Chapter 2

In Silico Systems Biology Approaches for the Identification of Antimicrobial Targets

Malabika Sarker, Carolyn Talcott, and Amit K. Galande

Abstract

Classical antibiotic discovery efforts have relied mainly on molecular library screening coupled with target-based lead optimization. The conventional approaches are unable to tackle the emergence of antibiotic resistance and are failing to provide understanding of multiple mechanisms behind drug actions and the off-target effects. These insufficiencies have prompted researchers to focus on a multidisciplinary approach of systems biology-based antibiotic discovery. Systems biology is capable of providing a big-picture view for therapeutic targets through interconnected networks of biochemical reactions derived from both experimental and computational techniques. In this chapter, we have compiled software tools and databases that are typically used for target identification through in silico analyses. We have also identified enzyme- and broad-spectrum metabolite-based drug targets that have emerged through in silico systems microbiology.

Key words Antimicrobials, Database, Metabolites, Omics, Software tools, Systems biology, Targets

1 Introduction

Antimicrobial research conducted in the second half of the twentieth century largely focused on two mainstream approaches. The first approach involved identifying leads for the next generation of antibiotics through screening diverse sets of molecular libraries (1), and the second approach consisted of identifying novel antimicrobial targets through reductionism. The reductionist approach, fueled by molecular biology methods, focuses on studying specific functions of individual genes, proteins, and cells separately to identify valid molecular targets for therapeutic intervention (2). While these two themes of "library screening" and "target-based discovery" are central to any drug research program, investigators in antimicrobial discovery were among the first to recognize the major deficit in these conventional approaches. Screening and target-based methods typically lack a "big picture" that shows molecular connectivity and provides global understanding of cellular physiology. This realization was mostly prompted by the rapid emergence of drug resistance along with the failure of lead candidates in the

Sandhya Kortagere (ed.), *In Silico Models for Drug Discovery*, Methods in Molecular Biology, vol. 993,
DOI 10.1007/978-1-62703-342-8_2, © Springer Science+Business Media, LLC 2013

late stages of drug development, owing especially to the off-target effects (3). One way to address the issues around these "isolated" drug discovery efforts is to obtain a global view of molecular mechanisms by studying cells as systems through the multidisciplinary and technology-driven approach of systems biology (4, 5).

Systems biology is the twenty-first century science that transformed the reductionist focus to a global view and added a new perspective to classical pharmaceutical research. In the context of drug discovery, systems biology involves integrating networks of biochemical reactions through experimental and computational approaches to provide a comprehensive understanding of the therapeutic target, including mechanisms of action. Systems biology emerged primarily as the result of the catalogue of genes provided by the multiple genome projects. The data from these genomics efforts were then utilized by other follow-up "omics" technologies—transcriptomics, proteomics, metabolomics, glycomics, among others—which were expected to provide a thorough understanding of the dynamics and interplay within biological systems. However, from the very beginning, the deluge of data produced by the experimental "omics" technologies proved overwhelming and created an immediate need to apply computational approaches for curation, pathway modeling, and bioinformatics (6, 7). Indeed, the past few years have seen monumental advances in this direction, and the in silico approaches have now taken center stage in systems biology (8, 9).

The availability of completed microbial genome sequences and the development of advanced microbial databases and high-performance software tools have opened up new opportunities for creating innovative computational methods for antimicrobial target identification. Traditionally, antimicrobial targets have been identified through knowledge of the function or essentiality of individual genes or proteins. Potential targets thus identified are generally taken through a validation process involving gene knockouts or site-directed mutagenesis experiments in whole cells or animals that lead to loss-of-function phenotypes. Experimental target validation can now be complemented with computational experiments such as in silico knockouts. In silico methods have the advantage of speed and low cost along with the ability to provide a systems view of the whole microbe at any given physiological stage. Consequently, in silico approaches are capable of generating hypotheses and questions that are unlikely to emerge through experimental methods (10).

In silico systems biology approaches are best used in combination with the experimental "omics" technologies. For example, when studying the proteome of *Mycobacterium tuberculosis* using mass spectrometry-based methods, we immediately realized that although the *M. tuberculosis* genome codes for about 4,000 proteins, several proteomics laboratories had identified only small subsets of this proteome. Also, because laboratories varied in sample preparation, chromatography, mass spectrometers, and bioinformatics, the types of proteins identified in these proteomics studies were significantly

different (11–16). Moreover, when we combined the lists of *M. tuberculosis* proteins identified through multiple proteomics experiments by several different laboratories, a consolidated list of only about 2,000 nonredundant proteins was generated, which is only 50 % coverage of the *M. tuberculosis* proteome (unpublished observation). Thus, if proteomics is being used as a primary technology for target identification, many of the target proteins of interest are probably not "visible" to the mass spectrometers. Consequently, a logical next step would be to evaluate observed proteomics data in the context of genome-scale microbial in silico models to fill the gaps in the experimental data sets.

Our laboratory recently conducted a comprehensive study (17) that further underscores the utility of in silico approaches for novel target identification. We recently reported in silico analysis of metabolic networks of a panel of representative gram-positive and gram-negative bacteria and provided valuable findings on metabolites that could be used as antimicrobial targets. In-depth literature mining was performed to identify metabolites that are essential for the growth and survival of a broad spectrum of bacteria as determined by direct experimental evidence. To identify potential targets among these essential metabolites, in silico pathway analysis was performed through the BioCyc Pathway/Genome Database (PGDB). A PGDB was automatically generated in BioCyc from the annotated genome sequence of that organism using BioCyc Pathway Tools software (developed by Peter D. Karp and coworkers at the Bioinformatics Research Group at SRI International—http://www.ecocyc.org/download.shtml). BioCyc has 2038 available PGDBs till date, each containing the predicted metabolic network of an organism, including metabolic pathways, enzymes (and the genes encoding them), metabolites (with structural details), and reaction details. BioCyc Pathway Tools software further produces a pathway-based visualization of cellular biochemical networks, called the cellular overview diagram, which supports interrogation and systems biology analyses of the whole organism. BioCyc provides overview diagrams for more than 3000 organisms from bacteria to humans. We used the cellular overview diagram for comparative analyses of the complete metabolic networks of two or more organisms. In the display of an overview for one organism, the software can highlight all reactions that are either shared or not shared with other combinations of organisms. In our work, the entire human metabolic network from HumanCyc was compared with the networks of pathogens of interest to search for metabolites that are absent from humans. Moreover, the reactions around the selected metabolites were compared for their presence or absence among the bacteria of choice to determine whether the metabolites are shared and hence can have broad-spectrum action. The essential metabolites that were absent in humans and present in multiple pathogens were chosen as potential targets. Additionally, metabolites were selected for which no alternative compensatory pathway was present.

