Systems biology model databases and resources

Carel van Gend* and Jacky L. Snoep*†1

*Department of Biochemistry, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa, and †Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K.

Abstract

Systems biology aims at a quantitative understanding of systemic behaviour as a function of its components and their interactions. In systems biology studies computer models play an important role: (i) to integrate the components’ behaviour and (ii) to analyse experimental data sets. With the growing number of kinetic models that are being constructed for parts of biological systems, it has become important to store these models and make them available in a standard form, such that these models can be combined, eventually leading to a model of a complete system. In the present chapter we describe database initiatives that contain kinetic models for biological systems, together with a number of other systems biology resources related to kinetic modelling.

Introduction

The ultimate predictive model in systems biology would simulate the phenotype of an organism on the basis of its DNA sequence. A core assumption of such a model would be that a system can be understood on the basis of its component characteristics and interactions. With all the components of the cell being coded for by the DNA, it should then in principle be possible to predict cellular

1To whom correspondence should be addressed (email jls@sun.ac.za).
behaviour from its genetic blueprint. Currently, we are still far away from giving a complete quantitative description of even a single cell. However, we are able to give a fairly complete map of the network structure (not kinetics) of metabolism in a number of model organisms [e.g. KEGG (Kyoto Encyclopedia of Genes and Genomes) database; http://www.genome.ad.jp/kegg/], on the basis of DNA sequence homologies [e.g. BiGG (biochemically, genetically and genomically structured) database; http://bigg.ucsd.edu/]. Note that such an analysis on the basis of homologies is essentially different from a first principle prediction of gene product function from the DNA sequence.

Remarkable progress has been made in the last century to characterize cellular components and to accumulate knowledge of cellular pathways and physiology. But most of these studies were focused on a specific aspect and did not attempt to integrate the knowledge on the parts to describe systemic behaviour; this is the subject of the relatively new field of systems biology.

One of the challenges of systems biology is to combine knowledge on all of the parts of a system to come to an understanding of its functioning as a whole. Systems biology could only develop as rapidly as it has done because so much knowledge had been accumulated on a genome scale. Starting with the sequencing of complete genomes, and later also with other complete systems’ measurements, large data sets have been obtained in so-called ‘omics’ studies: at the transcriptional level (DNA chip arrays), the translational level (proteomics) and the molecular level (metabolomics).

Computer models play an important role in systems biology approaches. Ideally one would construct a model using biochemical knowledge of the components and validate these models using the ‘omics’ data sets (see e.g. the Silicon Cell project; http://www.siliconcell.net). Thus one could scan the scientific literature for a biochemical characterization of the components in a system and cast this in a mathematical formulation of the system. However, many of the characteristics of a component (e.g. $K_m$ values of enzymes) are dependent on the conditions under which they are determined and the results in the literature will not necessarily reflect the same conditions for all components and the conditions used might not always be physiological. Typically one would need kinetic information on many components in a system and it can be very time consuming to find this information (and not all components might be characterized). A number of initiatives have been started to store kinetic information in databases, including the conditions under which they have been determined (e.g. Brenda; http://www.brenda-enzymes.info/; and SABIO-RK; http://sabio.villa-bosch.de/SABIORK/). In addition the importance for standardization of data sets and experimental conditions is being realized (e.g. STRENDA; http://www.strenda.org/).

Building a detailed kinetic model for a complete biological system such as a cell is an ambitious project. Even for a relatively simple micro-organism, such as *Escherichia coli*, several thousands of reactions are occurring simultaneously, and it will be very difficult to construct a detailed model for such a large system in a single step. Significant progress has been made in the modelling of
genome-scale metabolic systems using so-called flux (balance) analysis, where steady-state constraints on the metabolic network have been used to calculate fluxes over the network (e.g. [1]). However, such models do not contain kinetics and for large-scale kinetic models a modular approach might be more suitable [2]. In such an approach, models are constructed for parts of the system and, after validation, models can be combined into a larger model thereby gradually building a complete model of a cell. Such a modular approach has been suggested in the Silicon Cell project (SiC; http://www.siliconcell.net/), in which it has been realized that the models should be constructed in a specific way such that they can later be combined.

