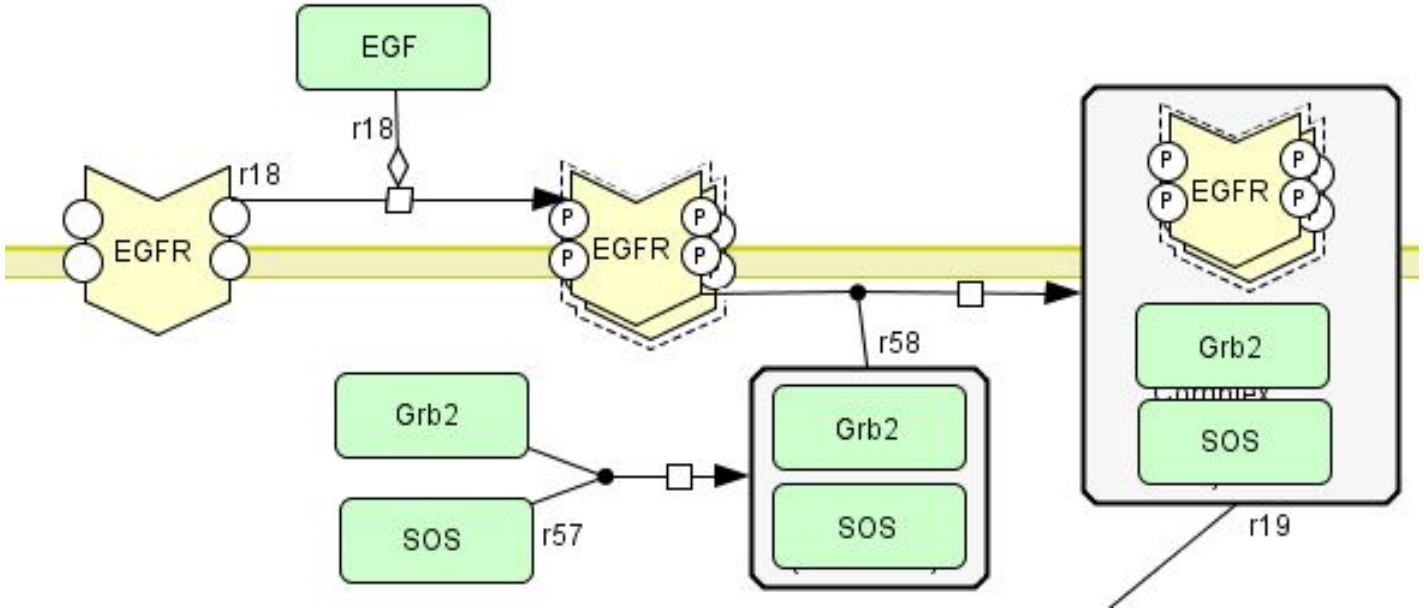
Rule-Based Modeling &

Open World SBGN

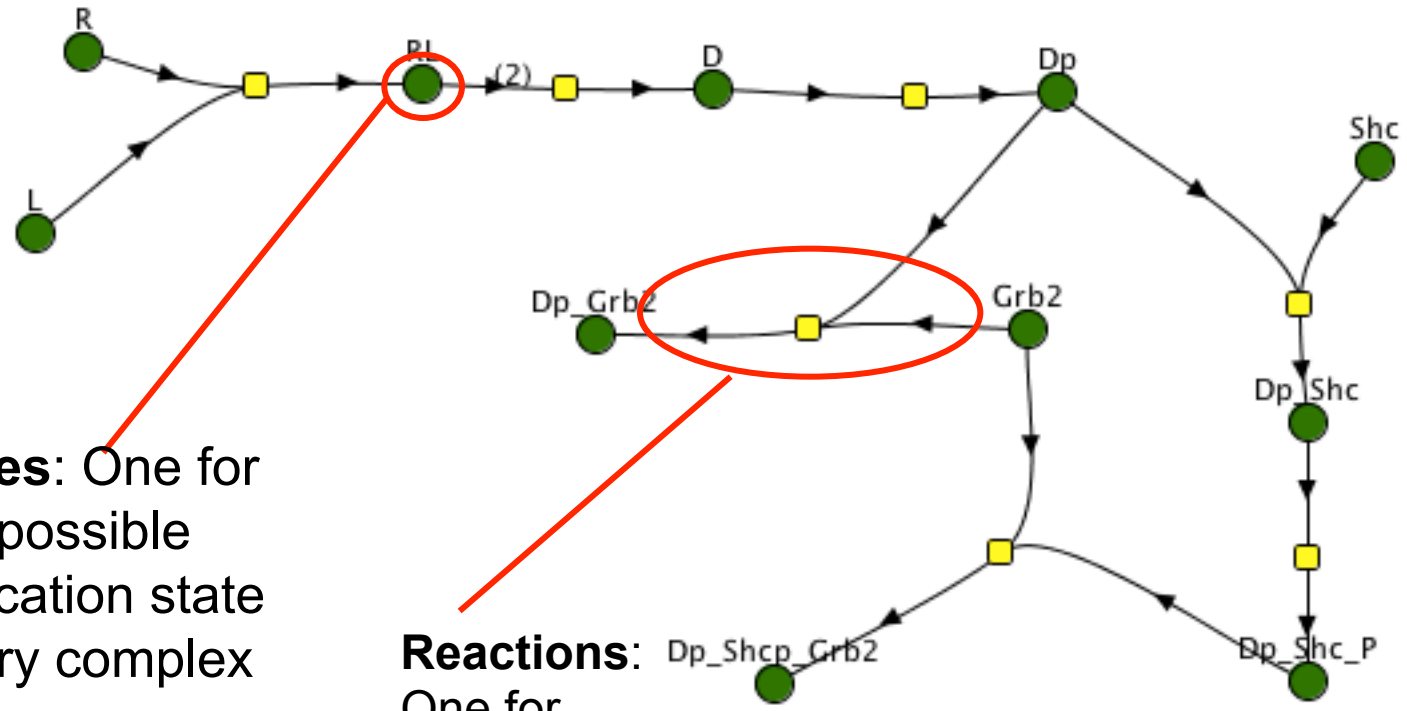
Michael Blinov
Center for Cell Analysis & Modeling
University of Connecticut Health Center
COMBINE, Los Angeles, August 19th, 2014

