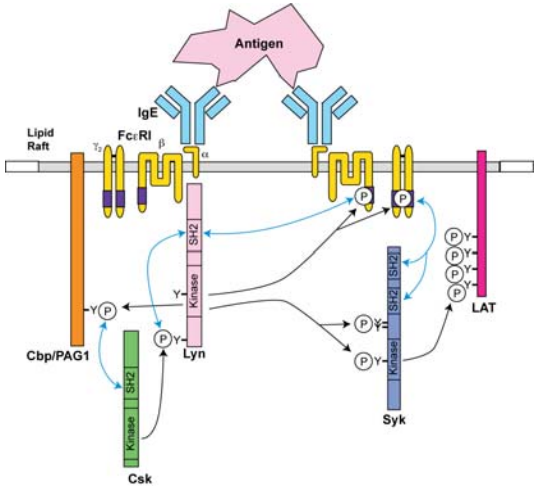
Rule-Based Kinetic Modeling using BIONETGEN



Michael Blinov
Center for Cell Analysis and Modeling
University of Connecticut Health Center

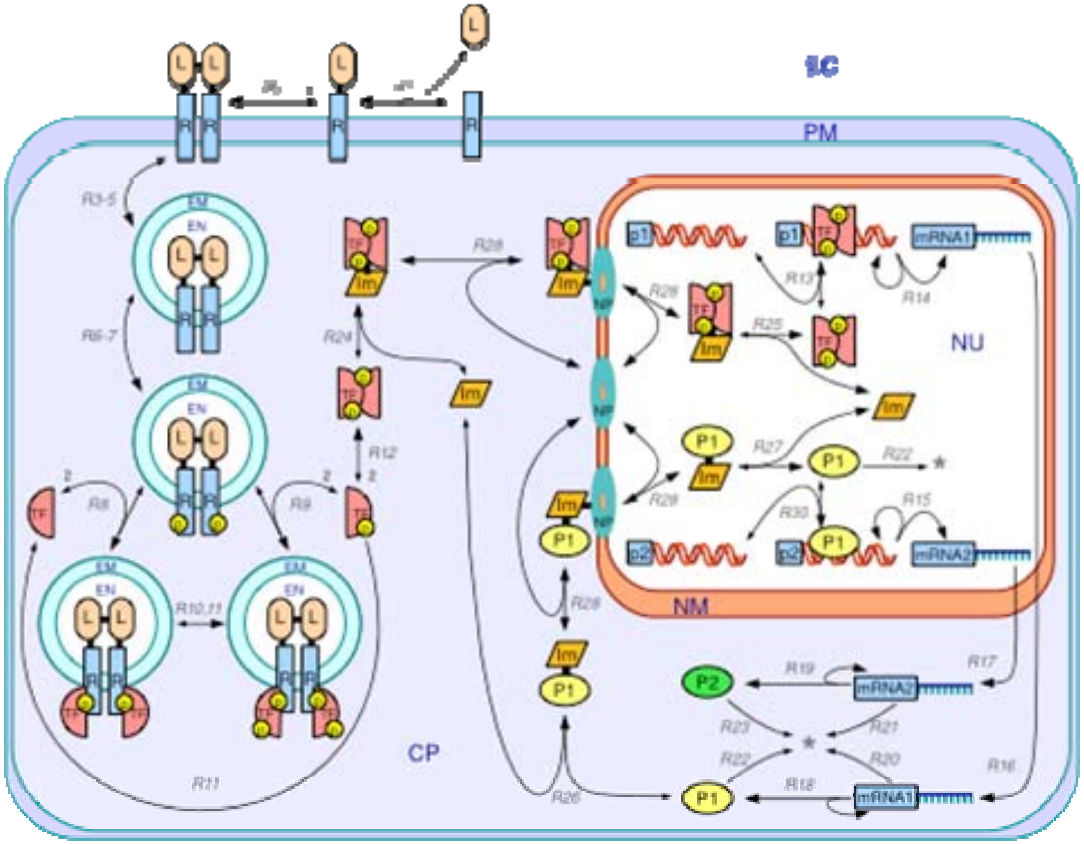
Lily Chylek
Biochemistry, Molecular & Cell Biology Graduate Program
Cornell University

Jim Faeder
Department of Computational and Systems Biology
University of Pittsburgh School of Medicine

Outline

- **Introduction / review of rule-based modeling and simulation**
- Overview of a BIONETGEN model
- Entering and running a simple model
- Some examples

Toward whole cell modeling



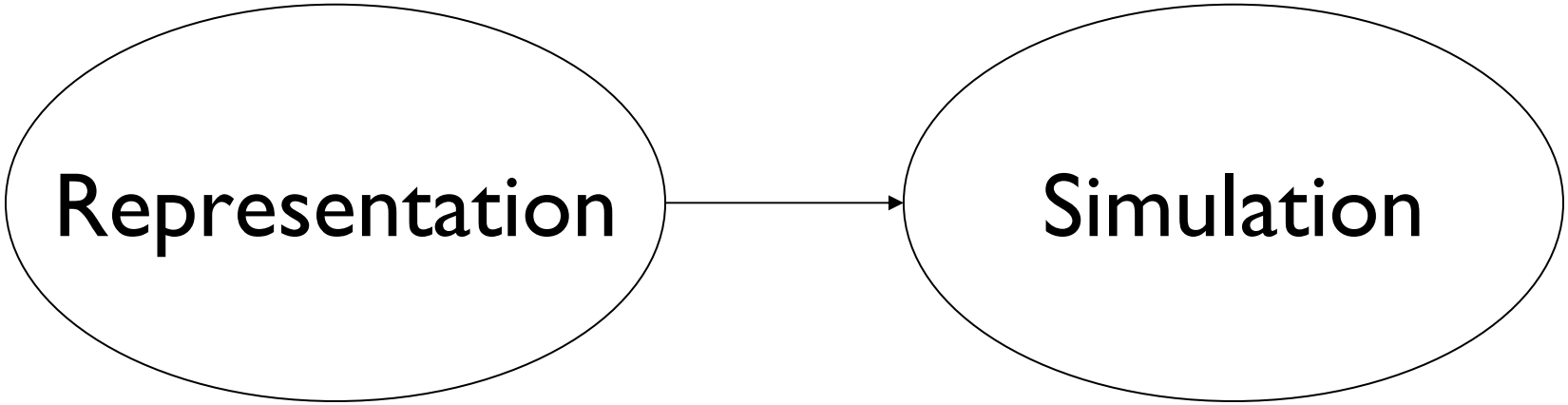
Justin Hogg

Key Questions:

- What parts of cell machinery are critical to the outcome?
- What controls the cell-to-cell variability?
- Can that variability be utilized by the organism to promote survival?

Modeling cell signaling

AIM: Model the biochemical machinery by which cells process information (and respond to it).



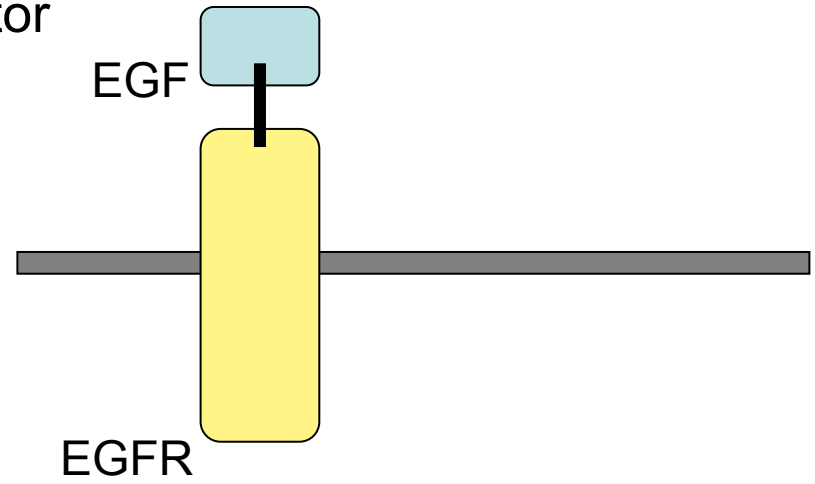
BIONETGEN Language
kappa
etc.

ODE, PDE
Stochastic Simulation Algorithm
Kinetic Monte Carlo
Brownian dynamics

Early events in EGFR signaling

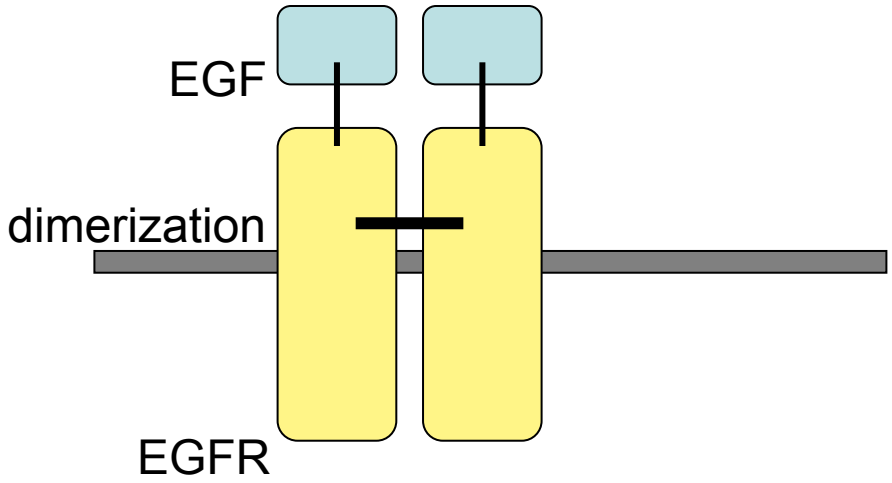
EGF = epidermal growth factor
EGFR = epidermal growth factor receptor

1. EGF binds EGFR



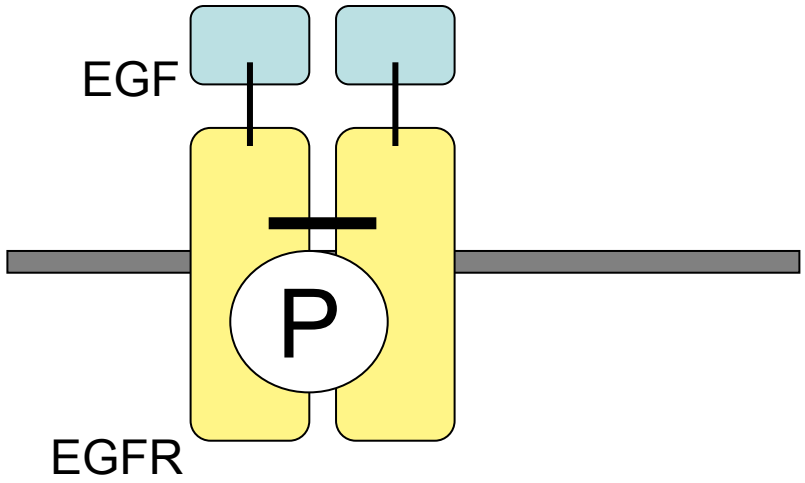
Early events in EGFR signaling

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



Early events in EGFR signaling

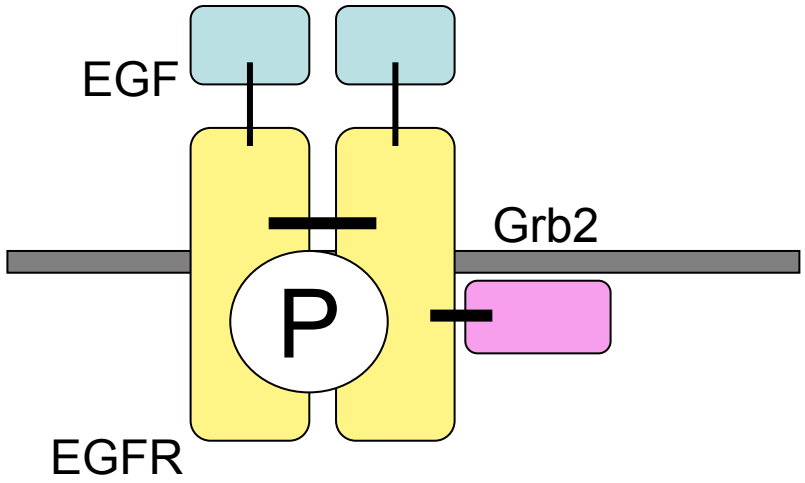
1. EGF binds EGFR
2. EGFR dimerizes
3. **EGFR transphosphorylates**



Early events in EGFR signaling

Grb2 pathway

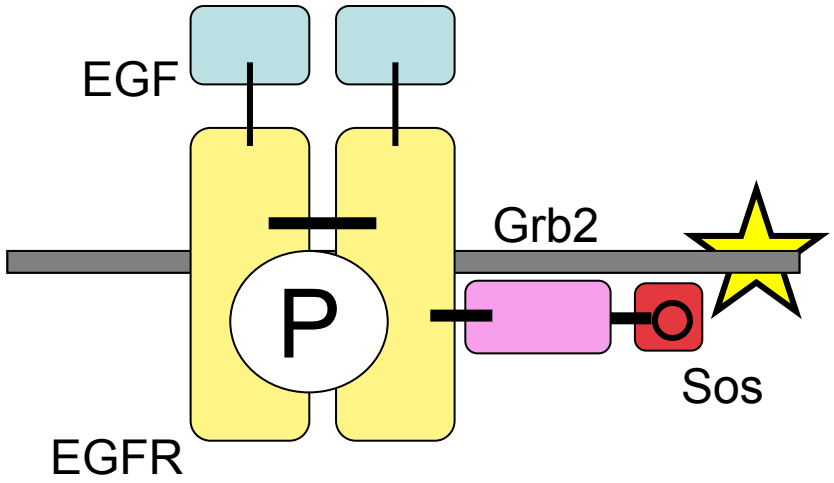
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Grb2 binds phospho-EGFR**



Early events in EGFR signaling

Grb2 pathway

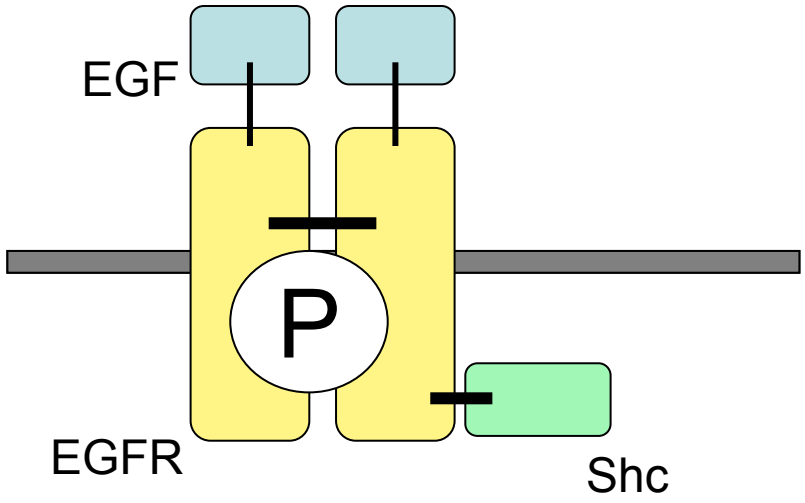
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Grb2 binds phospho-EGFR
5. **Sos binds Grb2 (Activation Path 1)**



Early events in EGFR signaling

Shc pathway

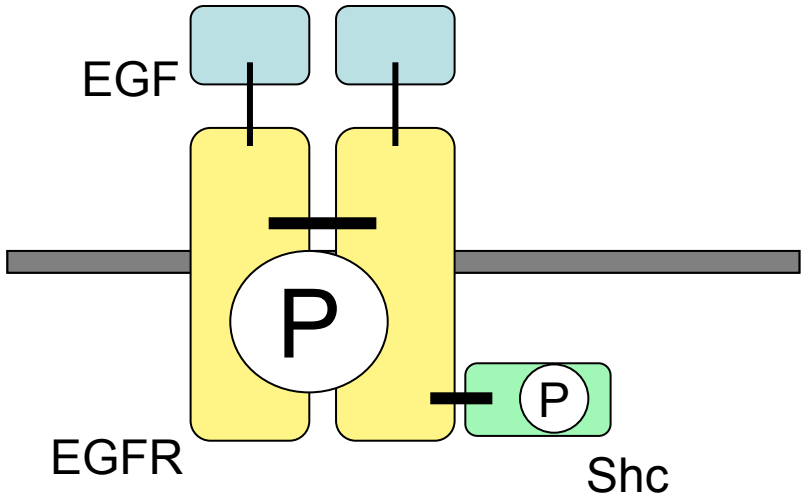
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. **Shc binds phospho-EGFR**



Early events in EGFR signaling

Shc pathway

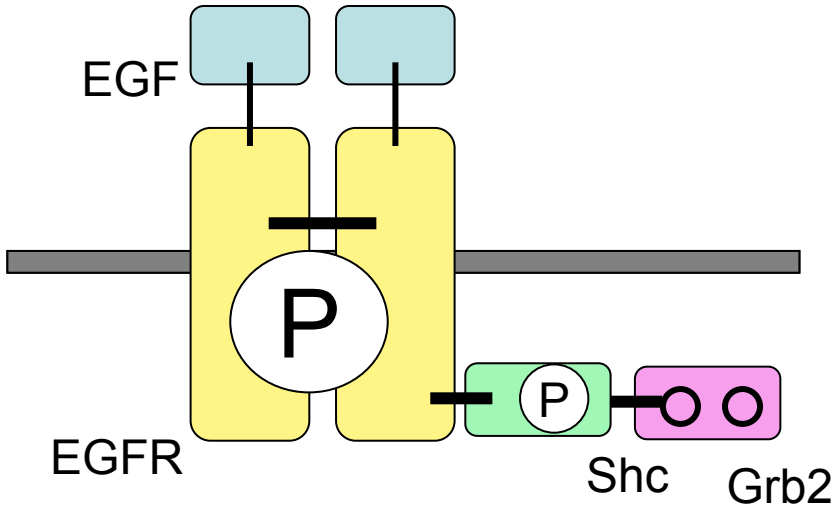
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. **EGFR transphosphorylates Shc**



Early events in EGFR signaling

Shc pathway

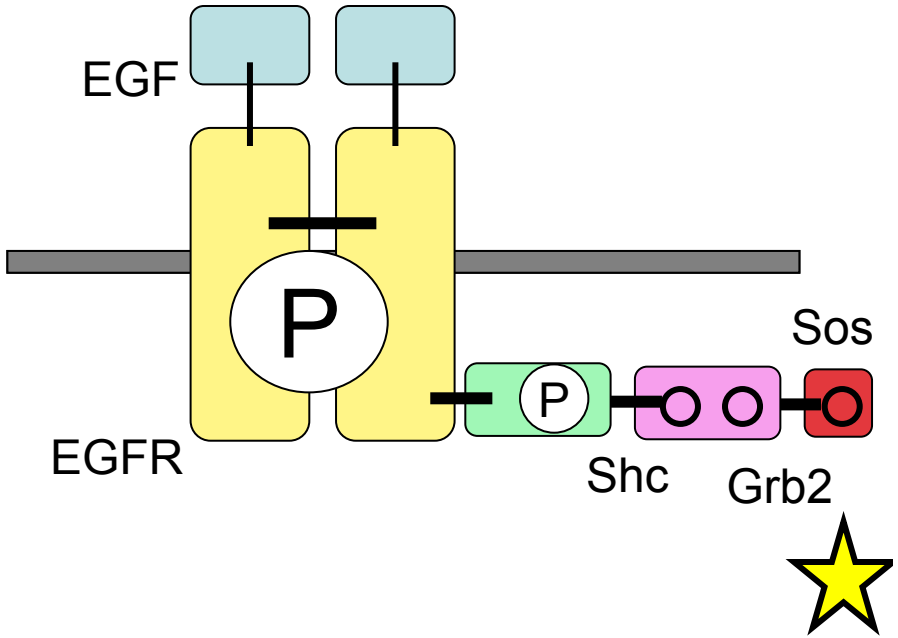
1. EGF binds EGFR
2. EGFR dimerizes
3. EGFR transphosphorylates
4. Shc binds phospho-EGFR
5. EGFR transphosphorylates Shc
6. **Grb2 binds phospho-Shc**



