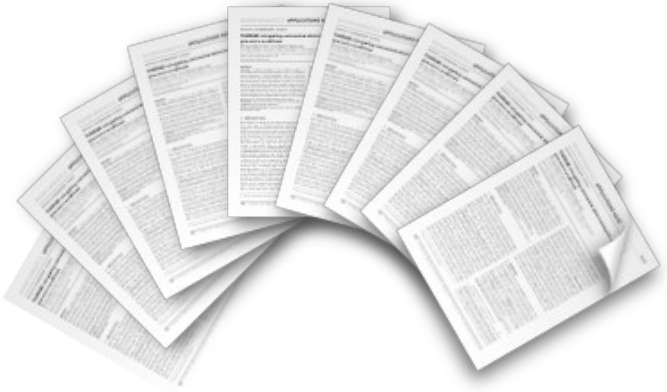
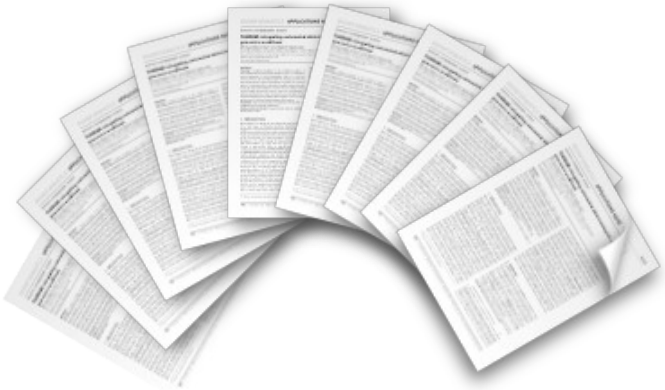
EMBL-EBI

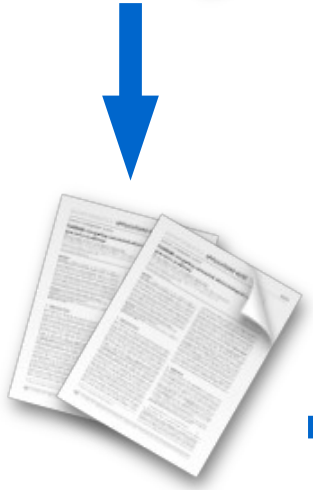
Large-scale generation of mathematical models from biological pathways

Nicolas Le Novère
EMBL-European Bioinformatics Institute
Babraham Institute

4th INCF Neuroinformatics conference, 10-12 Sept 2012, Munich



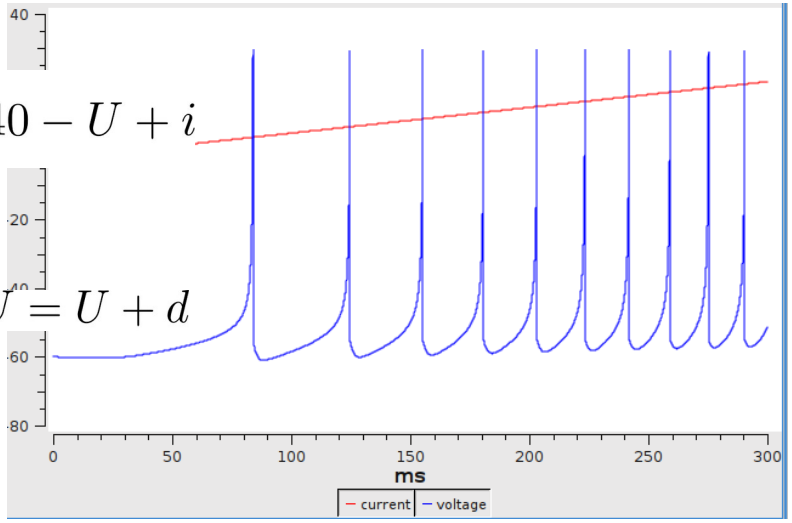


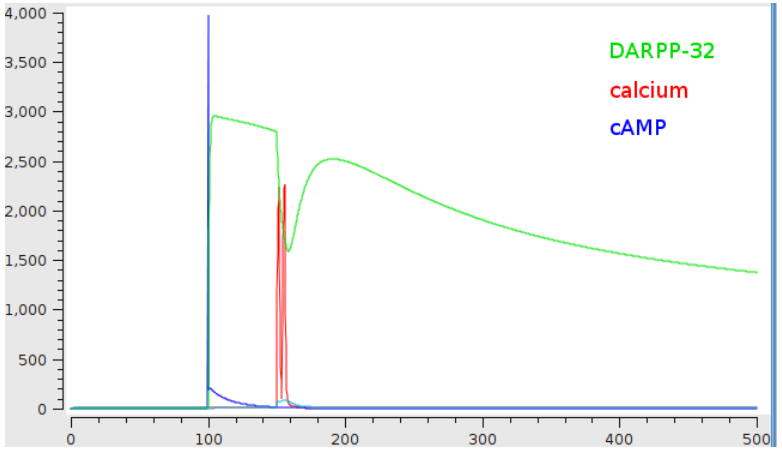


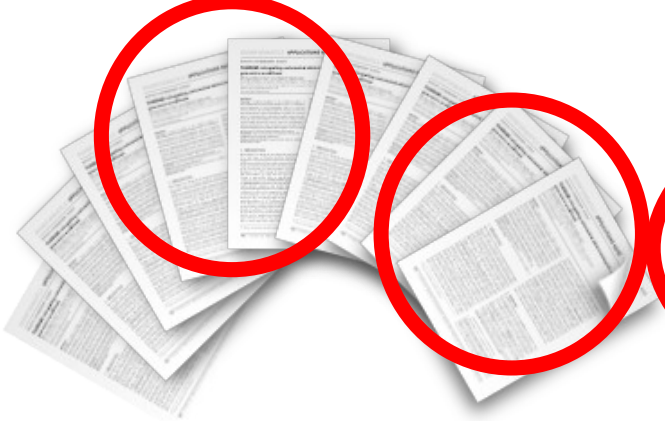
$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - U + i$$

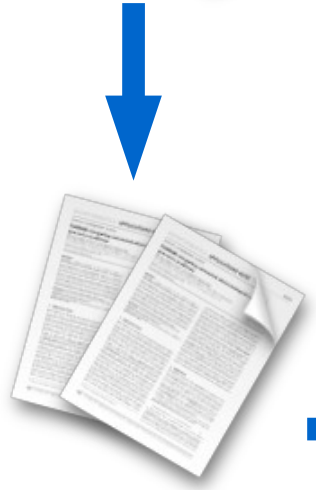
$$\frac{dU}{dt} = a(bv - U)$$

$$v > V_{\text{thresh}} \Rightarrow v = c; U = U + d$$





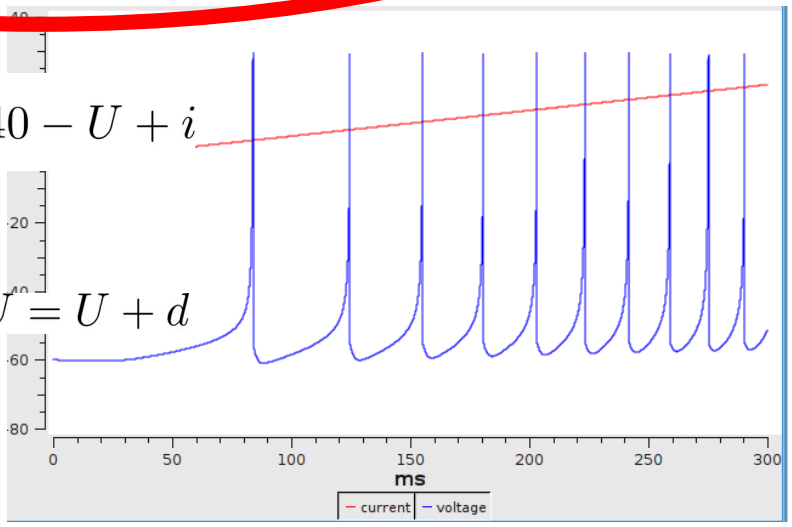


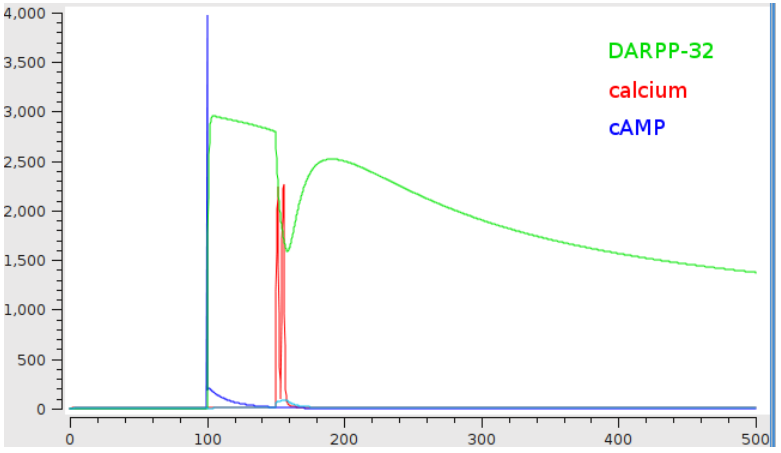


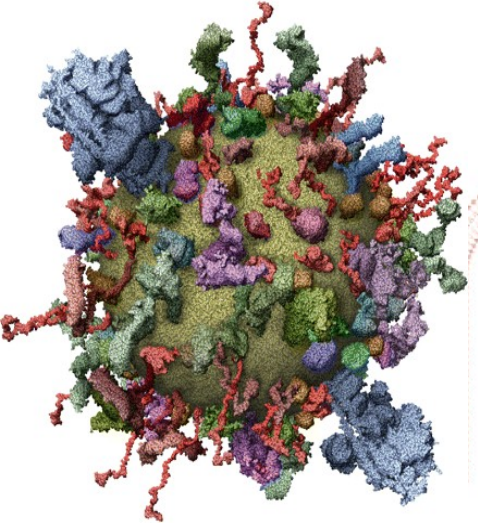
$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - U + i$$

$$\frac{dU}{dt} = a(bv - U)$$

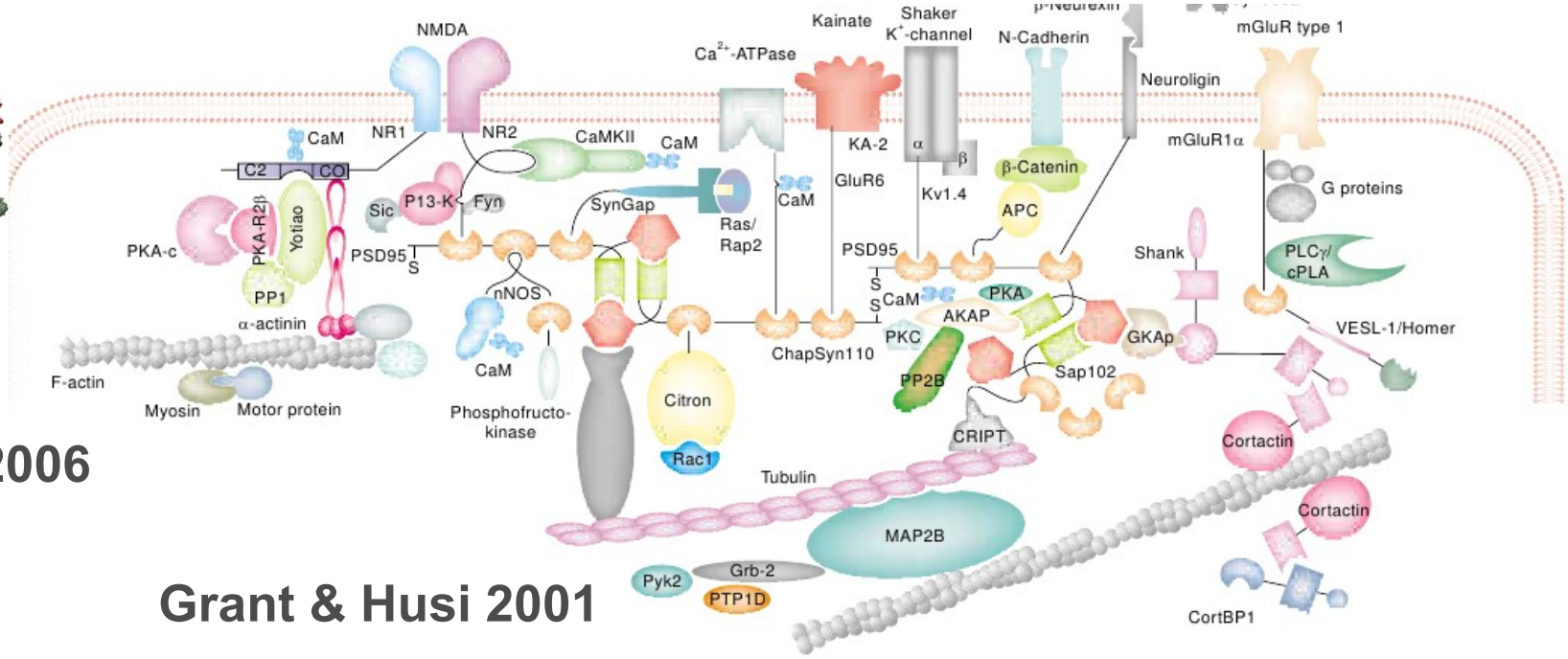
$$v > V_{\text{thresh}} \Rightarrow v = c; U = U + d$$



