JB Minireview—Bioinformatics & Systems Biology

J. Biochem. **137**, 653–657 (2005)
DOI: 10.1093/jb/mvi083

Electronic Data Sources for Kinetic Models of Cell Signaling

G. V. HarshaRani, Sharat J. Vayttaden and Upinder S. Bhalla*

National Centre for Biological Sciences, Tata Institute of Fundamental Research, GKVK Campus, Bangalore 560065, India

Received October 29, 2004; accepted March 16, 2005

Functional understanding of signaling pathways requires detailed information about the constituent molecules and their interactions. Simulations of signaling pathways therefore build upon a great deal of data from various sources. We first survey electronic data resources for cell signaling modeling and then based on the type of data representation the data sources are broadly classified into five groups. None of the data sources surveyed provide all required data in a ready-to-be-modeled fashion. We then put forward a "wish list" for the desired attributes for an ideal modeling centric database. Finally, we close with perspectives on how electronic data sources for cell signaling modeling have developed. We suggest that future directions in such data sources are largely model-driven and are hinged on interoperability of data sources.

Key words: biological databases, kinetics, models, signal transduction, simulation.

Abbreviations: AfCS, Alliance for Cellular Signaling; BBID, Biological Biochemical Image Database; BRENDA, Braunschweig Enzyme Database—the comprehensive enzyme information system; CellML, Cell Markup Language; DOQCS, Database of Quantitative Cellular Signaling; E-Cell, Electronic Cell; eMIM, Electronic Molecular Interaction Map; GIF, Graphics Interchange Format; iHOP, Information Hyperlinked over Proteins; JPEG, Joint Photographic Experts Group; JWS, Java Web Simulation; KDBI, Kinetic Data of Biomolecular Interactions; KEGG, Kyoto Encyclopedia of Genes and Genomes; MATLAB, Matrix Laboratory; ODE, Ordinary Differential Equation; PNG, Portable Network Graphics; ProTherm, Thermodynamic Database for Proteins and Mutants; SBML, Systems Biology Markup Language; STKE, Signal Transduction Knowledge Environment; SVG, Scalable Vector Graphics; TAIR, The Arabidopsis Information Resource; UML, Unified Modeling Language; URL, Uniform Resource Locator; XML, eXtensible Markup Language.

1. Introduction

Cellular information flow is largely mediated through a complex network of biochemical signaling events. These events have recently become the focus of much work in the broad field of systems biology. To a first approximation such events can be represented as biochemical reaction steps, which are amenable to relatively standard techniques for kinetic simulations. Kinetic modeling of biological signaling systems involves the steps of defining the system topology, obtaining kinetic parameters, and finally deciding on the model formalism. Each step requires a different type of information. In the first section of this review we look at the different types of data sources available for kinetic modeling and how helpful are these sources at different stages of modeling. The latter section identifies some important attributes of a model-centric database.

2. Why do we need databases?

In recent years Biology has begun to come to terms with the large amount of information being generated and documented through modern experimental techniques. This problem is especially acute in fields like systems biology where the chemistry among signaling molecules leads to combinatorial numbers of possible interactions. The problem of information overload often impels researchers to develop an excessively narrow focus on a few specific molecules, thus possibly missing critical interactions. A closely related problem for the systems modeler is the lack of specific, easily accessible information on quantitative biochemical parameters such as K_m, V_{max}, K_f, K_b, or local subcellular signaling volumes. Thus, it becomes important to use database management systems for highly reliable data retrieval and efficient information management (*1–3*). Use of databases also requires addressing the problems of synonymy and polysemy inherent to the signaling field.

Data requirements, of course, depend on research needs. This has been the motivation for setting up field-specific databases that present relevant information to practitioners of a specific field. For example, biologists refer to the National Institute of Health's genetic sequence database, GenBank® (http://www.ncbi.nlm.nih.gov/Genbank/index.html) for an annotated collection of all publicly available DNA sequences. Much more specialized data is available through efforts such as The Arabidopsis Information Resource (http://www.arabidopsis.org/) provides a comprehensive resource for the scientific community working with Arabidopsis thaliana. TAIR is a portal that integrates different types of information and tries to be a database resource for the Arabidopsis user community (*4*).

