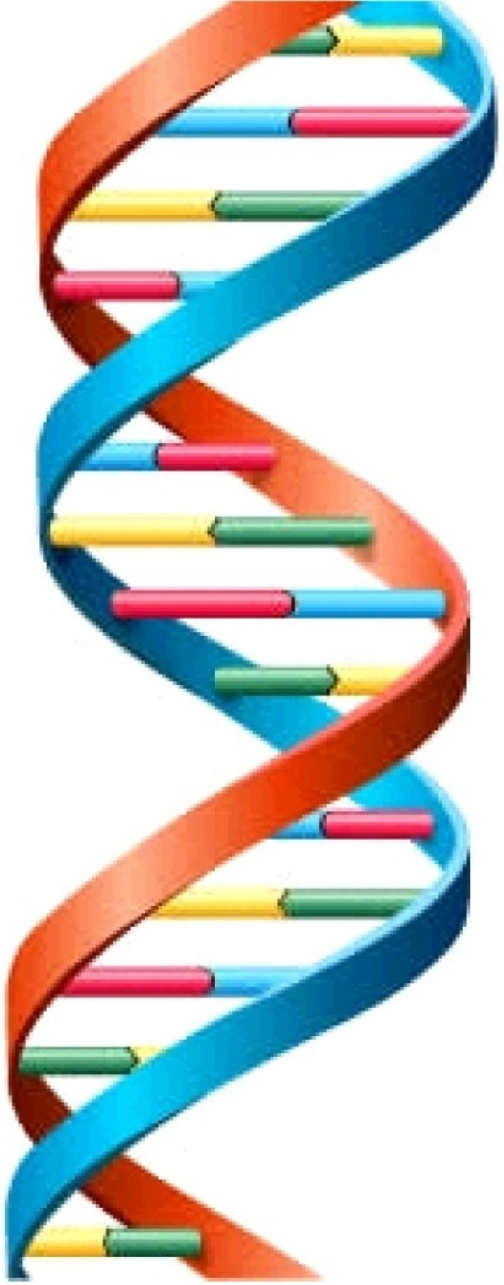


equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

1 Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).

2 Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

3 Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).

4 Ekman, V. W., *Arkiv. Mat. Astron. Fysik. (Stockholm)*, **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two sugar-phosphate chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the z-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at King's College, London. One of us (J. D. W.) has been aided by a fellowship from the National Foundation for Infantile Paralysis.

J. D. Watson
F. H. C. Crick
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems,
Cavendish Laboratory, Cambridge.
April 2.

1 Pauling, L., and Corey, R. B., *Nature*, **171**, 346 (1953); *Proc. U.S. Nat. Acad. Sci.*, **39**, 84 (1953).

2 Furberg, S., *Acta Chem. Scand.*, **6**, 634 (1952).

3 Chargaff, E., for references see Zamenhof, S., Brawerman, G., and Chargaff, E., *Biochim. et Biophys. Acta*, **9**, 402 (1952).

4 Wyatt, G. R., *J. Gen. Physiol.*, **36**, 201 (1952).

5 Astbury, W. T., Symp. Soc. Exp. Biol. 1, Nucleic Acid, 66 (Camb. Univ. Press, 1947).

6 Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

Molecular Structure of Deoxypentose Nucleic Acids

While the biological properties of deoxypentose nucleic acid suggest a molecular structure containing great complexity, X-ray diffraction studies described here (cf. Astbury[1]) show the basic molecular configuration has great simplicity. The purpose of this communication is to describe, in a preliminary way, some of the experimental evidence for the polynucleotide chain configuration being helical, and existing in this form when in the natural state. A fuller account of the work will be published shortly.

The structure of deoxypentose nucleic acid is the same in all species (although the nitrogen base ratios alter considerably) in nucleoprotein, extracted or in cells, and in purified nucleate. The same linear group of polynucleotide chains may pack together parallel in different ways to give crystalline[1–3], semi-crystalline or paracrystalline material. In all cases the X-ray diffraction photograph consists of two regions, one determined largely by the regular spacing of nucleotides along the chain, and the other by the longer spacings of the chain configuration. The sequence of different nitrogen bases along the chain is not made visible.

Oriented paracrystalline deoxypentose nucleic acid (structure *B*[1] in the following communication by Franklin and Gosling) gives a fibre diagram as shown in Fig. 1 (cf. ref. 4). Astbury suggested that the strong 3·4-A. reflexion corresponded to the internucleotide repeat along the fibre axis. The ~ 34 A. layer lines, however, are not due to a repeat of a polynucleotide composition, but to the chain configuration repeat, which causes strong diffraction as the nucleotide chains have higher density than the interstitial water. The absence of reflexions on or near the meridian immediately suggests a helical structure with axis parallel to fibre length.

Diffraction by Helices

It may be shown[1] (also Stokes, unpublished) that the intensity distribution in the diffraction pattern of a series of points equally spaced along a helix is given by the squares of Bessel functions. A uniform continuous helix gives a series of layer lines of spacing corresponding to the helix pitch, the intensity distribution along the nth layer line being proportional to the square of J_n, the nth order Bessel function. A straight line may be drawn approximately through the innermost maxima of each Bessel function and the origin. The angle this line makes with the equator is roughly equal to the angle between an element of the helix and the helix axis. If a unit repeats n times along the helix there will be a meridional reflexion (J_n) on the nth layer line. The helical configuration produces side-bands on this fundamental frequency, the effect[2] being to reproduce the intensity distribution about the origin around the new origin, on the nth layer line, corresponding to C in Fig. 2.

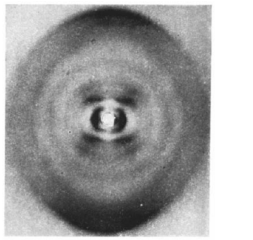


Fig. 1. Fibre diagram of deoxypentose nucleic acid from *B. coli*. Fibre axis vertical

We will now briefly analyse in physical terms some of the effects of the shape and size of the repeat unit or nucleotide on the diffraction pattern. First, if the nucleotide consists of a unit having circular symmetry about an axis parallel to the helix axis, the whole diffraction pattern is modified by the form factor of the nucleotide. Second, if the nucleotide consists of a series of points on a radius at right-angles to the helix axis, the phases of radiation scattered by the helices of different diameter passing through each point are the same. Summation of the corresponding Bessel functions gives reinforcement for the inner-

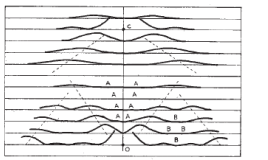


Fig. 2. Diffraction pattern of system of helices corresponding to structure of deoxypentose nucleic acid. The squares of Bessel functions are plotted about 0 on the equator and on the first, second, third and fifth layer lines for half of the nucleotide mass at 20 A. diameter and remainder distributed along a radius, the mass at a given radius being proportional to the radius. About *C* on the tenth layer line similar functions are plotted for an outer diameter of 12 A.

We wish to thank Prof. J. T. Randall for encouragement; Profs. E. Chargaff, R. Signer, J. A. V. Butler and Drs. J. D. Watson, J. D. Smith, L. Hamilton, J. C. White and G. R. Wyatt for supplying material without which this work would have been impossible; also Drs. J. D. Watson and Mr. F. H. C. Crick for stimulation, and our colleagues R. E. Franklin, R. G. Gosling, G. L. Brown and W. E. Seeds for discussion. One of us (H. R. W.) wishes to acknowledge the award of a University of Wales Fellowship.

M. H. F. Wilkins
Medical Research Council Biophysics Research Unit,
A. R. Stokes
H. R. Wilson
Wheatstone Physics Laboratory,
King's College, London.
April 2.

1 Astbury, W. T., Symp. Soc. Exp. Biol., 1, Nucleic Acid (Cambridge Univ. Press, 1947).

2 Riley, D. P., and Oster, G., *Biochim. et Biophys. Acta*, **7**, 526 (1951).

3 Wilkins, M. H. F., Gosling, R. G., and Seeds, W. E., *Nature*, **167**, 759 (1951).

4 Astbury, W. T., and Bell, F. O., Cold Spring Harb. Symp. Quant. Biol., **6**, 109 (1938).

5 Cochran, W., Crick, F. H. C., and Vand, V., *Acta Cryst.*, **5**, 581 (1952).

6 Wilkins, M. H. F., and Randall, J. T., *Biochim. et Biophys. Acta*, **10**, 192 (1953).

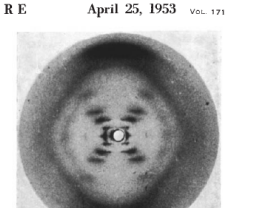
Molecular Configuration in Sodium Thymonucleate

Sodium thymonucleate fibres give two distinct types of X-ray diagram. The first corresponds to a crystalline form, structure *A*, obtained at about 75 per cent relative humidity; a study of this is described in detail elsewhere[1]. At higher humidities a different structure, structure *B*, showing a lower degree of order, appears and persists over a wide range of ambient humidity. The change from *A* to *B* is reversible. The water content of structure *B* fibres which undergo this reversible change may vary from 40–50 per cent to several hundred per cent of the dry weight. Moreover, some fibres never show structure *A*, and in these structure *B* can be obtained with an even lower water content.

The X-ray diagram of structure *B* (see photograph) shows in striking manner the features characteristic of helical structures, first worked out in this laboratory by Stokes (unpublished) and by Crick, Cochran and Vand[2]. Stokes and Wilkins were the first to propose such structures for nucleic acid as a result of direct studies of nucleic acid fibres, although a helical structure had been previously suggested by Furberg (thesis, London, 1949) on the basis of X-ray studies of nucleosides and nucleotides.

While the X-ray evidence cannot, at present, be taken as direct proof that the structure is helical, other considerations discussed below make the existence of a helical structure highly probable.

Structure *B* is derived from the crystalline structure *A* when the sodium thymonucleate fibres take up quantities of water in excess of about 40 per cent of their weight. The change is accompanied by an increase of about 30 per cent in the length of the fibre, and by a substantial re-arrangement of the molecule. It therefore seems reasonable to suppose that in structure *B* the structural units of sodium thymonucleate (molecules or groups of molecules) are relatively free from the influence of neighbouring molecules, each unit being shielded by a sheath of water. Each unit is then free to take up its least-energy configuration independently of its neighbours and, in view of the nature of the long-chain molecules involved, it is highly likely that the general form will be helical[3]. If we adopt the hypothesis of a helical structure, it is immediately possible, from the X-ray diagram of structure *B*, to make certain deductions as to the nature and dimensions of the helix.



Sodium deoxyribose nucleate from calf thymus. Structure *B*

The innermost maxima on the first, second, third and fifth layer lines lie approximately on straight lines radiating from the origin. For a smooth single-strand helix the structure factor on the nth layer line is given by:

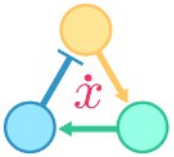
$$F_n = J_n(2\pi rR)\exp i\,n(\psi + \tfrac{1}{2}\pi),$$

where $J_n(u)$ is the nth-order Bessel function of u, r is the radius of the helix, and R and ψ are the radial and azimuthal co-ordinates in reciprocal space[4]; this expression leads to an approximately linear array of intensity maxima of the type observed, corresponding to the first maxima in the functions J_1, J_2, J_3, etc.

If, instead of a smooth helix, we consider a series of residues equally spaced along the helix, the transform in the general case treated by Crick, Cochran and Vand is more complicated. But if there is a whole number, m, of residues per turn, the form of the transform is as for a smooth helix with the addition, only, of the same pattern repeated with its origin at heights mc^*, $2mc^*$ . . . etc. (c is the fibre-axis period).

In the present case the fibre-axis period is 34 A. and the very strong reflexion at 3·4 A. lies on the tenth layer line. Moreover, lines of maxima radiating from the 3·4-A. reflexion as from the origin are visible on the fifth and lower layer lines, having a J_4 maximum coincident with that of the origin series on the fifth layer line. (The strong outer streaks which apparently radiate from the 3·4-A. maximum are not, however, so easily explained.) This suggests strongly that there are exactly 10 residues per turn of the helix. If this is so, then from a measurement of R_4, the position of the first maximum on the 4th layer line (for n 5<), the radius of the helix, can be obtained. In the present instance, measurements of R_1, R_2, R_3 and R_4 all lead to values of r of about 10 A.

HAPPY BIRTHDAY BEAUTIFUL



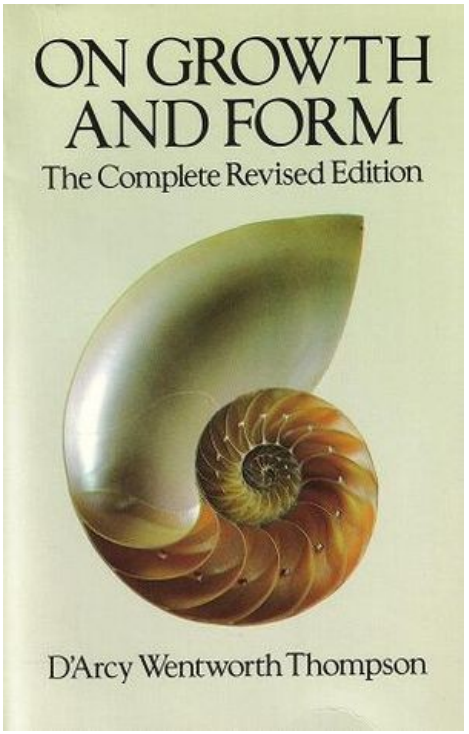
Large-scale harnessing of pathway knowledge to provide pump-priming models for all

Nicolas Le Novère

n.lenovere@gmail.com

Why using mathematical models?

