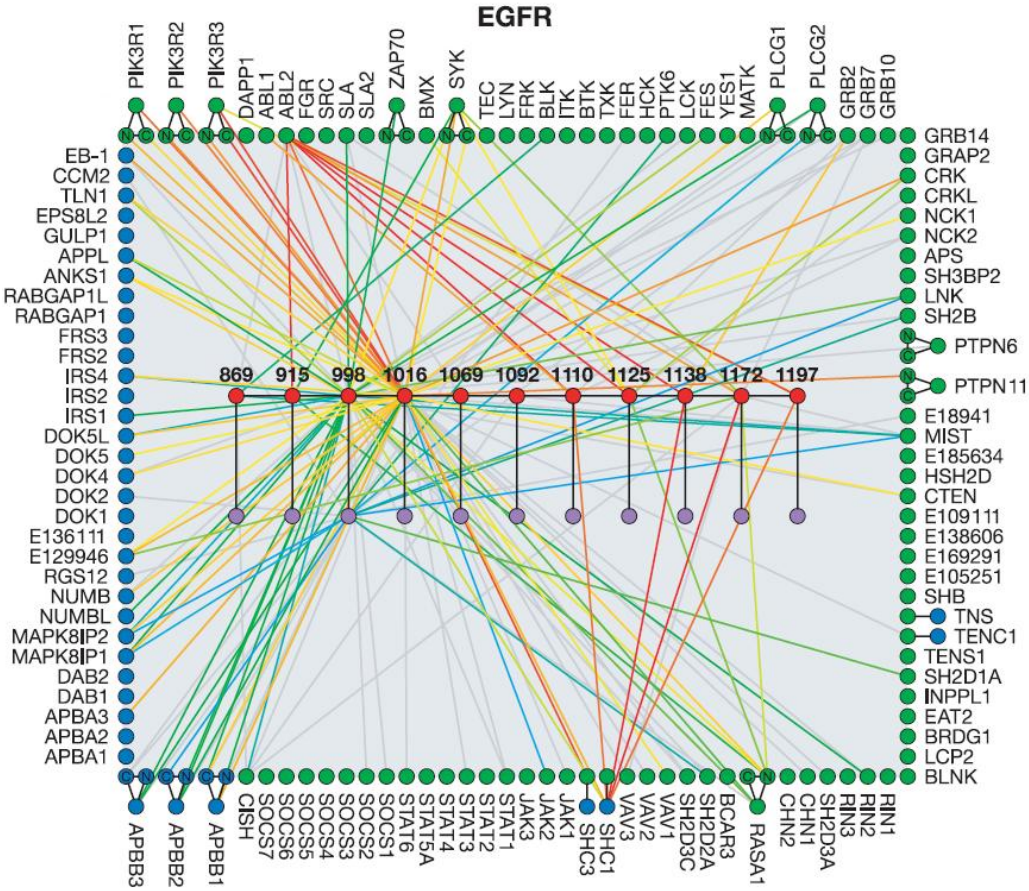
Experimental data

the kinetics of multiple phosphorylation sites



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

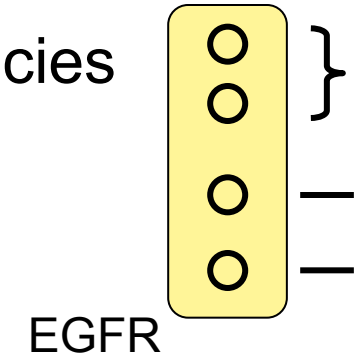
affinities for multiple binding partners



Richard B. Jones et al.,
Nature 439, 168-174 (2006).

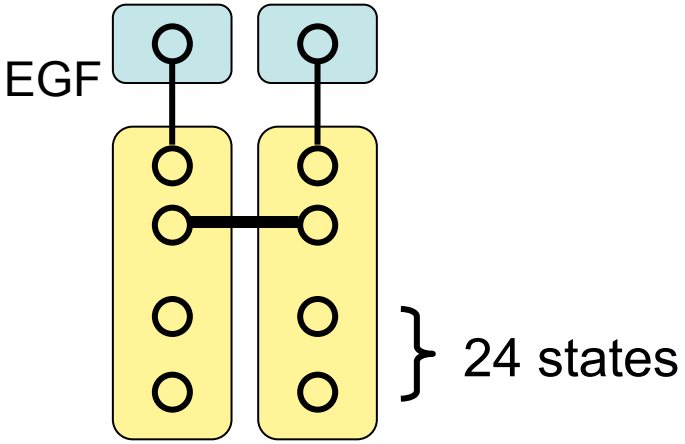
Combinatorial complexity of early events

Monomeric species



48 species

Dimeric species



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

The problem: multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

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

Each site has ≥ 1 binding partner

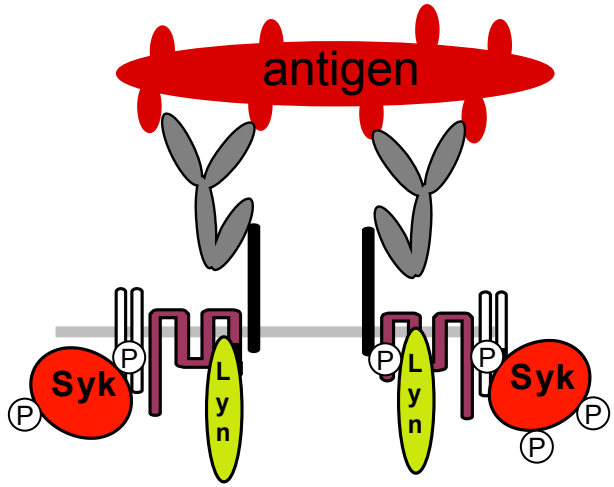
$\Rightarrow$ more than 3^9=19,683 total states

EGFR must form *dimers* to become active

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

Early Events in FcεRI receptor Signaling

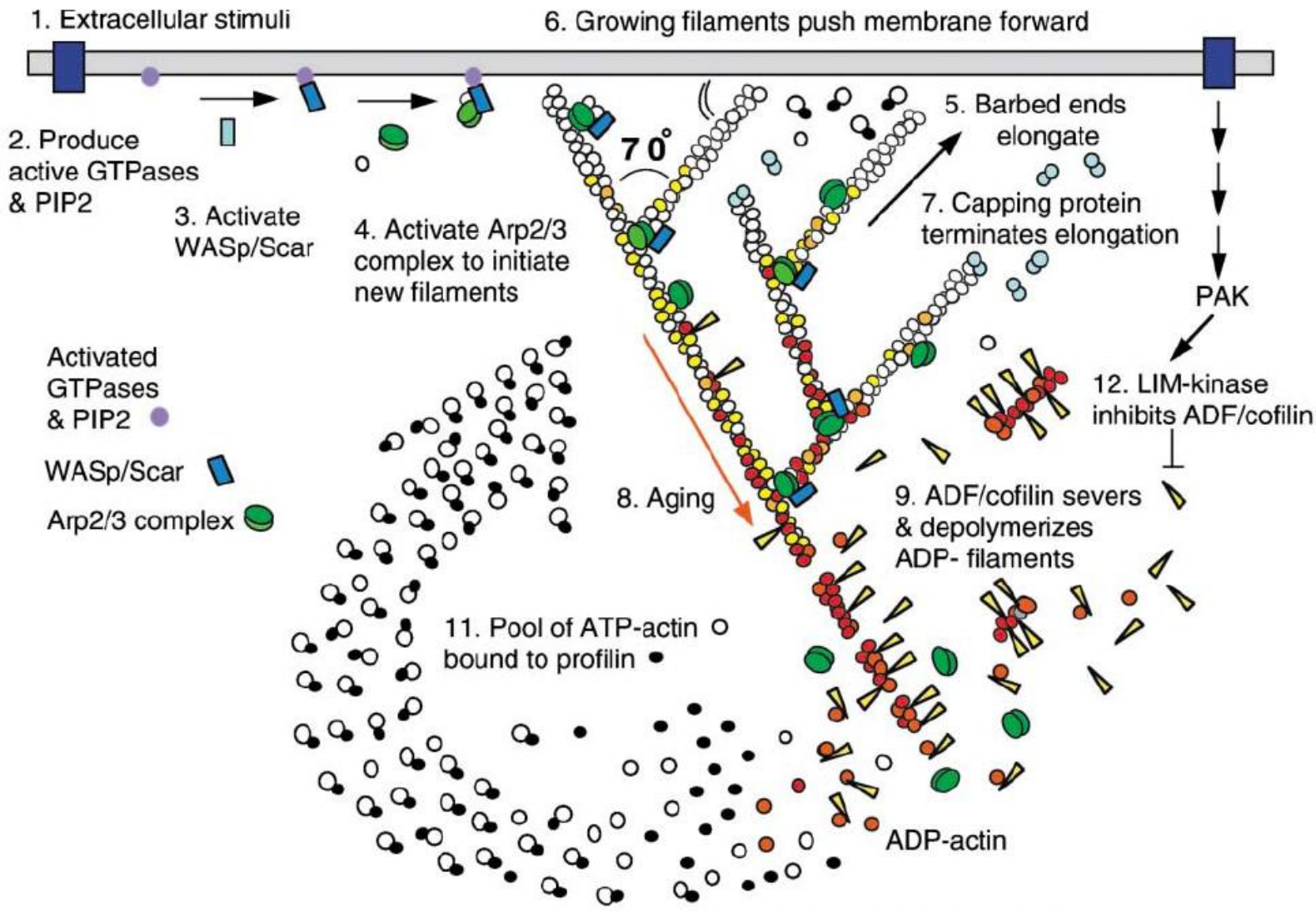
1. Multivalent antigen binds to IgE on cell surface forming aggregates
2. Tyrosine kinase Lyn associates with receptors and **transphosphorylates** ITAM tyrosines
3. Phosphorylated ITAMs recruit Syk and additional Lyn
4. Syk is transphosphorylated by Lyn or Syk
5. Phosphorylation of Syk is critical for downstream events ("activation")



