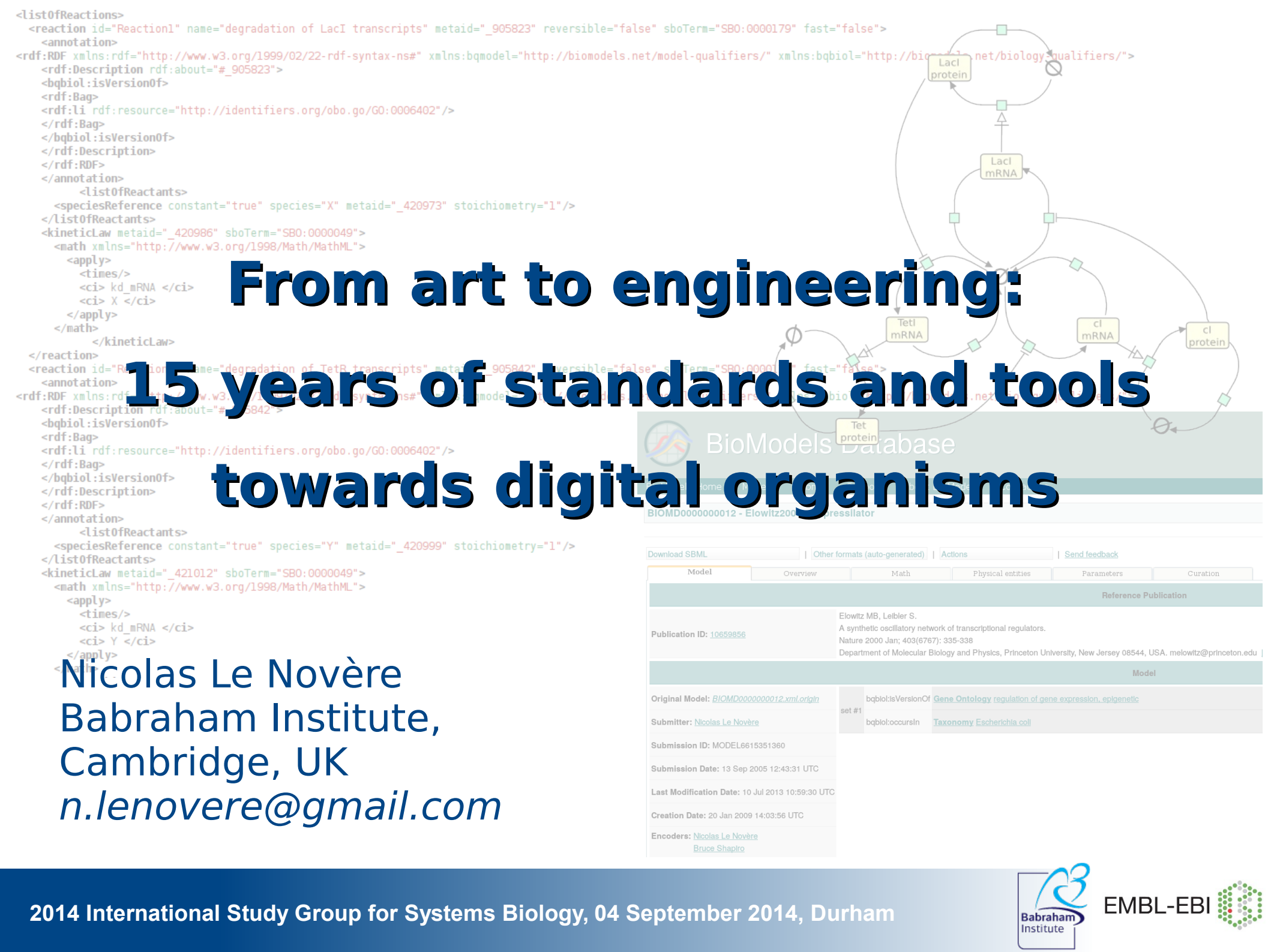


From art to engineering:

15 years of standards and tools towards digital organisms

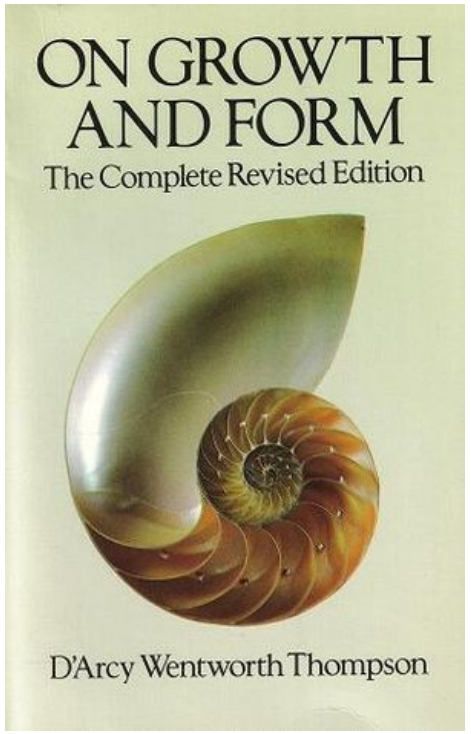
Nicolas Le Novère
Babraham Institute,
Cambridge, UK
n.lenovere@gmail.com

Babraham Institute

EMBL-EBI

Why using mathematical models?

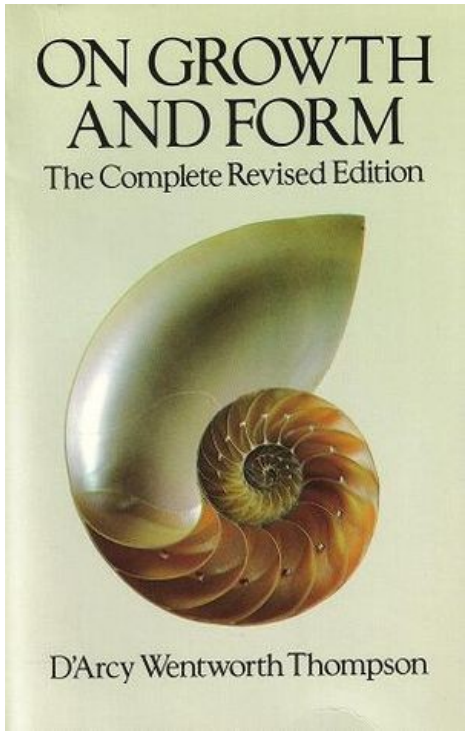
Describe



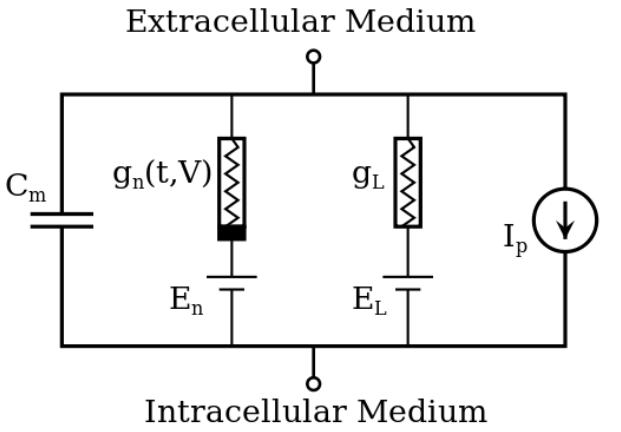
1917

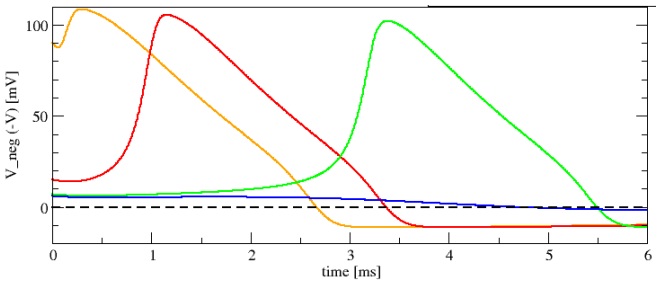
Why using mathematical models?

Describe	Explain



1917





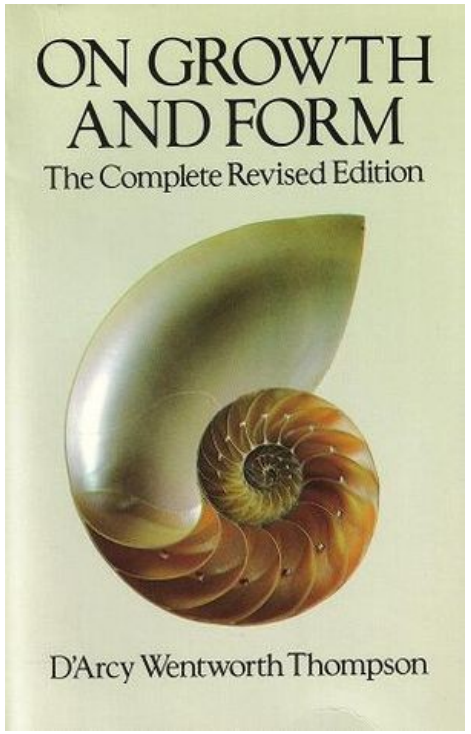
1952

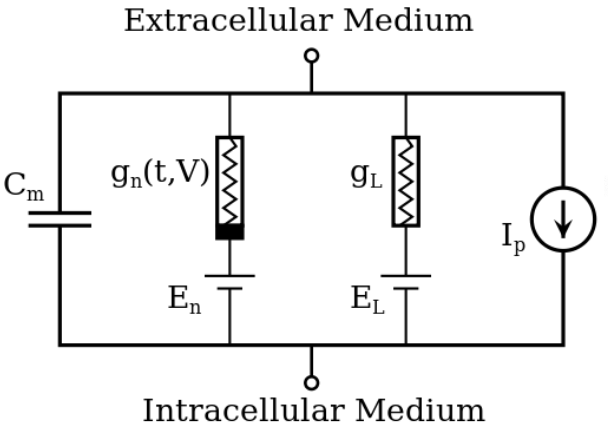
Babraham Institute

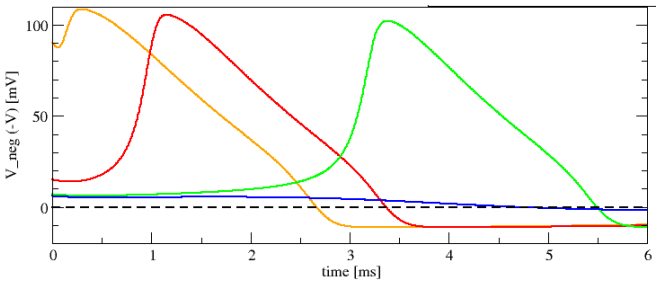
EMBL-EBI

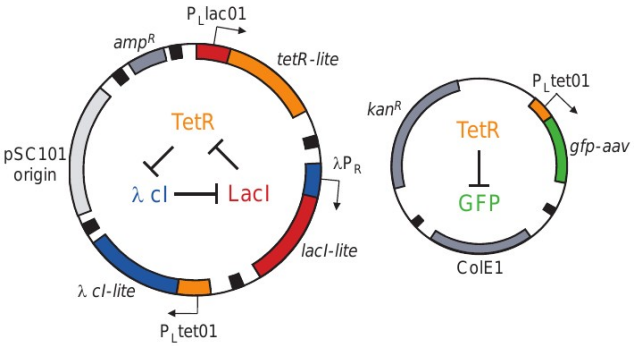
Why using mathematical models?

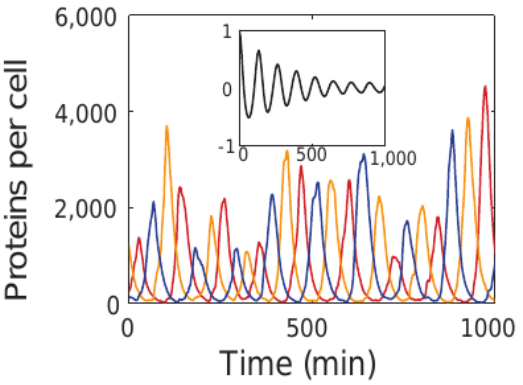
Describe	Explain	Predict











1917 **1952** **2000**

Babraham Institute

EMBL-EBI

What is a mathematical model?

Wikipedia (October 14th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

What is a mathematical model?

Wikipedia (October 14th 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 we want to know or compare with experiments

What is a mathematical model?

Wikipedia (October 14th 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 $\text{mass}_{t+\Delta t} = 0.5 \cdot \text{mass}$
$t_{1/2}$	

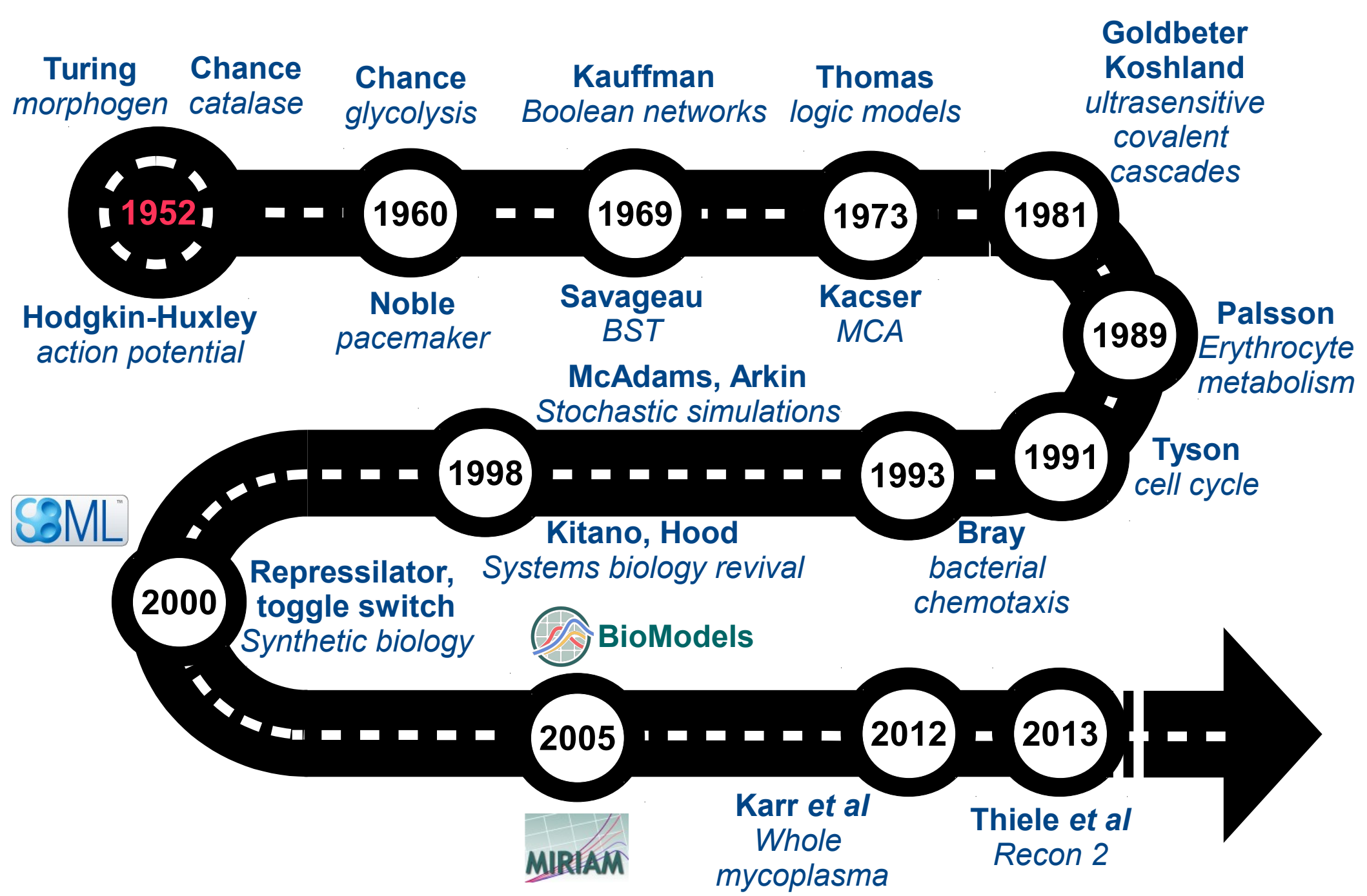
What we already know or want to test

What is a mathematical model?

Wikipedia (October 14th 2013): “A mathematical model is a description of a system using mathematical concepts and language.”

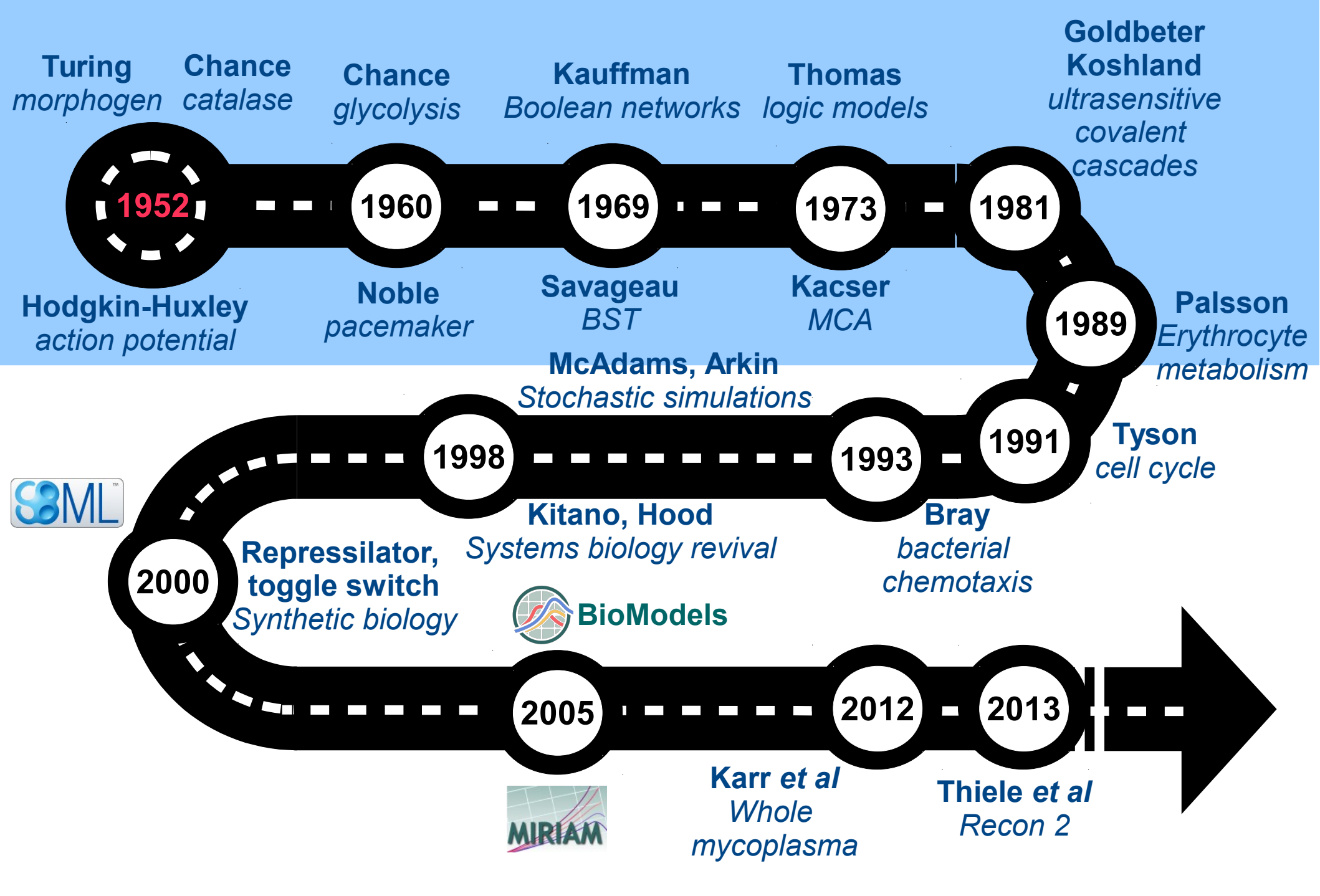
variables	relationships	constraints
[x]	$K_d = \frac{[A] \cdot [B]}{[AB]}$	[x]≥0
Vmax	$d[X]/dt = k \cdot [Y]^2$	Energy conservation
Kd	$\sum_i [X]_i - F(t) = 0$	Boundary conditions (v < upper limit)
EC_{50}	$k(t) \sim N(k, \sigma^2)$	Objective functions (maximise ATP)
length	If $\text{mass}_t > \text{threshold}$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

