

BioModels Database: A Repository of Mathematical Models of Biological Processes

Vijayalakshmi Chelliah, Camille Laibe, and Nicolas Le Novère

Abstract

BioModels Database is a public online resource that allows storing and sharing of published, peer-reviewed quantitative, dynamic models of biological processes. The model components and behaviour are thoroughly checked to correspond the original publication and manually curated to ensure reliability. Furthermore, the model elements are annotated with terms from controlled vocabularies as well as linked to relevant external data resources. This greatly helps in model interpretation and reuse. Models are stored in SBML format, accepted in SBML and CellML formats, and are available for download in various other common formats such as BioPAX, Octave, SciLab, VCML, XPP and PDF, in addition to SBML. The reaction network diagram of the models is also available in several formats. BioModels Database features a search engine, which provides simple and more advanced searches. Features such as online simulation and creation of smaller models (submodels) from the selected model elements of a larger one are provided. BioModels Database can be accessed both *via* a web interface and programmatically *via* web services. New models are available in BioModels Database at regular releases, about every 4 months.

Key words Controlled vocabularies, model interpretation, SBML, CellML, Reaction network diagram

1 Introduction

Mathematical modelling has been used for decades to help scientists understand the mechanisms and dynamics behind the complex biological processes. As models come from different modelers with different perspectives and are encoded in different formats, their reuse is not always straightforward. Reimplementation of models from scientific publications depends on how well the models are documented. Certain missing details can prevent the models from being reused, which consequently lead to a complete loss of the knowledge provided by the papers. The aim of BioModels Database (*see Note 1*) [1, 2] is to overcome these limitations by the following: (1) encoding and simulating the models in standard formats, (2) providing semantic annotation by linking the model elements

to controlled vocabularies and external data resources, and (3) providing a free, centralized, publicly accessible repository for storing, searching, and retrieving curated and annotated computational models. BioModels Database is part of the BioModels.net initiative (*see Note 2*) [3]. The following sections elaborate on its significant features, the structure, and use of the resource.

2 General Information About the Resource

BioModels Database offers a wide range of features and serves as an efficient model sharing platform. The increase in both the number and the complexity of the models since its origin in 2005 (*see Fig. 1*) demonstrates its recognition in the field of computational systems biology.

2.1 Models Provenance

Some models are implemented from articles published in peer-reviewed scientific journals by a team of curators. However, an increasing number of models are directly submitted by the modelers themselves. In the past, some models also came from collaboration with other repositories, such as the former SBML model repository (Caltech, USA), JWS Online (*see Note 3*) [4] the Database of Quantitative Cellular Signaling (DOQCS) (*see Note 4*), and the CellML repository (*see Note 5*) [5].

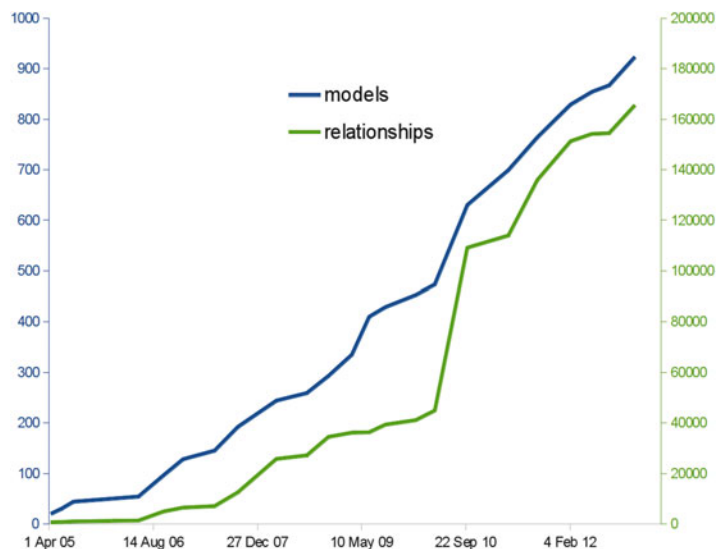


Fig. 1 Growth of BioModels Database: the total number of literature based models (*cyan*) and the total number of relationships (*yellow*) within these models are plotted here. The number of relationships includes SBML “reactions”, “rate rules”, “assignment rules”, “algebraic rules” and “events”. There has been approximately a 20-fold increase in the number of models, since the launch of the resource in 2005, with an average increase in complexity of the models (measured by the number of mathematical relationships) being increased 5 times in the same period.

