



The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks

Sudhir Sivakumaran, Sridhar Hariharaputran, Jyoti Mishra and Upinder S. Bhalla*

National Centre for Biological Sciences, GKVK Campus, Bangalore 560065, India

Received on February 15, 2002; revised on June 28, 2002; accepted on September 11, 2002

ABSTRACT

Motivation: Analysis of cellular signaling interactions is expected to pose an enormous informatics challenge, perhaps even larger than analyzing the genome. The complex networks arising from signaling processes are traditionally represented as block diagrams. A key step in the evolution toward a more quantitative understanding of signaling is to explicitly specify the kinetics of all chemical reaction steps in a pathway. Technical advances in proteomics and high-throughput protein interaction assays promise a flood of such quantitative data. While annotations, molecular information and pathway connectivity have been compiled in several databases, and there are several proposals for general cell model description languages, there is currently little experience with databases of chemical kinetics and reaction level models of signaling networks.

Results: The Database of Quantitative Cellular Signaling is a repository of models of signaling pathways. It is intended both to serve the growing field of chemical-reaction level simulation of signaling networks, and to anticipate issues in large-scale data management for signaling chemistry.

Availability: The Database of Quantitative Cellular Signaling is available at <http://doqcs.ncbs.res.in>. Links to the signaling model simulator, GENESIS/Kinetikit are at <http://www.ncbs.res.in/~bhalla/kkit/index.html> and are also provided from within the database. The database source code is available under the GNU Public License.

Contact: bhalla@ncbs.res.in

INTRODUCTION

Signaling networks are the computational and control system of the cell. The traditional view of signaling involves transduction of chemical signals at the cell surface, and their propagation via sequences of biochemical events involving proteins and second messengers (Stryer, 2001). In its broader sense, the cellular signaling network includes

genetic, cytoskeletal and cell trafficking elements. Qualitative analyses of such networks have been carried out using logical representations, for example in plant signaling (Mendoza *et al.*, 1999; reviewed in Genoud *et al.*, 2001). These studies are appropriate in many cases where general connectivity is understood, but kinetic details are uncertain. Mathematical and modeling methods are useful in gaining further insights into cellular function (Tyson *et al.*, 2001). Several recent studies have undertaken a quantitative analysis of cellular signaling at the level of mass-action kinetics of signaling pathways (Bhalla and Iyengar, 1999; Kuroda *et al.*, 2001; Lamb, 1994) and genetic interactions (Gillespie, 1977). Some studies include analysis of the three-dimensional, stochastic and cellular mechanical function (Stiles *et al.*, 1998; Shimizu *et al.*, 2000; Arkin and Ross, 1994). Current data sources, such as test-tube biochemistry, are clearly poor approximations to cellular conditions, nevertheless these are the best sources of data we currently have. It is anticipated that more biologically detailed descriptions will become increasingly feasible with new experimental techniques (Teruel and Meyer, 2001; Kierzek, 2001; Voytik-Harbin *et al.*, 2001). All such quantitative descriptions have mass-action chemistry as a common denominator. There is therefore a clear need for developing data management and analysis systems appropriate for such data.

A number of initiatives have come into being as part of this process. They can broadly be grouped into three categories: databases, simulators and model description languages. Although it is structured as a database, the DOQCS project draws its inspiration especially from simulation projects related to cellular signaling (e.g. Bhalla and Iyengar, 1999). These studies have given rise to many models incorporating detailed and explicit reaction schemes and parameters. DOQCS is a resource for such models that sets quantitative functional analysis of the data as a central consideration for database development. As this emphasis differs from other databases, a further goal of the DOQCS project was to identify and fulfill

*To whom correspondence should be addressed.

distinctive database requirements as the dataset expanded and usage patterns became clearer.

METHODS AND ALGORITHMS

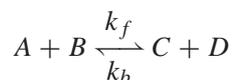
DOQCS primarily contains explicit chemical-reaction level models of signaling pathways, accompanied by annotations, information about data sources, and parameter derivations. Most of these models are based on published modeling studies, and the rest have been developed specifically for DOQCS. Prior to database entry, the functionality of each model is verified using the Kinetikit/GENESIS simulator (Bhalla, 1998). This initial filter ensures that all models are complete since all parameters must be defined in order to replicate published properties. Kinetikit features a database dump facility that automatically converts data to SQL commands suitable for upload to the database.

The data are stored using the MySQL database (<http://www.mysql.com/>) using PHP (<http://www.php.net>) as an interface, (Welling and Thomson, 2001). Displays are generated using the PHP interface to the GD library (<http://www.boutell.com/gd/>).