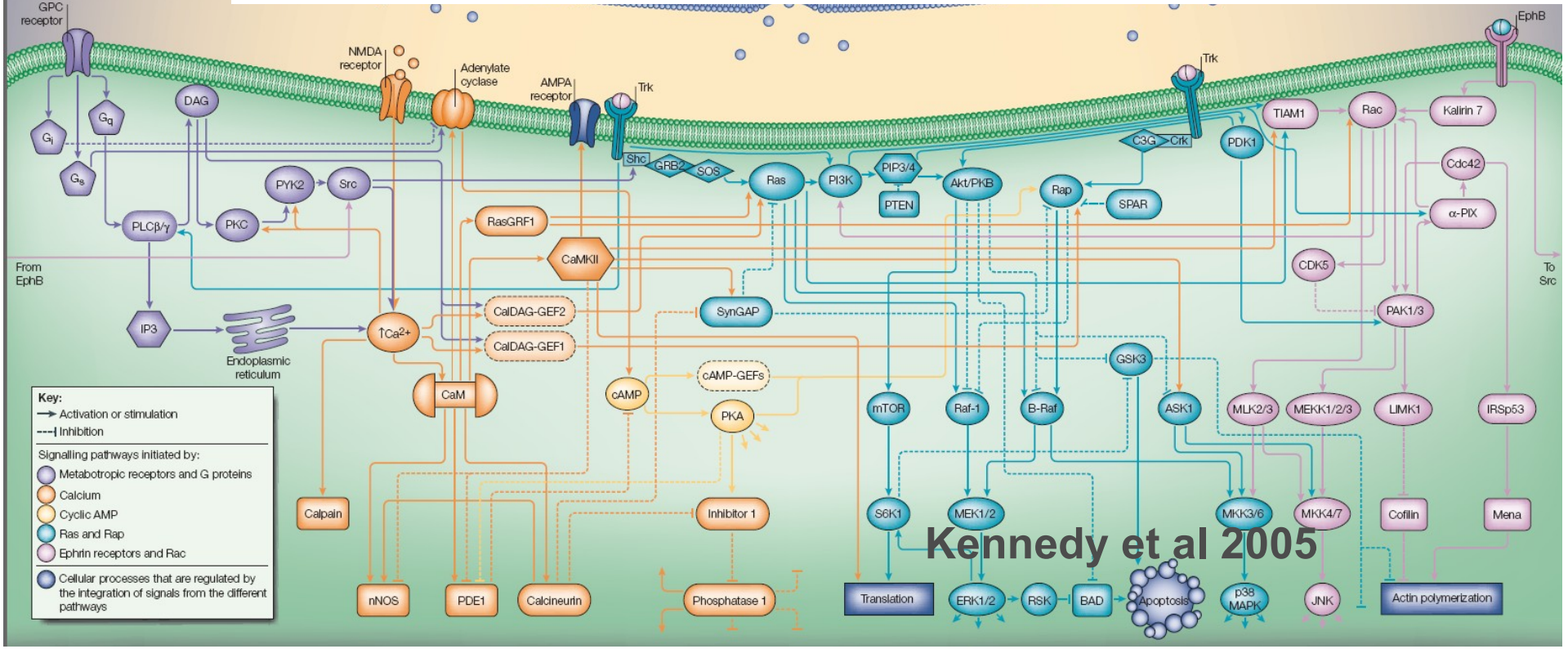




Takamori et al 2006



Grant & Husi 2001



Kennedy et al 2005

Models are becoming larger

Theory

Cell

A Whole-Cell Computational Model Predicts Phenotype from Genotype

Jonathan R. Karr,[1,4] Jayodita C. Sanghvi,[2,4] Derek N. Macklin,[2] Miriam V. Gutschow,[2] Jared M. Jacobs,[2] Benjamin Bolival, Jr.,[2] Nacyra Assad-Garcia,[3] John I. Glass,[3] and Markus W. Covert[2,*]

[1]Graduate Program in Biophysics
[2]Department of Bioengineering
Stanford University, Stanford, CA 94305, USA
[3]J. Craig Venter Institute, Rockville, MD 20850, USA
[4]These authors contributed equally to this work
*Correspondence: mcovert@stanford.edu
http://dx.doi.org/10.1016/j.cell.2012.05.044

SUMMARY

Understanding how complex phenotypes arise from individual molecules and their interactions is a primary challenge in biology that computational approaches are poised to tackle. We report a whole-cell computational model of the life cycle of the human pathogen *Mycoplasma genitalium* that includes all of its molecular components and their interactions. An integrative approach to modeling that combines diverse mathematics enabled the simultaneous inclusion of fundamentally different cellular processes and experimental measurements. Our whole-cell model accounts for all annotated gene functions and was validated against a broad

First, until recently, not enough has been known about the individual molecules and their interactions to completely model any one organism. The advent of genomics and other high-throughput measurement techniques has accelerated the characterization of some organisms to the extent that comprehensive modeling is now possible. For example, the mycoplasmas, a genus of bacteria with relatively small genomes that includes several pathogens, have recently been the subject of an exhaustive experimental effort by a European consortium to determine the transcriptome (Güell et al., 2009), proteome (Kühner et al., 2009), and metabolome (Yus et al., 2009) of these organisms.

The second limiting factor has been that no single computational method is sufficient to explain complex phenotypes in terms of molecular components and their interactions. The first approaches to modeling cellular physiology, based on ordinary differential equations (ODEs) (Atlas et al., 2008; Browning et al., 2004; Castellanos et al., 2004, 2007; Domach et al.,

Large-scale community curated pathways

http://www.nature.com/naturebiotechnology

A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology

Markus J Herrgård[1,19,20], Neil Swainston[2,3,20], Paul Dobson[3,4], Warwick B Dunn[3,4], K Yalçin Arga[5], Mikko Arvas[6], Nils Blüthgen[3,7], Simon Borger[8], Roeland Costenoble[9], Matthias Heinemann[9], Michael Hucka[10], Nicolas Le Novère[11], Peter Li[2,3], Wolfram Liebermeister[8], Monica L Mo[1], Ana Paula Oliveira[12], Dina Petranovic[12,19], Stephen Pettifer[2,3], Evangelos Simeonidis[3,7], Kieran Smallbone[3,13], Irena Spasić[2,3], Dieter Weichart[3,4], Roger Brent[14], David S Broomhead[3,13], Hans V Westerhoff[3,7,15], Betül Kırdar[5], Merja Penttilä[6], Edda Klipp[8], Bernhard Ø Palsson[1], Uwe Sauer[9], Stephen G Oliver[3,16], Pedro Mendes[2,3,17], Jens Nielsen[12,18] & Douglas B Kell*[3,4]

NATURE BIOTECHNOLOGY VOLUME 26 NUMBER 10 OCTOBER 2008

2153 molecular species

1857 reactions

15 compartments

5928 cross-references

Many spin-off studies

5th iteration: 2573 species, 2110 reactions

What is a “pathway”

Wikipedia (May 29th 2012): “In biochemistry, metabolic pathways are series of chemical reactions occurring within a cell. In each pathway, a principal chemical is modified by a series of chemical reactions. Enzymes catalyze these reactions [...]”

Different types: Signalling pathways, metabolic networks, gene regulatory networks ...

Many “pathway” databases:

Biocarta, Bio/MetaCyc, Ingenuity IPA, KEGG Pathway, Panther pathways, Reactome, STKE, Wikipathways etc.

➜Detailed representation of reality based on observation

What is a “model”

Wikipedia (May 29th 2012): “A mathematical model is a description of a system using mathematical concepts and language.”

A model is made up of variables, functions and constraints

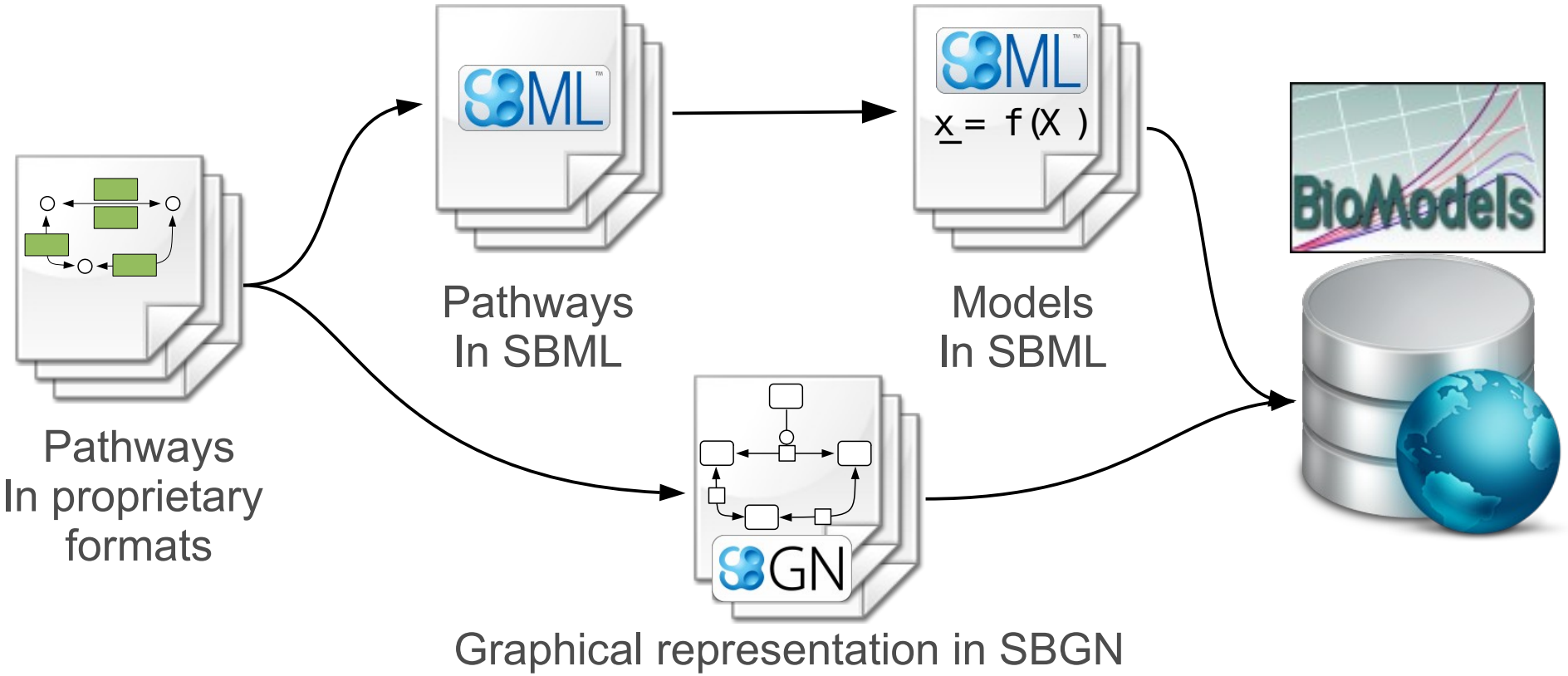
Different types: Dynamical models, logical models, rule-based models, multi-agent models, statistical models ...