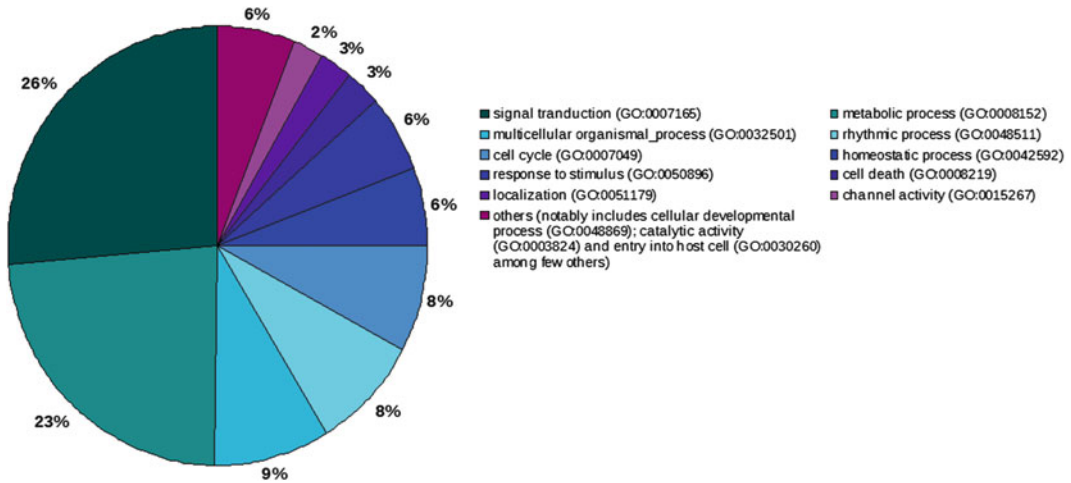


Fig. 2 Type of models: categorization of models in the curated branch of BioModels Database based on the GO terms present in the annotation of the models. This chart was generated by enumerating models in the database, whose annotations refer either to the GO terms listed here or to the children of the GO terms listed here

More than 300 scientific journals recommend submission of models to BioModels Database in their instructions for authors. These include journals from Nature Publishing Group (NPG), the Public Library of Science (PLoS), the Royal Society of Chemistry (RSC), and BioMed Central (BMC). Authors can quote the model identifier (obtained on submitting their model to BioModels Database) in their paper. This allows readers to access the model online and download it once the paper is published.

2.2 Diversity of Models Hosted

BioModels Database covers a wide range of models from several biological categories. It hosts models of simple biochemical reaction systems, larger and complex kinetic models, metabolic network models, and steady-state models. Figure 2 represents the categorization of models in the curated branch using the Gene Ontology (GO) [6] terms present in the model annotation.

2.3 Model Curation Pipeline

Models are not directly visible and retrievable by the public when they are submitted to BioModels Database. From submission to their public release, all models undergo various automated and manually performed curation and annotation procedures to ensure a consistent level of quality and accuracy. This succession of processes is depicted in Fig. 3.

2.4 Model Submission

Submission of models to BioModels Database is free and is open to everyone. Models can be submitted via an online interface and are accepted in two formats, Systems Biology Markup Language