It is to be expected that many kinetic models will be developed in systems biology projects and these models should be available to the scientific community. Coding the mathematical models from the literature is a time-consuming and frustrating job since many of the models are not completely described. Although many databases exist that contain resources relevant to systems biology, we shall limit ourselves in this chapter primarily to databases and resources which contain, or are associated with, kinetic models of biological systems (Table 1). We will focus on our own initiative, JWS Online [3] which was the first web-driven database that made a curated set of biological models available for simulation.

**Model databases**

**JWS Online**

JWS Online, developed in the Department of Biochemistry, University of Stellenbosch (South Africa), the Vrije Universiteit (Amsterdam, The Netherlands) and Manchester University (Manchester, U.K.), has three facets: it is a curated database of peer-reviewed models, it is a web-based system for the simulation of these models, and it provides a secure facility for the storage and testing of models under review by scientific journals [3].

JWS Online relies on a client–server architecture. It is accessed through a web-browser interface, with a remote server machine performing the calculations. Currently JWS Online is accessible via three URLs: http://jjj.biochem.sun.ac.za, http://jjj.bio.vu.nl and http://jjj.mib.ac.uk. The client–server model has the advantage that no more than a Java-enabled web browser is required by users wishing to access the functionality of JWS Online. No special software is required, since the ‘heavy lifting’ is done by dedicated software installed on the server machine; this includes Mathematica (http://www.wolfram.com), a powerful system capable of performing complicated calculations and integrations, the PostgreSQL (http://www.postgresql.org) database system, and various Python (http://www.python.org) scripts to process user requests, interface with the database and generate dynamic HTML pages to return to the user.

The web interface may be used to search for, query and select models. Searches may be made on various criteria, either by keyword (one of model
Table 1. A comparison between several web-based initiatives that store kinetic models and/or make models available for simulation

The initiatives are compared on their functionality with respect to whether they (i) store a collection of models (DB), (ii) allow simulations to be run on the site (Sim), (iii) curate the stored models (i.e. do the models show the same behaviour as the published model; Cur), (iv) annotate the models (Ann), (v) actively add more models (here copying from other initiatives is not considered active; Add) and (vi) support web services (WS).

<table>
<thead>
<tr>
<th>Initiative</th>
<th>URL</th>
<th>DB</th>
<th>Sim</th>
<th>Cur</th>
<th>Ann</th>
<th>Add</th>
<th>WS</th>
</tr>
</thead>
</table>
| JWS Online | http://jjj.biochem.sun.ac.za
http://jjj.bio.vu.nl
http://jjj.mib.ac.uk | Yes | Yes | Yes | No* | Yes | Yes |
| Biomodels | http://www.ebi.ac.uk/biomodels
http://jjj.mib.ac.uk | Yes | No† | Yes | Yes | Yes | Yes |
| Virtual Cell | http://www.nrcam.uchc.edu | Yes | Yes | No | No | No | No |
| WebCell | http://webcell.kaist.ac.kr | Yes | Yes | No | No | No | No |
| JSim | http://jsr.bioeng.washington.edu/ | Yes | Yes | No | No | Yes | No |
| CCDB | http://www.itb.cn.it/celldcyte/ | Yes | Yes | Yes | Yes | Yes | No |
| DOQCS | http://doqcs.ncbs.res.in/ | Yes | No | Yes | Yes | Yes | No |
| ModelDB | http://senselab.med.yale.edu/
senselab/ModelDB/ | Yes | No | Yes | No | Yes | No |
| SigPath | http://www.sigpath.org/ | Yes | No | No | Yes | No | No |
| CellML | http://www.cellml.org | Yes | No | No | Yes | No | No |
| SBML | http://sbml.org | No | No | Yes | No | Yes | No |

*Annotation of models will be added in 2008, in-line with MIRIAM standards, in collaboration with Biomodels.
†Most of the models have a link to JWS Online for simulation.
author name, title, journal in which published, model organism or category),
or by selecting the organism and category from a drop-down menu. In each
case, a list of models satisfying the selection criteria is returned to the user.
From this list, further information for any of the models may be requested;
this information includes the title, full list of authors and the abstract of the
published description of the model. Each model in the list also has a link to a
page from which SBML (Systems Biology Markup Language) [4] and PySCeS
(Python Simulator for Cellular Systems) [5] representations of the model may
be downloaded.