The context or what we want to ignore



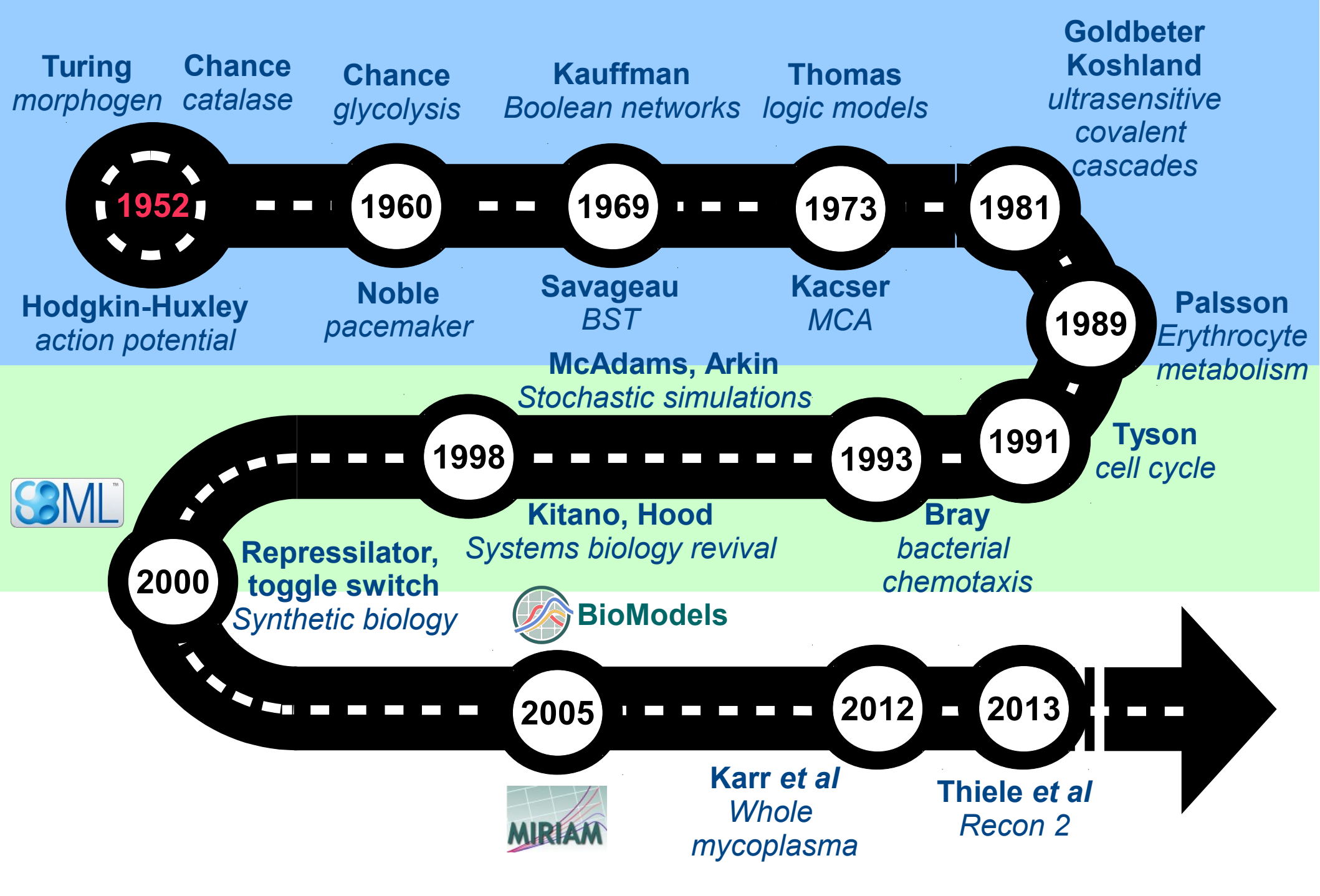
Babraham Institute

EMBL-EBI



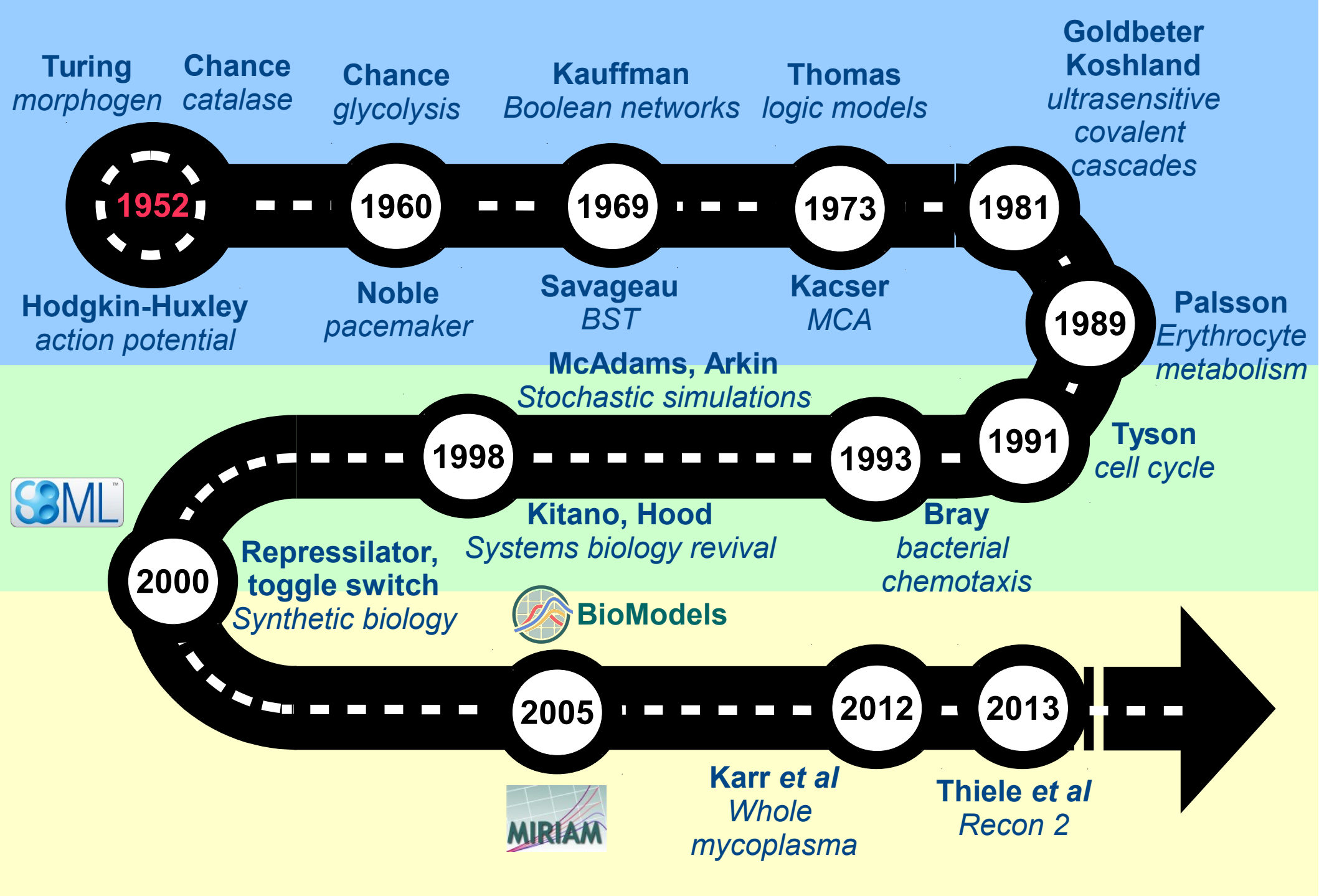
Babraham Institute

EMBL-EBI



Babraham Institute

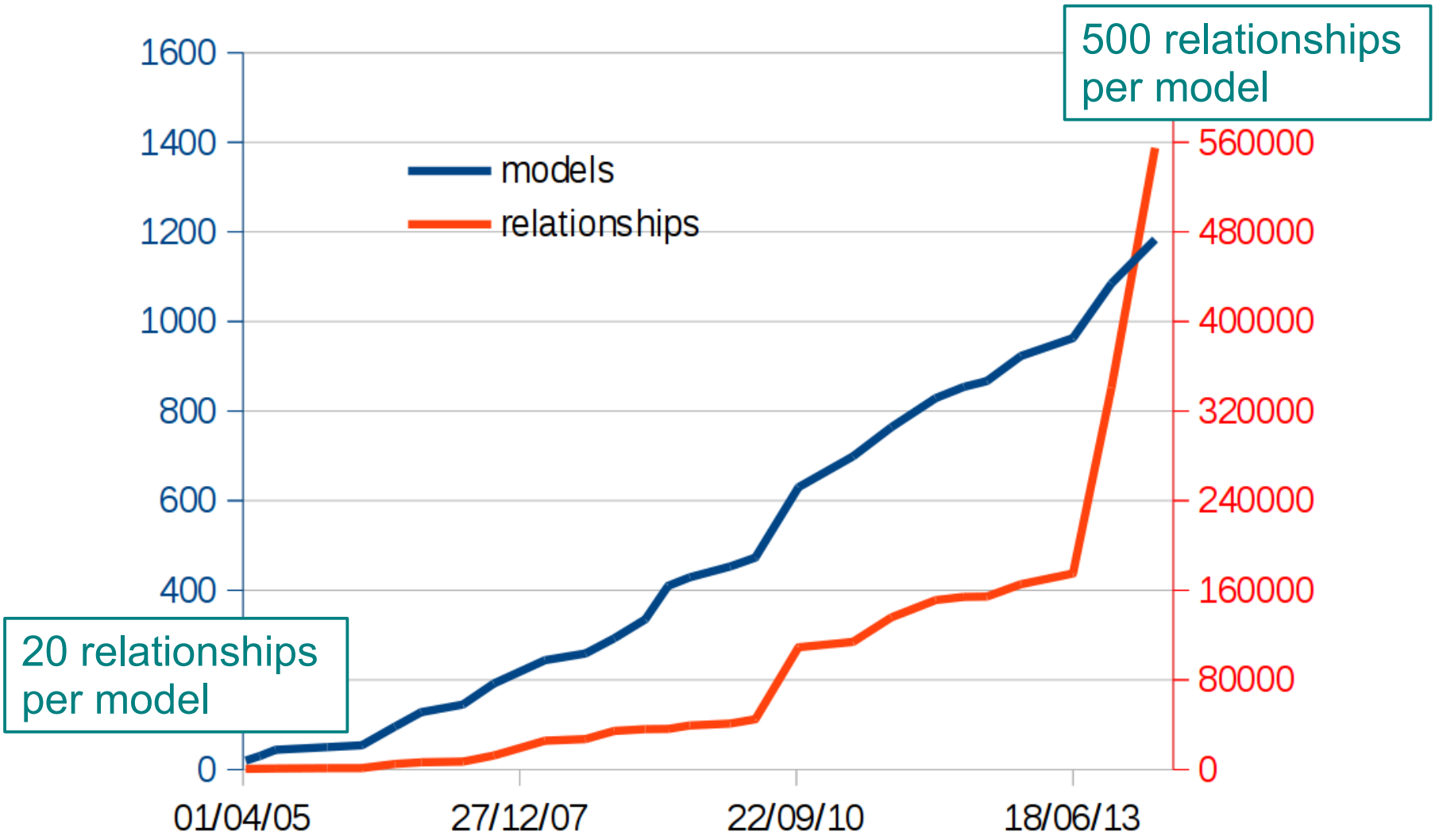
EMBL-EBI



Babraham Institute

EMBL-EBI

Computational models on the rise



BioModels Database growth (published models branch) since its creation

Improvised	Designed
One off	Many
Unique	Standard
Manually produced	Automated production
One or few artists	Collaboration
Produced in one go	Workflow
Fragile	Robust

Art.

Engineering

We need to

Verify

Re-use

Modify

Build upon

Integrate with

Therefore we need to share

Model descriptions

Simulation descriptions

Parametrisations

Biological meaning

Three types of standards

Minimal requirements **WHAT**	What to encode in order to share experiments and understand results MIBBI
Data-models **HOW**	How to encode the information defined above in a computer-readable manner combine
Terminologies	Structured representation of knowledge, with concept definitions and their relationships  OBO foundry

A language to describe computational models in biology

	Model descriptions
Data-models	SBML

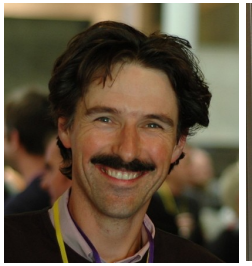
Born in Caltech 2000

John Doyle

Hiroaki Kitano

Hamid Bolouri

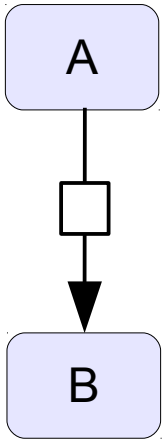
Mike Hucka



Andrew Finney

Herbert Sauro

Hucka et al. (2003)

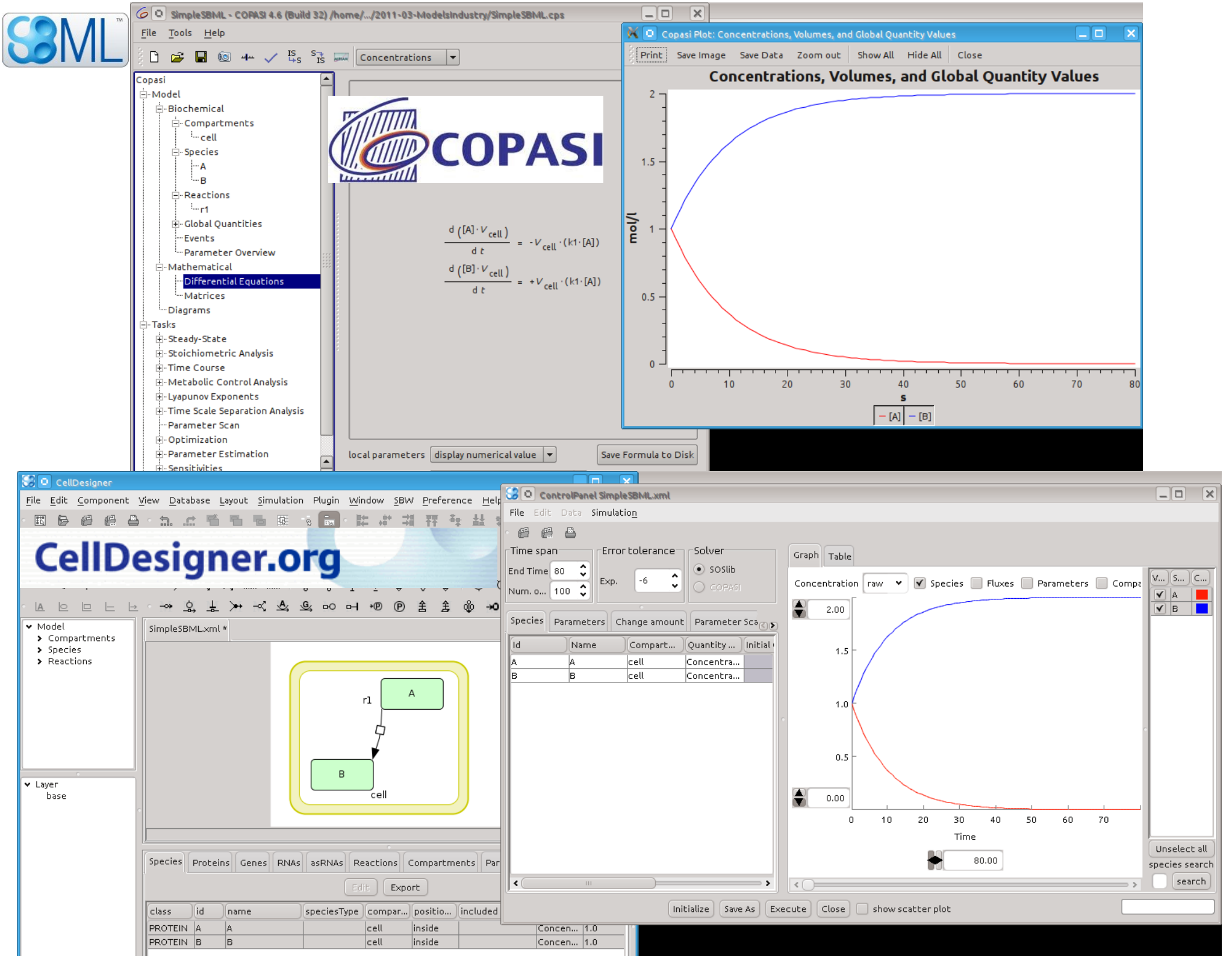


$$\frac{d[B]}{dt} = k1 \times [A]$$

```xml
<?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>
```

A very simple SBML file

http://sbml.org


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

_computational BIOLOGY

RESOURCE

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

A not so simple SBML file (Recon2)

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

MODEL1109130000

SBML Software Guide

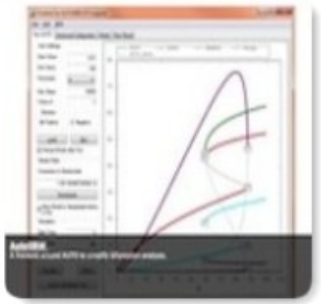
The following pages describe SBML-compatible software packages known to us. We offer different ways of viewing the information, all drawn from the same underlying data collected from the systems' developers via our **software survey**. The *Matrix* provides a table listing all known software and a variety of their features; the *Summary* provides general descriptions of most of the software; and the *Showcase* provides a sequential slideshow of a subset of the software.

Number of software packages listed in the matrix today: **263**.

Go to the SBML Software Matrix

Go to the SBML Software Summary

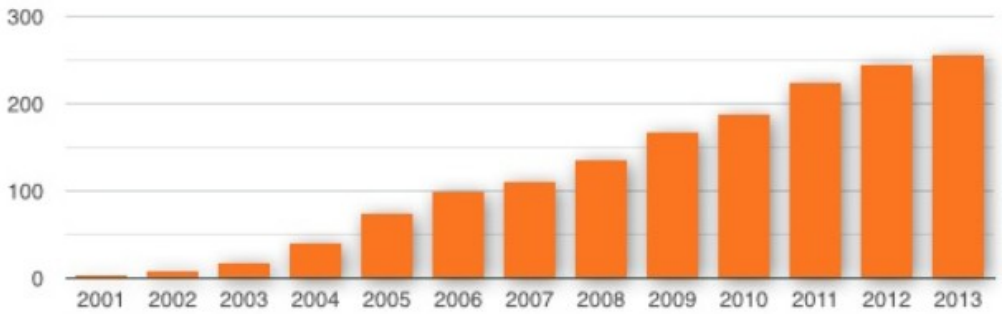
Go to the SBML Software Showcase



Please **tell us** about additions and updates.

Historical trend

The following graph shows the total number of known SBML-compatible software packages each year, as counted by the SBML Team. The counts shown are for approximately the middle of each year.

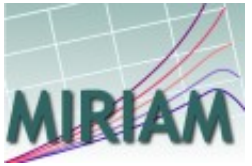


Adding the semantics to the syntax

	Model descriptions
Minimal requirements	MIRIAM
Data-models	SBML
Terminologies	SBO

Le Novère *et al.* (2005), Courtot *et al.* (2011)

Born in Heidelberg 2004



Encoded in public standard format	☑
Single reference description	☑
Reflect biological processes	☑
Simulatable and reproduce results	☑
Model creators details	☑
Time of creation and last modification	☑
Precise terms of distribution	☑
Model components identified	☑
Cross reference as triplets {collection, record, qualifier}	☑
Collection and record in one agreed-upon URI	☑

http://co.mbine.org/standards/miriam

MIRIAM

Encoded in public standard format	☑
Single reference description	☑
Reflect biological processes	☑
Simulatable and reproduce results	☑
Model creators details	☑
Time of creation and last modification	☑
Precise terms of distribution	☑
Model components identified	☑
Cross reference as triplets {collection, record, qualifier}	☑
Collection and record in one agreed-upon URI	☑

http://co.mbine.org/standards/miriam

identifiers.org (aka new MIRIAM URIs)

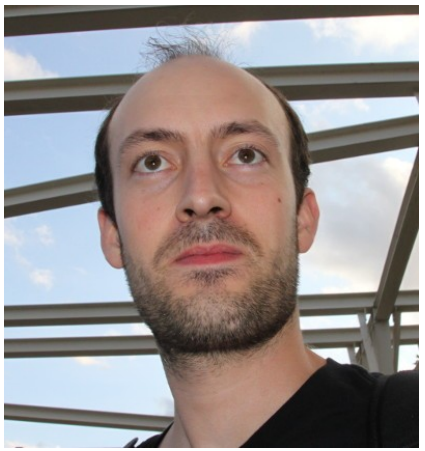
http://identifiers.org/namespace/identifier

protocol

type of URI

collection

record



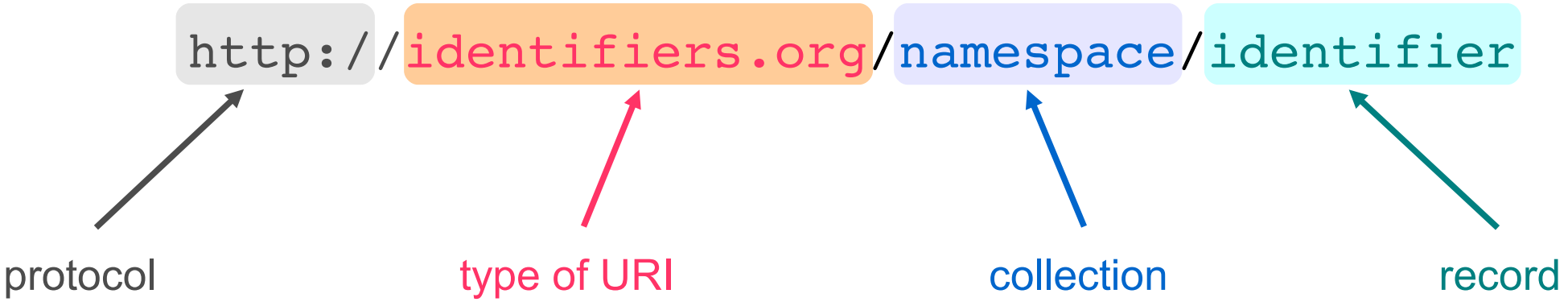
Camille Laibe

Nick Juty

Sarala Wimalaratne

Juty *et al.* (2012)

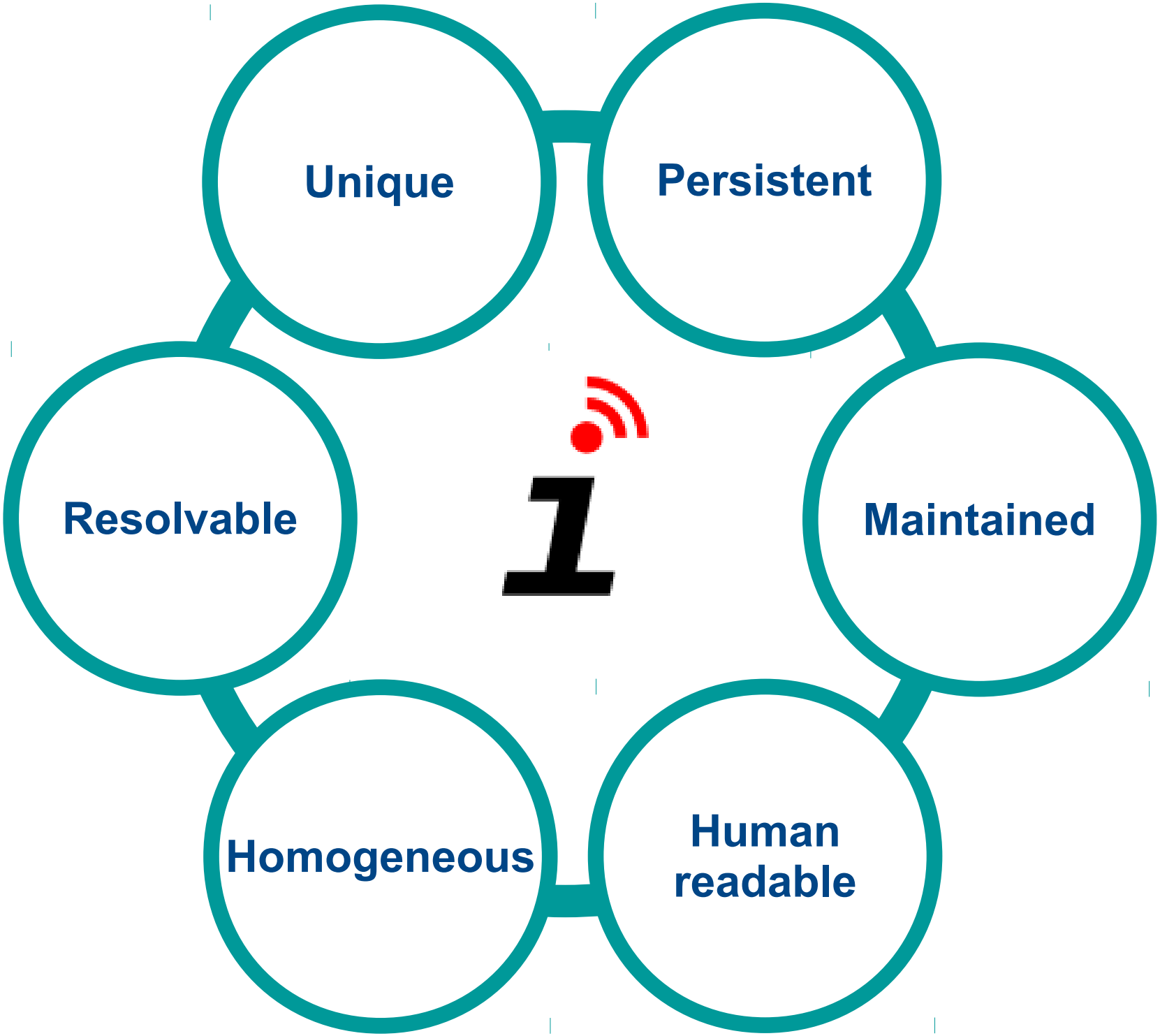
identifiers.org (aka new MIRIAM URIs)



http://identifiers.org/pubmed/22140103

http://identifiers.org/ec-code/1.1.1.1

http://identifiers.org/go/GO:0000186


Unique
Persistent
Resolvable
Maintained
Homogeneous
Human readable

MIRIAM Registry

Search

Examples: ontology, enzyme, Japan, EMBL

Categories & tags

Home | Browse | Download | Web services | Documentation | Contribute | Identifiers.org | About | Feedback

Persistent identification for life science data

The MIRIAM **Registry** provides a set of online services for the generation of unique and perennial identifiers, in the form of URIs. It provides the core data which is used by Identifiers.org.