Based on these analyses, we identified ten metabolites as potential candidates for developing novel antibiotics. These are lipid II, meso-2,6-diaminoheptanedioate (meso-DAP), pantothenate, biotin, shikimate, L-aspartyl-4-phosphate, deoxythymidine diphosphate (dTDP)-α-L-rhamnose, uridine diphosphate (UDP)-D-galacto-1,4-furanose, des-*N*-acetyl mycothiol, and siroheme. Previous identification of the first five metabolites as targets for antibiotic discovery validates our in silico approach and suggests that the latter five metabolites could be promising candidates as well. Identifying key metabolites and then developing metabolite scavengers as broad-spectrum antimicrobials is a relatively new approach that can benefit tremendously from in silico systems biology methods. Because conventional antimicrobial targets such as proteins and nucleic acids are amenable to mutations and antibiotic resistance, the relatively immutable metabolites provide a new frontier in antimicrobial drug discovery.

In addition to these efforts, here we have also performed in-depth data mining from publications that have reported antimicrobial targets derived through complementary in silico methods. We have classified these targets into two categories: metabolites (17–21) and metabolic enzymes (10, 19, 22–34). We observed that a few different in silico studies had independently arrived at the same metabolite or metabolic enzyme targets. Representative examples of such common targets are shown in Figs. 1 and 2. These targets not only present potential opportunities for developing broad-spectrum antimicrobials but also underscore the importance of conducting multiple and distinct in silico experiments for target identification.

Finally, whereas in silico approaches are showing great promise in antimicrobial drug discovery, the computational tools are scattered and lack standardization (35). One fundamental issue is the lack of a central resource that can provide comparative information on major microbial databases and software tools, which can be utilized for in silico systems biology, especially in the context of antimicrobial drug discovery. Accordingly, Tables 1, 2, and 3 in this chapter provide this information in a systematic tabular format to facilitate the selection of appropriate computational tools for researchers interested in in silico systems microbiology.

2 Microbial Databases

Systems-level investigation of genomic-scale information requires the development of integrated databases dealing with heterogeneous data, which can be queried for simple properties of genes or other database objects as well as for complex network-level properties, for the analysis and modeling of complex biological processes. Several database have been developed to provide valuable information from the bench chemist to biologist and from medical practitioner to pharmaceutical scientist, in a structured format. The advent of