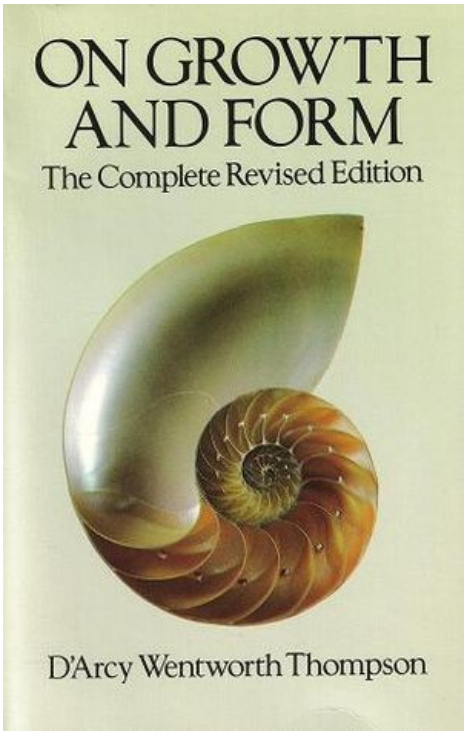
Describe



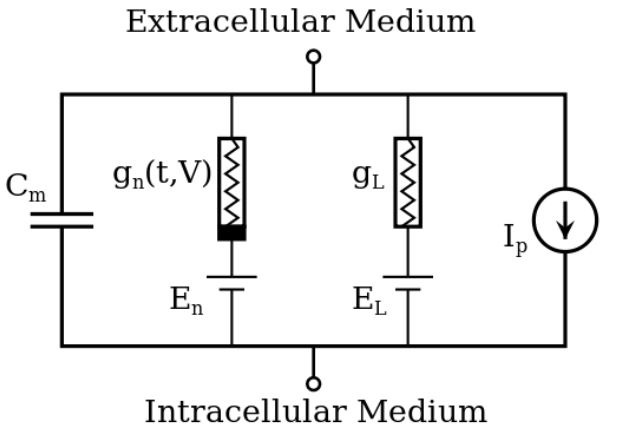
1917

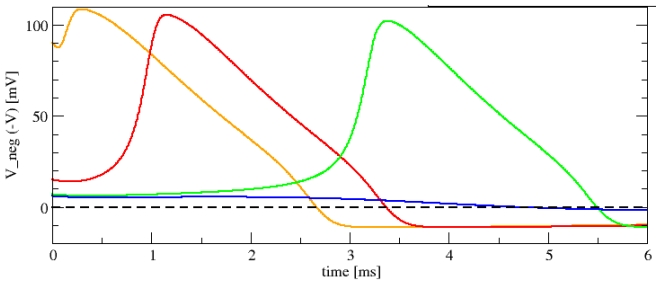
Why using mathematical models?

Describe	Explain



1917

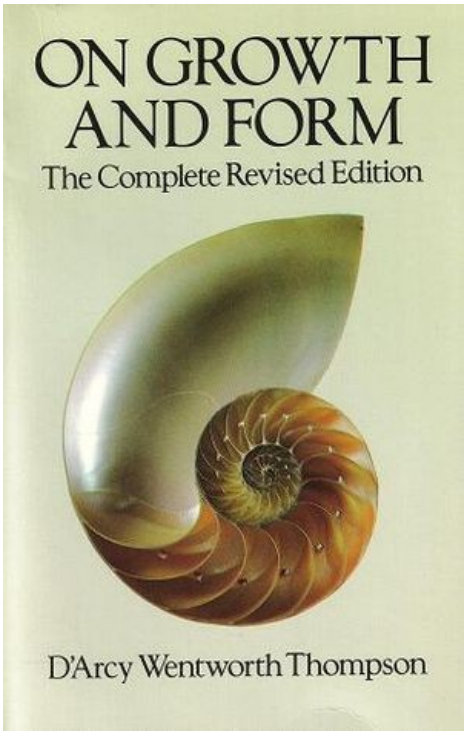


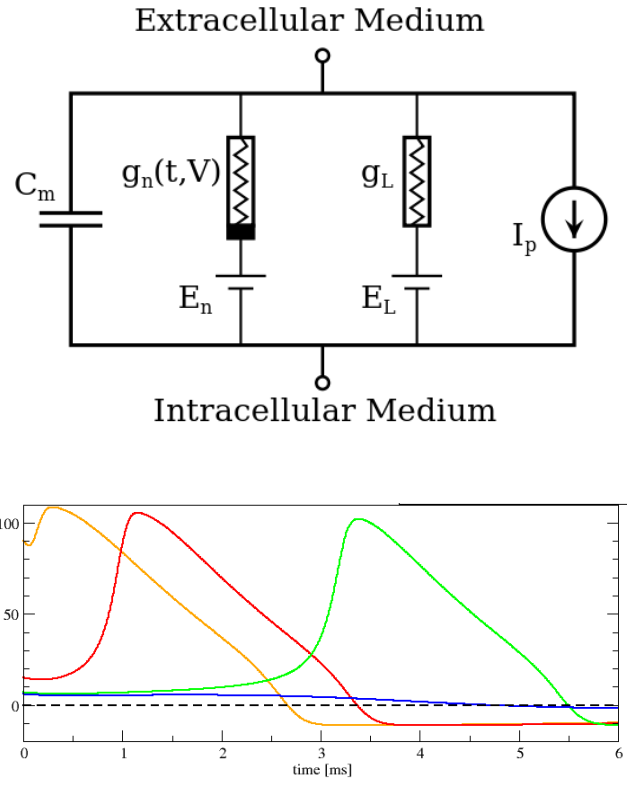


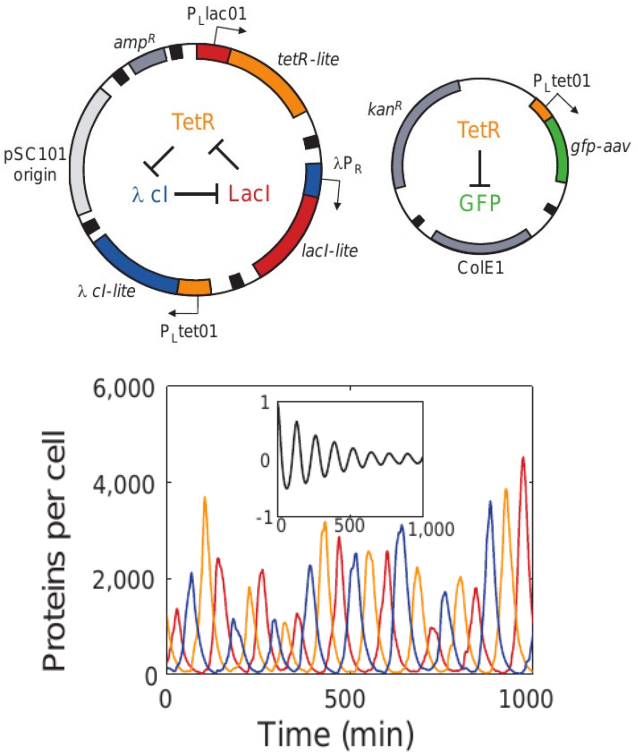
1952

Why using mathematical models?

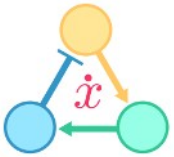
Describe	Explain	Predict





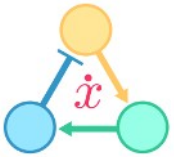


1917 1952 2000



What is a mathematical model?

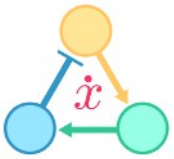
Wikipedia (April 17th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”



What is a mathematical model?

Wikipedia (April 17th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

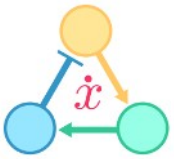
variables
[x]
Vmax
Kd
EC_{50}
length
$t_{1/2}$



What is a mathematical model?

Wikipedia (April 17th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

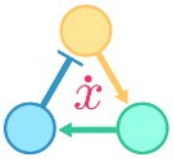
variables	relationships
[x]	$K_d = \frac{[A] \cdot [B]}{[AB]}$
Vmax	$d[X]/dt = k \cdot [Y]^2$
Kd	$\sum_i [X]_i - F(t) = 0$
EC_{50}	$k(t) \sim N(k, \sigma^2)$
length	If $\text{mass}_t > \text{threshold}$ then $\text{mass}_{t+\Delta t} = 0.5 \cdot \text{mass}$
$t_{1/2}$	



What is a mathematical model?

Wikipedia (April 17th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

variables	relationships	constraints
[x]	$K_d = \frac{[A] \cdot [B]}{[AB]}$	[x]>0
Vmax	$d[X]/dt = k \cdot [Y]^2$	Energy conservation
Kd	$\sum_i [X]_i - F(t) = 0$	Boundary conditions (v < upper limit)
EC_{50}	$k(t) \sim N(k, \sigma^2)$	Objective functions (maximise ATP)
length	If $\text{mass}_t > \text{threshold}$ then $\text{mass}_{t+\Delta t} = 0.5 \cdot \text{mass}$	Initial conditions
$t_{1/2}$		



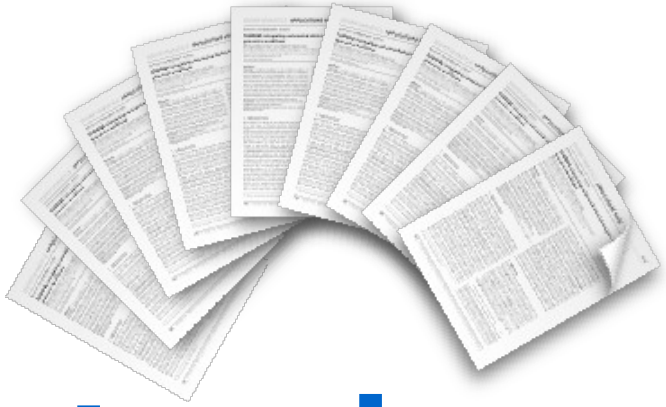
What is a mathematical model?

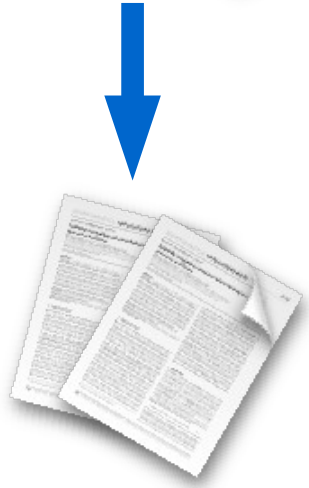
Wikipedia (April 17th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

variables	relationships	constraints
[x]	$K_d = \frac{[A] \cdot [B]}{[AB]}$	[x]>0
Vmax	$d[X]/dt = k \cdot [Y]^2$	Energy conservation
Kd	$\sum_i [X]_i - F(t) = 0$	Boundary conditions (v < upper limit)
EC_{50}	$k(t) \sim N(k, \sigma^2)$	Objective functions (maximise ATP)
length	If $\text{mass}_t > \text{threshold}$ then $\text{mass}_{t+\Delta t} = 0.5 \cdot \text{mass}$	Initial conditions
$t_{1/2}$		

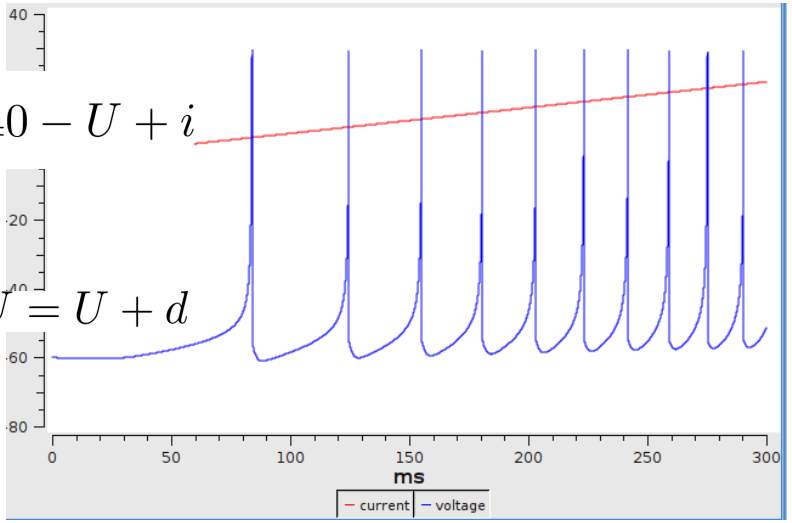
Different types: Dynamical models, logical models, rule-based models, multi-agent models, statistical models, etc.