Different levels of granularity for the variables and precision for the functions based on the questions being asked and the data available

- Abstract representation of reality based on needs

Aim of the project

- To provide a starting point to model as many biochemical pathways as possible in as many species as possible
- To provide pathways in a standard format readable by most systems biology software



The Systems Biology Markup Language

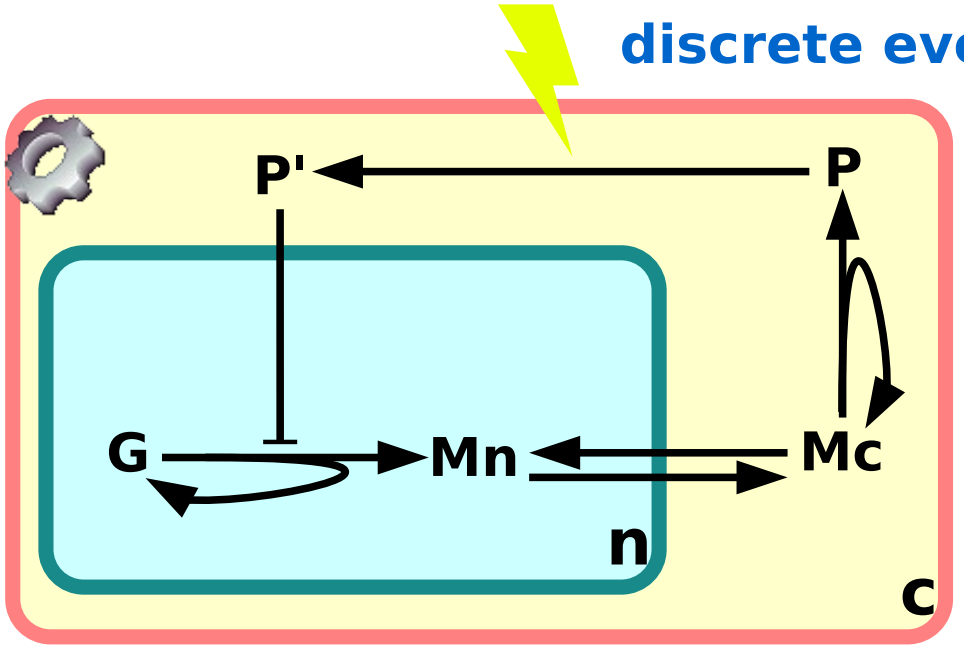
```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="3" version="1"·
      xmlns="http://www.sbml.org/sbml/level3/version1/core">
  <model>
    <listOfFunctionDefinitions> <!-- --> </listOfFunctionDefinitions>
    <listOfUnitDefinitions> <!-- --> </listOfUnitDefinitions>
    <listOfCompartments> <!-- --> </listOfCompartments>
    <listOfSpecies> <!-- --> </listOfSpecies>
    <listOfParameters> <!-- --> </listOfParameters>
    <listOfInitialAssignments> <!-- --> </listOfInitialAssignments>
    <listOfRules> <!-- --> </listOfRules>
    <listOfConstraints> <!-- --> </listOfConstraints>
    <listOfReactions> <!-- --> </listOfReactions>
    <listOfEvents> <!-- --> </listOfEvents>
  </model>
</sbml>
```

variables

relationships

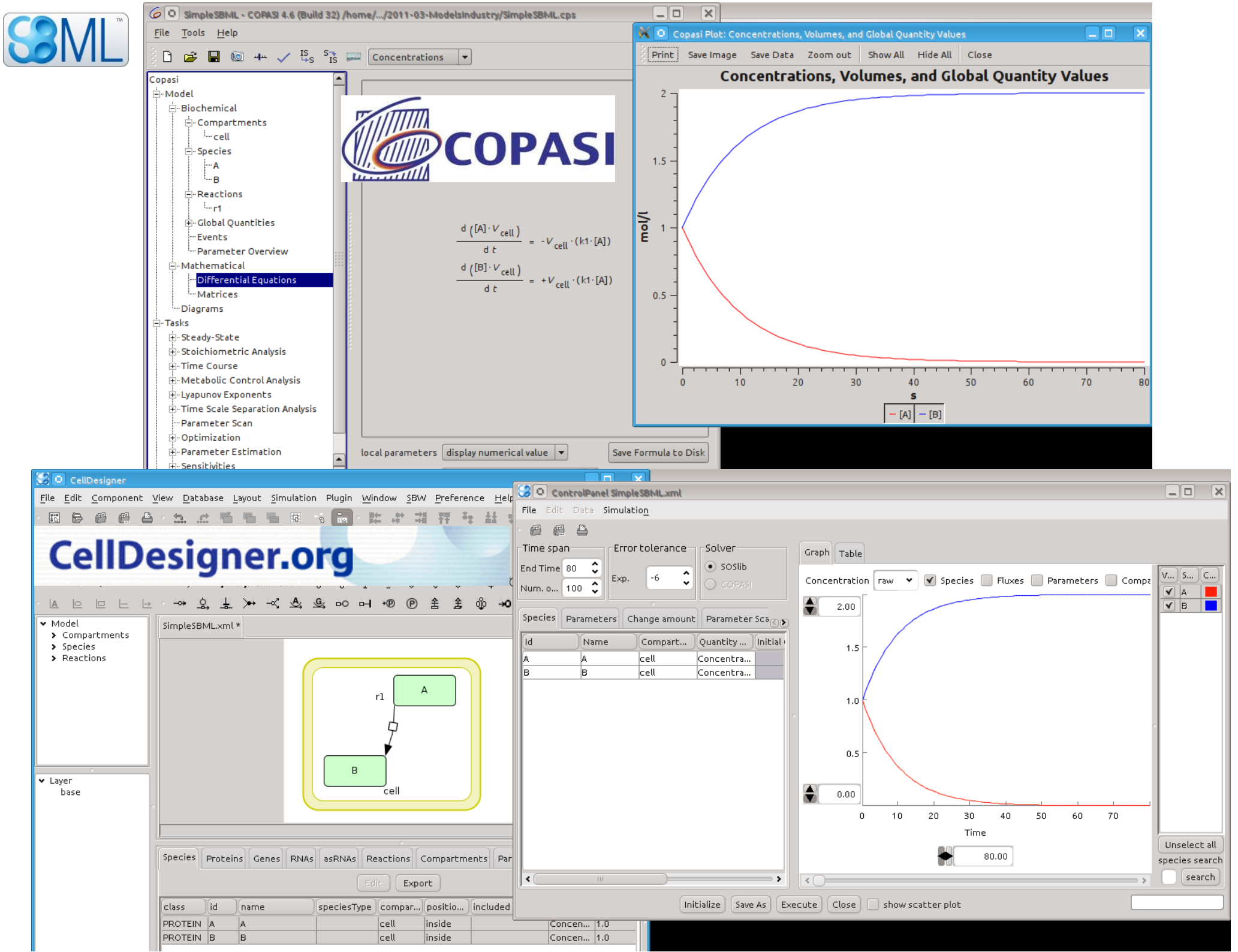
discrete events

arbitrary rules



A very simple SBML file (A ⟶ B)