Faeder et al., J. Immunol. (2003)

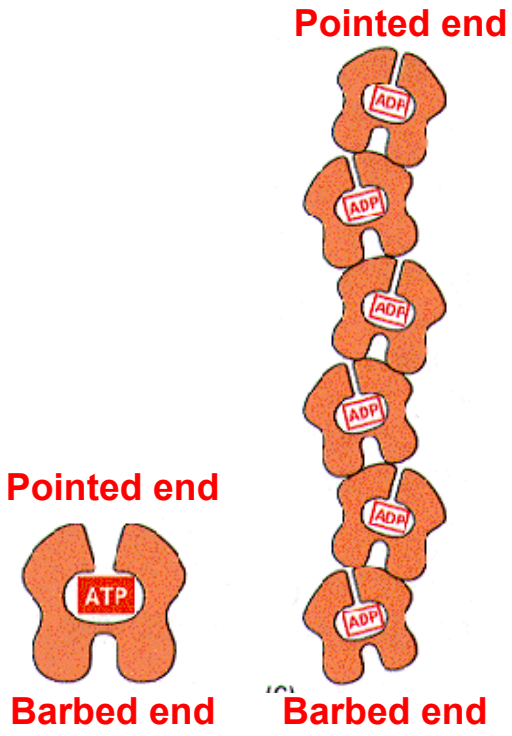
Not a pathway!

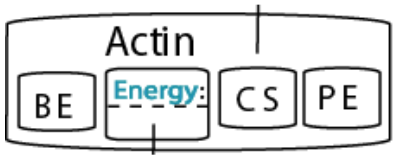
Actin Filaments Formation



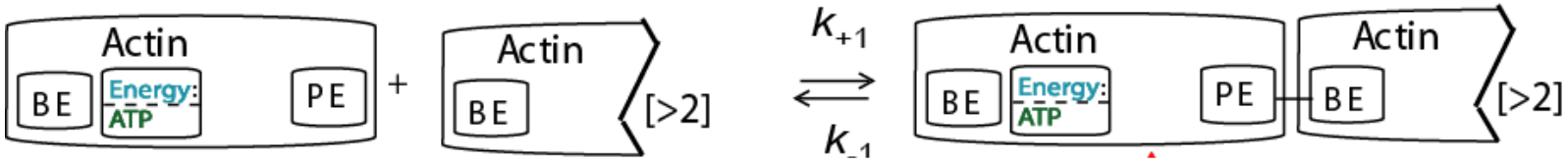
Pollard et al., Annual Rev Biphys Biomol Sruct (2000)

Infinite chains

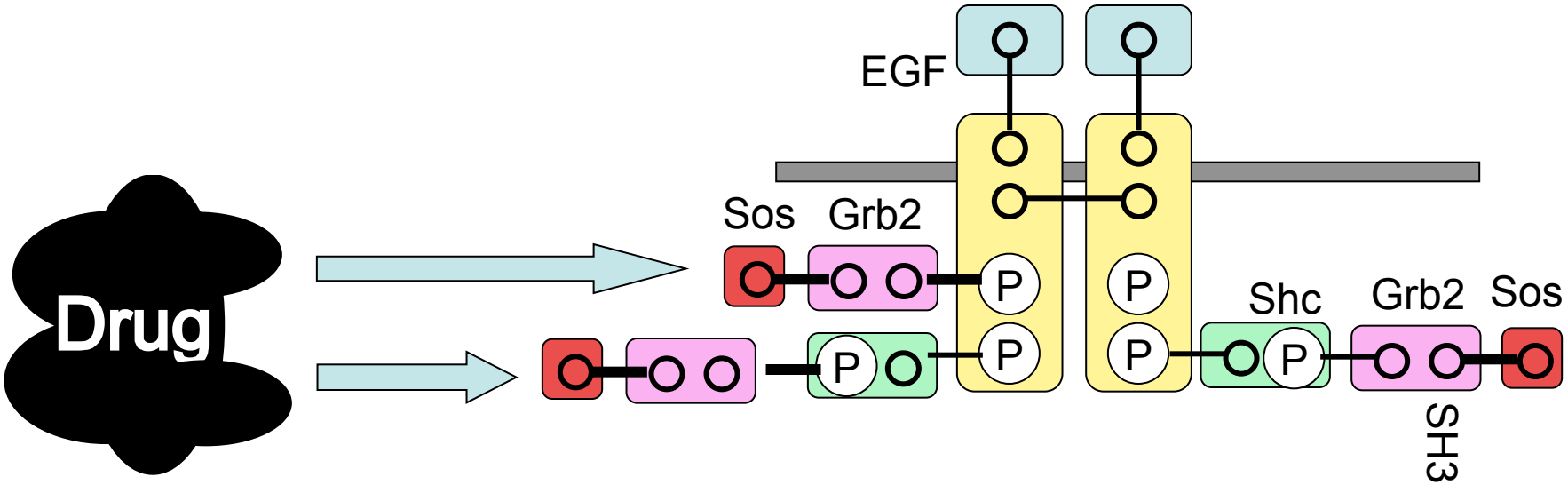




Barbed end of F-Actin of the length more than 2



Big promise???

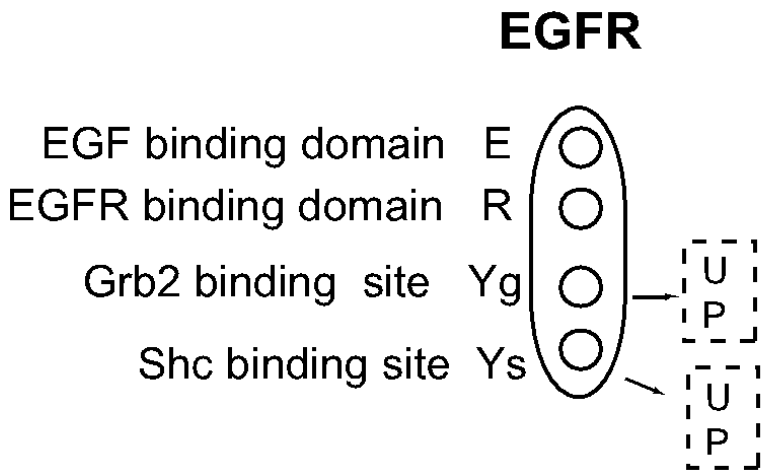


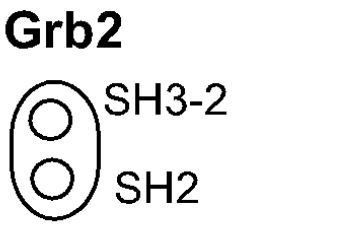
Understanding at this level of detail is critical to our ability to develop new therapies for disease

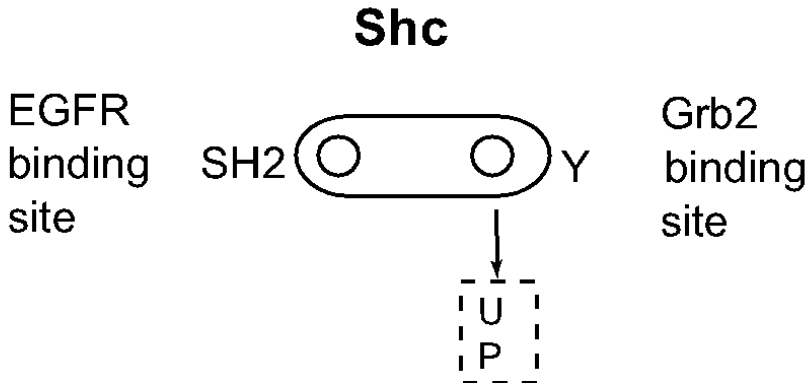
Graph-based representation

M. L. Blinov, et al (**2006**) Graph theory for rule-based modeling of biochemical networks. Lect. *Notes Comp. Sci 4230*