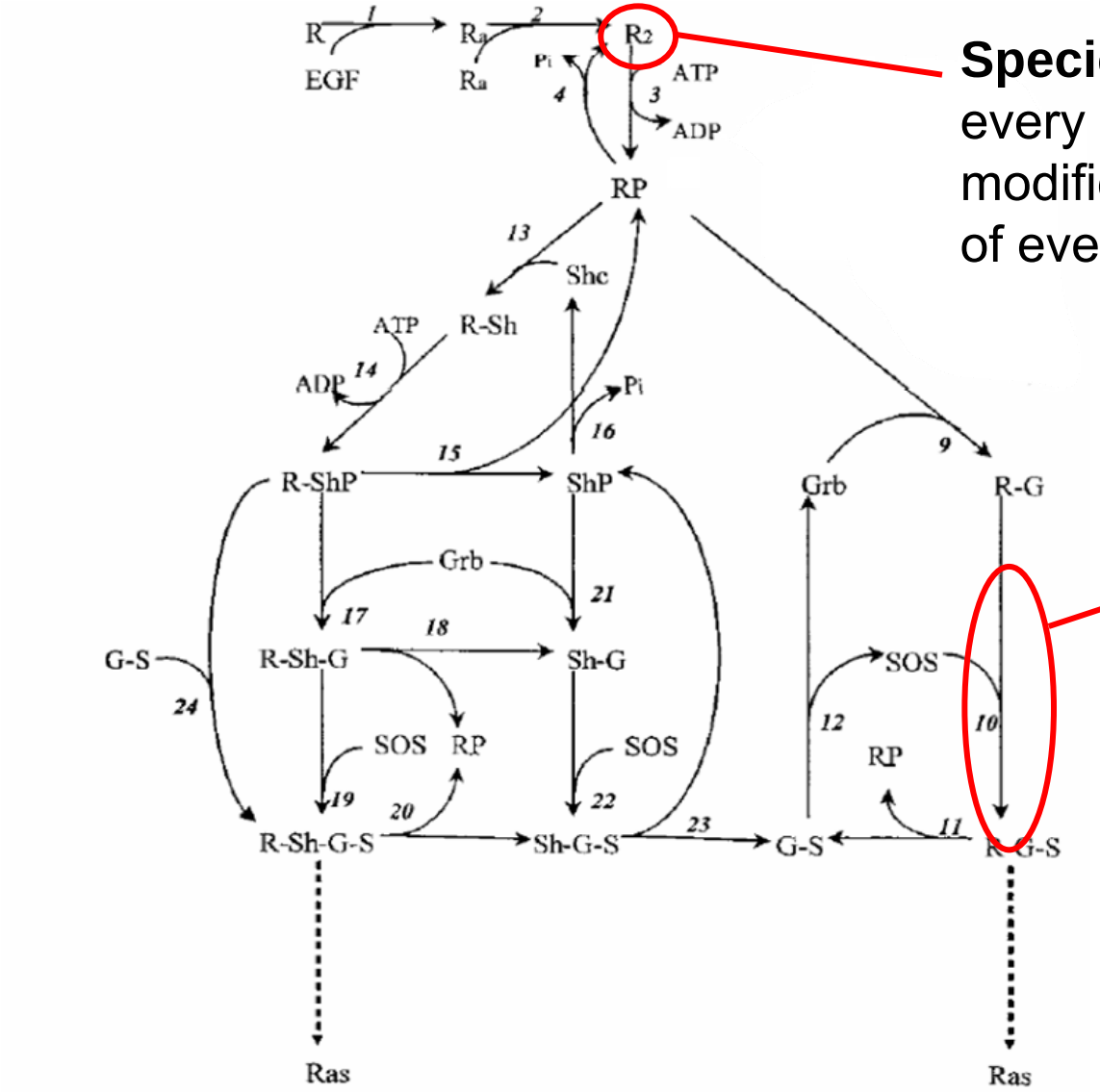
Early events in EGFR signaling

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)**



The next step: write down reaction network



Species: One for every possible modification state of every complex

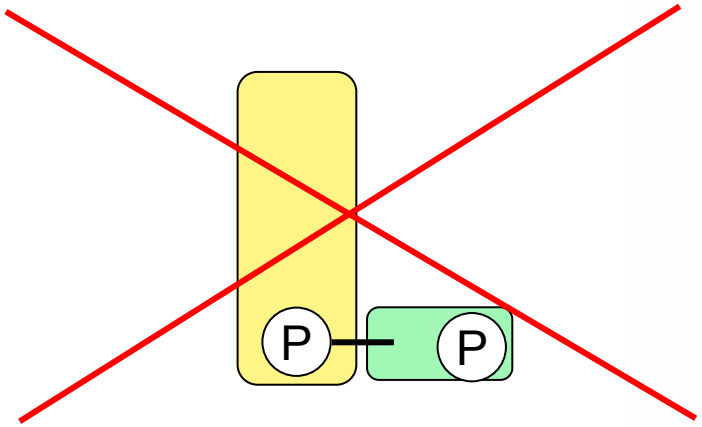
Reactions: One for every transition among species

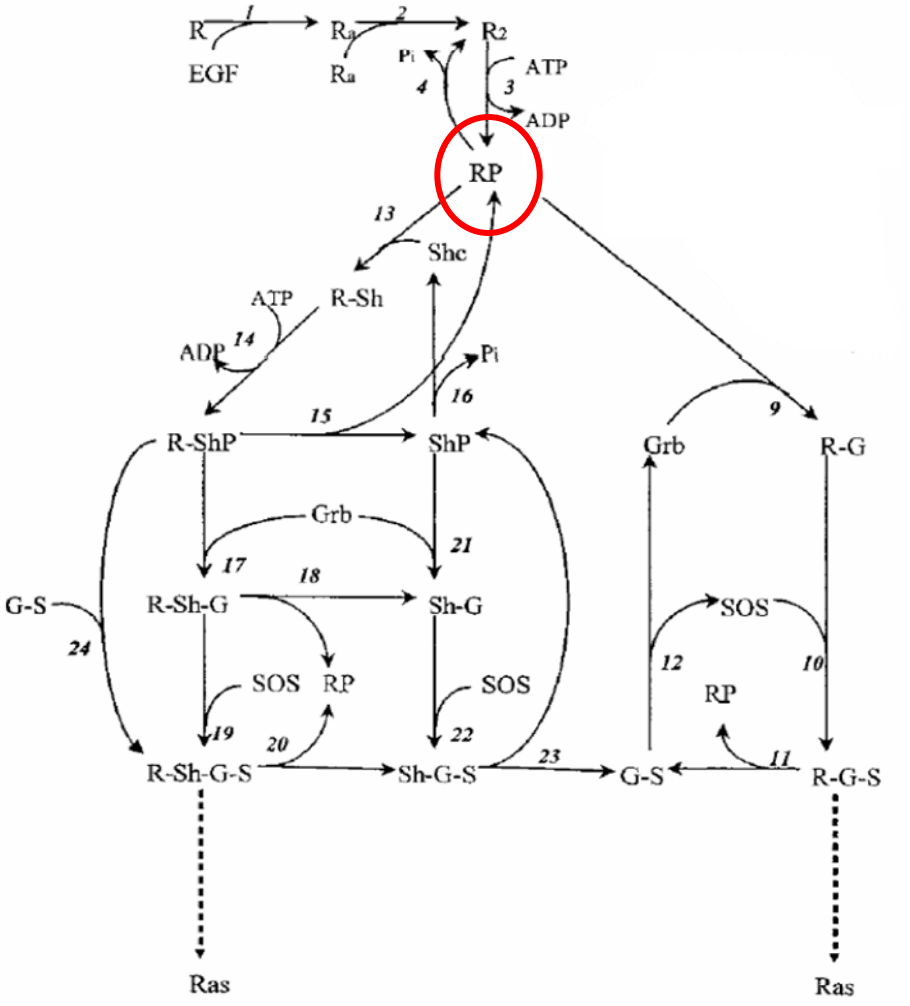
Kholodenko et al., J. Biol. Chem. **274**, 30169 (1999)

Assumptions made:

Phosphorylation inhibits dimer breakup

No modified monomers

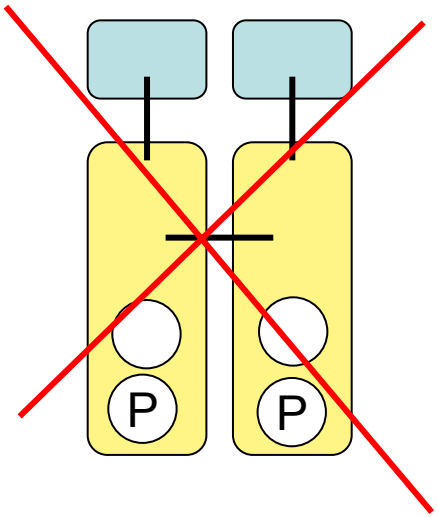


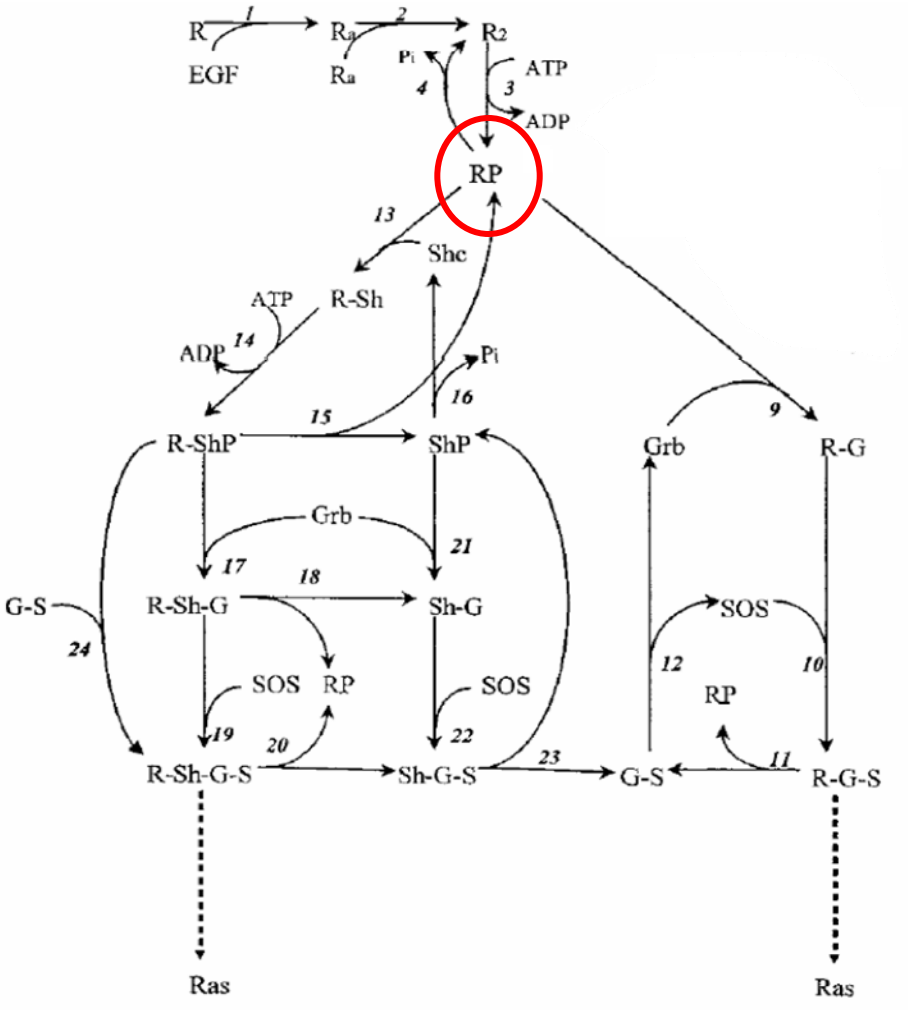


Assumptions made:

Phosphorylation is simultaneous

Same phosphorylation timecourses for all residues

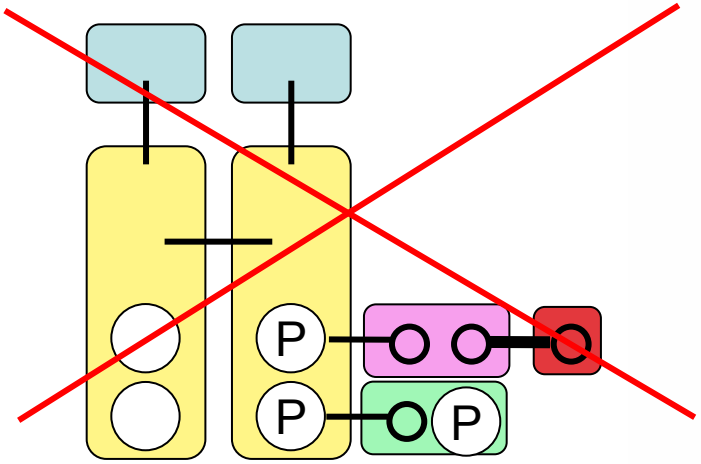


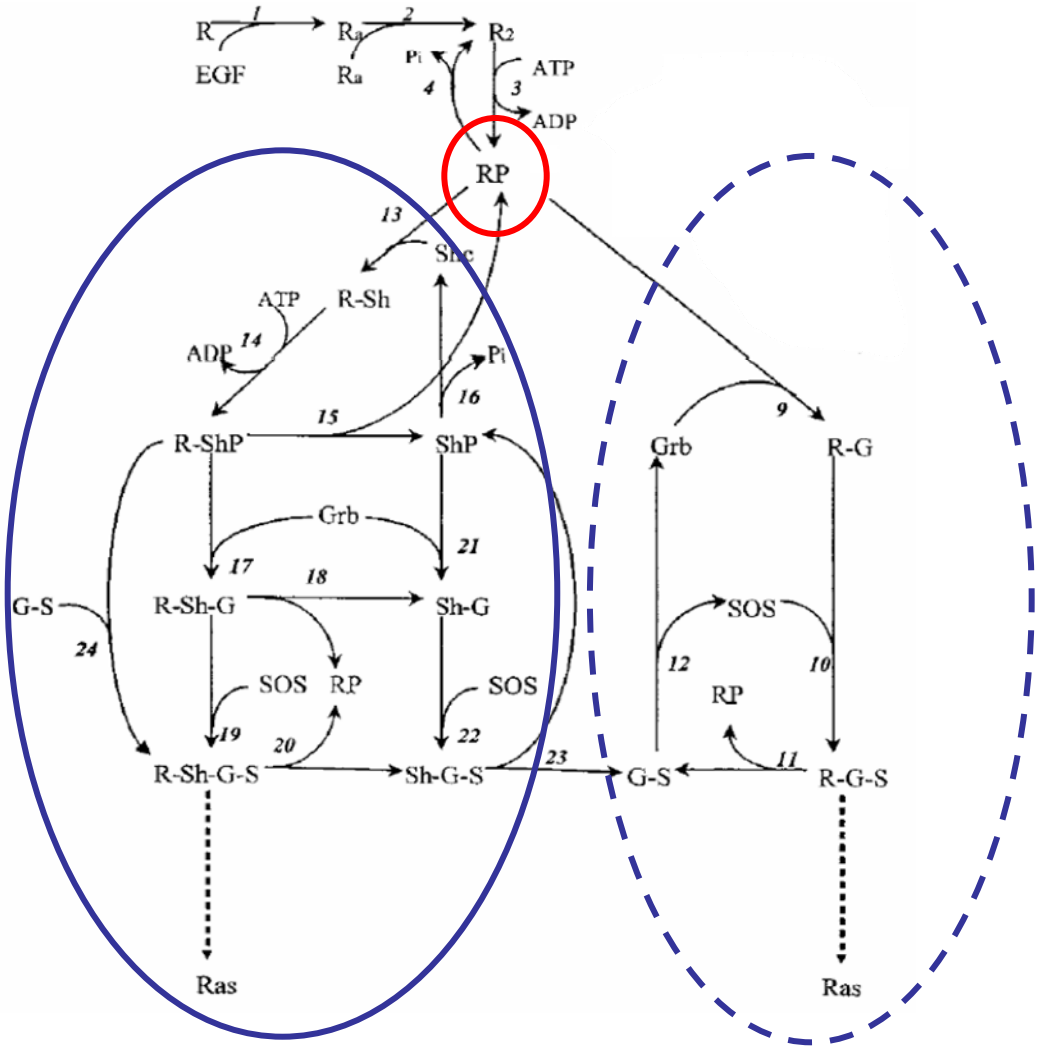


Assumptions made:

Adaptor binding is competitive

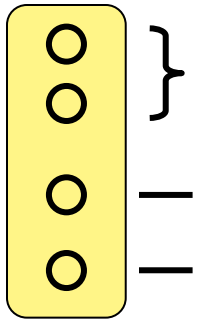
No dimers with more than one site modified





Combinatorial complexity of early events

Monomeric species



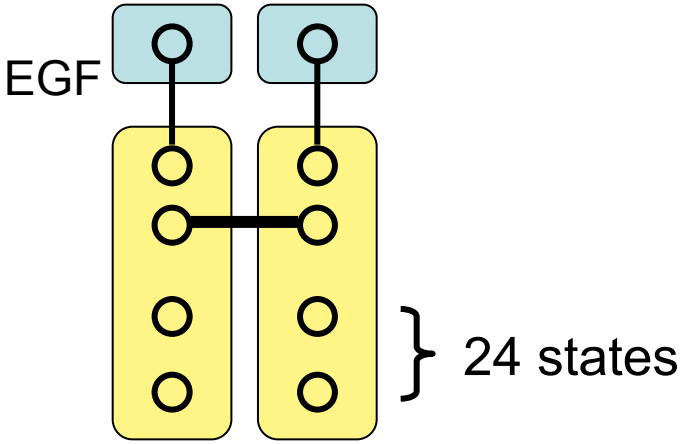
2 states: bound/unbound to ligand

4 states: un-p, p, p-Grb2, p-Grb2-Sos

6 states: ….

EGFR

Dimeric species



48 species

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

Problem

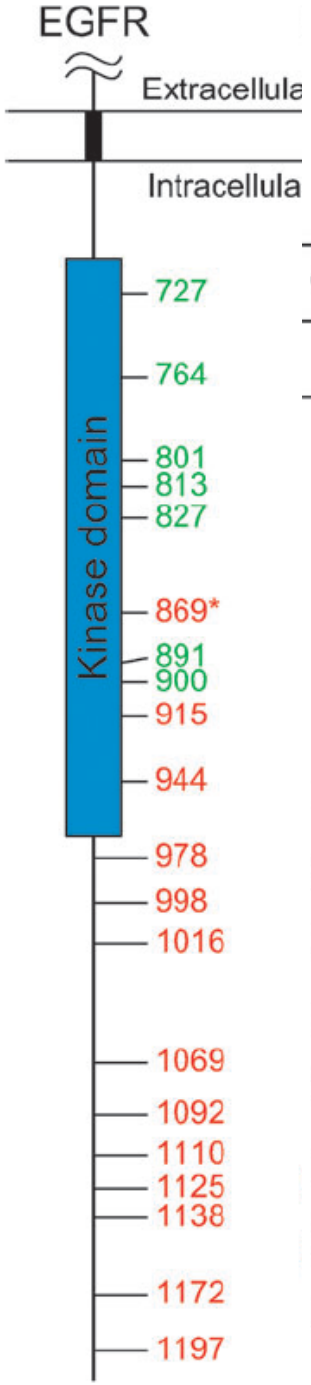
To **explicitly** specify all species and interactions, models are based on **implicit assumptions**, and thus

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

Solution

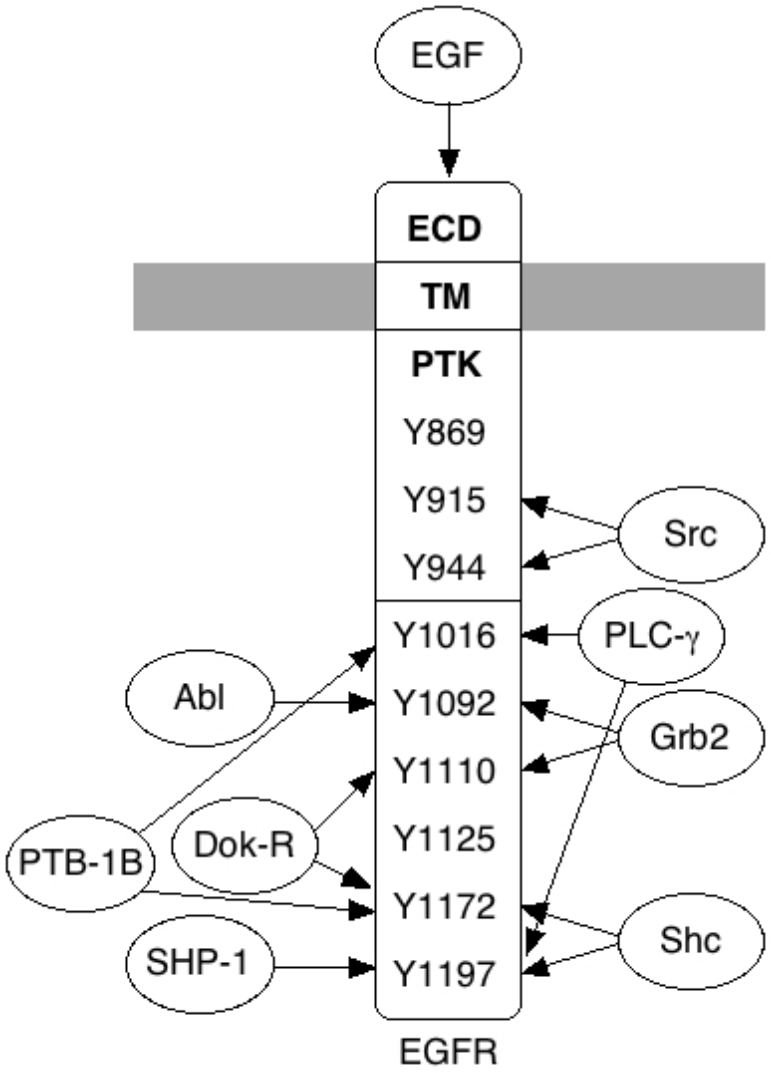
Specify model by **explicit assumptions**, but do not explicitly specify all species and interactions.

