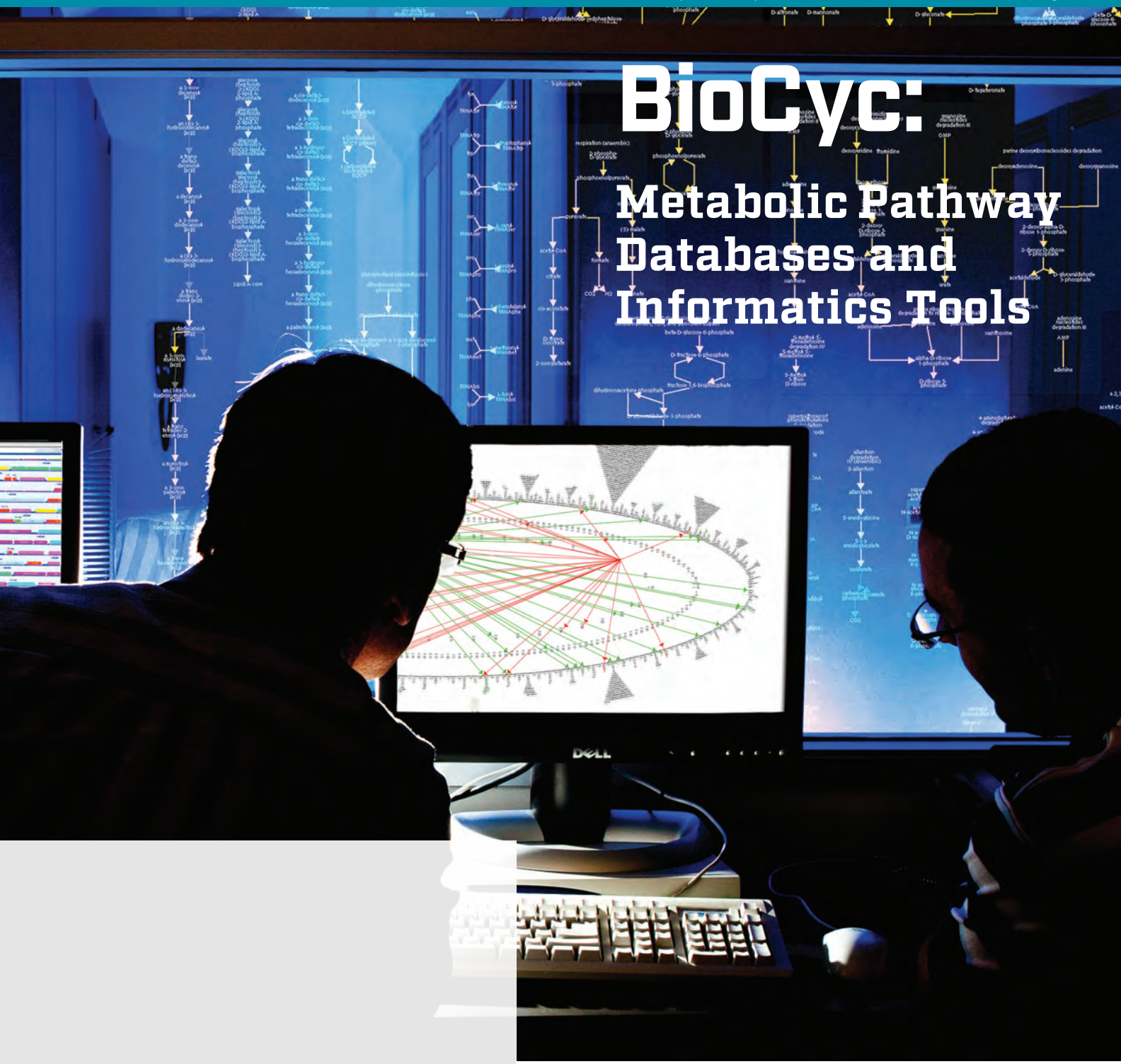
SIMB News

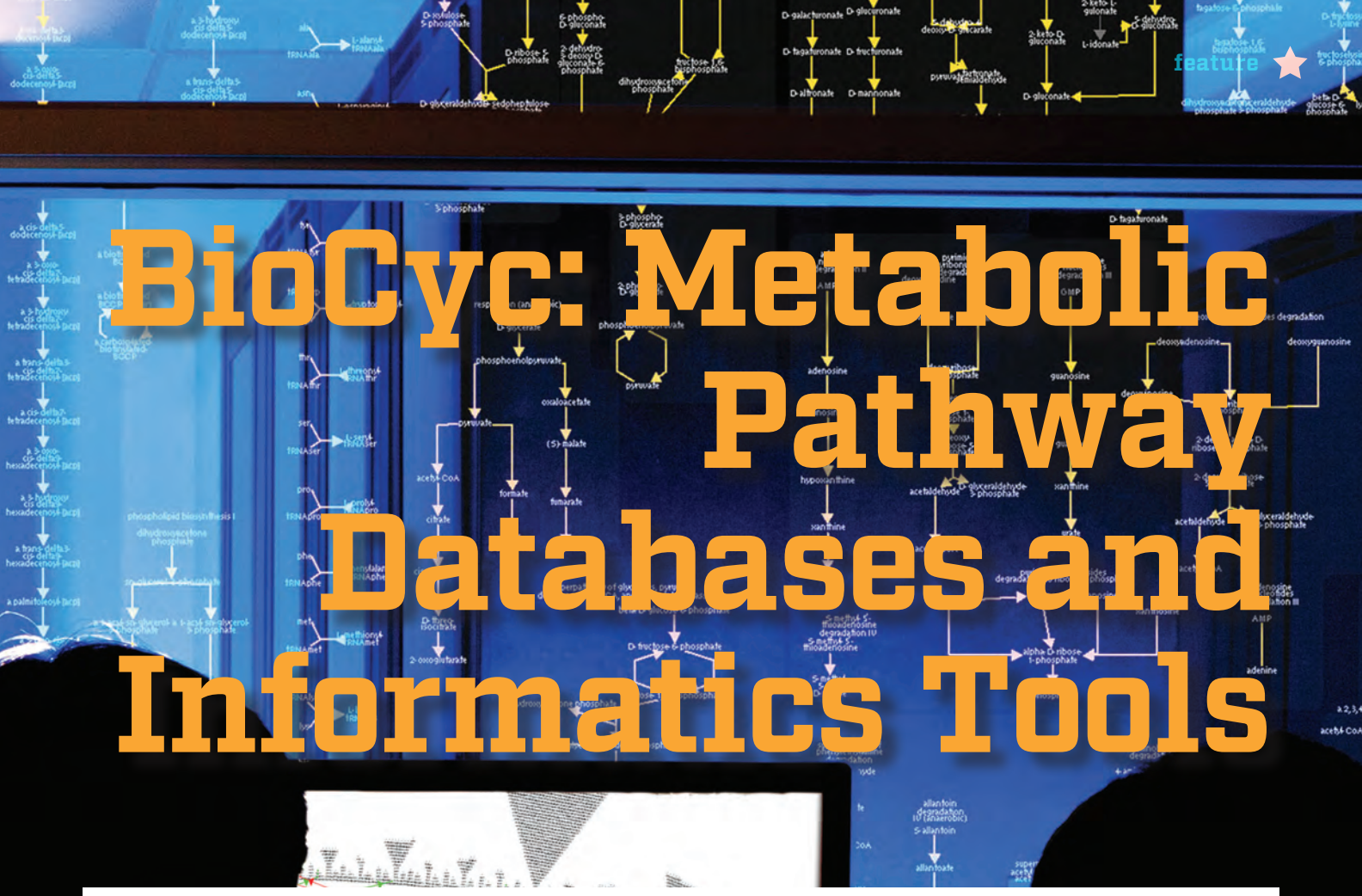
News Magazine of the Society for Industrial Microbiology and Biotechnology
January/February/March 2023 v.73 n.1 • *www.simbhq.org*