Molecular entity graph: examples

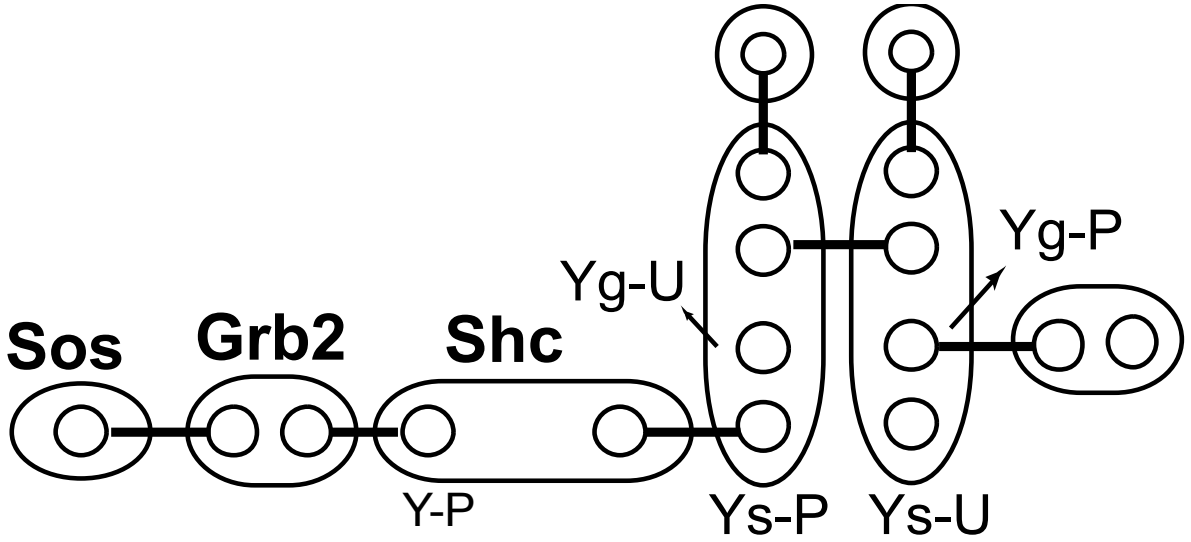




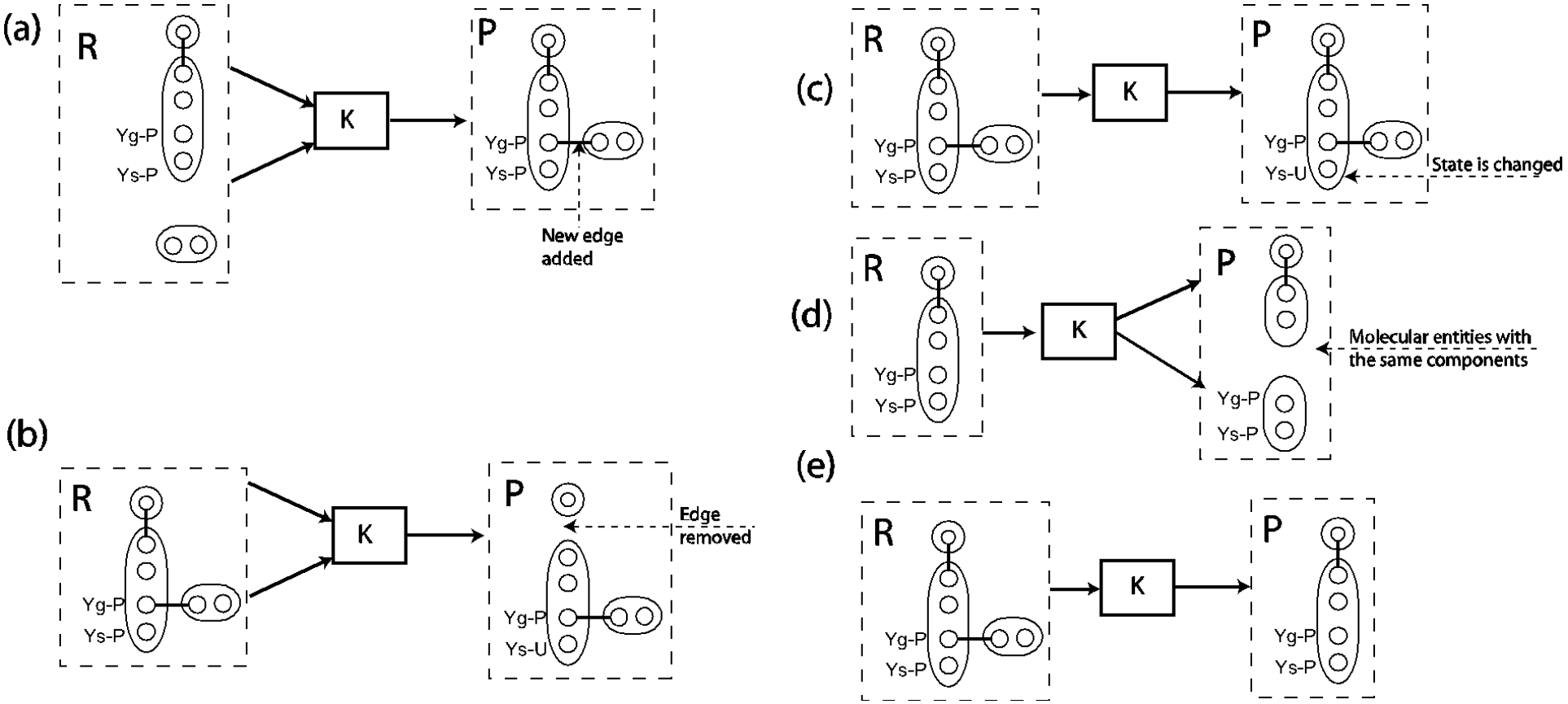


Chemical Species graph: definition

- A Chemical Species Graph C is a fully defined molecular entity or a set of molecular entities.
  - Any and all variable attributes taking specific values.



