CellDesigner

Akira Funahashi
Keio University, Japan
27th Feb. 2011

Keio University
1858
CALAMVS
GLADIO
FORTIOR

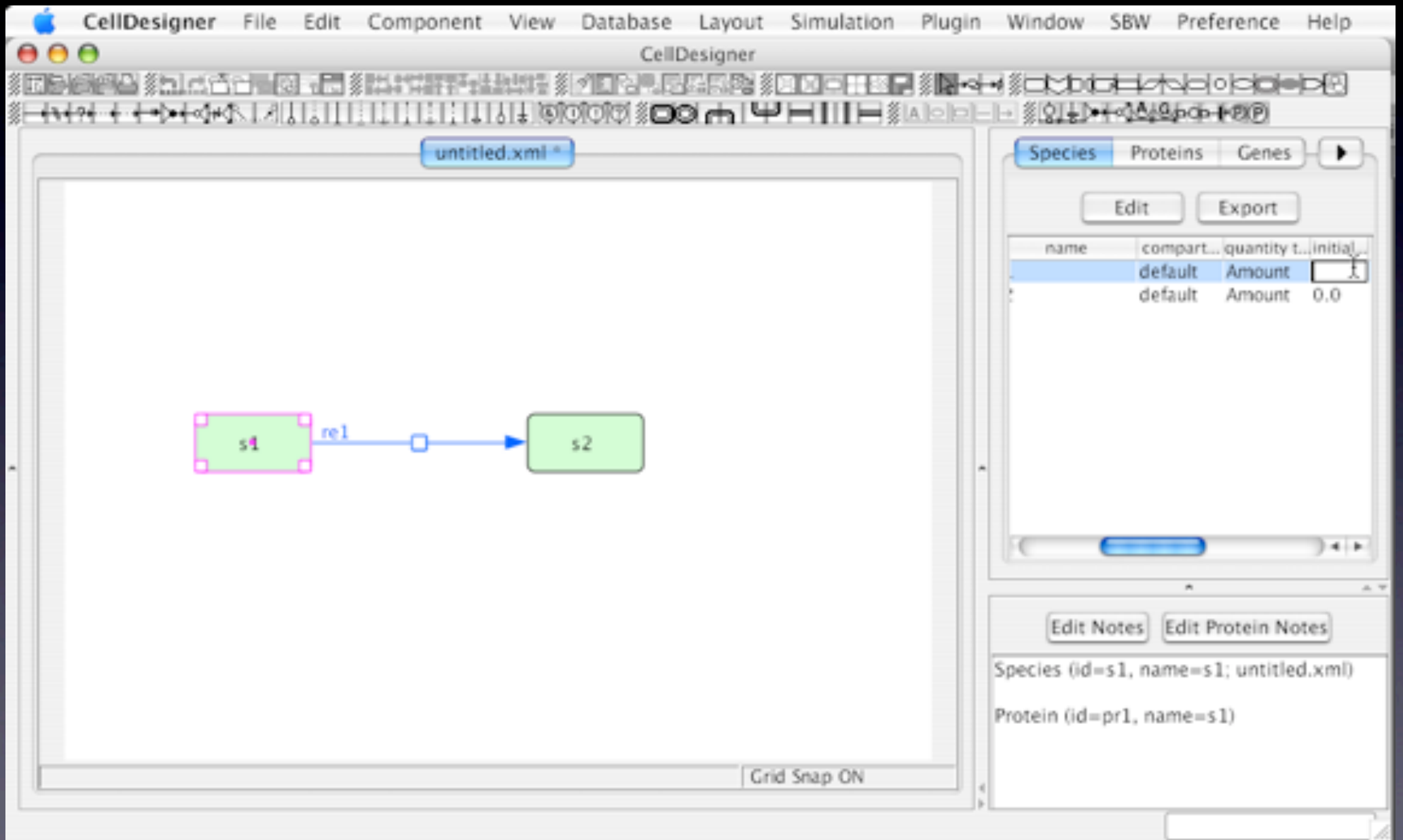
SBI The Systems Biology Institute

Overview

- Introduction of CellDesigner
  - What kind of model you can build
  - SBML (Systems Biology Markup Language)
  - SBGN (Graphical Notation)

- How to build a model with CellDesigner
  - Pathway map
  - Mathematical model

CellDesigner



Nature Molecular Systems Biology 4(173) 2008

Comprehensive pathway map

Molecular Systems Biology 4; Article number 173; doi:10.1038/msb.2008.7
Citation: ***Molecular Systems Biology*** **4:173**

www.molecularsystemsbiology.com

molecular systems biology

REVIEW

A comprehensive modular map of molecular interactions in RB/E2F pathway

Laurence Calzone[1,2,3,6], Amélie Gelay[1,5,6], Andrei Zinovyev[1,2,3,6,*], François Radvanyi[4,5,7] and Emmanuel Barillot[1,2,3,7]

[1] Institut Curie, Service Bioinformatique, Paris,
[2] INSERM, U900, Paris, F-75248 France,
[3] Ecole des Mines de Paris, ParisTech, Fontainebleau, F-77300 France,
[4] Institut Curie, Centre de Recherche, Paris, F-75248 France and
[5] Institut Curie, CNRS UMR144, Paris, France
[6] These authors contributed equally to this work.
[7] These two authors are the joint senior authors of this paper.
* Corresponding author: Institut Curie, Service Bioinformatique, 26 rue d'Ulm, Paris 75248, France. Tel. +33 1 53 10 70 50; Fax: +33 1 44 41 09 08; E-mail: andrei.zinovyev@curie.fr

Received 17.7.07; accepted 23.1.08

We present, here, a detailed and curated map of molecular interactions taking place in the regulation of the cell cycle by the retinoblastoma protein (RB/RB1). Deregulations and/or mutations in this pathway are observed in most human cancers. The map was created using Systems Biology Graphical Notation language with the help of CellDesigner 3.5 software and converted into BioPAX 2.0 pathway description format. In the current state the map contains 78 proteins, 176 genes, 99 protein complexes, 208 distinct chemical species and 165 chemical reactions. Overall, the map recapitulates biological facts from approximately 350 publications annotated in the diagram. The network contains more details about RB/E2F interaction network than existing large-scale pathway databases. Structural analysis of the interaction network revealed a modular organization of the network, which was used to elaborate a more summarized, higher-level representation of RB/E2F network. The simplification of complex networks opens the road for creating realistic computational models of this regulatory pathway.
Molecular Systems Biology 4 March 2008; doi:10.1038/msb.2008.7
Subject Categories: metabolic and regulatory networks; cell cycle
Keywords: cell-cycle regulation; E2F; RB pathway; RB1; systems-biology standards

Introduction

The cell cycle is the succession of four phases called G1, S, G2 and M. In dividing cells, DNA replication (S phase) and mitosis (M phase) alternate (Alberts *et al*, 1994), and are separated by two gap phases, G1 and G2 phases. In quiescent cells, the cells are considered to be in G0 phase. When they receive external signals, such as growth factors, a series of activations push the cell from a G0 to a G1 state and enters the cell cycle. The whole process of cell division is mainly orchestrated by complexes composed of two subunits, a kinase and a cyclin partner. These complexes phosphorylate a certain number of proteins, either activating or inhibiting them. Among them, the retinoblastoma tumour suppressor protein RB (RB1) is a key regulator in cell-cycle entry (transition G1/S). It sequesters a family of transcription factors, the E2Fs, responsible for the transcription of many genes involved in cell-cycle regulation, DNA replication and other functions like the activation of the apoptotic pathway (Muller *et al*, 2001). RB functions as a brake in the cell cycle, which is released when external signals trigger S-phase entry. The main targets of the external signals are the G1 cyclin/CDK complexes. Once active, the complexes, among them CycD1/CDK4,6, act as starters of the cell cycle (Novak *et al*, 2007) and phosphorylate RB, which then releases E2F (DeGregori, 2004).

RB is a member of a family of proteins called the pocket proteins (Knudsen and Wang, 1997). These proteins RB, p107 and p130, share sequence similarities, especially in the 'pocket domain' (Stevaux and Dyson, 2002), which is responsible for their repressor function. RB protein contains domains where the binding sites for co-repressors (E2F proteins and viral oncoproteins) are situated. These sites are subjected to most mutations.

RB is a tumour suppressor gene. Because of its implication in so many, if not all, cancers (Sherr and McCormick, 2002), the study of RB regulation requires a special attention.

More specifically, the RB/E2F pathway is commonly deregulated in cancer through genetic or epigenetic mechanisms, resulting in E2F activation. Several common oncogenes (involved in many cancer types) are the activators of the pathway, whereas several common tumour suppressor genes are inhibitors of the pathway. For example, cyclin D1 (CCND1), E2F3 and the two cyclin-dependent kinases CDK4 and CDK6 can be activated by translocation, amplification or mutation, whereas RB (RB1) and the cyclin-dependent kinase inhibitors p16INK4a (CNKN2A) and p15INK4b (CDKN2B) can be inactivated by point mutation, homozygous deletion or DNA methylation. In addition, RB can be inactivated by several oncogenic viral proteins including E7 from human papillomavirus, which is responsible for more than 90% of cervical carcinomas (Munger *et al*, 2001). Tumour suppressor gene inactivation is found not only in sporadic tumours but also in tumour-prone families. Germline mutations of RB1 results in retinoblastoma with a high penetrance early in young individuals and late in life in sarcomas and lung and bladder carcinomas (Knudson, 1971; Nevins, 2001; Giacinti and Giordano, 2006). Germline mutations of p16INK4a results in

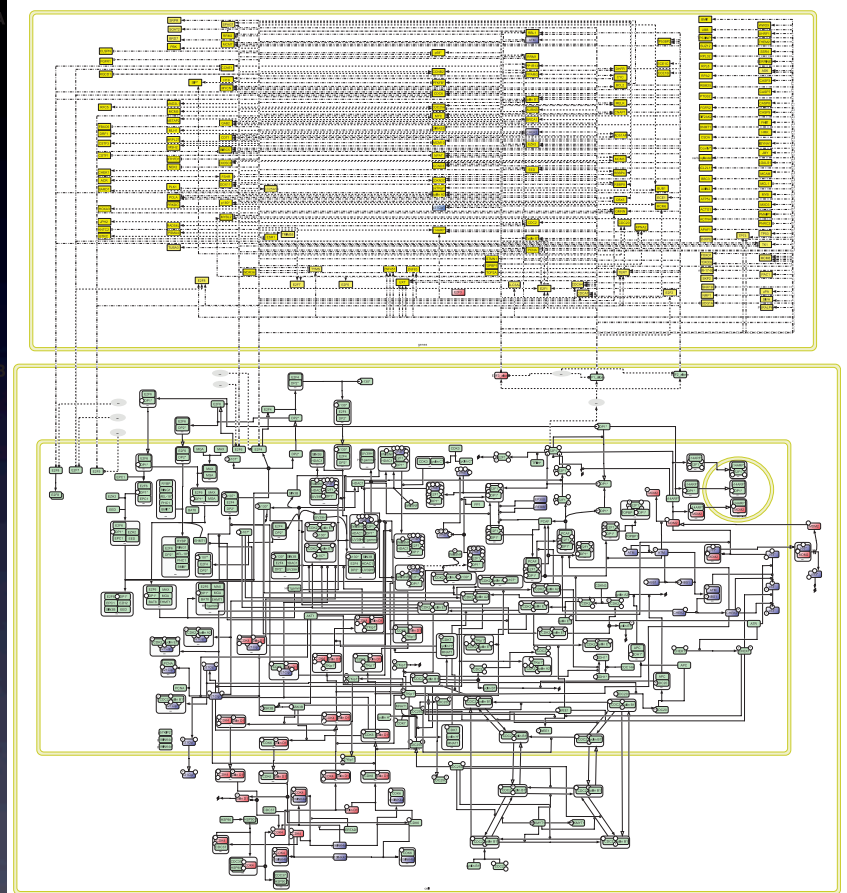


Figure 2 The textbook pathway of RB has been expanded by integrating data from the literature. The E2F transcription factors (represented here by single proteins in the nuclear compartment) are connected by activation and inhibition arrows to their gene targets. (**A**) Map of target genes of E2F transcription factors. Each E2F associates with different cofactors to activate or inhibit the transcription of many genes; pointed arrows mean activation and flat arrows mean inhibitions (**B**) Map of protein–protein interaction network. Each icon on the diagram represents distinct chemical species. See Kitano and co-workers' description of CellDesigner's standard notation (Kitano *et al*, 2005) for a detailed meaning of shapes. When the information is available (from Atlas Oncology web-page: www.atlasgeneticsoncology.org/), tumour suppressor genes and the corresponding proteins are coloured in blue and oncogenes in red, the other proteins are in green. To read and navigate through the map, visit our webpage: http://bioinfo-out.curie.fr/projects/rbpathway/. The map is clickable and allows easy access to all included information (such as literature references or standard protein ids) and hyperlinked to other databases.

are connected by 'activation' and 'inhibition' relations. The information about these relations is derived from the detailed diagram. For example, in the detailed map, E2F1 is phosphorylated by CycA2/CDK2 and is subsequently recognized for degradation, which is translated in the modular map by CycA2/CDK2 module inhibiting E2F1-3 module.

2011年3月2日水曜日

European journal of biochemistry / FEBS 267, no. 6 (2000): 1583-8.

Mathematical model

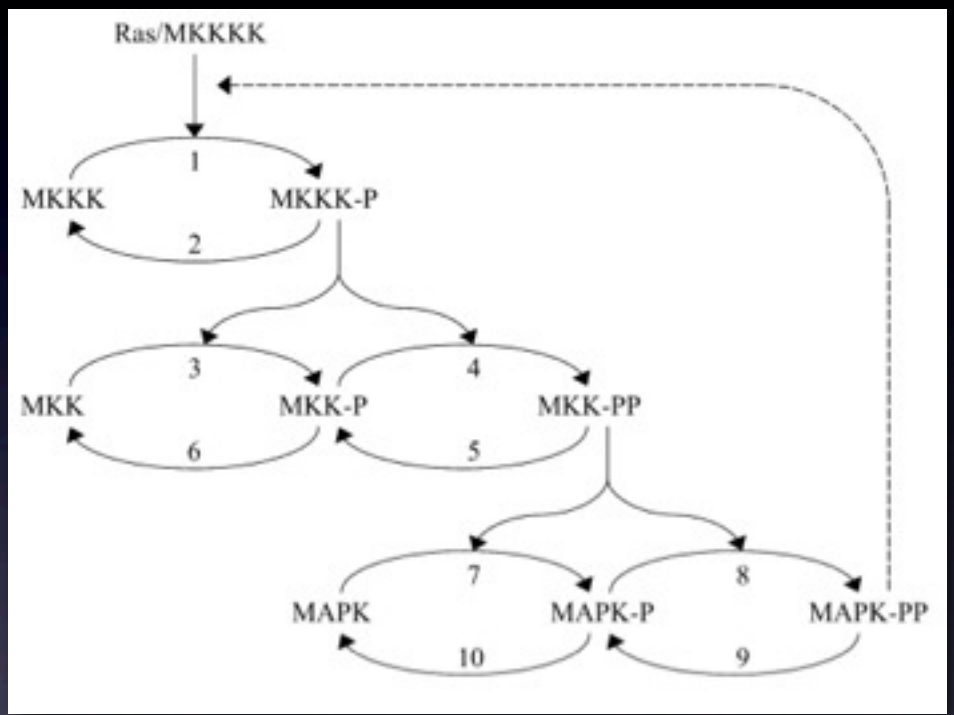
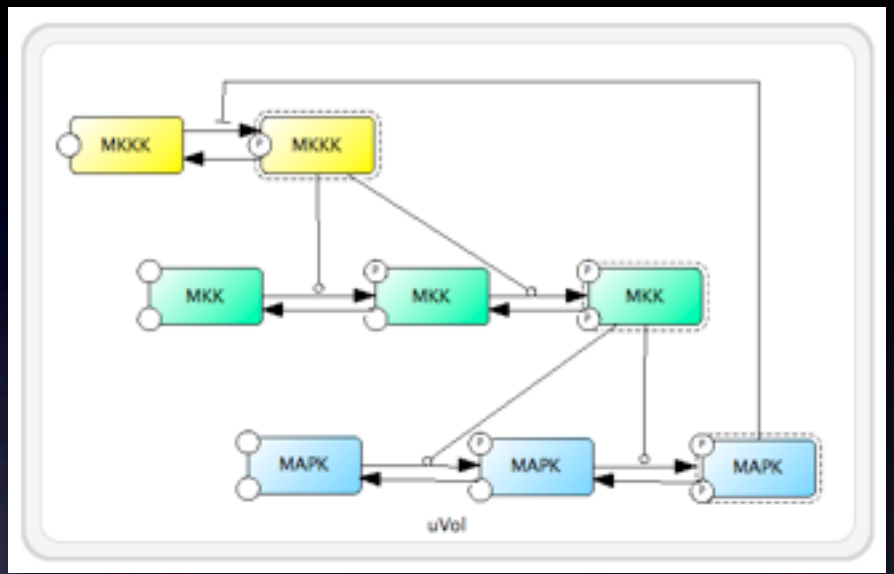
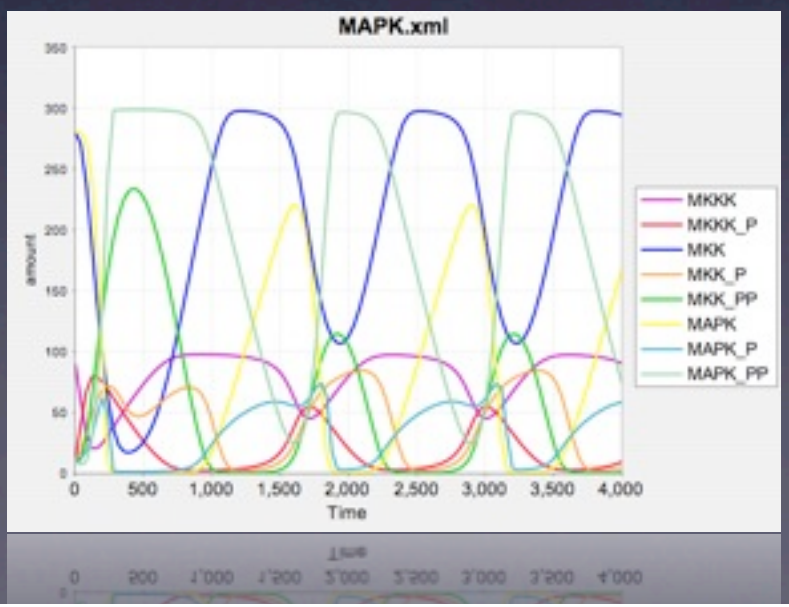
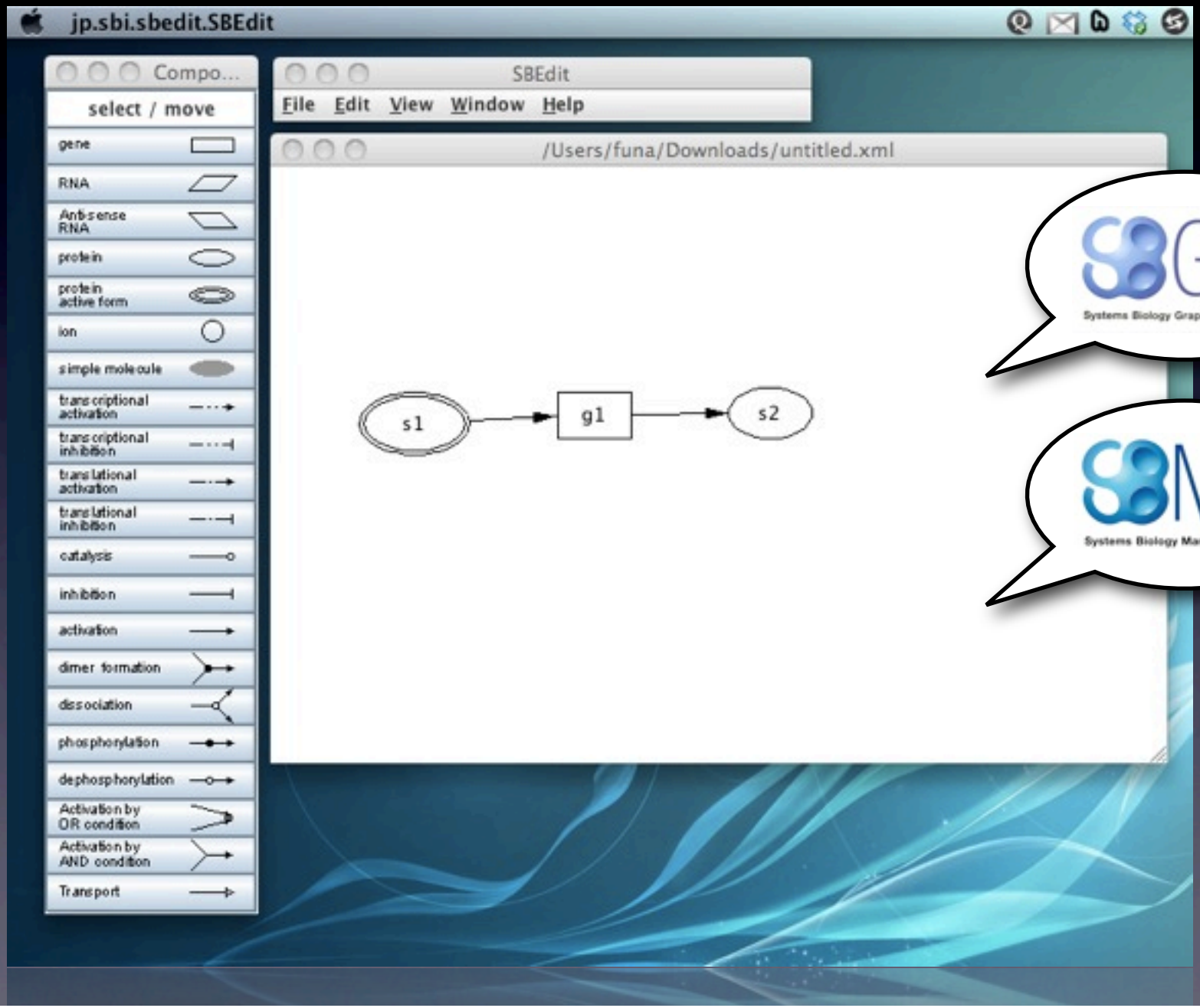


Fig. 1. Kinetic scheme of the MAPK cascade. Feedback effect of MAPK on the rate of MKKK phosphorylation is shown schematically by the dashed line. Numbering of individual steps corresponds to kinetic equations in Tables 1 and 2.



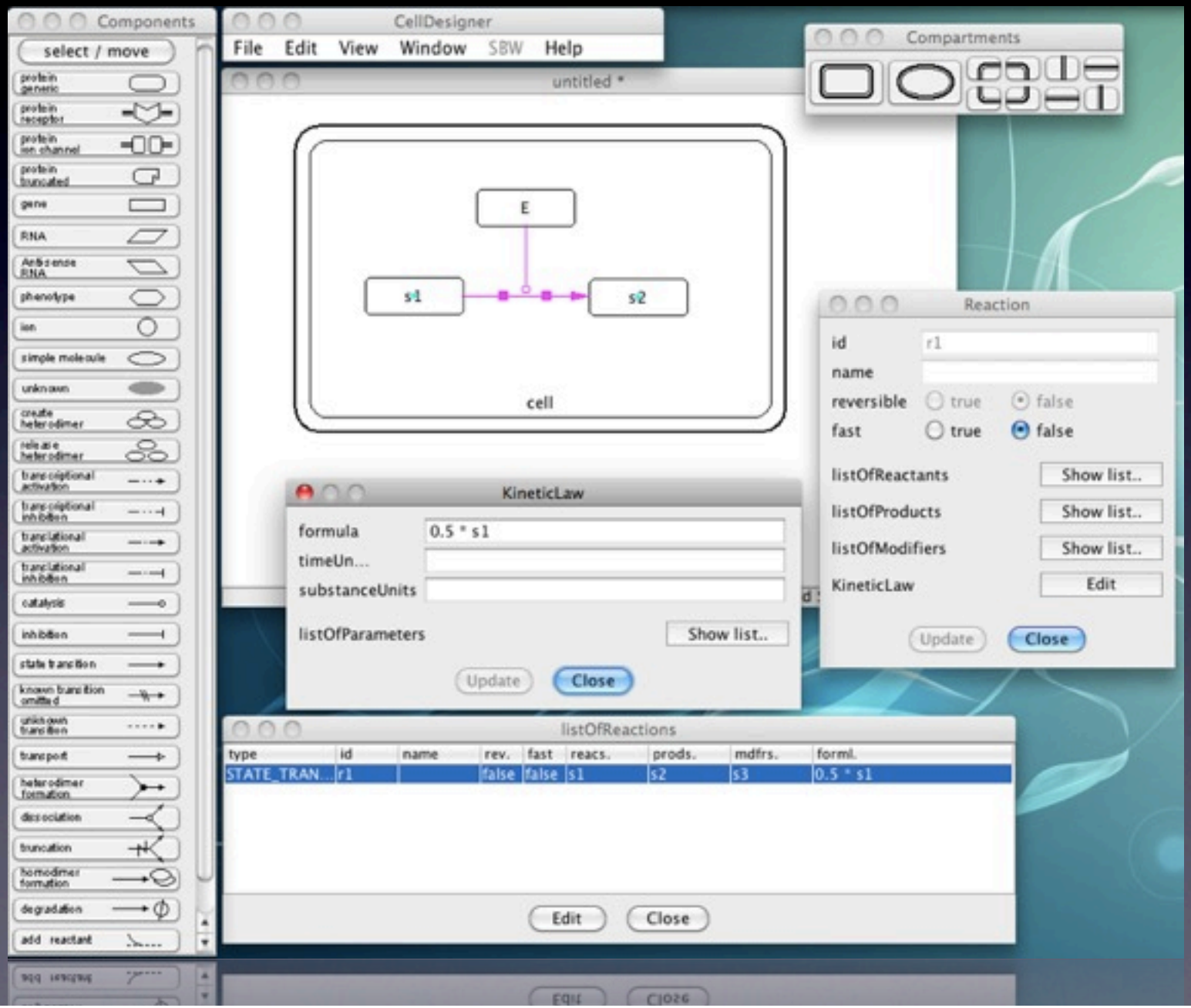


History (2002)



```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level1" version="1" level="1" xmlns:sbedit=
    <annotation>
        <sbedit:modelDisplay sizeX="600" sizeY="400"/>
    </annotation>
    <model name="untitled">
        <listOfSpecies>
            <specie name="s1">
                <annotation>
                    <sbedit:class>PROTEIN_ACTIVE_FORM</sbedit:class>
                </annotation>
            </specie>
            <specie name="s2">
                <annotation>
                    <sbedit:class>PROTEIN</sbedit:class>
                </annotation>
            </specie>
            <specie name="g1">
                <annotation>
                    <sbedit:class>GENE</sbedit:class>
                </annotation>
            </specie>
        </listOfSpecies>
        <annotation>
            <sbedit:listOfAliases>
                <sbedit:alias name="a1" specie="s1">
                    <sbedit:bounds x="59.0" y="151.0" w="74.0" h="42.0"/>
                </sbedit:alias>
                <sbedit:alias name="a2" specie="s2">
                    <sbedit:bounds x="306.0" y="148.0" w="56.0" h="34.0"/>
                </sbedit:alias>
                <sbedit:alias name="a3" specie="g1">
                    <sbedit:bounds x="192.0" y="151.0" w="49.0" h="31.0"/>
                </sbedit:alias>
            </sbedit:listOfAliases>
        </annotation>
        <listOfReactions>
                <reaction name="">
                    <annotation>
                        <sbedit:reactionType>ACTIVATION</sbedit:reactionType>
                    </annotation>
                    <listOfReactants>
```

History (2003)

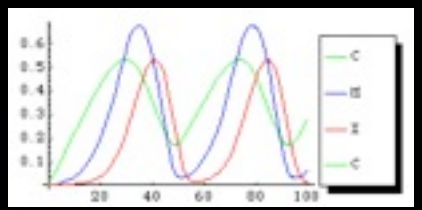


Motivation

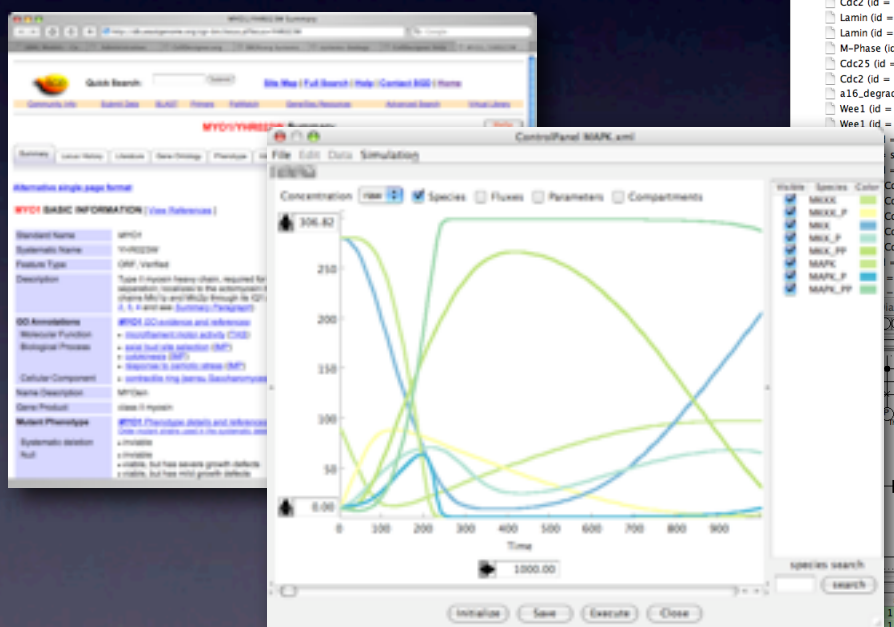
- Provide a software tool which uses:
  - SBML as a native file format
  - solidly defined graphical notation to represent biochemical networks
- Provide a software tool which can
  - integrate with existing resources / software tools
  - support many researchers as possible

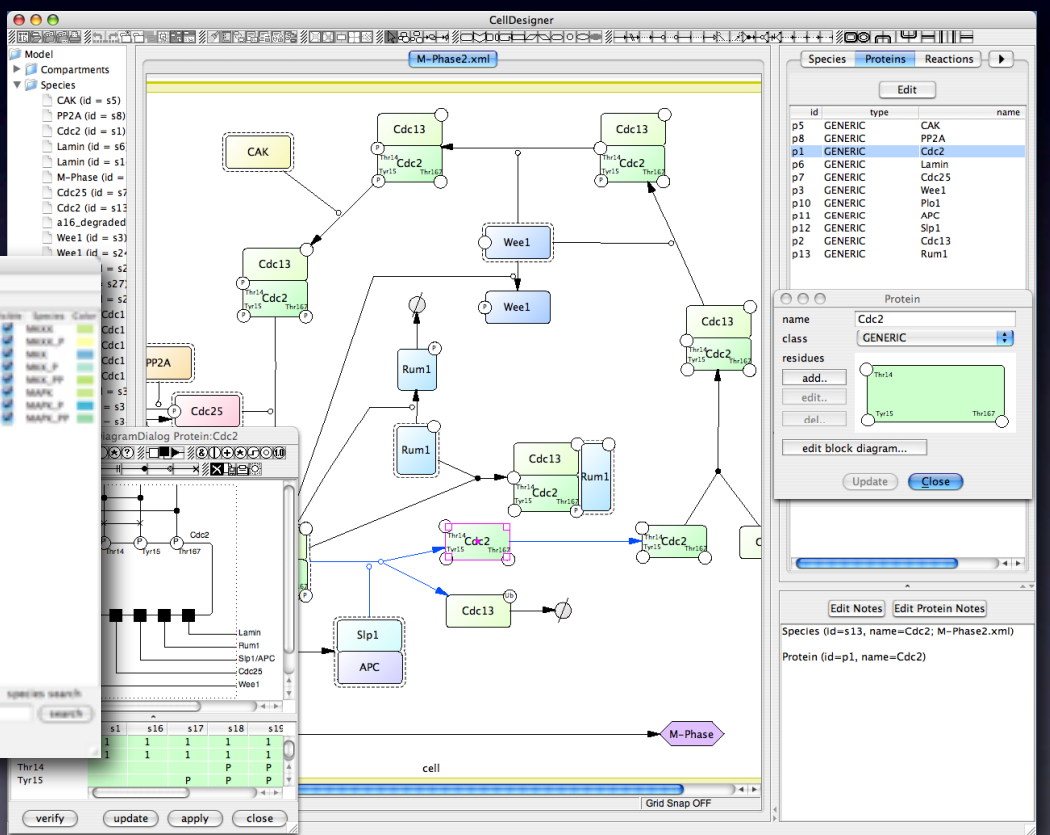
1 min. to run a simulation

CellDesigner

SBML + SBGN +  +

= CellDesigner

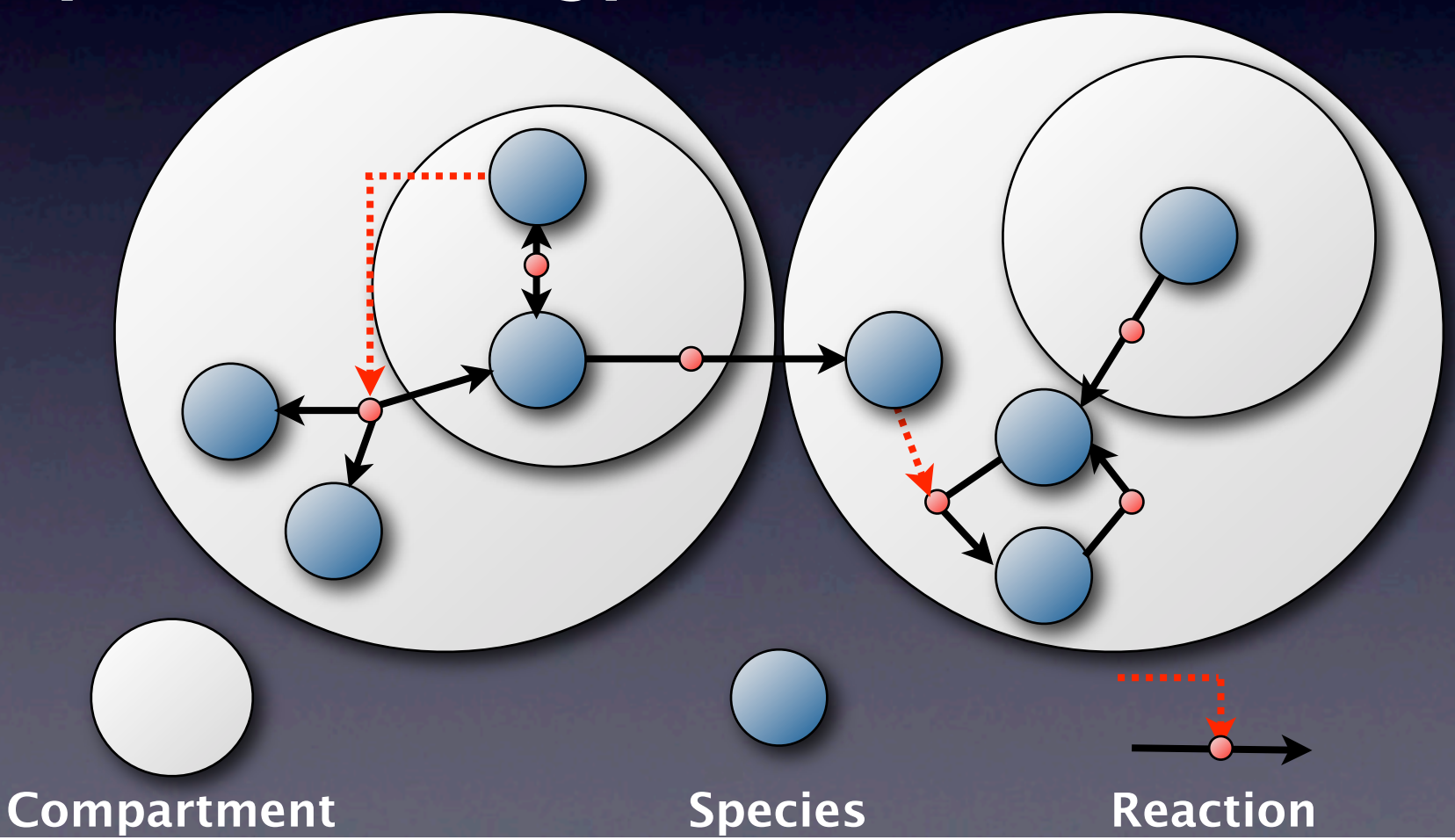




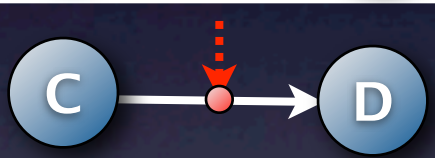
Modeling tool for biochemical and gene-regulatory network

SBML

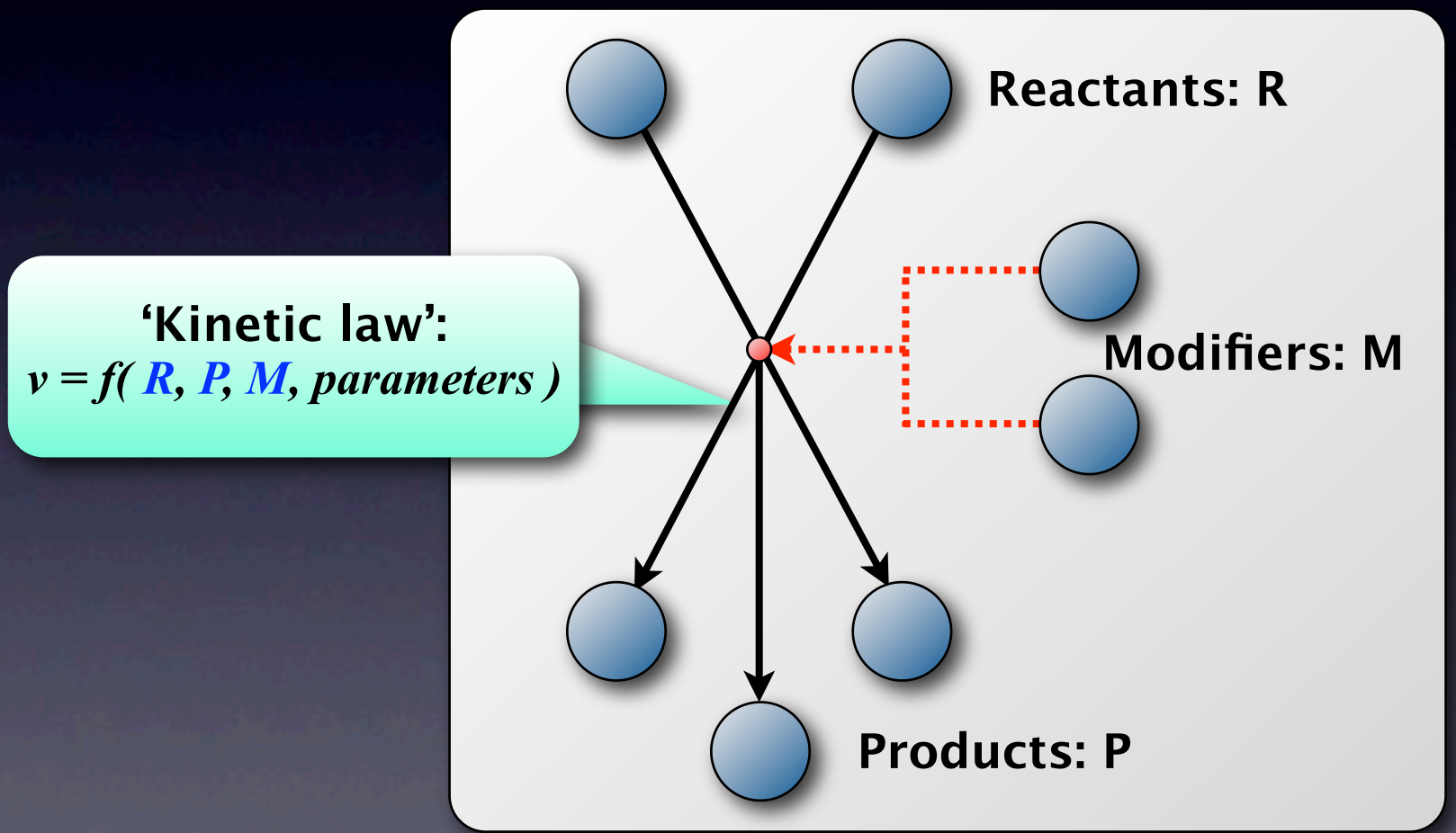
- SBML (Systems Biology Markup Language)
- A machine-readable format (XML) for representing computational models in systems biology



SBML model

listOfSpecies	listOfReactions	listOfCompartments
A		cytosol
B		nucleus
C		
D		

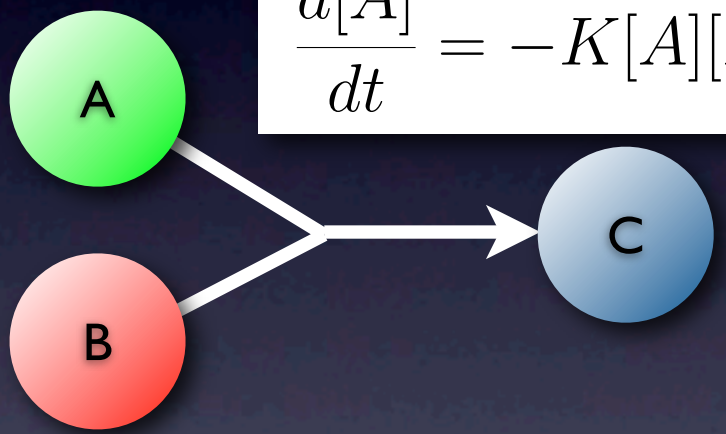
Reactions according to SBML



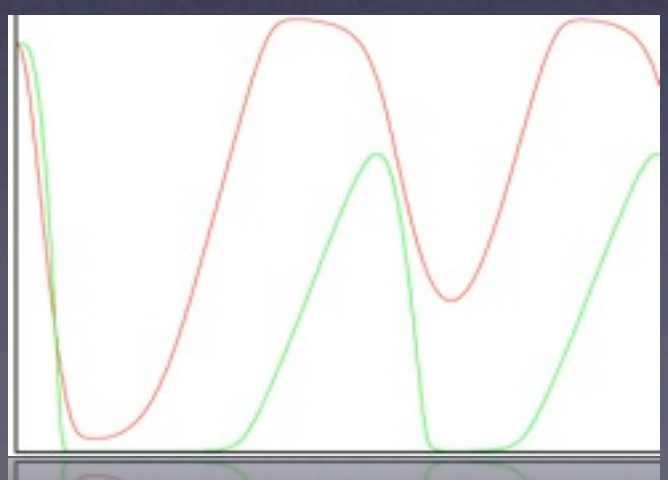
Kinetic law

- Describe the behavior of concentration, num. of molecules

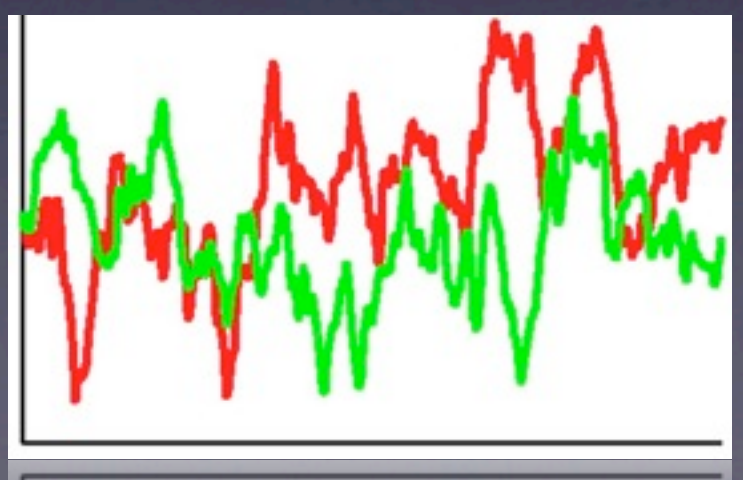
$$\frac{d[A]}{dt} = -K[A][B]$$



ODE



SSA



CellDesigner↔SBML

Biochemical reaction

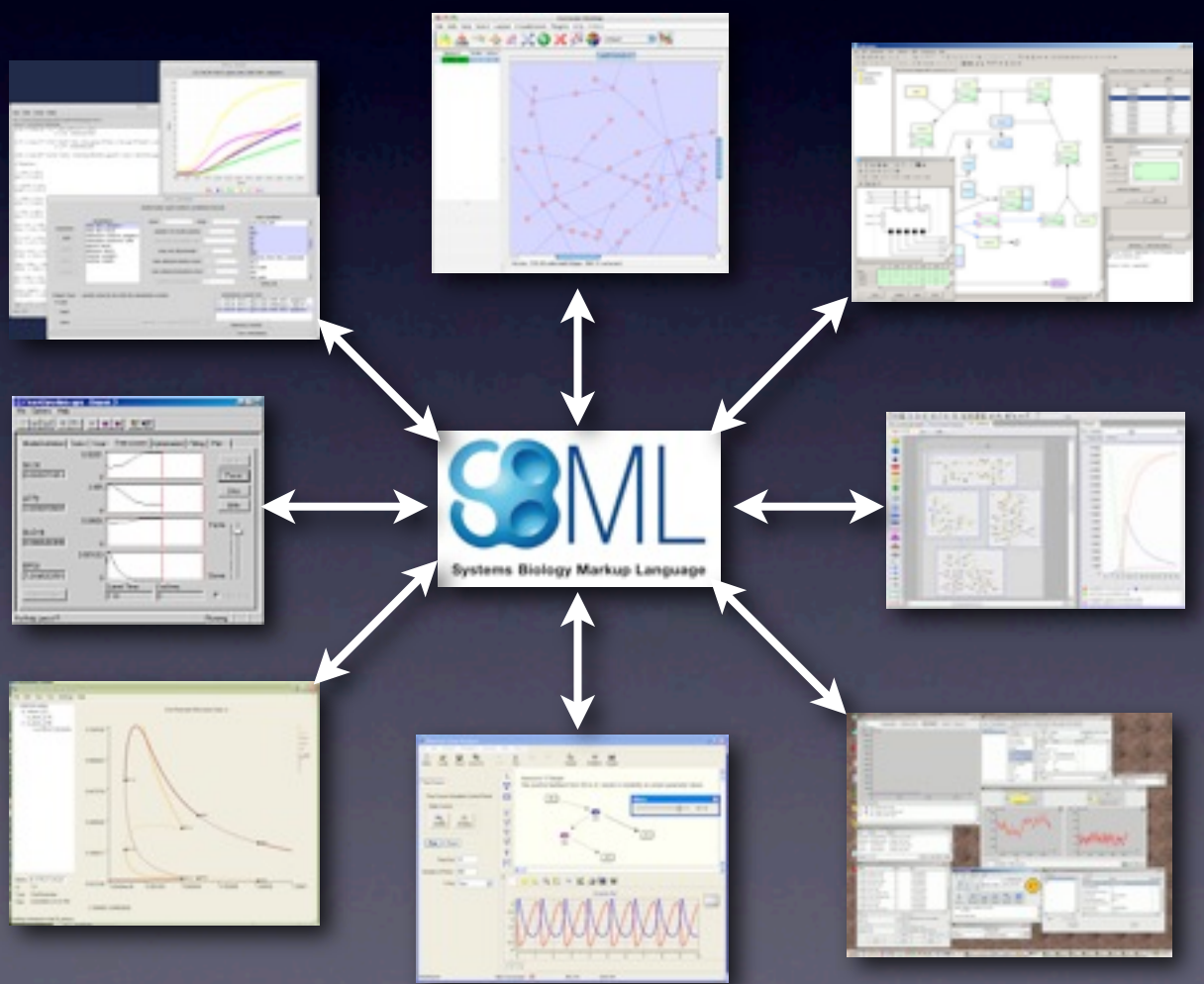

```
<listOfSpecies>
  <species id="s1" name="s1" compartment="default"
initialAmount="0" charge="0"/>
  <species id="s2" name="s2" compartment="default"
initialAmount="0" charge="0"/>
</listOfSpecies>
<listOfReactions>
  <reaction id="re1" reversible="false" fast="false">
    <listOfReactants>
      <speciesReference species="s1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="
    </listOfProducts>
    <kineticLaw formula="k*s1">
    </kineticLaw>
  </reaction>
</listOfReactions>
```

