

---

## Designing and encoding models for synthetic biology

Lukas Endler, Nicolas Rodriguez, Nick Juty, Vijayalakshmi Chelliah, Camille Laibe, Chen Li and Nicolas Le Novère

*J. R. Soc. Interface* published online 1 April 2009  
doi: 10.1098/rsif.2009.0035.focus

---

### Supplementary data

["Data Supplement"](#)

<http://rsif.royalsocietypublishing.org/content/suppl/2009/04/01/rsif.2009.0035.focus.DC1.html>

### References

[This article cites 88 articles, 35 of which can be accessed free](#)

<http://rsif.royalsocietypublishing.org/content/early/2009/03/27/rsif.2009.0035.focus.full.html#ref-list-1>

### P<P

Published online 1 April 2009 in advance of the print journal.

### Subject collections

Articles on similar topics can be found in the following collections

[synthetic biology](#) (13 articles)

### Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right-hand corner of the article or click [here](#)

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *J. R. Soc. Interface* go to: <http://rsif.royalsocietypublishing.org/subscriptions>

---

## REVIEW

# Designing and encoding models for synthetic biology

Lukas Endler\*, Nicolas Rodriguez, Nick Juty, Vijayalakshmi Chelliah,  
 Camille Laibe, Chen Li and Nicolas Le Novère

*EMBL—European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton,  
 Cambridge CB10 1SD, UK*

A key component of any synthetic biology effort is the use of quantitative models. These models and their corresponding simulations allow optimization of a system design, as well as guiding their subsequent analysis. Once a domain mostly reserved for experts, dynamical modelling of gene regulatory and reaction networks has been an area of growth over the last decade. There has been a concomitant increase in the number of software tools and standards, thereby facilitating model exchange and reuse. We give here an overview of the model creation and analysis processes as well as some software tools in common use. Using markup language to encode the model and associated annotation, we describe the mining of components, their integration in relational models, formularization and parametrization. Evaluation of simulation results and validation of the model close the systems biology ‘loop’.

**Keywords:** synthetic biology; systems biology markup language; computational biology

## 1. INTRODUCTION

Synthetic biology has seen an explosive development over the last decade (Sprinzak & Elowitz 2005; Andrianantoandro *et al.* 2006). The engineering of biological systems has been boosted by the availability and improvement of wet-lab techniques. Hand in hand with the development of these experimental approaches, the use of computational tools and methods to model and interpret biological systems has spread and become the mainstream. While at the end of last century, only a few software tools were specifically designed to build and interpret gene regulatory and biochemical reaction networks, today they number in the hundreds.

Biological systems are highly interconnected, and their behaviours most often cannot be derived solely from the properties of single components without the help of computers. While the dynamics of elementary interactions and simple regulatory motifs can sometimes be inferred without the help of mathematical models, the behaviour of more complex reaction networks cannot be unravelled merely by reasoning alone. Therefore, mathematical descriptions of biochemical processes and

computational analysis, in combination with experimental results, have become a necessity.

Computer simulations allow the prediction of system behaviour and can help to elucidate the mechanisms underlying biological phenomena. An initial model should be created which is reasonably faithful to the biological structure, and is capable of reproducing known behaviours to an acceptable extent. This can then be used to probe the effects of different environmental conditions, perturbations and variants or different designs. For instance, it is possible to use *in silico* experiments to screen for possible drug targets (Undrovinas *et al.* 2006), to optimize certain aspects of a system (e.g. Reijenga *et al.* 2001), to simulate the behaviour of mutants (e.g. Bray *et al.* 1993; Novak & Tyson 1997) and to analyse the evolvability and robustness of certain systems (von Dassow *et al.* 2000). Analyses at this speed and low cost could never be achieved solely using classical experimental methods. Finally, and maybe equally important, models can show us gaps in our understanding of biological processes and suggest which piece of the puzzle or experimental data is missing. The modelling process itself allows one to scrutinize not only the available data, but also the known or assumed mechanisms. This often leads to more questions than answers, and sometimes shows that mechanisms believed to be established do not actually fit into observed phenomena or are insufficient to explain experimental results.

\*Author for correspondence (lukas@ebi.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rsif.2009.0035.focus> or via <http://rsif.royalsocietypublishing.org>.

One contribution to a Theme Supplement ‘Synthetic biology: history, challenges and prospects’.

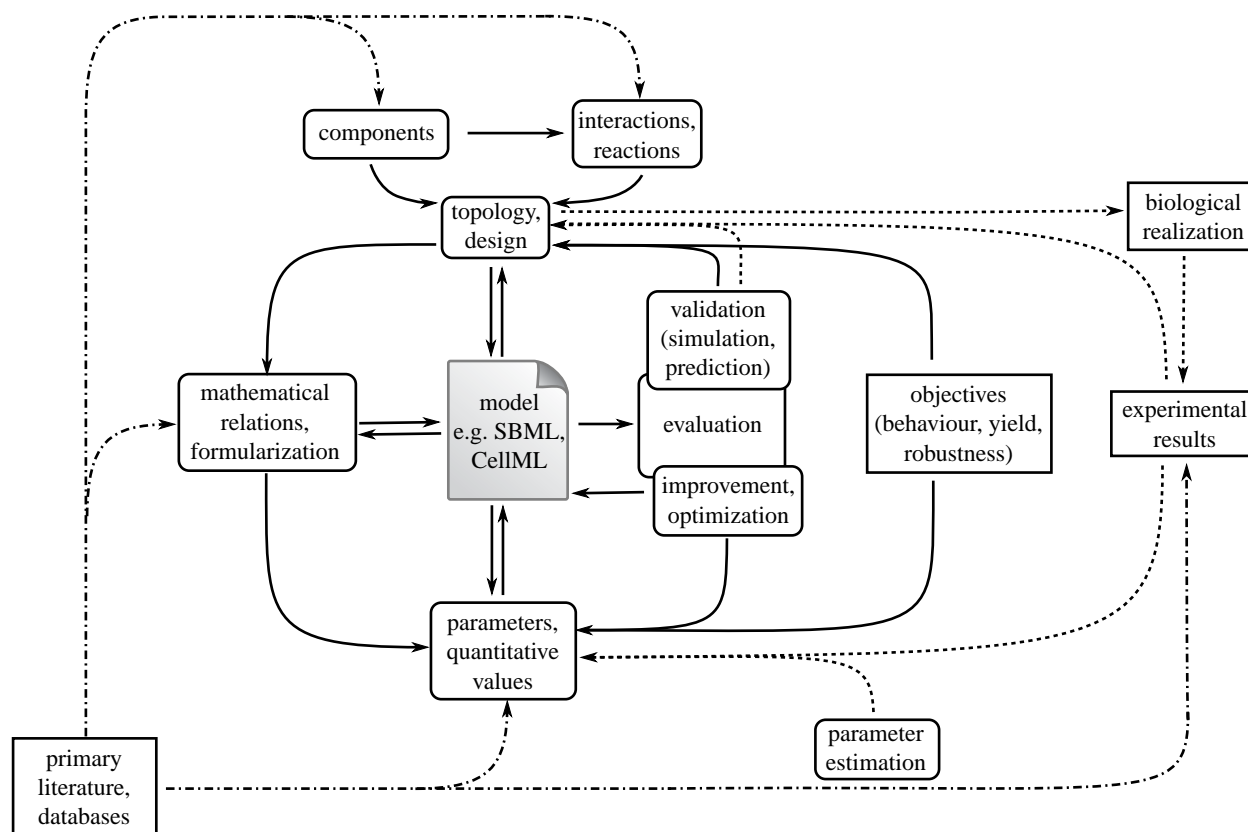


Figure 1. Diagram describing the model creation process, centred around the model file. External inputs come from literature, databases or experiments. Validation and parameter estimation give direct feedbacks on the model, while the predictions can also guide experimental design. The desired objectives for the system to be designed are part of a mostly *in silico* design cycle, the biological implementation and the fitting and comparison of the model to observations are part of the standard systems biology model development cycle.

Modelling has been an integral part of the synthetic biology work since its inception (Elowitz & Leibler 2000; Gardner *et al.* 2000). Synthetic biology, indeed, does not differ from any other engineering activity, where designing and testing of models are as important as the building step. This not only allows economy of time and effort, but also permits optimization of the final product. However, modelling in biology is most often an iterative process; the mathematical model has to be validated by comparison of the model analyses and simulation results with experimental measurements, and to be refined depending on these comparisons. This is especially true for kinetic modelling since a description of the system's final state is required together with its changes over a temporal range. Figure 1 illustrates the model creation process.

Driven by the needs of the research community, the availability of more usable software tools and the existence of community standards, the user population has shifted from dedicated experts in mathematical modelling to scientists with different backgrounds and interests. Computational models have now become another tool for use in parallel to various experimental approaches. Modelling, though, still remains a highly interdisciplinary task. First, it demands a detailed understanding of the biological and biochemical processes to be modelled. This biological knowledge alone is normally not enough, since encoding it into a computer readable form requires some mathematical

knowledge and, depending on the tools used, also needs some computer skills. Model manipulation, evaluation and prediction generation, on the other hand, require both computational and biological experience. So while more and more intuitively usable software tools emerge, modelling is still a collaborative effort of experts in different fields or the domain of scientists with a broader overview and training.