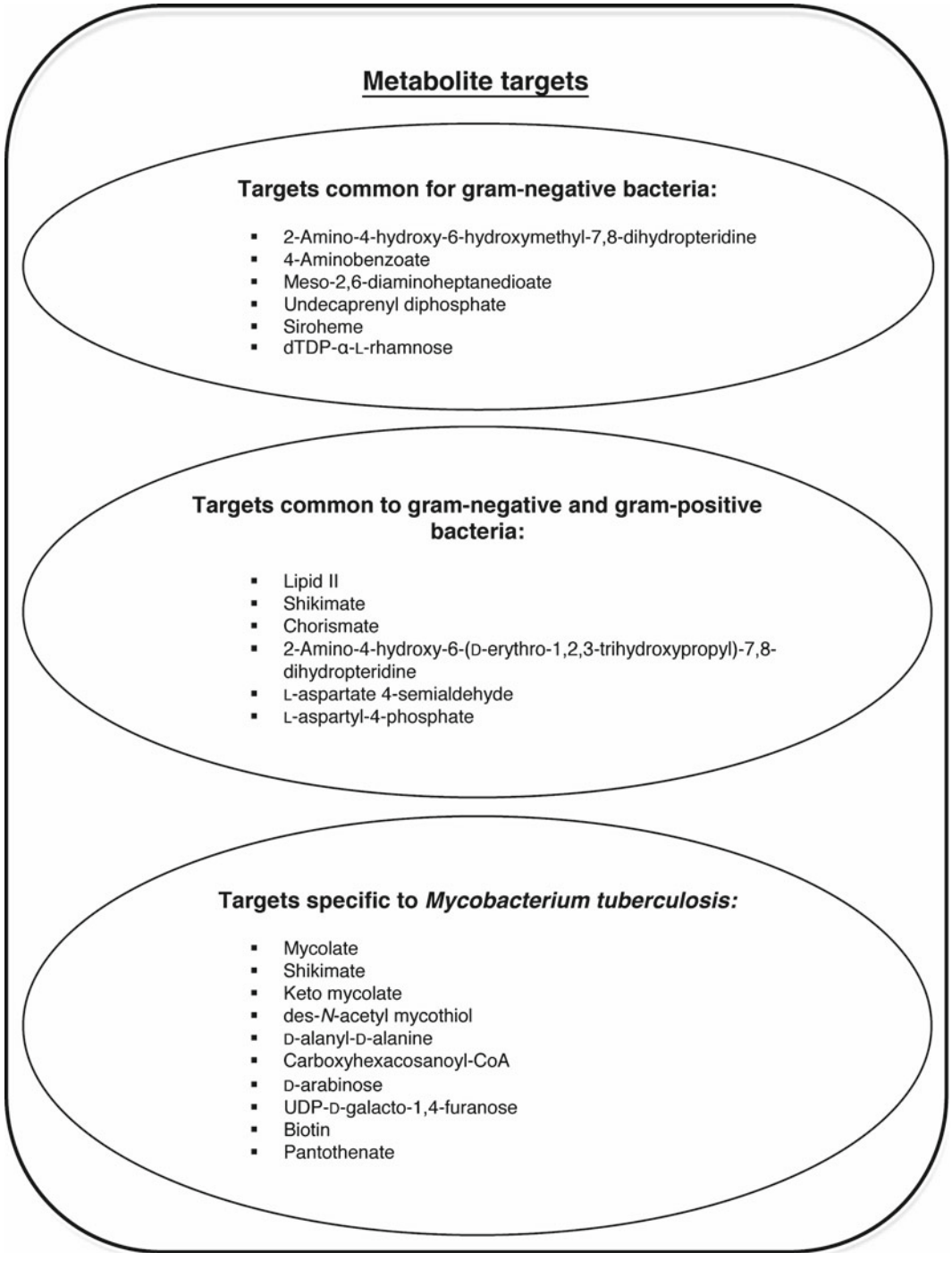

Fig. 1 Common microbial metabolites identified as drug targets. dTDP = deoxythymidine diphosphate, UDP = uridine diphosphate

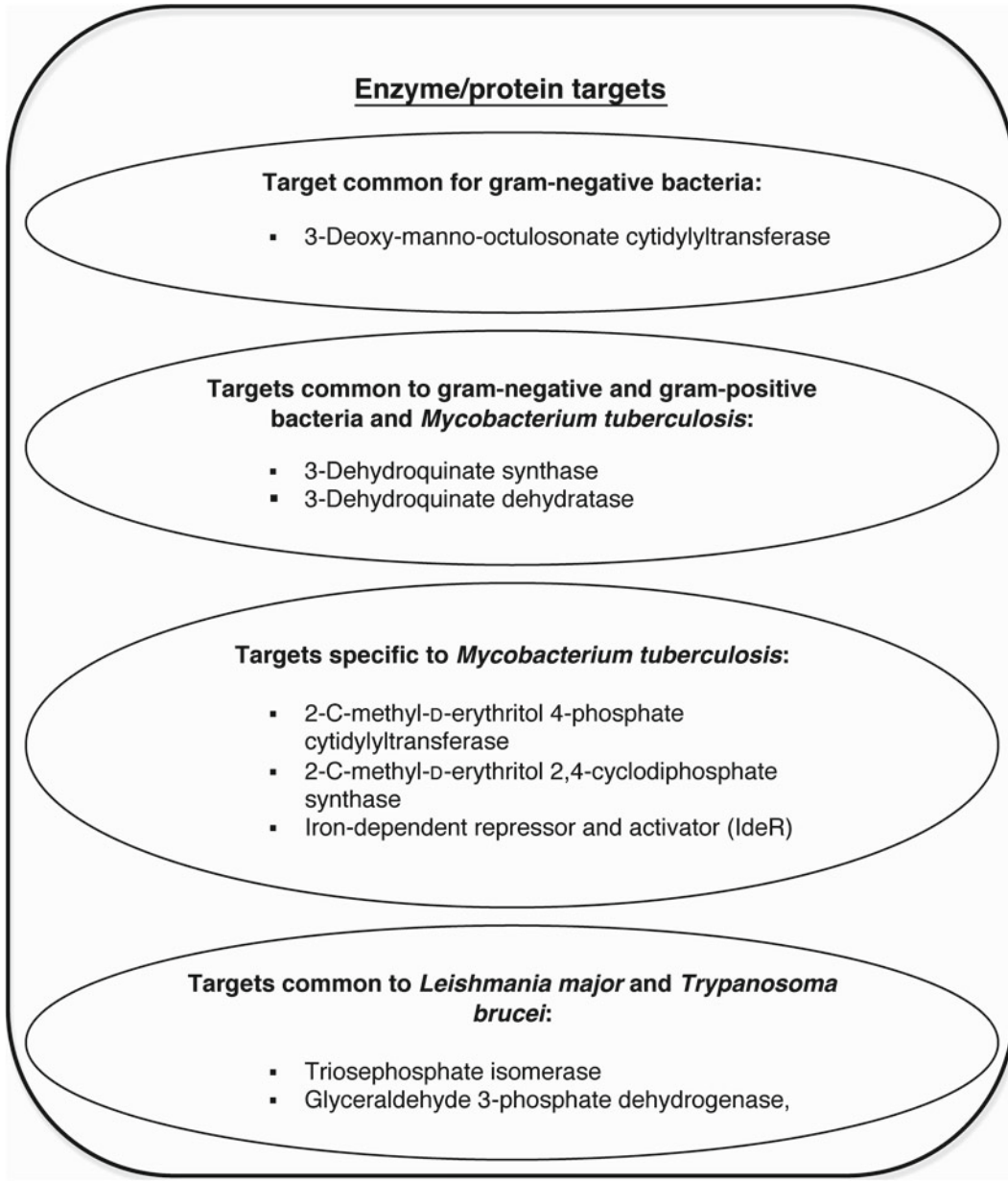

Fig. 2 Common microbial enzymes identified as drug targets

information technology and computational power enhanced the ability to access large volumes of data in the form of a database where one could do compilation, searching, archiving, analysis, and finally knowledge derivation (36).