MathML

SBML

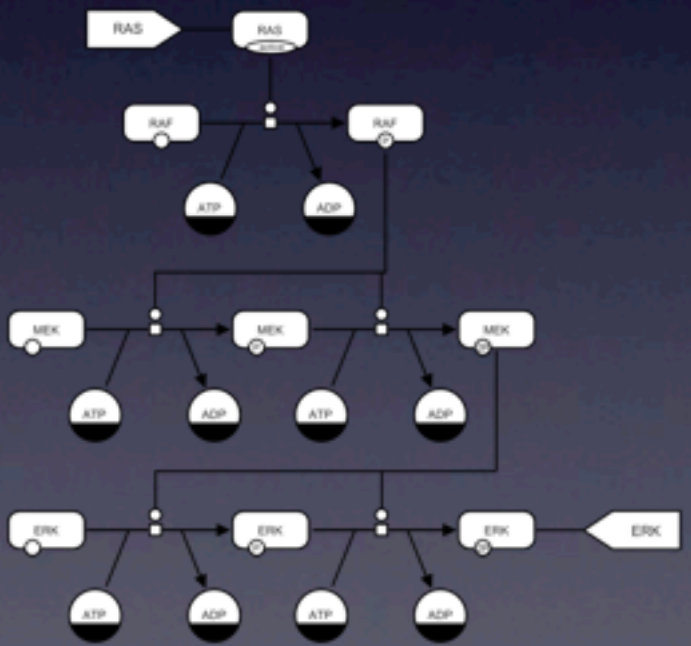
Applications supporting SBML

- Over 200 software packages support SBML
- http://sbml.org



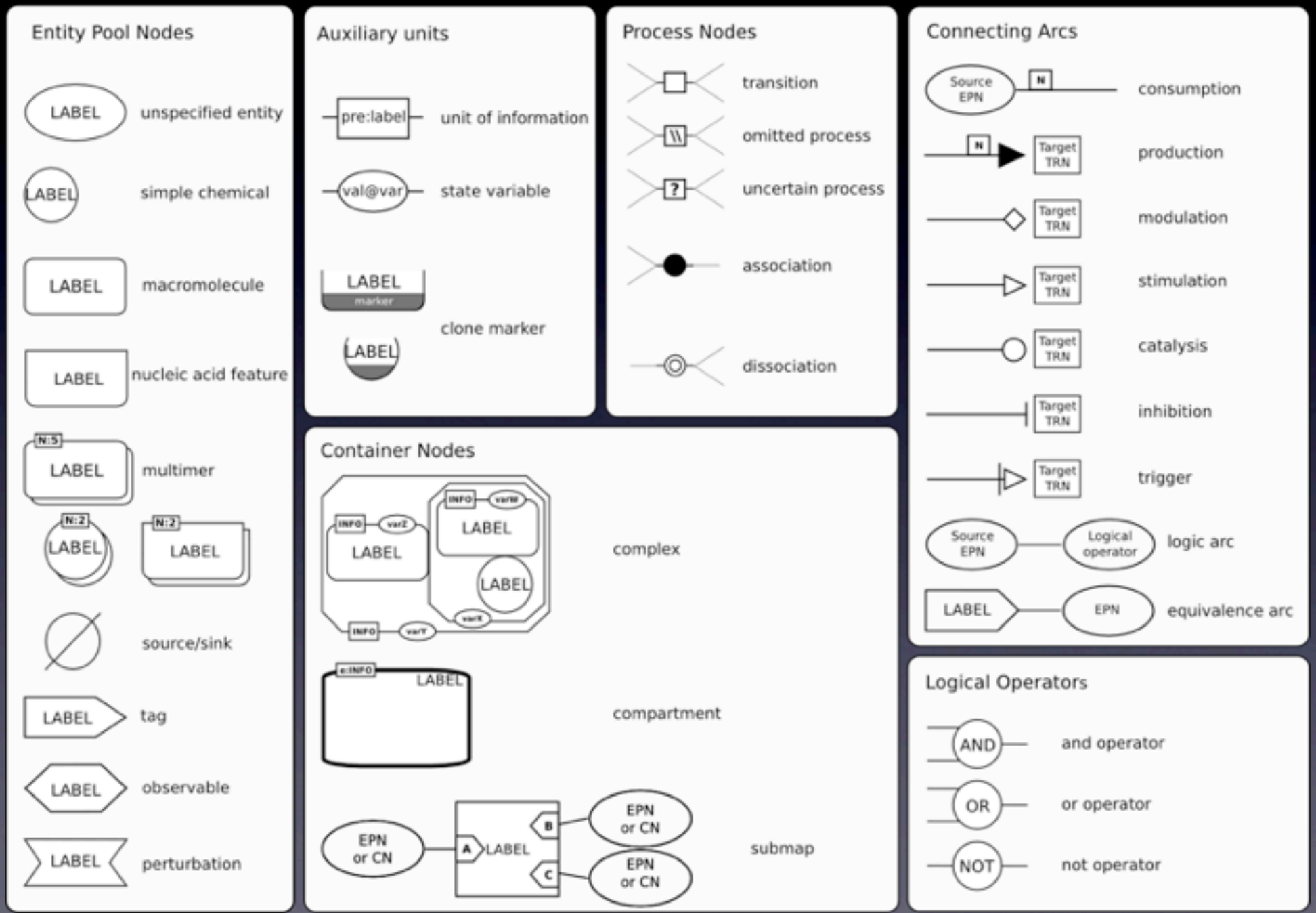
SBGN

- A Visual Notation for Network Diagrams in Biology
- Representation of Biochemical and Cellular Processes studied in Systems Biology http://sbgn.org

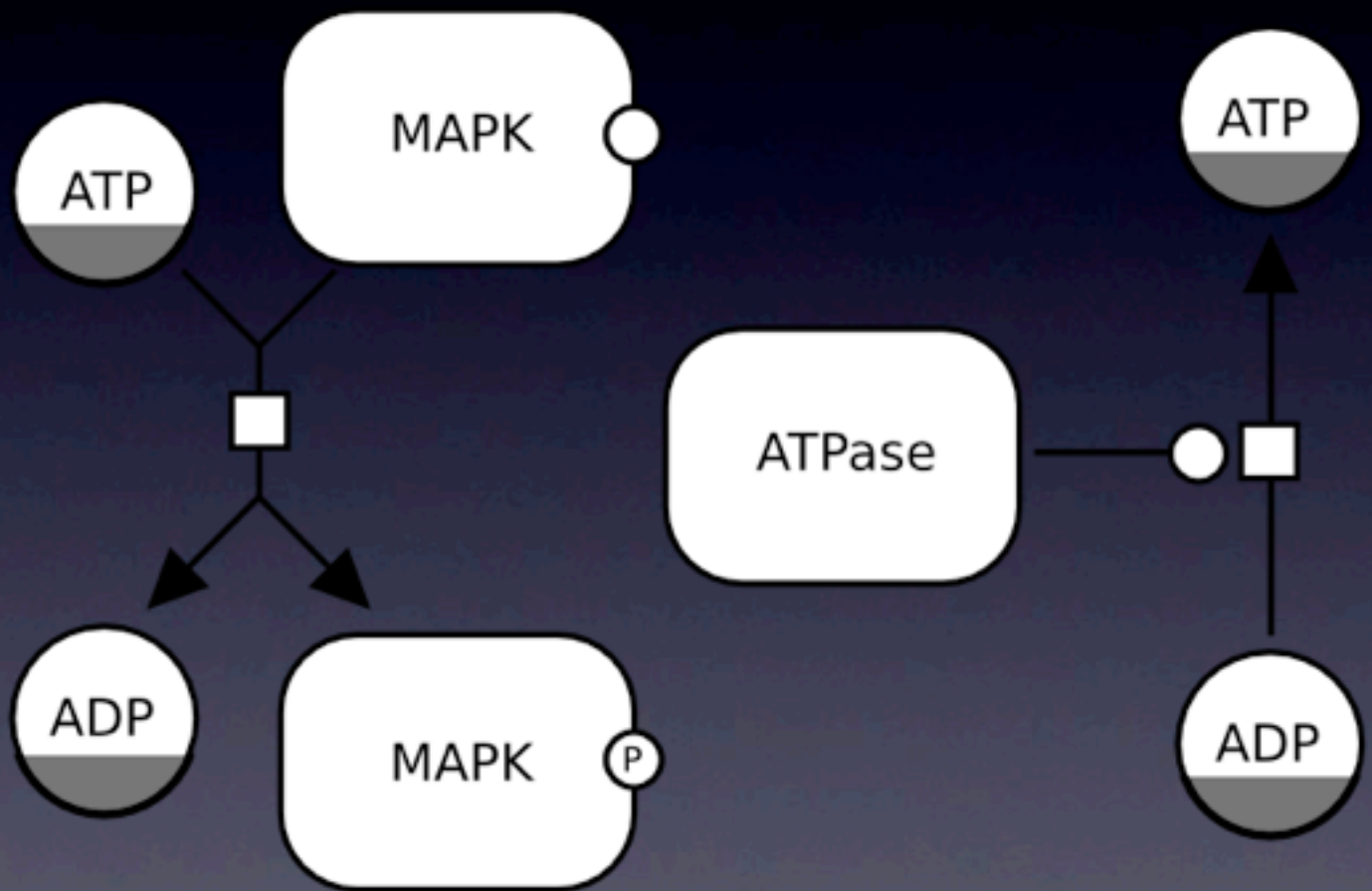


Le Novère N., et al.: The Systems Biology Graphical Notation, *Nature Biotechnology*, 27(8), pp.735-41, (2009)

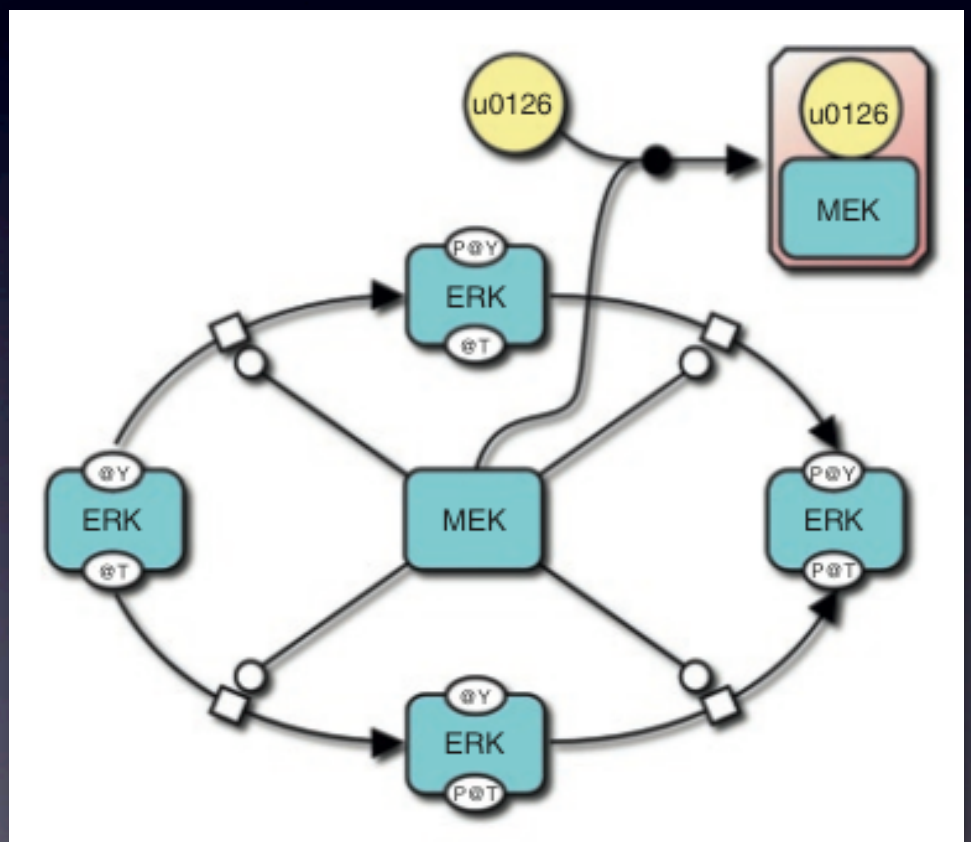
SBGN


Entity Pool Nodes
LABEL
unspecified entity
LABEL
simple chemical
LABEL
macromolecule
LABEL
nucleic acid feature
N:5
LABEL
multimer
N:2
LABEL
N:2
LABEL
source/sink
LABEL
tag
LABEL
observable
LABEL
perturbation
Auxiliary units
pre:label
unit of information
val@var
state variable
LABEL
marker
clone marker
LABEL
Process Nodes
transition
omitted process
?
uncertain process
association
dissociation
Connecting Arcs
Source EPN
N
consumption
N
Target TRN
production
Target TRN
modulation
Target TRN
stimulation
Target TRN
catalysis
Target TRN
inhibition
Target TRN
trigger
Source EPN
Logical operator
logic arc
LABEL
EPN
equivalence arc
Container Nodes
INFO
varZ
LABEL
INFO
varW
LABEL
LABEL
INFO
varY
varX
complex
e:INFO
LABEL
compartment
EPN or CN
A
LABEL
B
C
EPN or CN
EPN or CN
submap
Logical Operators
AND
and operator
OR
or operator
NOT
not operator

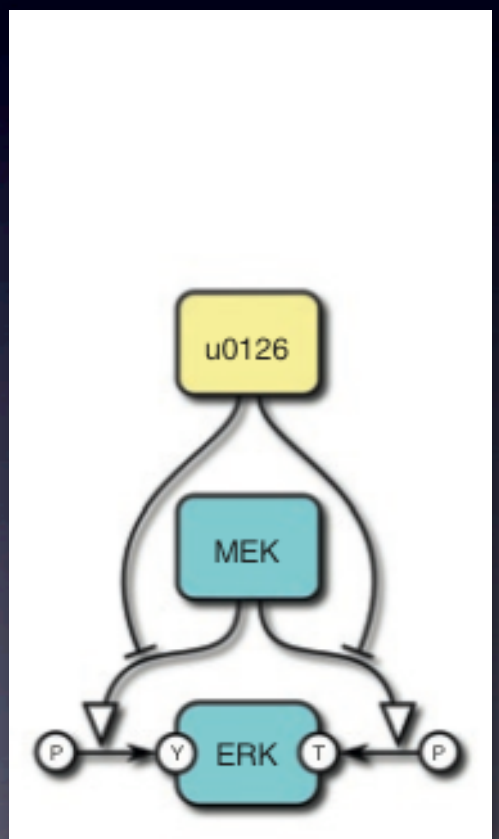
SBGN



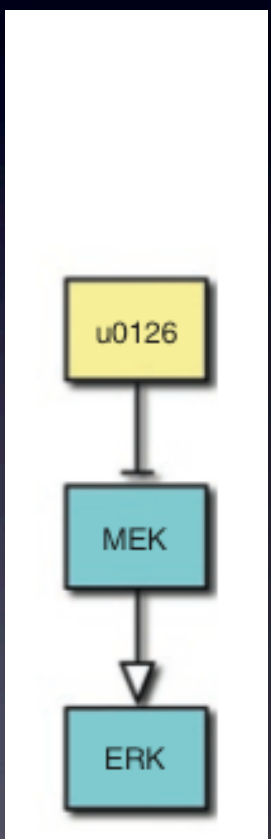
SBGN



Process Diagram

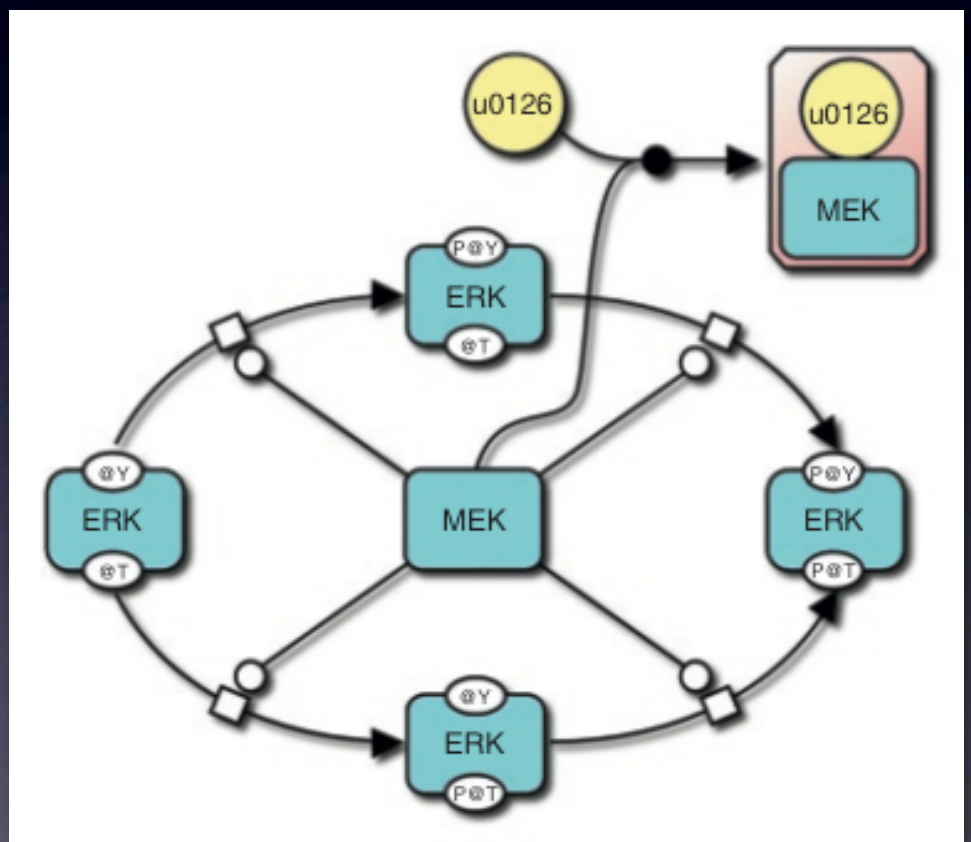


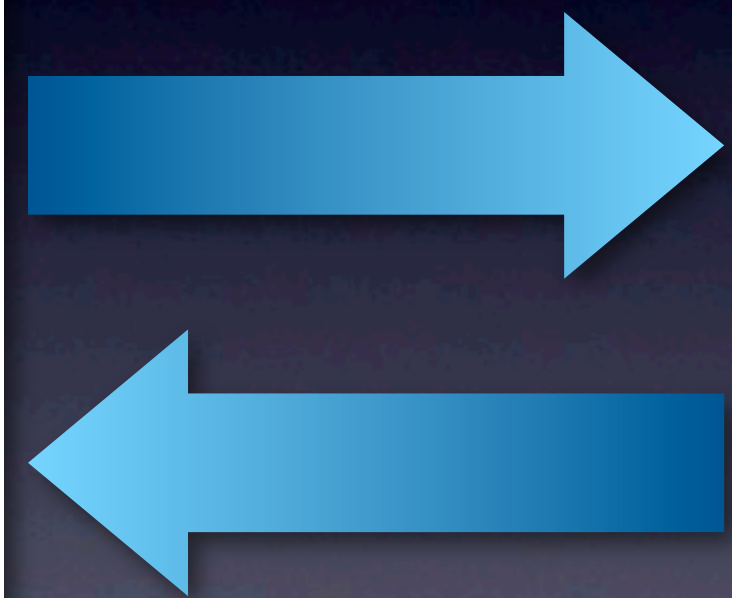
Entity Relationship

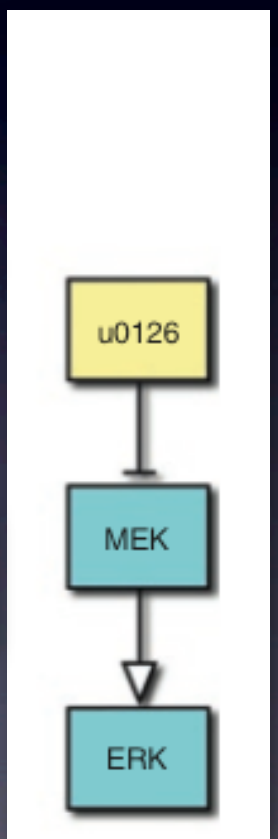


Activity Flow

SBGN





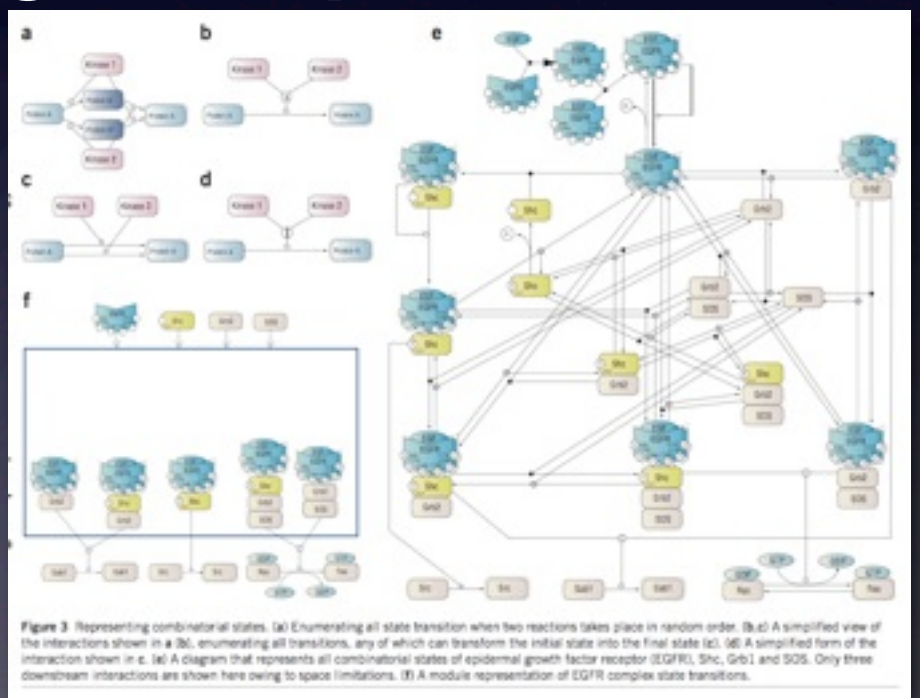


Process Diagram

Activity Flow

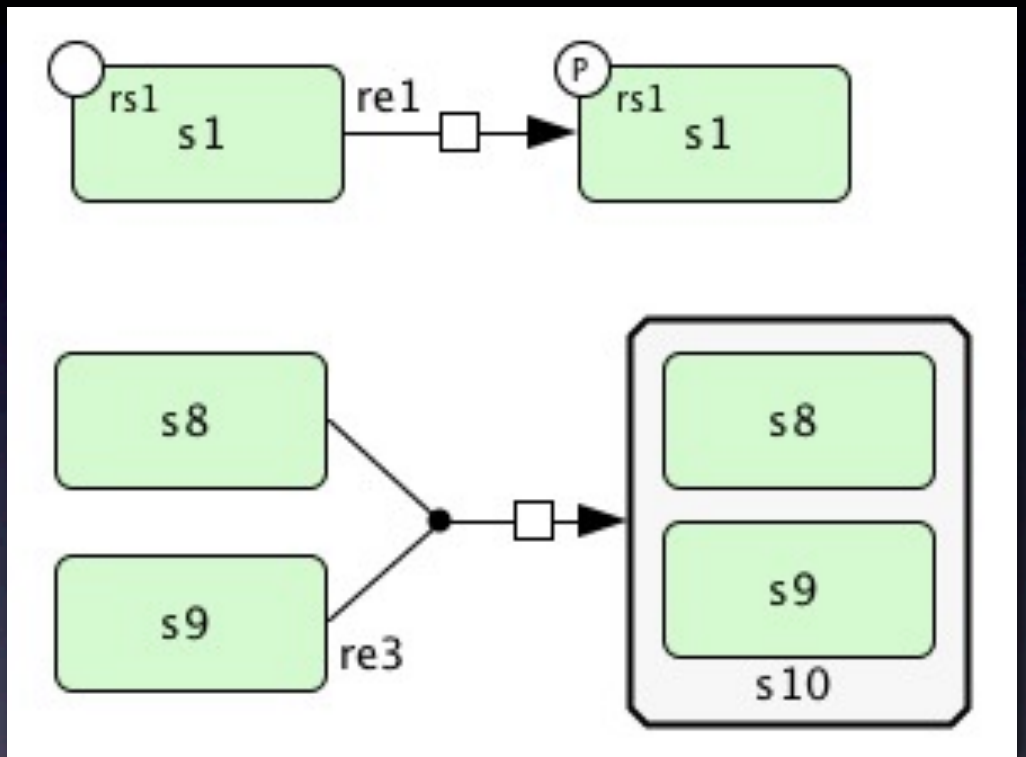
Graphical Notation

- Graphical Notation for representing biological interactions
- protein-protein interaction, gene regulatory networks



Kitano, H. et al. "Using process diagrams for the graphical representation of biological networks", Nature Biotechnology 23(8), 961 - 966 (2005)

State transition



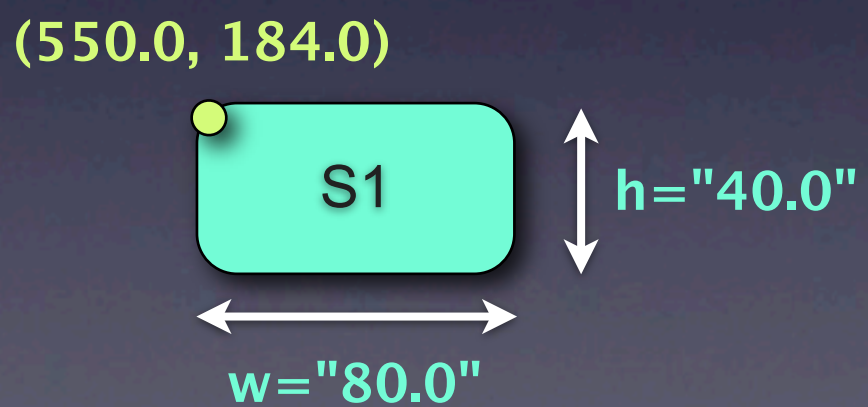
SBGN ⇄ SBML

- Species type, Reaction type is stored in <annotation> for each species, reactions
- Layout information is stored separately

```
<sbml>
 <model>
  <annotation>
    layout information
  </annotation>
  <listOfSpecies>
    <species>
      <annotation>species type</annotation>
    </species>
  </listOfSpecies>
 </model>
</sbml>
```

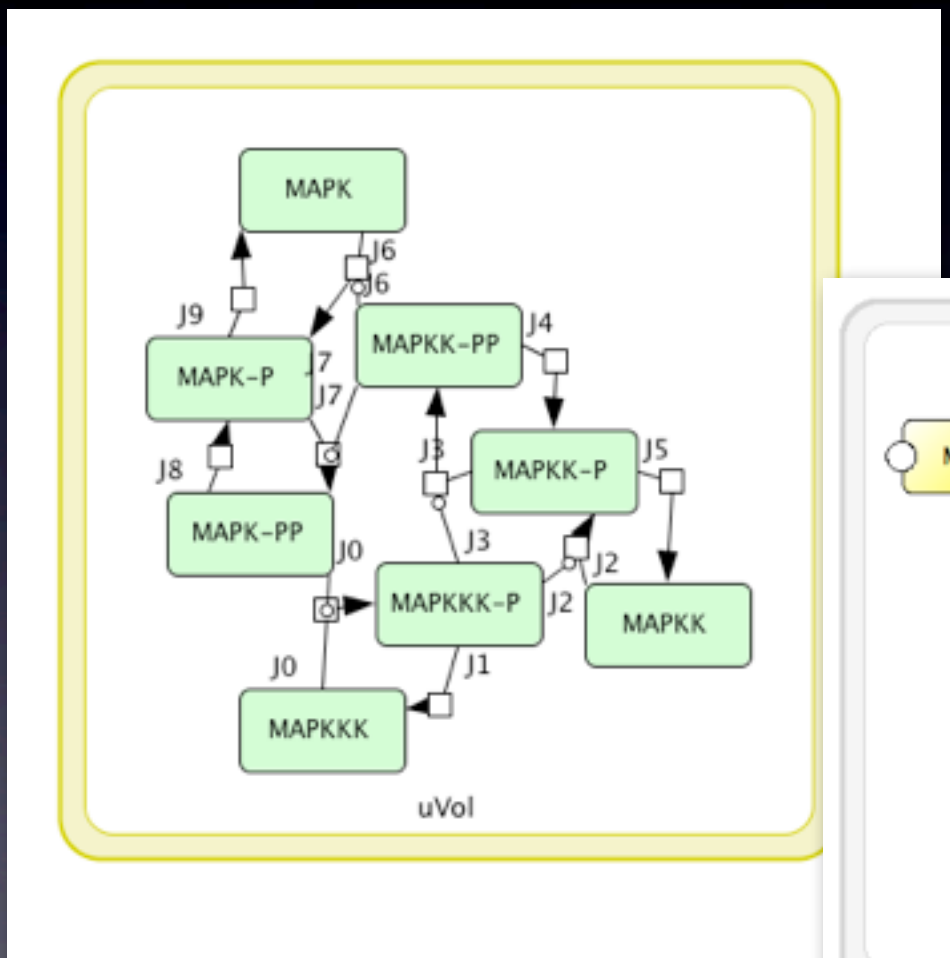
SBGN ⇄ SBML

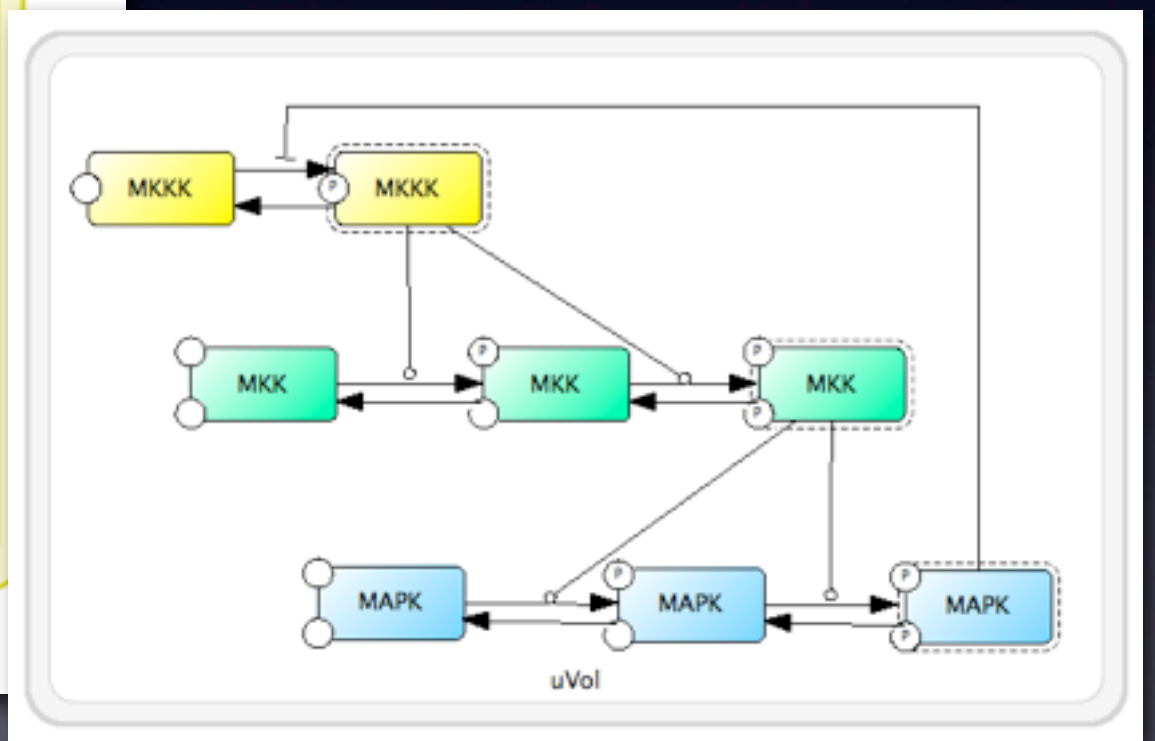
```
<celldesigner:speciesAlias compartmentAlias="ca3" id="a1" species="s1">
   <celldesigner:activity>active</celldesigner:activity>
   <celldesigner:bounds h="40.0" w="80.0" x="550.0" y="184.0">
   </celldesigner:bounds>
   <celldesigner:singleLine width="1.0"></celldesigner:singleLine>
   <celldesigner:paint color="ffb3d2ff" scheme="Gradation">
   </celldesigner:paint>
</celldesigner:speciesAlias>
```



SBGN ⇄ SBML

Pure SBML (w/o Graphical Notation)

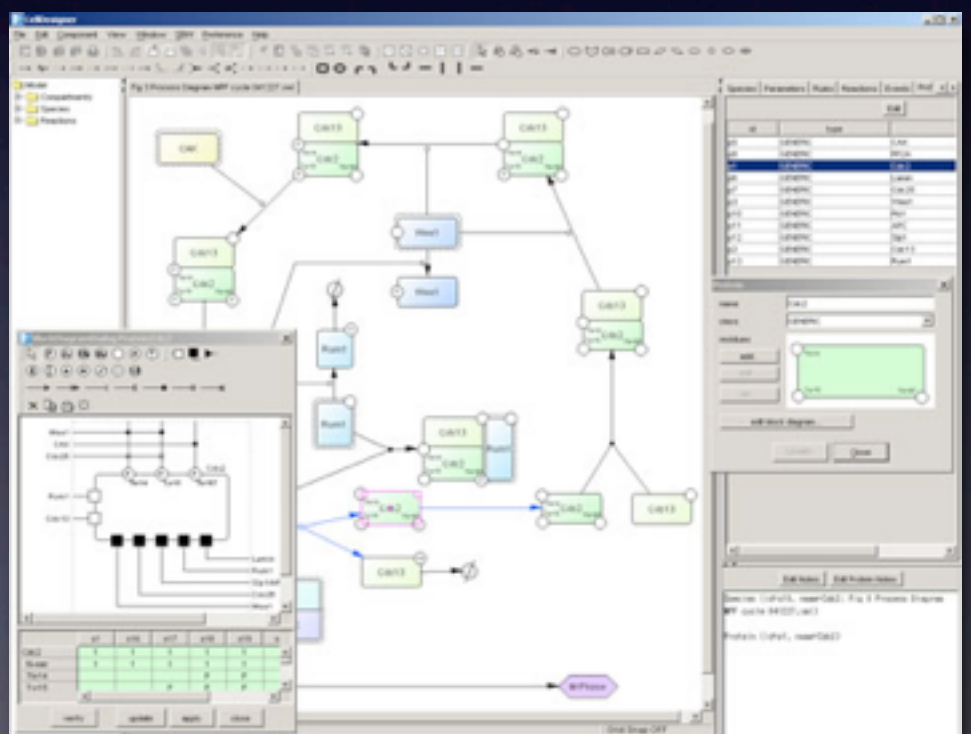




w/ Graphical Notation

CellDesigner 4.1

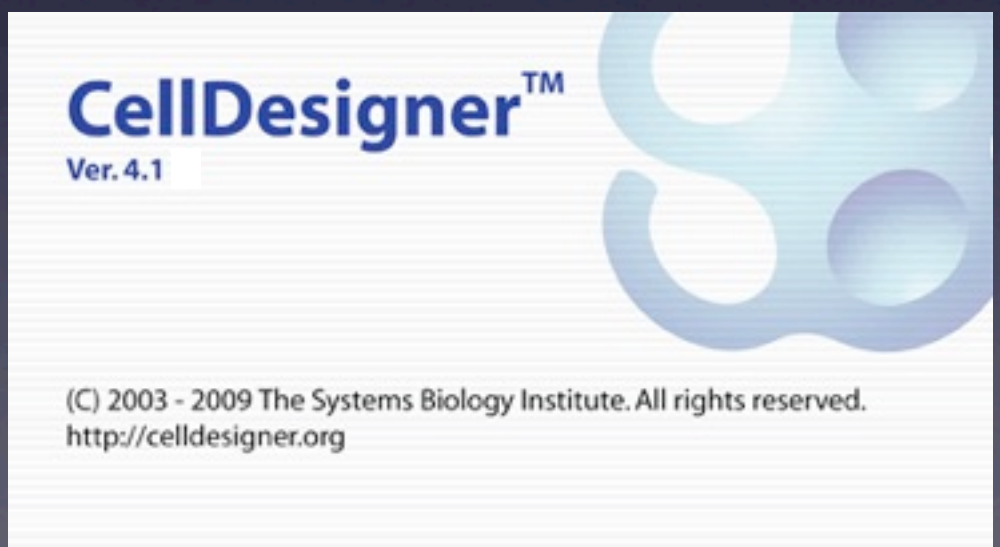
- SBML support
- Graphical notation (SBGN)
- Built-in simulator (SBML ODE Solver, COPASI)
- Integrate with Analysis tool, other simulators through SBW
- Database connection
- Export to PDF, PNG, etc.
- Freely available
- Supported Environment
  - Windows (XP or later)
  - Mac OS X (Tiger, Leopard)
  - Linux



http://celldesigner.org

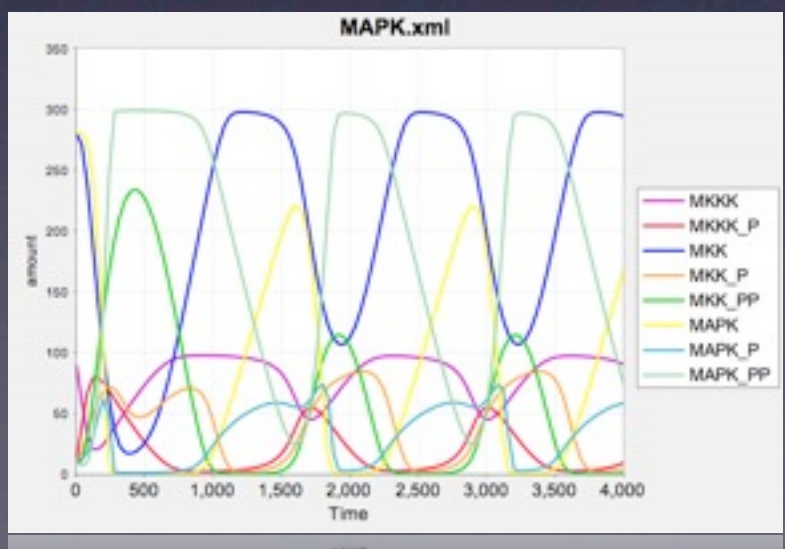
What's new

- SBML L2v4 support
- MIRIAM annotation
- SBOTerm
- SBGN Process Diagram Level-1
- Integration with SABIO-RK
- Connect to PANTHER
- new Plugin API
- GUI improvement



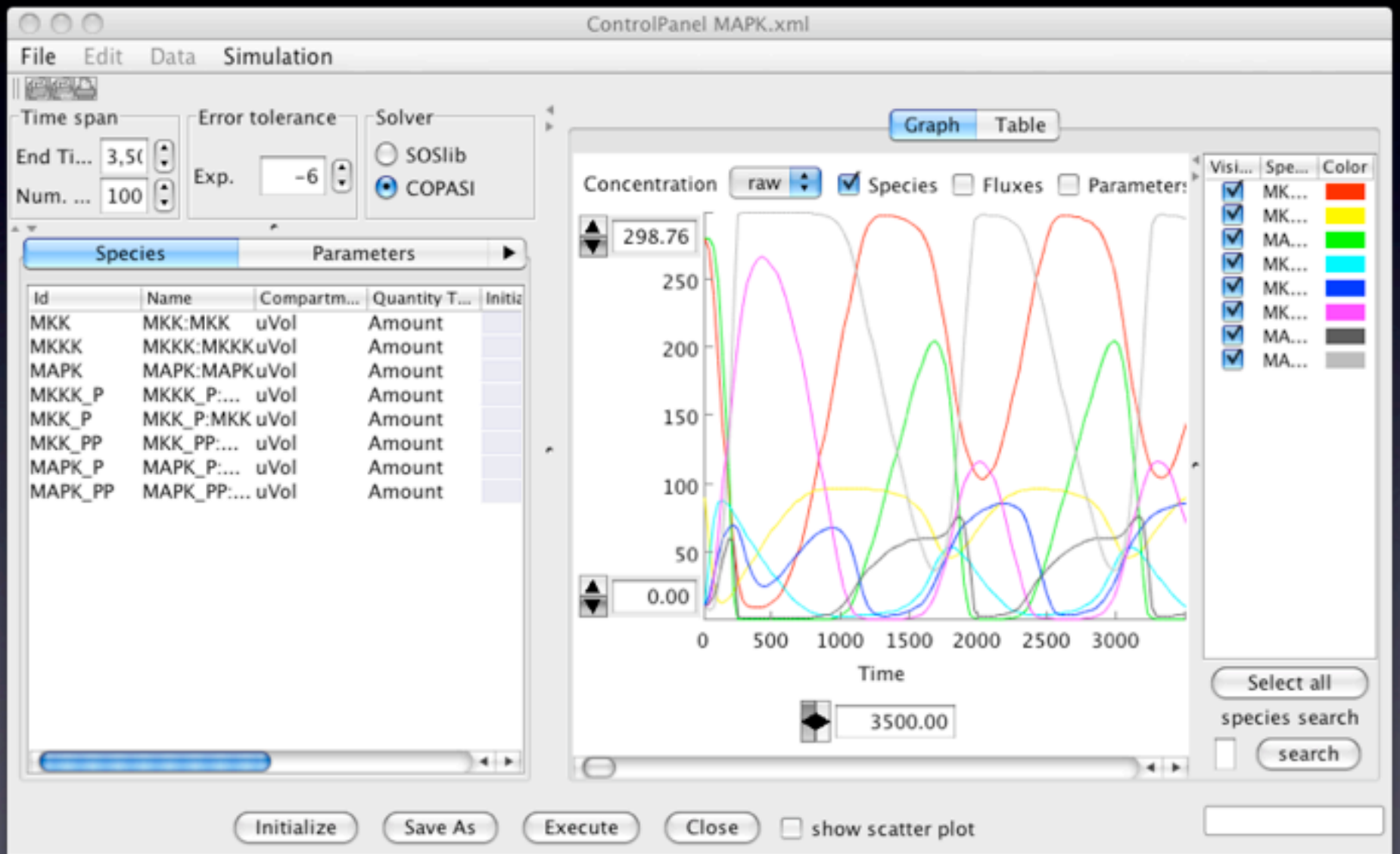
Simulator

- SOSLib: SBML ODE Solver Library (Univ. of Vienna) http://sbmlsolver.sf.net/
  - written in ANSI C
  - call CVODE for integration
- COPASI: (Univ. of Manchester, VBI, Univ. of Heidelberg) http://copasi.org/
  - ODE & Stochastic simulation
  - Language Bindings