Domain-domain interactions



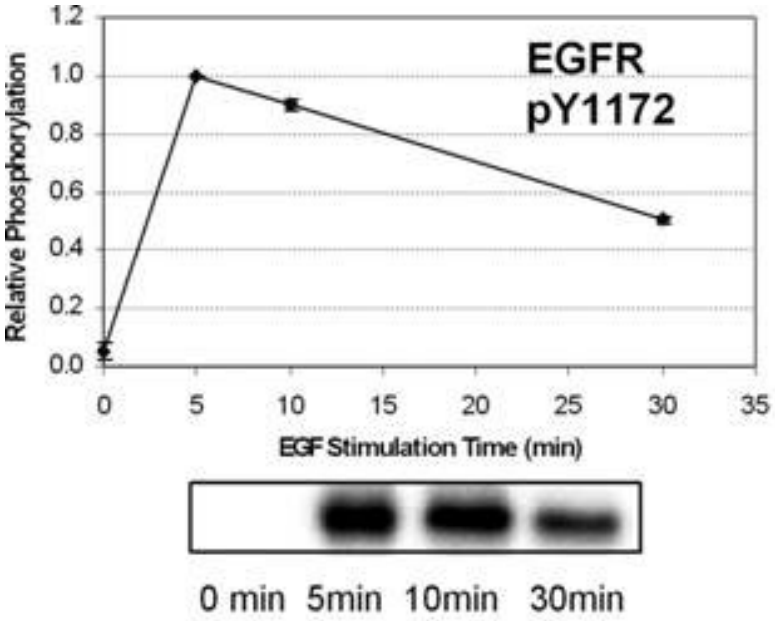
EGFR

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



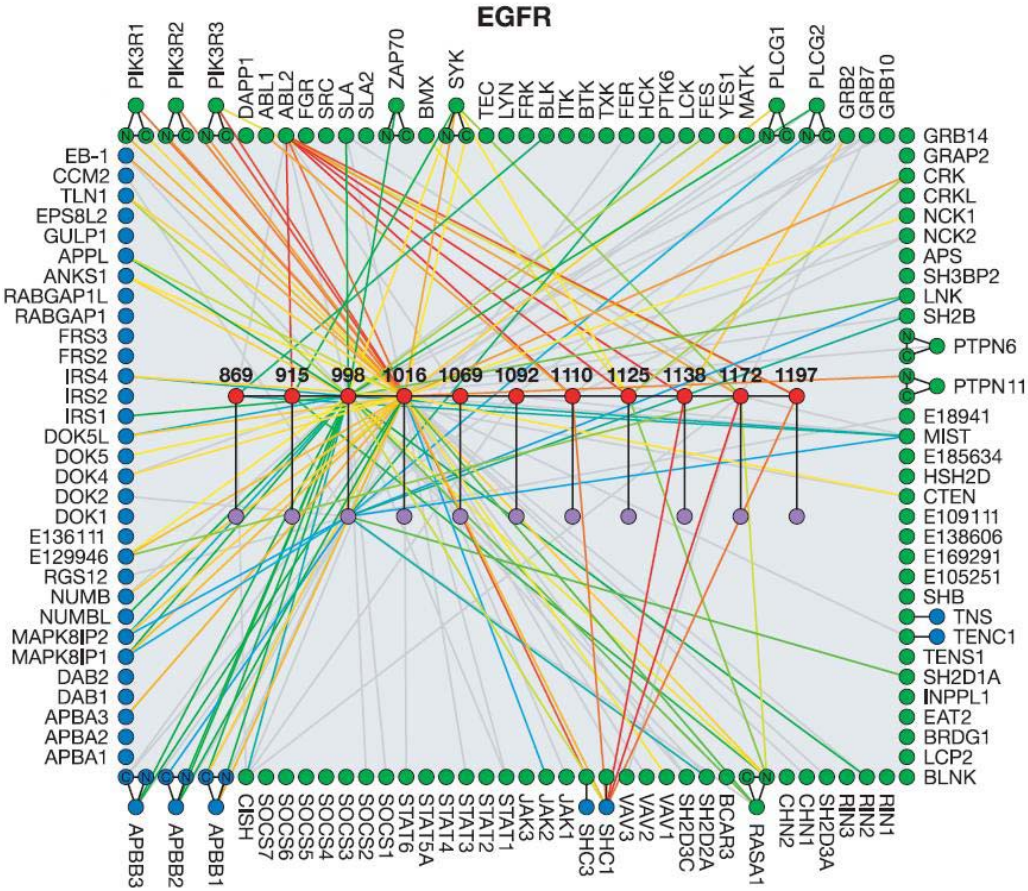
Experimental data

the kinetics of multiple phosphorylation sites



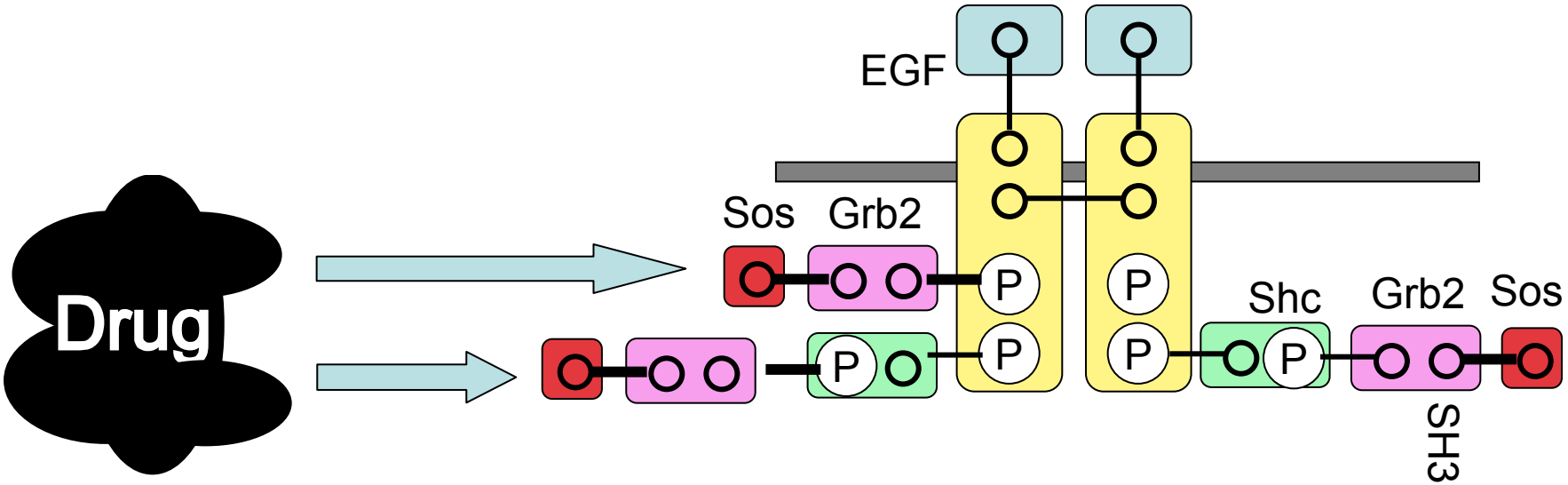
Zhang et al.,
Mol. Cell. Proteomics **4**, 1240 (2005).

affinities for multiple binding partners



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

Big promise???



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

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

Epidermal growth factor receptor (EGFR)

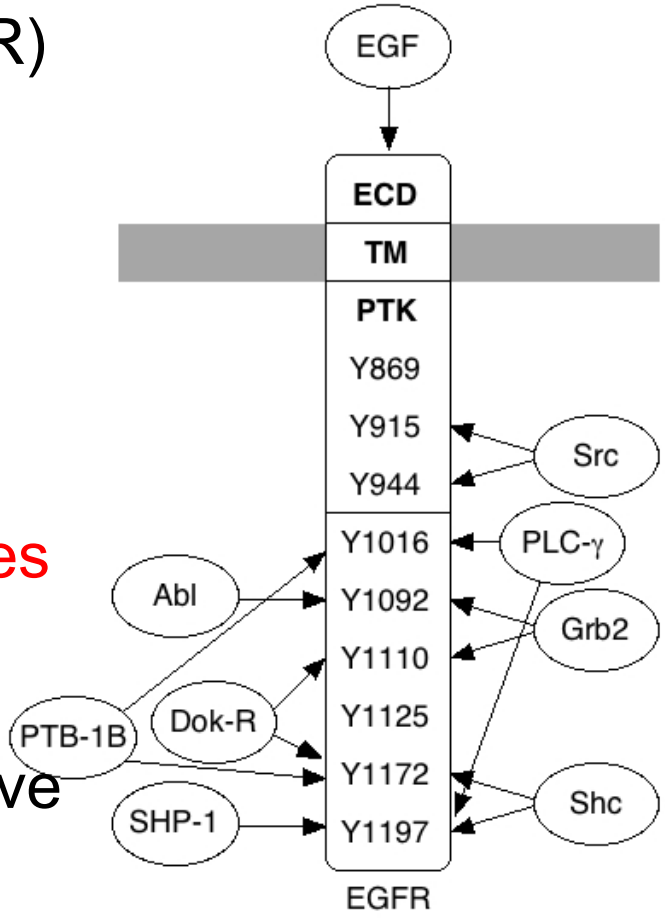
9 sites ⎩ 2^9=512 phosphorylation states

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



Simplest situation: ligand-induced dimerization of receptor-kinases, followed by binding of a protein and its phosphorylation

- May ligand bind to receptor in a dimer?
- May ligand dissociate from receptor in a dimer?
- May ligand dissociate from phosphorylated monomer?
- Can ligand dissociate from phosphorylated receptor in a dimer?
- May dimer break-up when one or two ligand are present?
- Does phosphorylation requires presence of one or two ligands in a dimer?
- Can phosphorylated dimer break-up?

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

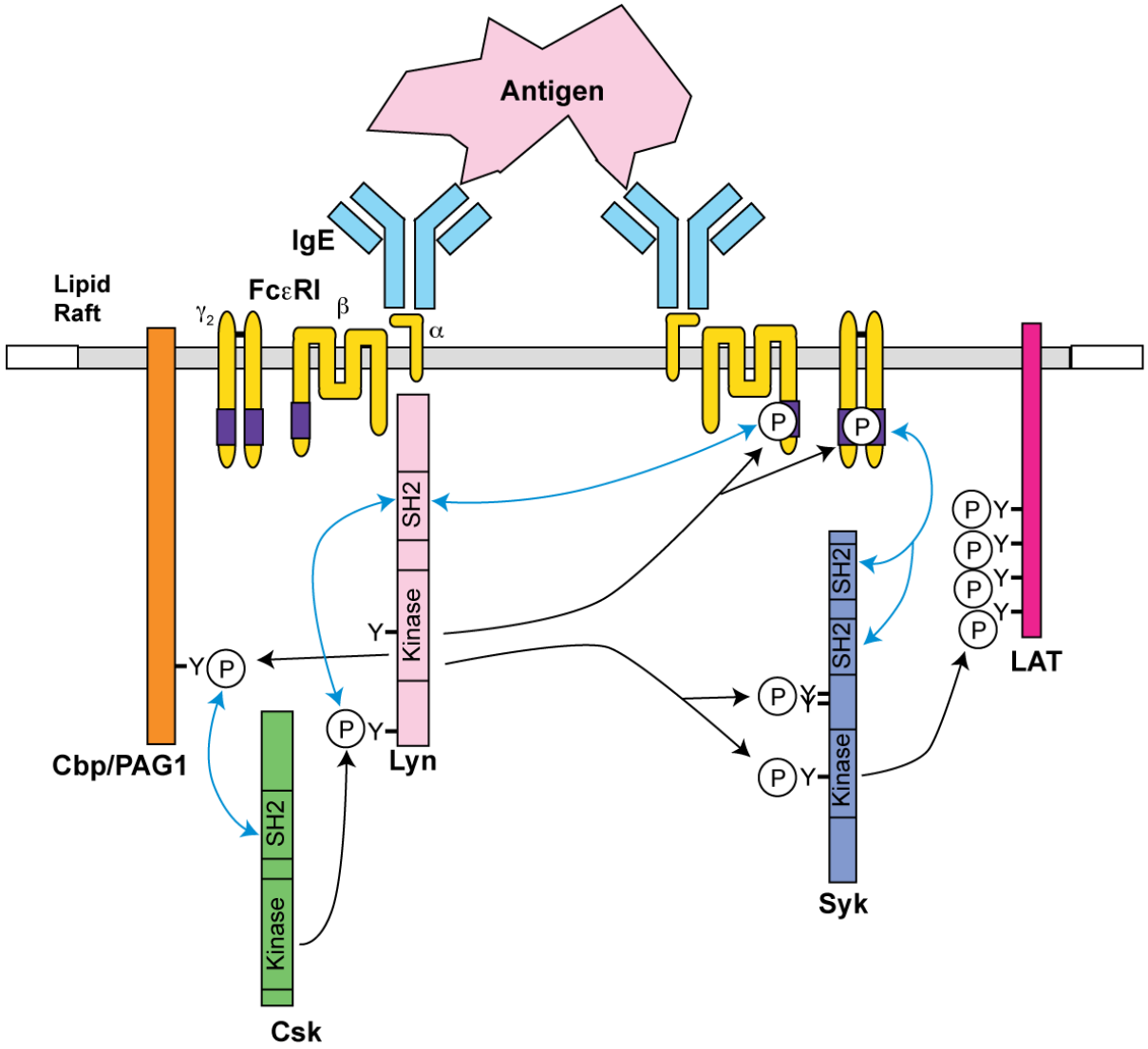
Principles of rule-based modeling

- 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

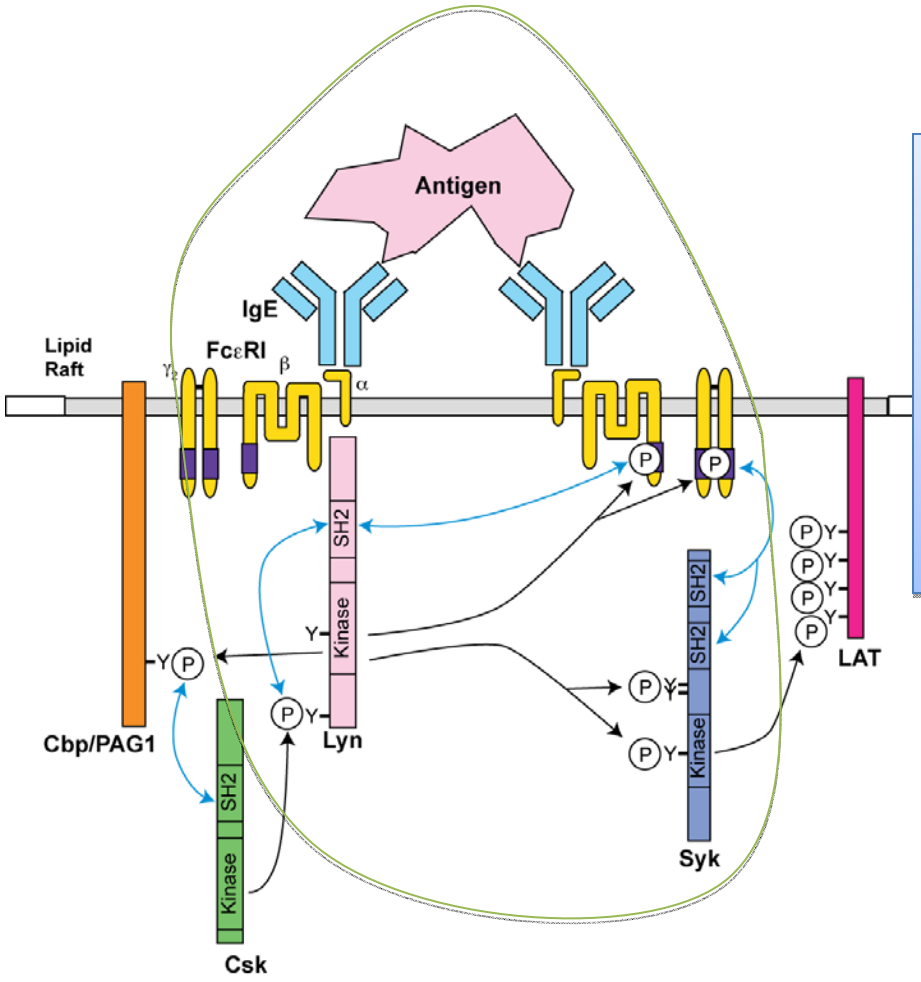
What do we gain

- 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.

Early events in FcεRI signaling



Syk activation model



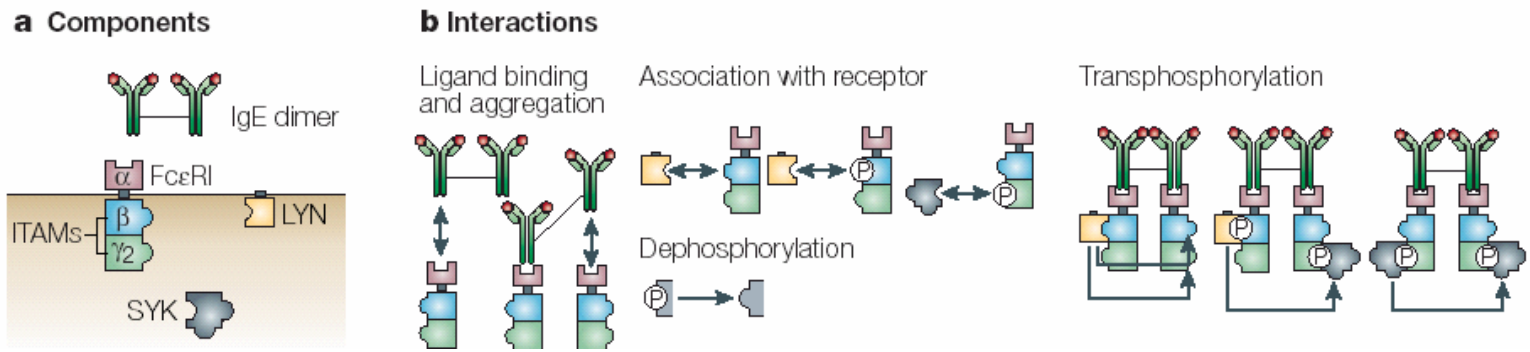
Key variables

- ligand properties
- protein expression levels
- multiple Lyn-FceRI interactions
- transphosphorylation

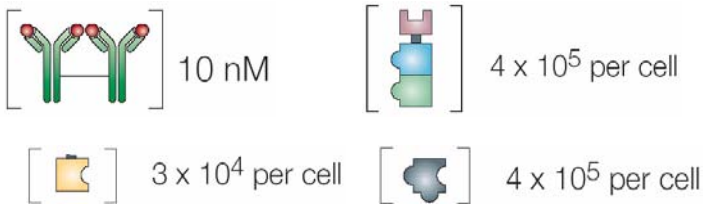
Mol. Immunol.,2002
J. Immunol., 2003

Standard modeling protocol

1. Identify components and interactions.



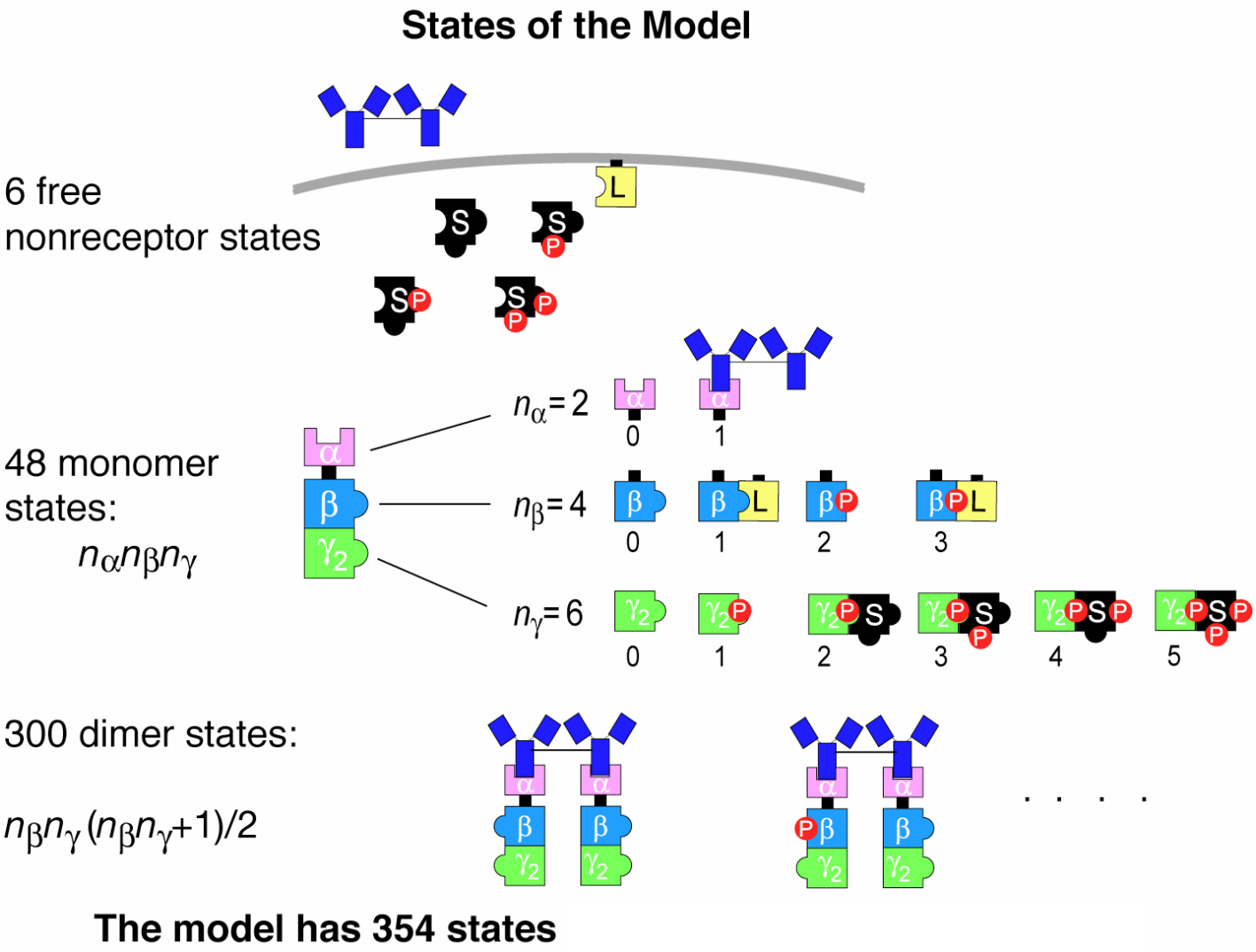
2. Determine **concentrations** and **rate constants**



