J. Microbiol. Biotechnol. (2006), **16**(6), 832–848

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Resources for Systems Biology Research

KIM, JIN SIK[1,2], HONGSEOK YUN[1,2,3], HYUN UK KIM[1], HYUNG SEOK CHOI[1], TAE YONG KIM[1], HAN MIN WOO[1], AND SANG YUP LEE[1,2]*

[1]*Metabolic and Biomolecular Engineering National Research Laboratory, Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea*
[2]*Department of BioSystems, BioProcess Engineering Research Center and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea*
[3]*Macrogen Corp., World Meridian Venture Center 10F, #60-24, Gasan-dong, Geumcheon-gu, Seoul, Korea*

Received: March 19, 2006
Accepted: May 8, 2006

Abstract Systems biology has recently become an important research paradigm that is anticipated to decipher the metabolic, regulatory, and signaling networks of complex living organisms on the whole organism level. Thus, various research outputs are being generated, along with the development of many tools and resources for systems biology research. Accordingly, this review provides a comprehensive summary of the current resources and tools for systems biology research that will hopefully be helpful to researchers involved in this field. The resources are categorized into the following five groups: genome information and analysis, transcriptome and proteome databases, metabolic profiling and metabolic control analysis, metabolic and regulatory information, and software for computational systems biology. A summary table and some future perspectives are also provided.

Key words: Systems biology, genome, transcriptome, proteome, static/dynamic simulation

As the number of complete genome sequences continues to increase, the information generated from sequencing projects is exploding [40]. In addition, the amount of subsequent data generated by utilizing genome sequences and related information, including transcriptome, proteome, metabolic, and regulatory networks, is also rapidly increasing. Thus, the combination of computational methods, mathematics, statistics, and modeling knowledge together with traditional and advanced biological methods has become indispensable to obtain biologically meaningful data from such a flood of information [34]. This kind of interdisciplinary research then enables a detailed understanding of the behavior of cells and their characteristics. Within this context, systems biology is a general term used for new biological research that combines wet and *in silico* experiments to decipher the complex behavior of cells and organisms at the whole system level.

The number of resources dedicated to systems biology research has increased significantly in recent years, ranging from internet resources for the sequencing and analysis of genomes to tools for the modeling and simulation of cellular and regulatory networks. These internet resources contain various tools and applications that enable researchers to use the Web site contents, plus information imported from other research groups.

Here, the resources and tools for systems biology research are reviewed based on five categories (Table 1): genome information and analysis, transcriptome and proteome databases, metabolic profiling and metabolic control analysis, metabolic and regulatory information (databases), and software for computational systems biology and static/dynamic simulation.

SYSTEMS BIOLOGY: WEB SITES, TOOLS, AND OTHER RESOURCES

Genome Information and Analysis

***Arabidopsis thaliana* Information Resource [48].** The Arabidopsis Information Resource (TAIR) is specifically for groups working with *Arabidopsis thaliana* and contains an extensive database for the efficient querying of information. This information can be viewed using an interactive MapViewer and analyzed using various tools introduced by the Web site. TAIR also provides several bioinformatic tools for sequence comparison, motif prediction, and graph visualization. For example, users can view the sequences of

*Corresponding author
Phone: 82-42-869-3930; Fax: 82-42-869-3910;
E-mail: leesy@kaist.ac.kr

Table 1. Web sites for systems biology research.

Site description	Web site address	Reference
Genome Information and Analysis		
Arabidopsis thaliana information resource TAIR	http://www.arabidopsis.org/	[48]
Broad Institute	http://www.broad.mit.edu/genome_bio/	[41]
ENSEMBL e! genome browser	http://www.ensembl.org/	[3]
Gramene	http://www.gramene.org/	[21]
Mouse (*Mus musculus*) genome informatics	http://www.informatics.jax.org/	[4]
PEC - Profiling of *E. coli* Chromosome	http://www.shigen.nig.ac.jp/ecoli/pec/index.jsp	-
PhenomicDB	http://www.phenomicdb.de/	[23]
PIR Georgetown University genomes	http://www-nbrf.georgetown.edu/pir/genome.html	[59]
SGD	http://www.yeastgenome.org	[18]
Stanford University genome resources	http://genome-www.stanford.edu/	[2]
The Institute of Genomic Research (TIGR)	http://www.tigr.org/	[35]
Wellcome Trust Sanger Institute	http://www.sanger.ac.uk/	-
U.S. Dept. of Energy Joint Genome Initiative	http://www.jgi.doe.gov/	-
Washington University Genome Center	http://genome.wustl.edu/	-
Transcriptome and Proteome Databases		
Arrayexpress, European Bioinformatics Institute	http://www.ebi.ac.uk/arrayexpress	[46]
GEO, Gene expression repository	http://www.ncbi.nlm.nih.gov/geo	[10]
MAD, Jackson Labs.	http://mad.jax.org	-
MGED - Microarray Gene Expression Data Society	http://www.mged.org/	[58]
Yale Microarray Database (YMD)	http://info.med.yale.edu/microarray/	[8]
ExPASy Proteomics Server	http://tw.expasy.org/	[14]
Pfam	http://pfam.wustl.edu/index.html	[11]
Protein Lounge	http://www.proteinlounge.com/	-
UCSF Protein Prospector	http://prospector.ucsf.edu/	[7]
Metabolic Profiling and Metabolic Control Analysis		
13C-FLUX Web Page	http://www.simtec.mb.uni-siegen.de/software_13cflux.0.html	-
Atomic Reconstruction of Metabolism	http://www.metabolome.jp/index.html	-
Mass Spectrometry and Biotechnology Resource	http://www.ionsource.com	-
Metabolic Control Analysis and Rate-Limiting Steps	http://www.bi.umist.ac.uk/users/mjfprb/2IRM/s003p007.asp	-
Metabolomics Society	http://www.metabolomicssociety.org/	-
Metabolic and Regulatory Information (Databases)		
aMAZE	http://www.amaze.ulb.ac.be/	[36]
BioCyc	http://biocyc.org/	[29]
BioSilico	http://biosilico.kaist.ac.kr	[19]
BRENDA	http://www.brenda.uni-koeln.de/	[51]
EcoCyc	http://ecocyc.org/	[26]
KEGG	http://www.kegg.com/	[24]
MetaCyc	http://metacyc.org/	[6]
PANTHER Pathways	https://panther.appliedbiosystems.com/pathway/	[44]
PathArt	http://jubilantbiosys.com/pd.htm	-
RegulonDB	http://www.cifn.unam.mx/Computational_Genomics/regulondb/	[49]

A. thaliana using a SeqViewer and analyze the biochemical pathways using the AraCyc developed based on BioCyc (http://biocyc.org) [29].

Broad Institute [41]. The genome biology program of the Broad Institute mainly focuses on the analysis of human, microbial, and fungal genomes and includes the genomes, genome regulation, and cell circuits. The Broad Institute has also been involved in several genome sequencing projects, including mammalian, fungal, and bacterial genomes, and provides the resulting genome information, single

Table 1. Continued.

Site description	Web site address	Reference
Software for Computational Systems Biology, Static/Dynamic Simulation		
BIOCHAM	http://contraintes.inria.fr/BIOCHAM/	-
BioCharon	http://www.cis.upenn.edu/biocomp/new_html/index.php3	-
Bioglyphics & BALSA	http://www.csi.washington.edu/teams/modeling/projects/BALSA/	-
BioModels.net	http://biomodels.net/	-
BioNetGen	http://cellsignaling.lanl.gov/bionetgen/	[5]
BioSens	http://www.chemengr.ucsb.edu/~ceweb/faculty/doyle/biosens/BioSens.htm	-
Bio-SPICE	https://biospice.org/	[13]
BioTapestry	http://labs.systemsbiology.net/bolouri/software/BioTapestry/	-
BioUML	http://www.biouml.org	-
BSTLab	https://bioinformatics.musc.edu/bstlab	-
CADLIVE	http://kurata21.bse.kyutech.ac.jp/cadlive	[30]
CellDesigner	http://www.celldesigner.org	-
Cellerator	*http://www-aig.jpl.nasa.gov/public/mls/cellerator*	[53]
CellML	http://www.cellml.org	[37]
Cellware	http://www.bii.a-star.edu.sg/research/sbg/cellware/index.asp	[9]
COPASI	http://www.copasi.org	-
CSB.DB	http://csbdb.mpimp-golm.mpg.de	[55]
Cytoscape	http://www.cytoscape.org/	[52]
Dizzy	http://labs.systemsbiology.net/bolouri/software/Dizzy/	[47]
E-Cell	http://ecell.sourceforge.net/	[56]
FluxAnalyzer	http://www.mpimagdeburg.mpg.de/en/research/projects/1010/1014/1020/mfaeng/fluxanaly.html	[28]
Gepasi	http://www.gepasi.org/	[43]
INSILICO Discovery	http://www.insilico-biotechnology.com/f_products.html	-
Jarnac: Systems Biology Software	http://www.cds.caltech.edu/~hsauro/Jarnac.htm	-
JWS Online	http://jjj.biochem.sun.ac.za/index.html	[45]
KINSOLVER	http://lsdis.cs.uga.edu/~aleman/kinsolver/	-
MesoRD	http://mesord.sourceforge.net/	[17]
MetaboLogica	http://www.metabologica.com/	-
MetaFluxNet	http://mbel.kaist.ac.kr/mfn/	[33]
MMT2	http://www.simtec.mb.uni-siegen.de/software_mmt2.0.html	-
Monod	http://www.molsci.org/Dispatch?action-WebdocWidget:4884-detail=1	-
Moleculizer	http://www.molsci.org/~lok/moleculizer/index.html	[38]
NetBuilder	http://strc.herts.ac.uk/bio/maria/NetBuilder/	-
Pathway Simulation Tool (runsbml)	http://www.ariadnegenomics.com/technology/simulation.html	-
PaVESy	http://pavesy.mpimp-golm.mpg.de/PaVESy.htm	[39]
PROTON	http://tunicata.techfak.uni-bielefeld.de/proton/web/main.jsp?dir=Home	-
Reactome	http://www.reactome.org	[22]
SigPath	http://www.sigpath.org/	-
SigTran	http://csi.washington.edu/teams/modeling/projects/sigtran/	-
SmartCell	http://www.embl.de/ExternalInfo/serrano/smartcell/	-
StochSim	http://www.anat.cam.ac.uk/~compcell/StochSim.html	[31]
STOCKS 2.0: Stochastic simulations of biochemical reaction networks	http://www.sysbio.pl/stocks/	[27]
The PNK 2e	http://page.mi.fu-berlin.de/~trieglaf/PNK2e/index.html	-
Virtual Cell	http://www.vcell.org/	[54]
Worldwide PDB (protein databank)	http://www.wwpdb.org/	-
WebCell	http://webcell.kaist.ac.kr/	[32]