Studies (*5*) have shown that scientists generally obtain information from three major sources: their own experiments, personal communication with other scientists, and textual material (print or electronic). De Groote *et al.* (*6*) have shown that users often prefer online resources to print, with convenience and full-text availability being key factors for selecting online resources. De Groote's study has suggested that databases without links to full

*To whom correspondence should be addressed. Phone: +91-80-23636-421, Fax: +91-80-23636-662, bhalla@ncbs.res.in

text and online journal collections without links from bibliographic databases tend to have lower use. This shift towards online data access also emphasizes the increasingly important role of electronic resources for biochemical modeling.

3. Databases as kinetic modeling resources for cell signaling

Databases like KEGG (http://www.genome.jp/kegg/pathway.html) (*7*) and BioCyc (http://www.biocyc.org/) (*8*) are valuable comprehensive resources for metabolic network modeling. Cellular signaling currently lacks equally extensive and well-curated databases, though information gateways like AfCS (http://www.signaling-gateway.org/) (*9*) are efforts towards this goal. Most cell signaling databases deal with a specific subset of information in the field.

Model building broadly involves three closely interlinked steps: defining the model components and interactions, obtaining parameters, and deciding on the level of detail to use for the model. The source of information at each step in model building is provided by various kinds of data resources. Based on the general types of information presented in different databases these data sources fulfill different roles in biological systems modeling and can broadly be classified as detailed in the sections below.

Representative examples of freely available databases classified according to this grouping are available at http://www.ncbs.res.in/~bhalla/modeling_resources/index.html as an online supplement. URL decay is a problem for cited online resources (*10*); the URLs listed in the classification were checked to be current and functional as of the time of writing.

3.1. Diagram resources. Databases that are diagram centric, range in their representations of pathway diagrams from block diagram depictions available in BioCarta (http://www.biocarta.com/genes/index.asp) to more detailed network depictions like the Kohn interaction maps (*11*). The diagrams range from static depictions on BBID (http://bbid.grc.nia.nih.gov/) (*12*) to interactive pathway diagram displays like eMIMs (http://discover.nci.nih.gov/mim/home.jsp) (*13*), *Science*'s STKE: The Connections Maps Database (http://stke.sciencemag.org/cm/) (*14*), and Boehringer Mannheim Biochemical Pathways Wall Chart (http://www.expasy.org/cgi-bin/search-biochem-index).

Images are mostly stored in bitmap graphic formats like GIF, JPEG, or the PNG (*15*). The disadvantage of such a representation is that it is a fixed-size representation and the text within the images cannot be queried. At least two XML based formats now exist for searchable diagrammatic representations of signaling pathways. A low-level diagram graphics specification level utilizes the SVG specification. SBML incorporates a reaction layout extension, which provides similar functionality at a higher level of abstraction. In addition to the ease of size reduction and enlargement, such formats allows for searchable text within the images. Resources like STKE Connections Maps and eMIMs use this format for pathway representation. It is likely that such searchable formats will become more important in online resources for biochemical modeling.

3.2. Model depositories. Several useful model collections are accessible through the web. Some are collections of pre-existing quantitative models maintained by developers of the model and are specific to a field. For example, the Millar Group Circadian Rhythm model collection (http://template.bio.warwick.ac.uk/staff/amillar/PEBrown/CircadianModelling/NewModels/NewModels.htm) is devoted to models of the cell cycle. Others are general collections of models that are based on published mathematical models taken from journals, conference proceedings, and from textbook defined pathways. These repositories are usually just a plain listing of different models *e.g.* CellML model repository (http://www.cellml.org/examples/repository/) (*16*, *17*), SBML model repository (http://sbml.org/models/) (*18*) and do not allow searching for parameters attached to individual reactions and molecules. The models are listed individually or are available in a compressed package. Though model depositories such as these are important resources of quantitative information they may be difficult to query for information on a specific reaction or molecule. Access to specific parameters may sometimes involve downloading the model and simulator and running it to explore model parameters.