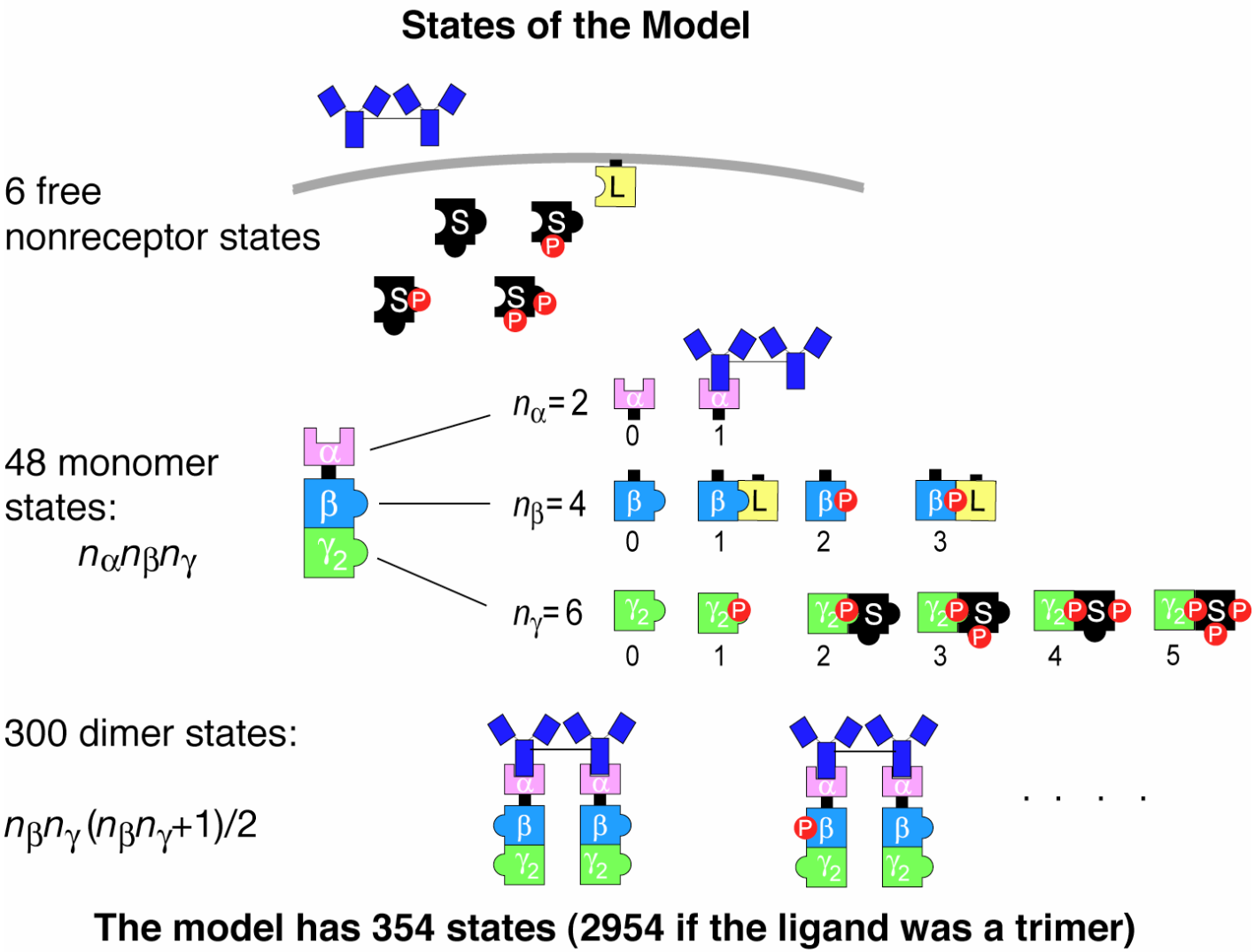
3. Write and solve model equations.

$$\dot{\mathbf{x}} = \mathbf{S} \cdot \mathbf{v}(\mathbf{x})$$

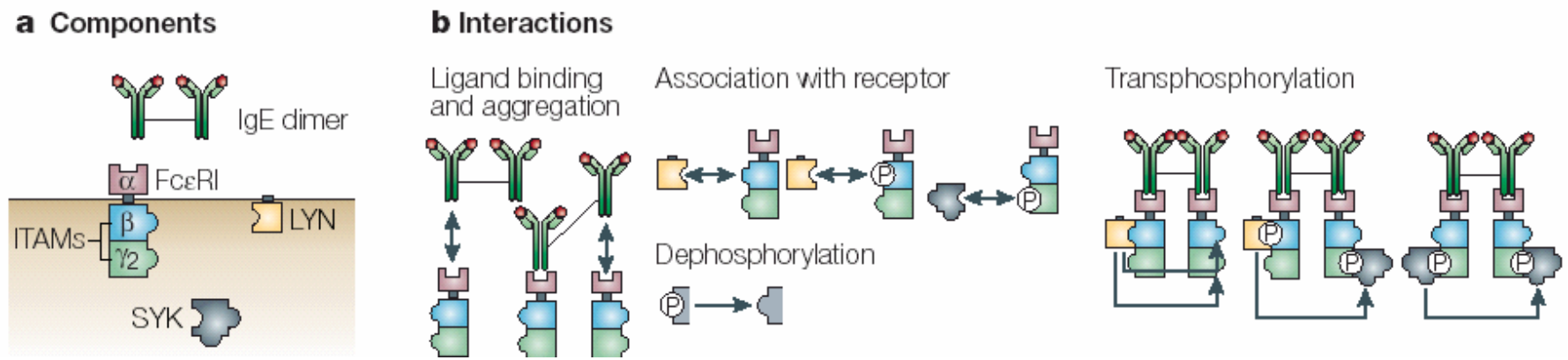
Combinatorial complexity



Combinatorial complexity



Addressing combinatorial complexity

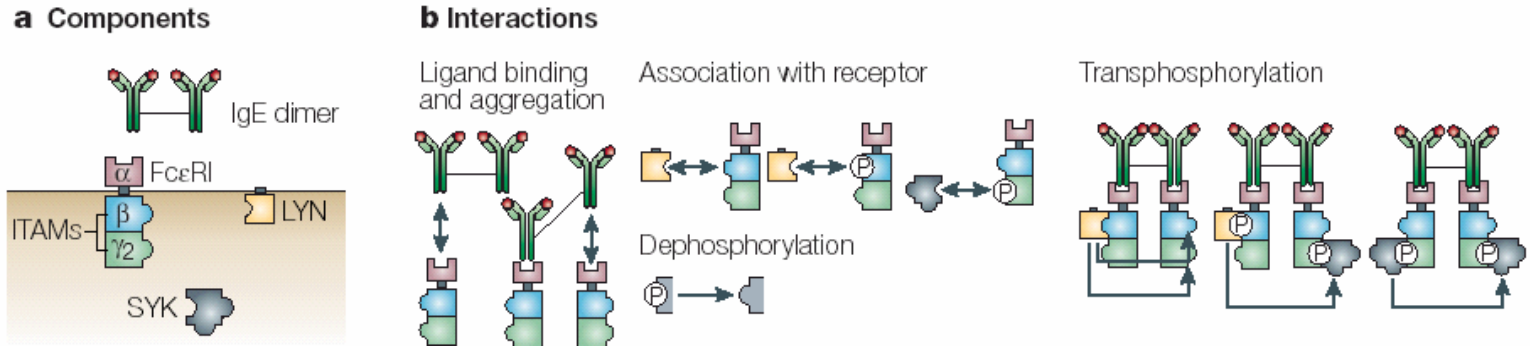


354 species / 3680 reactions

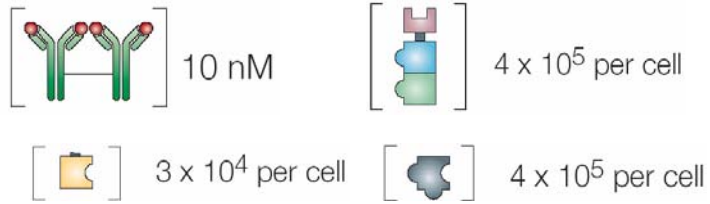
- **Standard approach** – writing equations by hand – won't work!
- **New approach**
  - Write model by describing interactions.
  - Automatically generate the equations.

Rule-based modeling protocol

1. Define components as *structured objects* and interactions as *rules*.



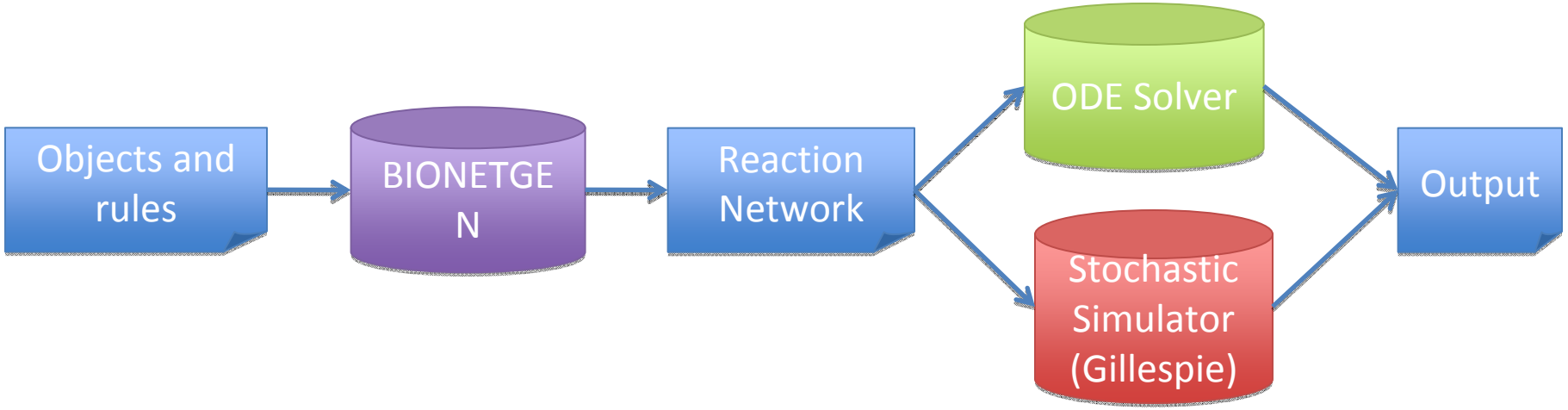
2. Determine **concentrations** and **rate constants**



3. Generate and simulate the model.

Rule-based modeling protocol

1. Define components as *structured objects* and interactions as *rules*.
2. Determine **concentrations** and **rate constants**
3. Generate and simulate the model.

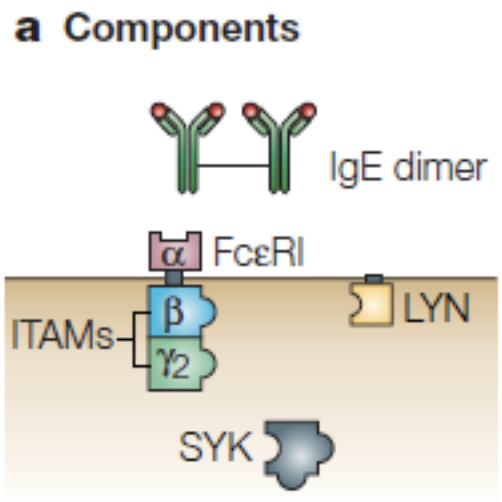


http://bionetgen.org

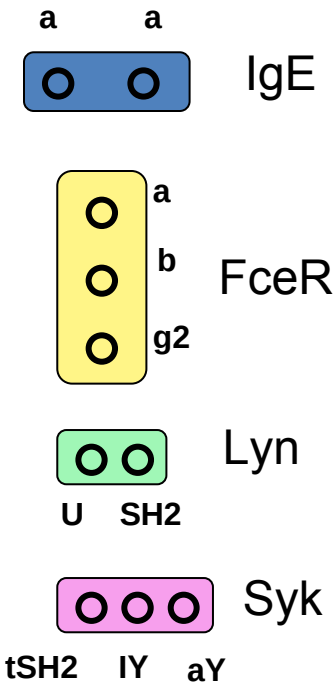
Faeder, Blinov, and Hlavacek, *Methods Mol. Biol.* (2009)

Defining Molecules

Biological Cartoon



Formal Graph Representation



BIONETGEN Description

IgE(a,a)

FceRI(a,b~U~P,g2~U~P)

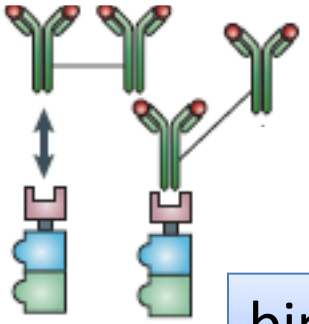
Lyn(U,SH2)

Syk(tSH2,IY~U~P,aY~U~P)

Defining Interaction Rules

Biological Cartoon

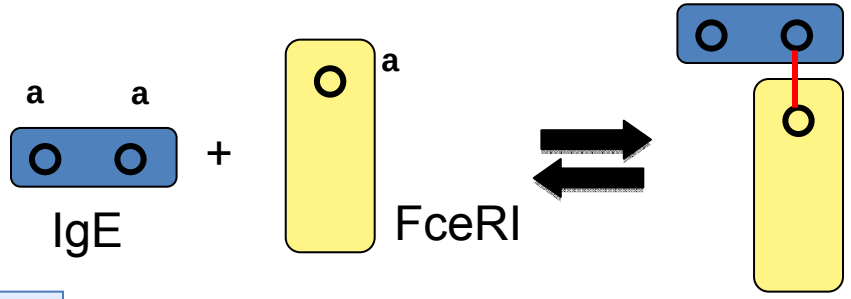
Ligand binding



binding and dissociation

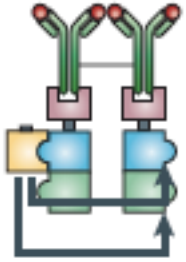
Formal graph representation

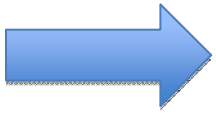
IgE(a,a) + FceRI(a)<-> IgE(a,a!1).FceRI(a!1)



BIONETGEN Description

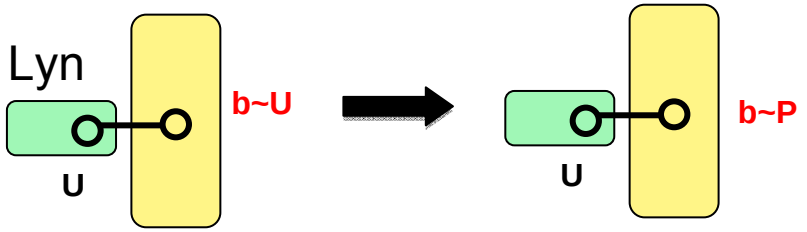
Transphosphorylation





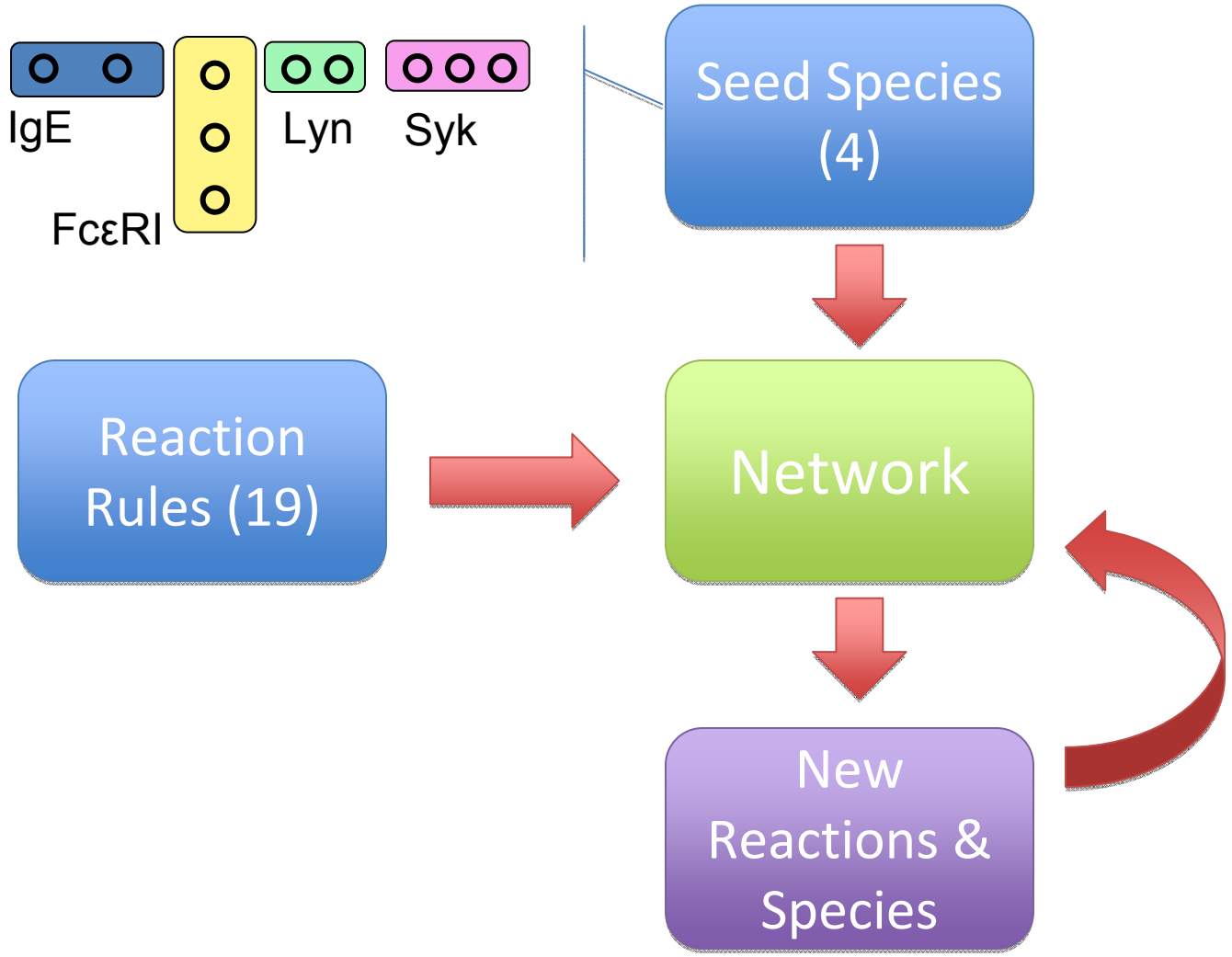
component state change

Lyn(U!1).FceRI(b!1).FceRI(b~U)-> \
Lyn(U!1).FceRI(b!1).FceRI(b~P)



