Visualization of Rule-based Modeling – Virtual Cell & Systems Biology Graphical Notations (SBGN) approaches

Michael Blinov

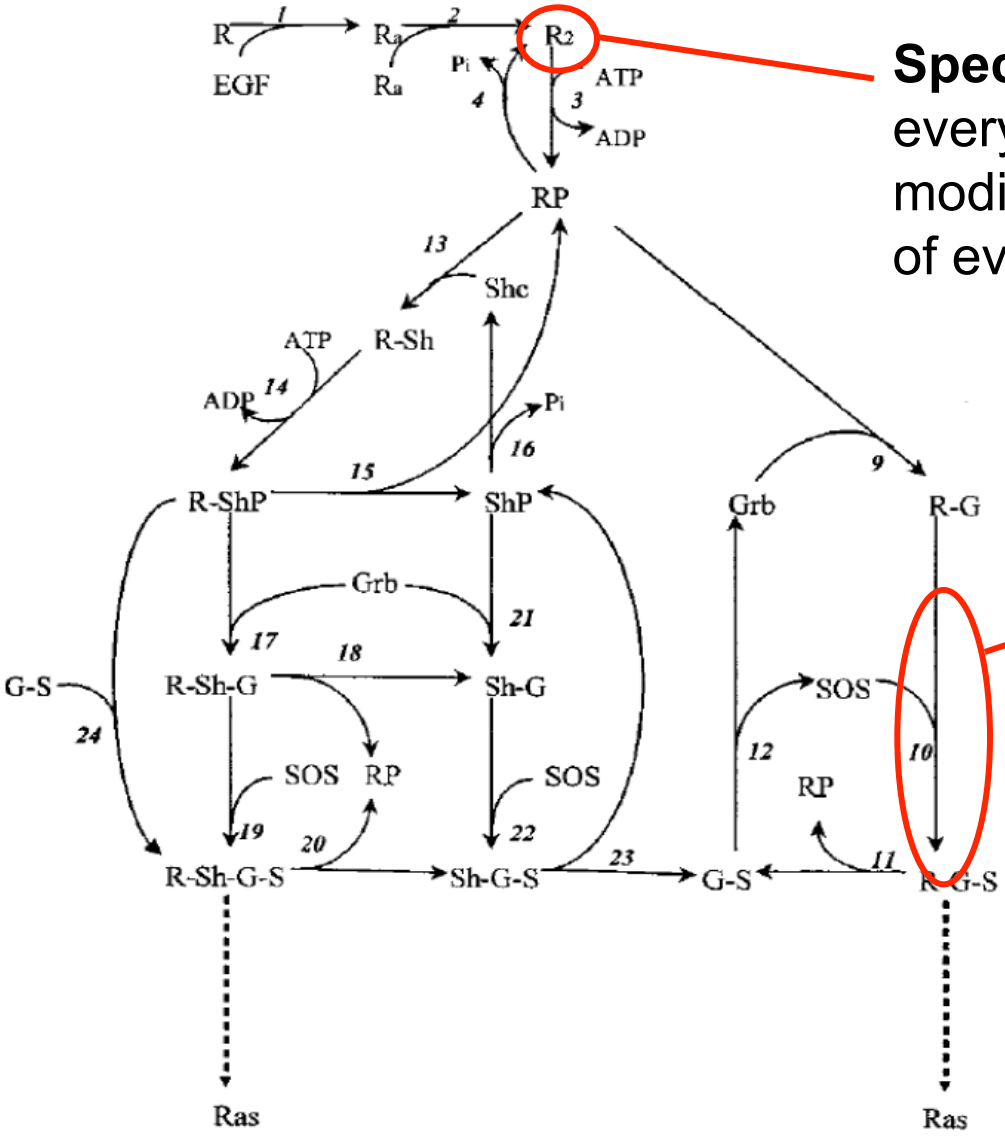
RIP CCAM, August 15th, 2014

Acknowledgement

- Jim Schaff & Ion Moraru for brainstorming sessions
- Fei Gao and Dan Vasilescu for code implementation
- Stuart Moody and Anatoly Sorokin for discussions at SBGN meetings
- CCAM team
- SBGN community

A typical reaction network

18 species
34 reactions

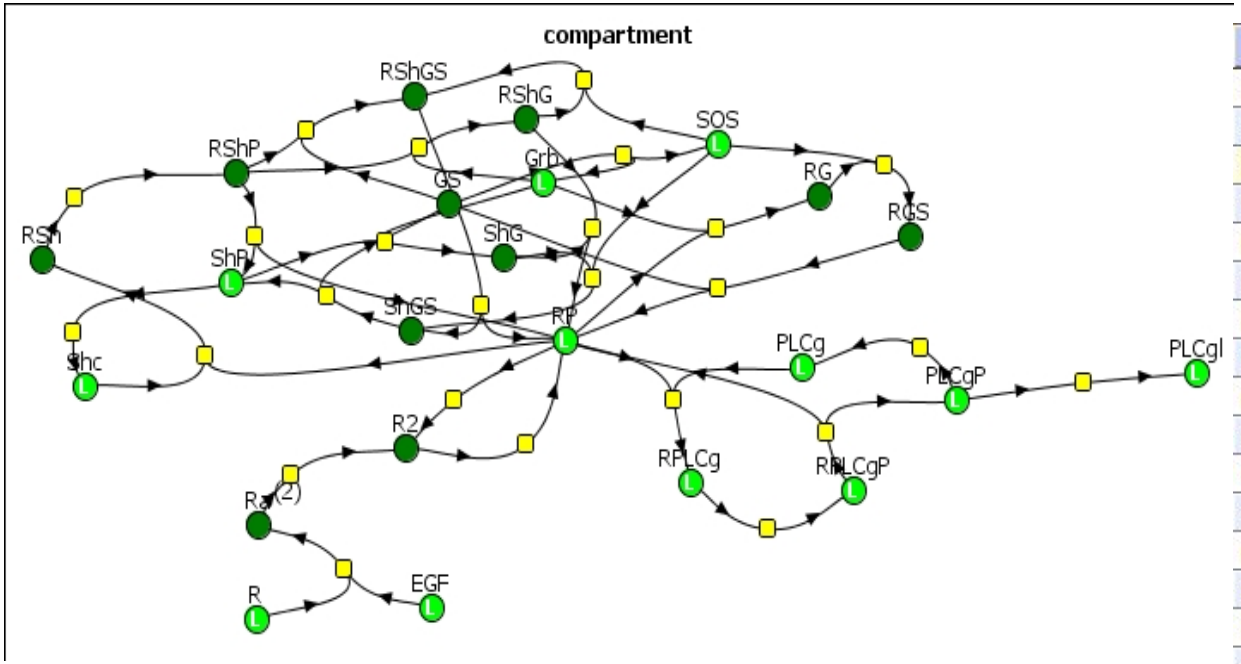


Species: One for every possible modification state of every complex

Reactions: One for every transition among species

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

Reaction network visualization



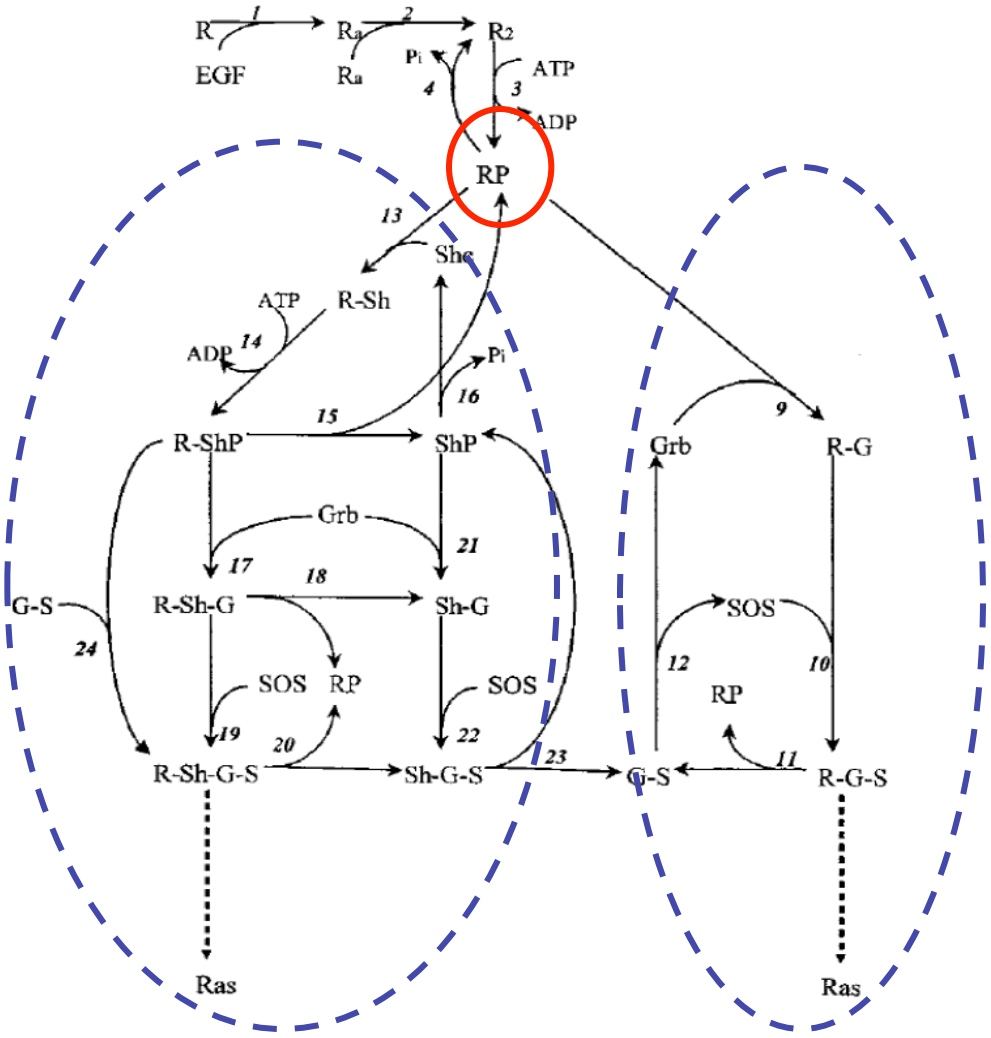
Reaction
EGF + R -> Ra
2Ra -> R2
R2 -> RP
RP -> R2
RP + PLCg -> RPLCg
RPLCg -> RPLCgP
RPLCgP -> RP + PLCgP
PLCgP -> PLCg
RP + Grb -> RG
RG + SOS -> RGS
RGS -> RP + GS
GS -> Grb + SOS
RP + Shc -> RSh
RSh -> RShP
RShP -> RP + ShP
ShP -> Shc
Grb + RShP -> RShG
RShG -> RP + ShG
SOS + RShG -> RShGS
RShGS -> RP + ShGS
Grb + ShP -> ShG
SOS + ShG -> ShGS
ShGS -> GS + ShP
GS + RShP -> RShGS
PLCgP -> PLCgI

VCell provides “minimalistic” description:

- –Bipartie graph (species and reaction nodes)
- –Single type of species nodes and two types of arrows
- –Many automatic layouts

Assumptions made

- **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



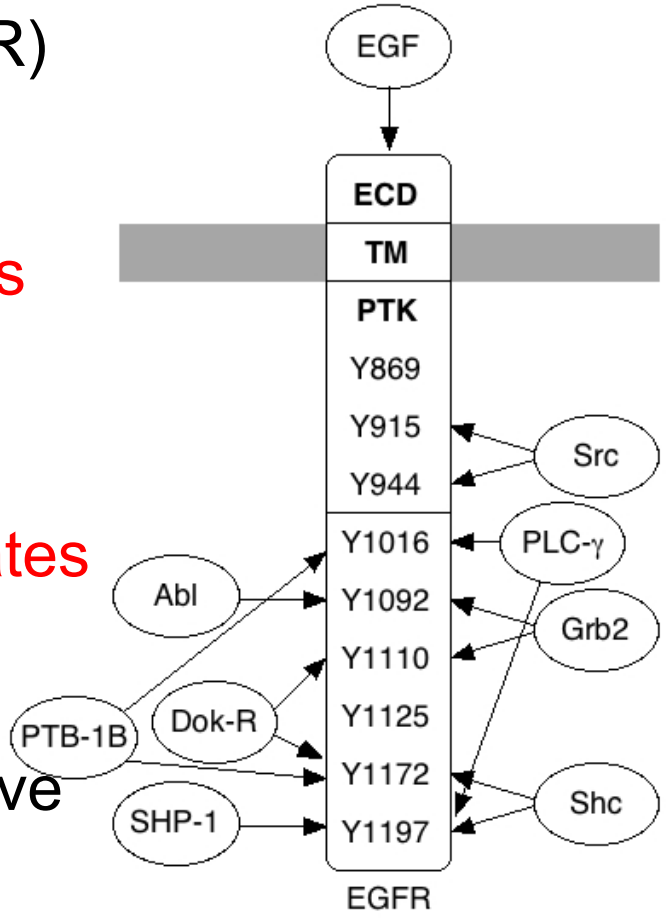
The problem: multiplicity of sites and binding partners gives rise to combinatorial complexity

Epidermal growth factor receptor (EGFR)