Table 1 describes all the microbe-specific databases, and Table 2 lists databases for the microbial community. These databases provide genomic sequence data, gene and protein information, gene expression data, metabolic reactions and pathways, interaction network,

Table 1
Microbe-specific databases

Microbes	Database or Web resource	Description (URL)
Aspergillus spp.	*Aspergillus* Genome Database	Stanford University: Genomic sequence data and gene and protein information for aspergilli (http://www.aspgd.org/)
	Aspergillus Comparative Database	Broad Institute of MIT and Harvard: Comparative and functional genomics of seven aspergilli spp. (http://www.broadinstitute.org/annotation/genome/aspergillus_group/MultiHome.html)
Bacillus subtilis	SubtiList	Institut Pasteur: Genome annotation and analysis of bacterium *B. subtilis* 168 (http://genolist.pasteur.fr/SubtiList/)
	NRSub	University Lyon 1: Nonredundant, fully annotated database of sequences of *B. subtilis* 168 (http://pbil.univ-lyon1.fr/nrsub/nrsub.html)
Escherichia coli	EcoCyc	SRI International: Comprehensive literature-based curation of the entire genome and of transcriptional regulation, transporters, and metabolic pathways for bacterium *E. coli* K-12 MG1655 (http://ecocyc.org/)
	Ecogene	University of Miami School of Medicine: *E. coli* K-12 genome and proteome sequences, including extensive gene bibliographies (http://www.ecogene.org/3.0/)
	Colibri	Institut Pasteur: Genome analysis of *E. coli* (http://genolist.pasteur.fr/Colibri/)
Francisella tularensis	*Francisella tularensis* group Database	Broad Institute of MIT and Harvard: Comparative genomics analysis and virulence mechanisms of bacteria *F. tularensis* (http://www.broadinstitute.org/annotation/genome/francisella_tularensis_group/MultiHome.html)
Helicobacter pylori	PyloriGene	Institut Pasteur: Annotation and comparative analysis of bacteria *H. pylori* strains: 26695 and J99 (http://genolist.pasteur.fr/PyloriGene/)
Mycobacterium leprae	Leproma	Institut Pasteur: Genome analysis of the leprosy (Hansen disease) bacillus *M. leprae* (http://genolist.pasteur.fr/Leproma/)
Mycobacterium tuberculosis	TB Database	Stanford University: Provides genomic data (for 28 annotated genomes) and several thousand microarray datasets from in vitro experiments and *M. tuberculosis*-infected tissues (http://www.tbdb.org/)

(continued)

Table 1
(continued)

Microbes	Database or Web resource	Description (URL)
	TubercuList	Institut Pasteur: Complete dataset of DNA and protein sequences derived from *M. tuberculosis* H37Rv, linked to annotations and functional assignments (http://genolist.pasteur.fr/TubercuList/)
	*web*TB	TB Structural Genomics Consortium: Provides *M. tuberculosis* genome, structure summary for all known tuberculosis proteins, the *M. tuberculosis* regulatory database of proteins up- or downregulated in TB, top 100 persistence targets in TB (http://www.webtb.org/)
	TBrowse	Institute of Genomics and Integrative Biology—India: Resource for the integrative analysis of the *M. tuberculosis* genome (http://tbrowse.osdd.net/)
	TB Drug Resistance Mutation Database	Harvard School of Public Health: Provides mutations associated with *M. tuberculosis* drug resistance (http://www.tbdreamdb.com/)
Mycoplasma pulmonis	MypuList	Institut Pasteur: Genome analysis of the bacterium *M. pulmonis* (http://genolist.pasteur.fr/MypuList/)
Plasmodium falciparum	PlasmoDB	University of Georgia: Functional genomic database for malaria parasites, *P. falciparum*, *P. vivax*, *P. yoelii*, *P. berghei*, *P. chabaudi*, and *P. knowlesi* (http://plasmodb.org/plasmo/)
	Plasmodium falciparum database	Broad Institute of MIT and Harvard: Comparative genomics analysis of *Plasmodium* spp. (http://www.broadinstitute.org/annotation/genome/plasmodium_falciparum_spp/MultiHome.html)
Pseudomonas aeruginosa	*Pseudomonas* Genome Database	Simon Fraser University: Comparative genomics of all *Pseudomonas* spp. (http://www.pseudomonas.com/)
Saccharomyces cerevisiae	YeastCyc	SRI International: Comprehensive literature-based curation of the entire genome and of transcriptional regulation, transporters, and metabolic pathways of the budding yeast *S. cerevisiae* (http://biocyc.org/YEAST/organism-summary?object=YEAST)
	Saccharomyces Genome Database	Stanford University: Complete *S. cerevisiae* genomic sequence, its genes and their products, the phenotypes of its mutants, and the literature supporting these data (www.yeastgenome.org)
	Comprehensive Yeast Genome Database	Max-Planck-Institut für Biochemie: Molecular structure and functional network of *S. cerevisiae* and comparative analysis for related yeasts (http://mips.helmholtz-muenchen.de/genre/proj/yeast/)

(continued)

Table 1
(continued)

Microbes	Database or Web resource	Description (URL)
Vibrio cholerae	*Vibrio cholerae* Database	Broad Institute of MIT and Harvard: Comparative genomic studies of the different strains of *V. cholerae* (http://www.broadinstitute.org/annotation/genome/vibrio_cholerae/MultiHome.html)
Influenza virus	Influenza Research Database	University of Texas Southwestern Medical Center: Comprehensive, integrated data about influenza virus genome sequences, virus phenotypic characteristics, and results from surveillance activities for the discovery and development of influenza virus vaccines, diagnostics, and therapeutics (http://www.fludb.org/brc/home.do?decorator=influenza)
HIV	HIV Databases	Los Alamos National Laboratory: Contain data on HIV genetic sequences, immunological epitopes, drug resistance-associated mutations, and vaccine trials (http://www.hiv.lanl.gov/content/index)