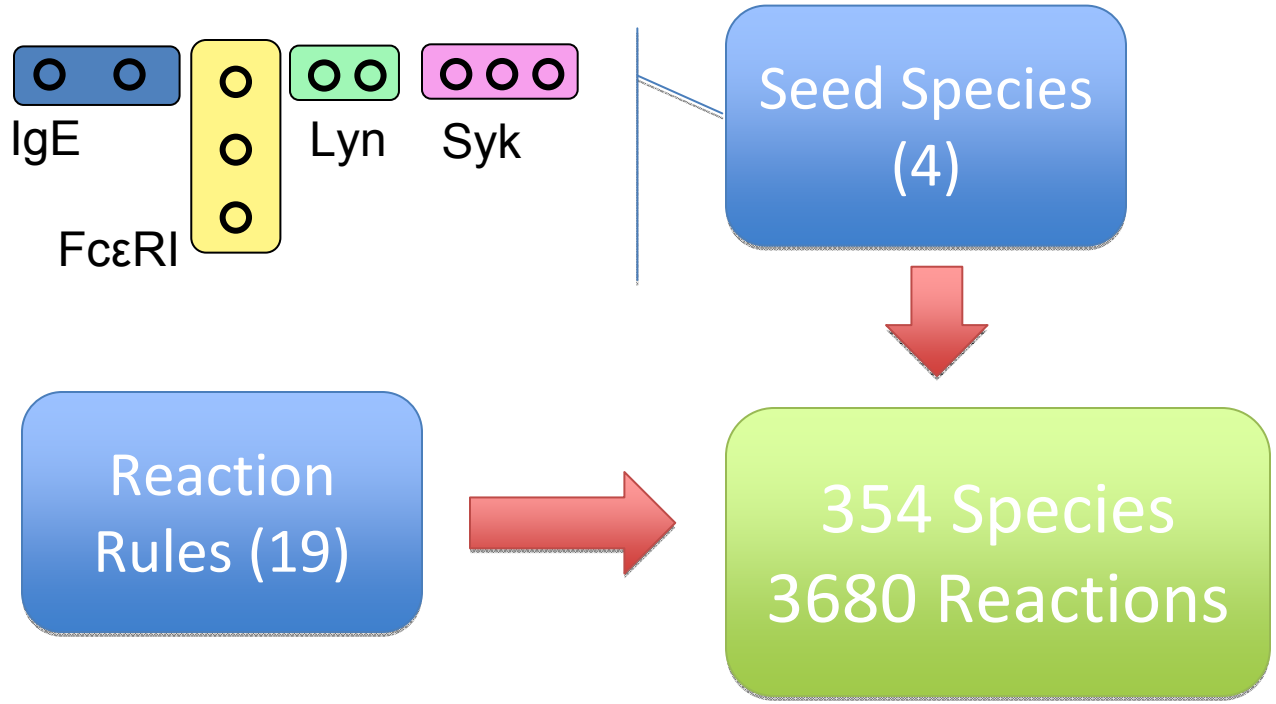
Automatic Network Generation

FcεRI Model



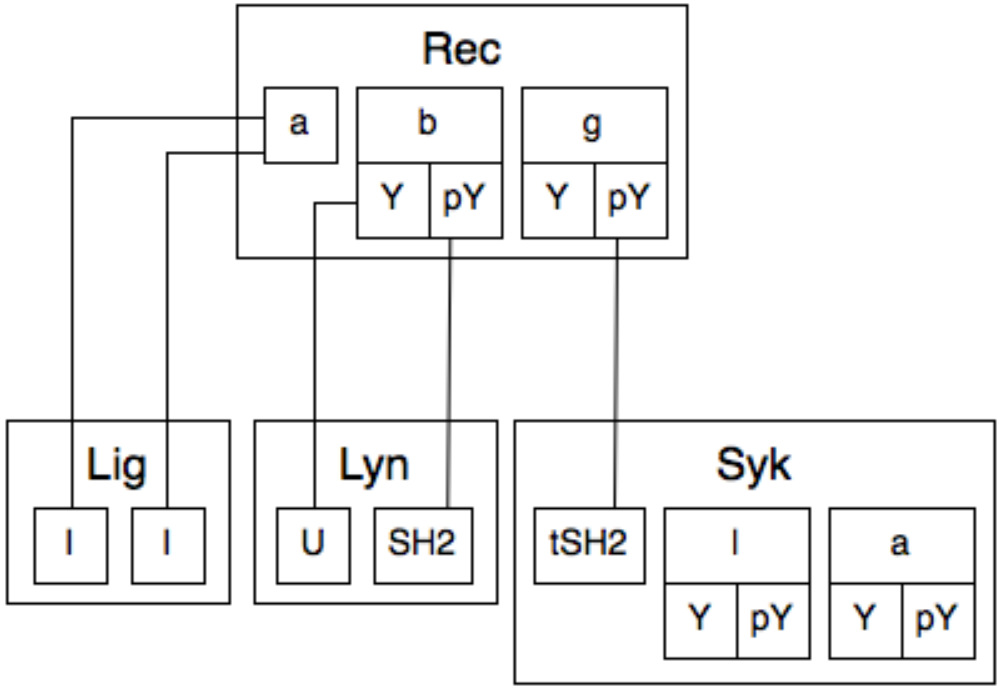
Automatic Network Generation

FcεRI Model



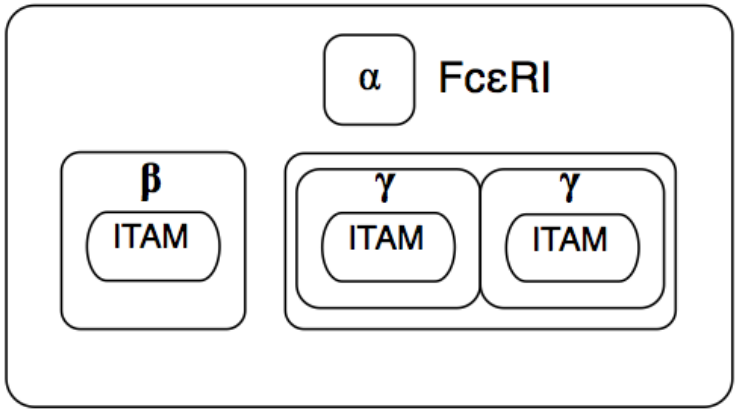
Lily

Contact map



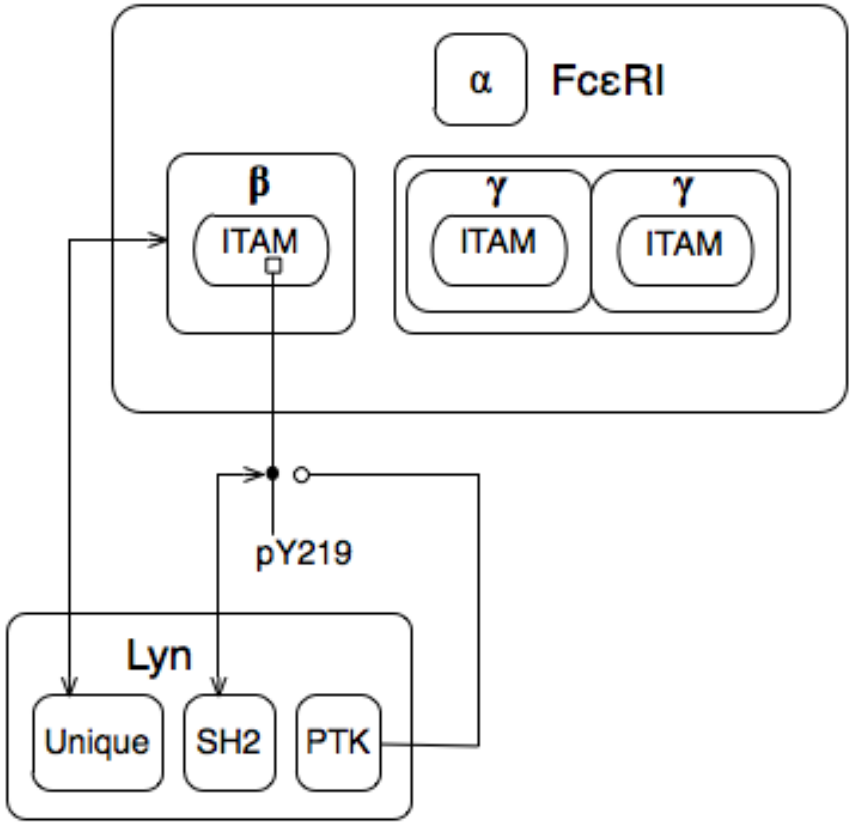
- Shows molecules, components, internal states, and binding interactions.
- Lacks clear representation of enzyme-substrate relationships and protein substructure.

Extended contact map



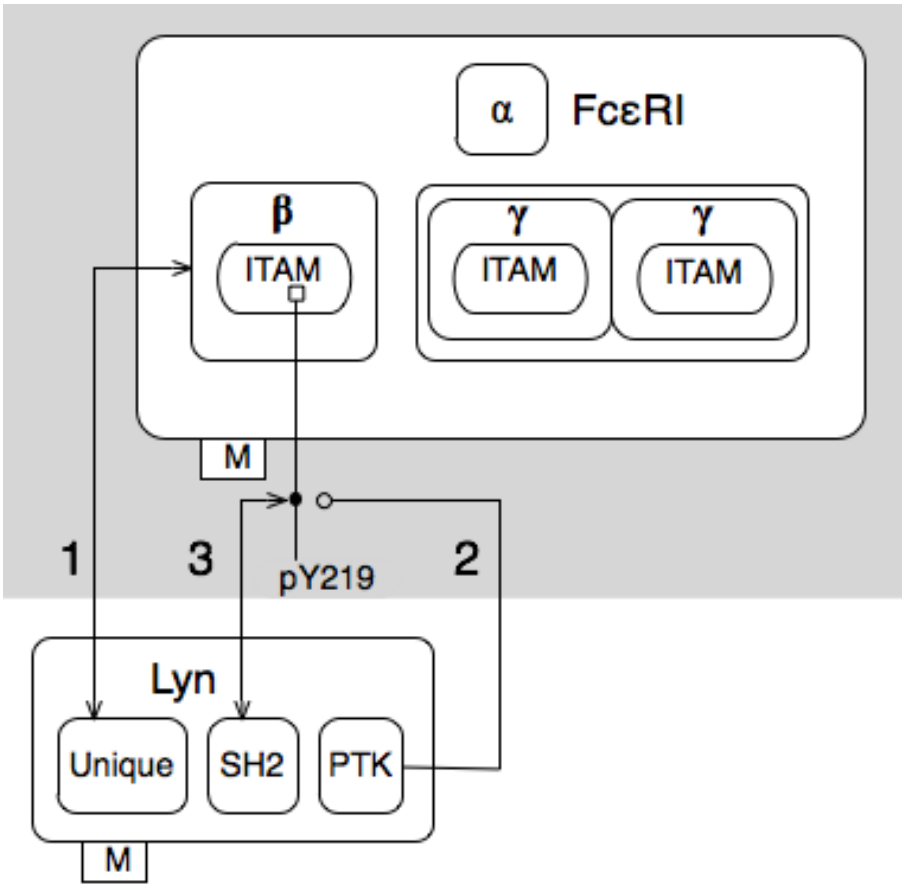
- Multiple levels of nesting to show protein substructure.

Extended contact map



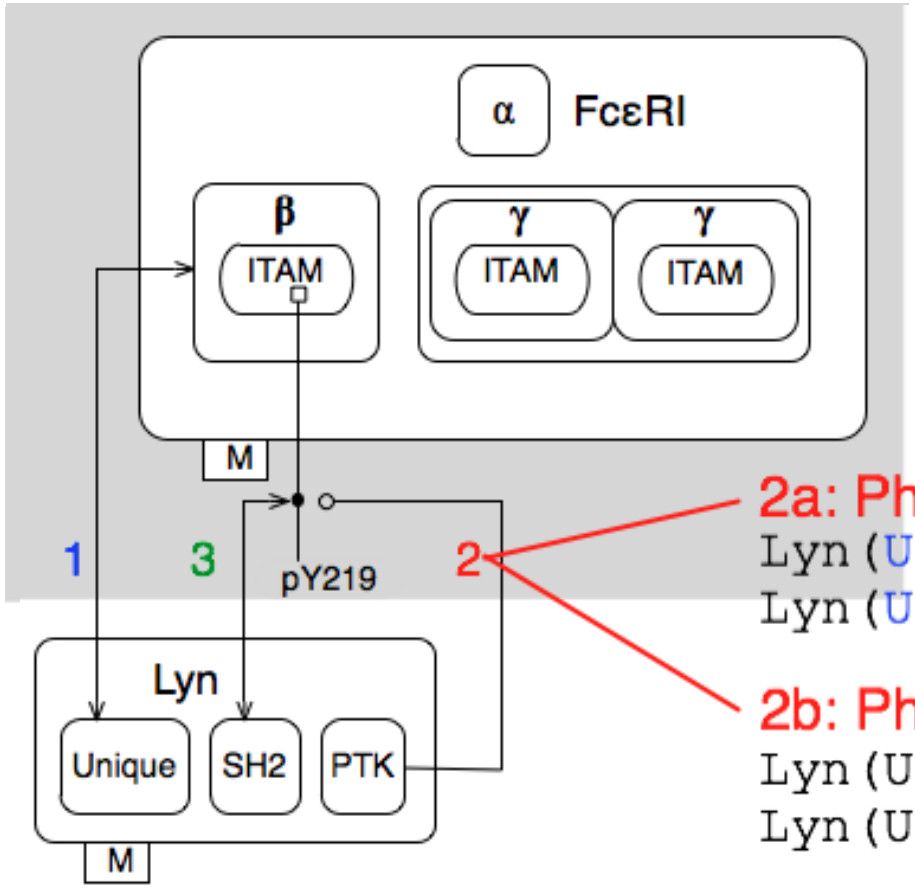
- Multiple levels of nesting to show protein substructure.
- Distinguish different types of interactions (binding vs. catalysis).
- Show sites of post-translational modification.

Extended contact map



- Use tags to show locations of molecules.
- Use shading to indicate hierarchy of molecules in signaling.
- Connect interactions to rules, where context is accounted for.

Extended contact map



2a: Phosphorylation by constitutive Lyn

```
Lyn(U!1,SH2).Rec(b~Y!1).Rec(b~Y)  ->
Lyn(U!1,SH2).Rec(b~Y!1).Rec(b~pY)
```

2b: Phosphorylation by SH2-bound Lyn

```
Lyn(U,SH2!3).Rec(b~pY!3).Rec(b~Y)  ->
Lyn(U,SH2!3).Rec(b~pY!3).Rec(b~pY)
```

Annotation: Model guide

1. Name

FcεRI, β subunit

2. Definition in Model

Molecule type definition (as part of receptor):
`Rec(a,b~Y~pY,g~Y~pY)`

3. Explanatory Note

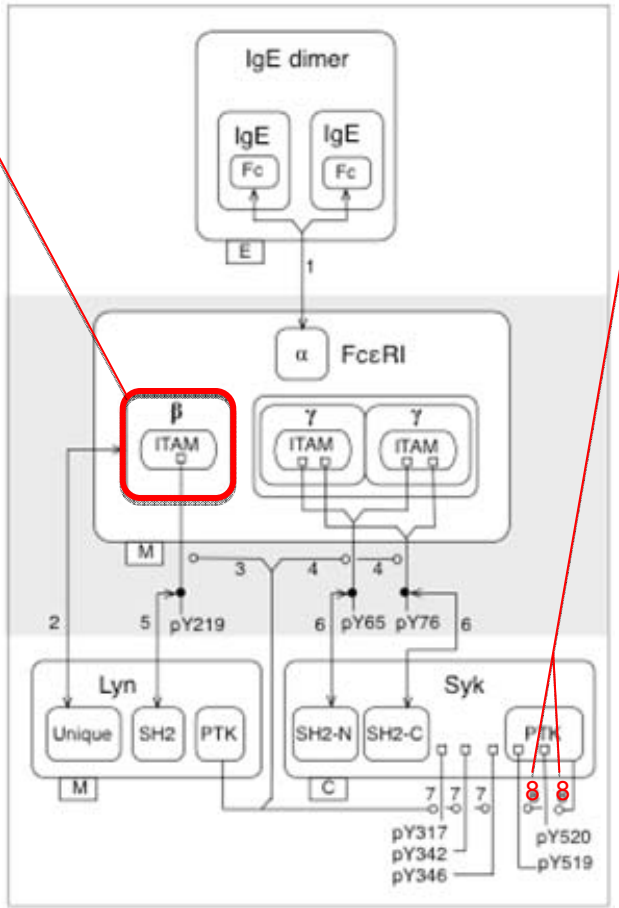
Experimental evidence indicates that the β subunit functions as an amplifier of signaling [1]. This subunit contains an unusual ITAM, being shorter than the canonical sequence and having an additional tyrosine between the N- and C-terminal tyrosines. Mutation of the N-terminal tyrosine (Y219) results in destabilization of the association between Lyn and β [2]. β is also unusual in that it spans the plasma membrane four times so that its N- and C- terminal regions are both in contact with the cytoplasm [3].

4. References

[1] Lin S, Cicala C, Scharenberg AM, Kinet JP: The Fc(epsilon)RIbeta subunit functions as an amplifier of Fc(epsilon)RIgamma-mediated cell activation signals. Cell 1996, 85:985-95.

[2] Oh M, Billingsley JM, Jouvin MH, Kinet JP: Molecular dissection of the FcRbeta signaling amplifier. J Biol Chem 2004, 279:45782-90.

[3] Kuster H, Zhang L, Brini AT, MacGlashan DW, Kinet JP: The gene and cDNA for the human high affinity immunoglobulin E receptor beta chain and expression of the complete human receptor. J Biol Chem 1992, 267:12782-7.



1. Brief Description

Syk transphosphorylates Y519 and Y520 in the activation loop of Syk.

2. Explanatory Note

Full activation of Syk, needed for degranulation of RBL cells, requires *trans*-autophosphorylation at the two activation loop tyrosines, Y519 and Y520 [1,2]. Phosphorylation of Syk by Syk is modeled using the following rules:

3. Rules

```
Lig(l11,l12).Syk(tSH213,a~Y).
Rec(a12,g~pY13).Rec(a11,g~pY14).
Syk(tSH214,a~Y) ->
Lig(l11,l12).Syk(tSH213,a~Y).
Rec(a12,g~pY13).Rec(a11,g~pY14).
Syk(tSH214,a~pY) pSS
```

4. Comments on Rules

The first rule characterizes phosphorylation by Syk that is not phosphorylated on its activation loop, and the second rule characterizes phosphorylation by Syk that is phosphorylated on its activation loop. The two residues are lumped together as a single site.

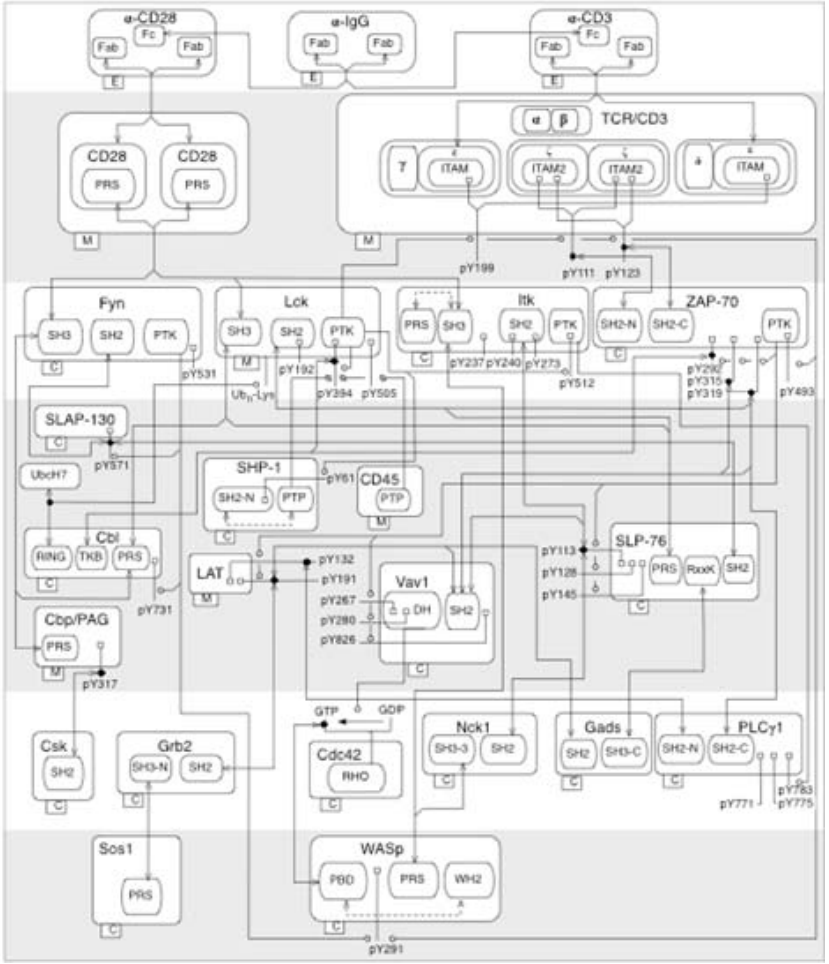
5. References

[1] Zhang J, Billingsley ML, Kincaid RL, Siraganian RP: Phosphorylation of Syk activation loop tyrosines is essential for Syk function. An in vivo study using a specific anti-Syk activation loop phosphotyrosine antibody. J Biol Chem 2000, 275:35442-7.

[2] Siraganian RP, Zhang J, Suzuki K, Sada K: Protein tyrosine kinase Syk in mast cell signaling. Mol Immunol 2002, 38:1229-33.

Limits of GF

- Large-scale model of T cell receptor signaling
- Number of species and reactions is effectively infinite.