9 sites ⇒ 2^9=512 phosphorylation states

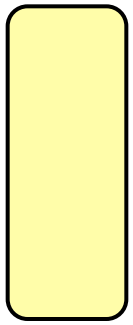
Each site has ≥ 1 binding partner
⇒ more than 3^9=19,683 total states

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

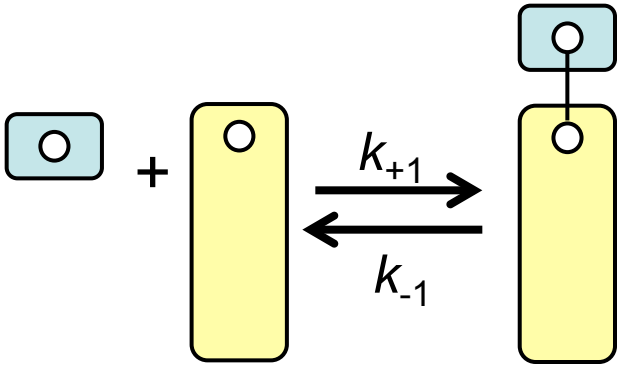


Molecules, components and rules

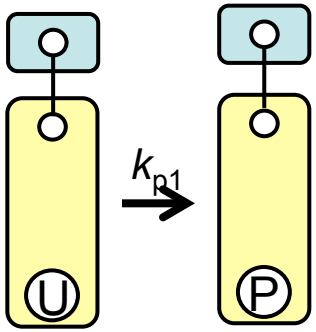
Molecules



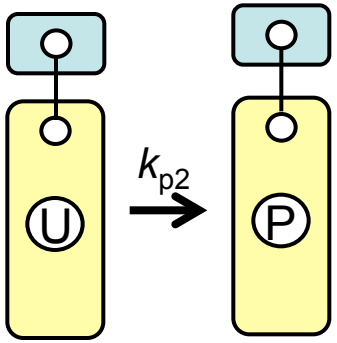
Rule 1



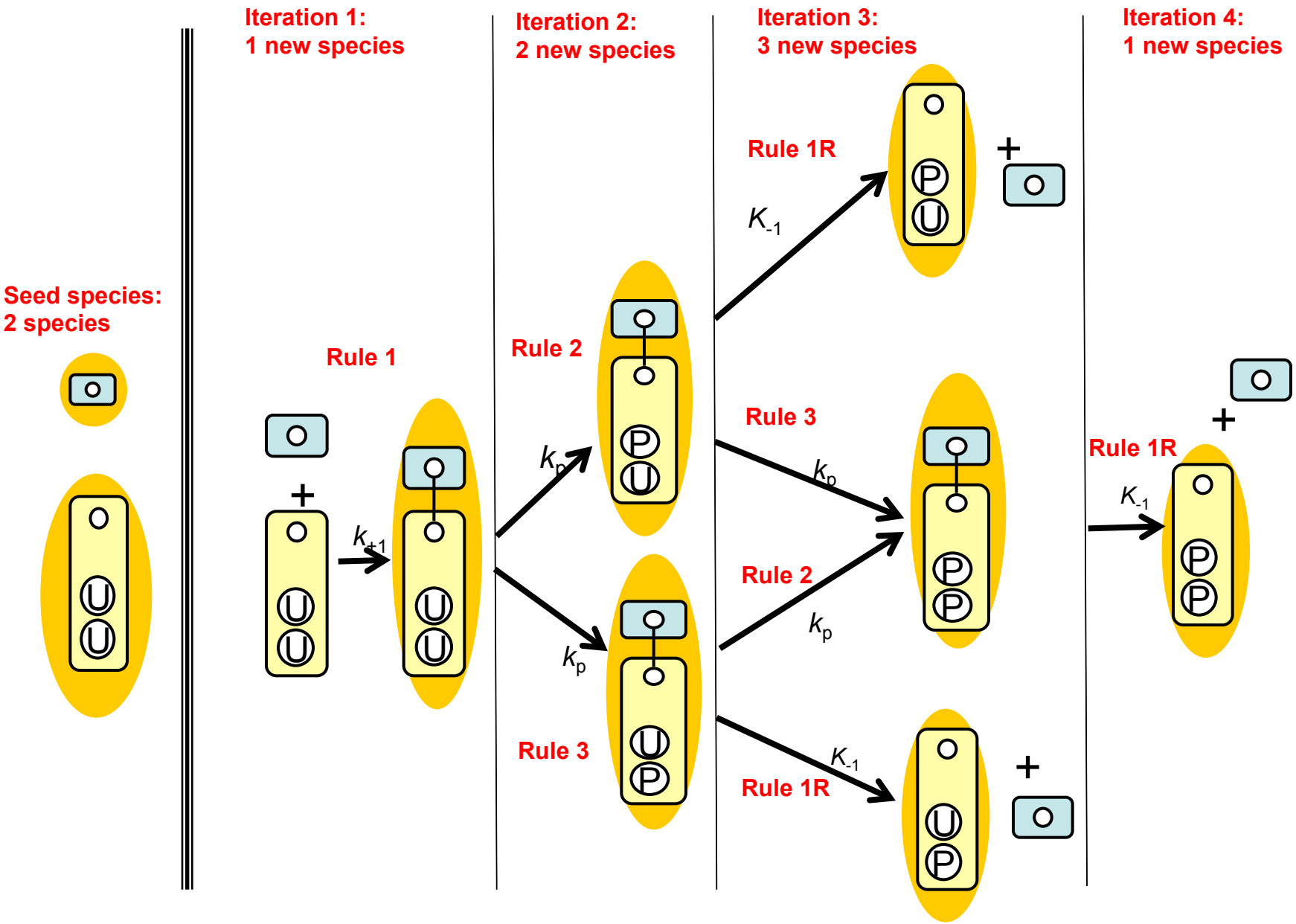
Rule 2



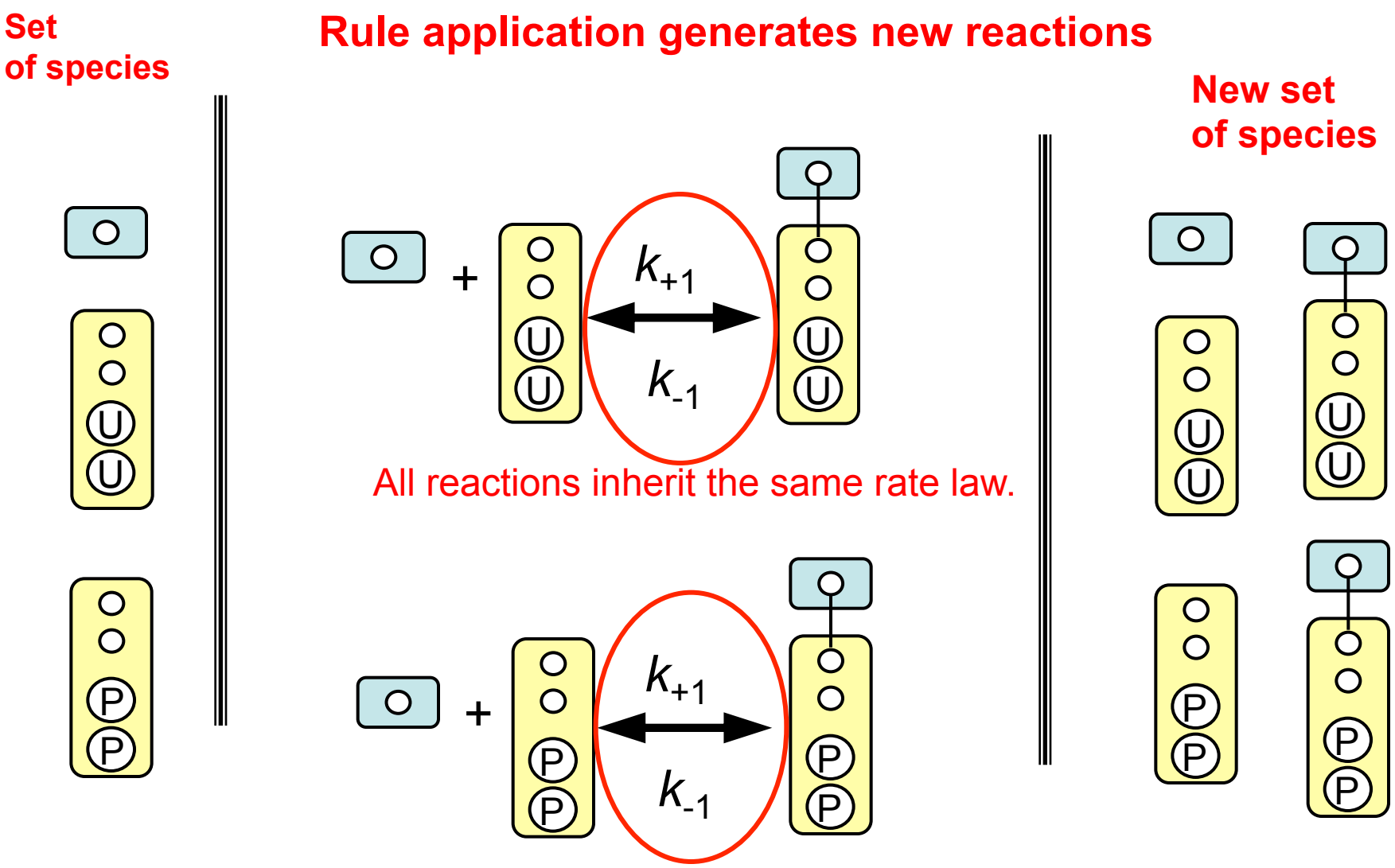
Rule 3



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

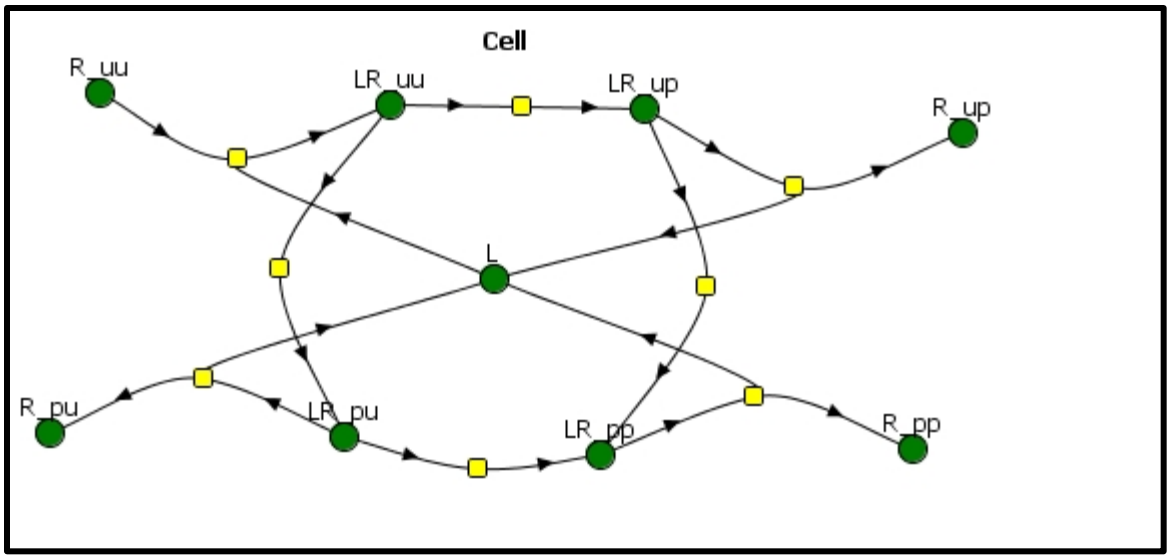
....