Selecting a particular model results in a Java applet and model reaction
scheme being downloaded to the user’s web browser. The scheme represents
all of the reactions constituting the model; hovering the pointer over a particu-
lar reaction displays the full kinetics of this reaction. Double clicking on the
reaction scheme produces a pop-up window with yet more information on the
reactions.

From the applet, the model parameters may be examined in detail and
altered if required. The applet allows a range of operations to be performed on
the model; these include time course simulations, steady-state analyses and
metabolic control analyses. Selecting a model simulation or analysis causes a
request to be sent to the server; here the request is performed by an instance
of a Mathematica kernel, and the results are returned to the user’s browser.
A screenshot of the JWS Online interface and a simulation result is shown in
Figure 1. Simulation results can be stored as a text or a CSV (comma separated
variable) file.

JWS Online may also be accessed as a web service. A WSDL (web service
description language) file describing the service is available at http://jjj.
biochem.sun.ac.za/axis/services/QueryJWS?wsdl. The web service is prima-
arily suited for machine-to-machine interactions, and may be incorporated in
workflows such as those facilitated by Taverna (http://taverna.sourceforge.
net/) or similar web service frameworks. The web service allows database que-
ries, model simulations and analyses to be performed, with the results being
returned as XML-wrapped data.

JWS Online at present contains over 80 models, each of them curated
to display the same simulation results as in the original manuscript in which
the model was published. In addition, JWS Online can be used as a simulator
for models from the Biomodels database (see below), via the link http://jjj.
biochem.sun.ac.za/biomodels or http://jjj.bio.vu.nl/biomodels.

An important service provided by JWS Online is the storage of models
under review. Participating journals transfer models contained in articles under
review to the secure area of JWS Online; reviewers are then able to access
and evaluate these models using the simulation. Password control prevents
these models being visible or available to non-designated users. If a model
is accepted for publication, it is moved from the secure area of JWS Online to
the main repository and is available to anyone.
BioModels is a database of published mathematical models of biological systems [6]. BioModels is hosted at the EMBL-EBI (European Molecular Biology Lab/European Bioinformatics Institute) in the U.K. and is a collaborative effort between EMBL-EBI, the SBML team at Caltech (The California Institute of Technology, Pasadena, CA, U.S.A.), DOQCS (Database of Quantitative Cellular Signalling) at the National Center for Biological Sciences (Bangalore, India), the Keck Graduate Institute (Claremont, CA, U.S.A.), the Systems Biology Institute (Shibuya, Tokyo, Japan) and JWS Online at the Universities of Stellenbosch (South Africa), Manchester (U.K.) and the Vrije Universiteit (Amsterdam, The Netherlands).

BioModels contains both curated and uncurated models. BioModels is actively adding new models, and also exchanges models with JWS Online using the SBML format. Models included in the database are checked to be compliant with MIRIAM (Minimum Information Requested In the Annotation of biochemical Models) [7] standards. Over 100 models are included in the curated database.

BioModels focuses on curation of models and annotation of these, providing links to data resources such as publications, databases of compounds and pathways and controlled vocabularies.

From the interface to BioModels, users may browse the curated or non-curated databases, search for models, or submit models to the database.
Models are available for download in a variety of formats. SBML L2v1 is the primary format, whereas other formats which may be available (depending on the nature of the model) are CellML, SciLab (http://www.scilab.org/), XPP (http://www.math.pitt.edu/~bard/xpp/xpp.html) and BioPax (Biological Pathway Exchange; http://www.biopax.org/index.html).

The submission of models may be in SBML or CellML formats, and an on-line validator may be used to check the validity of models submitted using SBML.

Finally, models in the database may be simulated by following a link to the JWS Online simulation and analysis engine. A screenshot of the Biomodels database for Biomodel#144 and its JWS Online interface and simulation result is shown in Figure 2.