DESIGN AND IMPLEMENTATION

Data model

Chemical kinetic simulations are performed by converting chemical equations of the general form



to systems of differential equations of the form

$$d[A]/dt = -k_f[A][B] + k_b[C][D]$$

and applying standard numerical integration methods (Bhalla, 1998) to calculate the time evolution of these reactions.

There are several ways of specifying chemical kinetic models. Many models are reported in terms of concise systems of differential equations, after applying mass conservation rules to eliminate redundant equations (e.g. Asthagiri and Lauffenburger, 2000). It is also common to assume equilibrium relationships between molecules to avoid solving multiple differential equations (Grzybowski *et al.*, 2000; Hecht *et al.*, 1990). Some models have also been described in terms of concentration-dependent rate constants (Kholodenko, 2000). In distinction to these abstractions and numerical simplifications, simulators such as GENESIS/Kinetikit take a more chemically detailed approach and require that every molecule, reaction, and enzyme activity be explicitly specified (Bhalla, 2002a). A key design decisions for DOQCS was to retain this explicit chemical-level description for

all models. DOQCS is implemented in MySQL, which is a relational database rather than an object-oriented database. Therefore molecules, reactions and enzymes are each represented in distinct tables. Each chemical object in the simulation maps onto a row in the appropriate table in DOQCS. For example, a protein kinase would be represented as an entry in the molecule table, and each of its catalytic reactions with distinct substrates as a distinct entry in the enzyme table. A stoichiometric reaction such as binding of regulatory subunit to catalytic subunit would be represented as one entry in the reaction table, plus five entries in the molecule table: one each for the substrate, product, regulatory subunit catalytic subunit, and the complex (Figure 1). This explicit mapping of table entries to the basic chemical concepts of molecules and reactions has several advantages:

- There is a direct correspondence between database entries and experimentally measurable quantities such as reactant concentrations.
- As there are no assumptions about equilibrium situations it is possible to apply reaction schemes to dynamic chemical situations on a time-scale shorter than the equilibrium time-scale.
- Stochastic chemical systems can be represented without any change to the database. It is sufficient to re-interpret the entries for kinetic rates as probabilities of reaction events. For example, a rate constant k_f in units of $\mu\text{M}/\text{sec}$ could be scaled to # of molecules/sec and this can be used to estimate reaction transition probabilities. The reaction scheme and other entries in the database would be unaffected.
- There is a logical extension into three-dimensional reaction-diffusion systems by addition of spatial distribution information. The chemical organization of the tables need not change.

DATABASE STRUCTURE

The table structure of the database is designed to represent three levels of data description: the chemistry of individual reactions, their organization into pathways, and the administrative level pertaining to accession information (Figure 1). As discussed above, there are separate tables for molecules, reactions, and enzymes. Each entry in these tables includes an identifier for the pathway and accession to which it belongs. Other common fields are notes and name. The remaining parameters are data-type specific. Lists of substrates and products in reaction and enzyme table entries specify chemical connectivity. Enzymes contain a further entry to specify the parent molecule. This is necessary because a specific protein may have several

