BRIEFINGS IN FUNCTIONAL GENOMICS AND PROTEOMICS. VOL 6. NO 3. 240–251 doi:10.1093/bfgp/elm027

Resources, standards and tools for systems biology

Christoph Wierling, Ralf Herwig and Hans Lehrach

Advance Access publication date 17 October 2007

Abstract

Modelling and simulation techniques are valuable tools for the understanding of complex biological systems. The design of a computer model necessarily has many diverse inputs, such as information on the model topology, reaction kinetics and experimental data, derived either from the literature, databases or direct experimental investigation. In this review, we describe different data resources, standards and modelling and simulation tools that are relevant to integrative systems biology.

Keywords: *Bioinformatics; systems biology; pathway databases; systems biology standards; modelling and simulation; modelling tools*

INTRODUCTION

For a long time, research in molecular biology has been focused on the analysis of relevant components of the cellular network (proteins, metabolites) in isolation. By this approach, thousands of genes have successfully been characterized and functionally annotated. But biological systems are complex and their characteristics are the result of a highly interwoven interaction network developing through time and space. Fundamental characteristics of living systems, such as the assimilation of nutrients, growth and reproduction, the perception of (environmental) signals and its processing can be narrowed down ultimately to the single unit that all living things are composed of: the cell. Thus, the understanding of the characteristics of cellular systems is essential, but this requires an approach that takes into account both interactions at the molecular level as well as physiological functions that are characteristics of the whole organism. In particular, in the light of understanding multigenic and complex diseases that cannot be pinned down to a single gene or component, systems approaches become increasingly important.

Systems biology explanations of physiology and disease should be multi-level; from molecular pathways and regulatory networks, through cells and organs, ultimately to the level of the whole organism. With the use of computer models for such processes, *insilico* predictions can be generated on the state of the disease or the effect of the individual therapy [1]. Models are partial representations and their aim is to explain which features of a system are necessary and sufficient to understand it [2]. The performance of a model is mainly defined by its predictive power. What is predicted depends on the task and of course is general.

Systems biology is going to revolutionize our knowledge of disease mechanisms and the interpretation of data from high-throughput technologies. Biological systems can be studied by (i) investigating the components of cellular networks and their interactions, (ii) applying experimental high-throughput and whole-genome techniques and (iii) integrating computational methods with experimental efforts [3]. This approach requires an integration of experimental and computational methods [4] and, thus,

Corresponding author. Christoph Wierling, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, 14195 Berlin, Germany. Tel: +49 30 8413 1272; Fax: +49 30 8413 1380; E-mail: wierling@molgen.mpg.de

Christoph Wierling is a Computational Biologist at the Max-Planck-Institute for Molecular Genetics. His research interests focus on modelling of biological systems and development of computational tools for systems biology.

Ralf Herwig has been working as group leader in Bioinformatics at the Max-Planck-Institute for Molecular Genetics since 2001 and works on several projects covering genomics, proteomics and systems biology.

Hans Lehrach is the Director of the Max-Planck-Institute for Molecular Genetics and was chairman of the German Human Genome Project. Research interests focus on functional genomics, technology development and systems biology.

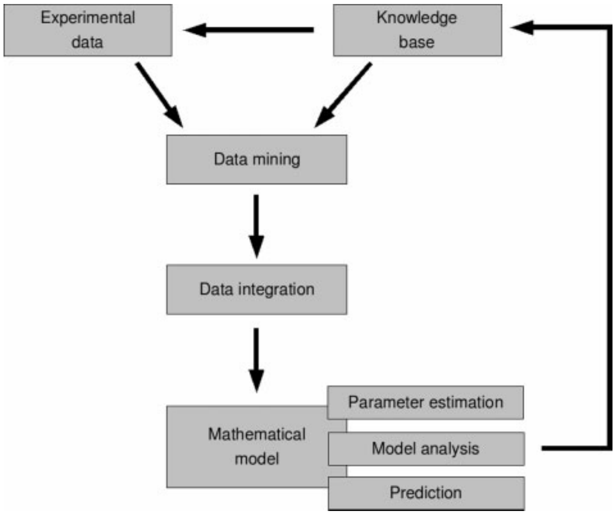


Figure 1: Modelling biological systems.

an iterative, integrated process of data mining and data gathering (e.g. from scientific literature, databases and experiments), data integration, computational modelling and analysis and finally validation of specific observations that were not explicit beforehand (Figure 1).

Using data mining steps, one agglomerates sufficient details for the generation of model prototypes of the biological system under investigation. Finally, using analysis methods, the mathematical model is refined, cross-validated with regard to internal and external features (for example using parameter estimation and validation methods) and being used for the formulation of new hypotheses that in turn are subject to further experimental investigation.

Systems biology methodology and approaches have evolved rapidly over the past few years, driven by emerging new high-throughput technologies. The most important boost was given by the large sequencing projects such as the human genome project that resulted in the full sequence of the human and other genomes [5, 6]. This knowledge builds the theoretical basis to compute gene regulatory motifs, to determine the exon–intron structure of genes and to derive the coding sequence of potentially all genes through many organisms. From the exact sequences, probes for whole genome DNA arrays have been constructed that allow monitoring of the transcriptome of a given cell- or tissue-type. Proteomics technologies have been used to identify translation status on a large scale (2D-gels, mass spectrometry). Protein–protein interaction data involving thousands of components have been measured to determine information on the proteome [7]. Multiple databases exist, a variety of experimental techniques have produced gene and proteome expression data from various tissues and samples and important disease-relevant pathways have been investigated. Information on promoter regions and transcription factors is available for many genes as well as sequence information. This information—although extremely helpful—cannot be utilized efficiently, because of the lack of integrative analysis tools.

To validate such data in the system-wide hierarchical context ranging from DNA to RNA to protein to interaction networks and further on to cells, one needs to correlate and integrate such information. Thus, an important part of systems biology is data integration.

In this work, we review key genomic and computational resources available for systems biology approaches. The first part of our review is devoted to databases and public repositories that store functional high-throughput data, for example, gene expression data, sequences, disease information, annotation of protein function and biological pathways. Creating a fundamental knowledgebase is the first essential step in the development of computer models for cellular processes and these data repositories can be used in order to derive the ‘parts list’ of the biological process under study. By combining these data and by adding functional relationships they deliver an initial topology (or reaction network) that is the basis for dynamic modelling.

In the second part of this review, we describe some important standards for exchanging relevant data. Experimental investigation in functional genomics and proteomics, e.g. of disease processes, typically involves several steps of experimental testing using heterogeneous data techniques. The integration of those data requires common schema for data storage, data representation and data transfer. For particular experimental techniques (e.g. in transcriptome and proteome research), such schema has already been established. On a more complex level, schema has also been defined for biological models and pathways such as SBML [8, 9], CellML [10] and BioPAX. Most of these repositories use an XML-based language style.