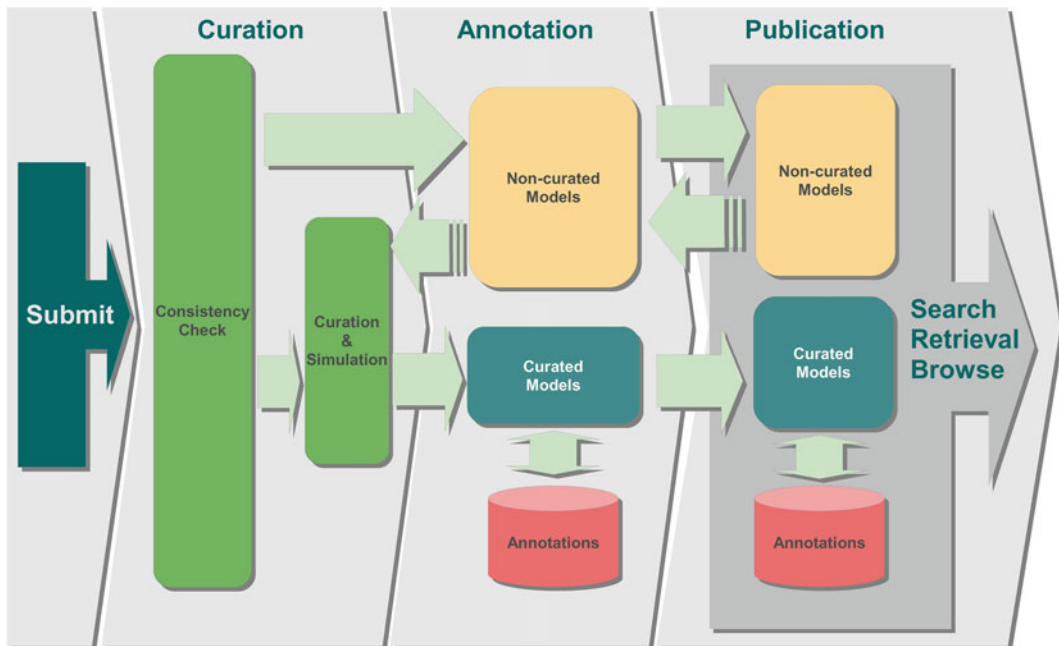


Fig. 3 BioModels Database pipeline: the figure shows the pipeline that handles each model, from its submission to its public release. This illustrates the sequence of steps involved in processing the models and encompasses both public branches of the database (curated and non-curated) as well as the possibility of curating and annotating models already published in the non-curated branch

(SBML) (*see Note 6*) [7] and CellML [5]. The reference of the publication describing the model can be provided (either as a PubMed Identifier (*see Note 7*) or a Digital Object Identifier (*see Note 8*) or an URL) if available at submission time.

While BioModels Database only distributes models that have been described in the peer-reviewed scientific literature, models can be submitted prior to the publication of their associated paper. However, these models are made publicly available only after the publication of the corresponding paper. At the time of submission, each model is assigned a unique and perennial identifier which allows users to access and retrieve it. This identifier can be used by authors as a reference in their publications.

2.5 Curation

The models that are submitted are queued into the curation pipeline and pass through several steps to get published. BioModels Database is composed of two branches: a curated and a non-curated one. Models in both branches are fully SBML compliant. Depending on their curation status, models are moved to one of the two branches.

Models that satisfy the Minimum Information Required in the Annotation of Models (MIRIAM) guidelines [8] progress to the curated branch. These models are thoroughly checked and

corrected for accuracy as they must match and reproduce the results published in the reference publication. A representative figure or table reproduced by the model (which is present in the reference publication) together with a description of how it was obtained is available for each model. To run the simulation experiments under the conditions described in the reference publication, a tool that is different from the one that is used by the authors is used. This is to check and demonstrate that the model is far from software-specific behaviors and to avoid hidden dependencies. The tools most commonly used are COPASI [9, 10], the SBML ODE Solver [11], or the facilities provided by the Systems Biology Workbench [12]. This ensures that the encoded form of the model that is provided corresponds to what was described in the paper.

There are several reasons for models to be in the non-curated branch. These are models that either do not satisfy the full requirements for MIRIAM compliance or have not been curated yet due to limited time and resources. Many of these models are pathway maps, network models, or steady-state models (such as Flux Balance Analysis models [13]), without sufficient quantitative results provided for validation. Others are only the subsets of the whole model described in the chapter, as some parts cannot yet be encoded in SBML. And finally, a small number of models could not be made to reproduce the published results due to untraceable errors in the implementation or typos in the publication (often even after contacting the authors of the article).

3 Annotation

All model elements in the curated branch are furthermore thoroughly annotated with cross-references to other database records and ontology terms. The annotations are included in the models using MIRIAM URIs [14]. As model elements are not always named precisely to relate directly to the corresponding biological processes or physical entity, annotations are necessary to enhance interpretability by both users and software tools. To date, model elements in BioModels Database are annotated using around 40 different external resources. Some of the predominantly used external resources for model annotations are Gene Ontology, ChEBI Ontology, Brenda Tissue Ontology, Systems Biology Ontology, Taxonomy, Reactome, KEGG, and UniProt. A collection of resources and their URIs can be obtained from MIRIAM Registry (*see Note 9*) [15].

3.1 Model Publication

Following the curation and annotation phases, the final stage in the model processing pipeline is the publication of model. Once the curation and annotations are completed, the model is tagged

as ready for publication and becomes available online during the next release of BioModels Database. New releases happen 2–4 times a year.

4 BioModels Database Web Interface

The resource provides several features to allow users to quickly search and locate their models of interest, analyze them, simulate them, extract submodels, or download them in various formats. These facilities are available *via* a web browser or can be accessed from other tools by using the accompanying web services.