Typical pathway description



http://PantherDB.org

Typical visualization of a reaction network

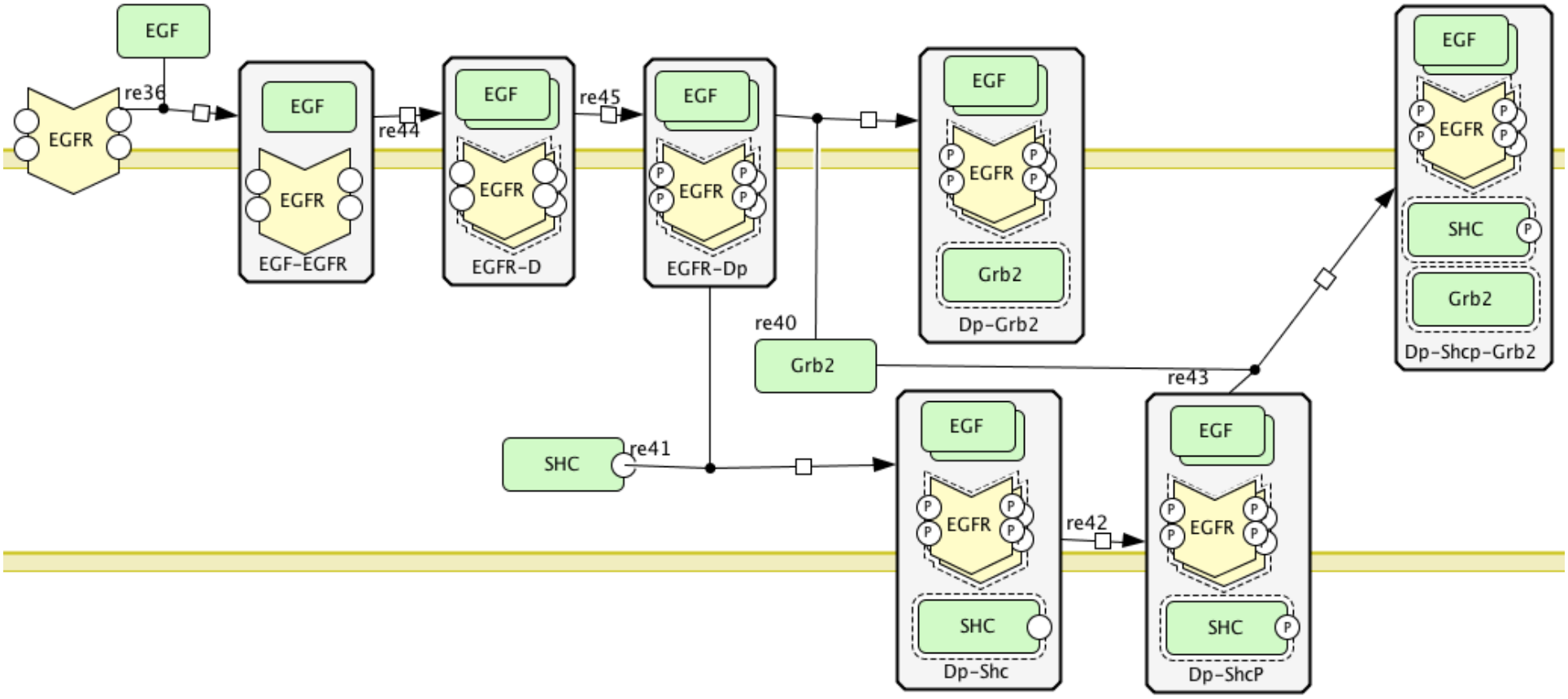


Species: One for every possible modification state of every complex

Reactions: One for every transition among species

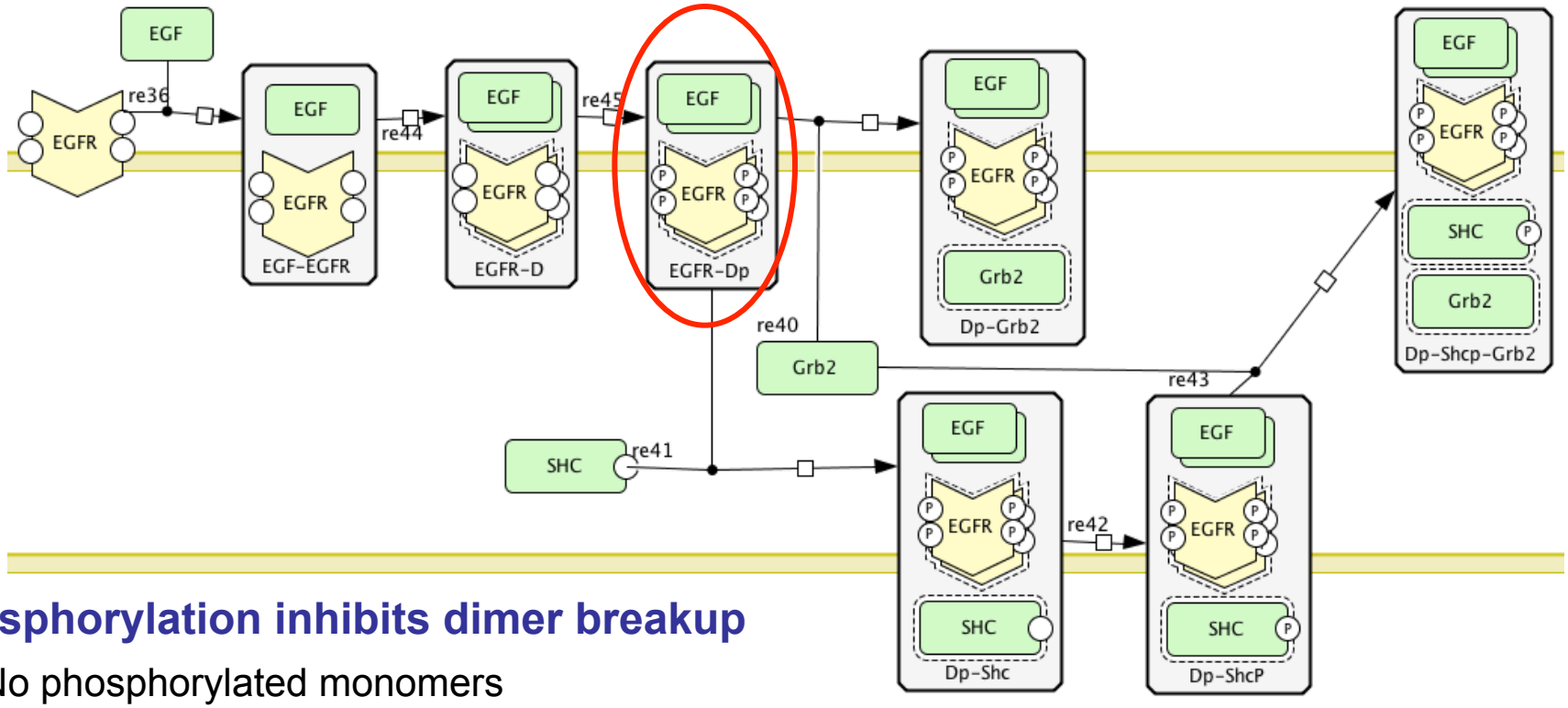
http://vcell.org

Close world" description: SBGN-PD (CellDesigner Drawing)



http://celldesigner.org

“Close world” description enforces very strong assumptions



- **Phosphorylation inhibits dimer breakup**
  - No phosphorylated monomers
  - No association of monomers with other proteins complexes
  - To become dephosphorylated, all proteins should dissociate and dimer should break up
- **No individual phosphosites are considered**
  - Phosphorylation is identical for all phosphosites
  - Same phosphorylation timecourses
  - At most single adapter protein can bind to a dimer

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

Epidermal growth factor receptor (EGFR)

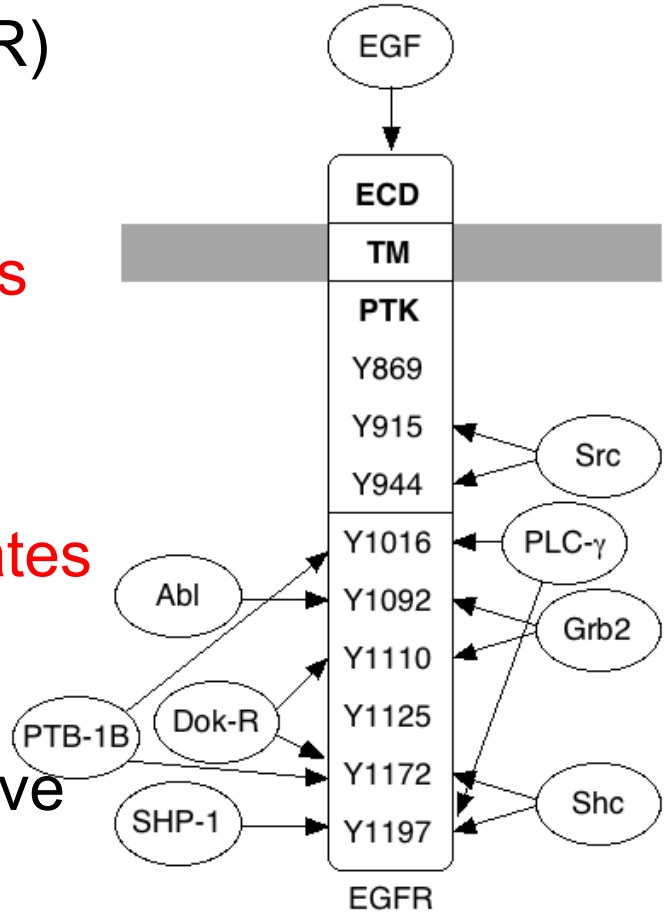
9 sites ⇒ 2^9=512 phosphorylation states

Each site has ≥ 1 binding partner

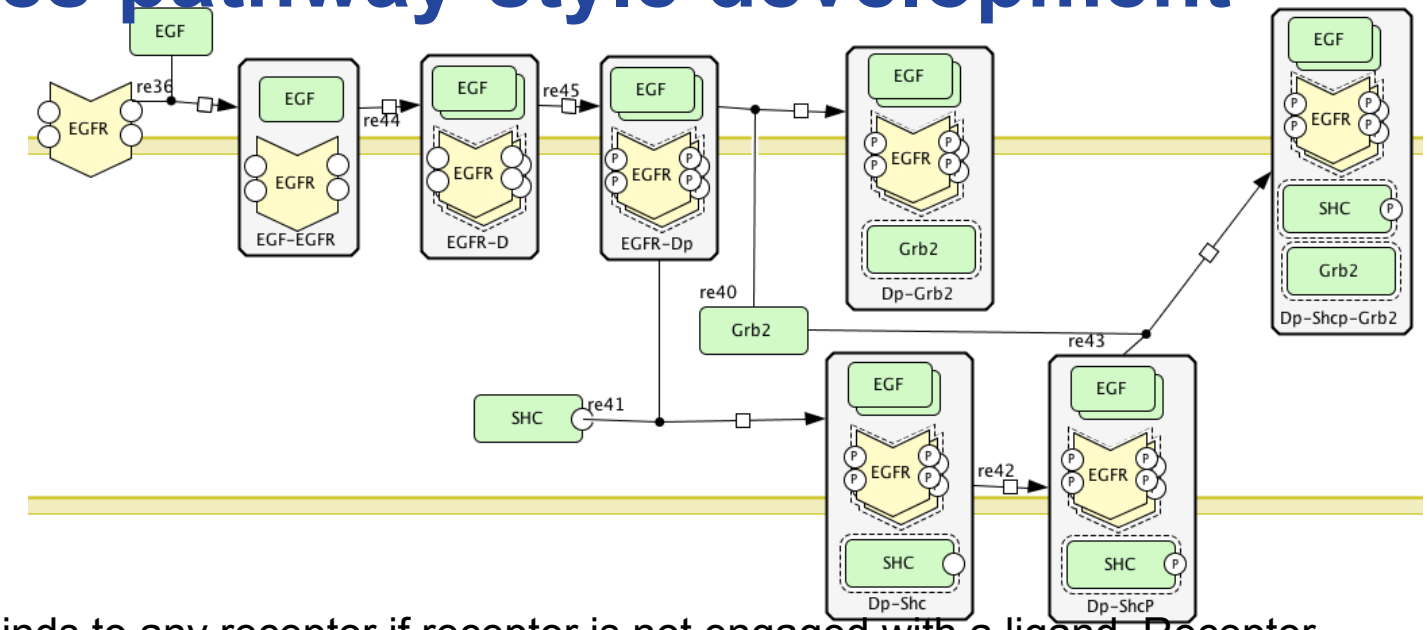
⇒ more than 3^9=19,683 total states

EGFR must form *dimers* to become active

⇒ more than 1.9×10^8 states



The problem: lifting assumptions eliminates pathway-style development



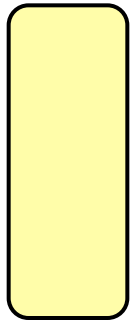
1. Ligand reversibly binds to any receptor if receptor is not engaged with a ligand. Receptor can be a monomer, a dimer, phosphorylated, associated with intracellular proteins, etc
2. Ligand-receptor monomers can dimerize. Monomers can be phosphorylated, unphosphorylated, etc.
3. Once two RTKs are in close proximity, they can transphosphorylate individual phosphosites of each other. Whether ligand is still present is irrelevant.
4. Each receptor phosphotyrosine can bind adapter protein, whether receptor is in monomer, dimer, …