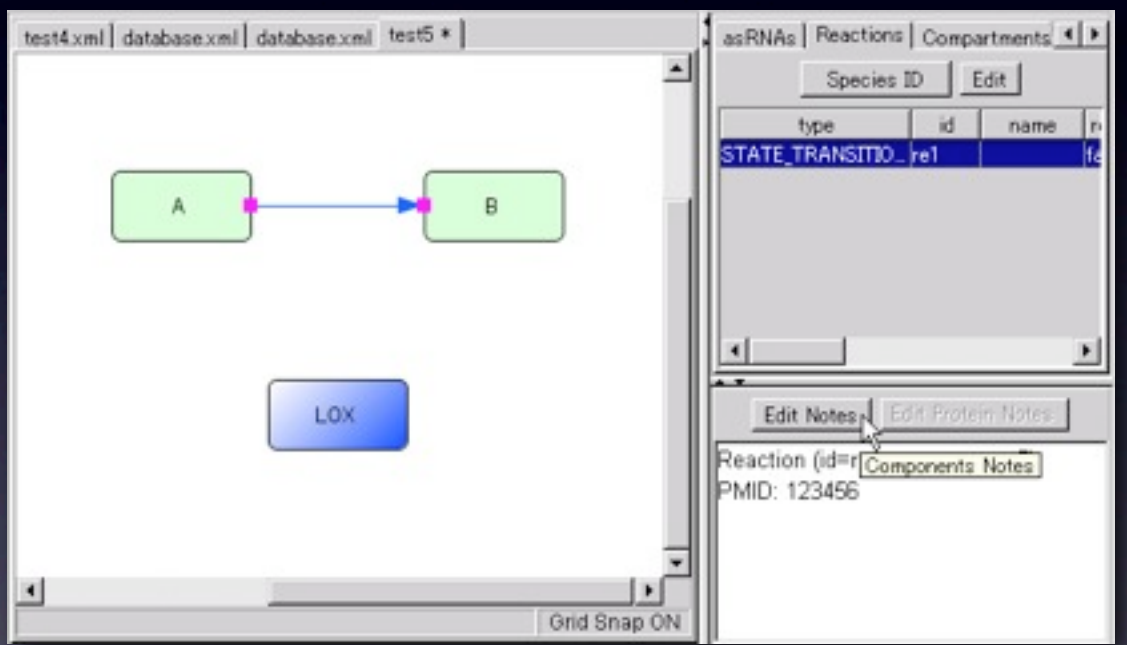
Simulation

- Can call SOSlib / COPASI as a solver



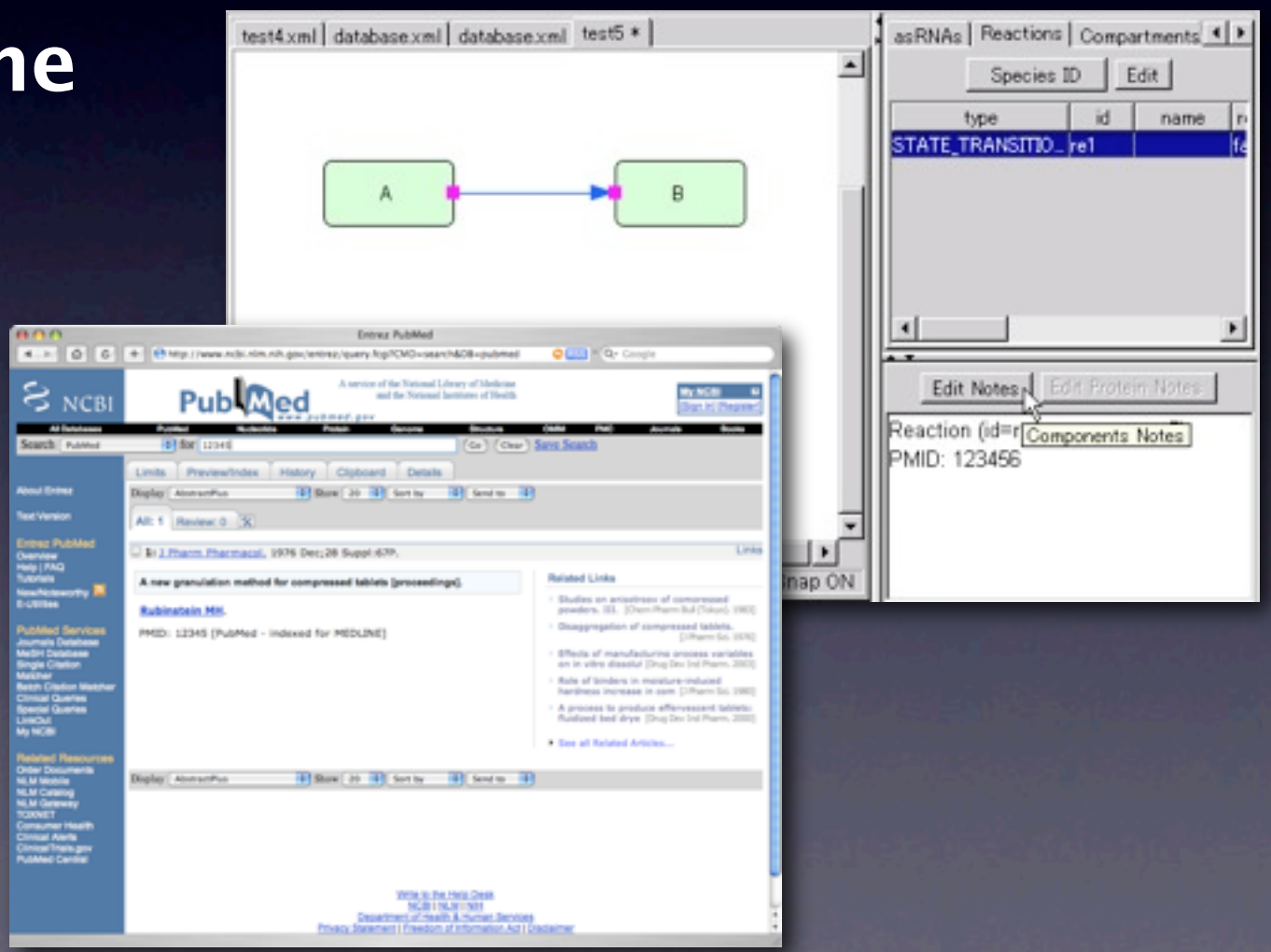
Database connection

- Search Database by Notes, Name:
  - PubMed: PMID: 123456
  - Entrez Gene
  - SGD
  - DBGET
  - iHOP



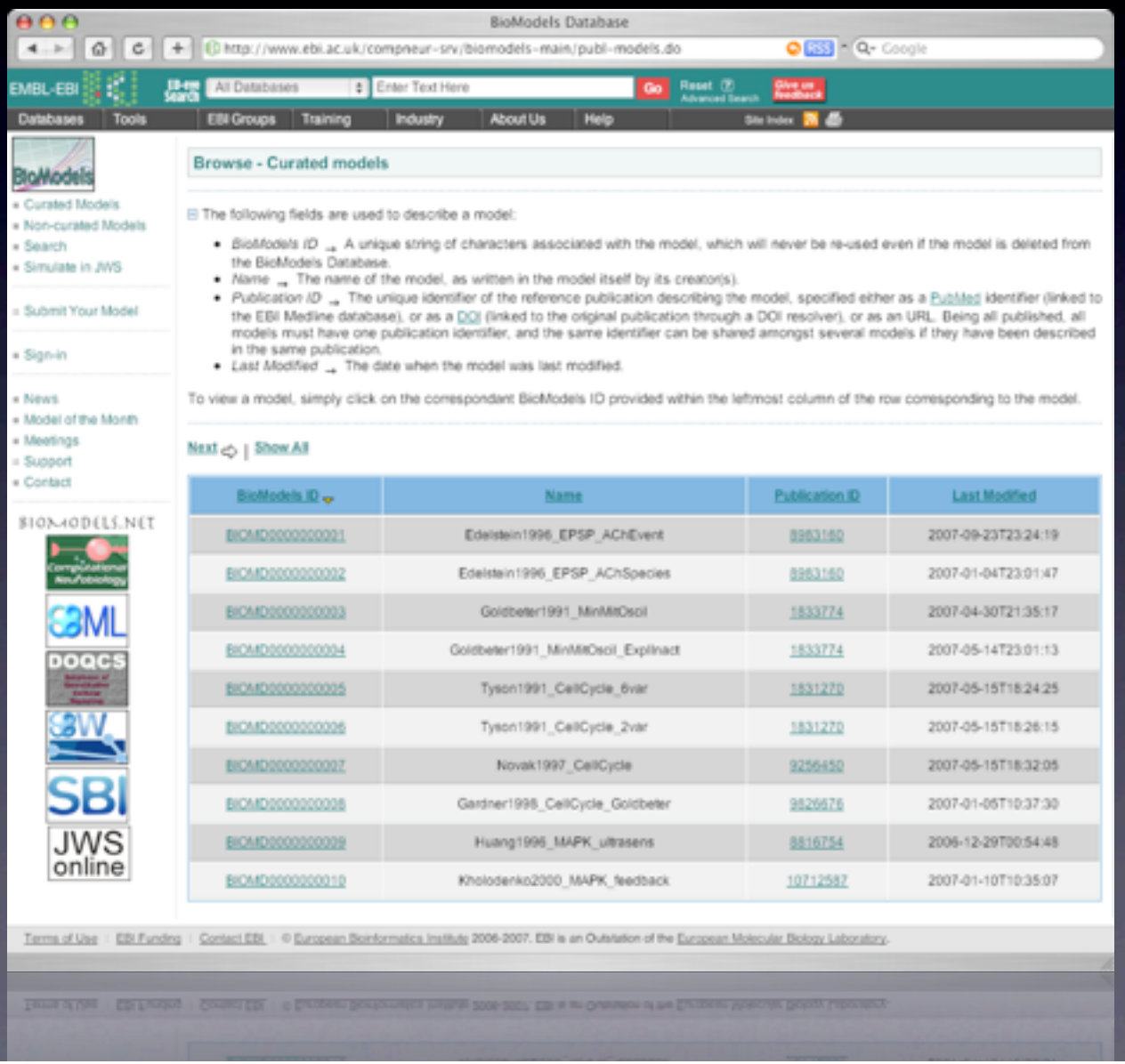
Database connection

- Search Database by Notes, Name:
  - PubMed: PMID: 123456
  - Entrez Gene
  - SGD
  - DBGET
  - iHOP



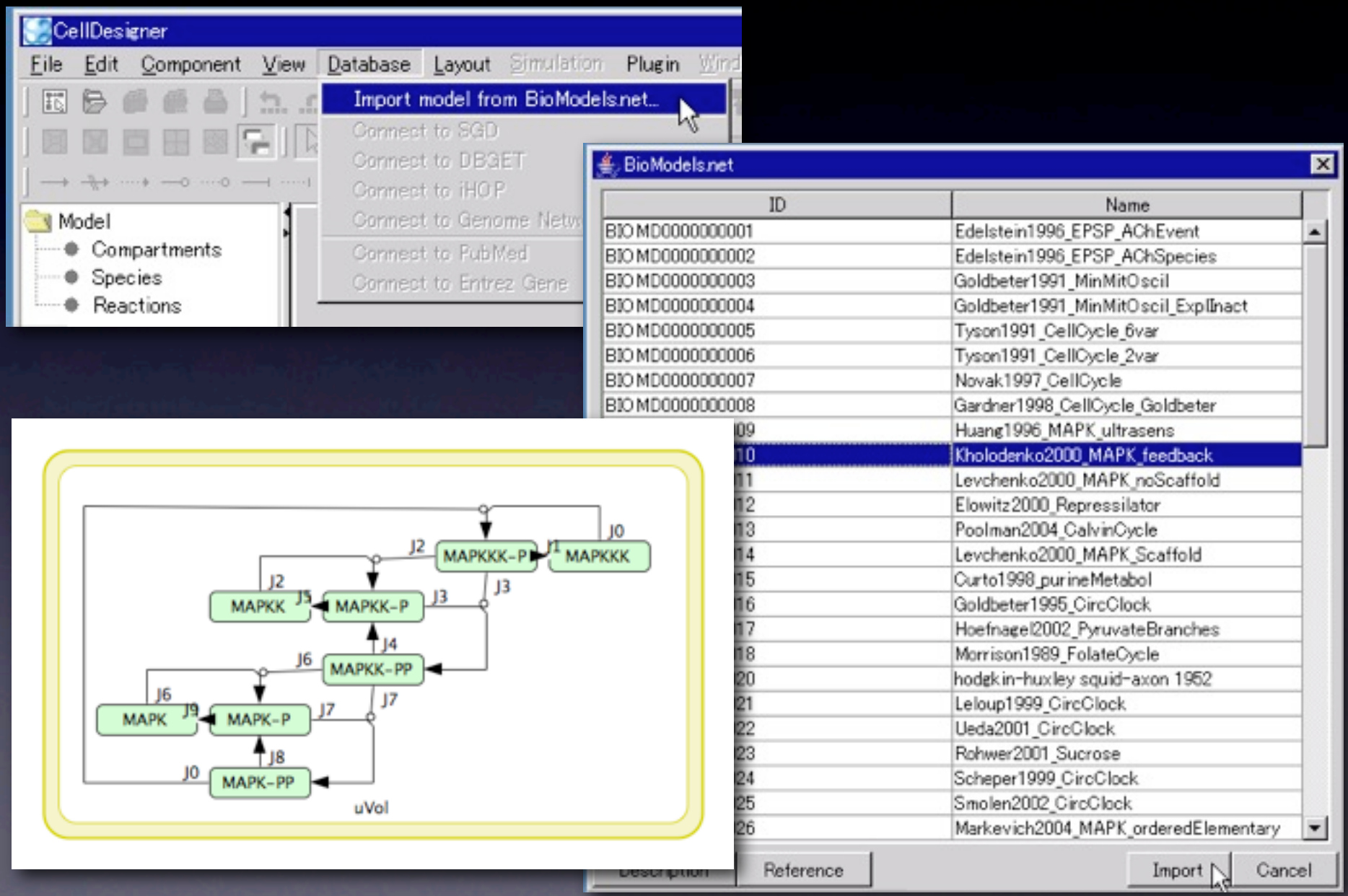
Database connection

- Import model from BioModels.net



Database connection

- Import model from BioModels.net

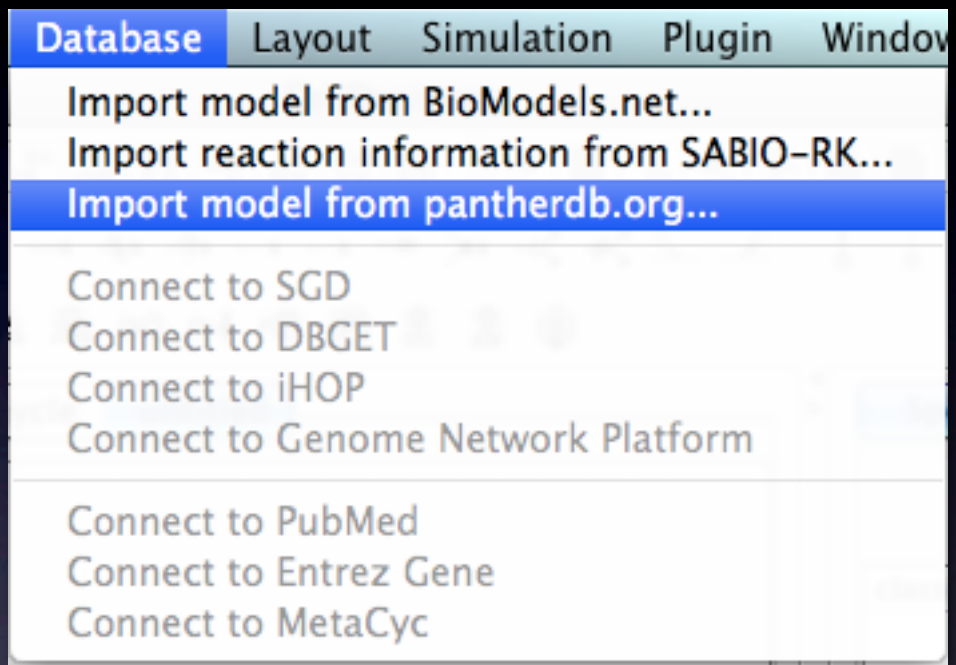


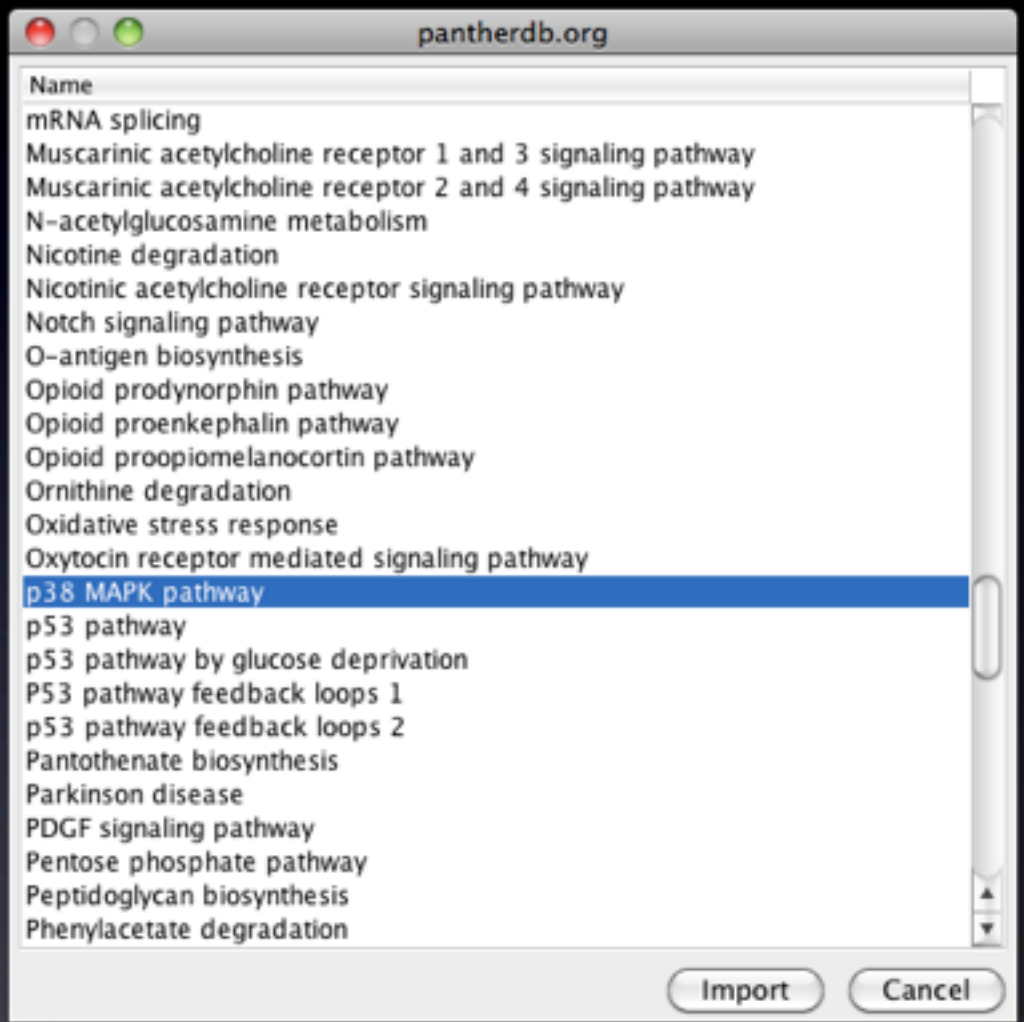
Database connection

- Import model from PANTHER

Database connection

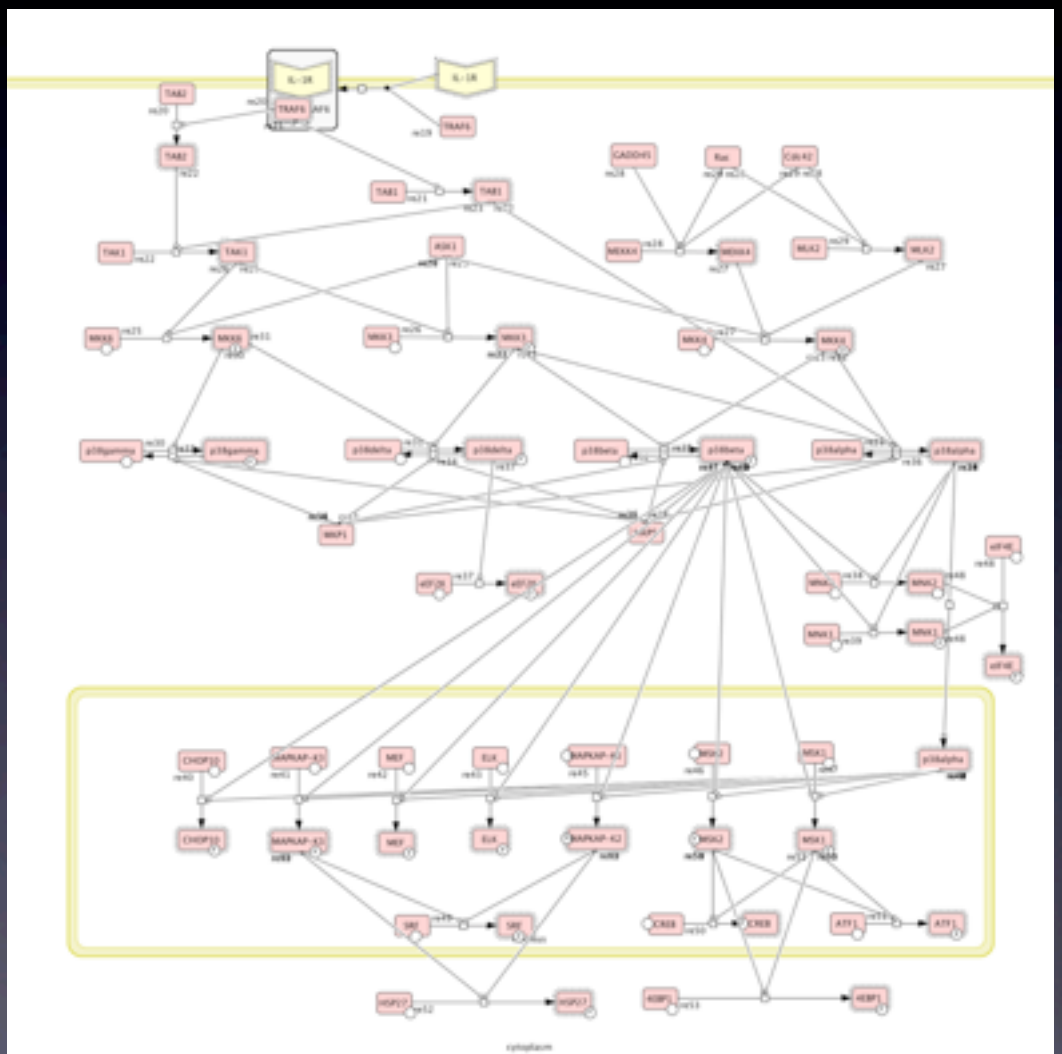
- Import model from PANTHER

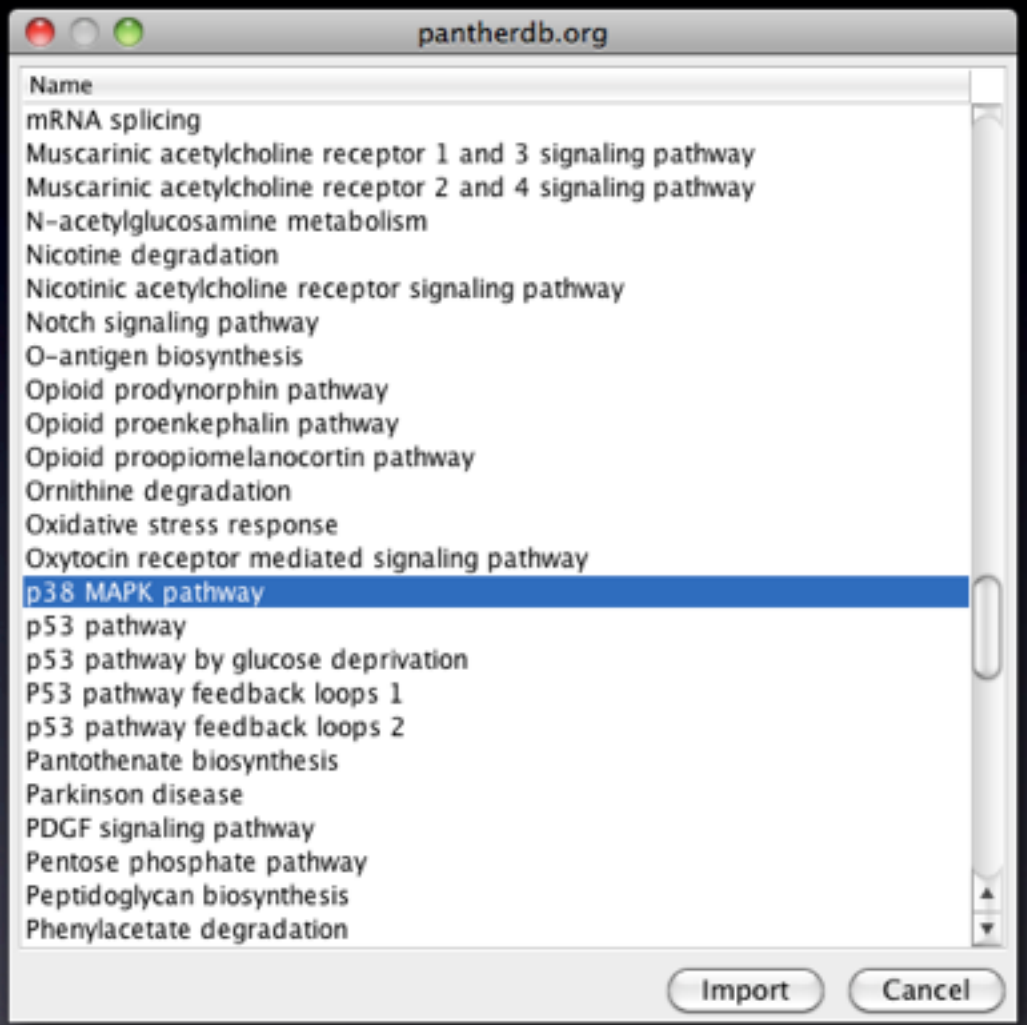




Database connection

- Import model from PANTHER





SABIO-RK

- Web-accessible database
- http://sabio.villa-bosch.de/
- Contains information about biochemical reactions, related kinetic equations and parameters

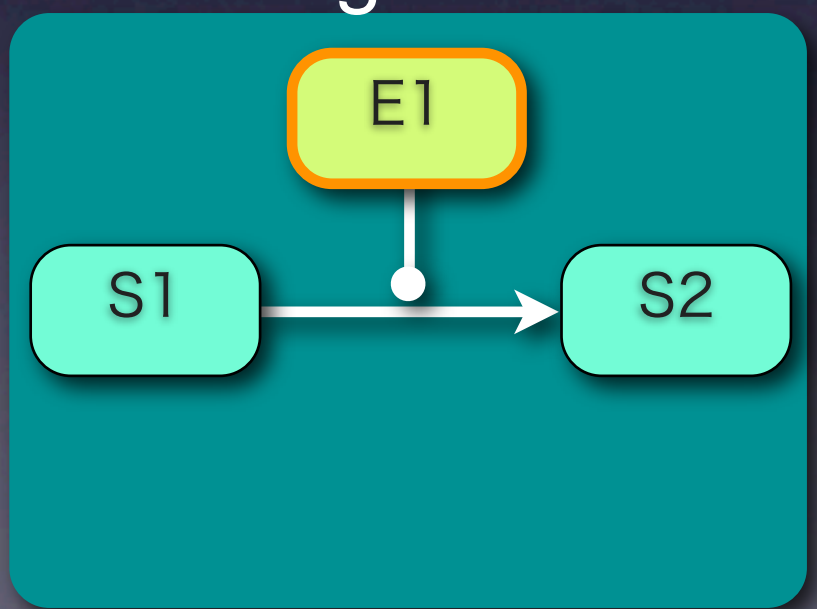
Heidelberg Institute for Theoretical Studies

HITS

CellDesigner ⇔ SABIO RK

- Users can import additional information to each object (reaction) on-the-fly
- SBML (Systems Biology Markup Language) is used to exchange information

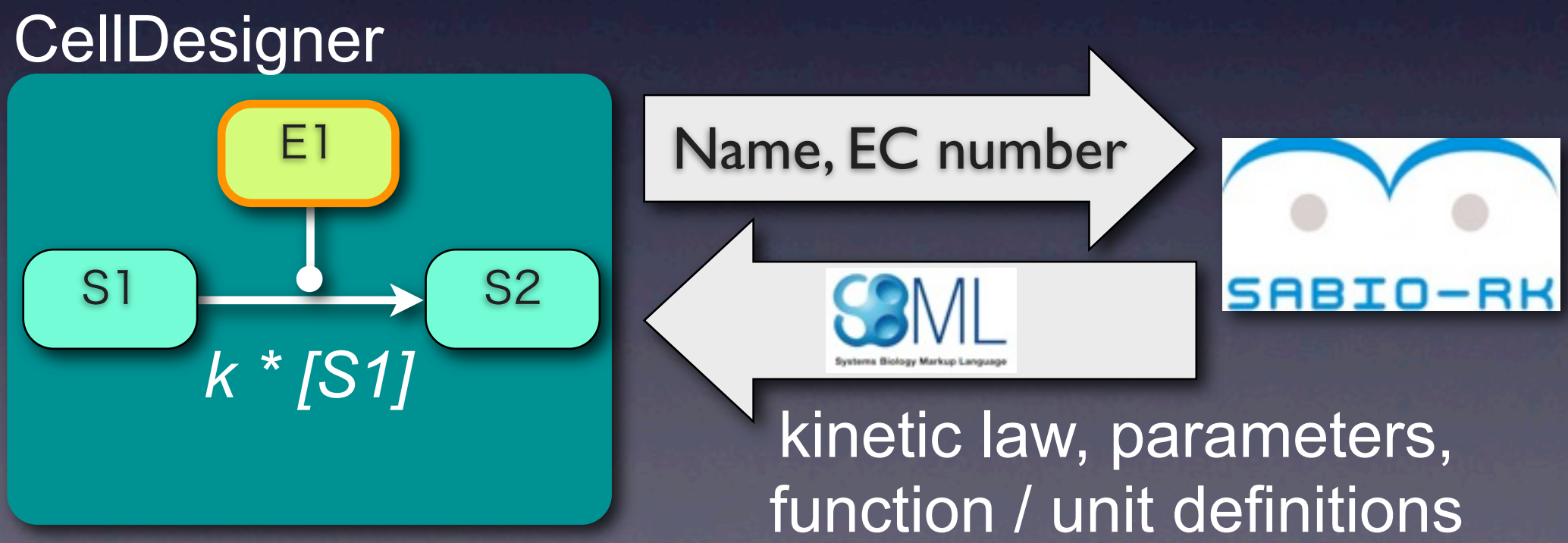
CellDesigner



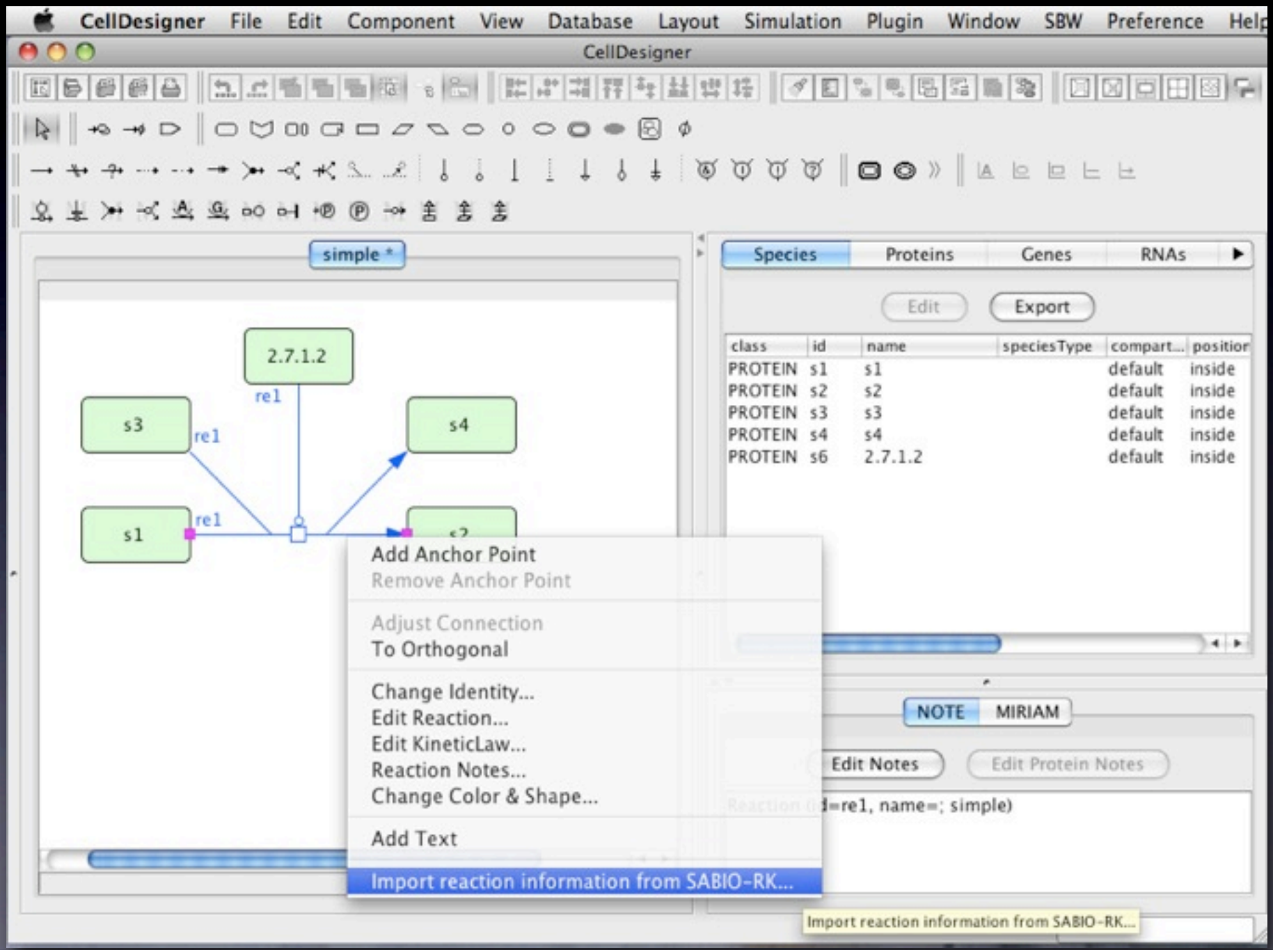
SABIO-RK

CellDesigner ⇔ SABIO RK

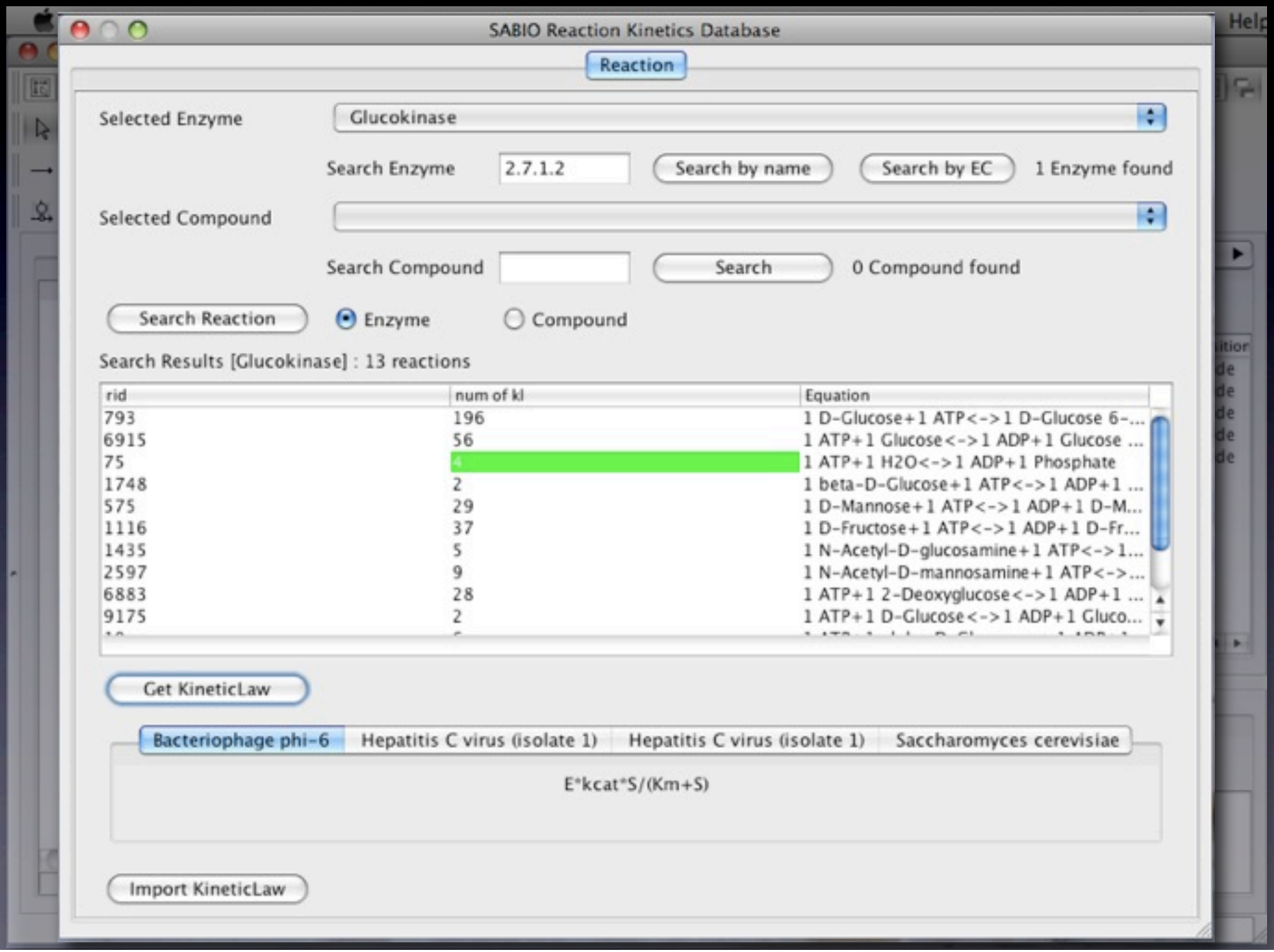
- Users can import additional information to each object (reaction) on-the-fly
- SBML (Systems Biology Markup Language) is used to exchange information



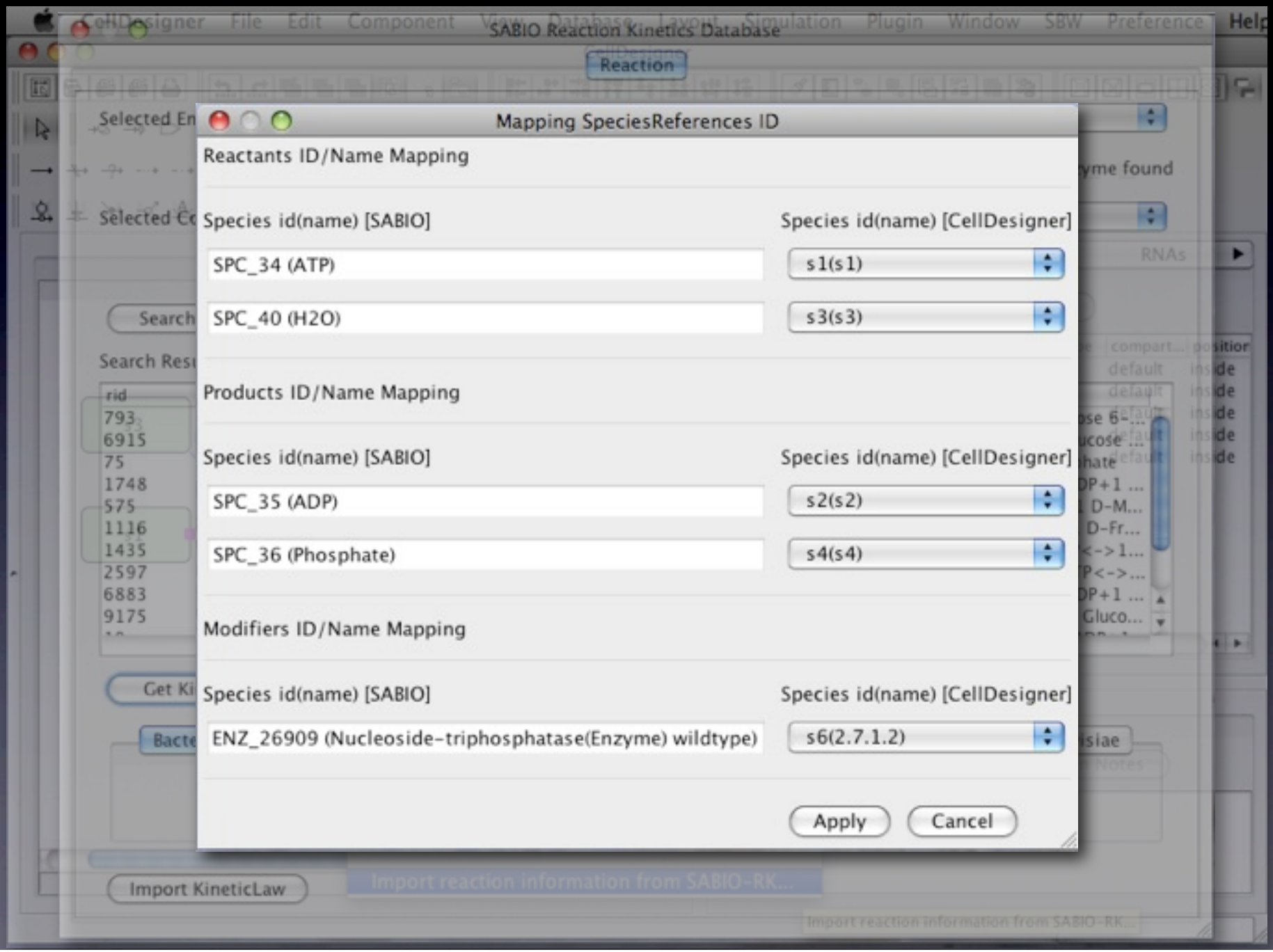
Example


CellDesigner File Edit Component View Database Layout Simulation Plugin Window SBW Preference Help
CellDesigner
simple *
2.7.1.2
re1
s3
re1
s4
s1
re1
Species
Proteins
Genes
RNAs
Edit
Export
class id name speciesType compart... positior
PROTEIN s1 s1 default inside
PROTEIN s2 s2 default inside
PROTEIN s3 s3 default inside
PROTEIN s4 s4 default inside
PROTEIN s6 2.7.1.2 default inside
Add Anchor Point
Remove Anchor Point
Adjust Connection
To Orthogonal
Change Identity...
Edit Reaction...
Edit KineticLaw...
Reaction Notes...
Change Color & Shape...
Add Text
Import reaction information from SABIO-RK...
NOTE
MIRIAM
Edit Notes
Edit Protein Notes
=re1, name=; simple)
Import reaction information from SABIO-RK...