BioCyc:
Metabolic Pathway Databases and Informatics Tools

Peter D. Karp and Ron Caspi, SRI International

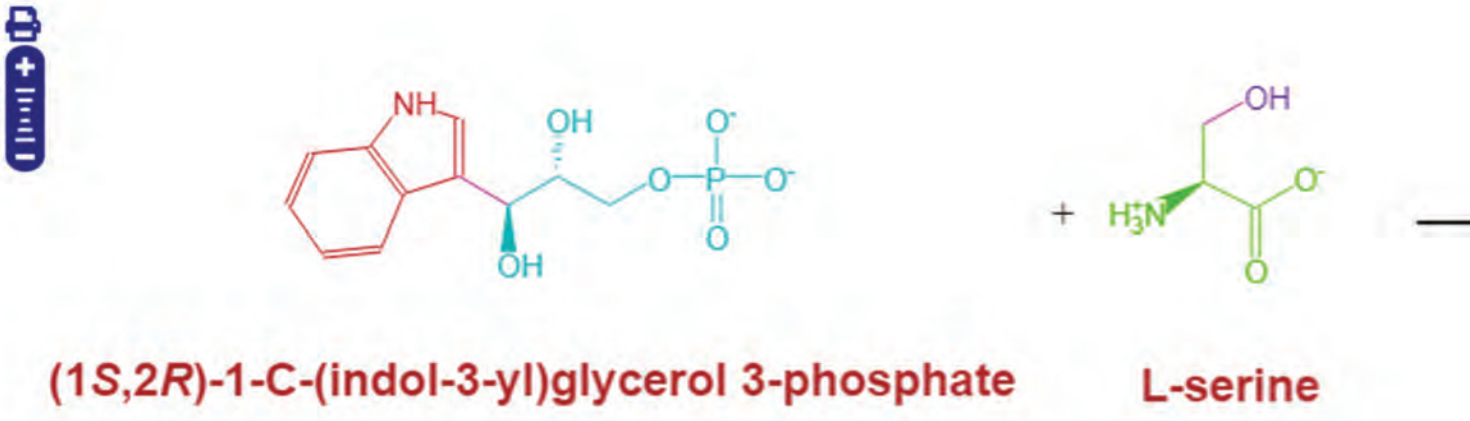
BioCyc: Metabolic Pathway Databases and Informatics Tools



Abstract

This article describes a coordinated set of bioinformatics databases and software tools designed to solve multiple problems faced by metabolic engineers and microbiologists related to metabolic pathways. Those problems include the following: (1) Answering basic questions about the metabolism of a given organism; that is, what pathways does a given bacterium possess and what enzymes and metabolites participate in a given pathway? (2) Predicting the metabolism of an organism from its genome sequence. (3) Engineering new pathways into an organism. (4) Predicting pathway activation levels from omics datasets. (5) Comparative analyses of metabolism. BioCyc.org provides 20,000 Pathway/Genome databases for sequenced microbes that describe their reconstructed metabolic networks. The MetaCyc DB provides a universal encyclopedia of metabolism across all domains of life. BioCyc computational tools provide search, comparison, and multiple analysis operations, including omics data analysis.

Image generated by SRI International Graphics Department



Introduction

The immense diversity of microbial metabolism presents microbiologists with great challenges: How do we capture and interrogate the huge mass of information about microbial metabolism? For any given microbe, how do we navigate and analyze its metabolic network to gain an in-depth understanding of its capabilities? And how do we compare the metabolism of multiple microbes to understand the differences in their networks? BioCyc, a resource that has benefited from more than 20 years of research and development, is designed to address these challenges (Karp et al., 2019).

BioCyc.org provides a collection of 20,000 microbial Pathway/Genome Databases (PGDBs), each of which describing the genome and metabolic network of a single sequenced organism. Each PGDB is based on a computational reconstruction of the organism's metabolism computed from its sequenced genome (Karp et al., 2011). For some organisms, that reconstruction is supplemented with manually curated information. The metabolic information present in each database includes the reactions, metabolites, enzymes, and pathways that make up each metabolic network.

BioCyc.org also provides dozens of computational tools for searching and analyzing genome and metabolic information. Search tools enable users to find a given metabolite, reaction, enzyme, or pathway. Visualization tools enable users to inspect search results, including information pages for metabolites, reactions, and pathways. Visualization tools also enable navigation through full metabolic networks. Analysis tools include RouteSearch, which searches for novel metabolic routes from a feedstock to a target compound, and tools for visualizing and analyzing transcriptomics and metabolomics data on individual pathway diagrams and on full metabolic map diagrams. A suite of comparison tools is available to differentiate the metabolic networks of two or more organisms.

The computational tools present in the BioCyc.org website are provided by the Pathway Tools software. A local Pathway Tools installation can serve as a full data management environment for genome and metabolism information for thousands of organisms. For example, Pathway Tools includes a genome browser, BLAST searches, sequence pattern searches, and a multiple sequence alignment tool. It also includes interactive editing tools for modifying a PGDB, such as adding a new metabolite, changing the function of a gene, adding a transcriptional regulatory interaction, and modifying a metabolic pathway.

A local Pathway Tools installation supports the analysis of genomes not present in the BioCyc website by enabling the generation of new PGDBs for any annotated genomes, including proprietary genomes or a multi-organism metagenome. In addition, a local Pathway Tools

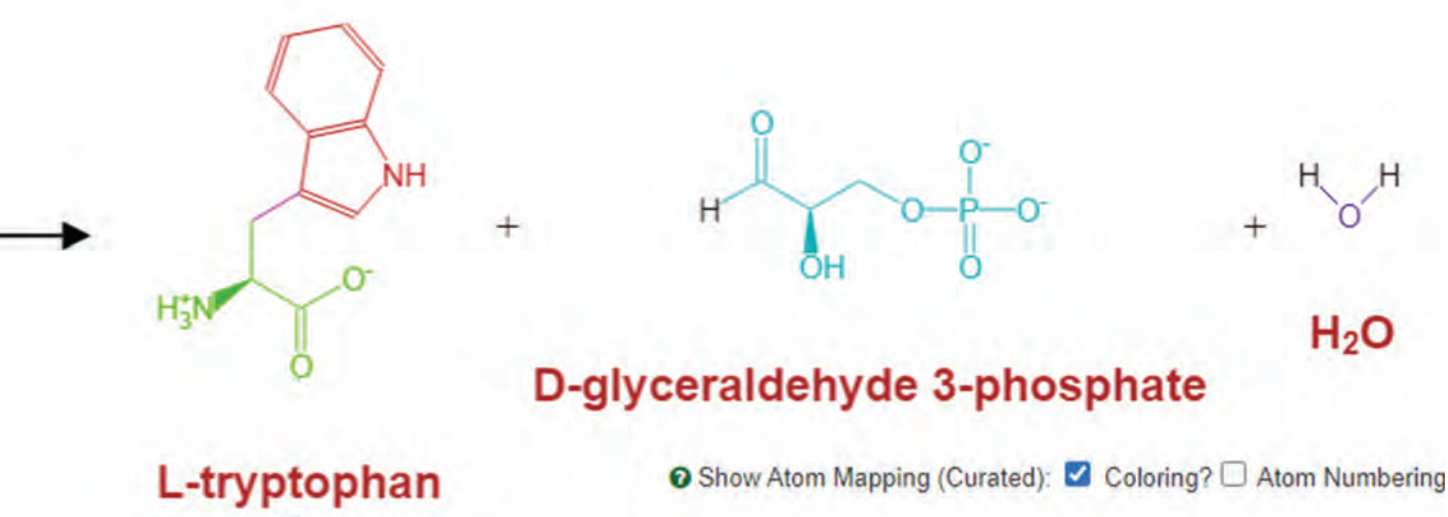


Figure 1

installation enables the RouteSearch tool to add new reactions from MetaCyc when generating routes.

BioCyc Databases

The MetaCyc Database

The MetaCyc DB (MetaCyc.org) is a key reference database for metabolism that contains experimentally studied metabolic pathways, reactions, enzymes, and metabolites from all domains of life. It has received extensive curation from 74,000 publications to ensure the accuracy and completeness of its information (Caspi et al., 2020).

In addition to serving scientists as an online encyclopedic reference on metabolism, MetaCyc plays a central role in several BioCyc computational tools. MetaCyc provides the reference metabolic information for BioCyc tools for computing metabolic reconstructions; those tools copy relevant metabolites, reactions, and pathways from MetaCyc into a new database containing the reconstruction. MetaCyc is also the source of exogenous reactions for the RouteSearch tool that designs novel pathways from a specified start-to-end metabolite. In addition, our gap-filling tool draws from MetaCyc reactions to fill gaps in metabolic models.

MetaCyc metabolite pages, such as that for L-tryptophan, include a list of chemical names and synonyms, molecular weight and monoisotopic mass, chemical formula and structure, unique identifiers (e.g., InChI and SMILES), Gibbs free energy of formation, and links to many other chemical databases.