Rule-based modeling

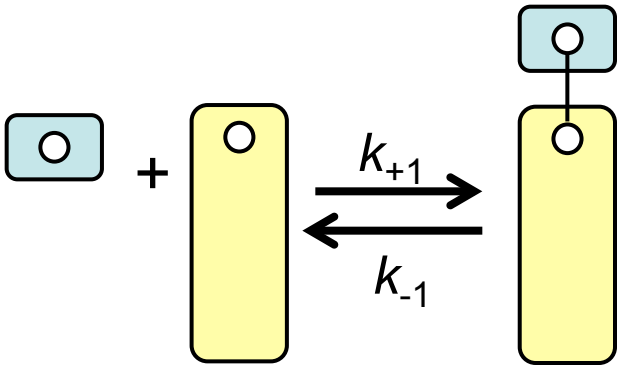
- BioNetGen/NFSim (http://bionetgen.org), Jose-Juan Tapia is here
- Simmune (http://simmune.org), Fengkai Zhang is here
- Virtual Cell/BioNetGen (http://vcell.org), Ion Moraru is here
- Kappa (http://kappalanguage.org), Anatoly Sorokin is here

Molecules, components and rules

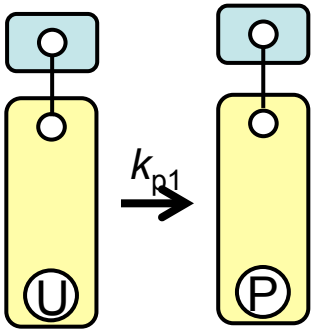
Molecules



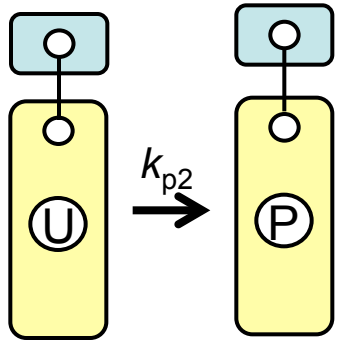
Rule 1



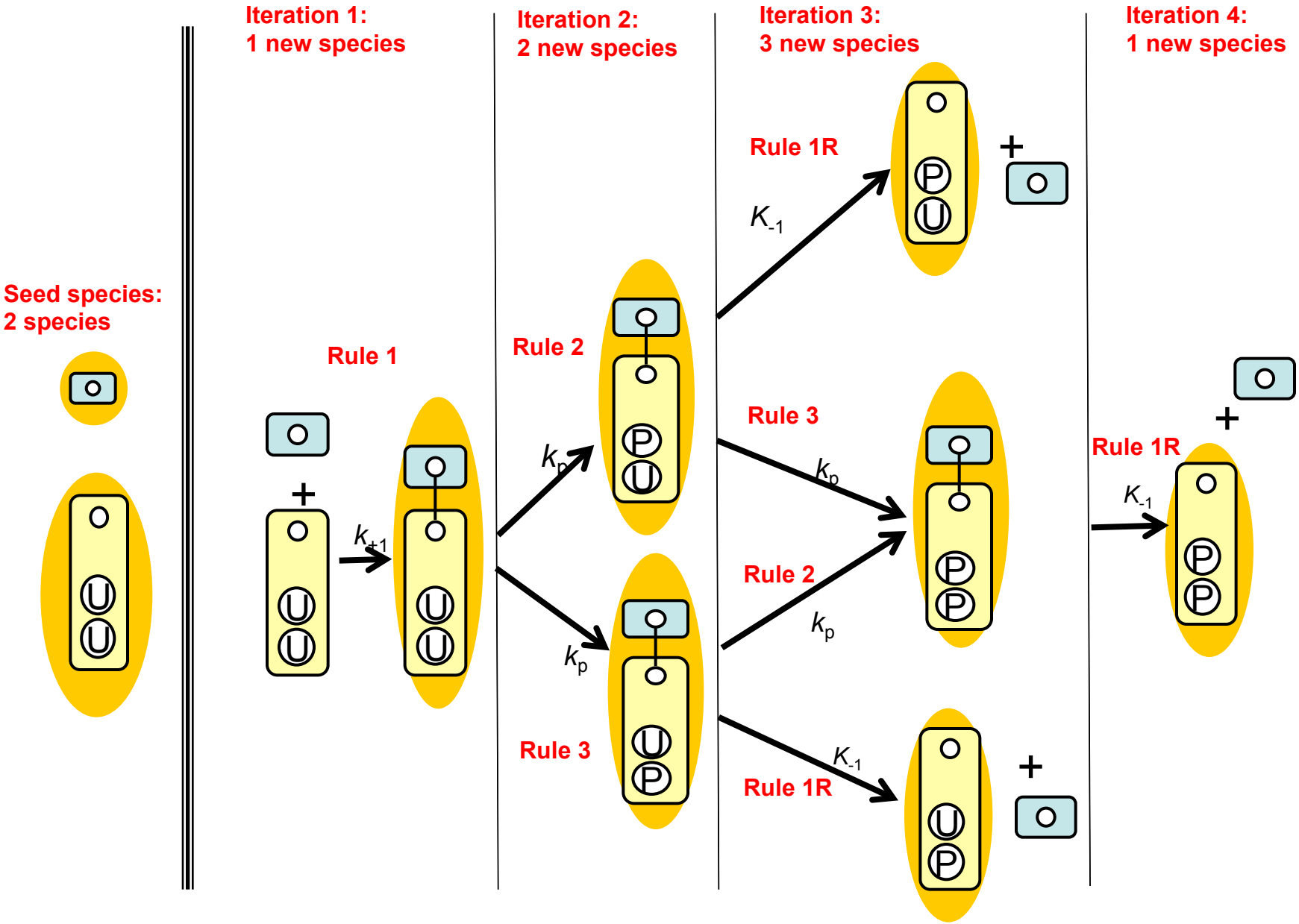
Rule 2



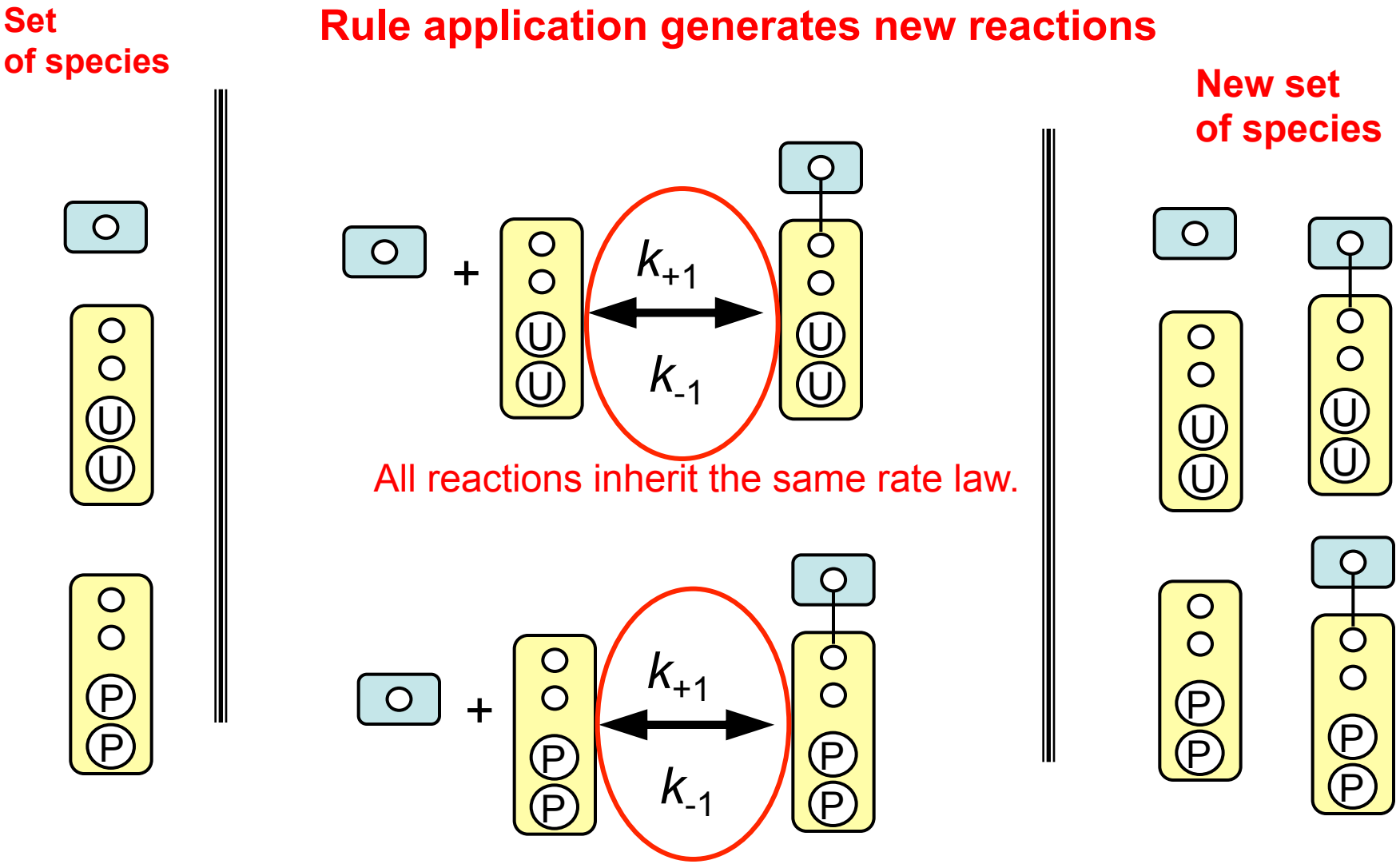
Rule 3



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

....