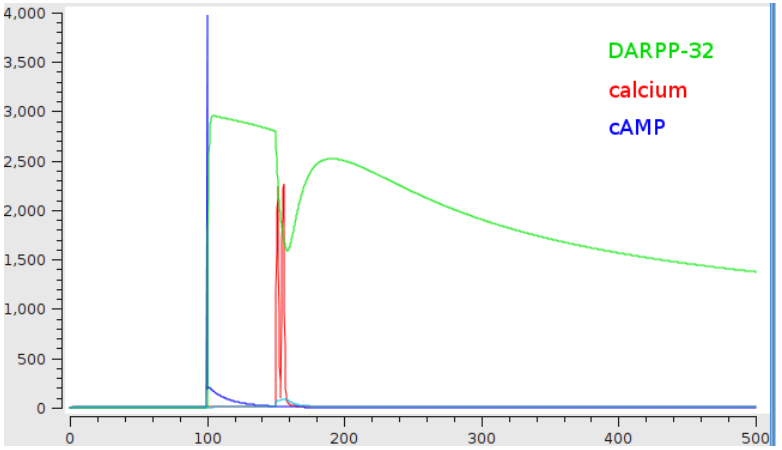
Knowledge-based model creation

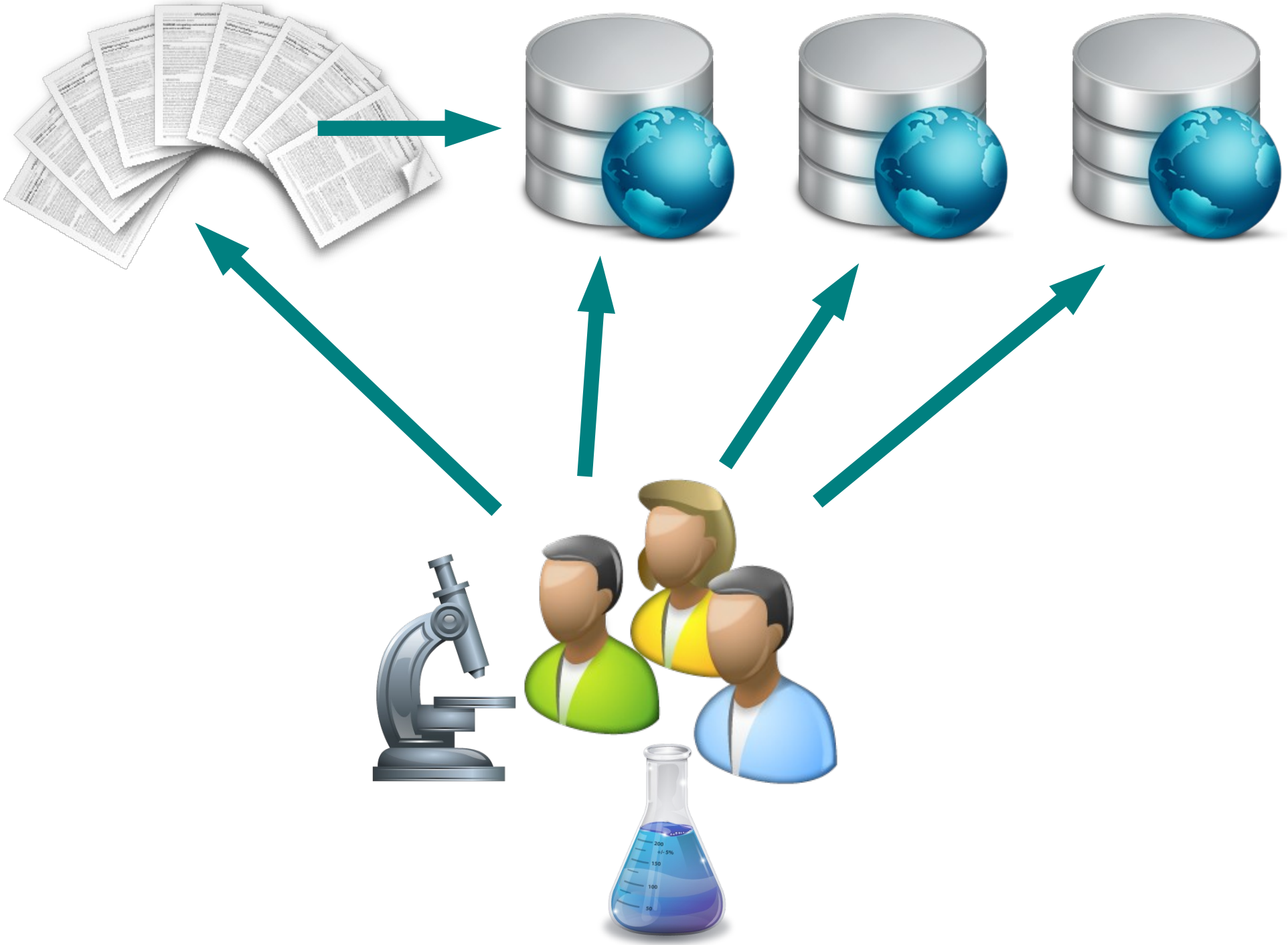


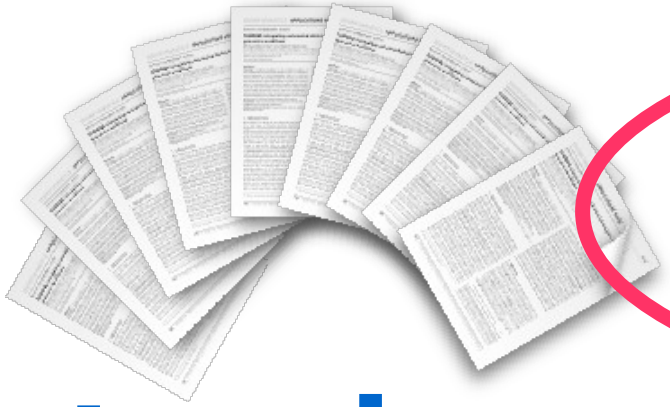


$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - U + i$$
$$\frac{dU}{dt} = a(bv - U)$$
$$v > V_{\text{thresh}} \Rightarrow v = c; U = U + d$$





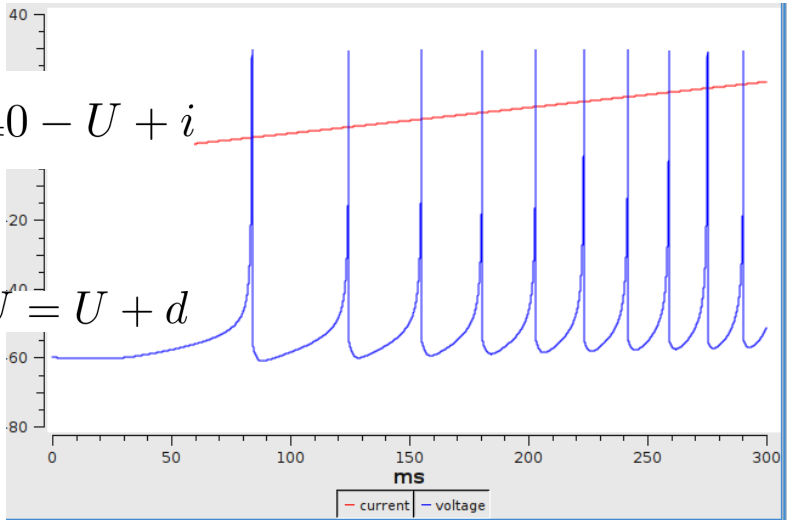


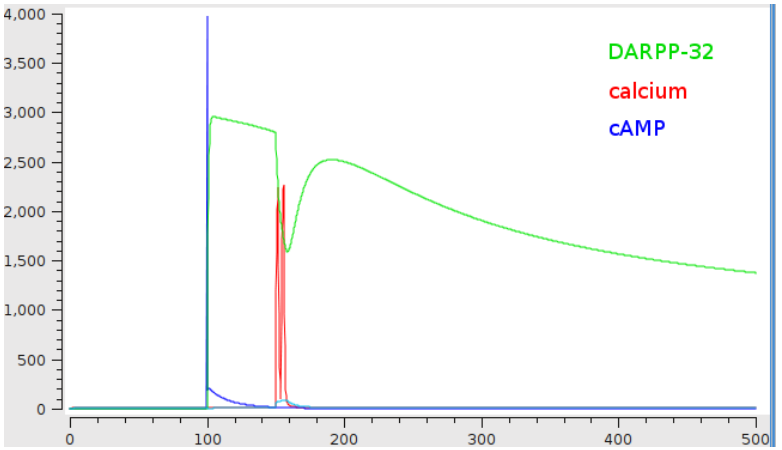


$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - U + i$$

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

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





EMBL-EBI Services Research

Training

Training Home | Train at EBI | Train outside EBI | Public events | Train Online | Abo

I want my pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes)

Joint EMBL-EBI-Wellcome Trust Course: In Silico Systems Biology

Venue:
European Bioinformatics Institute, Cambridge, CB10 1SD
United Kingdom
See map: Google Maps
Date: Tuesday, June 25, 2013 - Saturday, June 29, 2013
Organizers:
Laura Emery , EMBL-EBI, UK
Julio Saez-Rodriguez , EMBL-EBI
Nicolas Le Novère , EMBL-EBI
Admin support:
Wellcome Trust Advanced Courses , Wellcome Trust Advanced Courses, UK
Registration Opens Date: Thursday, August 23, 2012
Registration Deadline: Friday, March 8, 2013
Participation: Open application with selection

Overview | Programme | Registration | Trainers |

EMBL-EBI Services Research

Training

Training Home | Train at EBI | Train outside EBI | Public events | Train Online | Abo

I want my pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes)

Joint EMBL-EBI-Wellcome Trust Course: In Silico Systems Biology

Venue:
European Bioinformatics Institute, Cambridge, CB10 1SD
United Kingdom
See map: Google Maps
Date: Tuesday, June 25, 2013 - Saturday, June 29, 2013
Organizers:
Laura Emery , EMBL-EBI, UK
Julio Saez-Rodriguez , EMBL-EBI
Nicolas Le Novère , EMBL-EBI
Admin support:
Wellcome Trust Advanced Courses , Wellcom
Registration Opens Date: Thursday, Augu
Registration Deadline: Friday, March 8, 20
Participation: Open application with select

Overview | Programme | Registration | Traine

Starting on July 1, 2011 the KEGG FTP site for academic users will be transferred from GenomeNet at Kyoto University to NPO Bioinformatics Japan, and it will be available only to paid subscribers. The publicly funded portion, the medicus directory, will continue to be freely accessible at GenomeNet. The KEGG FTP site for commercial customers managed by Pathway Solutions will remain unchanged. The new FTP site is available for free trial until the end of June.

Please register to learn more about the KEGG FTP subscription.

Thank you!

Minoru Kanehisa

EMBL-EBI Services Research

Training

Training Home | Train at EBI | Train outside EBI | Public events | Train Online | Abo

Joint EMBL-EBI-Wellcome Trust Course: In Silico Systems Biology

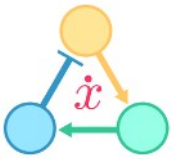
Venue:
European Bioinformatics Institute, Cambridge, CB10 1SD
United Kingdom
See map: Google Maps
Date: Tuesday, June 25, 2013 - Saturday, June 29, 2013
Organizers:
Laura Emery , EMBL-EBI, UK
Julio Saez-Rodriguez , EMBL-EBI
Nicolas Le Novère , EMBL-EBI
Admin support:
Wellcome Trust Advanced Courses
Registration Opens Date: Thurs
Registration Deadline: Friday, M
Participation: Open application wit

Overview | Progra | gistrat | Trainers |

I want my pathways from KEGG (Kyoto Encyclopedia of Genes and Genomes)

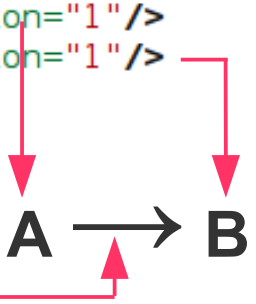
I want my pathways in SBML™

Babraham Institute



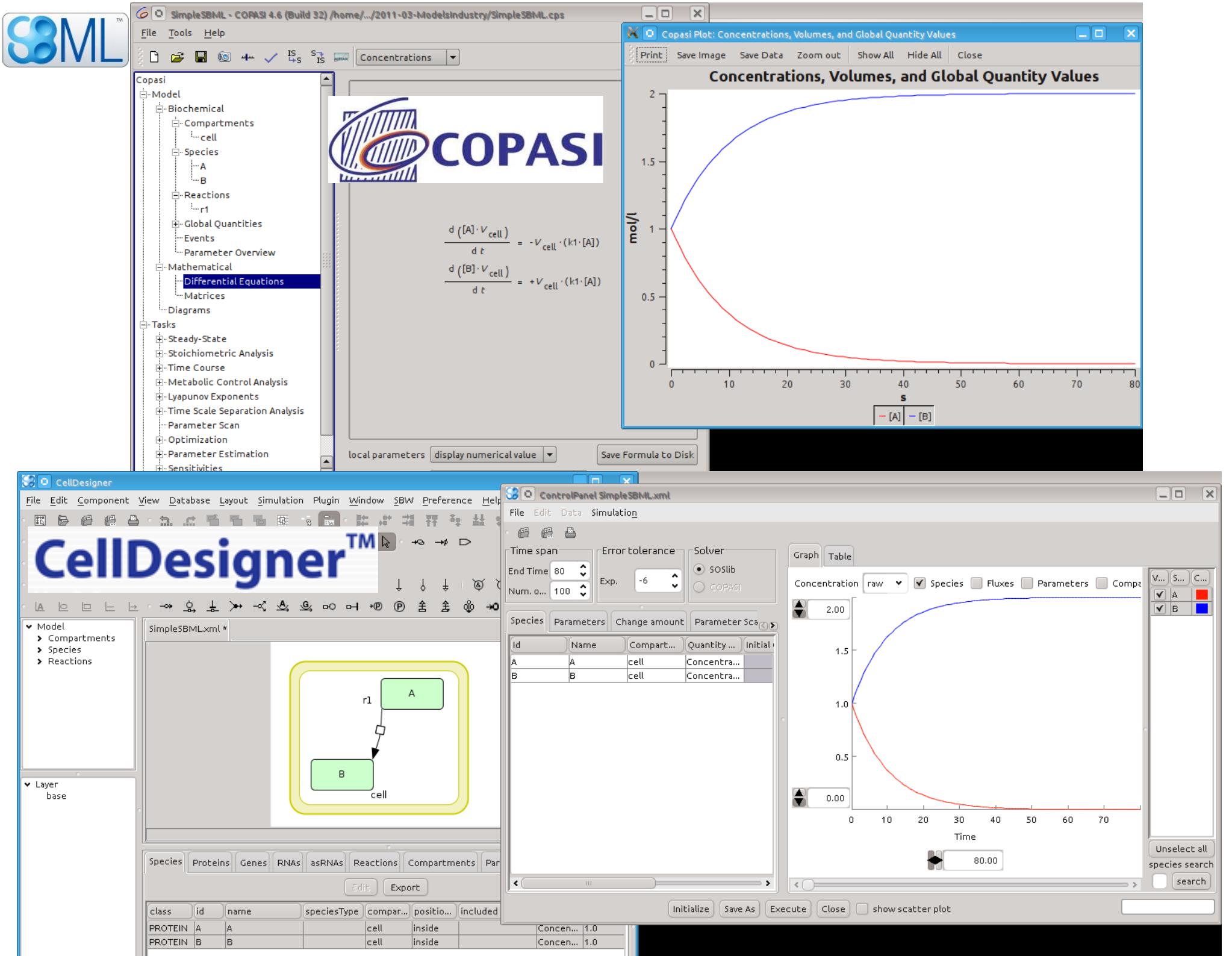
Systems Biology Markup Language

```
<?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>
```



$$v1 = k1 \times [A] \times V_{cell}$$

A very simple SBML file


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™
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