The Regulation tab lists the regulatory influences of the compound on enzymes within MetaCyc.

The Reactions tab lists the reactions known to produce, consume, and transport the compound across all domains of life, with associated pathway(s) indicated when relevant. Clicking on a reaction opens a MetaCyc reaction page, such as that for the L-tryptophan aminotransferase reaction. The reaction page shows the reaction equation with the chemical structures automatically colored to show substructures that are conserved between the reactants and the products and the bonds that are made and broken by this reaction, based on the atom-mapping information that we compute for most MetaCyc reactions (Figure 1). For most reactions, the reaction page lists one or more enzymes that catalyze them, the organisms from which those enzymes were isolated, and the metabolic pathways in which the reaction participates.

Clicking on a pathway name opens a MetaCyc pathway page (Figure 2). Different organisms have evolved alternative mechanisms for accomplishing a given biochemical transformation (e.g., degradation of L-tryptophan). We call these alternative mechanisms pathway variants, and append roman numerals to the pathway name to designate variants, such as L-tryptophan degradation VIII (to tryptophol). Pathway drawings are generated automatically by the BioCyc software, and the user can control which aspects of a pathway are included in the diagrams, such as metabolite chemical structures or allosteric regulation of pathway's enzymes. Pathway pages also list some specific organisms that are known to possess the pathway and list the taxonomic groups in which the pathway is expected to be found. Pathway pages contain a mini-review of the literature describing that pathway. Mini-reviews are found in other pages as well; those for pathway and enzyme pages tend to be the most extensive (Caspi et al., 2013).

Clicking on an enzyme name opens a MetaCyc gene/protein page, such as that for the TrpE subunit of anthranilate synthase of *E. coli*. Enzyme pages collect information regarding the reaction(s) catalyzed by the enzyme and the pathways in which they participate. Enzyme pages also list the cellular location(s), the activators, inhibitors, and cofactors affecting the enzyme, and, when relevant, the multimeric complex in which the protein is a subunit of. When available, the page lists kinetic parameters for the enzyme as well as its temperature and pH optima. Some proteins are annotated by GO terms and protein features (known regions and residues of interest within the protein, such as metal ion binding sites, active sites, and phosphorylation sites), which can be accessed via the Go Terms and Protein Features tabs, respectively.

The overall data content of MetaCyc is summarized and compared with the content of the KEGG pathway database in Table 1. MetaCyc contains more data (7.3 times as many pathways and 1.6 times as many reactions) and more in-depth summaries of the literature in its mini-reviews.

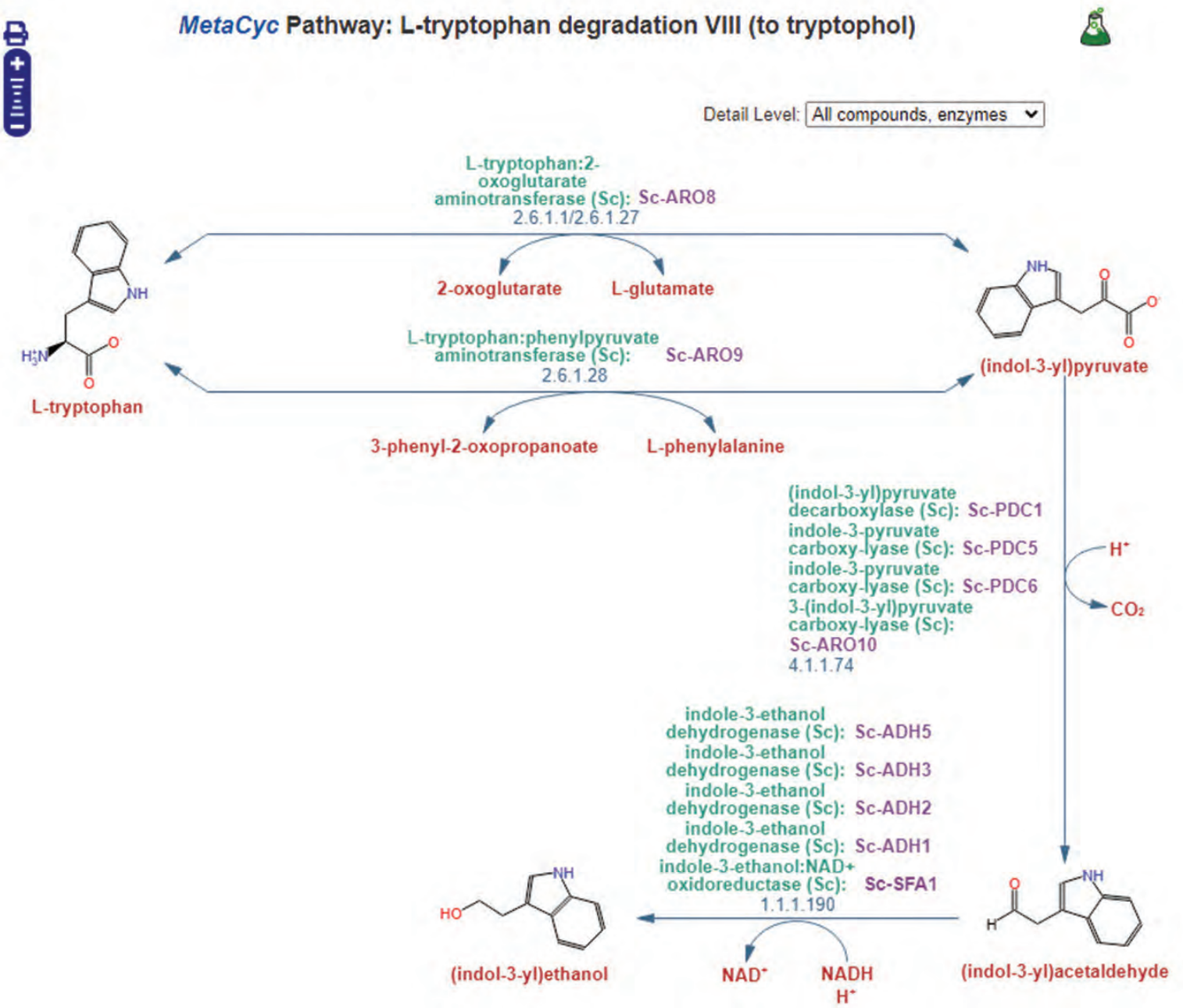
To clarify the relationship between MetaCyc and BioCyc, one major aim of MetaCyc is to integrate experimentally derived information on metabolism from all domains of life to provide a solid reference source for performing metabolic reconstructions and developing metabolic models. Thus, the collection of information for any one organism in MetaCyc will necessarily be limited to information from experimental studies. For example, MetaCyc contains 22 pathways recorded as being

	MetaCyc	KEGG
Genomes	20,028 (BioCyc)	8,611
Pathways	3,085	425
Reactions	18,391	11,860
Mini-Reviews (textbook pages)	10,392	1,557

Table 1

A comparison of MetaCyc version 26.5 (December 2022) with KEGG version 104.0+/0415 (December 2022). The first line compares the number of genomes in BioCyc versus those in the KEGG Genome database (not including viral genomes). Subsequent lines compare the number of pathways and reactions in MetaCyc with the number in KEGG (obtained from the KEGG PATHWAY and KEGG REACTION databases, respectively). To calculate the number of textbook pages in commentary, we collected the total lines of commentary in MetaCyc and in KEGG reference databases and converted them to the number of textbook pages by using the formula 3050 words = 1 page.

Figure 2


MetaCyc Pathway: L-tryptophan degradation VIII (to tryptophol)
Detail Level: All compounds, enzymes
L-tryptophan:2-oxoglutarate aminotransferase (Sc): Sc-ARO8
2.6.1.1/2.6.1.27
2-oxoglutarate
L-glutamate
L-tryptophan
L-tryptophan:phenylpyruvate aminotransferase (Sc): Sc-ARO9
2.6.1.28
3-phenyl-2-oxopropanoate
L-phenylalanine
(indol-3-yl)pyruvate
(indol-3-yl)pyruvate decarboxylase (Sc): Sc-PDC1
indole-3-pyruvate carboxy-lyase (Sc): Sc-PDC5
indole-3-pyruvate carboxy-lyase (Sc): Sc-PDC6
3-(indol-3-yl)pyruvate carboxy-lyase (Sc): Sc-ARO10
4.1.1.74
H+
CO2
indole-3-ethanol dehydrogenase (Sc): Sc-ADH5
indole-3-ethanol dehydrogenase (Sc): Sc-ADH3
indole-3-ethanol dehydrogenase (Sc): Sc-ADH2
indole-3-ethanol dehydrogenase (Sc): Sc-ADH1
indole-3-ethanol:NAD+ oxidoreductase (Sc): Sc-SFA1
1.1.1.190
(indol-3-yl)ethanol
NAD+
NADH
H+
(indol-3-yl)acetaldehyde

experimentally studied in *Pseudomonas putida* KT2440. In contrast, the BioCyc DB for *Pseudomonas putida* KT2440 contains 334 metabolic pathways (including those present in MetaCyc), most of which were predicted during the computational metabolic reconstruction for this strain.