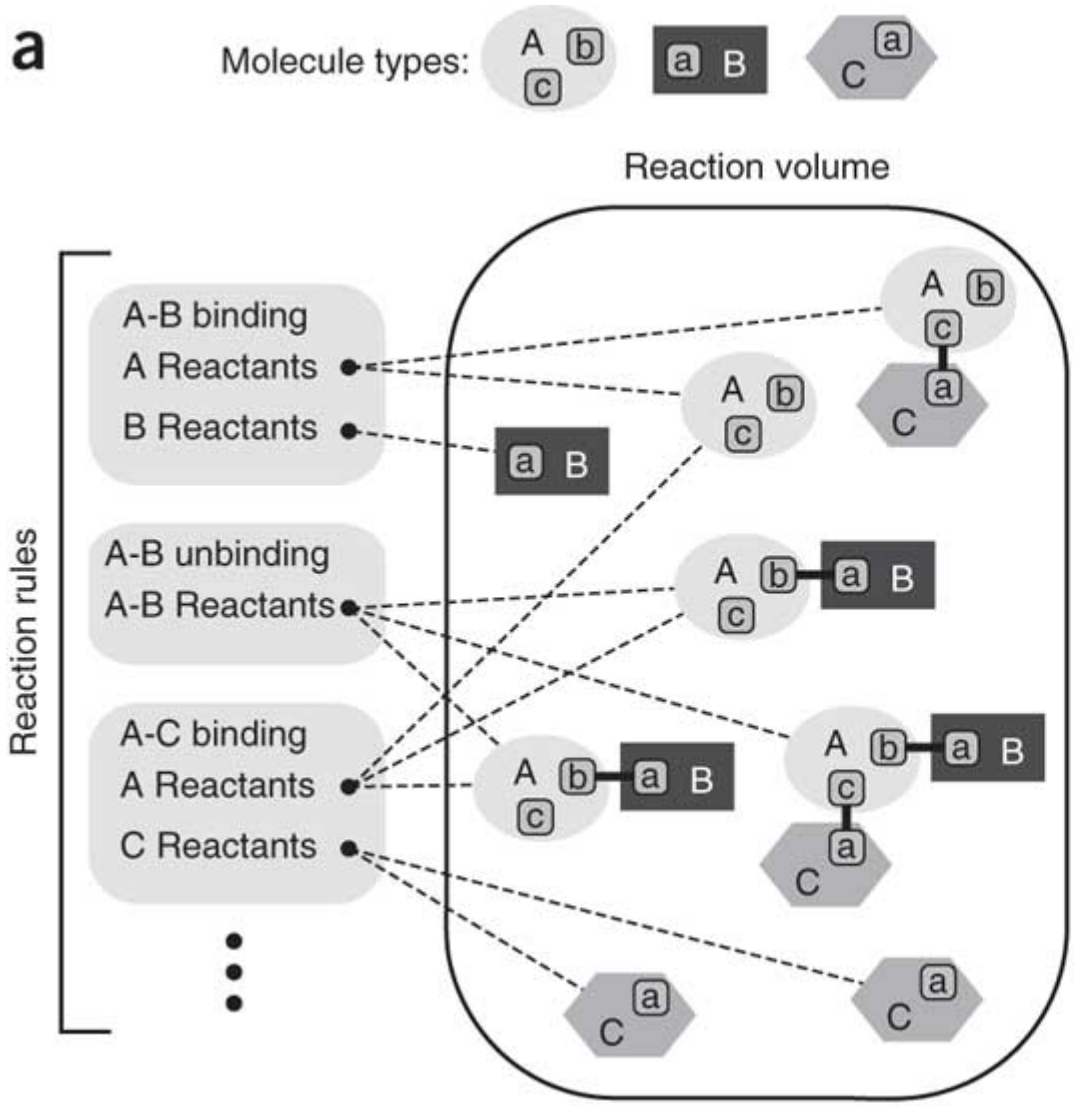
Rules application n $\mathcal{R}(S_n) = R_{n+1}, S_{n+1}$

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

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

Blinov et al. *Bioinformatics, 2004*

NFSim



b

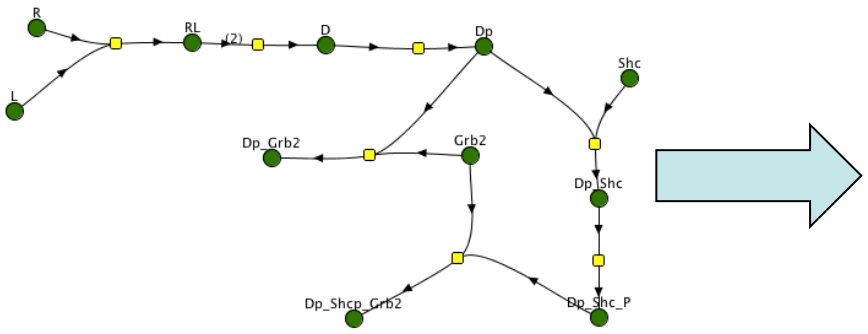
```
begin parameters
   # define constant parameters of the model
   k1 0.02
   . . .
end parameters

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
```

Sneddon et al. *Nat Methods,* **2011**

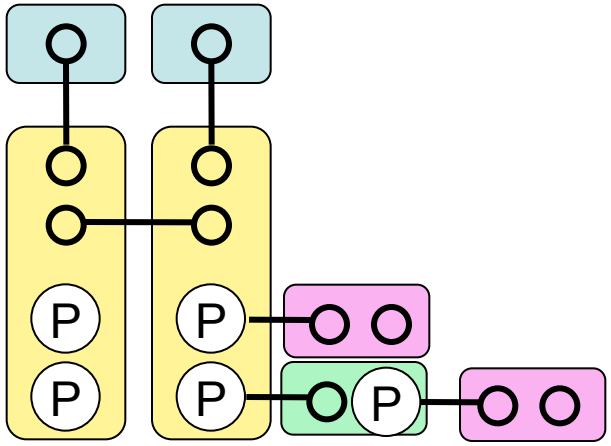
Expanded version of the pathway model

- 4 molecule types
- 11 reaction rules
- No new rate parameters (!)

11 species
7 reactions



~(2x3x5)^2 ~450 species
~1000 reactions



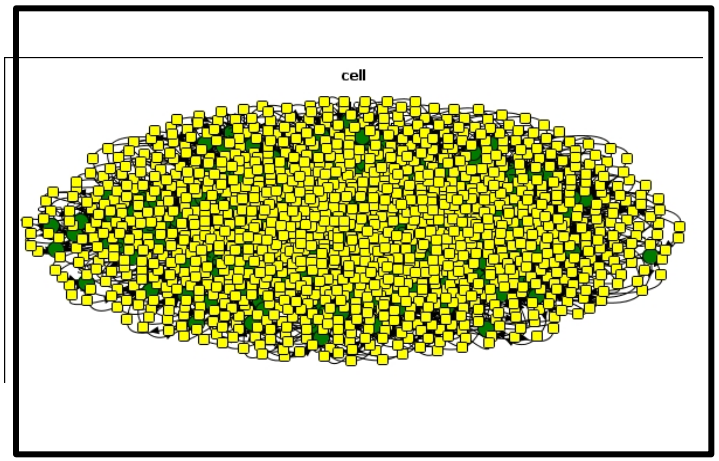
Blinov et al. *Biosystems* 2006

Visualization of rule-based models

Reaction Diagram | Reactions | Structures | Species | Species Types | Observables