The core of the *Registry* is a catalogue of data collections (corresponding to controlled vocabularies, databases, ...), their URIs and the corresponding physical URLs (or resources). These resources are monitored daily to ensure data accessibility and the validity of the resolution mechanism.

Access to the Registry's dataset is made available via exports (XML and RDF) and Web Services (SOAP and REST).

All provided data and services are free for use by all.

Access data

- Browse by data collection name
- Browse by types of data (categories & tags)
- Web services
- Download complete dataset (XML)
- Identifiers.org

Contribute

- Contact the team and community
- Edit existing data collection
- Request new data collection(s)
- Provide feedback

Learn & discover

- Getting started with the Registry
- Frequently Asked Questions
- Publications, presentations, posters, ...
- Review of URI based identification systems
- Documentation
- About the Registry

Latest publication

Identifiers.org and MIRIAM Registry: community resources to provide persistent identification.

Juty N., Le Novère N., Laibe C.
Nucleic Acids Research. 2012; 40 (Database issue): D58

Laibe *et al.* (2007)

Registry statistics

Published

Data collections:	521 (531)
Resources:	651 (704)
Last update:	Aug 12, 2014

Under curation

Data collections:	409
Resources:	415
Last update:	Jun 30, 2014

News

Dataset descriptor and RDF representations

August 2013

The Registry now provides a dataset descriptor and RDF representations of the whole Registry and individual data collections (in RDF/XML and Turtle formats). Read more...

Primary resources

July 2013

Identifiers.org and its Registry now highlight the "primary resource" for data collections. Read more...

Presentation at BioHackathon 2013

June 2013

Presentation "Identifiers.org: practical integration tool for heterogeneous datasets" at the BioHackathon 2013 Symposium in Tokyo, Japan (slides, PDF)

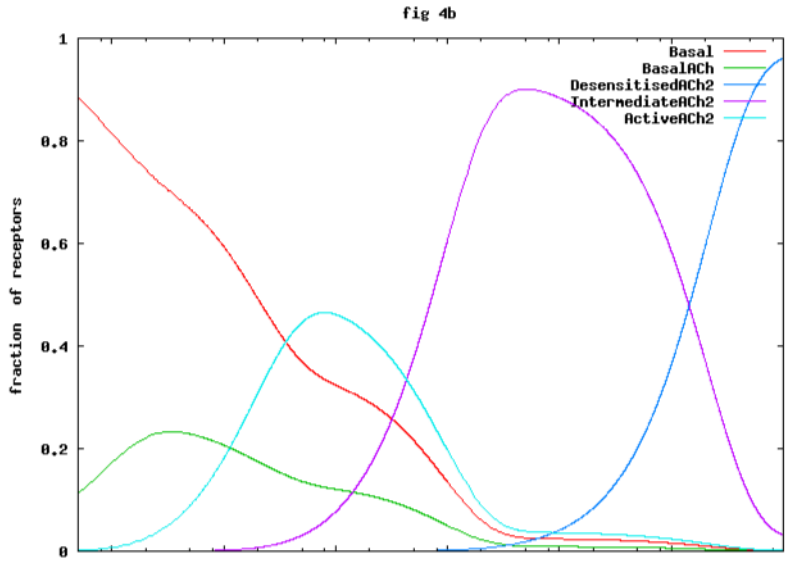
http://www.ebi.ac.uk/miriam/
http://identifiers.org/registry

SBML and Identifiers.org cross-references

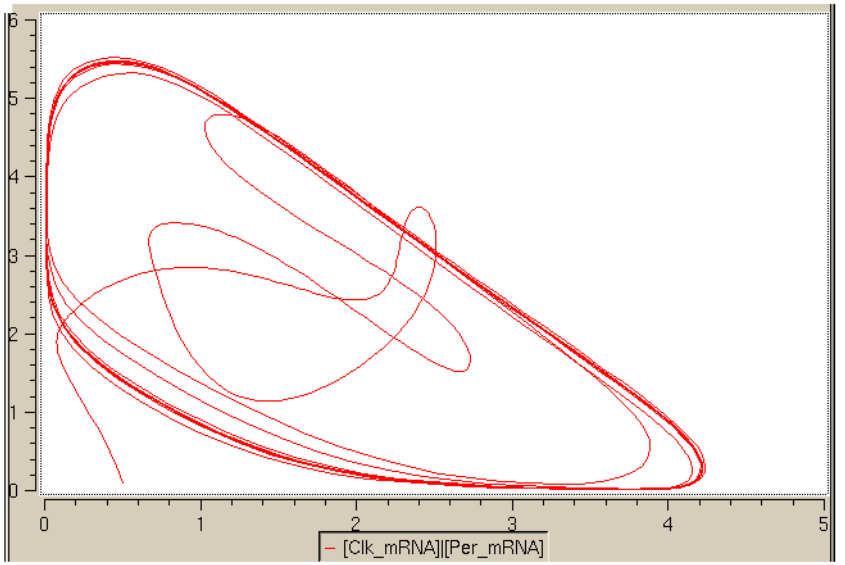
```
<species id="ca_calmodulin" metaid="cacam">
  <annotation>
    <rdf:RDF
        xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:bqbiol="http://biomodels.net/biology-qualifiers/">
      <rdf:Description rdf:about="#cacam">
        <bqbiol:hadPart>
          <rdf:Bag>
            <rdf:li rdf:resource="http://identifiers.org/uniprot/62158"/>
            <rdf:li rdf:resource="http://identifiers.org/chebi/CHEBI:29108"/>
          </rdf:Bag>
        </bqbiol:hasPart>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
</species>
```

Surely, this is enough?

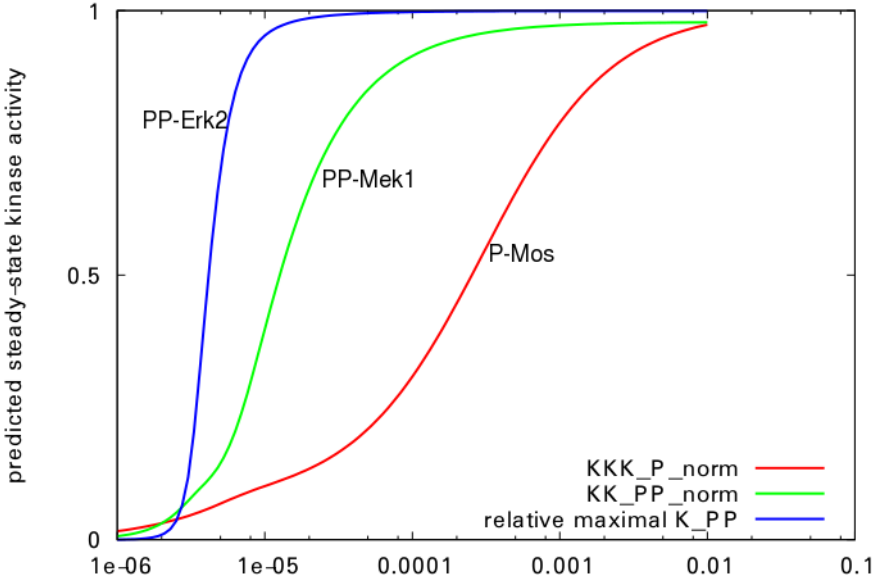
Simulation experiment = model + what to do with it



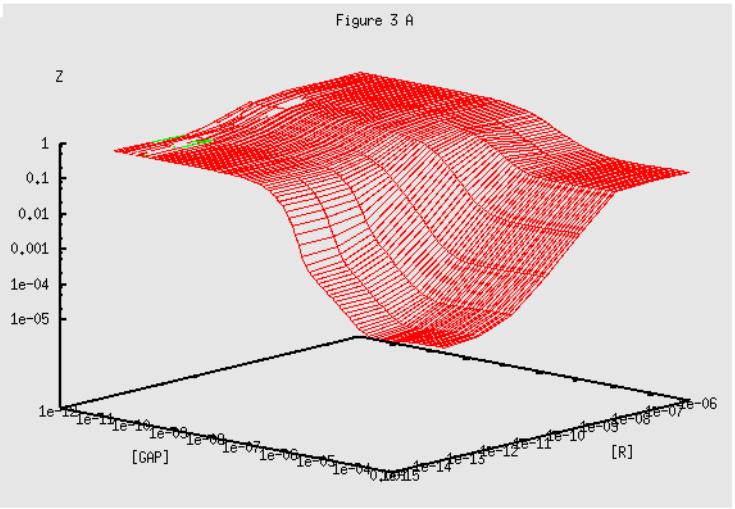
Edelstein et al 1996 (BIOMD0000000002)



Ueda, Hagiwara, Kitano 2001 (BIOMD0000000022)



Huang & Ferrell (BIOMD0000000009)

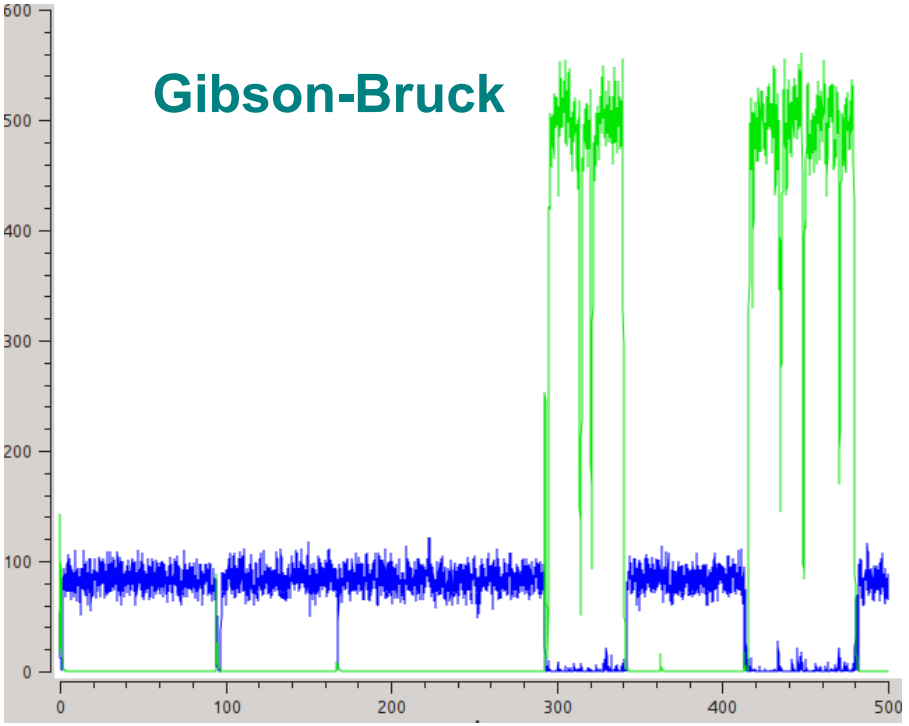


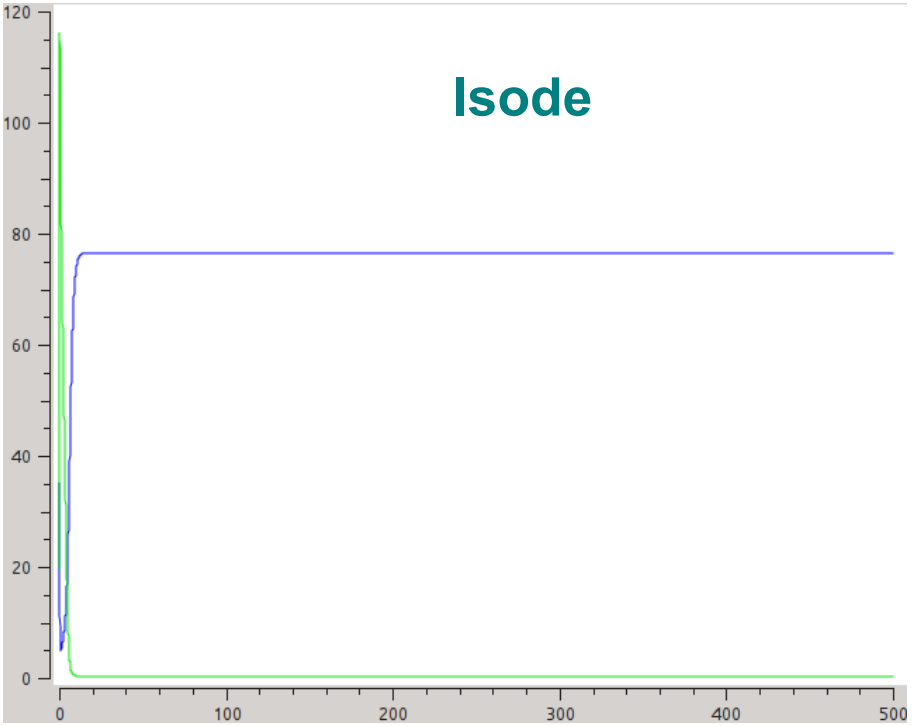
Bornheimer et al 2004 (BIOMD0000000086)

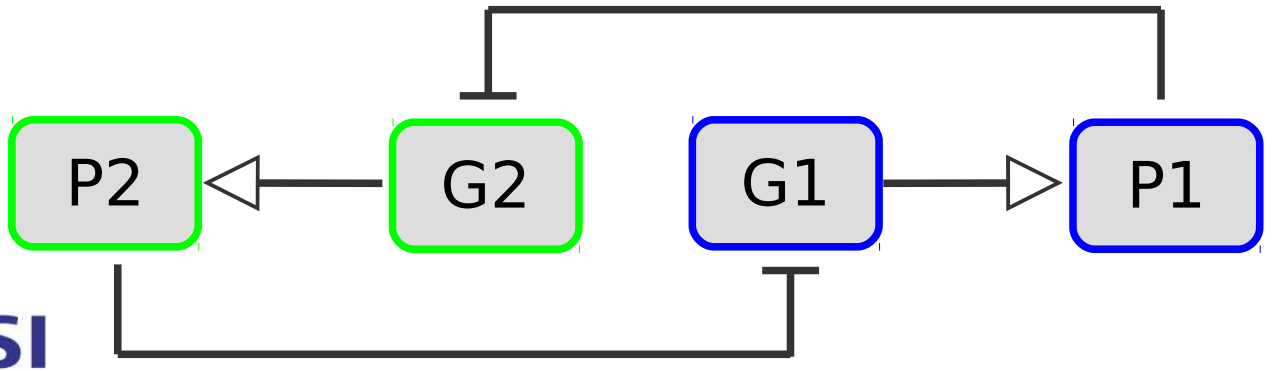
Babraham Institute

EMBL-EBI

Algorithm choice affects behaviour







COPASI

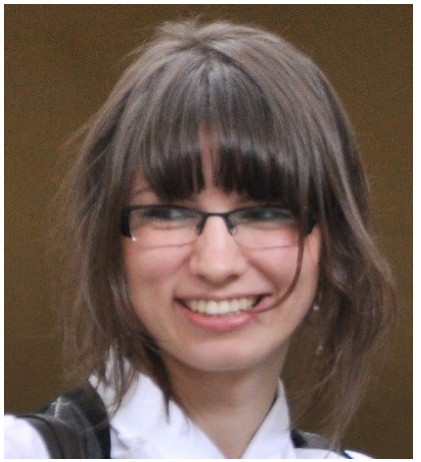
Babraham Institute

EMBL-EBI

Description of simulations and analyses

	Model descriptions	Simulations and analysis
Minimal requirements	MIRIAM	MIASE
Data-models	SBML, SBGN	SED-ML
Terminologies	SBO	KISAO

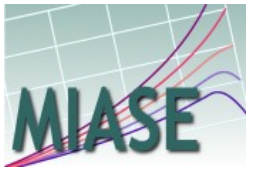
Dagmar Waltemath



Anna Zhukova

Born in Hinxton 2007

Waltemath *et al.* (2011, 2011), Courtot *et al* (2011)



- Provide models or mean of access ☑
- Equations, parameter values and necessary conditions ☑
- Standard formats, code available or full description ☑
- Modifications required before simulation ☑

- Simulation steps, algorithms, order, processing ☑
- Information for correct implementation of all steps ☑
- If not open source, all information to rewrite ☑
- If dependent on platform, how to use this platform ☑

- Post-processing steps to generate final results ☑
- How to compare results to get insights ☑

http://co.mbine.org/standards/miase

Simulation Experiment Description Markup Language

```
<?xml version="1.0" encoding="utf-8"?>
<sedML xmlns="http://sed-ml.org/"
       xmlns:math="http://www.w3.org/1998/Math/MathML"
       level="1" version="1">
  <listOfSimulations><!-- --> </listOfSimulations>
  <listOfModels>
    <model id="" source="">
      <listOfChanges><!-- --></listOfChanges>
    </model>
  </listOfModels>
  <listOfTasks><!-- --></listOfTasks>
  <listOfDataGenerators><!-- --></listOfDataGenerators>
  <listOfOutputs>
    <plot2D />
    <plot3D />
    <report />
  </listOfOutputs>
</sedML>
```

http://sed-ml.org

Flexible model use in SED-ML

Any XML

```
<listOfModels>
  <model id="model1"
         name="Regular_Spiking"
         language="http://identifiers.org/combine.specifications/sbml.level-2.version-4.release-1"
         source="http://identifiers.org/biomodels.db/BIOMD0000000127" />
  <model id="model2"
         name="chattering"
         source="model1">
    <listOfChanges>
      <changeAttribute target=
          "/sbml/model/listOfParameters/parameter[@id='c']/@value" newValue="-50">
      </changeAttribute>
      <changeAttribute target=
          "/sbml/model/listOfParameters/parameter[@id='d']/@value" newValue="42">
      </changeAttribute>
    </listOfChanges>
  </model>
</listOfModels>
```

Remote access

Modifications before simulations

That looks very useful. Where can I find those?

BioModels Database

Search

Advanced

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

BioModels Database is a repository of computational models of biological processes. Models described from literature are manually curated and enriched with cross-references. All models are provided in the Public Domain. More information about BioModels Database can be found in the FAQ.

Models published in the literature

Browse

Manually curated
(530 models)

Non curated
(655 models)

Alternative access

Gene Ontology classification (new)

Gene Ontology tree

Advanced search

Models automatically generated from pathway resources (Path2Models)

Browse

Metabolic (112,898 models)
Non-metabolic (27,531 models)
Whole genome metabolism (2,641 models)

Alternative access

Taxonomy

Ded

Contact us | Main instance at EMBL-EBI, UK | Mirror at Caltech, USA | Mode

Acknowledgements: EMBL BBSRC bioscience for the future

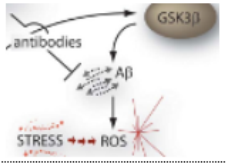
http://www.ebi.ac.uk/biomodels

Le Novère *et al.* (2006), Li *et al* (2010)

Model of the month

August, 2014

A mathematical model describing the molecular mechanisms involved in AD and the effect of immunisation against Aβ on soluble Aβ,



Marco Donizelli

el access

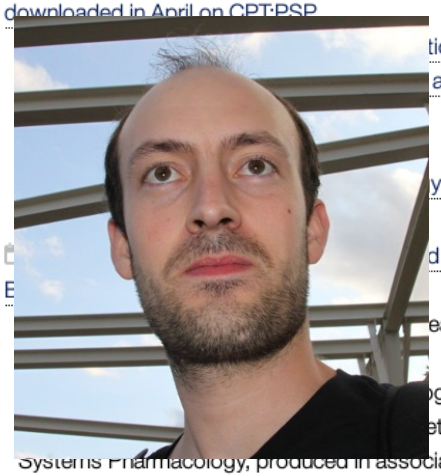
o provide a
. Also, a
r models
ave been

25 April 2014 Our recent diabetes review most downloaded in April on CPT:PSP

Viji Chelliah

tical
and

Camille Laibe



y website

d curator of

ead

gy
etrics &

Systems Pharmacology, produced in association with Nature Publishing Group (PSPod) is now online. Listen it.

