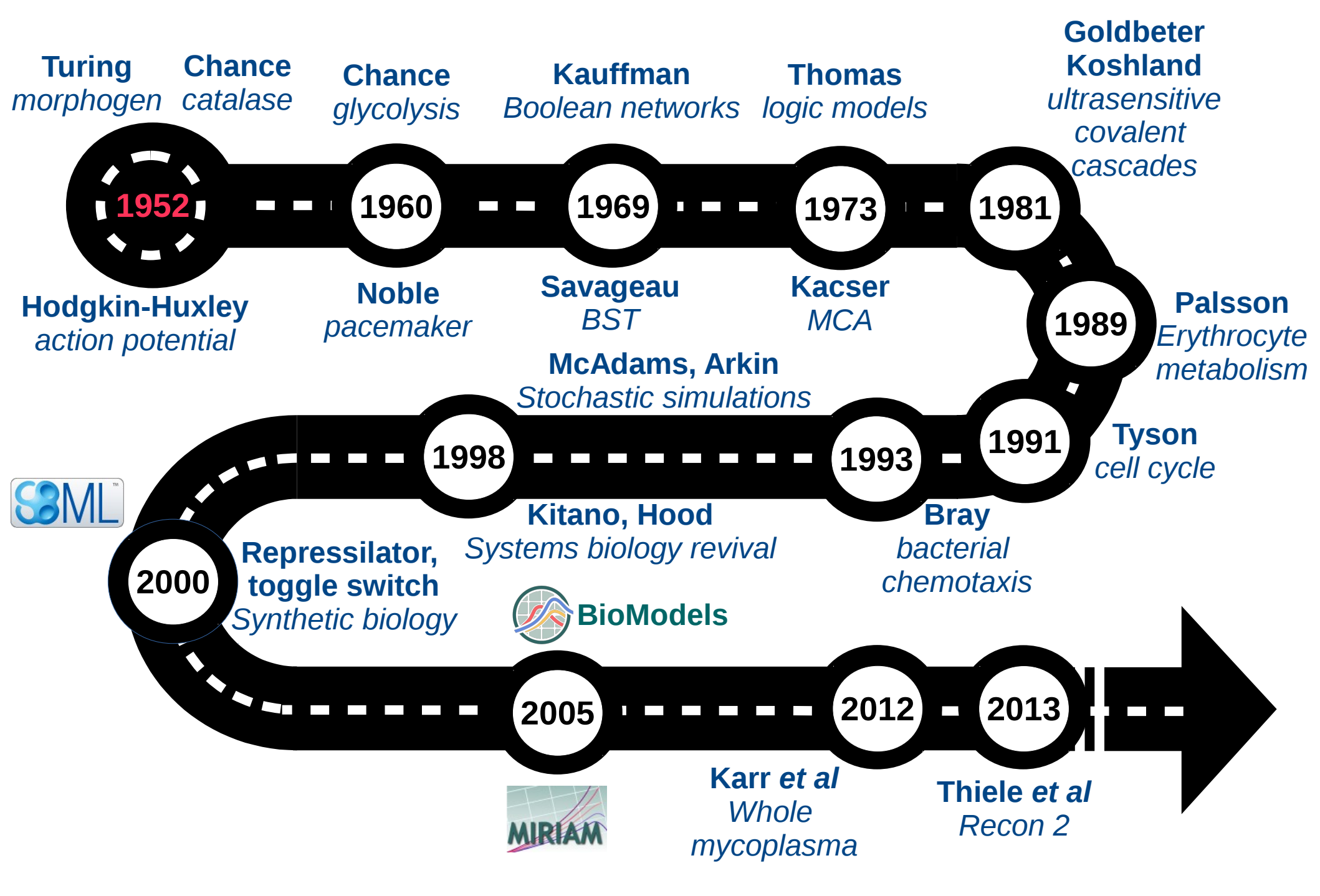
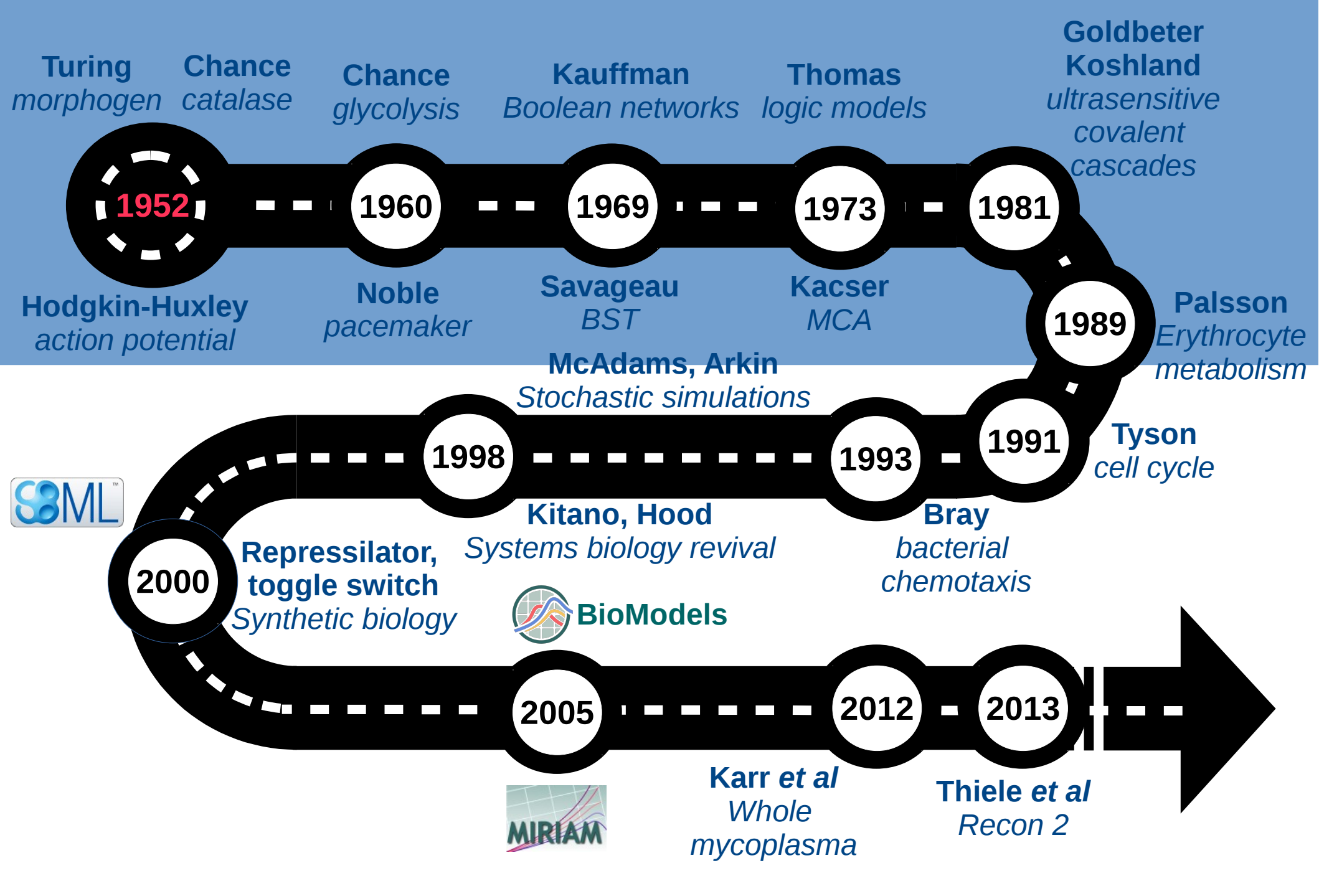
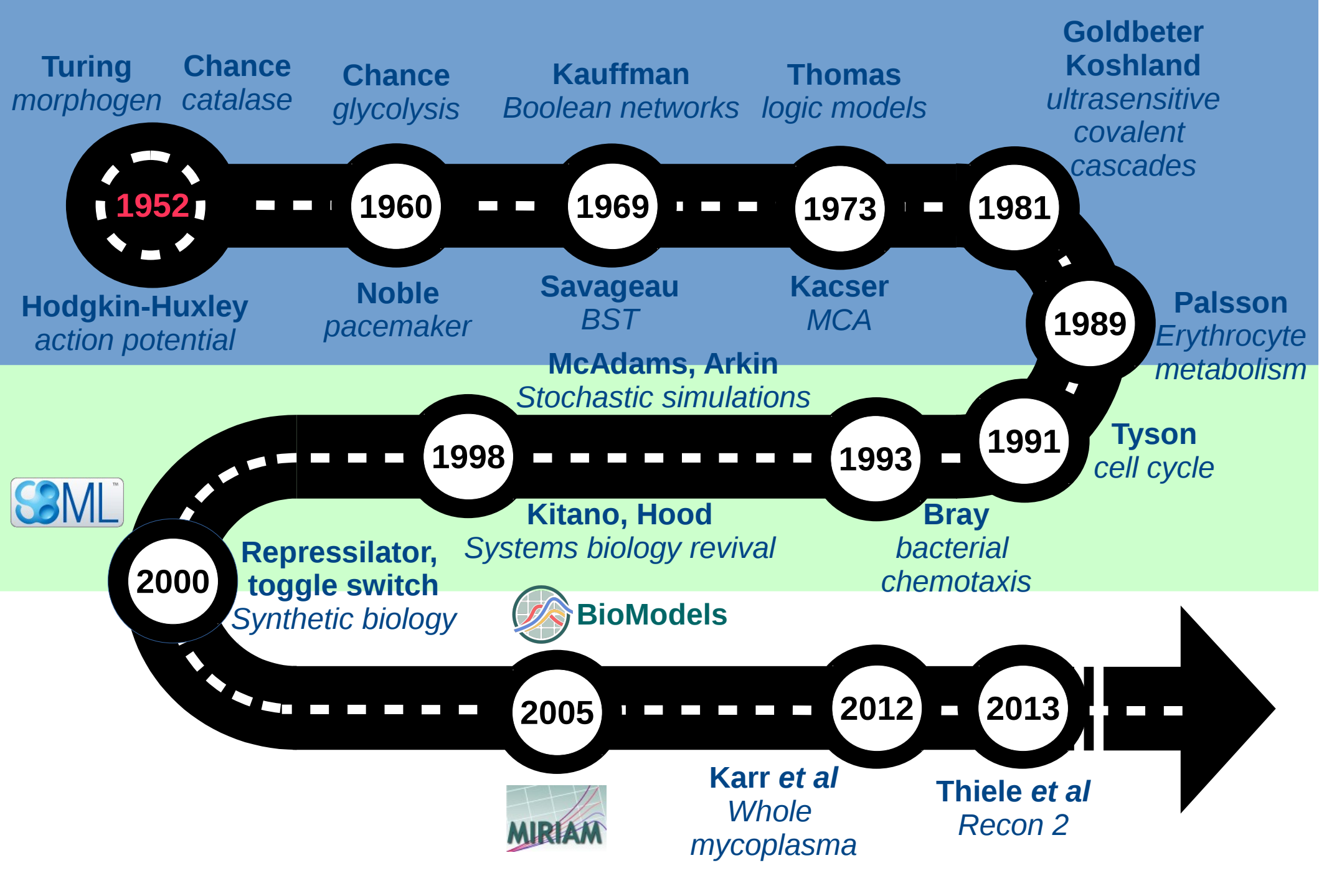
BioModels