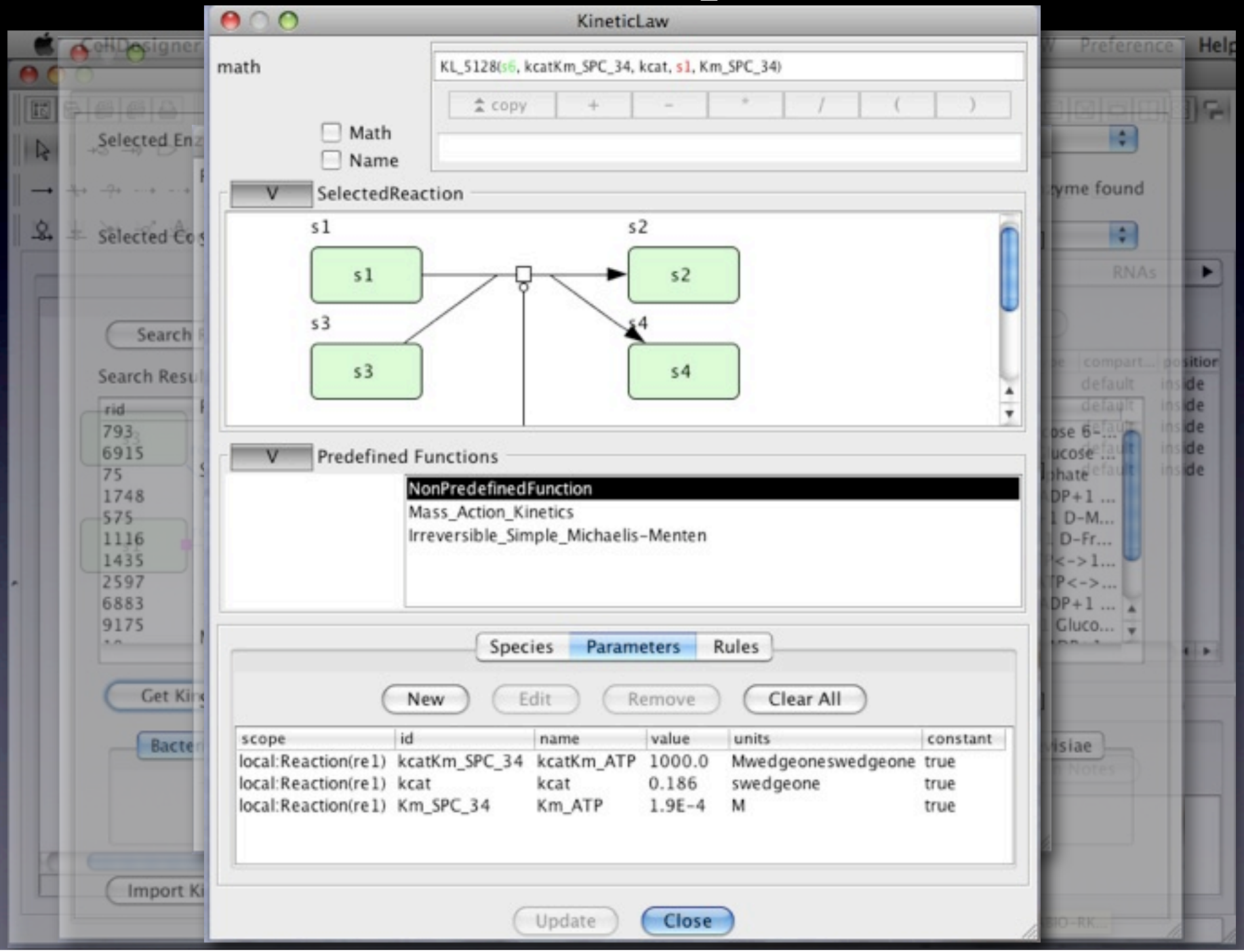
Example


SABIO Reaction Kinetics Database
Reaction
Selected Enzyme
Glucokinase
Search Enzyme
2.7.1.2
Search by name
Search by EC
1 Enzyme found
Selected Compound
Search Compound
Search
0 Compound found
Search Reaction
Enzyme
Compound
Search Results [Glucokinase] : 13 reactions
rid
num of kl
Equation
793
196
1 D-Glucose+1 ATP<->1 D-Glucose 6-...
6915
56
1 ATP+1 Glucose<->1 ADP+1 Glucose ...
75
4
1 ATP+1 H2O<->1 ADP+1 Phosphate
1748
2
1 beta-D-Glucose+1 ATP<->1 ADP+1 ...
575
29
1 D-Mannose+1 ATP<->1 ADP+1 D-M...
1116
37
1 D-Fructose+1 ATP<->1 ADP+1 D-Fr...
1435
5
1 N-Acetyl-D-glucosamine+1 ATP<->1...
2597
9
1 N-Acetyl-D-mannosamine+1 ATP<->...
6883
28
1 ATP+1 2-Deoxyglucose<->1 ADP+1 ...
9175
2
1 ATP+1 D-Glucose<->1 ADP+1 Gluco...
Get KineticLaw
Bacteriophage phi-6
Hepatitis C virus (isolate 1)
Hepatitis C virus (isolate 1)
Saccharomyces cerevisiae
E*kcat*S/(Km+S)
Import KineticLaw

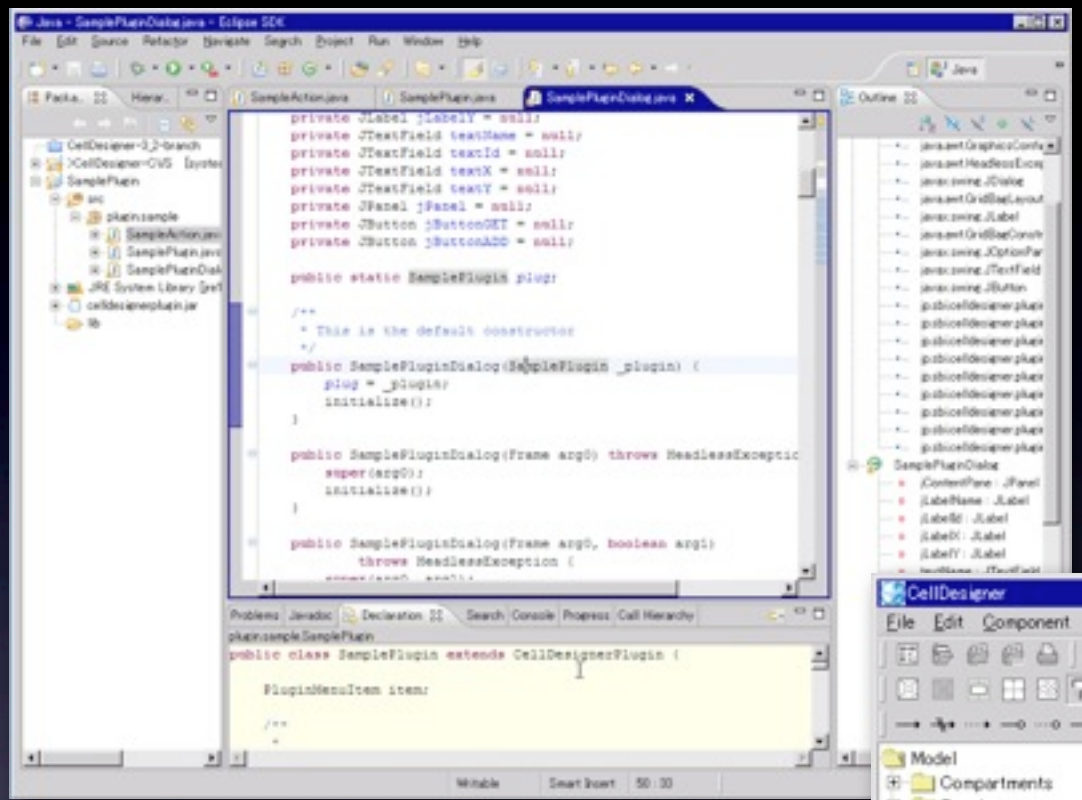
Example


SABIO Reaction Kinetics Database
Reaction
Mapping SpeciesReferences ID
Reactants ID/Name Mapping
Species id(name) [SABIO]
Species id(name) [CellDesigner]
SPC_34 (ATP)
s1(s1)
SPC_40 (H2O)
s3(s3)
Products ID/Name Mapping
Species id(name) [SABIO]
Species id(name) [CellDesigner]
SPC_35 (ADP)
s2(s2)
SPC_36 (Phosphate)
s4(s4)
Modifiers ID/Name Mapping
Species id(name) [SABIO]
Species id(name) [CellDesigner]
ENZ_26909 (Nucleoside-triphosphatase(Enzyme) wildtype)
s6(2.7.1.2)
Apply
Cancel
Import KineticLaw

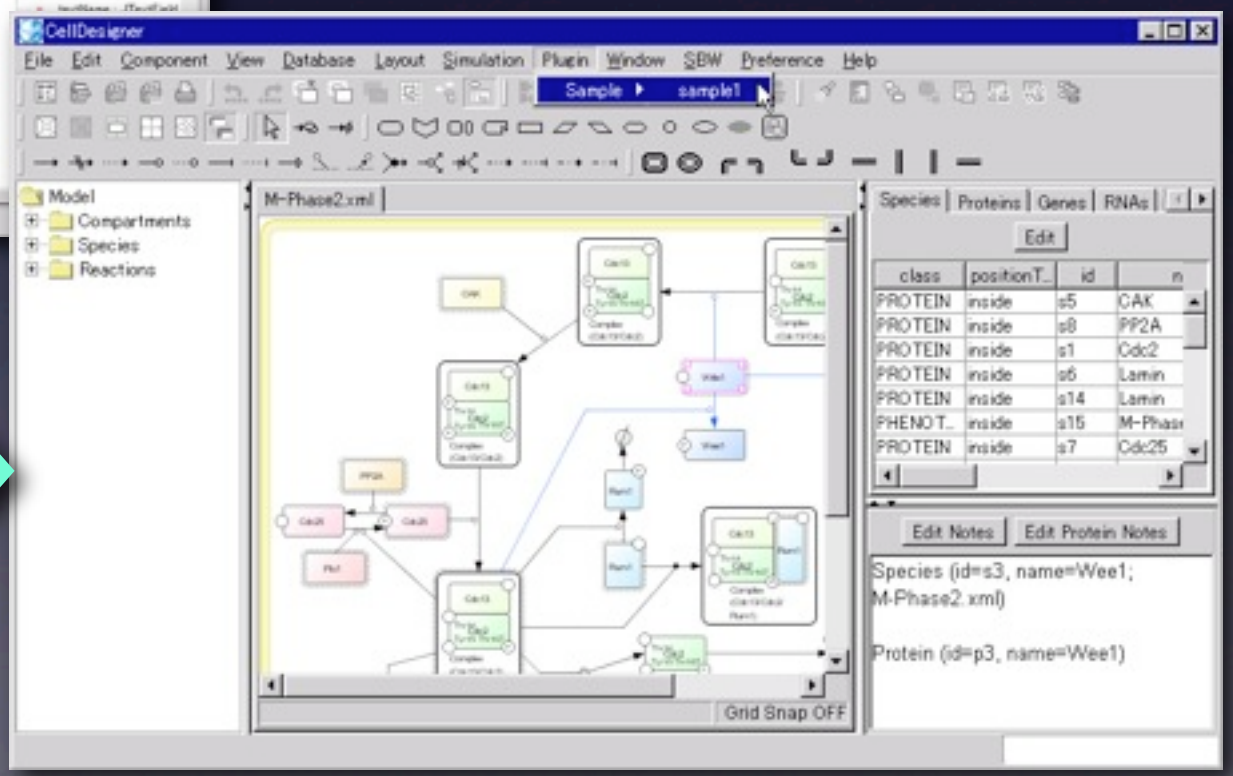
Example


KineticLaw
math
KL_5128(s6, kcatKm_SPC_34, kcat, s1, Km_SPC_34)
copy
+
-
*
/
(
)
Math
Name
SelectedReaction
s1
s2
s3
s4
Predefined Functions
NonPredefinedFunction
Mass_Action_Kinetics
Irreversible_Simple_Michaelis-Menten
Species
Parameters
Rules
New
Edit
Remove
Clear All
scope
id
name
value
units
constant
local:Reaction(re1)
kcatKm_SPC_34
kcatKm_ATP
1000.0
Mwedgeoneswedgeone
true
local:Reaction(re1)
kcat
kcat
0.186
swedgeone
true
local:Reaction(re1)
Km_SPC_34
Km_ATP
1.9E-4
M
true
Update
Close

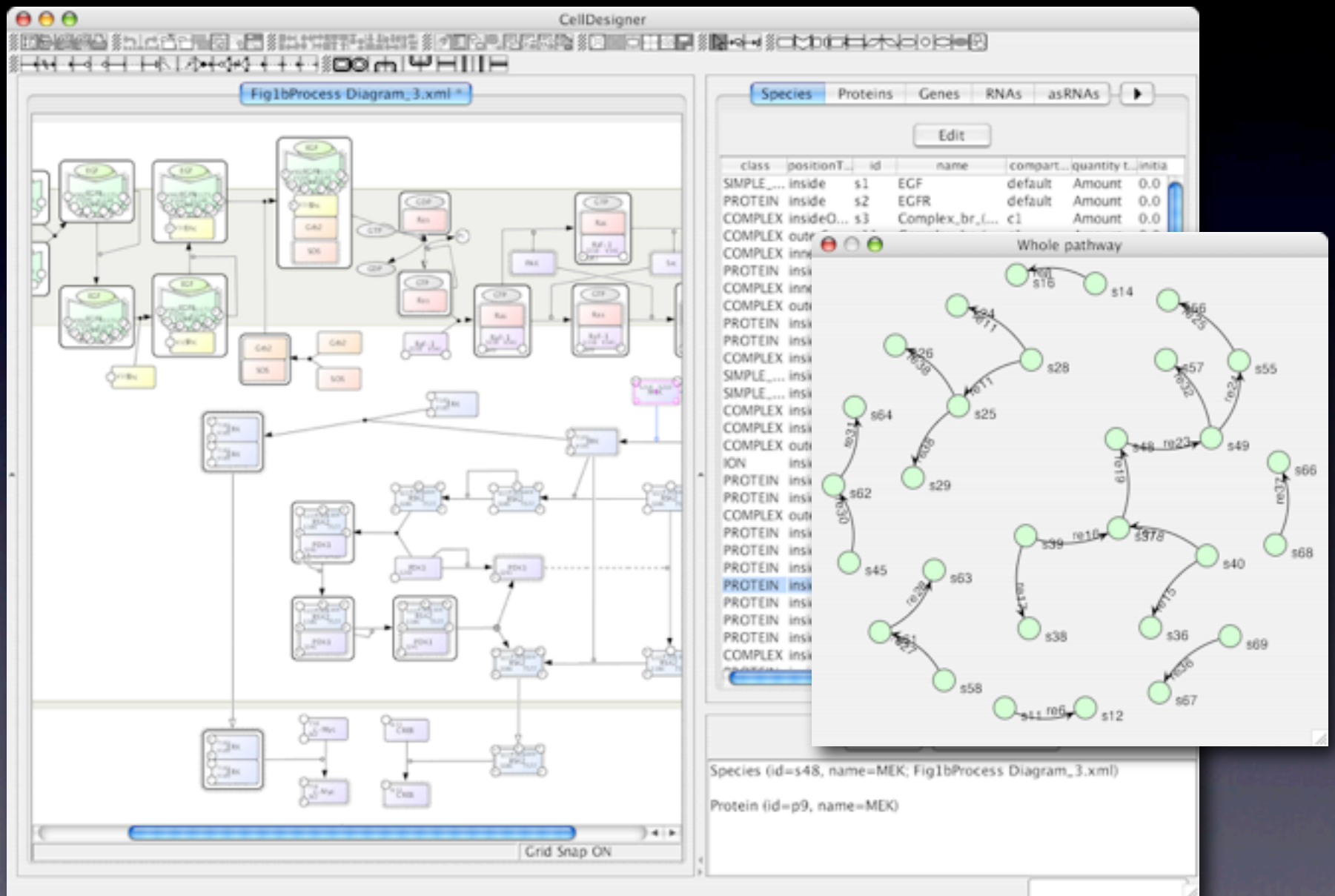
Plugin development



- Develop plugin on Eclipse
- Call plugin from [Plugin] menu on CellDesigner



Example Plugin



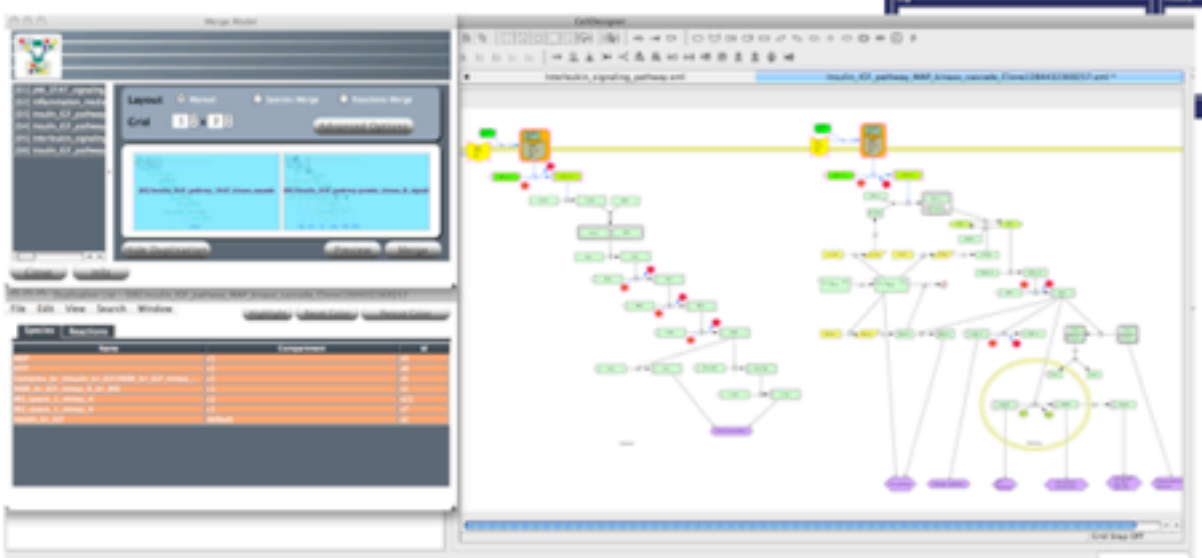
Example Plugin

- Merge Models & Payao Uploader plugin by Samik Ghosh

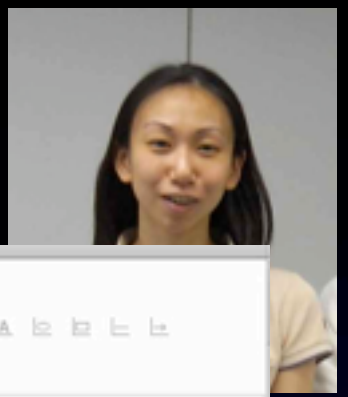
New CellDesigner Plugins

http://celldesigner.org/plugins/

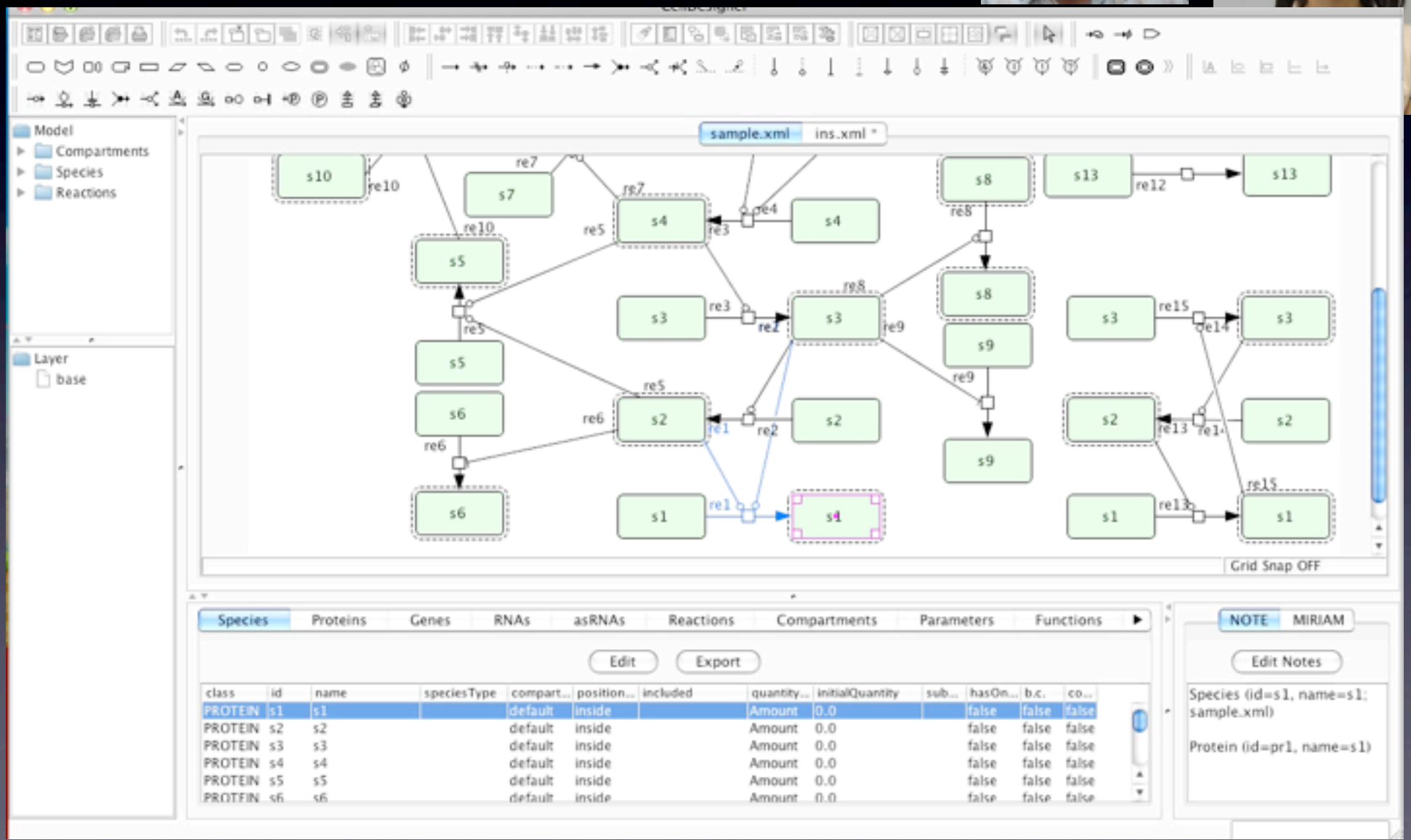
- Merge Models:
- Payao Uploader:



Example Plug



- Pathway Classification plugin by Hiromu Takizawa and Noriko Hiroi



Motivation

- Provide a software tool which uses:
  - SBML as a native file format
  - solidly defined graphical notation to represent biochemical networks
- Provide a software tool which can
  - integrate with existing resources / software tools
  - support many researchers as possible

Nature Biotechnology 24 (2) 2006

Examples

Depicting signaling cascades

To the editor:

In a paper in the August issue (*Nat. Biotechnol.* **23**, 961–966, 2005), Kitano *et al.* discuss the use of process diagrams to map signal-transduction cascades. They have used the formalism of process diagrams to specify pathway maps that are both readable and precise, and they have developed a map depicting hundreds of species and reactions involved in signaling by the epidermal growth factor receptor (EGFR)[1]. However, this map, as expansive as it is, omits the vast majority of species and reactions that could potentially be generated during signaling. We submit that comprehensive process diagrams for this, or any other signaling system, are very likely to be of unmanageable size.

The reason is combinatorial complexity, a hallmark of signal-transduction cascades[2–5]. Although Kitano *et al.* discuss this problem in their paper and suggest some solutions (e.g., modules for concise representation of subnetworks of a signaling system), we feel their solutions are inadequate in that explicit representation of all species at some level is still required.

Here, we wish to call attention to an alternative method of representation that we believe better addresses the problem of combinatorial complexity. This method involves the use of graphical reaction rules to represent the protein-protein interactions in a system and their consequences[6,7]. A rule illustrates features of species relevant for a particular type of reaction that can result from a protein-protein interaction, whereas a process diagram illustrates individual species and reactions.

Before discussing rules further, we should clarify the limitations of process diagrams. Let us consider the map of Figure 3e in the original Kitano *et al.* paper, which depicts 18 species and 32 reactions involved in EGFR signaling. These species and reactions correspond, more or less, to those included in the mathematical model of Kholodenko *et al.*[8], and they arise from interactions among five proteins: EGFR, its ligand epidermal growth factor (EGF), the adapters Grb2 and Shc, and the guanine nucleotide exchange factor Sos. The map, as we will elaborate shortly, presents an arguably oversimplified picture of signaling events. However, it is already challenging to decipher because a fairly large number of pictograms and intersecting arrows are needed to illustrate the various species and reactions. How complicated would the map be if it presented a more comprehensive picture of signaling?

Interactions of the proteins considered in Figure 3e of Kitano *et al.* can potentially generate not tens of species but hundreds to thousands of species, and even more reactions[4,9–11]. A focus on the 18 species of the map is appropriate only if several limiting assumptions hold true. These assumptions, upon which the model of Kholodenko *et al.*[8] (and derivative models such as that of Schoeberl *et al.*[12]) are based, include the following: first, simultaneous phosphorylation of tyrosines of both receptors in a ligand-induced receptor dimer; second, association of at most one adapter with a given receptor dimer at a time; and third, no dissociation of receptor dimers if receptors are phosphorylated.

In recent work[11], we discuss the validity of these assumptions and consider the impact of relaxing them. The result is an extended model for Sos activation that predicts the dynamics of a network of 356 species and 3,749 unidirectional reactions, all of which arise from protein-protein interactions underlying the map of Kitano *et al.*

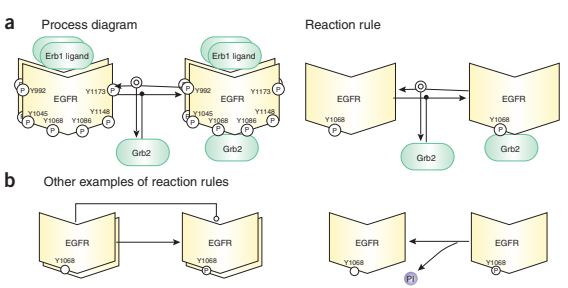


Figure 1 A process diagram and three graphical reaction rules drawn using CellDesigner[16]. (**a**) The process diagram illustrates Grb2 binding to a particular EGFR-containing species: three species and two unidirectional reactions are depicted. The adjacent reaction rule, drawn in a style consistent with the diagrammatic conventions of Kitano *et al.*, also pertains to Grb2 interaction with EGFR. It is one of the rules used to generate our model for EGFR signaling[11], and it indicates that Grb2-EGFR association via Y1068 in EGFR depends only on phosphorylation of this residue. By convention, it is assumed that the interaction represented in a rule is independent of all features not explicitly indicated. Thus, multiple species may qualify as reactants in a type of reaction defined by a rule. The exact number of reactions generated by the rule depends on the graph grammar of which the rule is a part (that is, the rule set and seed species that generate a model)[7]. Within the scope of our model[11], the rule shown here generates 312 distinct unidirectional reactions. (**b**) These reaction rules, which are also included in the rule set used to generate our model for EGFR signaling[11], represent transphosphorylation of one EGFR in a receptor dimer by the neighboring receptor and receptor dephosphorylation, which is catalyzed by phosphatases assumed to be present in excess. The left rule indicates that EGFR-catalyzed phosphorylation of Y1068 depends on dimerization of EGFR. In contrast, the right rule indicates that receptor dephosphorylation is spontaneous and independent of the state of EGFR aggregation. These rules generate 144 and 156 reactions, respectively, in our model for EGFR signaling[11].

We have found that consideration of this additional complexity is necessary if the model is to make accurate predictions about network dynamics and the role of specific components, such as individual sites of tyrosine phosphorylation[11,13].

Drawing a process diagram with 356 species to represent the interactions of only five proteins[11] would be inefficient and difficult to accomplish or read. Moreover, there are no obvious modules that could be introduced to simplify the process diagram, because the reaction network is highly branched[11]. In any case, a module has the drawback that protein-protein interactions are either altogether hidden (when the module is closed) or obscured by the possibly large number of species and reactions that can arise from the interactions (when the module is open).

Given that protein-protein interactions can generate myriad species and reactions for combinatorial reasons, what can be done to capture the essence of these interactions without ignoring their combinatorial complexity? To address this problem, we have proposed that protein-protein interactions and their effects be represented in the form of reaction rules that are generators of species and reactions[14,15]. More recently, we have introduced graphical reaction rules[6,7], in which graphs similar to the pictograms of process diagrams are used to represent features of proteins and protein complexes. Graphical rules were introduced to allow the connectivity of proteins in a complex to be explicitly represented, and they also provide a means to comprehensibly visualize protein–protein interactions, as illustrated in **Figure 1**.

In summary, process diagrams are useful for representing the individual species and reactions that can arise in a signaling system. However, representation at this microscopic level of detail may not be practical. In the face of combinatorial complexity, diagrams can be overly complicated or hide information about protein-protein interactions. An alternative approach is to represent not the species and reactions resulting from the interactions of proteins in a system but rather the interactions themselves. This task can be accomplished relatively easily using graphical reaction rules. A set of rules can be interpreted to obtain a mathematical model that accounts comprehensively for the species and reactions logically consistent with the rules, even when large numbers of species and reactions are possible[7,14,15]. We are currently extending the BioNetGen software package[14,15] to provide tools for drawing and interpreting graphical reaction rules (http://cellsignaling.lanl.gov/). In the future, we believe such model-generation tools will play an important role in obtaining a mechanistic understanding of cellular information processing and in manipulating signaling systems for therapeutic and biotechnological purposes.

Michael L. Blinov, Jin Yang, James R. Faeder & William S. Hlavacek

Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: wish@lanl.gov

1. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. *MMol. Syst. Biol.* published online 25 May 2005; doi:10.1038/msb4100014
2. Endy, D. & Brent, R. *Nature* **409**, 391–395 (2001).
3. Bray, D. *Science* **299**, 1189–1190 (2003).
4. Hlavacek, W.S., Faeder, J.R., Blinov, M.L., Perelson, A.S. & Goldstein, B. *Biotechnol. Bioeng.* **84**, 783–794 (2003).
5. Goldstein, B., Faeder, J.R. & Hlavacek, W.S. *Nat. Rev. Immunol.* **4**, 445–456 (2004).
6. Faeder, J.R., Blinov, M.L. & Hlavacek, W.S. In *Proc. 2005 ACM Symp. Appl. Computing* (Santa Fe, NM, March 13-17, 2005) Liebrock, L.M., Editor. ACM Press, New York, NY, pp. 133-140. doi:10.1145/1066677.1066712
7. Blinov, M.L., Yang, J., Faeder, J.R. & Hlavacek, W.S. *In Proc. BioCONCUR 2005* (Mishra, B. & Priami, C., eds) 1–19 (San Francisco, CA, August 27, 2005).
8. Kholodenko, B.N., Demin, O.V., Moehren, G. & Hoek, J.B. *J. Biol. Chem.* **274**, 30169–30181 (1999)
9. Conzelmann, H. *et al. Syst. Biol.* **1**, 159–169 (2004).
10. Borisov, N.M., Markevich, N.I., Hoek, J.B. & Kholodenko, B.N. *Biophys. J.* **89**, 951–966 (2005).
11. Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. *BioSystems* doi:10.1016/j.biosystems.2005.06.014, **in press** (2005).
12. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. & Müller, G. *Nat. Biotechnol.* **20**, 370–375 (2002).
13. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Syst. Biol.* **2**, 5–15 (2005).
14 Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. *Bioinformatics* **20**, 3289–3291 (2004).
15. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Complexity* **10**, 22–41 (2005)
16. Funahashi, A., Tanimura, N., Morohashi, M. & Kitano, H. *BIOSILICO* **1**, 159–162 (2003).

Kitano *et al.* respond:

The first issue raised by Blinov *et al.* suggests that pathway maps are too simplistic to represent the protein combinatorial explosion in signal cascades. They detail Figure 3e in our article to illustrate their point; however, this figure was used solely to demonstrate the look-and-feel of how to represent pathways as process diagrams. Therefore, we used part of the diagram in a Hanahan and Weinberg paper[1], which is also a pathway extensively used in simulation studies[2,3]. It was not argued that this was a comprehensive representation of the EGFR pathway. Our recent interaction map published in *Molecular Systems Biology*[4] was intended to be a comprehensive EFGR map of experimentally validated interactions. We did not enumerate all possible interactions and molecular states and recognize that there are interactions not listed in the map due to lack of experimental validation, despite theoretical and intuitive possibilities. The process diagram is neutral on what should be described in the map. It defines the graphical representation of an interaction map; thus, the oversimplification critique does not apply to the process diagram itself as construction of these maps relies on experimental evidence.

The second issue raised was that describing all combinatorial states of molecules and resulting complexes would result in a combinatorial explosion making a rule-based approach more appropriate for modeling. We would argue that this depends on the intended use of the map. The process diagram was motivated by an experimentalist's need partly to represent detailed interactions, including residue modification state, to improve experimental design, and partly to visualize their data in the context of a pathway map where each combinatorial state has been explicitly described, regardless of the level of complexity. It is imperative that software tools make such complex and large-scale maps accessible to users.

Although the rule-based approach has attracted much attention as a viable approach for dynamical simulation[5,6], it may not allow users to project experimental data on to each combinatorial state without expansion. As illustrated by Blinov *et al.* wherever the rule-based approach is shown to be effective, the process diagram can then be used to expand graphical notation to represent rules and the network generated from the rule. We would like to incorporate such features into the process diagram and are receptive to constructive critiques to create standard graphical notations; to this end, we have formed an international alliance to standardize graphical notation called Systems Biology Graphical Notation (http://www.sbgn.org/).

1. Hanahan, D. & Weinberg, R.A. *Cell* **100**, 57–70 (2000).
2. Kholodenko, B.N., Demin, O.V., Moehren, G. & Hoek, J.B. *J. Biol. Chem.* **274**, 30169–30181 (1999).
3. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. & Muller, G. *Nat. Biotechnol.* **20**, 370–375 (2002).
4. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. *Mol. Systems Biol.* doi: 10.1038/msb4100014 (25 May 2005).
5. Lok, L. & Brent, R. *Nat. Biotechnol.* **23**, 131–136 (2005).
6. Borisov, N.M., Markevich, N.I., Hoek, J.B. & Kholodenko, B.N. *Biosystems* doi:10.1016/j.biosystems.2005.03.006 (17 October 2005).

2011年3月2日水曜日

Examples

Nature Molecular Systems Biology 4(173) 2008

Molecular Systems Biology 4; Article number 173; doi:10.1038/msb.2008.7
Citation: ***Molecular Systems Biology*** **4:173**

www.molecularsystemsbiology.com

molecular systems biology

REVIEW

A comprehensive modular map of molecular interactions in RB/E2F pathway

Laurence Calzone[1,2,3,6], Amélie Gelay[1,5,6], Andrei Zinovyev[1,2,3,6,*], François Radvanyi[4,5,7] and Emmanuel Barillot[1,2,3,7]

[1] Institut Curie, Service Bioinformatique, Paris,
[2] INSERM, U900, Paris, F-75248 France,
[3] Ecole des Mines de Paris, ParisTech, Fontainebleau, F-77300 France,
[4] Institut Curie, Centre de Recherche, Paris, F-75248 France and
[5] Institut Curie, CNRS UMR144, Paris, France
[6] These authors contributed equally to this work.
[7] These two authors are the joint senior authors of this paper.
* Corresponding author: Institut Curie, Service Bioinformatique, 26 rue d'Ulm, Paris 75248, France. Tel. +33 1 53 10 70 50; Fax: +33 1 44 41 09 08; E-mail: andrei.zinovyev@curie.fr

Received 17.7.07; accepted 23.1.08

We present, here, a detailed and curated map of molecular interactions taking place in the regulation of the cell cycle by the retinoblastoma protein (RB/RB1). Deregulations and/or mutations in this pathway are observed in most human cancers. The map was created using Systems Biology Graphical Notation language with the help of CellDesigner 3.5 software and converted into BioPAX 2.0 pathway description format. In the current state the map contains 78 proteins, 176 genes, 99 protein complexes, 208 distinct chemical species and 165 chemical reactions. Overall, the map recapitulates biological facts from approximately 350 publications annotated in the diagram. The network contains more details about RB/E2F interaction network than existing large-scale pathway databases. Structural analysis of the interaction network revealed a modular organization of the network, which was used to elaborate a more summarized, higher-level representation of RB/E2F network. The simplification of complex networks opens the road for creating realistic computational models of this regulatory pathway.
Molecular Systems Biology 4 March 2008; doi:10.1038/msb.2008.7
Subject Categories: metabolic and regulatory networks; cell cycle
Keywords: cell-cycle regulation; E2F; RB pathway; RB1; systems-biology standards

Introduction

The cell cycle is the succession of four phases called G1, S, G2 and M. In dividing cells, DNA replication (S phase) and mitosis (M phase) alternate (Alberts *et al*, 1994), and are separated by two gap phases, G1 and G2 phases. In quiescent cells, the cells are considered to be in G0 phase. When they receive external signals, such as growth factors, a series of activations push the cell from a G0 to a G1 state and enters the cell cycle. The whole process of cell division is mainly orchestrated by complexes composed of two subunits, a kinase and a cyclin partner. These complexes phosphorylate a certain number of proteins, either activating or inhibiting them. Among them, the retinoblastoma tumour suppressor protein RB (RB1) is a key regulator in cell-cycle entry (transition G1/S). It sequesters a family of transcription factors, the E2Fs, responsible for the transcription of many genes involved in cell-cycle regulation, DNA replication and other functions like the activation of the apoptotic pathway (Muller *et al*, 2001). RB functions as a brake in the cell cycle, which is released when external signals trigger S-phase entry. The main targets of the external signals are the G1 cyclin/CDK complexes. Once active, the complexes, among them CycD1/CDK4,6, act as starters of the cell cycle (Novak *et al*, 2007) and phosphorylate RB, which then releases E2F (DeGregori, 2004).

RB is a member of a family of proteins called the pocket proteins (Knudsen and Wang, 1997). These proteins RB, p107 and p130, share sequence similarities, especially in the 'pocket domain' (Stevaux and Dyson, 2002), which is responsible for their repressor function. RB protein contains domains where the binding sites for co-repressors (E2F proteins and viral oncoproteins) are situated. These sites are subjected to most mutations.

RB is a tumour suppressor gene. Because of its implication in so many, if not all, cancers (Sherr and McCormick, 2002), the study of RB regulation requires a special attention.

More specifically, the RB/E2F pathway is commonly deregulated in cancer through genetic or epigenetic mechanisms, resulting in E2F activation. Several common oncogenes (involved in many cancer types) are the activators of the pathway, whereas several common tumour suppressor genes are inhibitors of the pathway. For example, cyclin D1 (CCND1), E2F3 and the two cyclin-dependent kinases CDK4 and CDK6 can be activated by translocation, amplification or mutation, whereas RB (RB1) and the cyclin-dependent kinase inhibitors p16INK4a (CNKN2A) and p15INK4b (CDKN2B) can be inactivated by point mutation, homozygous deletion or DNA methylation. In addition, RB can be inactivated by several oncogenic viral proteins including E7 from human papillomavirus, which is responsible for more than 90% of cervical carcinomas (Munger *et al*, 2001). Tumour suppressor gene inactivation is found not only in sporadic tumours but also in tumour-prone families. Germline mutations of RB1 results in retinoblastoma with a high penetrance early in young individuals and late in life in sarcomas and lung and bladder carcinomas (Knudson, 1971; Nevins, 2001; Giacinti and Giordano, 2006). Germinal mutations of p16INK4a results in

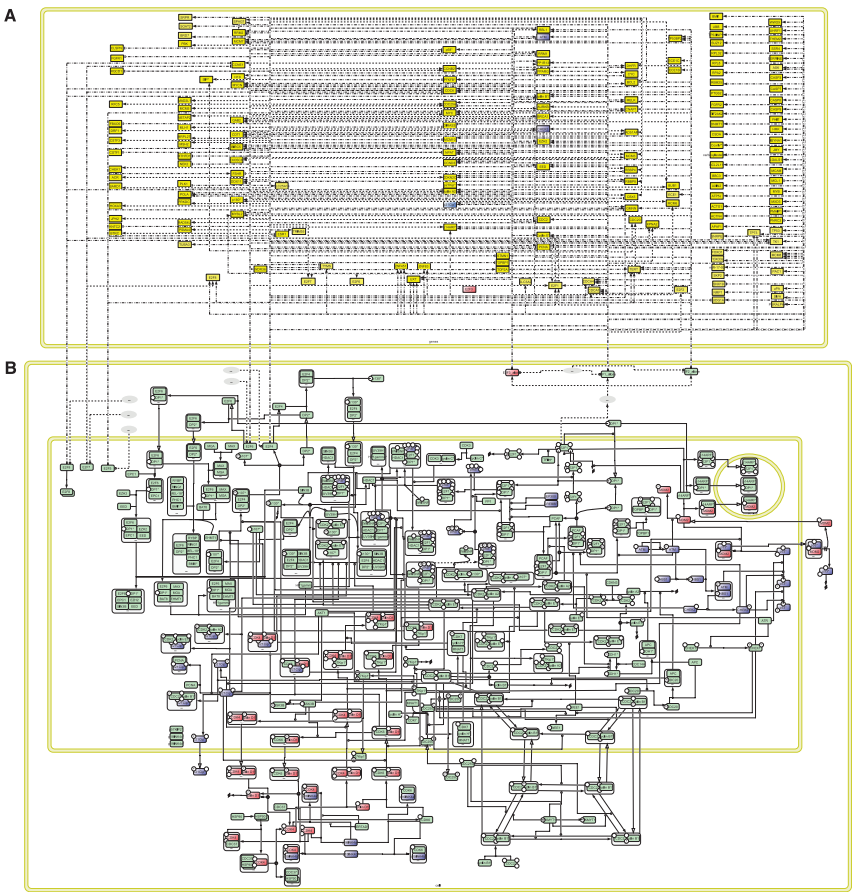


Figure 2 The textbook pathway of RB has been expanded by integrating data from the literature. The E2F transcription factors (represented here by single proteins in the nuclear compartment) are connected by activation and inhibition arrows to their gene targets. (**A**) Map of target genes of E2F transcription factors. Each E2F associates with different cofactors to activate or inhibit the transcription of many genes; pointed arrows mean activation and flat arrows mean inhibitions (**B**) Map of protein–protein interaction network. Each icon on the diagram represents distinct chemical species. See Kitano and co-workers' description of CellDesigner's standard notation (Kitano *et al*, 2005) for a detailed meaning of shapes. When the information is available (from Atlas Oncology web-page: www.atlasgeneticsoncology.org/), tumour suppressor genes and the corresponding proteins are coloured in blue and oncogenes in red, the other proteins are in green. To read and navigate through the map, visit our webpage: http://bioinfo-out.curie.fr/projects/rbpathway/. The map is clickable and allows easy access to all included information (such as literature references or standard protein ids) and hyperlinked to other databases.

are connected by 'activation' and 'inhibition' relations. The information about these relations is derived from the detailed diagram. For example, in the detailed map, E2F1 is phosphorylated by CycA2/CDK2 and is subsequently recognized for degradation, which is translated in the modular map by CycA2/CDK2 module inhibiting E2F1-3 module.

2011年3月2日水曜日

Examples

Nature Cell Biology 6(11), 2006

Physicochemical modelling of cell signalling pathways

Bree B. Aldridge, John M. Burke, Douglas A. Lauffenburger and Peter K. Sorger

Physicochemical modelling of signal transduction links fundamental chemical and physical principles, prior knowledge about regulatory pathways, and experimental data of various types to create powerful tools for formalizing and extending traditional molecular and cellular biology.

This review is aimed at biologists interested in mathematical modelling of biochemical pathways, but who are relatively unfamiliar with the topic. Our discussion focuses on pathways involving 'signals' rather than metabolites. In this context, physicochemical modelling is a natural extension of informal or conceptual pathway modelling. Formal modelling is much more powerful in putting molecular detail in a physiological context, uncovering principles of biological design and creating dynamic repositories of interpretable knowledge. However, to realize this power, challenges inherent in construction, verification, calibration, interpretation and publication of models must be addressed.

MATHEMATICAL MODELS IN MOLECULAR, CELLULAR AND DEVELOPMENTAL BIOLOGY

Contemporary molecular, cellular and developmental biology seeks to describe physiological processes in terms of gene functions and specific molecular mechanism. Medicine and drug discovery add the practical goals of understanding disease and developing treatments. The 'component identification' phase of modern biology is approaching completion, and the sheer size of the cellular 'parts list' highlights the importance of understanding function, not at the level of single genes, but rather at a higher level of abstraction, involving pathways and circuits. In many cases, conceptual modelling of biology is at the breaking point[1] — it is impossible mentally to juggle large pathways involving many components. The missing ingredient is mathematics. Used appropriately, mathematical models can represent pathways in a physically and biologically realistic manner, incorporate a wide variety of empirical observations, and generate novel and useful hypotheses. Pathway modelling has existed for some time, particularly in the field of prokaryotic metabolism[2,3], but it remains at an early stage of development. It is challenging to construct accurate models and establish rigorous links to experimental data (see accompanying article by Jaqaman *et al.* in *Nature Rev. Mol. Cell Biol.*). This commentary is based on the premise that useful models of critical mammalian pathways can nonetheless be constructed using an iterative modify–measure–mine–model procedure that closely integrates experiment and mathematics (Fig. 1).

Bree B. Aldridge, John M. Burke, Douglas A. Lauffenburger and Peter K. Sorger are in the Center for Cell Decision Processes, Department Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Bree B. Aldridge and Peter K. Sorger are in the Department of Systems Biology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115, USA.
e-mail: psorger@mit.edu

APPROACHES TO PHYSICOCHEMICAL MODELLING

Physicochemical modelling seeks to describe biomolecular transformations (such as covalent modification, intermolecular association and intracellular localization) in terms of equations derived from established physical and chemical theory[4–8]. These 'kinetic' or 'reaction' models use prior knowledge to make specific molecular predictions and work best with pathways in which components and connectivity are relatively well established. When prior knowledge is sparse, data-driven statistical models are more appropriate (see accompanying article by Janes *et al.* in *Nature Rev. Mol. Cell Biol.*). Equations in physicochemical models refer to identifiable processes (such as catalysis and assembly) and parameters have physical interpretation (such as concentration, binding affinity, and reaction rate). The models can be viewed as translations of familiar pathway maps into mathematical form — a process that should become easier and more transparent with the adoption of common schematic standards[9].

The correct mathematical form for a physicochemical model depends on the properties of the system being studied and the goals of the modelling effort. Ordinary and partial differential equations (ODEs and PDEs) are most commonly and both can be cast in either deterministic or stochastic form. Stochastic equations include effects arising from random fluctuation around the average behaviour. Currently, the most common means of representing biochemical pathways is through a set of coupled ODEs (an ODE network). ODE networks represent the rates of production and consumption of individual biomolecular species, $d[X_i]/dt$, in terms of mass action kinetics — an empirical law stating that rates of a reaction are proportional to the concentrations of the reacting species. Each biochemical transformation is therefore represented by an elementary reaction with forward and reverse rate constants. Changes in localization, a central feature of biological pathways, are represented by compartmentalization. Each species is allowed to inhabit one or more compartments and to move among the compartments through elementary reactions. Compartments are also used to represent assembly of macromolecular complexes and other non-enzymatic changes of state. Two fundamental assumptions of the compartmentalized ODE

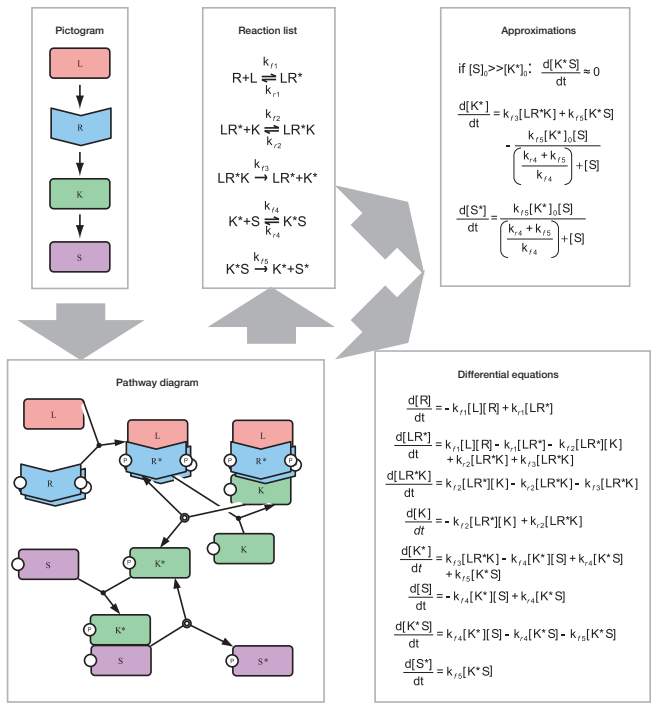


Figure 3 Steps in physicochemical modelling. A pathway map is a highly abstracted pictogram of biomolecules and their interactions. Here, a simple linear ligand–receptor–kinase–substrate pathway is depicted. Although the pictogram conveys the general information flow in the network, mechanistic details required for mathematical modelling are absent. A formal pathway diagram drawn with CellDesigner details the reaction network[40]. Instead of representing the kinase as one object (as in the pictogram), each form of the kinase, either in complex or alone, is depicted (K, K*, LR*K, and K*S). A key challenge in developing a pathway diagram is making choices about granularity in number of species and reactions (see text). In this example, the receptor is a dimer and each subunit has two phosphorylation sites, yielding 64 possible ligand–receptor dimer complexes. However, this complexity is represented simply by two species: non-active and unphosphorylated (R) and ligand-bound, fully phosphorylated (LR*). It should be noted that approaches such as rules-based modelling may be preferred to the use of pathway diagrams (see text for details). A complete list of reactions is generated from the pathway diagram. This list can be automatically produced with several specialized software tools (Box 1). For reversible reactions, both forward and backward rate constants must be indicated. From the list of reactions, a system of differential equations is enumerated using appropriate rate laws, such as mass action kinetics, which uses the product of a rate constant and the concentrations of the reactants to calculate the reaction rates. Simplifying assumptions can be made to reduce the complexity or size of a model. The Michaelis-Menten approximation to enzyme–substrate kinetics is often applied. This particular rate form assumes rapid equilibrium of an intermediate complex (K*S), so that an equilibrium assumption is imposed ($d[K^*S]/dt = 0$), thus reducing the number of species in the model. Because this is an approximation, its use can alter model behaviour, particularly when the intermediate complex does not reach equilibrium or the reaction is tightly coupled to other processes[25–27].

can be introduced as simplified 'lumped' rates. At the same time, metabolic and synthetic processes are themselves being subjected to quantitative modelling. Thus, hybrid models can be constructed in which specific biological processes are alternately modelled in detail or in aggregate. For example, a highly simplified 'lumped rate' representation of a detailed metabolic model could be embedded in a physicochemical model of signal transduction to yield a hybrid. Realistic regulation could be reproduced by adding an adjustable parameter to the grouped metabolic model that makes metabolism dependent on signalling.