Table 2
Databases and web resources for in silico systems microbiology

Database or Web resource	Description (URL)
ARDB	University of Maryland: Information on antibiotic resistance genes in sequenced bacteria (http://ardb.cbcb.umd.edu/)
BiGG	University of California San Diego: Knowledge base of large-scale metabolic reconstructions and high-quality curated metabolic models (http://bigg.ucsd.edu)
BioCyc	SRI International: Describes the genome and metabolic pathways of a single organism—total 1690 pathway/genome databases (http://biocyc.org/)
BioModels Database	EMBL-EBI: Repository of peer-reviewed, published, computational models (http://www.ebi.ac.uk/biomodels-main/)
CDD	Collaborative Drug Discovery: Repository of small-molecule libraries of more than 300,000 compounds derived from patents, literature, and high-throughput screening data shared by academic and pharmaceutical laboratories tested against *M. tuberculosis*; preliminary public antimalarial database from multiple sources on 30,000 public compounds (https://www.collaborativedrug.com/pages/public_access)

(continued)

Table 2
(continued)

Database or Web resource	Description (URL)
CSB.DB	Max Planck Institute of Molecular Plant Physiology: Presents the results of biostatistical analyses on gene expression data in association with additional biochemical and physiological knowledge (http://csbdb.mpimp-golm.mpg.de/csbdb/home/databases.html)
DOE Systems Biology Knowledgebase	U.S. Department of Energy: Community-driven cyberinfrastructure for sharing and integrating data and analytical tools for experimental design as well as modeling and simulation (http://genomicscience.energy.gov/compbio/index.shtml#page=news)
ERGO Light	Integrated Genomics: Curated database of public and proprietary genomic DNA with connected similarities, functions, pathways, functional models, and clusters (http://www.ergo-light.com/)
EuPathDB	Bioinformatics Resource Center: Provides genomic-scale datasets associated with the eukaryotic pathogens (http://eupathdb.org/eupathdb/)
GeneDB	The Wellcome Trust Sanger Institute: Genome database for prokaryotic and eukaryotic organisms (http://www.genedb.org/Homepage)
GTD	Institute of Integrative Omics and Applied Biotechnology—India: Provides putative genomic drug targets of most common human bacterial pathogens (http://iioab-dgd.webs.com/)
HAMAP	Swiss Institute of Bioinformatics: Database of completely sequenced microbial proteome sets and manually curated microbial protein families in UniProtKB/Swiss-Prot (http://hamap.expasy.org/)
HOGENOM	Université Claude Bernard : Database of homologous genes from fully sequenced organisms including bacteria (http://pbil.univ-lyon1.fr/databases/hogenom/acceuil.php)
KEGG	Kyoto University: Comprehensive database of biological systems, including genes, enzymes, metabolites, reactions, and pathways (http://www.genome.jp/kegg/)
MetaCyc	SRI International: Database of nonredundant, experimentally elucidated metabolic pathways—1,790 pathways from more than 2,216 different organisms (http://metacyc.org/)
MicrobesOnline	Virtual Institute for Microbial Stress and Survival: Community resource for comparative and functional genome analysis for over 1,000 complete genomes of bacteria, archaea, and fungi and thousands of expression microarrays from diverse organisms (http://www.microbesonline.org/)
MicroScope	LABGeM—Genoscope: Platform for microbial genome annotation and comparative genomics for 640 organisms (https://www.genoscope.cns.fr/agc/microscope/home/index.php)

(continued)

Table 2
(continued)

Database or Web resource	Description (URL)
MPIDB	J Craig Venter Institute: Provides all known physical microbial interactions (24,295 experimentally determined interactions) among proteins of 250 bacterial species/strains (http://www.jcvi.org/mpidb/about.php)
NIAID Systems Biology for Infectious Diseases Research	National Institute of Allergy and Infectious Diseases: Systems Biology Program for Infectious Disease Research comprising four centers: The TB Systems Biology Center, The Systems Virology Center, The Center for Systems Influenza, and The Center for Systems Biology for EnteroPathogens (http://www.niaid.nih.gov/labsandresources/resources/dmid/sb/Pages/default.aspx)
Pathema	J Craig Venter Institute: Core resource that supports basic research for a set of six target NIAID category A-C pathogens—genome sequencing and intergenomic comparisons (http://pathema.jcvi.org/Pathema/)
PATRIC	Virginia Bioinformatics Institute: Provides rich data and analysis tools for all bacterial species in the selected NIAID category A–C priority pathogens list (http://www.patricbrc.org/portal/portal/patric/Home)
RCBPR	Resource Center for Biodefense Proteomics Research: Proteomics and host-pathogen interactions for biodefense-related microorganisms (http://pir.georgetown.edu/pirwww/proteomics/)
TargetTrack	Protein Structure Initiative: Provides information on the experimental progress and status of target amino acid sequences selected for structural determination (http://sbkb.org/tt/)
TDR Targets Database	Universidad Nacional de General San Martín: Provides diverse datasets to facilitate the identification and prioritization of drugs and drug targets in neglected disease pathogens as both a Web site and a tool (http://tdrtargets.org/)
TransportDB	The Institute for Genomic Research: Describes the predicted cytoplasmic membrane transport protein complement for organisms whose complete genome sequence is available—includes 288 bacteria (http://www.membranetransport.org/)
VIDA	Virus Database at University College London: Contains a complete collection of homologous protein families derived from open reading frames from complete and partial virus genomes (http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA.html)
ViPR	NIAID Virus Pathogen Database and Analysis Resource: Provides a comprehensive data repository for all types of data related to 13 families of human pathogenic category A–C viral pathogens (http://www.viprbrc.org/brc/home.do?decorator=toga)
xBASE	University of Birmingham: Comprehensive resource for comparative bacterial genomics of 191 genomes (http://www.xbase.ac.uk/)