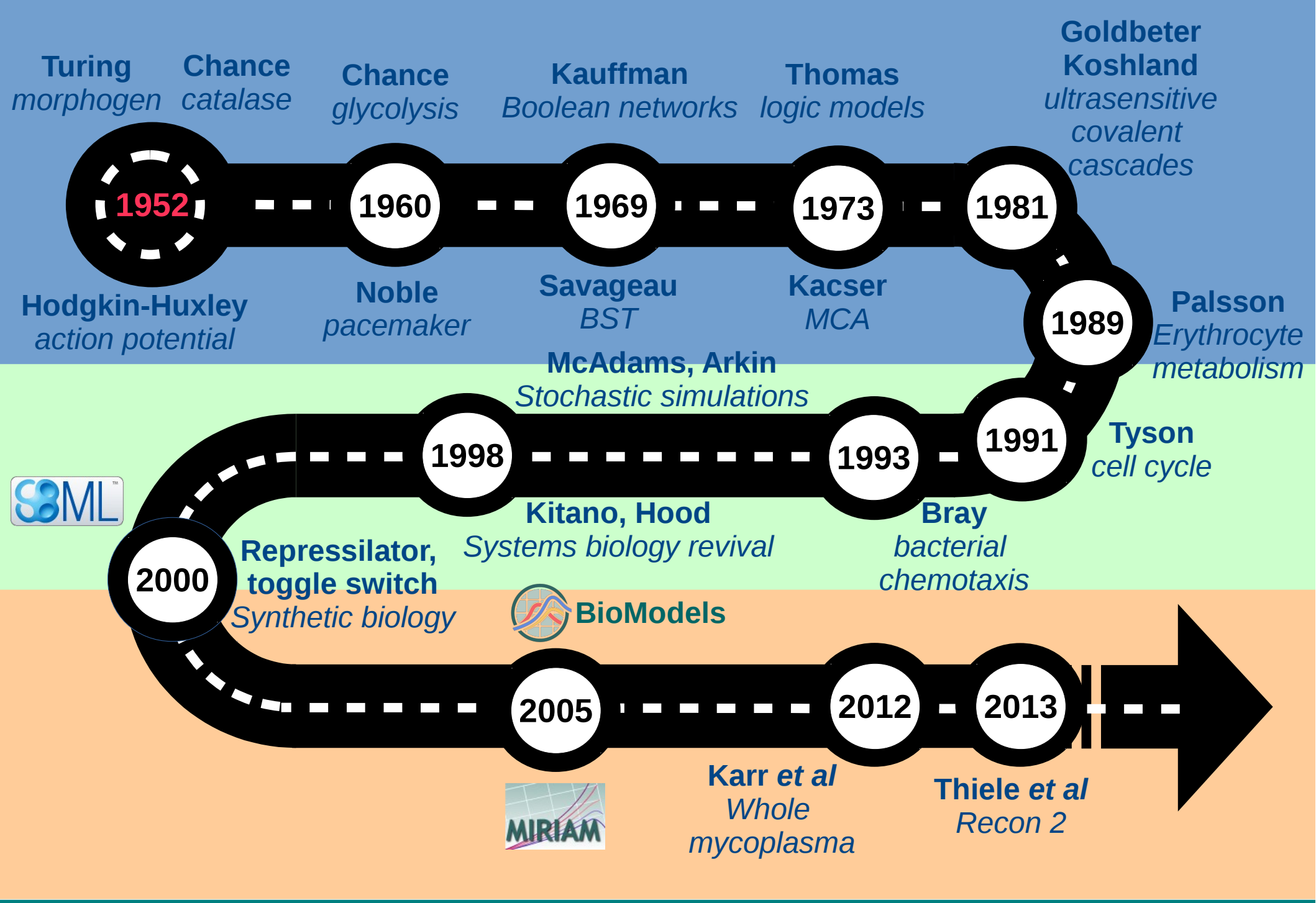
Sharing and re-using computational models of biological processes

Nicolas Le Novère, Babraham Institute, EMBL-EBI


Turing
morphogen
Chance
catalase
Chance
glycolysis
Kauffman
Boolean networks
Thomas
logic models
Goldbeter
Koshland
ultrasensitive
covalent
cascades
1952
1960
1969
1973
1981
Hodgkin-Huxley
action potential
Noble
pacemaker
Savageau
BST
Kacser
MCA
1989
Palsson
Erythrocyte
metabolism
McAdams, Arkin
Stochastic simulations
1998
1993
1991
Tyson
cell cycle
SBML
Kitano, Hood
Systems biology revival
Bray
bacterial
chemotaxis
2000
Repressilator,
toggle switch
Synthetic biology
BioModels
2005
2012
2013
MIRIAM
Karr et al
Whole
mycoplasma
Thiele et al
Recon 2


Turing
morphogen
Chance
catalase
Chance
glycolysis
Kauffman
Boolean networks
Thomas
logic models
Goldbeter
Koshland
ultrasensitive
covalent
cascades
1952
1960
1969
1973
1981
Hodgkin-Huxley
action potential
Noble
pacemaker
Savageau
BST
Kacser
MCA
1989
Palsson
Erythrocyte
metabolism
McAdams, Arkin
Stochastic simulations
1998
1993
1991
Tyson
cell cycle
SBML
Kitano, Hood
Systems biology revival
Bray
bacterial
chemotaxis
2000
Repressilator,
toggle switch
Synthetic biology
BioModels
2005
2012
2013
MIRIAM
Karr et al
Whole
mycoplasma
Thiele et al
Recon 2


Turing
morphogen
Chance
catalase
Chance
glycolysis
Kauffman
Boolean networks
Thomas
logic models
Goldbeter
Koshland
ultrasensitive
covalent
cascades
1952
1960
1969
1973
1981
Hodgkin-Huxley
action potential
Noble
pacemaker
Savageau
BST
Kacser
MCA
1989
Palsson
Erythrocyte
metabolism
McAdams, Arkin
Stochastic simulations
1998
1993
1991
Tyson
cell cycle
SBML
Kitano, Hood
Systems biology revival
Bray
bacterial
chemotaxis
2000
Repressilator,
toggle switch
Synthetic biology
BioModels
2005
2012
2013
MIRIAM
Karr et al
Whole
mycoplasma
Thiele et al
Recon 2


Turing
morphogen
Chance
catalase
Chance
glycolysis
Kauffman
Boolean networks
Thomas
logic models
Goldbeter
Koshland
ultrasensitive
covalent
cascades
1952
1960
1969
1973
1981
Hodgkin-Huxley
action potential
Noble
pacemaker
Savageau
BST
Kacser
MCA
1989
Palsson
Erythrocyte
metabolism
McAdams, Arkin
Stochastic simulations
1998
1993
1991
Tyson
cell cycle
SBML
Kitano, Hood
Systems biology revival
Bray
bacterial
chemotaxis
2000
Repressilator,
toggle switch
Synthetic biology
BioModels
2005
2012
2013
MIRIAM
Karr et al
Whole
mycoplasma
Thiele et al
Recon 2

The world before BioModels

- List of models on webpages. Generally a link to the model in some format
- Main repositories were ModelsDb (neuroscience), DOQCS (neuroscience), the CellML and the SBML repositories. Plus a handful of personal web pages
- 100% of the models in the SBML repository were wrong. Wrong structure, wrong parametrisation, or did not produce the correct result
- It was generally not possible to test the other repositories because of the lack of tools

SBML Hackathon 2004 @ EBI

SBML Hackathon 2004 @ EBI

Ruminations on Creating a Biomodels.net

Mike Hucka
mhucka@caltech.edu
The SBML Team
Control and Dynamical Systems, MC 107-81
California Institute of Technology, Pasadena, CA 91125, US
http://www.sbml.org/

DRAFT

May 14, 2004

1 An Introduction

There are currently several project threads that have been floating around and I'd like to propose a project that I think would bring some of them together into a unique and globally useful resource. These threads include the following:

DRAFT

May 14, 2004

BioModels Database

1 An Introduction

There are currently several project threads that have been floating around and I'd like to propose a project that I think would bring some of them together into a unique and globally useful resource. These threads include the following:

MIRIAM annotations

- People have been repeatedly voicing the desire for a better database of SBML models. The existing repository on `sbml.org` was never meant to be a full database, only a stop-gap collection until such time as we can find the resources to do better. It may be time to organize a real database.
- People want databases of models to connect to other databases, especially databases of molecular information, Medline, etc.
- We have connections now to EBI and other groups who have expertise in databases and an eagerness to connect them to computational modeling efforts in systems biology. The time is ripe for leveraging these connections.
- The SBML and CellML groups have a long-standing interest to connect not only to each other's efforts, but to others such as BioPAX.