A community-driven global reconstruction of human metabolism

Ines Thiele[1,2,37], Neil Swainston[3,4,37], Ronan M T Fleming[1,5], Andreas Hoppe[6], Swagatika Sahoo[1], Maike K Aurich[1], Hulda Haraldsdottir[1], Monica L Mo[7], Ottar Rolfsson[1], Miranda D Stobbe[8,9], Stefan G Thorleifsson[1], Rasmus Agren[10], Christian Bölling[6], Sergio Bordel[10], Arvind K Chavali[11], Paul Dobson[12], Warwick B Dunn[3,13], Lukas Endler[14], David Hala[15], Michael Hucka[16], Duncan Hull[4], Daniel Jameson[3,4], Neema Jamshidi[7], Jon J Jonsson[5], Nick Juty[17], Sarah Keating[17], Intawat Nookaew[10], Nicolas Le Novère[17,18], Naglis Malys[3,19,20], Alexander Mazein[21], Jason A Papin[11], Nathan D Price[22], Evgeni Selkov, Sr[23], Martin I Sigurdsson[1], Evangelos Simeonidis[22,24], Nikolaus Sonnenschein[25], Kieran Smallbone[3,26], Anatoly Sorokin[21,27], Johannes H G M van Beek[28–30], Dieter Weichart[3,31], Igor Goryanin[21,32], Jens Nielsen[10], Hans V Westerhoff[3,28,33,34], Douglas B Kell[3,35], Pedro Mendes[3,4,36] & Bernhard Ø Palsson[1,7]

Multiple models of human metabolism have been reconstructed, but each represents only a subset of our knowledge. Here we describe Recon 2, a community-driven, consensus 'metabolic reconstruction', which is the most comprehensive representation of human metabolism that is applicable to computational modeling. Compared with its predecessors, the reconstruction has improved topological and functional features, including ~2× more reactions and ~1.7× more unique metabolites. Using Recon 2 we predicted changes in metabolite biomarkers for 49 inborn errors of metabolism with 77% accuracy when compared to experimental data. Mapping metabolomic data and drug information onto Recon 2 demonstrates its potential for integrating and analyzing diverse data types. Using protein expression data, we automatically generated a compendium of 65 cell type–specific models, providing a basis for manual curation or investigation of cell-specific metabolic properties. Recon 2 will facilitate many future biomedical studies and is freely available at http://humanmetabolism.org/.

An understanding of metabolism is fundamental to comprehending the phenotypic behavior of all living organisms, including humans, where metabolism is integral to health and is involved in much of human disease. High quality, genome-scale 'metabolic reconstructions' are at the heart of bottom-up systems biology analyses and represent the entire network of metabolic reactions that a given organism is known to exhibit[1]. The metabolic-network reconstruction procedure is now well-established[2] and has been applied to a growing number of model organisms[3]. Metabolic reconstructions allow for the conversion of biological knowledge into a mathematical format and the subsequent computation of physiological states[1,4,5] to address a variety of scientific and applied questions[3,6]. Reconstructions enable network-wide mechanistic investigations of the genotype-phenotype relationship. A high-quality reconstruction of the metabolic network is thus

[1]Center for Systems Biology, University of Iceland, Reykjavik, Iceland. [2]Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, Reykjavik, Iceland. [3]Manchester Centre for Integrative Systems Biology, University of Manchester, Manchester Institute of Biotechnology, Manchester, UK. [4]School of Computer Science, University of Manchester, Manchester, UK. [5]Department of Biochemistry and Molecular Biology, University of Iceland, Reykjavik, Iceland. [6]Computational Systems Biochemistry Group, Charité–Universitätsmedizin Berlin, Berlin, Germany. [7]Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. [8]Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands. [9]Netherlands Bioinformatics Centre, Nijmegen, the Netherlands. [10]Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. [11]Department of Biomedical Engineering, University of Virginia, Charlottesville, Virginia, USA. [12]Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, UK. [13]Centre for Advanced Discovery and Experimental Therapeutics (CADET), Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. [14]Institute for Theoretical Chemistry, University of Vienna, Vienna, Austria. [15]Department of Biology, University of North Texas, Denton, Texas, USA. [16]Computing and Mathematical Sciences Department, California Institute of Technology, Pasadena, California, USA. [17]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK. [18]Babraham Institute, Babraham Research Campus, Cambridge, UK. [19]Faculty of Life Sciences, University of Manchester, Manchester, UK. [20]School of Life Sciences, Gibbet Hill Campus, University of Warwick, Coventry, UK. [21]School of Informatics, University of Edinburgh, Edinburgh, UK. [22]Institute for Systems Biology, Seattle, Washington, USA. [23]Genome Designs, Inc., Walnut Creek, California, USA. [24]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, Esch-sur-Alzette, Luxembourg. [25]School of Engineering and Science, Jacobs University Bremen, Bremen, Germany. [26]School of Mathematics, University of Manchester, Manchester, UK. [27]Institute of Cell Biophysics, Russian Academy of Sciences, Moscow region, Pushchino, Russia. [28]Department of Molecular Cell Physiology, Vrije Universiteit, Amsterdam, the Netherlands. [29]Section Medical Genomics, Department of Clinical Genetics, Vrije Universiteit University Medical Centre, Amsterdam, the Netherlands. [30]Netherlands Consortium for Systems Biology, Amsterdam, The Netherlands. [31]School of Dentistry, The University of Manchester, Manchester, UK. [32]Okinawa Institute Science and Technology, Okinawa, Japan. [33]School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, UK. [34]Swammerdam Institute for Life Sciences, Faculty of Science, University of Amsterdam, Amsterdam, The Netherlands. [35]School of Chemistry, The University of Manchester, Manchester, UK. [36]Virginia Bioinformatics Institute, Virginia Tech, Blacksburg, Virginia, USA. [37]These authors contributed equally to this work. Correspondence should be addressed to I.T. (ines.thiele@gmail.com).

Received 7 September 2012; accepted 19 December 2012; published online 3 March 2013; doi:10.1038/nbt.2488

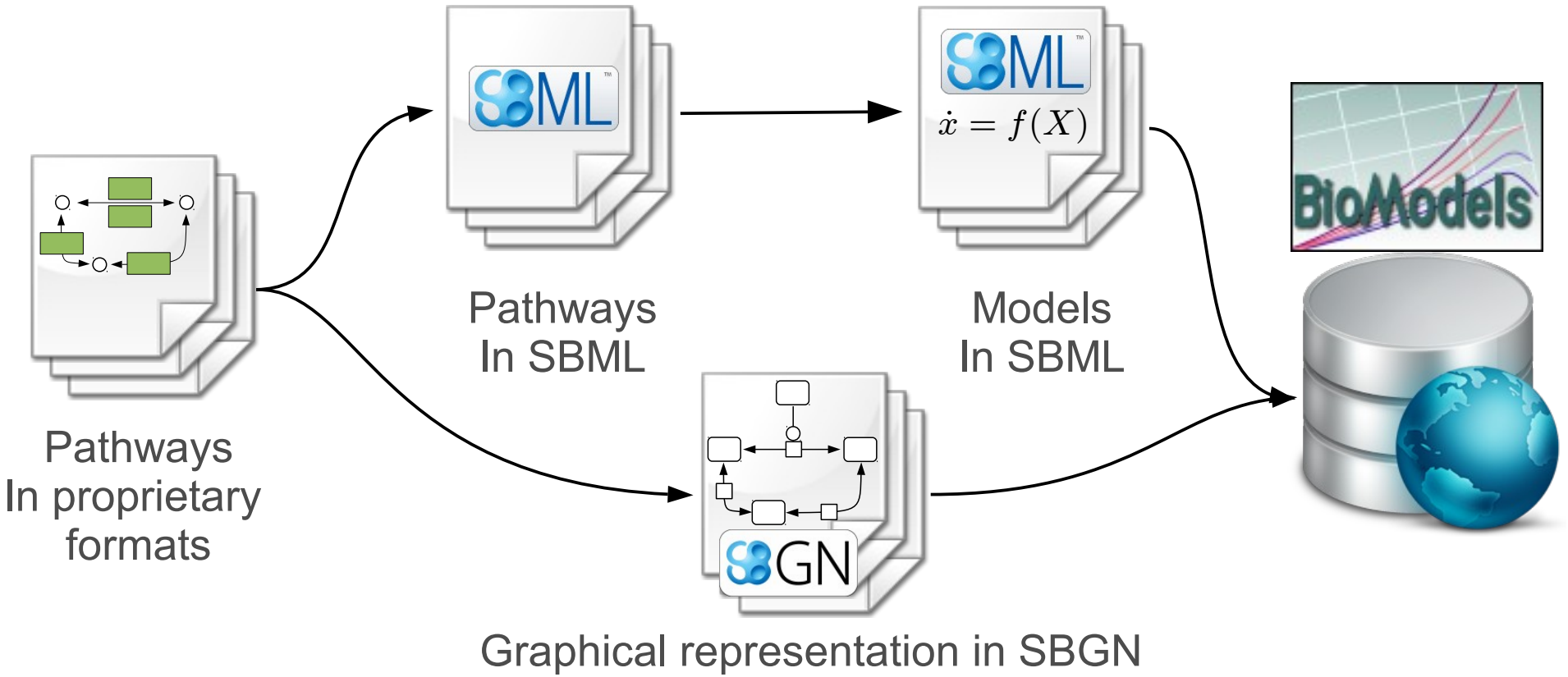
- 8 compartments
- 5 063 metabolites
- 2 194 proteins
- 7 440 reactions

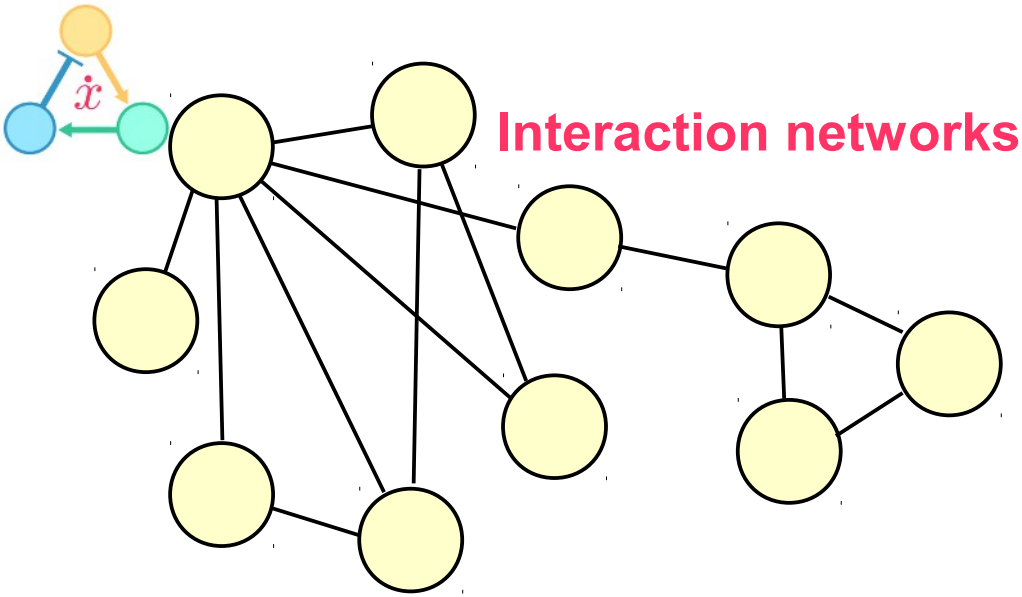
A not so simple SBML file (Recon2)