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level2/version4" level="2" version="4">
  <model name="Simple Model">
    <listOfCompartments>
      <compartment id="cell" size="1" />
    </listOfCompartments>
    <listOfSpecies>
      <species id="A" compartment="cell" initialConcentration="1"/>
      <species id="B" compartment="cell" initialConcentration="1"/>
    </listOfSpecies>
    <listOfParameters>
      <parameter id="k1" value="0.1"/>
    </listOfParameters>
    <listOfReactions>
      <reaction id="r1" reversible="false">
      <listOfReactants>
          <speciesReference species="A"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReference species="B"/>
        </listOfProducts>
        <kineticLaw>
          <math xmlns="http://www.w3.org/1998/Math/MathML">
            <apply>
              <times/>
              <ci> cell </ci>
              <ci> k1 </ci>
              <ci> A </ci>
            </apply>
          </math>
        </kineticLaw>
      </reaction>
    </listOfReactions>
  </model>
</sbml>
```

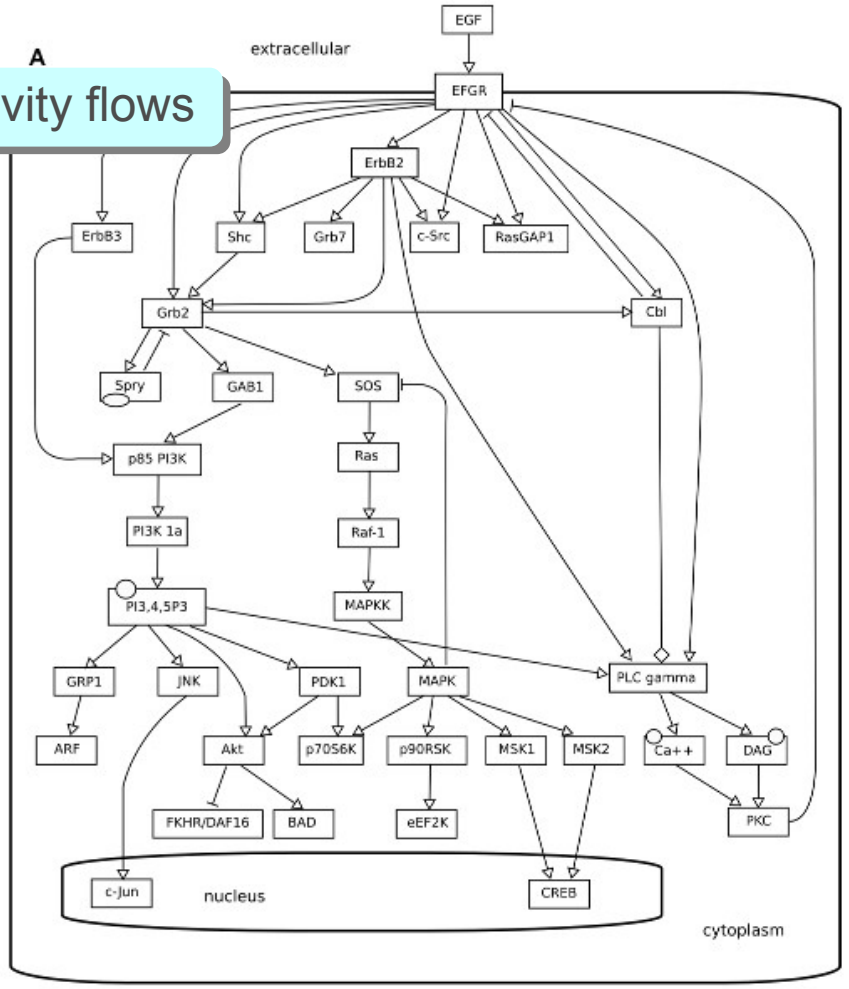

SBML
SimpleSBML - COPASI 4.6 (Build 32) /home/.../2011-03-ModelsIndustry/SimpleSBML.cps
File Tools Help
Concentrations
Copasi
Model
Biochemical
Compartments
cell
Species
A
B
Reactions
r1
Global Quantities
Events
Parameter Overview
Mathematical
Differential Equations
Matrices
Diagrams
Tasks
Steady-State
Stoichiometric Analysis
Time Course
Metabolic Control Analysis
Lyapunov Exponents
Time Scale Separation Analysis
Parameter Scan
Optimization
Parameter Estimation
Sensitivities
COPASI
d([A]·V_cell)/dt = -V_cell·(k1·[A])
d([B]·V_cell)/dt = +V_cell·(k1·[A])
local parameters
display numerical value
Save Formula to Disk
Copasi Plot: Concentrations, Volumes, and Global Quantity Values
Print Save Image Save Data Zoom out Show All Hide All Close
Concentrations, Volumes, and Global Quantity Values
mol/l
s
[A] [B]
CellDesigner
File Edit Component View Database Layout Simulation Plugin Window SBW Preference Help
CellDesigner.org
Model
Compartments
Species
Reactions
Layer
base
SimpleSBML.xml *
r1
A
B
cell
Species Proteins Genes RNAs asRNAs Reactions Compartments
Edit Export
class id name speciesType compar... positio... included
PROTEIN A A cell inside Concen... 1.0
PROTEIN B B cell inside Concen... 1.0
ControlPanel SimpleSBML.xml
File Edit Data Simulation
Time span
End Time 80
Num. o... 100
Error tolerance
Exp. -6
Solver
SOSlib
COPASI
Species Parameters Change amount Parameter Sca
Id Name Compart... Quantity ... Initial
A A cell Concentra...
B B cell Concentra...
Graph Table
Concentration raw
Species Fluxes Parameters Compa
2.00
0.00
Time
80.00
V... S... C...
A
B
Unselect all
species search
search
Initialize Save As Execute Close show scatter plot

SBML Level 3 packages

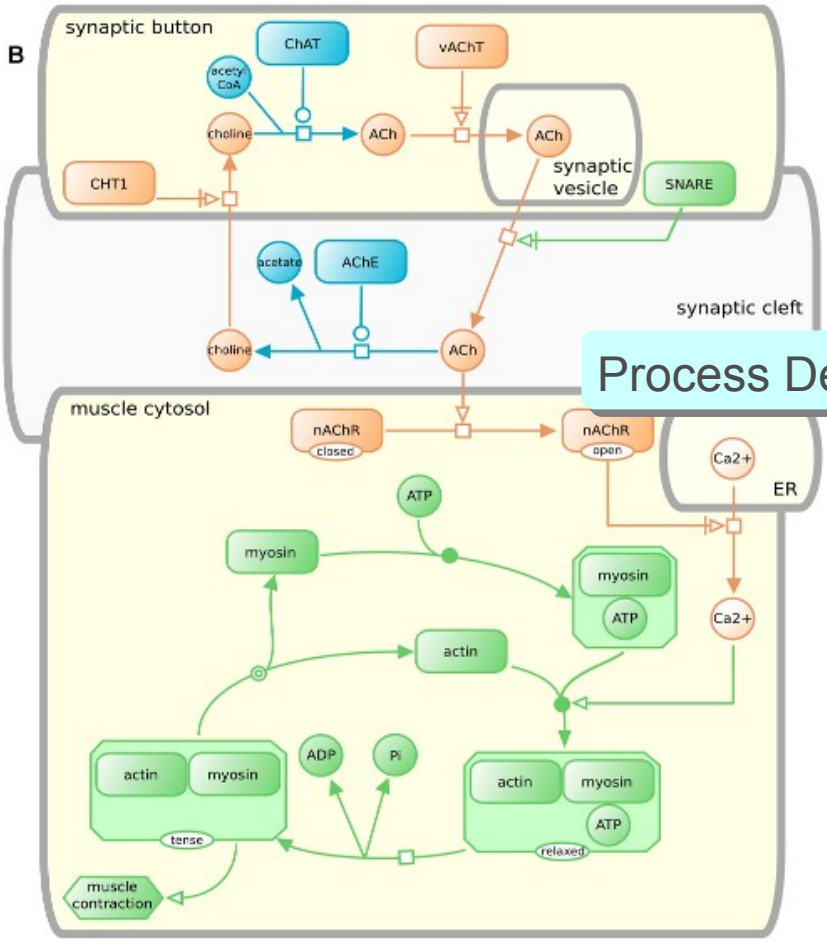
- Core package – public specification
- Graph Layout – specification finalised
- Graph rendering – specification finalised
- Complex species – specification finalised
- Groups - specification finalised
- Model composition – specification finalised
- Qualitative models – specification finalised