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

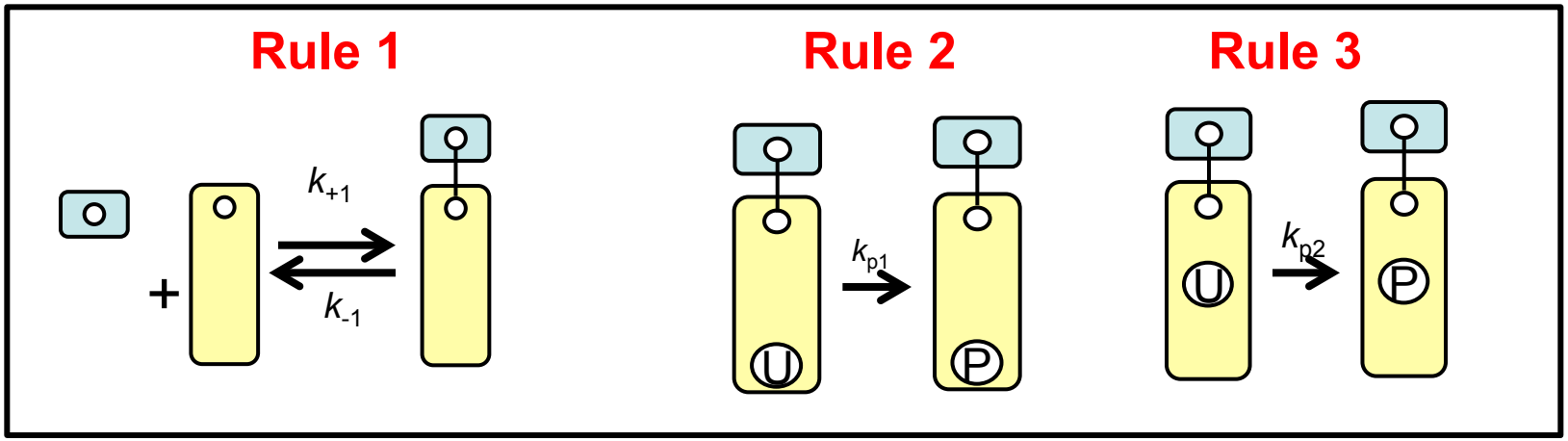
Termination Terminate if $S_n = S_{n+1}$

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

Visualization



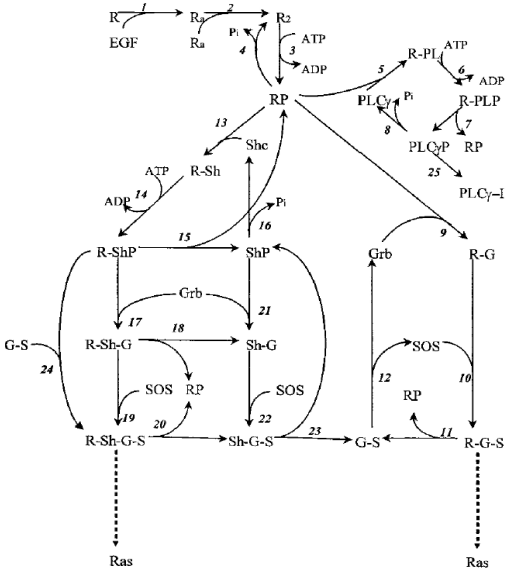
VS



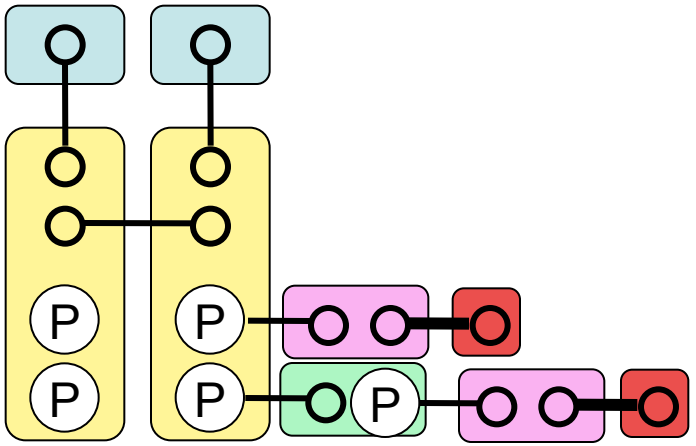
Expanded version of the Kholodenko model

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

18 species
34 reactions

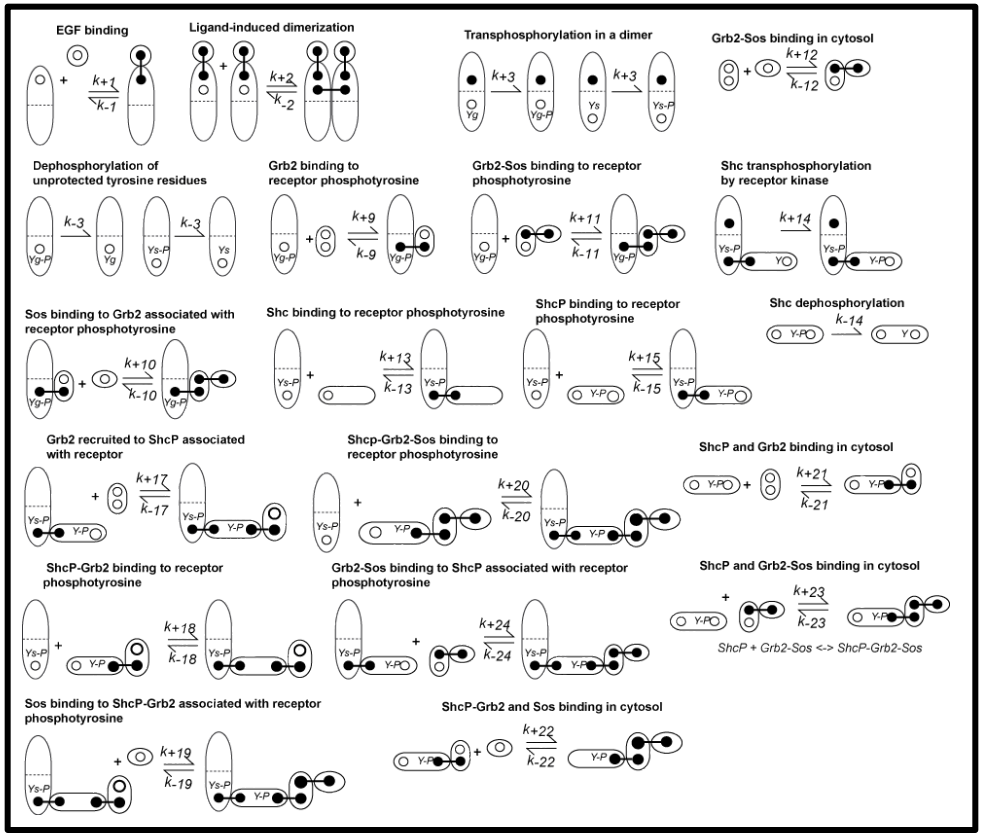


356 species
3,749 reactions

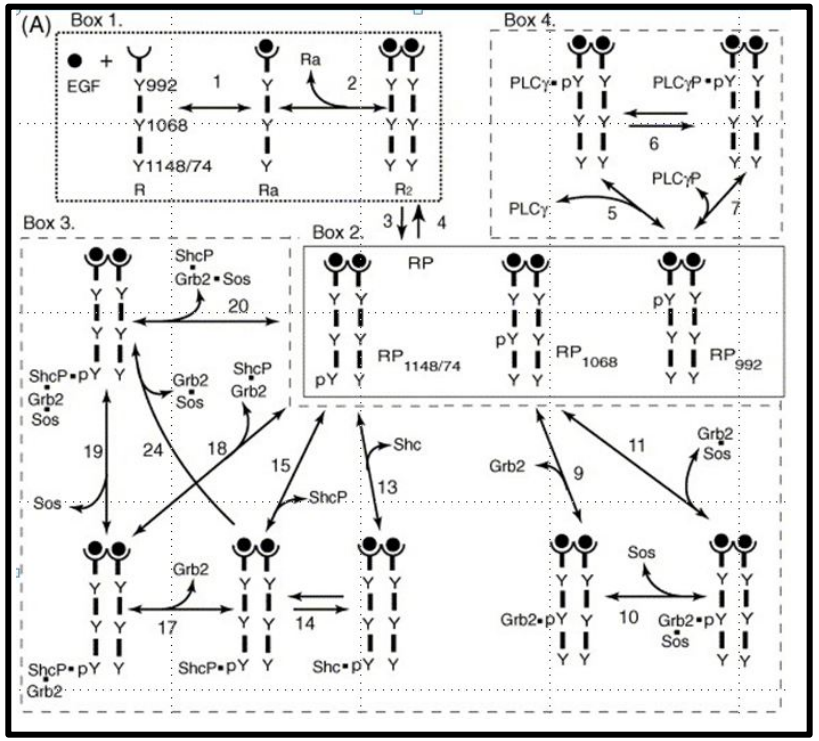


Blinov et al. *Biosystems* **83**, 136 (2006)

Visualization of rule-based models…



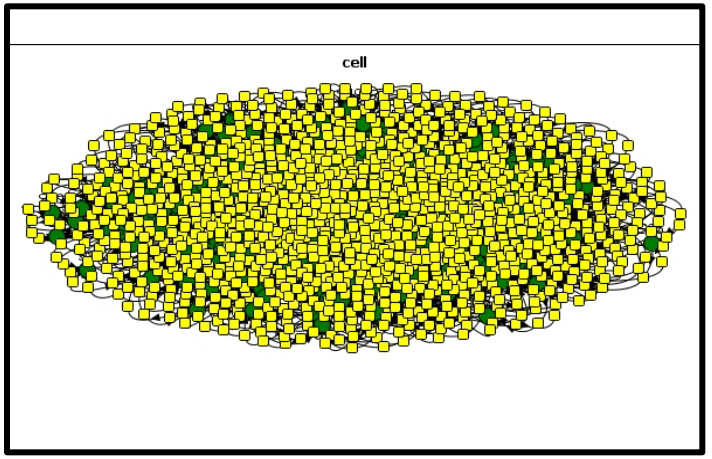
Drawing cartoons is not easy…

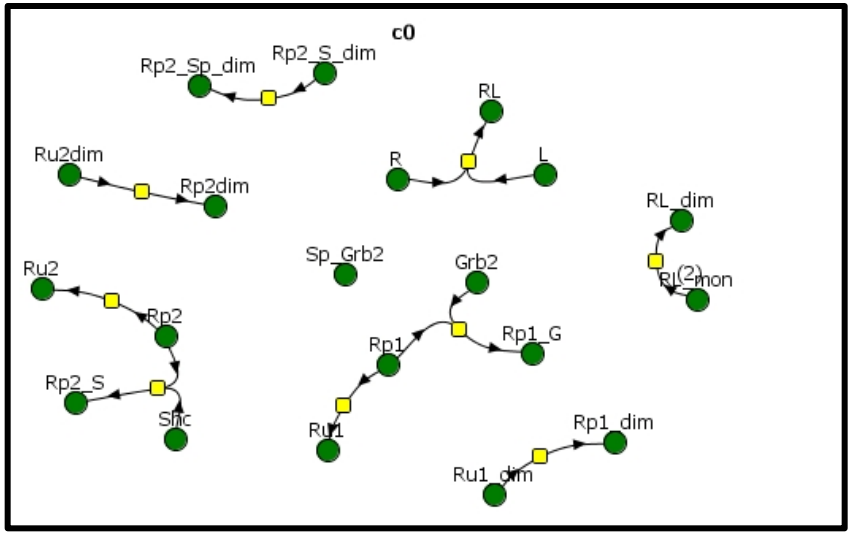


Visualization of rule-based models…

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

Reaction
egfr(l,r)+egf(r) <-> egfr(!1,r).egf(r!1)
egfr(!1,r)+egfr(!2,r) <-> egfr(!1,r!3).egfr(!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)