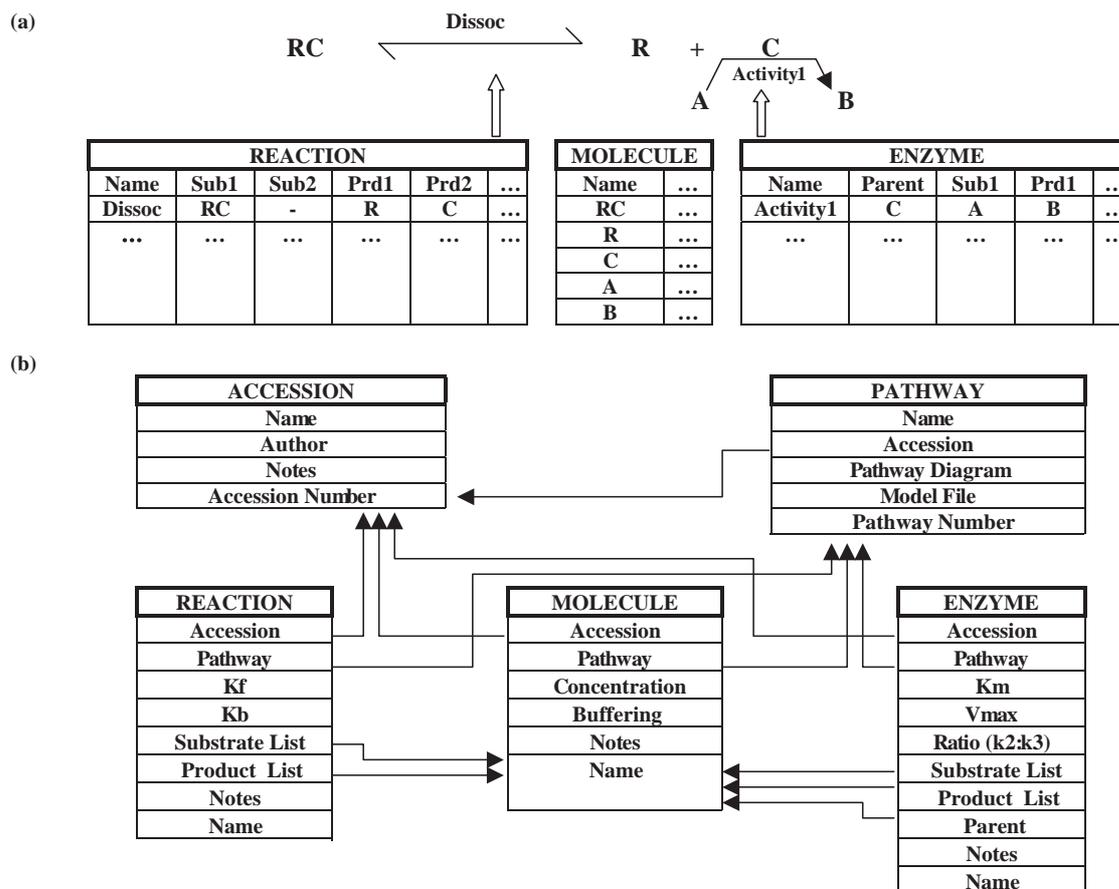


Fig. 1. (a) Representation of chemical reaction scheme in database. An illustrative kinetic scheme for enzyme regulation is shown. It involves five molecules, one reaction and one enzymatic activity. Each of the reactions, molecules and enzymatic activities maps onto one row of their respective tables in the database. The reaction and enzyme tables refer to the reactant molecules through substrate and product lists. The enzyme table has an additional entry for the parent molecule that possesses the enzyme activity. (b) Database structure. The accession table contains administrative information about each database entry, and the accession number is a unique index for a particular accession. The pathway table describes the structure of the pathway and also has a unique pathway number as an index. Multiple pathways may be grouped within an accession. Each pathway entry contains an identifier for the accession to which the pathway belongs. Each chemical entry in the reaction, molecule and enzyme tables has identifiers for the pathway and for the accession to which it belongs. Reactions and enzymes have substrate and product lists pointing to molecule entries. Enzyme tables additionally store the name of the ‘parent’ molecule that possesses the enzyme activity. The thesaurus table (not shown) maps molecule, reaction and enzyme names to canonical forms.

enzyme activities with different rates, acting on different substrates.

An accessory table named ‘thesaurus’ has been implemented because of the inconsistency in chemical naming schemes among different models. This provides a lookup for each molecule, reaction and enzyme name in terms of a consistent single name.

The next level of organization is the pathway. This maps closely onto individual pathway blocks in conventional signaling pathway representations. Chemical entries (reactions, molecules and enzymes) are grouped into pathways through the pathway identifier in these tables. Like the chemical tables, the pathway too contains accession,

name and notes fields. In addition it contains the pathway reaction diagram.

The top level of the tabular hierarchy is the accession. This primarily contains information pertaining to the administration of accessions to the database. Currently two accession types are implemented: individual pathways and networks. In the former there is a single pathway model in the accession. Such entries are intended as building blocks for larger signaling models. Network accessions, as the name implies, include several interacting signaling pathways, each pointing to the same accession. These entries raise an interesting organizational issue: how to represent interactions between pathways? Incorporation

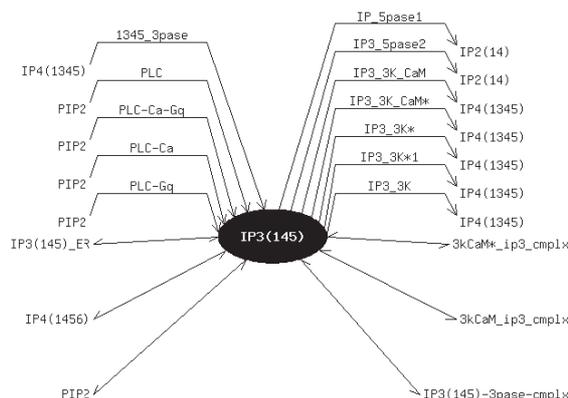


Fig. 2. Map of molecular connectivity. The display indicates each enzymatic action and reaction upstream and downstream of the selected molecule, $IP_3(145)$ in this example. In addition it also indicates any substrates and products if the selected molecule itself is an enzyme (not present for $IP_3(145)$). Each of the molecules in the display can be selected by clicking, following which the display shifts to display the selected molecule and its reactions. Similar maps are generated for molecule-to-pathway and pathway-to-pathway connectivity.

of interactions within individual pathways would be confusing, as this would mean that one pathway contains reactions and molecules pertaining to another. In object-oriented terms, the pathways would no longer be cleanly encapsulated, as they would have dependencies on the contents of other pathways. The current solution, borrowed from Kinetikit, is to define a basal pathway ‘kinetics’ that contains molecules and reactions that interact with more than one pathway.