nucleotide polymorphisms (SNPs), cancer datasets, and complete sets of the small inhibitory RNAs (RNAi). In addition, for the efficient analysis of genome-scale data sets, the Broad Institute has developed and distributed various programs, such as Arachne and Argo, which are used for assembling genome sequences from whole genome shotgun reads and for visualizing and manually annotating full genomes, software for the haplotype analysis and image generation of chromosome regions to elucidate genetic variations, and gene expression analysis and visualization tools for genome-wide expression and statistical validation of different experimental sets, which also relate the procedures from genome sequences to transcriptome data.

ENSEMBL e! Genome Browser [3]. Ensembl is a joint project between the European Molecular Biology Laboratory (EMBL) - European Bioinformatics Institute (EBI) and the Wellcome Trust Sanger Institute (WTSI) to develop a platform system that can automatically generate eukaryotic genomes. The objective of the Ensembl project is described as the automatic analysis of genome information, maintenance and analysis of currently deposited data, Web-based representation of the data, and distribution of the analysis results. The species included in the Ensembl project belong to three groups: mammals, chordates, and eukaryotes, and 17 species, such as *Homo sapiens*, *Mus musculus*, and *Caenorhabditis elegans*, are used for the analysis.

Gramene: A Comparative Mapping Resource for Grains [21]. Gramene is an open-source resource for the comparative analysis of genome information on grasses. The objective of this resource is to facilitate the study of homological relationships based on sequencing, structure and functional analysis, pathway analysis, and phenotypic characteristics. These objectives are achieved using the relationships among grasses, which can be queried and visualized using Web-based tools. A sequence-based comparative analysis can be performed using the genome browser adapted from Ensembl, comparative map viewer (CMap), and protein display tools.

Mouse (*Mus musculus*) Genome Informatics [4]. Mouse Genome Informatics (MGI) provides integrated access to data on the genomics and genetics of the lab mouse. More specifically, the Mouse Genome Database (MGD) includes data on gene characterization, nomenclatures, homologies, phenotypes, allelic variants, mutants, and so forth. Databases for genome sequencing and gene expression are also provided, along with the MGD database, thereby facilitating the integration of emerging mouse genomic sequence data and different types of gene expression information. In addition, the Gene Ontology (GO) project [15] and Mouse Tumor Biology (MTB) database provide a GO browser and facilitate the efficient choice of experimental cancer models, respectively.

PEC-Profiling of *E. coli* Chromosome. The profiling of the *Escherichia coli* chromosome (PEC) database was developed to connect any relevant information that could help characterize the *E. coli* genome, and is mainly focused on discovering the function of each gene. The database is intended to offer a user-friendly interface for researchers. For example, basic information is provided on each gene, such as the gene name, gene length, and location, plus the essentiality of each gene for cell growth is also available, based on reports and deletion studies, whereas sequence comparisons and motif searches present the relationships among the genes.

PhenomicDB [23]. PhenomicDB is a phenotype-genotype database for multiple organisms, including human, mouse, fly, and other model organisms. These data have been integrated into a single resource by coarse-grained semantic mapping of the phenotypic data fields, and by including the mostly popular gene indices and orthology relationships. As the gene indices of the NCBI gene and orthologs from the HomoloGene are included in this database, users can compare phenotypes over many organisms simultaneously and obtain a wide range of information, such as an ID, organism, symbol, sequence, gene ontology, orthologs, descriptions, references, and so on.

PIR Georgetown University Genomes [59]. The Protein Information Resource (PIR) is an integrated bioinformatics resource to support genomic and proteomic research, and has provided various protein databases and tools freely to users for 40 years. PIR currently offers a variety of resources that mainly assist with the propagation and standardization of protein annotation. The PIR superfamilies (PIRSF) protein classification system is a network with multiple levels of sequence diversity from superfamilies, reflecting the evolutionary relationship of proteins and domains. Meanwhile, the PIR-International Protein Sequence Database (PIR-PSD) was produced to record the comprehensive, timeless, nonredundant, and fully-classified properties of proteins. In addition, iProLink supports text mining in the area of literature-based database curation, named entity recognition, whereas iProClass contains value-added annotation reports for UniProt proteins.

SGD [18]. The *Saccharomyces* Genome Database (SGD) is a scientific database of the molecular biology and genetics of the yeast *Saccharomyces cerevisiae*. The database provides a variety of genomic and biological information, and maintains the gene name list of *S. cerevisiae*. Various tools for the analysis and comparison of yeast genes and gene products are provided by the SGD, including basic tools for sequence homology analysis, protein motif prediction, and a metabolic pathway viewer. In addition, the database provides gene expression profile data and a querying system for multiple microarray studies.

Stanford University Genome Resources [2]. The Stanford genome resources are mainly focused on gene expression profiling and its analysis. The Stanford Microarray Database (SMD) supports user-specified analyses of

microarray experimental data and provides quality assessment of experimental data, clustering, and other data analyses using Web-based and client programs. Registered users can also obtain graphical data using visualization tools. In addition, the SMD provides tools for data exchange, data organization or annotation, and array design, etc. Several other genome databases, including *Saccharomyces*, *Arabidopsis*, *Candida*, and *Tetrahymena*, are also provided on the site.

The Institute of Genomic Research [35]. The Institute of Genomic Research (TIGR), founded in 1992, is mainly focused on the comparative, functional, and structural analysis of genomes and gene products from a wide variety of organisms, including eukaryotes, eubacteria, and viruses. Starting with the first complete genome of a pathogen, *Haemophilus influenzae* [12], the Web site now contains more than 21 microbial genomes. The TIGR Web site mainly consists of genome databases, functional genomic tools, a microbial sequencing center, and software for bioinformatic analyses. The genome databases contain comprehensive microbial resources, which are tools that allow the user to access all the bacterial genome sequences completed to date. The functional genomics section, which mainly deals with DNA microarray resources, provides external links to a variety of resources, including protocols developed at TIGR and data associated with TIGR publications. Finally, the TIGR Web site also includes many free applications related to gene finding/annotation, alignment, sequencing/finishing, microarray, grid computing, and others.

Wellcome Trust Sanger Institute. The Sanger Institute remains one of the most important sites for the analysis of genome information, providing various genome data, bioinformatics software, and genome/genetics information on Sanger projects. The main purpose is to extend genome knowledge, particularly through large-scale genome sequencing and analysis. Ensemble genome data resources for the major eukaryotic species, such as human, mouse, rat, zebra fish, and fly, are also accessible from this site, along with a software and database section for production sequencing, physical mapping, and informatic analysis. Consequently, users can assemble and confirm the sequences, and also compare them with others using various types of informatic software. The informatic applications provide support in the following areas: sequencing and mapping operations, sequence analysis/annotation, and related research topics.

U.S. Dept. of Energy Joint Genome Initiative. The genome section of the U.S. Department of Energy (DOE) Genome Initiative provides querying and downloading services for more than 121 microbial and eukaryotic genomes. An advanced search option also allows researchers to locate genes, transcripts, and proteins by directly or indirectly searching for keywords characterizing various protein sequences that have the best alignments with the transcripts. The genome browsing tool gives the predicted genes along with each scaffold of the assembly processes. The genome tools can also be directly connected to external databases and tools for comparative analyses: e.g., the Clusters of Orthologous Group (COG) of the National Center for Biotechnology Information (NCBI), Kyoto Encyclopedia of Genes and Genomes (KEGG) of Kyoto University, Basic Local Alignment Search Tool (BLAST), and so forth [24, 42, 57]. By utilizing information generated from metabolic databases, users can collect information on gene distribution according to the functional categories defined in each database.

Washington University Genome Center. The Web site of the Genome Sequencing Center at Washington University provides general information, current research status, and external links to the related sites of 31 complete genomes, mainly consisting of metazoans, including *Caenorhabditis elegans*, *Homo sapiens*, and *Mus musculus*, etc. A sequence analysis of the internal databases at the Genome Sequencing Center via a local BLAST server enables researchers to find similarities to the sequences deposited in the St. Louis project.