In the third part of the review, we introduce state-of-the-art tools for dynamic computational modelling. For dynamical modelling, common approaches are based on systems of ordinary differential equations

Table I: Selected data resources and databases for systems biology

Data resource	URL	References
Ontology		
GO	http://www.geneontology.org/	[71]
Pathway databases		
KEGG	http://www.genome.jp/kegg/	[31]
Reactome	http://www.reactome.org/	[32]
BioCyc (including EcoCyc, MetaCyc, HumanCyc)	http://www.biocyc.org/	[33]
Pathway Interaction Database (PID)	http://pid.nci.nih.gov/	
BioCarta	http://www.biocarta.com/	
Spike	http://www.cs.tau.ac.il/~spike/	
IntAct	http://www.ebi.ac.uk/intact/	[40]
Database of interacting proteins (DIP)	http://dip.doe-mbi.ucla.edu/	[41]
Kinetics databases		
BRENDA	http://www.brenda.uni-koeln.de/	[46]
SABIO-RK	http://sabio.villa-bosch.de/SABIORK/	[47]
Expression data resources		
Gene Expression Omnibus (GEO)	http://www.ncbi.nlm.nih.gov/projects/geo/	[27, 28]
ArrayExpress	http://www.ebi.ac.uk/arrayexpress/index.html	[29]
Disease specific databases		
OMIM	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM	
Systems biology model repositories		
BioModels	http://www.biomodels.org/	[48]
JWS	http://jjj.biochem.sun.ac.za/	[49]

(ODEs) that describe biochemical reaction networks. Other possible deterministic approaches are e.g. neural networks, cellular automatons, Boolean or Petri nets [11, 12]. Computer tools allow the analysis of the dynamic behaviour of the reaction networks with the given model parameters. It is a very important feature of such systems to allow the estimation of these parameters from experimental data and the analysis of the behaviour of the system with respect to changes of these parameters. We give an overview on the features of several tools and highlight results from the PyBioS system developed in our laboratory.

DATA RESOURCES FOR SYSTEMS BIOLOGY

The development of mathematical models of cellular systems requires a lot of information on different aspects of the system. Data typically arises from several levels of cellular information quantified by different functional genomics technologies such as DNA, RNA or protein sequence data, gene expression data from array experiments, abundance data of proteins and metabolites from diverse experimental techniques (e.g. mass spectrometry, 2D-gels, blots), information on protein–protein interactions or protein modifications or kinetics of enzyme activities or binding affinities, among others. The most important resource for such information is the scientific literature and human expertise curated in public databases. In particular, for the development of mathematical models standardized resources that provide their data in a computational amenable and reusable manner are a preferable resource. Table 1 gives a brief list of some important databases. A large compilation of relevant database resources is given in [13]. Moreover, *Nucleic Acids Research* offers a yearly database issue in January, providing a broad overview of diverse databases.

Primary data resources

The National Center for Biotechnology Information (NCBI) [14] and the European Bioinformatics Institute (EMBL-EBI) [15] provide several databases that are widely used in biological research offering information about nucleotide sequences, proteins, genes, molecular structures and gene expression. Among the nucleotide sequence databases, the Genetic Sequence database (GenBank), the Reference Sequence Database (RefSeq) and UniGene can be found at the NCBI. Related databases at the EMBL-EBI are the EMBL Nucleotide Database or the Ensembl automatic genome annotation database. The Ensembl project is developing and maintaining a system for the management and presentation of genomic sequences and annotation for eukaryotic genomes [16–19]. Similarly to nucleotide sequence

Table 2: Numbers of overlapping reactions/interactions from different pathway databases that can be mapped to each other in respect of identical substrates and products

	Reactome	KEGG	HumanCyc	PID	Biocarta	Intact	Dip	Spike
Reactome	12 042							
KEGG	209	1498						
HumanCyc	93	199	1077					
PID	8	0	0	1064				
Biocarta	62	0	1	114	2160			
Intact	78	0	1	0	42	5690		
Dip	15	0	2	0	25	114	1152	
Spike	55	0	0	50	125	976	114	11 181

databases, Swiss-Prot, TrEMBL [20] and UniProt [21], provide information on protein sequences and annotations. Moreover, there are databases for protein families, domains and functional groups such as InterPro [22, 23] or those with a focus on protein structures like Protein Data Bank (PDB) [24]. Since a few years, also non-translated RNAs and microRNAs revealed to be highly important in the control of cellular systems and gave rise to the implementation of related databases, like RNAdb [25] or miRBase [26], with the objective of gathering current information. Microarray data provide a valuable resource in the interpretation of the transcriptome levels of genes. Large repositories store these data from multiple studies such as the Gene Expression Omnibus (GEO) [27, 28] at NCBI and the ArrayExpress [29] at EMBL-EBI. These databases provide free distribution and shared access to comprehensive gene expression datasets. Data include single and multiple channel microarray-based experiments measuring the abundance of mRNA, genomic DNA and protein molecules. Data from non-array-based high-throughput functional genomics and proteomics technologies are also archived, including SAGE and mass spectrometry peptide profiling.

Pathway and interaction databases

Pathway databases are particularly interesting for modelling approaches, since they offer a straightforward way of building network topologies by the annotated reaction systems [30]. These databases provide integrated representations of functional knowledge of the different components of a biological system and constitute a basis for the topology of mathematical models. The databases Kyoto Encyclopedia of Genes and Genomes (KEGG) [31], Reactome [32] and BioCyc [33] contain metabolic reactions and several signal transduction pathways. KEGG is a reference knowledgebase offering information about genes and proteins, biochemical compounds, reactions and pathways. It provides 317 reference pathways that are linked to genes and reactions of 38 eukaryotes and many microorganisms. It can be accessed via the web, FTP and web services. Reactome [32] is managed as a collaboration of the Cold Spring Harbor Laboratory, the EBI and the Gene Ontology Consortium. It uses a very precise specification (ontology) of components and interactions that comprises details on stoichiometry, localization, references to external databases, etc. This covers also processes like complex formation events or translocations of molecules. A further pathway database with a similar scope is BioCyc [33] that covers pathway data on *Escherichia coli* (EcoCyc), and predicted metabolic pathways of other microorganisms (MetaCyc) and human (HumanCyc). Databases with a specific focus on signalling events are BioCarta [34], Spike [35], Transpath [36], STKE [37], NetPath [38] and the Pathway Interaction Database (PID) [39]. An inherent aspect of the pathway concept is protein–protein interaction subject of the databases IntAct [40] or database of interacting proteins (DIP) [41]. Gene regulation processes and gene regulatory networks are not yet covered in as much detail as metabolic processes or signalling. However, there are databases that store information on transcription factor binding sites such as RegulonDB [42], TRED [43] and Transfac [44]. The lack of uniform data models and data access methods of the existing almost 224 interactions and pathway databases make data integration very difficult [30, 45]. Table 2 illustrates the overlap of several of these pathway resources in human.

Besides topological information about cellular reaction networks, kinetic data, such as kinetic laws and kinetic constants, are of particular interest for the generation of mathematical models. Two databases that are concerned with such data are BRENDA [46] and SABIO-RK [47].