## 2. FORMATS FOR MODEL REPRESENTATION

Biochemical or genetic regulatory networks can be represented in various forms and formats, ranging from a crude pathway or an interaction map on a napkin to a system of ordinary differential equations in a computer algebra system (plain text file), or to machine executable code producing integrated time courses of a specific system. Biological reaction networks have many different logical layers. The first layer comprises the interacting molecules, as well as all the state variables. The interactions between the components and the underlying reaction network define the topology of the system and form the next logical layer. The final layer comprises the mathematical expressions for the interactions, and the fluxes in and out of the system.

Synthetic biology activities require the incorporation of as many of these layers as possible in the model. Static network analysis or creation of deletion mutants

of a system requires knowledge of the stoichiometry of reactions. This cannot always be readily derived from, for example, ordinary differential equation representations without additional information. Furthermore, in some cases, a system has to be interpreted using different modelling frameworks. For example, in genetic regulatory networks, a continuous deterministic framework might be useful for bifurcation analysis of the qualitative behaviour, while a probabilistic discrete approach could be used to explore robustness and behaviour at low concentrations of some species (Elowitz & Leibler 2000).

As dedicated tools, targeted at different tasks, often use distinct entry formats, their use requires conversion or reimplementing of the original models into compatible formats. Manual conversion of a model from one format to another can be a very tedious and error-prone task, especially for bigger systems. The development of exchange formats for mathematical models facilitated the use of different programs and tools and facilitated their reuse and alteration. Two extensible markup language (XML) formats, the systems biology markup language (SBML) (Hucka *et al.* 2003) and cellular markup language (CellML; Lloyd *et al.* 2004), have gained broad support in the modelling community. Both languages store mathematical expressions using content MathML v. 2.0 (Ausbrooks *et al.* 2003) and use the resource description framework (RDF; Beckett 2004) to incorporate machine-readable metadata.

SBML has been developed as a community effort and consists of hierarchical lists of elements describing the compartments of the model, the pools of reacting entities and the interactions taking place. All elements can be annotated and commented in both human-readable and machine-interpretable forms, allowing extensive documentation and external references to be directly included in the model file. SBML has been widely adopted by the modelling community and is supported by hundreds of tools (see [http://sbml.org/SBML\\_Software\\_Guide](http://sbml.org/SBML_Software_Guide)). A key to the success and rapid adoption of SBML by developers has been the availability of a general application programming interface (API) library, LIBSBML (Bornstein *et al.* 2008). This allows easy access to all elements of a model and provides unit and semantic checking of the SBML file. Tools can therefore use their own internal representation of a mathematical model, and read and write the actual model file over the standardized interface. As LIBSBML also features bindings to common scripting languages, a computer-literate modeller can use the API for automatic batch manipulation or simple analysis of SBML files.

Here follows a brief description of the main SBML elements:

- *Compartment*. A well-stirred container in which species reside and reactions take place. This may be, for example, a two-dimensional membrane or a cell extending in three dimensions.
- *Species*. Entities of a certain type residing in a specific compartment. For example, ATP in the compartment mitochondrion would constitute a different species from ATP in the cytosol.
- *Parameter*. This includes all other kinds of named quantities. They can be constant or variable and altered by rules and events. For example, the proton-motive force or a chemical potential could be defined as a parameter in a model.
- *Reaction*. A transformation between one or more species with specified stoichiometries. This encompasses not only chemical reactions, but also transport between compartments, in- and out-flux and so on. It can contain an expression for a rate law.
- *Rule*. Mathematical expressions to define relations between species and parameters. They come in three different types: assignment rules directly assign the value of an expression to a variable, rate rules to its rate of change with time and algebraic rules define general algebraic equalities between different variables.
- *Function*. Definitions of reusable mathematical functions.
- *Event*. A discontinuous change taking place when a trigger condition changes from false to true. The change can be in all kinds of variables and apply to more than one.
- *Note*. Extensive remarks in extensible hypertext markup language (XHTML), meant to be of human readable form.
- *Annotation*. Software-specific metadata and controlled, machine-readable characterizations in an RDF format.

The other widely used XML format, CellML, provides a more modular structure. A model description consists of components and lists of connections between them. This leads to a greater flexibility, especially for multi-scale models, and allows for easier reuse of components at the expense of biochemical semantics. The language is mostly developed by the Physiome Project and Bioengineering Institute at the University of Auckland. To date, there have been fewer tools supporting CellML (<http://www.cellml.org/tools>) than SBML, but tools exist to interconvert between both formats (see <http://www.ebi.ac.uk/compneur-srv/sbml/convertors/SBMLConvertors.html>; Schilstra *et al.* 2006), offering users the possibility to choose from a broad range of evaluation software programs.

An important, though often neglected, part of mathematical models is descriptions and characterizations of its components. It can be very valuable to include references and comments directly in the model file, especially for bigger models. Here, entities and relations can become hard to assign to their corresponding biological species and processes, and the original sources can be hard to keep track of. Therefore, extensive documentation of the model, and the references for its components, should always be kept with the model file, especially if more than one person is involved in the modelling process. Traditionally, a model is described by an accompanying text physically separated from the file containing it. While it is essential to have a verbose description of the model, direct annotation of the elements is very useful for exchange and reuse of models, as well as during the creation process itself. A reporting standard for

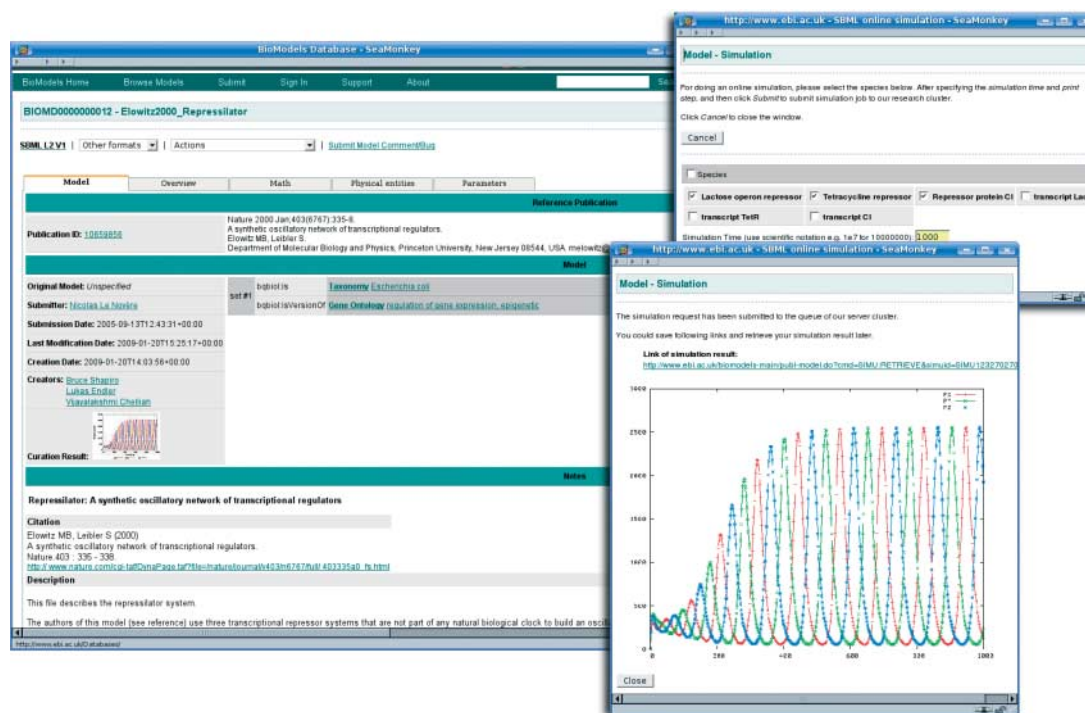


Figure 2. BioModels database is a resource offering curated and annotated versions of models published in peer-reviewed literature. The models can be browsed, downloaded or directly simulated online, as shown here for the repressilator.

the curation and validation of biological models—MIRIAM or minimal information requested in the annotation of biochemical models (Le Novère *et al.* 2005)—not only gives rules an encoded model to comply with, but also defines a form for controlled annotations. This scheme for annotations consists of a stable identifier in the form of a uniform resource identifier (URI; Berners-Lee *et al.* 2005) for each piece of information and a qualifier indicating the relation of this information to the annotated element. A catalogue of resources (such as controlled vocabularies or databases) is provided in MIRIAM resources (Laibe & Le Novère 2007), as well as a set of online services, which include creation and resolution of model annotations in URI form. The ability to include not only the mathematical description, but also reference and biological information in the same file simplifies exchange and reuse of models and eases collaborative work of groups of people on one model. Using a formal modelling language such as SBML, the process of modelling can easily become a collaborative effort, with different people using different tools to insert key players and reactions, annotations, kinetic rate laws, mathematical expressions and parameter values into a common model file.