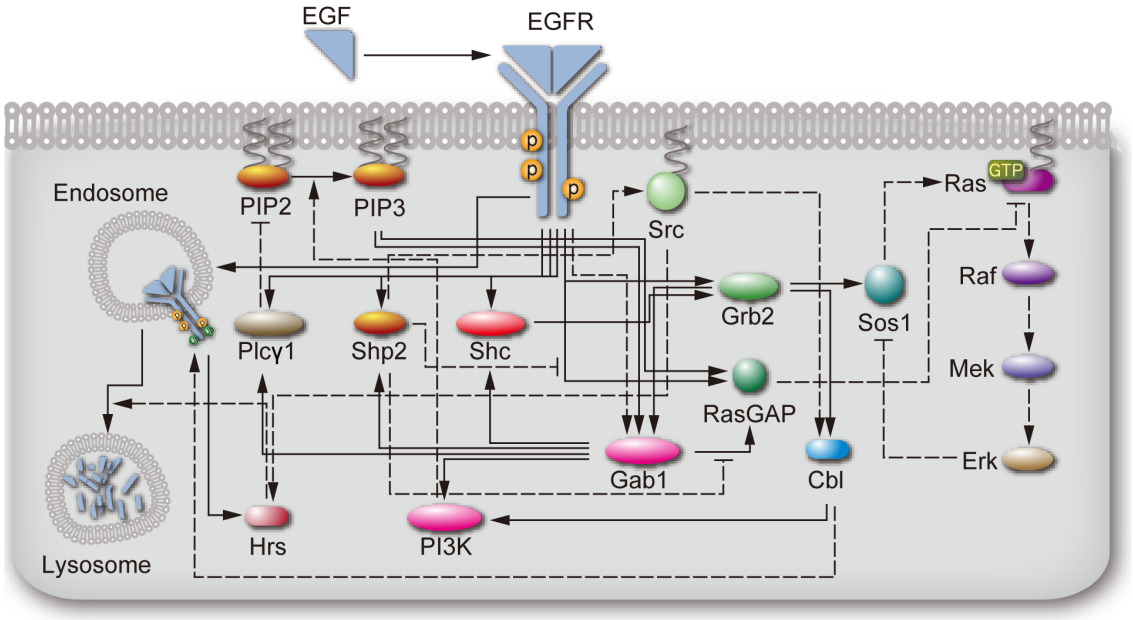


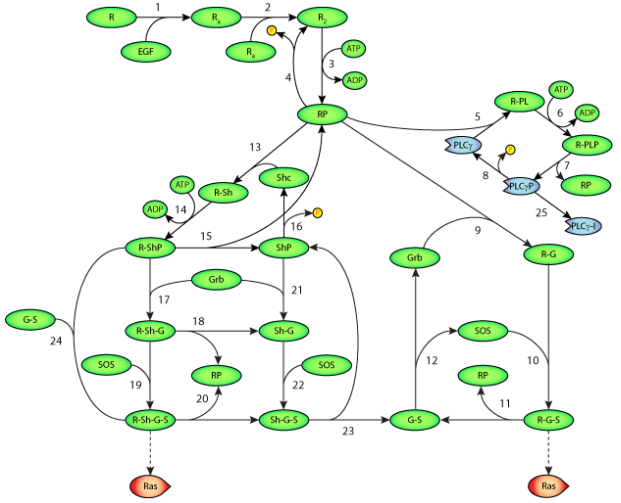


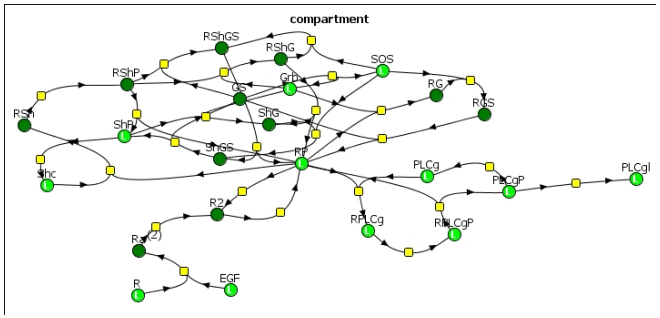
Reading/writing code is not easy…

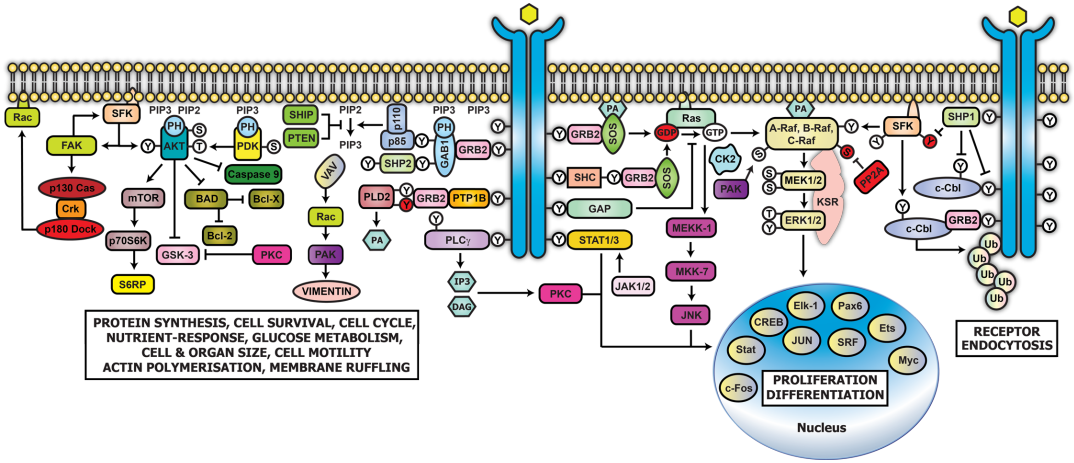
Almost all reactants and products in patterns are distinct, so cartoons can not be assembled into a network…

Variety of model visualizations: EGFR signaling



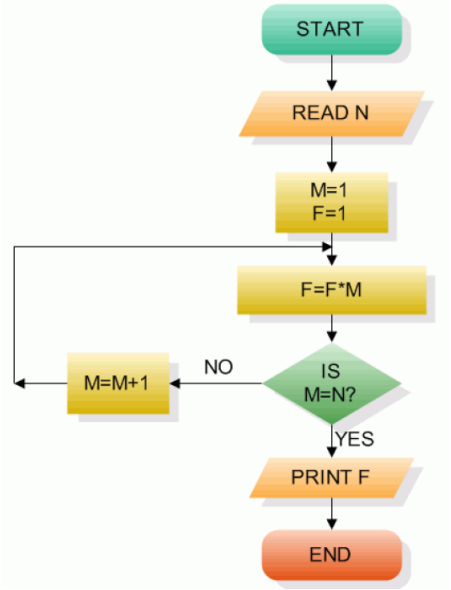




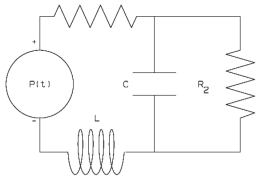


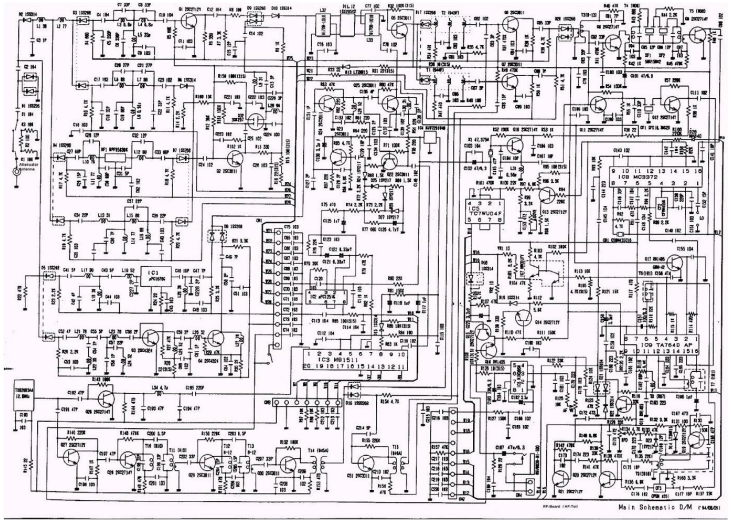
Visualization in Exact Sciences

Standard CS notations



Standard EE notations





Biology Electrical Engineering, Pathways Electronic Circuits

- Circuits and algorithms are designed by humans
  - Follow a defined set of rules
  - Boundaries well defined
- Pathways
  - Don't know all the "rules"
  - Boundaries not clear
  - Knowledge often incomplete

Ideal visual representation

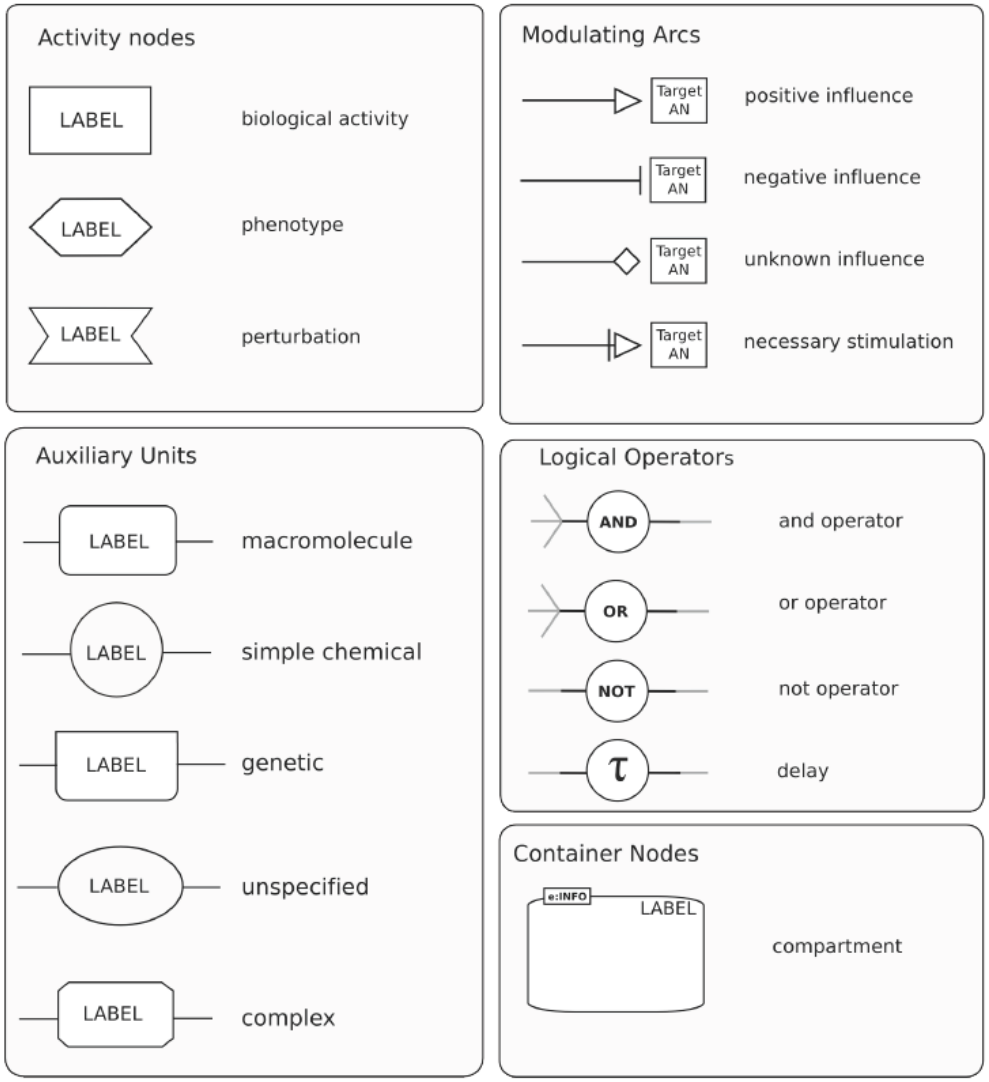
- Visual Language:
  - rich to capture adequately complex biological information
  - represent both mechanistic and qualitative relationships
  - support hierarchical representation, modularization and encapsulation
- Rules and guidelines ensure:
  - reader can unambiguously understand what writer meant
  - machine readable and writable, allow computer manipulation/analysis
- Has an underlying conceptual abstraction
  - a model of its world
  - vocabulary (symbols/**glyphs**)
  - **syntax** (basic assembly or glyphs)
  - **grammar** (rules based on meaning - **semantics**)

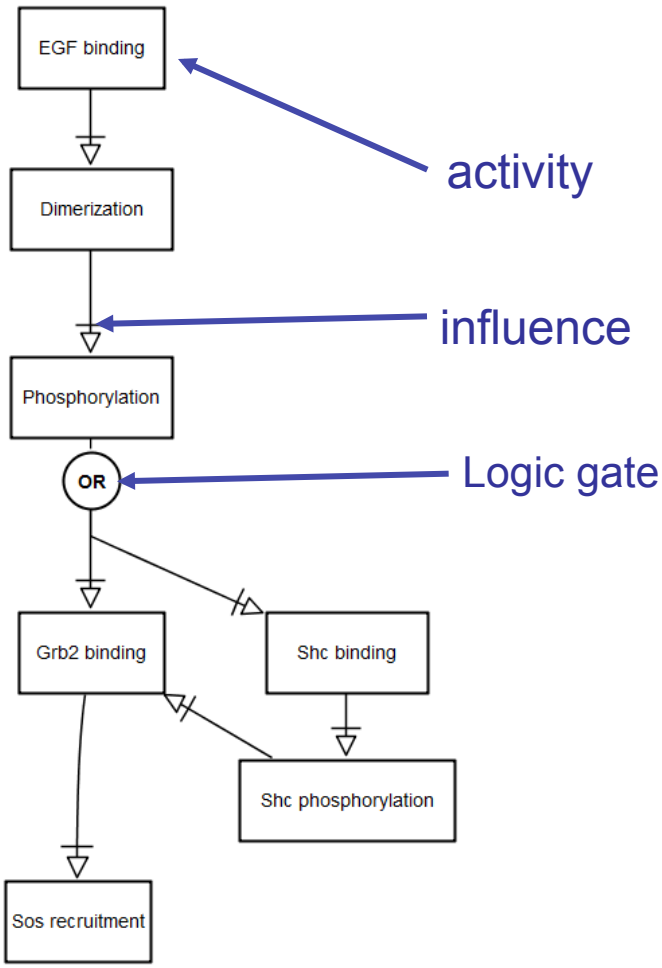
What is SBGN?

- A community effort initiated by Hiroaki Kitano in late 2005.
- Three orthogonal languages to unambiguously describe biochemical and cellular events in graphs:
  - Activity Flow (SBGN AF)
  - Process diagrams (SBN PD)
  - Entity Relationships (SBGN ER)
- Limited amount of symbols (32 in PD, 21 in ER)
- Can graphically represent quantitative models, biochemical pathways, genetic interactions, at different levels of granularity.