3.3. Specialized databanks. There are now many specialized electronic resources that contain data useful for modeling. BRENDA (http://www.brenda.uni-koeln.de) (*19*) provides quantitative information on enzymes, including parameters such as K_m, V_{max}, and specific activity. Kinetics of specific reactions is available on KDBI (http://xin.cz3.nus.edu.sg/group/kdbi/kdbi.asp) (*20*). Thermodynamic parameters are provided on ProTherm (http://gibk26.bse.kyutech.ac.jp/jouhou/protherm/protherm.html) (*21*). Information gateways are a distinct category of data sources for obtaining signaling specific information that could be used in modeling. Examples include AfCS-Nature Signaling Gateway and Science's STKE (http://stke.sciencemag.org/). These information gateways are different from specific databases because the information available here is from different databases and the gateway provides for common interface for dissemination of information.

3.4. Searchable model depositories. Searchable model depositories combine the features of specialized databanks with the availability of a working model. These allow the query for molecules and reactions and thus the specific quantitative parameters in the model are made available independent of downloading and running a model. Many models use a modular design to represent signaling pathways; databases can facilitate the reuse of parts of complex models by either making available the model in parts or by displaying model details in a modular fashion. DOQCS (http://doqcs.ncbs.res.in/) (*22*) is an example of a searchable model depository that provides search features to obtain biochemical interactions and parameters of reactions, enzymes and molecules comprising working models in the database. The design of models on DOQCS allows reuse of models both in parts and as a whole. In addition to being a database of current models, DOQCS is also a database of record for previously published models. SigPath (http://icb.med.cornell.edu/crt/SigPath/index.xml) (*23*) is an example of an information management system that similarly allows

search of parameters attached to models held in the repository.

3.5. Online modeling databases. Only a few databases allow online simulation of their models. Online simulation of models has the advantage that the user can explore model behavior without downloading the model and simulator. JWS Model Database (http://jjj.biochem.sun.ac.za/database/) (*24*) and Budding Yeast Cell Cycle Database (http://leibniz.biol.vt.edu/research/budding_yeast_model/pp/index.php) are examples of databases that allow online simulation of their models. It is still not possible on these databases to query for specific parameters without having to "see" the model though such flexibility would be useful. A slightly different focus is provided by the E-Cell (http://www.e-cell.org/) (*25*) and Virtual Cell (http://www.nrcam.uchc.edu/vcellR3/login/login.jsp) (*26*) projects, which have a collection of models available to users but are primarily on-line simulation resources.

Overall, it is clear that most current pathway databases have different niche strengths. Most of the databases provide information from only a small domain that can be used for modeling. No single online pathway database in the public domain has all the features required to be a single-point entry for signal transduction pathway modeling. Given the rapid diversification of the field, it is perhaps valuable to have a thriving ecosystem of different database specialties. Since no database fulfills all these requirements there is a need for database interoperability to facilitate modeling of signaling pathways.

4. What would make an ideal resource for modeling cell signaling?

What would a modeler like from a database? The best thing would be to have a suitable model already available. The next best would be to have ready access to the kinds of data needed to build the model: interaction topology, parameters, and detailed literature sources. In either case, the modeler would need an effective visualization mechanism to navigate the complexity of the signaling network. There are many existing data resources that have niche strengths in the areas mentioned above. A model centric database could allow for seamless interoperability with these resources. We consider some of these attributes below.

4.1. Model reusability. The model centric database should store models in a manner that facilitates easy reuse of both complete and part models. A Web-based simulator plugged in to the database would allow a user to query and run simulations from any remote location. It would be possible to run models from different simulation/analysis tools if the simulator uses standard formats like SBML (*18*), CellML (*16*, *17*), or UML (*27*) as an input or the database is capable of converting general formats like ODEs or MATLAB (MathWorks, Natick, MA; http://www.mathworks.com/) into the form that the simulator uses. An extensible design or data classes in a database allow addition of new data types from future models.

4.2. Efficient search. The usefulness of a database depends on the relevance of the results that it provides in response to a user query. URL based navigation is a fast way to achieve limited functionality but databases need to add support for complex queries (*28*). Form based queries are clearly desirable as they enable rapid execution of complex queries. A model centric database would need to handle different flavors of searches.

In order to successfully handle searches for parameters independent of the model, the database would need to address the problems in biomolecular nomenclature due to the tendency of modelers to use abbreviations or generic molecular names. To do this it would be useful for the database to map model entries to a standard ontology (*29–31*) for signaling pathways.