Transcriptome and Proteome Databases

Arrayexpress, European Bioinformatics Institute [46]. This Web page is part of the homepage of the European Bioinformatics Institute and is a comprehensive public repository of microarray data. Currently, this Web database contains information on 851 experiments, 557 arrays, and 4,453 protocols. Users can utilize text-based queries to look for appropriate array results by matching experiments, arrays, or protocols, and researchers with unique results from their microarray experiments can submit their data to the Arryexpress repository through File Transfer Protocol (FTP). The Web site also lists microarray standards and provides updates on microarray-related research news.

GEO, Gene Expression Repository [10]. The Gene Expression Omnibus (GEO) is a high-throughput gene expression data repository, as well as a curated gene expression resource for data access. Users can explore the gene expression data throughout the GEO navigator section. The GEO provides data on microarray experiments, along with a serial analysis of gene expression (SAGE) and mass spectrometry (MS) proteomic data. GEO data can be constructed using particular GEO records for the accession number and display bar or by querying all GEO data sets in a specific field corresponding to either Entrez GEO DataSets or Entrez GEO Profiles interfaces.

MAD, Jackson Labs. The Microarray Database (MAD) project was constructed based on experimental data from gene expression assays, and contains a data set from published data on the Web. The MAD data page can be retrieved for summary information about published projects. MAD data files are generated in a text format, which is delimited by tabs to enable imported data files to be converted into spreadsheet applications, such as Microsoft

Excel. The available MAD data files are also listed on the Web page.

MGED-Microarray Gene Expression Data Society [58]. The Microarray Gene Expression Data (MGED) Web site is dedicated to the efficient management of microarray data by focusing on the establishment of standards for the analysis and exchange of microarray data, and the development of microarray databases and related appropriate software. The Web site details six standardization projects: Minimum Information About a Microarray Experiment (MIAME), MicroArray and Gene Expression (MAGE), Ontologies (OWG), Transformations, Reporting Structure for Biological Investigations Working Groups (RSBI WGs), and Minimum Information Specification For *In Situ* Hybridization and Immunohistochemistry Experiments (MISFISHIE), where MIAME was established to promote unambiguous interpretation and reproduction of microarray data, MAGE facilitates the exchange of microarray data between different systems via establishing a data exchange format (MAGE-ML) and object model (MAGE-OM) for microarray experiments (standardization), OWG provide standard terms for the unambiguous description or annotation of microarray experiments, Transformations (The MGED Data Transformation and Normalization Working Group) facilitate the transformation and normalization of microarray data, RSBI WGs focus on applications of microarray techniques in the fields of toxicogenomics, nutrigenomics, and environmental genomics, and MISFISHIE standardizes the least amount of information that should be provided to interpret and verify results, and enables other research groups to reproduce and evaluate them.

YMD, Yale Microarray Database [8]. The Yale Microarray Database (YMD) was established based on the joint participation of laboratories and research centers at Yale University. The database is primarily dedicated to the storage and retrieval of microarray images and analysis data produced by hundreds of researchers. It also enables researchers to perform various statistical analyses of microarray data, some of which may or may not require software packages. Various features are available in this database, including a project management interface, the linking of query results to genome annotations, and aggregation of expression data, etc.

ExPASy Proteomics Server [14]. The ExPASy (Expert Protein Analysis System) is run by the Swiss Institute of Bioinformatics (SIB) and is a collection of databases and analytical tools in the fields of proteins and proteomics. The ExPASy databases include SWISS-PROT and TrEMBL, SWISS-2DPAGE, PROSITE, ENZYME, and the SWISS-MODEL Repository, where SWISS-PROT is a protein sequence database that includes annotated information, TrEMBL is a supplemental database of SWISS-MODEL that contains computer-annotated information, SWISS-2DPAGE is a database that contains protein information generated from 2D polyacrylamide gel electrophoresis (2D-PAGE), PROSITE is a database of protein domains and families that covers biologically meaningful patterns and profiles with respect to protein functions, ENZYME primarily deals with the nomenclature of enzymes, and SWISS-MODEL is a database of protein structure models.

In addition to these proteomic databases, ExPASy provides a number of sequence analysis tools and proteomics tools. For instance, BLAST allows similarity searches of a query protein sequence against a protein or nucleotide database, ScanProsite scans a sequence against the database and identifies all the relevant patterns, profiles, and rules, and SWISS-MODEL can generate a 3D structure of a protein with a similar sequence to that of another protein where the 3D structure is already known.

Some very important proteomics tools are also available, including FindMod for predicting the modification of a protein sequence, FindPept for identifying peptides from the unspecific cleavage of proteins, PeptideMass for calculating the theoretical masses of peptides from cleaved proteins, and PeptIdent (or TagIdent and MultiIdent) for identifying proteins using experimental information, etc. Moreover, the ExPASy databases are linked to a number of relevant databases, such as DNA Data Bank of Japan (DDBJ), Protein Data Bank (PDB), FlyBase, the Mouse Genome Database (MGD), the *Saccharomyces* Genome Database (SGD), and Pfam, which cover a wide range of molecular biology and biochemistry.

Pfam [11]. Pfam is a database of protein domain families, and contains multiple alignments of protein domains for each family. These alignments indicate evolutionarily conserved protein domains that bestow specific functions on the proteins. Profile Hidden Markov models (profile HMMs) can be very useful for categorizing an unknown protein into a family and predicting its function by searching the domain of the new protein sequence.

Protein Lounge. The Protein Lounge is a systems biology Web site that contains 8 different databases and 6 research tools that systematically organize gene and protein data. The databases are as follows: Pathway database for signal transduction and metabolic pathways, siRNA database for comprehensive information on small interfering RNA (siRNA) targets and siRNA cloning tool, Peptide-Antigen database for antigenic peptide targets in various organisms, Protein Interaction database for comprehensive information on protein-protein/protein-nonprotein interactions, Kinase-Phosphatase database for information on all kinases and phosphatases, as well as their associated substrates, Transcription Factors database for concrete information, such as classification, functions, and DNA sequences, Disease Genes database for disease-causing genes, and Protein database for information associated with proteins, including signal pathways, disease relation, and sequences, etc.

Meanwhile the research tools available on the Protein Lounge Web site are as follows: Pathways Builder Tool

for reconstruction and modification of pathways, Protein Hydroplotter for information on the hydrophilic and hydrophobic regions of a protein of interest, Peptide Finder for finding the best antigenic peptides of a given protein sequence, Easy siRNA for searching siRNA targets and cloning them online, Clone Easy for online PCR cloning, and Protein Vision for displaying the 3D protein structure.

UCSF Protein Prospector [7]. The ProteinProspector contains databases that can be searched for a protein sequence in combination with mass spectrometry experiments. The available tools are 12 Sequence Database Search programs, which seek to elucidate protein sequences using various data generated from mass spectrometry [e.g., MS-Fit that uses peptide-mass fingerprinting data from mass spectrometry (MS), and MS-Seq that uses sequence tag data from MS/MS], 5 Peptide/Protein MS Utility programs, which concentrate on elucidating various characteristics of peptides, such as their mass and amino acid composition, and a FASTA Database Manipulation/Information Tool for calculating useful statistics for a given database.

Metabolic Profiling and Metabolic Control Analysis

13C-FLUX Web Page. This Web site provides users with information on metabolic flux analysis using ^{13}C labeling experiments, including metabolic flux analysis, carbon labeling experiments (CLEs), the evaluation of CLEs, statistical analysis, and relevant literature. This Web site is focused on how to calculate intracellular fluxes using labeled experiments and obtain corresponding solutions. In addition, a full version of ^{13}C-Flux is downloadable if a license agreement is completed.

Atomic Reconstruction of Metabolism. This Web site presents the ARM (Atomic Reconstruction of Metabolism) project, powered by state-of-the-art algorithms in graph theory. The ARM digitizes the entire biochemical network, and provides solutions for metabolomic research and software technology, facilitating the investigation of metabolism. Corresponding software and documents are also available from this Web site. In addition, this Web site supports an online metabolic pathway search tool and is linked to useful Web sites for metabolome analysis, including mass spectra, structural, flux, and visualization resources.

Mass Spectrometry and Biotechnology Resource. This Web site provides total solutions for metabolomic research and resources. Users can gain access to official databases that contain basic mass spectra and protein information. The IonSource Web page contains recently updated information about carbohydrate and amino acid phosphorylations. In addition, it provides educational tutorials on the featured ion sources, metabolic information, and diverse equipment, such as mass analyzers or high-throughput technologies coupled with mass spectrometry. Useful links are also provided, along with protein database sites and daily updated conference news. This Web site is very useful as a source of mass spectrometry and biotechnology information.

Metabolic Control Analysis and Rate-Limiting Steps. This Web site is part of a Web-based tutorial section and explains different concepts with appropriate examples, providing a very helpful introduction to metabolic control analysis. Moreover, readers can find useful links throughout the Web site and simulate metabolic networks with online user-defined parameters for self-study. This Web site is very useful for students, especially in the areas of pathway structures and metabolic regulatory systems.

Metabolomics Society. The Metabolomics Society is a nonprofit and independent organization and has held annual meetings since 2004. The Web site provides society members with useful metabolomics news, including the development of metabolomics around the world, collaborative research work in metabolomics and related fields, and the promotion of noteworthy metabolomics research publications, etc.

Metabolic and Regulatory Information (Databases)

aMAZE [36]. This Web site is dedicated to the efficient management and analysis of complex information on biological systems generated by high-throughput techniques. It is composed of specialized databases that represent information on metabolic pathways, protein-protein interactions, gene regulation, and signal transduction.