- Flux balance constraint – specification finalised
- Distributions and ranges - specification under discussion
- Spatial diffusion – specification under discussion
- Enhanced metadata – specification under discussion
- Arrays and sets – specification proposed
- Dynamic structures - needed
- ???

Systems Biology Graphical Notation

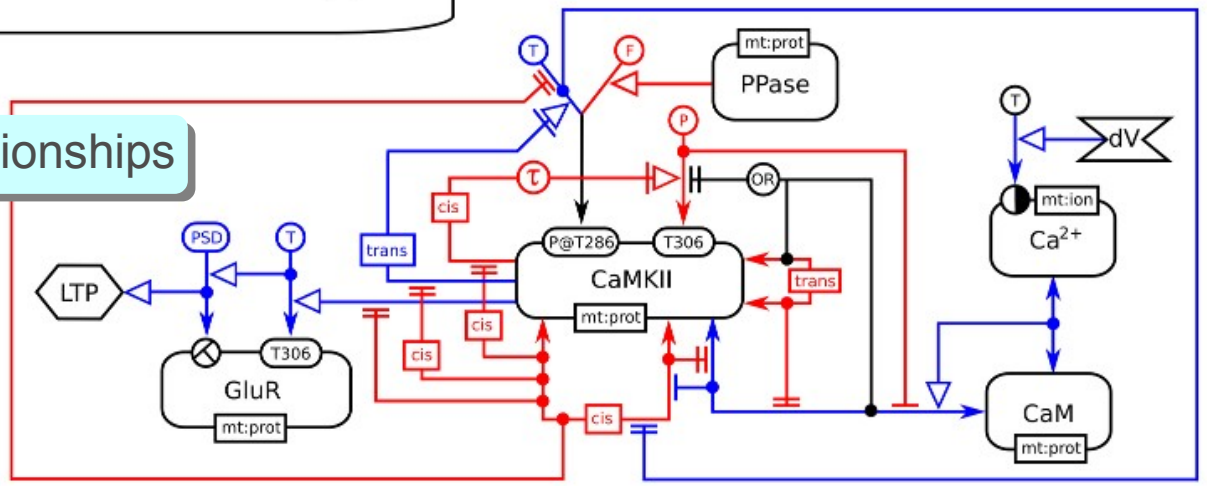
Activity flows



Process Descriptions



Entity Relationships



KEGG PATHWAY Database

Wiring diagrams of molecular interactions, reactions, and relations

Select prefix: map | Organism

Enter keywords: | Go | Help

Pathway Maps

KEGG PATHWAY is a collection of manually drawn pathway maps (see new maps, change history, and last updates) representing our knowledge on the molecular interaction and reaction networks for:

0. **Global Map**
1. **Metabolism**
   Carbohydrate Energy Lipid Nucleotide Amino acid Other amino acid Glycan
   Cofactor/vitamin Terpenoid/PK Other secondary metabolite Xenobiotics Overview
2. **Genetic Information Processing**
3. **Environmental Information Processing**
4. **Cellular Processes**
5. **Organismal Systems**
6. **Human Diseases**

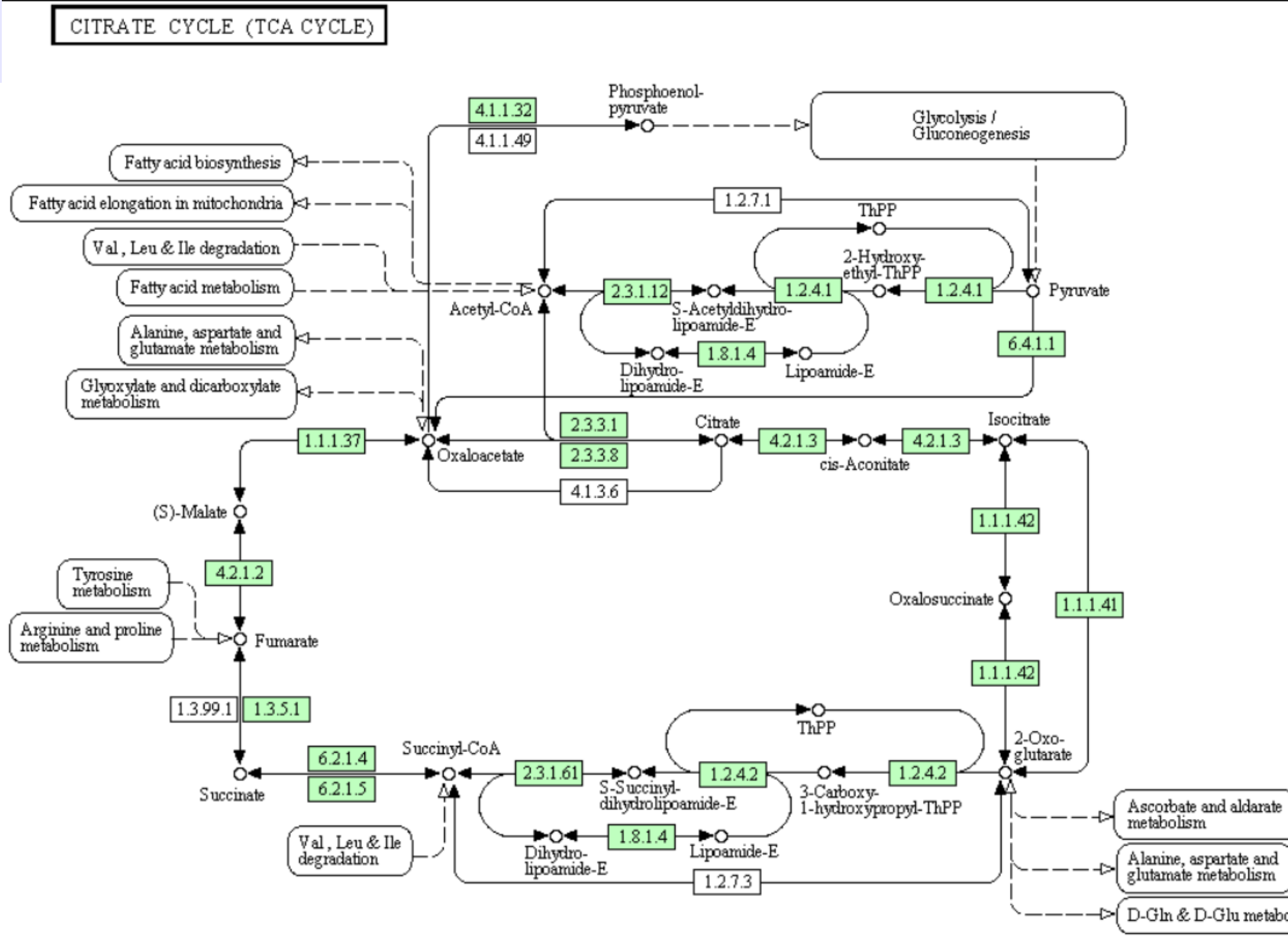
and also on the structure relationships (KEGG drug structure maps) in:

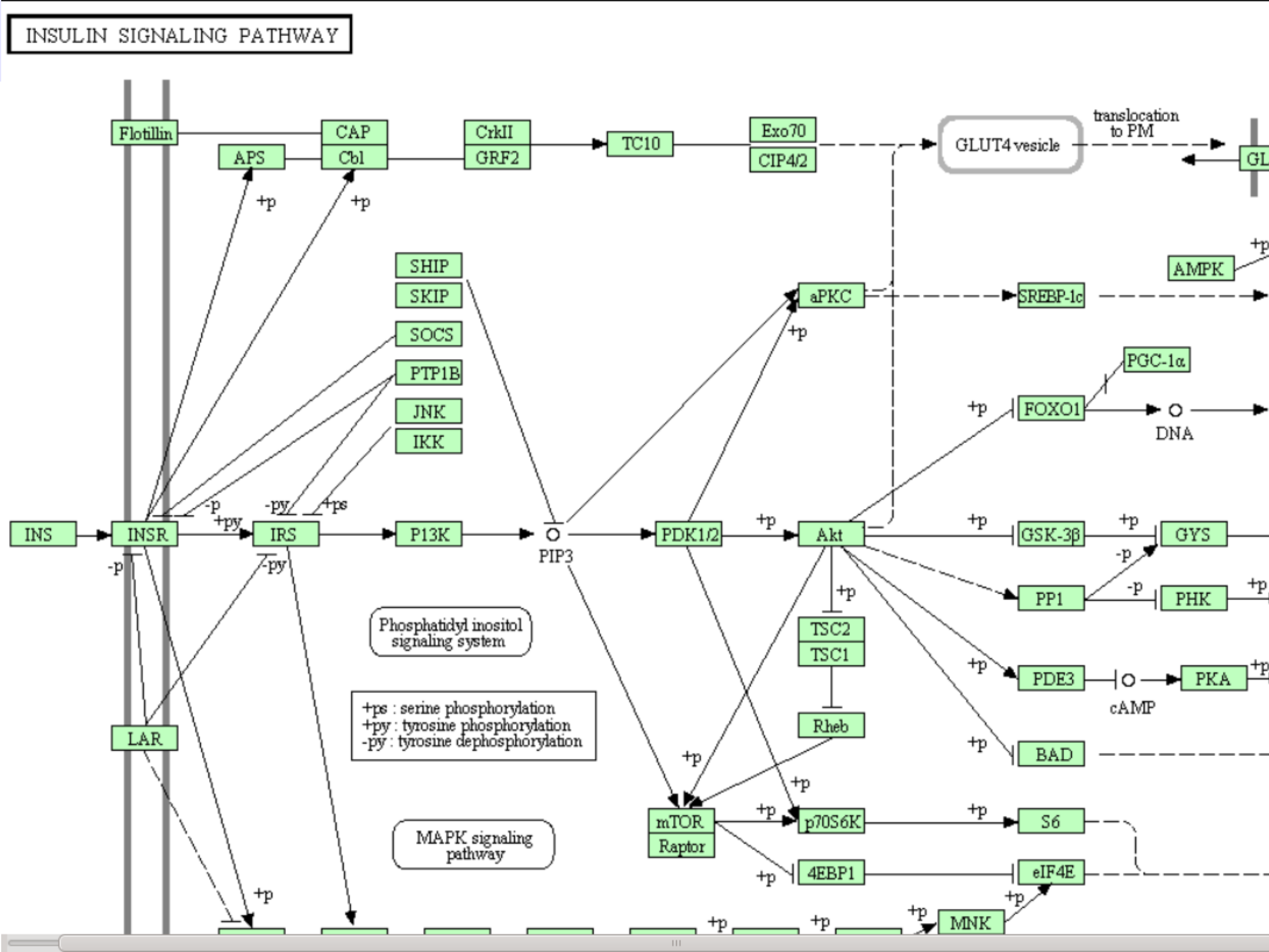
7. **Drug Development**

Pathway Mapping

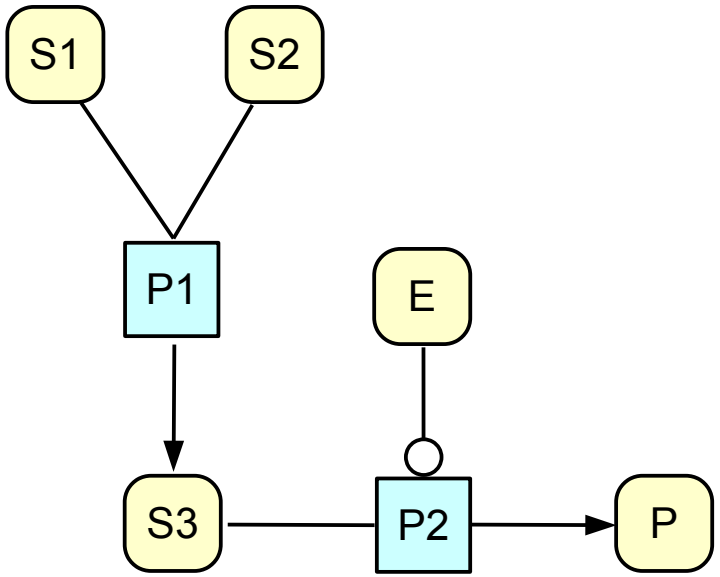
KEGG PATHWAY mapping is the process to map molecular datasets, especially large-scale datasets in genomics, transcriptomics, proteomics, and metabolomics, to the KEGG pathway maps for biological interpretaion of higher-level systemic functions.

- Search Pathway - basic pathway mapping tool
- Search&Color Pathway - advanced pathway mapping tool


CITRATE CYCLE (TCA CYCLE)
4.1.1.32
4.1.1.49
Phosphoenol-pyruvate
Glycolysis / Gluconeogenesis
Fatty acid biosynthesis
Fatty acid elongation in mitochondria
Val , Leu & Ile degradation
Fatty acid metabolism
Alanine, aspartate and glutamate metabolism
Glyoxylate and dicarboxylate metabolism
1.2.7.1
ThPP
2-Hydroxy-ethyl-ThPP
2.3.1.12
1.2.4.1
1.2.4.1
Pyruvate
Acetyl-CoA
S-Acetyldihydro-lipoamide-E
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
6.4.1.1
1.1.1.37
2.3.3.1
2.3.3.8
Oxaloacetate
Citrate
4.2.1.3
4.2.1.3
cis-Aconitate
Isocitrate
4.1.3.6
(S)-Malate
1.1.1.42
Tyrosine metabolism
4.2.1.2
Oxalosuccinate
1.1.1.41
Arginine and proline metabolism
Fumarate
1.1.1.42
1.3.99.1
1.3.5.1
ThPP
2-Oxo-glutarate
6.2.1.4
6.2.1.5
Succinyl-CoA
2.3.1.61
1.2.4.2
1.2.4.2
Succinate
S-Succinyl-dihydrolipoamide-E
3-Carboxy-1-hydroxypropyl-ThPP
Ascorbate and aldarate metabolism
Val , Leu & Ile degradation
1.8.1.4
Dihydro-lipoamide-E
Lipoamide-E
Alanine, aspartate and glutamate metabolism
1.2.7.3
D-Gln & D-Glu metabo
00020 6/28/11

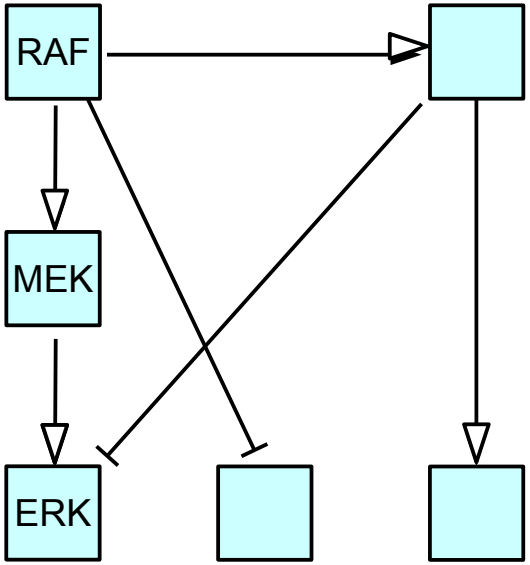

INSULIN SIGNALING PATHWAY
Flotillin
CAP
Cbl
APS
CrkII
GRF2
TC10
Exo70
CIP4/2
GLUT4 vesicle
translocation to PM
SHIP
SKIP
SOCS
PTP1B
JNK
IKK
aPKC
AMPK
SREBP-1c
PGC-1α
FOXO1
DNA
INS
INSR
IRS
P13K
PIP3
PDK1/2
Akt
GSK-3β
GYS
PP1
PHK
TSC2
TSC1
Rheb
PDE3
cAMP
PKA
BAD
LAR