BioCyc Organism-Specific Databases

BioCyc organism-specific databases contain the full genome of the organism; database objects describing every gene and protein in the organism; and a mixture of computationally predicted information about the metabolic reactions, pathways, and metabolites of this organism. These databases are created through a series of operations that include computational inferences, import of data from other bioinformatics databases, and, for selected organisms, manual curation.

Computational inferences

» Metabolic reconstruction (computational prediction of metabolic pathways and reactions)

» Prediction of pathway holes (genes likely to code for enzymes catalyzing pathway reactions with no assigned enzyme)

» Prediction of transport reactions

» Prediction of operons

» Prediction of protein complexes

» Determination of orthologs to other BioCyc genomes

» Data import

» Creation of database links to UniProt

» Import of Gene Ontology terms and protein features from UniProt

» Import of protein localization data from PSORTdb

» Import of gene essentiality data from OGEE

Manual curation

» Curation of gene functions from the literature, including enzymatic activities and regulatory functions

» Curation of protein complexes

» Deletion of incorrectly-predicted enzymatic activities and pathways

» Curation of missing pathways

Some of the BioCyc databases describe organisms that are used extensively in the fields of synthetic biology and metabolic engineering and may be of particular interest to researchers in these fields. The following databases received manual curation.

Escherichia coli K-12 MG 1655 (EcoCyc). K-12 strains of *E. coli* serve as the most common hosts for recombinant DNA technology. The EcoCyc DB is arguably the most comprehensive organism knowledgebase available. It has been manually curated by teams from multiple institutions for over 25 years from 42,000 publications. Its contents span protein function, metabolic pathways, transport, and regulation (Keseler et al., 2021).

Bacillus subtilis 168. Following *E. coli*, *B. subtilis* is the second most commonly used bacterium in synthetic biology. In addition to being the best-characterized Gram-positive bacterium, the organism can secrete proteins in the gram per liter range and does not produce any toxic by-products (van Dijl & Hecker, 2013).

Pseudomonas putida KT2440. *P. putida* is a well-established host for cloning and gene expression, displays solvent tolerance, has a remarkable capability to degrade aromatic compounds, and has a high tolerance to oxidative stress. It is considered a model organism for biodegradation studies and as production platform and is currently used for bioremediation, small molecule production, and bioplastics production (Nikel et al., 2016).

Saccharomyces cerevisiae S288c (YeastCyc). *S. cerevisiae* is used in brewing, baking, and as the main source of nutritional yeast. The yeast genome, which was the first eukaryotic genome to be completely sequenced, is highly accessible to manipulation, making it an excellent model for genome engineering and a preferred host for expression of eukaryotic proteins. The *S. cerevisiae* database was generated in 2002 by the *Saccharomyces* Genome Database (SGD) team and benefited from

significant curation by SGD curators. In 2012, the database was transferred to SRI International, where it is curated by members of the SRI Bioinformatics Research Group (Caspi et al., 2014).

Corynebacterium glutamicum ATCC 13032. *C. glutamicum* has been traditionally used industrially for large-scale production of amino acids. This "work horse" of the amino acid fermentation industry is nonpathogenic, does not produce any toxins, and has been used for over 70 years for multi-million-ton scale production of glutamate and lysine. It is now commonly engineered to allow the production of a wide range of industrially relevant compounds using a variety of carbon sources (Becker et al., 2018).

Lactobacillus plantarum WCSF1. *L. plantarum* is one of the most studied species that are extensively used in the food industry. The bacterium is widely used in the manufacture of dairy products, fermented foods, and bacteriocins (Behera et al., 2018). It is also considered a probiotic and serves as a source of commercially important enzymes such as esterases and lipases (Andersen et al., 1995; Kim et al., 2017; Uppada et al., 2017).

Lactobacillus rhamnosus GG. *L. rhamnosus* has been used for various health effects including the prevention and treatment of gastro-intestinal infections and diarrhea and stimulation of immune responses and is currently one of the most widely studied probiotic strains. It has antibacterial activity and anti-inflammatory effects on its host. It also produces and secretes proteins that reduce the inflammatory state and apoptosis of intestinal epithelial cells (Claes et al., 2012; Yan et al., 2007).

The following databases received no manual curation:

Clostridium acetobutylicum ATCC 824. Solventogenic clostridia are attractive hosts for anaerobic biosyntheses because they produce a broad spectrum of chemicals that can be used as precursors to or directly as biofuels and industrial chemicals (Tracy et al., 2012). *C. acetobutylicum* natively produces acetone, butanol, and ethanol and has been used for production of chemicals for the biofuels, flavoring, cosmetics, and plasticizers fields. However, clostridia-based bioproduction remains economically unfavorable on an industrial-scale because of the inefficient fermentation process, leading to intensive efforts to modify their metabolic network to remove bottlenecks (Cheng et al., 2019; Liao et al., 2018; Yang et al., 2016).

Shewanella oneidensis MR-1: Microbial fuel cells convert organic compounds to electricity (Logan & Rabaey, 2012). The process requires that the microorganisms transfer electrons to electrodes. This is enabled by microbial nanowires, electrically conductive filaments that facilitate long-range extracellular electron transfer. Bacterial nanowires are involved in several additional processes such as electromethanogenesis (Kato et al., 2012) and microbial electrosynthesis (Rabaey & Rozendal, 2010). S. oneidensis produces nanowires that conduct current by electron hopping between cytochromes surrounding a filament of an unspecified composition (El-Naggar et al., 2010; Pirbadian & El-Naggar, 2012).

Geobacter sulfurreducens PCA: Like *S. oneidensis*, *G. sulfurreducens* produces nanowires, but these nanowires are completely different, comprising pili that have metal-like conductivity attributed to overlapping pi–pi orbitals of aromatic amino acids (Malvankar & Lovley, 2014; Reguera et al., 2005).

Bacteroides thetaiotaomicron: This organism is a prevalent and stable resident of the human gut. It possesses an extensive collection of saccharolytic enzymes and serves as a primary fermenter of host-, diet-, or microbially derived polysaccharides. Genetic modification of this bacterium to sense and respond to stimuli in the gut has been suggested as providing a foundation for microbiome engineering (Mimee et al., 2015).

Deinococcus radiodurans R1: *Deinococcus* spp. are among the most radiation-resistant microorganisms, showing a remarkable resistance towards ionizing radiation, desiccation, UV radiation, and oxidizing agents. Several studies have used D. radiodurans for small molecule production or bioremediation of toxic compounds under stress conditions, such as radioactive environments (Brim et al., 2000; Gerber et al., 2015; Lange et al., 1998).