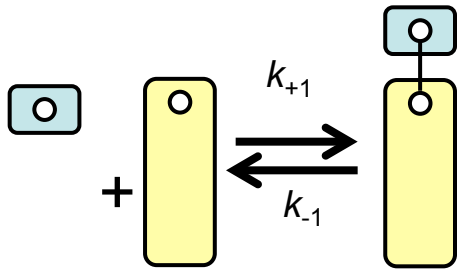
Reaction
egfr(l,r)+egf(r) <-> egfr(l!1,r).egf(r!1)
egfr(l!1,r)+egfr(l!2,r) <-> egfr(l!1,r!3).egfr(l!2,r!3)
egfr(r!1,Y1068~Y) -> egfr(r!1,Y1068~pY)
egfr(r!1,Y1148~Y) -> egfr(r!1,Y1148~pY)
egfr(Y1068~pY) -> egfr(Y1068~Y)
egfr(Y1148~pY) -> egfr(Y1148~Y)
egfr(r!2,Y1148~pY!1).Shc(PTB!1,Y317~Y) -> egfr(r!2,Y1148~pY!1).Shc(PTB!1,Y317~pY)
Shc(PTB!1,Y317~pY) -> Shc(PTB!1,Y317~Y)
egfr(Y1068~pY)+Grb2(SH2,SH3) <-> egfr(Y1068~pY!1).Grb2(SH2!1,SH3)
egfr(Y1068~pY)+Grb2(SH2,SH3!2) <-> egfr(Y1068~pY!1).Grb2(SH2!1,SH3!2)
egfr(Y1068~pY!1).Grb2(SH2!1,SH3)+Sos(dom) <-> egfr(Y1068~pY!1).Grb2(SH2!1,SH3!2).Sos(dom!2)
egfr(Y1148~pY)+Shc(PTB,Y317~Y) <-> egfr(Y1148~pY!1).Shc(PTB!1,Y317~Y)
egfr(Y1148~pY)+Shc(PTB,Y317~pY) <-> egfr(Y1148~pY!1).Shc(PTB!1,Y317~pY)
egfr(Y1148~pY)+Shc(PTB,Y317~pY!1).Grb2(SH2!1,SH3) <-> egfr(Y1148~pY!2).Shc(PTB!2,Y317~pY!1).Grb2(SH2!1,SH3)
egfr(Y1148~pY)+Shc(PTB,Y317~pY!1).Grb2(SH2!1,SH3!3).Sos(dom!3) <-> egfr(Y1148~pY!2).Shc(PTB!2,Y317~pY!1).Grb2(SH2!1,SH3!3).Sos(dom!3)
egfr(Y1148~pY!1).Shc(PTB!1,Y317~pY)+Grb2(SH2,SH3) <-> egfr(Y1148~pY!1).Shc(PTB!1,Y317~pY!2).Grb2(SH2!2,SH3)
egfr(Y1148~pY!1).Shc(PTB!1,Y317~pY)+Grb2(SH2,SH3!3).Sos(dom!3) <-> egfr(Y1148~pY!1).Shc(PTB!1,Y317~pY!2).Grb2(SH2!2,SH3!3).Sos(dom!3)
Shc(PTB!1,Y317~pY!2).Grb2(SH2!2,SH3)+Sos(dom) <-> Shc(PTB!1,Y317~pY!2).Grb2(SH2!2,SH3!3).Sos(dom!3)
Shc(PTB,Y317~pY)+Grb2(SH2,SH3) <-> Shc(PTB,Y317~pY!1).Grb2(SH2!1,SH3)
Shc(PTB,Y317~pY)+Grb2(SH2,SH3!2) <-> Shc(PTB,Y317~pY!1).Grb2(SH2!1,SH3!2)
Shc(PTB,Y317~pY) -> Shc(PTB,Y317~Y)
Grb2(SH2,SH3)+Sos(dom) <-> Grb2(SH2,SH3!1).Sos(dom!1)
Shc(PTB,Y317~pY!2).Grb2(SH2!2,SH3)+Sos(dom) <-> Shc(PTB,Y317~pY!2).Grb2(SH2!2,SH3!3).Sos(dom!3)

Reading/writing code is not easy…

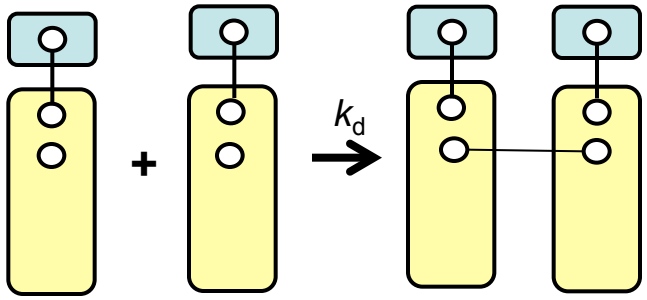


Visualization

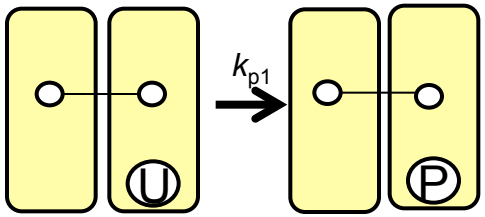
Rule 1: Ligand binds any receptor



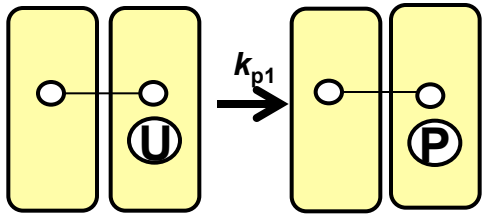
Rule 2: Any two ligand-associated monomers can dimerize



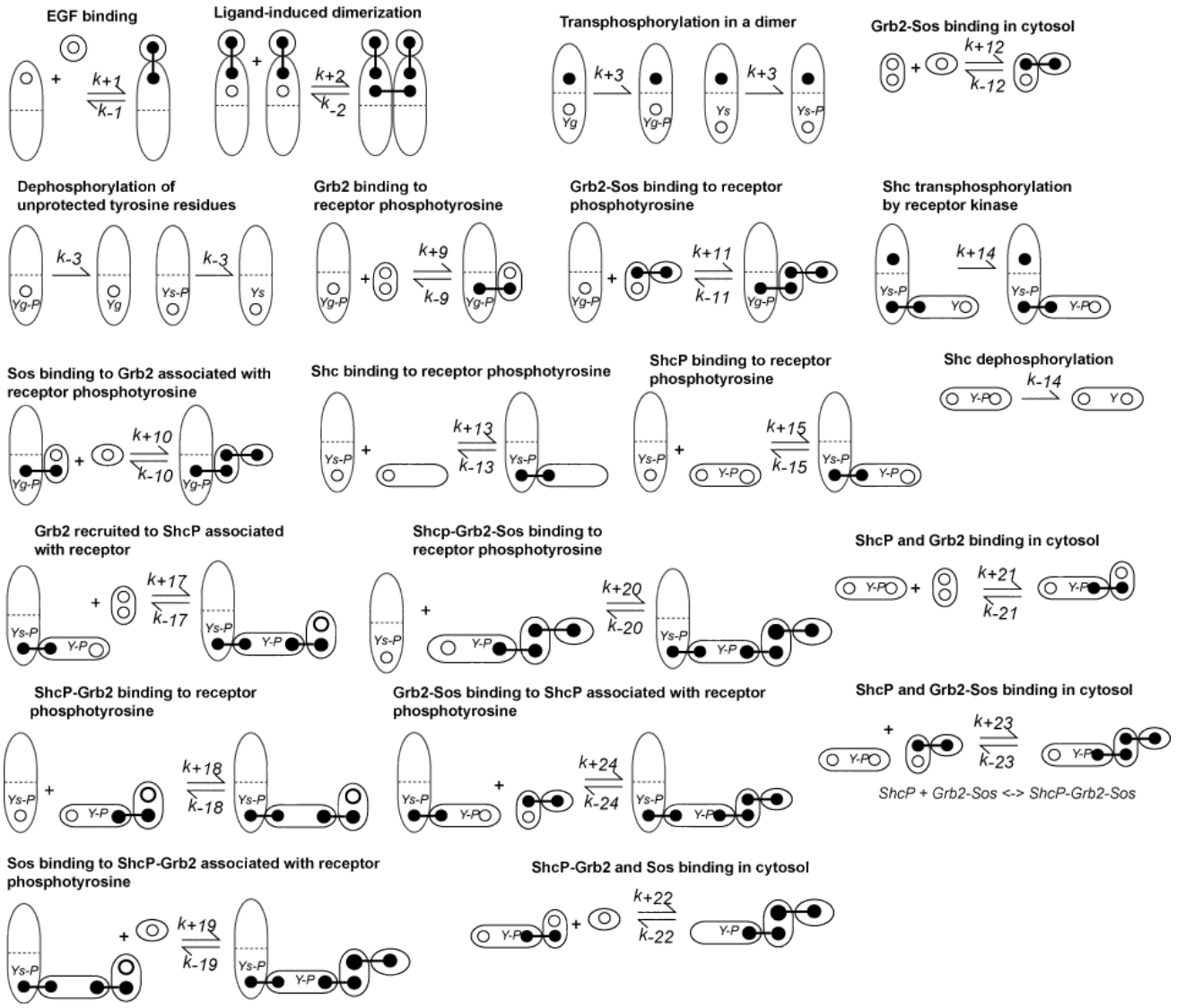
Rule 3: residue 1 becomes trahsphosphorylated in a dimer