>140 000 models

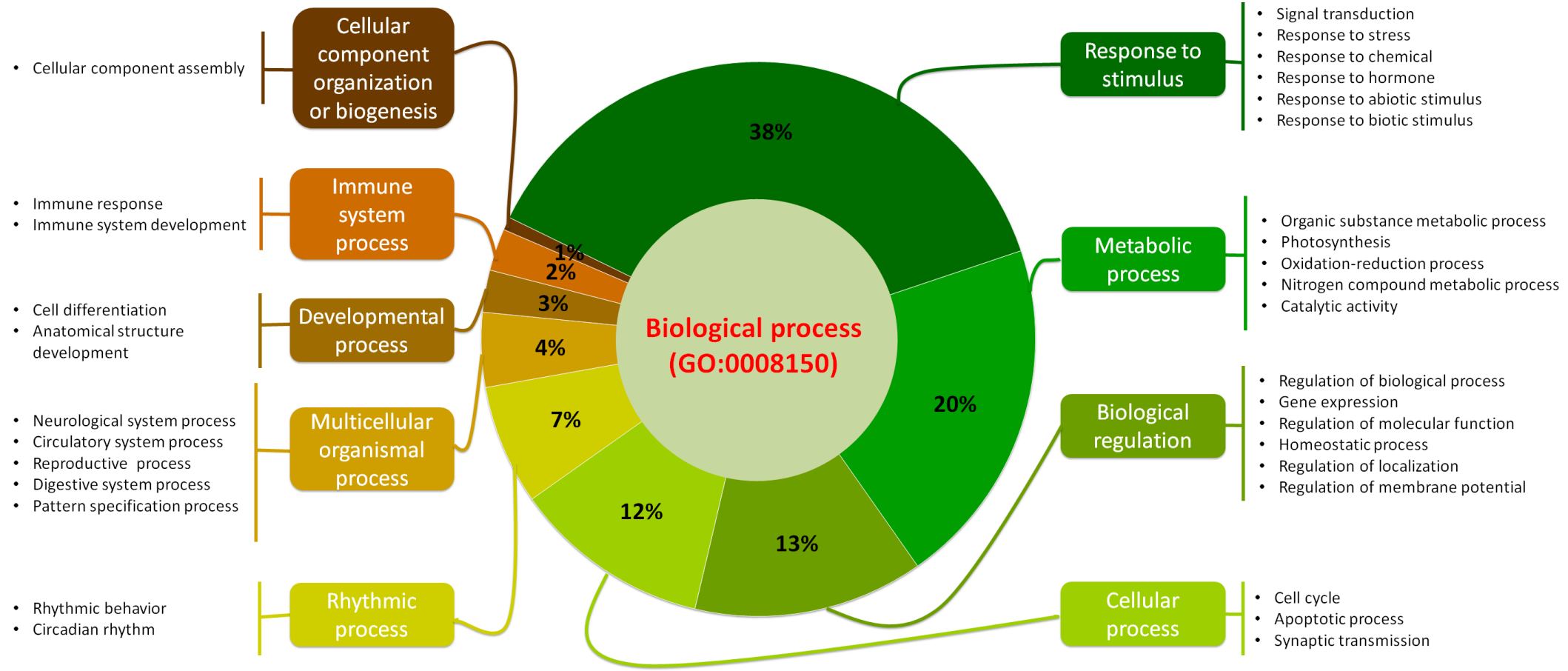
>10 millions math relations

~ 200 millions cross-refs

> 300 journals advise deposition

~ 1000 citations

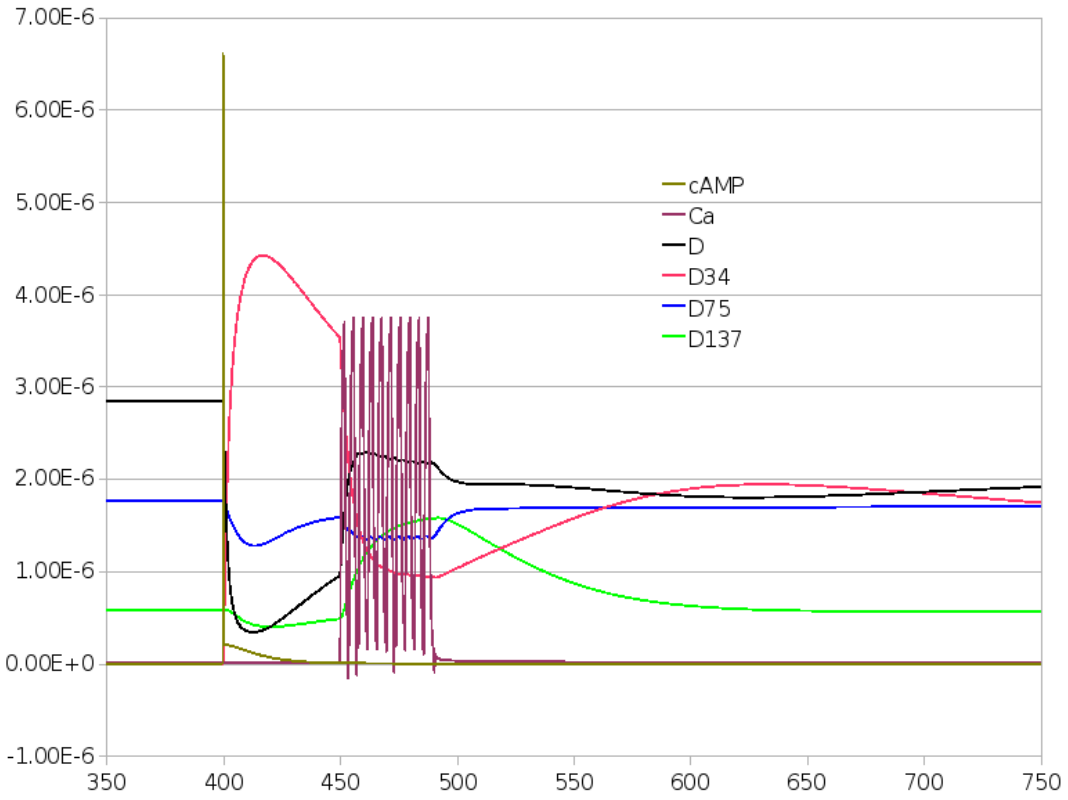
1 million page requests per year



Babraham Institute

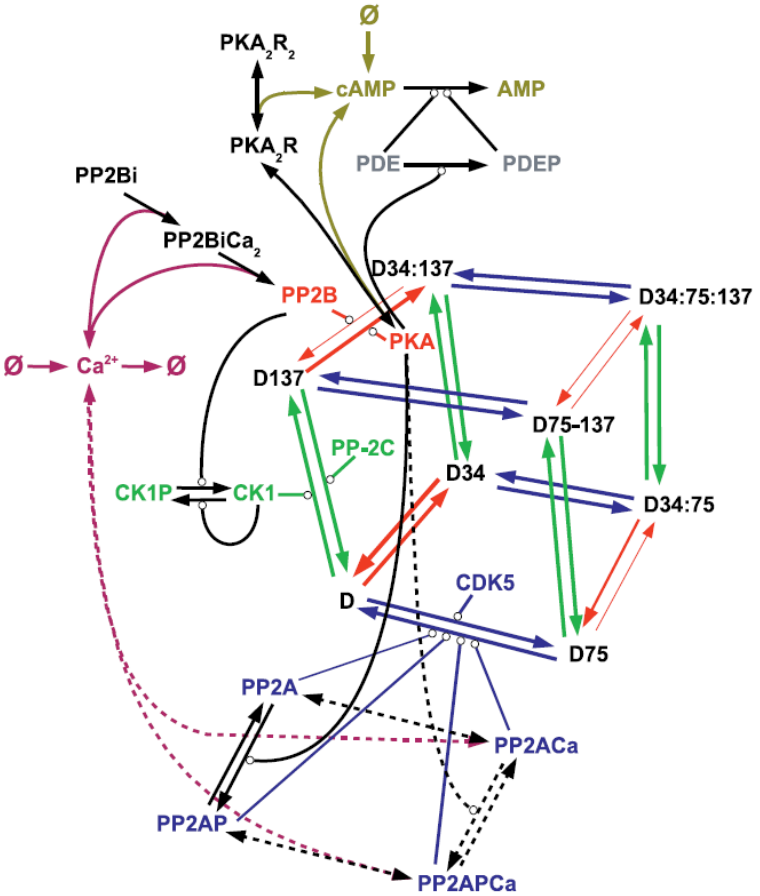
EMBL-EBI

Biochemical models



Fernandez et al. DARPP-32 is a robust integrator of dopamine and glutamate signals
PLoS Comput Biol (2006) 2: e176.

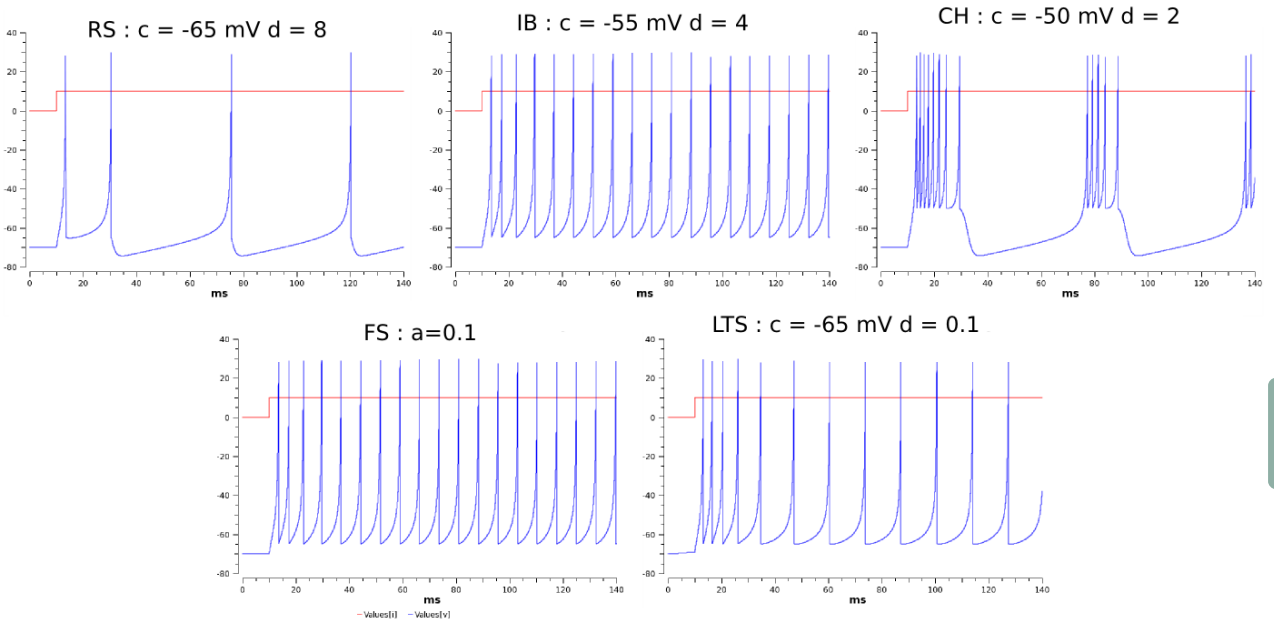
BIOMD0000000153



reaction:

$$v_{\text{on}} = k_{\text{on}} \times [\text{D}] \times [\text{CDK5}] \times \text{Vol}$$

Neuroscience models



Izhikevich EM. Simple model of spiking neurons. *IEEE Trans Neural Netw* (2003) 14(6):1569-1572.

BIOMD0000000127

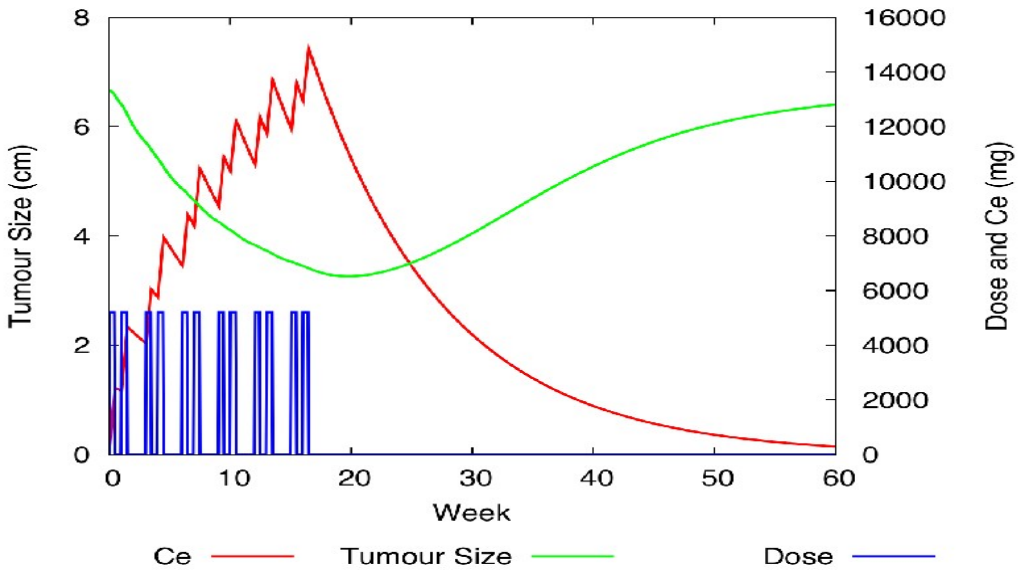
rate rule:

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

event:

$$\text{when } v > V_{\text{thresh}} \begin{cases} v = c \\ U = U + d \end{cases}$$

Pharmacometrics models



Tham et al (2008) A pharmacodynamic model for the time course of tumor shrinkage by gemcitabine + carboplatin in non-small cell lung cancer patients. *Clin Cancer Res*. 2008 14(13): 4213-8.

BIOMD0000000234

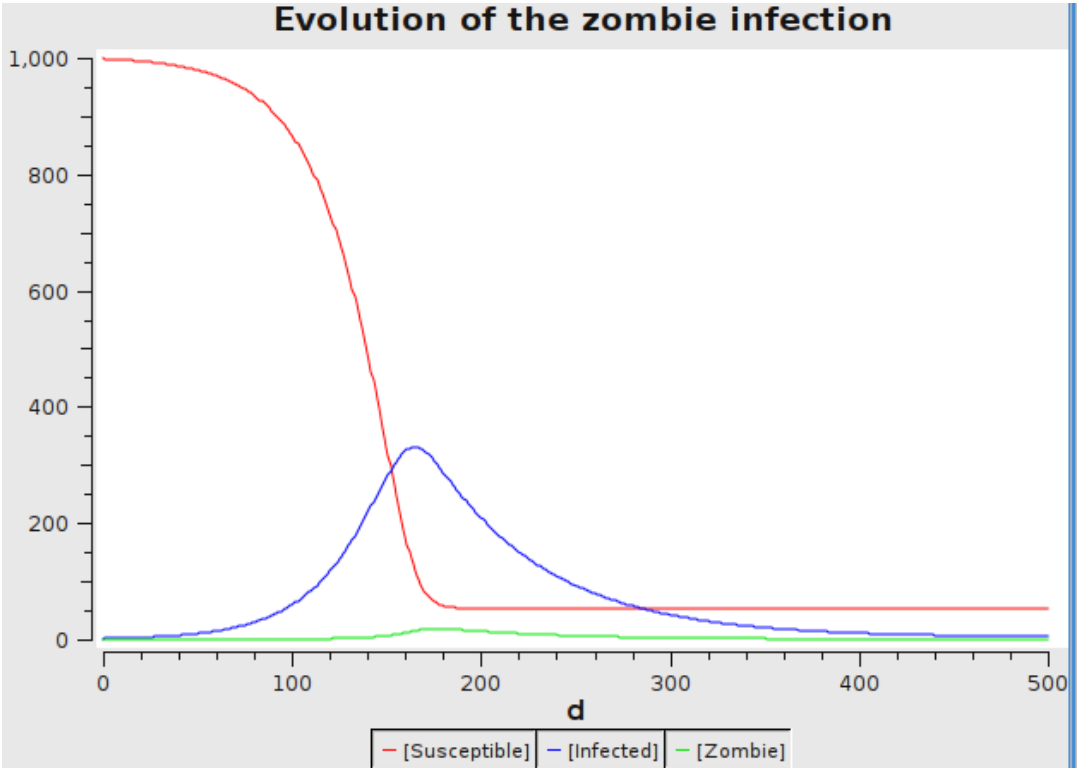
rate rule:

$$\frac{dSize}{dt} = (Rate_{in} \times Effect - K_{over} \times Size) \times Size$$

assignment rule:

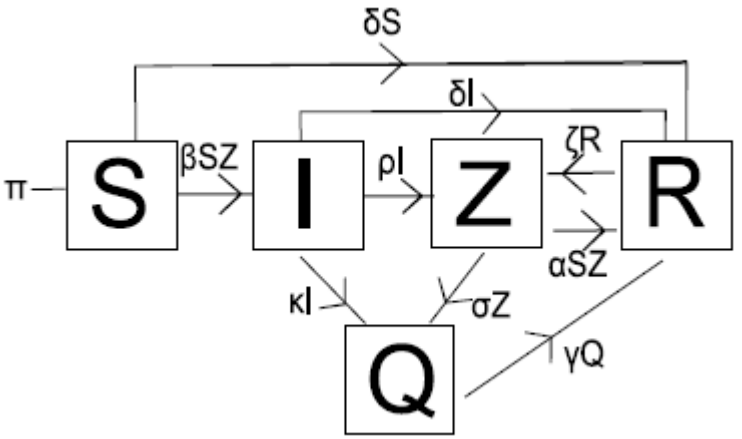
$$Effect = 1 - \frac{E_{max} \times Ce}{Amt_{50} + Ce}$$

Epidemiology models



Munz P et al. When zombies attack!: Mathematical modelling of an outbreak of zombie infection. in "Infectious Disease Modelling Research Progress", (2009)133-150

MODEL1008060001



Was it worth it?

"You should not develop standards and easy to use modelling software. This allows biologists to write models, and they don't know how to do it properly."

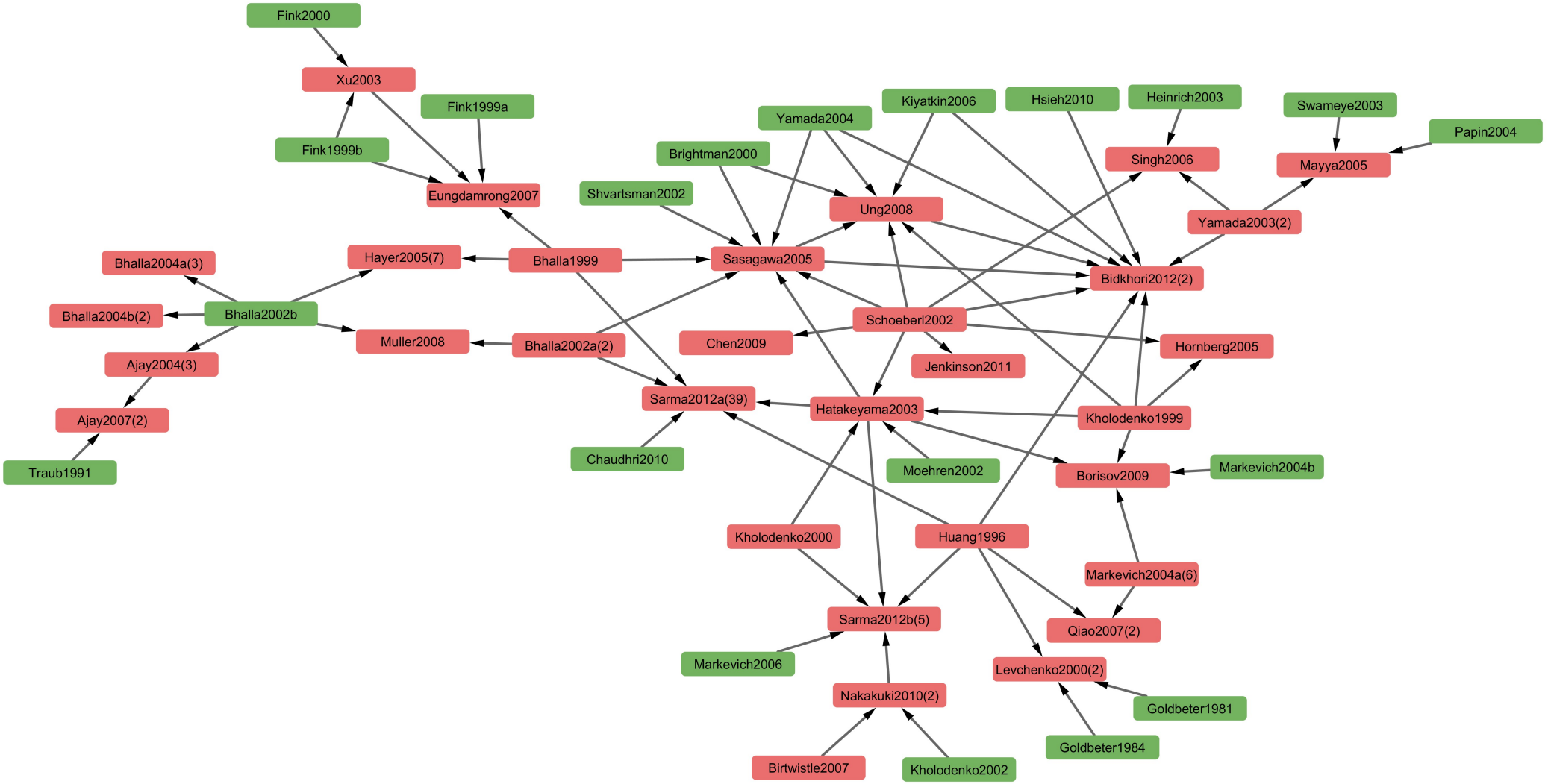
Biomathematician, 2007

"By developing BioModels you harmed the cause of modelling in biology. My students do not learn how to make a model, they download it ready to use instead"

Theoretical biologist, 2006

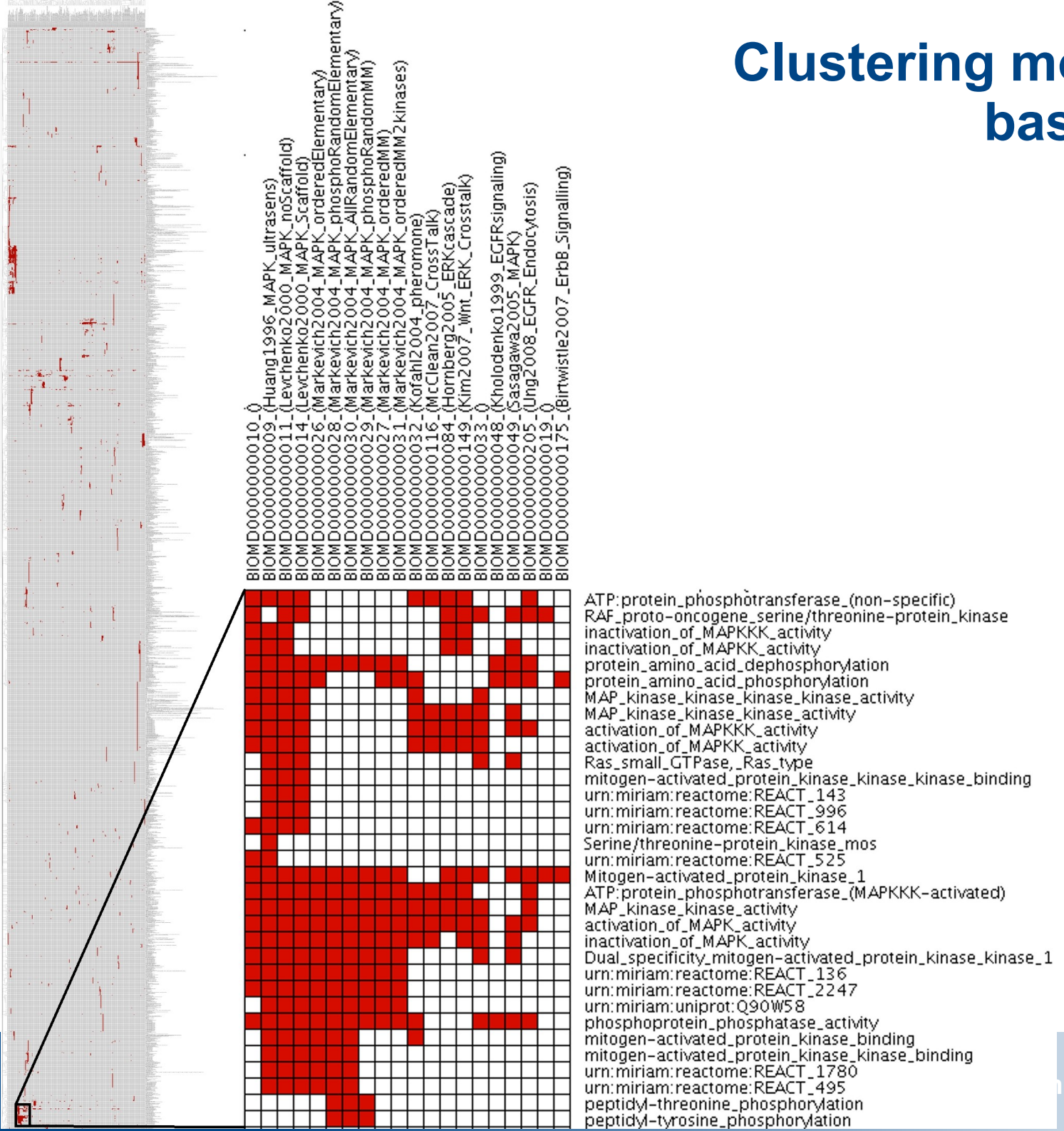
REUSE

Erb receptor signalling



DISCOVER

Clustering models (and data) based on metadata



Schulz *et al.* (2011)