The Web site employs a system called aMAZE Workbench for representing and analyzing molecular interactions and cellular processes. This workbench allows a wide range of cellular information to be stored, flexible querying and information visualization, advanced programmatic analyses, and the efficient generation of cellular process data.

Another two important features of aMAZE are aMAZE LightBench and SigTrans Bench, where the former is a Web-based resource that allows simple browsing of information on chemical reactions, genes, and enzymes within metabolic pathways, as well as transcriptional regulations of the corresponding genes, whereas the latter is a database on signal transduction.

BioCyc [29]. BioCyc is a collection of Pathway/Genome databases that provide information on metabolic pathways and genomes of different organisms. The databases within BioCyc can be divided into three tiers, depending on the number of reviews and updates they have received. The first tier includes databases that have undergone intensive manual reviews and are still being updated, such as EcoCyc, which describes *Escherichia coli* K-12 [26], and MetaCyc, which comprehensively describes enzymes and metabolic pathways from more than 300 organisms [6]. The BioCyc Open Chemical Database (BOCD) also belongs to this tier and includes descriptions of metabolites, enzyme activators, inhibitors, and cofactors. The second

tier currently includes 17 databases that have undergone a moderate number of reviews and updates, generally less than those in the first tier. The third tier includes 142 databases that have not undergone any reviews or updates.

BioCyc is particularly useful for studying individual biochemical reactions, metabolic pathways, and the related location(s) of the gene(s). Moreover, it provides comprehensive genomic information and an Omics Viewer designed for analyzing gene expression, proteomics, and metabolomics data.

BioSilico [19]. BioSilico is a Web-based database system that integrates several public databases mainly for metabolic pathways. Thus, it serves as a gateway to databases such as KEGG/LIGAND [16], ENZYME [1], BRENDA [51], EcoCyc [26], and MetaCyc [6]. Access to these databases is facilitated with common interfaces that have different data contents within the scope of the same topic.

BioSilico is regularly updated and developed, and will soon enable users to design metabolic pathways. Moreover, it will be linked to other Web-based analytical tools/software, including the metabolic flux analysis program package MetaFluxNet [33].

BRENDA [51]. BRENDA is a database that comprehensively covers enzyme functional data. The main contribution of this database is that it systematically collects enzyme data that are normally scattered among journals in various fields. The main features of the database include Nomenclature (Enzyme Name, EC Number, etc.), Reaction & Specificity (Pathway, Reaction Type, Substrates, Inhibitors, etc.), Functional Parameters (K_m Value, Specific Activity, pH, & Temperature Optimum, etc.), Organism-related Information, Enzyme Structure (Sequence/SwissProt Link, 3-D Structure, Posttranslational Modification, etc.), Isolation & Preparation (Purification, Crystallization, etc.), Stability (pH & Temperature Stability, etc.), Disease & References, and Application & Engineering.

EcoCyc [26]. EcoCyc is one of the databases within BioCyc and comprehensively describes the genome and metabolism of *E. coli* K-12 MG1655. The main features of EcoCyc can be organized into 4 areas; genome, transcriptional regulation, membrane transporters, and metabolism. EcoCyc contains the full genome sequence of *E. coli* and provides detailed descriptions of every gene and its location. The content is regularly updated based on published literature. For transcriptional regulation, it details the components involved in transcriptional regulation, such as operons, promoters, and transcriptional factors, along with their coordinated transcriptional regulatory mechanism and the genetic network of *E. coli*. This database also contains all known transport reactions, and metabolic and signaling pathways, as well as their associated enzymes.

Beyond the fundamental data content of *E. coli* K-12 MG1655, the database is also linked to other databases on different *E. coli* strains. Moreover, it can be used to annotate a newly sequenced bacterial genome via the BLAST search and its metabolic pathways via the PathoLogic Component of the Pathway Tools software.

KEGG: Kyoto Encyclopedia of Genes and Genomes [24]. KEGG is a collection of 4 large databases called PATHWAY, GENES, LIGAND, and BRITE, plus various software tools. The PATHWAY database is the most unique and extensive, and contains protein interaction networks, such as metabolic pathways, genetic information processing, signal transduction, and infectious diseases. The GENES database provides extensive cover of experimentally and/or computationally predicted genetic data originally derived from other public resources, as well as microarray gene expression profiles obtained from Japanese research groups. The LIGAND database is a source of detailed information on chemical compounds, enzymes, and glycans, plus their associated reactions. The BRITE database contains the KO (KEGG Orthology), which refers to the grouping of two genes as orthologs when they are mapped to the same KEGG pathway node, and is used to characterize unknown pathways. In addition, KEGG provides a number of computational services for sequence analysis, such as BLAST, FAST, MOTIF, and CLUSTALW.

MetaCyc [6]. MetaCyc is a general nonredundant metabolic database that contains comprehensive primary and secondary metabolic pathways and the corresponding compounds, proteins, enzymes, and genes. The pathways stored in this database have been retrieved from more than 240 different organisms, mostly microorganisms and plants. Because of the extensive information on metabolic pathways, this database is particularly useful for scientists and engineers who work on computational biology, systems biology, and metabolic engineering. MetaCyc does not deal with genetic sequence data, distinguishing it from EcoCyc. Although originally focused on the qualitative aspects of metabolic pathways, MetaCyc recently started to cover quantitative data as well, such as enzyme kinetics data.

PANTHER Pathways [44]. *Protein ANalysis THrough* Evolutionary Relationships (PANTHER) is a database that classifies proteins according to their families, subfamilies, molecular functions, and biological processes and pathways. This review only highlights the PANTHER Pathways database that contains more than 60 signaling pathways, into which protein subfamilies and sequences are mapped. Besides browsing a list of pathways and searching via keywords, researchers can also perform a gene expression analysis, which is particularly useful for the analysis of microarray data. The gene expression results are represented using a pathway diagram that eventually reveals the relationship between the gene and the proteins in the pathway.

PathArt. PathArt, a bioinformatic and chemoinformatic product from the company called Jubilant Biosys (http://

jubilantbiosys.com/index.htm), is a comprehensive database of metabolic and signaling pathways, and a very useful tool for various systems-biological analyses. For instance, this database enables the identification and comparison of pathways across physiologies, diseases, and organisms, while also viewing and building information on cross-talking pathways. An accompanying tool to PathArt is the Dynamic Pathway Articulator Component, which is a visualization tool that dynamically builds pathways and analyzes the database according to the user's input. Another important component is the Microarray Data Analysis Component. PathArt is also integrated with Sportfire, which allows the mapping of gene expression information onto pathways.

RegulonDB [49]. RegulonDB is a database of the transcriptional regulations, operon organization, and expression conditions for *E. coli* K-12. It shows the location of the regulatory components of the affected genes via a Java Interface Tool, making it possible to graphically view the location of genes of interest in a chromosome, or a set of affected genes according to functional classes and growth conditions. It also provides Gene Expression Analysis Tools (GETools) and Regulatory Sequence Analysis Tools (RSAT), where the former allows the analysis of the transcriptional regulation of global gene expression data generated from a microarray in comparison with a RegulonDB database, and the latter detects regulatory signals in noncoding gene sequences. Finally, RegulonDB contains several data sets, such as regulatory network interactions, promoters, transcription factors, and operons, which also enhance studies on the regulatory mechanisms of *E. coli*.

Software for Computational Systems Biology, Static/Dynamic Simulation

BIOCHAM. The Biochemical Abstract Machine (BIOCHAM) is a programming environment that supports the modeling, simulation, and querying of biochemical systems. The Contraints project focused on developing constraint programming languages for combinatorial optimization problems and bioinformatics. As such, BIOCHAM provides a rule-based language for modeling biochemical systems with concurrent transition systems and a query language based on the temporal Computation Tree Logic (CTL) for querying and the automatic validation of both quantitative and qualitative models in systems biology. BIOCHAM uses the symbolic model checker NuSMV and constraint-based model checker DMC for CTL queries. Once the model is formalized, a Machine Learning technique is applied to check the adequacy of a given set of CTL properties in the model. However, in spite of these features, BIOCHAM requires all the system parameters to be encoded in a finite vector of the data and control variables, making it difficult to use. Thus, BIOCHAM is essentially a model checking tool that is useful for querying and validating complex models in systems biology.

BioCharon. BioCharon is an integrated environment for modeling and simulating biomolecular networks. The main focus is to apply hybrid modeling and simulation techniques to biological systems under parametric and functional uncertainty. BioCharon consists of two programs, BioSketchPad and Charon. BioSketchPad uses graphic tools to design a biological model under the hybrid systems theory. Once the model is created, it is converted into a Charon model to facilitate further analysis and simulation. Charon employs state-space exploration techniques to analyze the model characteristics and allows users to determine the interrelationship between local biological variations and global behavior.

Bioglyphics & BALSA. The Biological Abstraction Layer for Simulation Analysis (BALSA) is a software environment for the rapid prototyping and testing of cell system hypotheses. The goal of Cell Systems Initiative (CSI) is to better understand the dynamic information mechanisms in all living cells, thus it has developed a modeling and simulation framework that consists of two main software systems, BALSA for modeling and SigTran for simulation. BALSA provides a symbol-based modeling language named BioGlyphics. The symbols used in BioGlyphics can be categorized into four components: entity, event, context, and punctuation. The entity component represents small molecules, macromolecules, and complexes in various states, whereas the event component describes the physical and chemical reactions between entities, including interactions, transformations, translocations, and abstract processes. The context component symbolizes the spatial locations and boundaries of a model, and the punctuation component groups statements into blocks, such as a module. BioGlyphics then helps BALSA build complex, multistate, and multicontextual molecular systems. Moreover, BALSA can change Bioglyphics models into a graphical network or the Systems Biology Markup Language (SMBL) [20].

