

The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases

Ron Caspi¹, Tomer Altman¹, Richard Billington¹, Kate Dreher², Hartmut Foerster³, Carol A. Fulcher¹, Timothy A. Holland¹, Ingrid M. Keseler¹, Anamika Kothari¹, Aya Kubo¹, Markus Krummenacker¹, Mario Latendresse¹, Lukas A. Mueller³, Quang Ong¹, Suzanne Paley¹, Pallavi Subhraveti¹, Daniel S. Weaver¹, Deepika Weerasinghe¹, Peifen Zhang² and Peter D. Karp^{1,*}

¹SRI International, 333 Ravenswood, Menlo Park, CA 94025, USA, ²Carnegie Institution, Department of Plant Biology, 260 Panama Street, Stanford, CA 94305, USA and ³Boyce Thompson Institute for Plant Research, Tower Road, Ithaca, New York 14853 USA

Received October 14, 2013; Accepted October 19, 2013

ABSTRACT

The MetaCyc database (MetaCyc.org) is a comprehensive and freely accessible database describing metabolic pathways and enzymes from all domains of life. MetaCyc pathways are experimentally determined, mostly small-molecule metabolic pathways and are curated from the primary scientific literature. MetaCyc contains >2100 pathways derived from >37 000 publications, and is the largest curated collection of metabolic pathways currently available. BioCyc (BioCyc.org) is a collection of >3000 organism-specific Pathway/Genome Databases (PGDBs), each containing the full genome and predicted metabolic network of one organism, including metabolites, enzymes, reactions, metabolic pathways, predicted operons, transport systems and pathway-hole fillers. Additions to BioCyc over the past 2 years include YeastCyc, a PGDB for *Saccharomyces cerevisiae*, and 891 new genomes from the Human Microbiome Project. The BioCyc Web site offers a variety of tools for querying and analysis of PGDBs, including Omics Viewers and tools for comparative analysis. New developments include atom mappings in reactions, a new representation of glycan degradation pathways, improved compound structure display, better coverage of enzyme kinetic data, enhancements of the Web Groups functionality, improvements to the Omics viewers, a new

representation of the Enzyme Commission system and, for the desktop version of the software, the ability to save display states.

INTRODUCTION

MetaCyc (MetaCyc.org) is a highly curated nonredundant reference database of small-molecule metabolism. It contains metabolic pathway and enzyme data that have been experimentally validated and reported in the scientific literature (1). Owing to its exclusively experimentally determined pathways and enzymes, intensive curation and tight integration of data and references, MetaCyc is a uniquely valuable resource for various fields including biochemistry, enzymology, genome and metagenome analysis and metabolic engineering. The metabolic pathways and enzymes in MetaCyc are derived from all domains of life.

In conjunction with its role as a general reference on metabolism, MetaCyc can be used as a reference database for the PathoLogic component of the Pathway Tools software (2) to computationally predict the metabolic network of any organism that has a sequenced and annotated genome (3). During this automated process, a predicted metabolic network is created in the form of a Pathway/Genome Database (PGDB). In addition to the automated creation of PGDBs, Pathway Tools has editing capabilities that enable scientists to improve and update these computationally generated PGDBs by manual curation. MetaCyc has been used by SRI International (SRI) to create >3000 PGDBs (as of October 2013), which are available through the BioCyc (BioCyc.org)

*To whom correspondence should be addressed. Tel: +1 650 859 4358; Fax: +1 650 859 3735; Email: pkarp@ai.sri.com

Web site. Interested scientists may adopt any of these PGDBs through the BioCyc Web site for further curation (biocyc.org/intro.shtml#adoption).

MetaCyc is also used by other scientists to create additional PGDBs, many of which are available to the public via the scientists' own Web sites. Together with BioCyc, these PGDBs form the *MetaCyc family of databases* (4).

More than 250 groups have used Pathway Tools and MetaCyc to create PGDBs for their organisms of interest, including important model organisms such as *Saccharomyces cerevisiae* (5), *Arabidopsis thaliana* (6), *Oryza sativa* (7), *Mus musculus* (8), *Bos taurus* (9), *Medicago truncatula* (10), *Populus trichocarpa* (11), *Dictyostelium discoideum* (12), *Leishmania major* (13), *Chlamydomonas reinhardtii* (14), several *Solanaceae* species (15), bioenergy-related organisms (BeoCyc) and many pathogenic organisms (16) (see <http://biocyc.org/otherpgdbs.shtml> for a more complete list). Examples of organisms that were studied during the previous 2 years using Pathway Tools include archaea (17,18), bacteria (19–49), fungi (50–54), a diatom (55), plants (56–59) and lower eukaryotes (60,61). In addition, Pathway Tools is used to analyze data from the Human Microbiome Project (62–66) and other metagenomic data sets (27,67,68).

A web server included in Pathway Tools enables the publishing of PGDBs through either the Internet or an internal network, and the Navigator component of Pathway Tools allows the browsing and analyzing of PGDBs, either locally or over the Internet. A detailed description of Pathway Tools can be found in (2).

PGDBs generated by Pathway Tools and MetaCyc are an excellent platform for the integration of genome information with many other types of data comprising metabolism, regulation and genetics. They provide powerful tools for analyzing omics data sets from experiments related to gene transcription, metabolomics, proteomics,

ChIP-chip analysis and other resources. During the past 2 years, we again significantly expanded the data content of MetaCyc and BioCyc, and added supporting enhancements to the Pathway Tools software and BioCyc Web site, as described in the following sections.

METACYC ENHANCEMENTS

Expansion of MetaCyc

All pathways in MetaCyc are curated from the experimental literature. Since the last *Nucleic Acids Research* publication (2 years ago) (1), we added 384 new base pathways (pathways comprised of reactions only, where no portion of the pathway is designated as a subpathway) and 33 superpathways (pathways composed of at least one base pathway plus additional reactions or pathways), and updated 154 existing pathways, for 538 new and revised pathways. The total number of base pathways grew by 17%, from 1790 (version 15.5) to 2097 (version 17.5) (the total increase is <384 pathways because some existing pathways were deleted from the database during this period). A comparison of MetaCyc 16.0 and a kyoto encyclopedia of genes and genomes (KEGG) version downloaded on 27 February 2012 showed that MetaCyc contained significantly more reactions and pathways than did KEGG, although the number of reactions occurring in pathways in the two databases was similar (69).

Along with the increase in pathway number, the number of enzymes, reactions, chemical compounds and citations in the database grew by 20, 19, 11 and 21%, respectively; the number of referenced organisms increased by 18% (currently at 2460). See Table 1 for a list of species with >20 experimentally elucidated pathways in MetaCyc, and Table 2 for the taxonomic distribution of all MetaCyc pathways.

Table 1. List of species with >20 experimentally elucidated pathways represented in MetaCyc (meaning that there is experimental evidence for the occurrence of these pathways in the organism)

Bacteria		Eukarya		Archaea	
<i>Escherichia coli</i>	312	<i>Arabidopsis thaliana</i>	328	<i>Methanocaldococcus jannaschii</i>	25
<i>Pseudomonas aeruginosa</i>	70	<i>Homo sapiens</i>	229	<i>Methanosarcina barkeri</i>	21
<i>Bacillus subtilis</i>	60	<i>Saccharomyces cerevisiae</i>	172	<i>Sulfolobus solfataricus</i>	21
<i>Pseudomonas putida</i>	50	<i>Rattus norvegicus</i>	81		
<i>Salmonella typhimurium</i>	41	<i>Glycine max</i>	62		
<i>Pseudomonas fluorescens</i>	31	<i>Solanum lycopersicum</i>	55		
<i>Mycobacterium tuberculosis</i>	31	<i>Pisum sativum</i>	55		
<i>Klebsiella pneumoniae</i>	26	<i>Mus musculus</i>	51		
<i>Enterobacter aerogenes</i>	25	<i>Nicotiana tabacum</i>	46		
<i>Agrobacterium tumefaciens</i>	23	<i>Zea mays</i>	45		
		<i>Solanum tuberosum</i>	43		
		<i>Oryza sativa</i>	41		
		<i>Hordeum vulgare</i>	27		
		<i>Spinacia oleracea</i>	27		
		<i>Catharanthus roseus</i>	26		
		<i>Triticum aestivum</i>	25		
		<i>Bos taurus</i>	21		
		<i>Petunia x hybrida</i>	21		

The species are grouped by taxonomic domain and are ordered within each domain based on the number of pathways (number following species name) to which the given species was assigned.