Language of SBGN: Activity flow

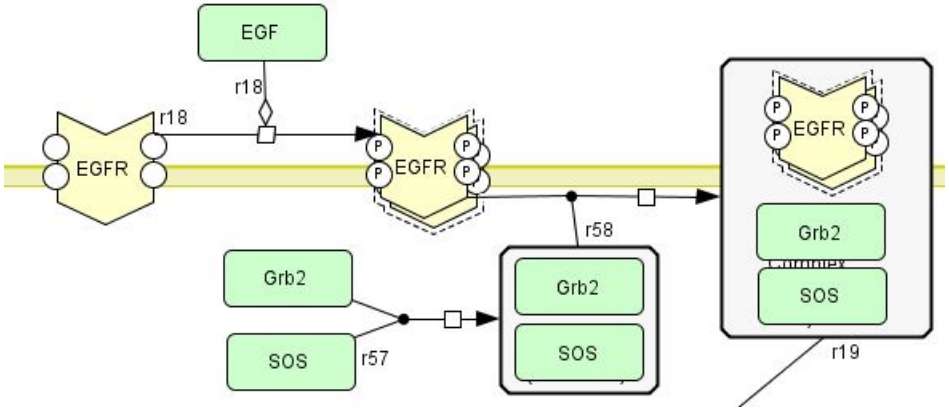
- Flux of information going from one entity to another



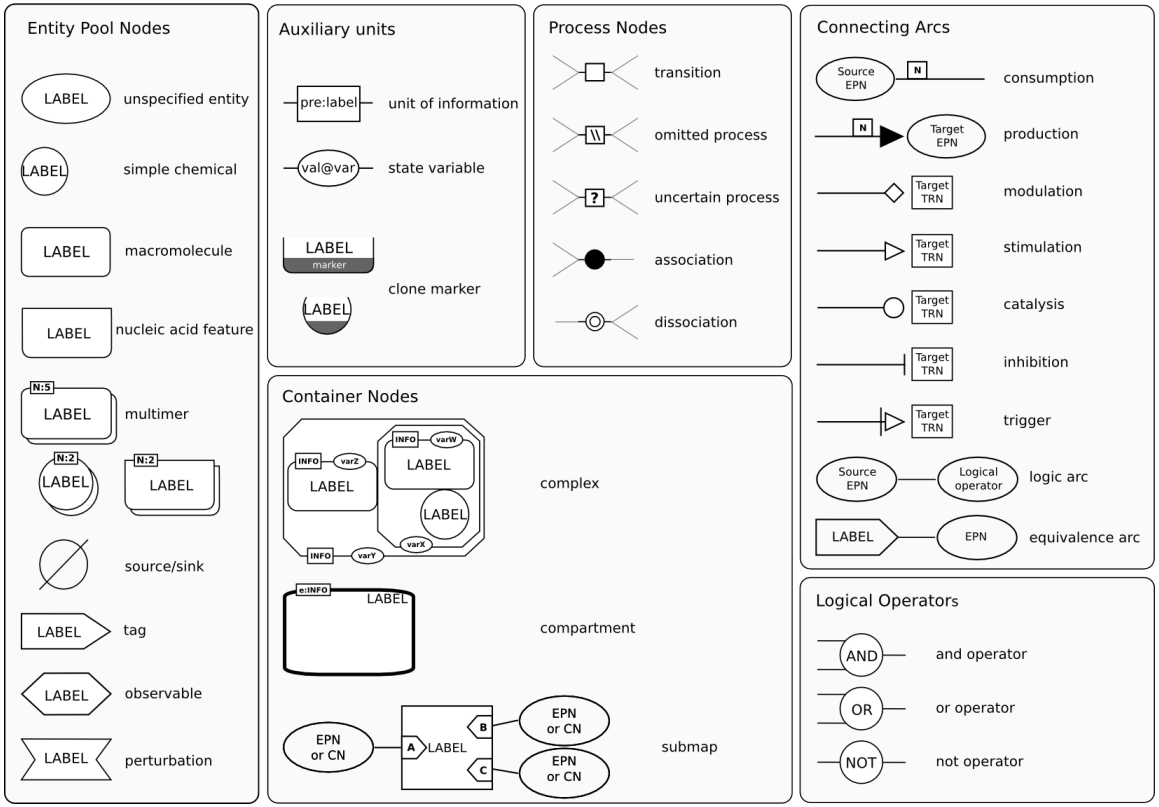


Language of SBML: Process Diagrams

- "Glorified" representation of reaction networks:
  - Multiple types of arrows and nodes
  - complexes
  - User-specified colors, multiple attributes, etc
- Adapted for machine-reading, e.g.
  - no dashed lines
  - Independent on colors and widths
  - Logic
- Used in CellDesigner and PantherDB
- Widely used for poster presentations

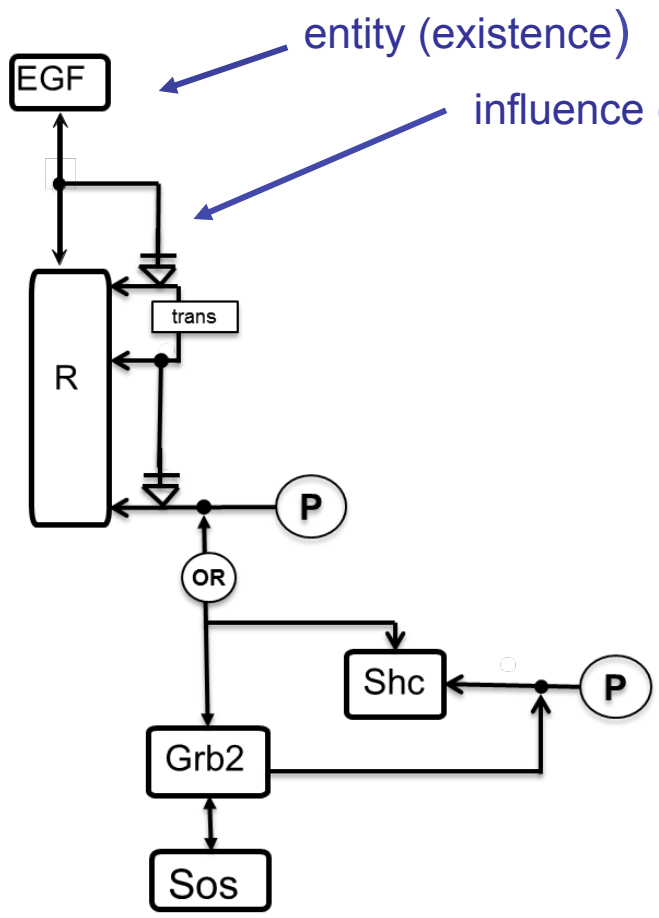


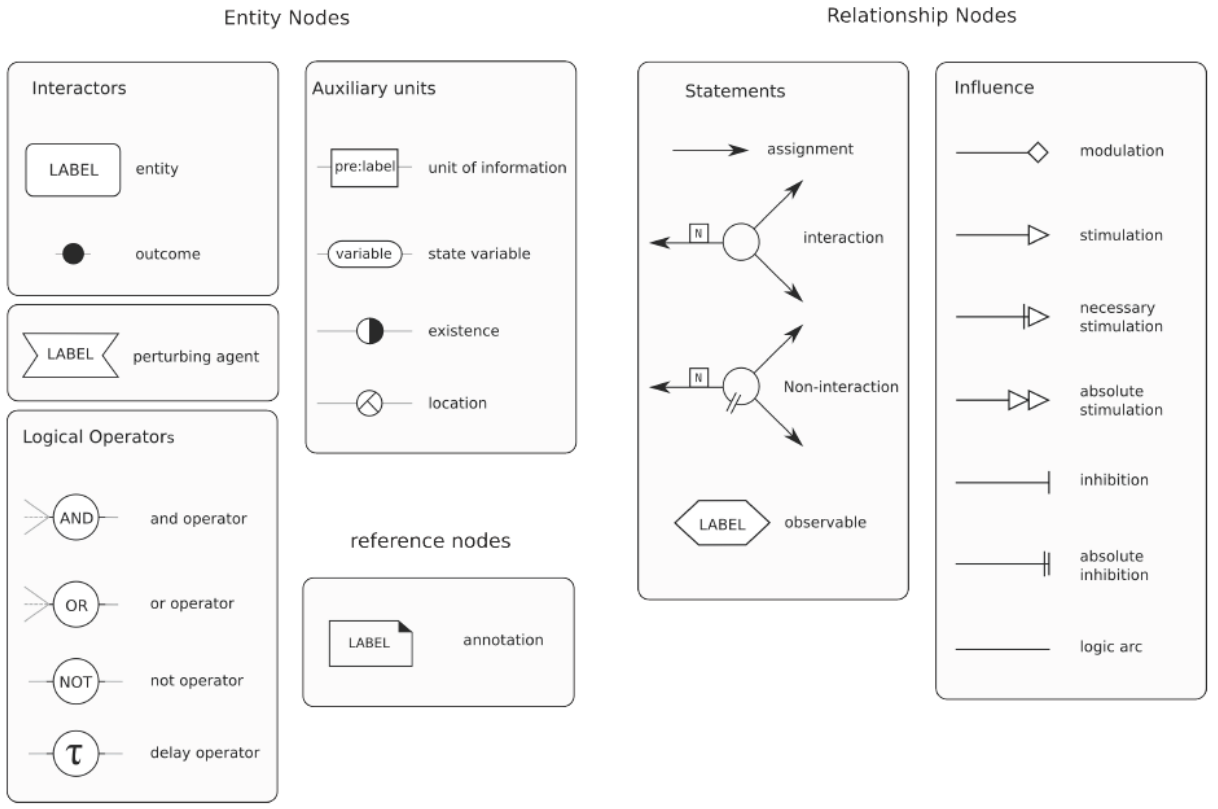
SYSTEMS BIOLOGY GRAPHICAL NOTATION REFERENCE CARD



Language of SBGN: Entity relationships

- Generalization of Molecular Interaction Maps (MIM)
- Describe interactions between entities and the rules that control them

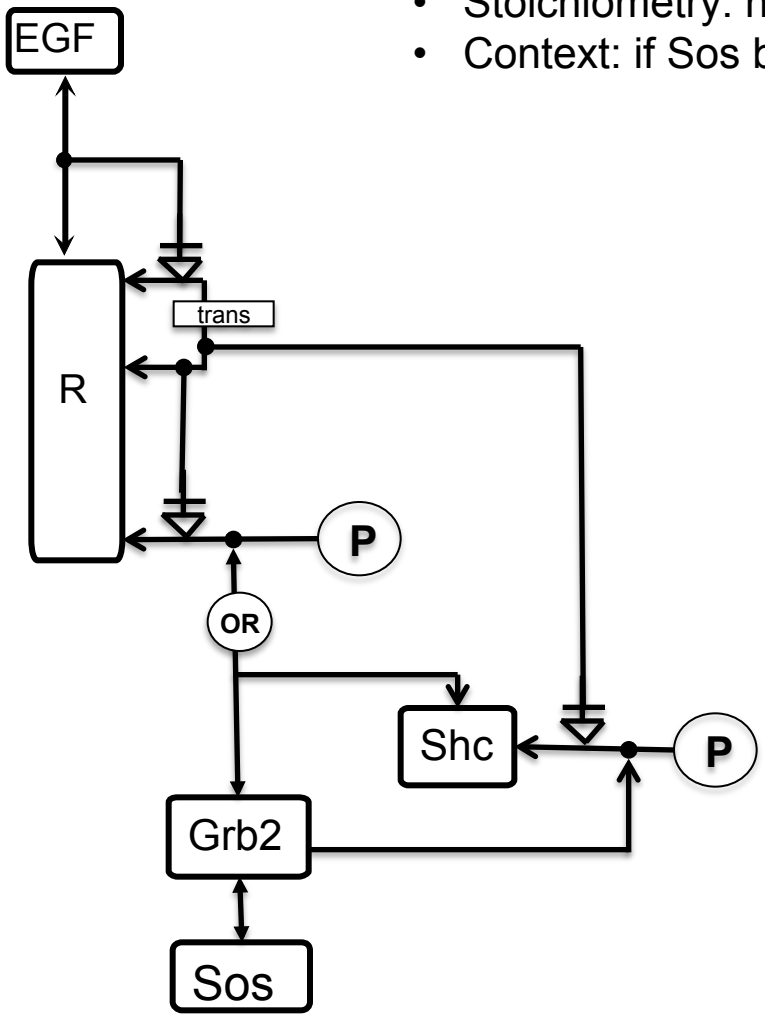




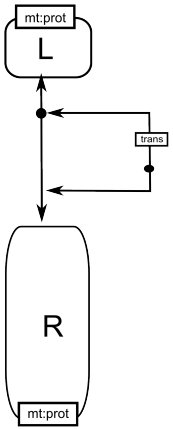
Problems of SBGN-ER or MIM

Not enough information to recover a model:

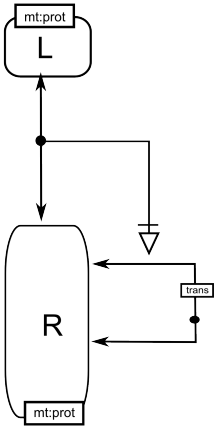
- Reverse interactions are not defined
- Stoichiometry: how many R in dimer are phosphorylated?
- Context: if Sos binds complexes Grb2 in complex with EGFR, Shc?