Mathematical models of a biochemical reaction system have been made available to the scientific community in form of publications often depicting a diagram of the reaction system or a list of the reaction equations, along with a mathematical description (e.g. as a differential equation system), and lists of kinetic parameters and concentrations of specific states. Recently, model databases have been installed such as the BioModels database [48] or JWS [49]. Both are public, centralized databases of curated, published, quantitative kinetic models of biochemical and cellular systems. For instance, the BioModels database currently provides 87 curated and 40 non-curated models.

Theme-specific databases

Whereas most of the above-mentioned databases are fairly general, there exist multiple databases with a specific focus. For instance, there are databases that are focused on a certain species, for example MGD for mouse [50], Flybase for *Drosophila melanogaster* [51], wormbase for *Caenorhabditis elegans* [52] or SGD for yeast [53], or they contain information on specific diseases, such as cancer (e.g. COSMIC [54]) and diabetes (e.g. T1DBase [55]), or they contain information on a specific subject such as chemical compounds found in biological systems (ChEBI [56], the Human metabolome database [57], PRIDE [58], LipidMaps [59], the Human serum metabolome project [60]).

Mining literature for systems biology

Finally, the integration of literature information is highly important. Literature is accessed in a derived form such as the concepts represented by the Medical Subject Headings (MeSH) and Gene Ontologies (GO). A further approach that is recently applied for building systems biology resources is text mining [61]. Text mining can either be used for pre-selection of appropriate literature or the automatic extraction of data from literature. In particular, systems biology can benefit significantly from the extraction of data on molecular interactions of cellular components and related information about the kinetics of the interactions [62]. However, text mining of scientific literature is still in its early phase and the precision of its results, as given by false-positive and false-negative rates, has to be improved. For further review on literature mining see [63–65].

STANDARDS USED IN SYSTEMS BIOLOGY

An important part of systems biology is data integration. Although data integration itself cannot explain the dynamical behaviour of biological systems, it is useful for increasing the information content of the individual experimental observation, enhancing the quality of the data and identifying relevant components in the model. On the basic level of complexity data integration consists of the integration of heterogeneous data resources and databases with the aim of parsing data from these databases, to query for information and to make it usable for modelling. Technically, database integration requires the definition of data-exchange protocols and languages and the development of parsers that interconnect the databases to a data layer that is able to display the heterogeneous data sources in a unified way.

A standard for representation, storage and exchange of data is a convention about the information items necessary to describe the experiment and the encoding of this information (e.g. expression data of microarray experiments or information about the relations between components and interactions of a pathway). The standard has to enable an unambiguous transfer and interpretation of the data and information. Developing a standard involves four steps: an informal design of a conceptual model, a formalization, the development of a data exchange format and the implementation of supporting tools [66].

Conceptual design

The first step, the conceptual model design, gives an informal description of the related domain and specifies its delimitation. The description should address the minimal number of most informative parameters but should still provide a common ground for all related applications [66]. For instance, for the microarray domain a conceptualization is provided by Minimum Information about a Microarray Experiment (MIAME) [67] and Minimum

Table 3: Examples of XML-based standards used in systems biology

Modelling system	Description	Application area	References
BioPAX	Biological Pathways Exchange	Description and exchange of biological pathway data	[74]
SBML	Systems Biology Markup Language	Describing and exchange of mathematical models	[8, 9]
SBGN	Systems Biology Graphical Notation	Visual notation for computational models of biological systems	[96]
CellML	Cell Markup Language	Describing and exchange of mathematical models	[10]
PSI-MI	Proteomics Standards Initiative Molecular Interaction	Description and exchange of protein-protein interactions	[97]

Information about a Proteomics Experiment (MIAPE) [68] which gives guidelines for the standardized collection, integration, storage and dissemination of proteomics data. Like specifications for experimental data also concepts for the description of mathematical models such as Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM) [69] have been elaborated.

Data representation formalisms and languages

The description of a given domain can be represented in any format, but the use of common representation formalisms and languages makes it easier to compare and interpret data from similar domains and it facilitates the integration, computational processing and comprehensive interpretation of that data. Controlled vocabularies are a prerequisite for a consistent data description. They contain sets of words or phrases representing particular entities, processes or abstract concepts [70]. Within a particular controlled vocabulary, individual terms are usually associated with a unique identifier, an unambiguous definition and occasionally also synonyms to prevent misinterpretations.

Furthermore, ontologies are used for conceptualization of a knowledge domain. An Ontology defines terms and relations along with a vocabulary of a topic area and thus, provides a common terminology over a certain domain. Relations are, for example, 'is-a' relations that describe a generalization, forming a term hierarchy. An example is the GO that builds the basis for a generalized functional annotation of genes and their products. The naming of genes and gene products is not necessarily systematic and genes having identical functions are given different names in different organisms or the verbal description of location and function might be different. To address this problem the GO was initiated as a collaborative effort [71]. GO terms have a parent–child relationship. GO defines three top-level categories, 'molecular function', 'biological process' and 'cellular component' and organizes all keywords in a hierarchical graph-like structure. The terms defined in GO form a directed acyclic graph. The power of the GO project lies in the fact that many applications have been developed that use GO terms to validate other data for functional information.

Data exchange formats

During the last years, the eXtensible Markup Language (XML) [72] has been proofed to be a flexible tool for the definition of standard formats not only for applications in different fields of information technology, but also for the management of data from diverse experimental platforms. One example designed for data from microarray experiments is MAGE-ML [73]. Others are, for instance, those dealing with pathway data and mathematical models. Table 3 gives an overview of some XML-based data exchange formats used in systems biology. SBML [8, 9], CellML [10] and BioPAX [74] have the potential to become *de facto* standards for their respective application area.

BioPAX is defined by the BioPAX working group and is designed for handling information on pathways and topologies of biochemical reaction networks. The Systems Biology Markup Language (SBML) is a format for 'describing models common to research in many areas of computational biology, including cell-signalling pathways, metabolic pathways, gene regulation and others' [8, 9]. Major releases of the SBML standard are called levels, where level 2 is the most recent. SBML defines list of species (entities of the model), compartments, parameters and reactions, among others. SBML is widely used—it is supported by over 110 software systems.

A comparison of SBML and BioPAX comes to the conclusion that, while the main structures of these formats are similar, SBML is tuned towards simulation models of molecular pathways. BioPAX turns out to be the most general and expressive format [75], even if it is lacking definitions for the representation of dynamic data such as kinetic laws and parameters.

It is argued that the syntactic and document-centric XML cannot achieve the level of interoperability required by the highly dynamic and integrated bioinformatics applications. Therefore, semantic web technology like resource-description framework (RDF) and the web ontology language (OWL) have been proposed as alternatives to current XML technology [76].

Using standards brings several advantages, e.g. an ontology along with a defined vocabulary is used that promotes an accurate description of the data and it provides a software-independent common representation of the data. One of the most important general problems in building standards for biology is that our understanding of living systems is not static but rather constantly developing what necessitates a regular update of these standards [66].