Reaction is a graph rewriting consistent with chemistry

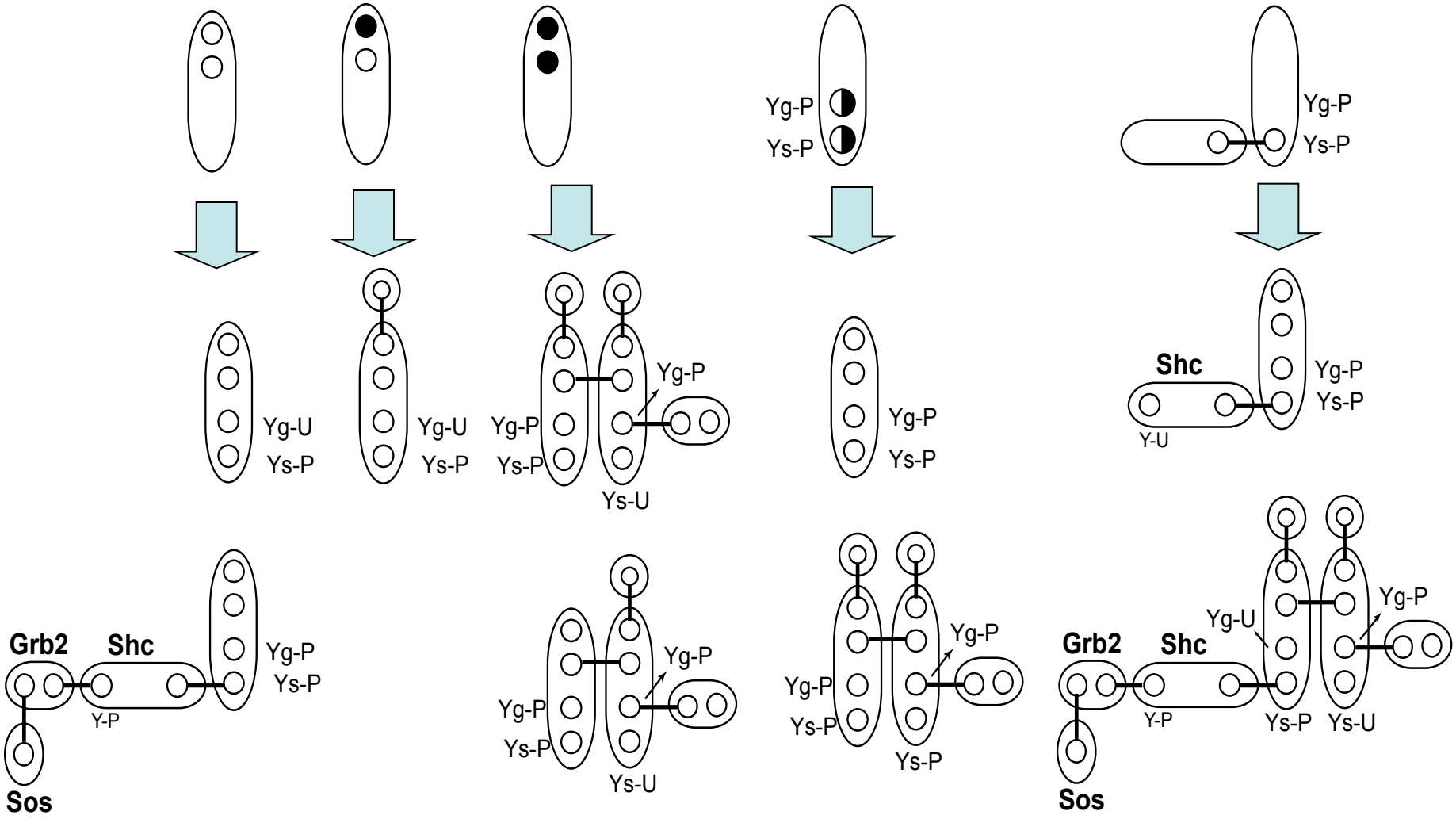


Rule-based description

M. L. Blinov, et al. Graph theory for rule-based modeling of biochemical networks. Lect. *Notes Comp. Sci 4230* (2006)

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

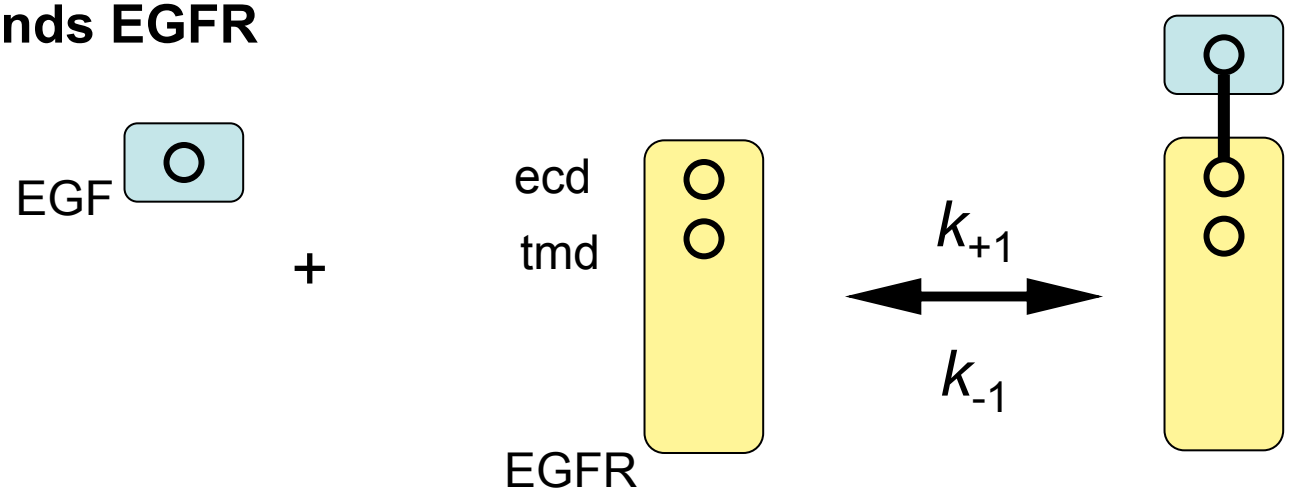
Chemical species selected by patterns



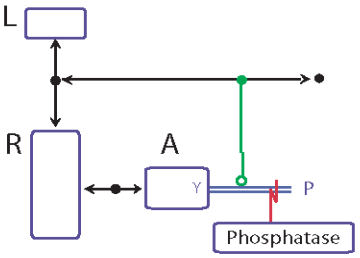
Reaction rules define individual reactions

- Each rule specifies some **experimentally-testable** feature of the system

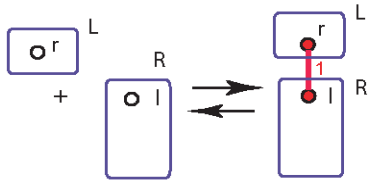
EGF binds EGFR



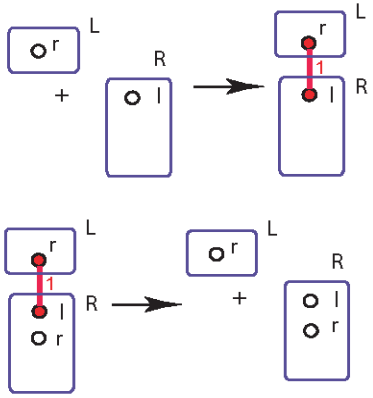
(a)



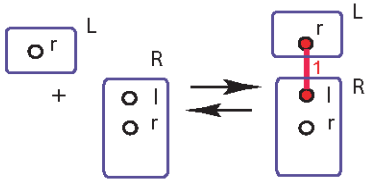
(b1) Ligand-binding independent on dimerization



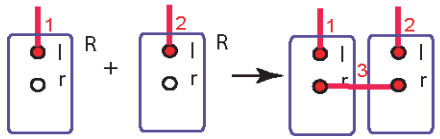
(b2) Ligand binds to any receptor, but can not dissociate in a dimer



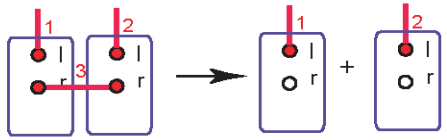
(b3) Ligand can interact with monomers only



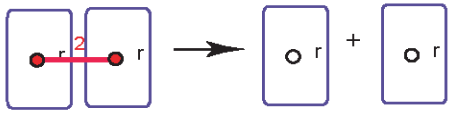
(c1) Dimer formation is ligand-induced



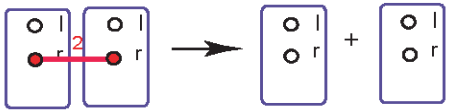
(c2) Dimer can break-up only when both ligands are present



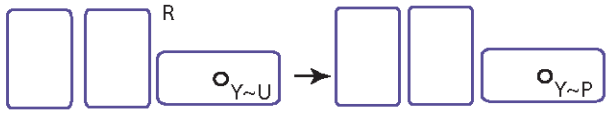
(c2) Dimer break-up is spontaneous



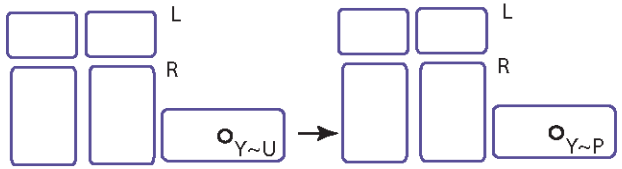
(c4) Dimer can break-up only after both ligand are gone.



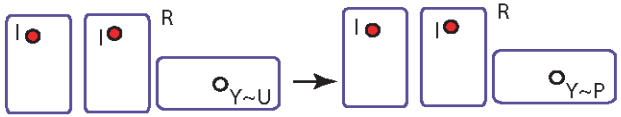
(d1) A is phosphorylated in a dimer



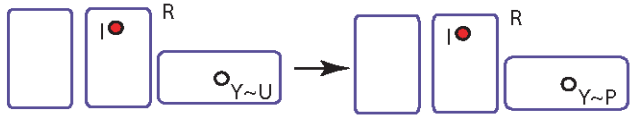
(c2) Phosphorylation requires 2 ligands L



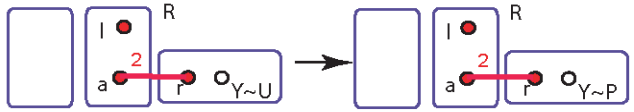
(d3) Phosphorylation requires two ligands



(d4) Phosphorylation requires at least one ligand



(d5) Explicit requirement which ligand is required



Rule-based modeling

Problem

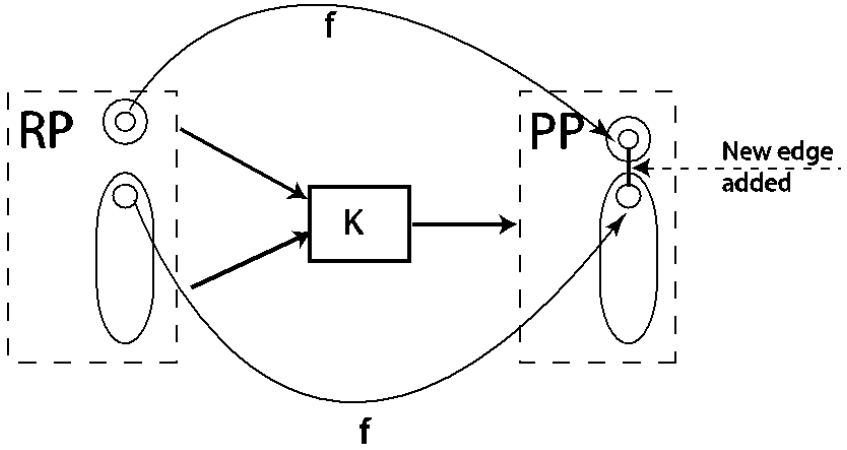
To **explicitly** specify all species and interactions, models are based on **implicit assumptions**, and thus

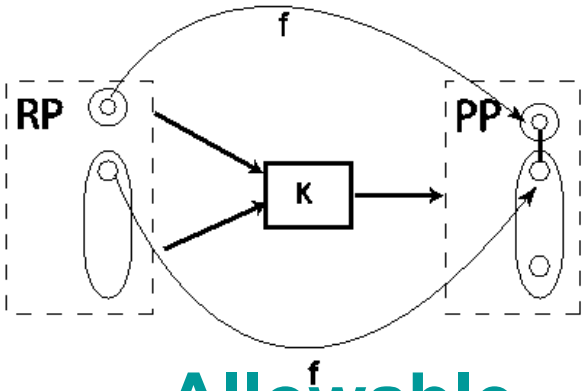
- Limit the number of species and interactions
- Do not allow investigation of different assumptions

Solution

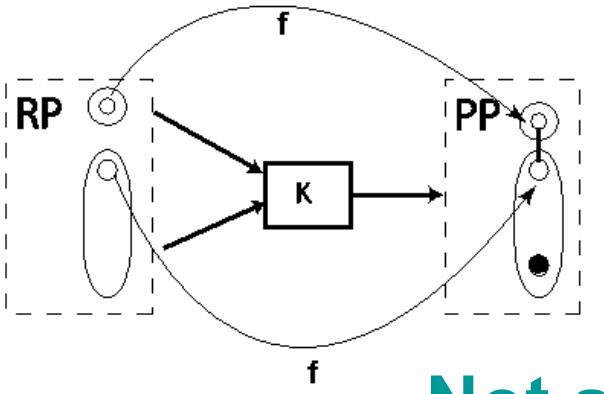
Specify model by **explicit assumptions**, but do not explicitly specify all species and interactions.