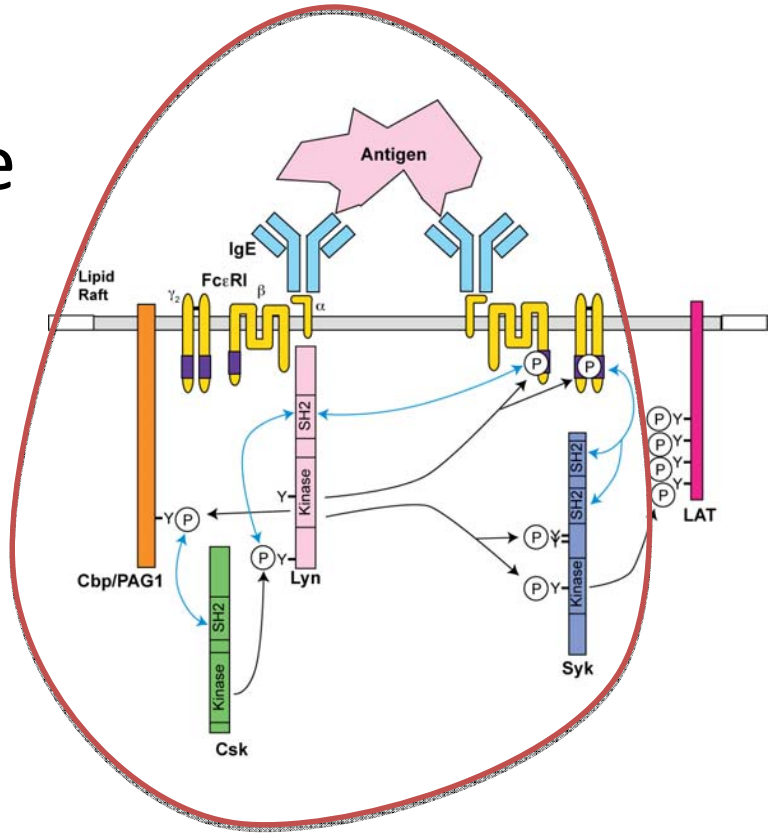


Hu, Chylek, and Hlavacek, in preparation.

Jim

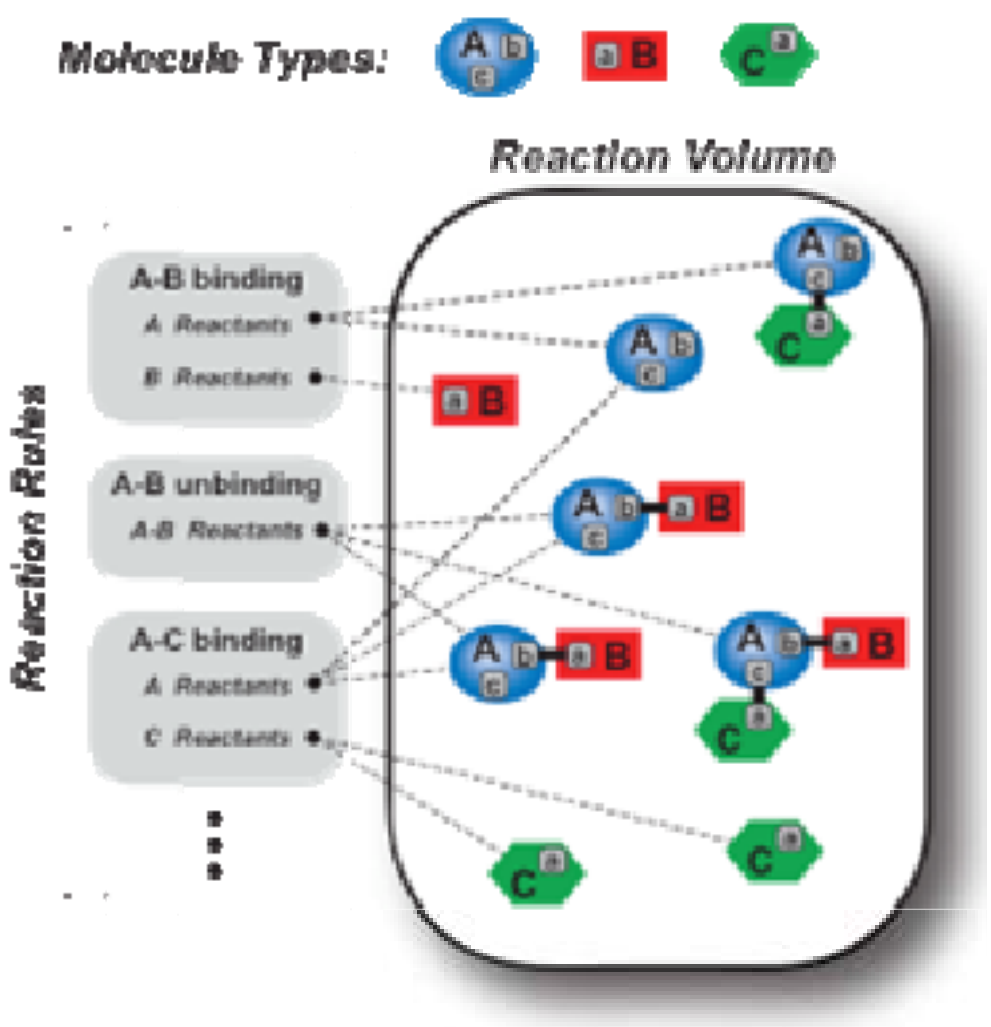
Limits of GF

- Extending model to include Lyn regulation results in >20,000 states.



Agent-based simulation approach

NFSIM
Molecules are instantiated as agents in well-mixed reactor.

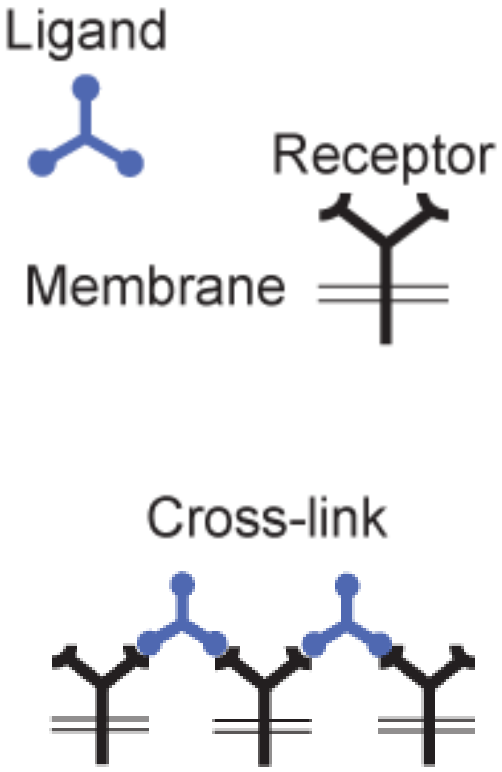


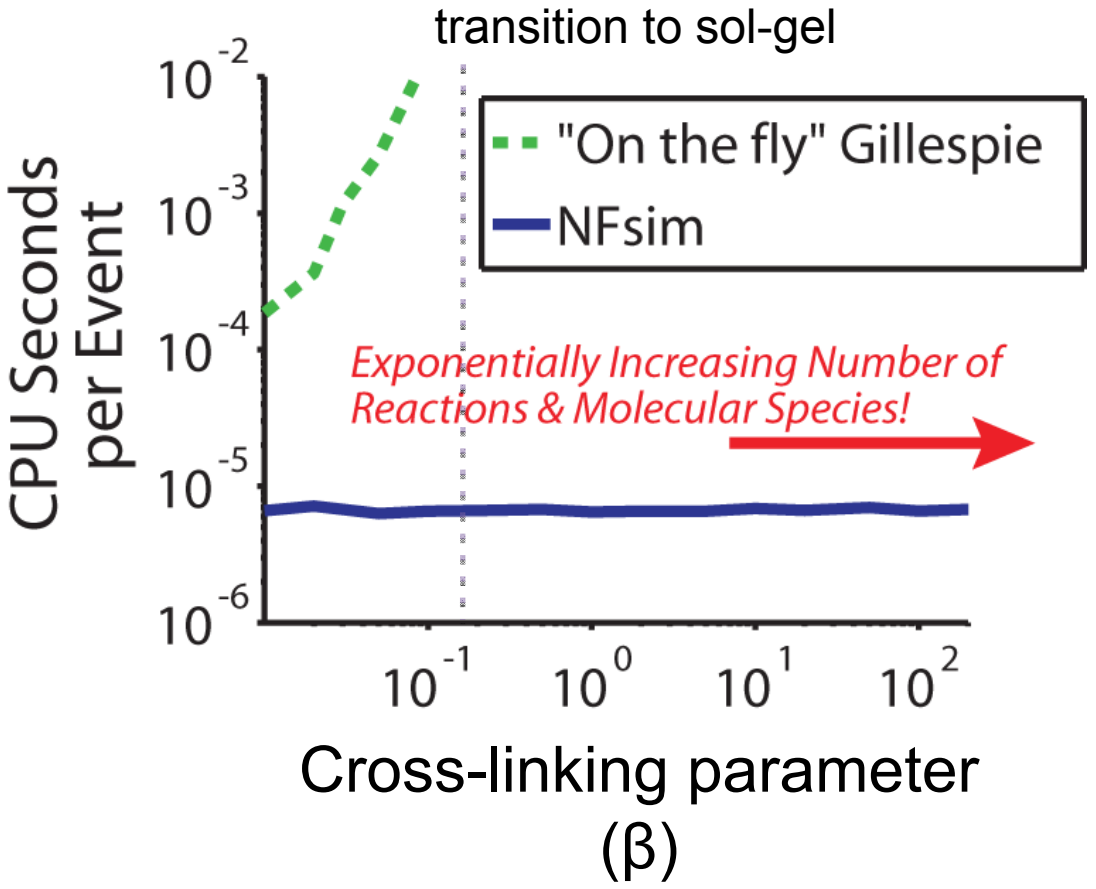
Sneddon, Faeder & Emonet (2010)

http://emonet.biology.yale.edu/nfsim/

NFSIM Performance

Trivalent Ligand, Bivalent Receptor (TLBR) System

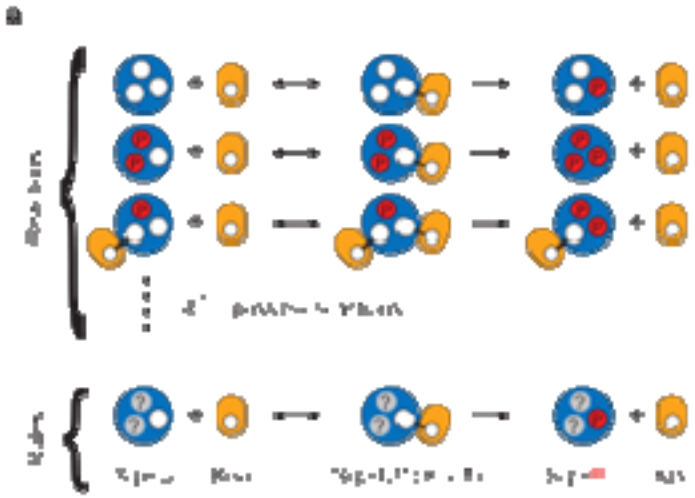


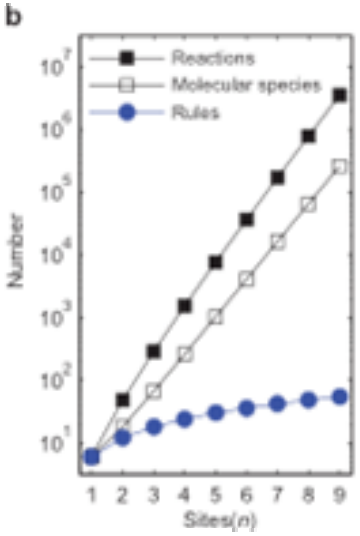


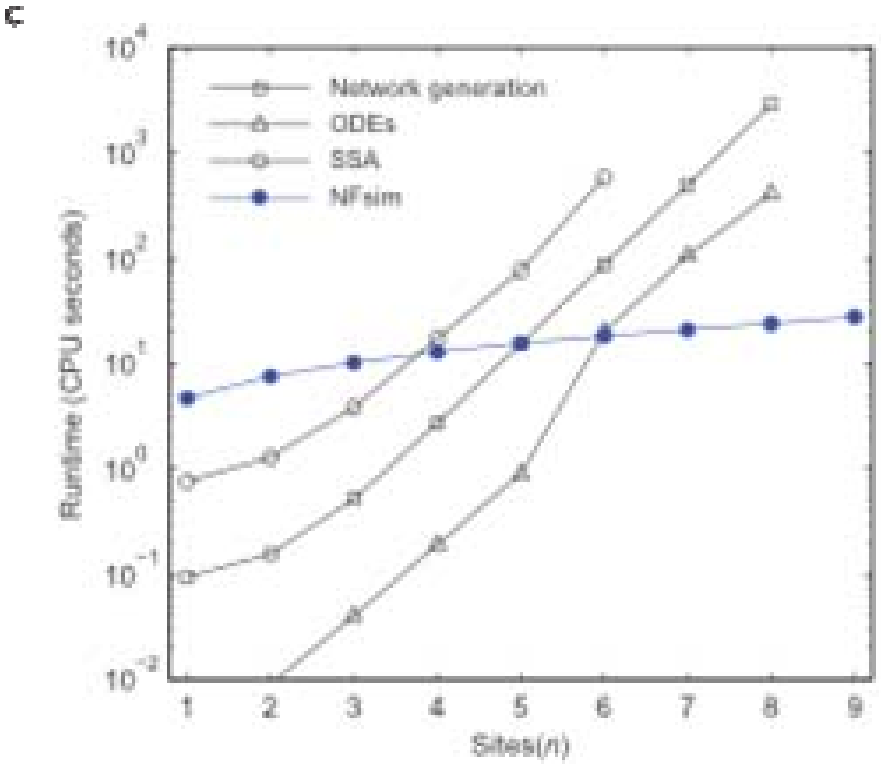
3000 Receptors, 10,000 Ligands

NFSIM Performance

n-site phosphorylation model

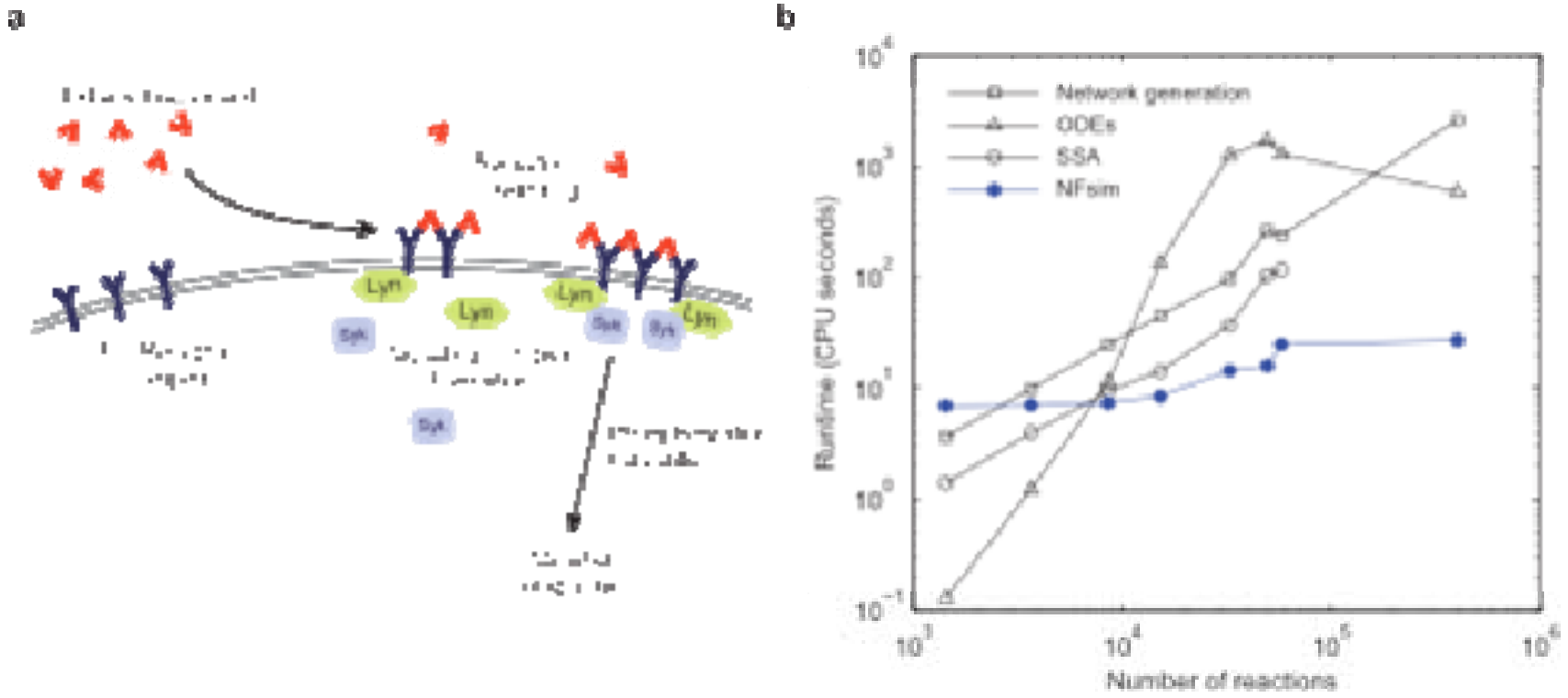






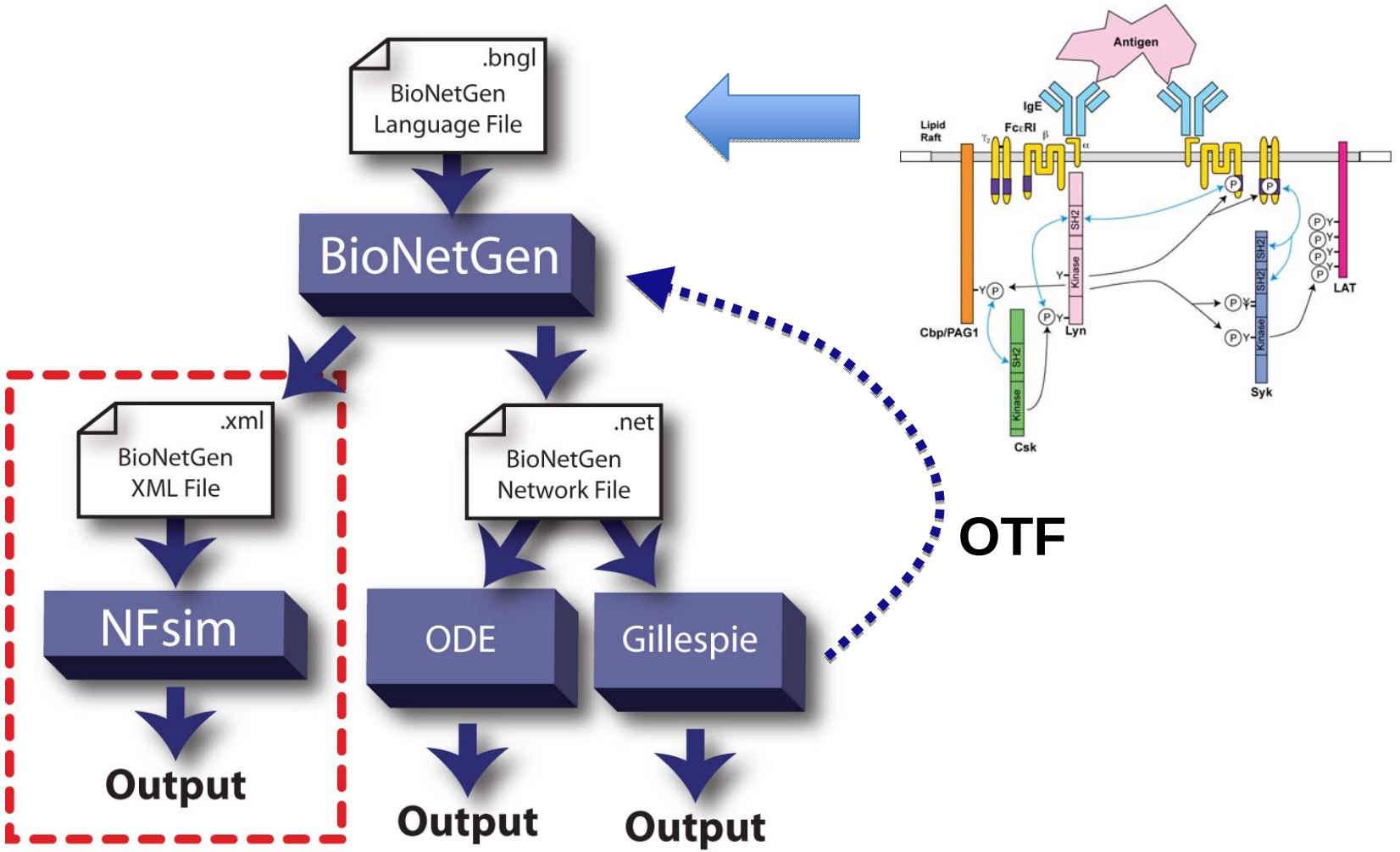
NFSIM Performance

FcεRI signaling models



Increasing complexity

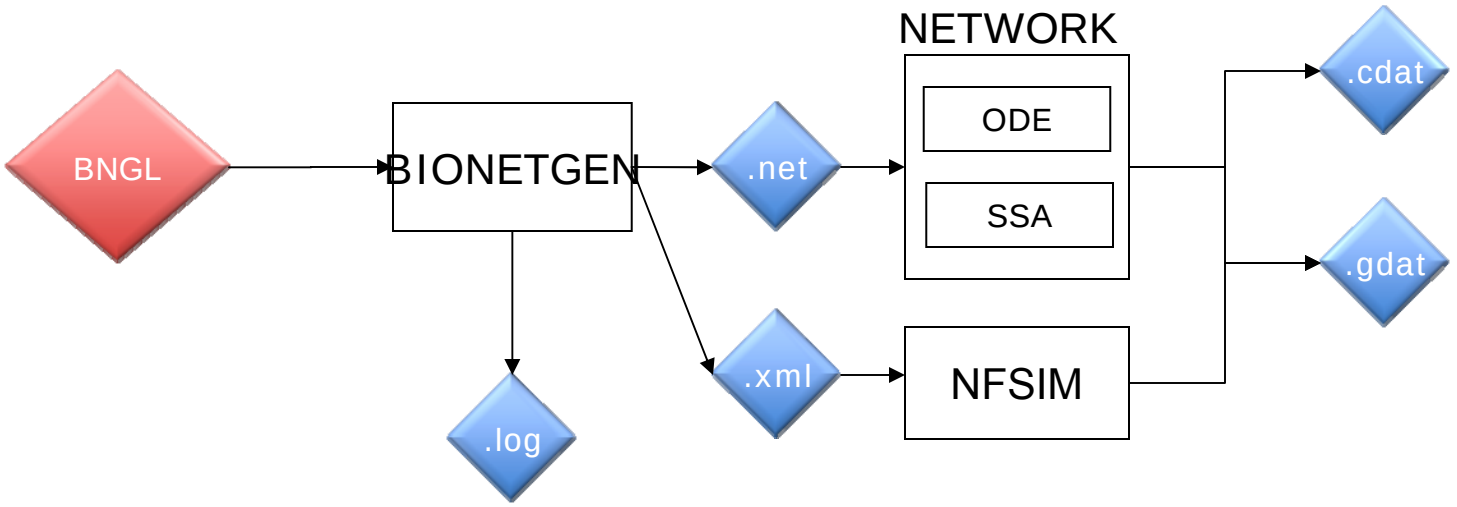
Integration with BIONETGEN



Outline

- Introduction / review of rule-based modeling and simulation
- **Overview of a BIONETGEN model**
- Entering and running a simple model
- Some examples