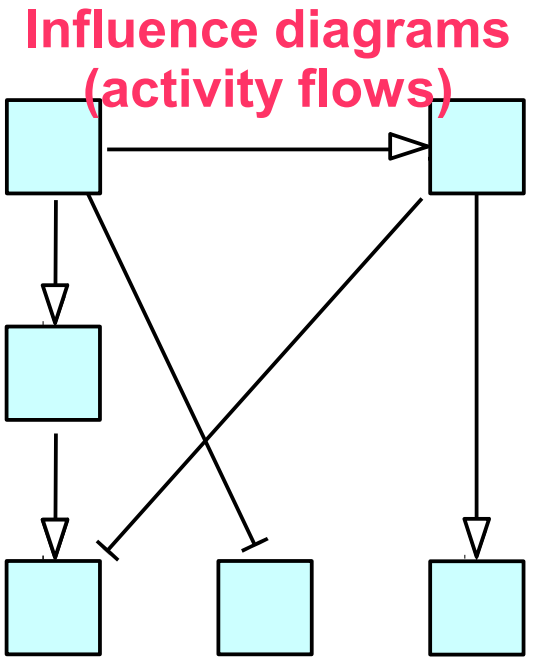
Babraham Institute

Aims of the Path2Models project

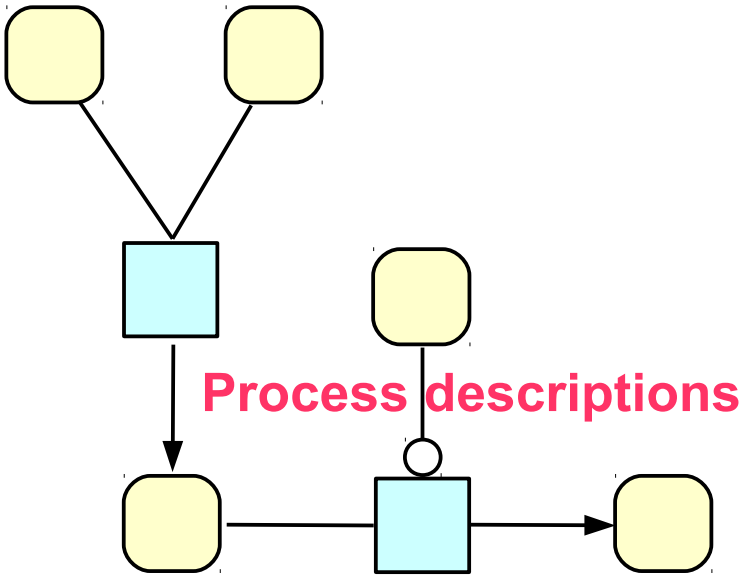
- To re-use existing pathway data to generate biochemically based models
- To provide a starting point to model as many biochemical pathways as possible in as many species as possible
- To provide models in a standard format readable by most systems biology software

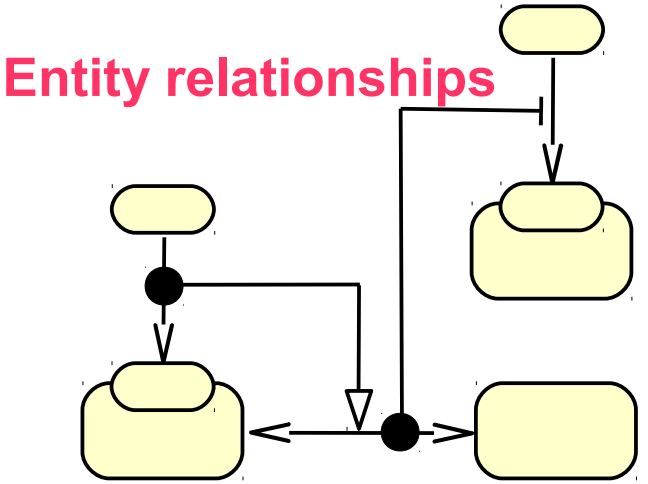




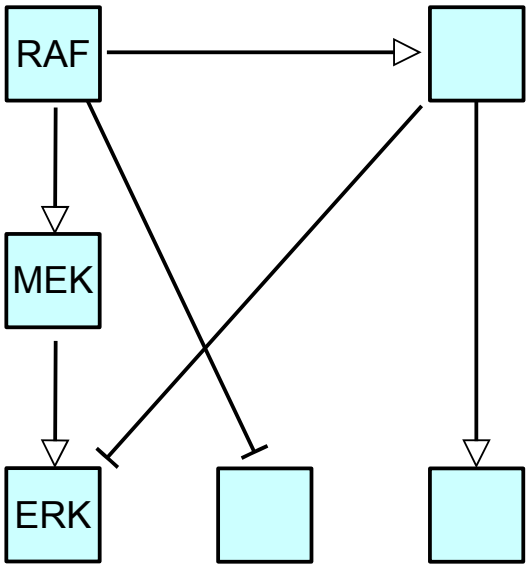


Four types of networks in systems biology

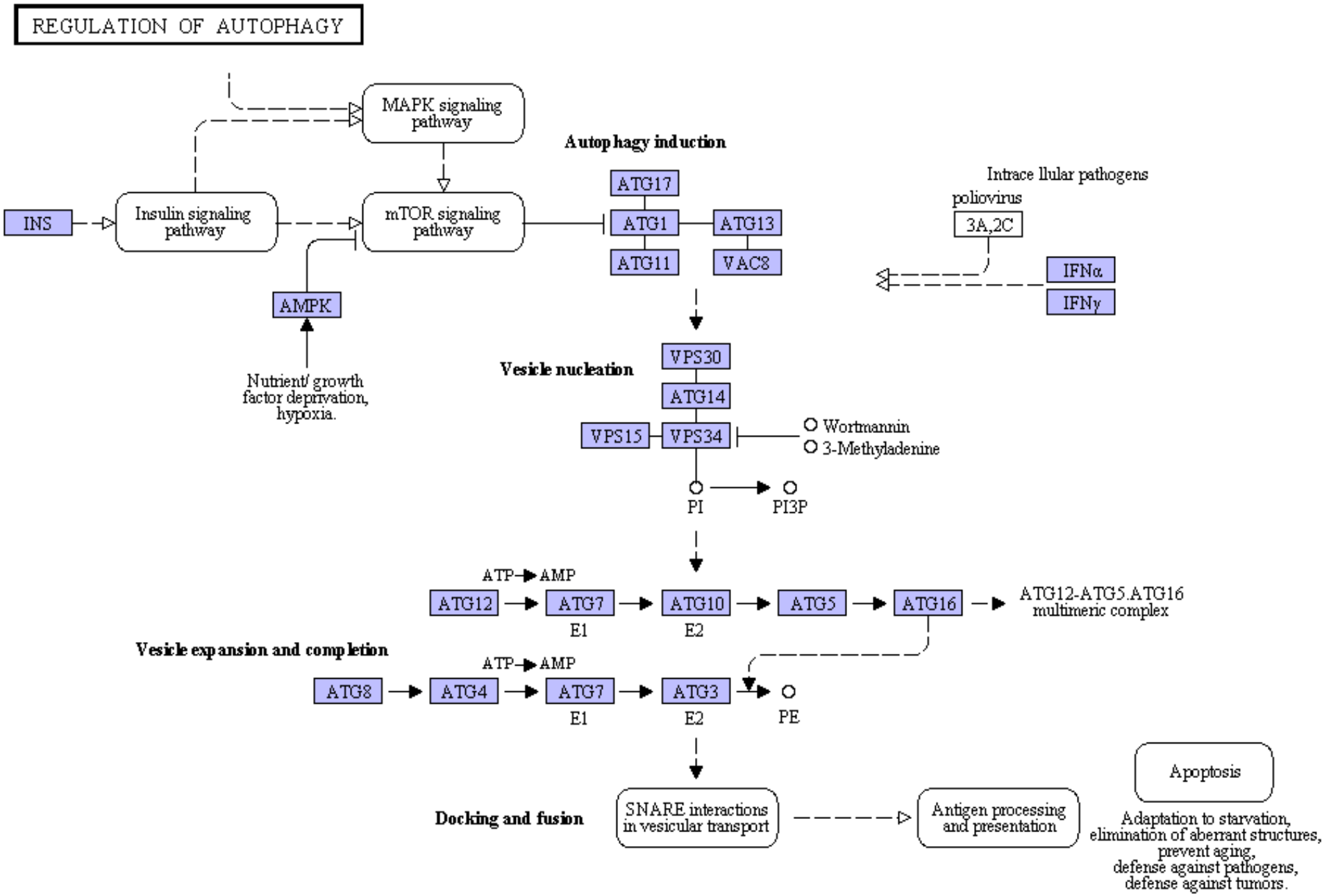


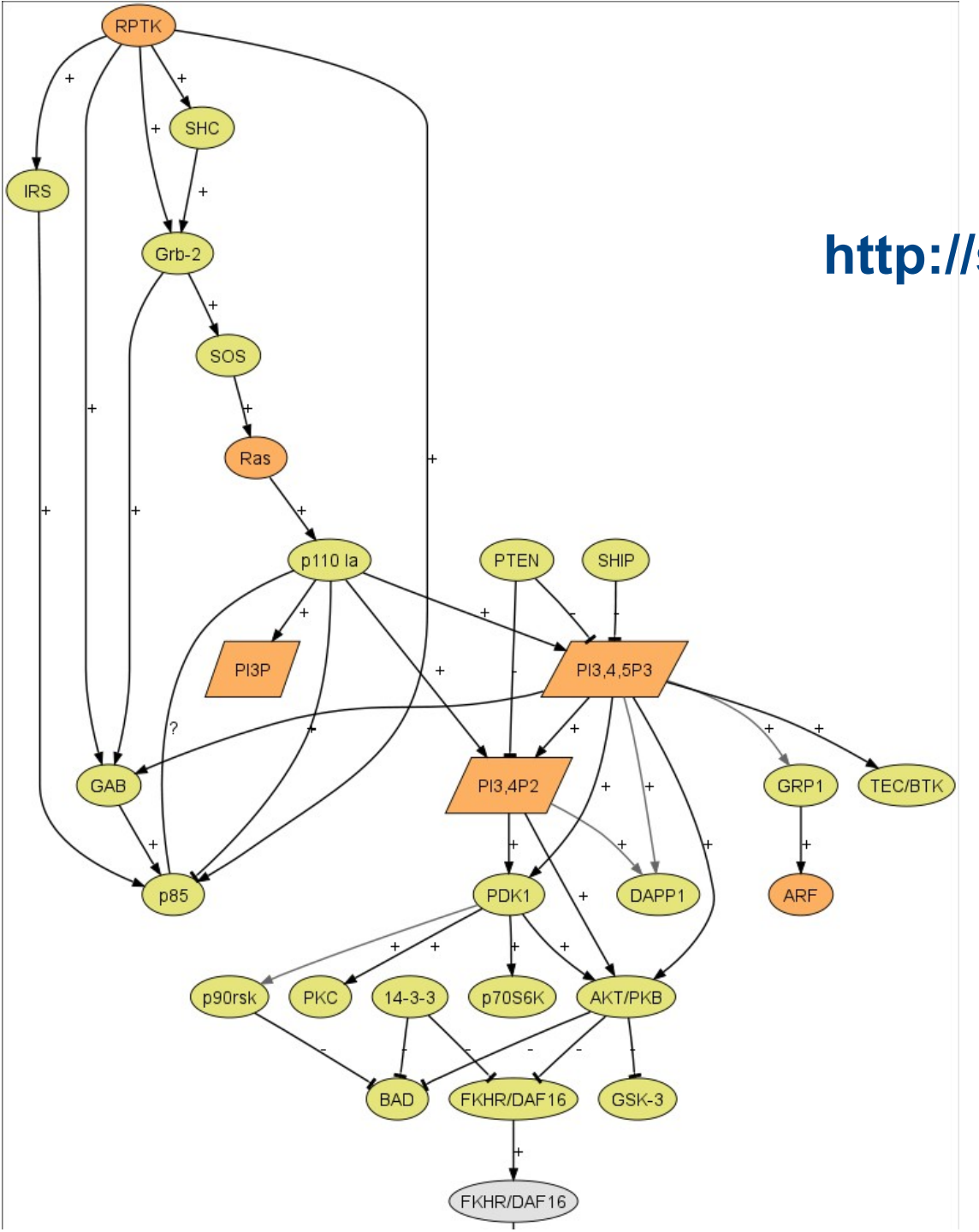


Influence diagrams (activity-flows)



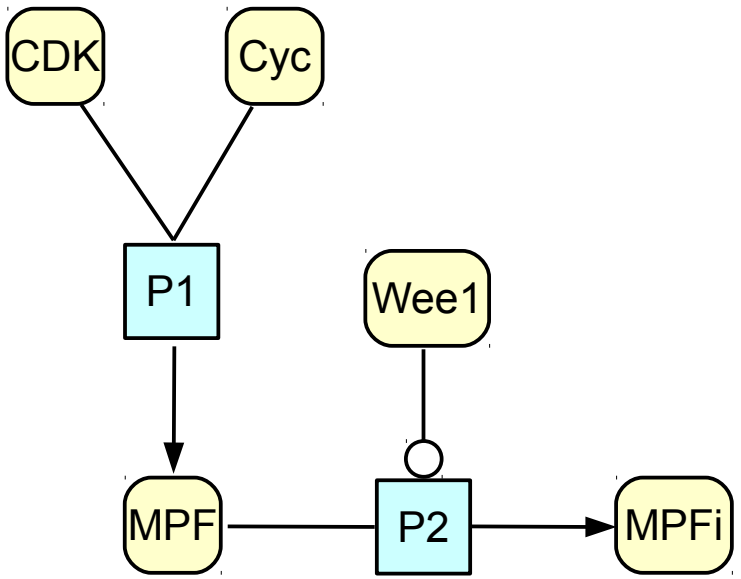
- Signalling pathways, gene regulatory networks
- Nodes represent activities
- Directional
- Sequential
- Non-mechanistic
- Logical modelling


REGULATION OF AUTOPHAGY
MAPK signaling pathway
Autophagy induction
INS
Insulin signaling pathway
mTOR signaling pathway
ATG17
ATG1
ATG13
ATG11
VAC8
Intracellular pathogens
poliovirus
3A,2C
IFNα
IFNγ
AMPK
Nutrient/ growth factor deprivation, hypoxia.
Vesicle nucleation
VPS30
ATG14
VPS15
VPS34
Wortmannin
3-Methyladenine
PI
PI3P
ATP
AMP
ATG12
ATG7
ATG10
ATG5
ATG16
E1
E2
ATG12-ATG5.ATG16 multimeric complex
Vesicle expansion and completion
ATG8
ATG4
ATG3
PE
Docking and fusion
SNARE interactions in vesicular transport
Antigen processing and presentation
Apoptosis
Adaptation to starvation,
elimination of aberrant structures,
prevent aging,
defense against pathogens,
defense against tumors.
04140 5/11/09
(c) Kanehisa Laboratories