Rule 4: residue 2 becomes trahsphosphorylated in a dimer



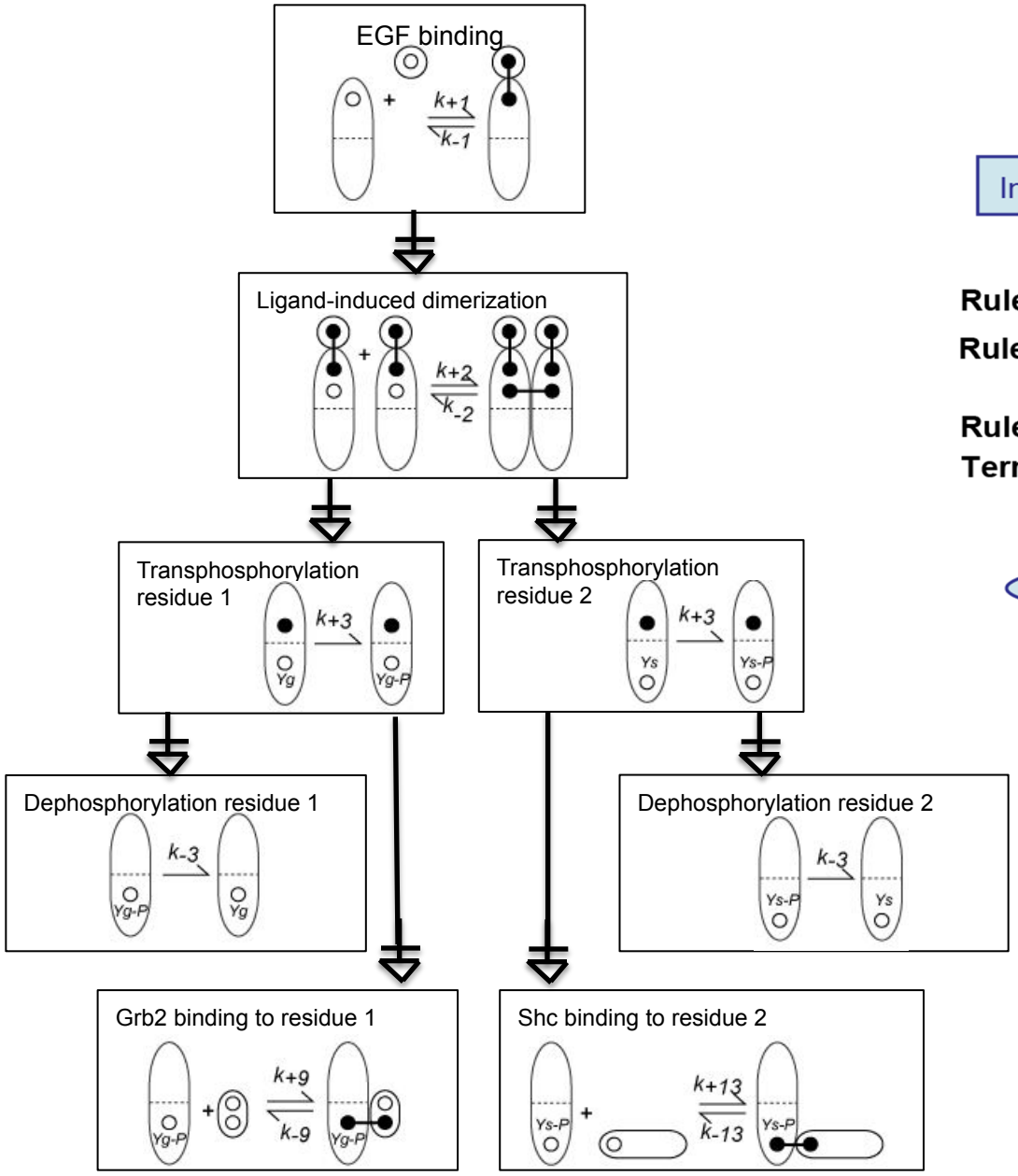
Visualization of RB models: cartoons



Compare SBGN Languages

	AF	PD	ER
Describes:	Influence of biological activities on each other	Conversion physical entities into other entities, change their states or location	Interactions between entities and the rules that control them
Level of description:	Conceptual description of influences	Mechanistic	Mechanistic
Building blocks:	Different activities of physical entities are represented separately	Different states of physical entities are represented separately	Physical entities are represented only once
Ambiguity:	Ambiguous interpretation in biochemical terms	Completely unambiguous	May be ambiguous
Temporality:	Sequential influences	Sequential	Non-sequential
Pitfalls:	Not suitable to represent mechanistic details	Sensitive to combinatorial explosion	Sequence of events is not easily recovered

Visualization of Rule-Based Modes: Activity Flow



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

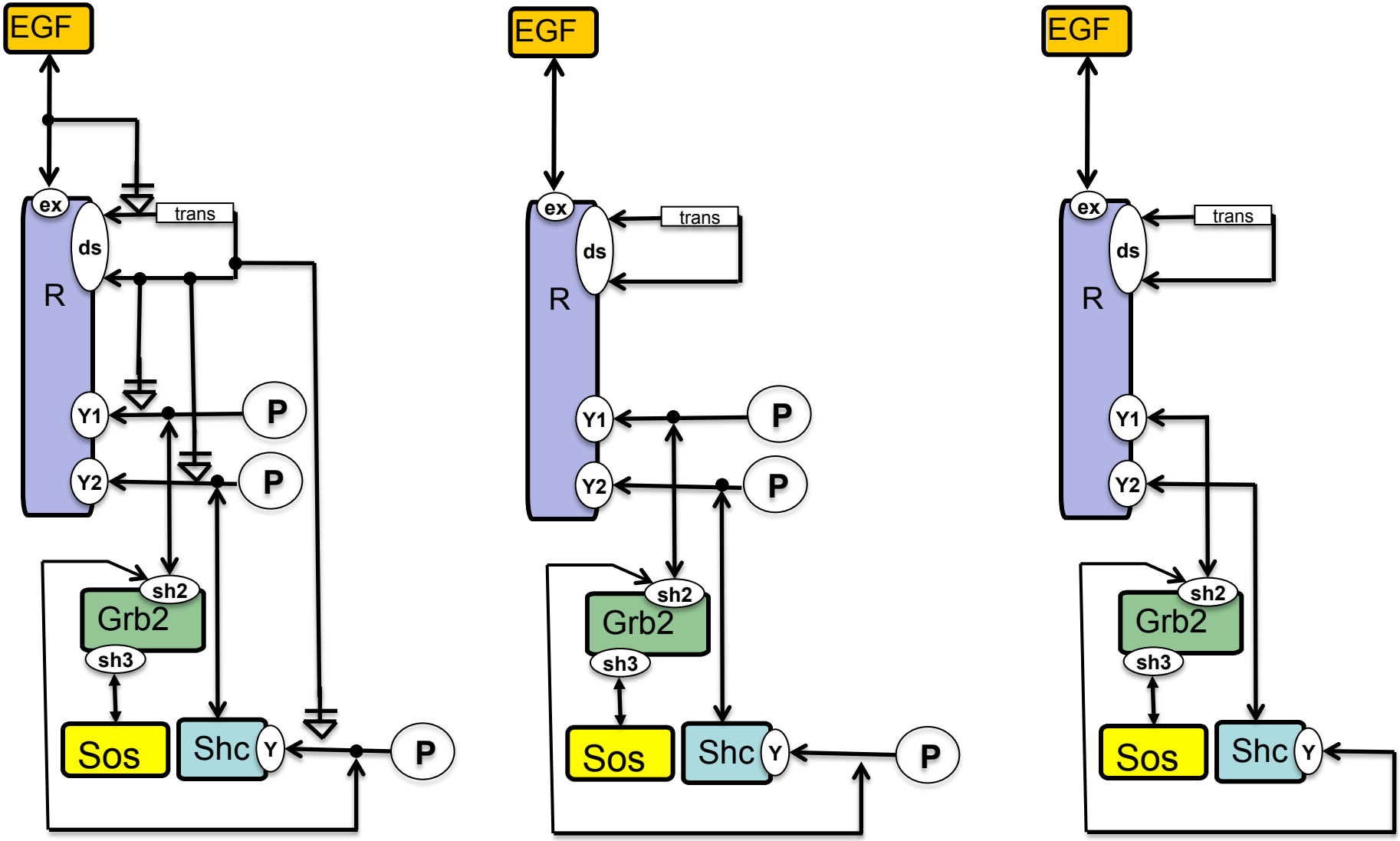
Termination Terminate if $S_n = S_{n+1}$