COMBINE

- We have technology now to translate a substantial portion of models from CellML to SBML. Technology exists for translating SBML Level 1 to CellML, and although it doesn't exist for SBML Level 2, there is good reason to believe it could be implemented. It is therefore highly likely that a significant portion of models in practice can be cross-translated.

SBO

- The CellML group has been working on ontologies for models which may ultimately be an ideal basis for a unifying database schema that could serve as a master schema for storing models in format-independent manner. (It is worth remembering that SBML was never designed to be a database representation. It may not be suitable for that role. Ditto for CellML. Internal to a database, it may make sense to use a root schema from which projections of the contents of a model are then translated out to SBML, BioPAX, CellML, whatever.)

Heidelberg Fall 2004

Unique database for models? No, thank you. But common curation guidelines

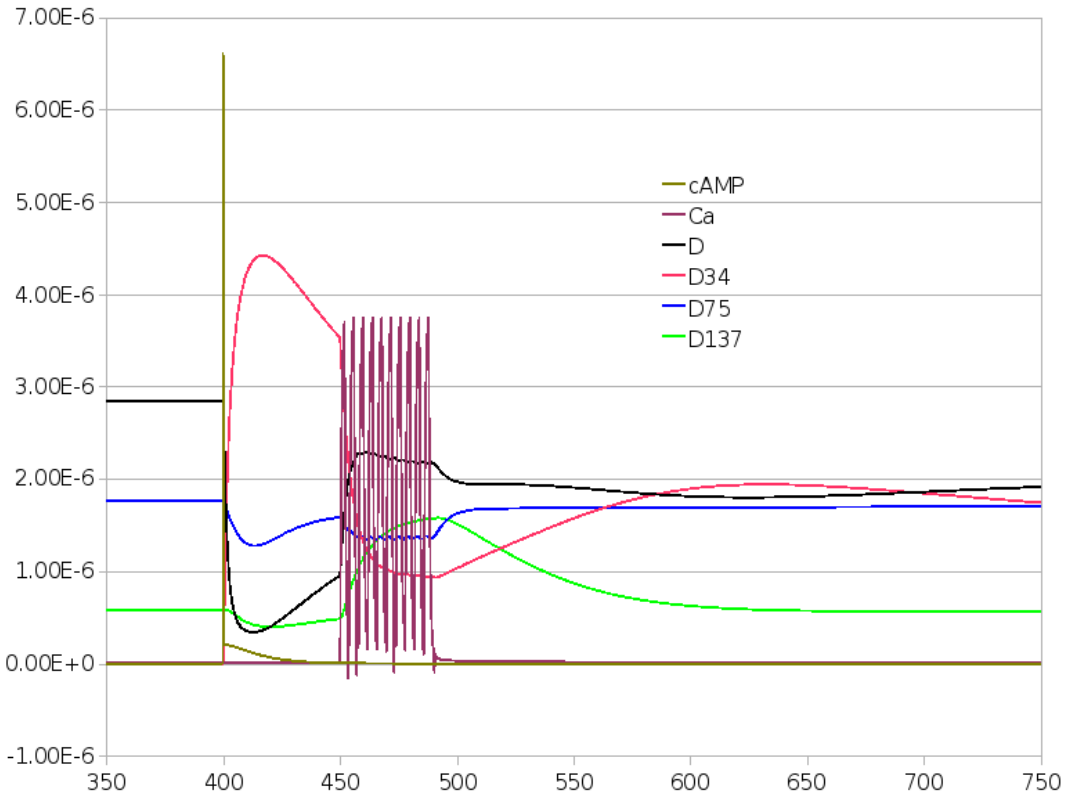
http://co.mbine.org/standards/miriam

MIRIAM

What is BioModels?

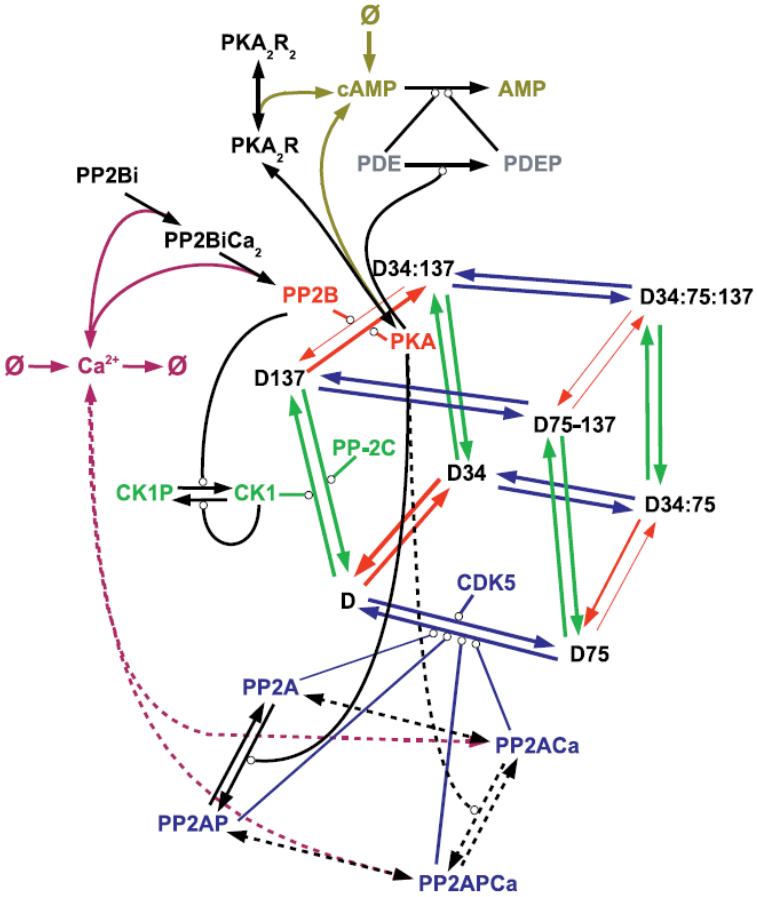
- Store and serve quantitative models of biological interest
- Models described in the peer-reviewed scientific literature + models automatically generated from pathway resources
- Models are curated: computer software check the syntax, while human curators check the semantics
- Models are simulated to ensure they provide the expected results
- Model components are annotated, to improve identification and retrieval
- Models are accepted in several formats, and served in several others

Process based biochemical models



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

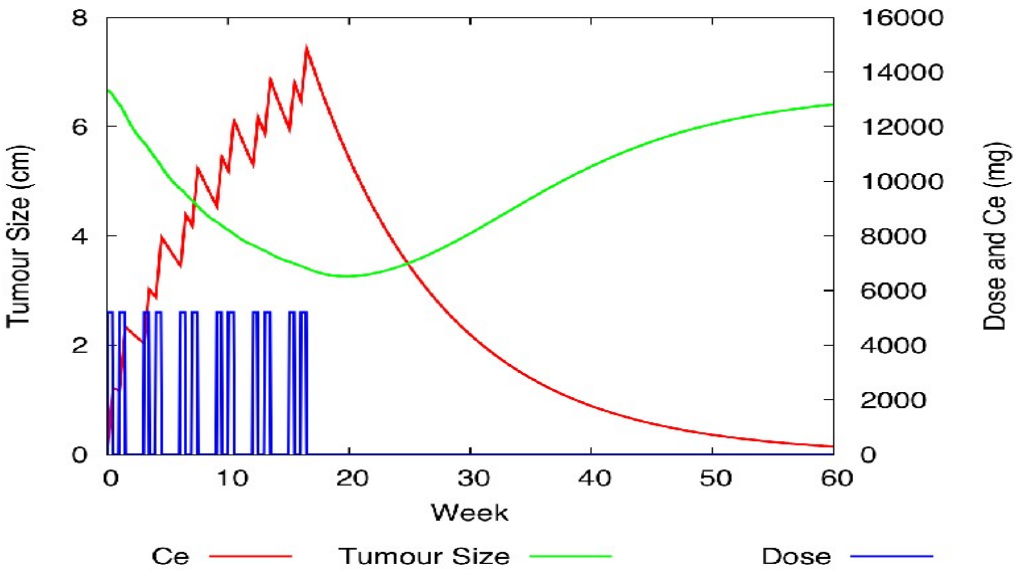
BIOMD0000000153



reaction:

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

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

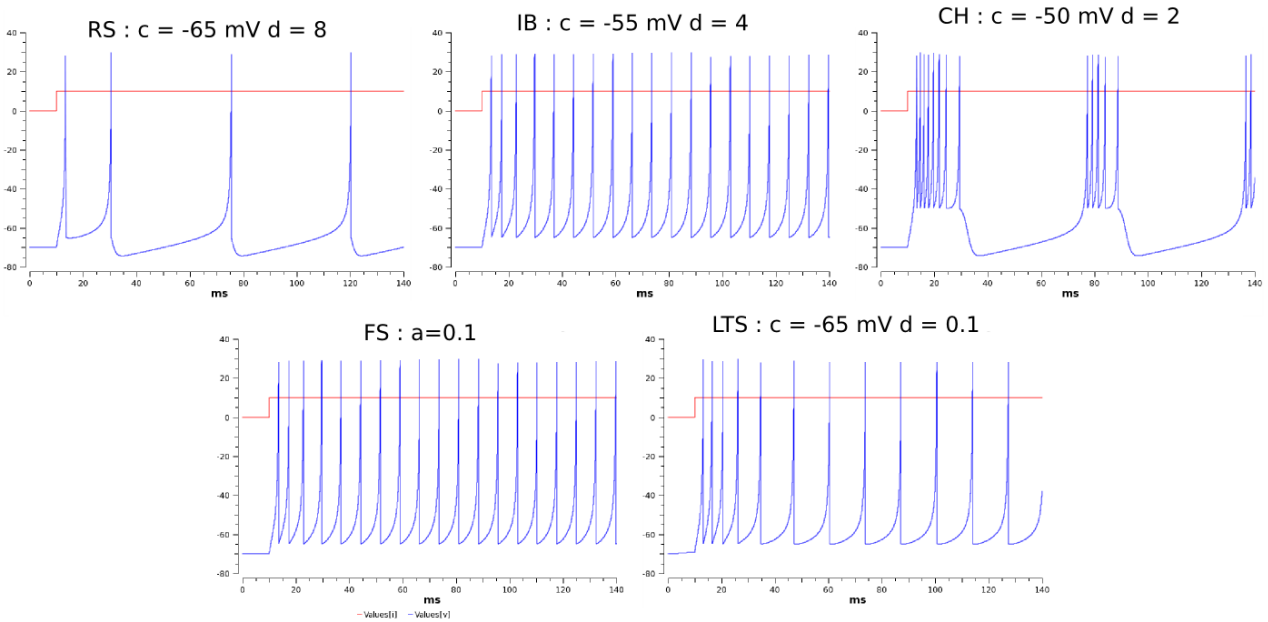
rate rule:

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

assignment rule:

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

Neuroscience models



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

BIOMD0000000127

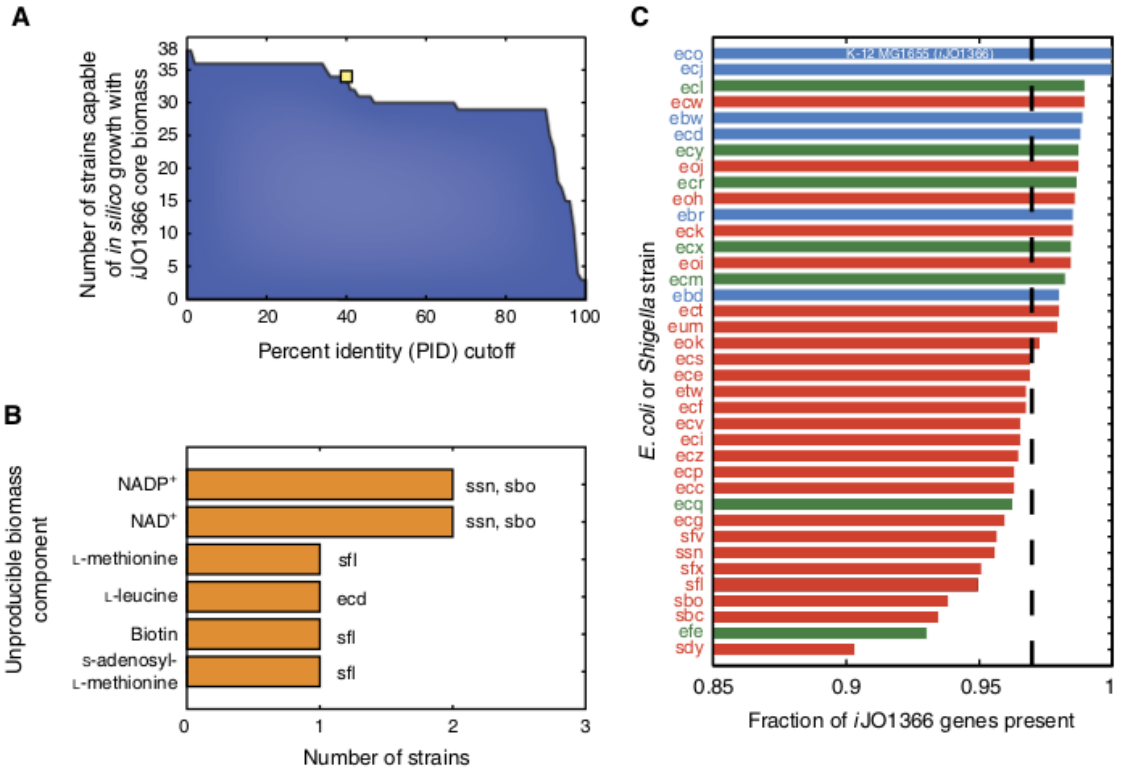
rate rule:

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

event:

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

Flux Balance Analysis models

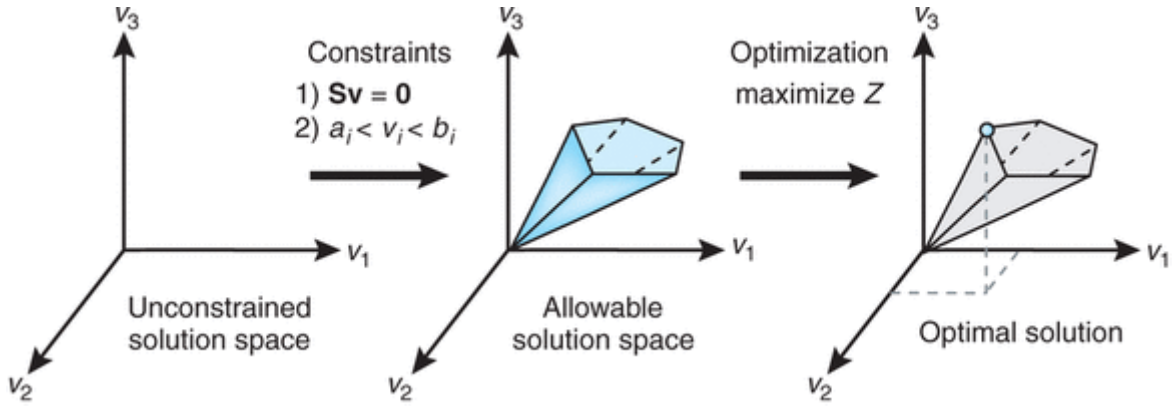


Orth et al. A comprehensive genome-scale reconstruction of Escherichia coli metabolism - 2011
Mol Syst Biol (2011);7:535

MODEL1108160000

reaction:

$$v_i = flux_i$$



BioModels is not limited to biochemistry!

Is it about bio[logy|chemistry|medicine]?