Table 2. The distribution of pathways in MetaCyc based on the taxonomic classification of associated species

Bacteria		Eukarya		Archaea	
Proteobacteria	1042	Viridiplantae	916	Euryarchaeota	137
Firmicutes	319	Fungi	343	Crenarchaeota	42
Actinobacteria	280	Metazoa	312	Thaumarchaeota	1
Bacteroidetes/Chlorobi	62	Euglenozoa	28		
Cyanobacteria	65	Alveolata	16		
Deinococcus-Thermus	28	Amoebozoa	10		
Tenericutes	18	Stramenopiles	7		
Thermotogae	25	Fornicata	4		
Aquificae	16	Rhodophyta	4		
Spirochaetes	12	Haptophyceae	4		
Chlamydiae -Verrucomicrobia	7	Parabasalia	3		
Planctomycetes	6				
Chloroflexi	5				
Fusobacteria	4				
Nitrospirae	2				
Thermodesulfobacteria	2				
Chrysiogenetes	1				

For example, the statement ‘Tenericutes 18’ means that there is experimental evidence for at least 18 MetaCyc pathways for their occurrence in members of this taxonomic group. Major Taxonomic groups are grouped by domain and are ordered within each domain based on the number of pathways (number following taxon name) associated with the taxon. A pathway may be associated with multiple organisms.

Atom mapping

A reaction atom mapping describes for each atom of a reactant (excluding hydrogens) its corresponding atom in the product. Implicitly, an atom mapping illustrates which bonds are broken and created during the reaction. Atom mapping information is depicted in PGDB reactions by coloring conserved chemical moieties within a reaction (currently available only in the Firefox and Chrome browsers). In addition, if the user hovers the mouse over an atom in a reactant, the corresponding atom is highlighted in the product (again, only in the Firefox and Chrome browsers).

The atom-mapping data of each reaction can also be downloaded from the MetaCyc Web site after the reaction page is displayed by selecting the ‘Download atom mapping(s) for this reaction’ command from the right side bar. All atom mappings for MetaCyc are stored in one flat file, atom-mappings.dat, and the Mol files for all compounds involved in the atom mappings are stored in MetaCyc-MOLfiles.tgz. The atom mapping encoding, as stored in atom-mappings.dat, is described at http://biocyc.org/PGDBConceptsGuide.shtml#node_sec_3.5. The atom mappings were computed by a technique described in (70). The error rate in computed atom mappings has been evaluated at <2%, although some atom mappings may possibly be incorrect due to specific enzyme activities that have not been taken into account.

In MetaCyc version 17.5, >10 100 reactions have computed atom mappings. The vast majority of reactions that lack atom-mapping data are either not completely mass balanced or they include substrates without an atomic structure. In a few rare cases, reactions were not processed because the computation would be too time consuming. In contrast, ~5% of the reactions have multiple atom mappings, often due to symmetries in the compound structure. We tried to eliminate such

duplicates, but we kept the cases where the enzyme might operate in more than one way.

The atom mappings are also used by the RouteSearch Tool (See Section below).

New representation of glycan-degradation pathways

The degradation of large and complex glycan polymers poses a challenge for standard pathway diagrams. Rather than a linear process, the degradation of such polymers often consists of multiple types of enzymes simultaneously attacking different types of bonds within the polymer. The enzymes work in parallel, resulting in the liberation of small fragments. Because no particular order exists for these attacks, an attempt to show the process as linear, involving specific intermediates, is misleading. To overcome this limitation, we developed a new type of pathway diagram that shows the precise location of the sites attacked by the different enzymes by using color-coded arrows pointing to the cleaved bonds within the polymer structure (Figure 1). To make these diagrams easier to comprehend, we simplified the polymer structure by using glycan monomers as the basic building blocks. These pathways are often shown as a complex reaction, with the initial polymer structure on the left, and the final small products on the right.

Pathway Tools now supports the symbolic representation of glycans recommended by the Consortium for Functional Glycomics (CFG) and uses the Glyco-CT format for the import/export of such structures. To enable the curation of glycan structures, we developed a Pathway Tools interface for the GlycanBuilder software (71,72). GlycanBuilder is a tool that enables fast and intuitive drawing of glycan structures and was originally developed as the main interface for structure searches and results display in the EUROCarbDB databases.

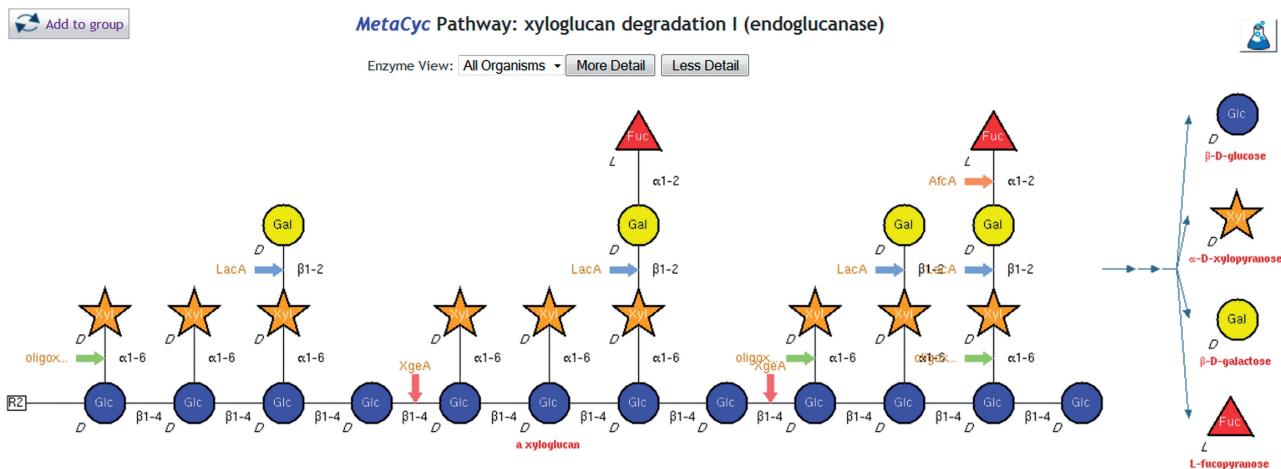


Figure 1. The new glycan degradation pathways use symbolic representation to illustrate the structures of complex glycan molecules. Colored arrows show the sites that are cleaved by enzymes and provide hyperlinks to those enzymes. The final products produced by the combined degradation of the polymer by all enzymes are listed on the right side of the diagram.

The introduction of the symbolic structures did not replace the existing atomic structures—glycan molecules in MetaCyc may contain both the regular atomic structure that is used for all chemical compounds, and the CFG symbolic representation.

Kinetic data in PGDBs

We have recently revised the types of enzyme kinetic data that can be captured in Pathway Tools PGDBs, the interface used to enter these data and the presentation of the data. When the reaction is reversible, capturing the optimal temperature and pH for each direction is now possible. K_m , V_{max} , K_{cat} and specific activity are now collected separately for each reactant, including for alternative substrates. K_i values for inhibitors are collected as before.

Version 17.5 of MetaCyc includes 3883 enzymatic reactions with K_m data (5965 K_m values), 242 enzymatic reactions with V_{max} values, 390 enzymatic reactions with K_{cat} values and 201 enzymatic reactions with specific activity values.

In addition, all kinetic data are now presented in a table format that makes it much easier to read (Figure 2).

New representation of the EC system in MetaCyc

The Enzyme Commission (EC) classifies enzymes based on the reaction(s) that they catalyze (see <http://www.chem.qmul.ac.uk/iubmb/enzyme/rules.html>). Since its creation, Pathway Tools has encoded this information by assigning the EC number to the reaction catalyzed by the enzyme, with a limitation that only one EC number could be assigned to each reaction (although multiple reactions could be assigned to the same number). However, this approach gave limited compatibility with the many-to-many relationship between EC numbers and reactions that is used by the EC system. To increase compatibility with the EC system we have implemented a new way of encoding EC numbers. A new object type (EC-number) was added to the database to represent EC numbers (Figure 3). These EC-number objects have their own

page, which contains the information drafted by the EC as well as links to several external databases. Any number of reactions can be linked to these EC-number objects, either as ‘official’ EC reactions, meaning that the reaction precisely matches the reaction(s) specified by the EC for this EC number, or as ‘unofficial’ EC reactions, meaning that while the reaction is not identical to the one used by the EC, it is implied to be catalyzed by this type of enzyme. When an enzyme has been assigned to all official reactions of a particular EC number, the software automatically recognizes that it fulfills the definition requirements for that EC number, and lists that enzyme in the EC-number page. Thus, we have implemented, in a dynamic computational manner, the principle of enzyme classification as defined by the EC, which is based on the reactions catalyzed by the enzyme.

Data integration with other databases

EC classification

MetaCyc is regularly updated with data from the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB), which includes new and modified EC entries. The data are retrieved from the ExplorEnz database (www.enzyme-database.org) (73). The EC entries at ExplorEnz and MetaCyc are linked to each other.