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

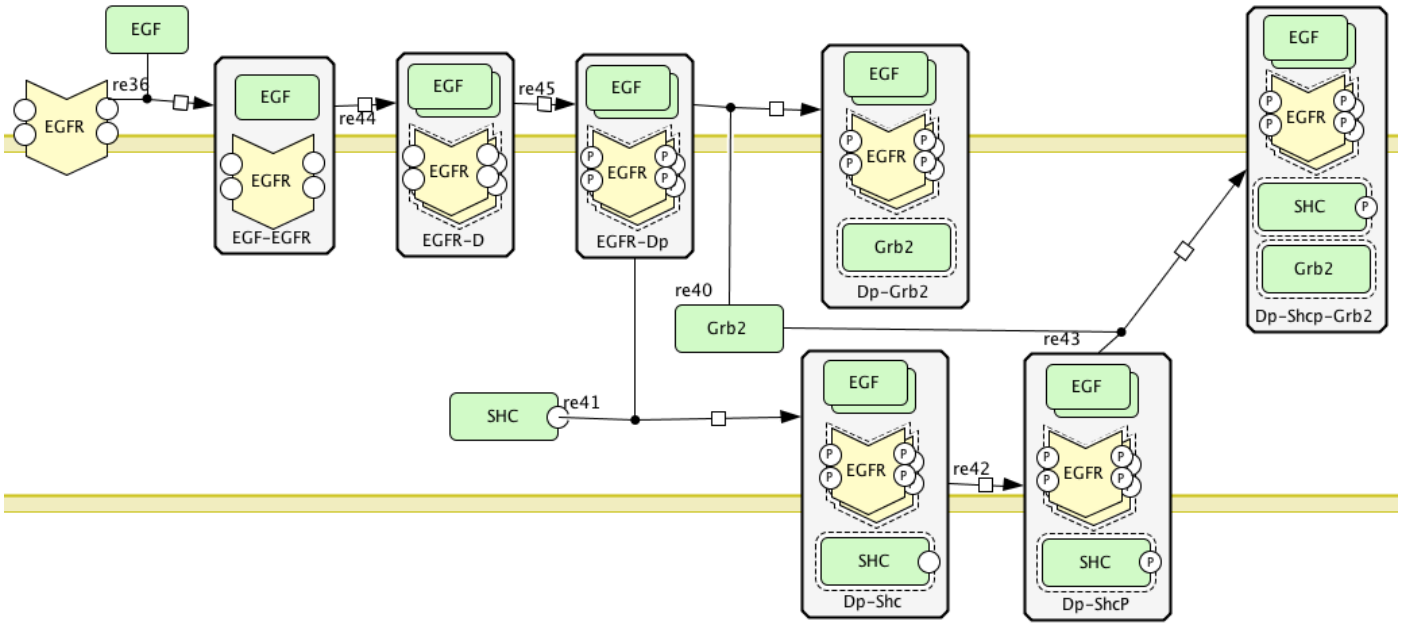
Drawing cartoons is not easy…

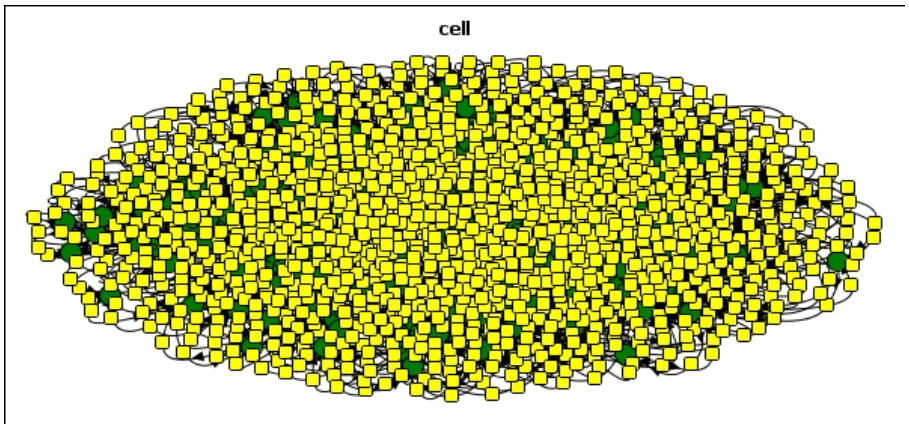
Such visualization can be done only after rules specification, not before