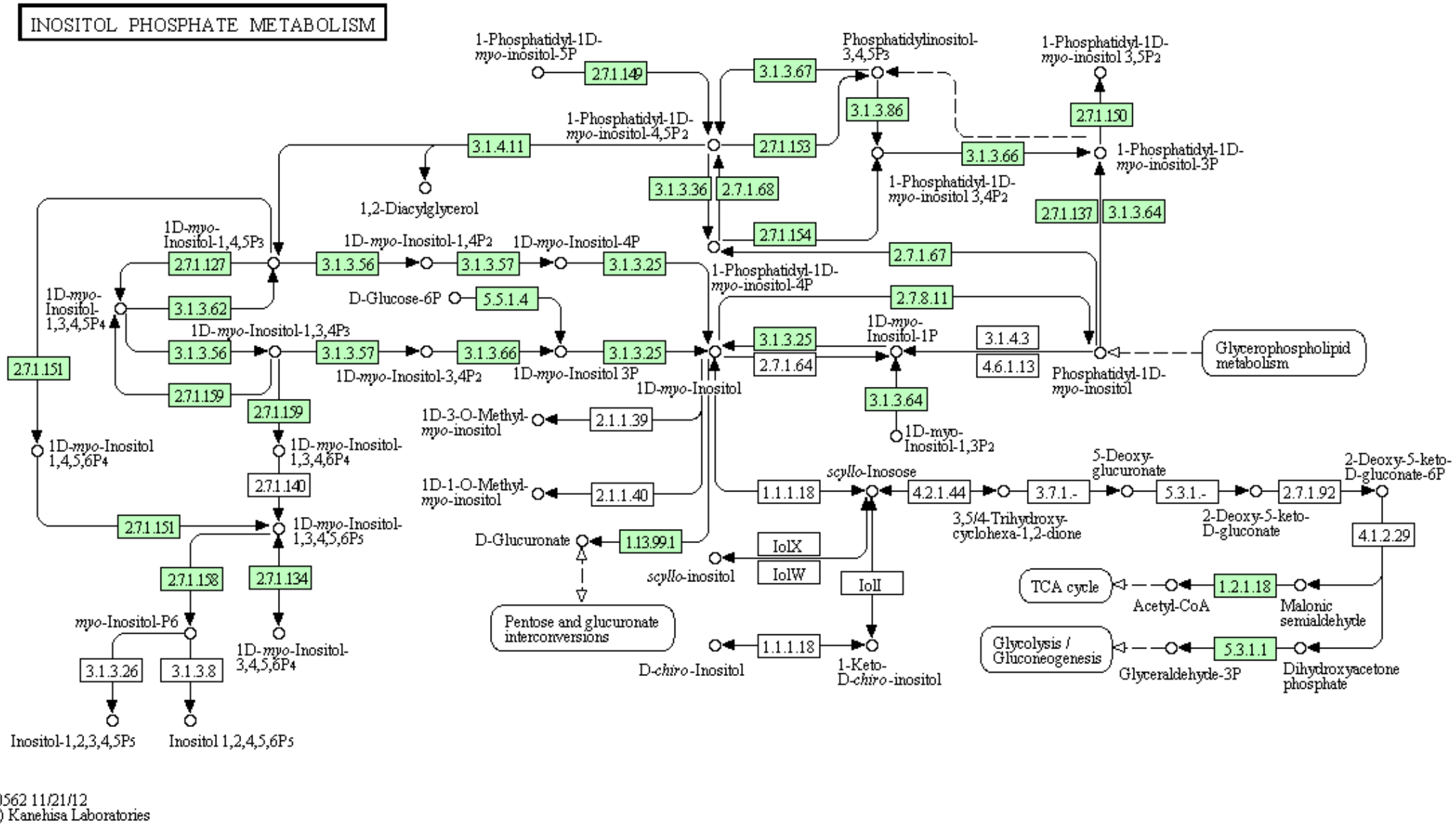
http://stke.sciencemag.org/cm/

Process Descriptions

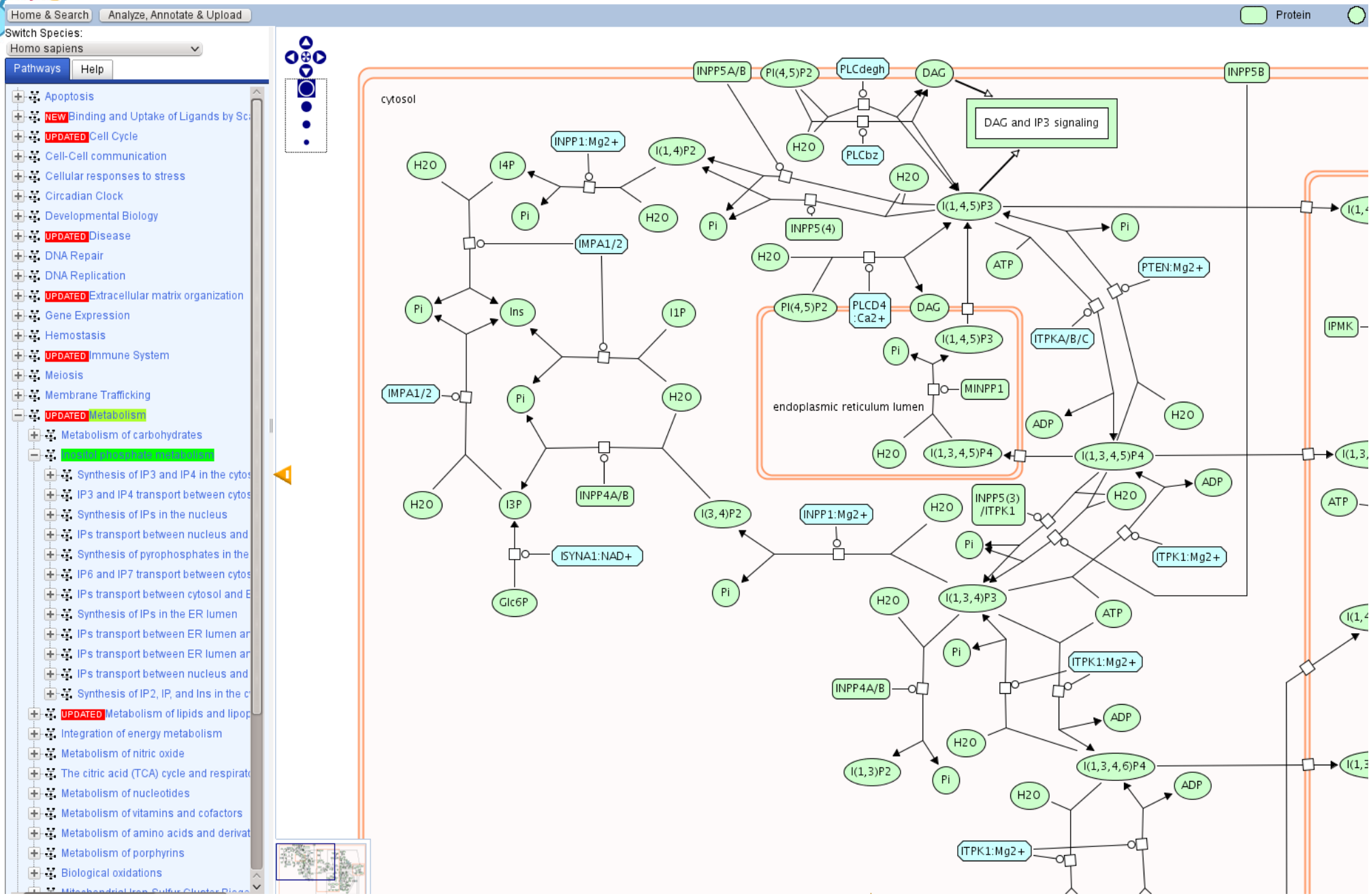


- Biochemistry, Metabolic networks
- Nodes represent populations of reactants and reactions
- Directional
- Sequential
- Mechanistic
- Subjected to combinatorial explosion
- Process modelling (ODE, SSA)

http://www.genome.jp/kegg/pathway.html



http://www.reactome.org



Three parallel workflows

Signalling Pathways

+

→

Logical models
of individual signalling pathways

Three parallel workflows

Signalling Pathways

+

→ **Logical models of individual signalling pathways**

Metabolic Networks

→ **Chemical kinetics models of individual metabolic pathways**

Three parallel workflows

Signalling Pathways

+

→ **Logical models of individual signalling pathways**

Metabolic Networks

→ **Chemical kinetics models of individual metabolic pathways**

Metabolic Networks

+

→ **Flux Balance Analysis of whole genome reconstructions**

From KEGG metabolic pathways to full kinetic models

KGML

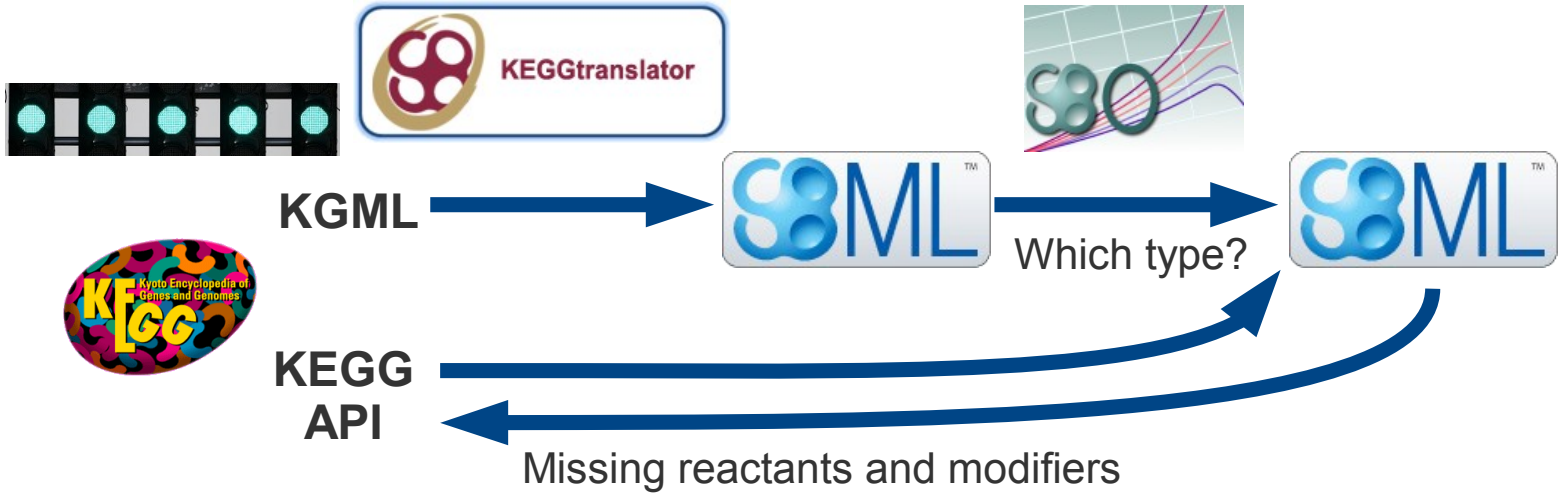
From KEGG metabolic pathways to full kinetic models