Ranking and retrieval of models

Model	BioModel	Similarity	*P*-value	Overlap	*P*-value
Huang1996_MAPK_ultrasens	BIOMD0000000009	1.000	<=1e-3	30	0.0e+00
Levchenko2000_MAPK_noScaffold	BIOMD0000000011	0.930	<=1e-3	28	0.0e+00
Levchenko2000_MAPK_Scaffold	BIOMD0000000014	0.874	<=1e-3	26	0.0e+00
Kholodenko2000_MAPK_feedback	BIOMD0000000010	0.830	<=1e-3	20	0.0e+00
Markevich2004_MAPK_orderedElementary	BIOMD0000000026	0.749	<=1e-3	16	2.9e-15
Markevich2004_MAPK_phosphoRandomElementary	BIOMD0000000028	0.692	<=1e-3	15	9.1e-14
Markevich2004_MAPK_AllRandomElementary	BIOMD0000000030	0.692	<=1e-3	15	9.1e-14
Markevich2004_MAPK_orderedMM	BIOMD0000000027	0.691	<=1e-3	12	9.8e-10
Markevich2004_MAPK_orderedMM2kinases	BIOMD0000000031	0.691	<=1e-3	12	9.8e-10
Markevich2004_MAPK_phosphoRandomMM	BIOMD0000000029	0.626	<=1e-3	11	1.6e-08
Hornberg2005_ERKcascade	BIOMD0000000084	0.523	<=1e-3	9	2.7e-06
McClean2007_CrossTalk	BIOMD0000000116	0.453	<=1e-3	8	2.7e-05
Kofahl2004_pheromone	BIOMD0000000032	0.441	<=1e-3	12	9.8e-10
Goldbeter1991_MinMitOscil_ExplInact	BIOMD0000000004	0.389	<=1e-3	3	1.5e-01
Brown2004_NGF_EGF_signaling	BIOMD0000000033	0.371	<=1e-3	9	2.7e-06
Ung2008_EGFR_Endocytosis	BIOMD0000000205	0.363	<=1e-3	8	2.7e-05
Kim2007_Wnt_ERK_Crosstalk	BIOMD0000000149	0.355	<=1e-3	10	2.2e-07
Goldbeter1991_MinMitOscil	BIOMD0000000003	0.349	<=1e-3	3	1.5e-01
Sasagawa2005_MAPK	BIOMD0000000049	0.339	<=1e-3	9	2.7e-06
Swat2004_Mammalian_G1_S_Transition	BIOMD0000000228	0.317	<=1e-3	2	4.0e-01
Tyson1991_CellCycle_6var	BIOMD0000000005	0.304	<=1e-3	4	4.2e-02
Goldbeter1995_CircClock	BIOMD0000000016	0.274	<=1e-3	4	4.2e-02
Novak1997_CellCycle	BIOMD0000000007	0.259	<=1e-3	1	7.6e-01
Novak2001_FissionYeast_CellCycle	BIOMD0000000111	0.255	<=1e-3	2	4.0e-01
Leloup1999_CircClock	BIOMD0000000021	0.246	<=1e-3	4	4.2e-02
Birtwistle2007_ErbB_Signalling	BIOMD0000000175	0.236	<=1e-3	1	7.6e-01
Neves2008_Cell_Shape	BIOMD0000000182	0.222	<=1e-3	6	1.6e-03
Leloup1998_CircClock_LD	BIOMD0000000171	0.219	<=1e-3	4	4.2e-02
Veening2008_DegU_Regulation	BIOMD0000000240	0.211	4.0e-03	2	4.0e-01
Chen2004_CellCycle	BIOMD0000000056	0.209	4.0e-03	4	4.2e-02
Fernandez2006_ModelA	BIOMD0000000152	0.207	5.0e-03	2	4.0e-01
Fisher2006_NFAT_Activation	BIOMD0000000123	0.199	7.0e-03	2	4.0e-01
Hatakeyama2003_MAPK	BIOMD0000000146	0.198	7.0e-03	1	7.6e-01

MAPK

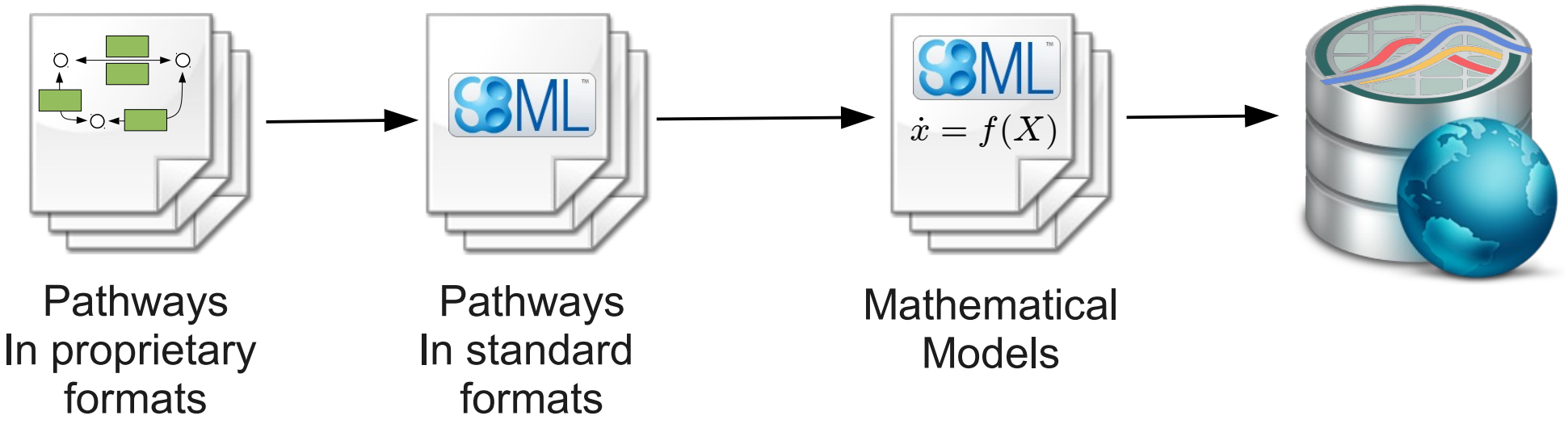
Retrieval of models using gene expression

Model	BioModel	Similarity	*P*-value	Overlap	*P*-value
Wolf2001 respiratory oscillations	BIOMD0000000090	0.207	<=1e-3	6	6.6e-09
Chassagnole2001_Threonine Synthesis	BIOMD0000000066	0.184	<=1e-3	4	1.5e-05
Curien2009_Aspartate_Metabolism	BIOMD0000000212	0.170	<=1e-3	5	3.6e-07
Curien2003_MetThr_synthesis	BIOMD0000000068	0.141	<=1e-3	2	1.0e-02
Proctor2007_ubiquitine	BIOMD0000000105	0.098	2.0e-03	1	1.4e-01
Curto1998_purineMetabol	BIOMD0000000015	0.063	1.1e-02	2	1.0e-02
Ibrahim2008_Spindle_Assembly_Checkpoint_dissociation	BIOMD0000000186	0.057	1.8e-02	0	1.0e+00
Ibrahim2008_Spindle_Assembly_Checkpoint_convey	BIOMD0000000187	0.057	1.8e-02	0	1.0e+00
Rodriguez-Caso2006_Polyamine_Metabolism	BIOMD0000000190	0.040	7.1e-02	1	1.4e-01
Nijhout2004_Folate_Cycle	BIOMD0000000213	0.032	1.1e-01	1	1.4e-01
Morrison1989_FolateCycle	BIOMD0000000018	0.030	1.3e-01	1	1.4e-01
Zatorsky2006_p53_Model3	BIOMD0000000154	0.023	2.5e-01	0	1.0e+00
Zatorsky2006_p53_Model6	BIOMD0000000155	0.023	2.5e-01	0	1.0e+00
Hunziker2010_p53_StressSpecificResponse	BIOMD0000000252	0.023	2.5e-01	0	1.0e+00
Zatorsky2006_p53_Model5	BIOMD0000000156	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model4	BIOMD0000000157	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model2	BIOMD0000000158	0.022	2.7e-01	0	1.0e+00
Zatorsky2006_p53_Model1	BIOMD0000000159	0.022	2.7e-01	0	1.0e+00
Proctor2008_p53_Mdm2_ATM	BIOMD0000000188	0.013	4.3e-01	0	1.0e+00
McClean2007_CrossTalk	BIOMD0000000116	0.012	4.7e-01	0	1.0e+00
Proctor2008_p53_Mdm2_ARF	BIOMD0000000189	0.012	4.9e-01	0	1.0e+00
Haberichter2007_cellcycle	BIOMD0000000109	0.011	5.0e-01	0	1.0e+00
Sasagawa2005_MAPK	BIOMD0000000049	0.006	5.5e-01	0	1.0e+00

BUILD

From pathways to models … path2models

- Provide pathways in a standard format
- Re-use existing pathway data to generate biochemically based models
- Provide starting points to build more quantitative models



Büchel *et al.* (2013)

Signalling Pathways

+

Logical models
of individual signalling pathways

Signalling Pathways

+

Logical models
of individual signalling pathways

Metabolic Networks

Chemical kinetics models
of individual metabolic pathways

Signalling Pathways

+

Logical models
of individual signalling pathways

Metabolic Networks

Chemical kinetics models
of individual metabolic pathways

Metabolic Networks

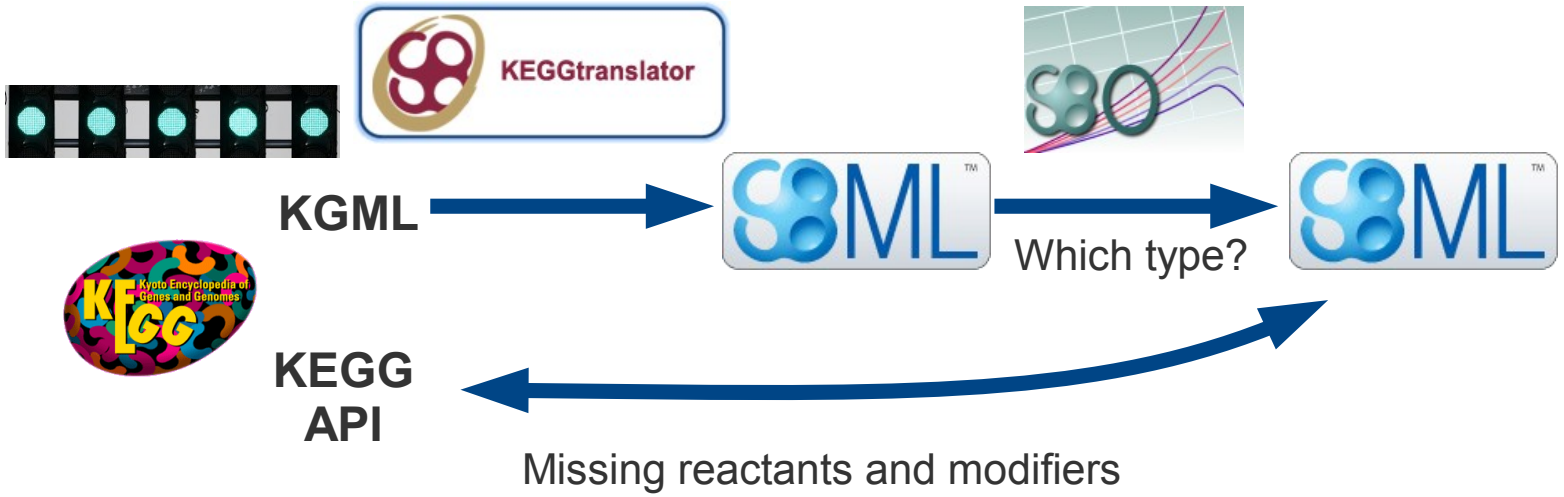
+

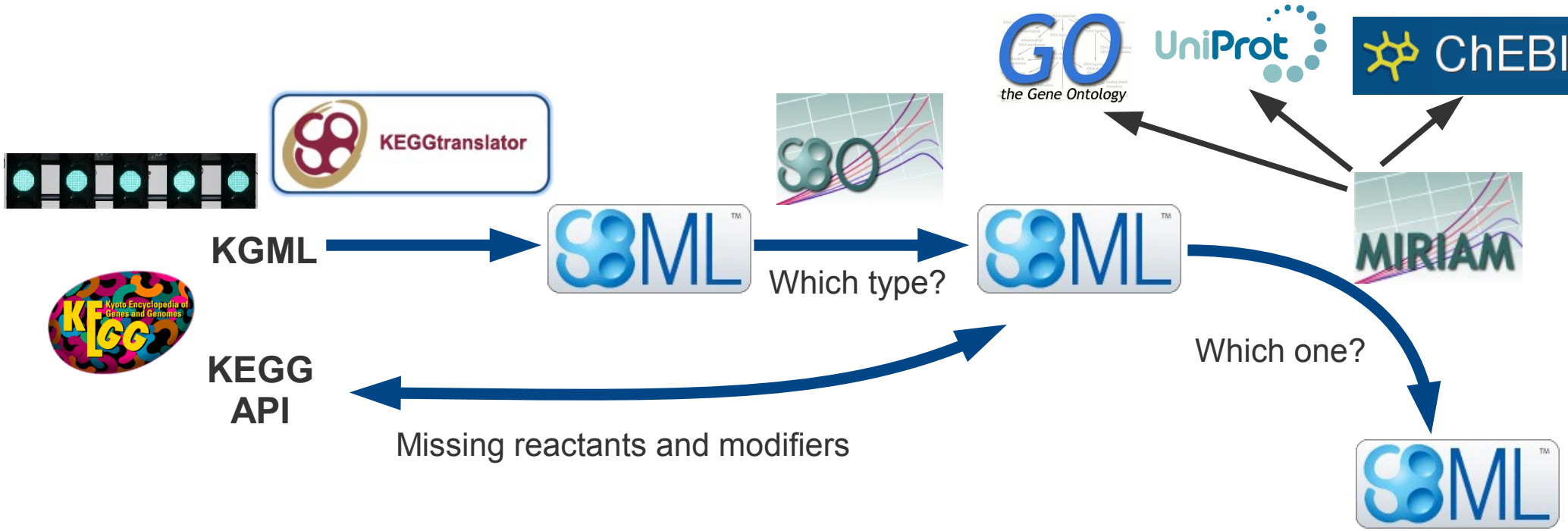
Flux Balance Analysis
of whole genome reconstructions

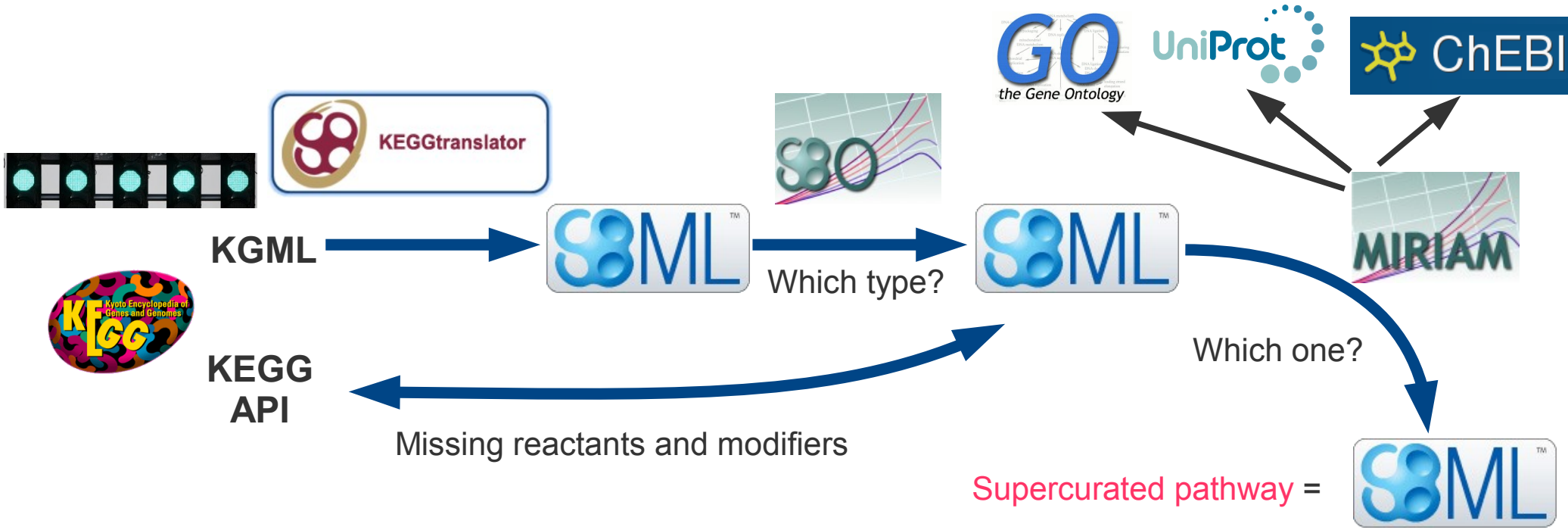
KEGGtranslator

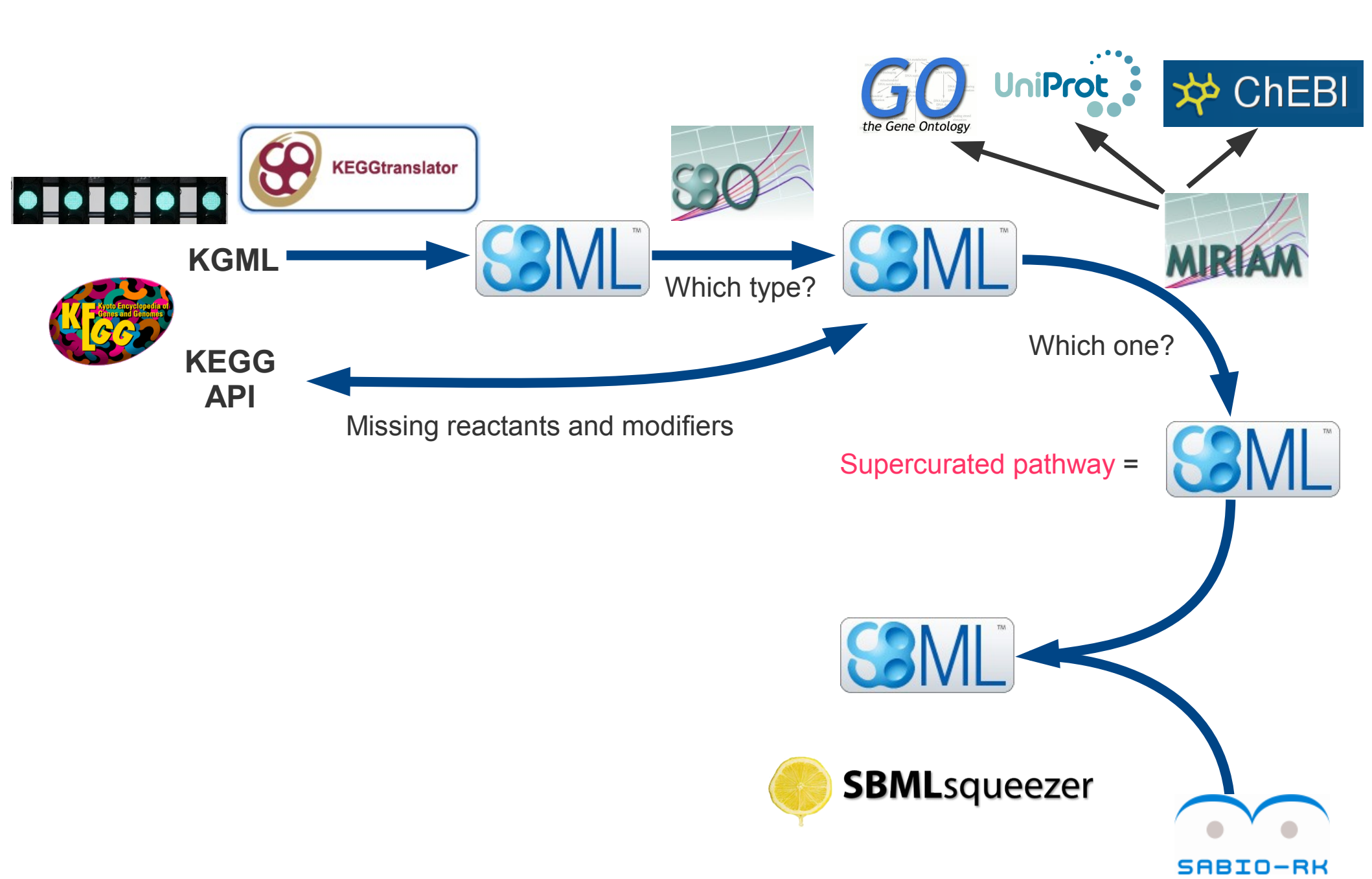
KGML → SBML

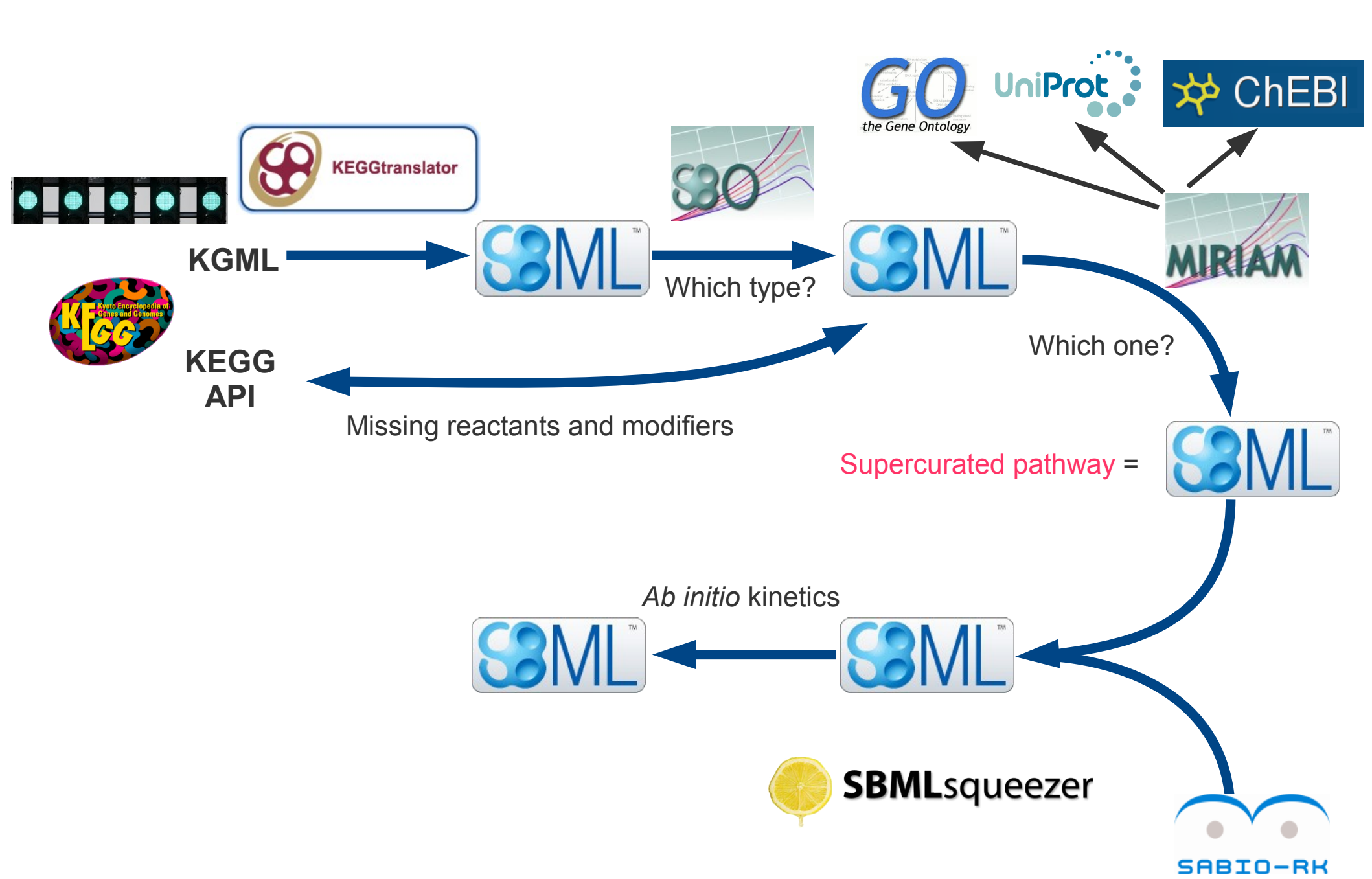

KEGGtranslator
KGML
Which type?