MODELLING TOOLS

Annotation tools

The first step in setting up a model is by summarizing in the computer, all reactions, interactions and processes that are relevant to the model, either as a list of reactions or as a diagram that is describing those processes and depicts the network structure. There are several software tools that can be used for this purpose, for example JDesigner [77] or the graphical and user-friendly interface of the CellDesigner software [78]. Another software that has a sophisticated graphical user interface and supports BioPAX for model exchange is Cytoscape [79]. A more advanced but comprehensive tool for annotation is the Reactome Curator Tool that can take advantage of the already existing data provided by a local copy of the Reactome MySQL database. Reaction systems designed by this tool have to be converted into appropriate interchange formats, like SBML or BioPAX.

Table 4: Selected modelling and simulation tools for systems biology

Modelling system	References
General Modelling Tools	
GEPASI	[82–84]
CellDesigner	[78]
E-Cell	[86, 87]
ProMoT/Diva	[88]
Virtual Cell	[89–91]
Systems Biology Workbench (SBW) and its add-ons	[92]
COPASI	[85]
PyBioS	[3, 95]
Model Visualization Tools	
BioTapestry	[98]
Cytoscape	[99]
VisANT	[100]

Modelling tools

Once the model topology is designed, a mathematical model can be created. If this is, for example, a kinetic model, further data on the kinetic laws and kinetic parameters has to be identified or appropriate assumptions have to be made. For this purpose, diverse software tools have been developed. One can use commercial tools like Mathematica or Matlab that are well elaborated and offer broad spectra of functionalities. One disadvantage of using these programs is that the differential equation system of the mathematical model has to be formulated explicitly by the user. Overviews of current software platforms and projects that face up to this as well as an overview about computational requirements for this purpose is given in [3, 80, 81]. Common systems among others—for this purpose are Gepasi [82–84], COPASI [85], E-Cell [86, 87], ProMoT/Diva [88], Virtual Cell [89–91] or the Systems Biology Workbench (SBW) and its add-ons [92]. Table 4 summarizes some modelling and visualization tools for systems biology. A comprehensive list of modelling and simulation tools is also given in [93] that reports the results of an online survey of systems biology standards. This report identified CellDesigner [78] as the most popular stand-alone application in respect to its graphical functionalities.

Gepasi and COPASI come up with user-friendly interfaces for the simulation and analysis of biochemical systems. They support the definition of compartments. Common kinetic types as well as user-defined kinetic types are available. They provide time-course simulation and steady-state calculation and the ability to explore the behaviour of the model over a wide range of parameter values using a parameter scan that runs one simulation for each parameter combination. Gepasi and COPASI can characterize steady states using metabolic control analysis (MCA) and linear stability analysis and they are capable of doing parameter estimation with experimental data and optimization.

E-Cell is based on the modelling theory of the object-oriented Substance–Reactor Model. Models are constructed with three object classes, substance, reactor and system. Substances represent state-variables, reactors describe operations on state variables and systems represent logical or physical compartments. Time-course calculation is done by the use of a simulation engine. Numerical integration is supported by first-order Euler or fourth-order Runge–Kutta method.

ProMoT/Diva consists of the modelling tool ProMoT and the simulation environment Diva. The workbench deals with modular models and can handle Differential Algebraic Equation (DAE) systems. Modelling is supported with a graphical user interface and a modelling language. The modelling tool provides the possibility to use existing modelling entities out of knowledge bases.

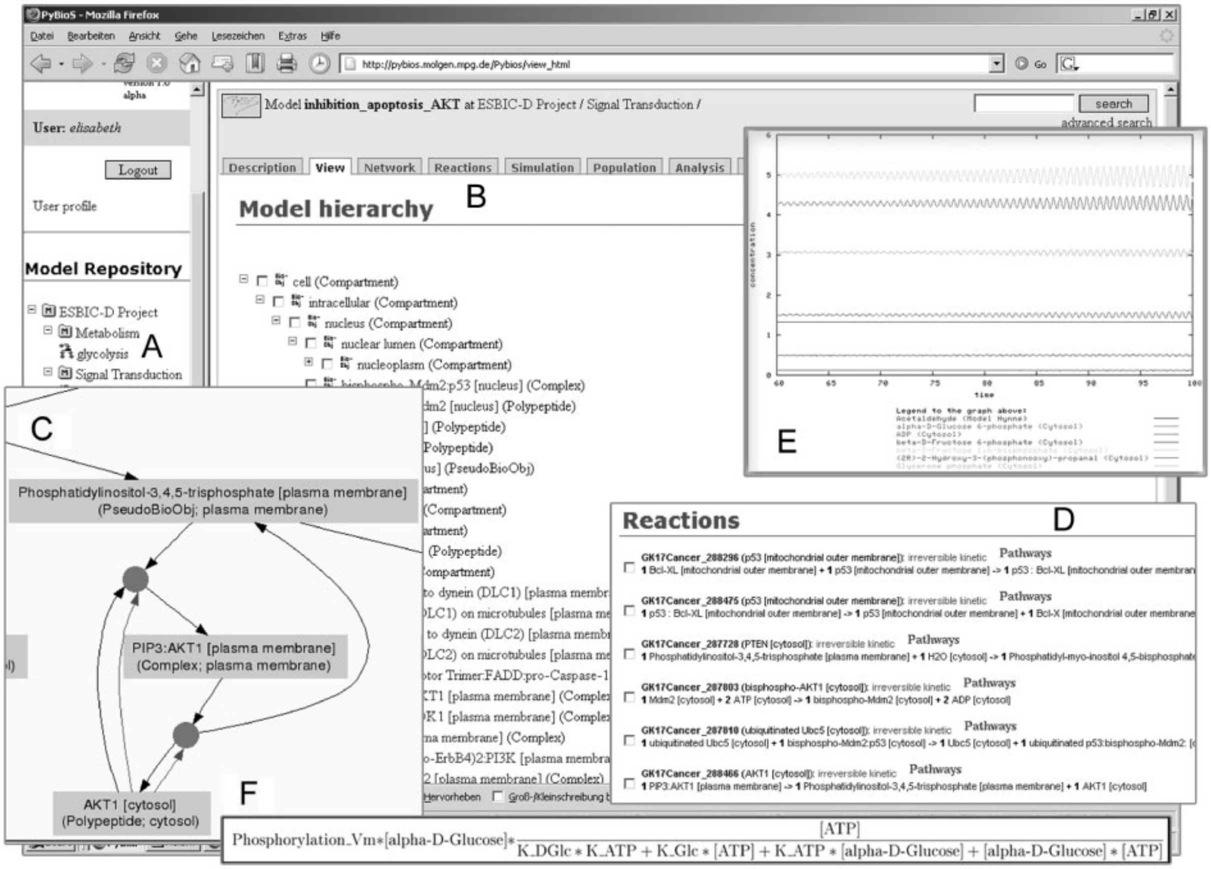
The Virtual Cell is a web-based client-server architecture with central databases of user models. It provides a formal framework for modelling biochemical, electrophysiological and transport phenomena while considering the sub-cellular localization of the molecules that take part in them [91].

