



## BioCyc.org Genome and Metabolic Pathway Web Portal and Pathway Tools Software

SRI offers innovative tools for modeling and analyzing genomes, metabolic pathways, and regulatory networks to support activities in drug discovery, synthetic biology, biotechnology, and basic research. These tools accelerate research and support novel analyses, leading to **better science, faster**.

SRI's BioCyc.org web portal contains 20,000 Pathway/ Genome Databases (PGDBs) for sequenced genomes. A PGDB contains the entire genome of an organism, as well as its biochemical pathways and (when curated) its regulatory network. Two BioCyc databases, EcoCyc<sup>1</sup> and MetaCyc<sup>2</sup> are derived from more than three decades of literature-based curation of genome and pathway data. The BioCyc.org website is used each year by more than 600,000 unique users worldwide.

The downloadable Pathway Tools software,<sup>3</sup> licensed to date to 15,500 groups, is a pathway/genome data management system that enables creation of a private BioCyc portal and/ or creation of BioCyc-like portals containing custom genome data. The software can ingest proprietary genome data, compute a metabolic reconstruction, then make the genome accessible through search, visualization, and comparative-analysis tools. Data can be accessed using a web browser and through a desktop application.

“ For Ginkgo scientists, BioCyc and Pathway Tools are the go-to resources of knowledge and software for exploring ideas, answering questions, and analyzing data.”

—Ginkgo Bioworks

“ I'm impressed with the level of detail and care in BioCyc annotations. It is a tremendous resource for pathway analysis in metabolomics.”

— Prof. Art Edison, University of Georgia

### BIOCYC FEATURES

- 20,000 Pathway/Genome Databases
- 68 databases curated from 146,000 publications
- Extensive mini-reviews for genes and pathways
- Most comprehensive pathway database MetaCyc
- BioCyc.org Tools:
  - Search and visualization of genomes and pathways
  - Transcriptomics data analysis
  - Metabolomics data analysis
  - Design novel pathways

### BioCyc Benefits: Better Science, Faster

BioCyc provides major advantages over free databases (detailed comparison: <https://biocyc.org/biocyc-free-comparison.pdf>)

- BioCyc curators distill the literature, saving users time in searching, reading, synthesizing, and reconciling errors in the literature
- BioCyc integrates a vast array of data, reducing time to learn and search other databases
- Gene functions are more fully understood in their pathway and regulatory context

BioCyc's accurate, comprehensive data collection and unique analysis tools are a foundation for better science:

- Higher quality publications and grant proposals
- Faster, more complete pathway analysis of metabolomics, transcriptomics, and multi-omics data
- Tool for design of optimized pathways for synthetic biology and metabolic engineering

### Applications of BioCyc and Pathway Tools

BioCyc and Pathway Tools span genome informatics, pathway informatics, and regulatory informatics. Applications include:

- Enterprise Genome Data Management: Extensive search tools to speed information finding; multiple visualization tools expedite user uptake of information
- Computational Inferences: Predict metabolic pathways, genes coding for missing enzymes in metabolic pathways, protein complexes, and operons
- Microbiome Analysis: Create PGDBs from metagenome-assembled genomes (MAGs); search and compare community members.



Pathway Tools is free to academic groups for research and teaching.



▲ Figure 1: The Omics Dashboard provides a visual readout of the cellular state within an omics dataset.

- Omics Data Analysis: Enrichment analysis, and display gene expression, metabolomics, and proteomics data in the Omics Dashboard (Figure 1), and on a metabolic map diagram configured for each organism (Figure 2)
- Drug Discovery:
  - Target selection: Finds drug targets by identifying essential genes using chokepoints and metabolic modeling, identifying enzymes present in multiple pathways, and identifying previously uncharacterized genes filling holes in the metabolic network.
  - Lead generation: Extensive data on enzyme inhibitors in EcoCyc and MetaCyc
  - Target and lead evaluation: Improved analysis of omics data
- Synthetic Biology:
  - Extensive characterization of cellular metabolism for common chassis organisms
  - MetaCyc database catalogs 3,100 metabolic pathways, 19,000 reactions, and 14,000 enzymes from all domains of life, gathered from 76,000 publications

- RouteSearch Tool designs optimal pathways that convert a source metabolite to a target metabolite by combining organism reactions with MetaCyc reactions
- Quantitative metabolic modeling guides alternative strain designs

- Comparative Genome and Pathway Analyses

## The Biocyc Database Collection

Each BioCyc PGDB describes the genome and metabolic network of a single organism, and in some cases its regulatory network. PGDBs integrate data imported from external databases, computational inferences, and curated information for 68 of the databases.