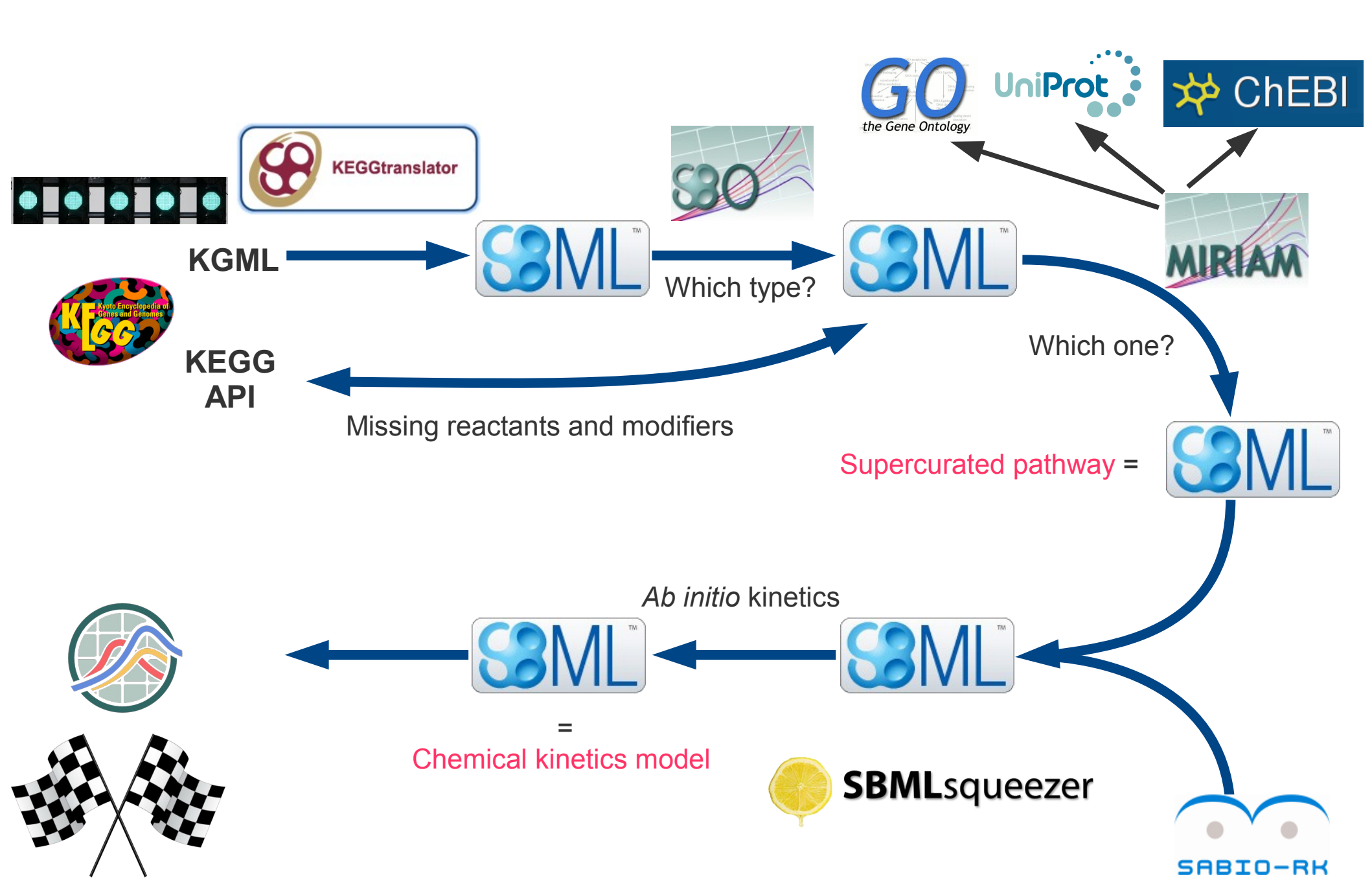

KEGGtranslator
KGML
SBML
Which type?
SBML
KEGG
API
Missing reactants and modifiers


GO
the Gene Ontology
UniProt
ChEBI
KEGGtranslator
SBO
KGML
SBML
Which type?
SBML
MIRIAM
KEGG
Kyoto Encyclopedia of Genes and Genomes
KEGG API
Missing reactants and modifiers
Which one?
SBML

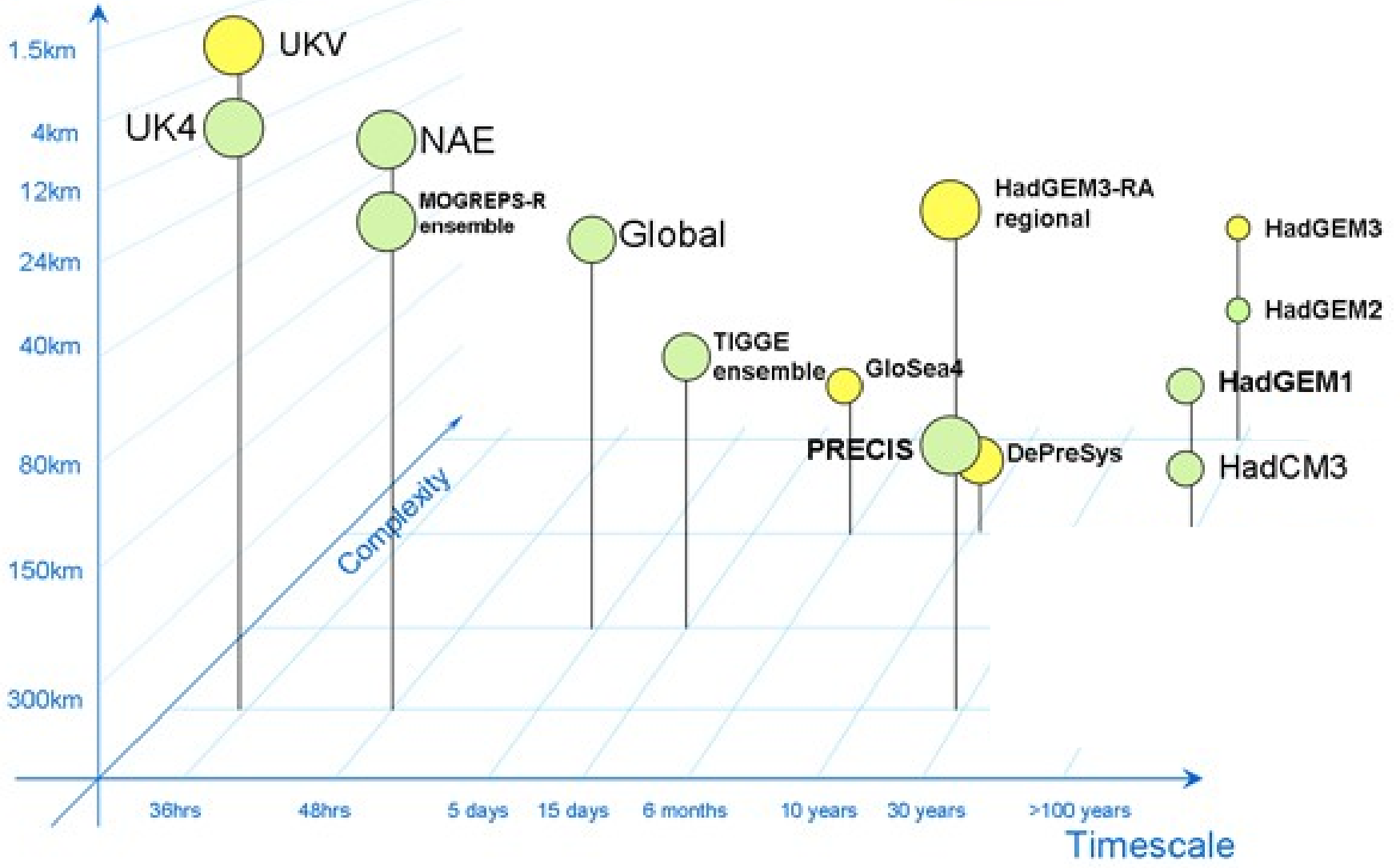

GO
the Gene Ontology
UniProt
ChEBI
KEGGtranslator
SBO
KGML
SBML
Which type?
SBML
MIRIAM
KEGG
Kyoto Encyclopedia of Genes and Genomes
KEGG API
Which one?
Missing reactants and modifiers
Supercurated pathway =
SBML


GO
the Gene Ontology
UniProt
ChEBI
KEGGtranslator
SBO
MIRIAM
KGML
SBML
Which type?
SBML
Which one?
KEGG
API
Missing reactants and modifiers
Supercurated pathway =
SBML
SBML
SBMLsqueezer
SABIO-RK

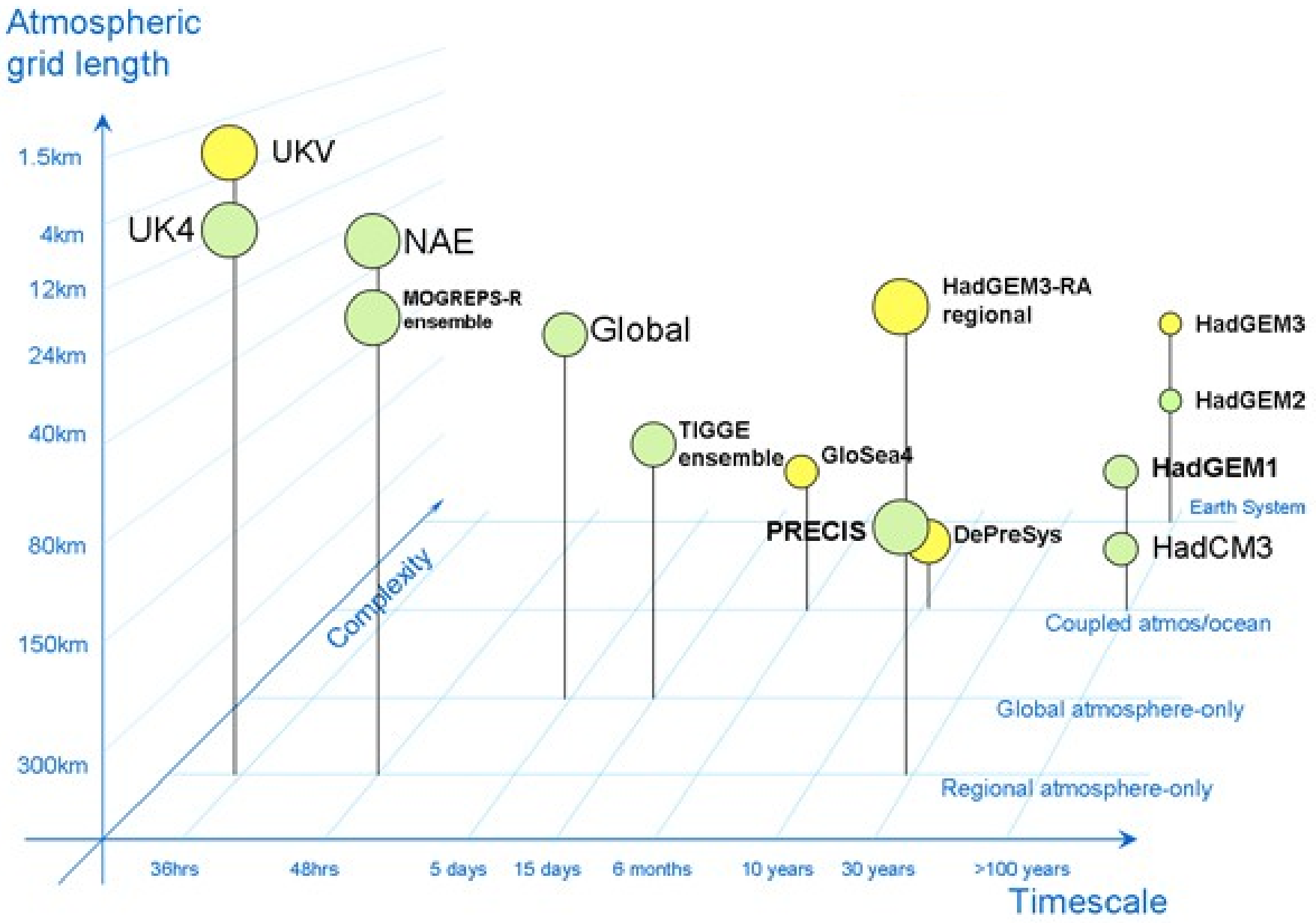

GO
the Gene Ontology
UniProt
ChEBI
KEGGtranslator
SBO
MIRIAM
KGML
SBML
Which type?
SBML
Which one?
KEGG
API
Missing reactants and modifiers
Supercurated pathway =
SBML
Ab initio kinetics
SBML
SBML
SBMLsqueezer
SABIO-RK


GO
the Gene Ontology
UniProt
ChEBI
KEGGtranslator
SBO
KGML
SBML
Which type?
SBML
MIRIAM
KEGG
API
Which one?
Missing reactants and modifiers
Supercurated pathway =
SBML
Ab initio kinetics
SBML
SBML
=
Chemical kinetics model
SBMLsqueezer
SABIO-RK

Are-we ready to develop comprehensive models of whole cells, organs and organisms?


1.5km
4km
12km
24km
40km
80km
150km
300km
UKV
UK4
NAE
MOGREPS-R ensemble
Global
TIGGE ensemble
GloSea4
PRECIS
HadGEM3-RA regional
DePreSys
HadGEM3
HadGEM2
HadGEM1
HadCM3
Complexity
36hrs
48hrs
5 days
15 days
6 months
10 years
30 years
>100 years
Timescale

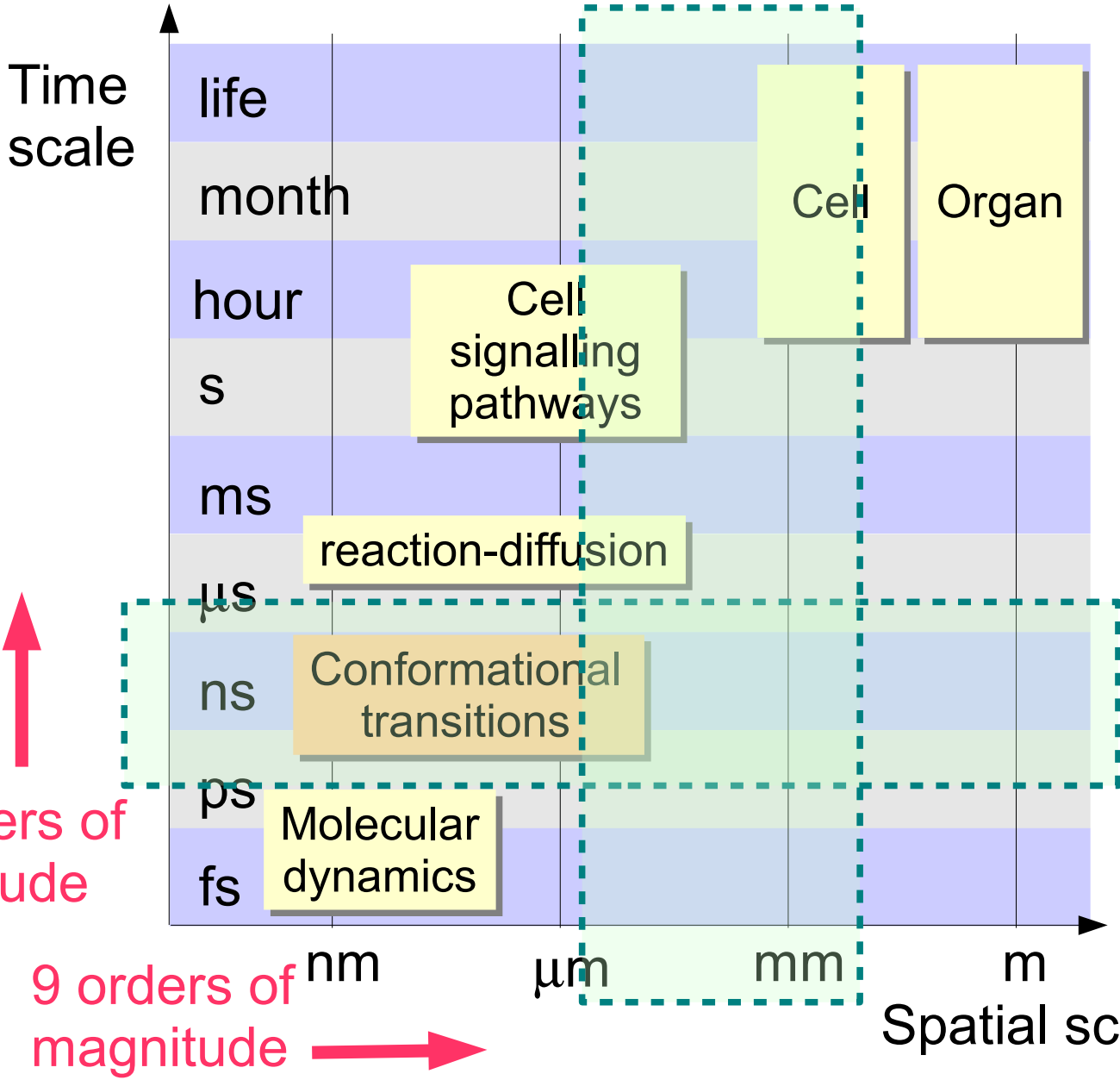
Met Office Seamless Unified Model



Babraham Institute

EMBL-EBI

Challenge 1: scales



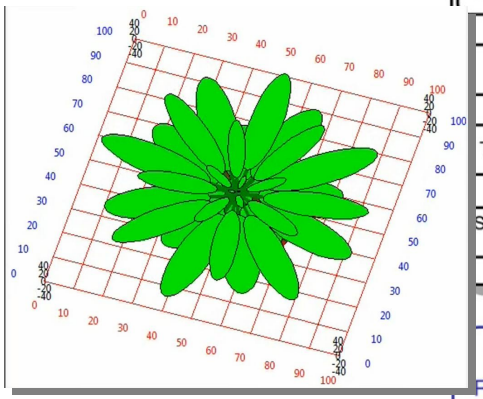
Coupled multi-scale models

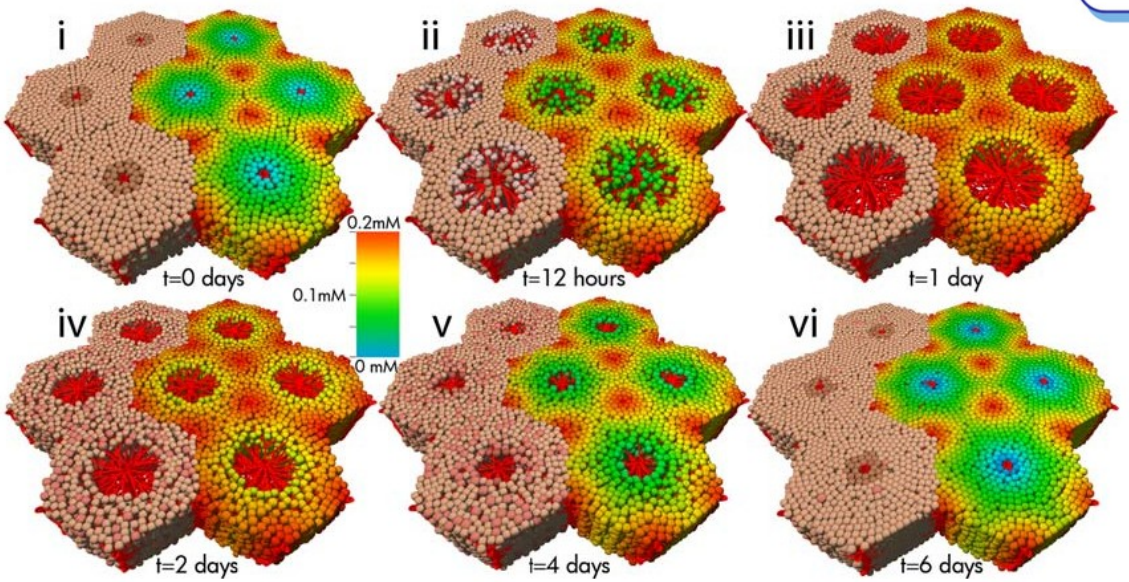
Theoretical and Computational roadblockers