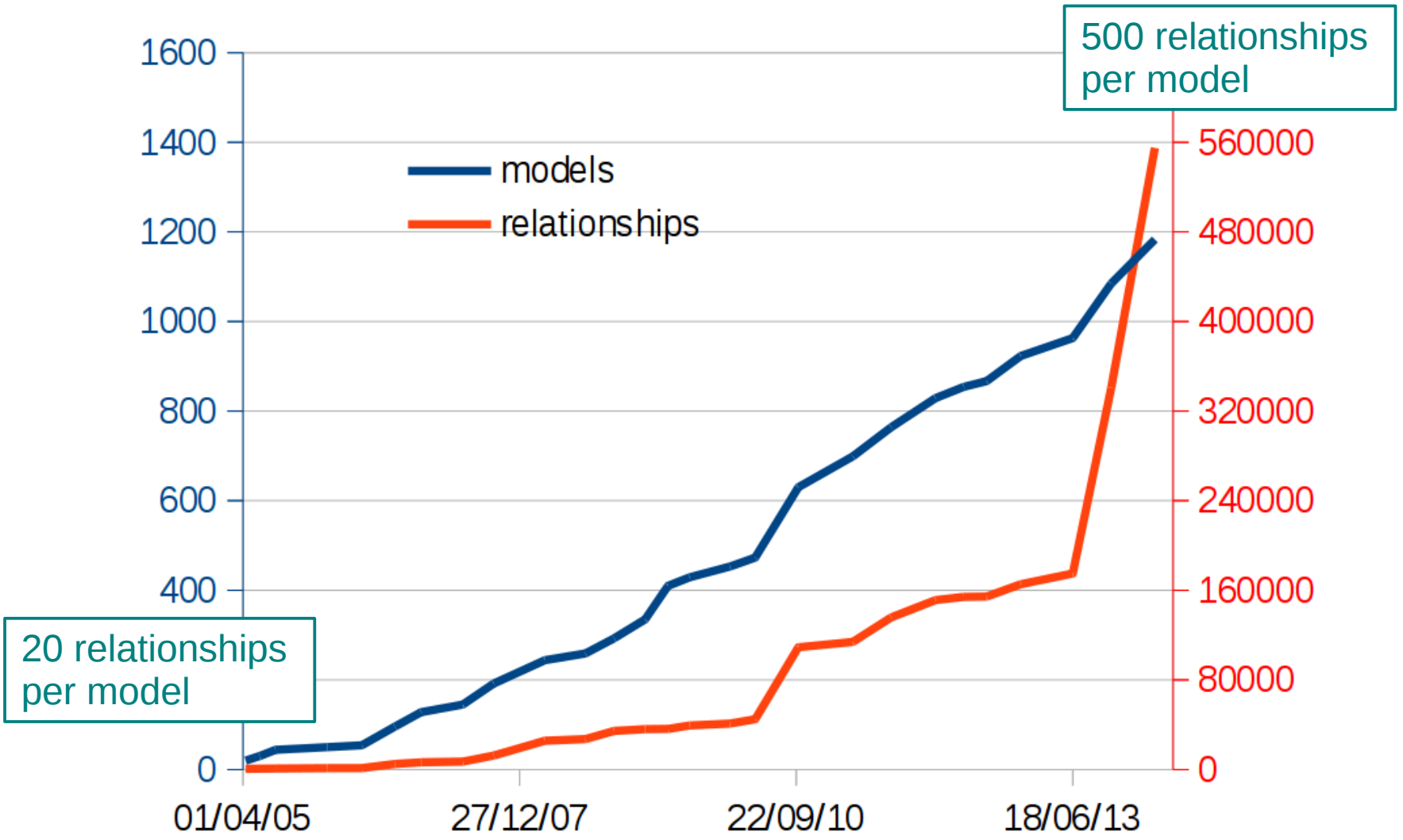
Is it a computational model?

Is it published?

Then it should be in BioModels

Was it worth it?

Complexity of shared models on the rise



BioModels growth (published models branch) since its creation

"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

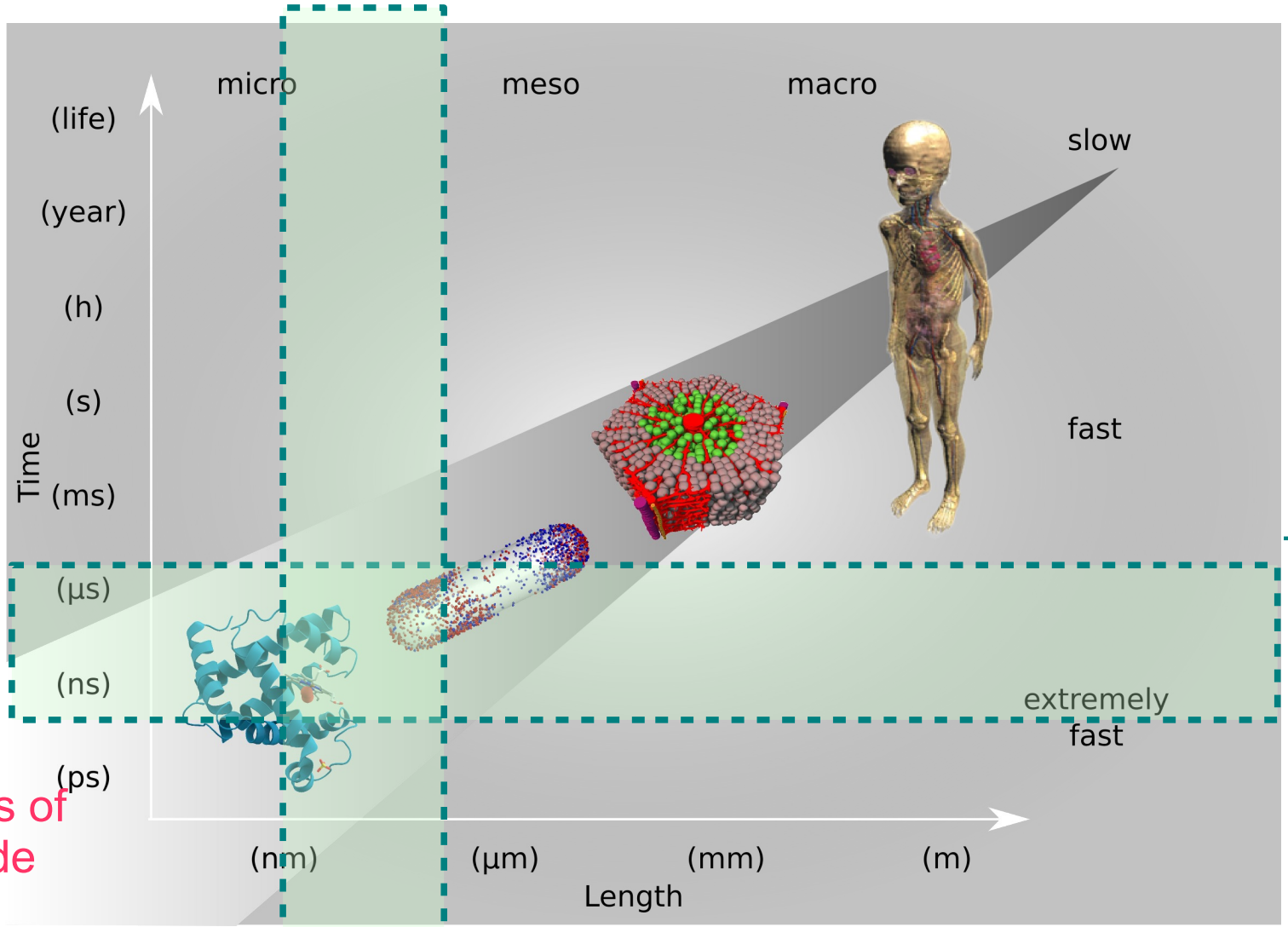
What now?

Expand the focus further

- Domain
  - Synthetic biology
  - Cell biology
  - Development
  - Physiology
  - Neuroscience
- Types of models
  - Spatial models
  - Multi-agent models
  - Modular models
  - Statistical models

- More formats
- Flexible infrastructure
- Evolution of what is a “model”
- More important role of metadata, past the cross-references
- Active links with other communities

Multi-scale modelling



Coupled multi-scale models

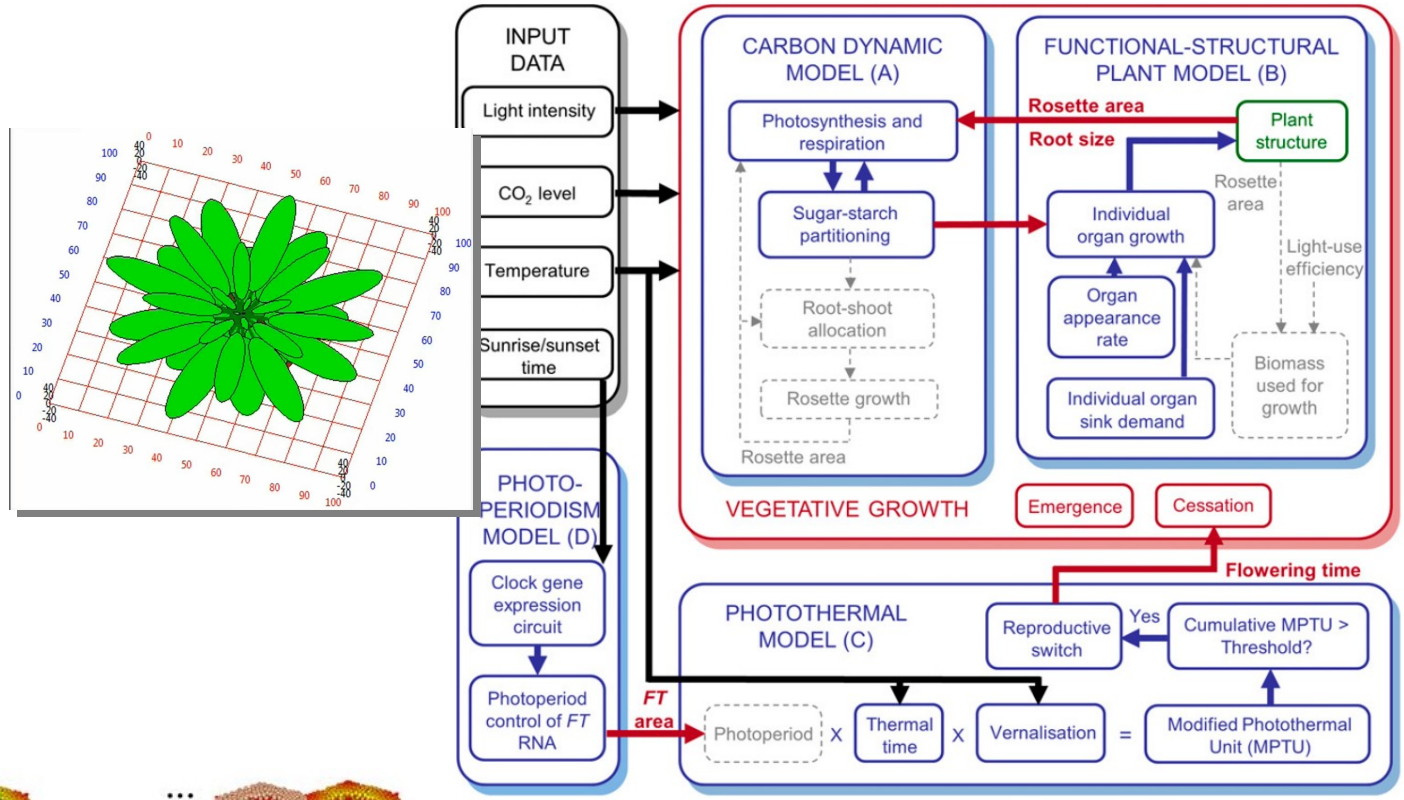
Theoretical and Computational roadblockers