KEGGtranslator

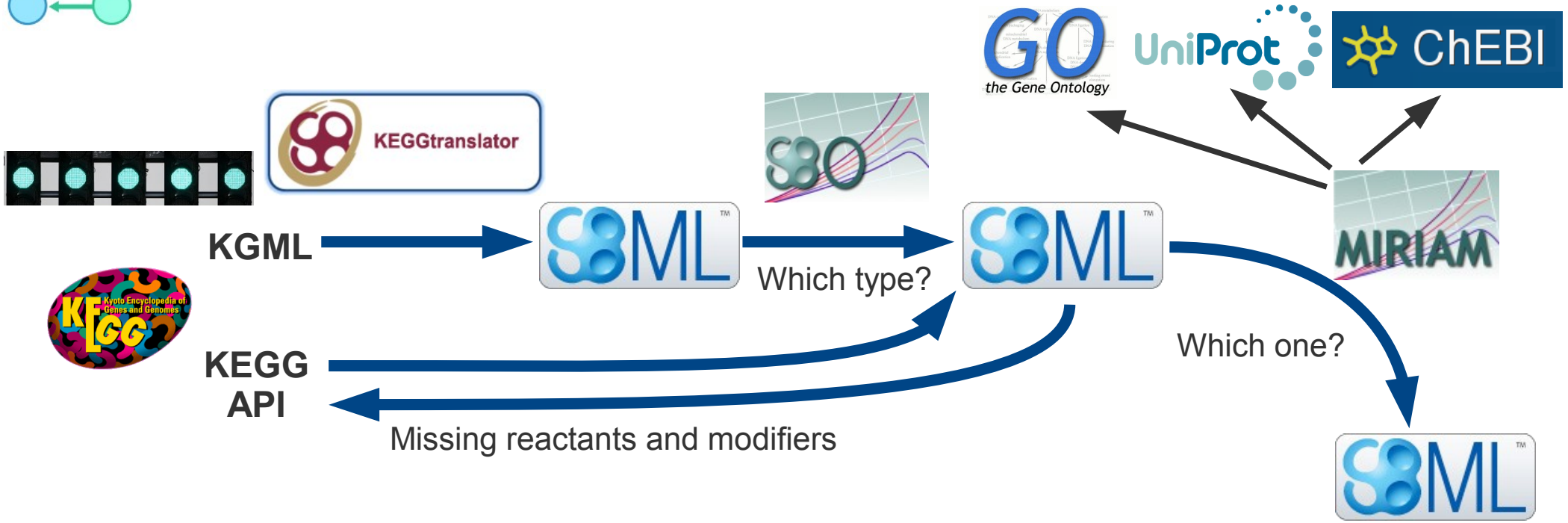
KGML → SBML

From KEGG metabolic pathways to full kinetic models

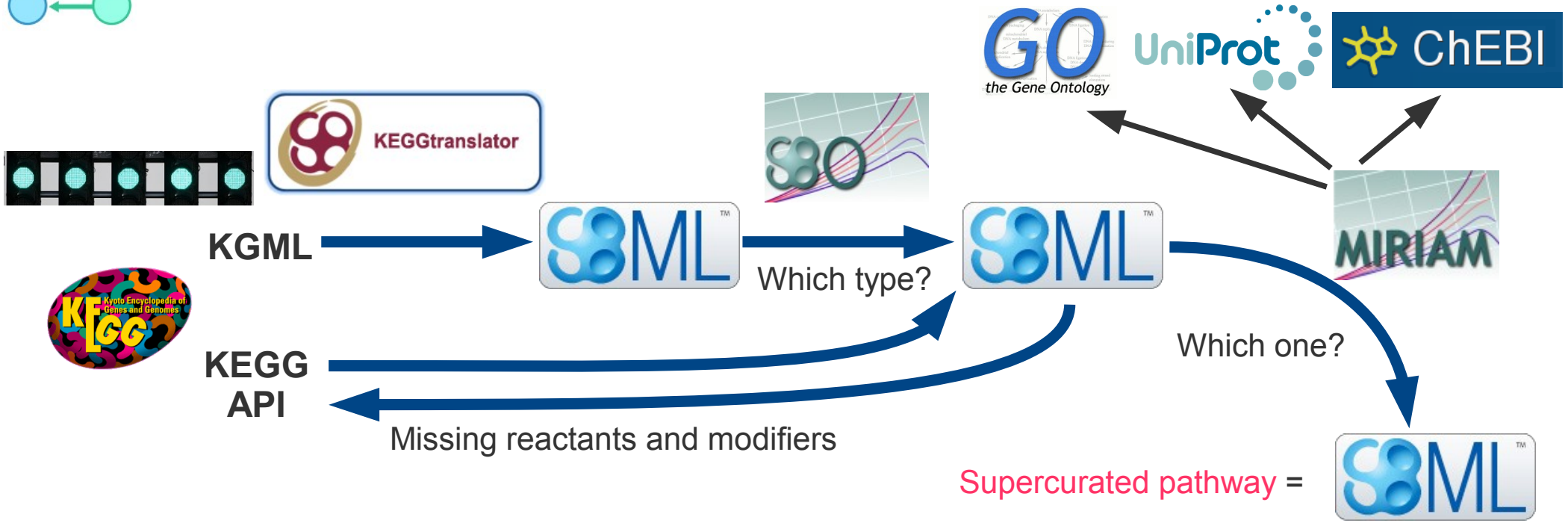
From KEGG metabolic pathways to full kinetic models



From KEGG metabolic pathways to full kinetic models



From KEGG metabolic pathways to full kinetic models



From KEGG metabolic pathways to full kinetic models

KEGGtranslator

KGML

KEGG API

Which type?

Which one?

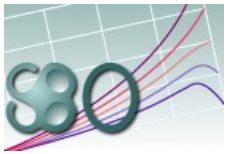
Missing reactants and modifiers

Supercurated pathway =

SBMLsqueezer

From KEGG metabolic pathways to full kinetic models

KEGGtranslator



GO the Gene Ontology

UniProt

ChEBI

KGML

SBML

Which type?

SBML

MIRIAM

KEGG API

Missing reactants and modifiers

Which one?

Supercurated pathway = SBML

Ab initio kinetics

SBML

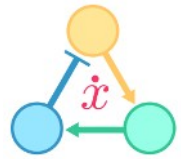
SBML

SBMLsqueezer

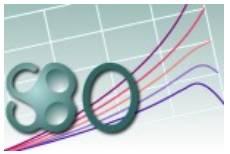
SABIO-RK

Babraham Institute

From KEGG metabolic pathways to full kinetic models



KEGGtranslator



GO the Gene Ontology

UniProt

ChEBI

KGML

SBML

Which type?

SBML

MIRIAM

KEGG API

Missing reactants and modifiers

Which one?

Supercurated pathway =

SBML

Ab initio kinetics

SBML

SBML

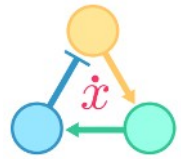
=

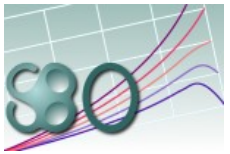
Chemical kinetics model

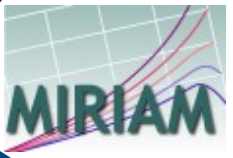
SBMLsqueezer

SABIO-RK

From KEGG metabolic pathways to full kinetic models







KGML

Which type?

KEGG API

Missing reactants and modifiers

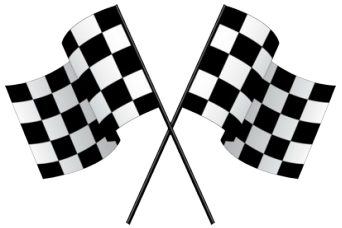
Which one?

Supercurated pathway =

Ab initio kinetics

=

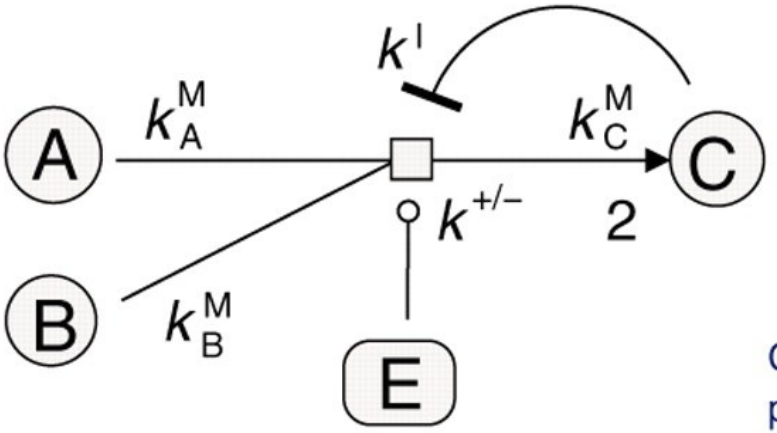
Chemical kinetics model



SBMLsqueezer

SABIO-RK

Metabolic reactions: Common modular rate-law



Reaction rate

Enzyme level

Stoichiometry, 3 parametrizations

$$v = u\, f\, \frac{T}{D + D^{reg}}$$

Complete or partial regulation

5 rate laws

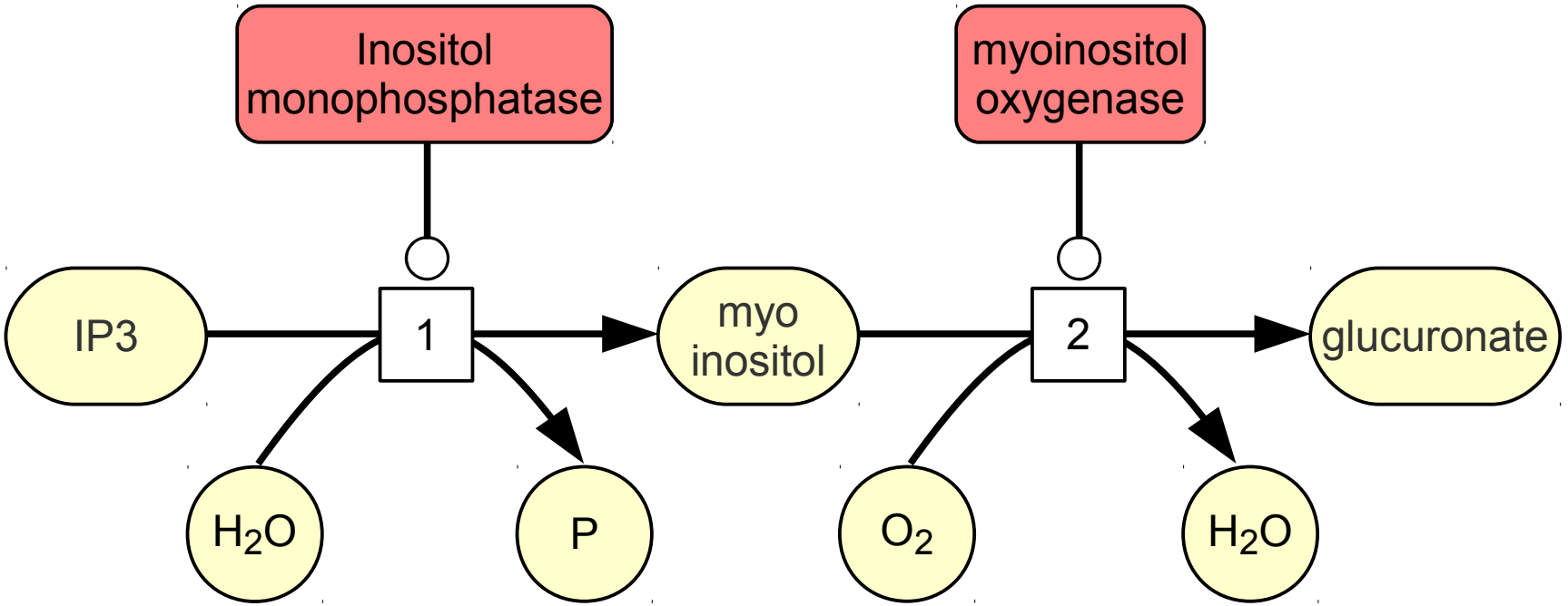
Specific regulation

- 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

Unparametrised Vs. known kinetics



$$v1 = [IMPA] \times k_{\text{cat}} \times \frac{[\text{IP3}]}{K_{\text{m}} + [\text{IP3}]}$$

To estimate

$$v2 = V_{\text{max}} \times \frac{[\text{myoinositol}]}{K_{\text{m}} + [\text{myoinositol}]}$$

$3.5 \cdot 10^{-5}\,\text{mol} \cdot \text{s}^{-1} \cdot \text{g}^{-1}$

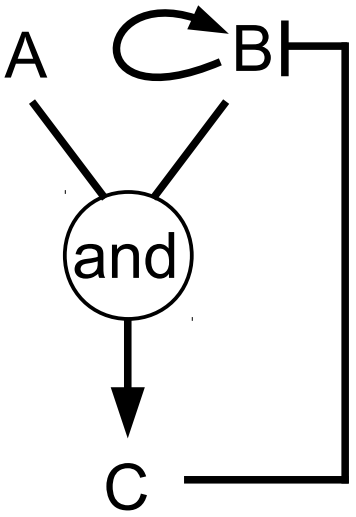
$0.0033\,\text{M}$

SABIO-RK

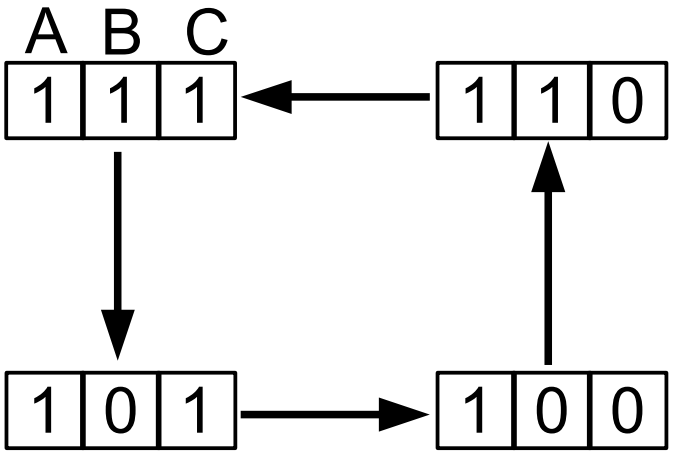
Model BMID000000038685 "Inositol phosphate metabolism"

Signalling pathways: Logical models

- 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
- One can add delays, inputs etc.