Table 3
Software tools for in silico systems biology

Software Tools	Description (URL)
Automated metabolic network reconstruction tools	
The SEED	University of Chicago: Develop comparative genomics environment and curated genomic data (http://theseed.uchicago.edu/FIG/index.cgi)
YANAvergence	Universität Würzburg: Provides a software framework for rapid network assembly, network overview, and network performance analysis (http://www.bioinfo.biozentrum.uni-wuerzburg.de/computing/yana)
Pathway tools software	SRI International: Pathway Tools software supports creation, editing, querying, visualization, analysis, and publishing of Pathway/Genome Database (http://bioinformatics.ai.sri.com/ptools/)
Metabolic network reconstruction software	
ERGO	Integrated Genomics, Inc.: Supports both automatic and manual genome-wide curation (https://ergo.integratedgenomics.com/)
SimPheny	Intrexon Corporation: Enables the development of predictive computer models of organisms, from bacteria to humans (http://g6g-softwaredirectory.com/bio/cross-omics/agent-based/20629-GT-Life-Sci-Genomatica-SimPheny.php)
Metabolic network analysis tools	
BioSDP	Universität Stuttgart: Matlab component specially designed for the analysis of uncertain biochemical networks via semidefinite programming (http://biosdp.sourceforge.net/)
geWorkbench	MAGNet: A Java-based open-source platform for integrated genomics (http://wiki.c2b2.columbia.edu/workbench/index.php/Home)
Machine learning tool	Technische Universität Wien: Does latent class analysis, short-time Fourier transform, fuzzy clustering, support vector machines, shortest path computation, bagged clustering, naive Bayes classifier, etc. (http://cran.r-project.org/web/packages/e1071/index.html)
NeAT (Network Analysis Tools)	Université Libre de Bruxelles: Toolbox for the analysis of biological networks, clusters, classes, and pathways (http://rsat.bigre.ulb.ac.be/rsat/index_neat.html)
Modeling, simulation, and analysis software	
CellDesigner	Systems Biology Institute—Japan: Structured diagram editor for drawing gene regulatory and biochemical networks with links to simulation and other analysis packages (http://www.celldesigner.org/)
Cellware	Bioinformatics Institute—Singapore: Grid-based modeling and simulation tool that conducts modeling and simulation of gene regulatory and metabolic pathways (http://www.bii.a-star.edu.sg/achievements/applications/cellware/index.php)
COPASI	Virginia Bioinformatics Institute & EML Research: Software application for simulation and analysis of biochemical networks and their dynamics (http://www.copasi.org/tiki-index.php)
Dizzy	Stephen Ramsey—Institute for Systems Biology: Chemical kinetics stochastic simulation software package written in Java (http://magnet.systemsbiology.net/software/Dizzy/)

(continued)

Table 3
(continued)

Software Tools	Description (URL)
Dynetica	California Institute of Technology: Simulator of dynamic networks written in Java that does model building for systems expressed as reaction networks (http://www.duke.edu/~you/Dynetica_page.htm)
E-Cell	Keio University: Object-oriented software suite for modeling, simulation, and analysis of large-scale complex systems such as biological cells (http://www.e-cell.org/models/)
Simmune	NIAID Laboratory of Systems Biology: Suite of software tools tool for simulating and analyzing immune system behavior (http://www.niaid.nih.gov/LabsAndResources/labs/aboutlabs/lsb/Pages/simmuneproject.aspx)
VCell	University of Connecticut Health Center: Supports complex models with a Web-based Java interface to specify compartmental topology and geometry, molecular characteristics, and relevant interaction parameters (http://www.nrcam.uchc.edu/)
Simulation software (flux balance analysis)	
Clp	Coin-or-linear programming: Open-source linear programming solver written in C++ to find solutions of mathematical optimization (https://projects.coin-or.org/Clp)
COBRA Toolbox	University of California San Diego: MATLAB package for constraint-based reconstruction and analysis methods to simulate, analyze, and predict a variety of metabolic phenotypes using genome-scale models (http://opencobra.sourceforge.net/openCOBRA/Welcome.html)
Fluxor	Jeremy Zucker—Harvard Medical School: Python command-line tool that takes a metabolic network specified in Systems Biology Markup Language (SBML) and performs flux balance analysis using the GNU Linear Programming Kit (GLPK) and SWIG (http://fluxor.sourceforge.net/)
GAMS	GAMS Development Corporation: High-level modeling system for mathematical programming and optimization (http://www.gams.com/)
GPLK	GNU Project: Set of routines written in ANSI C for solving large-scale linear programming (http://www.gnu.org/s/glpk/)
GLPKMEX	GNU Project: Matlab MEX interface for the GLPK library for solving linear programming (http://glpkmex.sourceforge.net/)
ILOG CPLEX 8.100	IBM: Provides flexible, high-performance mathematical programming solvers for linear programming (http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/)
Matlab	MathWorks: High-level language and interactive environment that enables simulation of biochemical networks using integrated flux balance analysis, regulatory flux balance analysis, and ordinary differential equations (http://www.mathworks.com/products/matlab/)
MetaFluxNet	Korea Advanced Institute of Science and Technology: Program package for quantitatively analyzing metabolic fluxes (http://mbel.kaist.ac.kr/lab/mfaml/main.html?page=metafluxnet.html)
Yices	SRI International: Constraint solver that can handle flux balance analysis (http://yices.csl.sri.com/)

(continued)

Table 3
(continued)