### 3. GATHERING BITS AND PIECES

Prior to the creation of a detailed kinetic model, one needs to list the main elements. These are the key interacting species and other observables of the system one wants to model (and synthesize afterwards). In addition, one must gather their known regulatory interactions and chemical reactions from the scientific literature and databases. The kind of databases to mine

depends very much on the system to be modelled. Apart from the general list given in the yearly nucleic acid research database issue (Galperin & Cochrane 2009), an overview of resources more specific to modelling is given in Ng *et al.* (2006) and Wierling *et al.* (2007). For biological pathways and reactions, the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa *et al.* 2008), Reactome (Vastrik *et al.* 2007; Matthews *et al.* 2008), Panther (Mi *et al.* 2007) and the Meta- and BioCyc (Caspi *et al.* 2008) databases are especially useful, as they either directly export in generally used computer-readable formats (such as SBML) or their outputs can be converted to them.

Complementary to searching through databases, text mining tools can help to confine the number of articles to screen and also reveal valuable information on interactions or relationships between components. Among the tools available, iHOP (<http://www.ihop-net.org/>; Hoffmann & Valencia 2004) and Chilibot (<http://www.chilibot.net>; Chen & Sharp 2004) have been valuable for us in identifying potential protein and gene-regulatory interactions. A general overview of text mining and some of the available tools is given in Krallinger *et al.* (2008) and Ananiadou *et al.* (2006), for example.

Another possibility is to start with existing models that one can access through databases and repositories such as the BioModel database (Le Novère *et al.* 2006; figure 2), Java web simulation (JWS) online (Snoep & Olivier 2002) or the CellML model repository (Lloyd *et al.* 2008). For more specific models, although still quite remote from synthetic biology projects, one can use dedicated repositories such as the database of Quantitative Cellular Signalling (Sivakumaran *et al.* 2003) and ModelDB (Hines *et al.* 2004). The models

stored in these databases not only give an idea of the interactions and species one may need to model a given biological process, but also provide an overview of the mathematical relations used by other researchers. The models, or parts of them, can then be used as starting points or modules for new modelling efforts. Some of these databases, such as Biomodels database, directly allow creation and retrieval of complete submodels based on user-selected components. Another possibility for creating models from existing ones is to use an online modelling environment such as WEBCELL (Lee *et al.* 2006). In addition to offering multiple analysis methods, WEBCELL also allows import and modification of models from Biomodels database and JWS online. An interesting approach is adopted by SYCAMORE (Weidemann *et al.* 2008), which merges a database of quantitative enzyme kinetics—SABIO-RK (Wittig *et al.* 2006)—with the ability to create models via a web front end. These can subsequently be downloaded in the SBML format.

To keep track of the resources and references used for model components, as mentioned above, they can be stored in a separate text file or as metadata in the XML file directly. SBML offers two possibilities for the latter: (i) the notes and (ii) annotation elements. Both provide further information regarding model elements in human-readable or a machine-friendly format, respectively. Notes are encoded in XHTML, providing flexibility and human readability. Annotations, on the other hand, contain machine-readable metadata in various XML formats, which may be software-specific information or standard pointers, such as MIRIAM annotation in RDF. Many SBML-supporting tools, such as SBML editor (Rodriguez *et al.* 2007), COMPLEX PATHWAY SIMULATOR (COPASI, Hoops *et al.* 2006) and CELLDESIGNER (Funahashi *et al.* 2003), allow notes or annotations to be kept.

#### 4. MODEL CREATION

Having identified the key components of the system, reacting entities and topology, one needs to design the biochemical structure of the model. While the former steps have dealt with biology, the current one deals with chemistry. This step is still independent of a specific mathematical framework or formulation used. While stoichiometry and qualitative influence of the species can be sketched at this stage, no mathematical expressions and quantitative information are necessary for creating a static model of the system. Still, this kind of a model has already been used to do basic structural analysis, for example for building steady-state models of metabolism (for reviews see Klamt & Stelling 2003; Llaneras & Picó 2008; Planes & Beasley 2008), studying logical networks (Glass & Kauffman 1973; Thomas 1973) or looking for network motifs in regulatory networks (Alon 2007).

SBML supports this stage, especially with the reactions element. This element is not solely destined for simple chemical reactions, but encompasses all forms of transformations between physical entities as well as transport, in- and outflows. Each reaction needs to have at least one reactant or product with a given

stoichiometry. Furthermore, it can also contain a species called modifiers, which are not consumed by the reaction, but they influence its rate in some form. To further characterize the role of involved reactants and modifiers, it is helpful to use controlled vocabularies. SBML allows the inclusion of identifiers pointing to terms from the systems biology ontology (SBO; Le Novère *et al.* 2007) to add a layer of semantics on the components of a model. SBO is a set of six controlled vocabularies that cover, in the context of systems biology, elements such as interactions, mathematical expressions, modelling frameworks and quantitative parameters. In the repressilator model, for e.g., *LacIp* might be labelled as an inhibitor (SBO:0000020) in the transcription of the *tetR* gene.

Concerning electrical circuit diagrams, a graphical representation of a biochemical system can be very helpful at this stage. Unlike the engineering disciplines, however, the biological community has yet to adopt a unique and standardized graphical language for the display of biological networks. The Systems Biology Graphical Notation (SBGN) is a recently proposed visual language for the representation of biological networks (Le Novère *et al.* 2008). It allows unambiguous representation of a process diagram of the reaction network. A number of modelling tools allow one to lay out systems graphically and store them in SBML. CELLDESIGNER (Funahashi *et al.* 2003), for example, is a platform-independent graphical editor for biological networks, into which existing models in SBML format can be imported and modified. Notes can also be added and mathematical equations entered for the different interactions. Substrates, products and modifiers can be added either directly by drawing and setting the appropriate connections on the canvas or by entering them in the corresponding lists. Information about the graphical layout is stored in a CELLDESIGNER-specific format in the annotation elements of the newly created SBML file. A possible SBGN compliant layout of the repressilator and CELLDESIGNER's interface are shown in figure 3.

While CELLDESIGNER is mainly an editor, it has some built-in deterministic integration capability features, based on a third party ODE SOLVER, SBMLODESOLVER (Machné *et al.* 2006). For a more thorough analysis, such as stochastic interpretations or structural analysis, the simulation package COPASI (Hoops *et al.* 2006) and the large group of tools contained in the systems biology workbench (SBW; Sauro *et al.* 2003) can be used directly from the program. Using SBML as its native format CELLDESIGNER generates models that can be used by the whole family of SBML aware tools.

While being intuitive and visually appealing, graphical representations can quickly become cluttered and confusing for large or densely connected networks. Another possibility for creating and editing models in SBML format is purely form-based editors, which provide scalability while retaining precision. We would just wish to mention a few different approaches at this point. For instance, JIGCELL (Vass *et al.* 2004), developed at the Virginia Polytechnic Institute & State University, uses a model builder in the form of a spreadsheet. This gives a nice overview even in cases of complex and dense networks. Meanwhile COPASI (Hoops *et al.* 2006) uses a tree-like representation of

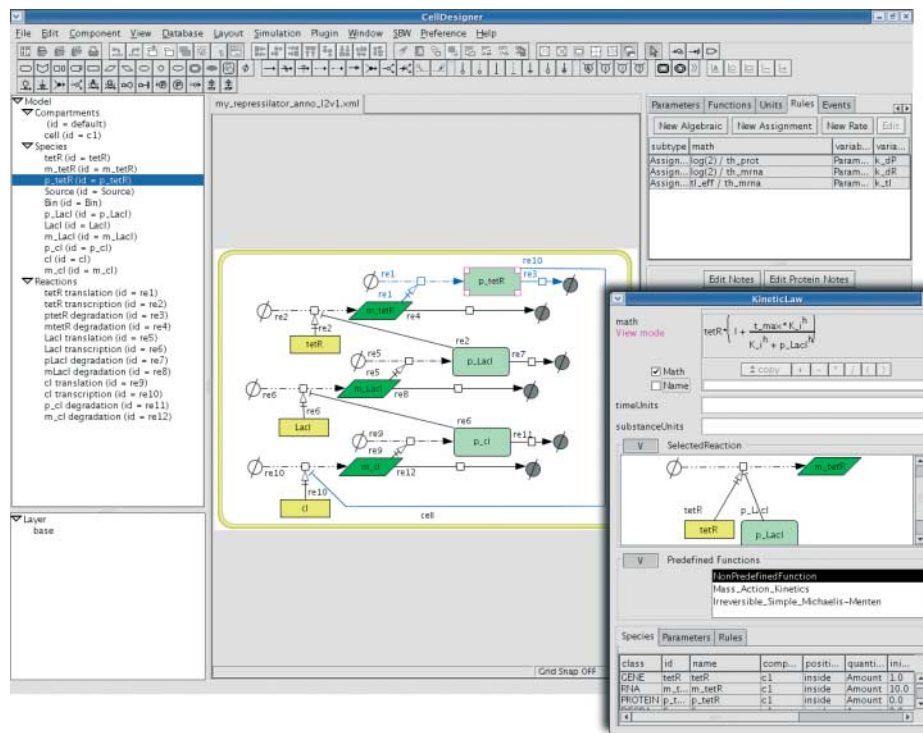


Figure 3. The CELLDESIGNER interface provides an ample selection of graphical elements and editing options to design biochemical and mathematical models. The repressilator is represented in CELLDESIGNER's graphical notation, a derivative of SBGN. The raised window is the editing form for the kinetic law of a single reaction.