Virtual Cell

The Virtual Cell Modeling and Simulation Framework (http://www.vcell.org), released by the National Resource for Cell Analysis and Modeling at the University of Connecticut Health Center, is a platform that facilitates the construction and simulation of models of cellular biological processes [8]. These include biochemical, electrophysiological and transport phenomena.

Both non-spatial models and models which explicitly take into account the spatial distribution of the molecular participants in these processes may be constructed. The former are represented mathematically as sets of ODEs (ordinary differential equations), whereas the latter are represented as sets

![Figure 2. Screenshot of the Biomodels database](image)

The Biomodels database contains 150 curated models (September 2007 release). The interface to Biomodel#144 is shown, together with the JWS Online simulation tool for the model and a simulation result. The screenshot is shown with permission.
of partial differential equations. Both model types are solved numerically. A screenshot of the Virtual Cell user interface and simulation result is shown in Figure 3.

The framework follows a client–server architecture, with the client being available either as an applet in a web browser, or as a free-standing Java application accessed using the Java Web Start technology.

The user interface of the Virtual Cell framework is presented as three components: the BioModel document, the Geometry document and Math document. The first contains a physiological description of the model, whereas the second describes the geometrical details of the model. The Application section of the BioModel document links the model with a particular Geometry. Together, the BioModel and Geometry are used to generate the Math document, which contains a mathematical representation of the model.

Models may be imported from or exported to SBML [4], CellML [9] and Matlab (http://www.mathworks.com). Models may also be retrieved from, or saved to, the server and made available for sharing with others. There is no active effort by the maintainers of the Virtual Cell to add models to the database, and models in the database are not curated.

**WebCell**

WebCell is hosted at KAIST (Korea Advanced Institute of Science and Technology). It provides a simulation environment for models of cellular networks, and may be used for the exploration of their steady-state and dynamic behaviour [10]. WebCell uses a client–server architecture, running Java Servlet pages on the server side and Java applets in a web browser on the
client machine. New models can be created on the client side or models in SBML format may be imported and stored in the database. Models may also be exported and saved in various formats, including SBML or Matlab.

Models stored in the WebCell database are not curated, but are largely models taken from the JWS Online, Biomodels and SBML repositories. The simulation functionality is similar to JWS Online. After selecting the use of an ODE solver or a DAE (differential algebraic equation) solver, time-course simulations on the model can be run. Results can be returned as a plot or as a CSV text file, a screenshot of a simulation result is presented in Figure 4. In addition, structural pathway analyses and metabolic control analyses may be performed.

**JSim**

JSim (http://physiome.org/jsim) is a Java-based simulation environment that can be used for the construction and analysis of quantitative numerical models. It is hosted by the NSR (National Simulation Resource) project at the Department of Bioengineering, University of Washington (Washington, U.S.A.) and is associated with the Physiome Project. JSim may be run as a freestanding executable on the user’s computer, or as a Java applet interfacing to a remote server on which the calculations are performed. In either case a graphical user interface is presented to the user from which models can be imported, configured, run and saved.
JSim uses its own MML (Mathematical Modelling Language) for model specification. Models may be edited in the GUI (graphical user interface) and are compiled before running to check for consistency. The results of model simulations are available as both a graph and text. A screenshot of the user interface and a simulation result is shown in Figure 5.

JSim includes a repository of models, to which models may be submitted.

**CCDB**
The CCDB (Cell Cycle Database) is hosted at the Institute for Biomedical Technologies in Italy (http://www.itb.cnr.it/cellcycle/index.html) [11]. It enables searches to be made of genes and proteins involved in the human and yeast cell cycles. CCDB is also creating a collection of published cell-cycle models based on ODEs. Many models in the collection have associated SBML files describing the model, and a link which enables a simulation of this using XPPAUT. A simulation result is shown in Figure 6.

**DOQCS**
DOQCS is hosted at the NCBS (National Centre for Biological Sciences) and is part of the Tata Institute of Fundamental Research in Bangalore [12]. DOQCS is a repository of models of signalling pathways, including reaction schemes, concentrations, rate constants, as well as annotations on the models. It furthermore provides functions for search, reaction and pathway navigation, and comparison.