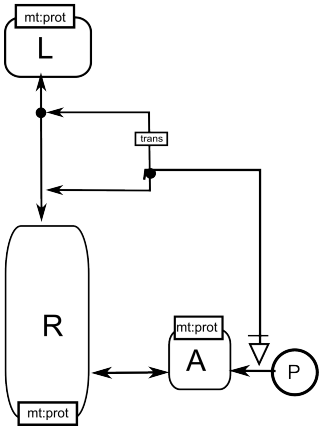
Two ligands are required for dimerization



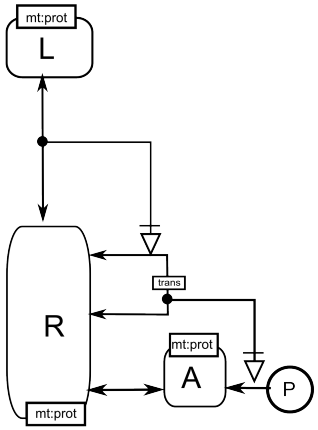
At least one ligand is required for dimerization



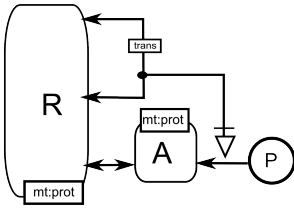
Two ligands are necessary for phosphorylation



One ligands is required for phosphorylation



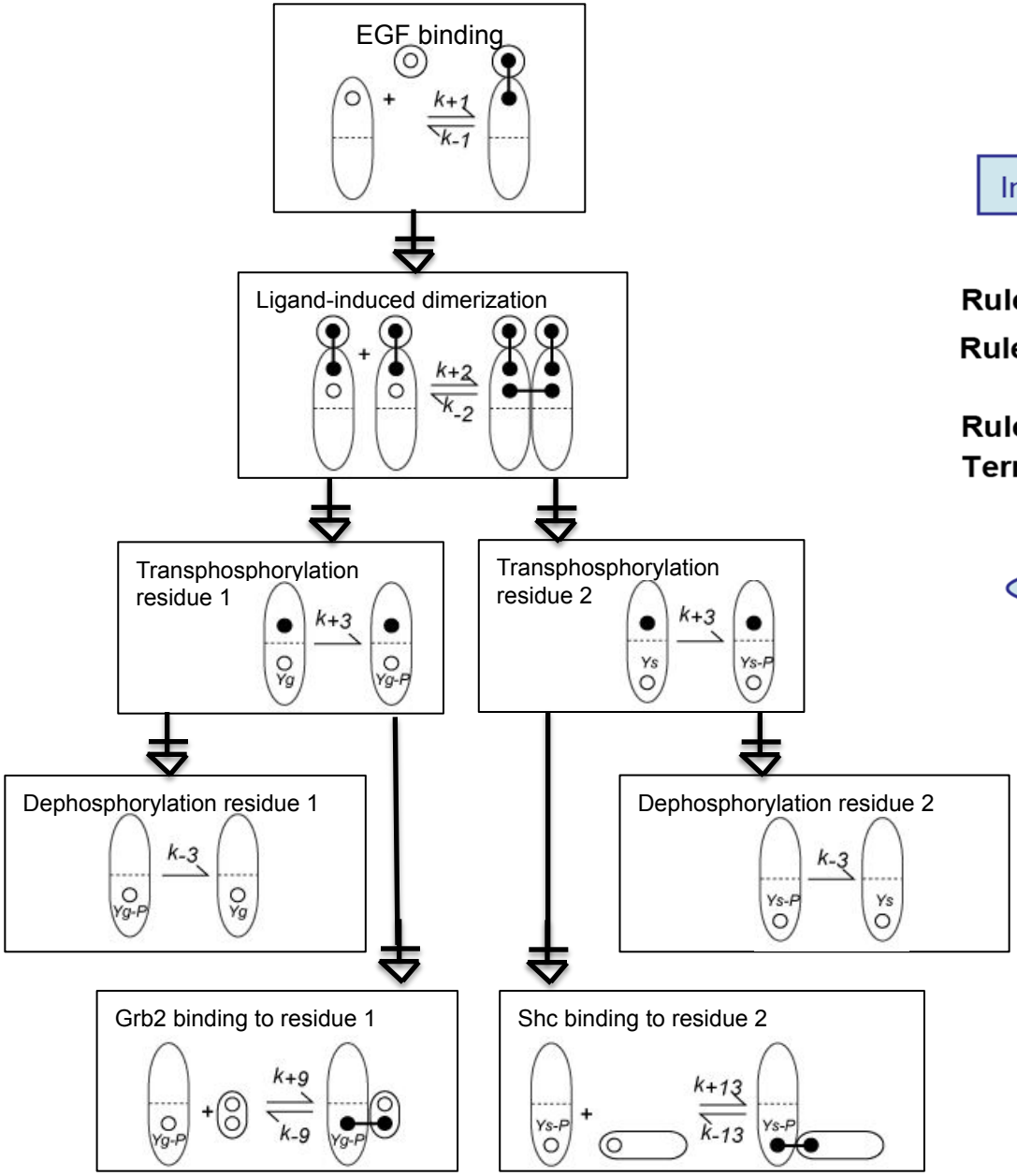
Two receptors are required for phosphorylation



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	Ambiguous
Temporality:	Sequential influences	Sequential	Non-sequential
Pitfalls:	Not suitable to represent association, dissociation, multi-state entities	Sensitive to combinatorial explosion	Creation, destruction, translocation and stoichiometry are not easily represented

Rule-based Activity Flow Diagram



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

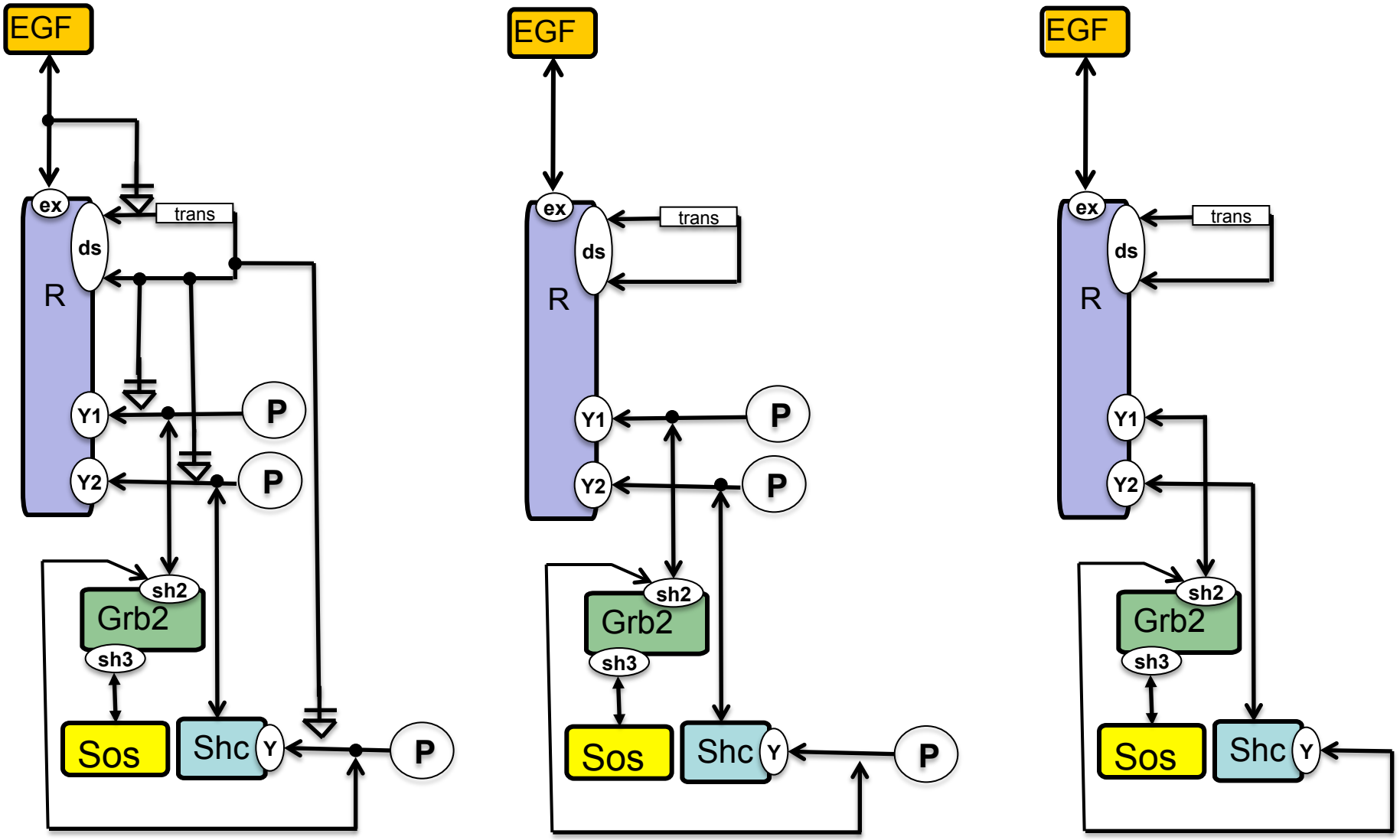
Termination Terminate if $S_n = S_{n+1}$

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

Drawing cartoons is not easy…

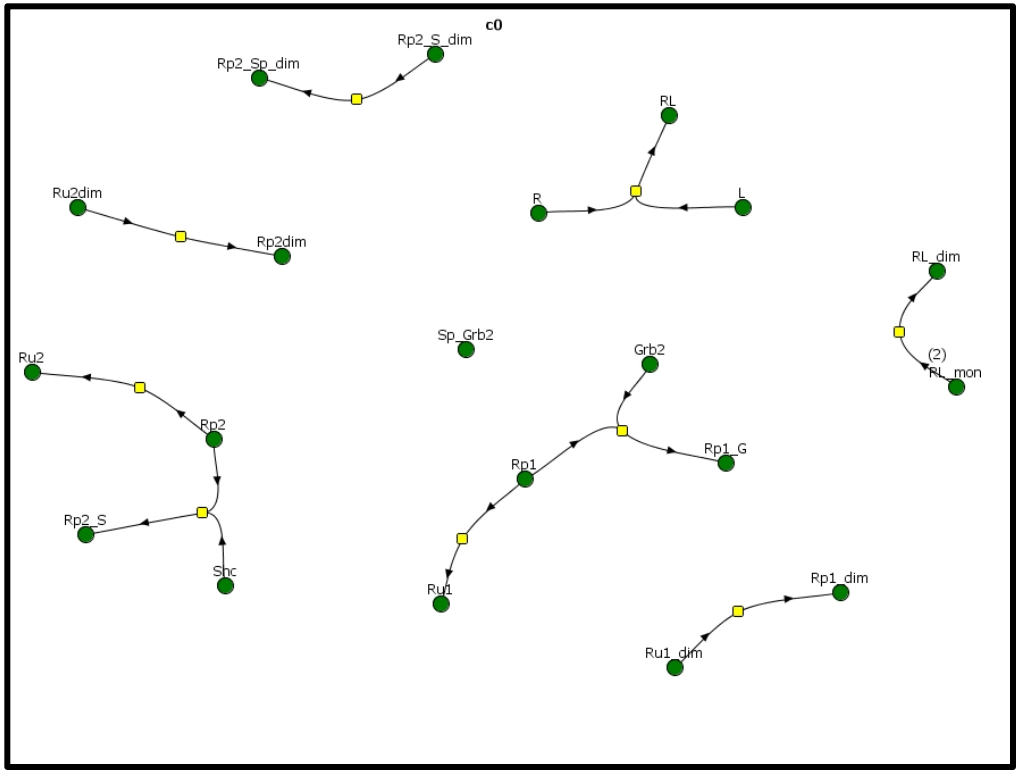
Such visualization can be done only after rules specification, not before