the different models elements, spreadsheets to quickly list elements of a given type and more detailed forms to precisely alter each element. While potentially slower than entering reactions in a pure spreadsheet, it facilitates the definition of complex reactions and performs extensive checking in the background. A third approach is taken by SBMLEditor (Rodriguez *et al.* 2007; figure 4), a low-level SBML editor, currently without any simulation or evaluation capabilities. SBMLEditor is more closely based on the SBML format itself and supports the entire language. It has been up to now one of the few that permit direct entering of MIRIAM and SBO annotations.

Lately, some tools specifically aimed for synthetic biology have been created. One approach has been to create gene-regulatory networks from reusable modules similar to those stored in the MIT Registry of Standard Biological Parts (Endy 2005). An effort using quantitative modules in CellML format has been presented at the SysBioSys 2007 conference (Rouilly *et al.* 2007). Some recent tools have allowed the export in SBML format, such as the command line tool ASPARTS (Rodrigo *et al.* 2007), synthetic biology software suite (Hill *et al.* 2008), as well as an extension (Marchisio & Stelling 2008) written for the ProMOT (Ginkel *et al.* 2003) suite. These tools and efforts are still in the early stages, but they have already seemed to be quite usable, even if they do not offer the full range of biological parts available in the registry.

## 5. ENTERING MATHEMATICAL RELATIONS AND NUMERICAL VALUES

Once a potential network structure is created, a mathematical framework has to be chosen to develop

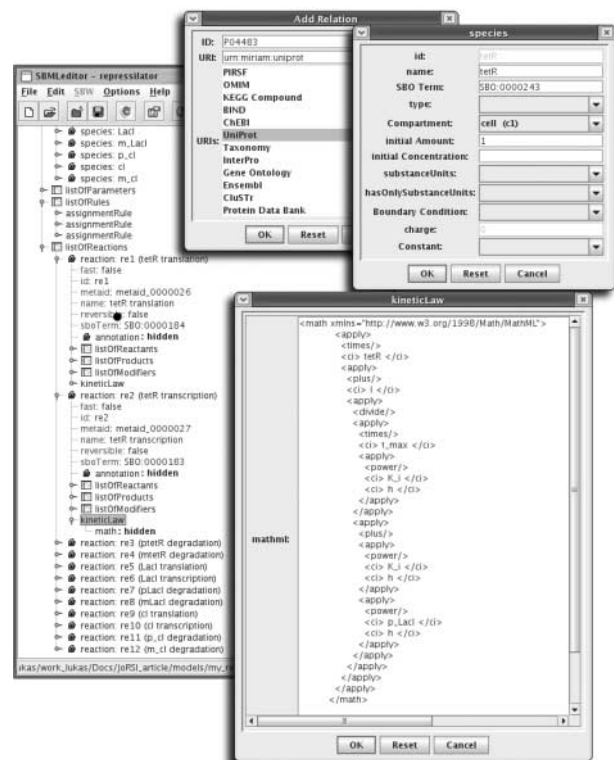


Figure 4. SBMLEditor allows detailed manipulation of all elements of SBML. Here the main overview, the species editing form, a MIRIAM annotation dialogue and the MathML code of the kinetic law shown in figure 2 are displayed.

a kinetic version. As different frameworks allow for complementary analysis and interpretation, it is often useful to keep the model interpretable by more than

one formalism. The simultaneous use of alternative frameworks has proved beneficial, especially for gene regulatory networks. This can be seen in one of the landmark papers of synthetic biology describing the repressilator (Elowitz & Leibler 2000). A deterministic approach was used to analyse the qualitative behaviour in dependence of key parameters, while a stochastic version permitted testing of the robustness of the design to transcriptional noise. In this section, we will concentrate on these two frameworks.

As chemical and biological processes are inherently stochastic, it seems natural to take this into account for modelling. However, stochastic simulations are generally computationally much more intensive, and for bigger models or greater numbers of interacting components, this burden quickly becomes limited. For metabolic processes, and in general reactions involving more than a few hundred molecules, the continuous deterministic approach provides a fairly good approximation.

The adequate forms of the mathematical expressions describing interactions and reaction velocities are often difficult to find at first, and sometimes different kinetic laws have to be tried out. In general, one can apply expression derived from first principles or, if the mechanism of reaction is not well defined or completely unknown, resort to generic or empirical rate laws. One of the most general approaches is to employ rate laws derived according to the law of mass action kinetics. In these equations, reaction rates are directly proportional to the activities of the reactants (e.g. concentrations in dilutions and partial pressures in gas phase) to a power, called the order of the reaction for this reactant, which is equal to its stoichiometry. This allows automatic generation of kinetic expressions for bigger reaction networks with only stoichiometric information, by fitting experimental datasets (sometimes leading to reaction orders that differ from stoichiometries). While this approach is widely applied in chemical systems and also has been quite successfully used in signal transduction, it leads to a high number of parameters and intermediary steps, and needs the explicit inclusion of catalysing enzymes. For enzyme-catalysed reactions with a known reaction mechanism, mechanistic rate laws using quasi steady-state or rapid equilibrium assumptions can be applied. The most common rate laws can be looked up in reference books (e.g. Segel (1993) or Cornish-Bowden (2004)) or derived using methods such as the graph-based one derived by King & Altman (King & Altman 1956; Chou 1989). However, for this approach a detailed knowledge of the reaction mechanisms and modifying effectors is essential. Each reaction has also to be treated individually, which might not be feasible for larger models. As an alternative, generic expressions can be used, which show the general behaviour of enzyme-catalysed reactions for varying ranges around reference states of the system. These allow the inclusion of experimentally derived parameters, activators or inhibitors to some extent, even if the exact mechanisms are not known. The simplest forms of these are the irreversible Michaelis–Menten and Hill-like rate laws. However, as many reactions cannot be assumed to be mostly unidirectional (Cornish-Bowden & Cardenas 2001), generic reversible forms for the most

commonly used biochemical rate laws have been proposed. Examples are the convenience rate law (Liebermeister & Klipp 2006) and the reversible Hill equation (Rohwer *et al.* 2007). These rate laws also allow one to include thermodynamic constants, which are easier to measure than kinetic constants, and to easily adopt experimentally determined parameters, while displaying behaviours similar to more detailed mechanisms over a broad range of concentrations.

Some modelling tools offer predefined rate laws. For instance, COPASI and WEBCELL have a wide range of enzyme and general kinetic laws which can be easily incorporated into the model or used to create more elaborate ones if necessary. Another interesting feature offered by COPASI is the ability to transform all reversible reactions in a model into pairs of irreversible ones. While this of course only works for simple rate laws, it facilitates the use of subsequent stochastic evaluations, as these require the use of probabilities for elementary molecular events.

A convenient way to try out different rate laws for larger models is shown by SBMLSQUEEZER (Dräger *et al.* 2008), a plug-in for CELLDISIGNER. It allows one to choose single or groups of reactions, and to apply different kinetic laws from a list of formalisms. In order to assign the adequate rate law and use the correct reactants, products and modifiers, it analyses the diagram created by CELLDISIGNER (a derived form of SBGN) and takes SBO annotations of the species, parameters and reactions into account. While this still needs some user interaction, it is a commendable improvement that greatly facilitates the creation of larger and complex models.

Various mathematical formalisms are employed to model gene-regulatory networks (reviewed in de Jong 2002). While there is evidence for the stochastic nature of transcriptional regulation (Fiering *et al.* 2000; Elowitz *et al.* 2002), deterministic approaches are often chosen due to their easier evaluation, and for simplicity of qualitative analysis. In our example, the repressilator, the authors used a Hill-type regulation function to simulate transcriptional repression, and first-order mass action rate laws for modelling translation, transcription and decay. Hill-type functions are often used to model gene expression or signal transduction as they are a subset of logistic functions, and they provide a sigmoid or step-like dose response to modulators, as found in some experimental results (Yagil & Yagil 1971; Rosenfeld *et al.* 2005; Kaplan *et al.* 2008). Sometimes, the Hill coefficient can be justified by cooperative binding behaviour—as in the case of the repressilator. In other cases, it can be due to dimensional restriction or more complex mechanisms, subsumed into a single step. Transcription and translation, while actually being composed of multiple reactions, are modelled as single first-order reactions. For the stochastic simulations, repressor binding had to be unravelled into single steps. COPASI allows one to interpret the model using different algorithms. Both a sample stochastic and deterministic time-course evaluation is shown in figure 5.