The issue of model granularity also arises with equations representing elementary reactions. For example, when a reaction is a hundred times or more faster than other reactions, it can be assumed that the fast process

2011年3月2日水曜日

Examples

Nucleic Acids Research 34, 2006

Nucleic Acids Research, 2006, Vol. 34, Web Server issue W645–W650
doi:10.1093/nar/gkl229

Applications for protein sequence–function evolution data: mRNA/protein expression analysis and coding SNP scoring tools

Paul D. Thomas*, Anish Kejariwal, Nan Guo, Huaiyu Mi, Michael J. Campbell, Anushya Muruganujan and Betty Lazareva-Ulitsky

Evolutionary Systems Biology Group, SRI International, 333 Ravenswood Ave., Menlo Park CA 94025, USA

Received February 14, 2006; Revised March 6, 2006; Accepted March 27, 2006

ABSTRACT

The vast amount of protein sequence data now available, together with accumulating experimental knowledge of protein function, enables modeling of protein sequence and function evolution. The PANTHER database was designed to model evolutionary sequence–function relationships on a large scale. There are a number of applications for these data, and we have implemented web services that address three of them. The first is a protein classification service. Proteins can be classified, using only their amino acid sequences, to evolutionary groups at both the family and subfamily levels. Specific subfamilies, and often families, are further classified when possible according to their functions, including molecular function and the biological processes and pathways they participate in. The second application, then, is an expression data analysis service, where functional classification information can help find biological patterns in the data obtained from genome-wide experiments. The third application is a coding single-nucleotide polymorphism scoring service. In this case, information about evolutionarily related proteins is used to assess the likelihood of a deleterious effect on protein function arising from a single substitution at a specific amino acid position in the protein. All three web services are available at http://www.pantherdb.org/tools.

INTRODUCTION

The continued improvements in DNA sequencing technology are rapidly expanding our knowledge of the genomes and, by inference (through the genetic code and prediction of open reading frames), the proteomes of extant species. These DNA and protein sequences provide detailed information about molecular evolution. Combined with information about protein function derived from biochemical and genetic experiments, the molecular evolution data can shed light on the relationship between protein sequence and function. The PANTHER database (1,2) was designed to model the relationships between protein sequence and function for all major protein families, using molecular taxonomy tree building combined with human biological interpretation of the resulting trees. The trees are used to locate functional divergence events within protein families that define subfamilies of proteins of shared function.

The current version of PANTHER (6.0) contains trees for over 5000 protein families, divided into over 30 000 functional subfamilies. For each family and subfamily group, a multiple sequence alignment is constructed that aligns 'equivalent' positions (i.e. descended from the same ancestral codon) in each of the proteins in the group. Each multiple sequence alignment is then represented as a hidden Markov model (HMM) that summarizes, for each position, the probabilities of each of the 20 amino acids appearing (or of insertions and deletions) at that position in the given group of related sequences.

The resulting HMM parameters can be used in a number of scientific applications. We discuss two here. The first is classification of new sequences. The match between a sequence and an HMM is given a score by calculating the probability that the sequence was 'generated' by that HMM, and comparing it with the probability that the sequence was generated by a random HMM of the same length (3). For a new sequence, this HMM 'score' can be calculated for each of the family and subfamily HMMs, and the sequence is classified as belonging to same group as the best-scoring HMM (provided that the score is also statistically significant). In PANTHER, because each HMM is classified by the functions of its constituent proteins, protein sequences can be assigned to functional

*To whom correspondence should be addressed. Tel: +1 650 859 2434; Fax: +1 650 859 3735; Email: paul.thomas@sri.com

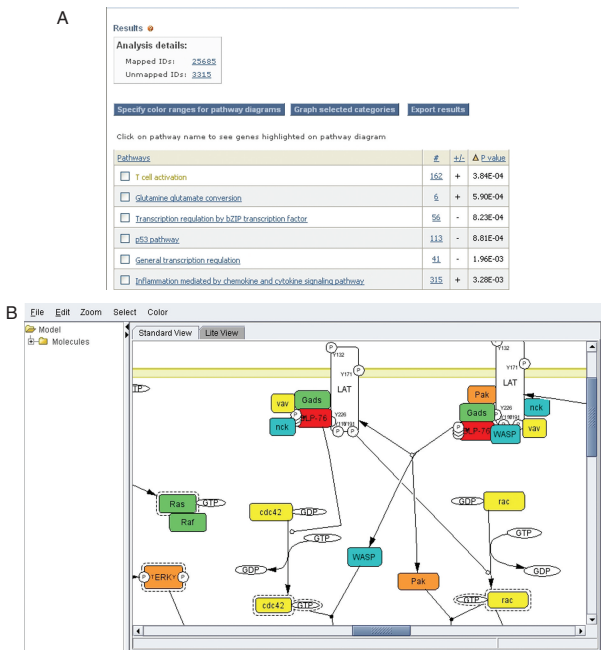


Figure 3. Expression data analysis and visualization on the PANTHER website. (**A**) Mann–Whitney *U*-test results, and (**B**) CellDesigner (15) diagram of the T-cell activation signaling pathway from the PANTHER Pathway database (accession P00053, author Adam Douglass). This applet colors proteins according to a 'heat map' calculated from user-input values. Protein components are mapped to PANTHER HMMs. Active forms (dashed-line boxes) and phosphorylated forms (small circles around the letter 'P') of proteins are clearly indicated in the diagram. A total of 107 pathways (mostly signaling pathways) are currently available.

the uploaded file, depending on the desired source of the PANTHER classification data: either the pre-calculated classifications available on the PANTHER site, or a user-generated file. For using the pre-calculated PANTHER data, the file must contain two columns: the first is the gene or protein identifier, and the second is the numerical value. For user-specified data, the file must contain three columns: an arbitrary tracking identifier (e.g. a UniProt identifier or gene symbol); the PANTHER HMM identifier indicating the classification of the gene/protein; and the numerical value.

The output of the tool is a list of *P*-values for each comparison between a functional category distribution and the reference distribution. Each distribution, and how it compares with the reference distribution, can be viewed graphically from the output page. We find that this is critical for interpreting the any deviation between the functional category distribution and the overall distribution. The genes/proteins in each category can also be viewed from the output page by clicking on the listed counts. In addition, for pathways, clicking on the pathway name will bring up an interactive Java applet that colors the pathway using a 'heat map' derived from the input values (Figure 3).

Coding SNP scoring service

The non-synonymous SNP scoring service is available at http://www.pantherdb.org/tools/csnpScoreForm.jsp. The methodology used to generate the scores is described in detail in (1) and summarized in (14). Briefly, the method uses a multiple alignment of a family of protein sequences, together with information about functional subfamilies within that family, to estimate the probabilities of different amino acids occurring at different positions in the protein family. High probability amino acids are likely to result in a functional protein, while low probability amino acids are likely to have a deleterious effect on protein function. We quantify the likely functional effect with a substitution position-specific evolutionary conservation (subPSEC) score, calculated as simply the log of the ratio of the probabilities of the two substituted amino acids: $\ln(P_{sub}/P_{wt})$, where P_{sub} is the probability of the substituted amino acid and P_{wt} is the probability of the wild-type amino acid. Smaller (more negative) subPSEC scores indicate a higher likelihood of being deleterious. We have recently added a third parameter to the subPSEC score: the number of independent counts n_{ic}, a measure of

Examples

EMBO J. 26, 2007

The EMBO Journal (2007) 26, 1055–1067 | © 2007 European Molecular Biology Organization | All Rights Reserved 0261-4189/07
www.embojournal.org

THE EMBO JOURNAL

c-Myc primed mitochondria determine cellular sensitivity to TRAIL-induced apoptosis

Anni I Nieminen, Johanna I Partanen, Annika Hau and Juha Klefstrom*

Cancer Cell Circuitry Laboratory, Institute of Biomedicine/Biochemistry and Molecular Cancer Biology Program, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland

Oncogenic c-Myc renders cells sensitive to TRAIL-induced apoptosis, and existing data suggest that c-Myc sensitizes cells to apoptosis by promoting activation of the mitochondrial apoptosis pathway. However, the molecular mechanisms linking the mitochondrial effects of c-Myc to the c-Myc-dependent sensitization to TRAIL have remained unresolved. Here, we show that TRAIL induces a weak activation of procaspase-8 but fails to activate mitochondrial proapoptotic effectors Bax and Bak, cytochrome *c* release or downstream effector caspase-3 in non-transformed human fibroblasts or mammary epithelial cells. Our data is consistent with the model that activation of oncogenic c-Myc primes mitochondria through a mechanism involving activation of Bak and this priming enables weak TRAIL-induced caspase-8 signals to activate Bax. This results in cytochrome *c* release, activation of downstream caspases and postmitochondrial death-inducing signaling complex -independent augmentation of caspase-8-Bid activity. In conclusion, c-Myc-dependent priming of the mitochondrial pathway is critical for the capacity of TRAIL-induced caspase-8 signals to activate effector caspases and for the establishment of lethal caspase feedback amplification loop in human cells.

The EMBO Journal (2007) **26,** 1055–1067. doi:10.1038/sj.emboj.7601551; Published online 1 February 2007
Subject Categories: signal transduction; differentiation & death
Keywords: apoptosis; Bcl-2 family; c-Myc; TRAIL

Introduction

The TNF-family death ligands TNF-α, CD95L/FasL and TRAIL promote apoptosis in many types of tumor and virus-infected cells. Remarkably, primary or non-transformed cells are often resistant to the death receptor-induced apoptosis, yet they express functional receptors for these ligands (LeBlanc and Ashkenazi, 2003; Fesik, 2005). The death receptor ligands can promote regression of tumors *in vivo*, which in part is attributable to the selective tumor cell killing by these agents. However, only TRAIL induces tumor regression without exhibiting significant systemic toxicity *in vivo*, and currently both TRAIL and agonistic antibodies to TRAIL receptors are in clinical trials for the treatment of cancer (Fesik, 2005). Despite years of intense study, only little is known about mechanisms that render tumor cells sensitive to TRAIL.

TRAIL kills cancer cells by binding specific cell surface death receptors, which are TNFRSF10A (DR4) and TNFRSF10B (DR5). After binding TRAIL, these transmembrane receptors form trimeric complexes and assemble a death-inducing signaling complex (DISC) in their cytosolic parts (LeBlanc and Ashkenazi, 2003). DISC is a primary signaling complex in which an adaptor protein FADD mediates recruitment of cysteine protease procaspase-8 to the receptors. FADD interacts with procaspase-8 through homotypic death effector domain interactions and promotes oligomerization-dependent auto-activation of this caspase. The active caspase-8 can initiate an apoptotic caspase cascade. In addition, TRAIL may also induce activation of other signaling pathways. For example, JNK, p38 MAPK and IKK/NF-κB kinase pathways are activated downstream of DISC assembly and caspase-8 activation (Varfolomeev *et al*, 2005). The kinase pathways may control apoptosis but also mediate non-apoptotic, for example proliferative or inflammatory, effects of death ligands (Algeciras-Schimnich *et al*, 2002; LeBlanc and Ashkenazi, 2003; Wajant *et al*, 2003; Secchiero *et al*, 2005). The DISC-activated caspase-8 connects to the downstream apoptotic death machinery in two ways. In certain cells (Type I), DISC assembly generates large amounts of active caspase-8, which is sufficient to directly cleave and activate downstream effector caspases, such as caspases-3, -6 and -7, that execute the apoptotic death program. In other cell types (Type II), death receptor-induced apoptosis requires engagement of the mitochondria-mediated pathway into the process of cell death (Scaffidi *et al*, 1998, 1999; Fulda *et al*, 2002; Rudner *et al*, 2005). The active caspase-8 can cleave proapoptotic Bcl-2 family protein Bid into an active form called truncated Bid (tBid). The tBid, in turn, recruits the mitochondrial pathway by activating the distal proapoptotic Bcl-2 family proteins Bax and Bak at the mitochondrial membranes, which leads to the release of apoptosis promoting factors such as holocytochrome *c* (cyt *c*), Smac/DIABLO and Omi/HtrA2 from mitochondria to the cytosol (Lowe *et al*, 2004). Once released into the cytosol, cyt *c* activates via APAF-1/caspase-9 complex effector caspases that execute apoptosis. TRAIL-induced apoptosis is often crucially dependent on the intact mitochondrial pathway (Deng *et al*, 2002; LeBlanc *et al*, 2002).

Activation of c-Myc renders primary and non-transformed cells sensitive to TNF-α, CD95L and TRAIL-induced apoptosis (Hueber *et al*, 1997; Klefstrom *et al*, 1997; Ricci *et al*, 2004; Wang *et al*, 2004). The molecular mechanisms underlying this apoptotic sensitization are not well understood, but they may involve an inhibitory action of c-Myc towards TNF-induced NF-κB activation, which normally counteracts the apoptotic action of TNF (Klefstrom *et al*, 1997; You *et al*, which specifically occurred in the cells with c-Myc is due to formation of high-order Bak complexes (Mikhailov *et al*, 2003; Ruffolo and Shore, 2003). Therefore, the weak c-Myc-induced immunostaining may indicate formation of preactive Bak mono- or oligomers (Zhang *et al*, 2004). It is notable, that previous studies have already implicated a role for c-Myc in

*Corresponding author. Cancer Cell Circuitry Laboratory, Institute of Biomedicine/Biochemistry and Molecular Cancer Biology Program, Biomedicum Helsinki, University of Helsinki, Haartmaninkatu 8, Helsinki 00014, Finland. Tel.: +358 9 1912 5493; Fax: +358 9 1912 5444; E-mail: juha.klefstrom@helsinki.fi

Received: 29 June 2006; accepted: 18 December 2006; published online: 1 February 2007

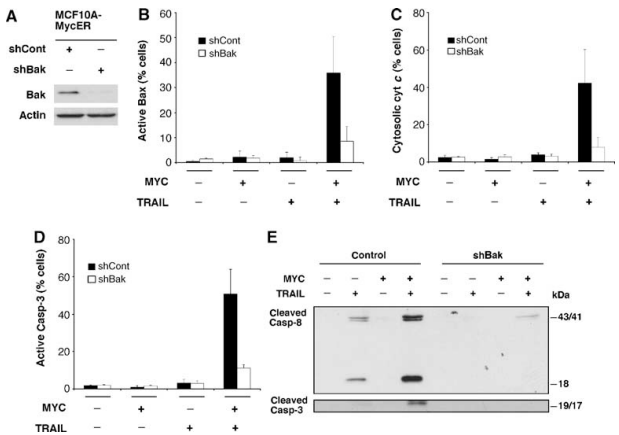


Figure 8 Bak is required for apoptotic events induced by c-Myc and TRAIL. (**A**) Western immunoblot demonstrating lentiviral shRNA silencing of endogenous Bak in MCF10A-MycER^{tm} cells. Lysates were made after hygromycin selection. (**B–D**) Bax activation, cyt *c* release and capase-3 activation were quantitated as described in Figure 1. The graph values in (B, C) represent mean ± s.d. of three independent experiments and of two experiments in (D). (**E**) The immunoblot shows that c-Myc fails to augment TRAIL-induced processing of procaspase-8 in Bak-deficient cells. The analyses were performed as in Figure 4.

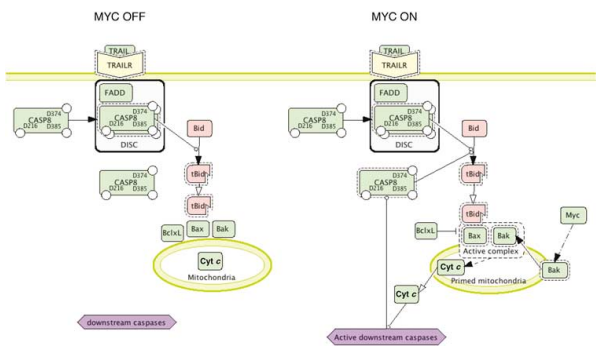


Figure 9 Mitochondria-priming model. Binding of TRAIL to its cognate receptors induces sublethal level of caspase-8 activity in human epithelial and fibroblast cells. In healthy cells, this weak caspase-8 activity is insufficient to trigger Bid-mediated Bax or Bak activation. However, oncogenic c-Myc or specific drugs can preactivate Bak and in these conditions even a weak activation of caspase-8 and Bid fully activates formation of Bak/Bax complexes, which triggers release of cyt *c* and subsequently recruits downstream effector caspases. Downstream effector caspases cleave vital cellular substrates and generate substantially more caspase-8 activity through interchain cleavage, which results in strong Bid activation. We propose that the onset of such caspase feedback loop represents a phase transition, where apoptosis becomes a TRAIL-independent cell autonomous process. These mechanisms may have evolved to ensure that the progression of late stage apoptosis is not dependent on the extracellular availability of death ligands. The model is illustrated in the figure as a process diagram with graphical notation system (Kitano *et al*, 2005). The symbols are: closed arrow, state transition; open arrow, translocation; closed arrow and dotted line, unknown transition; circle-headed line, promotion of transition; solid line surrounding protein complexes, known protein complexes; dotted line surrounding protein complexes, hypothetical protein complexes; dotted line surrounding single protein, active protein.

2011年3月2日水曜日

Examples

J. Biochem. 142, 2007

J. Biochem. 142, 183–191 (2007)
doi:10.1093/jb/mvm121

Model-based Definition of Population Heterogeneity and Its Effects on Metabolism in Sporulating *Bacillus subtilis*

Mineo Morohashi[1], Yoshiaki Ohashi[1,2,*], Saeka Tani[2], Kotaro Ishii[2,†], Mitsuhiro Itaya[2], Hideaki Nanamiya[3,†], Fujio Kawamura[3], Masaru Tomita[1,2] and Tomoyoshi Soga[1,2]

[1]Human Metabolome Technologies, Inc., Tsuruoka, Yamagata 997-0052, Japan; [2]Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017, Japan; and [3]College of Science, Rikkyo (St Paul's) University, Toshima-ku, Tokyo 171-8501, Japan

Received April 1, 2007; accepted May 11, 2007; published online June 1, 2007

The soil bacterium *Bacillus subtilis* forms dormant, robust spores as a tactic to ensure survival under conditions of starvation. However, the sporulating culture includes sporulating and non-sporulating cells, because a portion of the cell population initiates sporulation in wild-type strain. We anticipated that the population effect must be considered carefully to analyse samples yielding population heterogeneity. We first built a mathematical model and simulated for signal transduction of the sporulation cue to see what mechanisms are responsible for generating the heterogeneity. The simulated results were confirmed experimentally, where heterogeneity is primarily modulated by negative feedback circuits, resulting in generation of a bistable response within the sporulating culture. We also confirmed that mutants relevant to negative feedback yield either sporulating or non-sporulating subpopulations. To see the effect of molecular mechanism between sporulating and non-sporulating cells in distinct manner, metabolome analysis was conducted using the above mutants. The metabolic profiles exhibited distinct characteristics with time regardless of whether sporulation was initiated or not. In addition, several distinct characteristics of metabolites were observed between strains, which was inconsistent with previously reported data. The results imply that careful consideration must be made in the interpretation of data obtained from cells yielding population heterogeneity.

Key words: Bacillus subtilis, heterogeneity, metabolome, sporulation.

Abbreviations: ANOVA, analysis of variance; CE-TOFMS, capillary electrophoresis time-of-flight mass spectrometry; PCA, principal component analysis.

Phenotypic heterogeneity in clonal populations has been found in some species of bacteria under certain circumstances (*1–3*). The soil bacterium *Bacillus subtilis* can form population heterogeneity during sporulation under conditions of starvation (*4–7*). This is achieved by functions of positive- and negative-feedback loops in sporulation signal transduction. In cells receiving a sporulation signal, the phosphorylation of kinases (such as KinA) is stimulated and the phosphate group is transferred to Spo0A *via* phosphorelay (*8*). Phosphorylated Spo0A (Spo0A∼P) is a master regulator of sporulation, acting as a transcriptional factor for sporulation-associated genes. This signal transduction system is regulated by a complex mechanism involving multiple positive/negative-feedback loops (*9*).

Recent theoretical and experimental studies suggest that intrinsic characteristics of the biological system generate population heterogeneity [see (*10*) for a review]. Voigt *et al*. (*11*) investigated the dynamics of *sin* operon using a mathematical model, and showed that combining genes from a regulatory protein and its antagonist within the same operon could lead to diverse regulatory functions such as bistability, oscillation and pulse generation. In addition, Iber *et al*. (*12*, *13*) used the *spo*IIA operon as an example to show similar results while de Jong *et al*. (*9*) performed a qualitative simulation, reproducing qualitative characteristics consistent with these experimental results. Involvement of each genetic feedback loop is unquestionable, but how to modulate the scale of subpopulations is still unclear.

It has been a long time since omic approaches were introduced to investigate cellular dynamics. However, influence of population heterogeneity on omic data has never been discussed. We consider that lack of understanding regarding population heterogeneity mislead and complicate the interpretations of omic data.

Here we indicate that the population heterogeneity cannot be ignored in sporulation of *B. subtilis* population. At first, we employed a mathematical model to elucidate the dynamics of Spo0A∼P, including the involvement of both positive and negative feedbacks. Although experimental data cannot be obtained from quantitative

*To whom correspondence should be addressed. Tel: +81-235-25-1447, Fax: +81-235-25-1450,
E-mail: ohashi@humanmetabolome.com/ohashi@sfc.keio.ac.jp
†Present addresses: Kotaro Ishii, The University of Tokyo, Kashiwa, Chiba 277-0000, Japan; Hideaki Nanamiya, Ehime University, Matsuyama, Ehime 790-8577, Japan.

The sporulation fraction was defined in terms of colony-forming units (CFU) per millilitre.

Microscopy and Data Processing—An aliquot (∼20 μl) of culture medium at the sporulation phase was briefly centrifuged and the supernatant removed. Cells were washed once in MilliQ water then re-suspended in 2 μl of component A from the SlowFade-Antifade Kit (Molecular Probes, Inc., OR, USA). A 1 μl aliquot of the cell suspension was then inoculated onto an agarose layer on a glass slide and covered with a coverslip. Microscopic analyses were conducted using an AxioskopMOT 2 microscope (Carl Zeiss, Göttingen, Germany) and a CoolSNAP fx CCD camera (Roper Scientific, Inc., AZ, USA). To detect the fluorescence of GFPuv, Filter Set 17 (Carl Zeiss) was used. Images were obtained 40 s after UV excitation. The fluorescence intensity of individual cells was calculated using MetaMorph Ver. 4.6 software (Universal Imaging, Co., PA, USA).

Instrumentation—All capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS) experiments were performed using an Agilent CE Capillary Electrophoresis System G1600A (Agilent Technologies, CA, USA), and an Agilent TOFMS System G1969A. For system control and data acquisition we used G2201AA Agilent ChemStation software for CE and Analyst QS for Agilent TOFMS software.

CE-TOFMS Conditions for Cation Analysis—Samples were prepared as described in (*22*). Separations were carried out on a fused silica capillary (50 μm i.d. ×100 cm total length) using 1 M formic acid. Samples were injected with a pressure injection of 50 mbar for 3 s. The applied voltage was set at +30 kV and the sheath liquid was prepared as 50% (v/v) MeOH/H_2O. For TOFMS, ions were examined successively to cover the whole range of *m/z* values from 50 through 1,000. The fragmentor voltage was set at 75 V and the skimmer and Oct RFV voltages at 50 V and 125 V, respectively. The capillary voltage was set at 4,000 V (*23*).

CE-TOFMS Conditions for Anion/Nucleotide Analysis—Samples were prepared as described in (*22*). Separations were carried out on a fused silica capillary (50 μm i.d. ×100 cm total length) using 50 mM ammonium acetate (pH8.5 for anion, and pH7.5 for nucleotide, respectively). Samples were injected with a pressure injection of 50 mbar for 30 s. The applied voltage was set at +30 kV and the sheath liquid was prepared as 5-mM ammonium acetate 50% (v/v) MeOH/H_2O. For TOFMS, ions were examined successively to cover the whole range of *m/z* values from 50 through 1,000. The fragmentor voltage was set at 100 V and the skimmer and Oct RFV voltages at 50 V and 200 V for anion, and 75 V and 200 V for nucleotide, respectively. The capillary voltage was set at 3,500 V (*23*).

Data Processing—Peak extraction was carried out using our proprietary software (Sugimoto, unpublished data) and peak pre-processing was performed according to the P-BOSS method (*24*) using Excel 2003 (Microsoft, WA, USA). Mathematical simulation was conducted using XPP-AUTO (*25*). Statistical analyses were performed *via* MATLAB (Mathworks, MA, USA).

RESULTS AND DISCUSSION

The Negative-feedback Loop Dominates the Threshold of Sporulation Switch—In cells initiating sporulation, expression of *spo0H*, which encodes sporulation-specific σ^H, was induced by a reduction in the AbrB level (Fig. 1). The RNA polymerase that contains σ^H stimulated the expression of phosphorelay components, *kinA*, *spo0F* and *spo0A*, which constitute multiple points of the positive-feedback loop. Negative-feedback regulation was also observed in *B. subtilis* phosphorelay. Expression of the *spo0E* gene, which encodes Spo0A∼P-specific phosphatase, is induced by a reduction in the AbrB level at the sporulation onset (*26*). Accordingly, it has been suggested that phosphorelay is negatively regulated by a solo feedback system (*27*).

A mathematical model was created, and the dynamics of the model were simulated (see Supplementary Data for detail). The system was characterized by varying two parameters, the sporulation signal (ϕ) and concentration of Spo0A∼P, as illustrated in Fig. 2. The amount of stimulus required for sporulation switch increased as the ratio of negative and positive feedback loops, $r = f_N/f_P$, increased (Fig. 2A). Comparing the system characteristics by varying the feedback coefficients (f_N and f_P) revealed that as the value of f_N increased, the bistability region shifted its operating region dramatically towards a larger region against the sporulation signal (Fig. 2B), while f_P did not change its operating region sufficiently (Fig. 2C). These findings indicate that negative feedback, which is achieved by expression of the *spo0E* gene, primarily modulates bistability behaviour.

The function of Spo0E in population heterogeneity suggested in our mathematical model (Fig. 2A) was further demonstrated using BEST12014 (*spo0E*::*cat*), in which the negative-feedback loop created by *spo0E* is destroyed ($r = 0$ in our model). In this strain, distribution is excessively biased towards the sporulating subpopulation at T_3 (Fig. 3A and C), resulting in sporulation of >95% of the cells. This was consistent with the sporulation frequency at T_{24}. Next, we constructed a strain able

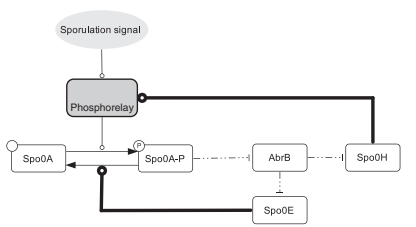


Fig. 1. **Schematic representation of the phosphorelay network required for initiation of sporulation in *B. subtilis*.** The diagram was illustrated using CellDesigner 3.5.1 (*37*) (http://celldesigner.org), and the notation follows that proposed by Kitano *et al*. (*38*). The networks downstream of AbrB are simply categorized into positive and negative feedback loops, the regulation of which is represented by a bold arrow from Spo0H, and a bold arrow from Spo0E, respectively.

2011年3月2日水曜日

Examples

Journal of Theoretical Biology
248, 2007

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 248 (2007) 130–144
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Signaling perturbations induced by invading *H. pylori* proteins in the host epithelial cells: A mathematical modeling approach

William Dampier, Aydin Tozeren*

Center for Integrated Bioinformatics, Bossone 714, School of Biomedical Engineering, Science and Health Systems, Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA

Received 25 October 2006; received in revised form 17 January 2007; accepted 6 March 2007
Available online 18 March 2007

Abstract

Helicobacter pylori (*H. pylori*), a gram-negative bacterium, infects the stomach of approximately 50% of the world population. *H. pylori* infection is a risk factor for developing chronic gastric ulcers and gastric cancer. The bacteria produce two main cytotoxic proteins: Vacuolating cytotoxin A (*VacA*) and Cytotoxin-Associated gene A (*CagA*). When these proteins enter the host cell they interfere with the host MAP Kinase and Apoptosis signaling pathways leading to aberrant cell growth and premature apoptosis. The present study expanded existing quantitative models of the MAP Kinase and Apoptosis signaling pathways to take into account the protein interactions across species using the CellDesigner tool. The resulting network contained hundreds of differential equations in which the coefficients for the biochemical rate constants were estimated from previously published studies. The effect of *VacA* and *CagA* on the function of this network were simulated by increasing levels of bacterial load. Simulations showed that increasing bacterial load affected the MAP Kinase signaling in a dose dependant manner. The introduction of *CagA* decreased the activation time of *mapK* signaling and extended activation indefinitely despite normal cellular activity to deactivate the protein. Introduction of *VacA* produced a similar response in the apoptosis pathway. Bacterial load activated both pathways even in the absence of external stimulation. Time course of emergence of transcription factors associated with cell division and cell death predicted by our simulation showed close agreement with that determined from a publicly accessible microarray data set of *H. pylori* infected stomach epithelium. The quantitative model presented in this study lays the foundation for investigating the affects of single nucleotide polymorphisms (SNPs) on the efficiency of drug treatment.

Keywords: *H. pylori*; Computational model; MapK; EGF signaling; FasL signaling

1. Introduction

Helicobacter pylori (*H. pylori*) is a gram-negative bacterium that inhabits the acidic conditions of the human stomach in 50% of the world's population (Hatakeyama and Brzozowski, 2006). Chronic *H. pylori* infection is a major risk factor for gastric ulcers and gastric cancer (Hatakeyama and Brzozowski, 2006).

H. pylori effects host stomach epithelial cells by producing two unique cytotoxic proteins: Vacuolating cytotoxin A (*VacA*) and Cytotoxin-Associated gene A (*CagA*) (Cover and Blanke, 2005; Hatakeyama and Higashi, 2005). In previously published experimental studies, *VacA* has been implicated in hyper-vacuolization, small molecule leakage and apoptosis (Cover and Blanke, 2005). *CagA* has been implicated in morphological changes ("hummingbird" phenotype) and induction of the MAP Kinase pathway (Hatakeyama and Higashi, 2005).

VacA acts as both a trans-membrane channel and a protein activator (Galmiche et al., 2000; Cover et al., 2003). *VacA* associates with planar lipid membranes, such as the outer cell membrane, vesicles and mitochondrial membranes (Galmiche et al., 2000). When *VacA* associates with vesicle membranes it creates a leak channel, which eventually leads to vesicle swelling (Cover and Blanke, 2005). This hyper-vacuolization is a drastic phenotypic

*Corresponding author. Tel.: +1 215 895 6148.
E-mail addresses: william.dampier@drexel.edu (W. Dampier), aydin.tozeren@drexel.edu (A. Tozeren).

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.03.014

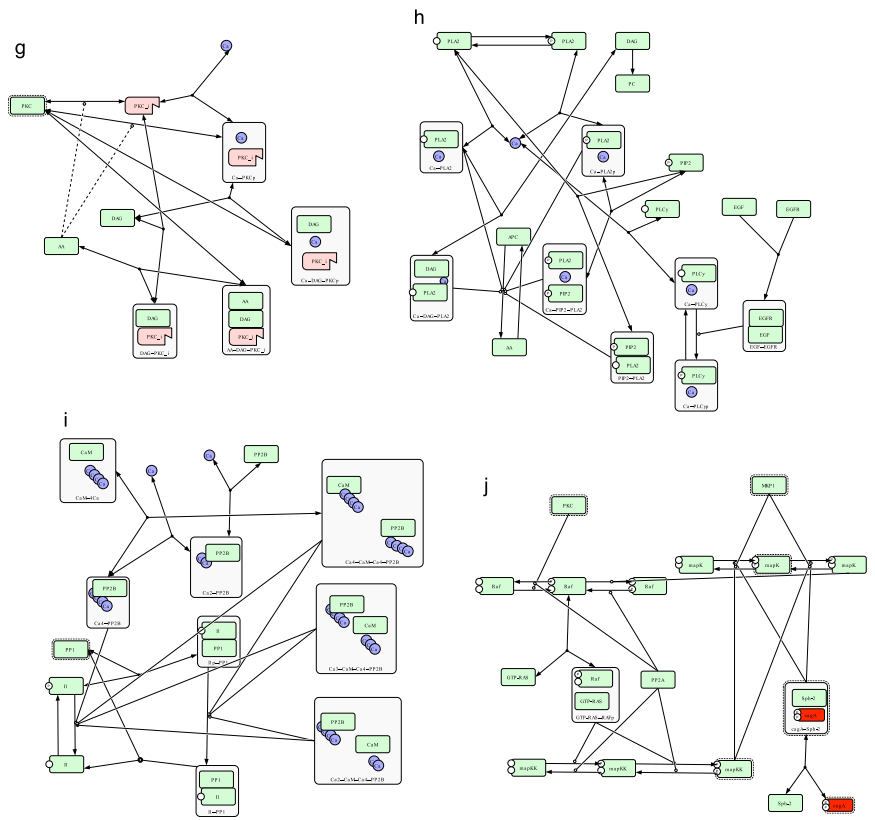


Fig. 1. (*Continued*)

2.2. Map Kinase signaling pathway

The Map Kinase signaling pathway equations used in our study were previously presented by (Bhalla and Iyengar, 1995). Their mathematical model contained 120 nodes (proteins in various phosphorylation states) and 200 reactions between the nodes. The rate equations used in the model consisted of simple association reactions and Michaelis–Menten enzyme kinetics. We used Locus Link ID comparison to confirm that the nodes in Bhalla and Iyengar (1995) were actually present in the KEGG Map Kinase pathway (Kanehisa and Goto, 2000). Moreover, we were able to confirm the interaction connections between these nodes of the Map Kinase pathway using the ingenuity pathway analysis (IPA [www.ingenuity.com]). The rate equations governing the network nodes and their interactions with *H. pylori* proteins are presented in the supplementary information. The Map Kinase parameters and equations appearing in the supplementary information were previously published by Bhalla and Iyengar (1995) and more recently used by Pant and Ghosh (2005a, b).

The equations governing the interaction of the host signaling networks with the invading *H. pylori* proteins used in this study are described in Eqs. (1)–(7) in Table 1. In these equations the bacterial protein *CagA* interacts with *GBR-2* and *Shp-2* within the Map Kinase pathway (Censini et al., 2001) and also *CagA* is phosphorylated at EPIYA motifs by the host protein *SRC* Kinase (Naito et al., 2006).

Acknowledgement

SBML
* SBML community

SBGN
* SBGN community

SBW (Univ. of Washington)
* Frank Bergmann
* Herbert Sauro

SOSlib (Univ. of Vienna)
* Rainer Machne
* Christoph Flamm

PANTHER (SRI)
* Anushya Muruganujan
* Huaiyu Mi

COPASI (Univ. of Heidelberg)
* Ralph Gauges
* Sven Sahle
* Ursula Kummer

SABIO-RK (HITS gGmbH)
* Saqib Mir (EBI)
* Martin Golebiewski
* Wolfgang Mueller
* Isabel Rojas

CellDesigner
* Yukiko Matsuoka (SBI)
* Samik Ghosh (SBI)
* Hiroaki Kitano (SBI)
* Noriko Hiroi (Keio Univ.)
* Hiromu Takizawa (Keio Univ.)
* Akiya Jouraku (NIAS)
* Norihiro Kikuchi (MKI)

CellDesigner Practice

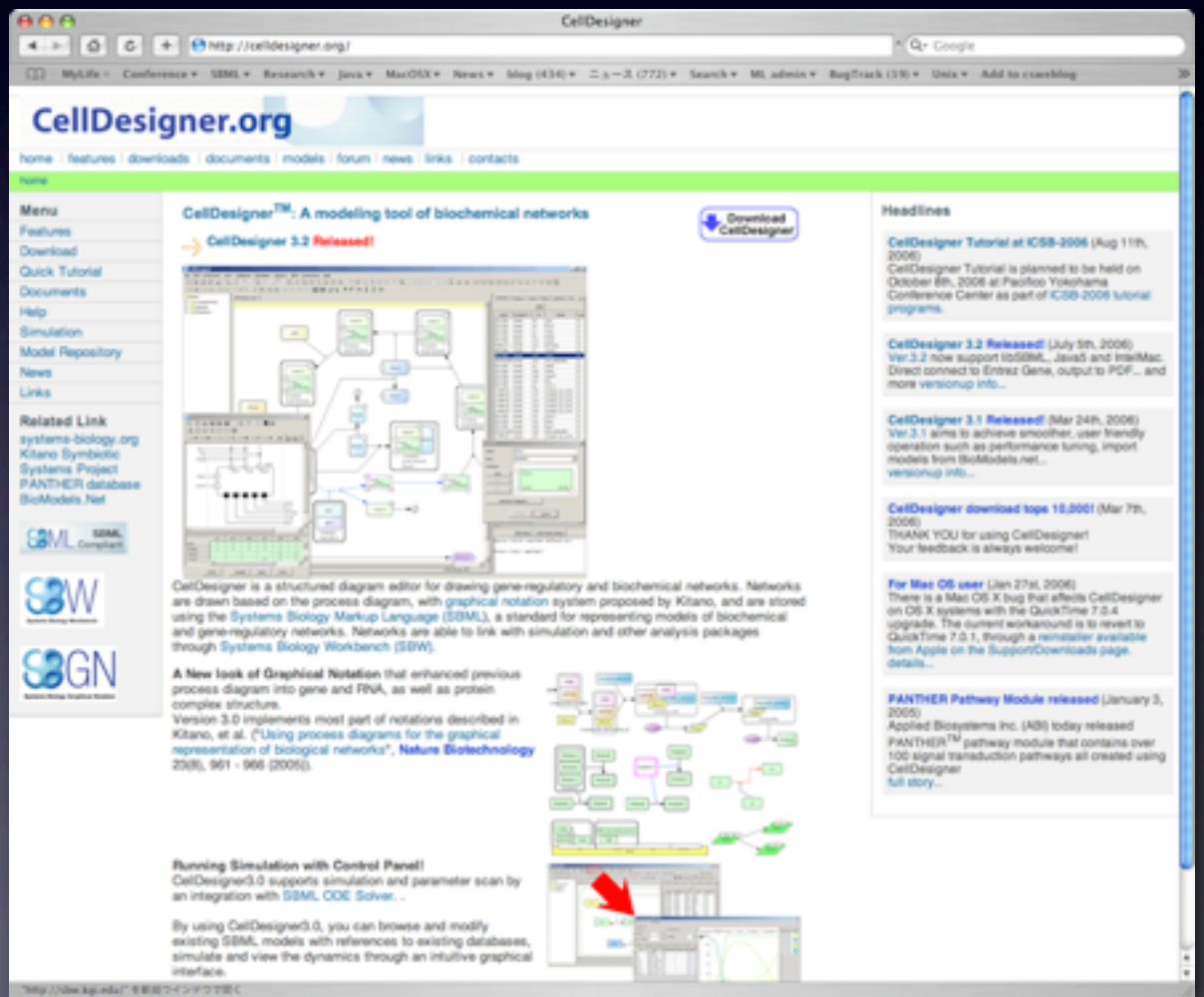
Akira Funahashi
Keio University, Japan
28th Feb. 2011

Keio University
1858
CALAMVS
GLADIO
FORTIOR

SBI The Systems Biology Institute

Before we start

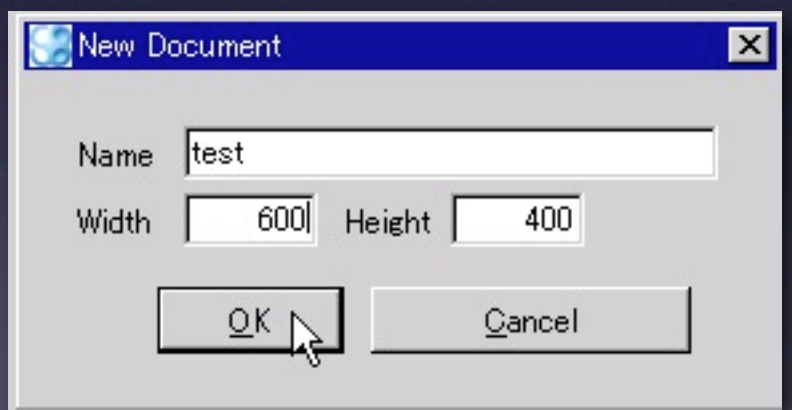
- Please download / install CellDesigner 4.1 from http://celldesigner.org

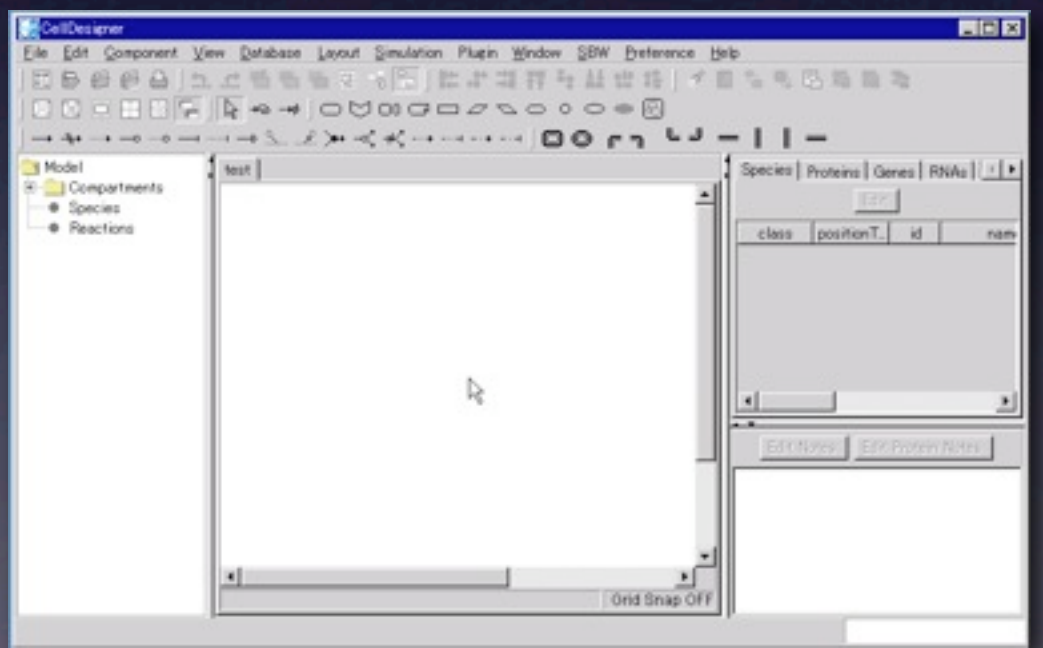


Installation

Demonstration

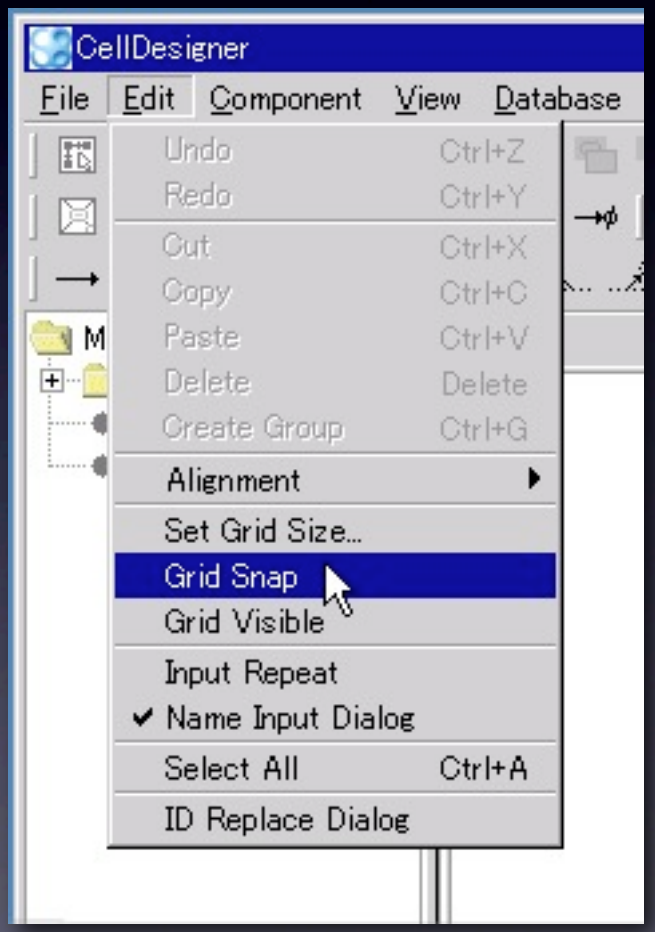
- Create new model:
  - [File] → [New] → input title → [OK]