NCBI taxonomy

The full NCBI Taxonomy database (74) is integrated into Pathway Tools, enabling specification of taxa using NCBI Taxonomy, and allowing taxonomic querying of MetaCyc pathways and enzymes. We continue to update the taxonomy entries with each major release of MetaCyc.

Gene ontology

The mapping between MetaCyc reactions and Gene Ontology (GO) process and function terms (75) is being continuously maintained by the GO Editorial Office at the EBI. An updated file is at <http://www.geneontology.org/external2go/metacyc2go>.

Enzymatic reaction of: myo-inositol 2-dehydrogenase

EC Number: 1.1.1.18



The reaction direction shown, that is, $A + B \leftrightarrow C + D$ versus $C + D \leftrightarrow A + B$, is in accordance with the direction of enzyme catalysis.

This reaction is reversible. [Ramaley, 1979]

Alternative Substrates for *myo*-inositol: [α-D-glucose \[Ramaley, 1979\]](#) , [α-D-xylopyranose \[Ramaley, 1979\]](#) , [D-pinitol \[Morinaga, 2006\]](#)

In Pathways: [myo-, chiro- and scillo-inositol degradation](#) , [myo-inositol degradation I](#)

Kinetic Parameters:

Substrate	K_m (μM)	k_{cat} (sec^{-1})	k_{cat}/K_m ($\text{sec}^{-1} \mu\text{M}^{-1}$)	V_{max} ($\mu\text{mol mg}^{-1} \text{min}^{-1}$)	Specific Activity (U/mg)	Citations
α-D-glucose	56000.0					[Ramaley, 1979]
NADH	36.0					[Daniellou, 2006]
scyllo-inosose	1000.0	13.0	0.013	19.2		[Daniellou, 2006]
myo-inositol	1100.0	9.0	0.0082	13.5	34.4	[Daniellou, 2006; Ramaley, 1979]
NAD⁺	230.0					[Ramaley, 1979]

$\text{pH}_{(\text{opt})}$ (forward direction): 9.5 [Ramaley, 1979]

$\text{pH}_{(\text{opt})}$ (reverse direction): 7 [Ramaley, 1979]

Figure 2. This figure illustrates some of the different types of enzymatic kinetic data that can be captured and presented by Pathway Tools. The software lets the curator enter the data using the units reported in a paper, and converts them automatically to the standard units. When possible, the catalytic efficiency is computed automatically and included in the table. Temperature and pH optima can be captured differently for the two directions of a reversible reaction.

Links to other databases

During the past 2 years we added new links from MetaCyc to several external databases, which are listed in Table 3.

EXPANSION OF BIOCYC

The BioCyc databases are organized into three tiers.

- Tier 1 PGDBs have received at least 1 year of manual curation. Although some Tier 1 PGDBs (e.g. MetaCyc and EcoCyc) have received decades of manual curation and are updated continuously, others are less well curated and are still in need of significant curation.
- Tier 2 PGDBs have received moderate amounts of review (less than a year), and may or may not be updated on an ongoing basis.
- Tier 3 PGDBs were created computationally and received no subsequent manual review or updating.

During the past 2 years, the number of BioCyc PGDBs increased from 1129 (version 15.1) to 2988 (version 17.1). The Tier 1 PGDB YeastCyc (*S. cerevisiae*), which has been curated for many years by the saccharomyces genome database, is now hosted at BioCyc.org and has undergone significant curation in the past year. The number of pathways in YeastCyc has grown from 154 in December 2012 to 259 in October 2013. The curation of fungal pathways will be one of our priorities for the next few years.

The HumanCyc PGDB (*Homo sapiens*, curated by SRI), the AraCyc PGDB (*Arabidopsis thaliana*, curated by the Plant Metabolic Network) and the LeishCyc PGDB (*Leishmania major* strain *Friedlin*, curated by a team from the University of Melbourne) have been upgraded to Tier 1 status, bringing the total of Tier 1 PGDBs to six (along with EcoCyc, MetaCyc and YeastCyc). As of version 17.1, Tier 2 includes 35 PGDBs, and Tier 3 includes 2947 PGDBs. Some Tier 2 PGDBs were provided by groups outside SRI. The

Sites ▾ Search ▾ Genome ▾ Metabolism ▾ Analysis ▾ Help ▾

MetaCyc EC 2.7.4.6 -- nucleoside-diphosphate kinase

Parent Class: [EC-Numbers](#) → [2](#) -- [Transferases](#) → [2.7](#) -- [Transferring phosphorous-containing groups](#) → [2.7.4](#) -- [Phosphotransferases with a phosphate group as acceptor](#)

Synonyms: nucleoside 5'-diphosphate kinase, nucleoside diphosphate (UDP) kinase, nucleoside diphosphokinase, nucleotide phosphate kinase, UDP kinase, uridine diphosphate kinase

Systematic Name: ATP:nucleoside-diphosphate phosphotransferase

Unification Links: [BRENDA:2.7.4.6](#) , [ENZYME:2.7.4.6](#) , [IUBMB-ExplorEnz:2.7.4.6](#)

Reaction:
 $\text{a nucleoside diphosphate} + \text{ATP} \rightarrow \text{a nucleoside triphosphate} + \text{ADP}$

Unofficial Reactions:
 $\text{CDP} + \text{ATP} \rightarrow \text{CTP} + \text{ADP}$,
 $\text{dADP} + \text{ATP} \rightarrow \text{dATP} + \text{ADP}$,
 $\text{dCDP} + \text{ATP} \rightarrow \text{dCTP} + \text{ADP}$,
 $\text{dGDP} + \text{ATP} \rightarrow \text{dGTP} + \text{ADP}$,
 $\text{dTDP} + \text{ATP} \rightarrow \text{dTTP} + \text{ADP}$,
 $\text{dUDP} + \text{ATP} \rightarrow \text{dUTP} + \text{ADP}$,
 $\text{GDP} + \text{ATP} \rightarrow \text{GTP} + \text{ADP}$,
 $\text{GDP} + \text{ADP} = \text{GTP} + \text{AMP}$,
 $\text{ATP} + \text{IDP} = \text{ITP} + \text{ADP}$,
 $\text{ATP} + \text{dIDP} = \text{ADP} + \text{dITP}$,
 $\text{UDP} + \text{ATP} \rightarrow \text{UTP} + \text{ADP}$

Enzymes and Genes:
nucleoside diphosphate kinase : [NME1](#) , [NME2](#) (*Homo sapiens*)
nucleoside-diphosphate kinase 1 : [NDPK1](#) (*Arabidopsis thaliana* col)
nucleoside diphosphate kinase : [YNK1](#) (*Saccharomyces cerevisiae*)
nucleoside diphosphate kinase : [ndk](#) (*Escherichia coli* K-12 substr. MG1655)

Summary:
Many nucleoside diphosphates can act as acceptors, while many ribo- and deoxyribonucleoside triphosphates can act as donors.

Citations: [[Berg54](#), [Ayengar56](#), [Kirkland59](#), [Krebs53](#), [Nakamura66](#), [Ratliff64](#)]

References

[Ayengar56](#): Ayengar P, Gibson DM, Sanadi DR (1956). "Transphosphorylations between nucleoside phosphates." *Biochim Biophys Acta* 21(1);86-91. PMID: 13363863

[Berg54](#): Berg P, Joklik WK (1954). "Enzymatic phosphorylation of nucleoside diphosphates." *J Biol Chem* 210(2);657-72. PMID: 13211603

[Kirkland59](#): Kirkland RJ, Turner JF (1959). "Nucleoside diphosphokinase of pea seeds." *Biochem J* 72;716-20. PMID: 14409347

[Krebs53](#): Krebs HA, Hems R (1953). "Some reactions of adenosine and inosine phosphates in animal tissues." *Biochim Biophys Acta* 12(1-2);172-80. PMID: 13115426

[Nakamura66](#): Nakamura H, Sugino Y (1966). "Metabolism of deoxyribonucleotides. 3. Purification and some properties of nucleoside diphosphokinase of calf thymus." *J Biol Chem* 241(21);4917-22. PMID: 5925862

[Ratliff64](#): Ratliff RL, Weaver RH, Lardy HA, Kuby SA (1964). "Nucleoside triphosphate-nucleoside diphosphate transphosphorylase (nucleoside diphosphokinase). I. Isolation of the crystalline enzyme from brewers' yeast." *J Biol Chem* 239;301-9. PMID: 14114857

Figure 3. EC numbers are now database objects that have their own pages. An EC-Number page includes all of the information defined by the EC, and additional information that includes a list of unofficial reactions (see text for details) and a list of enzymes determined by the software to fit the definition of the EC number.

database authors are identified on the database summary page (Analysis → Summary Statistics).