Some problems cannot be sufficiently described without taking spatial inhomogeneities into account. Although many intracellular processes have been



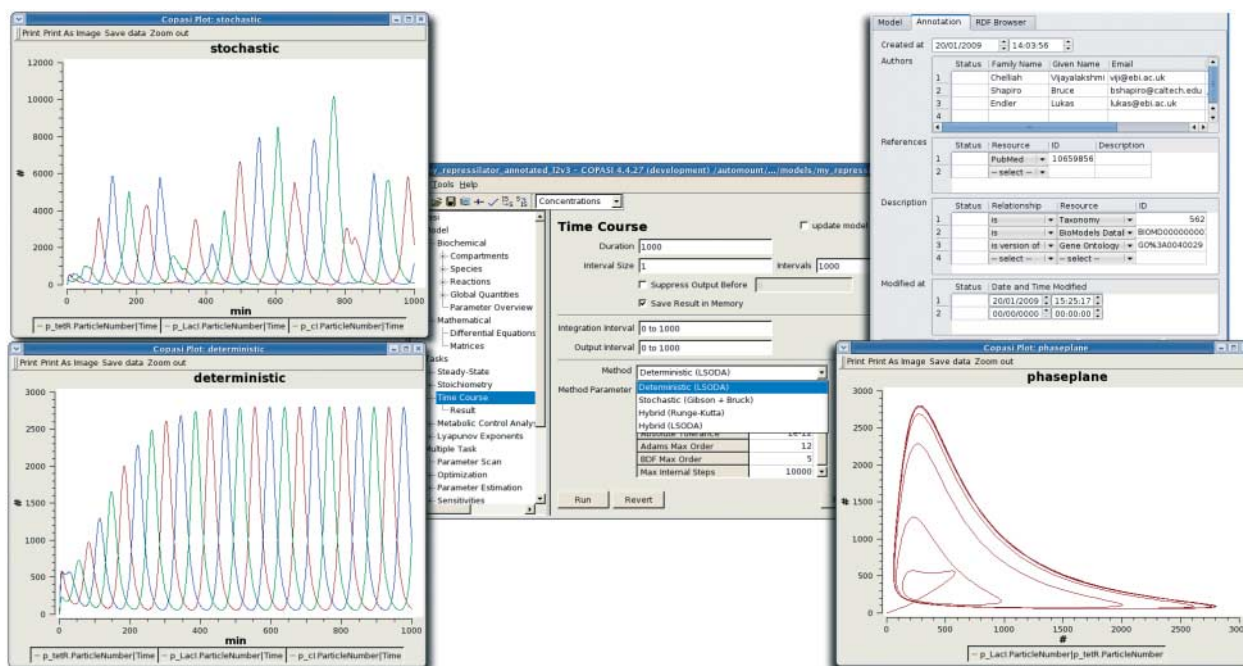


Figure 5. COPASI offers various methods for evaluating dynamic models and displaying the results. Time courses of the protein numbers of the repressor are shown here, calculated using a deterministic (LSODA) and a stochastic (Gibson–Bruck) method. The deterministic solution is also represented as the phase plane of the tetR and the LacI proteins, showing the convergence to a limit cycle. In a separate window, the MIRIAM annotations of the model element are shown.

described with the assumption of a well-stirred reaction environment, this approach is insufficient for processes involving intracellular or extracellular gradients or heterogeneous populations of molecules (Kholodenko 2006). While some tools exist to help with the creation and simulation, there are no standard formats for interchange of models involving diffusional processes up to now. An exhaustive survey of the various spatial modelling approaches is beyond the scope of this paper. Different software tools and formalisms commonly employed are reviewed in Lemerle *et al.* (2005), Takahashi *et al.* (2005) and Tolle & Le Novère (2006).

Another problem plaguing modelling, in particular of signalling processes, is the combinatorial explosion resulting from alternative non-covalent binding of proteins, and molecular entities existing under different states (conformations, covalent modifications, etc.) One approach to avoid the problem is to use agent-based models of populations. In these methods, interactions between autonomous individuals are simulated. They have been successfully employed to simulate, for instance, bacterial chemo-taxis (Shimizu *et al.* 2003), tissue formation and developmental processes (reviewed in Thorne *et al.* 2007) and cancer growth (Wang *et al.* 2007; Zhang *et al.* 2009). A related approach is the use of rule-based models, where actual reactions are not described, but only the rules to generate them during simulations (e.g. Blinov *et al.* 2004; Lok & Brent 2005). As with spatial modelling, there are still few software tools supporting these frameworks, specifically for biology.

## 6. FINDING AND FITTING PARAMETERS

To create a quantitative model, values for the various constants and parameters used in mathematical

relations have to be derived. While there exist a vast amount of experimentally derived values in scientific literature, it is often hard to find the relevant ones in the multitude of publications.

For enzyme-catalysed reactions, there exist various databases helping to identify the appropriate values. Two databases providing kinetic parameters are BRENDA (Chang *et al.* 2008) and the above mentioned SABIO-RK (Wittig *et al.* 2006). Both offer a wide range of parameters and reactions extracted from primary literature, with powerful search options. SABIO-RK additionally offers the mechanism assumed in the original source and the ability to export reactions in SBML format. Help with directly searching the primary literature is offered by KMedDB (<http://sysbio.molgen.mpg.de/KMedDB>; Hakenberg *et al.* 2004). It allows PubMed abstracts to be searched for various kinetic parameters in combination with compound, organism or enzyme reaction identifiers. Further information on the thermodynamics of biological reactions is available at the TECRDB (Goldberg *et al.* 2004). Another helpful source of general interaction parameters is given by the Kinetik Data of Biomolecular Interaction database (Kumar *et al.* 2008).

Another common way of finding quantitative information is going through papers describing modelling efforts in the relevant fields. These can be a valuable source of pointers to relevant primary literature, and can also help with deriving or adapting experimental parameters to the form needed for the model. The above-mentioned databases and repositories of models are quite useful in this respect, as they not only give an overview of existing parametrized models, but also offer links to the primary literature.

Most of the parameters derived from the literature can nevertheless be taken only as guideline values for

modelling. If measured time courses or steady-state data exist for the system to be described, several algorithms have been implemented for parameter estimation and refinement. For a short overview and comparison of some of these methods see Moles *et al.* (2003) and Rodriguez Fernandez *et al.* (2006). An intuitive and useful interface that features numerous global and local estimation methods is offered by COPASI. It allows simultaneous fitting of a subset of parameters to different sets of experimental values. Unfortunately, COPASI up to now has not included support for events. To estimate parameters in models containing events, the tool PET (<http://mpf.biol.vt.edu/pet>), closely connected with the JIGCELL suite, offers a convenient graphical interface also allowing multiple time series to be fitted simultaneously. Another tool supporting events is the command line-driven SBML-PET (Zi & Klipp 2006). For users of the commercial Matlab (The MathWorks, MA, USA) environment, SBTOOLBOX2 (Schmidt & Jirstrand 2006), a free package with SBML support, offers various estimation and optimization methods.

A completely different approach to search for adequate parameters is subsumed under the term optimization. While similar in its methods to parameter estimation, it differs in that it tries to reach a global goal rather than fit a given set of parameters, and is widely used in the engineering of metabolic systems. For a comprehensive review see Banga (2008). Again COPASI offers a range of different algorithms for minimization of a given target objective function.

Another interesting approach is inverse bifurcation analysis (Lu *et al.* 2006). It can be used to find parameter values exhibiting certain qualitative behaviours. For example, it can help in finding regimes that display certain kinds of switching behaviour, or creating more robust oscillators with a given system. Unfortunately, the tools available for this approach up to now have used MATHEMATICA (Wolfram Research, Inc., IL, USA) or Matlab, and have still required quite some expert knowledge and skill. Hopefully, more user-friendly tools will be developed for this promising approach in the future.

## 7. VALIDATION AND EVALUATION OF THE MODEL

Once a model has been created that can be run, it still has to prove its ability to reproduce experimental results up to a required accuracy and predict interesting and non-trivial observables.