The database entries directly represent the entities taking part in the reactions and the reactions themselves, with tables containing molecules, reactions and enzymes. Entries also include identifiers to associate them with particular

---

Figure 5. Screenshot of the JSim initiative
A model of the Biomodels database is loaded in JSim and a simulation result is shown. The screenshot is shown with permission.
pathways. Database reports may be made as a simple listing of parameters, a listing including annotations, as GENESIS scripting language files for use by Kinetikit (http://www.ncbs.res.in/~bhalla/kkit/index.html) to simulate the model, or as differential equations which may be converted into CellML or SBML.

**ModelDB**
ModelDB is a repository for the storage and retrieval of published computational neuroscience models [13]. It is hosted by the SenseLab project at Yale University (New Haven, CT, U.S.A.), and is tightly coupled with NeuronDB, a database of neuronal properties. ModelDB contains over 100 published models. These include models of networks, neurons, synapses, neuromuscular junctions, axons and ion channels. From the web interface, users may browse the database or search for models by model name or author. They may also search for publications, or upload new models.

ModelDB is not associated with any particular modelling language; the models are available in the format in which they were submitted to the database (e.g. Fortran, NEURON).

**SigPath**
SigPath is hosted at the Weill Medical College of Cornell University (Ithaca, NY, U.S.A.), and at the Mount Sinai School of Medicine (Manhattan, NY, U.S.A.) [14].
It is an information management system set up to facilitate the studies of the signalling pathways and networks of the cell. Biochemical information is stored with sufficient detail to enable quantitative models of these systems to be developed. SigPath comprises the SigPath ontology (a structured representation of the information contained in the system), a web-based interface supporting the query, editing, submission and review of information, a mechanism to link reaction information with information about components contained in other databases, and an interface allowing the assembly and export of models.

Models may be exported in the Kinetikit, JSim and SBML formats. Several uncurated models and pathways are available in the database.

**CellML**

CellML is hosted at the University of Auckland, Auckland, New Zealand. It is an XML-based modelling language aimed at describing systems at the cellular level. It includes information on model structure, the mathematics describing the model and metadata.

There is also a repository of CellML models at the host site, containing over 200 models; these include electrophysiological, reaction pathway and qualitative models [9].

**SBML**

SBML is hosted at Caltech (The California Institute of Technology, Pasadena, CA, U.S.A). It is a modelling language used for representing models of biochemical reaction networks and can be applied to, amongst others, metabolic networks, cell signalling pathways and regulatory networks [4].

The components of SBML include species, the entities participating in a reaction, compartments, the bounded volumes in which species interact, reactions, representing the interaction between species, as well as means of representing constraints, parameters and units.

SBML no longer maintains its own repository of models, but has instead transferred curatorship of its models to the BioModels database.

**Conclusions**

Systems biology is a multi-disciplinary field and a successful integration of our knowledge of components for a systemic understanding will be dependent on collaborations between research groups and on good software tools. There are many resources available to systems biologists, including genomics, proteomics, metabolic pathway and signal transduction databases. In this chapter we have described a number of kinetic model databases and focused on our own research interests and projects, JWS Online and its collaborations with numerous other research groups. Clearly, the initiatives we have described do not form a complete list and many more resources are available; a few systems biology websites with a large number of links are:
Often data sources for systems biology projects are stored in different databases and integration of such data sets can involve many steps in a workflow. The development of Webservices has made it possible to automate such workflows and several of the database-initiatives described above are accessible for such services. Taverna is a workflow execution and development environment that is particularly easy to use and for which many workflows have been written for the systems biology community [15]. Some of the many advantages of workflows include the ease of automation, the ability to rapidly interoperate services and the explicit definition of used methodologies, which are all-important in systems biology projects and will contribute to webservices and environments such as Taverna becoming indispensable tools.

Summary

- **Kinetic models are important tools in systems biology projects.**
- **A number of initiatives have been started to store kinetic models and make them available via web repositories.**
- **Several of these initiatives and their functionalities are described in the present chapter.**
- **For systems biology projects to be successful, a good integration of experimental data and kinetic models is essential and good software tools will play an important role in this.**

C.v.G. and the JWS Online project are funded by the South African National Bioinformatics Network.

References


© The Authors Journal compilation © 2008 Biochemical Society