Visualization of Rule Based models: Entity Relationships Diagram

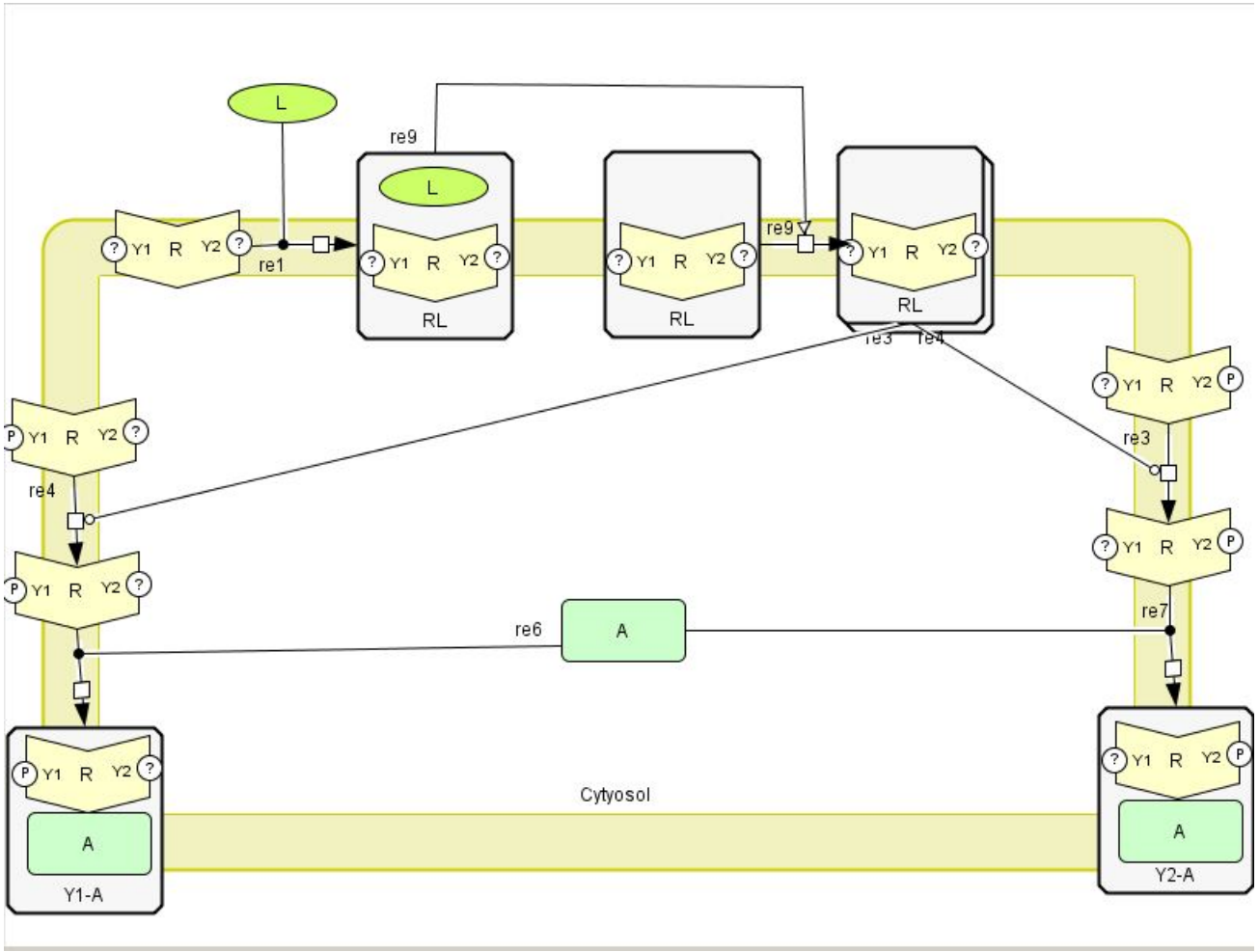


Visualization of RB models: Process Diagrams

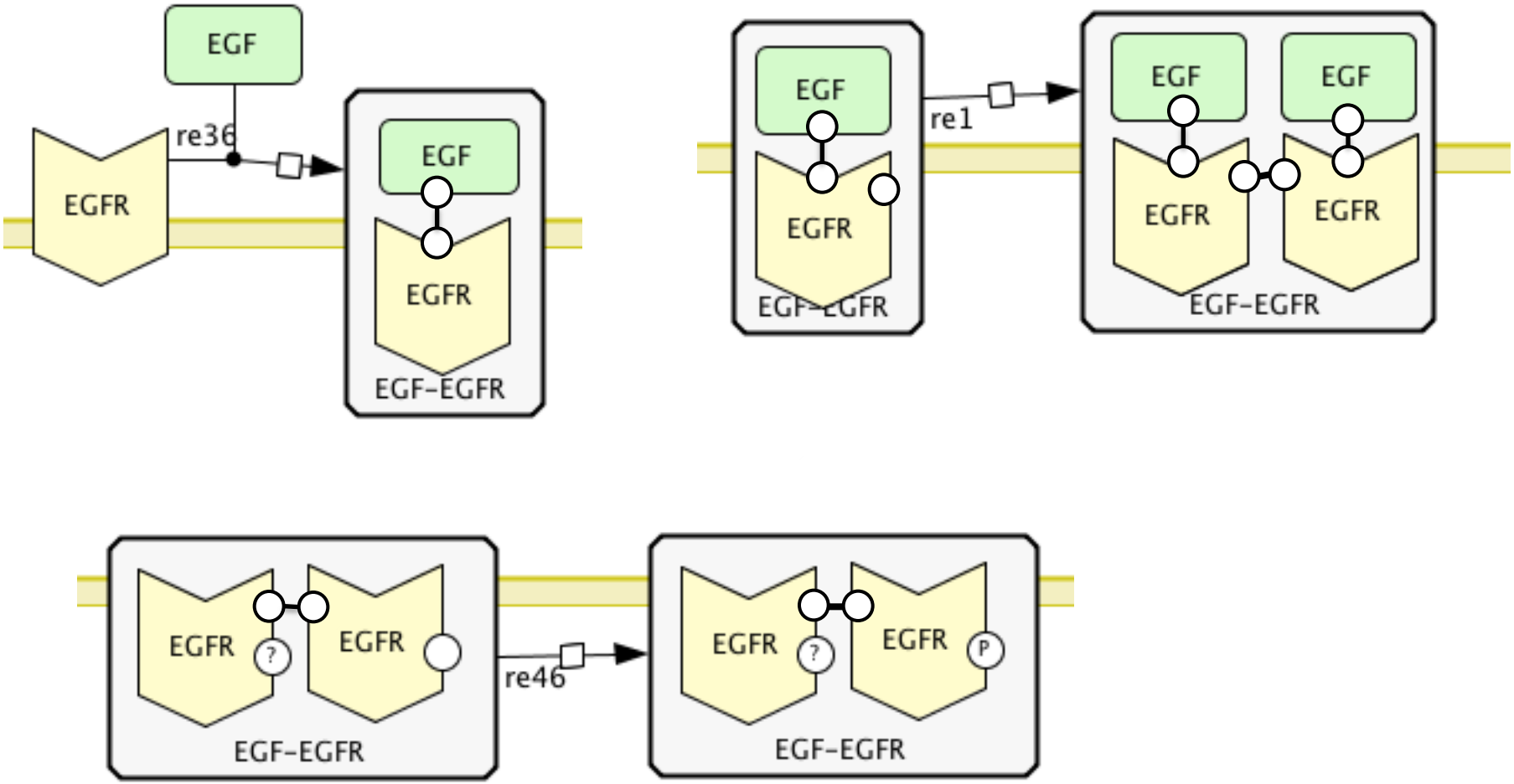




Process Diagrams: naive cartoons

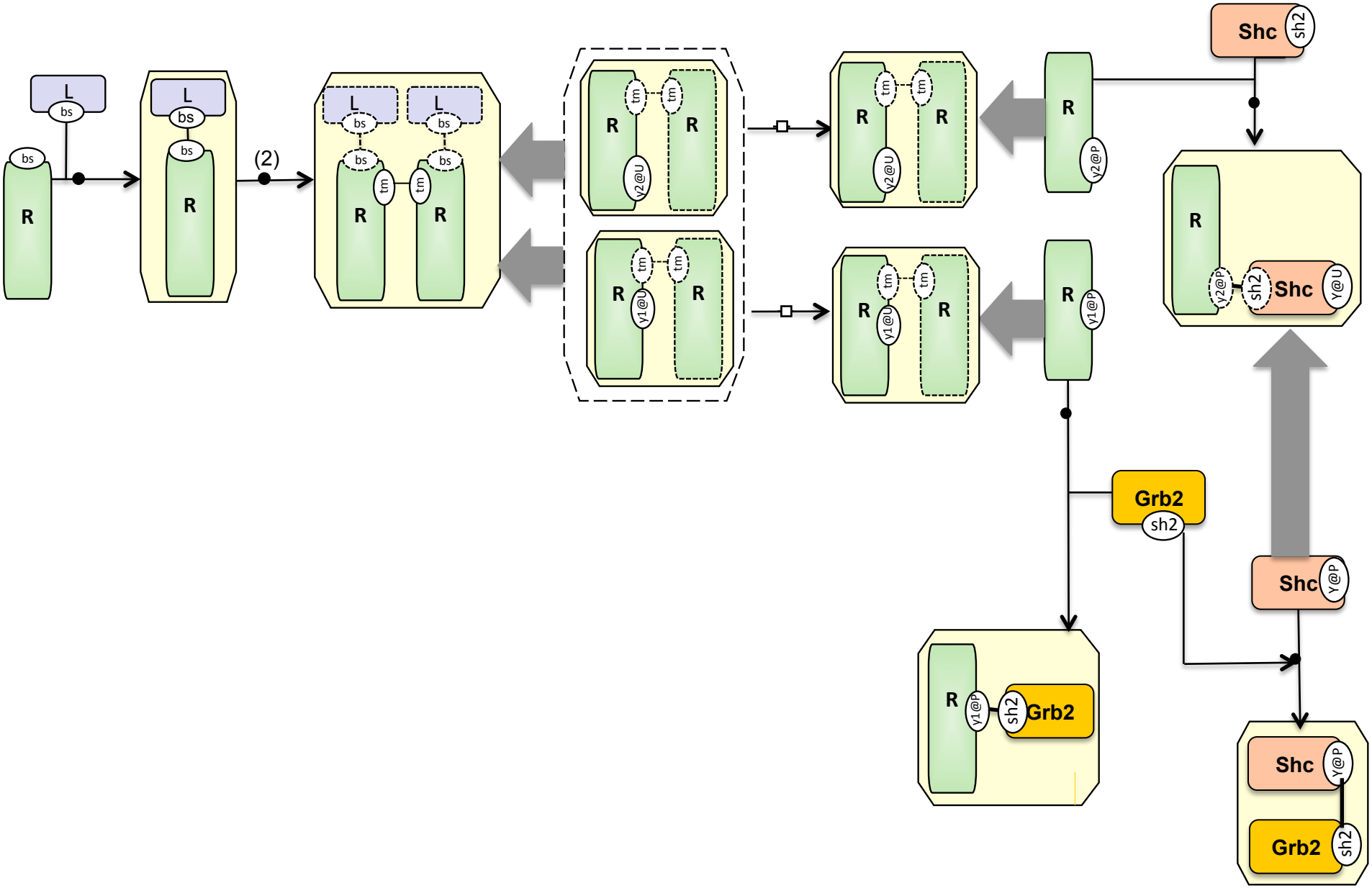


Open-world Process Diagram

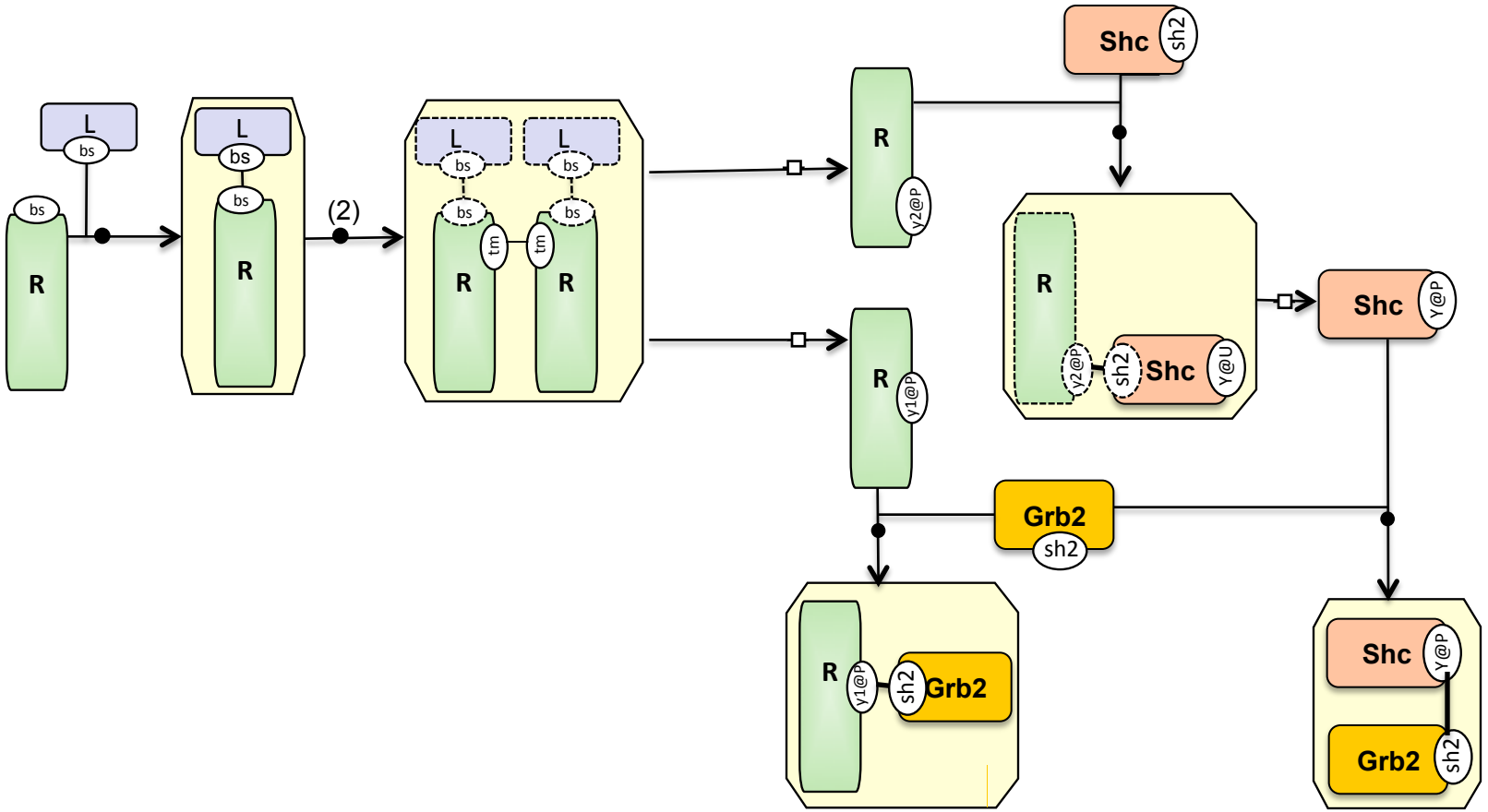


Blinov et al., Nature Biotech 2006

Open-world Process Diagram: subsets



Open-world Process Diagram: short notations



Acknowledgement

- SBGN team (http://sbgn.org)
- Virtual Cell team (http://vcell.org), Ion Moraru is here
- BioNetGen/NFSim team (http://bionetgen.org), Jose-Juan Tapia is here
- Simmune Team (http://simmune.org), Fengkai Zhang is here