The SBW provides a server that acts as a broker between different modelling and simulation tools (clients) via a common interface. These clients (add-ons) cover graphical tools for model population, deterministic and stochastic simulators and analysis tools like the integration of MetaTool [94]. Closely related to the SBW is the development of SBML that is used for communication by SBW.

A modelling and simulation platform for systems biology that is developed in our laboratory is PyBioS [3, 95]. PyBioS has a web-based user interface and provides functionalities for the model development, simulation and analysis (Figure 2). Compared to other systems biology modelling and simulation tools PyBioS provides interfaces to external pathway data resources that can be searched and directly be used for the generation of the model topology in mind. The system can handle large ODEs with thousands of reactions. Furthermore, PyBioS provides an interface for the upload of experimental data that can directly be used as initial values for model components such as mRNA, protein, metabolite or enzyme abundances. Appropriate kinetics can be chosen from a repository of pre-defined kinetic laws. From this information—the model topology and the reaction kinetics along with the respective parameters—PyBioS automatically creates a mathematical model that can be used subsequently for simulation and analysis. Besides plotting graphs of time-course data (concentrations or fluxes versus time), PyBioS also provides a visualization of the interaction network graph comprising coloured nodes according to simulation results. The latter is very helpful for the interpretation of simulation results with respect to the network structure. For the creation of a first model prototype, PyBioS has an interface to external pathway databases, such as Reactome, that can be used for the automatic generation of a model based on the respective network topology. Via the web-interface one can search for specific reactions or pathways and by this enrich the model automatically with reactions from pathway databases (Figure 3).

SUMMARY AND FUTURE NEEDS

The new functional genomics techniques will elevate our knowledge on biological networks to a great extent. Systems biology will play a key role in future research in the interpretation of such data. The information that we can gain about a biological system (for example a disease process) appears in practice as an experimental observation, and research is restricted to the granularity and the precision of the experimental techniques in use. It is very likely that the range of experimental granularity will increase in the next years utilizing heterogeneous techniques that target a biological question of interest at different points so that data integration


PyBioS - Mozilla Firefox
Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe
http://pybios.molgen.mpg.de/Pybios/view_html
version 1.0 alpha
User: *elisabeth*
Logout
User profile
Model Repository
ESBIC-D Project
Metabolism
glycolysis
A
Signal Transduction
Model **inhibition_apoptosis_AKT** at ESBIC-D Project / Signal Transduction /
search
advanced search
Description
View
Network
Reactions
Simulation
Population
Analysis
Model hierarchy
B
cell (Compartment)
intracellular (Compartment)
nucleus (Compartment)
nuclear lumen (Compartment)
nucleoplasm (Compartment)
bisphospho-Mdm2:p53 [nucleus] (Complex)
dm2 [nucleus] (Polypeptide)
] (Polypeptide)
Polypeptide)
us] (PseudoBioObj)
artment)
(Compartment)
artment)
(Polypeptide)
ompartment)
to dynein (DLC1) [plasma membr
DLC1) on microtubules [plasma me
to dynein (DLC2) [plasma membr
DLC2) on microtubules [plasma me
ptor Trimer:FADD:pro-Caspase-1
T1 [plasma membrane] (Complex
K1 [plasma membrane] (Complex
na membrane] (Complex)
o-ErbB4)2:PI3K [plasma membra
2 [plasma membrane] (Complex)
Hervorheben
Groß-/Kleinschreibung b
C
Phosphatidylinositol-3,4,5-trisphosphate [plasma membrane] (PseudoBioObj; plasma membrane)
PIP3:AKT1 [plasma membrane] (Complex; plasma membrane)
AKT1 [cytosol] (Polypeptide; cytosol)
F
concentration
0
1
2
3
4
5
6
60
65
70
75
80
85
90
95
100
time
Legend to the graph above:
Acetaldehyde (Model Hynne)
alpha-D-Glucose 6-phosphate (Cytosol)
ADP (Cytosol)
beta-D-Fructose 6-phosphate (Cytosol)
beta-D-Fructose 1,6-bisphosphate (Cytosol)
(2R)-2-Hydroxy-3-(phosphonooxy)-propanal (Cytosol)
Glycerone phosphate (Cytosol)
E
Reactions
D
Pathways

☐ **GK17Cancer_288296** (p53 [mitochondrial outer membrane]): irreversible kinetic
1 Bcl-XL [mitochondrial outer membrane] + **1** p53 [mitochondrial outer membrane] -> **1** p53 : Bcl-XL [mitochondrial outer membran

☐ **GK17Cancer_288475** (p53 [mitochondrial outer membrane]): irreversible kinetic
1 p53 : Bcl-XL [mitochondrial outer membrane] -> **1** p53 [mitochondrial outer membrane] + **1** Bcl-X [mitochondrial outer membrane

☐ **GK17Cancer_287728** (PTEN [cytosol]): irreversible kinetic
1 Phosphatidylinositol-3,4,5-trisphosphate [plasma membrane] + **1** H2O [cytosol] -> **1** Phosphatidyl-myo-inositol 4,5-bisphosphate

☐ **GK17Cancer_287803** (bisphospho-AKT1 [cytosol]): irreversible kinetic
1 Mdm2 [cytosol] + **2** ATP [cytosol] -> **1** bisphospho-Mdm2 [cytosol] + **2** ADP [cytosol]

☐ **GK17Cancer_287810** (ubiquitinated Ubc5 [cytosol]): irreversible kinetic
1 ubiquitinated Ubc5 [cytosol] + **1** bisphospho-Mdm2:p53 [cytosol] -> **1** Ubc5 [cytosol] + **1** ubiquitinated p53:bisphospho-Mdm2: [c

☐ **GK17Cancer_288466** (AKT1 [cytosol]): irreversible kinetic
1 PIP3:AKT1 [plasma membrane] -> **1** Phosphatidylinositol-3,4,5-trisphosphate [plasma membrane] + **1** AKT1 [cytosol]

$$\mathrm{Phosphorylation\_Vm} * [\text{alpha-D-Glucose}] * \frac{[\mathrm{ATP}]}{\mathrm{K\_DGlc} * \mathrm{K\_ATP} + \mathrm{K\_Glc} * [\mathrm{ATP}] + \mathrm{K\_ATP} * [\text{alpha-D-Glucose}] + [\text{alpha-D-Glucose}] * [\mathrm{ATP}]}$$

Figure 2: The PyBioS simulation environment. A particular model can be selected from the model repository (**A**) and its hierarchical model structure can be inspected (**B**). A graphical representation of the model is provided by an automatically generated network diagram (**C**). An overview of all reactions of a model is given by an appropriate listing (**D**). Simulation is based on an automatically generated mathematical model that is derived from its corresponding object-oriented model. Simulation results can be depicted either as graphs of the concentration time course data (**E**) or as coloured nodes in the network graph. Reaction kinetics can be displayed in a user friendly manner (**F**).