DATABASE INTERFACE AND UTILITIES

Searches

The multi-level organization of the data in DOQCS necessitates a corresponding multi-level search approach. The Search form in the web interface provides for selection on any or all of the five key tables in the database: Accession, Pathway, Molecule, Reaction and Enzyme. The results are formatted appropriately for each table. For example, search results for molecules, reactions and enzymes provide links back to the accession and pathway. However, enzymes are identified both by the name of the enzyme activity as well as the name of the enzyme molecule. This specific additional information has to be presented for enzyme search results. Each level of the search results includes the appropriate annotations for the specific component being searched for, as this turns out to be one of the most useful entries for interpreting searches.

Reaction and Pathway navigation

DOQCS provides a qualitative, graphical representation of pathway interactions by displaying a map of connectivity of signaling pathways and molecules. In this utility, a specified molecule or pathway is displayed with all incoming and outgoing interactions from the database shown graphically using arrows. The display can be set to three levels of detail. At the most detailed level, molecule-to-molecule connectivity information is extracted from the database. All molecular inputs and outputs of a selected molecule are displayed (Figure 2). Distinct arrows represent which molecules bind to, are formed from, or undergo enzymatic conversion from the molecule of interest. The connectivity map provides hyperlinks to each molecule in the diagram, which facilitates navigation through the chemistry of the signaling pathway. At an intermediate level, the map displays molecule-to-pathway connectivity. All pathway inputs and outputs to a given molecule are displayed. Clicking on these inputs and output pathways selects a key molecule in that pathway to be the new center of the display. At the highest level, the database displays pathway-to-pathway connectivity, again, each pathway is accessible through a mouse click. A summary of information and links to the detailed quantitative data in the database is extracted and tabulated below the connectivity map at each level of the hierarchy.

Comparisons

As discussed below, many pathway entries turn out to be closely related. This necessitated the development of various methods for comparing pathways. The current comparisons are of three kinds: component, parameter, and notes comparisons. The general approach is to match up corresponding molecules, reactions and enzymes from the respective pathways, calculate a similarity index for each entry, and average over all matches (Figure 3). The results are summarized graphically in a comparison tree that depicts the similarity between different models of a specific pathway. This tree represents very similar pathway models in close proximity and less similar models farther apart. Additional details regarding the percentage of similarity and the number of matched components are presented in tabular form below the comparison tree map.

The calculations for similarity are done as follows. Molecule names from each pathway are first converted to a canonical form using the ‘thesaurus’ table. A similar lookup is done for names of reactions and enzymes. This ensures inclusion of chemically identical but differently named entries in the comparison, which increases the number of comparisons and improves the reliability of the results. The fraction of matches among these canonical

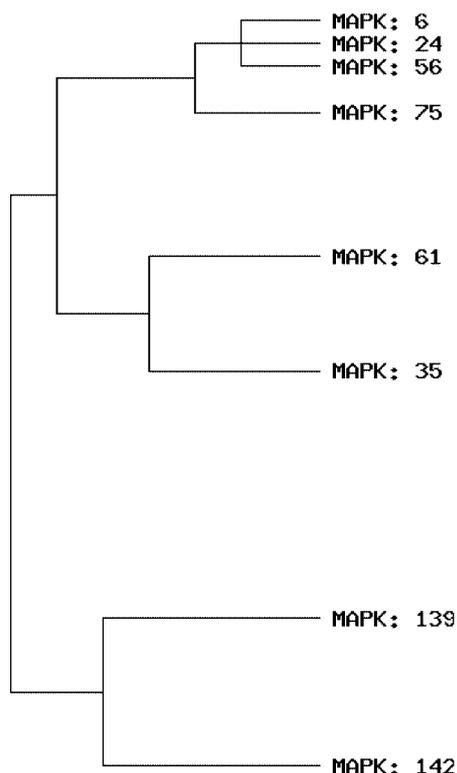


Fig. 3. Comparison tree generated to show the relatedness of several MAPK pathway models in the database. The terminal branches show the name of the pathway and its pathway identifier, and can be clicked to bring up details of the selected pathway. The distance between names indicates the difference between models, calculated as discussed in the text.

names is then calculated:

$$\text{Match \%} = 200 * M / (nA + nB)$$

where M is the number of name matches between pathway A and pathway B, having nA and nB components respectively.