4.1 Model Browsing, Searching, and Retrieval

Several features are available through the web interface to facilitate efficient usage of the models. Models can be browsed, independently in the two branches. They can also be located using a pruned Gene Ontology (GO) browser based on the model annotations.

BioModels Database incorporates a powerful search engine that allows users to retrieve models of interest. The search can either be a simple keyword search or a more advanced one. The search engine performs a sequential search by querying metadata, publication information, annotations, SBML file content and supplementary information from external data resources. The system also performs some post-processing of search results in order to deliver the most relevant results.

For example, a search with a taxonomic term traces the whole hierarchy in order to find related models. A search for the term “mammalia” returns models that are not only annotated with *Mammalia* but also with the children and parent taxonomic ranks. Indeed a biological process that happens in *Homo sapiens*, or of *Rattus norvegicus*, occurs in *Mammalia* as well. Similarly, a biological process that happens in *Metazoa* also occurs in *Mammalia* (Fig. 4).

4.2 Model Display

Each model is presented in a tabbed form, providing access to all the information stored by the database about the model. The model elements are hyperlinked between different tabs and the annotations are hyperlinked to their original resource page. The detailed description about the model is separated into six categories, namely, *Model*, *Overview*, *Math*, *Physical entities*, *Parameters*, and *Curation*, which are all accessible via dedicated tabs. This section is highlighted as area “A” in Fig. 5.

- The *Model tab* displays general information about the model such as its reference publication and general annotation to external resources. It also provides access to the submitted (original) version of the model file, as well as information about the encoders, and date and time of the model creation and last modification.

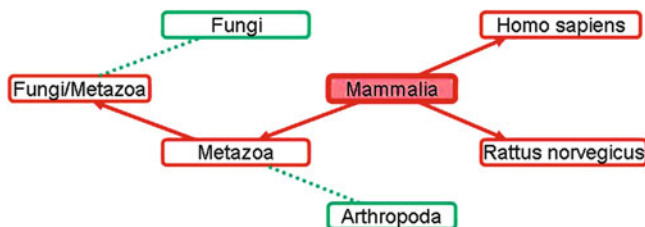


Fig. 4 Taxonomic search: the search engine considers the entire taxonomic tree, when the search is based on a taxonomic term. For example, a search based on the taxonomic term “Mammalia” returns models that are annotated with the term *Mammalia* and related ones, such as *Metazoa*, *Homo sapiens*, or *Rattus norvegicus* (represented as red in this figure)

[Download SBML](#) | [Other formats \(auto-generated\)](#) | [Actions](#) | [Submit Model Comment/Bug](#)

Model	Overview	Math	Physical entities	Parameters	Curation
Reference Publication					
Vinod PK, Freire P, Rattani A, Ciliberto A, Uhlmann F, Novak B. Computational modelling of mitotic exit in budding yeast: the role of separase and Cdc14 endocycles. J R Soc Interface 2011 Aug 8(61):1128-41. Department of Biochemistry, Oxford Centre for Integrative Systems Biology, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [more]					
Model					
Original Model:	BIOMD0000000370.xml.orgin	qbqbiol.ocoursIn	Taxonomy Saccharomyces cerevisiae		
Submitter:	Bela Novak	set #1 qbqbiol.isVersionOf	Cell Cycle Ontology CCO_P0000038		
Submission ID:	MODEL1111030000	qbmodel.isDerivedFrom	BioModels Database BIOMD0000000409		
Submission Date:	03 Nov 2011 16:25:35 UTC				
Last Modification Date:	02 Feb 2012 11:26:22 UTC				
Creation Date:	10 Nov 2011 16:53:58 UTC				
Encoders:	Vijayalakshmi Chelliah Bela Novak				
Notes					
This model is from the article: Computational modelling of mitotic exit in budding yeast: the role of separase and Cdc14 endocycles Vinod PK, Freire P, Rattani A, Ciliberto A, Uhlmann F, Novak B. J R Soc Interface 2011 Aug 7;8(61):1128-41. Epub 2011 Feb 2. 21288956 . Abstract: The operating principles of complex regulatory networks are best understood with the help of mathematical modelling rather than by intuitive reasoning. Hereby, we study the dynamics of the mitotic exit (ME) control system in budding yeast by further developing the Queraal's model. A comprehensive systems view of the network regulating ME is provided based on classical experiments in the literature. In this picture, Cdc20-APC is a critical node controlling both cyclin (Clb2 and Clb5) and phosphatase (Cdc14) branches of the regulatory network. On the basis of experimental situations ranging from single to quadruple mutants, the kinetic parameters of the network are estimated. Numerical analysis of the model quantifies the dependence of ME control on the proteolytic and non-proteolytic functions of separase. We show that the requirement of the non-proteolytic function of separase for ME depends on cyclin-dependent kinase activity. The model is also used for the systematic analysis of the recently discovered Cdc14 endocycles. The significance of Cdc14 endocycles in eukaryotic cell cycle control is discussed as well. This model originates from BioModels Database: A Database of Annotated Published Models (http://www.ebi.ac.uk/biomodels/). It is copyright (c) 2005-2012 The BioModels.net Team. For more information see the terms of use . To cite BioModels Database, please use: L J C, Donzelli M, Rodriguez N, Dharni H, Enfinger L, Chelliah V, Li L, He F, Henry A, Stefan M, Snoep JL, Hucka M, Le Novère N, Laibe C (2010) BioModels Database: An enhanced, correlated and annotated resource for published quantitative kinetic models. BMC Syst Biol. 4:92.					