Key Points

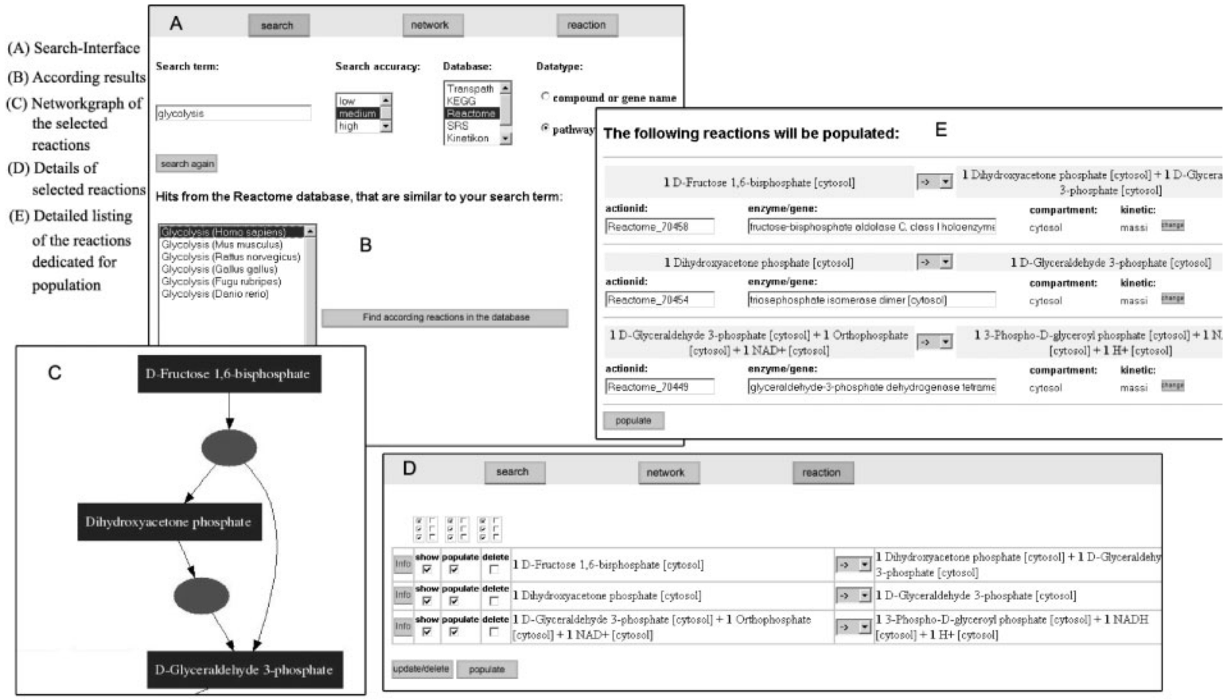
- Studying cellular processes involves large amounts of heterogeneous data.
- Systems biology approaches try to assemble these data in a unified way and combine it with methods from computational modelling and bioinformatics.
- Systems biology approaches try to develop computer models for the cellular processes, trained with the experimental data that are able to reproduce fundamental features of these processes.
- Systems biology is in its early stages and needs data integration and standardization.

remains a major challenge of future biomedical research.

In the case of complex disease conditions, it is clear that such integrated approaches are required in order to link clinical, genetic, behavioural and environmental data with diverse types of molecular phenotype information and to identify correlative associations. Such correlations, if found, are the key to identifying biomarkers and processes that are either causative or indicative of the disease.

Acknowledgements

Authors thank Atanas Kamburov for providing the statistics in Table 2, Elisabeth Maschke-Dutz for providing Figures 2 and 3 and Hendrik Hache and the anonymous reviewers for critical proof-reading of the article and their helpful comments. This work was supported by the EU FP6 grants LSHG-CT-2003-503269 and LSHB-CT-2006-037712 and the Max Planck Society.


(A) Search-Interface
(B) According results
(C) Networkgraph of the selected reactions
(D) Details of selected reactions
(E) Detailed listing of the reactions dedicated for population
A
search
network
reaction
B
C
D-Fructose 1,6-bisphosphate
Dihydroxyacetone phosphate
D-Glyceraldehyde 3-phosphate
D
E
search again
Find according reactions in the database
populate
update/delete
Info
change

Search term: glycolysis

Search accuracy: low / medium / high (medium selected)

Database: Transpath / KEGG / Reactome / SRS / Kinetikon (Reactome selected)

Datatype: ○ compound or gene name ● pathway

Hits from the Reactome database, that are similar to your search term:

- Glycolysis (Homo sapiens)
- Glycolysis (Mus musculus)
- Glycolysis (Rattus norvegicus)
- Glycolysis (Gallus gallus)
- Glycolysis (Fugu rubripes)
- Glycolysis (Danio rerio)

The following reactions will be populated:

1 D-Fructose 1,6-bisphosphate [cytosol]	->	1 Dihydroxyacetone phosphate [cytosol] + 1 D-Glycera 3-phosphate [cytosol]
actionid: Reactome_70458	enzyme/gene: fructose-bisphosphate aldolase C, class I holoenzyme	compartment: cytosol; kinetic: massi
1 Dihydroxyacetone phosphate [cytosol]	->	1 D-Glyceraldehyde 3-phosphate [cytosol]
actionid: Reactome_70454	enzyme/gene: triosephosphate isomerase dimer [cytosol]	compartment: cytosol; kinetic: massi
1 D-Glyceraldehyde 3-phosphate [cytosol] + 1 Orthophosphate [cytosol] + 1 NAD+ [cytosol]	->	1 3-Phospho-D-glyceroyl phosphate [cytosol] + 1 N [cytosol] + 1 H+ [cytosol]
actionid: Reactome_70449	enzyme/gene: glyceraldehyde-3-phosphate dehydrogenase tetrame	compartment: cytosol; kinetic: massi

	show	populate	delete			
Info	☑	☑	□	1 D-Fructose 1,6-bisphosphate [cytosol]	->	1 Dihydroxyacetone phosphate [cytosol] + 1 D-Glyceraldehy 3-phosphate [cytosol]
Info	☑	☑	□	1 Dihydroxyacetone phosphate [cytosol]	->	1 D-Glyceraldehyde 3-phosphate [cytosol]
Info	☑	☑	□	1 D-Glyceraldehyde 3-phosphate [cytosol] + 1 Orthophosphate [cytosol] + 1 NAD+ [cytosol]	->	1 3-Phospho-D-glyceroyl phosphate [cytosol] + 1 NADH [cytosol] + 1 H+ [cytosol]

Figure 3: The PyBioS modelling and simulation system provides an interface to pathway databases, such as Reactome, that can be used for the automatic generation of a model based on the network topology obtained from the respective pathway database. Via the web-interface one can search for specific reactions or pathways and by this enrich the model automatically with reactions from pathway databases. For instance, via the search interface (**A**) one can look for specific pathways, genes, compounds, etc. (**B**), build a network graph (**C**), inspect details of each selected reaction (**D**), and choose appropriate kinetic laws from a kinetics repository (**E**) and finally generate of those reactions a model in PyBioS that can subsequently be used for simulation and further analysis.