Next, parameter comparisons are performed on the subset of components with matching canonical names. The similarity index is given by the ratio of the smaller to the larger value of the parameter.

$$\begin{aligned} \text{Match \%} &= 100 && \text{if } pA = pB \\ &= pA/pB && \text{if } pA < pB \\ &= pB/pA && \text{if } pB < pA \end{aligned}$$

Where pA is a parameter from pathway A and pB is the corresponding parameter from pathway B.

Finally, the lengths of strings containing annotations for each pathway are compared. These lengths are used as parameters in the same manner as above. This comparison is weighted at one-tenth the other two in the tree display.

Reports

In keeping with its primary motivation in facilitating signaling model development, the report formats for the database provide parameter listings in human as well as machine-readable form. Currently four such formats have been implemented: (1) A simple tabular text listing of parameters. (2) A text listing including parameter annotations, for supplementary material for publications. (3) The Kinetikit model description file format. (4) Differential equations. Converters for other simulators and the CellML (<http://www.cellml.org/>) and SBML (ERATO/SBML-Systems Biology Workbench) (<http://www.cds.caltech.edu/erato/>) (Hucka *et al.*, 2001; Hucka, 2001) formats are planned.

RESULTS AND DISCUSSION

The database currently contains 26 accessions with 146 pathways and over 2500 molecules, reactions and enzymes. Based on a literature search, we estimate that the current size of the database represents approximately one-third of the published models of signaling pathways at the level of chemical kinetics.

This does not include a much larger set of metabolic models, and published models of calcium dynamics in the context of neuronal biophysics, that are outside the current scope of the database.

At this time, all accessions are done in-house pending development of a system for on-line entries and their curation. It is anticipated that the data entry rate will increase when the on-line system is implemented. We also expect the number of available models from the literature to increase rapidly with technological improvements in parameter determination.

From our database and other studies (Juty *et al.*, 2001) it is apparent that the naming of molecules, reactions, and enzymes is often inconsistent between models. On the one hand modelers prefer to use abbreviations in their models since formal molecular and enzyme names are cumbersome. On the other hand, naming differences complicate comparisons and may lead to ambiguity. Often these ambiguities are deliberate: generic models of Protein Kinase C (Bhalla and Iyengar, 1999) or Adenylyl Cyclase (Bhalla, 2002a) in the database, are conscious attempts to represent multiple isoforms with a single, averaged model. There are a number of possible solutions to naming inconsistencies. For example, part of the annotation process could involve construction of a matrix to identify equivalent molecules in different pathway models. We have implemented this by creating a thesaurus of canonical names, as previously discussed by Juty *et al.* (2001), but this entails additional curation effort. There are also possible heuristic approaches to matching molecules, reactions and enzymes based on their positions

in reaction graphs (Patel *et al.*, 2001; Ettinger, 2002). Each of these approaches presents drawbacks. For example, in the MAPK pathway certain key phosphatases act upon multiple molecules at multiple sites. When should such reactions be grouped together for comparison, and when should they be treated distinctly? Further, a phosphatase target may be represented as a single molecule in some models, but may exist in multiple states in other models. How should the database handle the mapping of molecular identity in such cases? Since DOQCS lays emphasis on providing original models, the original naming is retained in all cases. Some approaches to model comparisons are discussed below.

The current dataset suggests a clear tendency for signaling pathway models to be implemented in a number of variants, sharing most pathways but having unique versions of others. Not surprisingly, there is also a tendency for an evolutionary relationship between models developed over a period of time. For example, DOQCS currently contains eight versions of the MAPK pathway. Four of these are identical except for annotations. Two are more recent expanded models based on the earlier ones. The remaining two are distinct models based on original literature. How similar are these models? We have implemented one approach to similarity calculation based on parameter comparisons between equivalent molecules and reactions. As discussed above, this works well with the addition of a thesaurus (Juty *et al.*, 2001) to establish correspondence between model components. The similarity calculation approach breaks down when models describe fundamentally different reaction schemes.

A possible better, longer-term approach is suggested by detailed point-by-point comparisons between models (Blüthgen and Herzog, 2001). The strategy in this approach is to compare the functional behavior of the models rather than their internal details. To the extent that functional analysis and prediction is the key goal of model building, this may be a more appropriate way of comparing models. It would also sidestep many naming and model structure issues already discussed. From the viewpoint of database comparisons, this would require that the database either includes model simulation results, or generates them online, and performs comparisons between these results. To our knowledge no database currently implements such comparisons.