Synechocystis sp. PCC 6803: Photosynthetic cyanobacteria attract significant attention as a promising alternative to traditional hosts due to their ability to use solar irradiation

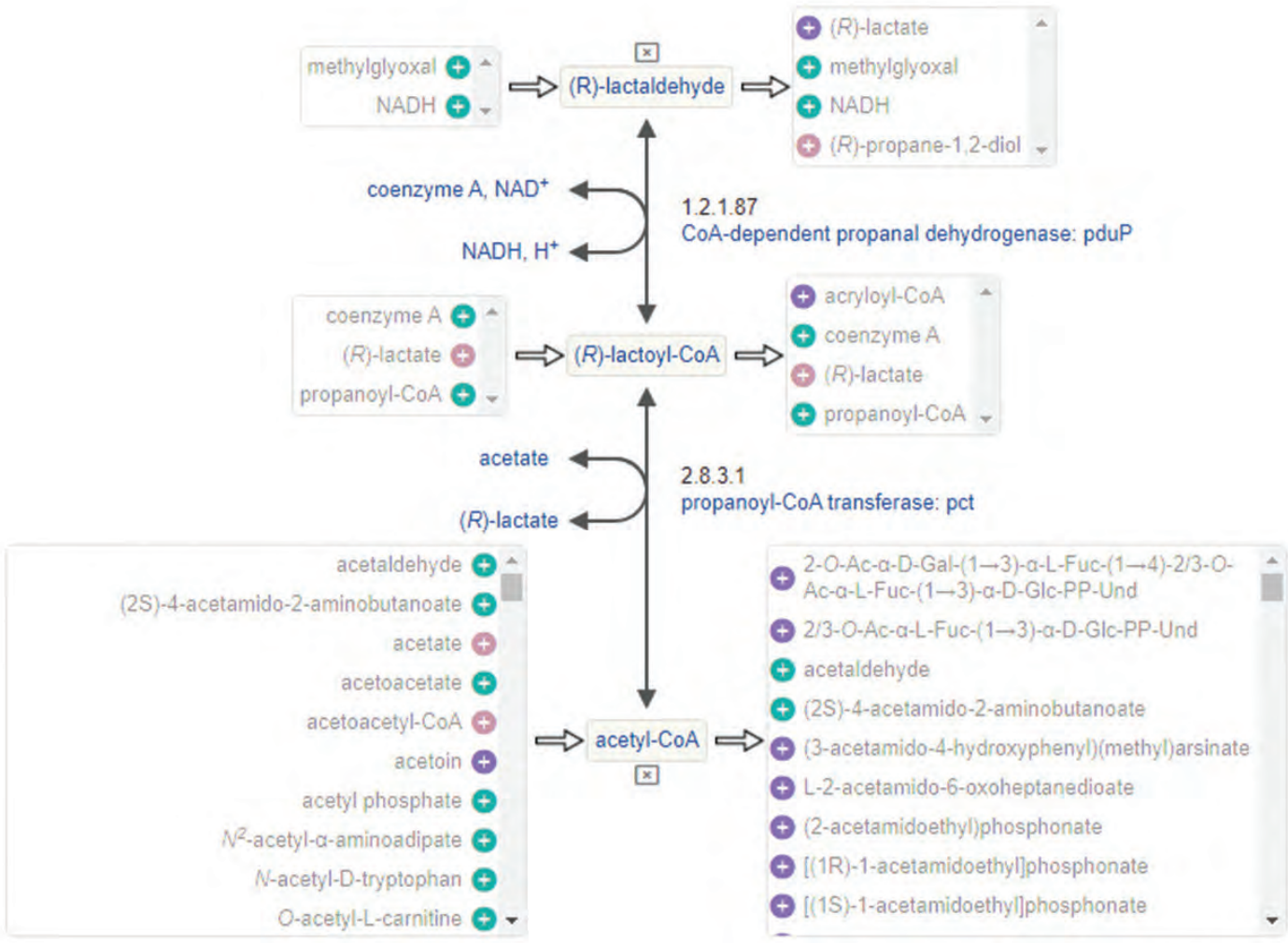


Figure 3

and CO2 as their sole energy and carbon sources, respectively. Cyanobacteria have been successfully engineered to produce more than 20 fuels and small molecules directly from CO2. Synechocystis sp. PCC 6803 is one of the cyanobacterial species with the most developed genetic tools and is a favorite cyanobacterium for genetic engineering (Wang et al., 2013; Wang et al., 2016; Yu et al., 2013).

Klebsiella pneumoniae: This organism natively produces large amounts of 2,3-butanediol (2,3-BD), a compound that has many industrial applications. The organism is easy to cultivate, grows rapidly in a simple medium, and can metabolize all the major sugars in hemicellulose and cellulose hydrolysates into 2,3-BD (Ji et al., 2011). Despite a certain risk of opportunistic infection by the organism, it has been used extensively for the production of 2,3-BD from many substrates, including wood hemicellulose (Yu & Saddler, 1982; Yu et al., 1985); sugar cane juice (Berbert-Molina et al., 2001); starch (Zheng et al., 2008); Jerusalem artichoke tubers (Sun et al., 2009); and corncob molasses (Wang et al., 2010).

BioCyc Software Tools

Search Tools and Information Seeking

One major task for industrial users of BioCyc is finding information about the genome or metabolism of their organism of interest or related organisms. Multiple search tools are available for finding information in BioCyc databases. The quick search box near the top of most BioCyc pages enables name-based searches across the

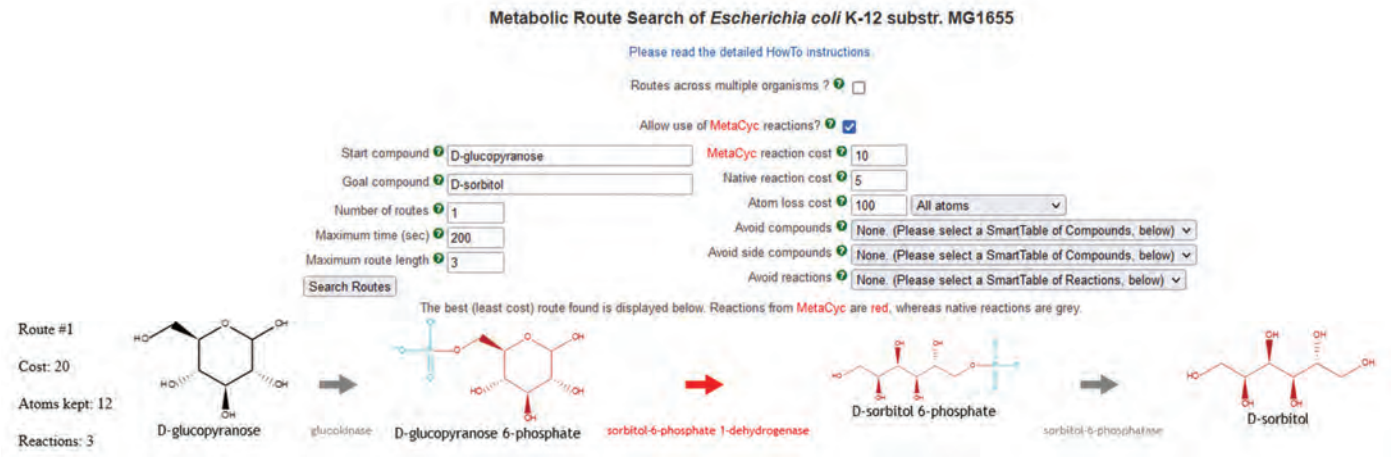


Figure 4

currently selected database, whether MetaCyc or a BioCyc organism-specific database. Searching for a term such as "pyruvate" or "utilization" produces a list of entities within that database whose name contains the search term, sorted by the type of entity. For example, results from the pyruvate search would include pathways such as pyruvate fermentation to acetate, enzymes such as pyruvate carboxylase, metabolites such as 3-hydroxypyruvate, and reactions involving a metabolite with pyruvate in its name.

Additional search tools are available under the Tools > Search menu. For example, the command Tools > Search > Search Genes, Proteins, or RNAs would enable searching a BioCyc database for enzymes with any combination of the following specified properties: molecular weight range, pI range, number of subunits, protein features, small molecule regulator, cofactor, substrate, or ligand.

The full metabolic network browser generates an organism-specific metabolic chart from an organism PGDB. This diagram can be interactively zoomed in real time in a web browser. As the diagram enlarges, it progressively depicts metabolites, enzyme names, and gene names. The user can search the diagram for the names of metabolites, enzymes, genes, and pathways. The diagram can also be overlayed with omics data and with predicted reaction fluxes, and it can be used for comparing two or more metabolic networks.

Metabolic Engineering Tools

Metabolic Network Explorer provides interactive exploration of the metabolic network of an organism starting with a compound of interest (see **Figure 3**). Once the user selects the initial compound, the tool shows all the reactions in the organism's metabolic network that either produce or consume that compound. Selecting one of those reactions generates a short pathway consisting of that reaction. The user can add reactions either leading to the first compound of the pathway or continuing from the pathway's last compound. Adding more reactions in this fashion builds a novel pathway based on enzymes available in the organism.

When executing this tool in the MetaCyc DB, the user can design pathways that consist of enzymes from multiple organisms. In that scenario, the tool may alert the designer to the presence of metabolic transformations that they are not familiar with and are otherwise difficult to find (Paley & Karp, 2021).