References

1. Herwig R, Lehrach H. Expression profiling of drug response – from genes to pathways. *Dialogues Clin Neurosci* 2006;**8**:283–93.
2. Noble D. The rise of computational biology. *Nat Rev* 2002; **3**:460–63.
3. Klipp E, Herwig R, Kowald A, *et al*. *Systems Biology in Practice. Concepts, Implementation and Application*. 2005. Weinheim: WILEY-VCH Verlag GmbH & Co. KgaA, 2005.
4. Kitano H. Computational systems biology. *Nature* 2002;**420**: 206–10.
5. Lander ES, Linton LM, Birren B, *et al*. Initial sequencing and analysis of the human genome. *Nature* 2000;**409**: 860–921.
6. Venter JC, Adams MD, Myers EW, *et al*. The sequence of the human genome. *Science* 2001;**291**:1304–51.
7. von Mering C, Krause R, Snel B, *et al*. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002;**417**:399–403.
8. Hucka M, Finney A, Sauro HM, *et al*. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;**19**:524–31.
9. Hucka M, Finney A, Bornstein BJ, *et al*. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Syst Biol* 2004;**1**:41–53.
10. Lloyd CM, Halstead MDB, Nielsen PF. CellML: its future, present and past. *Prog Biophys Mol Biol* 2004;**85**:433–50.
11. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comp Biol* 2002;**9**:67–103.
12. Assmus H, Herwig R, Cho KH, *et al*. Understanding the dynamics of biological systems: roles in medical research. *Expert Rev Mol Diagn* 2006;**6**:891–902.
13. Galperin MY. The molecular biology database Collection: 2007 update. *Nucleic Acids Res* 2007;**35**(Database issue):D3–4.
14. National Center for Biotechnology (NCBI). http://www.ncbi.nlm.nih.gov (30 March 2007, date last accessed).
15. European Bioinformatics Institute (EMBL-EBI). http://www.ebi.ac.uk/Databases (30 March 2007, date last accessed).
16. Hubbard T, Barker D, Birney E, *et al*. The Ensembl genome database project. *Nucleic Acids Res* 2002;**30**:38–41.
17. Birney E, Andrews D, Bevan P, *et al*. Ensembl 2004. *Nucleic Acids Res* 2004;**32**(Database issue):D468–70.
18. Birney E, Andrews TD, Bevan P, *et al*. An overview of Ensembl. *Genome Res* 2004;**14**:925–8.

19. Hammond MP, Birney E. Genome information resources – developments at Ensembl. *Trends Genet* 2004;**20**:268–72.
20. Boeckmann B, Bairoch A, Apweiler R, *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* 2003;**31**:365–70.
21. Apweiler R, Bairoch A, Wu CH, *et al*. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2004; **32**(Database issue):D115–9.
22. Biswas M, O'Rourke JF, Camon E, *et al*. Applications of InterPro in protein annotation and genome analysis. *Brief Bioinform* 2002;**3**:285–95.
23. Mulder NJ, Apweiler R, Attwood TK, *et al*. The InterPro database 2003 brings increased coverage and new features. *Nucleic Acids Res* 2003;**31**:315–8.
24. Berman HM, Westbrook J, Feng Z, *et al*. The protein data bank. *Nucleic Acids Res* 2000;**28**:235–42.
25. Pang KC, Stephen S, Engstrom PG, *et al*. RNAdb – a comprehensive mammalian noncoding RNA database. *Nucl. Acids Res* 2005;**33**(Database issue):D125–30.
26. Griffiths-Jones S, Grocock RJ, van Dongen S, *et al*. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* 2006;**34**(Database issue):D1404.
27. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;**30**:207–10.
28. Barrett T, Troup DB, Wilhite SE, *et al*. NCBI GEO: mining tens of millions of expression profiles – database and tools update. *Nucleic Acids Res* 2007;**35**(Database issue):D760–5.
29. Brazma A, Parkinson H, Sarkans U, *et al*. ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 2003;**31**:68–71.
30. Pathguide – The Pathway Resource List. http://www.pathguide.org/ (30 March 2007, date last accessed).
31. Kanehisa M, Goto S, Hattori M, *et al*. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006;**34**:D354–7.
32. Joshi-Tope G, Gillespie M, Vastrik I, *et al*. Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res* 2005;**33**(Database issue):D428–32.
33. Karp PD, Ouzounis CA, Moore-Kochlacs C, *et al*. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. *Nucleic Acids Res* 2005;**19**: 6083–89.
34. BioCarta. http://www.biocarta.com/ (30 March 2007, date last accessed).
35. Spike. http://www.cs.tau.ac.il/~spike/ (30 March 2007, date last accessed).
36. Schacherer F, Choi C, Götze U, *et al*. The TRANSPATH Signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics* 2001;**17**:1–6.
37. STKE. http://stke.sciencemag.org/ (30 March 2007, date last accessed).
38. NetPath. http://www.netpath.org/ (30 March 2007, date last accessed).
39. Pathway Interaction Database. http://pid.nci.nih.gov/ (30 March 2007, date last accessed).
40. Hermjakob H, Montecchi-Palazzi L, Lewington C, *et al*. IntAct – an open source molecular interaction database. *Nucleic Acids Res* 2004;**32**:D452–55.
41. Xenarios I, Rice DW, Salwinski L, *et al*. DIP: The database of interacting proteins. *Nucleic Acids Res* 2000;**28**: 289–91.
42. Salgado H, Gama-Castro S, Peralta-Gil M, *et al*. RegulonDB (version 5.0): Escherichia coli K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* 2006;**34**(Database issue):D394–7.
43. Zhao F, Xuan Z, Liu L, *et al*. TRED: a transcriptional regulatory element database and a platform for in silico gene regulation studies. *Nucleic Acids Res* 2005;**33**:D103–107.
44. Matys V, Kel-Margoulis OV, Fricke E, *et al*. TRANSFAC® and its module TRANSCompel®: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res* 2006; **34**(Database issue):D108–10.
45. Cary MP, Bader GD, Sander C. Pathway information for systems biology. *FEBS Lett* 2005;**579**:1815–20.
46. Schomburg I, Chang A, Ebeling C, *et al*. BRENDA, the enzyme database: updates and major new developments. *Nucleic Acids Res* 2004;**32**(Database issue):D431–3.
47. Wittig U, Golebiewski, M, Kania, R, et al. SABIO-RK: integration and curation of reaction kinetics data. In *Proceedings of the 3rd International Workshop on Data Integration in the Life Sciences 2006 (DILS'06), Hinxton, UK. Lecture Notes in Computer Science*, 2006;**4075**:94–103.
48. Le Novère N, Bornstein B, Broicher A, *et al*. BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res* 2006;**34**:D689–91.
49. Olivier BG, Snoep JL. Web-based kinetic modelling using JWS Online. *Bioinformatics* 2004;**20**:2143–4.
50. Eppig JT, Bult CJ, Kadin JA, *et al*. The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res* 2005;**33**: D471–5.
51. Crosby MA, Goodman JL, Strelets VB, *et al*. FlyBase: genomes by the dozen. *Nucleic Acids Res* 2007;**35**:D486–91.
52. WormBase web site. http://www.wormbase.org (30 March 2007, date last accessed).
53. Cherry JM, Ball C, Weng S, *et al*. Genetic and physical maps of Saccharomyces cerevisiae. *Nature* 1997;**387**(6632 Suppl):67–73.
54. Catalogue Of Somatic Mutations In Cancer (COSMIC). http://www.sanger.ac.uk/genetics/CGP/cosmic/ (30 March 2007, date last accessed).
55. Hulbert EM, Smink LJ, Adlem EC, *et al*. T1DBase: integration and presentation of complex data for type 1 diabetes research. *Nucleic Acids Res* 2007;**35**(Database issue):D742–6.
56. ChEBI. http://www.ebi.ac.uk/chebi/ (30 March 2007, date last accessed).
57. Human metabolome database. http://www.metabolomics.ca (5 September 2007, date last accessed).
58. PRIDE. http://www.ebi.ac.uk/pride (5 September 2007, date last accessed).
59. LipidMaps. http://www.lipidmaps.org (5 September 2007, date last accessed).
60. Human serum metabolome project. http://www.husermet.org (5 September 2007, date last accessed).
61. Roberts PM. Mining literature for systems biology. *Brief Bioinf* 2006;**7**:399–406.