Signaling networks have become the focus of many database projects, with a wide diversity of goals. The TRANSPATH database (Schacherer *et al.*, 2001) has extensive searchable maps of signaling pathways to follow signaling interactions. One of the objectives of the database is to facilitate simulations, and its data structures can accommodate chemical reaction details. The bulk of the data in TRANSPATH at the current time is block-diagram rather than chemical-reaction level. The related

database CSNDB (Takai-Igarashi *et al.*, 1998) also has molecular information but not described in terms of reaction kinetics. The databases DIP (Xenarios *et al.*, 2002), GeneNet (Kolpakov *et al.*, 1998) and BIND (Bader and Hogue, 2000) describe molecules involved in pathways, and provide reaction maps, but most do not have details on reaction kinetics. Biocarta (<http://www.biocarta.com>) has attractive annotated diagrams of many signaling pathways at the block-diagram level. GeneNet currently has two web-enabled genetic models and a schema for implementing kinetic models, but is not at this time a repository of signaling models. The database of the Alliance for Cellular Signaling (AFCS) has well-annotated protein lists and a selection of signaling maps, and has the stated goal of including quantitative details on signaling (Gilman, 2000) (<http://www.cellularsignaling.org/>). EMP (Selkov *et al.*, 1996) and BRENDA (Schomburg *et al.*, 2002) are excellent sources of enzyme kinetics, but include little or no regulatory details and lack the context of complete pathway models with upstream and downstream interactions.

The Virtual Cell project (Loew and Schaff, 2001) has several signaling models accessible on-line. The numerical model descriptions are embodied in a 'Mathematical Model' using the Virtual Cell Model Description Language (VCMDL). The current emphasis in the Virtual Cell project is simulations rather than maintaining a repository of signaling models. Several databases provide protein interaction data, e.g. KEGG (Kanehisa *et al.*, 2002), ASPD (Valuev *et al.*, 2002) and SPiD (Hoebeker *et al.*, 2001). Though such data are becoming increasingly quantitative, they are not sufficient to formulate reaction schemes and complete kinetics for signaling pathways.

DOQCS is therefore distinctive in being a repository of functional simulation models, based on chemical kinetics. The search and comparison facilities that have grown out of this specialization are also different from other databases. DOQCS does not currently provide simulator functionality, and is therefore not comparable to any of several simulator packages, e.g. V-Cell (Loew and Schaff, 2001), DBsolve (Juty *et al.*, 2001), Jarnac (Sauro, 2000), E-Cell (Tomita *et al.*, 1999), MCell (Stiles *et al.*, 1998) and Kinetikit (Bhalla, 1998). However, the models in DOQCS may be useful for these simulators as conversion tools appear. The XML-based model description languages SBML (Hucka *et al.*, 2001) (<http://www.cds.caltech.edu/erato/>) and CellML (<http://www.cellml.org>) offer the prospect of facilitating this interconversion.

Biological models are rapidly evolving out of the 'word model' phase into more quantitative, predictive and useful simulations. Data remains a limiting factor. Most current data is rather qualitative, and appropriate modeling methods have been developed for such data

(reviewed in Genoud *et al.* (2001)). Quantitative modeling of signaling pathways is currently a rather small niche in biology, but many stimuli are now converging to accelerate the process of quantification of cell biology, and thus greatly expand the scope for signaling models. Most current chemical kinetic models of signaling interactions incorporate a limited degree of compartmentalization, but exclude spatial detail. The schema for DOQCS is currently at this level. There are clear trends to include more cellular detail in simulators and signaling models. Such details include spatial information, stochasticity in single-molecule interactions, cytoskeletal effects and cellular mechanics, and integration of all these with genetic interactions (Gillespie, 1977; Arkin *et al.*, 1998; Stiles *et al.*, 1998). In addition to these model details, it would also be desirable to extend the DOQCS schema to support reaction modifiers, more general rate laws, and stimulus descriptions designed to replicate specific experiments.

Beyond greater biological realism in models there is a clear trend towards closer integration of database, simulator, and model description formats. XML-based languages seem to be one approach to facilitating this integration (Achard *et al.*, 2001; Lamont, 2001; Williams, 2002) but there are currently very few simulators that can read these model description languages (e.g. DBsolve6, Juty *et al.* (2001)). It is also currently difficult to combine a large number of individual pathway models into a signaling network model. New simulators and object-oriented descriptions of signaling networks are starting to address these challenges (Bhalla, 2002b; ECELL (Tomita *et al.*, 1999), MCell (Stiles *et al.*, 1998) (<http://www.mcell.psc.edu>)). The current DOQCS project is envisioned as one step in this highly desirable convergence of technologies where predefined individual pathway models selected from a database can be linked together using graphical methods and simulated in an appropriate cellular spatial context.