Inclusion of Human Microbiome Project genomes

As fully sequenced and annotated genomes become available from the Human Microbiome Project (http://www.hmpdacc.org/catalog/grid.php?dataset=genomic&project_status=Complete), they are integrated into the BioCyc collection. Version 17.5 includes 891 such genomes.

SOFTWARE AND WEB SITE ENHANCEMENTS

The following sections describe significant enhancements to Pathway Tools (which powers the BioCyc Web site) during the past 2 years.

Object-specific sidebar on BioCyc web pages

A new right-sidebar appears on BioCyc web pages (Figure 4). This sidebar contains operations specific to the currently displayed BioCyc web page. For example, when a metabolic pathway page is displayed, the sidebar includes operations such as customizing the layout of the pathway and painting omics data on it. When a gene page is displayed, the sidebar includes operations such as displaying the gene sequence and producing a comparative genome browser view of that gene alongside specified orthologs.

Improved compound structure graphics

The graphic display of the chemical structures on the compound and reaction pages has been re-implemented

Table 3. During the past 2 years we added new links from MetaCyc to the following external databases

Database name	Description	URL
Direct links		
dictyBase	A <i>Dictyostelium discoideum</i> model organism database	dictybase.org
DIP	A database of interacting proteins	dip.doe-mbi.ucla.edu
DisProt	A database of protein disorder	disprot.org
EuPathDB	A eukaryotic pathogen database	eupathdb.org
Expression atlas	A database of analyzed ArrayExpress Archive results	www.ebi.ac.uk/gxa
FlyBase	A <i>Drosophila melanogaster</i> model organism database	flybase.org
MINT	A molecular interaction database	mint.bio.uniroma2.it
PDB	A database of 3D structures of large biological molecules	rcsb.org/pdb
PDBsum	A pictorial database of PDB structures	www.ebi.ac.uk/pdbsum
PhosphoSitePlus	A database for protein post-translational modifications	Phosphosite.org
PRIDE	A proteomics identifications database	ebi.ac.uk/pride
Protein model portal	A database of protein models computed by comparative modeling methods	proteinmodelportal.org
Rhea	A manually annotated database of chemical reactions	ebi.ac.uk/rhea
STRING	A database of known and predicted protein-protein interactions	string-db.org
Swiss-model repository	A database of annotated three-dimensional comparative protein structure models generated by Swiss-Model	swissmodel.expasy.org/repository
'In-family' type links		
CAZy	A carbohydrate-active enzymes database	cazy.org
InterPro	A protein sequence functional analysis database	ebi.ac.uk/interpro
PANTHER	A database for protein analysis through evolutionary relationships	pantherdb.org
Pfam	A protein families database	pfam.sanger.ac.uk
PRINTS-S	A database of protein family fingerprints	bioinf.manchester.ac.uk/dbbrowser/sprint
ProDom	A database of protein domain families	prodom.prabi.fr
PROSITE	A database of protein domains, families and functional sites	prosite.expasy.org
SMART	A simple modular architecture research tool	smart.embl.de

using the scalable vector graphics web standard, resulting in higher quality graphics. This improvement is currently visible only when using the recent versions of the Firefox and Chrome web browsers.

Web Groups enhancements

A Web Group is a spreadsheet-like structure that can contain both Pathway Tools objects and other values such as numbers or strings. Like a spreadsheet, it is organized by rows and columns, and the user can add or delete any of them. A typical group contains a set of Pathway Tools objects in the first column (e.g. a set of genes generated by a search). The other columns contain properties of the object (e.g. the chromosomal position of each gene), or the result of a transformation (e.g. the reactions catalyzed by the gene products, or the corresponding genes from a different organism).

Web Groups can be created from search results, by importing data from external text files, and by adding objects individually from either their web pages or from another group. For example, a Web Group can contain a column of genes and columns of gene expression values, and the contents of the group can be painted onto a BioCyc metabolic map diagram using the Cellular Omics Viewer. Web Groups can be shared either publicly or with selected users.

Group transformations facilitate converting an existing group into a new group or into a new column in an existing group. Many new transformations have been added during the past 2 years, including several regulation-related transformations for genes (e.g. transforming a list of genes into a list of the transcription factors that

regulate their expression), the ability to transform a group of genes into a group of the upstream promoters of those genes, to transform a protein into a list of regulatory DNA sites it binds to and to transform a compound into a list of proteins it is known to either bind to, activate or inhibit.

A relatively recent innovation is the ability to incorporate nucleotide and amino-acid sequence data as group objects. Such sequence data can be automatically added to groups that contain genes or proteins. Genes, promoters and transcription-factor binding sites can be transformed not only into their sequence but also into a list of their coordinates in the genome. A list of DNA regions or point locations (e.g. mutation locations) can be imported from a file to form a group, which could then be transformed into the set of genes nearest those regions.

The Web Groups interface also enables users to apply an enrichment/depletion analysis to the contents of a group (e.g. given a list of genes, the user can easily compute whether that list is statistically overrepresented for genes within specific metabolic pathways, or for genes that are regulated by particular transcriptional regulators).

Metabolic RouteSearch

RouteSearch is a new web-based tool (accessible from the top menu command Metabolism → Metabolic Route Search) that generates reaction pathways connecting starting and ending metabolites specified by the user. Optional parameters include the number of routes to return, the maximum route length, the cost of using a native reaction (a reaction already found in the metabolic

Sites Search Genome Metabolism Analysis Groups Help

[Add to group](#)

Escherichia coli K-12 substr. MG1655 Enzyme: 4-hydroxy-tetrahydrodipicolinate synthase

Gene: [dapA](#) Accession Numbers: EG10205 (EcoCyc), b2478, ECK2474

Synonyms: DHDP5, dihydrodipicolinate synthase

Regulation Summary Diagram: ?

Summary:

4-Hydroxy-tetrahydrodipicolinate synthase, historically called dihydrodipicolinate synthase (DHDP5, DapA) is the first enzyme unique to lysine biosynthesis, catalyzing the condensation of pyruvate and (S)-aspartate B-semialdehyde. This is thought to be the rate-limiting step in lysine biosynthesis after aspartate kinase III [Laber92]. The product of the reaction catalyzed by DapA was identified as (4S)-4-hydroxy-2,3,4,5-tetrahydro-(2S)-dipicolinate (HTPA) [Blickling97].

The reaction proceeds via a ping-pong bi-bi mechanism; pyruvate initially binds to the enzyme via a Schiff base to the ϵ -amino group of the active site Lys161 residue [Laber92]. This is followed by addition of L-aspartate semialdehyde and transamination leading to cyclization and dissociation of HTPA [Blickling97]. The kinetic mechanism was refined using initial velocity and dead-end inhibition studies at both high and low pH, confirming the ping-pong reaction mechanism of the enzyme [Karsten97]. Surprisingly, Lys161 is not absolutely essential for catalysis [Soares10].

Crystal structures of the apo-enzyme and in complexes with substrates, substrate analogs and inhibitors, as well as of mutant enzymes have been solved [Mirwaldt95, Blickling97, Dobson04b, Dobson05, Dobson05a, Griffin08, Dobson08, Pearce08, Devenish08, Dobson09, Soares10, Boughton12]. DapA is a homotetramer that can be characterized as a dimer of dimers; each monomer consists of an N-terminal (B/ α)₈-barrel domain and three C-terminal α -helices. Site-directed mutants in the catalytic triad residues Tyr133, Thr44 and Tyr107 provided evidence for their functional importance [Dobson04b, Dobson09]. Arg138 plays a role in substrate binding [Dobson05a], and Ile203 may play a role in catalysis [Dobson08]. Although the active site of a dimeric variant of DapA is not disturbed, it shows reduced activity [Griffin08]. A Y107W mutant exists as a mixture of monomers and tetramers in solution and has reduced catalytic activity [Pearce08], and a L197D/Y107W double mutant is monomeric, has reduced catalytic efficiency and is not inhibited by lysine [MuscroftTaylor10a]. Molecular dynamics simulations indicate that the tetrameric form is relatively rigid, while the dimeric form is more flexible and shows disorder in the active site [Reboul2]. The C-terminal domain is required for maintenance of quaternary structure of the enzyme and thus for catalysis [Guo09]. The chaperone GroE appears to be required for folding of DapA [McLennan98].

hide

Escherichia coli K-12 substr. MG1655

Enzyme:
4-hydroxy-tetrahydrodipicolinate synthase

OPERATIONS

- ▶ Show this gene in another database
- ▶ Change organisms/databases for comparison operations
- ▶ Search for this gene in other databases
- ▶ Show orthologs (with operon diagrams) in multiple databases
- ▶ Align in Multi-Genome Browser

Sequences

- ▶ Protein Sequence
- ▶ Nucleotide Sequence

PortEco Links

- ▶ PORTECO: [dapA](#)
- ▶ ECOLIWIKI: [b2478](#)
- ▶ Significant Gene Expression Conditions
- ▶ Gene Expression Profiles
- ▶ Significant Knockout Phenotype Conditions

Figure 4. The new right-sidebar on BioCyc web pages contains operations that are specific to the currently displayed page. The operations and links available on the sidebar change depending on the type of object that is currently displayed. In this example, the operations and links are relevant to an *Escherichia coli* gene/protein page. Operations and links that are not specific to a particular object type are available from the menu bar at the top of the page and do not change.

network of the organism, as opposed to a reaction that has to be imported from MetaCyc), the cost of losing an atom along the way and the atom species to take into account. Specifying the maximum amount of time allowed for searching for routes is also possible.