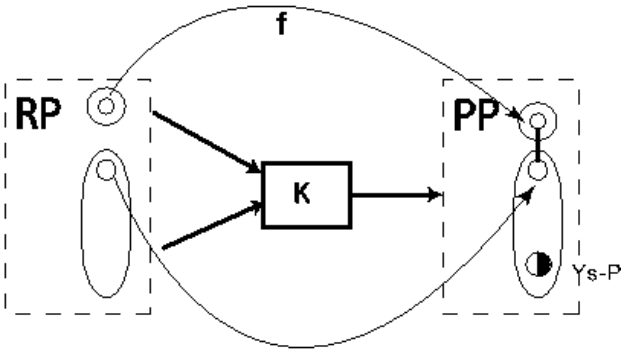
Reaction rule: graph transformation on patterns





Allowable

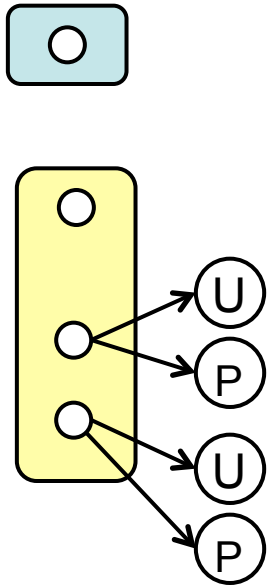




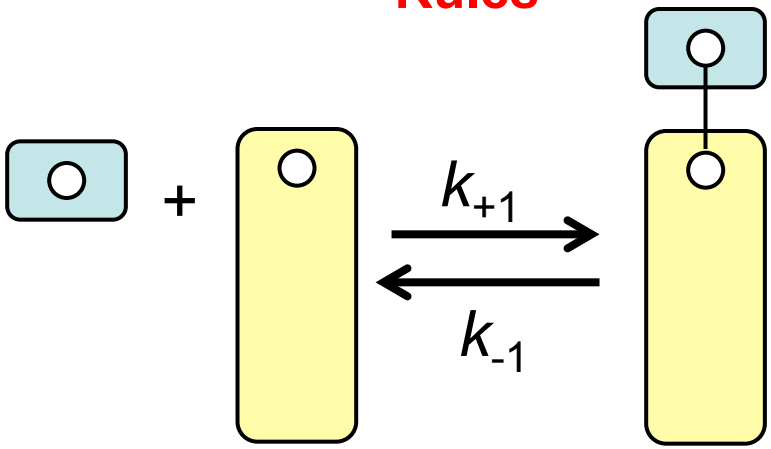
Not allowable

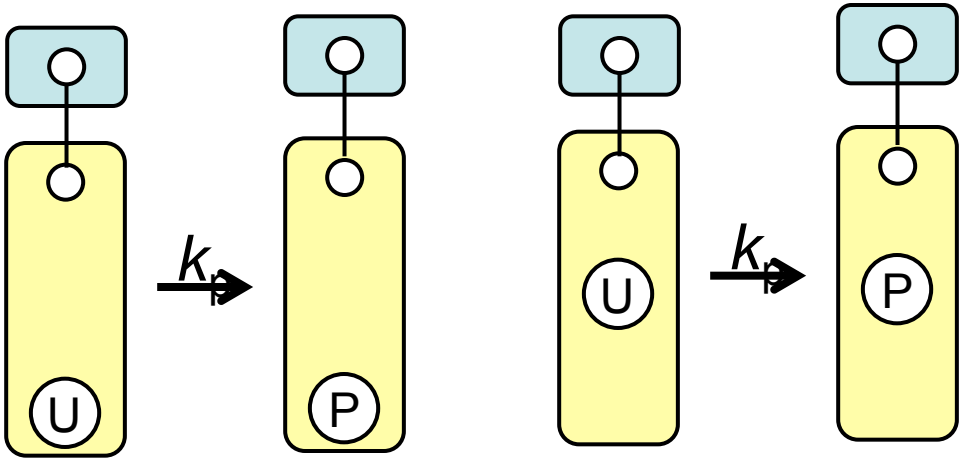
Molecules, components and rules

Molecules, binding sites, components and states

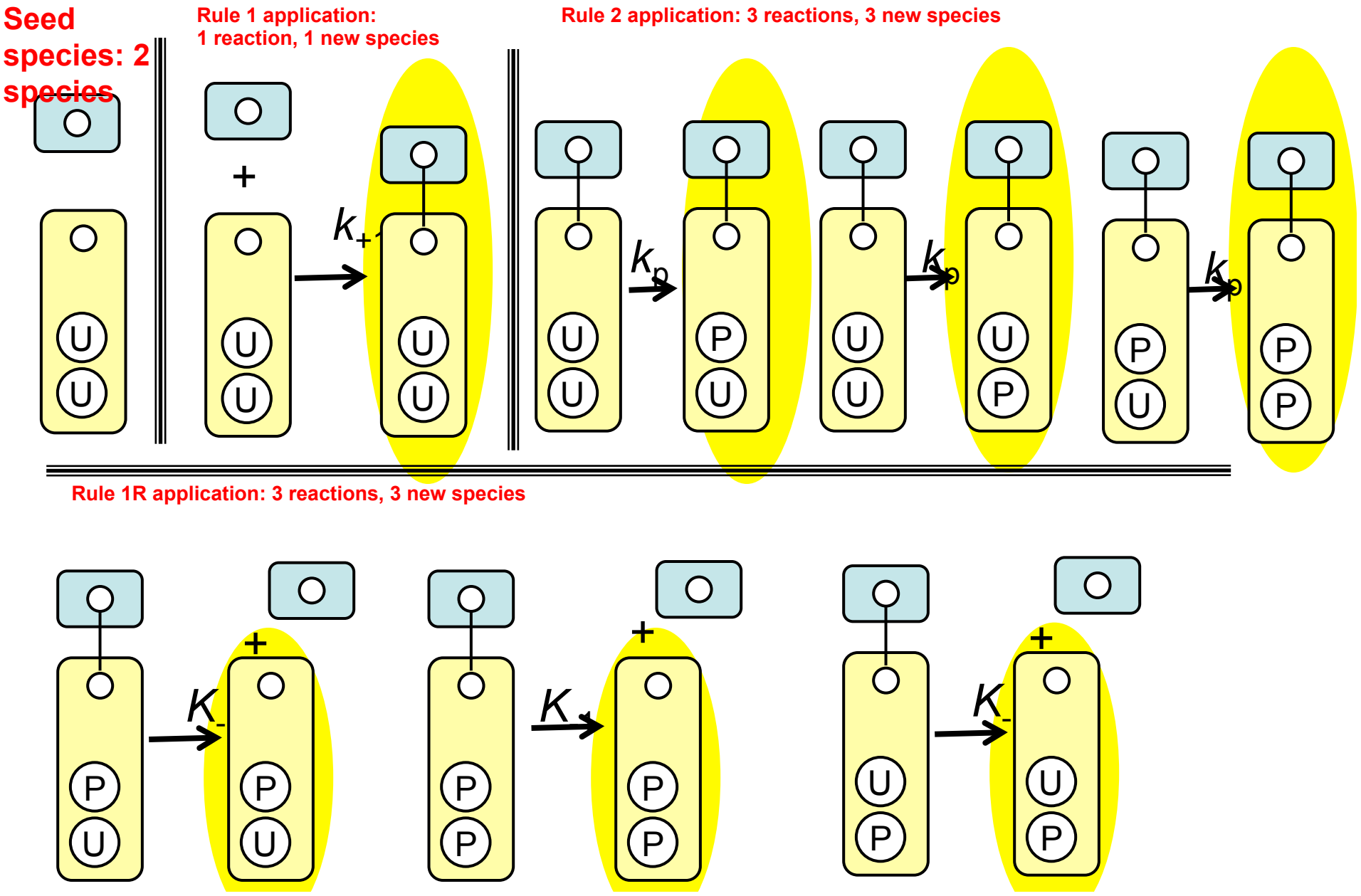


Rules

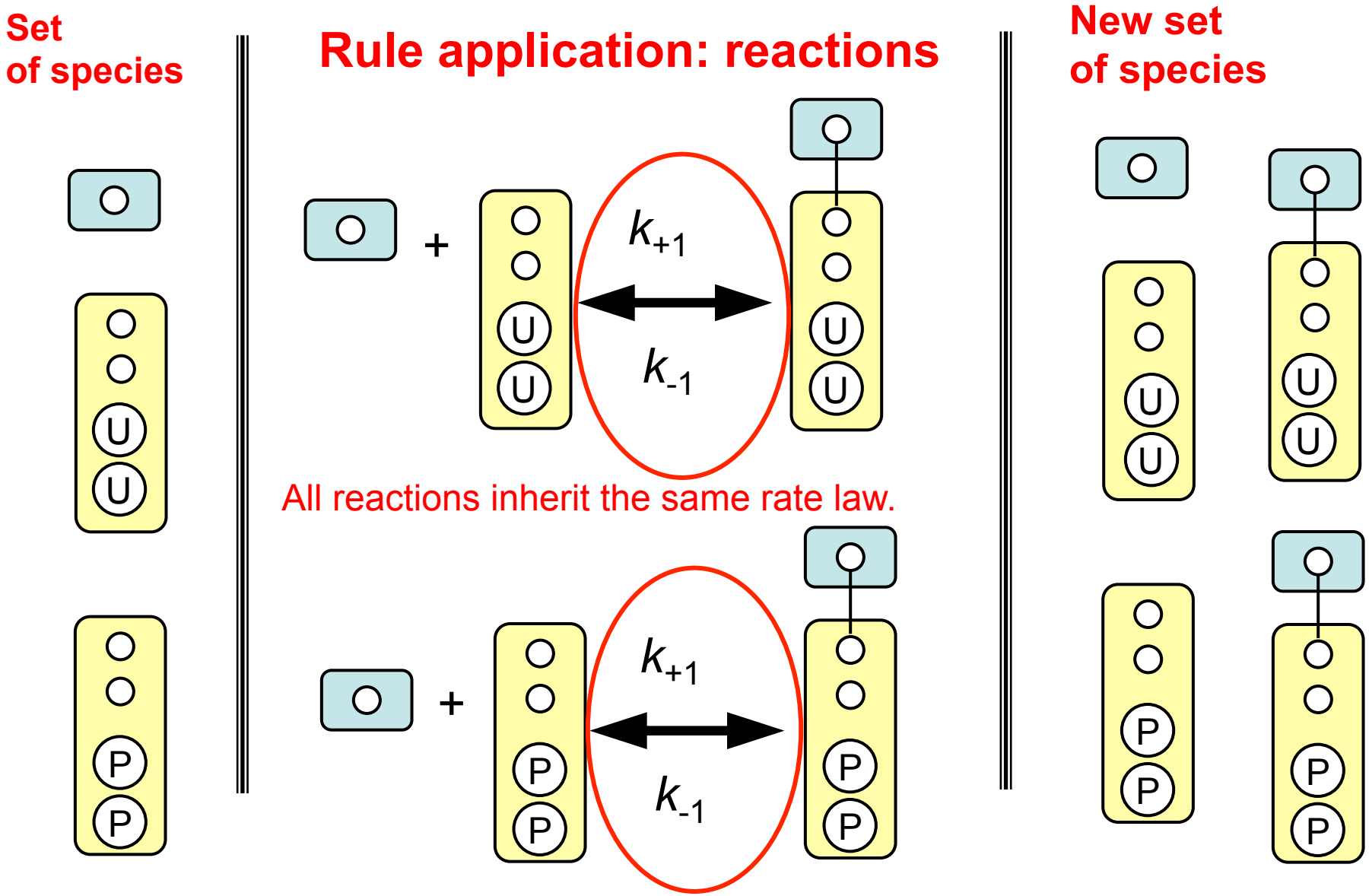




Rules generate reactions and species



Rules generate reactions and new chemical species



Rule-based model generation

Input: initial species S_0

Input: reaction rules $\mathcal{R}$

Rules application 1 $\mathcal{R}(S_0) = R_0, S_1$

Rules application 2 $\mathcal{R}(S_0 \cup S_1) = R_1, S_2$

....

Rules application n $\mathcal{R}(S_n) = R_{n+1}, S_{n+1}$

Termination Terminate if $S_n = S_{n+1}$

Model: species S_n and reactions R_{n+1}

Evolution of modeling

- Model variables described by mathematical equations
- Model species and interactions described by reaction networks - can be reduced to math equations
- Model properties of the biological systems, described by rules – can be reduced to reaction networks

Principles of rule-based modeling

- Based on the **assumption of proteins modularity**: interactions depend on a limited set of features of signalling molecules.
- **Logically consistent**: it accounts for all molecular species implied by user-specified activities, potential modifications and interactions of the domains of signaling molecules.
- Number of parameters is equal to the number of model features (**not big**!)
- Parameters **are well-defined**: no lumping, no coupling

BioNetGen Language (BNGL)

Faeder JR et al, *Methods Mol Biol*. 2009

BNGL

Molecules

A(b)

B(a,Y~U~P,loc~Cyt~Nuc)

Patterns

B_tot	B()
B_unbound	B(a)
B_bound	B(a!+)
B_phospho_all	B(Y~P!?)
B_phospho_unbound	B(Y~P)
B_phospho_bound	B(Y~P!+)
A_B_complex	A().B()

Reaction rules

A(b) + B(a) -> A(b!1).B(a!1) p

a bond between two components

B(Y~P) -> B(Y~U) d

Structure of the BNGL file

file.bngl

Define named variables.

```
begin parameters
end parameters
```

Define molecular types.

```
begin molecule types
end molecule types
```

Define initial species and concentrations.

```
begin species
end species
```

Define reaction types.

```
begin reaction rules
end reaction rules
```

Define observables.