REFERENCES

- Achard, F., Vaysseix, G. and Barillot, E. (2001) XML, bioinformatics and data integration. *Bioinformatics*, **17**, 115–125.
- Arkin, A. and Ross, J. (1994) Computational functions in biochemical reaction networks. *Biophys. J.*, **67**, 560–578.
- Arkin, A., Ross, J. and McAdams, H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, **149**, 1633–1648.
- Asthagiri, A.R. and Lauffenburger, D.A. (2000) Bioengineering models of cell signaling. *Annu. Rev. Biomed. Eng.*, **2**, 31–53.
- Bader, G.D. and Hogue, C.W.V. (2000) BIND-A Data specification for storing and describing biomolecular interactions, molecular complexes and pathways. *Bioinformatics*, **16**, 465–477.
- Bhalla, U.S. (1998) A network within: signaling pathways. In Bower, J.M. and Beeman, D. (eds), *The Book of Genesis. Exploring Realistic Neural Models with the GEneral NEural Simulation System*. Springer, New York, pp. 169–191.
- Bhalla, U.S. (2002a) Use of Kinetikit and GENESIS for modeling signaling pathways. *Meth. Enzymol.*, **345**, 3–23. (<http://doqcs.ncbs.res.in/>)
- Bhalla, U.S. (2002b) The chemical organization of signaling interactions. *Bioinformatics*, **18**, 855–863.
- Bhalla, U.S. and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science*, **283**, 381–387.
- Blüthgen, N. and Herzog, H. (2001) MAP-kinase-cascade: switch, amplifier or feedback controller. In Gauges, R., van Gend, C. and Kummer, U. (eds), *2nd Workshop on Computation of Biochemical Pathways and Genetic Networks*. Logos, Berlin, pp. 55–62.
- Ettinger, M. (2002) The complexity of comparing reaction systems. *Bioinformatics*, **18**, 465–469.
- Genoud, T., Trevino Santa Cruz, M.B. and Métraux, J.-P. (2001) Numerical simulation of plant signaling networks. *Plant Physiology*, **126**, 1430–1437.
- Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.
- Gilman, A. (2000) A letter to the Signaling community: AfCS. At http://afcs.swmed.edu/afcs/letter_to_community.htm.
- Grzybowski, B.A., Anderson, J.R., Colton, I., Brittain, S.T., Shakhnovich, E.I. and Whitesides, G.M. (2000) Modeling the kinetics of acylation of insulin using a recursive method for solving the systems of coupled differential equations. *Biophys. J.*, **78**, 652–661.
- Hecht, J.P., Nikonov, J.M. and Alonso, G.L. (1990) A BASIC program for the numerical solution of the transient kinetics of complex biochemical models. *Comput. Methods. Programs Biomed.*, **33**, 13–20.
- Hoebeke, M., Chiapello, H., Noirot, P. and Bessieres, P. (2001) SPiD: a subtilis protein interaction database. *Bioinformatics*, **17**, 1209–1212.
- Hucka, M. (2001) *SBML: Systems Biology Markup Language. Presentation at the First International Symposium on Computational Cell Biology*. Lenox, Massachusetts.
- Hucka, M., Finney, A., Sauro, H. and Bolouri, H. (2001) *Systems Biology Markup Language (SBML) Level 1: Structures and Facilities for Basic Model Definitions*.
- Juty, N.S., Spence, H.D., Hotz, H.R., Tang, H., Goryanin, I. and Hodgman, T.C. (2001) Simultaneous modelling of metabolic, genetic and product-interaction networks. *Bioinformatics*, **2**, 223–232.
- Kanehisa, M., Goto, S., Kawashima, S. and Nakaya, A. (2002) The KEGG databases at Genomenet. *Nucleic Acid Res.*, **30**, 42–46.
- Kholodenko, B.N. (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, **267**, 1583–1588.
- Kierzek, A.M. (2001) Dependence of magnitude of stochastic fluctuations in prokaryotic gene expression on the transcription and translation initiation rates. In Yi, T.M., Hucka, M., Morohashi, M. and Kitano, H. (eds), *Proceedings of the International Conference of Systems Biology*. Omnipress, USA, pp. 70.
- Kolpakov, F.A., Anako, E.A., Kolesov, G.B. and Kolchanov, N.A. (1998) GeneNet: a gene network database and its automated visualization. *Bioinformatics*, **14**, 529–537.
- Kuroda, S., Schweighofer, N. and Kawato, M. (2001) Exploration of signal transduction in cerebellar long-term depression by kinetic simulation. *J. Neurosci.*, **21**, 5693–5702.