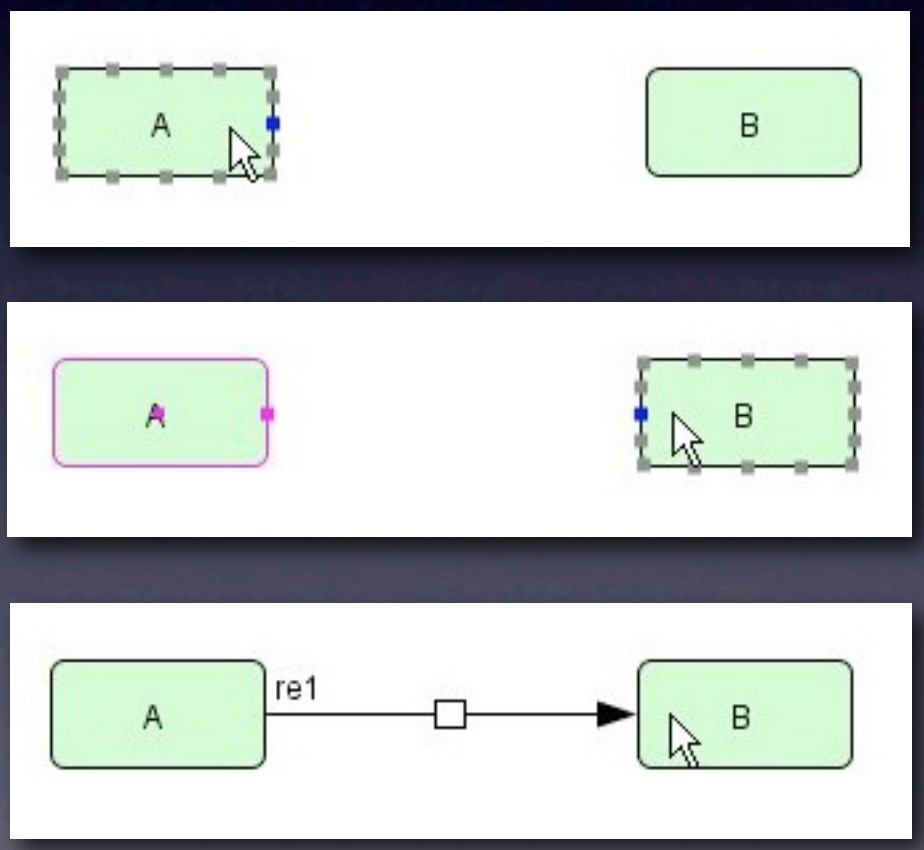
Tips

- Enable [Grid Snap] will help you draw your model much easier



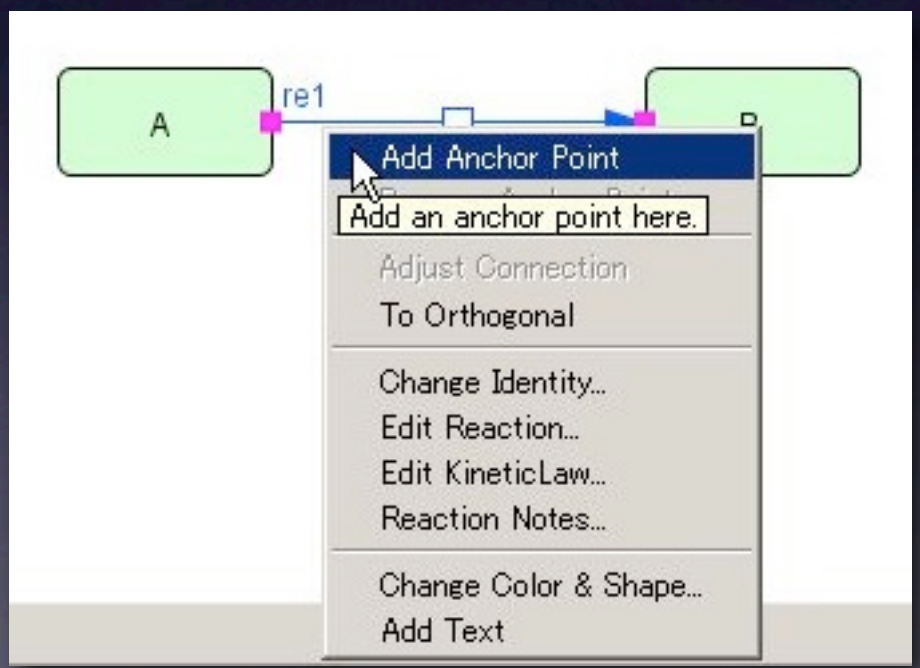
Create Reaction

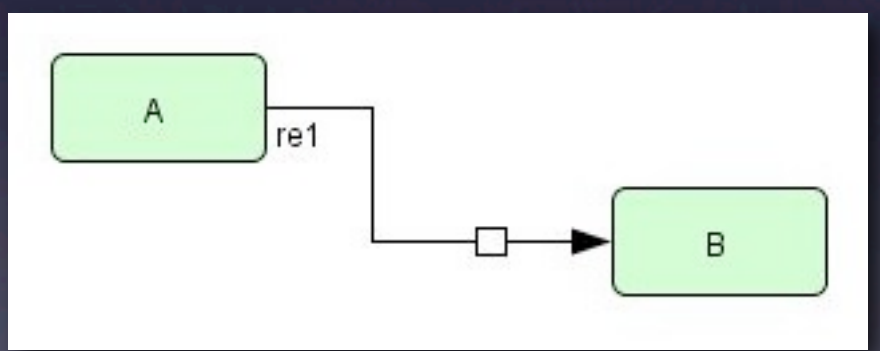
- Create Protein “A” and “B”
- Draw “State transition” arrow from “A” to “B”



Add Anchor Point

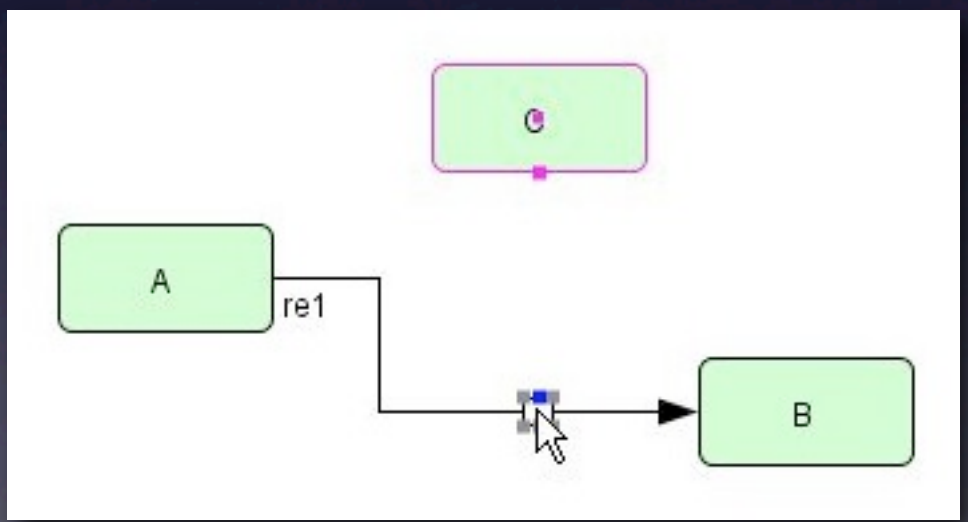
- Add 2 anchor points to reaction
- Drag reaction and anchor point to change its shape

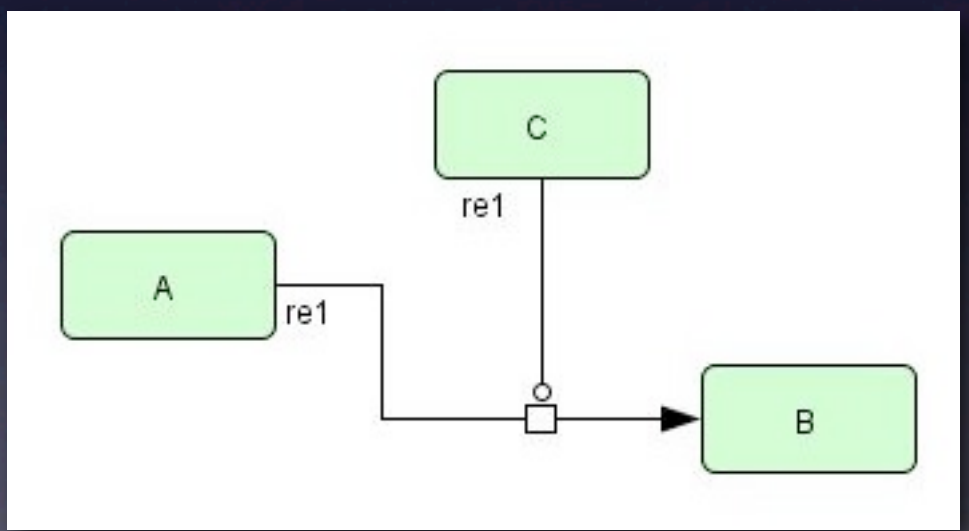




Add Catalysis

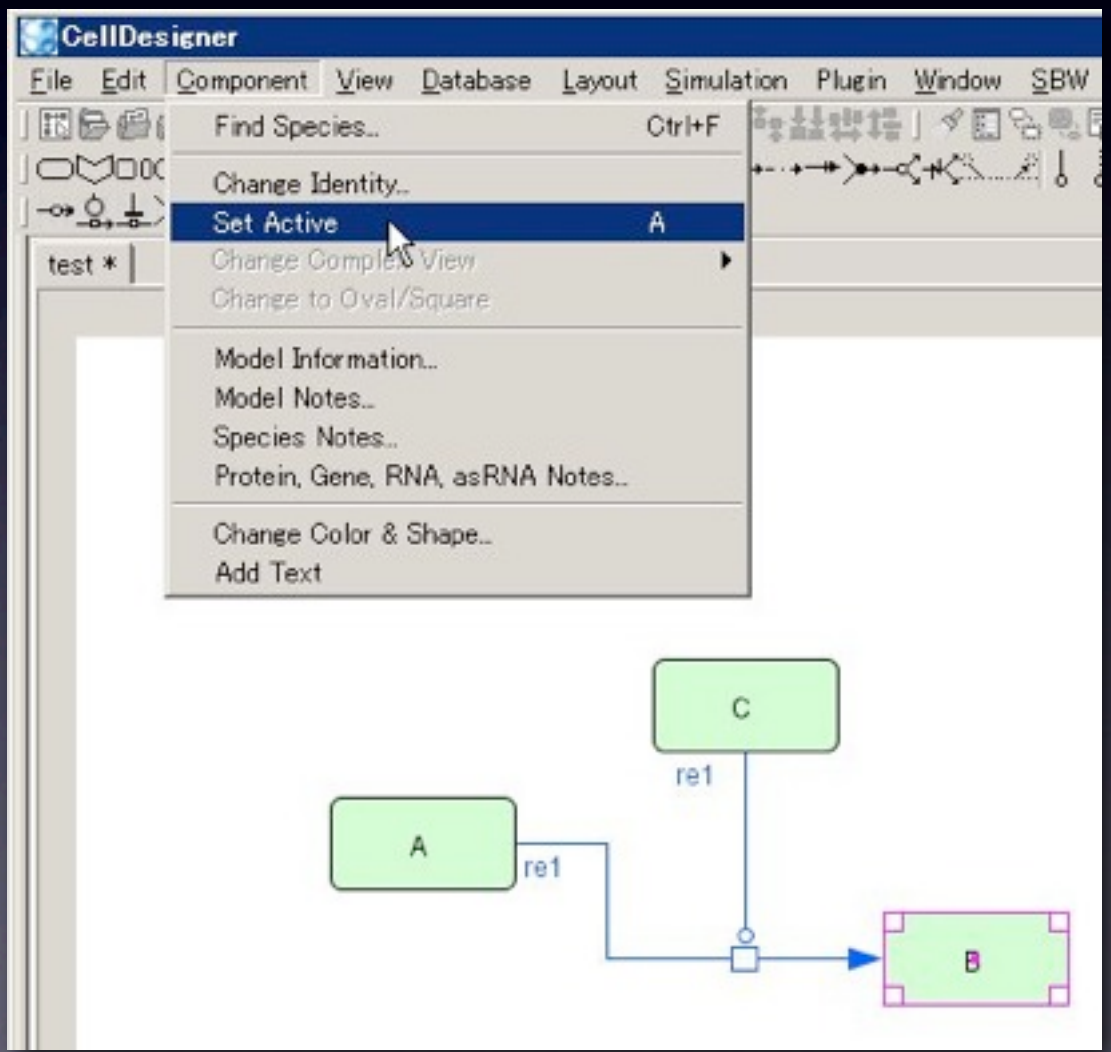
- Add Protein “C”
- Add Catalysis reaction from “C” to the reaction

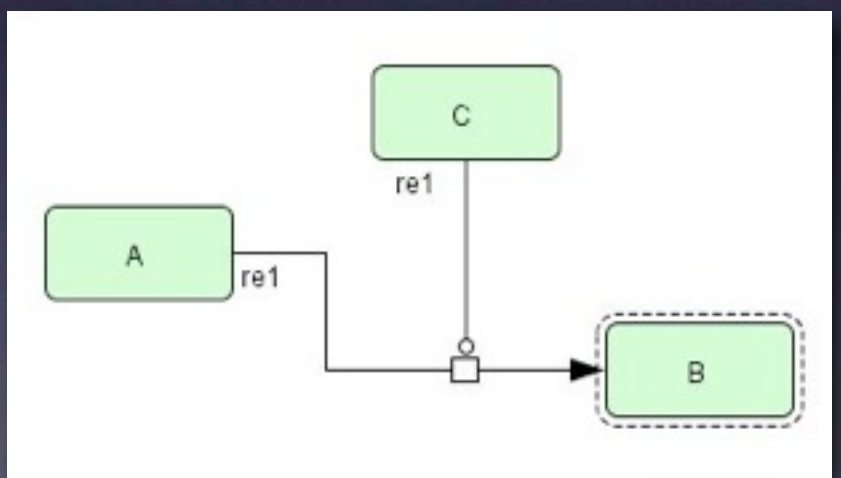




Set Active State

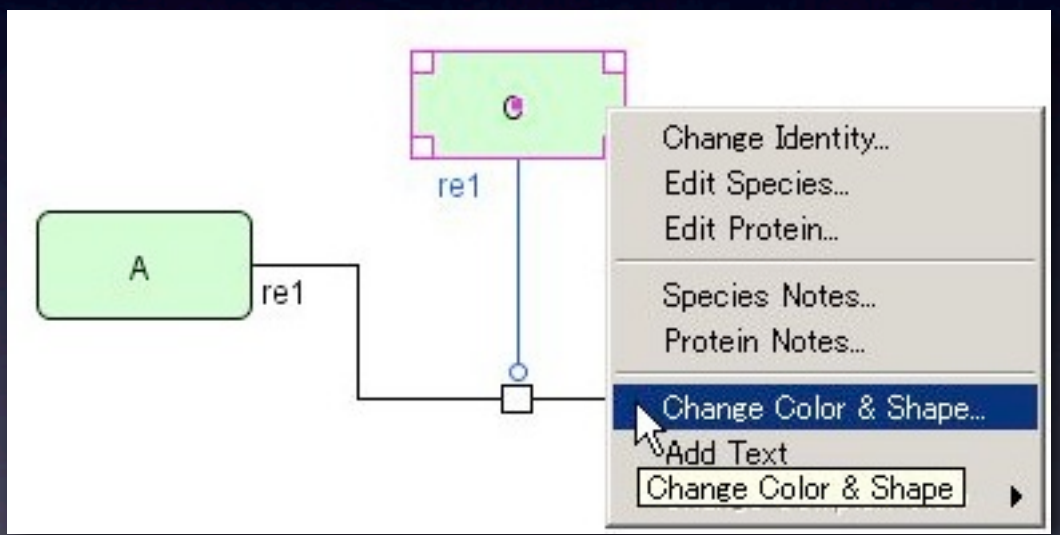
- Select Protein “B”
- [Component] → [Set Active]

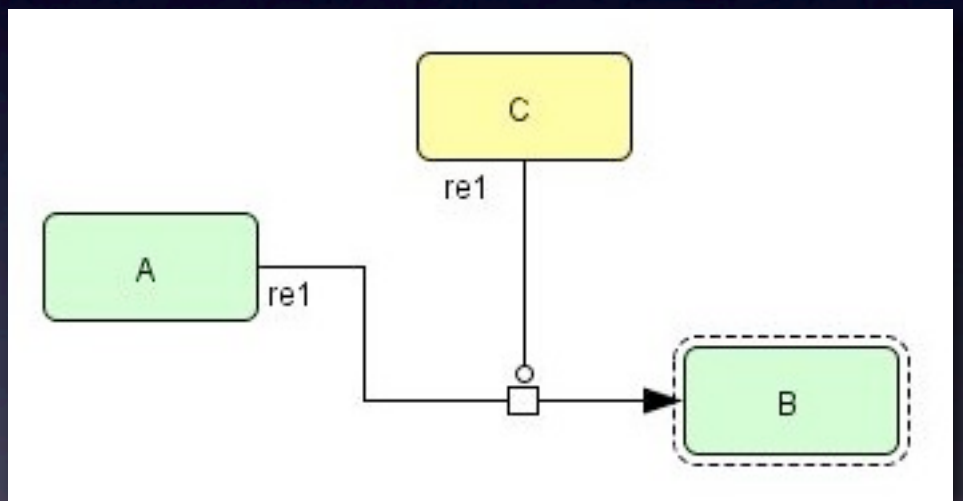




Change Color

- Right click on Protein “C”
- Select [Change Color & Shape...]

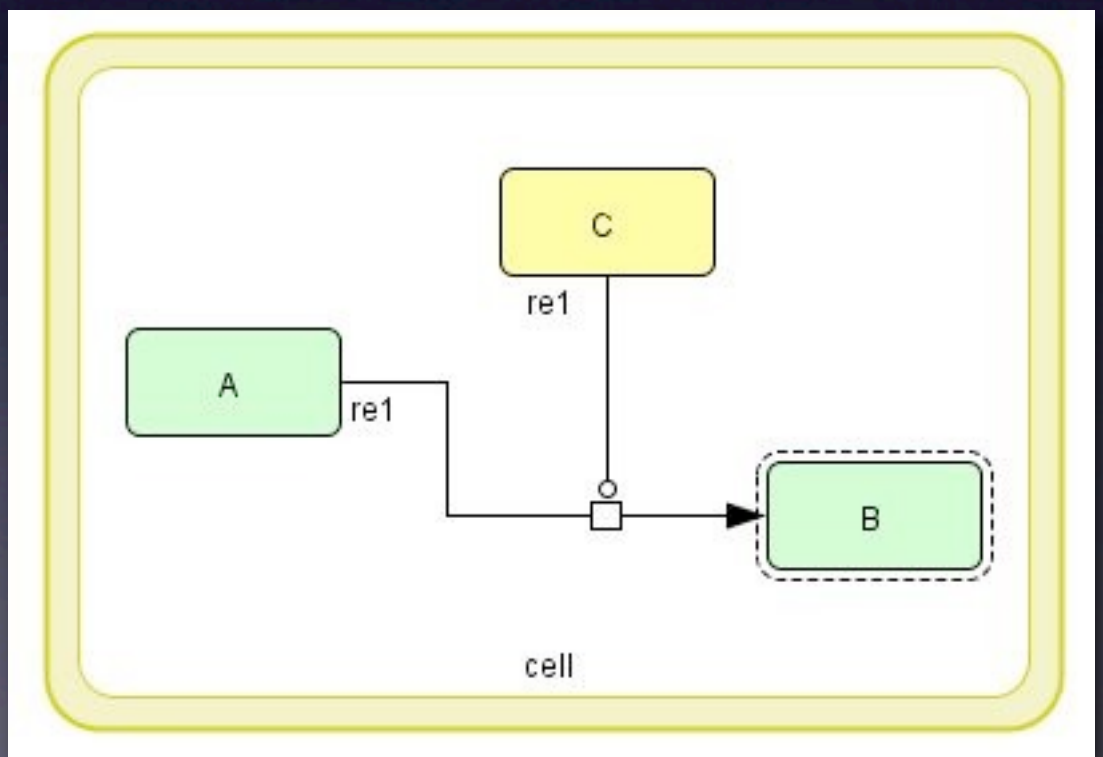




Compartment

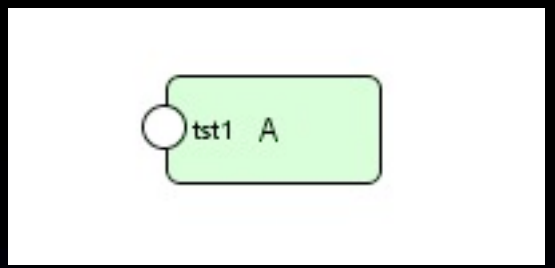
- Click [Compartment] icon

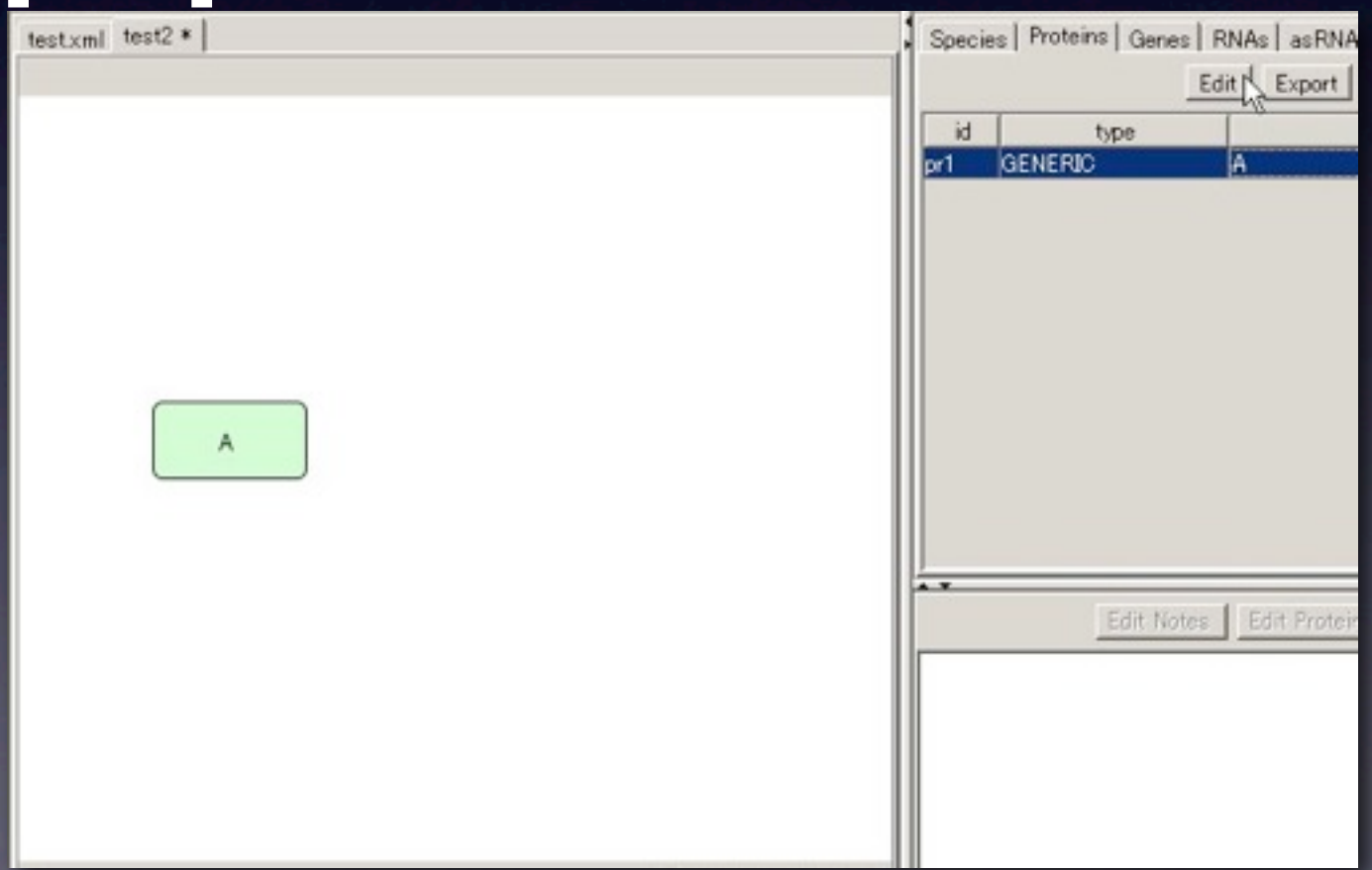
- Drag mouse cursor to specify its area
- Input name of compartment



Add Residue to Protein

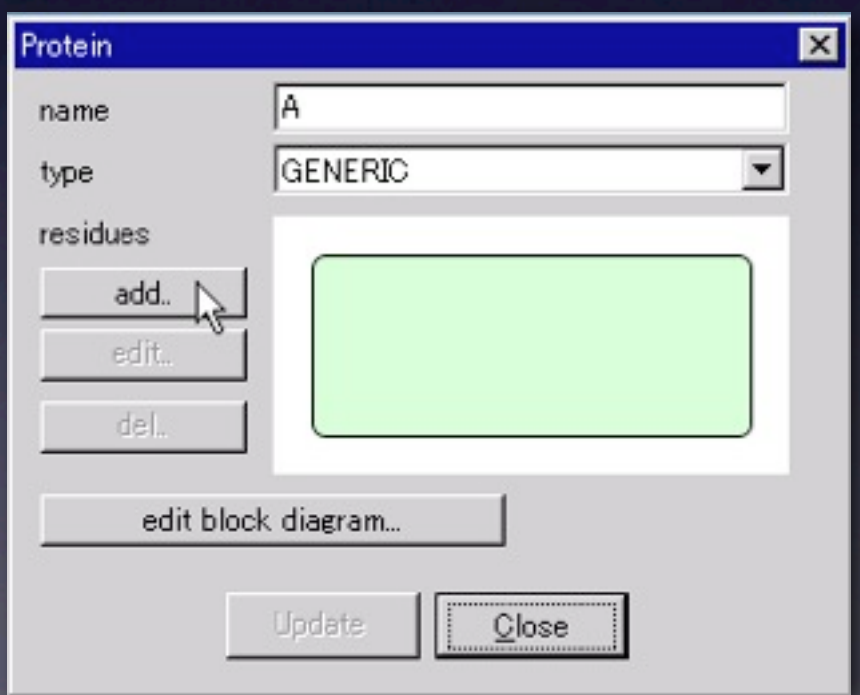
- Create new model (test2)
- Create Protein “A”
- Select Protein “A” in [Proteins] Tab
- Click [Edit] button

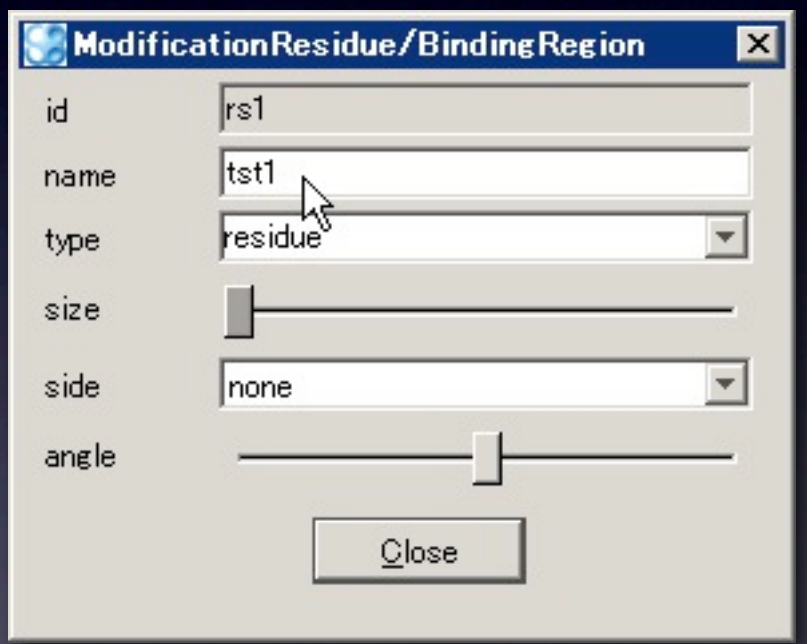


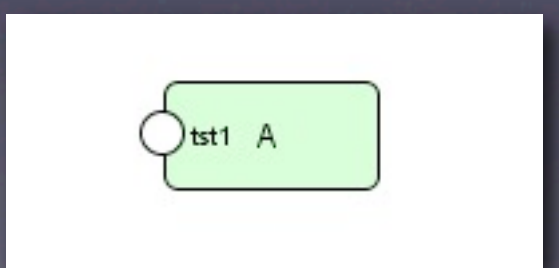


Add Residue to Protein

- Click [add] button on [Protein] dialog
- Input name for the residue (tst1)
- Click [Close] button
- Click [Update] Button

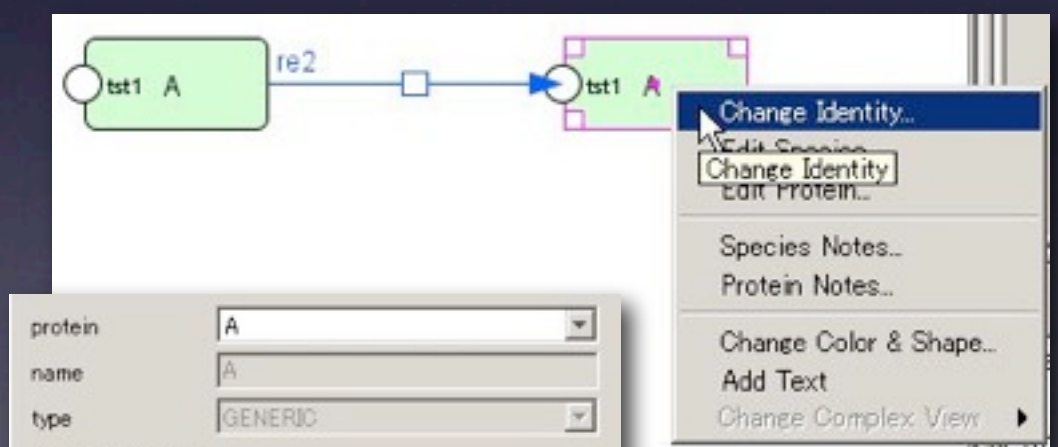


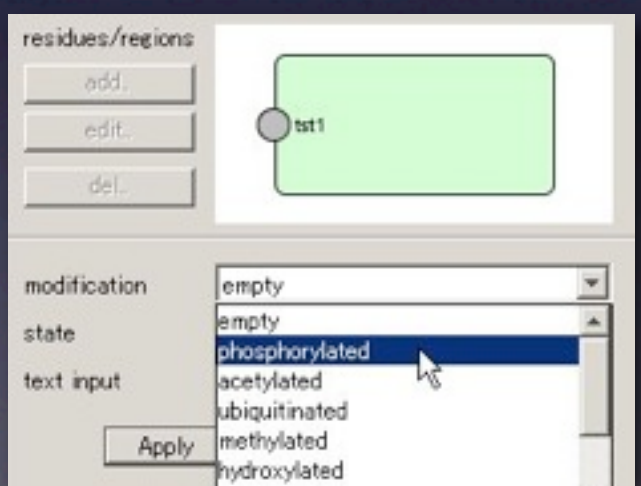


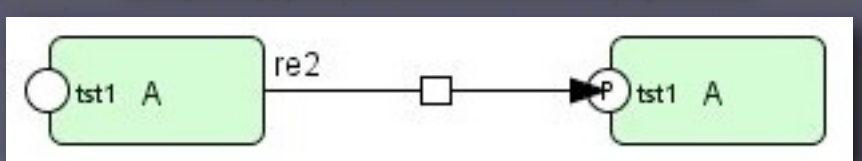


Add Residue to Protein

- Copy & Paste Protein “A” and then draw “State Transition” arrow
- Right Click on “A” (right side) and select [Change Identity...]
- Click residue “tst1” in Dialog
- Select [phosphorylated] in modification

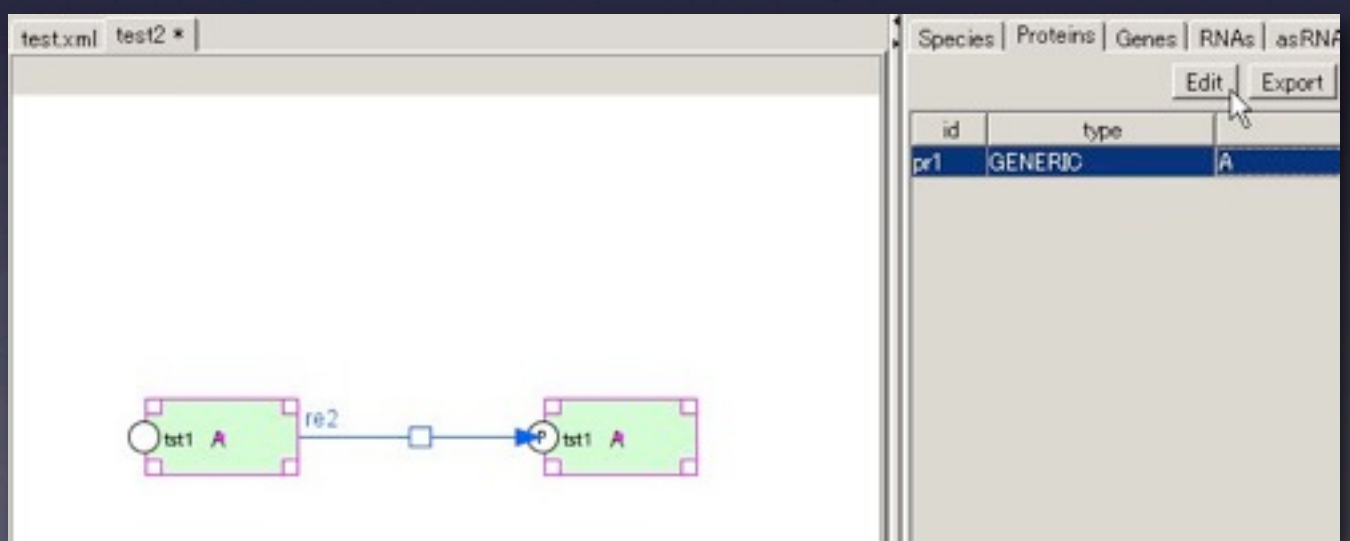


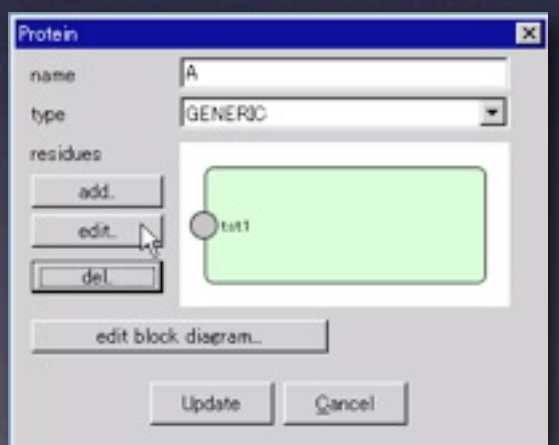




Change Position of Residue

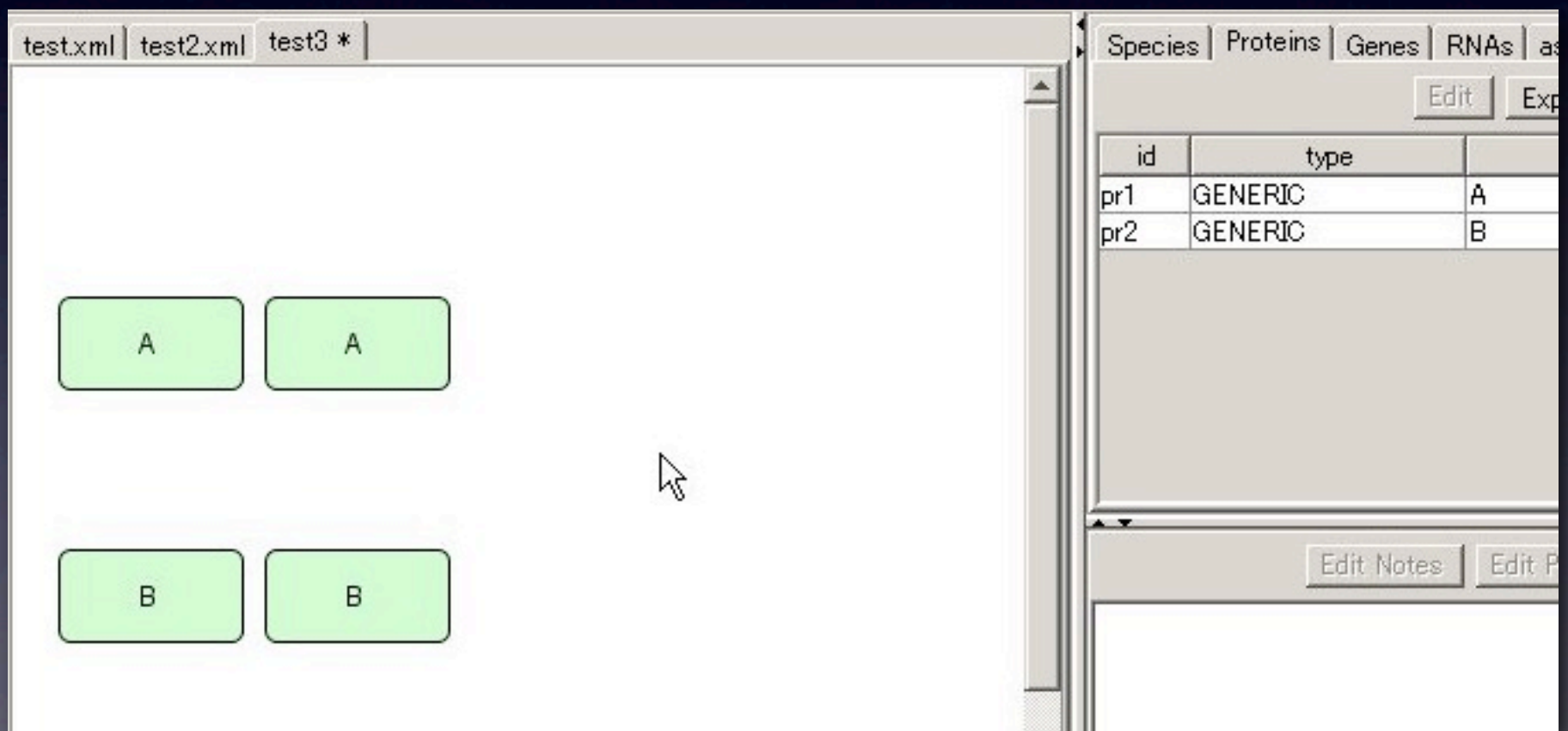
- Select Protein “A” in [Proteins] Tab
- Click [Edit] button
- Click residue “tst1” in Dialog
- Drag residue to your favorite position





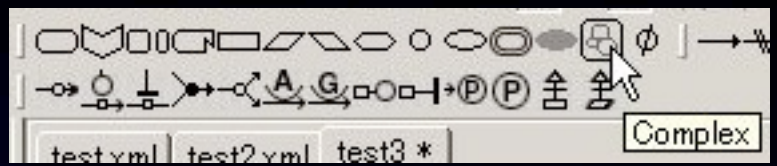
Complex

- Create new model (test3)
- Create Proteins “A” and “B”
- Copy & Paste both “A” and “B”

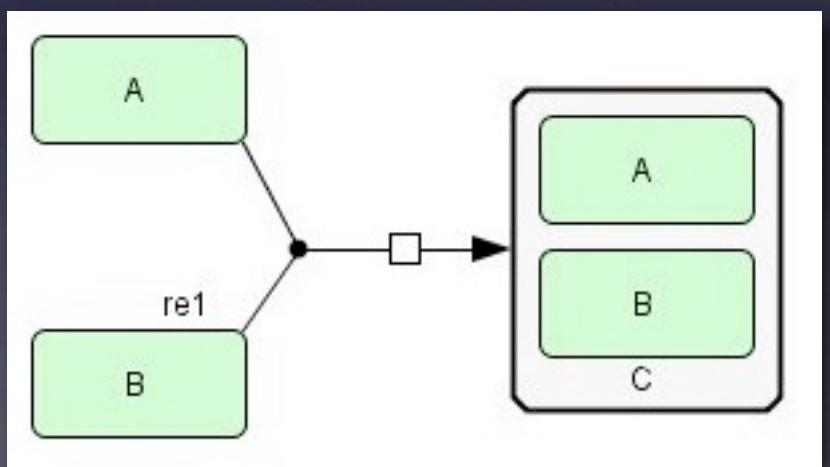


Complex

- Click [Complex] icon and create complex “C”

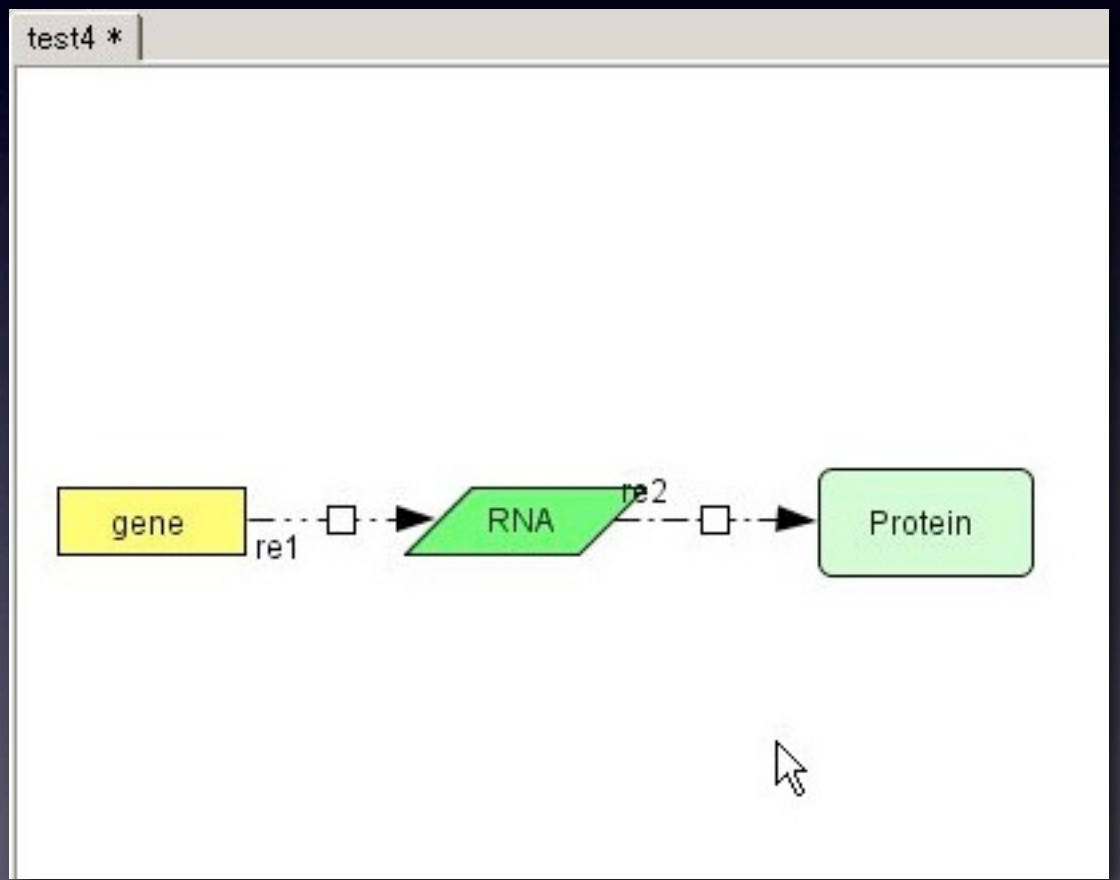


- Drag Protein “A” and “B” into complex C
- Draw “Heterodimer Association” arrow



Gene & RNA

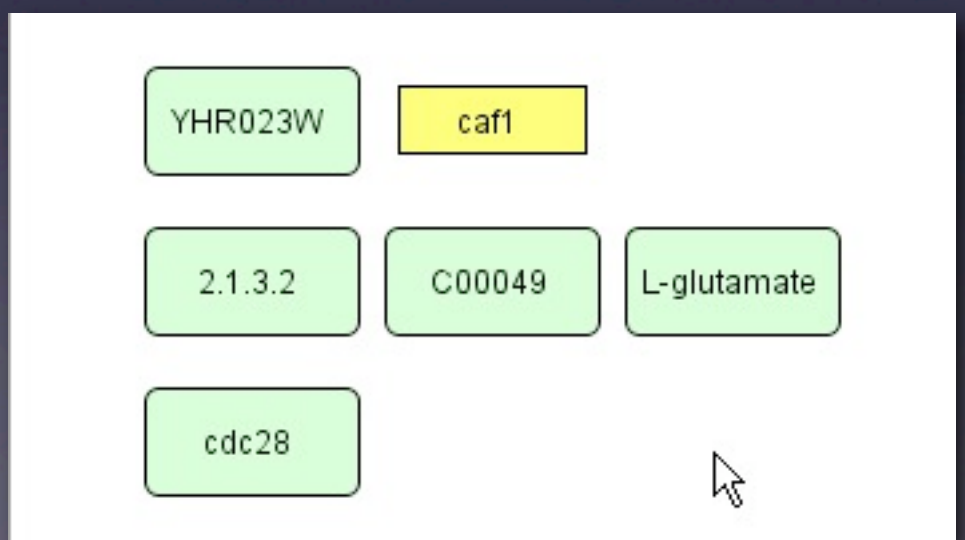
- Create new model (test4)
- Create gene, RNA and Protein
- Draw “Transcription” and “Translation”