Software Tools	Description (URL)
Visualization software	
Cytoscape	Institute of Systems Biology: Open-source software platform for network data integration, analysis, and visualization (http://cytoscape.org/)
GraphViz	Stephen C. North—AT&T Labs Research: Open-source graph visualization software (http://www.graphviz.org/)
Paintomics	Centro de Investigaciones Príncipe Felipe: Web tool for the integration and visualization of transcriptomics and metabolomics data (http://www.paintomics.org/)
VisANT	Boston University: Integrative visual analysis tool for biological networks and pathways (http://visant.bu.edu/)
Network layout tools	
EPE	University of Edinburgh: Visual editor designed for annotation, visualization, and presentation of wide variety of biological networks, including metabolic, genetic, and signal transduction pathways (http://epe.sourceforge.net/SourceForge/EPE.html)
JDesigner	ERATO project—Caltech: Win32 application, which allows one to draw a biochemical network and export the network in the form of SBML (http://sbw.kgi.edu/software/jdesigner.htm)
Pathway Projector	Keio University: Provides integrated pathway maps that are based upon the KEGG Atlas, with the addition of nodes for genes and enzymes, implemented as a scalable, zoomable map utilizing the Google Maps API (http://www.g-language.org/PathwayProjector/)
yEd	yWorks: Powerful diagram editors that can be used to quickly and effectively generate high-quality drawings of diagrams (http://www.yworks.com/en/products_yed_about.html)
Microarray data analysis tools	
GO.tools	The Gene Ontology: Tools for analysis of microarray data (http://www.geneontology.org/GO.tools.microarray.shtml)
Mayday	University of Tübingen: Graphical user interface for visualization, analysis, and storage of microarray data (http://www.microarray-analysis.org/mayday)
Microarray DB	Keio University: Tool for mapping transcriptome data onto KEGG pathways and for creating a Web-based database with an overview of the entire pathway (http://www.g-language.org/data/marray/)
SAM	Stanford University: Supervised learning software for genomic expression data mining (http://www-stat.stanford.edu/~tibs/SAM/)
General "omics" data analysis tools	
BL-SOM	Platform for Riken Metabolomics: An integrated analytical tool for a range of "omics" data (http://prime.psc.riken.jp/?action=blsom_index)
DAnTE	Pacific Northwest National Laboratory: Allows users to perform various downstream data analysis, normalization, data reduction, and hypothesis testing steps (http://omics.pnl.gov/software/DAnTE.php)

(continued)

Table 3
(continued)

Software Tools	Description (URL)
Pathway Tools Omics Viewer	EcoCyc—SRI International: Paints data values from the user's high-throughput and other experiments onto the cellular overview diagram for an organism (http://biocyc.org/expression.html)
VANTED	Leibniz Institute of Plant Genetics and Crop Plant Research—Germany: A tool for the visualization and analysis of networks with related experimental data (http://vanted.ipk-gatersleben.de/)
Metabolomics-specific data analysis tools	
MathDAMP	Keio University: Allows visualization of differences between metabolite profiles acquired by hyphenated mass spectrometry techniques (http://mathdamp.iab.keio.ac.jp/)
metAlign	Wageningen UR: Computer software tool for the analysis, alignment, and comparison of full-scan mass spectrometry datasets (http://www.metalign.wur.nl/UK/Download+and+publications/)
MetATT	University of Alberta: A web-based tool for time-series and two-factor metabolomic data analysis (http://metatt.metabolomics.ca/MetATT/)
MetaboAnalyst	University of Alberta: A web-based analytical pipeline for high-throughput metabolomics studies (http://www.metaboanalyst.ca/MetaboAnalyst/faces/Home.jsp)
metaP-Server	Helmholtz Zentrum München: Automates data analysis for the processing of metabolomics experiments (http://metabolomics.helmholtz-muenchen.de/metap2/)
MSFACTs	The Samuel Roberts Noble Foundation: Metabolomics spectral formatting, alignment, and conversion tools (http://www.noble.org/PlantBio/Sumner/msfacts/index.html)
MZmine 2	VTT Technical Research Centre of Finland: Toolbox for processing and visualization of mass spectrometry-based molecular profile data (http://mzmine.sourceforge.net/download.shtml)
SpectConnect	Massachusetts Institute of Technology: Systematic identification of conserved metabolites in gas chromatography/mass spectrometry data for metabolomics (http://spectconnect.mit.edu/)
SpinAssign	Platform for Riken Metabolomics: Provides batch annotations of a large number of metabolites against user nuclear magnetic resonance peaks (http://prime.psc.riken.jp/?action=nmr_search)
XCMS	Scripps Center for Metabolomics: Software for processing liquid chromatography–mass spectrometry-based metabolomics data (http://metlin.scripps.edu/xcms/)
Statistical computing software	
SAS software	SAS: An integrated system of software products for statistical analysis (http://www.sas.com/technologies/analytics/statistics/stat/)
SPSS	IBM: A computer program used for data mining and statistical analysis (http://www-01.ibm.com/software/analytics/spss/)

(continued)

Table 3
(continued)

Software Tools	Description (URL)
STATISTICA	StatSoft: Provides a comprehensive and integrated set of tools and solutions for data visualization, graphical data analysis, visual data mining, visual querying (http://www.statsoft.com/unique-features/statistica-general-overview/)
R Project	The R Foundation: A free software environment for statistical computing and graphics (http://www.r-project.org/)

comparative and functional genomics, information on mutation, virulence and drug resistance, libraries of small-molecule lead compounds, high-throughput experimental data based on transcriptomics, proteomics, and metabolomics. The microbes include gram-negative and gram-positive bacteria, protozoa, and viruses.

3 Software Tools for Microbial Systems Analysis

System-level studies are often built on molecular and genetic findings and "omics" studies including genomics, proteomics, and metabolomics. The main challenges in systems biology are the complexity of the systems, the vast quantities of data, and the scattered pieces of knowledge, all of which must be integrated; therefore, systematic computational tools are crucially important. Understanding complex biological systems requires extensive support from software tools. Such tools are needed at each step of a systems biology computational workflow, which typically consists of data handling, network inference, deep curation, dynamical simulation, and model analysis (37).

Table 3 lists the major advanced computational software tools that are currently used for data analysis, visualization, modeling, simulation, and statistical computing, especially for microbial metabolic networks, models, and "omics" experiments. The given selection while intended to cover currently available software in this field is subjective, and the reader should consider available literature to focus on the specialized aspects and specific applications of the listed databases and software tools.