Metabolic Route Search (MRS, see **Figure 4**) aims to find possible metabolic routes connecting a specified starting molecule to a specified product metabolite. The version of the tool that is available at BioCyc.org can search a single organism-specific database or a user-defined combination of organism-specific databases. The version of the tool available by installing a local copy of the Pathway Tools software also enables reactions to be drawn from the broad collection of metabolic reactions present in

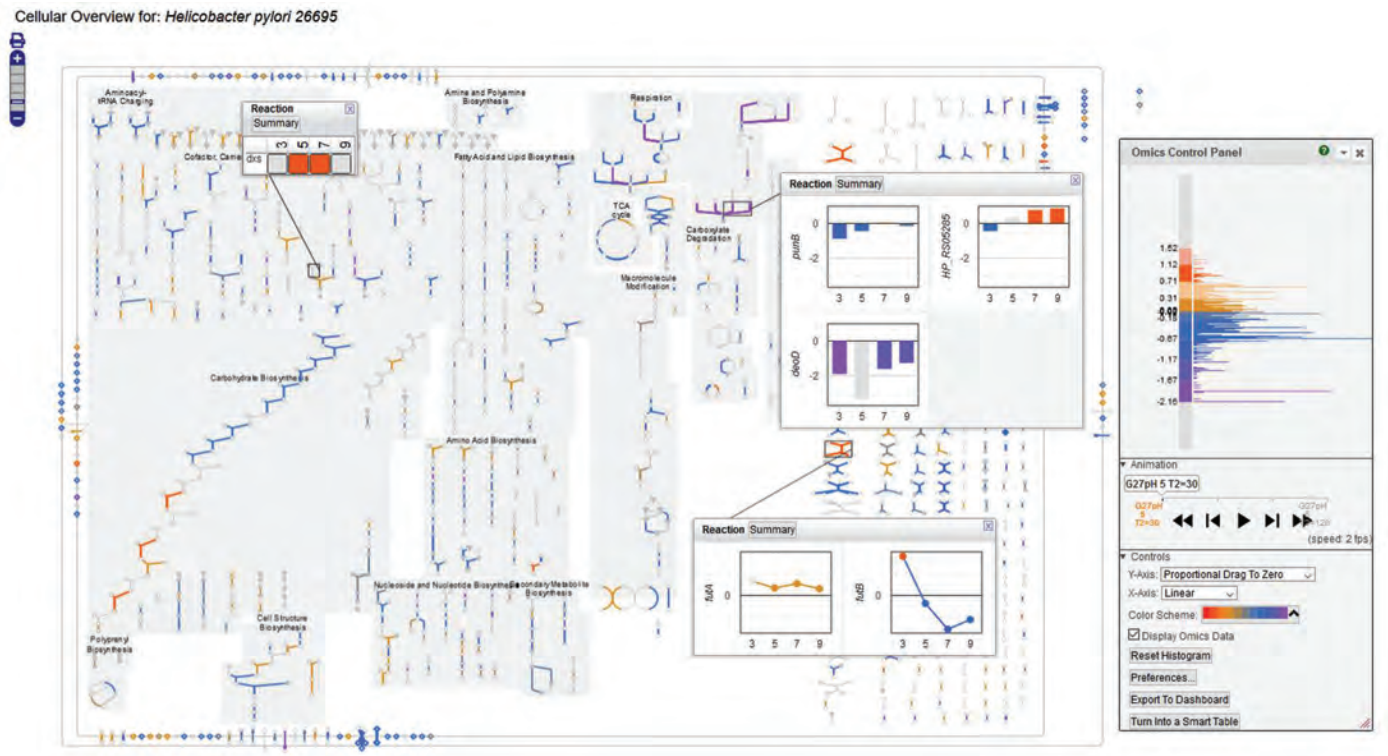


Figure 5

MetaCyc, which covers enzymes that are characterized from thousands of organisms (Krummenacker et al., 2019; Latendresse et al., 2014).

In addition to specifying the feedstock and goal compounds, the user can specify the number of routes to compute and the maximum route length (number of reactions). The software searches for minimal cost routes where the cost of a route depends on the included number of reactions (reactions imported from MetaCyc can be assigned a higher cost), and the number of feedstock atoms lost along the route (which are computed from MetaCyc atom mappings). The user can also specify compounds or reactions to avoid.

The RouteSearch tool can search only fully balanced reactions in which all the components have complete chemical structures. As a result, polymeric reactions are currently not included in the tool's results.

Metabolic Modeling: When installed locally, the Pathway Tools software enables the user to generate quantitative metabolic flux models. The modeling toolkit includes functions such as flux balance analysis, flux variability analysis, and reaction gap filling.

Omics Data Analysis

In many industrial applications, scientists generate transcriptomics and metabolomics data to better understand the response of their organisms to conditions of interest. BioCyc contains multiple tools that facilitate analysis of these data.

» Transcriptomics data analysis tools include the following capabilities:

» Visually overlay transcriptomics data onto individual pathway diagrams

» Visually overlay transcriptomics data onto multi-pathway diagrams

» Visually overlay transcriptomics data onto full metabolic network diagrams (Figure 5)

» Sort pathways according to a pathway activation score computed from the transcriptomics data