[Computational Systems Neurobiology Group](#) · [European Bioinformatics Institute](#) · [Terms of Use](#) · [Contact Us](#) · Developed by the [BioModels.net Team](#)

Fig. 5 Model display: this screen image shows the display of an example model in order to illustrate how the model information is displayed on the web interface

- The *Overview* tab provides access to all model components, namely, the mathematical expressions, physical entities, parameters, rules and other elements that constitute the model. Each model element is provided with a hyperlink, which leads to further details displayed on a different tab. Smaller models (submodels) can also be created by selecting a few model elements and followed by using the “Create a submodel with selected elements” command. This generates a valid SBML model by extracting all the selected elements and the additional elements required to have a valid model. The resulting model is

displayed in the new *submodel tab* and can be downloaded in SBML.

- The *Math tab* lists all the mathematical constructs, which include reactions, rules and events. Each mathematical expression is associated with a rendering of the mathematical equation and hyperlinked annotations.
- The *Physical entity tab* lists all the entity pools and compartments included in the model, along with their initial quantities and annotations.
- The *Parameter tab* lists all the parameter used in the mathematical expressions. Parameters whose values are determined by mathematical expressions are linked to the associated model element of the *Math tab*.
- The *Curation tab* displays a representative curation result (in the form of graphical plots or tables), that matches with that of the reference publication, and which was reproduced by simulating the model.

4.3 Model Exports

This section is highlighted as area “B” in Fig. 5. The menu *Download SBML* allows users to download the model in various versions of SBML [7]. The version of the model that was checked by the curators and used to produce the curation figures is indicated as “curated.” The other SBML versions are automatically generated and provided for convenience, as certain tools support only some specific levels or versions of SBML.

The menu *Other formats (auto-generated)* provides access to other model representation formats, such as BioPAX (*see Note 10*) (level 2 and level 3) [16] and the Virtual Cell Markup Language (VCML) [17]. BioModels Database also provides downloadable configuration files for open tools such as XPPAUT [18], SciLab (*see Note 11*), and Octave (m-file) (*see Note 12*). In addition, a human-readable summary of the models in the Portable Document Format (PDF), generated using SBML2LaTeX tool [19], is also available from this menu.

The menu *Actions* grants access to the graphical representations of the model’s reaction networks, following the Systems Biology Graphical Notation (SBGN) [20] and available in PNG (Portable Network Graphics (*see Note 13*)) and SVG (Scalable Vector Graphics (*see Note 14*)) formats. The reaction networks are also provided in a dynamic way via an interactive Java applet.

The menu *Actions* also provide access to the online simulation tools. BioModels Database embeds SOSlib [11] to provide a basic online simulation tool. The simulation results are returned both in graphical and textual form. For many models, an additional and more flexible simulation tool is available using JWS Online [4]. Some models have an additional feature called *Model of the Month* (*see Note 15*) which is a brief article that discusses the biological background, significance, structure and results of the model.

The *Models of the Month* are also accessible through BMC Systems Biology Gateway (*see Note 16*).

4.4 Web Services

BioModels Database features programmatic access *via* web services [21] (*see Note 17*). They allow, for example, direct retrieval of complex searches for models and the creation of submodels. The available services are described in a Web Services Description Language (WSDL) (*see Note 18*) file that enables software to understand available functions and their usage. The web services use the Simple Object Access Protocol (SOAP) (*see Note 19*) to encode requests and responses. The complete list of available methods, as well as a Java library and the associated documentation, is provided on the BioModels Database website.