References

1. Bevan P, Ryder H, Shaw I (1995) Identifying small-molecule lead compounds: the screening approach to drug discovery. Trends Biotechnol 13:115–121
2. Kitano H (2002) Systems biology: a brief overview. Science 295:1662–1664
3. Gwynn MN, Portnoy A, Rittenhouse SF et al (2010) Challenges of antibacterial discovery revisited. Ann N Y Acad Sci 1213:5–19
4. Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. Nat Biotechnol 22:1249–1252

5. Butcher EC, Berg EL, Kunkel EJ (2004) Systems biology in drug discovery. Nat Biotechnol 22:1253–1259
6. Davidov E, Holland J, Marple E et al (2003) Advancing drug discovery through systems biology. Drug Discov Today 8:175–183
7. Aderem A (2005) Systems biology: its practice and challenges. Cell 121:511–513
8. Palsson B (2000) The challenges of in silico biology. Nat Biotechnol 18:1147–1150
9. Kitano H (2002) Computational systems biology. Nature 420:206–210
10. Raman K, Yeturu K, Chandra N (2008) TargetTB: a target identification pipeline for *Mycobacterium tuberculosis* through an interactome, reactome and genome-scale structural analysis. BMC Syst Biol 2:109
11. Gu S, Chen J, Dobos KM et al (2003) Comprehensive proteomic profiling of the membrane constituents of a *Mycobacterium tuberculosis* strain. Mol Cell Proteomics 2:1284–1296
12. Bahk YY, Kim SA, Kim JS et al (2004) Antigens secreted from *Mycobacterium tuberculosis*: identification by proteomics approach and test for diagnostic marker. Proteomics 4: 3299–3307
13. Mawuenyega KG, Forst CV, Dobos KM et al (2005) *Mycobacterium tuberculosis* functional network analysis by global subcellular protein profiling. Mol Biol Cell 16:396–404
14. Mattow J, Siejak F, Hagens K et al (2007) An improved strategy for selective and efficient enrichment of integral plasma membrane proteins of mycobacteria. Proteomics 7:1687–1701
15. Malen H, Berven FS, Fladmark KE et al (2007) Comprehensive analysis of exported proteins from *Mycobacterium tuberculosis* H37Rv. Proteomics 7:1702–1718
16. Gonzalez-Zamorano M, Mendoza-Hernandez G, Xolalpa W et al (2009) *Mycobacterium tuberculosis* glycoproteomics based on ConA-lectin affinity capture of mannosylated proteins. J Proteome Res 8:721–733
17. Sarker M, Chopra S, Mortelmans K et al (2011) In silico pathway analysis predicts metabolites that are potential antimicrobial targets. J Comput Sci Syst Biol 4:021–026
18. Munger J, Bennett BD, Parikh A et al (2008) Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nat Biotechnol 26:1179–1186
19. Kim HU, Kim TY, Lee SY (2010) Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen Acinetobacter baumannii AYE. Mol Biosyst 6:339–348
20. Kim TY, Kim HU, Lee SY (2010) Metabolite-centric approaches for the discovery of antibacterials using genome-scale metabolic networks. Metab Eng 12:105–111
21. Kim HU, Kim SY, Jeong H et al (2011) Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. Mol Syst Biol 7:460
22. Schilling CH, Palsson BO (2000) Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. J Theor Biol 203:249–283
23. Yeh I, Hanekamp T, Tsoka S et al (2004) Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. Genome Res 14:917–924
24. Rahman SA, Schomburg D (2006) Observing local and global properties of metabolic pathways: 'load points' and 'choke points' in the metabolic networks. Bioinformatics 22:1767–1774
25. Jamshidi N, Palsson BO (2007) Investigating the metabolic capabilities of Mycobacterium tuberculosis H37Rv using the in silico strain iNJ661 and proposing alternative drug targets. BMC Syst Biol 1:26
26. Chavali AK, Whittemore JD, Eddy JA et al (2008) Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. Mol Syst Biol 4:177
27. Mazumdar V, Snitkin ES, Amar S et al (2009) Metabolic network model of a human oral pathogen. J Bacteriol 191:74–90
28. Oberhardt MA, Goldberg JB, Hogardt M et al (2010) Metabolic network analysis of *Pseudomonas aeruginosa* during chronic cystic fibrosis lung infection. J Bacteriol 192:5534–5548
29. Raghunathan A, Shin S, Daefler S (2010) Systems approach to investigating host-pathogen interactions in infections with the biothreat agent Francisella. Constraints-based model of *Francisella tularensis*. BMC Syst Biol 4:118
30. Crowther GJ, Shanmugam D, Carmona SJ et al (2010) Identification of attractive drug targets in neglected-disease pathogens using an in silico approach. PLoS Negl Trop Dis 4:e804
31. Plata G, Hsiao TL, Olszewski KL et al (2010) Reconstruction and flux-balance analysis of the Plasmodium falciparum metabolic network. Mol Syst Biol 6:408
32. Navratil V, De Chassey B, Combe CR et al (2011) When the human viral infectome and diseasome networks collide: towards a systems biology platform for the aetiology of human diseases. BMC Syst Biol 5:13
33. Fatumo S, Plaimas K, Adebiyi E et al (2011) Comparing metabolic network models based on genomic and automatically inferred enzyme information from Plasmodium and its human host to define drug targets in silico. Infect Genet Evol 11:708–715

34. Fang K, Zhao H, Sun C et al (2011) Exploring the metabolic network of the epidemic pathogen *Burkholderia cenocepacia* J2315 via genome-scale reconstruction. BMC Syst Biol 5:83
35. Ng A, Bursteinas B, Gao Q et al (2006) Resources for integrative systems biology: from data through databases to networks and dynamic system models. Brief Bioinform 7:318–330
36. Jagarlapudi SA, Kishan KV (2009) Database systems for knowledge-based discovery. Methods Mol Biol 575:159–172
37. Ghosh S, Matsuoka Y, Asai Y et al (2011) Software for systems biology: from tools to integrated platforms. Nat Rev Genet 12:821–832