RouteSearch only returns linear pathways from the starting to the ending metabolite. Along that linear pathway, it computes the weighted sum of the atoms lost, based on the cost of atom loss provided by the user. To do so, it uses the atom-mapping data already computed for MetaCyc (see section Atom Mapping) because the atom mappings define which atoms are transferred between compounds. The objective of RouteSearch is to minimize this sum. RouteSearch also simultaneously minimizes the length of the pathways found, as the weighted sum of all reactions used to reach the ending metabolite adds to the overall cost.

Notice that searching for such optimal routes may not return well-known routes from a starting to an ending

metabolite because the objective is to minimize both the number of reactions used and the number of atoms lost.

RouteSearch is a new tool and is still undergoing evaluation and potential redesign.

Improvements to omics data display on the web cellular overview

The Cellular Omics Viewer enables users to paint an organism-specific metabolic map diagram for any BioCyc PGDB with multicolor highlighting to represent large-scale omics data sets, as well as to animate the highlighting to show temporal changes in the data. Any type of data that can be mapped to a compound, a reaction, an enzyme or a gene is supported, although the most common data types are gene-expression, reaction flux and metabolomics data. During the previous year, we reengineered the web-based Omics Viewer to improve its performance. Both speed and browser resource use were

improved by one to two orders of magnitude. Currently the initial times to load data and display the resulting highlights are on the order of a few to the low tens of seconds.

Generating omics pop-ups

New functionality makes it possible to display per-node omics data in a pop-up window as a column chart, an x-y plot or a heat map. This new functionality is invoked by hovering the mouse over a reaction or metabolite of interest and selecting the ‘Omics’ option in the menu of the resulting pop-up window, which will graph the omics data for that object in column chart (bar) mode. Clicking different tabs converts the graph to an x-y plot or a heat map. Customizing the data labels is also possible.

The same type of pop-up can be generated to show all the data for a given pathway. Right-clicking a reaction within the pathway opens a pop-up that includes the option ‘Display Omics Data for Every Node in Pathway’ (Figure 5).

Customizing a pathway diagram with omics data

A new function enables the painting of omics data on a full-scale pathway diagram. The new functionality is invoked from the right-side bar by using the command ‘Customize Pathway Diagram’, which displays a window that includes an option for painting omics data onto the pathway diagram (Figure 6).

Generating a table of pathways with omics data

Generating a table displaying omics data painted onto small diagrams of all individual pathways is now possible. A ‘Show data’ selector has been added to the

Omics Viewer dialog, which enables users to select whether they want the omics data painted on the cellular overview, the table of individual pathways or both.

MetaFlux enhancements

The Flux Balance Analysis (FBA) module of Pathway Tools, called MetaFlux, enables the creation of steady-state quantitative metabolic flux models. MetaFlux is capable of solving FBA models, performing multiple gap-filling and performing multiple gene or reaction knockouts. Its latest enhancements include (i) the ability to specify compartments for the metabolites in the biomass reaction, the list of nutrients and the list of secreted metabolites; (ii) a much faster instantiation of generic reactions; (iii) some enhancements to the graphical user interface; and (iv) a new development mode called Fast-Gap.

In the regular development mode, gap-filling can be done simultaneously on reactions, nutrients, secretions and the biomass reaction. This regular development mode uses Mixed-Integer Linear Programming, which can be computationally time-consuming: it may require several hours of computing. The new Fast-Gap mode is limited to gap-filling only reactions, but it executes in a short time, typically in less than 1 min (at most a few minutes). Therefore, Fast-Gap can be used instead of the regular development mode when a fast answer is desired. Fast-Gap can also provide, in some cases, a more meaningful reaction gap-filling solution than the regular mode due to its use of a different optimization technique.

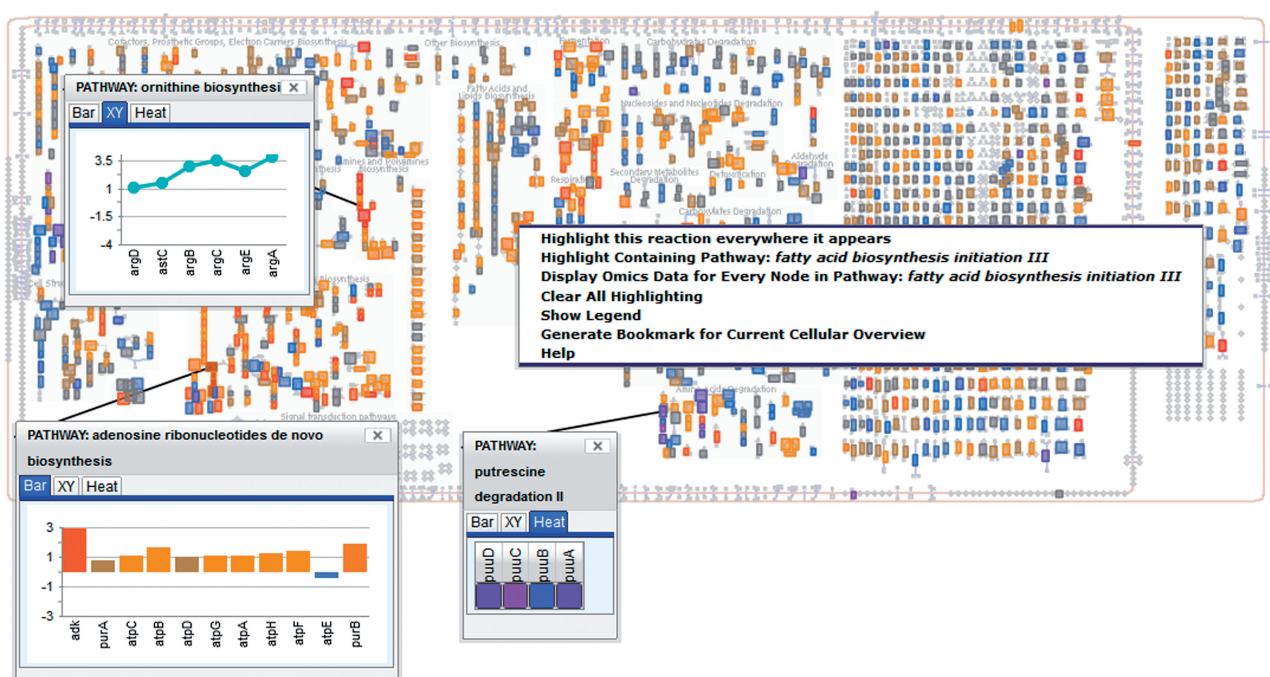


Figure 5. The Cellular Omics Viewer allows the user to paint omics data over the cellular Overview. New functionality enables the display of per-node omics data in a pop-up window as a column chart, an x-y plot or a heat map. The pop-ups can also be generated to show all the data for a given pathway. This figure also shows the pop-up that appears on right-clicking a reaction.