Phosphatidyl inositol signaling system
+ps : serine phosphorylation
+py : tyrosine phosphorylation
-py : tyrosine dephosphorylation
MAPK signaling pathway
mTOR
Raptor
p70S6K
S6
4EBP1
eIF4E
MNK

Process Descriptions



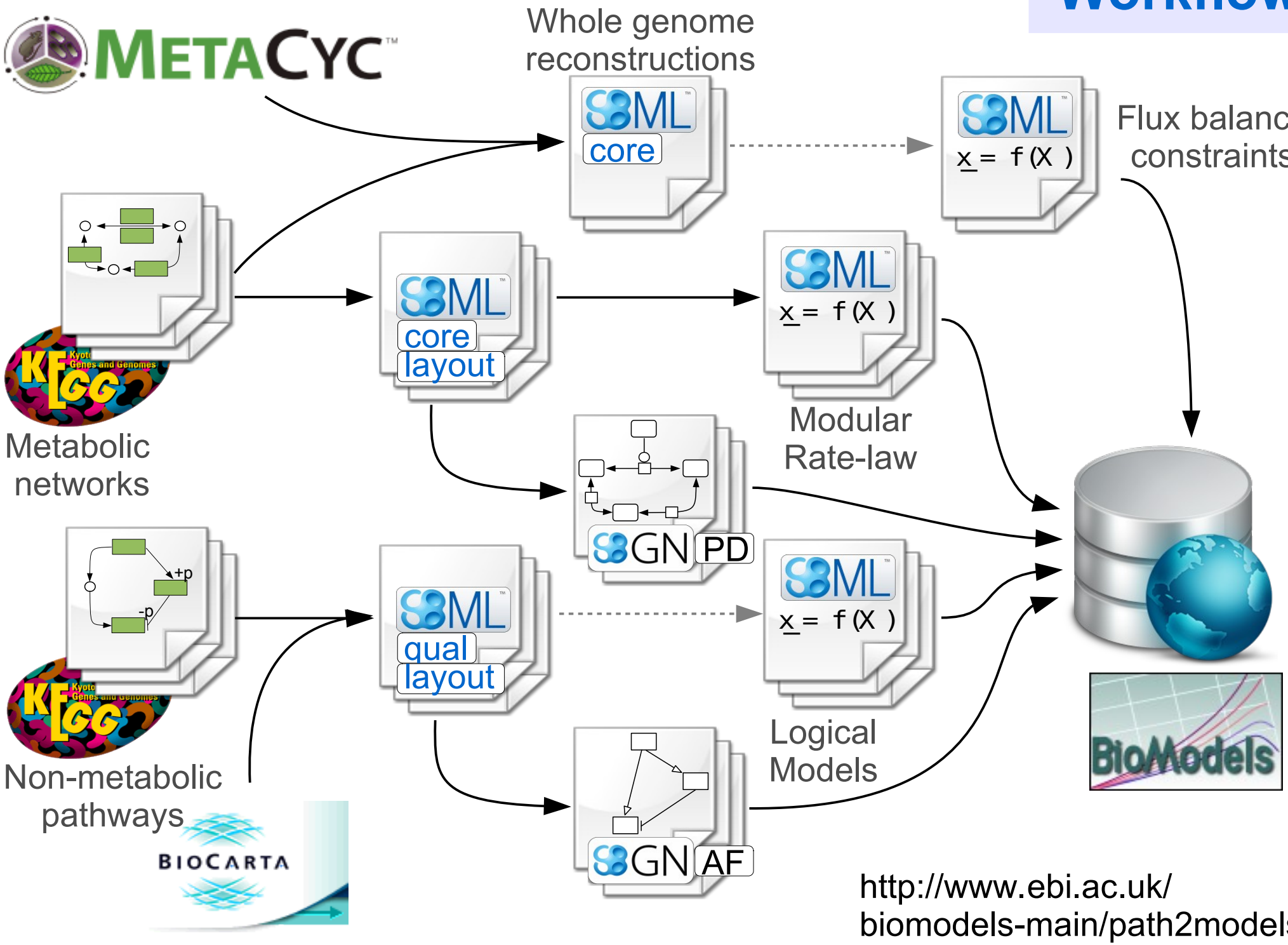
- Directional
- Sequential
- Mechanistic
- Subjected to combinatorial explosion
- Process modelling
- Biochemistry, Metabolic networks
- KEGG, Reactome
- SBML core

Activity-Flows



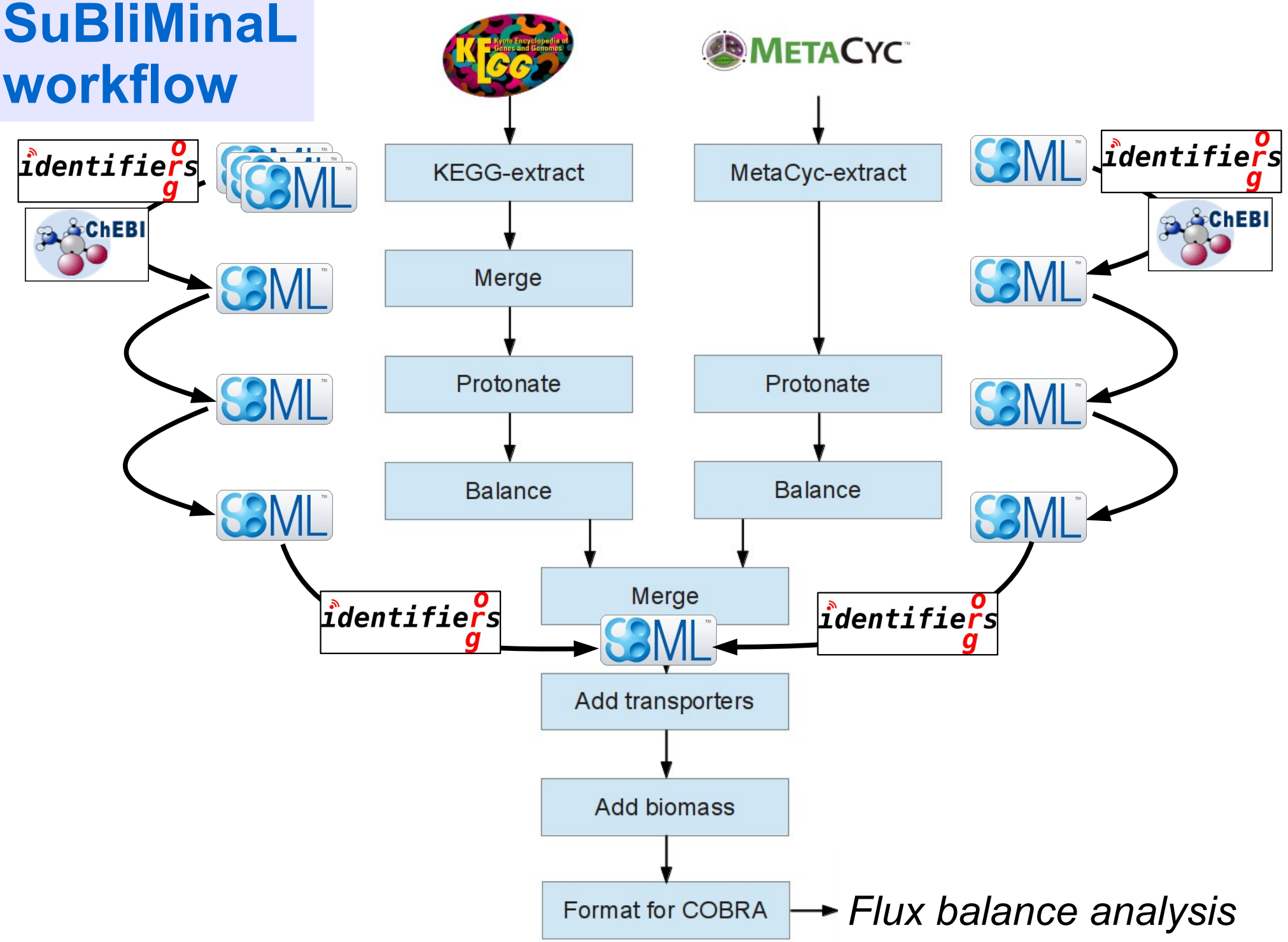
- Directional
- Sequential
- Non-mechanistic
- Logical modelling
- Signalling pathways, gene regulatory networks
- KEGG, STKEs
- SBML qual

Workflow



http://www.ebi.ac.uk/biomodels-main/path2models

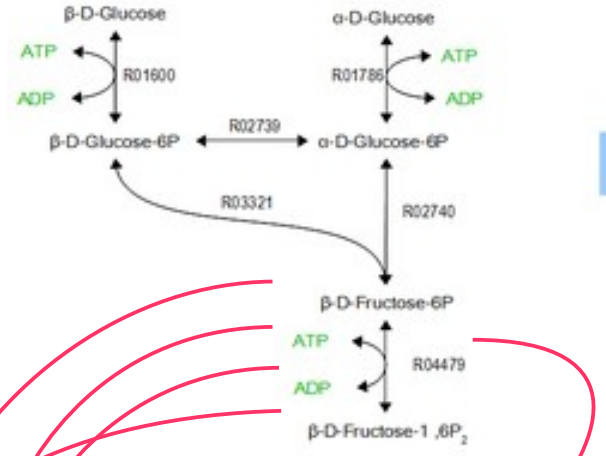
SuBliMinaL workflow



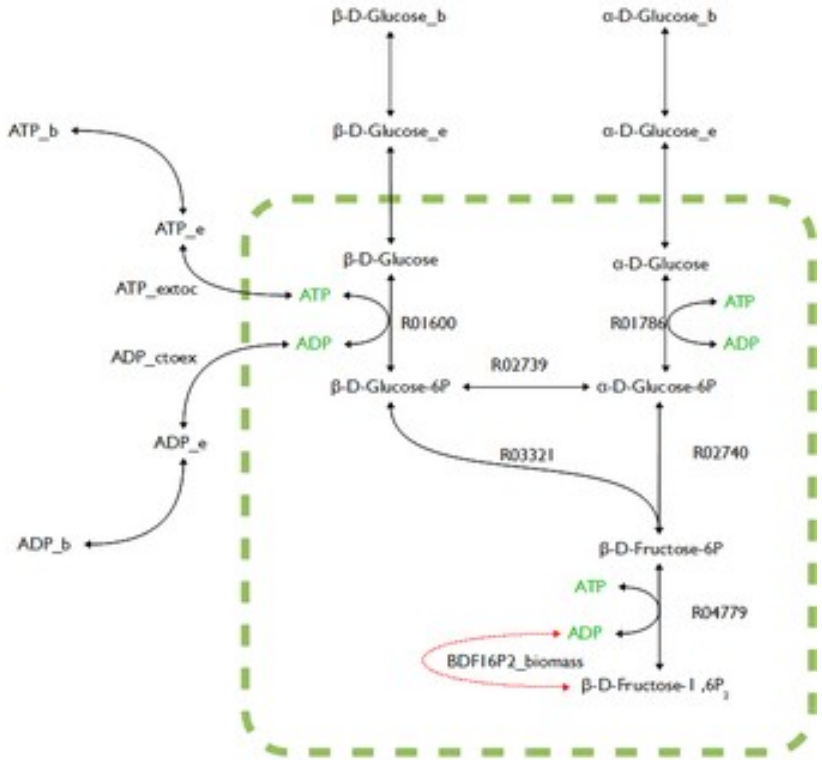
Flux balance analysis models

- For any connected metabolic network, one can build the stoichiometry matrix S. m rows are metabolites, n columns are reactions. Sij is positive for products, negative for substrates and null of metabolite not affected by the reaction.

Network and stoichiometric matrix



Prepare for FBA

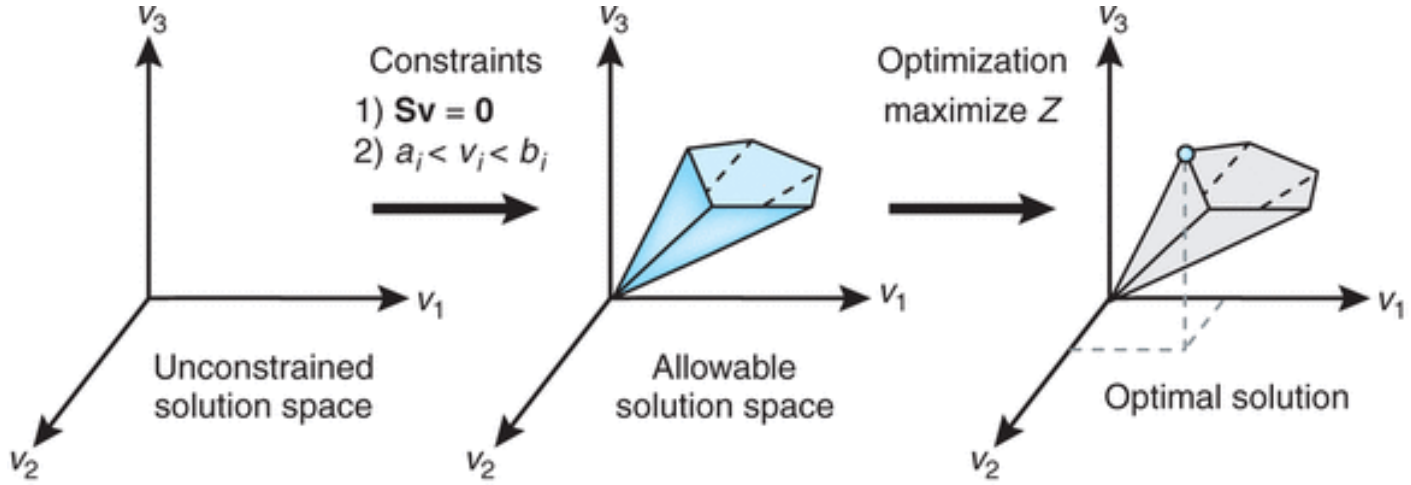


	R02739	R03321	R02740	R01600	R01786	R04779
C00002	0	0	0	-1	-1	-1
C00008	0	0	0	1	1	1
C00221	0	0	0	-1	0	0
C00267	0	0	0	0	-1	0
C00668	-1	0	-1	0	1	0
C01172	1	-1	0	1	0	0
C05345	0	1	1	0	0	-1
C05378	0	0	0	0	0	1

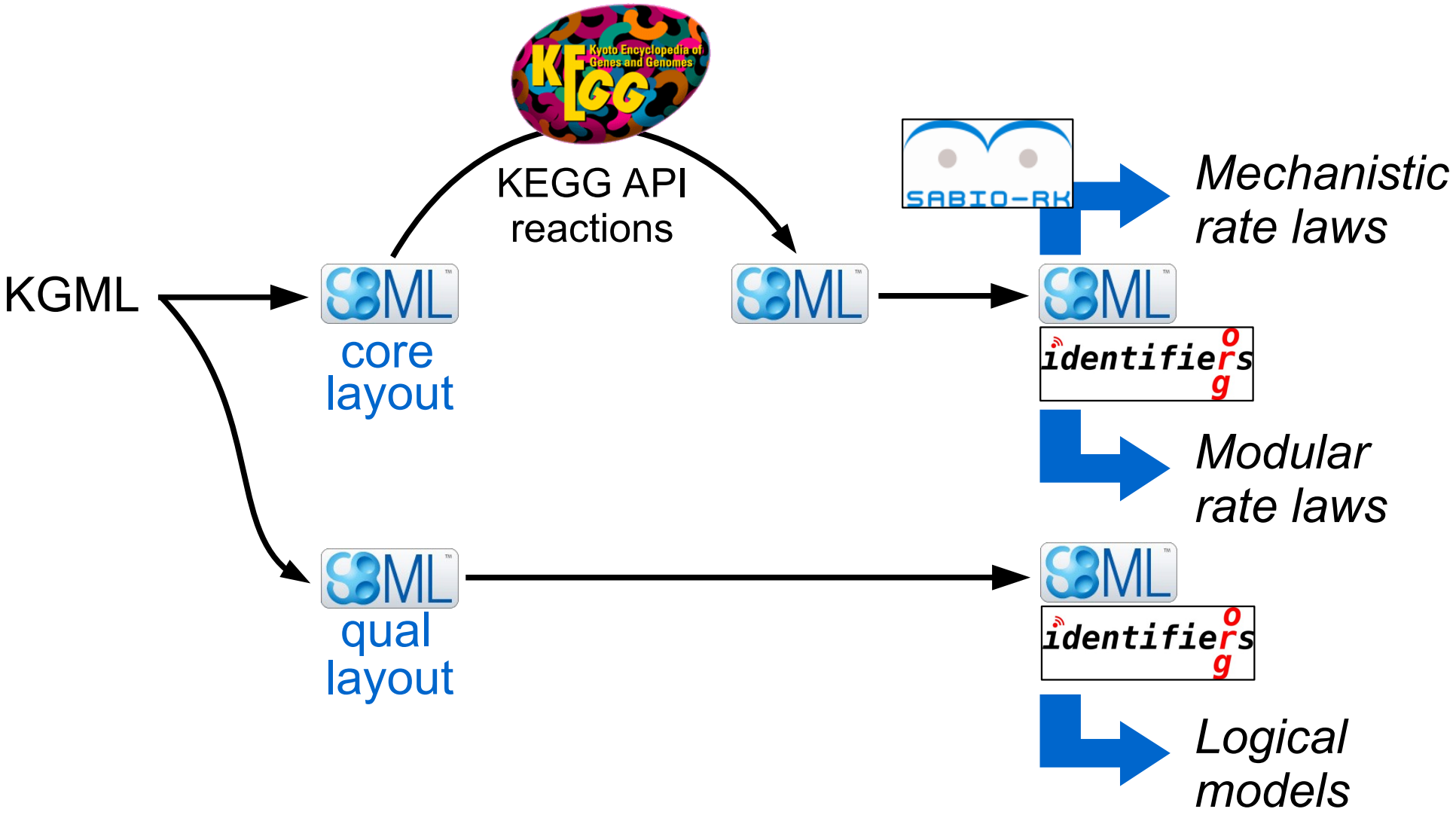
	R02739_c	R03321_c	R02740_c	R01600_c	R01786_c	R04779_c	ATP_extoc_tr	ADP_ctoex_tr	BDG_extoc_tr	ADG_extoc_tr	BDF16P2_biomass_tr
C00002[c]	0	0	0	-1	-1	-1	1	0	0	0	0
C00002[e]	0	0	0	0	0	0	-1	0	0	0	0
C00008[c]	0	0	0	1	1	1	0	-1	0	0	1
C00008[e]	0	0	0	0	0	0	0	1	0	0	0
C00221[c]	0	0	0	-1	0	0	0	0	1	0	0
C00221[e]	0	0	0	0	0	0	0	0	-1	0	0
C00267[c]	0	0	0	0	-1	0	0	0	0	1	0
C00267[e]	0	0	0	0	0	0	0	0	0	-1	0
C00668[c]	-1	0	-1	0	1	0	0	0	0	0	0
C01172[c]	1	-1	0	1	0	0	0	0	0	0	0
C05345[c]	0	1	1	0	0	-1	0	0	0	0	0
C05378[c]	0	0	0	0	0	1	0	0	0	0	-1

Flux balance analysis models

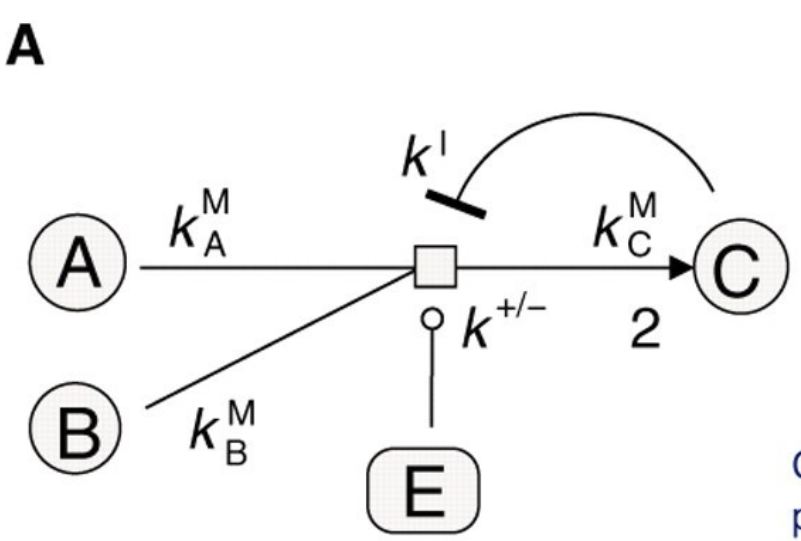
- For any connected metabolic network, one can build the stoichiometry matrix S. m rows are metabolites, n columns are reactions. Sij is positive for products, negative for substrates and null of metabolite not affected by the reaction.
- V is the vector of velocities for all the reactions. Vi is constrained by lower and upper bounds. No need to know the rate-laws.
- The solutions of S.V = 0 (that is the set of chemical kinetics differential equations) provide the steady-states of the system. In general n>>m, resulting in a continuum of solutions
- One can add objective functions to find out single optimal fluxes. E.g. maximum growth rate, or maximum ATP production.
- The system is solved by linear programming and the result is one vector of velocities.

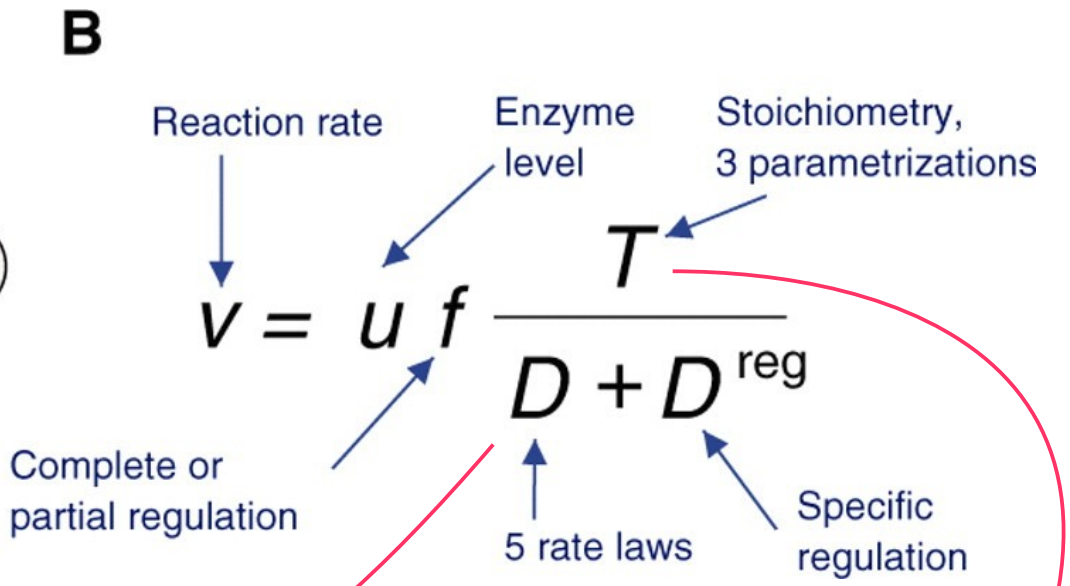


KEGGtranslator workflow



Common modular rate-law





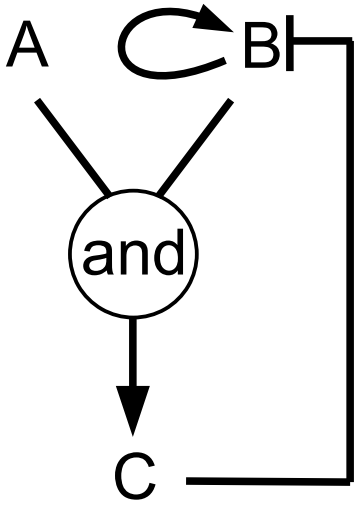
Common rate law
Direct binding rate law
Simultaneous binding rate law
Force dependent rate law
Power Law

$$T = k_r^+ \prod_i \left(\frac{c_i}{k_{ri}^{M}} \right)^{m_{ri}^+} - k_r^- \prod_i \left(\frac{c_i}{k_{ri}^{M}} \right)^{m_{ri}^-}$$