24 orders of magnitude

9 orders of magnitude

Chew *et al.* (2014)



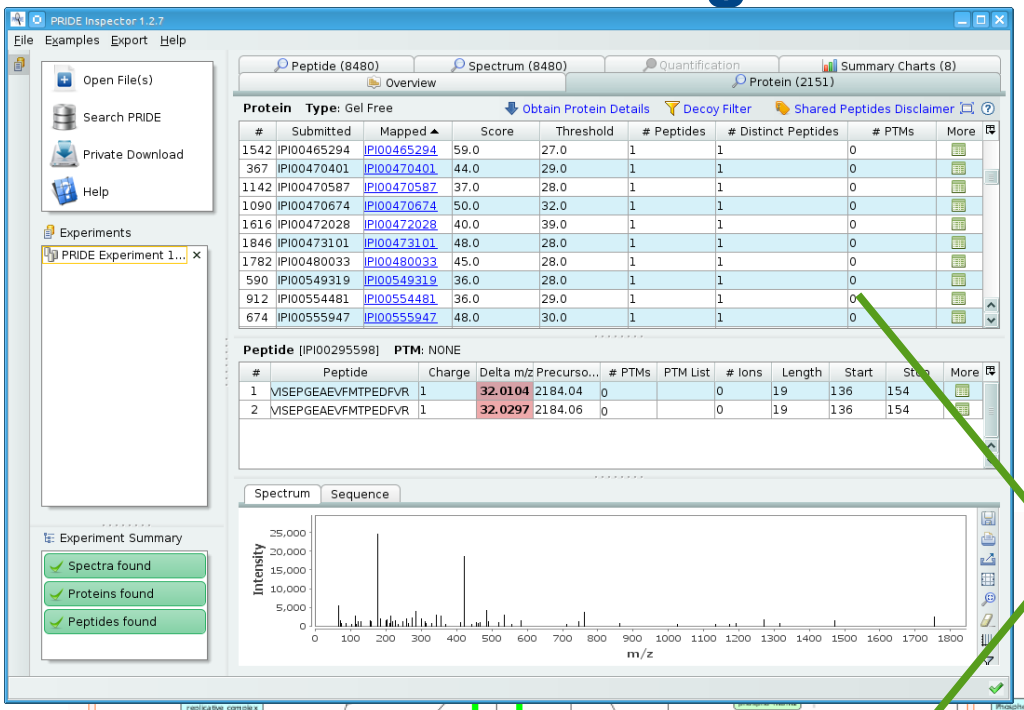


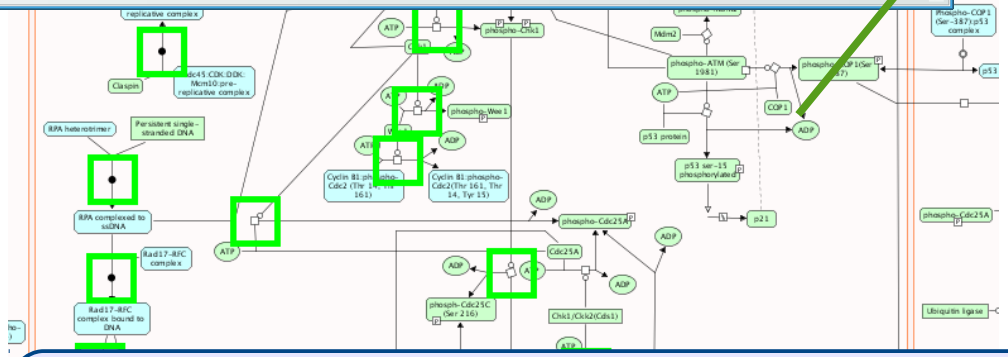
Schliess *et al.* (2014)

Babraham Institute

EMBL-EBI

Challenge 2: automatic generation





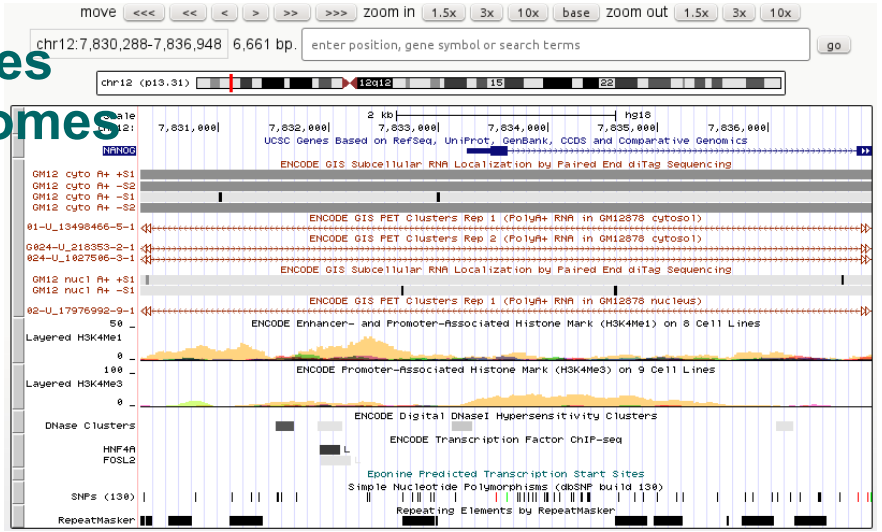
```
<assignmentRule metaid="metaid_0000023" variable="CT">
  <math xmlns="http://www.w3.org/1998/Math/MathML">
    <apply>
      <plus/>
      <ci> C2 </ci>
      <ci> CP </ci>
      <ci> M </ci>
      <ci> pM </ci>
    </apply>
  </math>
</assignmentRule>
</listOfRules>
<listOfReactions>
<reaction id="Reaction1" name="cyclin_cdc2k dissociation" metaid="_000010" reversible="false">
  <annotation>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:bqmodel="http://biom
      <rdf:Description rdf:about="#_000010">
        <bqbiol:hasVersion>
          <rdf:Bag>
            <rdf:li rdf:resource="http://identifiers.org/reactome/REACT_6308"/>
          </rdf:Bag>
        </bqbiol:hasVersion>
        <bqbiol:isVersionOf>
          <rdf:Bag>
            <rdf:li rdf:resource="http://identif                    000079"/>
          </rdf:Bag>
        </bqbiol:isVersionOf>
      </rdf:Description>
    </rdf:RDF>
  </annotation>
      <listOfReactants>
    <speciesReference species="M" metaid="_712369"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="C2" metaid="_712382"/>
    <speciesReference species="YP" metaid="_712394"/>
  </listOfProducts>
  <kineticLaw metaid="_712406">
    <math xmlns="http://www.w3.org/1998/Math/MathML">
      <apply>
        <times/>
        <ci> cell </ci>
        <ci> k6 </ci>
        <ci> M </ci>
```

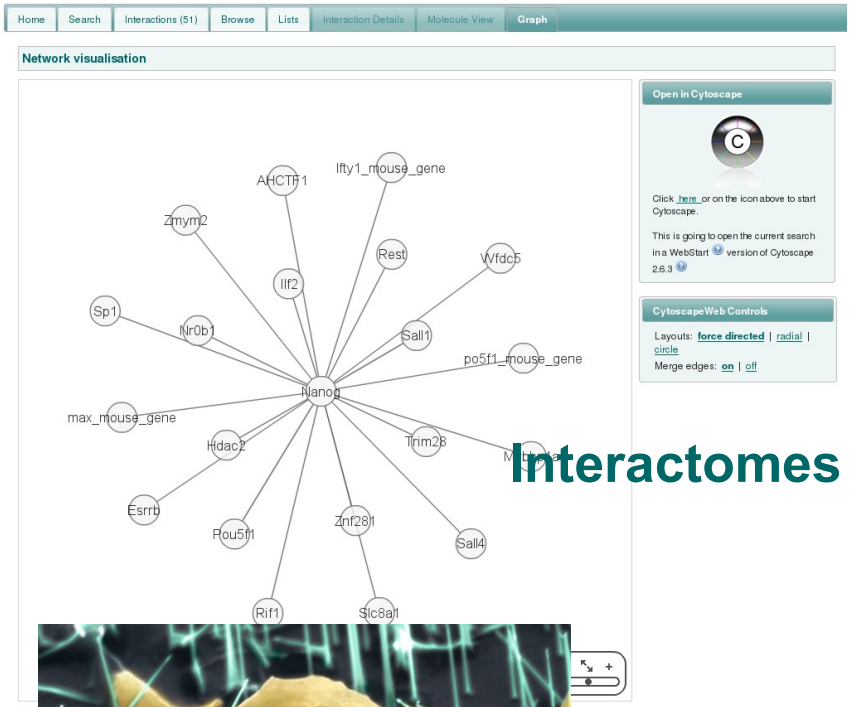
$$
\begin{aligned}
d[C2]/dt &= k_6[M] - k_8[{\sim}P][C2] + k_9[CP] \\
d[CP]/dt &= -k_3[CP][Y] + k_8[{\sim}P][C2] - k_9[CP] \\
d[pM]/dt &= k_3[CP][Y] - [pM]F([M]) + k_5[{\sim}P][M] \\
d[M]/dt &= [pM]F([M]) - k_5[{\sim}P][M] - k_6[M] \\
d[Y]/dt &= k_1[aa] - k_2[Y] - k_3[CP][Y] \\
d[YP]/dt &= k_6[M] - k_7[YP]
\end{aligned}
$$

Need to generate model automatically based on omics datasets. Known on small scale. The larger, the more human intervention needed.

Challenge 3: data heterogeneity

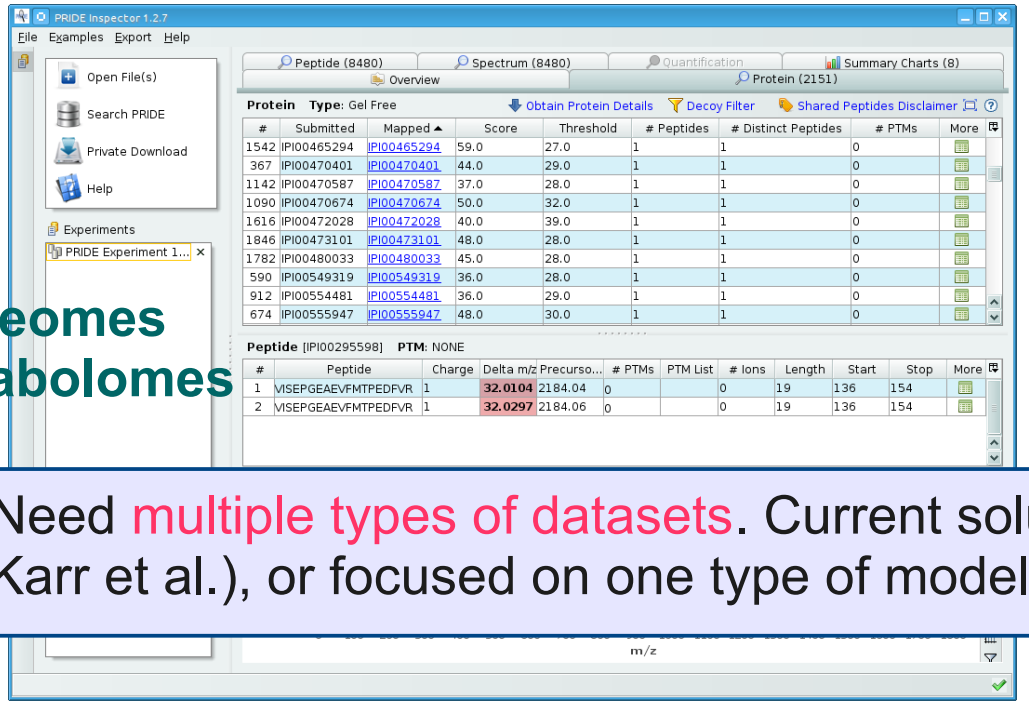
Genomes
epigenomes

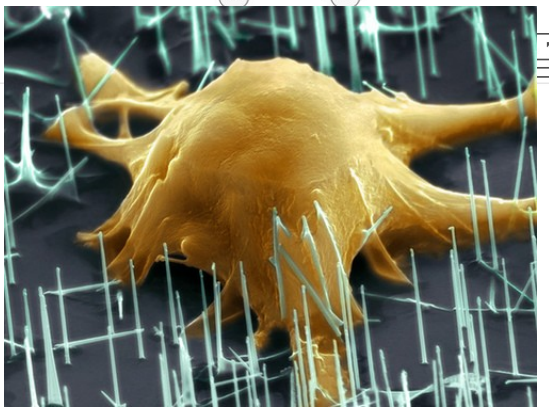




Interactomes

Proteomes
metabolomes

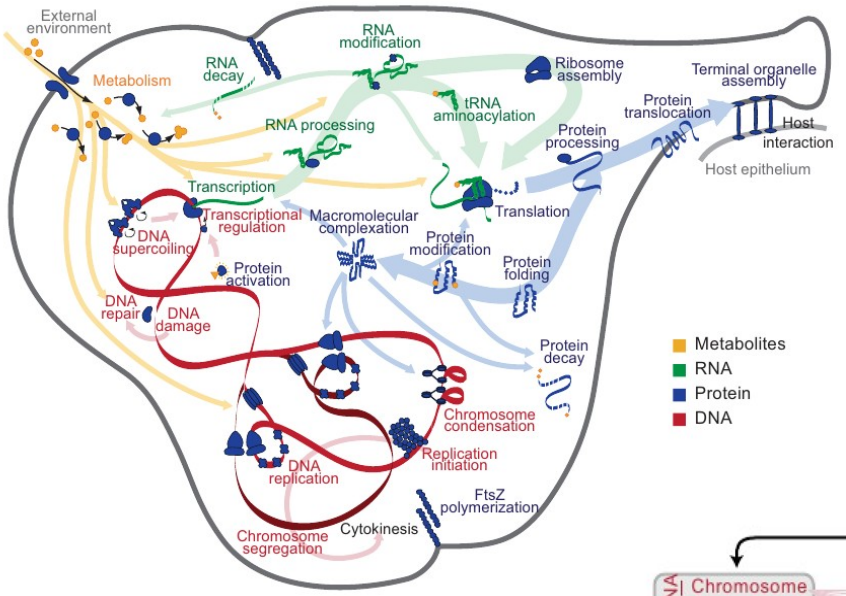




Imaging

Need multiple types of datasets. Current solutions either for small models (e.g Karr et al.), or focused on one type of model and data (e.g. metabolic models)

Modularity is mandatory



Theory

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

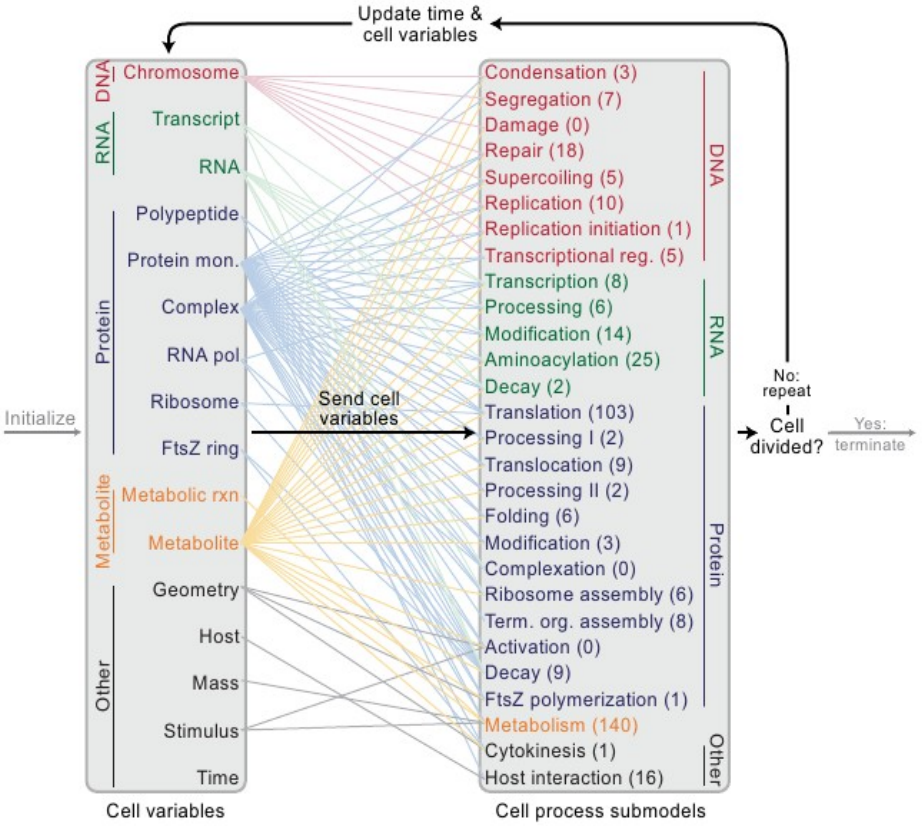
http://dx.doi.org/10.1016/j.cell.2012.05.044

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 range of data. The model provides insights into many previously unobserved cellular behaviors, including in vivo rates of protein-DNA association

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., 1984; Tomita et al., 1999), were limited by the difficulty in obtaining the necessary model parameters. Subsequently, alternative



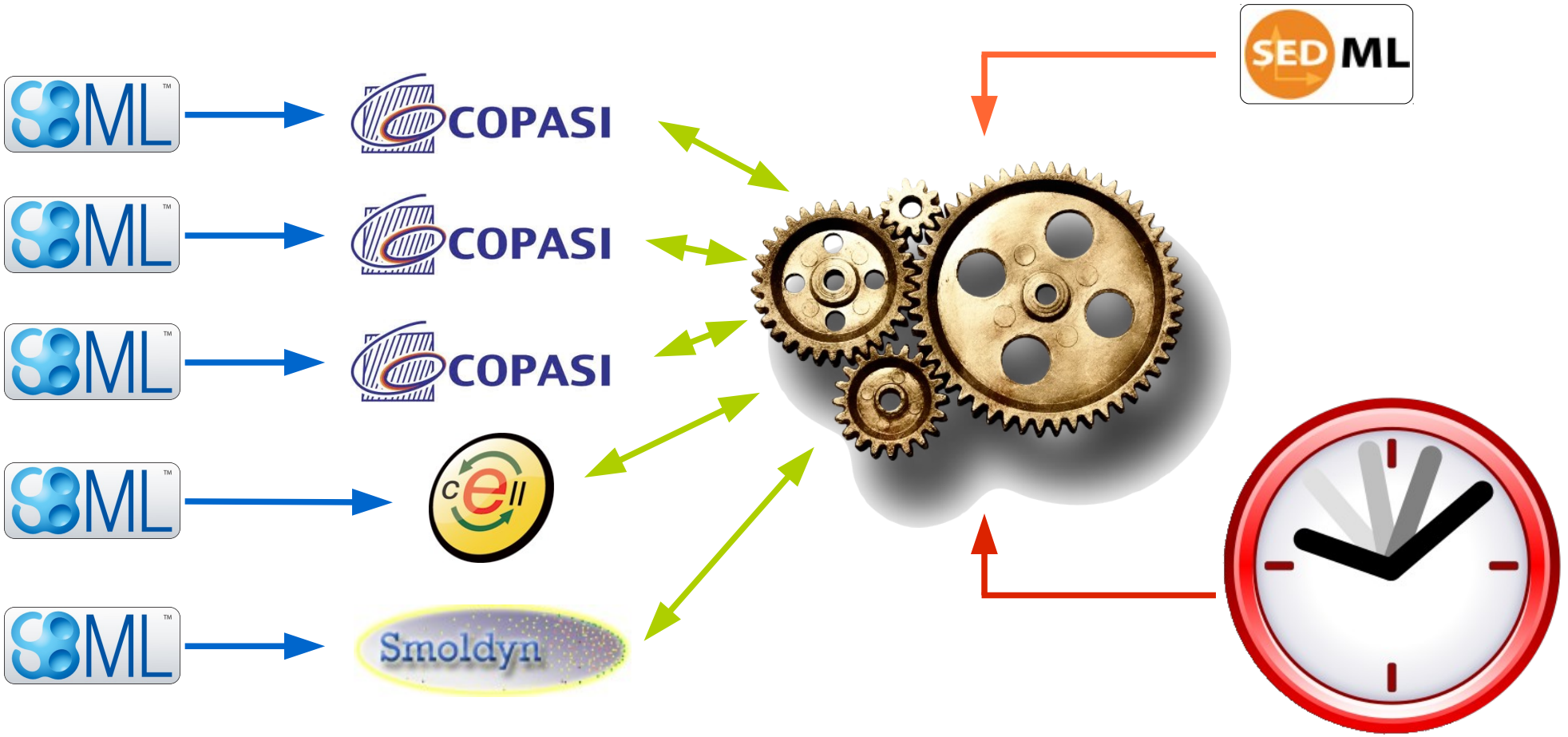
Babraham Institute

EMBL-EBI

Why is modularity important?

- Different parts can be modelled with different approaches
- Alternative modules representing the same biological process with different granularities or different modelling approaches
- Model families with alternative parts: avoid combinatorial explosion
- Modules can be developed at their own pace; easier versioning
- Distributed maintenance. No single individual or group can master the entirety of the project.