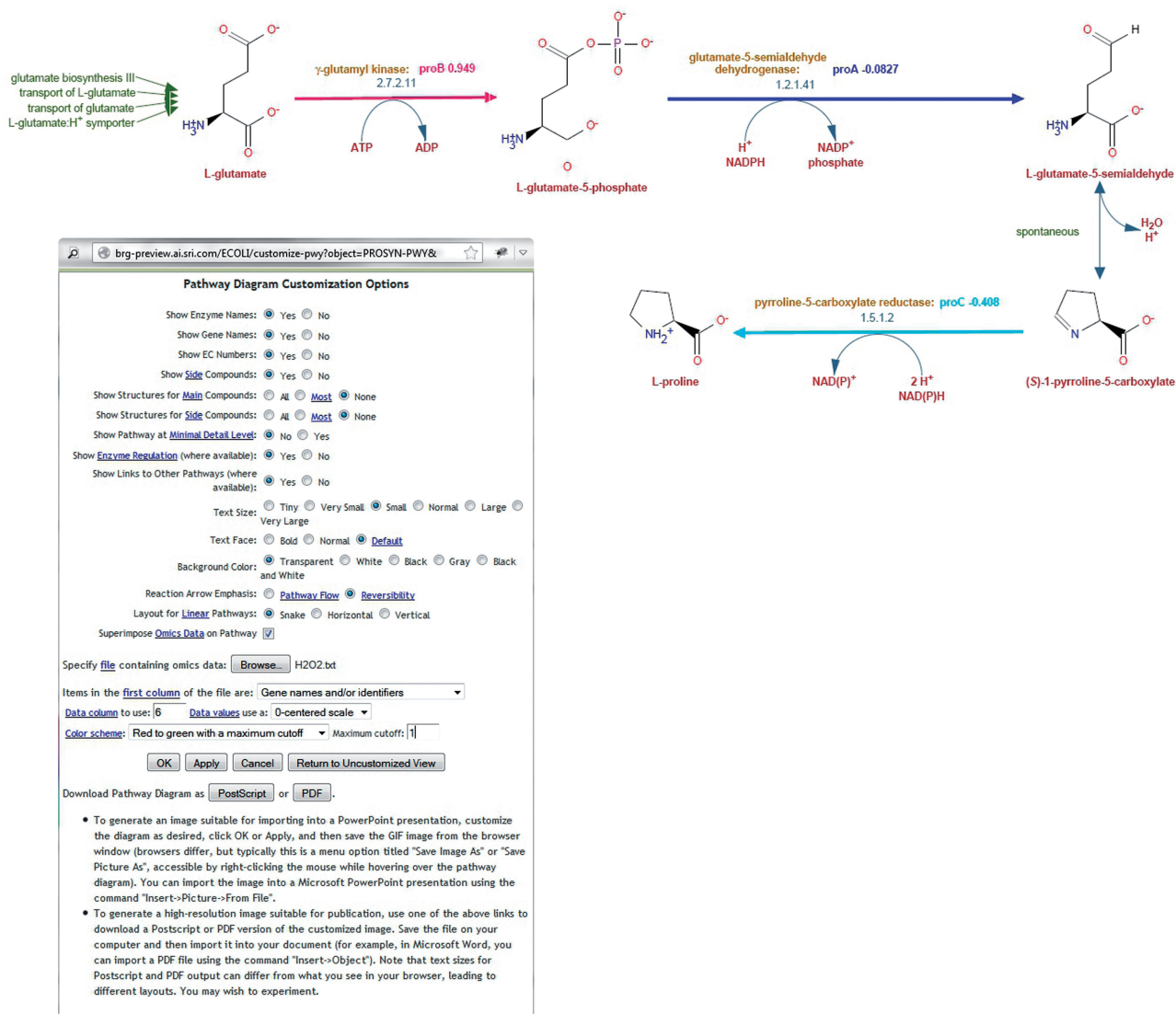


Figure 6. Pathway diagram customization is available via the web interface, and lets the user control many aspects of the pathway diagram. A new option allows painting user-supplied Omics data directly to the pathway. The modified diagram can be exported to a pdf or postscript format file for incorporation in presentations or manuscripts.

ENHANCEMENTS TO THE DESKTOP VERSION OF PATHWAY TOOLS

The following enhancements only apply to the desktop version of the Pathway Tools software.

Ability to save display state

The display state of Pathway Tools can now be saved to a file, which could be used for later restoration. Examples of display states that can be saved include the state of the omics viewers (including omics pop-ups), genome-browser tracks and cloned windows. The display-state file can be e-mailed to another user, who could then restore the exact same state on a different computer. Saving a display state to a file is invoked via the command **File** → **Save Display State to File**.

Improved interface for the PGDB registry

Users that install Pathway Tools on their computer can download and install any of the PGDBs available on the BioCyc Web site by using an embedded utility called PGDB Registry (accessible from the command Tools → Browse PGDB Registry). This utility enables downloading and installing a PGDB with a few mouse clicks. However, the proliferation in the number of PGDBs available for download had resulted in a major slowdown of the utility. The interface of the utility has been completely redesigned, so that finding PGDBs within the registry is now fast.

HOW TO LEARN MORE ABOUT METACYC AND BIOCYC

The BioCyc.org and MetaCyc.org Web sites provide several informational resources, including an online

BioCyc guided tour (76), a guide to the BioCyc database collection (77), a guide for MetaCyc (78), a guide for EcoCyc (79), a guide to the concepts and science behind PGDBs (80) and instructional webinar videos that describe the usage of BioCyc and Pathway Tools (81). We routinely host workshops and tutorials (on site and at conferences) that provide training and in-depth discussion of our software for both beginning and advanced users. To stay informed about the most recent changes and enhancements to our software, please join the BioCyc mailing list at <http://biocyc.org/subscribe.shtml>. A list of our publications is available online (82).

DATABASE AVAILABILITY

The MetaCyc and BioCyc databases are freely and openly available to all. See <http://biocyc.org/download.shtml> for download information. New versions of the downloadable data files and of the BioCyc and MetaCyc Web sites are released three or four times per year.

ACKNOWLEDGEMENTS

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the preceding agencies.

FUNDING

National Institute of General Medical Sciences of the National Institutes of Health (NIH) [GM080746, GM077678, GM088849 and GM75742]; Department of Energy [DE-SC0004878 toward Bioenergy-related pathway curation]; National Science Foundation [IOS-1026003 and DBI-0640769 toward MetaCyc curation performed by the Plant Metabolic Network]. Funding for open access: NIH.

Conflict of interest statement. None declared.

REFERENCES

- Caspi,R., Altman,T., Dreher,K., Fulcher,C.A., Subhraveti,P., Keseler,I.M., Kothari,A., Krummenacker,M., Latendresse,M., Mueller,L.A. *et al.* (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **40**, D742–D753.
- Karp,P.D., Paley,S.M., Krummenacker,M., Latendresse,M., Dale,J.M., Lee,T.J., Kaipa,P., Gilham,F., Spaulding,A., Popescu,L. *et al.* (2010) Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform.*, **11**, 40–79.
- Dale,J.M., Popescu,L. and Karp,P.D. (2010) Machine learning methods for metabolic pathway prediction. *BMC Bioinformatics*, **11**, 15.
- Karp,P.D. and Caspi,R. (2011) A survey of metabolic databases emphasizing the MetaCyc family. *Arch. Toxicol.*, **85**, 1015–1033.
- Christie,K.R., Weng,S., Balakrishnan,R., Costanzo,M.C., Dolinski,K., Dwight,S.S., Engel,S.R., Feierbach,B., Fisk,D.G., Hirschman,J.E. *et al.* (2004) Saccharomyces genome database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.*, **32**, D311–D314.
- Mueller,L.A., Zhang,P. and Rhee,S.Y. (2003) AraCyc: a biochemical pathway database for *Arabidopsis*. *Plant Physiol.*, **132**, 453–460.
- Liang,C., Jaiswal,P., Hebbard,C., Avraham,S., Buckler,E.S., Casstevens,T., Hurwitz,B., McCouch,S., Ni,J., Pujar,A. *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.
- Evsikov,A.V., Dolan,M.E., Genrich,M.P., Patek,E. and Bult,C.J. (2009) MouseCyc: a curated biochemical pathways database for the laboratory mouse. *Genome Biol.*, **10**, R84.
- Seo,S. and Lewin,H.A. (2009) Reconstruction of metabolic pathways for the cattle genome. *BMC Syst. Biol.*, **3**, 33.
- Urbanczyk-Wochniak,E. and Sumner,L.W. (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. *Bioinformatics*, **23**, 1418–1423.
- Zhang,P., Dreher,K., Karthikeyan,A., Chi,A., Pujar,A., Caspi,R., Karp,P., Kirkup,V., Latendresse,M., Lee,C. *et al.* (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. *Plant Physiol.*, **153**, 1479–1491.
- Fey,P., Gaudet,P., Curk,T., Zupan,B., Just,E.M., Basu,S., Merchant,S.N., Bushmanova,Y.A., Shaulsky,G., Kibbe,W.A. *et al.* (2009) dictyBase - a Dictyostelium bioinformatics resource update. *Nucleic Acids Res.*, **37**, D515–D519.
- Doyle,M.A., MacRae,J.I., De Souza,D.P., Saunders,E.C., McConville,M.J. and Likic,V.A. (2009) LeishCyc: a biochemical pathways database for *Leishmania major*. *BMC Syst. Biol.*, **3**, 57.
- May,P., Christian,J.O., Kempa,S. and Walther,D. (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. *BMC Genomics*, **10**, 209.
- Bombarely,A., Menda,N., Tecle,I.Y., Buels,R.M., Strickler,S., Fischer-York,T., Pujar,A., Leto,J., Gosselin,J. and Mueller,L.A. (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. *Nucleic Acids Res.*, **39**, D1149–D1155.
- Snyder,E.E., Kampanya,N., Lu,J., Nordberg,E.K., Karur,H.R., Shukla,M., Soneja,J., Tian,Y., Xue,T., Yoo,H. *et al.* (2007) PATRIC: the VBI pathoSystems resource integration center. *Nucleic Acids Res.*, **35**, D401–D406.
- Benedict,M.N., Gonnerman,M.C., Metcalf,W.W. and Price,N.D. (2012) Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *J. Bacteriol.*, **194**, 855–865.
- Ulas,T., Riemer,S.A., Zaparty,M., Siebers,B. and Schomburg,D. (2012) Genome-scale reconstruction and analysis of the metabolic network in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *PLoS One*, **7**, e43401.
- Chen,J.W., Scaria,J. and Chang,Y.F. (2012) Phenotypic and transcriptomic response of auxotrophic *Mycobacterium avium* subsp. paratuberculosis leuD mutant under environmental stress. *PLoS One*, **7**, e37884.
- Swithers,K.S., Petrus,A.K., Secinaro,M.A., Nesbo,C.L., Gogarten,J.P., Noll,K.M. and Butzin,N.C. (2012) Vitamin B(12) synthesis and salvage pathways were acquired by horizontal gene transfer to the Thermotogales. *Genome Biol. Evol.*, **4**, 730–739.
- Reina-Bueno,M., Argandona,M., Nieto,J.J., Hidalgo-Garcia,A., Iglesias-Guerra,F., Delgado,M.J. and Vargas,C. (2012) Role of trehalose in heat and desiccation tolerance in the soil bacterium *Rhizobium etli*. *BMC Microbiol.*, **12**, 207.
- Lolas,I.B., Chen,X., Bester,K. and Nielsen,J.L. (2012) Identification of triclosan-degrading bacteria using stable isotope probing, fluorescence in situ hybridization and microautoradiography. *Microbiology*, **158**, 2796–2804.
- Park,J.M., Song,H., Lee,H.J. and Seung,D. (2013) Genome-scale reconstruction and in silico analysis of *Klebsiella oxytoca* for 2,3-butanediol production. *Microb. Cell Fact.*, **12**, 20.
- Farrugia,D.N., Elbourne,L.D., Hassan,K.A., Eijkelkamp,B.A., Tetu,S.G., Brown,M.H., Shah,B.S., Peleg,A.Y., Mabbutt,B.C. and Paulsen,I.T. (2013) The complete genome and phenotype of a community-acquired *Acinetobacter baumannii*. *PLoS One*, **8**, e58628.
- Belda,E., Sekowska,A., Le Fevre,F., Morgat,A., Mornico,D., Ouzounis,C., Vallenet,D., Medigue,C. and Danchin,A. (2013) An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology*, **159**, 757–770.