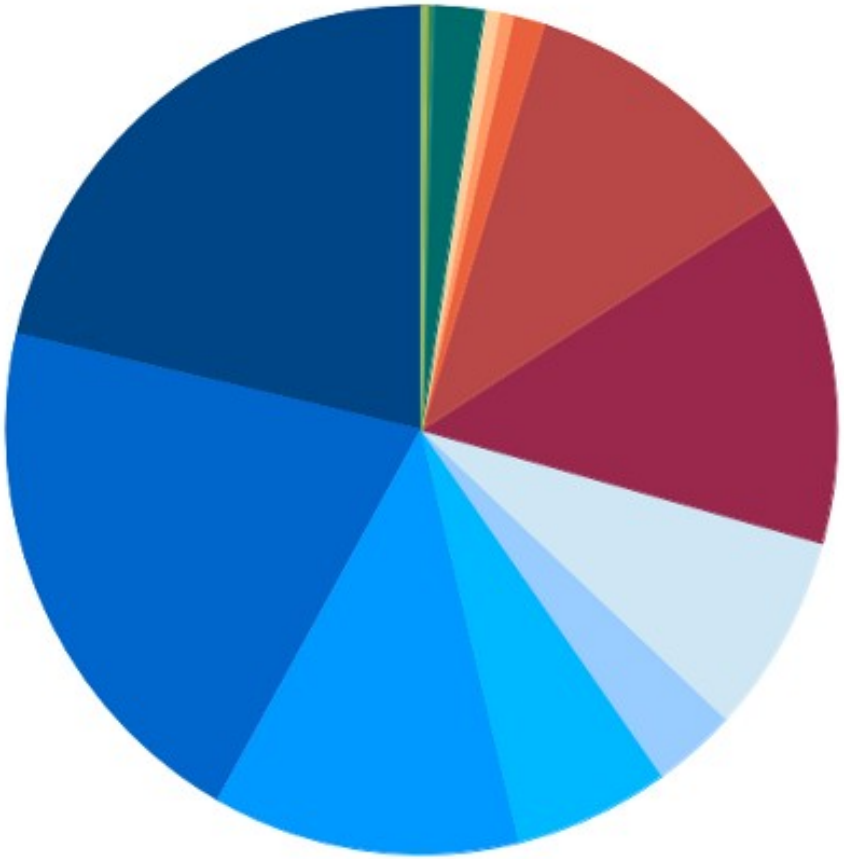


Influence diagram



state diagram

Gene Ontology Coverage



23nd BioModels Database release

- 11th August 2012 – 2nd release of Path2Models data
- 112 898 common modular rate law models of metabolic networks
- 27 306 qualitative models of signalling pathways
- 1 846 whole genome flux balance analysis models
- 239 models for human, 234 models for mouse
- 10 547 589 mathematical relations
- 444 024 053 cross references
- http://www.ebi.ac.uk/biomodels-main/path2models

Inositol phosphate metabolism - Mus musculus

Download SBML | Additional files | Send feedback

Model information

Identifier: BMID000000038685
Format: SBML L3 V1 (Layout)
Project: path2models
Categories: metabolic
Submission: 17 May 2012 22:37:29 UTC
Last modified: 10 Dec 2012 03:55:58 UTC
Published: 19 May 2012 23:49:21 UTC

Annotations

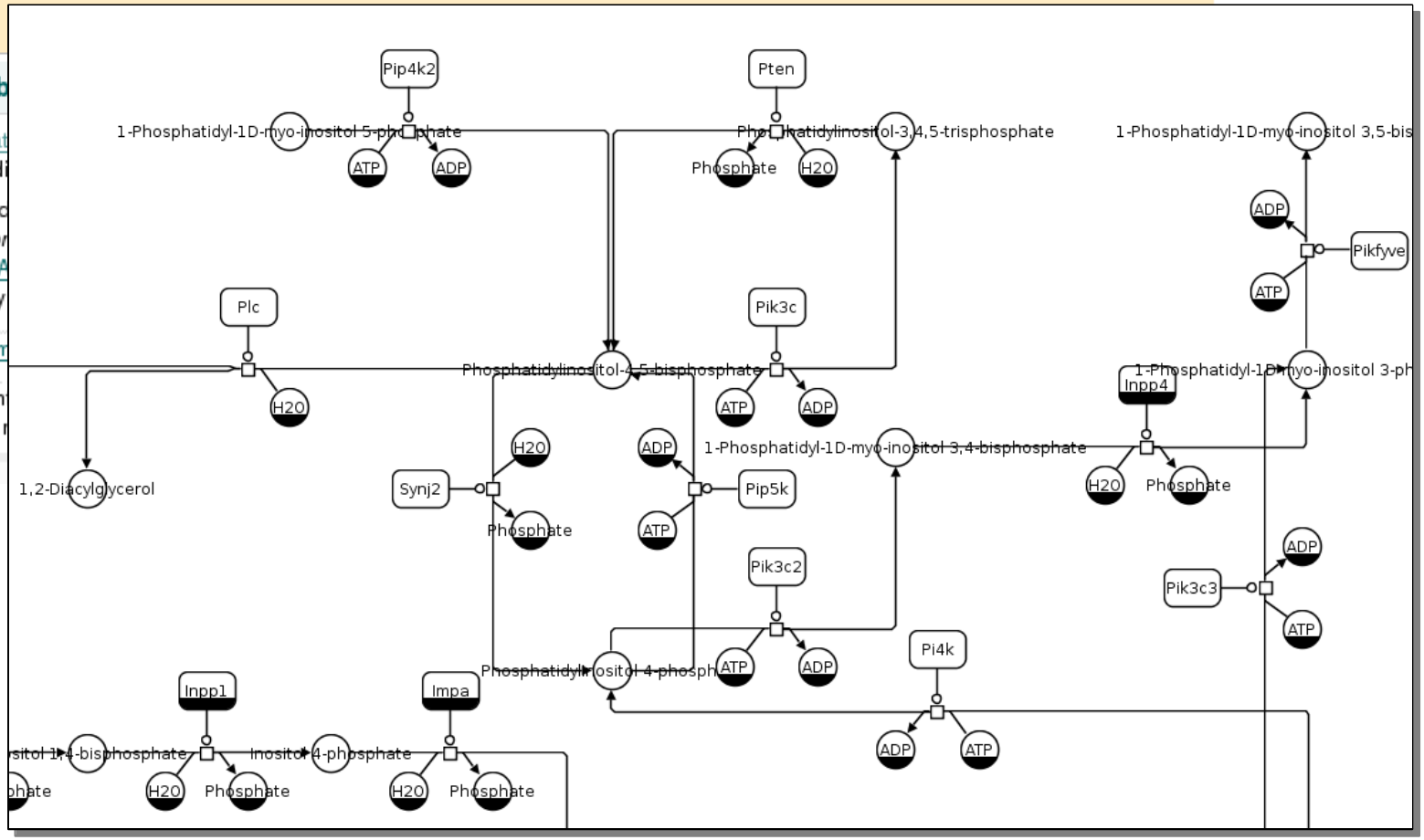
isDescribedBy	inositol metabolic process	*Gene Ontology*
occursIn	Mus musculus	*Taxonomy*
isDerivedFrom	Inositol phosphate metabolism	*KEGG Pathway*

Model of "Inositol phosphate metab

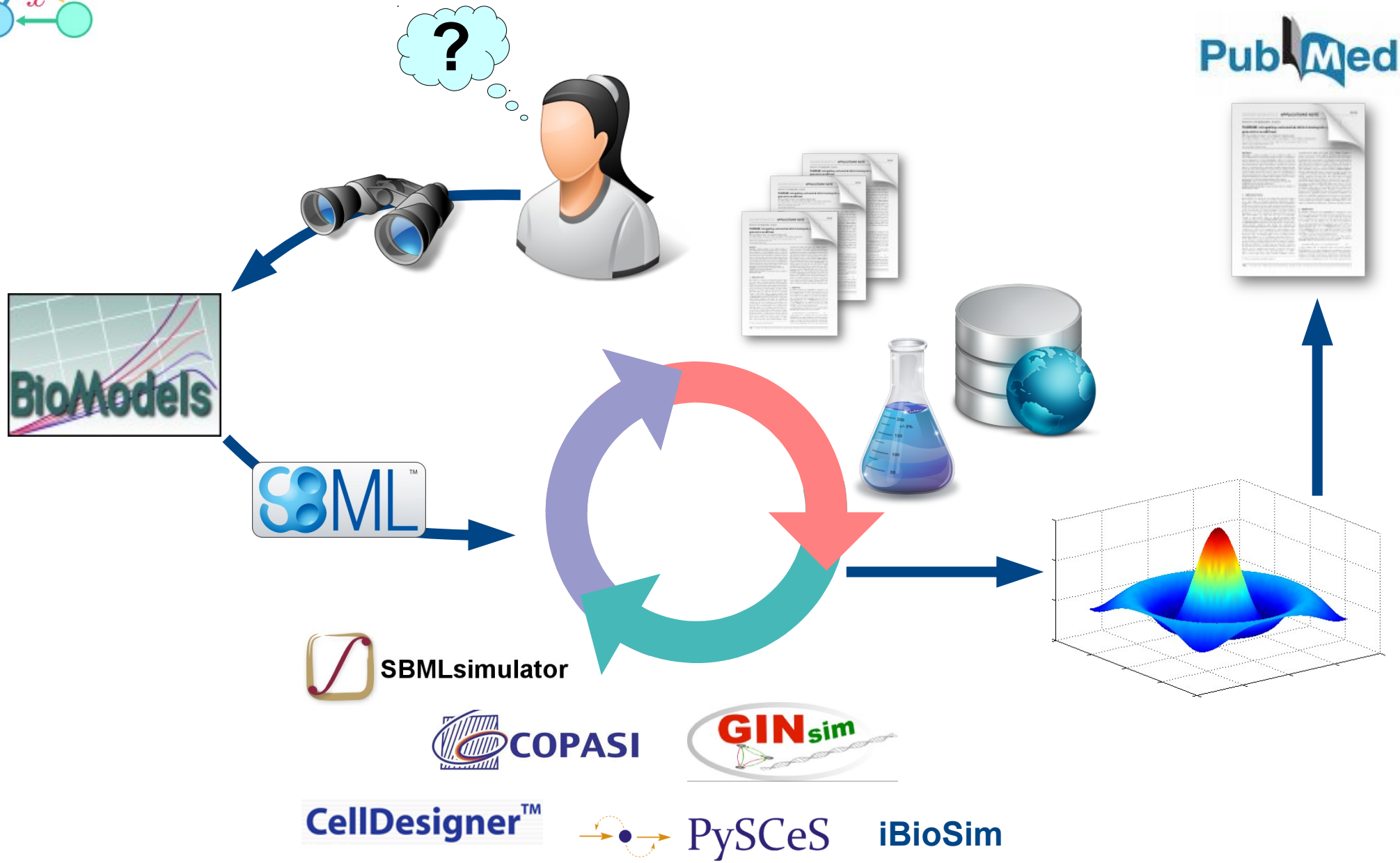
Graphical representation of 'Inositol phosphat
(PNG image hosted by the Kyoto Encyclopedi

This model has been automatically generated
formats. Wrzodek C, Dräger A, Zell A. *Bioinfor*
Some kinetic information were added from SA
The missing kinetic equations were added by

This model has been produced by the path2m



What can I do with Path2Models?



EMBL-EBI

Sarah Keating
Camille Laibe
Nicolas Rodriguez
Julio Saez-Rodriguez
Martijn van Iersel
Michael Schubert

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Andreas Dräger
Roland Keller
Matthias Rall
Clemens Wrzodek
Andreas Zell
Finja Bünel
Florian Mittag

MANCHESTER
1824

Douglas Kell
Pedro Mendes
Neil Swainston

IPK
GATERSLEBEN

Falk Schreiber
Tobias Czauderna

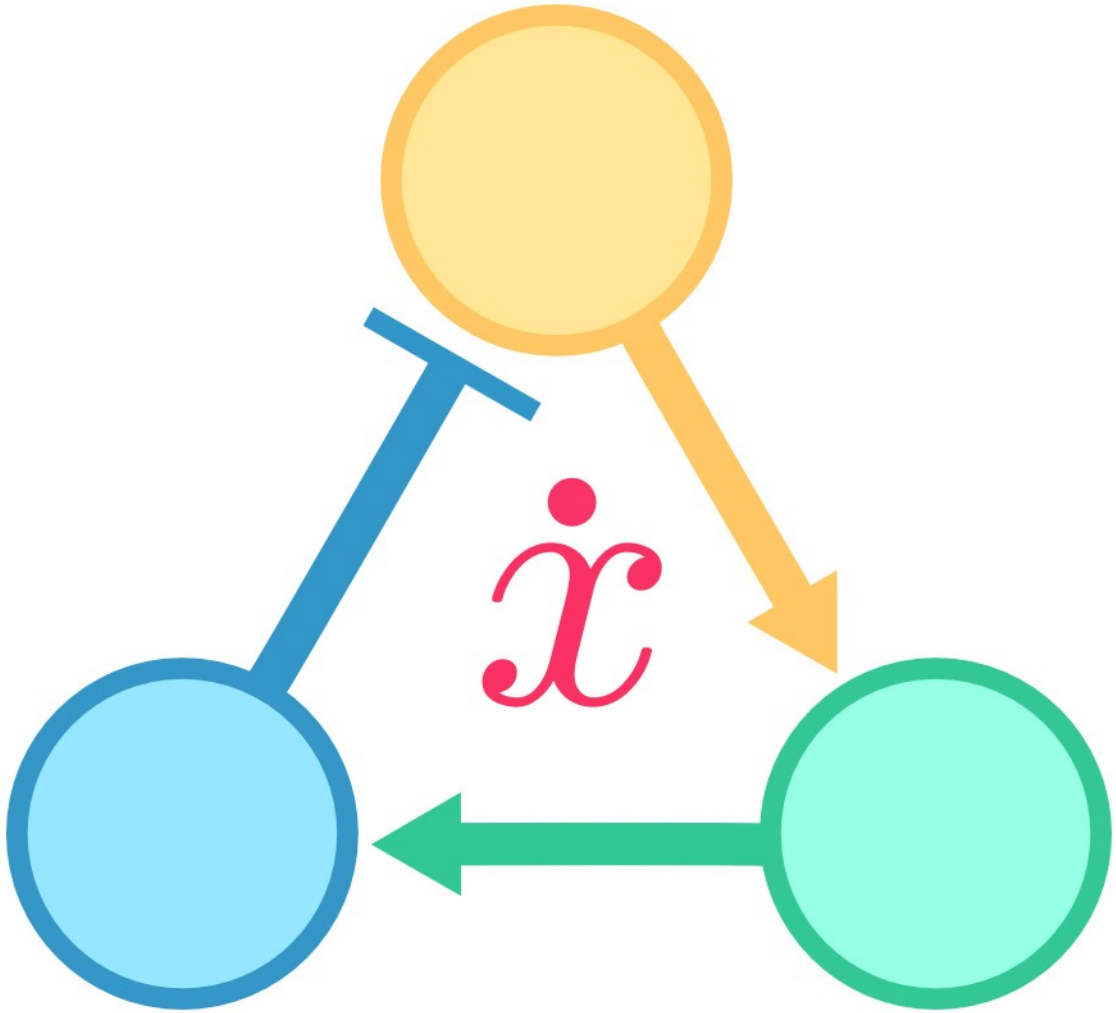
Heidelberg Institute for
Theoretical Studies

HITS

Martin Golebiewski
Wolfgang Müller

INSTITUTO GULBENKIAN DE CIÊNCIA

Claudine Chaouyia


x