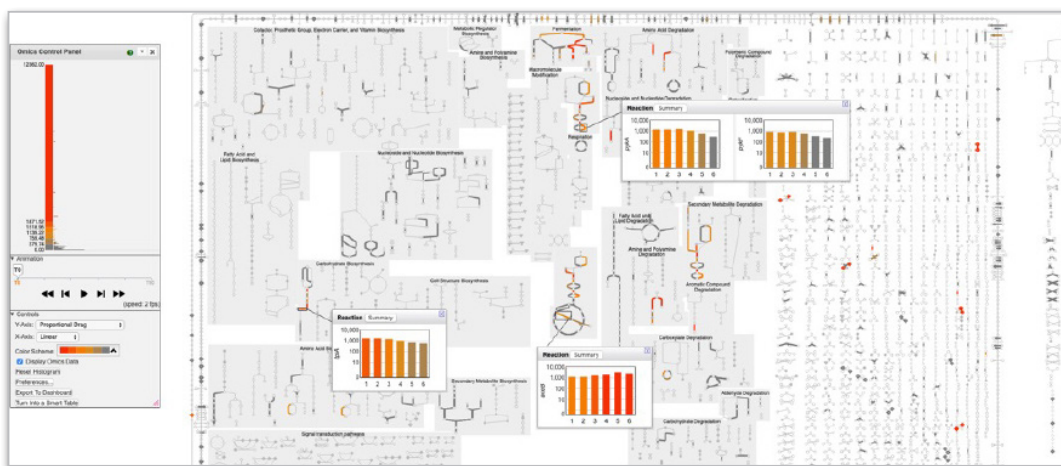
The most highly curated databases in BioCyc are:

- EcoCyc: Pathway/Genome Database for *Escherichia coli* K-12 MG1655. EcoCyc data have been gathered during two decades of literature-based curation from more than 44,000 articles.<sup>1</sup> EcoCyc provides the equivalent of 4,000 textbook pages of mini-review summaries for 4,172 *E. coli* genes.
- HumanCyc: HumanCyc includes literature-based curation of human enzymes and metabolic pathways and is frequently used for metabolomics data analysis.
- Additional curated databases include *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Mycobacterium tuberculosis*, *Staphylococcus aureus*, *Clostridioides difficile*, *Streptococcus pneumoniae*, *Helicobacter pylori*, and *Vibrio cholerae*.

All PGDBs include the genome, predicted metabolic pathways, predicted pathway hole fillers (genes coding for missing (unannotated) enzymes in metabolic pathways) and, for bacteria, predicted operons.

Additional types of data integrated in BioCyc include:

- Protein subcellular locations; enzyme activators, inhibitors, cofactors; enzyme kinetics data; protein features; Gene Ontology terms; predicted Pfam domains
- Regulatory information including promoters, terminators, operons, transcription-factor binding sites, regulatory RNAs
- Gene essentiality data



◀ Figure 2: The Cellular Omics viewer paints omics datasets onto a diagram of the cellular biochemical network. Reaction lines can be colored with gene expression, proteomics, or reaction flux data; compound nodes can be colored with metabolomics data. Multi-omics data can be analyzed by coloring data onto reactions and metabolites simultaneously. Omics pop-ups graph omics data values using bar graphs, heat maps, or X-Y plots.

- Reaction atom mappings
- Gibbs free energies of formation for metabolites
- Growth media
- Ortholog relationships to other BioCyc genomes
- Database links (e.g., to UniProt and RefSeq)