Validation of a model can be performed in two stages. Observation of the qualitative behaviour of the model can be very informative in models having multiple steady states and showing switch-like or oscillatory behaviours. The qualitative behaviour can be studied either over small parts of the parameter space, by simply scanning over defined ranges of parameters and initial conditions, or by doing global bifurcation analyses. While the first procedure is easily accessible in many software packages, the second one requires more dedicated tools. CELLDISIGNER offers simple scanning over a range of parameters through SBMLODESOLVER, while COPASI

offers more sophisticated possibilities. Using loops and combinations of tasks, scans over more than one condition or parameter are possible. For analysing the global qualitative behaviour of a model, numerical bifurcation analysis is one of the most widely used methods. Among the tools available, the free tools XPP-AUT (Ermentrout 2002) and OSCILL8 (<http://oscill8.sourceforge.net/>), both using the AUTO continuation library (Doedel 1981), have been used in biological modelling (e.g. Csikász-Nagy *et al.* 2006). Converters from SBML to the format used by XPP-AUT and OSCILL8 are integrated in JIGCELL, COPASI, SBW, SBTOOLBOX2 and Biomodels database. OSCILL8 can also be integrated into the SBW, allowing to seamlessly analyse models in this framework.

Model checking, originally used in computer science (Clarke & Emerson 1982; Queille & Sifakis 1982), is a different approach to validate the qualitative behaviour of systems. It allows one to test whether the system can fulfil certain objectives, for example reachability of certain states, the consecutive temporal activation of certain species or oscillatory behaviour. While support for quantitative dynamical models is still insufficient, BIOCHAM (Calzone *et al.* 2006) has limited SBML support and offers model checking capabilities. ROVER GENE (Batt *et al.* 2007), a free add-on to the commercial Matlab environment, has been more specifically aimed at the needs of synthetic biology. Using piecewise affine, or for a better fit multi-affine, functions to model transcriptional regulation, it allows users to check whether a genetic regulatory system can exhibit a desired dynamical property or behaviour in a given range of parameters and initial conditions. More relevant to the design of networks, the tool can also be used to find parameters showing a desired behaviour and to test the robustness of the behaviour around these parameter values.

Qualitative analysis can also give hints as to which parameters offer the best success in achieving a desired behaviour or whether a certain design can exhibit the wanted function at all. Identifying the most promising parameters to change, of course, depends not only on the mathematical analysis, but also on the biological feasibility. While some characteristics such as promoter strength, transcript and protein stability are quite variable, enzymatic activities, for example, might be harder to tweak. Also as changes in the characteristics of biological components can at best be qualitative, it is important to find parameter ranges that show behaviour robust to variations. In the repressor example mentioned above, the qualitative analysis led to the identification of a few key properties important for obtaining stable oscillations—strong promoters with tight cooperative repression and comparable mRNA and protein half-lives, with the protein half-lives mainly determining the period length. Apart from helping to choose the right biological components, these criteria also led the authors of the paper to introduce tags for proteases into the repressor sequences.

Model validity can also be checked by comparison of the results of simulation runs with quantitative experimental data, such as time courses or steady-state concentrations and fluxes. These can sometimes

be derived from the literature, or retrieved from databases, for example, quantification of mRNA or metabolites. If the model satisfactorily reproduces experimental results and displays the desired behaviours, it can further be tested by experimental verification of its predicted results and behaviour. JIGCELL features a dedicated tool, the COMPARTOR (described in Allen *et al.* 2003), to automatize this comparison. Using user-defined objective functions, it compares a model's results and transformations thereof with given sets of experimental data and assertions on model variables. It also allows different models or versions of a model to be tested to the same data and assertions and then to compare their performance.

## 8. CONCLUSION AND OUTLOOK

The tools already available for model creation are quite sufficient for most applications. Use of inter-convertible formats such as SBML or CellML endows scientists with a rich toolkit supporting nearly all aspects of model evaluation, interpretation and refinement. Most of these tools are freely available for academic users, allowing scientists to try and use more than one program for each task. The ample annotation possibilities of these XML-based formats allow the models to be thoroughly described, which helps in collaborative model building and interpretation, as well as in model exchange and reuse.

Nevertheless, as most of the tools have been designed for systems biology purposes, with a slightly different modelling process in mind, they lack some of the features desirable in synthetic biology. Firstly, they hardly support the modular building process used in some efforts to create synthetic gene regulatory systems. Some tools try to include the available resources targeting synthetic biology—such as the MIT's registry of standard biological parts—as model building resources. While this looks quite promising, they currently offer only a few modules of the available biological elements. Although some of them support general exchange formats and thereby allow for easy export to other tools, they are restricted to only a small number of mathematical formulations, and lack integration of metabolism and signal transduction cascades.

Another feature barely supported up to now is the ability to use qualitative behaviour prediction to guide the design process. Tools that deduce possible behaviours and suggest feasible layouts could be very helpful in the first stages of a synthetic biology project. BIOCHAM offers this facility to some extent, but more tools for deterministic and stochastic kinetic modelling would be desirable. While parameter estimation and optimization are integrated in many programs with intuitive interfaces, qualitative analysis capabilities are mostly found in separate dedicated tools, requiring more mathematical skills and expertise. Efforts to make these tools more accessible to mathematically less proficient scientists would be very useful for both the systems and synthetic biology communities.

A general problem in both systems and synthetic biology is the lack of tools supporting standardized annotation of model elements and biological entities.

Some standards have only recently been agreed upon by a broader part of the community, so some time may still be needed until widespread adoption. We believe the use of such a standardized annotation is a necessary development for the interchange and reuse of models and modules by different scientists and tools. One very successful effort in developing a standard, publicly available database of biological parts for synthetic biology has been created in the BioBrick registry of standard biological parts (<http://partsregistry.org/>). In a recent publication, Canton *et al.* (2008) proposed a promising way to represent quantitative characteristics of biological devices in the form of a data sheet and demonstrate it on a device composed of BioBrick parts. Mathematical modelling and *in silico* design of systems would profit very much from the adoption of such standards by the community.

As the vibrant and quickly growing communities of synthetic and systems biology have produced a plethora of tools and algorithms, we are optimistic that the features mentioned above will be implemented soon. The use of open standard formats and community-based development has proved to be of great value in biological modelling and hopefully this successful trend will continue and widen to newly emerging fields.

The authors are thankful to Dominic Tolle for discussions. This work was partly supported by the British Biotechnology and Biological Sciences Research Council.

## REFERENCES

- Allen, N., Shaffer, C., Vass, M., Ramakrishnan, N. & Watson, L. 2003 Improving the development process for eukaryotic cell cycle models with a modeling support environment. *Simulation* **79**, 674–688. (doi:10.1177/0037549703040944)
- Alon, U. 2007 Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461. (doi:10.1038/nrg2102)
- Ananiadou, S., Kell, D. B. & Tsujii, J. 2006 Text mining and its potential applications in systems biology. *Trends Biotechnol.* **24**, 571–579. (doi:10.1016/j.tibtech.2006.10.002)
- Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. 2006 Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, msb4100073-E1–msb4100073-E14. (doi:10.1038/msb4100073)
- Ausbrooks, R. *et al.* 2003 Mathematical markup language (MathML), v. 2.0, 2nd edn. Technical report, W3C.
- Banga, J. R. 2008 Optimization in computational systems biology. *BMC Syst. Biol.* **2**, 47. (doi:10.1186/1752-0509-2-47)
- Batt, G., Yordanov, B., Weiss, R. & Belta, C. 2007 Robustness analysis and tuning of synthetic gene networks. *Bioinformatics* **23**, 2415–2422. (doi:10.1093/bioinformatics/btm362)
- Beckett, D. 2004 RDF/XML syntax specification (revised). Technical report, W3C.
- Berners-Lee, T., Fielding, R. & Masinter, L. 2005 Uniform resource identifier (URI): generic syntax. Technical report, The Internet Society.
- Blinov, M., Faeder, J., Goldstein, B. & Hlavacek, W. 2004 BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289–3291. (doi:10.1093/bioinformatics/bth378)