BioModels Database is developed under the GNU General Public License, and the software is freely available from its SourceForge repository (*see Note 20*).

4.5 Conclusions

BioModels Database has significantly grown both in size and number of the models, since its origin in 2005 and serves as an efficient model sharing platform. The submission of models by modelers/authors themselves is increasing rapidly, and this demonstrates the popularity and recognition of BioModels Database in the community. The resource helps modelers to reuse already existing models or model components, to modify them by implementing their own theory, to publish articles describing new models, and to submit those new models to BioModels Database. For example, BIOMD0000000176 and BIOMD0000000177 are derived from BIOMD0000000172 which in turn is derived from BIOMD0000000064. BioModels Database is also used as a source of trusted models for benchmarking model simulation software packages.

BioModels Database has the potential to serve as a comprehensive repository for computational systems biology models, similar to the functionality of GenBank and Protein Data Bank (PDB), the data resources for genes and protein 3D structures.

5 Notes

1. <http://www.ebi.ac.uk/biomodels>
2. <http://biomodels.net>
3. <http://jjj.biochem.sun.ac.za/>
4. <http://doqcs.ncbs.res.in/>
5. <http://models.cellml.org/cellml>
6. <http://sbml.org/>
7. <http://www.pubmed.gov/>

8. <http://www.doi.org>
9. <http://www.ebi.ac.uk/miriam/>
10. <http://www.biopax.org/>
11. <http://www.scilab.org/>
12. <http://www.gnu.org/software/octave/>
13. <http://tools.ietf.org/html/rfc2083>
14. <http://www.w3.org/Graphics/SVG/>
15. <http://www.ebi.ac.uk/biomodels-main/modelmonth>
16. <http://www.biomedcentral.com/gateways/systemsbiology>
17. <http://www.ebi.ac.uk/biomodels-main/webservices>
18. <http://www.w3.org/TR/wsdl/>
19. <http://www.w3.org/TR/soap/>
20. <http://sourceforge.net/projects/biomodels/>

References

1. Le Novère N, Bornstein B, Broicher A (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* 34:D689–D691
2. Li C, Donizelli M, Rodriguez N et al (2010) BioModels Database: an enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Syst Biol* 4:92
3. Le Novère N (2006) Model storage, exchange and integration. *BMC Neurosci* 7(Suppl 1):S11
4. Snoep JL, Olivier BG (2002) Java Web Simulation (JWS); a web based database of kinetic models. *Mol Biol Rep* 29:259–263
5. Lloyd CM, Halstead MDB, Nielsen PF (2004) CellML: its future, present and past. *Prog Biophys Mol Biol* 85:433–450
6. Ashburner M, Ball CA, Blake JA et al (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25:25–29
7. Hucka M, Finney A, Sauro HM (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19:524–531
8. Le Novère N, Finney A, Hucka et al (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 23:1509–1515
9. Hoops S, Sahle S, Gauges (2006) COPASI—a COmplex PAtHway SIMulator. *Bioinformatics* 22:3067–3074
10. Mendes P, Hoops S, Sahle S (2009) Computational modeling of biochemical networks using COPASI. *Methods Mol Biol* 500(17–59): 2009
11. Machné R, Finney A, Müller S (2006) The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* 22(1406–1407): 2006
12. Bergmann FT, Sauro HM (2006) SBW—a modular framework for systems biology. pp. 37–1645.
13. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28:245–248
14. Juty N, Le Novère N, Laibe C (2012) Identifiers.org and MIRIAM Registry: community resources to provide persistent identification. *Nucleic Acids Res* 40:D580–D586
15. Laibe C, Novère NL (2007) MIRIAM Resources: tools to generate and resolve robust cross-references in systems biology. *BMC Syst Biol* 1:58
16. Demir E, Cary MP, Paley S et al (2010) The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 28:935–942
17. Moraru II, Schaff JC, Slepchenko BM (2008) Virtual cell modelling and simulation software environment. *IET Syst Biol* 2:352–362
18. Ermentrout, B. (2002) Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students

19. Dräger A, Planatscher H, Moutsou W et al (2009) SBML2L(A)T(E)X: conversion of SBML files into human-readable reports. *Bioinformatics* 25:1455–1456
20. Le Novère N, Hucka M, Mi H, Moodie S et al (2009) The systems biology graphical notation. *Nat Biotechnol* 27:735–741
21. Li C, Courtot M, Le Novère N, Laibe C (2010) BioModels.net web services, a free and integrated toolkit for computational modelling software. *Brief Bioinform* 11: 270–277