```
begin observables
end observables
```

Generate, equilibrate, and simulate network.

```
command1
…
```

Defining parameters

```
[index] parameter_name parameter_value
```

```
begin parameters
  1 R0   1
  2 kp1  0.5
  3 km1  0.1
  4 kp2  1e-3
  5 km2  0.1
  6 p1  10
  7 d1   5
  8 kpA  1-e4
  9 kmA  0.02
end parameters
```

Tips on Units

Consistent use of units in BNG is the user's responsibility. Any consistent set will work, but for switching between ODE and stochastic simulation methods, number per cell is the most convenient.

To get parameters in these units:

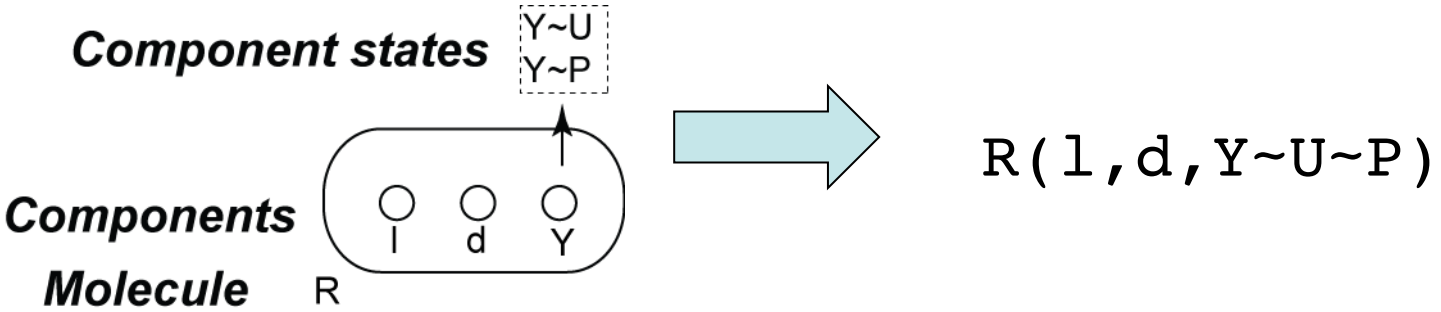
Concentrations: Multiply by $N_a \times V$, where V is $1/\rho_{cell}$ for extracellular ligands, V_{cell} for other components.

Uni-molecular rate constants: No conversion.

Bi-molecular rate constants: Divide by $N_a \times V$, where V is $1/\rho_{cell}$ extracellular ligand binding, V_{cell} reactions involving 1 or more cytoplasmic proteins, and χV_{cell} for reactions occurring in the plasma membrane.

Defining molecules

```
Molecule(comp1~s1~s2,…)
```



Components represent domains of proteins. May be binding sites, have conformational states, or both.

Defining initial species

```
[index] species_string [initial conc.]
```

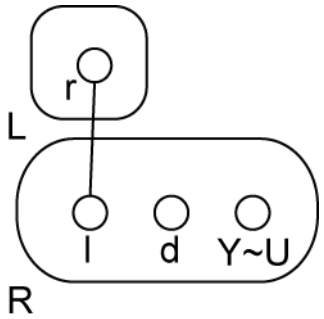
```
begin species
   1  L(r)         L0
   2  R(l,d,Y~U)   R0
   3  A(SH2)       A0
end species
```

Key points

1. No spaces in species strings
2. States for components that take states
3. Initial concentration may be number or parameter

Bonds and complexes

Bonds are indicated by edges in the species graph.
Bonds are indicated by an `!<number>`, where `<number>` is the index of the bond.



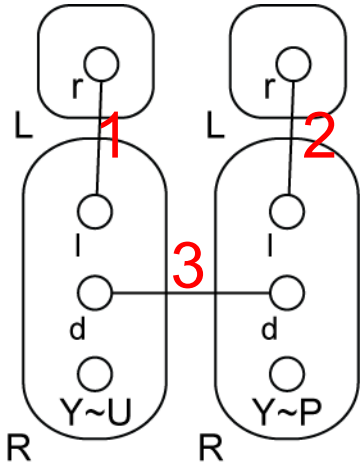
bond

```
L(r!1).R(l!1,d,Y~U)
```

'.' used to join molecules

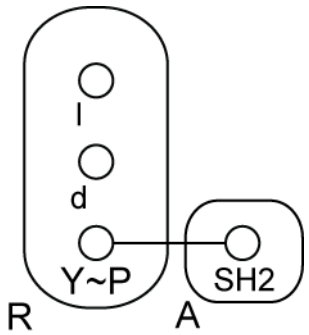
Note: bond index is used only to identify bond endpoints. All bonds are otherwise equivalent.

A more complex example




```
L(r!1).R(l!1,d!3,Y~U).L(r!2).R(l!2,d!3,Y~P)
```

Mixing states and edges

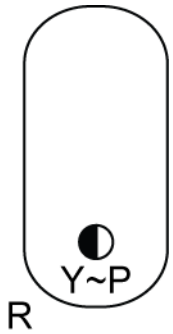



```
R(l,d,Y~P!1).A(SH2!1)
```

Patterns

Definition. A **pattern** is a graph in which some elements may be unspecified or may represent a range of values.

Patterns are used to select sets of species with common attributes on which to perform operations.