- Bornstein, B. J., Keating, S. M., Jouraku, A. & Hucka, M. 2008 LIBSBML: an API library for SBML. *Bioinformatics* **24**, 880–881. (doi:10.1093/bioinformatics/btn051)
- Bray, D., Bourret, R. B. & Simon, M. I. 1993 Computer simulation of the phosphorylation cascade controlling bacterial chemotaxis. *Mol. Biol. Cell* **4**, 469–482.
- Calzone, L., Fages, F. & Soliman, S. 2006 BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. *Bioinformatics* **22**, 1805–1807. (doi:10.1093/bioinformatics/btl172)
- Canton, B., Labno, A. & Endy, D. 2008 Refinement and standardization of synthetic biological parts and devices. *Nat. Biotechnol.* **26**, 787–793. (doi:10.1038/nbt1413)
- Caspi, R. et al. 2008 The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* **36**, D623–D631. (doi:10.1093/nar/gkm900)
- Chang, A., Scheer, M., Grote, A., Schomburg, I. & Schomburg, D. 2008 BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. *Nucleic Acids Res.* **37**, D588–D592. (doi:10.1093/nar/gkn820)
- Chen, H. & Sharp, B. M. 2004 Content-rich biological network constructed by mining PubMed abstracts. *BMC Bioinform.* **5**, 147. (doi:10.1186/1471-2105-5-147)
- Chou, K. C. 1989 Graphic rules in steady and non-steady state enzyme kinetics. *J. Biol. Chem.* **264**, 12 074–12 079.
- Clarke, E. & Emerson, E. 1982 Design and synthesis of synchronization skeletons using branching time temporal logic. In *Lecture Notes in Computer Science* (ed. D. Kozen), pp. 52–71. London, UK: Springer. (doi:10.1007/BFb0025774)
- Cornish-Bowden, A. 2004 *Fundamentals of enzyme kinetics*. London, UK: Portland Press.
- Cornish-Bowden, A. & Cardenas, M. L. 2001 Information transfer in metabolic pathways. Effects of irreversible steps in computer models. *Eur. J. Biochem.* **268**, 6616–6624. (doi:10.1046/j.0014-2956.2001.02616.x)
- Csikász-Nagy, A., Battogtokh, D., Chen, K. C., Novák, B. & Tyson, J. J. 2006 Analysis of a generic model of eukaryotic cell-cycle regulation. *Biophys. J.* **90**, 4361–4379. (doi:10.1529/biophysj.106.081240)
- de Jong, H. 2002 Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* **9**, 67–103. (doi:10.1089/10665270252833208)
- Doedel, E. J. 1981 AUTO: a program for the automatic bifurcation analysis of autonomous systems. *Congressus Numerantium* **30**, 265–284.
- Dräger, A., Hassis, N., Supper, J., Schröder, A. & Zell, A. 2008 SBMLsqueezer: a CELLDISIGNER plug-in to generate kinetic rate equations for biochemical networks. *BMC Syst. Biol.* **2**, 39. (doi:10.1186/1752-0509-2-39)
- Elowitz, M. B. & Leibler, S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338. (doi:10.1038/35002125)
- Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. 2002 Stochastic gene expression in a single cell. *Science* **297**, 1183–1186. (doi:10.1126/science.1070919)
- Endy, D. 2005 Foundations for engineering biology. *Nature* **438**, 449–453. (doi:10.1038/nature04342)
- Ermentrout, B. 2002 *Simulating, analyzing, and animating dynamical systems: a guide to XPPAUT for researchers and students*. Philadelphia, PA: SIAM.
- Fiering, S., Whitelaw, E. & Martin, D. I. 2000 To be or not to be active: the stochastic nature of enhancer action. *Bioessays* **22**, 381–387. (doi:10.1002/(SICI)1521-1878(200004)22:4<381::AID-BIESS>3.0.CO;2-E)
- Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. 2003 CELLDISIGNER: a process diagram editor for gene-regulatory and biochemical networks. *BIOSSILICO* **1**, 159–162. (doi:10.1016/S1478-5382(03)02370-9)
- Galperin, M. Y. & Cochrane, G. R. 2009 Nucleic acids research annual database issue and the NAR online molecular biology database collection in 2009. *Nucleic Acids Res.* **37**, D1–D4. (doi:10.1093/nar/gkn942)
- Gardner, T. S., Cantor, C. R. & Collins, J. J. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342. (doi:10.1038/35002131)
- Ginkel, M., Kremling, A., Nutsch, T., Rehner, R. & Gilles, E. D. 2003 Modular modeling of cellular systems with ProMoT/Divi. *Bioinformatics* **19**, 1169–1176. (doi:10.1093/bioinformatics/btg128)
- Glass, L. & Kauffman, S. 1973 The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* **39**, 103–129. (doi:10.1016/0022-5193(73)90208-7)
- Goldberg, R. N., Tewari, Y. B. & Bhat, T. N. 2004 Thermodynamics of enzyme-catalyzed reactions—a database for quantitative biochemistry. *Bioinformatics* **20**, 2874–2877. (doi:10.1093/bioinformatics/bth314)
- Hakenberg, J., Schmeier, S., Kowald, A., Klipp, E. & Leser, U. 2004 Finding kinetic parameters using text mining. *OMICS* **8**, 131–152. (doi:10.1089/1536231041388366)
- Hill, A. D., Tomshine, J. R., Weeding, E. M. B., Sotiropoulos, V. & Kaznessis, Y. N. 2008 SynBioSS: the synthetic biology modeling suite. *Bioinformatics* **24**, 2551–2553. (doi:10.1093/bioinformatics/btn468)
- Hines, M. L., Morse, T., Migliore, M., Carnevale, N. T. & Shepherd, G. M. 2004 MODELDB: a database to support computational neuroscience. *J. Comput. Neurosci.* **17**, 7–11. (doi:10.1023/B:JCNS.0000023869.22017.2e)
- Hoffmann, R. & Valencia, A. 2004 A gene network for navigating the literature. *Nat. Genet.* **36**, 664. (doi:10.1038/ng0704-664)
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. & Kummer, U. 2006 COPASI—a complex pathway simulator. *Bioinformatics* **22**, 3067–3074. (doi:10.1093/bioinformatics/btl485)
- Hucka, M. et al. 2003 The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531. (doi:10.1093/bioinformatics/btg015)
- Kanehisa, M. et al. 2008 KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484. (doi:10.1093/nar/gkm882)
- Kaplan, S., Bren, A., Zaslaver, A., Dekel, E. & Alon, U. 2008 Diverse two-dimensional input functions control bacterial sugar genes. *Mol. Cell* **29**, 786–792. (doi:10.1016/j.molcel.2008.01.021)
- Kholodenko, B. 2006 Cell-signalling dynamics in time and space. *Nat. Rev. Mol. Cell Biol.* **7**, 165–176. (doi:10.1038/nrm1838)
- King, E. & Altman, C. 1956 A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J. Phys. Chem.* **60**, 1375–1378. (doi:10.1021/j150544a010)
- Klamt, S. & Stelling, J. 2003 Two approaches for metabolic pathway analysis? *Trends Biotechnol.* **21**, 64–69. (doi:10.1016/S0167-7799(02)00034-3)
- Krallinger, M., Valencia, A. & Hirschman, L. 2008 Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome Biol.* **9**(Suppl. 2), S8. (doi:10.1186/gb-2008-9-s2-s8)
- Kumar, P. et al. 2008 Update of KDBI: kinetic data of biomolecular interaction database. *Nucleic Acids Res.* **37**, D636–D641. (doi:10.1093/nar/gkn839)