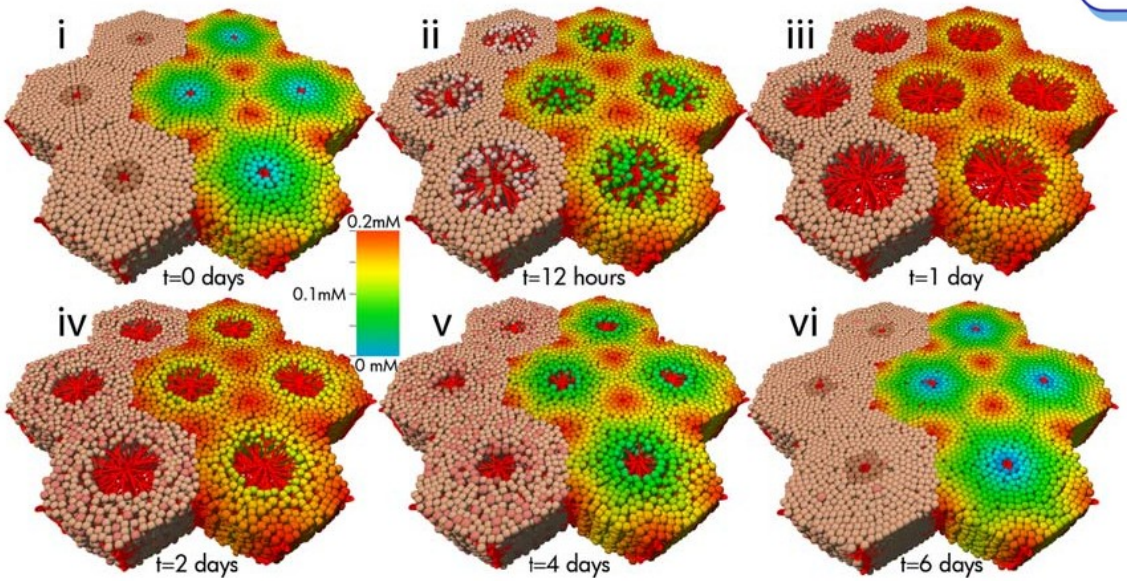
24 orders of magnitude

9 orders of magnitude

Adapted from Castiglione *et al.* (2014)

Chew *et al.* (2014)

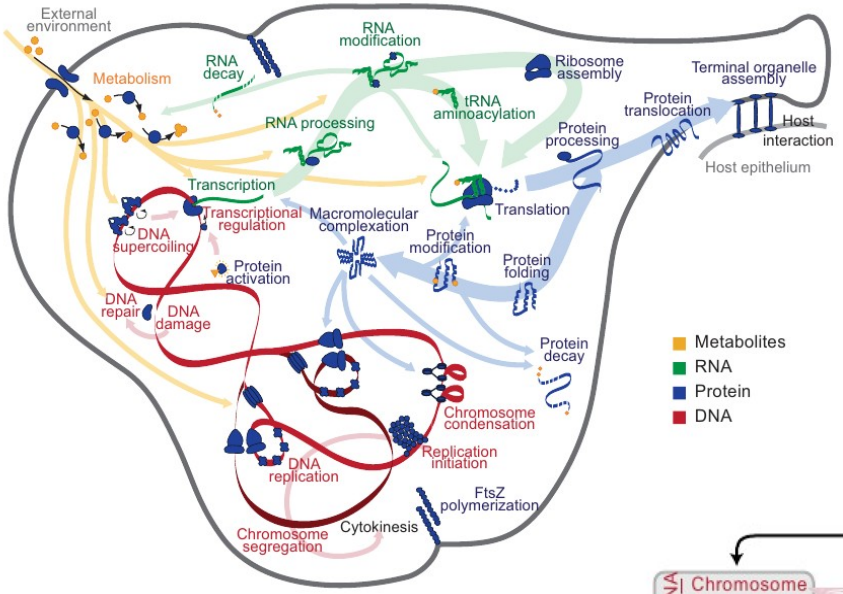




Schliess *et al.* (2014)

Why is modularity important?

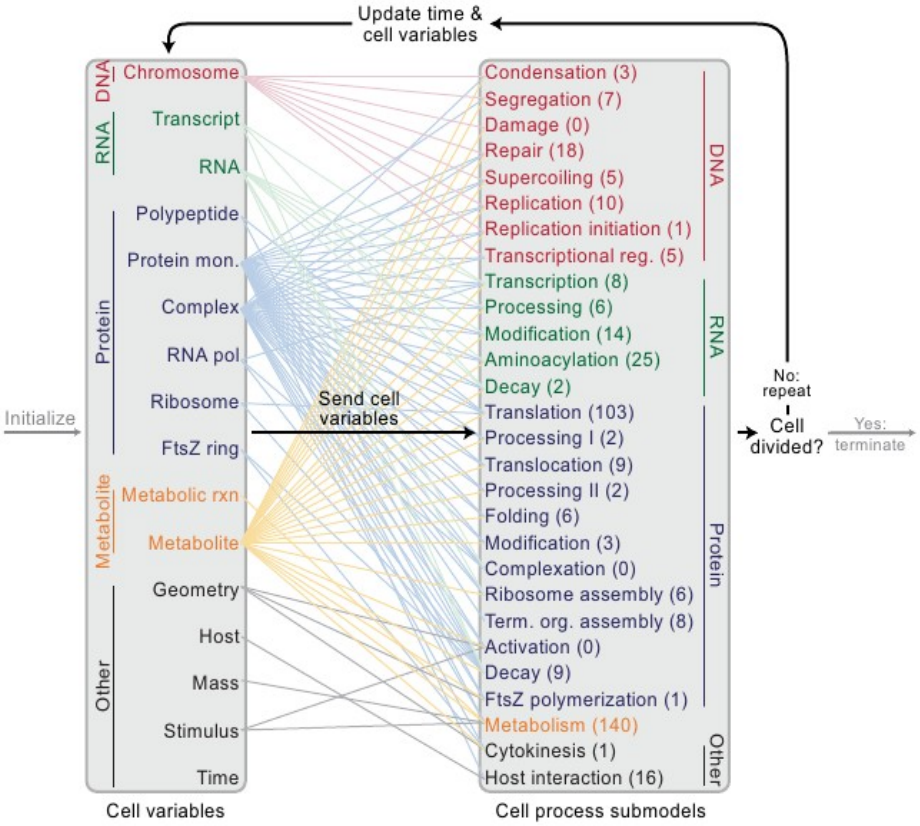
- Different parts can be modelled with different approaches
- Alternative modules representing the same biological process with different granularities or different modelling approaches
- Model families with alternative parts: avoid combinatorial explosion
- Modules can be developed at their own pace;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