26. Wagner, A., Segler, L., Kleinstuber, S., Sawers, G., Smidt, H. and Lechner, U. (2013) Regulation of reductive dehalogenase gene transcription in *Dehalococcoides mccartyi*. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, **368**, 20120317.
27. Herlemann, D.P., Lundin, D., Labrenz, M., Jurgens, K., Zheng, Z., Aspeborg, H. and Andersson, A.F. (2013) Metagenomic de novo assembly of an aquatic representative of the verrucomicrobial class Spartobacteria. *MBio.*, **4**, e00569–e00512.
28. Pentjuss, A., Odzina, I., Kostromins, A., Fell, D.A., Stalidzans, E. and Kalnenieks, U. (2013) Biotechnological potential of respiring *Zymomonas mobilis*: a stoichiometric analysis of its central metabolism. *J. Biotechnol.*, **165**, 1–10.
29. Belanger, L. and Charles, T.C. (2013) Members of the *Sinorhizobium meliloti* ChvI regulon identified by a DNA binding screen. *BMC Microbiol.*, **13**, 132.
30. Pastor, J.M., Bernal, V., Salvador, M., Argandona, M., Vargas, C., Csonka, L., Sevilla, A., Iborra, J.L., Nieto, J.J. and Canovas, M. (2013) Role of central metabolism in the osmoadaptation of the halophilic bacterium *Chromohalobacter salexigens*. *J. Biol. Chem.*, **288**, 17769–17781.
31. Gangaiah, D., Zhang, X., Fortney, K.R., Baker, B., Liu, Y., Munson, R.S. Jr and Spinola, S.M. (2013) Activation of CpxRA in *Haemophilus ducreyi* primarily inhibits the expression of its targets, including major virulence determinants. *J. Bacteriol.*, **195**, 3486–3502.
32. Ghignone, S., Salvioli, A., Anca, I., Lumini, E., Ortu, G., Petiti, L., Cruveiller, S., Bianciotto, V., Piffanelli, P., Lanfranco, L. et al. (2012) The genome of the obligate endobacterium of an AM fungus reveals an interphylum network of nutritional interactions. *ISME J.*, **6**, 136–145.
33. Kim, S.H., Harzman, C., Davis, J.K., Hutcheson, R., Broderick, J.B., Marsh, T.L. and Tiedje, J.M. (2012) Genome sequence of *Desulfitobacterium hafniense* DCB-2, a gram-positive anaerobe capable of dehalogenation and metal reduction. *BMC Microbiol.*, **12**, 21.
34. Song, Y., Matsumoto, K., Yamada, M., Gohda, A., Brigham, C.J., Sinskey, A.J. and Taguchi, S. (2012) Engineered *Corynebacterium glutamicum* as an endotoxin-free platform strain for lactate-based polyester production. *Appl. Microbiol. Biotechnol.*, **93**, 1917–1925.
35. Liu, X., Gao, B., Novik, V. and Galan, J.E. (2012) Quantitative proteomics of intracellular *Campylobacter jejuni* reveals metabolic reprogramming. *PLoS Pathogens*, **8**, e1002562.
36. Li, J. and Wang, N. (2012) The *gpsX* gene encoding a glycosyltransferase is important for polysaccharide production and required for full virulence in *Xanthomonas citri* subsp. *citri*. *BMC Microbiol.*, **12**, 31.
37. Adler, P., Song, H.S., Kastner, K., Ramkrishna, D. and Kunz, B. (2012) Prediction of dynamic metabolic behavior of *Pediococcus pentosaceus* producing lactic acid from lignocellulosic sugars. *Biotechnol. Prog.*, **28**, 623–635.
38. Vongsangnak, W., Figueiredo, L.F., Forster, J., Weber, T., Thykaer, J., Stegmann, E., Wohleben, W. and Nielsen, J. (2012) Genome-scale metabolic representation of *Amycolatopsis balhimycina*. *Biotechnol. Bioeng.*, **109**, 1798–1807.
39. Hanreich, A., Heyer, R., Benndorf, D., Rapp, E., Pioch, M., Reichl, U. and Klocke, M. (2012) Metaproteome analysis to determine the metabolically active part of a thermophilic microbial community producing biogas from agricultural biomass. *Can. J. Microbiol.*, **58**, 917–922.
40. Zou, W., Liu, L., Zhang, J., Yang, H., Zhou, M., Hua, Q. and Chen, J. (2012) Reconstruction and analysis of a genome-scale metabolic model of the vitamin C producing industrial strain *Ketogulonicigenium vulgare* WSH-001. *J. Biotechnol.*, **161**, 42–48.
41. Nuxoll, A.S., Halouska, S.M., Sadykov, M.R., Hanke, M.L., Bayles, K.W., Kielian, T., Powers, R. and Fey, P.D. (2012) CcpA regulates arginine biosynthesis in *Staphylococcus aureus* through repression of proline catabolism. *PLoS Pathogens*, **8**, e1003033.
42. Sadykov, M.R., Thomas, V.C., Marshall, D.D., Wenstrom, C.J., Moormeier, D.E., Widhelm, T.J., Nuxoll, A.S., Powers, R. and Bayles, K.W. (2013) Inactivation of the Pta-AckA pathway causes cell death in *Staphylococcus aureus*. *J. Bacteriol.*, **195**, 3035–3044.
43. Zou, W., Zhou, M., Liu, L. and Chen, J. (2013) Reconstruction and analysis of the industrial strain *Bacillus megaterium* WSH002 genome-scale in silico metabolic model. *J. Biotechnol.*, **164**, 503–509.
44. Takami, H., Noguchi, H., Takaki, Y., Uchiyama, I., Toyoda, A., Nishi, S., Chee, G.J., Arai, W., Nunoura, T., Itoh, T. et al. (2012) A deeply branching thermophilic bacterium with an ancient acetyl-CoA pathway dominates a subsurface ecosystem. *PLoS One*, **7**, e30559.
45. Khudyakov, J.I., D'Haeseleer, P., Borglin, S.E., Deangelis, K.M., Woo, H., Lindquist, E.A., Hazen, T.C., Simmons, B.A. and Thelen, M.P. (2012) Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, *Enterobacter lignolyticus*. *Proc. Natl Acad. Sci. USA*, **109**, E2173–E2182.
46. Klanchui, A., Khannapho, C., Phodee, A., Cheevadhanarak, S. and Meechai, A. (2012) iAK692: a genome-scale metabolic model of *Spirulina platensis* C1. *BMC Syst. Biol.*, **6**, 71.
47. Aklujkar, M., Haveman, S.A., DiDonato, R. Jr, Chertkov, O., Han, C.S., Land, M.L., Brown, P. and Lovley, D.R. (2012) The genome of *Pelobacter carbinolicus* reveals surprising metabolic capabilities and physiological features. *BMC Genomics*, **13**, 690.
48. Cavalheiro, J.M., de Almeida, M.C., da Fonseca, M.M. and de Carvalho, C.C. (2012) Adaptation of *Cupriavidus necator* to conditions favoring polyhydroxyalkanoate production. *J. Biotechnol.*, **164**, 309–317.
49. Flynn, C.M., Hunt, K.A., Gralnick, J.A. and Srien, F. (2012) Construction and elementary mode analysis of a metabolic model for *Shewanella oneidensis* MR-1. *Biosystems*, **107**, 120–128.
50. Balagurunathan, B., Jonnalagadda, S., Tan, L. and Srinivasan, R. (2012) Reconstruction and analysis of a genome-scale metabolic model for *Scheffersomyces stipitis*. *Microb. Cell Fact.*, **11**, 27.
51. Liu, T., Zou, W., Liu, L. and Chen, J. (2012) A constraint-based model of *Scheffersomyces stipitis* for improved ethanol production. *Biotechnol. Biofuels*, **5**, 72.
52. Xu, N., Liu, L., Zou, W., Liu, J., Hua, Q. and Chen, J. (2013) Reconstruction and analysis of the genome-scale metabolic network of *Candida glabrata*. *Mole. Biosyst.*, **9**, 205–216.
53. Wu, M., Zheng, M., Zhang, W., Suresh, S., Schlecht, U., Fitch, W.L., Aronova, S., Baumann, S., Davis, R., St Onge, R. et al. (2012) Identification of drug targets by chemogenomic and metabolomic profiling in yeast. *Pharmacogenet. Genomics*, **22**, 877–886.
54. Gianoulis, T.A., Griffin, M.A., Spakowicz, D.J., Dunican, B.F., Alpha, C.J., Sboner, A., Sismour, A.M., Kodira, C., Egholm, M., Church, G.M. et al. (2012) Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus *Ascocoryne sarcoides*. *PLoS Genet.*, **8**, e1002558.
55. Fabris, M., Matthijs, M., Rombauts, S., Vyverman, W., Goossens, A. and Baart, G.J. (2012) The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J.*, **70**, 1004–1014.
56. Dharmawardhana, P., Ren, L., Amarasinghe, V., Monaco, M., Thomason, J., Ravenscroft, D., McCouch, S., Ware, D. and Jaiswal, P. (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. *Rice*, **6**.
57. Giorgi, F.M., Del Fabbro, C. and Licausi, F. (2013) Comparative study of RNA-seq- and microarray-derived coexpression networks in *Arabidopsis thaliana*. *Bioinformatics*, **29**, 717–724.
58. Van Moerkercke, A., Fabris, M., Pollier, J., Baart, G.J., Rombauts, S., Hasnain, G., Rischer, H., Memelink, J., Oksman-Caldentey, K.M. and Goossens, A. (2013) CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. *Plant Cell Physiol.*, **54**, 673–685.
59. Fatima, T., Snyder, C.L., Schroeder, W.R., Cram, D., Datla, R., Wishart, D., Weslake, R.J. and Krishna, P. (2012) Fatty acid composition of developing sea buckthorn (*Hippophae rhamnoides* L.) berry and the transcriptome of the mature seed. *PLoS One*, **7**, e34099.
60. Desjardins, C.A., Cerqueira, G.C., Goldberg, J.M., Dunning Hotopp, J.C., Haas, B.J., Zucker, J., Ribeiro, J.M., Saif, S., Levin, J.Z., Fan, L. et al. (2013) Genomics of *Loa loa*, a Wolbachia-free filarial parasite of humans. *Nat. Genet.*, **45**, 495–500.
61. Creek, D.J., Anderson, J., McConville, M.J. and Barrett, M.P. (2012) Metabolomic analysis of trypanosomatid protozoa. *Mol. Biochem. Parasitol.*, **181**, 73–84.