The database could feature search algorithms that make it possible to extract biological relationships directly from bibliographic databases like PubMed, full journal articles or even a user defined collection of unstructured text (*32*). Automated text mining has progressed from recognition of protein names (*33*, *34*) and protein interactions to identifying cellular location (*35*, *36*) and even identification of kinetic parameters (*37*). Resources like iHOP (*38*) are useful for extracting protein interaction data from PubMed indexed abstracts through automated text-mining. The user can then choose to custom build a protein interaction network on iHOP.

It would be useful to have an interaction data search on a pathway modeling database to allow for identification of putative pathways and molecular interactions (*39–42*) from in-house datasets as well as publicly available resources. There are various (*43*, *44*) strategies that are commonly used for interaction prediction from genomic or proteomic data alone and by a combined search of expression data and literature (*45*). Covert et al (*46*) have reconstructed, on the basis of information derived from literature and databases, an integrated genome-scale computational model of a transcriptional regulatory and metabolic network in *E. coli*.

4.3. Pathway visualization. The pathway visualization tool on the database depicts biological interactions at different levels of details. Switching of the pathway depiction from block-diagram representation where the detailed knowledge of all possible reaction paths is not required, to detailed reaction-level diagrams that are totally unambiguous and suitable for simulation allows easy change of perspectives. At present there is no consensus on the drawing of such schemes. These vary from BioCarta style cartoon representations, to wiring diagrams like the Kohn interaction diagrams that borrow graphical notations from electronics, to 3-D representations of networks (*47*, *48*). The challenge for an accurate, complete, and comprehensible organization of signaling pathways is non-trivial.

5. Conclusions

Electronic data resources for biochemical modeling have come a very long way in a short time. The current survey reveals an exuberant diversity of database types, representations, and target audiences. There are two main directions that seem to emerge from the survey. The first is that there is a strong move to provide for interoperability at many levels. These include model description languages like SBML and CellML, interlinking of databases, ontologies for reaction nomenclature and a few database portals that combine many resources.

The second direction is the strong model-driven nature of database efforts. A few databases already have attached modeling resources, and conversely several

modeling efforts and simulation environments have associated databases. This is an extremely fruitful direction in a field where technical advances in experiments and modeling must drive the data handling capabilities of databases. For example, it is likely that the emergence of high-resolution quantitative microscopy methods and corresponding 3-dimensional simulation techniques will provide an impetus to new data handling and search techniques for such data.

Model centric databases bring data from multiple and independent sources together. This makes it possible to query the behavior of biological processes at a systems level. These databases help in organizing of the huge volumes of data into information in a manner tractable to the human mind.

Web-extras as supplement for the paper: Representative examples of freely available databases classified according to the grouping discussed in the paper will be made available at http://www.ncbs.res.in/~bhalla/modeling_resources/index.html as an online supplement. We thank Sriram M. Ajay for comments on the manuscript. USB is a Senior Research Fellow of the Wellcome Trust. We acknowledge support from the Wellcome Trust, NCBS and Biophase Simulations, Inc. to HRGV, SJV and USB.