BioModels.net. BioModels.net collects computational models and adopts an SBML format for the common representation and exchange of data. Each model is carefully cataloged and linked to relevant data resources, such as publications, compounds, or pathways. BioModels.net also supports browsing, searching, cross-referencing, and visualization of the models. In addition to collecting models, BioModels.net is also involved with two other projects, Minimal Information Requested in the Annotation of Biochemical Models (MIRIAM) for the standardization of models, and Systems Biology Ontologies (SBO) for the development of controlled vocabularies and ontologies.

BioNetGen [5]. BioNetGen is software for rule-based deterministic and/or stochastic modeling of cellular signaling networks. BioNetGen focuses on developing mathematical

models, related software, and database resources for complex biological systems, particularly for signal transduction systems. In support of these efforts, BioNetGen designs the models under the domain and rule-based approach, which facilitates the automatic generation of chemical species according to specified components, as well as the generation of reactions among these species according to specified rules in the models. As a result, BioNetGen can be used to observe interactions among components to investigate particular characteristics of the system. Basically, BioNetGen is text-based and command line software. Once the model is complete, a biochemical network is generated according to the rules of the model specifications and can be converted to SBML for further analyses.

BioSens. BioSens is a simulation and sensitivity analysis toolkit. For proper running and simulations, BioSens requires a Bio-SPICE dashboard [13] with a BioMat Bridge, Matlab, and one of the two ordinary differential equation (ODE)-solvers XXP or DASPK. BioSens provides a better understanding of how parameter perturbation can affect a system output by calculating the system sensitivity with respect to model parameters and initial conditions. In addition, Fisher Information Matrix (FIM)-based sensitivity measurements provide information on the robustness/fragility trade-off and optimal measurement set for maximum parameter identification and accuracy. For user convenience, BioSens supports a Matlab graphical user interface.

Bio-SPICE [13]. Bio-SPICE is an open source community that is attempting to develop a framework for systems biology. This framework includes modeling, simulation, visualization tools, and computational models, such as pathways, interaction networks, differential equations, and gene expressions. The Bio-SPICE framework has Dashboard as its core, since it functions as a basis for the integration of various tools. In other words, Bio-SPICE provides a very flexible architecture, within which users can integrate various tools, according to their purpose, by downloading them onto the Dashboard, thereby enabling various combinations of software environments for modeling, analysis, and simulation. Bio-SPICE currently consists of 18 groups and 41 tools: 14 model analysis tools, 10 data analysis tools, 3 databases, 9 model composition & visualization tools, and 5 simulators.

BioTapestry. BioTapestry is an interactive program for the design and simulation of genetic regulatory networks. The focus of Institute for Systems Biology (ISB) is to analyze the biological complexity and understand the functional aspects of biological systems. BioTapestry provides an interactive graphical network editor for the automatic layout of tabular interaction data, and highlighting multiple direct and indirect paths between any two genes in a network. Thus, large-scale models can be handled with consistency and clarity. Moreover, BioTapestry can represent systems that exhibit an increasing complexity over time through the use of temporal and spatial expressions and input data tables. As such, this feature allows BioTapestry to show the network evolution on an hourly basis. For increased flexibility, BioTapestry uses a Dizzy engine for simulation and SBML for model sharing.

BioUML. BioUML, an extensive software framework for systems biology written in Java, is designed to define biological systems using unified modeling language (UML) methodology. BioUML facilitates access to databases with experimental data, and the formalizing, visualization, and simulation of biological systems. Currently, BioUML has seven sections: Meta model, Viewer, Editor, Search engine, Modeler, Standard diagram and data types, and Database modules.

BSTLab. BSTLab is a software toolbox compatible with Matlab for biochemical system theory. The activities are focused on identifying theories and developing algorithms with a particular emphasis on systems biology. BSTLab has three major components: numeric analysis, symbolic analysis, and SBML conversion utility. When using the numeric and symbolic components, the system models are represented in the form of kinetic order matrices for the fluxes and vectors for the rate constants. BSTLab provides various functions to compute the steady state of the system, eigenvalues, sensitivities of dependent variables and fluxes with respect to rate constants and kinetic orders, and logarithmic gains of dependent variables and fluxes with respect to independent variables. It also provides automatic generation of ordinary differential equations for a given state that can be calculated via available Matlab solvers, thereby allowing controlled mathematical comparisons of alternative system structures.

CADLIVE [30]. CADLIVE is an interactive design tool for large-scale biological networks that uses a Graphic User Interface (GUI). Various compartments can be built, including the environment, cytoplasm, nucleoplasm, endoplasmic reticulum (ER), mitochondria, golgi, chloroplast, membrane, and others. Species, reaction, and gene symbols are also provided for drawing biological networks. When using these symbols, various states of transcription, biochemical reactions, and complex formations can be described between species, including DNA, RNA, proteins, metabolites, cofactors, ions, and others. In addition, the expressed model can be saved as regulator reaction equations in a database in a format compatible to a simulator.

CellDesigner. CellDesigner is a modeling software for biochemical networks that uses a GUI. CellDesinger supports the process diagrams, graphical notations, and symbols proposed by Kitano using a structured diagram editor. Each symbol can be connected according to its name or ID specified in its notes. The drawn models are stored in SBML format and can be linked to simulation and other analysis packages through Systems Biology Workbench (SBW) [50].

Cellerator [53]. Cellerator is a Mathematica package for the biological modeling of signal transduction networks, and supports the automatic generation of differential equations. Cellerator represents biochemical and transcriptional interactions with arrow-based notations, and automatically converts them into differential equations that can be solved numerically using Mathematica or other programs. As a Mathemtica package, Cellerator can easily extend to larger data structures.

CellML [37]. CellML is an open modeling language based on the eXtensible Markup Language (XML). CellML is designed as a standard method for representing and exchanging biological models. It consists of 3 main components: model structure for the relationships among different parts of the model, mathematics for the equations describing the biological processes, and metadata for additional information. In addition, CellML offers a model repository and software. Currently, 230 models are listed under 10 categories: 43 Signal Transduction Pathway Models, 33 Metabolic Pathway Models, 55 Cardiac Electrophysiological Models, 40 Calcium Dynamics Models, 25 Immunology Models, 12 Cell Cycle Models, 3 Simplified Electrophysiological Models, 12 Other Cell-Type Electrophysiological Models, and 7 Mechanical Models and Constitutive Laws. Besides the model repository, various other models are also available in the Tool Test Repository. For more convenient usage of CellML, several tools are provided, such as CellML API, the CellML editor tool suite, and Validator

Cellware [9]. Cellware is a modeling and simulation environment for cellular transactions, and consists of a modeling component, application component, and simulation component. When using these components, Cellware can design gene regulation and metabolic networks, while also supporting various analyses, including network analysis, deterministic and stochastic simulations, and parameter estimation. In addition, Cellware provides a grid environment for dealing with large models.

COPASI. The Complex Pathway Simulator (COPASI) is a simulation and analysis tool for biochemical networks. COPASI was developed in response to GEPASI's false calculation of Jacobians and is still in its test build phase. COPASI provides various analysis tools together with a model generator; thus, in addition to stochastic and deterministic time course simulations, it also offers a steady-state analysis, metabolic control analysis, elementary mode analysis, and mass conservation analysis. Moreover, a parameter scanning functionality was incorporated in the latest version.

CSB.DB [55]. The Comprehensive Systems Biology Database (CSB.DB) contains the results of biostatistical analyses of data from transcript and metabolite profiling experiments. CSB.DB also provides information about functionally interacting genes by searching for the best gene-to-gene correlations among the models in the database. It also helps identify pairs and larger units of genes. CSB.DB currently consists of three major parts: co-response database, metabolic and analytical database, and CSB.DB-associated Web pages. The co-response database maintains transcriptional information on the gene-to-gene correlations of *E. coli* (EcoCor@CSB.DB), *S. cerevisiae* (SceCoR@CSB.DB), and *A. thaliana* (AthCoR@CSB.DB). The metabolic and analytical database focuses on analytical techniques and contains comprehensive information on metabolomics (GMD@CSB.DB). The associated Web pages contain a manual collection of gene functions and experimentally analyzed results, such as the Plant Subtilase Database (PSDB@CSB.DB).

Cytoscape [52]. Cytoscape is software that visualizes molecular interaction networks and integrates these interactions with various data, such as gene expression profiles. Some of the important features of Cytoscape include building molecular interaction networks from raw interaction files (SIF format) containing lists of protein-protein and/or protein-DNA interaction pairs, as well as supporting user-defined interaction types, supporting Graph Markup Language, inputting mRNA expression profiles from tab- or space-delimited text files, importing gene functional annotations from the Gene Ontology (GO) [15] and KEGG databases [25], supporting plugins for network and molecular profile analysis, and filtering a network to select subsets of nodes and/or interactions based on current data (e.g., users can select nodes involved in a threshold number of interactions, nodes that share a particular GO annotation, or nodes whose gene expression levels change significantly under one or more conditions according to p-values loaded with gene expression data.). Furthermore, active subnetworks/pathway modules can be identified, where a network is screened against gene expression data to identify connected sets of interactions, i.e., *interaction subnetworks*, whose genes show particularly high levels of differential expression. The interactions contained in each subnetwork can then provide hypotheses for the regulatory and signaling interactions in control of the observed expression changes. Finally, clusters (highly interconnected regions) can be identified in any network loaded into Cytoscape. Depending on the type of network, clusters can mean different things. For instance, clusters in a protein-protein interaction network have been shown to be protein complexes and parts of pathways, whereas clusters in a protein similarity network represent protein families.