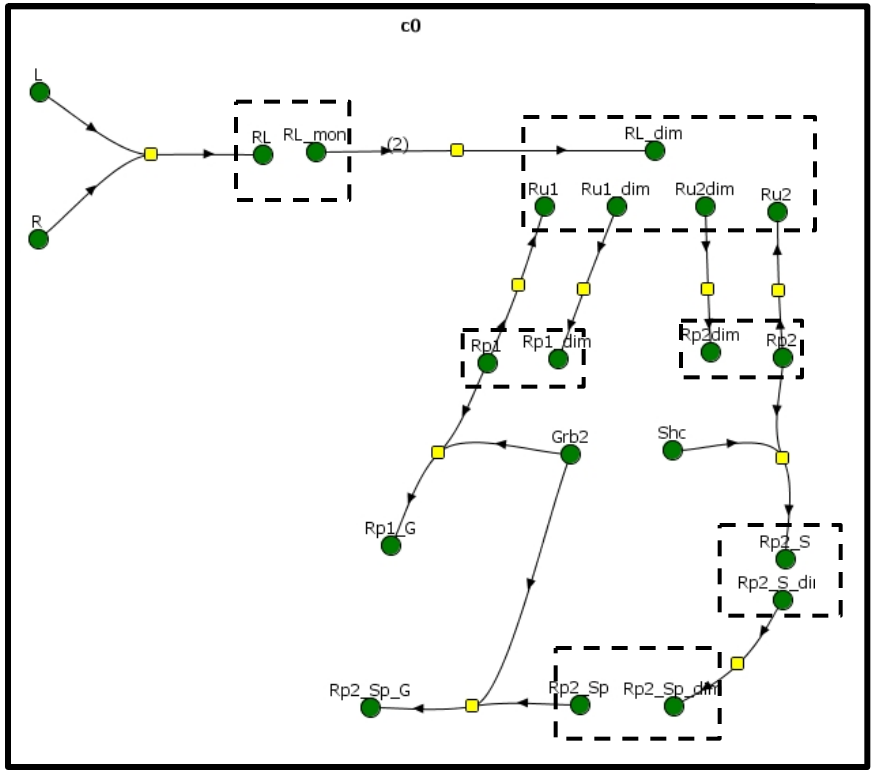
Rule Based Entity Relationships Diagram



Reaction-network visualization of rule-based models…

Approach: Species templates and rules as graph nodes



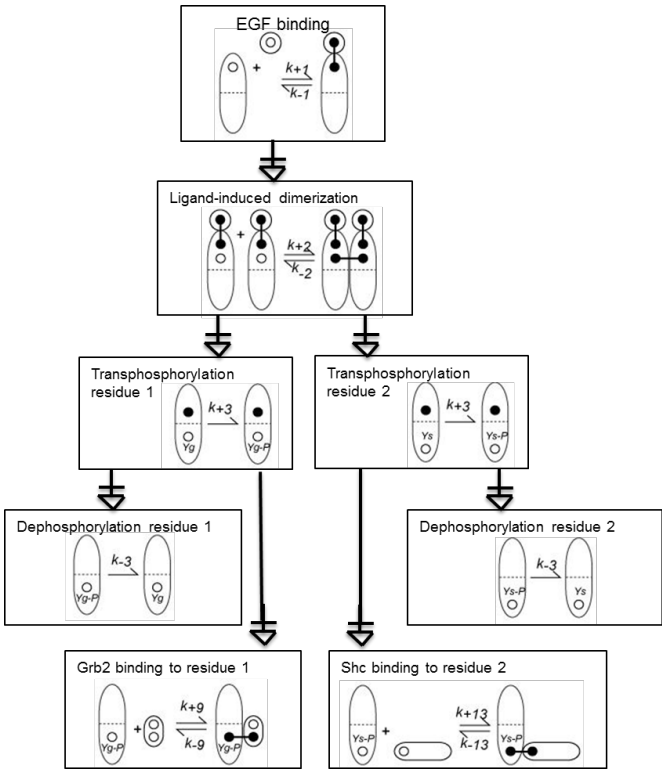


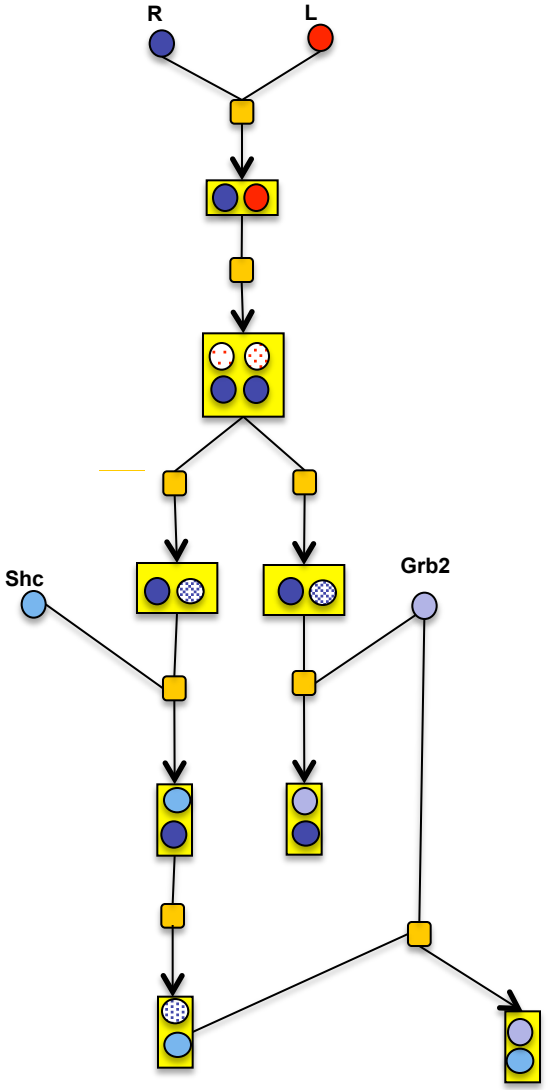
Options:

- user-specified grouping
- Algorithm to find similarities among patterns and automatically group them

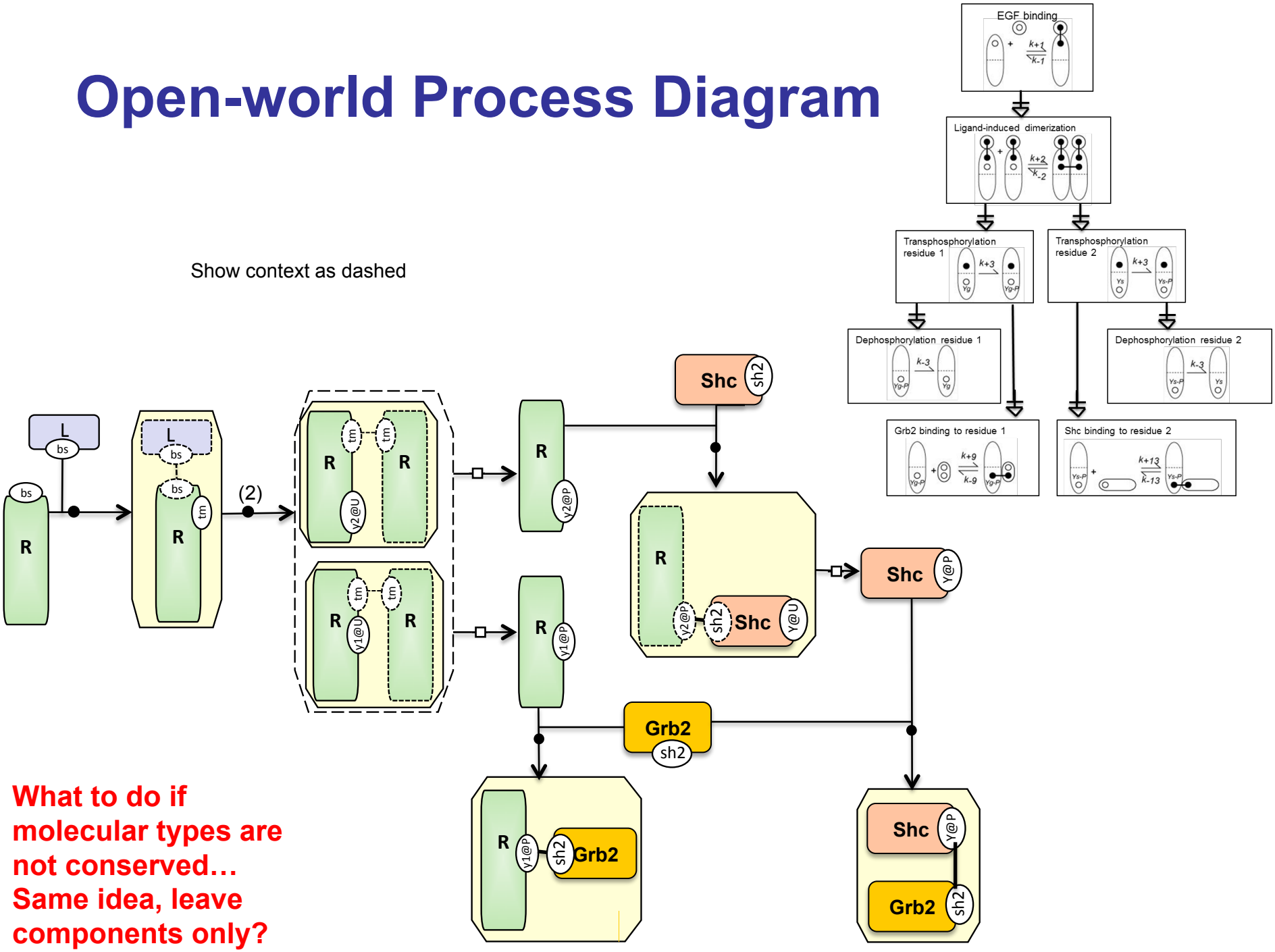
Still postprocessing!

VCell reaction diagram proposal – drop molecular types

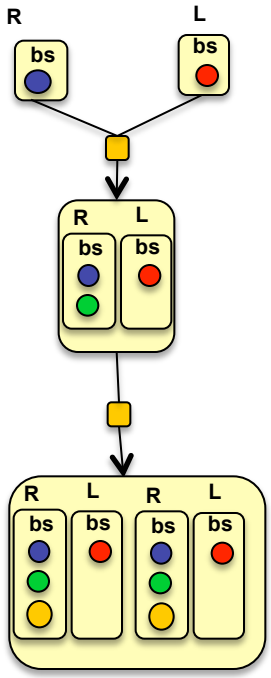


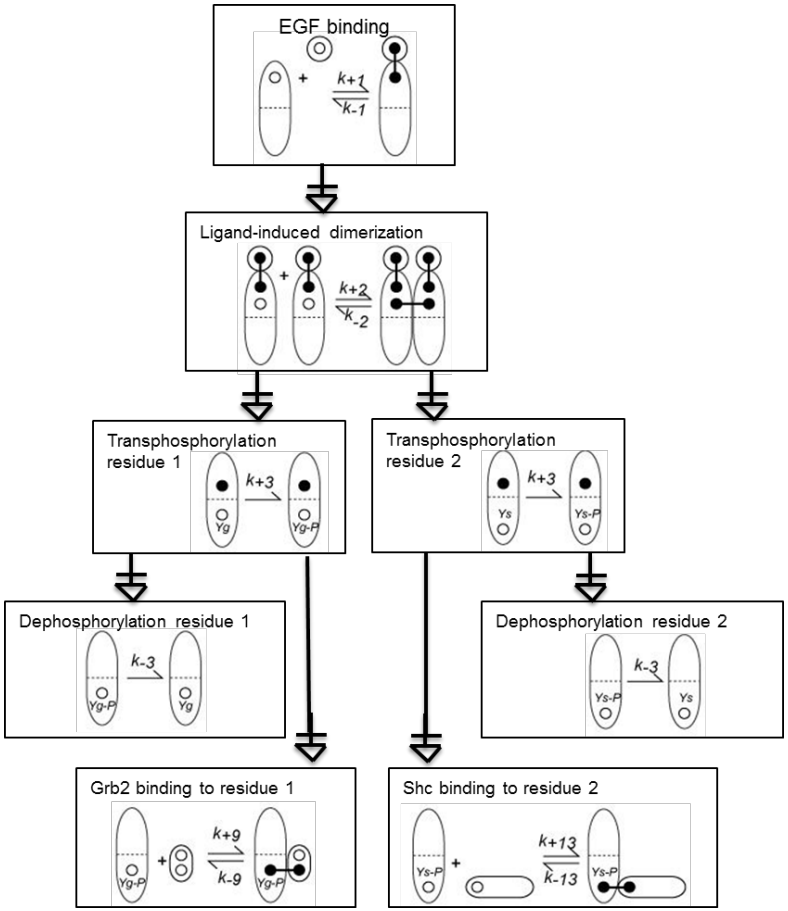


Open-world Process Diagram



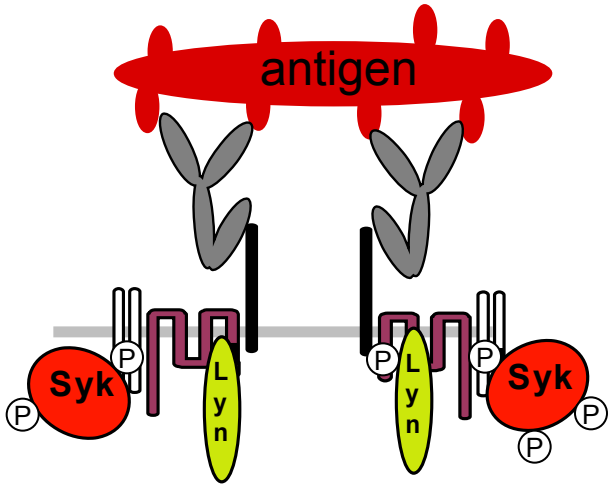
VCell reaction diagram proposal - drop components