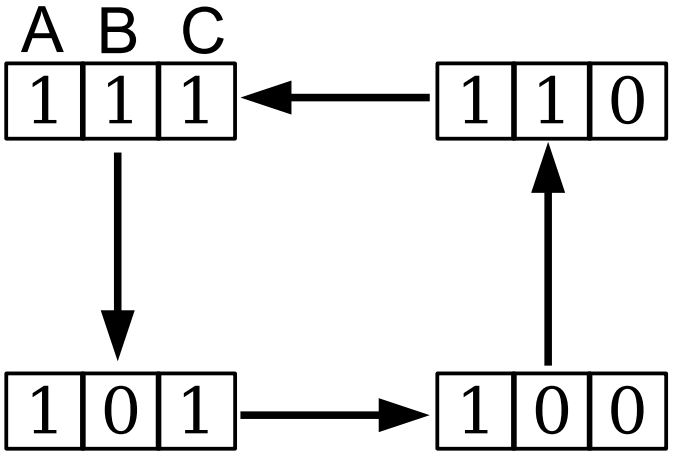
Liebermeister, Uhlendorf, Klipp (2010) Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics* 26: 1528-1534

Logical models in biology

- Variables can take a discrete number of values, at least 2
- Transitions of output are expressed as logical combinations of input values
- Simulations can be:
  synchronous: all the nodes are updated at once
  asynchronous: nodes are updated one after the other
  mixed
- One can add delays, inputs etc.



Influence diagram



state diagram

22nd BioModels Database release

- 20th May 2012 – first release of Path2Models data
- 112 898 common modular rate law models of metabolic networks
- 27 306 qualitative models of signalling pathways
- 1 836 whole genome flux balance analysis models
- 239 models for human, 234 models for mouse
- 444 133 925 cross references

http://www.ebi.ac.uk/biomodels-main/path2models

BioModels Home | Models | Submit | Support | About BioModels | Contact us | Search

Path2Models

The path2models project aims at the large scale generation of quantitative models form pathways.

Browse models

Models from this project are classified in 3 distinct categories:

- metabolic models
- non-metabolic models
- whole genome metabolism models

One can also browse those models by organism:

- list of all organisms

Search models

The following search will only look for models coming from the path2models project:

Search

Help about the search

- The keywords *AND* or *OR* (in upper cases) are available to refine the search. By default, if more than one word is present in a query, *OR* will be used to combine them.
- Double quotes (") can be used to force the search engine to match a whole expression containing several words.
- The colon character (:) must be escaped in the queries; one can use a backslash for this purpose (\:).

Download all models

- Archives of all models (from the latest release)

Glutamatergic synapse - Homo sapiens

Download SBML | Additional file(s) | Send feedback

Model information

Identifier: BMID000000017781
Format: SBML L3 V1 (Layout, Qualitative Models)
Project: path2models
Categories: non-metabolic
Submission: 17 May 2012 16:59:41 UTC
Last modified: 17 May 2012 16:59:41 UTC
Published: 19 May 2012 23:49:21 UTC

Annotations

occursIn	Homo sapiens	*Taxonomy*
isDescribedBy	regulation of synaptic transmission, glutamatergic	*Gene Ontology*
isDescribedBy	Glutamatergic synapse	*KEGG Pathway*

Notes

Model of "Glutamatergic synapse" in "Homo sapiens (human)"

Glutamate is the major excitatory neurotransmitter in the mammalian central nervous system(CNS). Glutamate is packaged into synaptic vesicles in the presynaptic terminal. Once released into the synaptic cleft, glutamate acts on postsynaptic ionotropic glutamate receptors (iGluRs) to mediate fast excitatory synaptic transmission. Glutamate can also act on metabotropic glutamate receptors (mGluRs) and exert a variety of modulatory effects through their coupling to G proteins and the subsequent recruitment of second messenger systems. Presynaptically