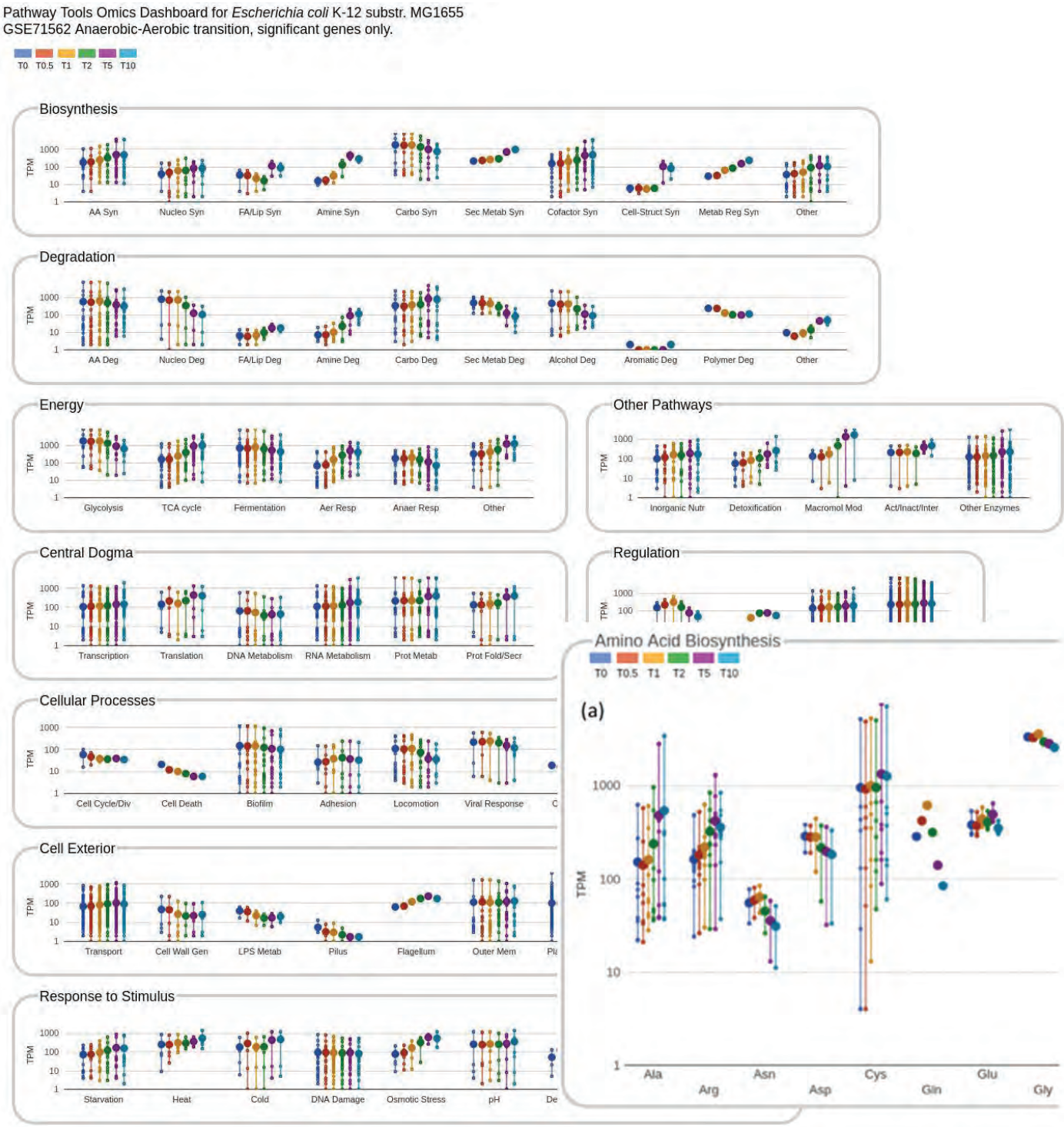

Pathway Tools Omics Dashboard for Escherichia coli K-12 substr. MG1655
GSE71562 Anaerobic-Aerobic transition, significant genes only.
T0 T0.5 T1 T2 T5 T10
Biosynthesis
AA Syn
Nucleo Syn
FA/Lip Syn
Amine Syn
Carbo Syn
Sec Metab Syn
Cofactor Syn
Cell-Struct Syn
Metab Reg Syn
Other
Degradation
AA Deg
Nucleo Deg
FA/Lip Deg
Amine Deg
Carbo Deg
Sec Metab Deg
Alcohol Deg
Aromatic Deg
Polymer Deg
Other
Energy
Glycolysis
TCA cycle
Fermentation
Aer Resp
Anaer Resp
Other
Other Pathways
Inorganic Nutr
Detoxification
Macromol Mod
Act/Inact/Inter
Other Enzymes
Central Dogma
Transcription
Translation
DNA Metabolism
RNA Metabolism
Prot Metab
Prot Fold/Secr
Regulation
Amino Acid Biosynthesis
T0 T0.5 T1 T2 T5 T10
(a)
TPM
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
Cellular Processes
Cell Cycle/Div
Cell Death
Biofilm
Adhesion
Locomotion
Viral Response
Cell Exterior
Transport
Cell Wall Gen
LPS Metab
Pilus
Flagellum
Outer Mem
Response to Stimulus
Starvation
Heat
Cold
DNA Damage
Osmotic Stress
pH

Figure 6

» Analyze data using the interactive Omics Dashboard tool (Figure 6)

» Perform gene-set enrichment analysis

» Metabolomics data analysis tools include the following capabilities:

» Search BioCyc for metabolites based on monoisotopic mass and/or chemical formula

» Visually overlay metabolomics data onto individual pathway diagrams

» Visually overlay metabolomics data onto multi-pathway diagrams

» Visually overlay metabolomics data onto full metabolic network diagrams

» Sort pathways according to a pathway activation score computed from the metabolomics data

» Analyze data using the interactive Omics Dashboard tool

» Calculate metabolomics-based pathway-covering sets

» Perform metabolite-set enrichment analysis

Comparative Pathway Analysis tools include

» A comparative analysis tool that visually highlights shared reactions on a full metabolic network diagram.

» A tool that generates tables of the pathways that are shared by a set of organisms. In this tool, the user selects a pathway category of interest, such as amino acid biosynthesis or carbohydrate catabolism pathways; the comparison table is based all pathways within that category.

Availability

BioCyc.org is available by subscription to all users. The *E. coli* database at *EcoCyc.org* is freely available.

The Pathway Tools software is freely available to academics and is available for a fee to commercial institutions.

Funding

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under award number R01GM080746, and by the National Institute of Allergy and Infectious Diseases of the NIH under award number R01AI160719.

References Cited

Andersen, H. J., Østdal, H., & Blom, H. 1995. Partial purification and characterisation of a lipase from *Lactobacillus plantarum* MF32. *Food Chem* 53(4):369-373.

Becker, J., Rohles, C. M., & Wittmann, C. 2018. Metabolically engineered *Corynebacterium glutamicum* for bio-based production of chemicals, fuels, materials, and healthcare products. *Metab Eng* 50:122-141.

Behera, S. S., Ray, R. C., & Zdolec, N. 2018. *Lactobacillus plantarum* with Functional Properties: An Approach to Increase Safety and Shelf-Life of Fermented Foods. *Biomed Res Int* 9361614.

Berbert-Molina, M. A., Sato, S., & Silveira, M. M. 2001. Ammonium phosphate as a sole nutritional supplement for the fermentative production of 2,3-butanediol from sugar cane juice. *Z Naturforsch C J Biosci* 56(9-10):787-791.

Brim, H., McFarlan, S. C., Fredrickson, J. K., Minton, K. W., Zhai, M., Wackett, L. P., & Daly, M. J. 2000. Engineering *Deinococcus radiodurans* for metal remediation in radioactive mixed waste environments. *Nat Biotechnol* 18:85-90.

Caspi, R., Altman, T., Billington, R., Dreher, K., Foerster, H., Fulcher, C. A., Holland, T. A., Keseler, I. M., Kothari, A., Kubo, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Ong, Q., Paley, S., Subhraveti, P., Weaver, D. S., Weerasinghe, D., Zhang, P., & Karp, P. D. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids* Res 42:D459-471.

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., Ong, W. K., Paley, S., Subhraveti, P., & Karp, P. D. 2020. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids* Res 48(D1):D445-D453.

Caspi, R., Dreher, K., & Karp, P. D. 2013. The challenge of constructing, classifying, and representing metabolic pathways. *FEMS Microbiol Lett* 345(2):85-93.

Cheng, C., Bao, T., & Yang, S. T. 2019. Engineering *Clostridium* for improved solvent production: recent progress and perspective. *Appl Microbiol and Biotechnol* 103(14):5549-5566.

Claes, I. J., Schoofs, G., Regulski, K., Courtin, P., Chapot-Chartier, M. P., Rolain, T., Hols, P., von Ossowski, I., Reunanen, J., de Vos, W. M., Palva, A., Vanderleyden, J., De Keersmaecker, S. C., & Lebeer, S. 2012. Genetic and biochemical characterization of the cell wall hydrolase activity of the major secreted protein of *Lactobacillus rhamnosus* GG. *PLoS One* 7(2):e31588.

El-Naggar, M. Y., Wanger, G., Leung, K. M., Yuzvinsky, T. D., Southam, G., Yang, J., Lau, W. M., Nealson, K. H., & Gorby, Y. A. 2010. Electrical transport along bacterial nanowires from *Shewanella oneidensis* MR-1. *Proceedings of the National Academy of Sciences of the United States of America* 107(42):18127-18131.

Gerber, E., Bernard, R., Castang, S., Chabot, N., Coze, F., Dreux-Zigha, A., Hauser, E., Hivin, P., Joseph, P., Lazarelli, C., Letellier, G., Olive, J., & Leonetti, J. P. 2015. *Deinococcus* as new chassis for industrial biotechnology: biology, physiology and tools. *J Appl Microbiol* 119(1):1-10.

Ji, X. J., Huang, H., & Ouyang, P. K. 2011. Microbial 2,3-butanediol production: a state-of-the-art review. *Biotechnol Adv* 29(3):351-364.

Karp, P. D., Billington, R., Caspi, R., Fulcher, C. A., Latendresse, M., Kothari, A., Keseler, I. M., Krummenacker, M., Midford, P. E., Ong, Q., Ong, W. K., Paley, S. M., & Subhraveti, P. 2019. The BioCyc collection of microbial genomes and metabolic pathways. *Brief Bioinform* 20(4):1085-1093.

Karp, P. D., Latendresse, M., & Caspi, R. 2011. The pathway tools pathway prediction algorithm. *Stand Genomic Sci* 5(3):424-429.

Kato, S., Hashimoto, K., & Watanabe, K. 2012. Methanogenesis facilitated by electric syntrophy via (semi)conductive iron-oxide minerals. *Environ Microbiol* 14(7):1646-1654.

Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martinez, C., Caspi, R., Kothari, A., Krummenacker, M., Midford, P. E., Muniz-Rascado, L., Ong, W. K., Paley, S., Santos-Zavaleta, A., Subhraveti, P., Tierrafria, V. H., Wolfe, A. J., Collado-Vides, J., Paulsen, I. T., & Karp, P. D. 2021. The EcoCyc Database in 2021. *Front Microbiol* 12:711077.

Kim, Y., Ryu, B. H., Kim, J., Yoo, W., An, D. R., Kim, B. Y., Kwon, S., Lee, S., Wang, Y., Kim, K. K., & Kim, T. D. 2017. Characterization of a novel SGNH-type esterase from *Lactobacillus plantarum*. *Int J Biol Macromol* 96:560-568.

Krummenacker, M., Latendresse, M., & Karp, P. D. 2019. Metabolic route computation in organism communities. *Microbiome* 7(1):89.