REFERENCES

1. Silberschatz, A., Korth, H.F., and Sudarshan, S. (2001) *Database Systems Concepts*, McGraw-Hill Columbus, OH
2. Date, C.J. (2003) *An Introduction to Database Systems*, Addison Wesley Boston, MA
3. Ramakrishnan, R. and Gehrke, J. (2002) *Database Management Systems*, McGraw-Hill Columbus, OH
4. Huala, E., Dickerman, A.W., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, M., Huang, W., Mueller, L.A., Bhattacharyya, D., Bhaya, D., Sobral, B.W., Beavis, W., Meinke, D.W., Town, C.D., Somerville, C., and Rhee, S.Y. (2001) The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. *Nucleic Acids Res.* **29**, 102–105
5. Grefsheim, S., Franklin, J., and Cunningham, D. (1991) Biotechnology awareness study, Part 1: Where scientists get their information. *Bull. Med. Lib. Assoc.* **79**,36–44
6. De Groote, S.L. and Dorsch, J.L. (2003) Measuring use patterns of online journals and databases. *J. Med. Lib. Assoc.* **91**, 231–240
7. Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32**, Database issue, D277–D280
8. Karp, P.D. and Riley, M. (1993) Representations of metabolic knowledge. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1**, 207–215
9. Gilman, A.G., Simon, M.I., Bourne, H.R., Harris, B.A., Long, R., Ross, E.M., Stull, J.T., Taussig, R., Arkin, A.P., Cobb, M.H., Cyster, J.G., Devreotes, P.N., Ferrell, J.E., Fruman, D., Gold, M., Weiss, A., Berridge, M.J., Cantley, L.C., Catterall, W.A., Coughlin, S.R., Olson, E.N., Smith, T.F., Brugge, J.S., Botstein, D., Dixon, J.E., Hunter, T., Lefkowitz, R.J., Pawson, A.J., Sternberg, P.W., Varmus, H., Subramaniam, S., Sinkovits, R.S., Li, J., Mock, D., Ning, Y., Saunders, B., Sternweis, P.C., Hilgemann, D., Scheuermann, R.H., DeCamp, D., Hsueh, R., Lin, K.M., Ni, Y., Seaman, W.E., Simpson, P.C., O'Connell, T.D., Roach, T., Choi, S., Eversole-Cire, P., Fraser, I., Mumby, M.C., Zhao, Y., Brekken, D., Shu, H., Meyer, T., Chandy, G., Heo, W.D., Liou, J., O'Rourke, N., Verghese, M., Mumby, S.M., Han, H., Brown, H.A., Forrester, J.S., Ivanova, P., Milne, S.B., Casey, P.J., Harden, T.K., Doyle, J., Gray, M.L., Michnick, S., Schmidt, M.A., Toner, M., Tsien, R.Y., Natarajan, M., Ranganathan, R., and Sambrano, G.R. (2002) Overview of the Alliance for Cellular Signaling. *Nature* **420**, 703–706
10. Wren, J.D. (2004) 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics* **20**, 668–672
11. Kohn, K.W. (1998) Functional capabilities of molecular network components controlling the mammalian G1/S cell cycle phase transition. *Oncogene* **16**, 1065–1075
12. Becker, K.G., White, S.L., Muller, J., and Engel, J. (2000) BBID: the biological biochemical image database. *Bioinformatics* **16**, 745–746
13. Aladjem, M.I., Pasa, S., Parodi, S., Weinstein, J.N., Pommier, Y., and Kohn, K.W. (2004) Molecular interaction maps–a diagrammatic graphical language for bioregulatory networks. *Sci. STKE* **2004**, pe8
14. Gough, N.R. (2002) Science's signal transduction knowledge environment: the connections maps database. *Ann. N. Y. Acad. Sci.* **971**, 585–587
15. Wiggins, R.H., 3rd, Davidson, H.C., Harnsberger, H.R., Lauman, J.R., and Goede, P.A. (2001) Image file formats: past, present, and future. *Radiographics* **21**, 789–798
16. Cuellar, A.A., Nielsen, P.F., Bullivant, D.P., and Hunter, P.J. (2003) CellML 1.1 for the Definition and Exchange of Biological Models. in *Conf. Proc. 2003 IFAC Symposium on Modelling and Control in Biomedical Systems* (Feng, D. and Carson, E., eds.) 2003 Nov., pp. 451–456, Elsevier
17. Lloyd, C.M., Halstead, M.D., and Nielsen, P.F. (2004) CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* **85**, 433–450
18. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, II, Hedley, W.J., Hodgman, T.C., Hofmeyr, J.H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., and Wang, J. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531
19. 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
20. Ji, Z.L., Chen, X., Zhen, C.J., Yao, L.X., Han, L.Y., Yeo, W.K., Chung, P.C., Puy, H.S., Tay, Y.T., Muhammad, A., and Chen, Y.Z. (2003) KDBI: Kinetic Data of Bio-molecular Interactions database. *Nucleic Acids Res.* **31**, 255–257
21. Gromiha, M.M., An, J., Kono, H., Oobatake, M., Uedaira, H., and Sarai, A. (1999) ProTherm: Thermodynamic Database for Proteins and Mutants. *Nucleic Acids Res.* **27**, 286–288
22. Sivakumaran, S., Hariharaputran, S., Mishra, J., and Bhalla, U.S. (2003) The Database of Quantitative Cellular Signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics* **19**, 408–415
23. Campagne, F., Neves, S., Chang, C.W., Skrabanek, L., Ram, P.T., Iyengar, R., and Weinstein, H. (2004) Quantitative information management for the biochemical computation of cellular networks. *Sci. STKE* **2004**, pl11
24. Snoep, J.L. and Olivier, B.G. (2002) Java Web Simulation (JWS); a web based database of kinetic models. *Mol. Biol. Rep.* **29**, 259–263
25. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C.A. (1997) E-CELL: Software Environment for Whole Cell Simulation. *Genome Inform. Ser. Workshop Genome Inform.* **8**, 147–155
26. Schaff, J., Fink, C.C., Slepchenko, B., Carson, J.H., and Loew, L.M. (1997) A general computational framework for modeling cellular structure and function. *Biophys. J.* **73**, 1135–1146