localized Group II and Group III mGluRs are thought to represent the classical inhibitory autoreceptor mechanism that suppresses excess glutamate release. After its action on these receptors, glutamate can be removed from the synaptic cleft by EAATs located either on the presynaptic terminal, neighboring glial cells, or the postsynaptic neuron. In glia, glutamate is converted to glutamine, which is then transported back to the presynaptic terminal and converted back to glutamate.

Graphical representation of 'Glutamatergic synapse' (PNG image hosted by the Kyoto Encyclopedia of Genes and Genomes, KEGG)

Original pathway (from the KEGG PATHWAY Database)

This model has been generated by the path2models project and is currently hosted on BioModels Database and identified by: BMID000000017781.

Glutamatergic synapse - Homo sapiens

Download SBML | Additional file(s) | Send feedback

Model information

Identifier: BMID000000017781 **Project:** path2models **Submission:** 17 May 2012 16:59:41 UTC

Format: SBML L3V1 (Layout, Qualitative Models) **Categories:** non-metabolic **Last modified:** 17 May 2012 16:59:41 UTC

occursIn
isDescribedBy
isDescribedBy

Model of "Glutamat
Glutamate is the n
vesicles in the pre
(iGluRs) to mediat
a variety of modula
Presynaptically lo
suppresses exces
located either on t
which is then trans

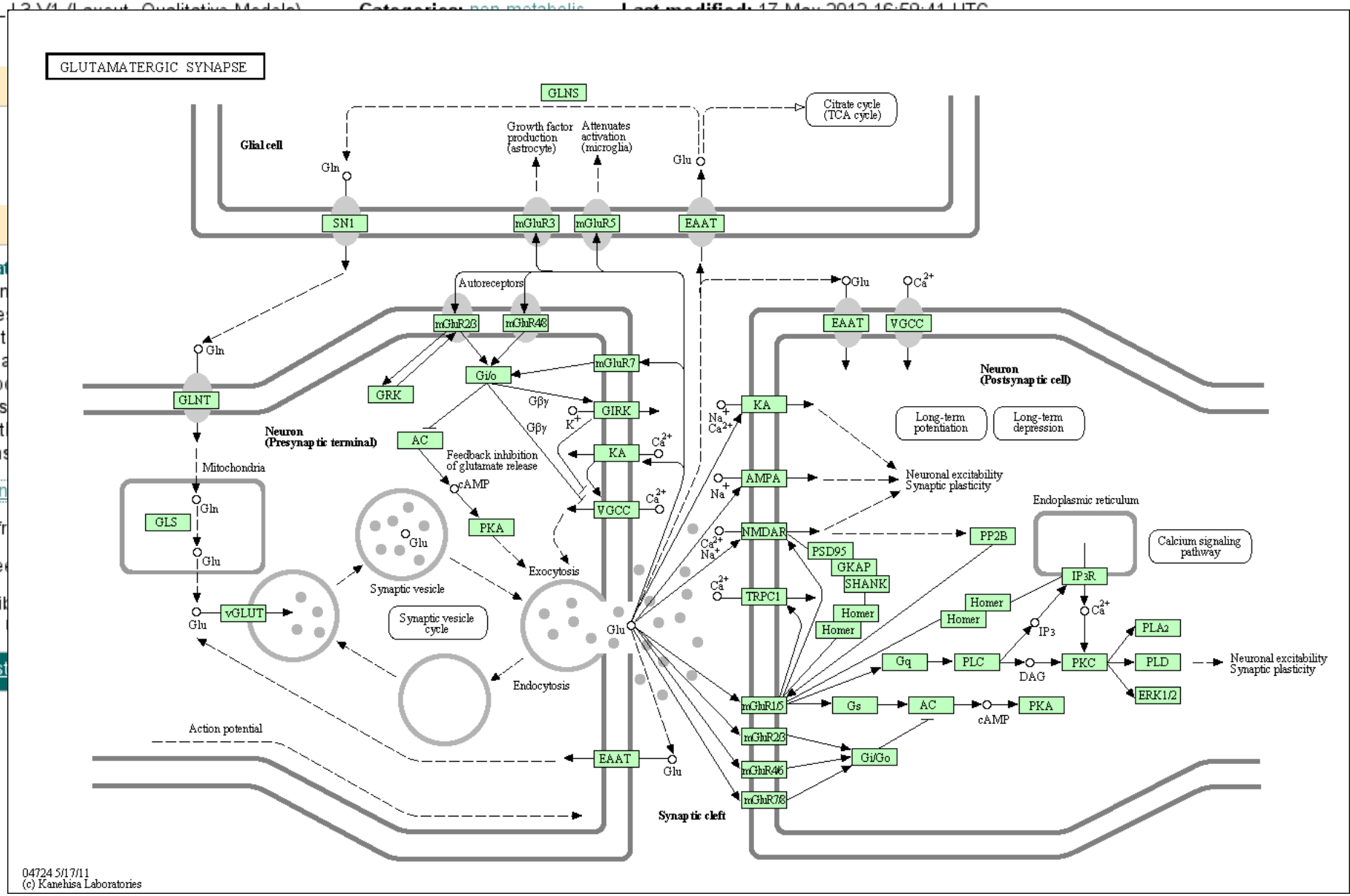
Graphical represen

Original pathway (fr

This model has be

To the extent possib
worldwide. Please

Computational Syst
Team



Acknowledgements

Neil Swainston

Clemens Wrzodek

Tobias Czauderna

Andreas Dräger

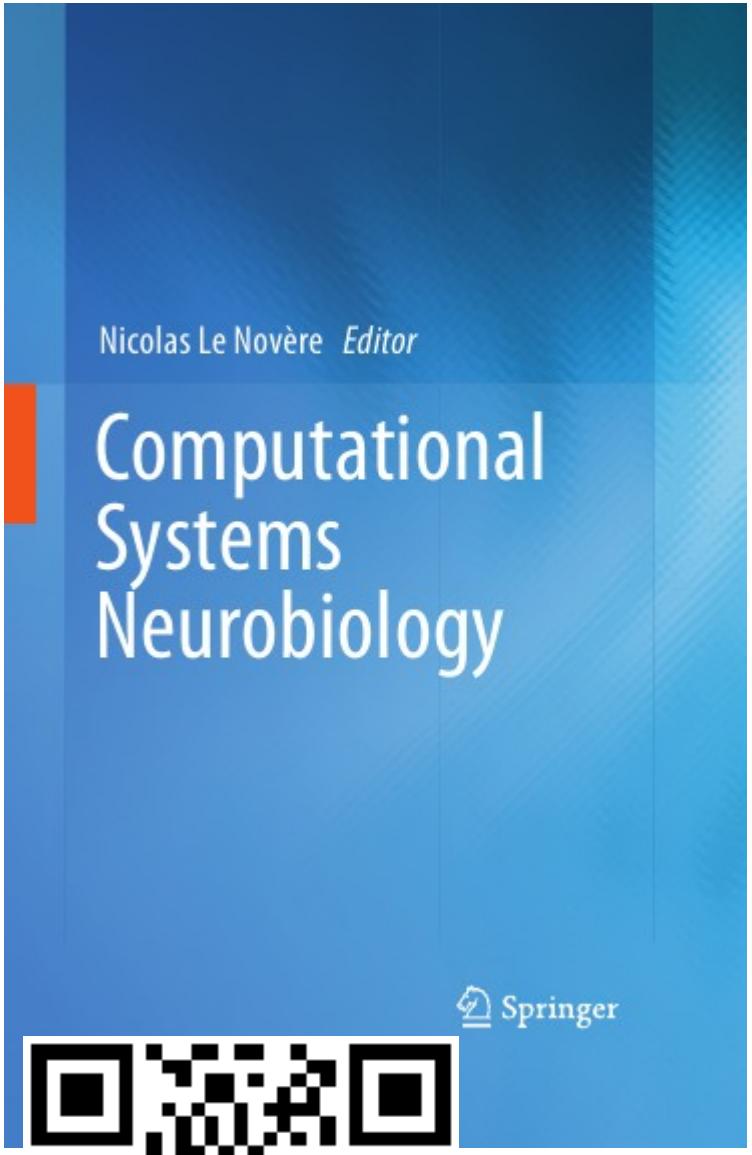
Finja Büchel

Michael Shubert

Camille Laibe

Acknowledgements

- EMBL-EBI (Cambridge): Sarah Keating, Camille Laibe, Nicolas Le Novère, Nicolas Rodriguez, Julio Saez-Rodriguez, Martijn van Iersel
  +Finja Bünel, Florian Mittag, Michael Schubert
- University of Tuebingen: Andreas Dräger, Roland Keller, Matthias Rall, Clemens Wrzodek, Andreas Zell
  +Finja Bünel, Florian Mittag
- University of Manchester: Neil Swainston
- Leibniz Institute of Plant Genetics and Crop Plant Research (Gatersleben): Falk Schreiber, Tobias Czauderna
- Instituto Gulbenkian de Ciência (Lisbon): Claudine Chaouyia



Gabriela Antunes, Douglas J Armstrong, Giorgio A. Ascoli, Upinder S Bhalla, Ingo Bojak, Sherry-Ann Brown, Sharon Crook, Erik De Schutter, Markus Diesmann, Stuart Edelstein, Steven J Eglen, Lukas Endler, Bard Ermentrout, Aldo Faisal, Brett L. Foster, Wulfram Gerstner, Todd A. Gillette, Padraig Gleeson, Geoffrey Goodhill, Kevin Gurney, Raquell Holmes, Mark Humphries, Sarah Keating, Anders Lansner, Nicolas Le Novère, David T.J. Liley, Les Loew, Duncan Mortimer, Richard Naud, Andrew Pocklington, Marie-Claude Potier, Isabelle Rivals, Fidel Santamaria, R. Angus Silver, Hugh Simpson, Melanie I Stefan, Volker Steuber, Lysimachos Zografos

Transcriptomics, proteomics

Biochemical systems, spatial modeling

Single compartment and multi-compartment neurons

Neuronal networks

Development of neuronal systems

Software and standards for modeling neuronal systems