Challenge 4: Single software not sufficient



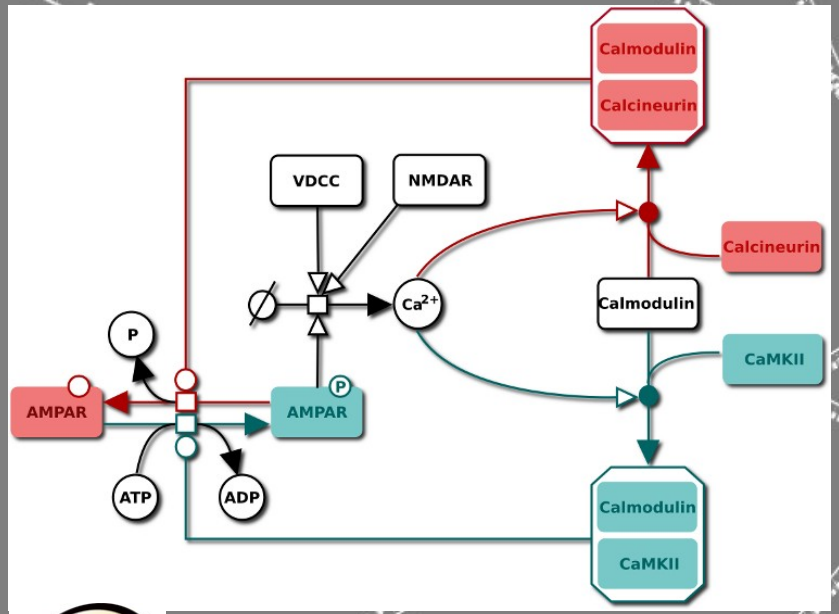
modular model

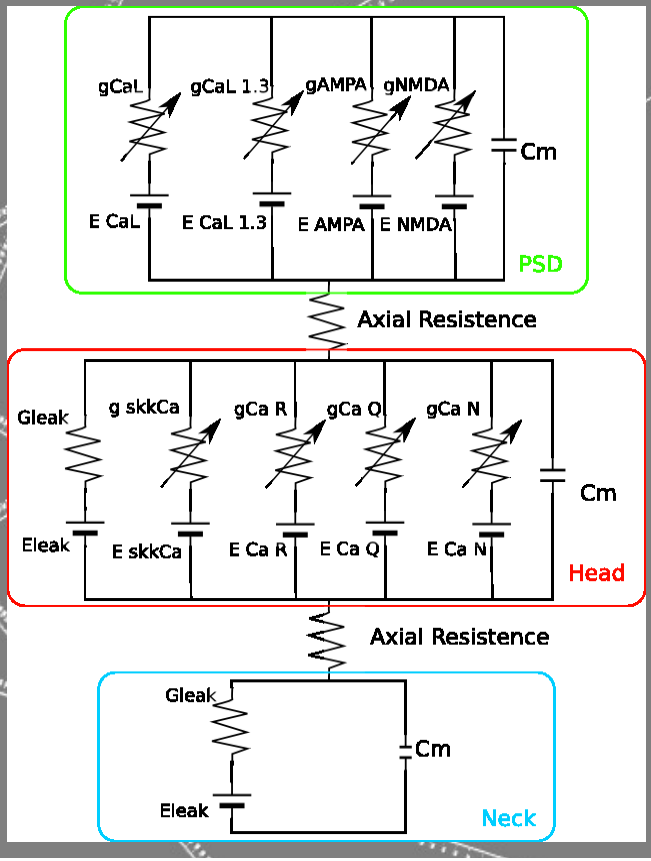
hybrid simulation system

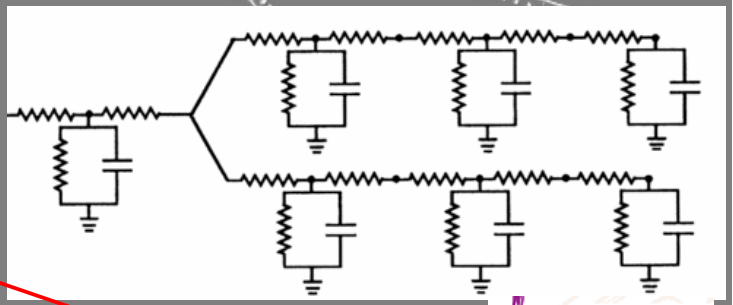
discrete event detection

adaptive synchronization

Whole cell: electro-biochemical models



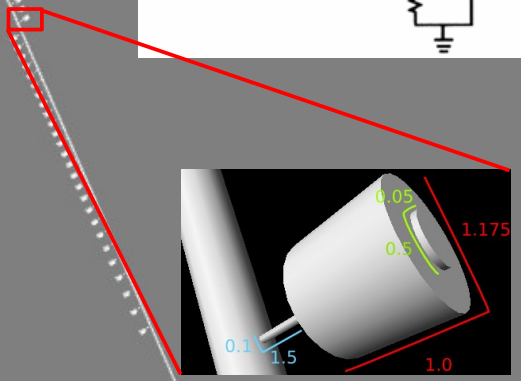




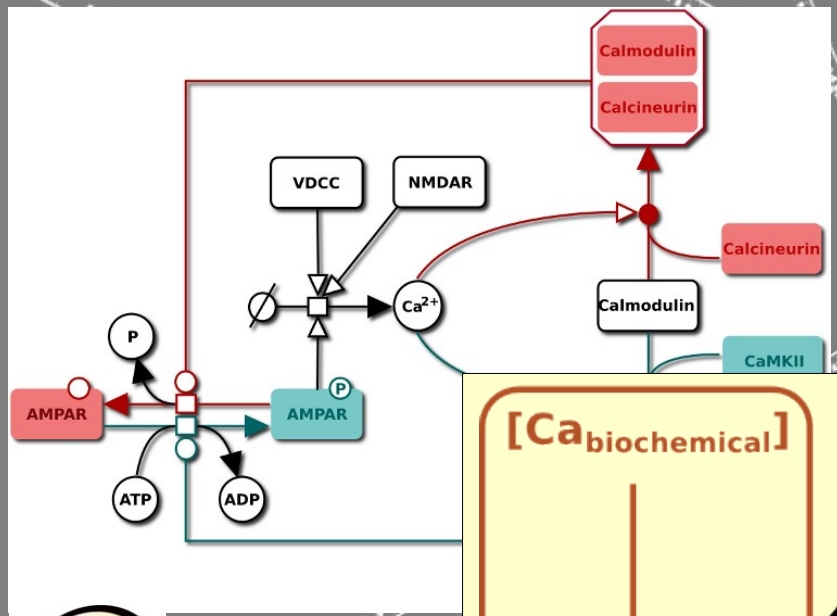
1504 spines

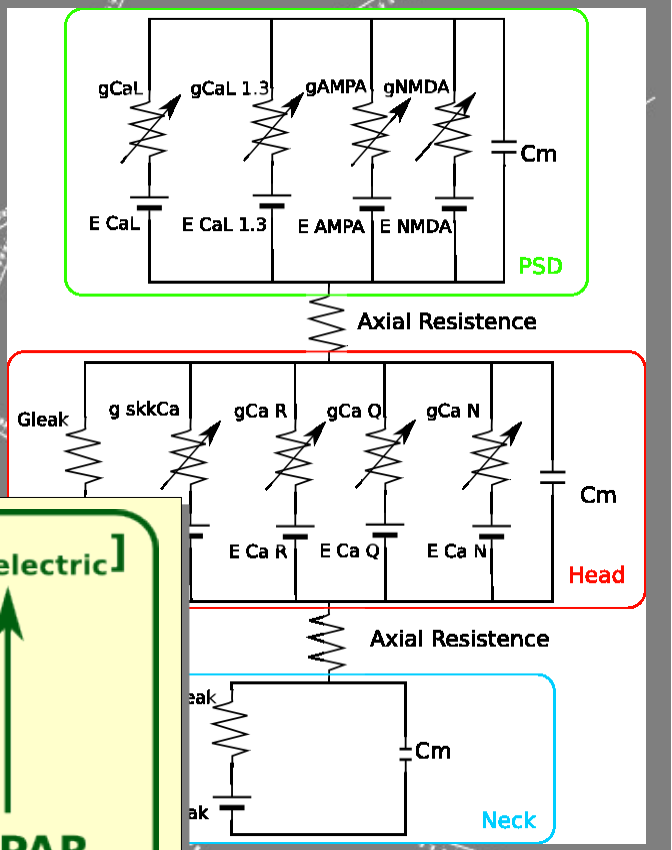
4761 compartments

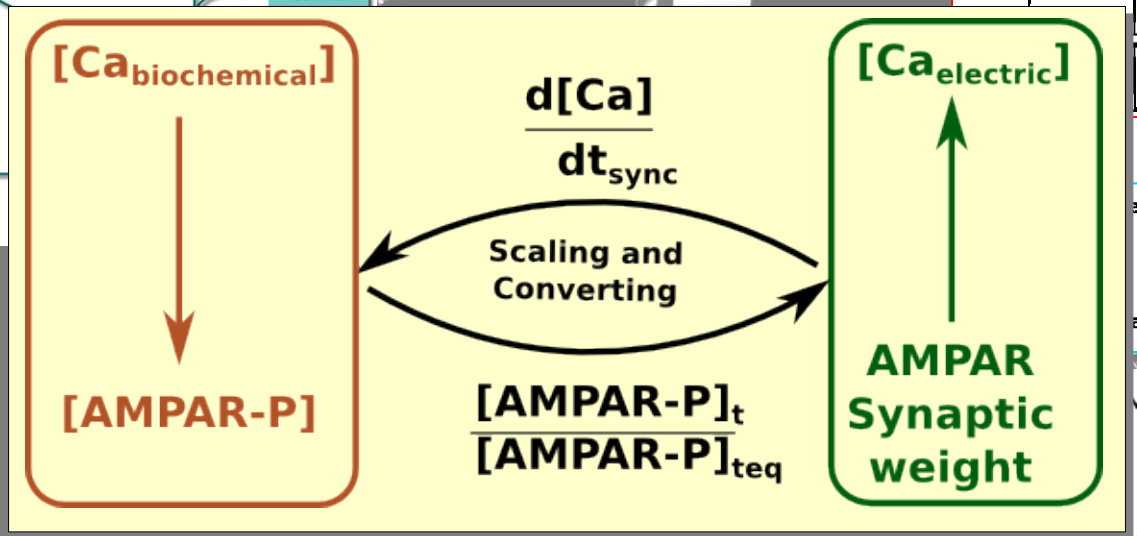
16362 channels

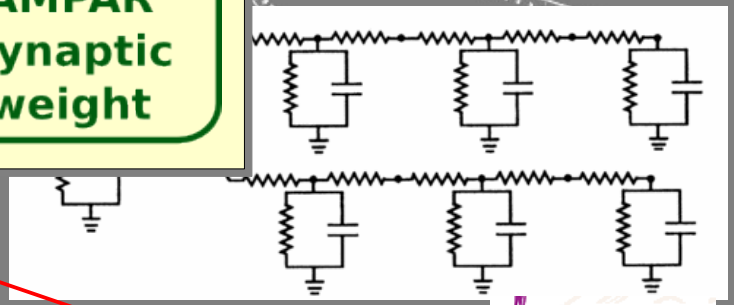


Mattioni et al (2012, 2013)





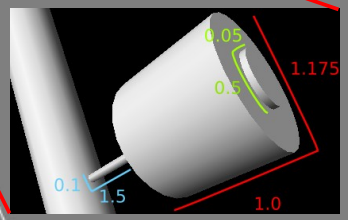




1504 spines

4761 compartments

16362 channels

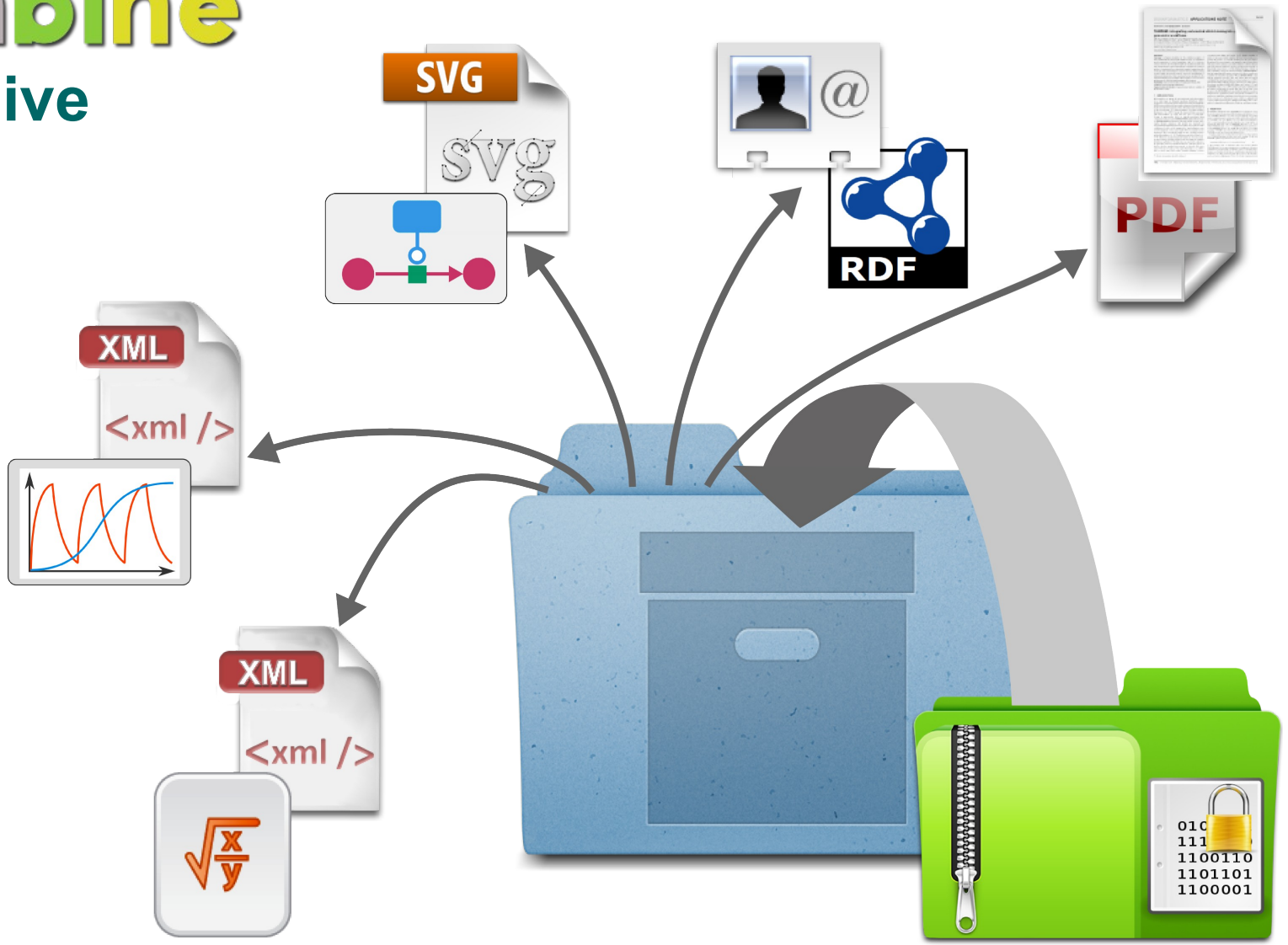


Mattioni et al (2012, 2013)

Challenge 5: virtualisation

Models must be self-contained and portable for everyone to reuse and customise → Virtual machines and modules (e.g. Rocks systems)

combine archive



http://co.mbine.org/standards/omex

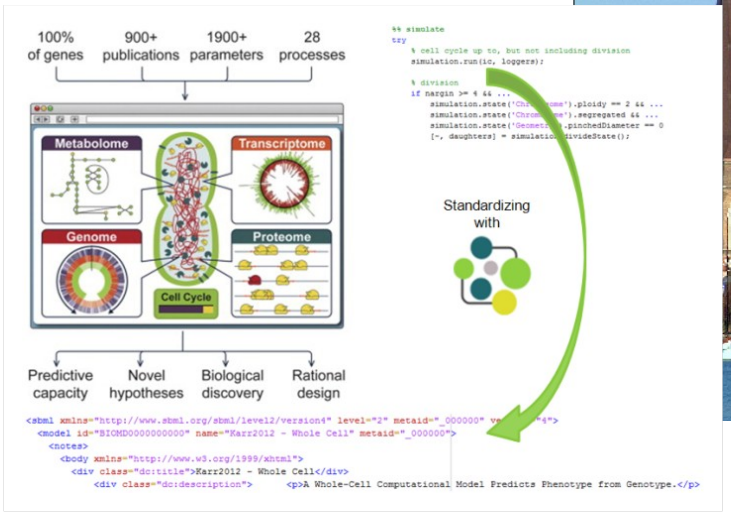
Bergman *et al. submitted*

Combining standards for today‘s models

@ITMZ Rostock

9th – 13th of March 2015, Rostock (Germany)
10 tutors & 50 students
fully funded

VolkswagenStiftung



@ITMZ Rostock

We need you to help us encode the whole-cell model in COMBINE standards!

Join us: http://bit.ly/wholecell


Frank Bergmann


Laurence Calzone


Viji Chelliah


Jonathan Karr


Wolfram Liebermeister


Nicolas Le Novère


Pedro Mendes


Chris Myers


Falk Schreiber


Dagmar Waltemath

Expert panel “digital organism” 2012-2013

Formation of a UK community network on Multi-Scale Biology

Google group MSB-net

https://groups.google.com/forum/#!forum/msb-net

Application for a network to be submitted to the BBSRC in 2014

→ 2/3 years
→ meetings
→ training
→ fostering collaborations
→ initiating grant applications

BioModels team

Ishan Ajmera
Raza Ali
Duncan Berenguier
Alexander Broicher
Vijayalakshmi Chelliah
Melanie Courtot
Marco Donizelli
Marine Dumousseau
Lukas Endler
Yvon Florent
Mihai Glont
Enuo He
Arnaud Henry
Henning Hermjakob
Gael Jalowicki
Nick Juty
Sarah Keating
Christian Knüpfer
Nicolas Le Novère
Chen Li
Camille Laibe
Stuart Moodie
Kedar Nath Natarajan
Jean-Baptiste Pettit
Michael Schubert
Maciej Swat
Karim Tazibt
Dagmar Waltemath
Sarala Wimalaratne
Anna Zhukova

Standards' editors

Richard Adams
Frank Bergman
Hamid Bolouri
Jonathan Cooper
Tobias Czauderna
Emek Demir
Andrew Finney
Stefan Hoops
Mike Hucka
Sarah Keating
Hiroaki Kitano
Nicolas Le Novère
Yukiko Matsuoka
Huaiyu Mi
Andrew Miller
Stuart Moodie
Ion Moraru
Chris Myers
David Nickerson
Brett Olivier
Sven Sahle
Herbert Sauro
Jim Schaff
Falk Schreiber
Lucian Smith
Anatoly Sorokin
Alice Villeger
Katja Wegner
Darren Wilkinson
Dagmar Waltemath

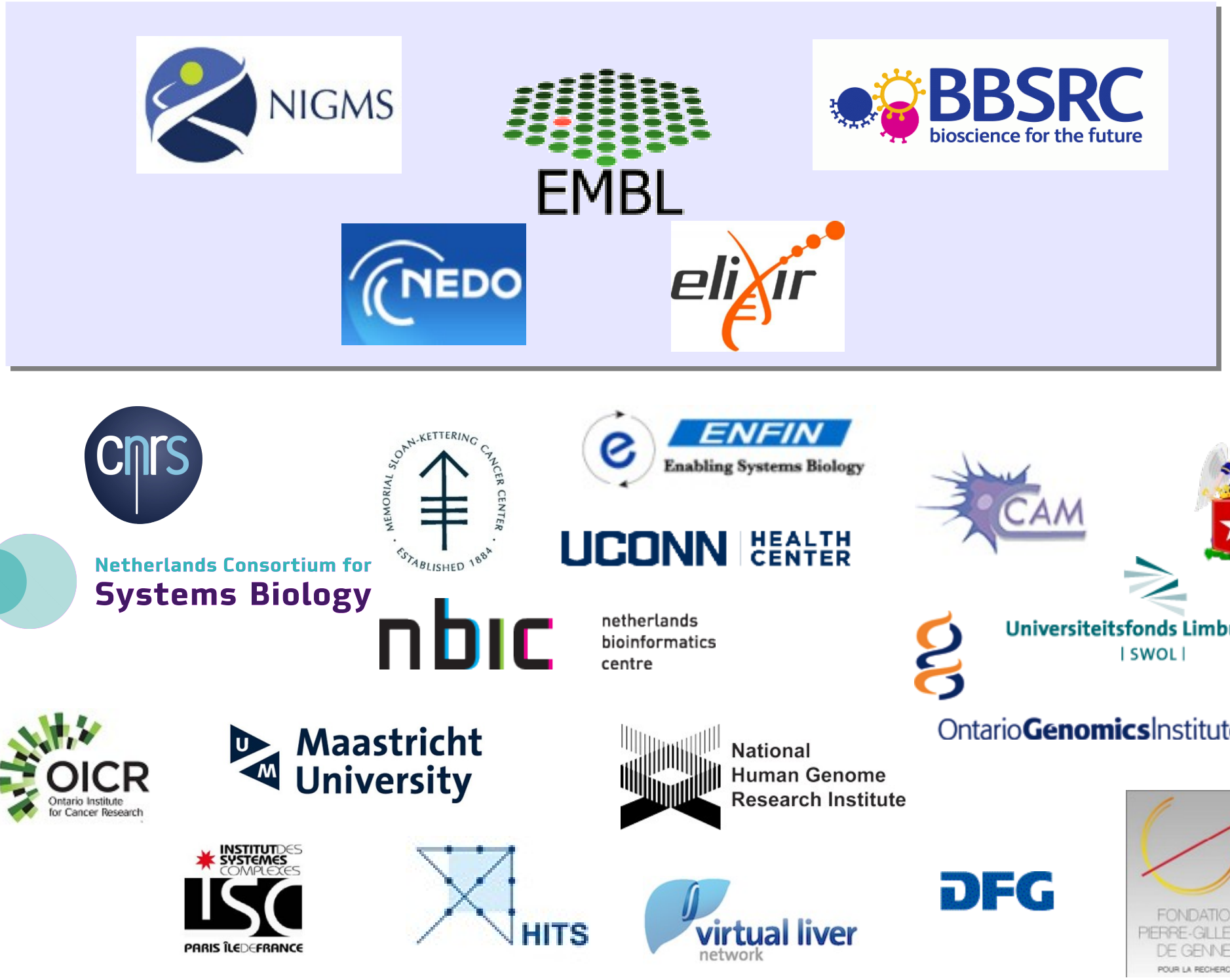
Path2Models

Finja Büchel
Claudine Chaouiya
Tobias Czauderna
Andreas Dräger
Mihai Glont
Martin Golebievski
Henning Hermjakob
Mike Hucka
Douglas Kell
Roland Keller
Camille Laibe
Nicolas Le Novère
Pedro Mendes
Florent Mittag
Wolfang Müller
Matthias Rall
Nicolas Rodriguez
Julio Saez-Rodriguez
Michael Schubert
Falk Schreiber
Neil Swainston
Martijn van Iersel
Clemens Wrzodek
Andreas Zell

The numerous scientists who participated to discussions

The community of Computational Systems Biology

You!


NIGMS
EMBL
BBSRC
bioscience for the future
NEDO
elixir
cnrs
MEMORIAL SLOAN-KETTERING CANCER CENTER
ESTABLISHED 1884
ENFIN
Enabling Systems Biology
CAM
UCONN HEALTH CENTER
Netherlands Consortium for
Systems Biology
nbic
netherlands
bioinformatics
centre
Universiteitsfonds Limburg
| SWOL |
OICR
Ontario Institute
for Cancer Research
Maastricht
University
National
Human Genome
Research Institute
OntarioGenomicsInstitute
INSTITUT DES SYSTEMES COMPLEXES
ISC
PARIS ÎLE DE FRANCE
HITS
virtual liver
network
DFG
FONDATION
PIERRE-GILLES
DE GENNES
POUR LA RECHERCHE