Dizzy [47]. Dizzy is a software package for the stochastic simulation of chemical kinetics and allows a user to define the model and implement Gillespie, Gibson-Bruck, and Tau-Leap stochastic algorithms. Dizzy can also import and export SBML model definition language, and display models graphically using the Cytoscape software system.

E-Cell [56]. E-Cell is a virtual cell system for computers that enables cell-level simulation. E-Cell is object-oriented software designed for the modeling, simulation, and analysis of large-scale complex systems, such as a microorganism cell. E-Cell Simulation Environment version 3, the core of E-Cell, integrates many differentially existing components. The E-Cell System basically consists of the E-Cell Simulation Environment, E-Cell Modeling Environment, and E-Cell Analysis Toolkit.

FluxAnalyzer [28]. FluxAnalyzer is a MATLAB-based software package that can analyze a stoichiometrically built metabolic network. Interactive flux maps play a crucial role for visualization and user interaction, where an abstract network model is linked with network graphics. Thereafter, several mathematical analyses can be performed, including a metabolic flux analysis, flux optimization, topology characterization, pathway analysis, and more. FluxAnalyzer is thus useful for metabolic engineering and systems biology.

GEPASI [43]. GEPASI software allows dynamic modeling of biochemical systems, covering the construction of biochemical models using kinetic parameters, optimizing the models, and simulating metabolic control analyses and linear stability analyses. Some of the specific features of GEPASI include the characterization of steady states using a metabolic control analysis and linear stability analysis, a scan utility for the advanced exploration of a model's behavior in multidimensional parameter space, data fitting or parameter estimation with experimental data, visualization of simulation results in 2D or 3D directly from the program, and the support of SBML level 1 for data interchange with other systems biology software.

INSILICO Discovery by INSILICO Biotechnology. INSILICO Discovery is a computational tool for the design and analysis of biochemical networks, and is a very powerful technology for integrating a wide range of systems biological data (genome, transcriptome, proteome, or physiological data) into *in silico* models. INSILICO Discovery also includes newly developed and efficient algorithms for optimizing metabolic networks. In addition, it facilitates the visualization of large-scale networks, loading of comprehensive pathways and reactions, mathematical analyses, network consistency checks, automated transfer of graphical representations into mathematical models, and algorithms for the incorporation of functional genomics for pathway analysis.

Jarnac: Systems Biology Software. Jarnac is a language used to describe and perturb biochemical systems, such as gene, metabolic, and signal transduction networks. Its main purpose is to construct and simulate metabolic networks. The features of Jarnac include metabolic control analysis, dynamic simulation (using an LSODA integrator), matrix arithmetic, and steady-state analysis (using an NLEQ solver).

JWS Online [45]. JWS Online is a database of kinetic models developed for the analysis of biochemical systems. The mathematical analyses are conducted online. The JWS Online Web site contains currently published models, as well as their SBML versions for downloading. Requests from the database currently lead to metabolic reaction schemes along with their kinetic equations. Eventually, time simulations, steady-state analysis, and metabolic control analysis will also be available.

KINSOLVER. KINSOLVER is a simulator designed to integrate chemical reaction networks based on solving multiplicative equations for the balance of mass using the parameters, rate constants, and initial conditions. Five standard solvers, Euler, Modified Euler, Runge Kutta (RK), Adaptive RK-Fehlberg, and LSODES, are used for the numerical calculations. Among the 5 standard solvers, the LSODES method consistently shows a relatively large error tolerance of 0.01 and with a small error tolerance of 10^{-6}.

MesoRD [17]. The Mesoscopic Reaction Diffusion Simulator (MesoRD) is a tool designed for the stochastic simulation of reactions and diffusions in three-dimensional space. As such, MesoRD efficiently implements the Next Subvolume Method, which makes use of the structure of a reaction-diffusion master equation.

MetaboLogica. MetaboLogica is systems biology software that can be used to visualize metabolic data, perform metabolic flux analyses (MFA), and design NMR/GC-MS experiments. It can also be used to construct kinetic models and simulate time profiles. The main features of MetaboLogica are a friendly GUI for graphically building metabolic networks, MFA toolbox (for calculating the steady-state flux distribution of a network and finding all the optimal flux distributions simultaneously), NMR/GC-MS Simulation toolbox, and Dynamic Analysis toolbox.

MetaFluxNet [33]. MetaFluxNet is a program package for reconstructing metabolic networks and performing various steady-static simulations. This tool enables users to observe the metabolic behavior of the *in silico* organism of interest in response to environmental and/or genetic perturbations. Consequently, the program can help reveal the metabolism of an organism and promote the systematic design of metabolic engineering experiments.

MMT2. MMT2 is stand-alone software specifically designed for metabolic modeling, and is very similar to Gepasi and Jarnac. In particular, the main features of MMT2 that are not characterized by other tools include its ability to cover both stationary and nonstationary experiments, various algorithms for model choice and falsification, and the visualization of dynamic networks and sensitivity data. In addition, the program supports SBML for flexible data exchange.

Monod. Monod, "the Modeler's notebook and data store," is a Web-based program that manipulates a reaction

network and thereby constructs models for the study of complex biochemical systems. It provides detailed information on molecular species, biological processes, and/or reactions based on corresponding literature linked to PubMed. In addition, this tool contains robust systems for maintaining previous versions of data, as well as permission control. The resulting data can also be exported into other simulation and visualization tools.

Moleculizer [38]. Moleculizer is based on stochastic calculation for the simulation of intracellular biochemical systems and is run along with Monod. This tool is particularly useful for the study of protein complexes. SBML and XML are also supported.

NetBuilder. NetBuilder is a stand-alone program that designs and simulates gene regulatory circuits, such as transcription and translation processes. It aims to quantitatively analyze gene regulatory networks so as to test experimental hypotheses. This is particularly useful for modeling and simulating the interactions taking place between transcription factors and genes/other transcription apparatuses.

Pathway Simulation Tool. The Pathway Simulation Tool was developed based on the need for modeling approaches that simplify the system to be studied in order to reduce the calculation burden (number of parameters), while maintaining the core characteristics. Kinetic models are read in an SBML format and reformatted in an internal data format.

PaVESy [39]. The major component of the Pathway Visualization and Editing System (PaVESy) is a relational database system. The integration of biological data, such as genes, proteins, and metabolites, can be performed during the database design. The construction of the relations among the components requires the assembly of metabolic and regulatory networks. The system supports a java-based class library for the database programming interface, allowing entries, annotations, interactions, subsets, and folders for the implementation of the concepts of object persistence. The data can be exported to various programs that support an SBML or Geographic Markup Language (GML) format.

PROTON. PROTON is an Integrative Modeling System (IMS) that utilizes integrative methods for modeling large-scale systems. PROTON makes use of molecular databases for the design of biochemical networks and can incorporate several relational databases, such as Oracle or Mysql, for the extraction of related information. Using information extracted from the database, biochemical processes and large-scale charts can be automatically constructed. Eventually, biochemical networks can be found from the biochemical processes and charts for the construction of dynamic systems.

Reactome [22]. Reactome is a knowledge base of biological processes in humans. The database deals with the biological processes of metabolic networks to high-level processes. Although Reactome was developed for the analysis of biological pathways in humans, it also includes various biological reactions obtained from organisms other than humans. Reactome provides extended search tools for querying various database objects, such as domains, physical entities, and pathways. Users can select options to restrict the field of query. Moreover, the PathFinder modules can generate pathways from given compounds that represent the starting and ending positions of the network. The entire contents, including the databases, can be obtained from the Web site.

SigPath by Institute for Computational Biomedicine (ICB). SigPath is an information management system (IMS) dedicated to the quantitative study of signaling pathways and networks. This IMS stores biochemical information, such as cellular components and their interactions, as well as basic reactions governing those interactions, thereby facilitating the construction of pathways and large-scale networks. Moreover, SigPath employs user-friendly ways of efficient data handling, such as the use of background information (e.g., names of molecules, aliases, and accession codes) for easy data submission, permitting multiple users to refine information over time, and development of data representation that stores both qualitative and quantitative information. Lastly, models constructed in SigPath can be exported to several programs, such as Kinetikit/Genesis, Virtual Cell, Jarnac/JDesigner, and JSim.

SigTran by Cell Systems Initiative. SigTran is cell signaling simulation software, specifically designed for large-scale simulations and the analysis of signal transduction networks. It is currently being employed by the UW/CSI modeling team to simulate T-cell signal transduction events. SigTran also supports the simulation of various kinetics, model debugging, statistics, and graphical displays.

SmartCell. SmartCell was developed for the modeling and simulation of diffusion reaction networks in a cell. It basically uses a stochastic reaction model and considers cell geometry, cell compartmentalization, localization of species, membrane diffusion and multistep reactions, transcription and translation processes, and cell growth. An improved graphical user interface for model development is also available in the package.