Lange, C. C., Wackett, L. P., Minton, K. W., & Daly, M. J. 1998. Engineering a recombinant *Deinococcus radiodurans* for organopollutant degradation in radioactive mixed waste environments. *Nat Biotechnol* 16(10):929-933.

Latendresse, M., Krummenacker, M., & Karp, P. D. 2014. Optimal metabolic route search based on atom mappings. *Bioinformatics* 30(14):2043-2050.

Liao, Z., Suo, Y., Xue, C., Fu, H., & Wang, J. 2018. Improving the fermentation performance of *Clostridium acetobutylicum* ATCC 824 by strengthening the VB1 biosynthesis pathway. *Appl Microbiol Biotechnol* 102(18):8107-8119.

Logan, B. E., & Rabaey, K. 2012. Conversion of wastes into bioelectricity and chemicals by using microbial electrochemical technologies. *Science* 337(6095):686-690.

Malvankar, N. S., & Lovley, D. R. 2014. Microbial nanowires for bioenergy applications. *Current Opinion in Biotechnology* 27:88-95.

Mimee, M., Tucker, A. C., Voigt, C. A., & Lu, T. K. 2015. Programming a Human Commensal Bacterium, *Bacteroides thetaiotaomicron*, to Sense and Respond to Stimuli in the Murine Gut Microbiota. *Cell Syst* 1(1):62-71.

Nikel, P. I., Chavarria, M., Danchin, A., & de Lorenzo, V. 2016. From dirt to industrial applications: *Pseudomonas putida* as a Synthetic Biology chassis for hosting harsh biochemical reactions. *Curr Opin Chem Biol* 34:20-29.

Paley, S., & Karp, P. D. 2021. The BioCyc Metabolic Network Explorer. *BMC Bioinformatics*, 22(1):208.

Pirbadian, S., & El-Naggar, M. Y. 2012. Multistep hopping and extracellular charge transfer in microbial redox chains. *Phys Chem Chem Phys* 14(40):13802-13808.

Rabaey, K., & Rozendal, R. A. 2010. Microbial electrosynthesis - revisiting the electrical route for microbial production. *Nat Rev Microbiol* 8(10):706-716.

Reguera, G., McCarthy, K. D., Mehta, T., Nicoll, J. S., Tuominen, M. T., & Lovley, D. R. 2005. Extracellular electron transfer via microbial nanowires. *Nature* 435(7045):1098-1101.

Sun, L. H., Wang, X. D., Dai, J. Y., & Xiu, Z. L. 2009. Microbial production of 2,3-butanediol from Jerusalem artichoke tubers by *Klebsiella pneumoniae*. *Appl Microbiol Biotechnol* 82(5):847-852.

Tracy, B. P., Jones, S. W., Fast, A. G., Indurthi, D. C., & Papoutsakis, E. T. 2012. *Clostridia*: the importance of their exceptional substrate and metabolite diversity for biofuel and biorefinery applications. *Curr Opin Biotechnol* 23(3):364-381.

Uppada, S. R., Akula, M., Bhattacharya, A., & Dutta, J. R. 2017. Immobilized lipase from *Lactobacillus plantarum* in meat degradation and synthesis of flavor esters. *J Genet Eng Biotechnol* 15(2):331-334.

van Dijl, J. M., & Hecker, M. 2013. *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb Cell Fact* 12:3.

Wang, A., Wang, Y., Jiang, T., Li, L., Ma, C., & Xu, P. 2010. Production of 2,3-butanediol from corncob molasses, a waste by-product in xylitol production. *Appl Microbiol Biotechnol* 87(3):965-970.

Wang, B., Pugh, S., Nielsen, D. R., Zhang, W., & Meldrum, D. R. 2013. Engineering cyanobacteria for photosynthetic production of 3-hydroxybutyrate directly from CO2. *Metab Eng* 16:68-77.

Wang, Y., Sun, T., Gao, X., Shi, M., Wu, L., Chen, L., & Zhang, W. 2016. Biosynthesis of platform chemical 3-hydroxypropionic acid (3-HP) directly from CO2 in cyanobacterium *Synechocystis* sp. PCC 6803. *Metab Eng* 34:60-70.

Yan, F., Cao, H., Cover, T. L., Whitehead, R., Washington, M. K., & Polk, D. B. 2007. Soluble proteins produced by probiotic bacteria regulate

intestinal epithelial cell survival and growth. *Gastroenterology* 132(2):562-575.

Yang, Y., Lang, N., Yang, G., Yang, S., Jiang, W., & Gu, Y. 2016. Improving the performance of solventogenic clostridia by reinforcing the biotin synthetic pathway. *Metab Eng* 35:121-128.

Yu, E. K., & Saddler, J. N. 1982. Enhanced Production of 2,3-Butanediol by *Klebsiella pneumoniae* Grown on High Sugar Concentrations in the Presence of Acetic Acid. *Appl Environ Microbiol* 44(4):777-784.

Yu, E. K. C., Deschatelets, L., Tan, L. U. L., & Saddler, J. N. 1985. A simple method for xylanase preparation used for the hydrolysis and fermentation of hemicellulose to butanediol. *Biotechnol Lett* 7(6):425-430.

Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y. J., & Zhang, F. 2013. Development of *Synechocystis* sp. PCC 6803 as a phototrophic cell factory. *Mar Drugs* 11(8):2894-2916.

Zheng, Y., Zhang, H., Zhao, L., Wei, L., Ma, X., & Wei, D. 2008. One-step production of 2,3-butanediol from starch by secretory over-expression of amylase in *Klebsiella pneumoniae*. *J Chem Technol Biotechnol* 83(10):1409-1412.

About the Authors

Peter Karp (*pkarp@ai.sri.com*; SRI International, 333 Ravenswood Ave, Menlo Park, CA 94025) directs the Bioinformatics Research Group at SRI International. He earned his PhD in Computer Science from Stanford University. His research interests include computational genomics, computational metabolic pathways, scientific databases, and scientific visualization. Karp has authored more than 190 publications in computer science and bioinformatics. He is an SRI Fellow, a Fellow of the International Society for Computational Biology, and a Fellow of the American Association for the Advancement of Science.

Ron Caspi (*ron.caspi@sri.com*) is a curator at the Bioinformatics Research Group at SRI International. He received his PhD in Marine Microbiology from the Scripps Institution of Oceanography at the University of California, San Diego. He is also the secretary of the IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and a member of the Enzyme Commission (EC).

Figure Legends

- Figure 1. BioCyc shows full structures for compounds in compound, reaction, and pathway pages; users can specify the level of details in pathway diagrams. Compound structures are uniformly protonated to the typical cellular pH of 7.3. All balanceable reactions are balanced, and an optional atom-mapping coloring helps users follow the chemical transformations. The control panel on the left allows zooming in/out and printing the diagram to a PDF file.
- Figure 2. Pathway diagrams show compounds, reactions, EC numbers, enzymes catalyzing the reactions, genes encoding the enzymes, and links to upstream and/or downstream pathways. The flask icon in the upper right corner indicates experimental evidence for the pathway. Every object can be clicked to navigate to its own page, providing additional details. A Detail Level button allows the user to select which features are displayed.
- Figure 3. The Metabolic Network Explorer provides interactive exploration of the metabolic network of an organism. This figure shows a user constructed route consisting of two reactions. The route can be expanded or modified by clicking on the circled plus symbols flanking the reactions.
- Figure 4. The Metabolic Route Search tool aims to find possible metabolic routes connecting a specified starting molecule to a specified product metabolite. Reactions imported from MetaCyc (indicating the enzymes catalyzing them are not present in the explored organism) are shown in red.
- Figure 5. The Cellular Overview Omics Viewer allows painting omics data on a diagram that shows all pathways and enzymatic reactions of an organism. The diagram can accept many types of omics data, such as transcriptomics, proteomics, and metabolomics. A control panel allows the user to control many aspects of the diagram. When multiple time points are recorded, the tool supports animation, moving automatically from one time point to another. Sophisticated zooming controls display more information as the zooming level increases. Popups can show information for individual enzymes in multiple plot types.
- Figure 6. The Omics Dashboard consists of a series of panels that summarize omics data for different cellular systems. Each panel contains a set of plots representing one subsystem, with large dots showing the average of multiple values. Clicking on a panel opens a new panel with more data (insert at bottom right corner) enabling to user to move from a high-level view representing the whole organism into individual enzymes with only a few clicks.