`R(Y~P!?)`

'?' indicates that bonding state is unspecified

Examples of matches

`R(l,d,Y~P)`

`L(r!1).R(l!1,d,Y~P)`

`R(l,d,Y~P!1).A(SH2!1)`

`R(l,d!1,Y~P).R(l,d!1,Y~P)`

`R(l,d!1,Y~P).R(l,d!1,Y~P)`

} **Two matches for same species**

Pattern conventions

1. Any unspecified component may take on any internal or binding state. In `R(Y~P!?)` both l and r are unspecified.
2. If a component is specified without an internal state, it may take on any internal state.
3. There are two edge wildcards:

   `!?` means may or may not be bound

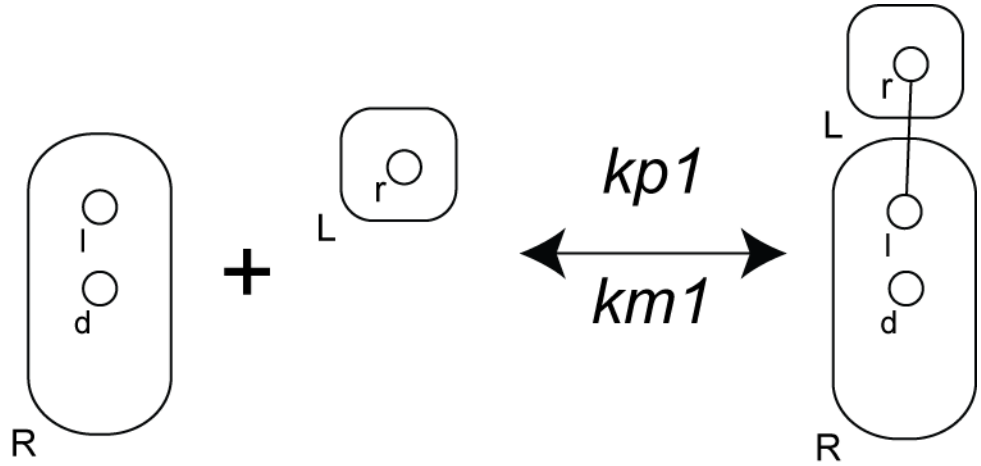
   `!+` means one or more additional bonds must be present

Reaction rules

Reaction rules consist of reactant and product patterns that are used to specify a transformation

```
reac_1 + … + reac_M -> prod_1 + … + prod_N k_f

reac_1 + … + reac_M <-> prod_1 + … + prod_N k_f, k_r
```




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

Application of the reaction rule

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

Forward

```
L(r) + R(l,d,Y~U) -> L(r!1).R(l!1,d,Y~U) kp1
L(r) + R(l,d,Y~P) -> L(r!1).R(l!1,d,Y~P) kp1
L(r) + R(l,d,Y~P!1).A(SH2!1) -> \
                                L(r!2).R(l!2,d,Y~P!1).A(SH2!1) kp1
```

Reverse

```
L(r!1).R(l!1,d,Y~U) -> L(r) + R(l,d,Y~U) km1
```

Observables

Definition. An **observable** is the sum of concentrations over a set of species selected by one or more patterns.

```
[type of observable] <observable name> patt1, …, patt_N
```

```
Rdim   R(d!+)
```
Selects receptors with dimerization domain bound

```
Rphos  R(Y~P!?)
```
Selects receptors with phosphorylated tyrosine

Observables

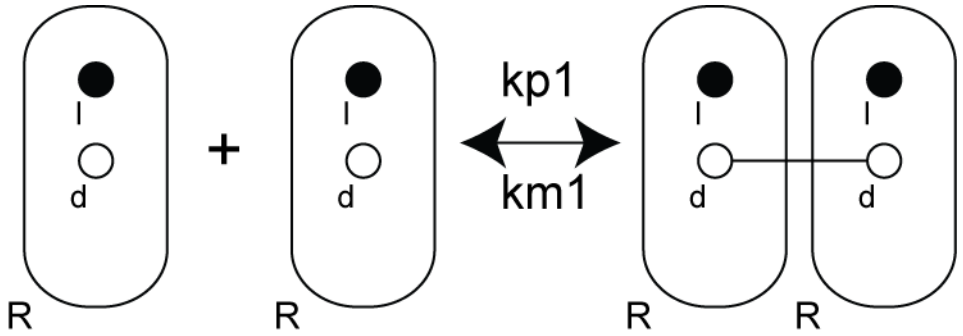
```
begin observables
  Molecules R_dim  R(d!+)
  Molecules R_phos R(Y~P!?)
  Molecules A_R    A(SH2!1).R(Y~P!1)
  Molecules A_phos A(Y~P!?)
end observables
```

`Molecules` keyword indicates that each species concentration is multiplied by the number of matches.

`Species` keyword indicates that concentration of each species is only added once.

Example of symmetric reaction

```
# Receptor aggregation
 2  R(l!+,d) + R(l!+,d) <-> \
        R(l!+,d!2).R(l!+,d!2) kp2, km2
```



Symmetry of reactant `R` molecules is preserved under this transformation. Rate constants are multiplied by factor of 1/2 to give correct rate, assuming `kp2` and `km2` are for single bond.

Commands

```
generate_network({overwrite=>1});
```

Apply reaction rules iteratively to generate species and reactions.

```
writeSBML();
```

Write reaction network to SBML Level 2 file.

```
simulate_ode({t_end=>5,n_steps=>50});
```

Solve ODE's to obtain time course for species concentrations and observables.

```
simulate_ssa({t_end=>5,n_steps=>50});
```

Solves using Gillesbie stochastic algorithm

See tutorial file for more details on command parameters.

VCell export

writeSBML()

#%VC% mergeReversibleReactions
#%VC% speciesRenamePattern("\." , "_") #replace . with _
#%VC% speciesRenamePattern("[\(,][a-zA-Z]\w*", "") #remove any text after (or ,
#%VC% speciesRenamePattern("~|!\d*", "") #remove ~ or ! and any digit after that
#%VC% speciesRenamePattern("\(", "") #remove (
#%VC% speciesRenamePattern("\)", "") # remove)
#%VC% speciesRenamePattern("EGFR", "r") # rename EGFR with r
#%VC% setUnit("all", "default")
#%VC% compartmentalizeSpecies("loc~endo", "3",
"Endosome","EndosomeMembrame")
#%VC% compartmentalizeSpecies("loc~endom", "2", "EndosomeMembrame",
"Cytoplasm")
#%VC% compartmentalizeSpecies("loc~cyt", "3", "Cytoplasm","Membrane")
#%VC% compartmentalizeSpecies("loc~cytm", "2", "Membrane", "Extracellular")
#%VC% compartmentalizeSpecies("loc~ext", "3", "Extracellular", "")

Output

```
BioNetGen version 2.0.19+
Reading from file example1.bngl
Read 13 parameters.
Read 3 species.
Read 4 observable(s).
Adding P as allowed state of component Y of molecule R
Adding P as allowed state of component Y of molecule A
Read 7 reaction rule(s).
WARNING: Removing old network file example1.net.
Iteration   0:     3 species      0 rxns 0.00e+00 CPU s 0.00e+00 (4.01e+00) Mb real (virtual) memory.
Iteration   1:     4 species      1 rxns 2.00e-02 CPU s 4.03e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   2:     5 species      3 rxns 1.00e-02 CPU s 4.04e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   3:     6 species      5 rxns 4.00e-02 CPU s 4.06e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   4:     9 species      9 rxns 5.00e-02 CPU s 4.09e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   5:    12 species     20 rxns 1.10e-01 CPU s 4.14e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   6:    14 species     32 rxns 1.10e-01 CPU s 4.17e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   7:    15 species     37 rxns 8.00e-02 CPU s 4.19e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   8:    19 species     42 rxns 8.00e-02 CPU s 4.24e+00 (2.94e+01) Mb real (virtual) memory.
Iteration   9:    21 species     64 rxns 2.30e-01 CPU s 4.28e+00 (2.94e+01) Mb real (virtual) memory.
Iteration  10:    21 species     71 rxns 6.00e-02 CPU s 4.28e+00 (2.94e+01) Mb real (virtual) memory.
```

Toy network has **21 species** and **71 reactions**.