27. Roux-Rouquie, M., Caritey, N., Gaubert, L., and Rosenthal-Sabroux, C. (2004) Using the Unified Modelling Language (UML) to guide the systemic description of biological processes and systems. *Biosystems* **75**, 3–14
28. Karp, P.D. (1996) Database links are a foundation for interoperability. *Trends Biotechnol.* **14**, 273–279
29. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29
30. Schulze-Kremer, S. (2002) Ontologies for molecular biology and bioinformatics. *In Silico Biol.* **2**, 179–193
31. Takai-Igarashi, T. and Mizoguchi, R. (2004) Cell signaling networks ontology. *In Silico Biol.* **4**, 81–87
32. Hirschman, L., Park, J.C., Tsujii, J., Wong, L., and Wu, C.H. (2002) Accomplishments and challenges in literature data mining for biology. *Bioinformatics* **18**, 1553–1561
33. Tanabe, L. and Wilbur, W.J. (2002) Tagging gene and protein names in biomedical text. *Bioinformatics* **18**, 1124–1132
34. Yu, H., Hatzivassiloglou, V., Friedman, C., Rzhetsky, A., and Wilbur, W.J. (2002) Automatic extraction of gene and protein synonyms from MEDLINE and journal articles. *Proc. AMIA Symp.* 919–923
35. Lu, Z., Szafron, D., Greiner, R., Lu, P., Wishart, D.S., Poulin, B., Anvik, J., Macdonell, C., and Eisner, R. (2004) Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics* **20**, 547–556
36. Stapley, B.J., Kelley, L.A., and Sternberg, M.J. (2002) Predicting the sub-cellular location of proteins from text using support vector machines. *Pac. Symp. Biocomput.* 374–385
37. Hakenberg, J., Schmeier, S., Kowald, A., Klipp, E., and Leser, U. (2004) Finding kinetic parameters using text mining. *Omics* **8**, 131–152
38. Hoffmann, R. and Valencia, A. (2004) A gene network for navigating the literature. *Nat. Genet.* **36**, 664
39. Akutsu, T., Miyano, S., and Kuhara, S. (2000) Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* **16**, 727–734
40. Li, Z. and Chan, C. (2004) Inferring pathways and networks with a Bayesian framework. *FASEB J.* **18**, 746–748
41. Genoud, T., Trevino Santa Cruz, M.B., and Metraux, J.P. (2001) Numeric simulation of plant signaling networks. *Plant Physiol.* **126**, 1430–1437
42. Liao, J.C., Boscolo, R., Yang, Y.L., Tran, L.M., Sabatti, C., and Roychowdhury, V.P. (2003) Network component analysis: reconstruction of regulatory signals in biological systems. *Proc. Natl Acad. Sci. USA* **100**, 15522–15527
43. Huynen, M., Snel, B., Lathe, W., 3rd, and Bork, P. (2000) Predicting protein function by genomic context: quantitative evaluation and qualitative inferences. *Genome Res.* **10**, 1204–1210
44. Sun, N. and Zhao, H. (2004) Genomic approaches in dissecting complex biological pathways. *Pharmacogenomics* **5**, 163–179
45. Sohler, F., Hanisch, D., and Zimmer, R. (2004) New methods for joint analysis of biological networks and expression data. *Bioinformatics* **20**, 1517–1521
46. Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., and Palsson, B.O. (2004) Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 92–96
47. Han, K. and Byun, Y. (2004) Three-dimensional visualization of protein interaction networks. *Comput. Biol. Med.* **34**, 127–139
48. Jenster, G. (2004) A visualisation concept of dynamic signalling networks. *Mol. Cell. Endocrinol.* **218**, 1–6