- Lamb, T.D. (1994) Stochastic simulation of activation in the G-protein cascade of phototransduction. *Biophys. J.*, **67**, 1439–1454.
- Lamont, J. (2001) Data capture: extending the net to new capabilities. In *KMWorld*. pp. 8–9.
- Loew, L.M. and Schaff, J.C. (2001) The Virtual Cell: a software environment for computational cell biology. *Trends Biotechnol.*, **19**, 401–406.
- Mendoza, L., Thieffry, D. and Alvarez-Buylla, E.R. (1999) Genetic control of flower morphogenesis in *Arabidopsis thaliana*: a logical analysis. *Bioinformatics*, **15**, 593–606.
- Patel, S., Rabone, J., Russell, S., Tissen, J. and Klaffke, W. (2001) Iterated reaction graphs: simulating complex Maillard reaction pathways. *J. Chem. Inf. Comput. Sci.*, **41**, 926–933.
- Sauro, H.M. (2000) Jarnac: a system for interactive metabolic analysis. In Hofmeyr, J.H.S., Rohwer, J.M. and Snoep, J.C. (eds), *Animating the Cellular map: Proceedings of the 9th International Meeting on Biothermokinetics*. Stellen Bosch University Press, ISBN 0-7972-0776-7, <http://members.lycos.co.uk/sauro/biotech.htm>.
- Schacherer, F., Choi, C., Gotze, U., Krull, M., Pistor, S. and Wingender, E. (2001) The TRANSPATH signal transduction database: a knowledge base on signal transduction networks. *Bioinformatics*, **17**, 1053–1057.
- Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem Sci.*, **27**, 54–56.
- Selkov, E., Basmanova, S., Gaasterland, T., Goryanin, I., Gretchkin, Y., Maltsev, N., Nenashev, V., Overbeek, R., Panyushkina, E., Pronevitch, L. et al. (1996) The metabolic pathway collection from EMP: the enzymes and metabolic pathways database. *Nucleic Acids Res.*, **24**, 26–28.
- Shimizu, T.S., Le Novere, N., Lavin, M.D., Beavil, A.J., Sutton, B.J. and Bray, D. (2000) Molecular model of a lattice of signaling proteins involved in bacterial chemotaxis. *Nat. Cell Biol.*, **2**, 792–796.
- Stiles, J.R., Bartol, T.M., Salpeter, E.E. and Salpeter, M.M. (1998) Monte Carlo simulation of neurotransmitter release using MCell, a general simulator of cellular physiological processes. In Bower, J.M. (ed.), *Computational Neuroscience*. Plenum, New York, pp. 279–284.
- Stryer, L. (2001) Signal transduction cascades. *Biochemistry*, 5th edn, WH Freeman, New York, pp. 325–360.
- Takai-Igarashi, T., Nadaoka, Y. and Kaminuma, T. (1998) A database for cell signaling networks. *J. Comp. Biol.*, **5**, 747.
- Teruel, M.N. and Meyer, T. (2001) Highly parallel single cell monitoring of receptor-triggered membrane translocation of a calcium sensing protein module. In Yi, T.M., Hucka, M., Morohashi, M. and Kitano, H. (eds), *Proceedings of the International Conference of Systems Biology*. Omnipress, USA, pp. 62.
- Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T.S., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C. and Hutchison, C.A. (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84.
- Tyson, J.J., Chen, K. and Novak, B. (2001) Network dynamics and cell physiology. *Nat. Rev. Mol. Cell Biol.*, **2**, 908–916.
- Valuev, V.P., Afonnikov, D.A., Ponomarenko, M.P., Milanesi, L. and Kolchanov, N.A. (2002) ASPD (Artificially Selected Proteins/Peptides Database): a database of proteins/peptides evolved *in vitro*. *Nucleic Acid Res.*, **30**, 200–202.
- Voytik-Harbin, S.L., Rajwa, B. and Robinson, J.P. (2001) Three-dimensional imaging of extracellular matrix and extracellular matrix-cell interactions. *Methods Cell Biol.*, **63**, 583–597.
- Welling, L. and Thomson, L. (2001) *PHP and MySQL Development*, Sams, ISBN 0-6723-1784-2.
- Williams, K. (2002) XML for Data: Modeling Many-To-Many Relationships. Tips and Techniques to Help You Create More Flexible XML Models. From *IBM DeveloperWorks*, XML Zone.
- Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S. and Eisenberg, D. (2002) DIP: the Database of Interacting Proteins. a research tool for studying cellular networks of protein interactions. *Nucleic Acid Res.*, **30**, 303–305.