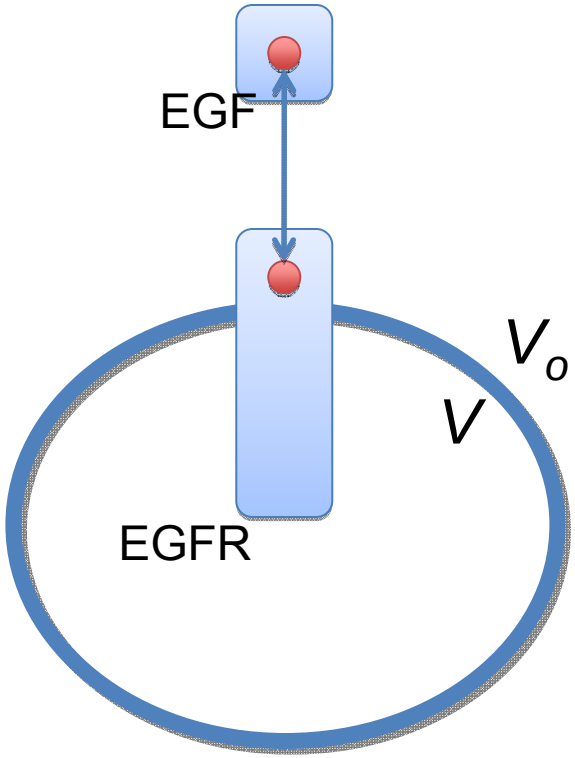
Overview of inputs and outputs



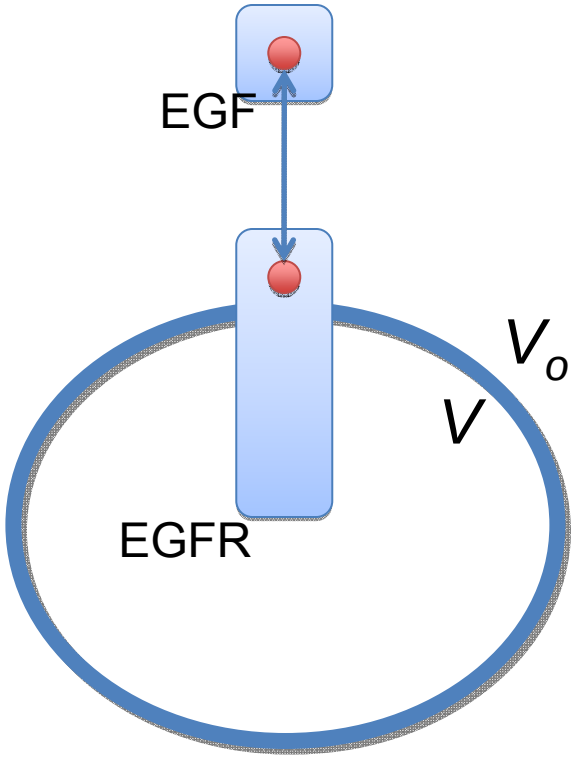
Elements of a BNGL file

- parameters
- molecule types
- seed species
- reaction rules
- observables
- actions

A simple example: Ligand-receptor binding



A simple example: parameters

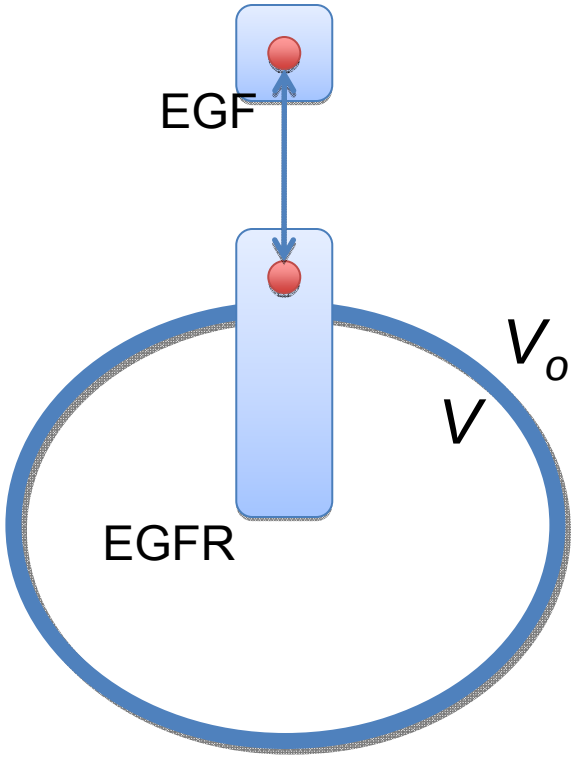



```
begin parameters
  # Physical and geometric constants
     NA     6.0e23 #Avogadro's num
     f          1 #scaling factor
  Vo      f*1e-10 # L
  V       f*3e-12 # L

 # Initial concentrations
  EGF0   2e-9*NA*Vo #nM
  EGFR0  f*1.8e5    #copy per cell

  # Rate constants
  kp1    9.0e7/(NA*Vo) #input /M/sec
  km1    0.06          #/sec
end parameters
```

A simple example: parameters



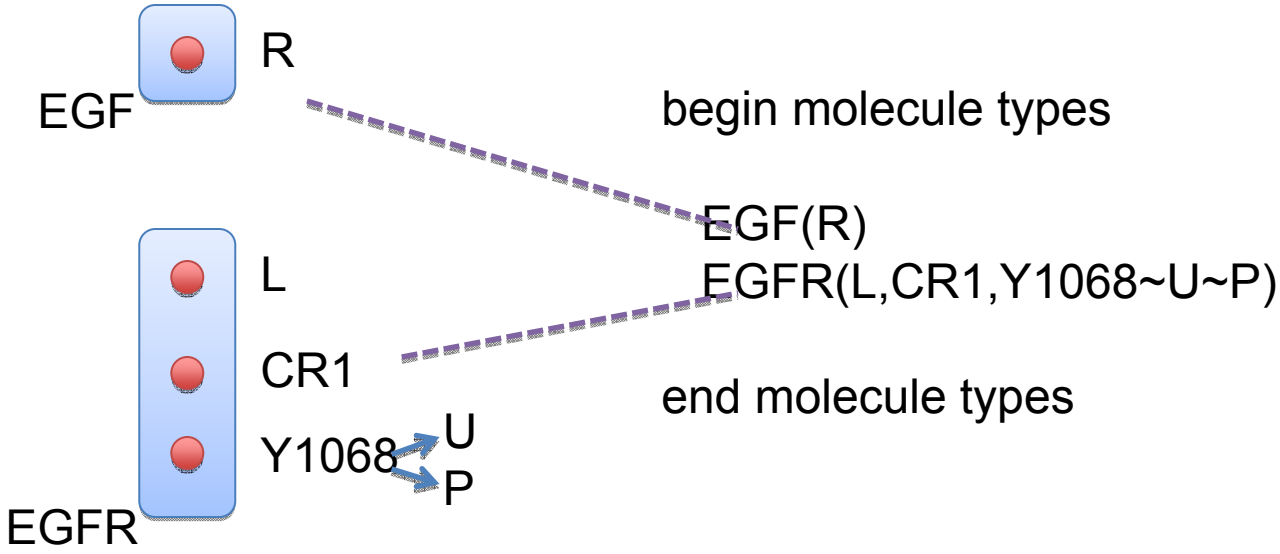
Summary

Concentration units : copies per cell

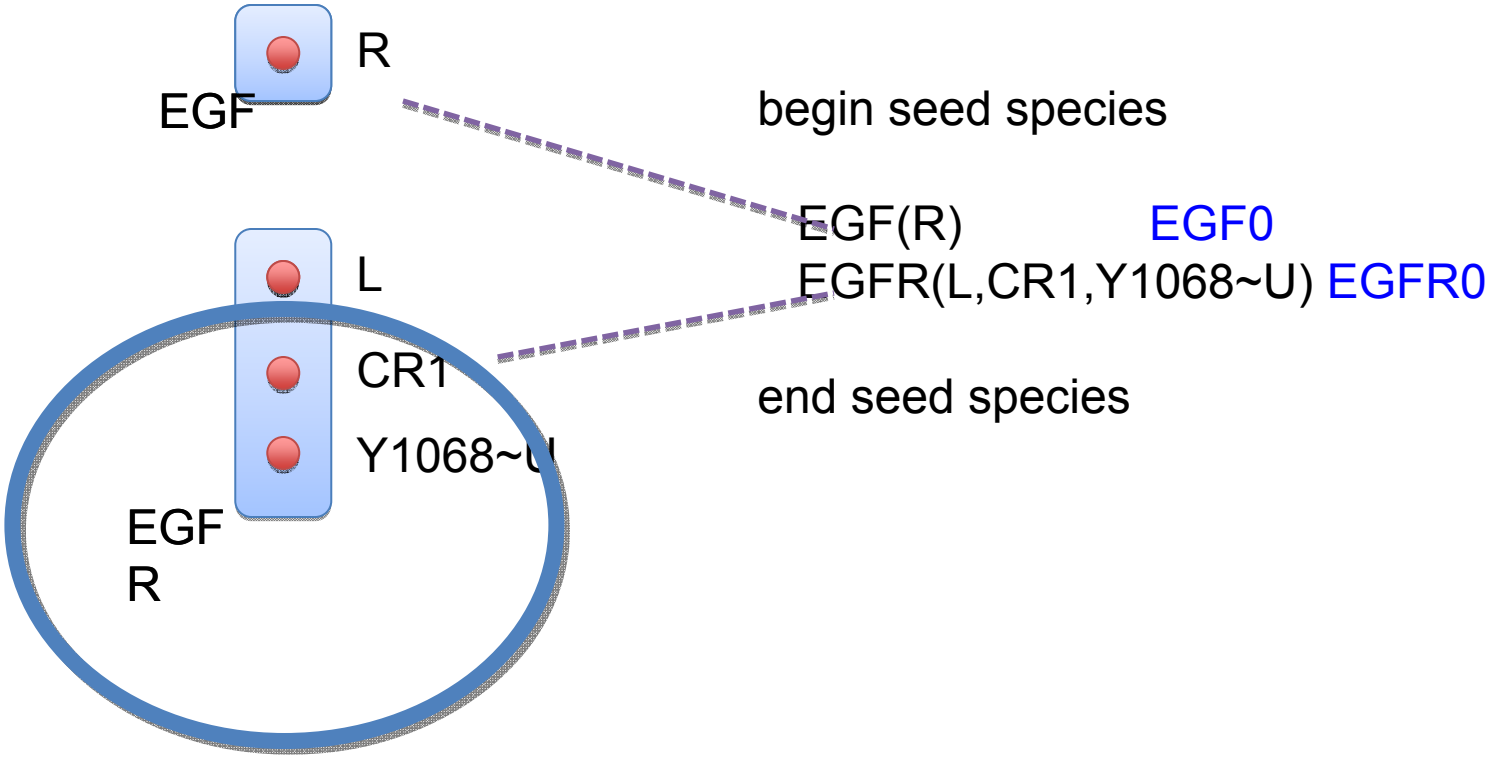
- *Multiply* concentrations by $NA*V_{rxn}$
- Don't scale first order rate constants
- *Divide* second order rate constants by $NA*V_{rxn}$

Following this recipe allows switching between ODE and stochastic simulation without parameter modification.

A simple example: molecule types

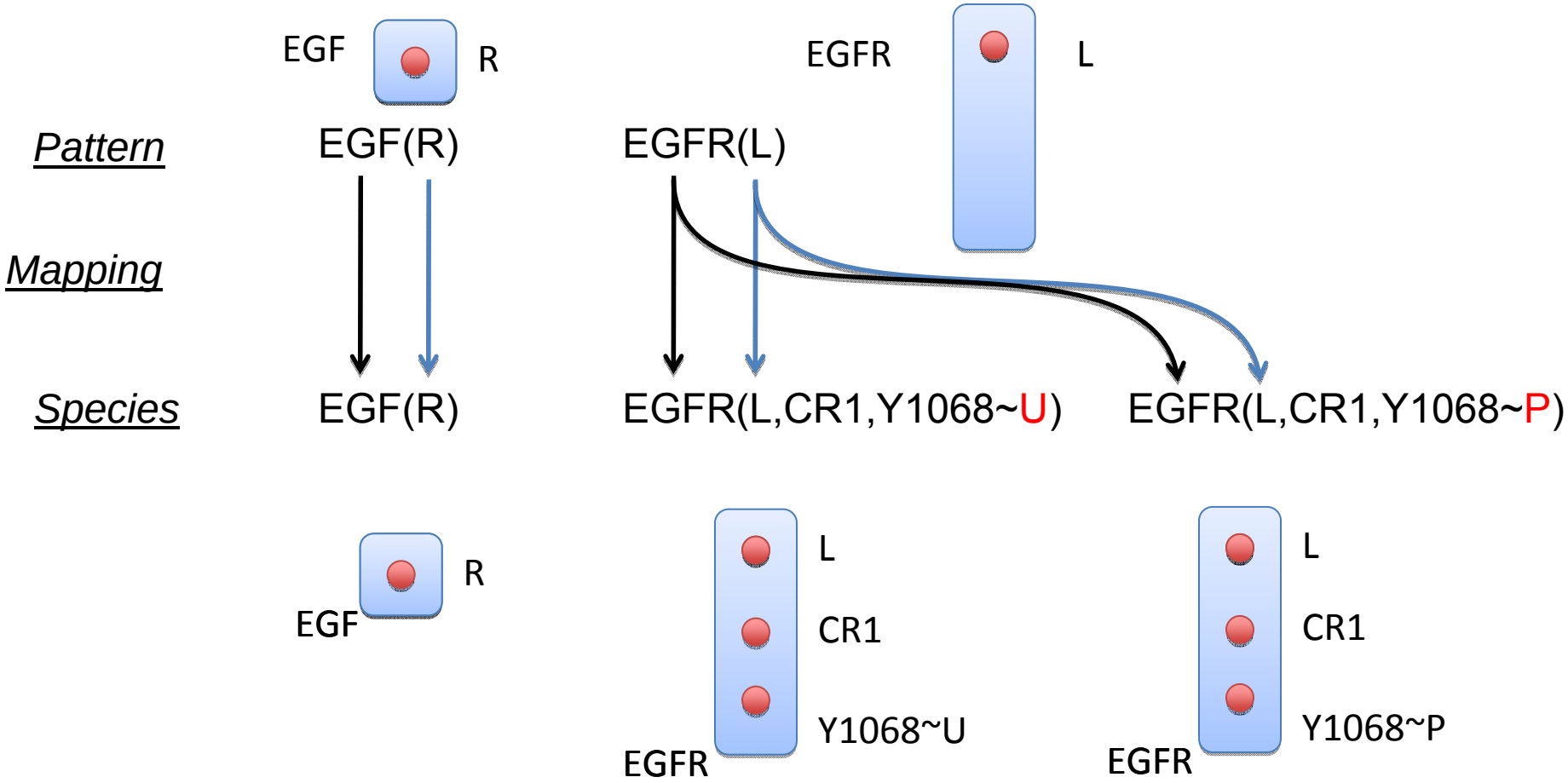


A simple example: seed species



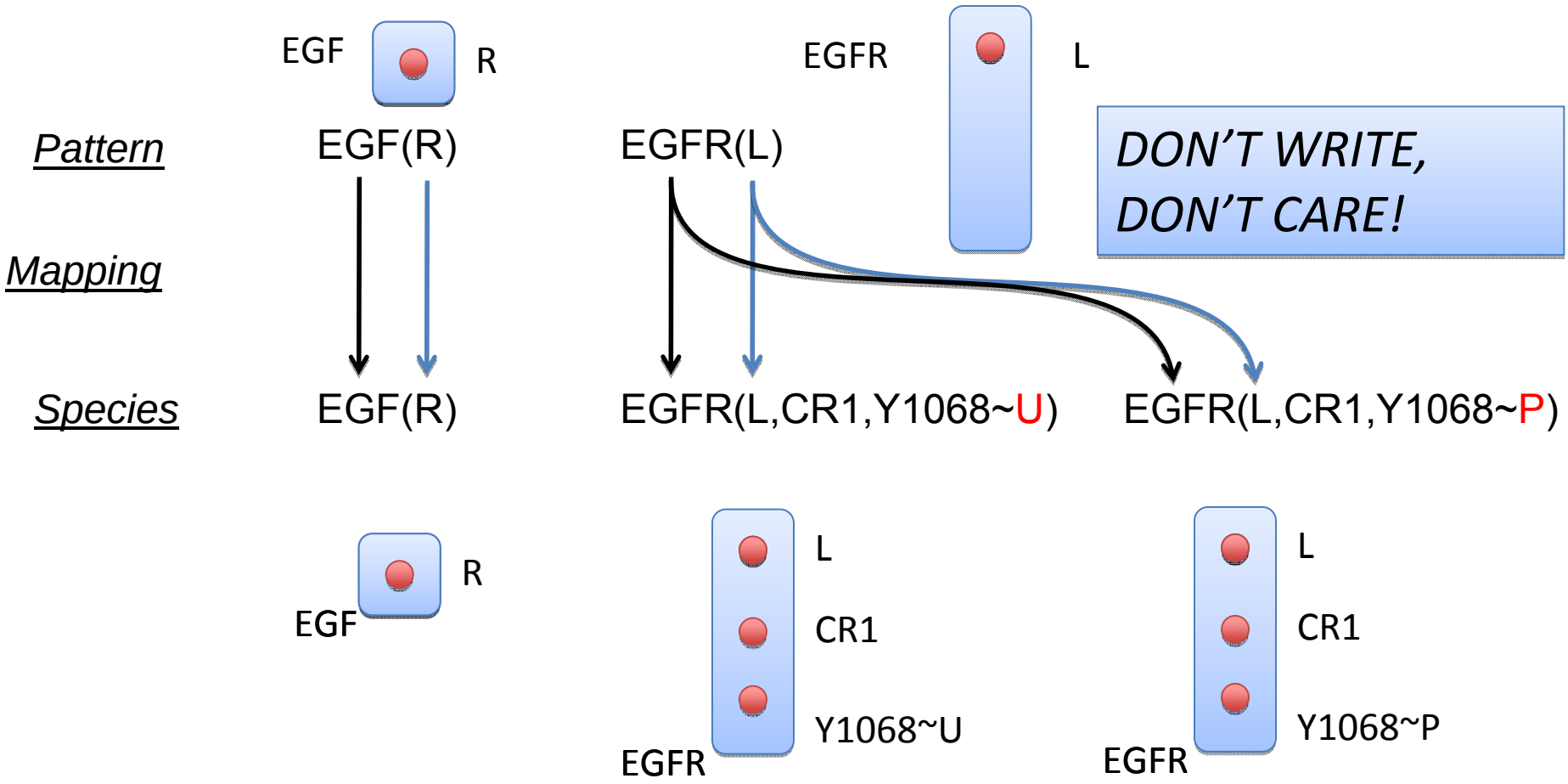
Pattern basics

Definition: A set of molecules that may be partially specified and selects a set of species through its allowed mappings.