See “geneRNA41.xml” for more examples

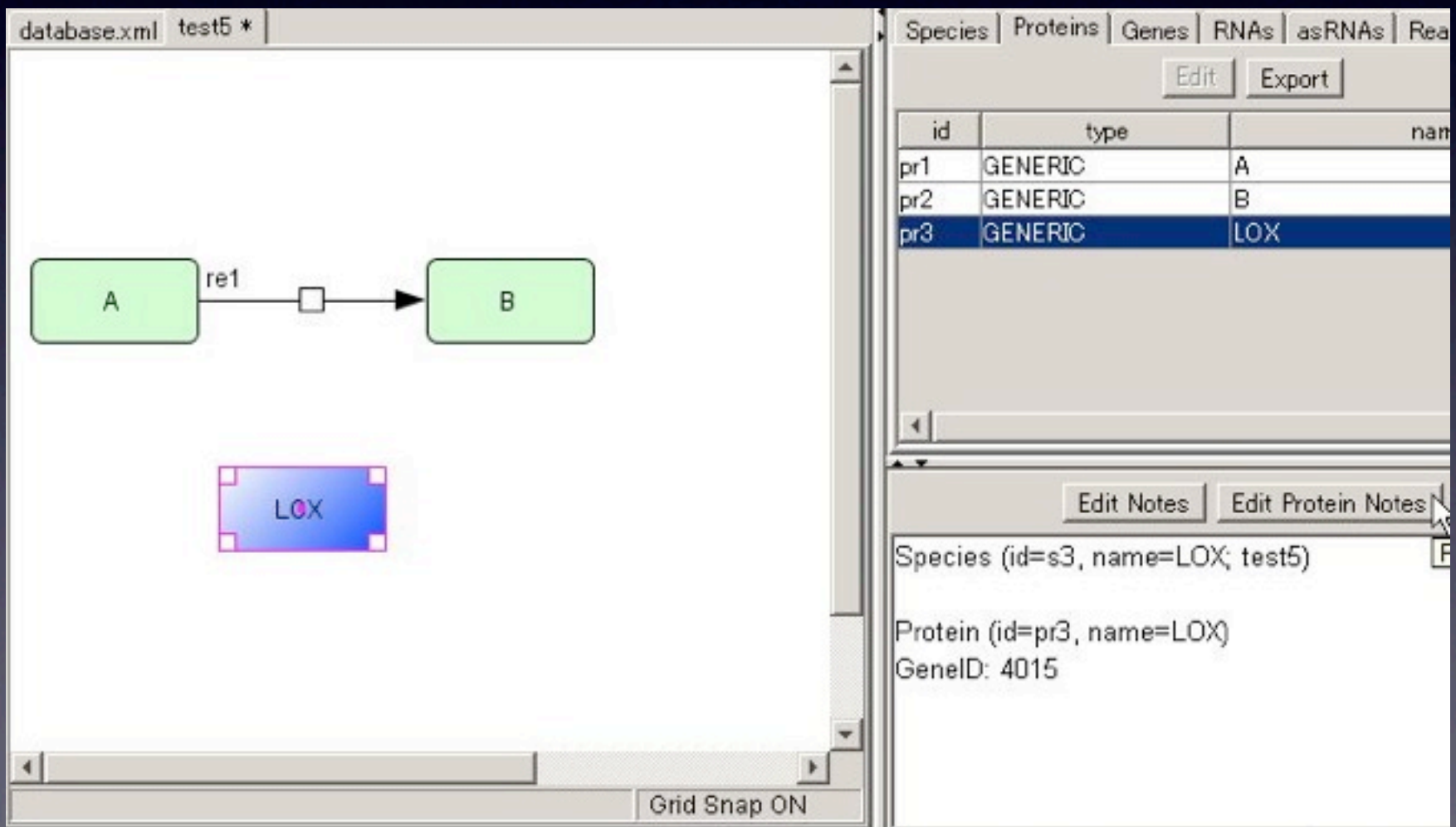
Database Connection

- Search Database by Name:
  - SGD
  - DBGET
  - iHOP
  - Entrez Gene



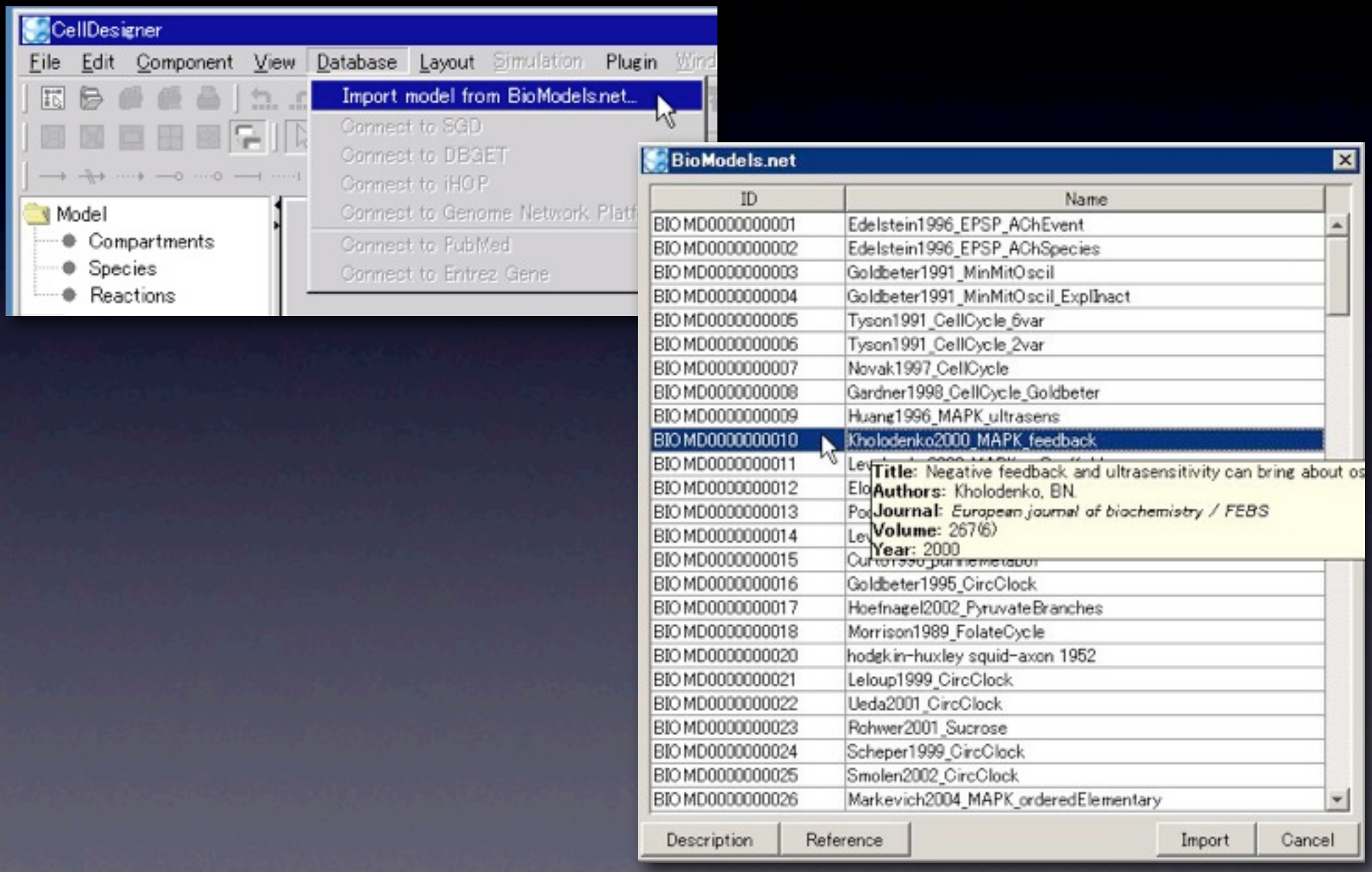
Database Connection

- Search Database by Notes:
  - PubMed: PMID: 123456
  - Entrez Gene: GeneID: 4015



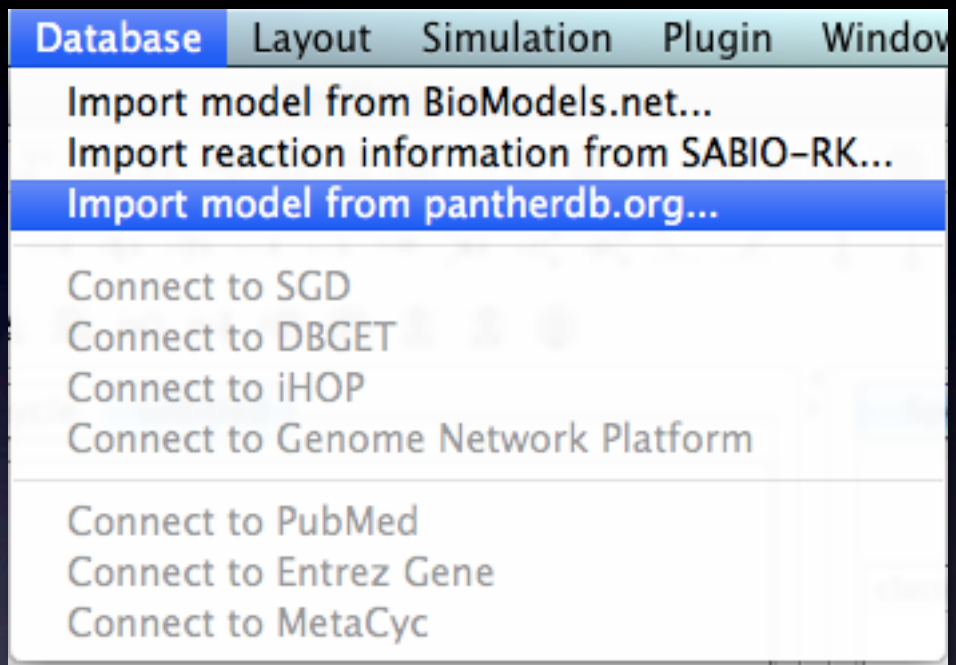
Database Connection

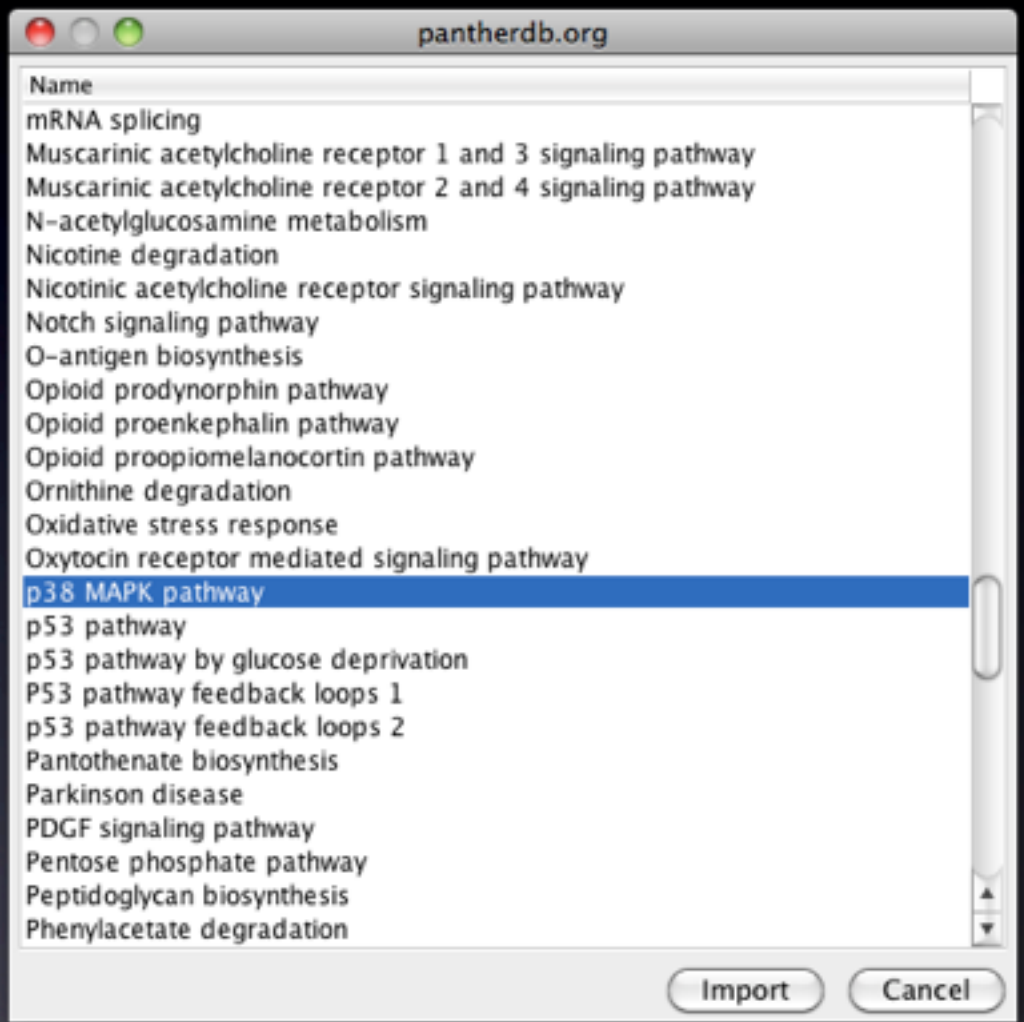
- Import model from BioModels.net



Database connection

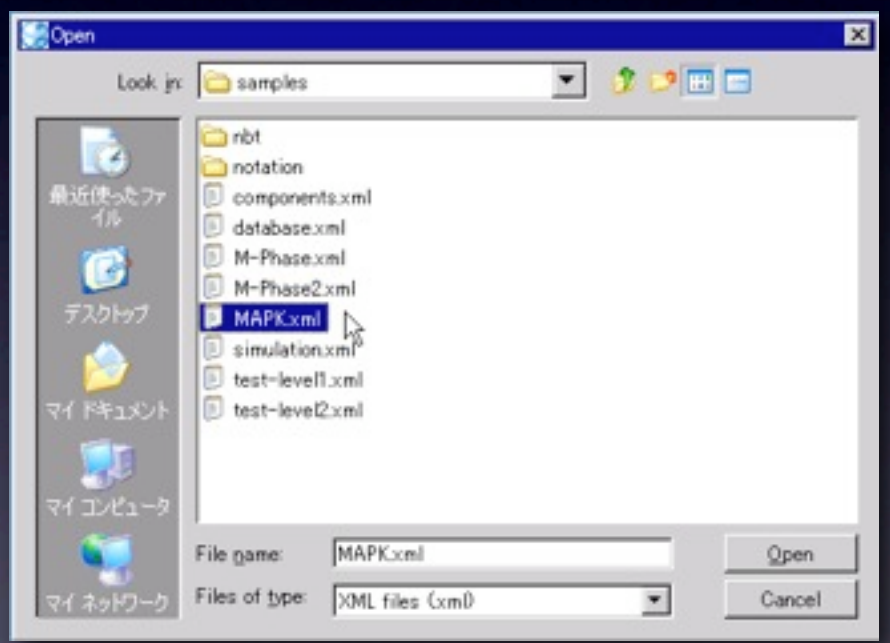
- Import model from PANTHER

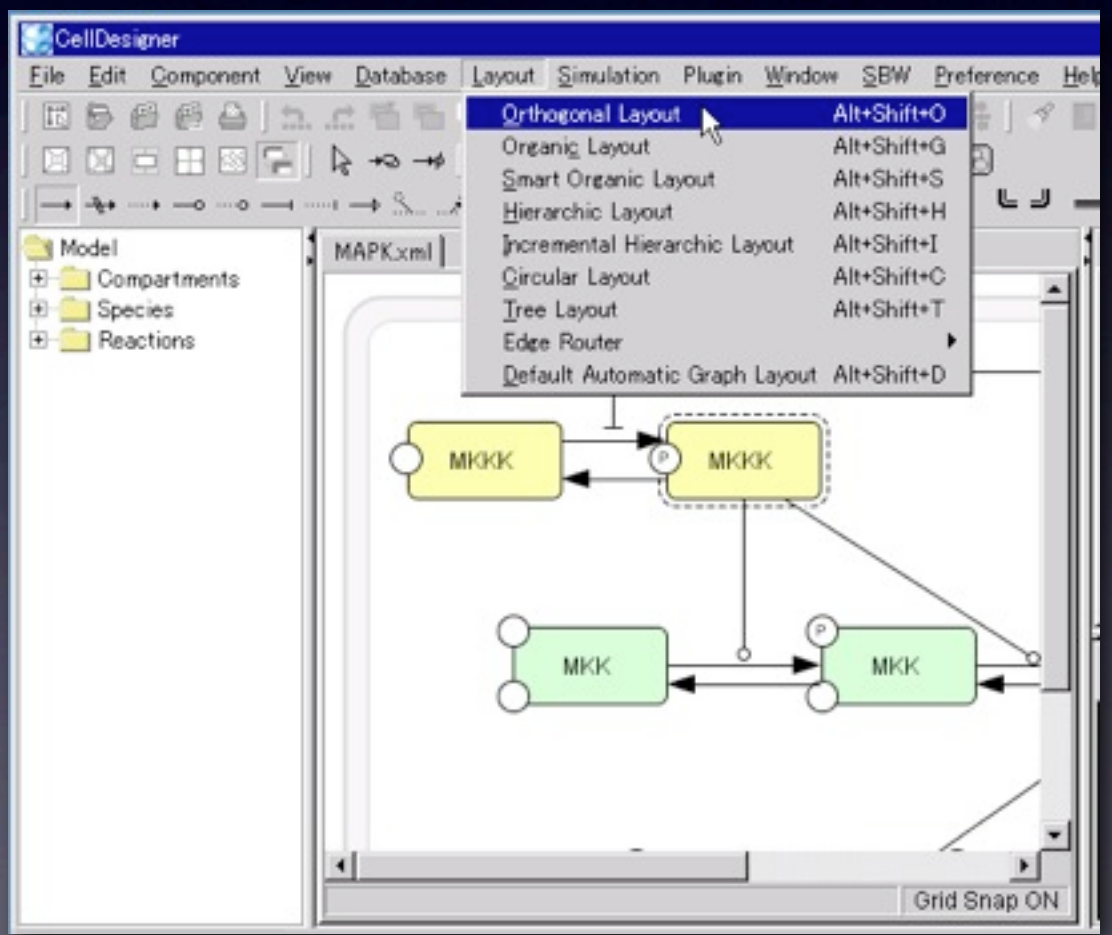




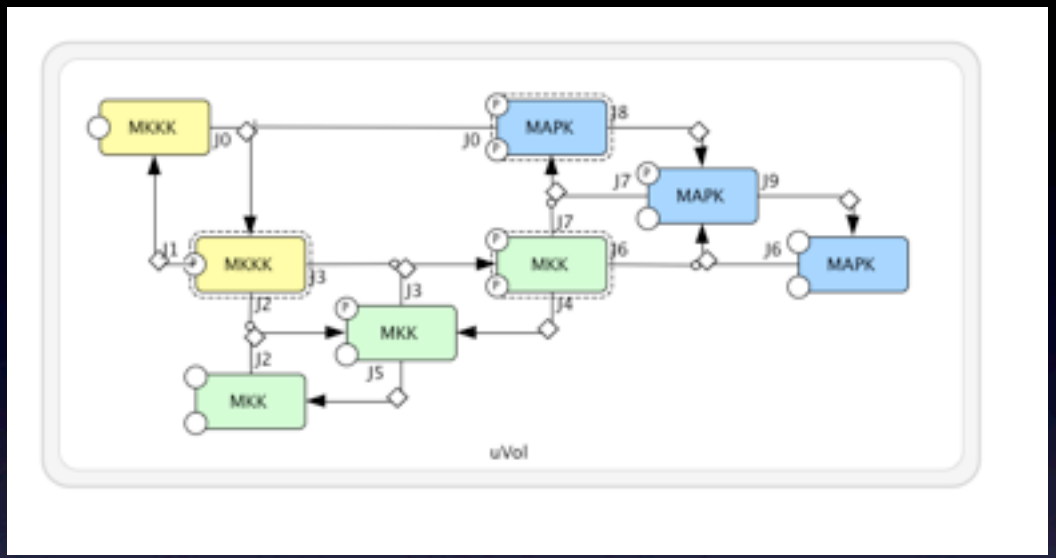
Auto Layout

- [File] → [Open] → samples/MAPK.xml
- [Layout] → [Orthogonal Layout]

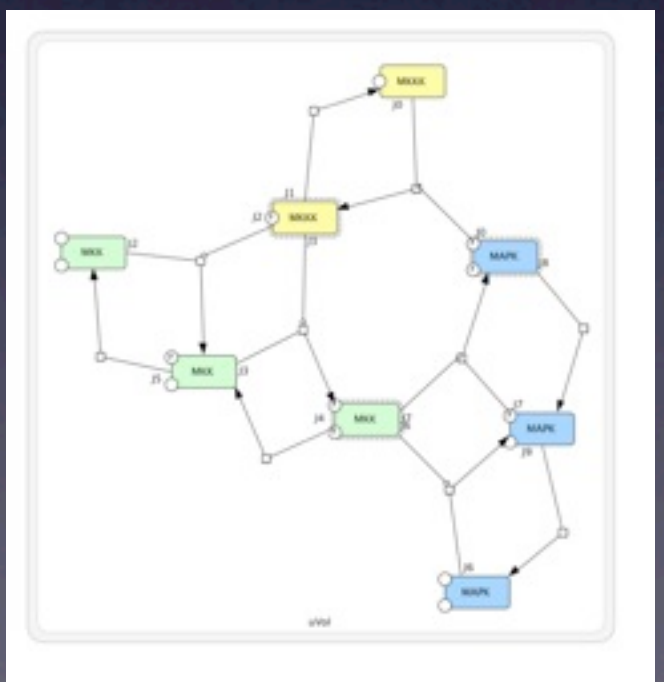


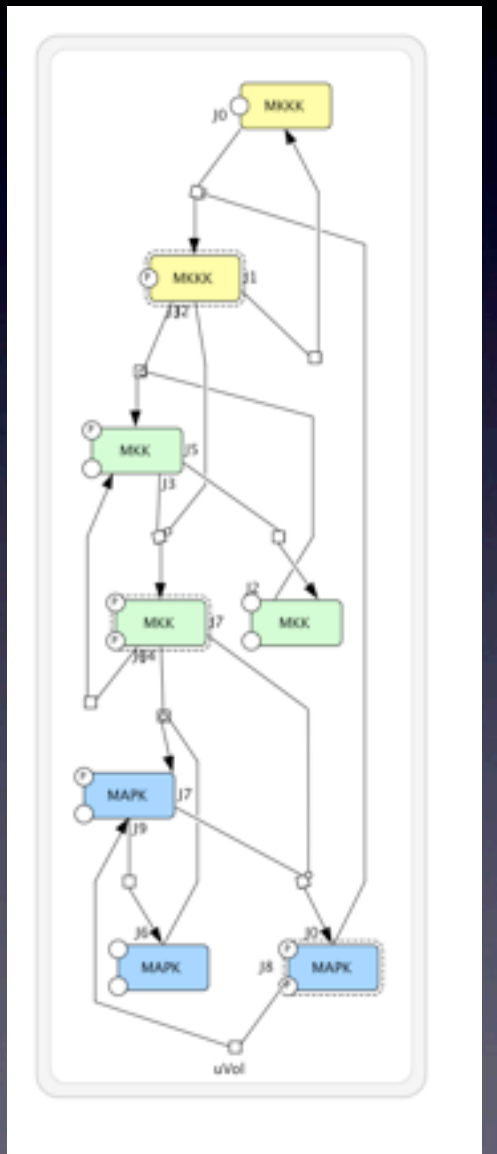


Auto Layout


MKKK
MAPK
MAPK
MKKK
MKK
MAPK
MKK
MKK
uVol

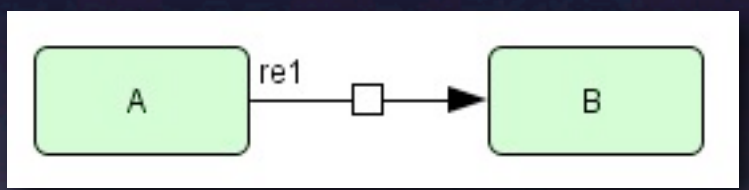

MAPK
MAPK
MKKK
MAPK
MKKK
MKK
MKK
MKK
uVol

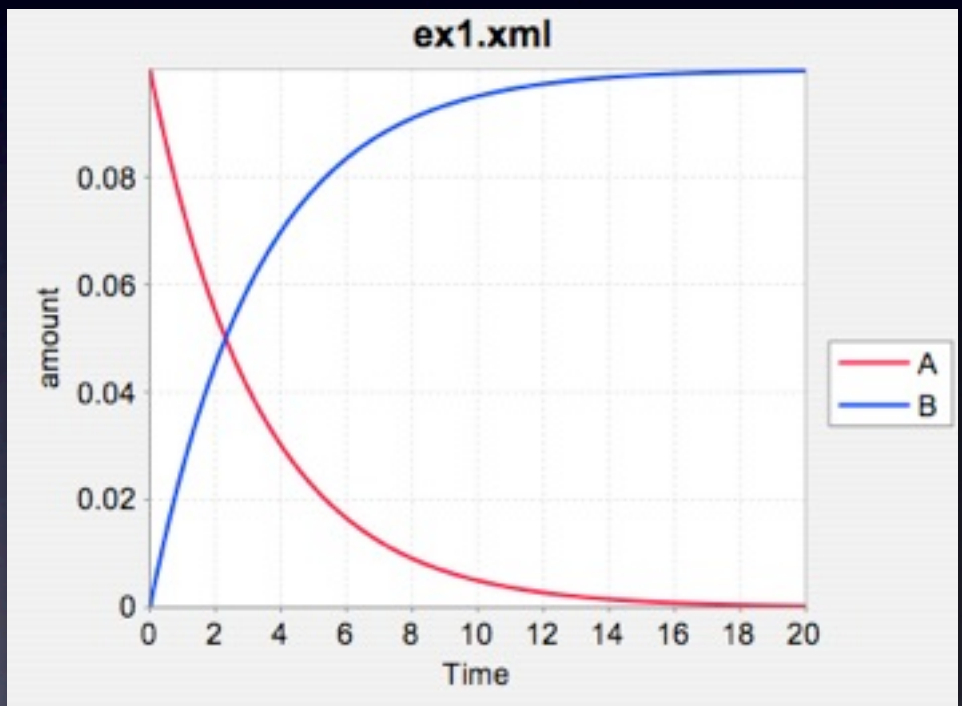



MKKK
MKKK
MKK
MKK
MKK
MAPK
MAPK
MAPK
uVol

Simulation (ex1)

- Create following biochemical reaction
- Click [Simulation] → [ControlPanel] and call SBML ODE Solver



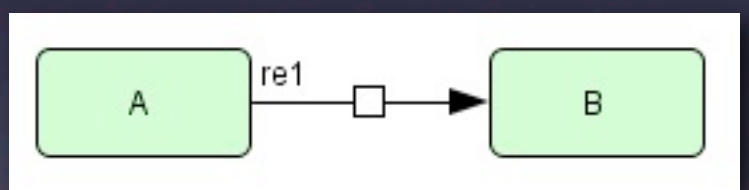


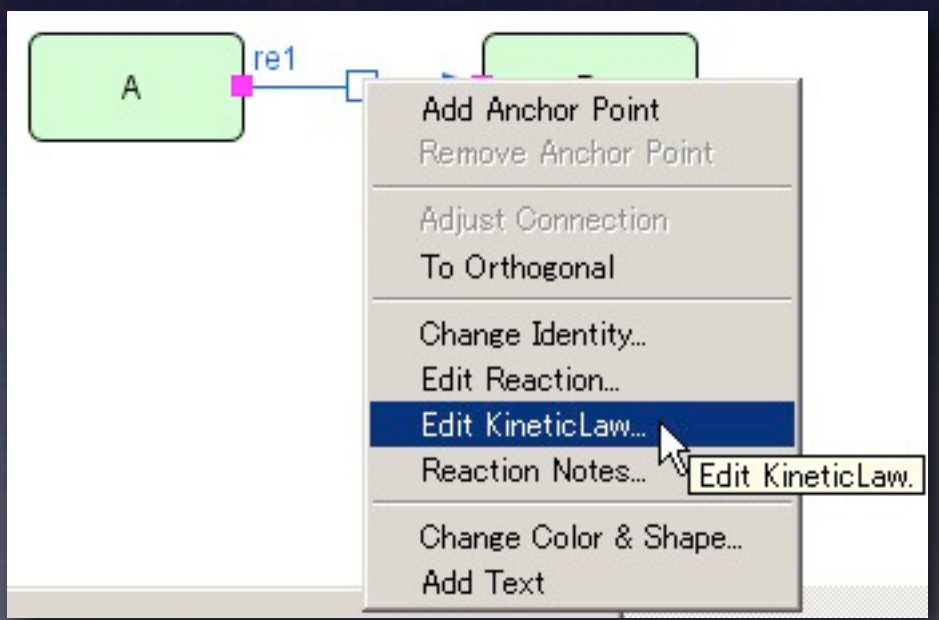
$$\frac{d[B]}{dt} = k[A]$$

k = 0.3
A = 0.1
B = 0

Simulation (ex1)

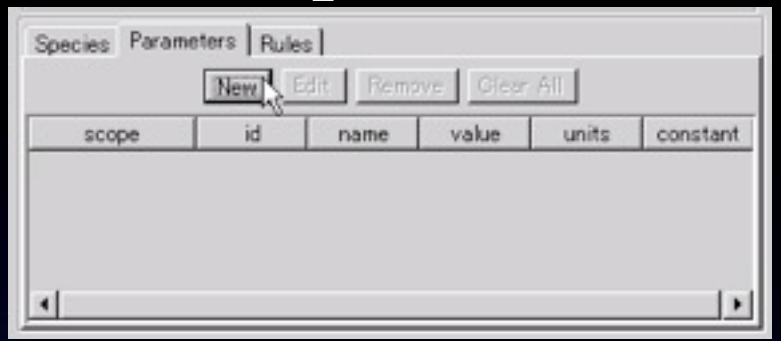
- Create new model (ex1)
- Create reaction
- Right click on the reaction and select [Edit KineticLaw...]



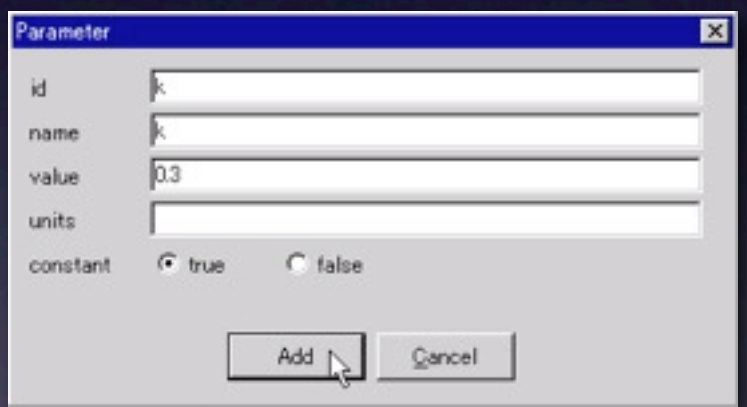


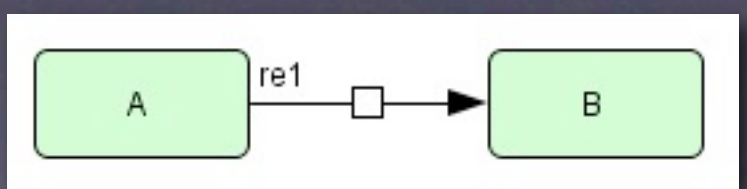
Simulation (ex1)

- Click [New] button on [Parameters] tab



- Input values as follows:
  - id: k
  - name: k
  - value: 0.3





$$\frac{d[B]}{dt} = k[A]$$

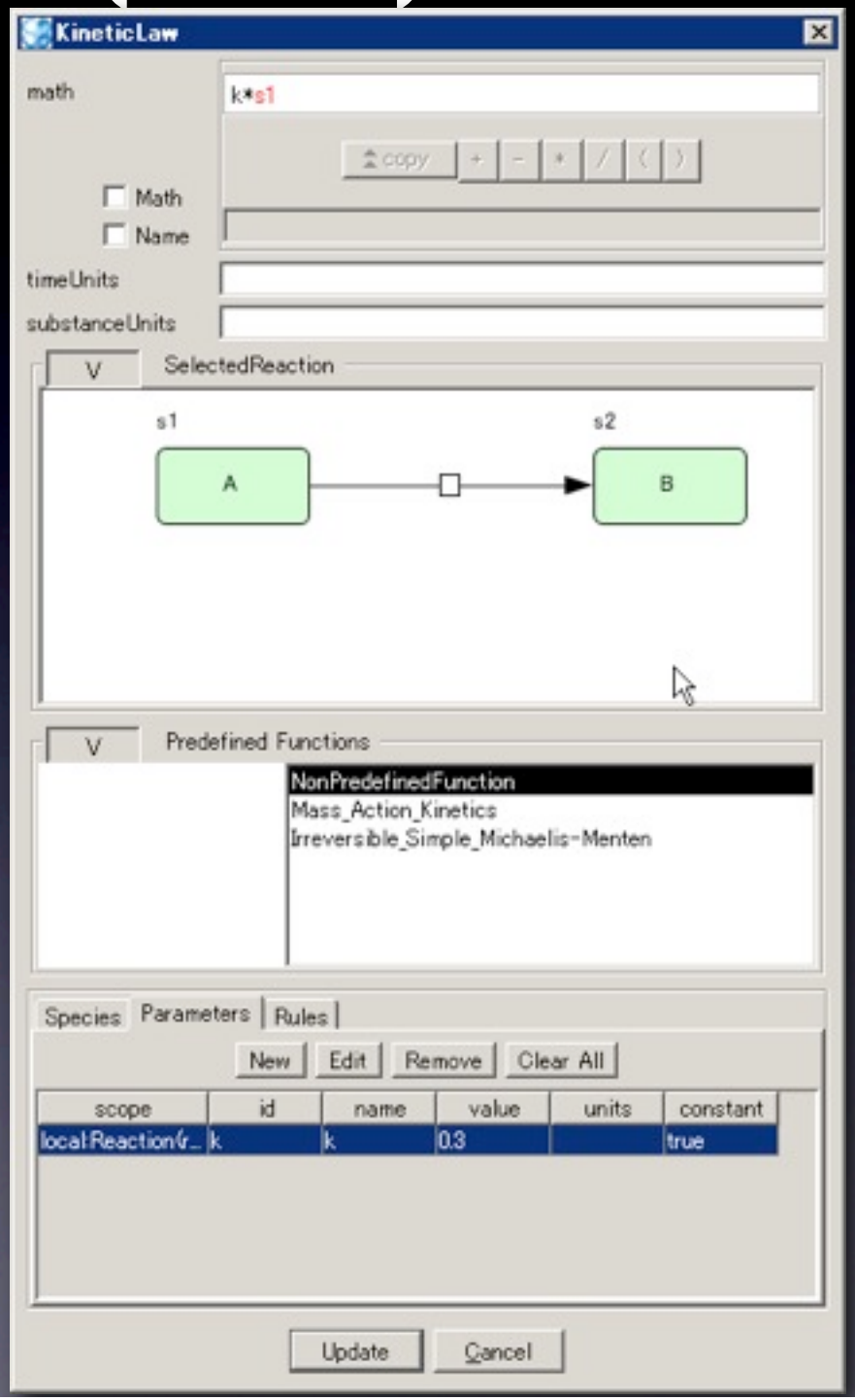
k = 0.3
A = 0.1
B = 0

Simulation (ex1)

- Select parameter “k”
- Click top most text field
- Click [copy] button
- Click [*] button
- Select Protein “A”
- Click top most text field
- Click [copy] button

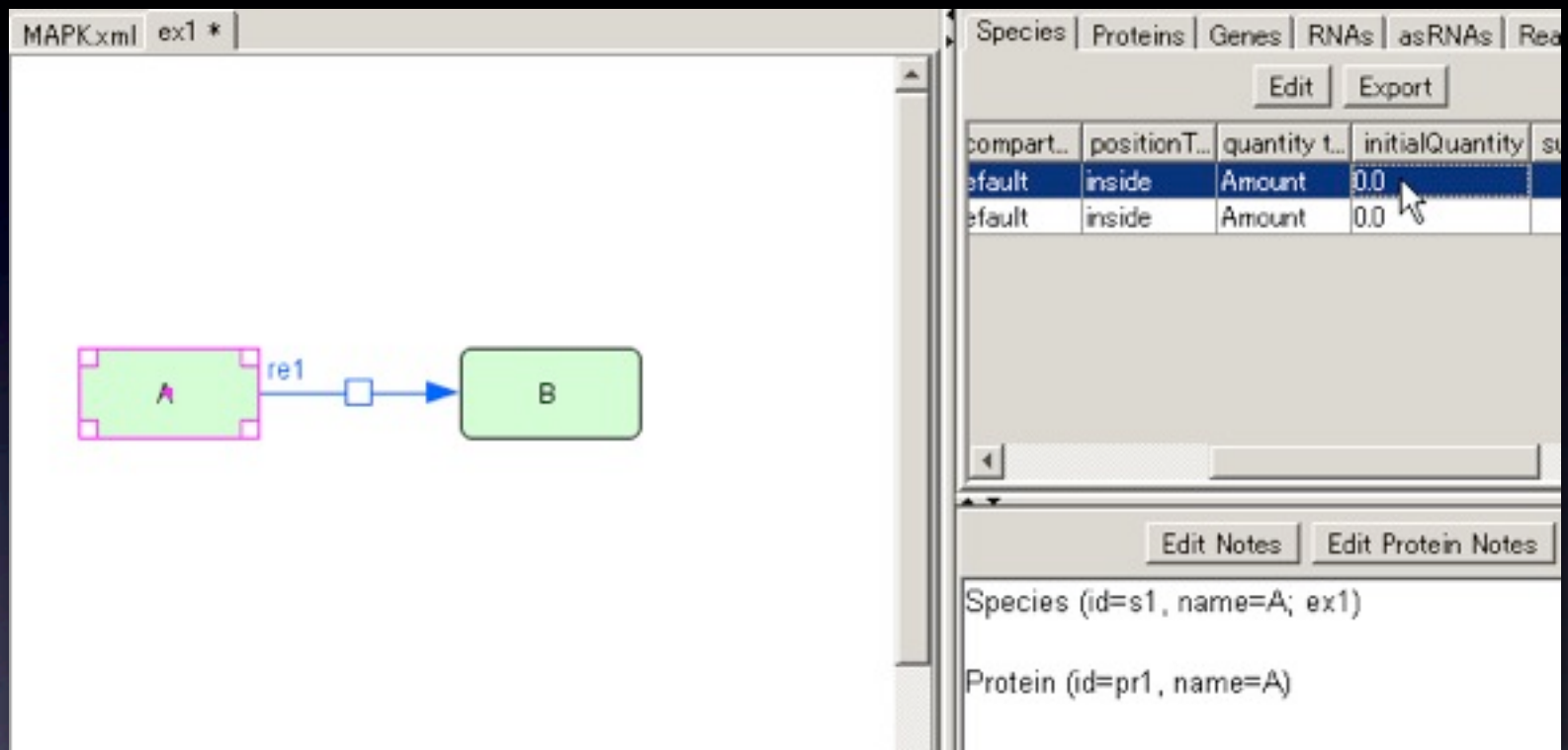
$$\frac{d[B]}{dt} = k[A]$$

k = 0.3
A = 0.1
B = 0



Simulation (ex1)

- Double click [initialQuantity] column for Protein “A”



- Set value as 0.1

$$d[B]/d[t] = k * [A]$$

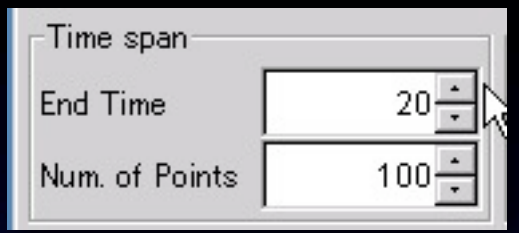
$k = 0.3$

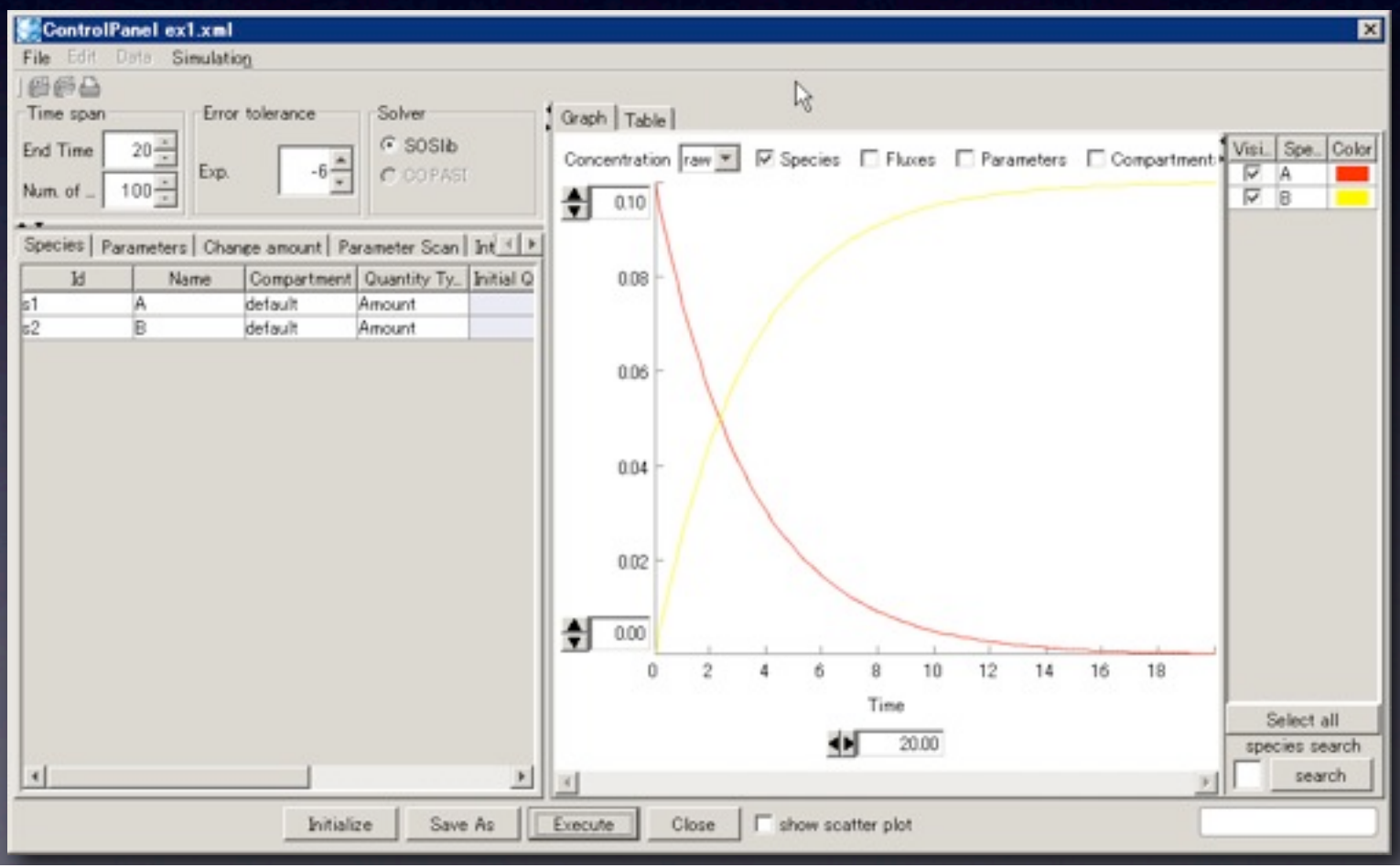
$A = 0.1$

$B = 0$

Simulation (ex1)

- Click [Simulation] → [ControlPanel]
- Set [End Time] to 20
- Click [Execute] button