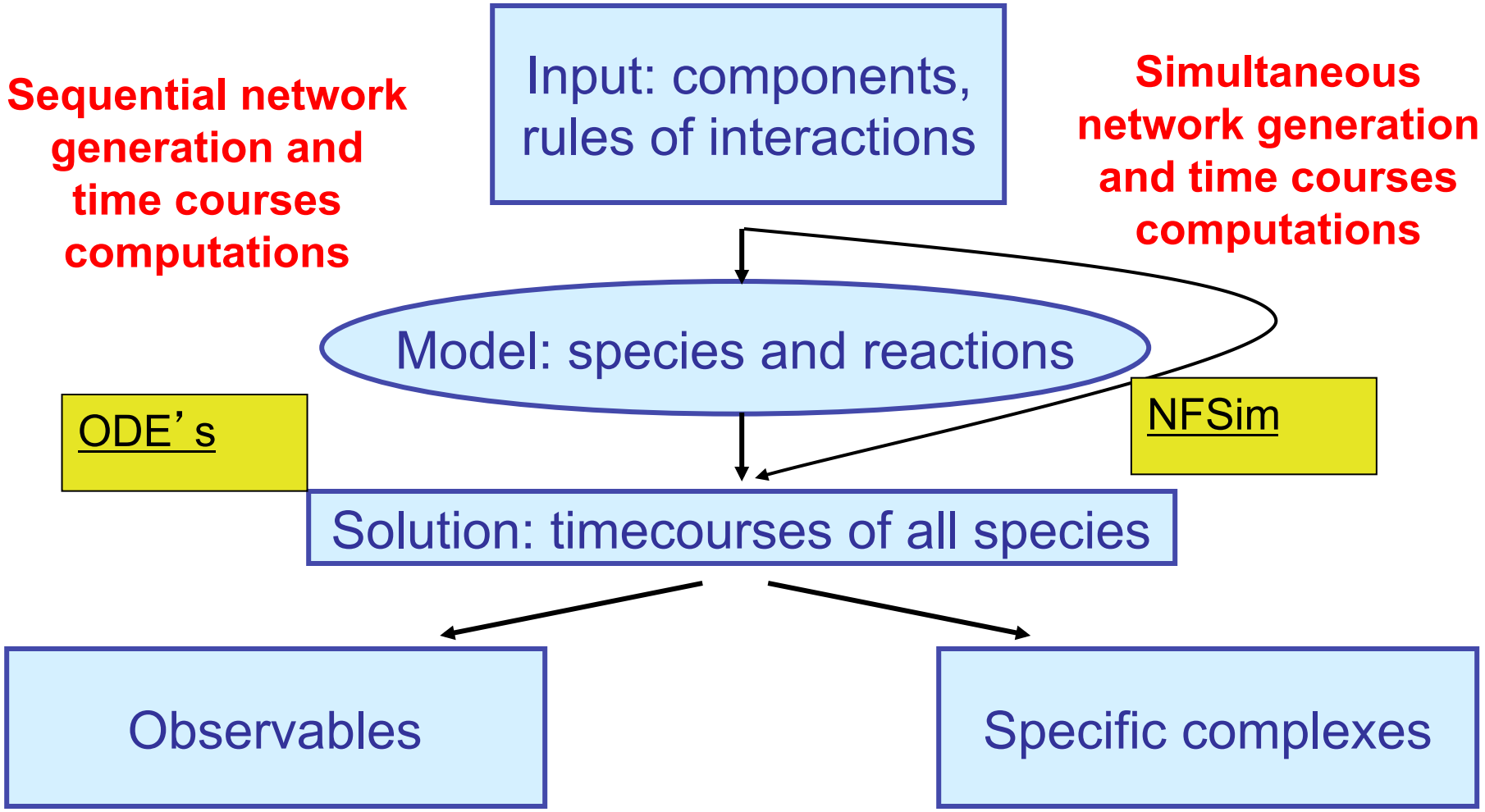
Rule-based modeling software

Blinov et al., Bioinformatics 2004

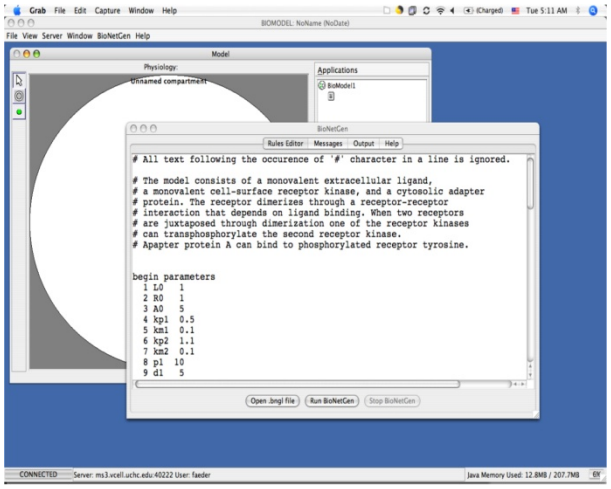
Faeder et al., Methods Mol Biol. 2009

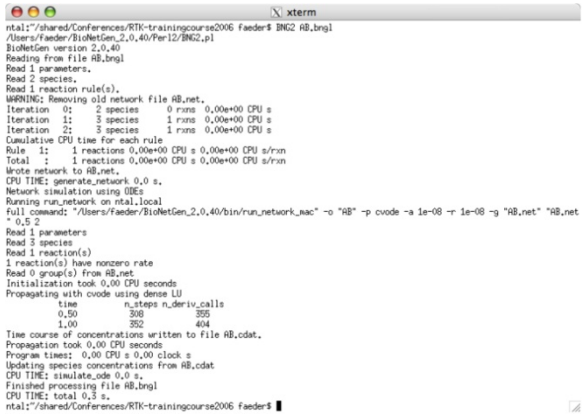
Sneddon et al., Nature Methods 2011

Rule-based modeling



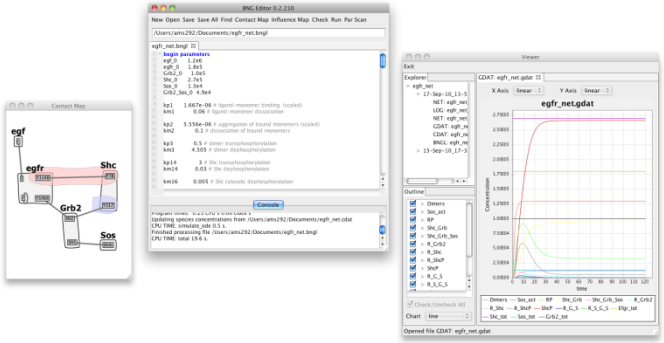
Rule-based modeling software tools





Web interface
(text-based input)

http://vcell.org/bionetgen



Stand-alone
RuleBender

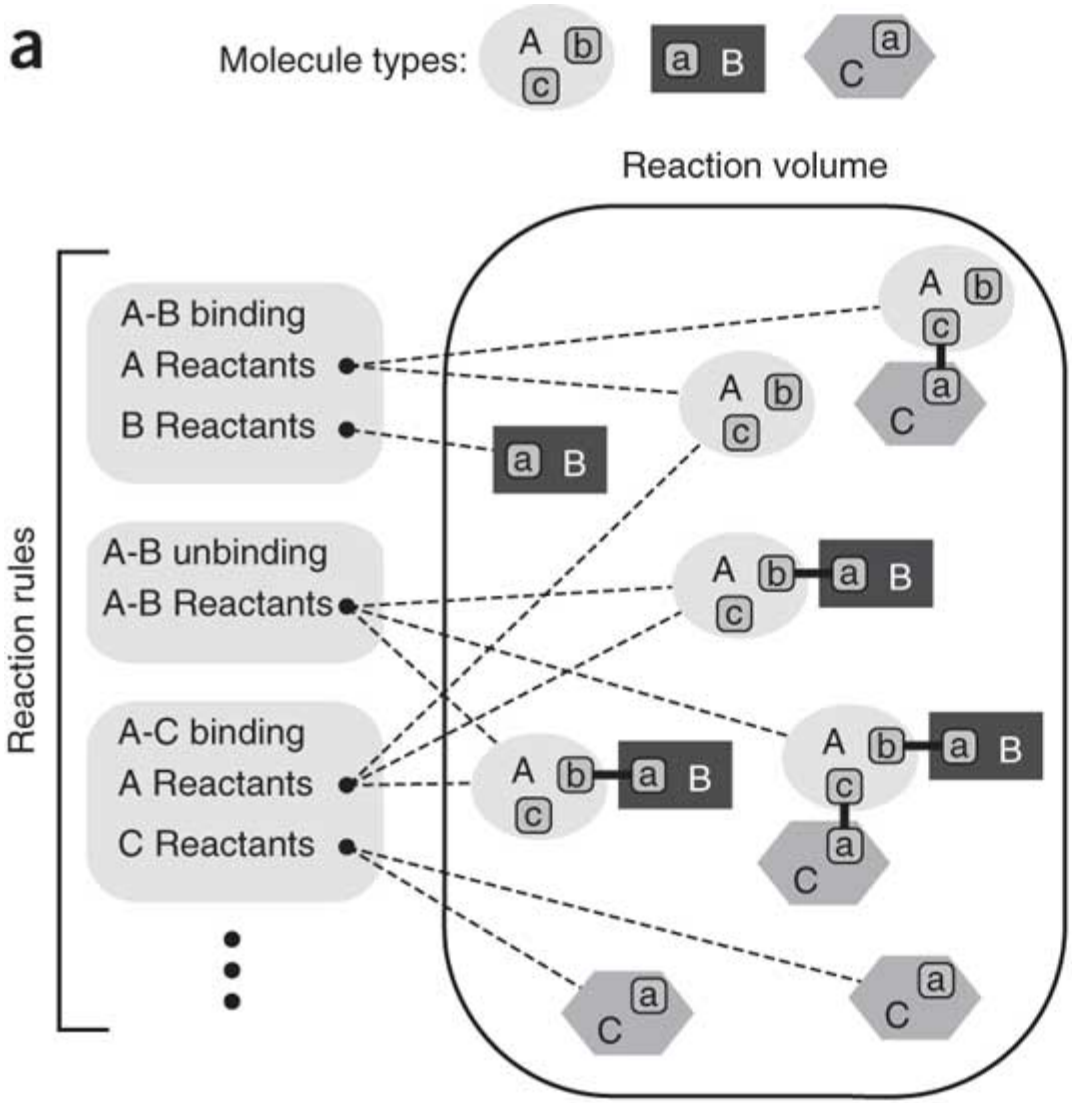
http://bionetgen.org

Network-free
simulation

http://nfsim.org

http://kappalanguage.org/

NFSim



b

```
begin parameters
   # define constant parameters of the model
   k1 0.02
   . . .
end parameters

begin molecule types
   # declare the types of molecules that exist
   A(b,c)
   B(a)
   C(a)
end molecule types

begin observables
   # patterns define basic simulation output
   Molecules AB_complex A(b!1).B(a!1)
end observables

begin reaction rules
   # rules define molecular interactions
   A(b) + B(a) -> A(b!1).B(a!1) k1
   A(b!1).B(a!1) -> A(b) + B(a) k2
   A(c) + C(a) -> A(c!1).C(a!1) k3
   . . .
end reaction rules
```

What do we gain

- New quantitative predictions about specific domains, complexes, and interactions, in contact with kind of experiments biologists do (monitoring levels, knocking out and over-expression of specific domains).
- New qualitative predictions (tracing reaction sequences, dominant molecular species).
- Testing hypotheses about signalling mechanisms, e.g. competitive versus non-competitive protein binding.
- Testing effects of specific genetic manipulations, e.g. effects of knock-outs.