- Laibe, C. & Le Novère, N. 2007 MIRIAM resources: tools to generate and resolve robust cross-references in systems biology. *BMC Syst. Biol.* **1**, 58. (doi:10.1186/1752-0509-1-58)
- Lee, D.-Y., Yun, C., Cho, A., Hou, B. K., Park, S. & Lee, S. Y. 2006 WebCell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. *Bioinformatics* **22**, 1150–1151. (doi:10.1093/bioinformatics/btl091)
- Lemerle, C., Ventura, B. D. & Serrano, L. 2005 Space as the final frontier in stochastic simulations of biological systems. *FEBS Lett.* **579**, 1789–1794. (doi:10.1016/j.febslet.2005.02.009)
- Le Novère, N. *et al.* 2005 Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515. (doi:10.1038/nbt1156)
- Le Novère, N. *et al.* 2006 BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**, D689–D691. (doi:10.1093/nar/gkj092)
- Le Novère, N., Courtot, M. & Laibe, C. 2007 Adding semantics in kinetics models of biochemical pathways. In *Proc. 2nd Int. Symp. on Experimental Standard Conditions of Enzyme Characterizations (ESEC 2006)*, Beilstein Institute, Frankfurt am Main, Germany, pp. 137–153.
- Le Novère, N. *et al.* 2008 Systems biology graphical notation: process diagram level 1. Technical report. (doi:10.10101/npre.2008.2320.1)
- Liebermeister, W. & Klipp, E. 2006 Bringing metabolic networks to life: convenience rate law and thermodynamic constraints. *Theor. Biol. Med. Model.* **3**, 41. (doi:10.1186/1742-4682-3-41)
- Llaneras, F. & Picó, J. 2008 Stoichiometric modelling of cell metabolism. *J. Biosci. Bioeng.* **105**, 1–11. (doi:10.1263/jbb.105.1)
- Lloyd, C. M., Halstead, M. D. B. & Nielsen, P. F. 2004 CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* **85**, 433–450. (doi:10.1016/j.pbiomolbio.2004.01.004)
- Lloyd, C. M., Lawson, J. R., Hunter, P. J. & Nielsen, P. F. 2008 The CellML model repository. *Bioinformatics* **24**, 2122–2123. (doi:10.1093/bioinformatics/btn390)
- Lok, L. & Brent, R. 2005 Automatic generation of cellular reaction networks with Molecuizer 1.0. *Nat. Biotechnol.* **23**, 131–136. (doi:10.1038/nbt1054)
- Lu, J., Engl, H. W. & Schuster, P. 2006 Inverse bifurcation analysis: application to simple gene systems. *Algorithms Mol. Biol.* **1**, 11. (doi:10.1186/1748-7188-1-11)
- Machné, R., Finney, A., Müller, S., Lu, J., Widder, S. & Flamm, C. 2006 The SBML ODE Solver library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* **22**, 1406–1407. (doi:10.1093/bioinformatics/btl086)
- Marchisio, M. A. & Stelling, J. 2008 Computational design of synthetic gene circuits with composable parts. *Bioinformatics* **24**, 1903–1910. (doi:10.1093/bioinformatics/btn330)
- Matthews, L. *et al.* 2008 Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622. (doi:10.1093/nar/gkn863)
- Mi, H., Guo, N., Kejariwal, A. & Thomas, P. D. 2007 PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.* **35**, D247–D252. (doi:10.1093/nar/gkl869)
- Moles, C. G., Mendes, P. & Banga, J. R. 2003 Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* **13**, 2467–2474. (doi:10.1101/gr.1262503)
- Ng, A., Bursteinas, B., Gao, Q., Mollison, E. & Zvebil, M. 2006 Resources for integrative systems biology: from data through databases to networks and dynamic system models. *Brief. Bioinform.* **7**, 318–330. (doi:10.1093/bib/bbl036)
- Novak, B. & Tyson, J. J. 1997 Modeling the control of DNA replication in fission yeast. *Proc. Natl Acad. Sci. USA* **94**, 9147–9152. (doi:10.1073/pnas.94.17.9147)
- Planes, F. J. & Beasley, J. E. 2008 A critical examination of stoichiometric and path-finding approaches to metabolic pathways. *Brief. Bioinform.* **9**, 422–436. (doi:10.1093/bib/bbn018)
- Queille, J. & Sifakis, J. 1982 Specification and verification of concurrent systems in CESAR. In *International Symposium on Programming* (eds M. Dezani-Ciancaglini & U. Montanari). Lecture Notes in Computer Science, vol. 137, pp. 337–351. Berlin, Germany: Springer. (doi:10.1007/3-540-11494-7\_22)
- Reijenga, K. A., Snoep, J. L., Diderich, J. A., van Verseveld, H. W., Westerhoff, H. V. & Teusink, B. 2001 Control of glycolytic dynamics by hexose transport in *Saccharomyces cerevisiae*. *Biophys. J.* **80**, 626–634. (doi:10.1016/S0006-3495(01)76043-2)
- Rodrigo, G., Carrera, J. & Jaramillo, A. 2007 Asmparts: assembly of biological model parts. *Syst. Synth. Biol.* **1**, 167–170. (doi:10.1007/s11693-008-9013-4)
- Rodríguez, N., Donizelli, M. & Novère, N. L. 2007 SBML EDITOR: effective creation of models in the systems biology markup language (SBML). *BMC Bioinform.* **8**, 79. (doi:10.1186/1471-2105-8-79)
- Rodríguez Fernandez, M., Egea, J. A. & Banga, J. R. 2006 Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinform.* **7**, 483. (doi:10.1186/1471-2105-7-483)
- Rohwer, J., Hanekom, A. J. & Hofmeyr, J.-H. 2007 A universal rate equation for systems biology. In *Proc. 2nd Int. Symp. on Experimental Standard Conditions of Enzyme Characterizations (ESEC 2006)*, Beilstein Institute, Frankfurt am Main, Germany, pp. 175–187.
- Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Belowitz, M. 2005 Gene regulation at the single-cell level. *Science* **307**, 1962–1965. (doi:10.1126/science.1106914)
- Rouilly, V., Canton, B., Nielsen, P. & Kitney, R. 2007 Registry of BioBricks models using CellML. In *BioSysBio 2007: Systems Biology, Bioinformatics, Synthetic Biology, Meeting Abstracts*. *BMC Systems Biology* **1**(Suppl. 1), 79. (doi:10.1186/1752-0509-1-s1-p79)
- Sauro, H. M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J. & Kitano, H. 2003 Next generation simulation tools: the systems biology workbench and BioSPICE integration. *OMICS* **7**, 355–372. (doi:10.1089/15362310322637670)
- Schilstra, M. J., Li, L., Matthews, J., Finney, A., Hucka, M. & Le Novère, N. 2006 CellML2SBML: conversion of CellML into SBML. *Bioinformatics* **22**, 1018–1020. (doi:10.1093/bioinformatics/btl047)
- Schmidt, H. & Jirstrand, M. 2006 Systems biology toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* **22**, 514–515. (doi:10.1093/bioinformatics/bti799)
- Segel, I. H. 1993 *Enzyme kinetics: behavior and analysis of rapid equilibrium and steady-state enzyme systems*. New York, NY: Wiley.
- Shimizu, T., Aksenov, S. & Bray, D. 2003 A spatially extended stochastic model of the bacterial chemotaxis signalling pathway. *J. Mol. Biol.* **329**, 291–309. (doi:10.1016/S0022-2836(03)00437-6)
- Sivakumaran, S., Hariharaputran, S., Mishra, J. & Bhalla, U. S. 2003 The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **19**, 408–415. (doi:10.1093/bioinformatics/btf860)

- Snoep, J. L. & Olivier, B. G. 2002 Java web simulation (JWS); a web based database of kinetic models. *Mol. Biol. Rep.* **29**, 259–263. (doi:10.1023/A:1020350518131)
- Sprinzak, D. & Elowitz, M. B. 2005 Reconstruction of genetic circuits. *Nature* **438**, 443–448. (doi:10.1038/nature04335)
- Takahashi, K., Arjunan, S. N. V. & Tomita, M. 2005 Space in systems biology of signaling pathways: towards intracellular molecular crowding *in silico*. *FEBS Lett.* **579**, 1783–1788. (doi:10.1016/j.febslet.2005.01.072)
- Thomas, R. 1973 Boolean formalisation of genetic control circuits. *J. Theor. Biol.* **42**, 565–583. (doi:10.1016/0022-5193(73)90247-6)
- Thorne, B. C., Bailey, A. M. & Peirce, S. M. 2007 Combining experiments with multi-cell agent-based modeling to study biological tissue patterning. *Brief. Bioinform.* **8**, 245–257. (doi:10.1093/bib/bbm024)
- Tolle, D. & Le Novère, N. 2006 Particle-based stochastic simulation in systems biology. *Curr. Bioinform.* **1**, 315–320. (doi:10.2174/157489306777827964)
- Undrovinas, A. I., Belardinelli, L., Undrovinas, N. A. & Sabbah, H. N. 2006 Ranolazine improves abnormal repolarization and contraction in left ventricular myocytes of dogs with heart failure by inhibiting late sodium current. *J. Cardiovasc. Electrophysiol.* **17**(Suppl. 1), S169–S177. (doi:10.1111/j.1540-8167.2006.00401.x)
- Vass, M., Allen, N., Shaffer, C. A., Ramakrishnan, N., Watson, L. T. & Tyson, J. J. 2004 The JIGCELL model builder and run manager. *Bioinformatics* **20**, 3680–3681. (doi:10.1093/bioinformatics/bth422)
- Vastrik, I. *et al.* 2007 Reactome: a knowledge base of biologic pathways and processes. *Genome Biol.* **8**, R39. (doi:10.1186/gb-2007-8-3-r39)
- von Dassow, G., Meir, E., Munro, E. M. & Odell, G. M. 2000 The segment polarity network is a robust developmental module. *Nature* **406**, 188–192. (doi:10.1038/35018085)
- Wang, Z., Zhang, L., Sagotsky, J. & Deisboeck, T. S. 2007 Simulating non-small cell lung cancer with a multiscale agent-based model. *Theor. Biol. Med. Model.* **4**, 50. (doi:10.1186/1742-4682-4-50)
- Weidemann, A. *et al.* 2008 SYCAMORE—a systems biology computational analysis and modeling research environment. *Bioinformatics* **24**, 1463–1464. (doi:10.1093/bioinformatics/btn207)
- Wierling, C., Herwig, R. & Lehrach, H. 2007 Resources, standards and tools for systems biology. *Brief. Funct. Genomic Proteomic* **6**, 240–251. (doi:10.1093/bfgp/elm027)
- Wittig, U., Golebiewski, M., Kania, R., Krebs, O., Mir, S., Weidemann, A., Anstein, S., Saric, J. & Rojas, I. 2006 SABIO-RK: integration and curation of reaction kinetics data. In *Data integration in the life sciences*, vol. 4075 (eds U. Leser, F. Neumann & B. Eckman). Lecture Notes in Computer Science, pp. 94–103. Berlin, Germany: Springer. (doi:10.1007/11799511\_9)
- Yagil, G. & Yagil, E. 1971 On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* **11**, 11–27. (doi:10.1016/S0006-3495(71)86192-1)
- Zhang, L., Wang, Z., Sagotsky, J. A. & Deisboeck, T. S. 2009 Multiscale agent-based cancer modeling. *J. Math. Biol.* **58**, 545–559. (doi:10.1007/s00285-008-0211-1)
- Zi, Z. & Klipp, E. 2006 SBML-PET: a systems biology markup language-based parameter estimation tool. *Bioinformatics* **22**, 2704–2705. (doi:10.1093/bioinformatics/bt443)