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



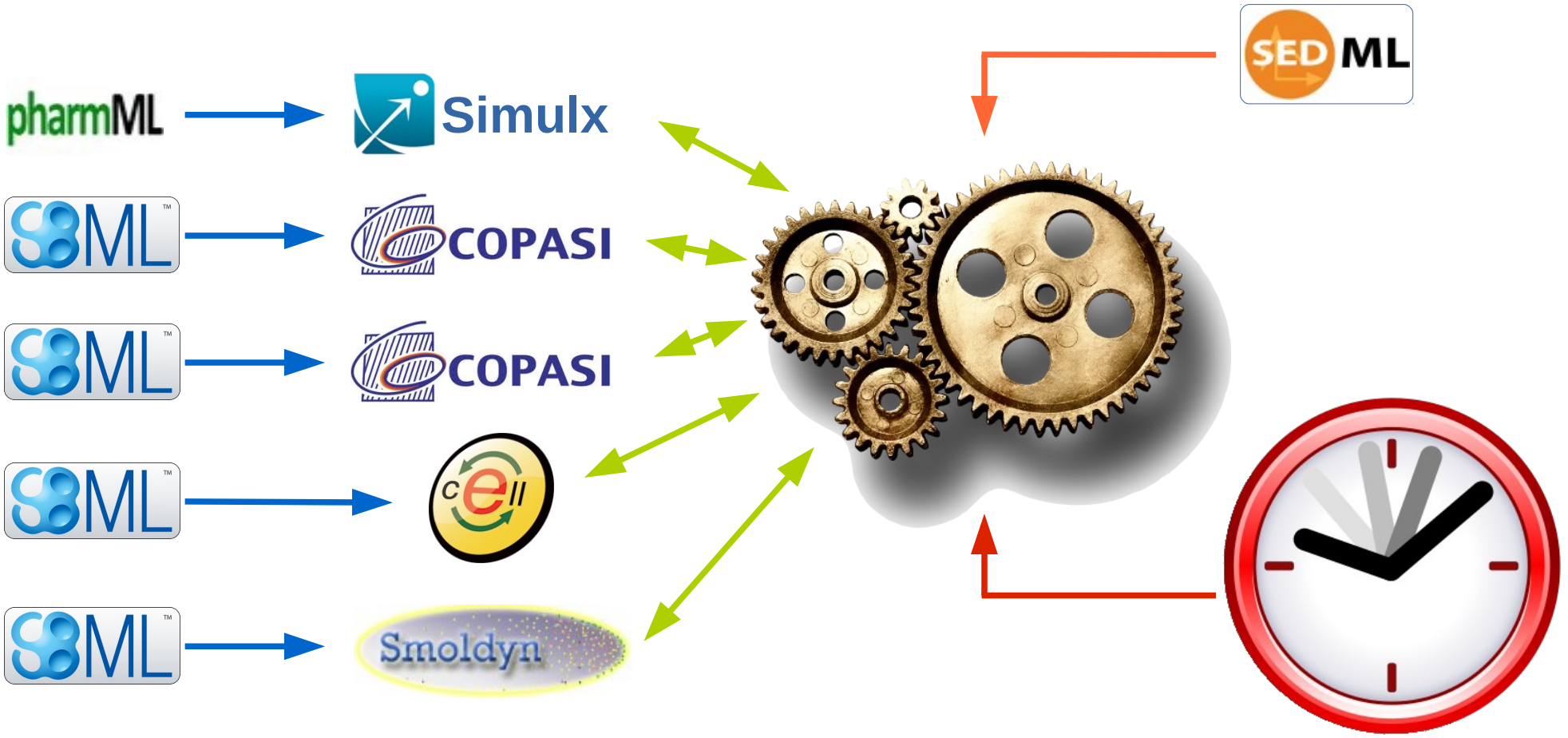
SUMMARY

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

Single software not sufficient



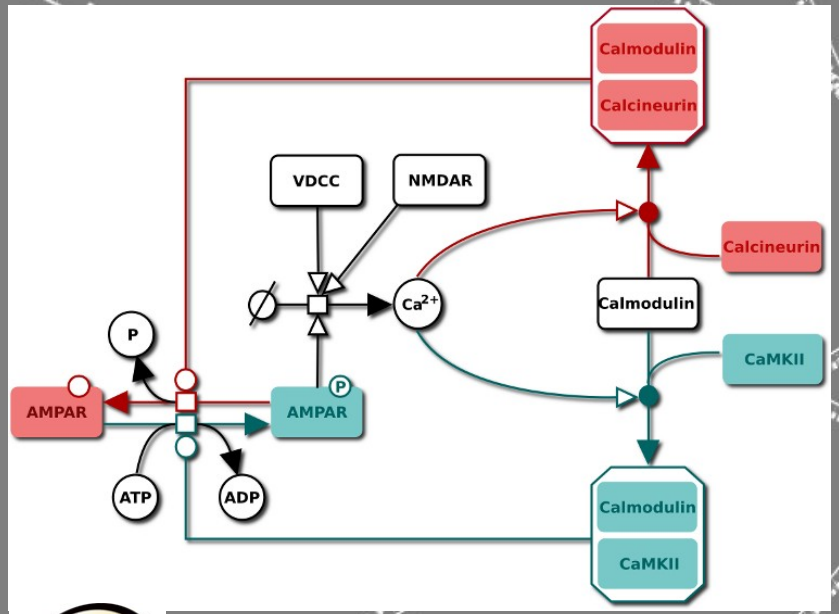
modular model

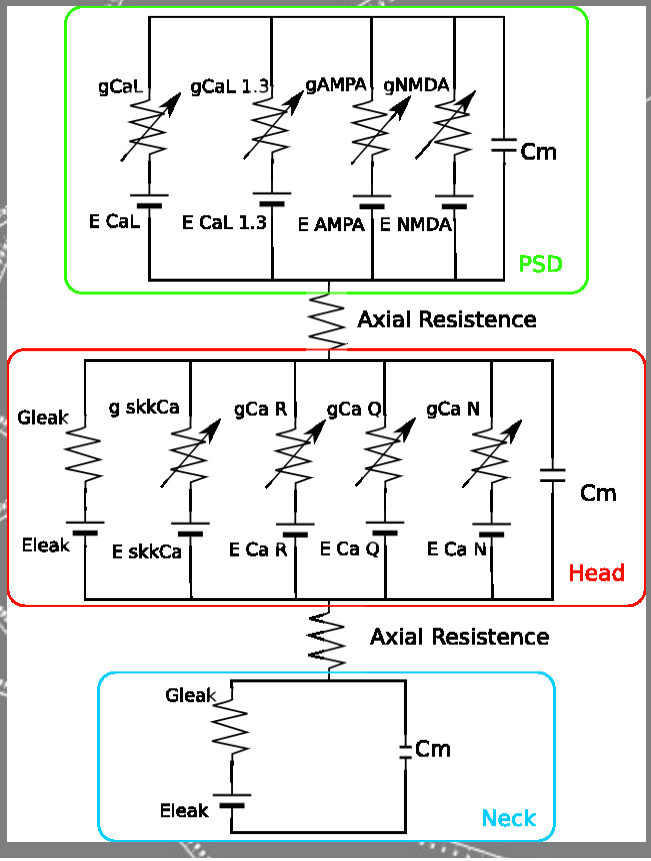
hybrid simulation system

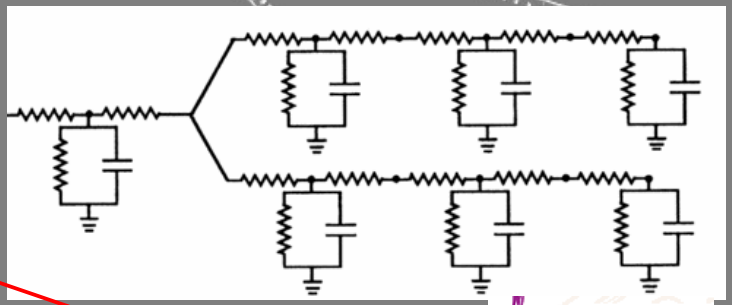
discrete event detection

adaptive synchronization

Whole cell: electro-biochemical models



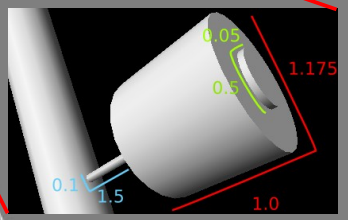




1504 spines

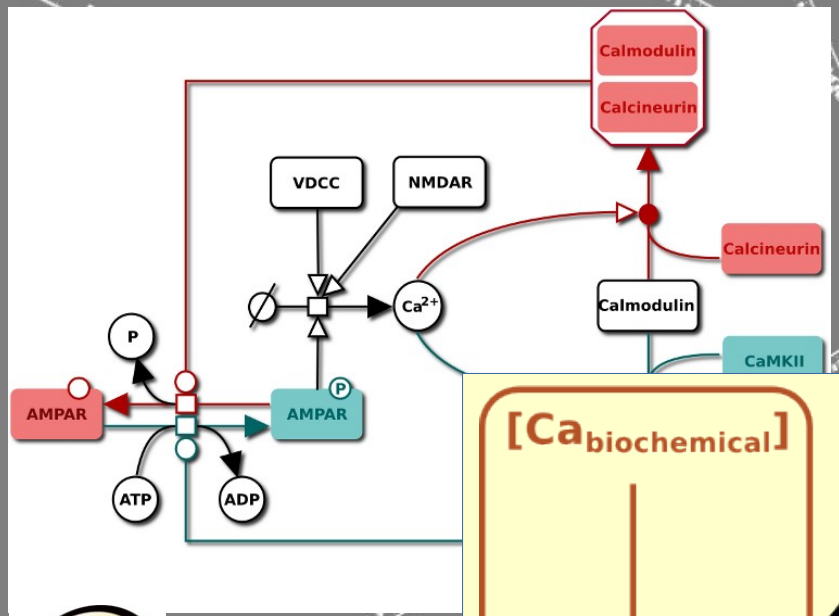
4761 compartments

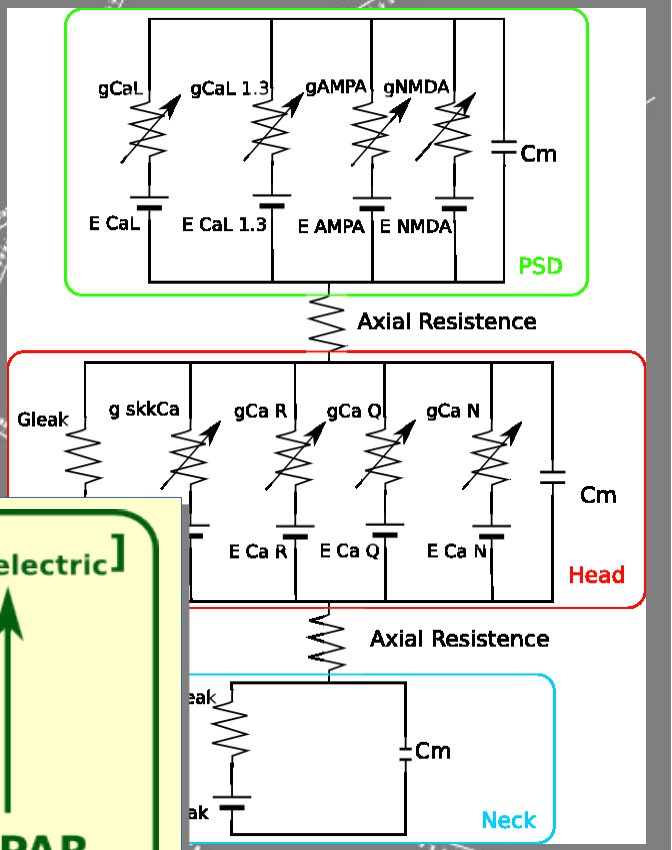
16362 channels

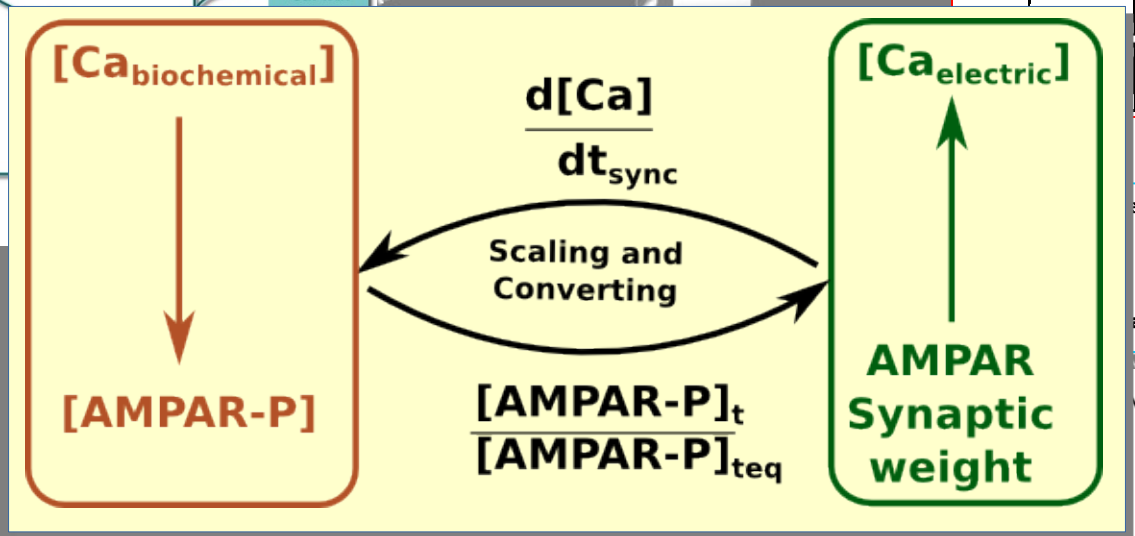


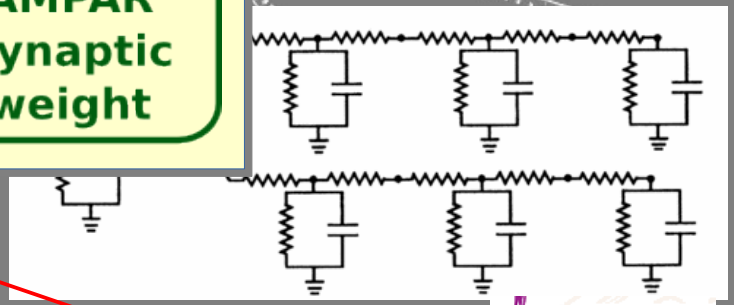
Mattioni et al (2012, 2013)

Whole cell: electro-biochemical models





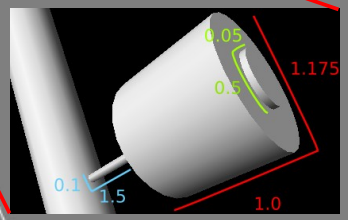




1504 spines

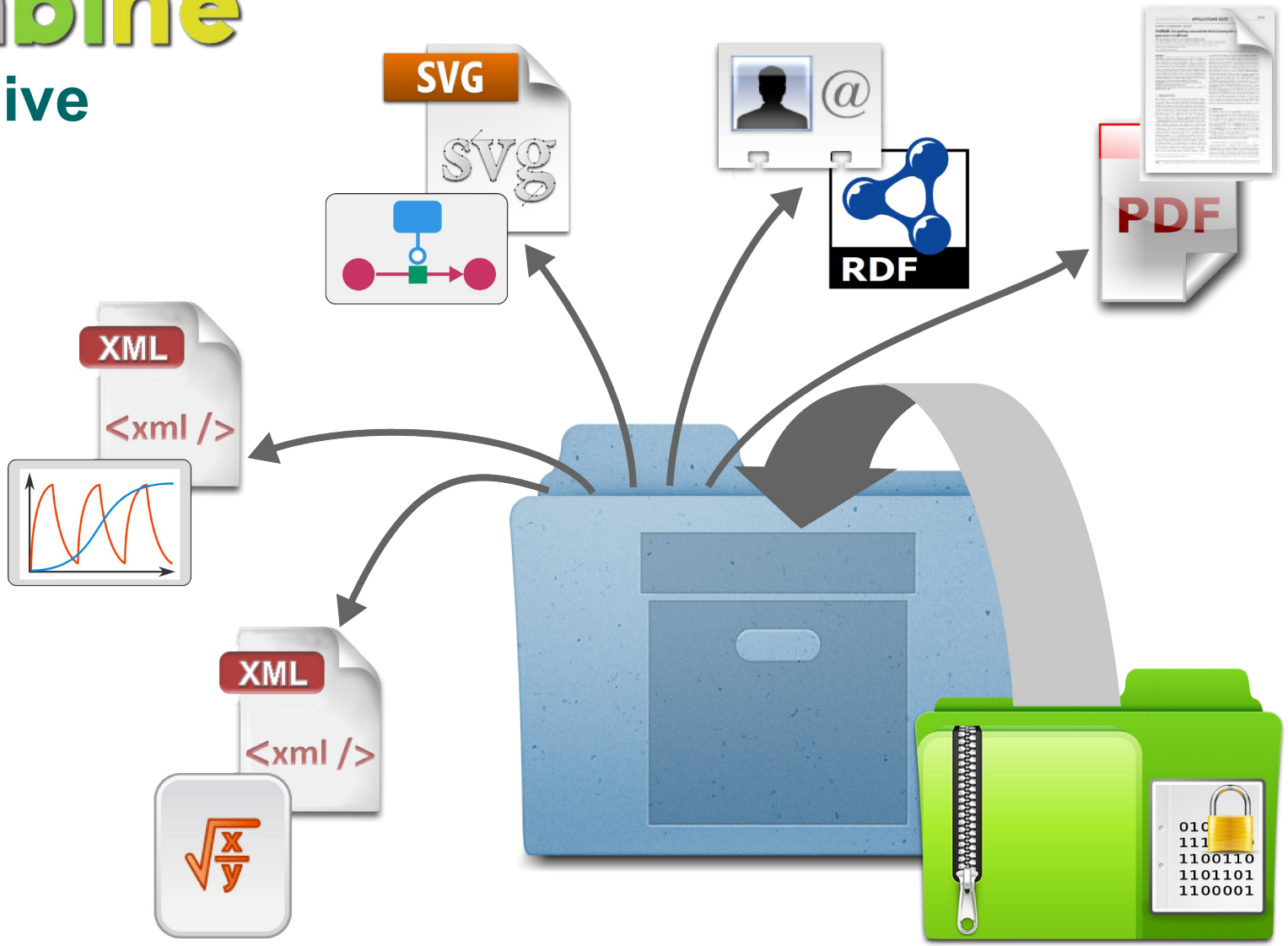
4761 compartments

16362 channels



Mattioni et al (2012, 2013)

combine archive



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

Bergman *et al. In the press*

EMBL-EBI

- **Developers**