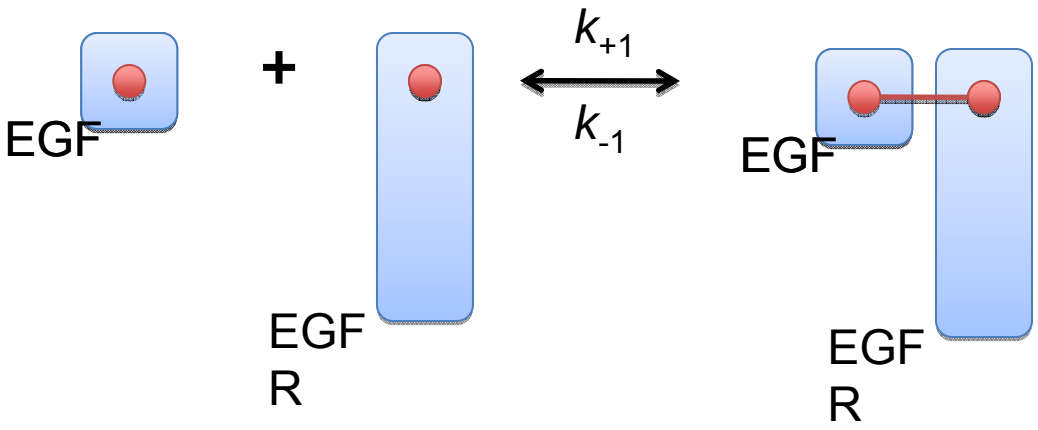


Pattern basics

Definition: A set of molecules that may be partially specified and selects a set of species through its allowed mappings.



A simple example: reaction rules




```
begin reaction rules

 EGF(R) + EGFR(L) <-> EGF(R!1).EGFR(L!1) kp1, km1

end reaction rules
```

A simple example: actions

```
begin parameters
  …
end parameters

begin molecule types
   EGF(R)
   EGFR(L,CR1,Y1068~U~P)
end molecule types

begin seed species
   EGF(R)              EGF0
   EGFR(L,CR1,Y1068~U) EGFR0
end seed species

begin reaction rules
 EGF(R) + EGFR(L) <-> EGF(R!1).EGFR(L!1) kp1, km1
end reaction rules

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

Generates network of species and reactions by iterative application of rules starting with seed species

A simple example: Application of the forward binding rule

Reaction rule

EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

Add bond

Step 1: Match patterns onto reactant species.
Step 2: Copy selected species to product side.
Step 3: Apply transformation specified by rule.

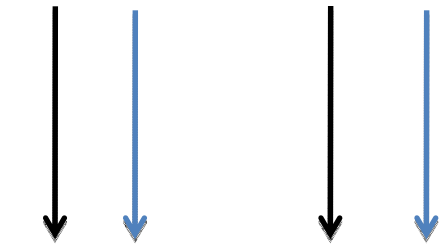
A simple example: Application of the forward binding rule

Reaction rule

EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

Add bond

Step 1: Match patterns onto reactant species.



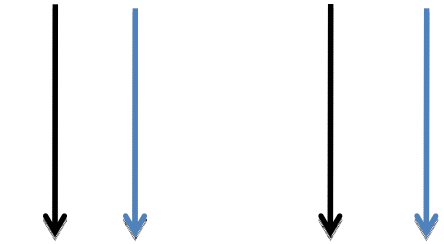
EGF(R) + EGFR(L,CR1,Y1068~U)

A simple example: Application of the forward binding rule

Reaction rule

EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

Add bond



Step 1: *Match patterns onto reactant species.*

EGF(R) + EGFR(L,CR1,Y1068~U)

Step 2: Copy selected species to product side.

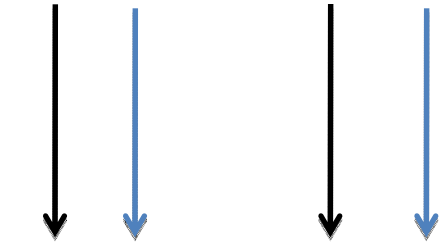
-> EGF(R) + EGFR(L,CR1,Y1068~U)

A simple example: Application of the forward binding rule

Reaction rule

EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

Add bond



Step 1: Match patterns onto reactant species.

EGF(R) + EGFR(L,CR1,Y1068~U)

Step 2: Copy selected species to product side.

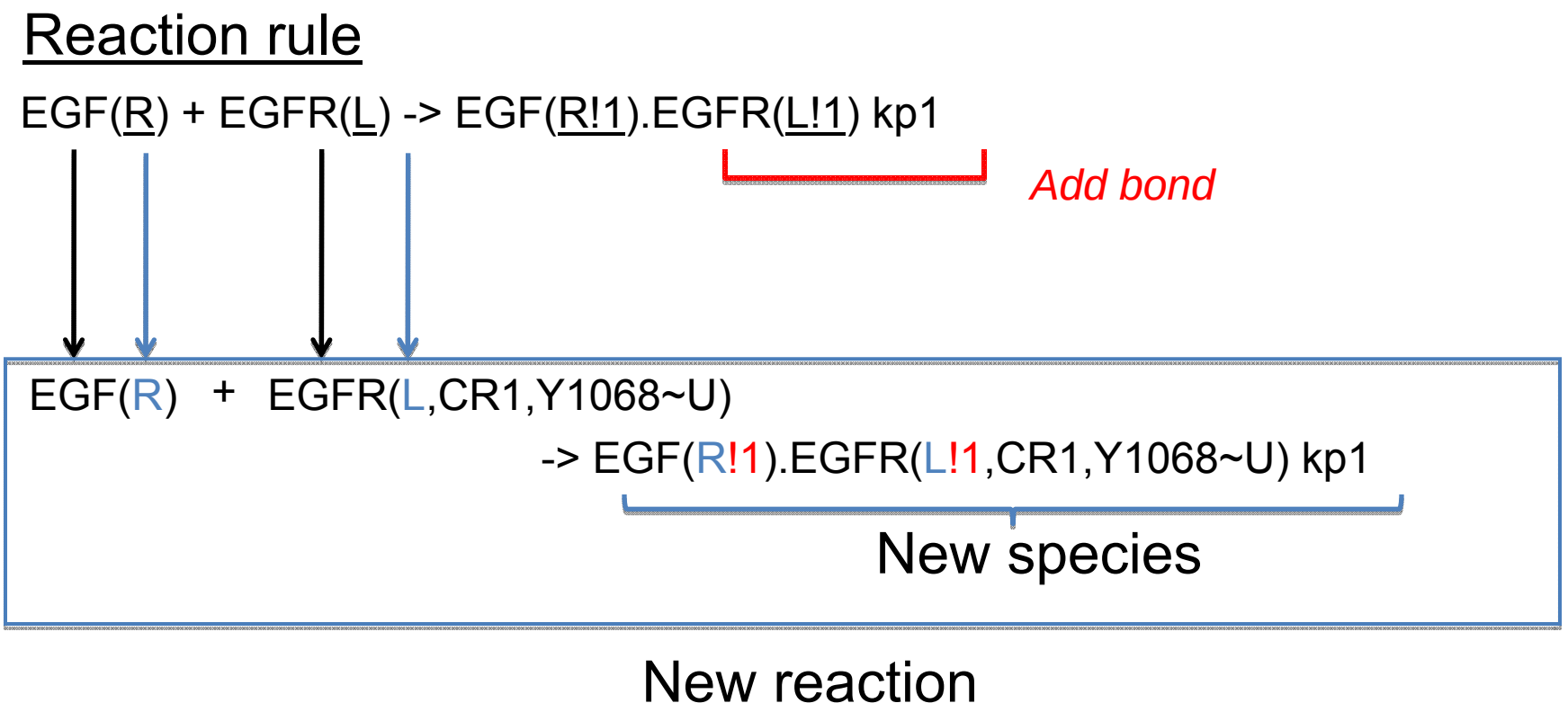
-> EGF(R) + EGFR(L,CR1,Y1068~U)

Step 3: Apply transformation specified by rule.

-> EGF(R!1).EGFR(L!1,CR1,Y1068~U) kp1

New species

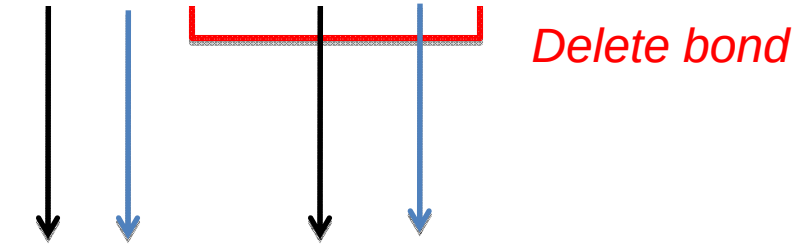
A simple example: Application of the forward binding rule



A simple example:
Application of the reverse binding rule

Reaction rule

EGF(R!1).EGFR(L!1) -> EGF(R) + EGFR(L) km1

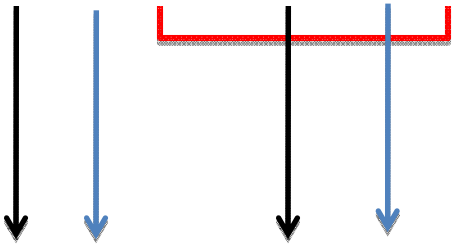


EGF(R!1).EGFR(L!1,CR1,Y1068~U)

A simple example:
Application of the reverse binding rule

Reaction rule

EGF(R!1).EGFR(L!1) -> EGF(R) + EGFR(L) km1



Delete bond

EGF(R!1).EGFR(L!1,CR1,Y1068~U)

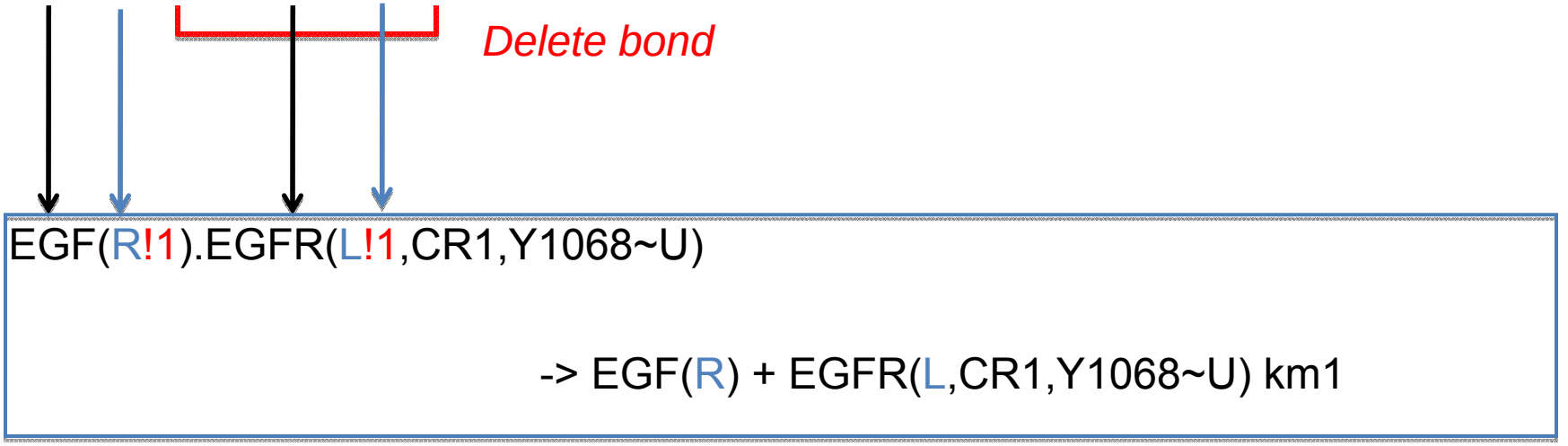
-> EGF(R!1).EGFR(L!1,CR1,Y1068~U)

A simple example:
Application of the reverse binding rule

Reaction rule

EGF(R!1).EGFR(L!1) -> EGF(R) + EGFR(L) km1

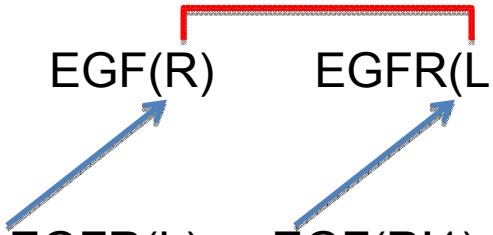
Delete bond



New reaction2

A simple example: Overview of generate_network

seed species EGF(R) EGFR(L,CR1,Y1068~U)



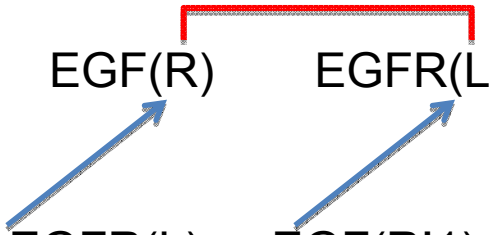
Iteration 1:

rule1f: EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

A simple example: Overview of generate_network

A simple example: Overview of generate_network

seed species EGF(R) EGFR(L,CR1,Y1068~U)



Iteration 1:

rule1f: EGF(R) + EGFR(L) -> EGF(R!1).EGFR(L!1) kp1

new reaction: EGF(R) + EGFR(L,CR1,Y1068~U) ->
EGF(R!1).EGFR(L!1,CR1,Y1068~U) kp1

new species: EGF(R!1).EGFR(L!1,CR1,Y1068~U)

Network after Iteration 1

species
1 EGF(R)
2 EGFR(L,CR1,Y1068~U)
3 EGF(R!1).EGFR(L!1,CR1,Y1068~U)

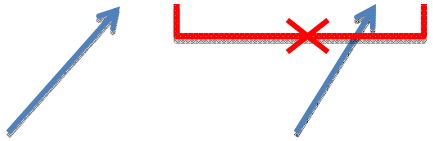
reactions
1 1,2 3 kp1

A simple example: Overview of generate_network

From iteration 1

species
1 EGF(R)
2 EGFR(L,CR1,Y1068~U)
3 EGF(R!1).EGFR(L!1,CR1,Y1068~U)

Iteration 2:



rule1r: EGF(R!1).EGFR(L!1) -> EGF(R) + EGFR(L) km1

new reaction: EGF(R!1).EGFR(L!1,CR1,Y1068~U)
-> EGF(R) + EGFR(L,CR1,Y1068~U) km1

Network after Iteration 2

species
1 EGF(R)
2 EGFR(L,CR1,Y1068~U)
3 EGF(R!1).EGFR(L!1,CR1,Y1068~U)

reactions
1 1,2 3 kp1 # rule1f
2 3 1,2 km1 # rule1r

A simple example: Overview of generate_network

From iteration 1

species
1 EGF(R)
2 EGFR(L,CR1,Y1068~U)
3 EGF(R!1).EGFR(L!1,CR1,Y1068~U)

Iteration 2:

rule1r: EGF(R!1).EGFR(L!1) -> EGF(R) + EGFR(L) km1

new reaction: EGF(R!1).EGFR(L!1,CR1,Y1068~U)
-> EGF(R) + EGFR(L,CR1,Y1068~U) km1

Network after Iteration 2 - final

species
1 EGF(R)
2 EGFR(L,CR1,Y1068~U)
3 EGF(R!1).EGFR(L!1,CR1,Y1068~U)

reactions
1 1,2 3 kp1 # rule1f
2 3 1,2 km1 # rule1r

Outline

- Introduction / review of rule-based modeling and simulation
- Overview of a BIONETGEN model
- **Entering and running a simple model**
- Some examples

Getting the BNGEditor

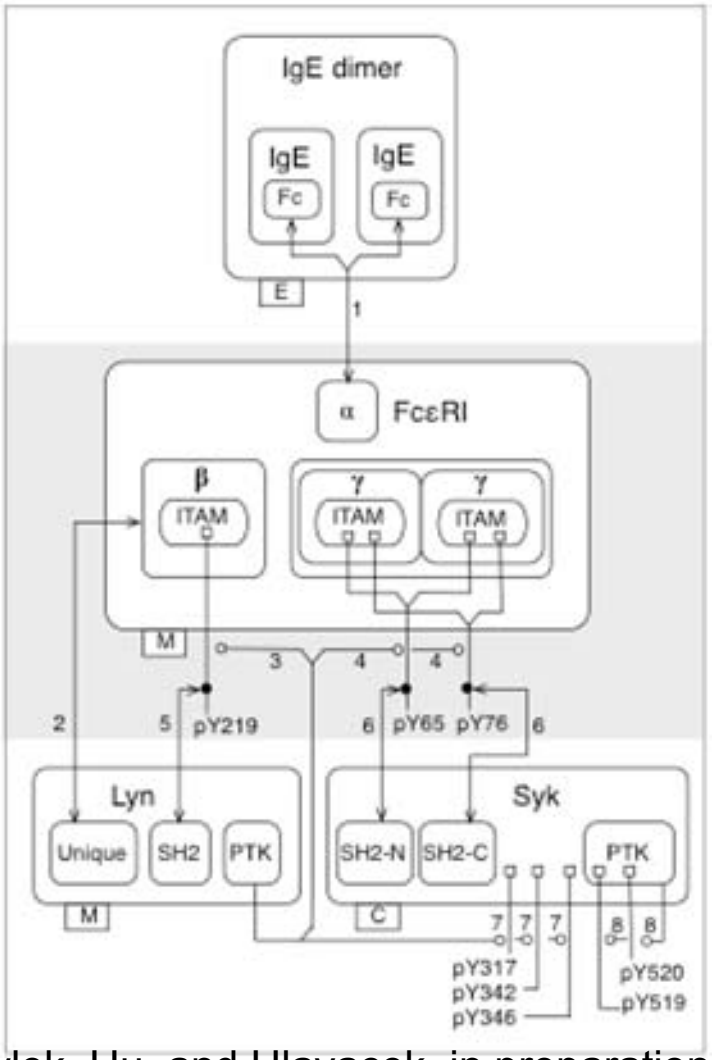
- http://bionetgen.org/index.php/BNGEditor
- To get future versions, you will have to register by sending email to bionetgen@lanl.gov, requesting a Wiki username and password.

Tasks for Simple Model

1. Open the BNGEditor.
2. Open the Simple.bngl file and review its structure.
3. View the contact map.
4. Add observables for bound ligand to the file using both Species and Molecules observables.
5. Generate the network and use Result viewer to examine the .net file.
6. Run a simulation using simulate_ode and examine the resulting .cdat and .gdat files.
7. Run a simulation using simulate_ssa and compare results with ode.
8. Run a parameter scan (make sure to comment out actions).

Syk Activation Model

“Extended Contact Map”



Chylek, Hu, and Hlavacek, in preparation.

1. Write rules for ligand-receptor binding.
2. Write rules for ligand-induced aggregation.

Theoretical Estimate of Crosslinking Constant

$$K_X R_T = K C_e$$

$$C_e = \frac{\#\text{ accessible receptors in disk}}{\text{volume of hemisphere}}$$

$$= \frac{4\pi\delta\bar{d}R_T}{4\pi\delta\bar{d}^2}$$

$$= R_T / \bar{d}$$

$$K_X = K / \bar{d}$$

$$K_X^{molec} = K_X / (N_A A)$$

$$= K / (N_A A\bar{d})$$

$$= K / (N_A V_{\text{mem}})$$

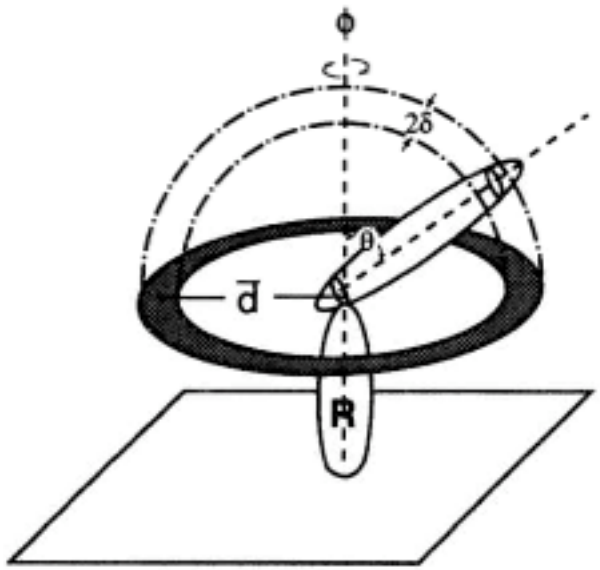


Figure 4. Hemispherical shell sampled by the free binding site on a bivalent ligand of average length d bound to a receptor R.

B. Goldstein and C. Wofsy, *Lect. Math. Life Sci.*, **24**, 70 (1994).

Next Steps for Syk Activation Model

1. Define molecule types and seed species for Lig and Rec.
2. Write rules for ligand-receptor binding and cross-linking.
3. Create observables for bound ligand and receptors in dimers.
4. Run a simulation and plot the results.
5. Make a dose-response curve and observe interesting features.
6. How does K_2 affect overall ligand binding?

More Advanced Topics

- Virtual Cell with advanced examples (Michael)
  - Fluorescent labeling
  - Basic chemistry example with exclude
  - Actin Filaments
- Network Visualization Approaches (Lily)
- NFsim models (Jim)
  - poly
  - multisite phosphorylation
- Compartmental BNGL (Jim)
  - Dimerization model
  - TF activation and gene regulation