StochSim [31]. StochSim focuses on the study of bacterial chemotaxis by considering the stochastic characteristics of the signaling pathway and its constituting individual reactions. StochSim provides a biochemical simulator for the expression of molecules or molecular complexes, and it consists of a platform-independent core engine and graphical interface. The algorithm implemented in StochSim is known to be slower than the Gillespie algorithm for the calculation of a small set of biochemical reactions. However, the performance of the simulator is better, as the molecules can be present in multiple states.

STOCKS 2.0: Stochastic Simulations of Biochemical Reaction Networks [27]. STOCKS 2.0 is a stochastic

simulator that implements the maximal time step method. This method enables a stochastic simulation with a given set of metabolic and regulatory networks involving small numbers of molecules. STOCKS 2.0 supports SBML with program-specific parameters as its input format. As an application of the simulator, the random variation of gene expression on metabolic networks and its propagation to the level of metabolic processes can be shown. Furthermore, the simultaneous modeling of gene expressions, transports, enzyme activities, and signaling can also be achieved using this tool.

PNK 2e. The Petri Net Kernel (PNK) 2e uses Stochastic Petri Nets (SPNs) for the modeling and simulation of biological systems. It is based on the Petri Net Kernel and involves several advanced features, such as extended stochastic Petri Nets and the exchange of models in an SBML format. In addition, it can perform both deterministic and stochastic simulations via a Systems Biology Workbench, as well as a structural analysis of the Petri Net.

Virtual Cell [54]. The virtual cell was developed to facilitate modeling and simulation for cell biologists. It is largely based on both a biological and mathematical framework. Therefore, users can construct biological models from the mathematical code generated by the software for simulation. Furtheremore, mathematicians can input their own mathematical description using the math framework. In addition, BioModel, Geometry, and Math Model documents constitute the core components of the software.

WebCell [32]. WebCell is a Web-based simulation environment for the analysis of cellular behavior. The environment is convenient for the construction, analysis, and visualization of network models. WebCell is developed based on the Java Server Page (JSP) and Java applet-sevlet technologies. WebCell supports the widely used Systems Biology Markup Language (SBML) for the efficient exchange of model information. The dynamic model behavior can be simulated by utilizing ordinary differential equations (ODEs) or differential algebraic equations (DAEs). WebCell also supports metabolic control analysis (MCA) for a sensitivity analysis of a model system.

Worldwide PDB (Protein Data Bank). The Worldwide Protein Data Bank (wwPDB) consists of three member organizations. The founding members include RCSB PDB (U.S.A.), MSD-EBI (Europe), and PDBj (Japan). The wwPDB was initiated to maintain a Protein Data Bank Archive for macromolecular structural data, which can be accessible from all around the world. This site provides resources on the services and projects of the ww PDB.

CONCLUSION AND FUTURE PERSPECTIVES

Systems biology research has been receiving much attention, as it facilitates the examination of a systems-level holistic view of cells and organisms. As a result, many research groups and institutes dedicated to systems biology research have recently been founded, and become involved in developing various methodologies and tools for the quantitative and qualitative analysis of biological systems at the systems level. The research areas covered by systems biology are also rapidly expanding beyond original expectations in terms of size, complexity, and scope. Thus, instead of starting from scratch, it is essential to follow-up on the developments already made by other researchers. Therefore, it is hoped this review will serve such a function. Reflecting the complexity and great need for the interdisciplinary aspects of systems biology research, the foundation of large-scale organizations, such as BIO-SPICE, BioModel.net, and the SBML project, has become a new trend. Standardization of the representation and exchange of information extracted from large-scale biological systems is also becoming more and more important. Thus, SBML is now widely accepted and supported by more than 80 software tools worldwide.

Furthermore, high-quality, accurate, and extensive data sets are required to fuel the progress of systems biology research. Yet, one current limitation is the difficulty of acquiring quantitative data, such as kinetic parameters, for dynamic simulation. Thus, fundamental research based on system-level approaches should be performed to supplement the presently insufficient data. In addition, advances in technology for generating various omics data and new computational methods should also help to overcome the existing limitations, and to make it possible to accelerate the development of systems biology.

Acknowledgments

This work was supported by the Korean Systems Biology Research Project (M10309020000-03B5002-00000) of the Korean Ministry of Science and Technology, and by the Brain Korea 21 Project. Further support from the LG Chem Chair Professorship, Microsoft, and IBM SUR program is also appreciated.

REFERENCES

1. Bairoch, A. 2000. The ENZYME database in 2000. *Nucleic Acids Res.* **28:** 304–305.
2. Ball, C. A., I. A. Awad, J. Demeter, J. Gollub, J. M. Hebert, T. Hernandez-Boussard, H. Jin, J. C. Matese, M. Nitzberg, F. Wymore, Z. K. Zachariah, P. O. Brown, and G. Sherlock. 2005. The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res.* **33:** D580–D582.
3. Birney, E., D. Andrews, M. Caccamo, Y. Chen, L. Clarke, G. Coates, T. Cox, F. Cunningham, V. Curwen, T. Cutts,

T. Down, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, S. Graf, M. Hammond, J. Herrero, K. Howe, V. Iyer, K. Jekosch, A. Kahari, A. Kasprzyk, D. Keefe, F. Kokocinski, E. Kulesha, D. London, I. Longden, C. Melsopp, P. Meidl, B. Overduin, A. Parker, G. Proctor, A. Prlic, M. Rae, D. Rios, S. Redmond, M. Schuster, I. Sealy, S. Searle, J. Severin, G. Slater, D. Smedley, J. Smith, A. Stabenau, J. Stalker, S. Trevanion, A. Ureta-Vidal, J. Vogel, S. White, C. Woodwark, and T. J. Hubbard. 2006. Ensembl 2006. *Nucleic Acids Res.* **34:** D556–D561.

4. Blake, J. A., J. T. Eppig, C. J. Bult, J. A. Kadin, J. E. Richardson, and Mouse Genome Database Group. 2006. The Mouse Genome Database (MGD): Updates and enhancements. *Nucleic Acids Res.* **34:** D562–D567.
5. Blinov, M. L., J. R. Faeder, B. Goldstein, and W. S. Hlavacek. 2004. BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20:** 3289–3291.
6. Caspi, R., H. Foerster, C. A. Fulcher, R. Hopkinson, J. Ingraham, P. Kaipa, M. Krummenacker, S. Paley, J. Pick, S. Y. Rhee, C. Tissier, P. Zhang, and P. D. Karp. 2006. MetaCyc: A multiorganism database of metabolic pathways and enzymes. *Nucleic Acids Res.* **34:** D511–D516.
7. Chalkley, R. J., P. R. Baker, L. Huang, K. C. Hansen, N. P. Allen, M. Rexach, and A. L. Burlingame. 2005. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell Proteomics* **4:** 1194–1204.
8. Cheung, K. H., K. White, J. Hager, M. Gerstein, V. Reinke, K. Nelson, P. Masiar, R. Srivastava, Y. Li, J. Li, H. Zhao, J. Li, D. B. Allison, M. Snyder, P. Miller, and K. Williams. 2002. YMD: A microarray database for large-scale gene expression analysis. *Proc. AMIA Symp.* 140–144.
9. Dhar, P., T. C. Meng, S. Somani, L. Ye, A. Sairam, M. Chitre, Z. Hao, and K. Sakharkar. 2004. Cellware -- A multi-algorithmic software for computational systems biology. *Bioinformatics* **20:** 1319–1321.
10. Edgar, R., M. Domrachev, and A. E. Lash. 2002. Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30:** 207–210.
11. Finn, R. D., J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, E. L. Sonnhammer, and A. Bateman. 2006. Pfam: Clans, Web tools and services. *Nucleic Acids Res.* **34:** D247–D251.
12. Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269:** 496–512.
13. Garvey, T. D., P. Lincoln, C. J. Pedersen, D. Martin, and M. Johnson. 2003. BioSPICE: Access to the most current computational tools for biologists. *OMICS* **7:** 411–420.
14. Gasteiger, E., A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel, and A. Bairoch. 2003. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31:** 3784–3788.
15. Gene Ontology Consortium. 2006. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* **34:** D322–D326.
16. Goto, S., Y. Okuno, M. Hattori, T. Nishioka, and M. Kanehisa. 2002. LIGAND: Database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* **30:** 402–404.
17. Hattne, J., D. Fange, and J. Elf. 2005. Stochastic reaction-diffusion simulation with MesoRD. *Bioinformatics* **21:** 2923–2924.
18. Hirschman, J. E., R. Balakrishnan, K. R. Christie, M. C. Costanzo, S. S. Dwight, S. R. Engel, D. G. Fisk, E. L. Hong, M. S. Livstone, R. Nash, J. Park, R. Oughtred, M. Skrzypek, B. Starr, C. L. Theesfeld, J. Williams, R. Andrada, G. Binkley, Q. Dong, C. Lane, S. Miyasato, A. Sethuraman, M. Schroeder, M. K. Thanawala, S. Weng, K. Dolinski, D. Botstein, and J. M. Cherry. 2006. Genome Snapshot: A new resource at the *Saccharomyces* Genome Database (SGD) presenting an overview of the *Saccharomyces cerevisiae* genome. *Nucleic Acids Res.* **34:** D442–D445.
19. Hou, B. K., J. S. Kim, J. H. Jun, D. Y. Lee, Y. W. Kim, S. Chae, M. Roh, Y. H. In, and S. Y. Lee. 2004. BioSilico: An integrated metabolic database system. *Bioinformatics* **20:** 3270–3272.
20. Hucka, M., A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, J. Wang, and SBML Forum. 2003. The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* **19:** 524–531.
21. Jaiswal, P., J. Ni, I. Yap, D. Ware, W. Spooner, K. Youens-Clark, L. Ren, C. Liang, W. Zhao, K. Ratnapu, B. Faga, P. Canaran, M. Fogleman, C. Hebbard, S. Avraham, S. Schmidt, T. M. Casstevens, E. S. Buckler, L. Stein, and S. McCouch. 2006. Gramene: A bird's eye view of cereal genomes. *Nucleic Acids Res.* **34:** D717–D723.
22. Joshi-Tope, G., M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, G. R. Gopinath, G. R. Wu, L. Matthews, S. Lewis, E. Birney, and L. Stein. 2005. Reactome: A knowledge base of biological pathways. *Nucleic Acids Res.* **33:** D428–D432.
23. Kahraman, A., A. Avramov, L. G. Nashev, D. Popov, R. Ternes, H. D. Pohlenz, and B. Weiss. PhenomicDB: A multi-species genotype/phenotype database for comparative phenomics. *Bioinformatics* **21:** 418–420.
24. Kanehisa, M., S. Goto, M. Hattori, K. F. Aoki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, and M. Hirakawa. 2006. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* **34:** D354–D357.

25. Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. 2004. The KEGG resource for deciphering the genome. *Nucleic Acids Res.* **32:** D277–D280.
26. Keseler, I. M., J. Collado-Vides, S. Gama-Castro, J. Ingraham, S. Paley, I. T. Paulsen, M. Peralta-Gil, and P. D. Karp. 2005. EcoCyc: A comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res.* **33:** D334–D337.
27. Kierzek, A. M. 2002. STOCKS: STOChastic kinetic simulations of biochemical systems with *Gillespie algorithm*. *Bioinformatics* **18:** 470–481.
28. Klamt, S., J. Stelling, M. Ginkel, and E. D. Gilles. 2003. FluxAnalyzer: Exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics* **19:** 261–269.
29. Krummenacker, M., S. Paley, L. Mueller, T. Yan, and P. D. Karp. 2005. Querying and computing with BioCyc databases. *Bioinformatics* **21:** 3454–3455.
30. Kurata, H., K. Masaki, Y. Sumida, and R. Iwasaki. 2005. CADLIVE dynamic simulator: Direct link of biochemical networks to dynamic models. *Genome Res.* **15:** 590–600.
31. Le Novere, N. and T. S. Shimizu. 2001. STOCHSIM: Modelling of stochastic biomolecular processes. *Bioinformatics* **17:** 575–576.
32. Lee, D. Y., C. Yun, A. Cho, B. K. Hou, S. Park, and S. Y. Lee. 2006. WebCell: A Web-based environment for kinetic modeling and dynamic simulation of cellular networks. *Bioinformatics* [in press].
33. Lee, D. Y., H. Yun, S. Park, and S. Y. Lee. 2003. MetaFluxNet: The management of metabolic reaction information and quantitative metabolic flux analysis. *Bioinformatics* **19:** 2144–2146.
34. Lee, S. Y., D. Y. Lee, and T. Y. Kim. 2005. Systems biotechnology for strain improvement. *Trends Biotechnol.* **23:** 349–358.
35. Lee, Y., J. Tsai, S. Sunkara, S. Karamycheva, G. Pertea, R. Sultana, V. Antonescu, A. Chan, F. Cheung, and J. Quackenbush. 2005. The TIGR gene indices: Clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.* **33:** D71–D74.
36. Lemer, C., E. Antezana, F. Couche, F. Fays, X. Santolaria, R. Janky, Y. Deville, J. Richelle, and S. J. Wodak. 2004. The aMAZE LightBench: A Web interface to a relational database of cellular processes. *Nucleic Acids Res.* **32:** D443–D448.
37. Lloyd, C. M., M. D. Halstead, and P. F. Nielsen. 2004. CellML: Its future, present and past. *Prog. Biophys. Mol. Biol.* **85:** 433–450.
38. Lok, L. and R. Brent. 2005. Automatic generation of cellular reaction networks with Moleculizer 1.0. *Nat. Biotechnol.* **23:** 131–136.
39. Ludemann, A., D. Weicht, J. Selbig, and J. Kopka. 2004. PaVESy: Pathway visualization and editing system. *Bioinformatics* **20:** 2841–2844.
40. Maglott, D., J. Ostell, K. D. Pruitt, and T. Tatusova. 2005. Entrez gene: Gene-centered information at NCBI. *Nucleic Acids Res.* **33:** D54–D58.
41. McCarthy, A. A. 2005. Broad Institute: Bringing genomics to real-world medicine. *Chem. Biol.* **12:** 717–718.
42. McGinnis, S. and T. L. Madden. 2004. BLAST: At the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32:** W20–W25.
43. Mendes, P. 1997. Biochemistry by numbers: Simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.* **22:** 361–363.
44. Mi, H., B. Lazareva-Ulitsky, R. Loo, A. Kejariwal, J. Vandergriff, S. Rabkin, N. Guo, A. Muruganujan, O. Doremieux, M. J. Campbell, H. Kitano, and P. D. Thomas. 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33:** D284–D288.
45. Olivier, B. G. and J. L. Snoep. 2004. Web-based kinetic modelling using JWS online. *Bioinformatics* **20:** 2143–2144.
46. Parkinson, H., U. Sarkans, M. Shojatalab, N. Abeygunawardena, S. Contrino, R. Coulson, A. Farne, G. G. Lara, E. Holloway, M. Kapushesky, P. Lilja, G. Mukherjee, A. Oezcimen, T. Rayner, P. Rocca-Serra, A. Sharma, S. Sansone, and A. Brazma. 2005. ArrayExpress -- A public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **33:** D553–D555.
47. Ramsey, S., D. Orrell, and H. Bolouri. 2005. Dizzy: Stochastic simulation of large-scale genetic regulatory networks. *J. Bioinform. Comput. Biol.* **3:** 415–436.
48. Rhee, S. Y., W. Beavis, T. Z. Berardini, G. Chen, D. Dixon, A. Doyle, M. Garcia-Hernandez, E. Huala, G. Lander, M. Montoya, N. Miller, L. A. Mueller, S. Mundodi, L. Reiser, J. Tacklind, D. C. Weems, Y. Wu, I. Xu, D. Yoo, J. Yoon, and P. Zhang. 2003. The *Arabidopsis* Information Resource (TAIR): A model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. *Nucleic Acids Res.* **31:** 224–228.
49. Salgado, H., S. Gama-Castro, M. Peralta-Gil, E. Diaz-Peredo, F. Sanchez-Solano, A. Santos-Zavaleta, I. Martinez-Flores, V. Jimenez-Jacinto, C. Bonavides-Martinez, J. Segura-Salazar, A. Martinez-Antonio, and J. Collado-Vides. 2006. RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res.* **34:** D394–D397.
50. Sauro, H. M., M. Hucka, A. Finney, C. Wellock, H. Bolouri, J. Doyle, and H. Kitano. 2003. Next generation simulation tools: The systems biology workbench and BioSPICE integration. *OMICS* **7:** 355–372.
51. Schomburg, I., A. Chang, C. Ebeling, M. Gremse, C. Heldt, G. Huhn, and D. Schomburg. 2004. BRENDA, the enzyme database: Updates and major new developments. *Nucleic Acids Res.* **32:** D431–D433.
52. Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13:** 2498–2504.
53. Shapiro, B. E., A. Levchenko, E. M. Meyerowitz, B. J. Wold, and E. D. Mjolsness. 2003. Cellerator: Extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19:** 677–678.

54. Slepchenko, B. M., J. C. Schaff, I. Macara, and L. M. Loew. 2003. Quantitative cell biology with the virtual cell. *Trends Cell Biol.* **13:** 570–576.
55. Steinhauser, D., B. Usadel, A. Luedemann, O. Thimm, and J. Kopka. 2004. CSB.DB: A comprehensive systems-biology database. *Bioinformatics* **20:** 3647–3651.
56. Takahashi, K., N. Ishikawa, Y. Sadamoto, H. Sasamoto, S. Ohta, A. Shiozawa, F. Miyoshi, Y. Naito, Y. Nakayama, and M. Tomita. 2003. E-Cell 2: Multi-platform E-Cell simulation system. *Bioinformatics* **19:** 1727–1729.
57. Tatusov, R. L., D. A. Natale, I. V. Garkavtsev, T. A. Tatusova, U. T. Shankavaram, B. S. Rao, B. Kiryutin, M. Y. Galperin, N. D. Fedorova, and E. V. Koonin. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22–28.
58. Whetzel, P. L., H. Parkinson, H. C. Causton, L. Fan, J. Fostel, G. Fragoso, L. Game, M. Heiskanen, N. Morrison, P. Rocca-Serra, S. A. Sansone, C. Taylor, J. White, and C. J. Stoeckert Jr. 2006. The MGED Ontology; A resource for semantics-based description of microarray experiments. *Bioinformatics* [in press].
59. Wu, C. H., L. S. Yeh, H. Huang, L. Arminski, J. Castro-Alvear, Y. Chen, Z. Hu, P. Kourtesis, R. S. Ledley, B. E. Suzek, C. R. Vinayaka, J. Zhang, and W. C. Barker. 2003. The Protein Information Resource. *Nucleic Acids Res.* **31:** 345–347.