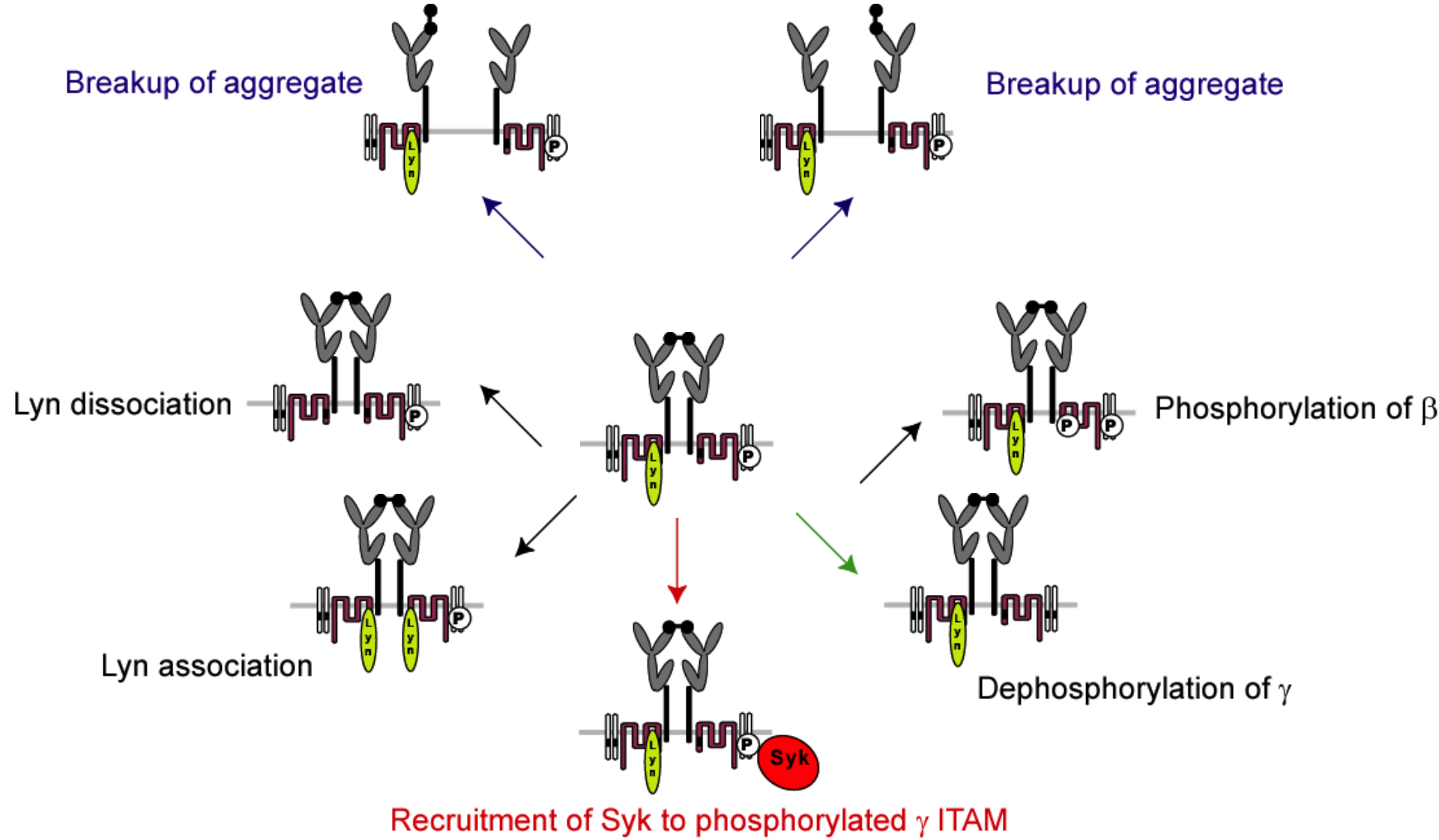


Early Events in FcεRI receptor Signaling

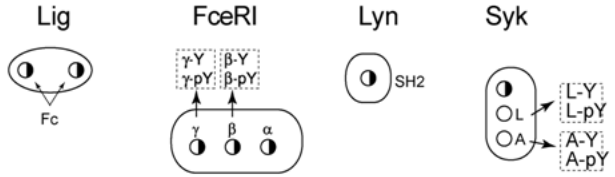
1. Multivalent antigen binds to IgE on cell surface forming aggregates
2. Tyrosine kinase Lyn associates with receptors and **transphosphorylates** ITAM tyrosines
3. Phosphorylated ITAMs recruit Syk and additional Lyn
4. Syk is transphosphorylated by Lyn or Syk
5. Phosphorylation of Syk is critical for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



Not a pathway!

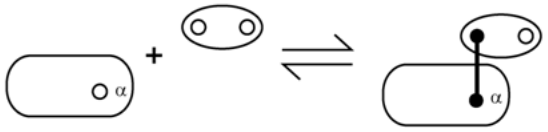


Molecules



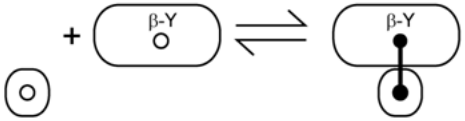
Reaction Rules

1. Ligand binding

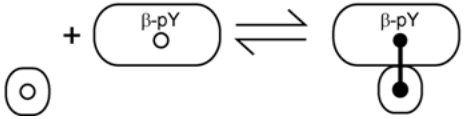


2. Ligand-induced aggregation

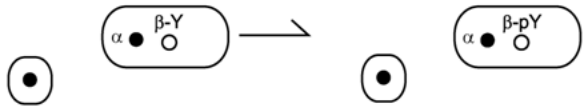
3. Binding of Lyn to unphosphorylated receptor



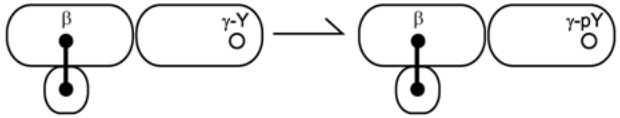
4. Binding of Lyn to phosphorylated receptor



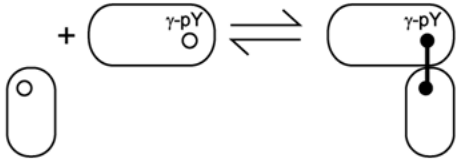
5. Transphosphorylation of β by Lyn



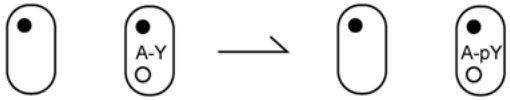
6. Transphosphorylation of γ by Lyn



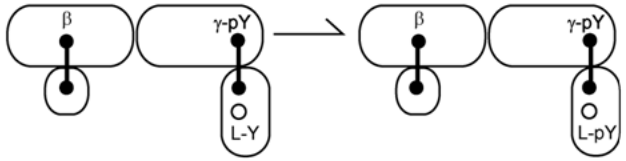
7. Binding of Syk to phosphorylated receptor

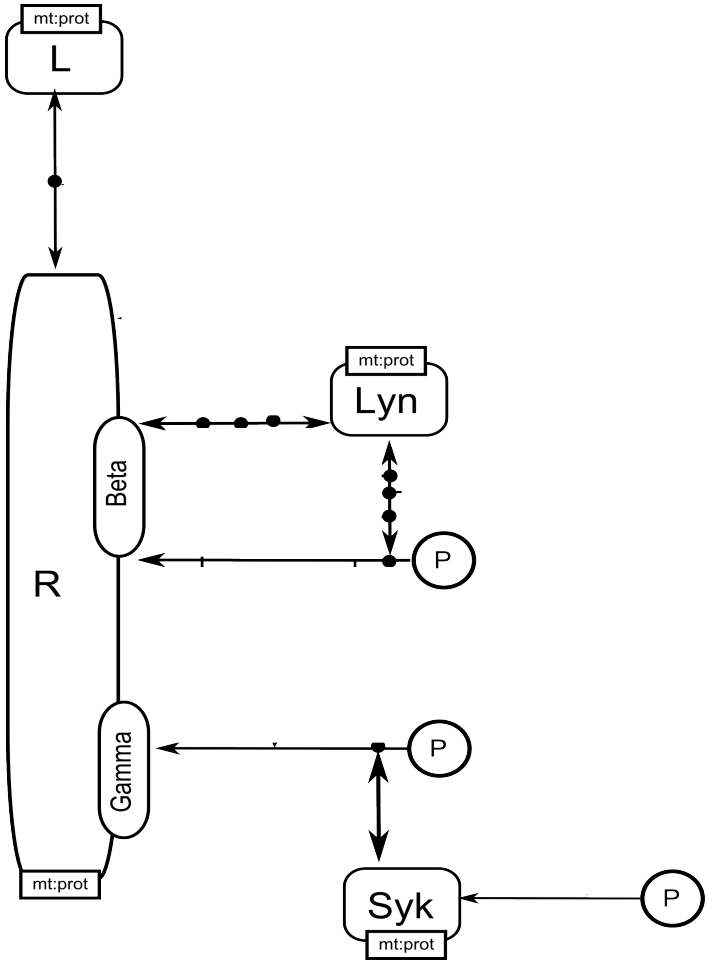


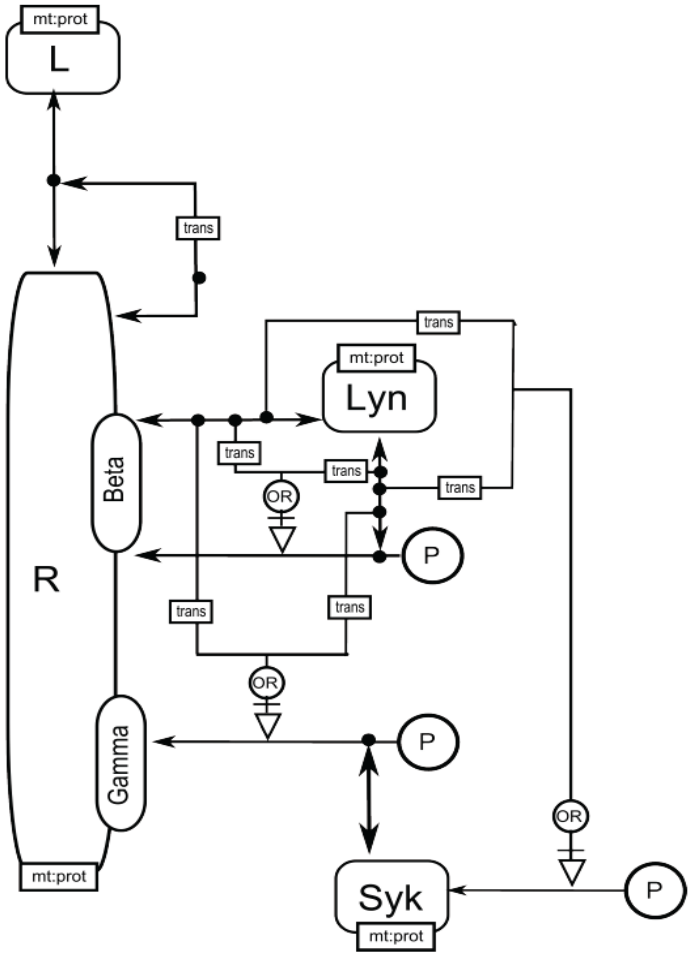
8. Transphosphorylation of Syk by Syk

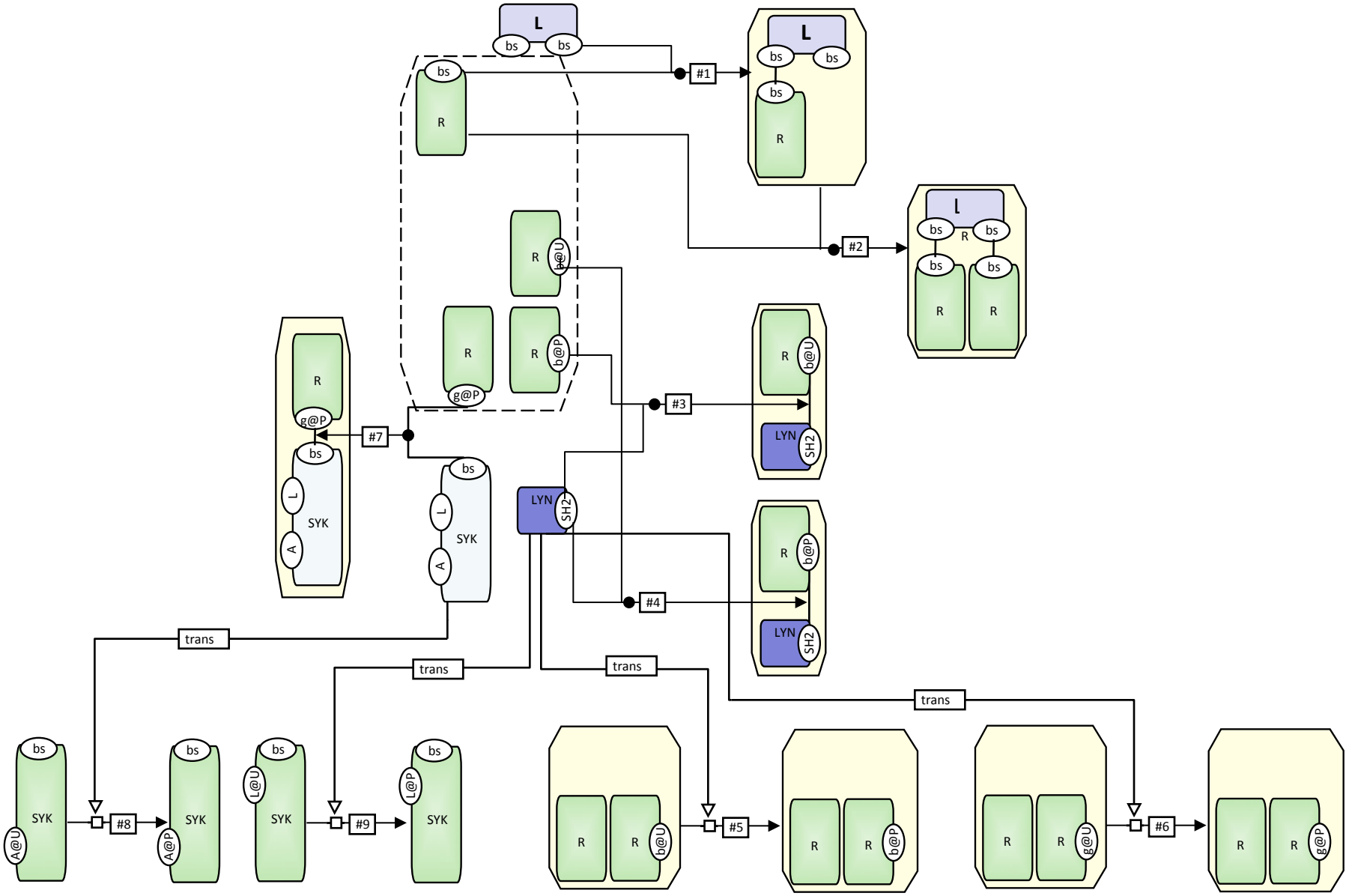


9. Transphosphorylation of Syk by Lyn




mt:prot
L
R
Beta
Gamma
mt:prot
Lyn
P
P
mt:prot
Syk
mt:prot
P


mt:prot
L
trans
R
Beta
Gamma
mt:prot
trans
mt:prot
Lyn
trans
trans
trans
OR
P
trans
trans
OR
P
OR
mt:prot
Syk
mt:prot
P


L
bs
bs
R
#1
#2
b@U
b@P
g@P
#3
LYN
SH2
#4
#7
SYK
L
A
trans
#8
#9
A@U
A@P
L@U
L@P
#5
#6
g@U