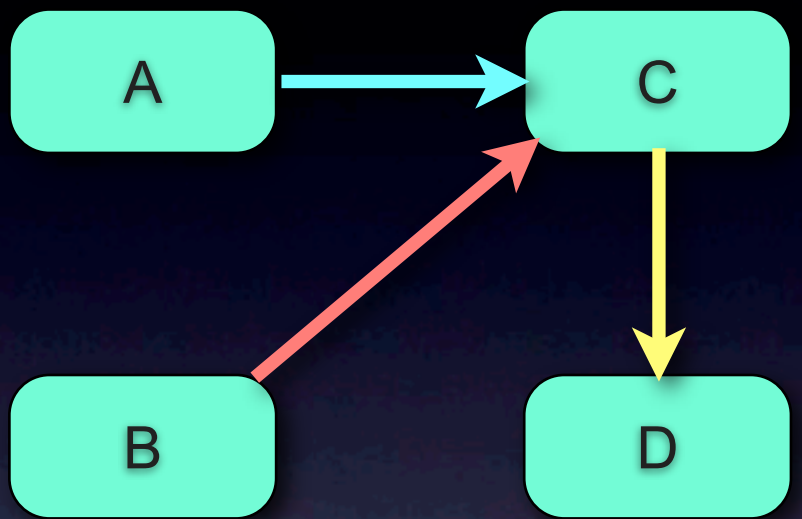
Equation → Network

$$\frac{dA}{dt} = -k_1 A$$

$$\frac{dB}{dt} = -k_2 B$$

$$\frac{dC}{dt} = k_1 A + k_2 B - k_3 C$$

$$\frac{dD}{dt} = k_3 C$$



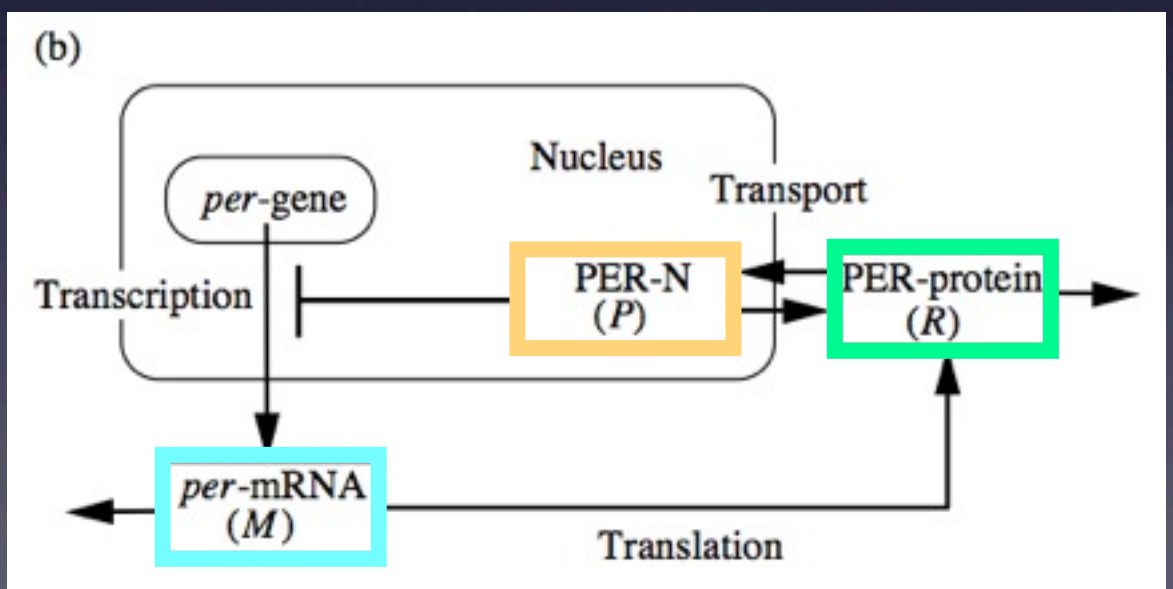
Circadian clock model

- Protein (P) inhibits transcription of mRNA (M)
- M is translated to Protein (R)
- P / R will be transported to cytosol / nucleus

$$\frac{dM}{dt} = \frac{1}{1 + (P/h)^n} - aM,$$

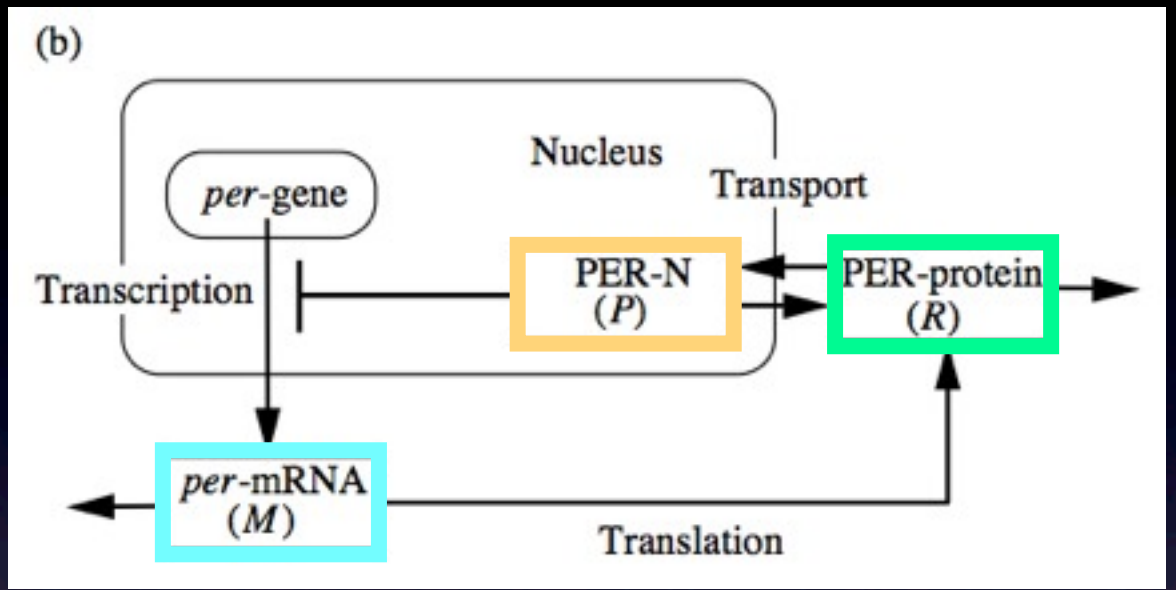
$$\frac{dR}{dt} = sM - (d + u)R + vP,$$

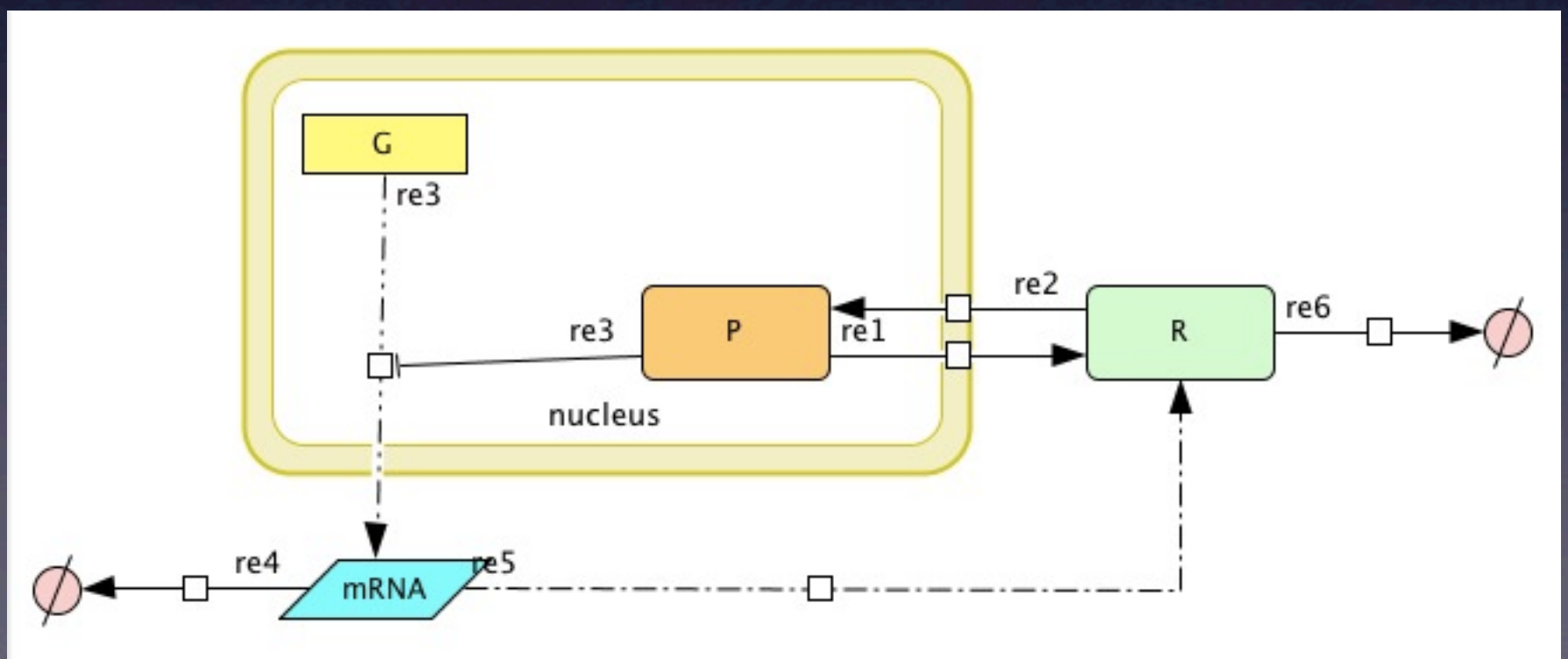
$$\frac{dP}{dt} = uR - vP.$$



J. theor. Biol. (2002) 216, 193–208

Circadian clock model





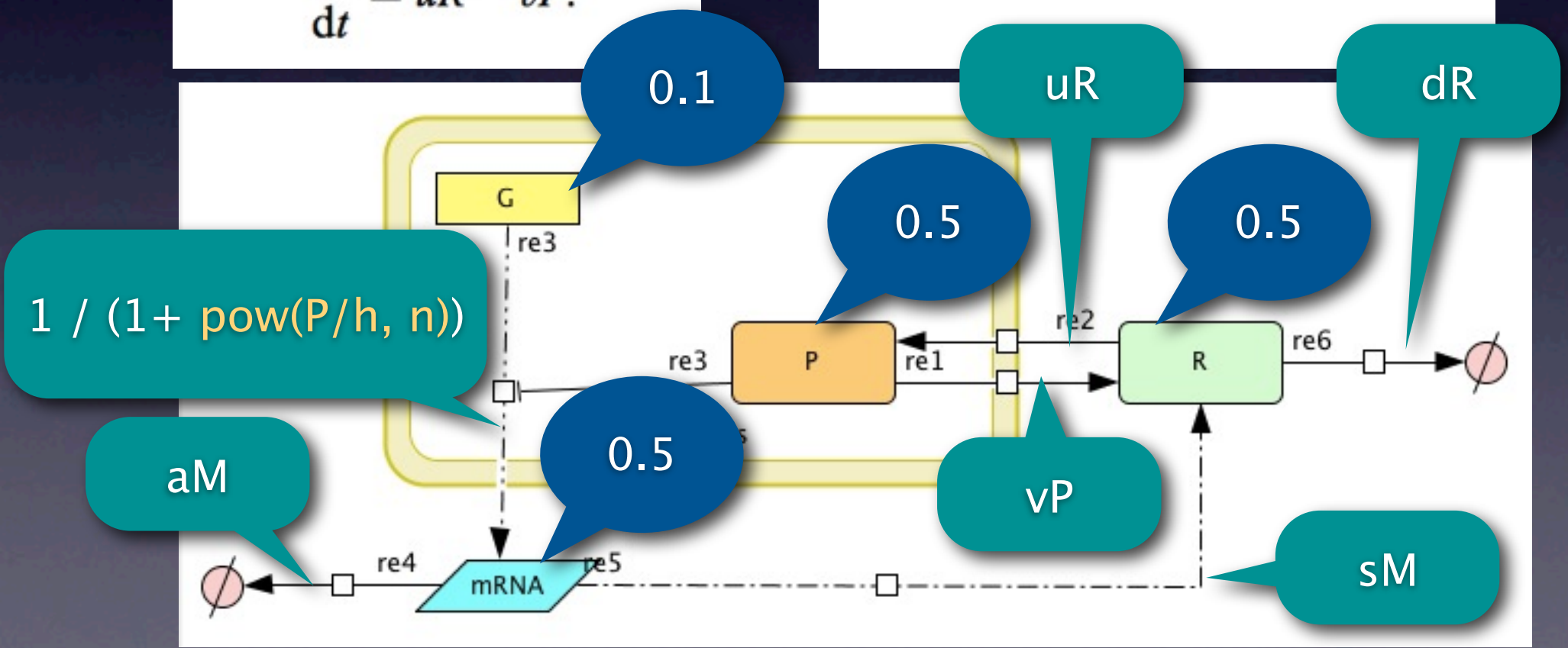
Circadian clock model

$$\frac{dM}{dt} = \frac{1}{1+(P/h)^n} - aM,$$

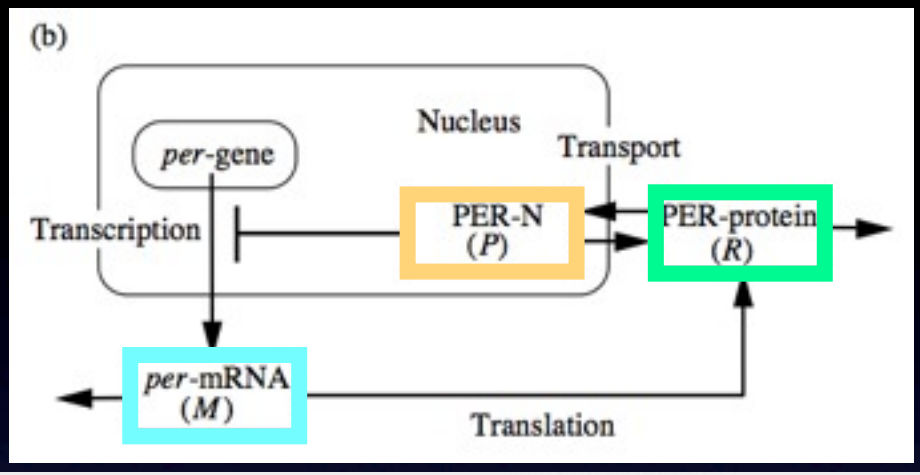
$$\frac{dR}{dt} = sM - (d+u)R + vP,$$

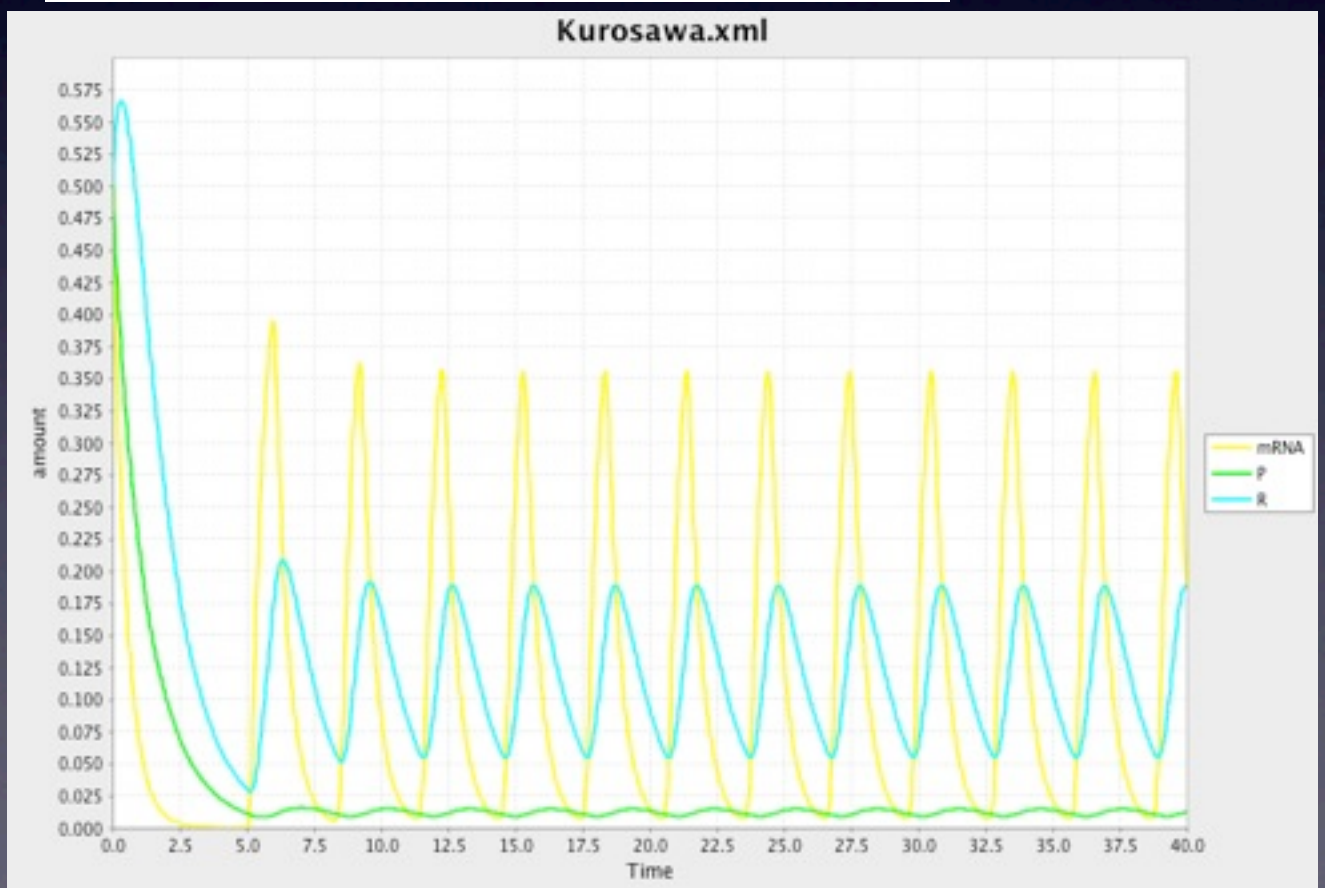
$$\frac{dP}{dt} = uR - vP.$$

$$\begin{aligned} a = s = d = v &= 1.0 \\ u &= 0.1 \\ h &= 0.01 \\ n &= 40 \end{aligned}$$

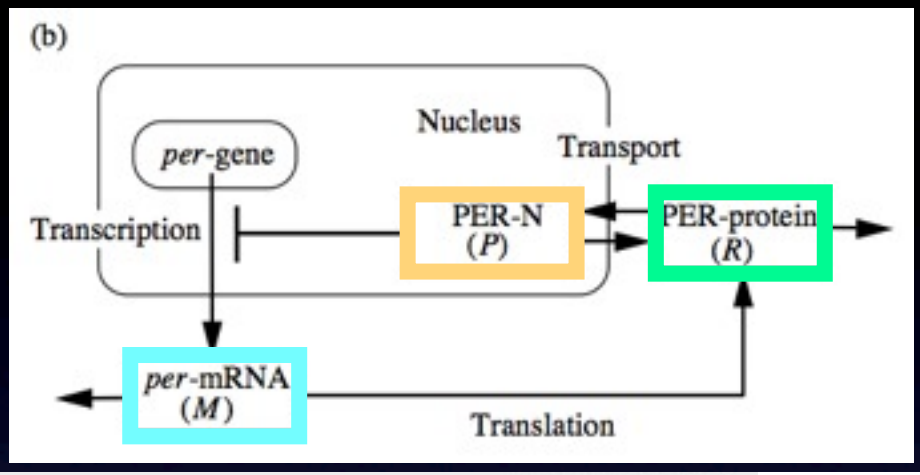


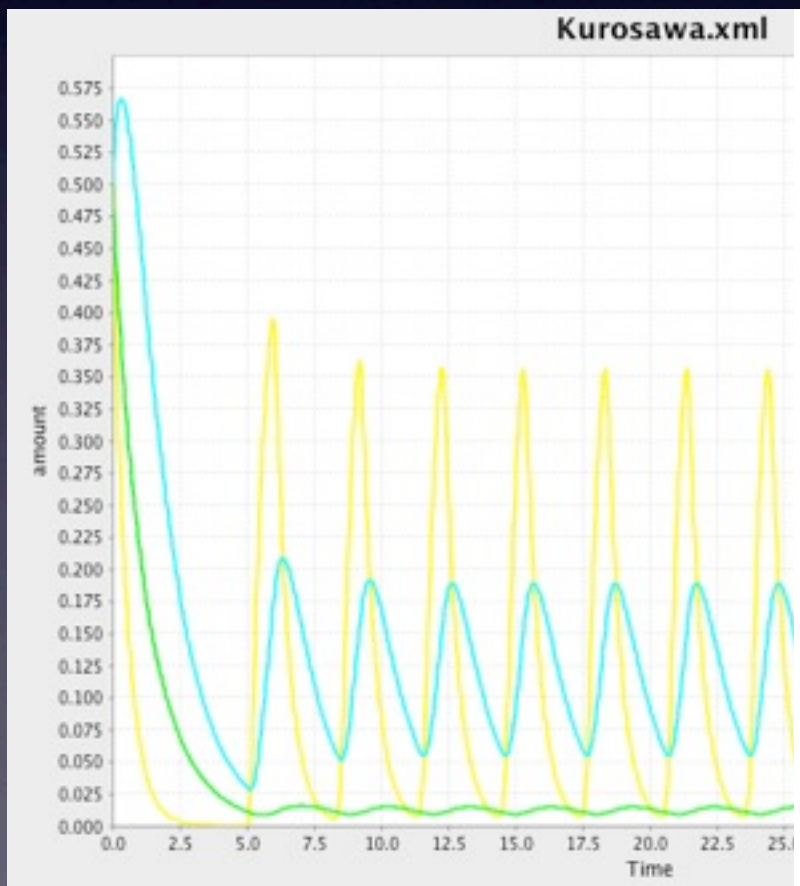
Circadian clock model

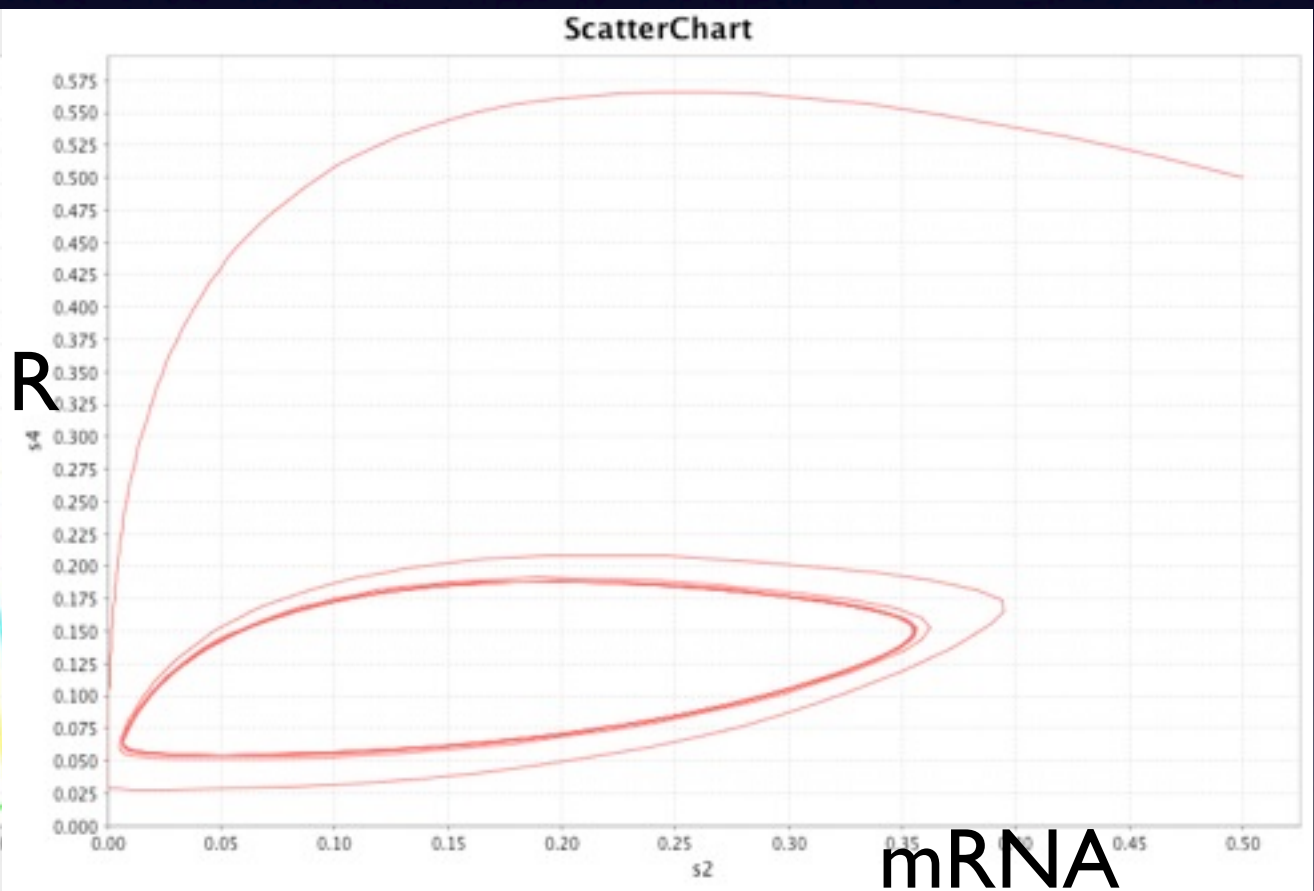




Circadian clock model







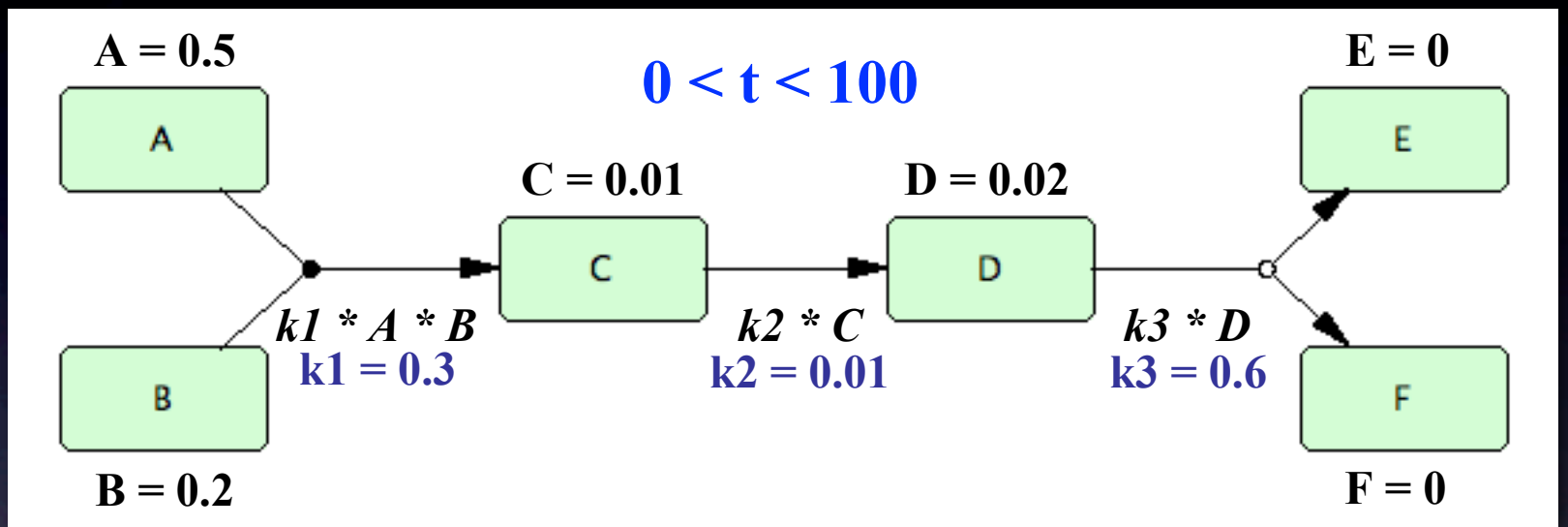
Summary

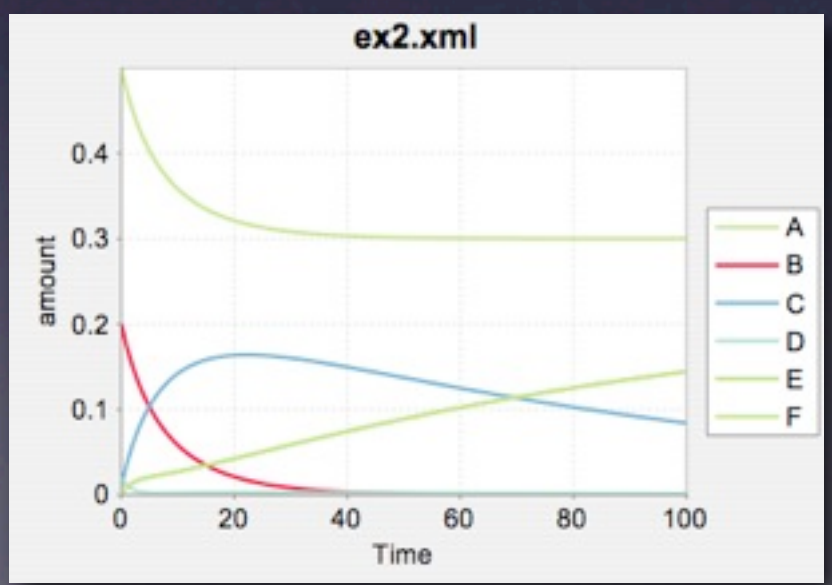
- Introduction of CellDesigner
  - What kind of model you can build
  - SBML (Systems Biology Markup Language)
  - SBGN (Graphical Notation)

- How to build a model with CellDesigner
  - Pathway map
  - Mathematical model

Simulation (ex2)

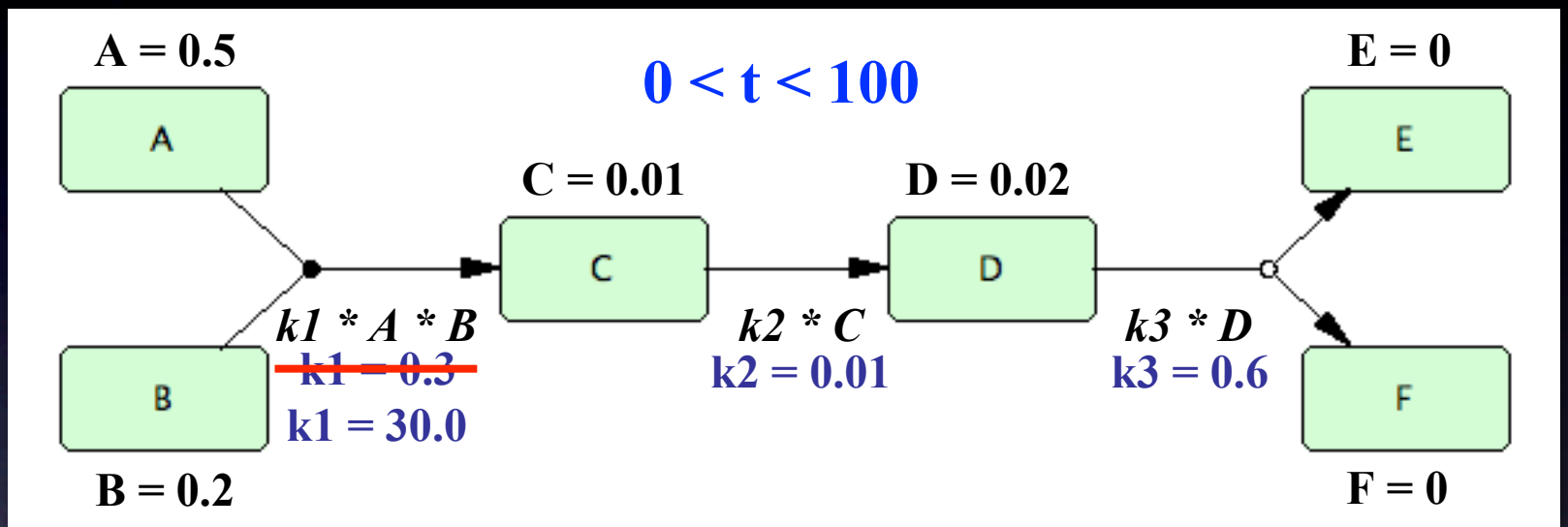
- Create following biochemical reactions
- Execute simulation from [ControlPanel]

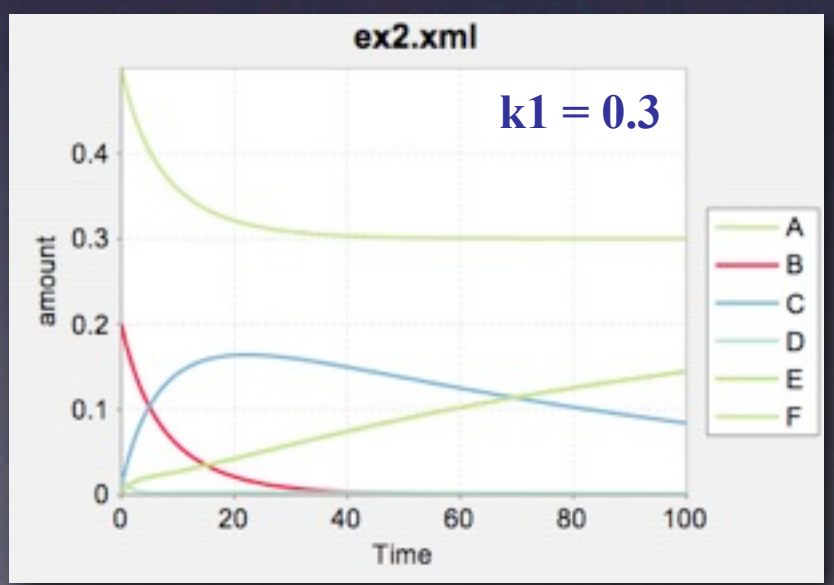


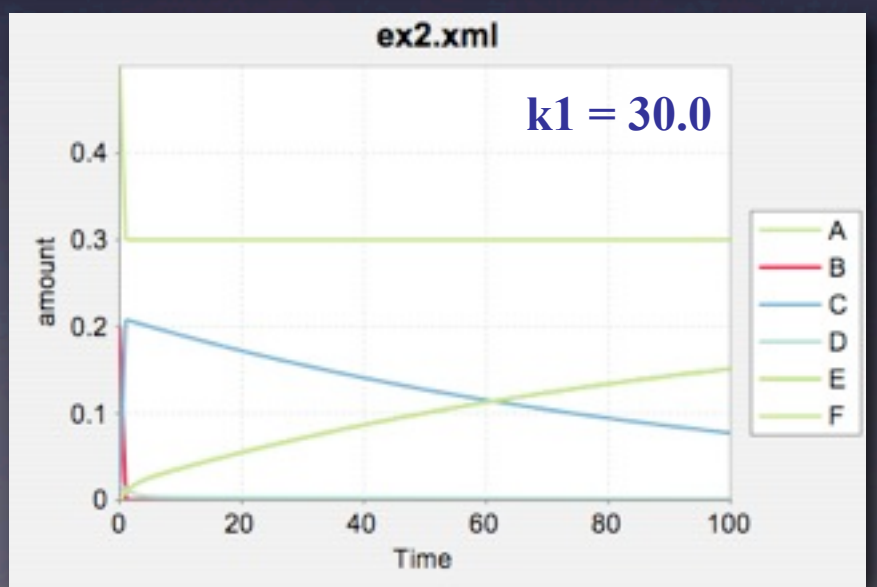


Simulation (ex2)

- Change parameter k1 to 30.0

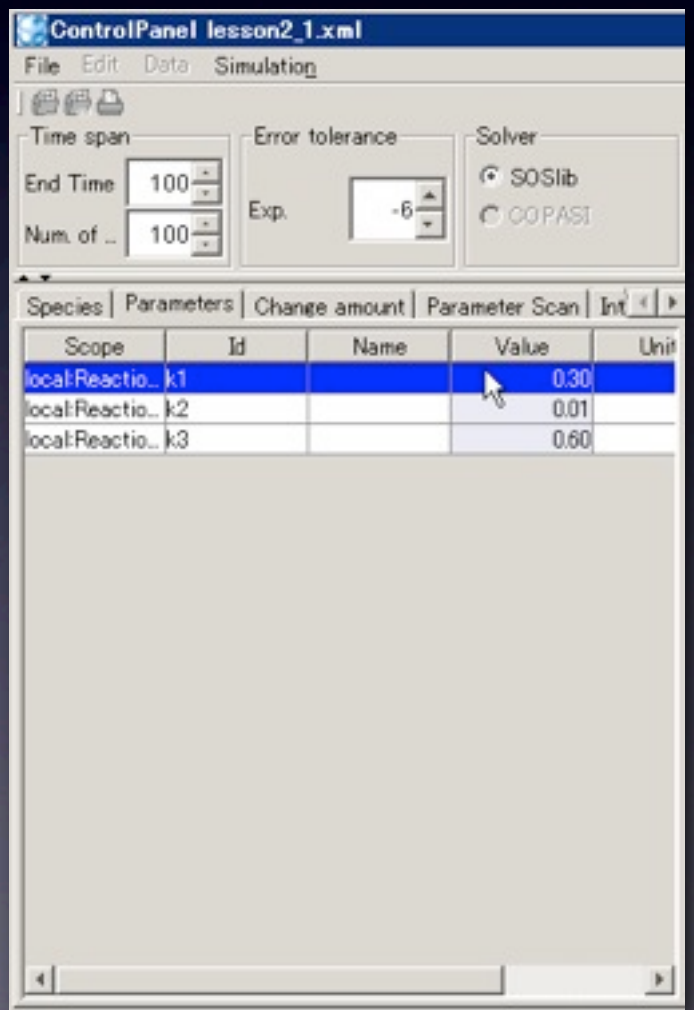


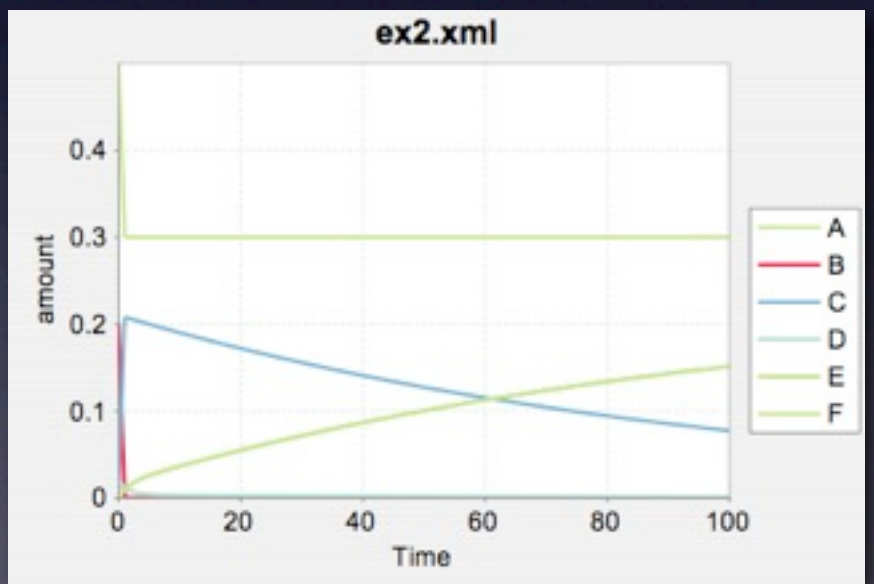




Simulation (ex2)

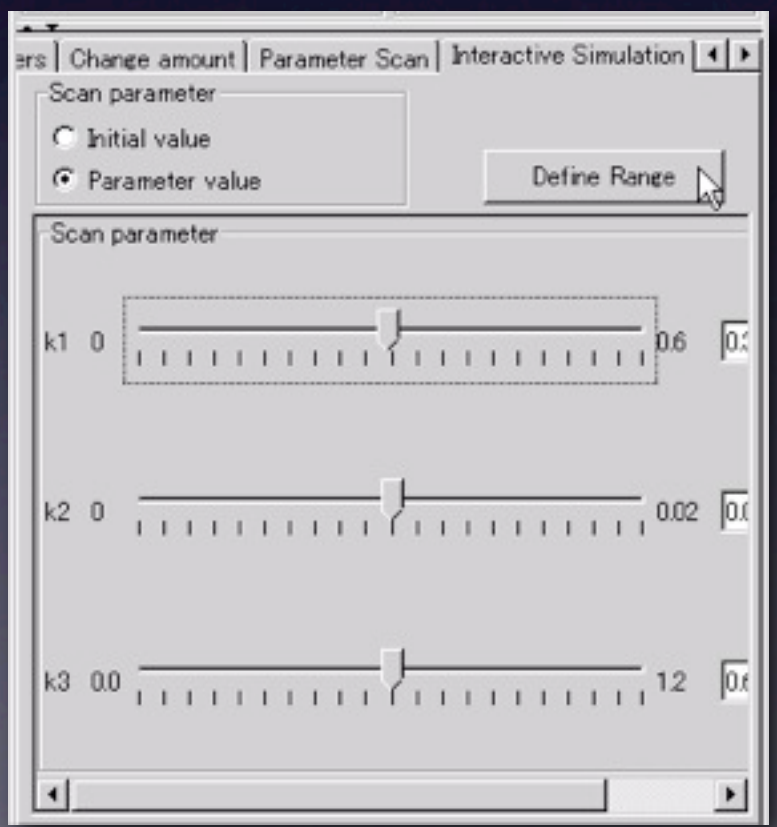
- Click [Parameters] tab
- Double click [Value] column for k1
- Change parameter k1 to 30.0

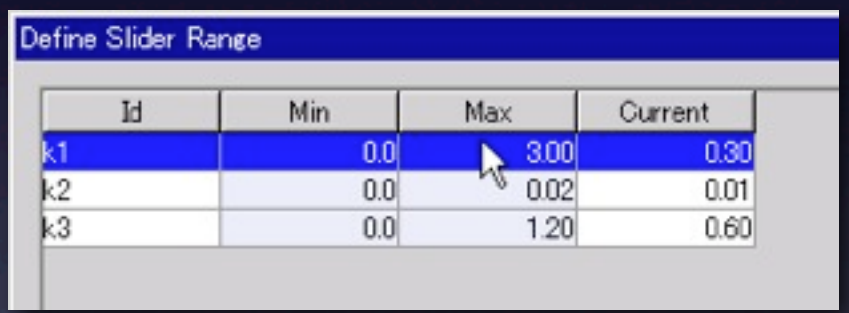




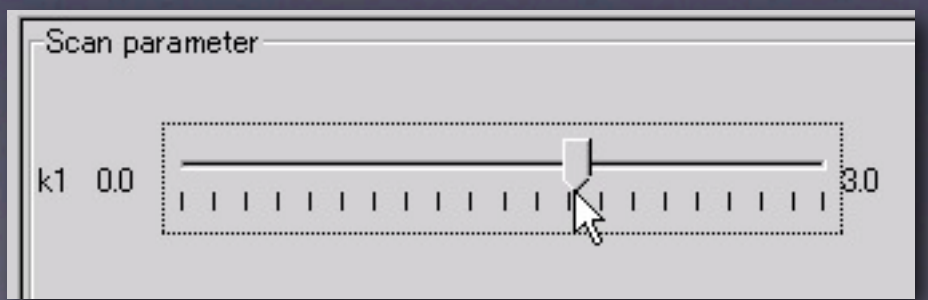
Simulation (ex2)

- Click [Interactive Simulation] tab
- Click [Parameter value] radio button
- Click [Define Range] button
- Click [Max] column for k1 and set value as 3.0



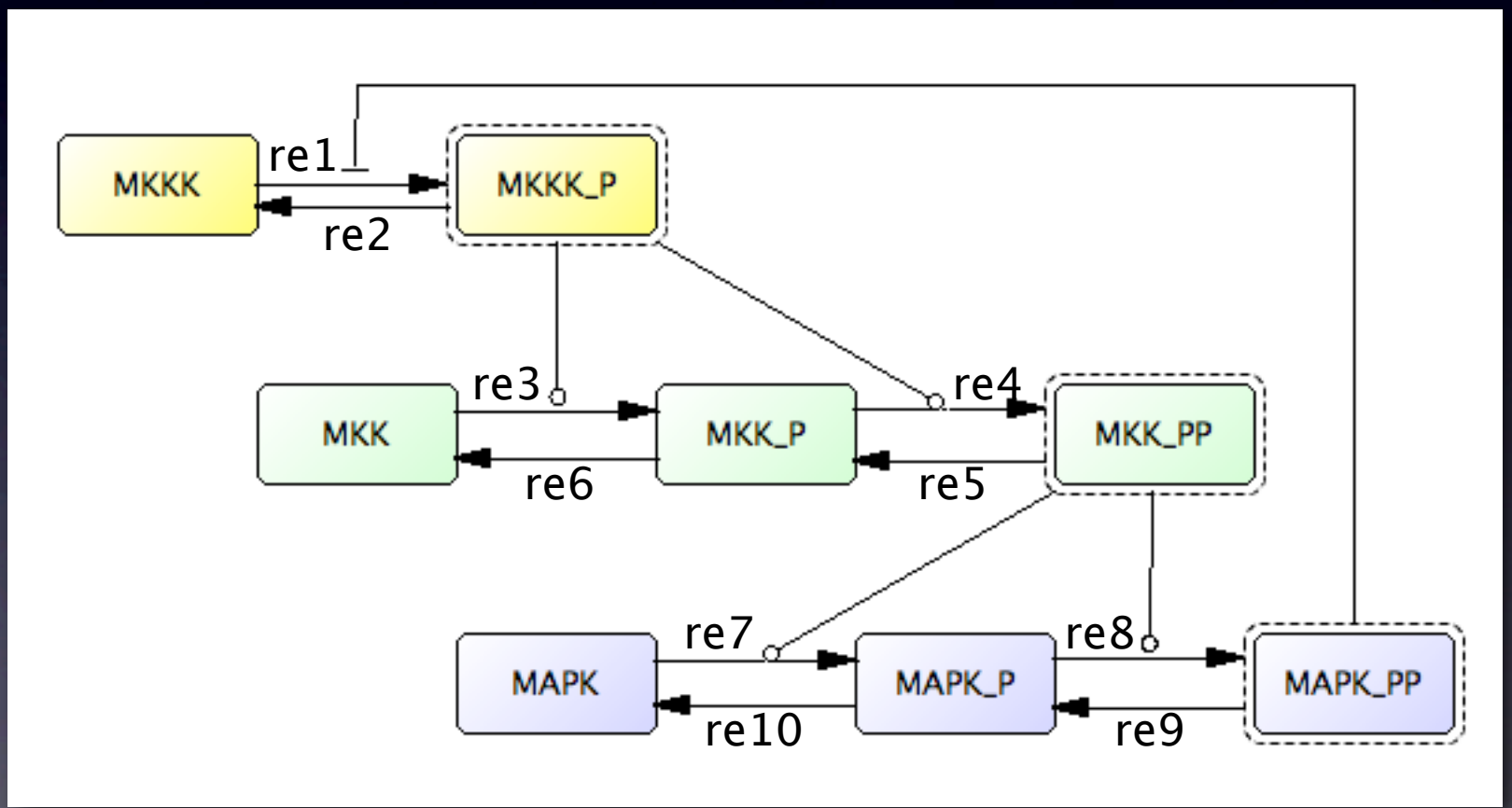


Drag sliderbar for k1



Exercise

- Create following model on CellDesigner



Kinetic Law

Reaction	Rate
re1	V1 * MKKK / ((1 + MAPK_PP / Ki) * (k1 + MKKK))
re2	V2 * MKKK_P / (KK2 + MKKK_P)
re3	k3 * MKKK_P * MKK / (KK3 + MKK)
re4	k4 * MKKK_P * MKK_P / (KK4 + MKK_P)
re5	V5 * MKK_PP / (KK5 + MKK_PP)
re6	V6 * MKK_P / (KK6 + MKK_P)
re7	k7 * MKK_PP * MAPK / (KK7 + MAPK)
re8	k8 * MKK_PP * MAPK_P / (KK8 + MAPK_P)
re9	V9 * MAPK_PP / (KK9 + MAPK_PP)
re10	V10 * MAPK_P / (KK10 + MAPK_P)

Initial Value & Parameters

Species	value
MKKK	90
MKKK_P	10
MKK	280
MKK_P	10
MKK_PP	10
MAPK	280
MAPK_P	10
MAPK_PP	10

Parameter	value
V1	2.5
Ki	9.0
k1	10.0
V2	0.25
KK2	8.0
k3	0.025
KK3	15.0
k4	0.025
KK4	15.0
V5	0.75
KK5	15.0

Parameter	value
V6	0.75
KK6	15.0
k7	0.025
KK7	15.0
k8	0.025
KK8	15.0
V9	0.5
KK9	15.0
V10	0.5
KK10	15.0

Simulation Result

- End Time: 4000