  Mélanie Courtot, Alexander Broicher, Finja Büchel, Marco Donizelli, Marine Dumousseau, Gael Jalowicki, Mihai Glont, Ron Henkel, Arnaud Henri, Sarah Keating, Christian Knuepfer, Camille Laibe, Chen Li, Lu Li, Florian Mittag, Kedar Nath Natarajan, Jean-Baptiste Pettit, Nicolas Rodriguez, Karim Tazibt, Martijn van Iersel, Dagmar Waltemath, Sarala Wimalaratne, Yangyang Zhao, Anna Zhukova
- **Curators**
  Ishan Ajmera, Vijayalakshmi Chelliah, Harish Dharuri, Lukas Endler, Enuo He, Nick Juty, Vincent Knight-Schrijver, Audald Lloret, Michael Schubert, Melanie Stefan
- Henning Hermjakob, Janet Thornton

External contributions

- **Collaborators**
  Upi Bhalla, Andreas Dräger, Jürgen Eils, Mike Hucka, Hiroaki Kitano, Ion Moraru, Julio Saez-Rodriguez, Herbert Sauro, Maria Schilstra, Falk Schreiber, Jacky Snoep, Neil Swainston
- **Scientific Advisory Board**
  Upi Bhalla, Carole Goble, Thomas Lemberger, Pedro Mendes, Ion Moraru, Wolfgang Müller, Philippe Sanseau, Herbert Sauro, Jacky Snoep
- All the computational systems Biology community, in particular: Rainer Machne, Bruce Shapiro, Kieran Smallbone

Current member coordinator

Generous funders