62. Hakenberg J, Schmeier S, Kowald A, *et al*. Finding kinetic parameters using text mining. *OMICS: J Int Biol* 2004;**8**: 131–52.

63. Ananiadou S, Kell DB, Tsujii J. Text mining and its potential applications in systems biology. *Trends Biotechnol* 2006;**24**:571–9.

64. Jensen LJ, Saric J, Bork P. Literature mining for the biologist: from information retrieval to biological discovery. *Nat Rev Genet* 2006;**7**:119–29.

65. Krallinger M, Valencia A. Text-mining and information-retrieval services for molecular biology. *Genome Biology* 2005;**6**:224.

66. Brazma A, Krestyaninova M, Sarkans U. Standards for systems biology. *Nat Rev Genet* 2006;**7**:593–605.

67. Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;**29**: 365–71.

68. Taylor CF, Paton NW, Lilley KS, *et al*. The minimum information about a proteomics experiment (MIAPE). *Nat Biotechnol* 2007;**25**:887–93.

69. Le Novère N, Finney A, Hucka M, *et al*. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 2005;**23**:1509–15.

70. Taylor CF. Standards for reporting bioscience data: a forward look. *Drug Discov Today* 2007;**12**:527–33.

71. The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nat Genet* 2000;**25**:25–9.

72. XML – extensible Markup Language. http://www.w3.org/XML (30 March 2007, date last accessed).

73. MAGE-ML. http://xml.coverpages.org/mageML.html (30 March 2007, date last accessed).

74. BioPAX. http://www.biopax.org/ (30 March 2007, date last accessed).

75. Strömbäck L, Lambrix P. Representations of molecular pathways: an evaluation of SBML, PSI MI and BioPAX. *Bioinformatics* 2005;**21**:4401–7.

76. Wang X, Gorlitsky R, Almeida JS. From XML to RDF: how semantic web technologies will change the design of 'omic' standards. *Nat Biotechnol* 2005;**23**:1099–103.

77. JDesigner. http://jdesigner.org/ (30 March 2007, date last accessed).

78. Funahashi A, Tanimura N, Morohashi M, *et al*. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* 2003;**1**: 159–62.

79. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**: 2498–504.

80. Takahashi K, Yugi K, Hashimoto K, *et al*. Computational challenges in cell simulation. *IEEE Intell Syst* 2002;**17**:64–71.

81. Alves R, Antunes F, Salvador A. Tools for kinetic modelling of biochemical networks. *Nat Biotech* 2006;**24**: 667–72.

82. Mendes P. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput Appl Biosci* 1993;**9**:563–71.

83. Mendes P. Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem Sci* 1997;**22**:361–3.

84. Mendes P, Kell D. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 1998;**14**:869–83.

85. Hoops S, Sahle S, Gauges R, *et al*. COPASI – a COmplex PAthway SImulator. *Bioinformatics* 2006;**22**:3067–74.

86. Tomita M, Hashimoto K, Takahashi K, *et al*. E-CELL: software environment for whole-cell simulation. *Bioinformatics* 1999;**15**:72–84.

87. Takahashi K, Ishikawa N, Sadamoto Y, *et al*. E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics* 2003;**19**:1727–9.

88. Ginkel M, Kremling A, Nutsch T, *et al*. Modular modeling of cellular systems with ProMot/Diva. *Bioinformatics* 2003; **19**:1169–76.

89. Schaff J, Fink CC, Slepchenko B, *et al*. A general computational framework for modeling cellular structure and function. *Biophys J* 1997;**73**:1135–46.

90. Loew LM, Schaff JC. The virtual cell: a software environment for computational cell biology. *Trends Biotechnol* 2001;**19**:401–6.

91. Slepchenko BM, Schaff JC, Macara I, *et al*. Quantitative cell biology with the virtual cell. *Trends Cell Biol* 2003;**13**:570–6.

92. Hucka M, Finney A, Sauro HM, et al. The ERATO Systems Biology Workbench: enabling interaction and exchange between software tools for computational biology. In *Proceedings of the Pacific Symposium on Biocomputing* 2002;450–61.

93. Klipp E, Liebermeister W, Helbig A, *et al*. Systems biology standards – the community speaks. *Nat Biotechnol* 2007;**25**: 390–1.

94. Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, *et al*. METATOOL: for studying metabolic networks. *Bioinformatics* 1999;**15**:251–7.

95. PyBioS. http://pybios.molgen.mpg.de/ (27 September 2007, date last accessed).

96. SBGN – Systems Biology Graphical Notation. http://sbgn.org (4 September 2007, date last accessed).

97. Hermjakob H, Montecchi-Palazzi L, Bader G, *et al*. The HUPO PSI's molecular interaction format–a community standard for the representation of protein interaction data. *Nat Biotechnol* 2004;**22**:177–83.

98. Longabaugh WJR, Davidson EH, Bolouri H. Computational representation of developmental genetic regulatory networks. *Dev Biol* 2005;**283**:1–16.

99. Shannon P, Markiel A, Ozier O, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;**13**:2498–504.

100. Hu Z, Mellor J, Wu J, *et al*. VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res* 2005;**33**:W352–7.