62. Abubucker,S., Segata,N., Goll,J., Schubert,A.M., Izard,J., Cantarel,B.L., Rodriguez-Mueller,B., Zucker,J., Thiagarajan,M., Henrissat,B. *et al.* (2012) Metabolic reconstruction for metagenomic data and its application to the human microbiome. *PLoS Comput. Biol.*, **8**, e1002358.
63. Borenstein,E. (2012) Computational systems biology and in silico modeling of the human microbiome. *Brief Bioinform.*, **13**, 769–780.
64. Morgan,X.C. and Huttenhower,C. (2012) Chapter 12: Human microbiome analysis. *PLoS Comput. Biol.*, **8**, e1002808.
65. Gevers,D., Pop,M., Schloss,P.D. and Huttenhower,C. (2012) Bioinformatics for the human microbiome project. *PLoS Comput. Biol.*, **8**, e1002779.
66. Jacobsen,U.P., Nielsen,H.B., Hildebrand,F., Raes,J., Sicheritz-Ponten,T., Kouskoumvekaki,I. and Panagiotou,G. (2013) The chemical interactome space between the human host and the genetically defined gut metatypes. *ISME J.*, **7**, 730–742.
67. Hildebrand,F., Ebersbach,T., Nielsen,H.B., Li,X., Sonne,S.B., Bertalan,M., Dimitrov,P., Madsen,L., Qin,J., Wang,J. *et al.* (2012) A comparative analysis of the intestinal metagenomes present in guinea pigs (*Cavia porcellus*) and humans (*Homo sapiens*). *BMC Genomics*, **13**, 514.
68. Sharon,I., Bercovici,S., Pinter,R.Y. and Shlomi,T. (2011) Pathway-based functional analysis of metagenomes. *J. Comput. Biol.*, **18**, 495–505.
69. Altman,T., Travers,M., Kothari,A., Caspi,R. and Karp,P.D. (2013) A systematic comparison of the MetaCyc and KEGG pathway databases. *BMC Bioinformatics*, **14**, 112.
70. Latendresse,M., Malerich,J.P., Travers,M. and Karp,P.D. (2012) Accurate atom-mapping computation for biochemical reactions. *J. Chem. Inf. Model.*, **52**, 2970–2982.
71. Ceroni,A., Dell,A. and Haslam,S.M. (2007) The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol. Med.*, **2**, 3.
72. Damerell,D., Ceroni,A., Maass,K., Ranzinger,R., Dell,A. and Haslam,S.M. (2012) The GlycanBuilder and GlycoWorkbench glycoinformatics tools: updates and new developments. *Biol. Chem.*, **393**, 1357–1362.
73. McDonald,A.G., Boyce,S. and Tipton,K.F. (2009) ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res.*, **37**, D593–D597.
74. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the national center for biotechnology information. *Nucleic Acids Res.*, **37**, D5–D15.
75. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
76. BioCyc online guided tour. <http://biocyc.org/samples.shtml> (29 October 2013, date last accessed). SRI International.
77. Guide to the BioCyc database collection. <http://biocyc.org/BioCycUserGuide.shtml> (29 October 2013, date last accessed). SRI International.
78. Guide to the MetaCyc database. <http://www.metacyc.org/MetaCycUserGuide.shtml> (29 October 2013, date last accessed). SRI International.
79. Guide to the EcoCyc Database. <http://biocyc.org/ecocyc/EcoCycUserGuide.shtml> (29 October 2013, date last accessed). SRI International.
80. Pathway/Genome Database Concepts Guide. <http://biocyc.org/PGDBConceptsGuide.shtml> (29 October 2013, date last accessed). SRI International.
81. BioCyc webinars. <http://biocyc.org/webinar.shtml> (29 October 2013, date last accessed). SRI International.
82. BioCyc publication list. <http://biocyc.org/publications.shtml> (29 October 2013, date last accessed). SRI International.