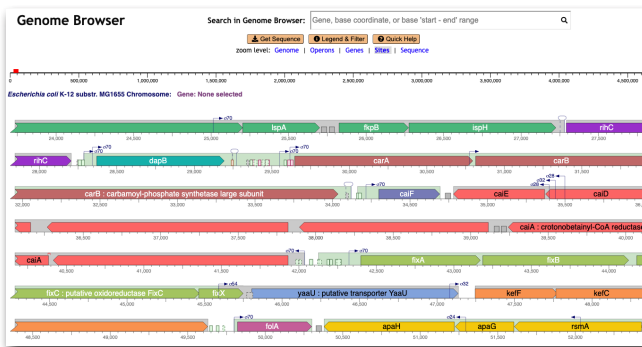
### Genome Informatics Tools

- Find genomes by name, taxonomy, phenotypic properties
- Gene information page
  - Retrieve amino-acid sequence and nucleotide sequence of arbitrary genome region
  - Search genes by name, accession number, sequence length, GO terms, and protein properties (pI, MW, protein features, subcellular location, ligand)
- The Genome Browser (Figure 3) was re-implemented in 2023 with ultra-fast zooming. It depicts genomic regions at user-selected resolution with semantic zooming. Visible features include pseudo-genes, promoters, transcription-factor binding sites, repeats, terminators, and nucleotide sequence. Genome posters can be generated.

BLAST search, sequence-pattern search, show effects of SNPs on translation, multiple sequence alignment, sequence retrieval.

### Pathway and Regulatory Informatics Tools

- Reaction, metabolite, and pathway searches
  - Pathway information page includes zoomable pathway diagrams with chemical structures
- Reaction information page
- Metabolite information page
  - Search metabolites by name, substructure, mass, monoisotopic mass, elemental composition
- Customize pathway diagrams for publication or assemble groups of pathways into a pathway collage
- Cellular Overview diagrams (Figure 2) are organism-specific depictions of metabolic and transporter networks that are zoomable and searchable
- Predict metabolic network from annotated genome
- Metabolic route search for metabolic engineering and organism communities
- The Regulatory Overview presents the genetic regulatory network stored in a PGDB



▲ Figure 3. Genome browser depiction of a region of the *E. coli* chromosome. Gene colors indicate operon organization. Promoters and terminators are depicted when known. Pseudogenes are marked with X's.

### Analysis Tools for Gene Expression and Metabolomics Data

- SmartTables store lists of genes, metabolites, pathways, and more. Browse database attributes, share with colleagues, transform to pathway lists, perform enrichment analysis
- Enrichment analysis for gene and metabolite sets
- The Omics Dashboard presents a visual readout of the expression status of all cellular systems to facilitate a rapid top-down user survey of cellular responses (Figure 1)
- Paint omics data onto individual pathways, pathway collage, and full metabolic networks (Figure 2) (example animation at <http://biocyc.org/ov-expr.shtml>)

### Microbiome Informatics

- Calculate pathway abundances across metagenome samples
- Create PGDBs for an organism community
- Search across multiple PGDBs, compare PGDBs
- Analysis of meta-omics data

### Comparative Genomics Tools

- Comparative genome browser (Figure 4) aligns chromosomal regions from multiple genomes at orthologous genes
- Compare pathway, reaction, metabolite, and protein complements of specified organisms
- Find genes, metabolites, and pathways across organisms

### Advanced Database Access

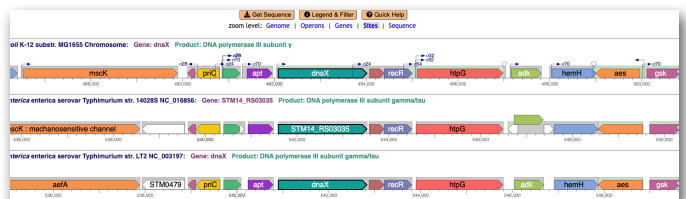
- Web services API; Python, Perl, Java, and Lisp APIs
- Author advanced queries of SQL power using the intuitive Structured Advanced Query Form
- Export PGDBs to BioPAX, SBML, GFF, and Genbank formats

### Pathway Tools Software

The Pathway Tools software powers the BioCyc.org web portal. Bring Pathway Tools in-house to create a private BioCyc portal containing selected BioCyc PGDBs, and to create private PGDBs for proprietary genomes.

### PATHWAY TOOLS FEATURES

- Enterprise pathway/genome data management
- Desktop and web operation
- Computational prediction of metabolic pathways, operons, protein complexes
- Incorporate selected BioCyc databases
- Quantitative metabolic modeling
- Includes BioCyc tools



▲ Figure 4. Comparative genome browser showing alignments with respect to the *yhhY* gene in one *E. coli* genome, a *Shigella flexneri* genome, and a *Salmonella enterica* genome. Colors indicate orthologs; there is strong conservation to the left of the *yhhY* gene, but not to the right of that gene (genes in white are not conserved).

## Metabolic Modeling in Pathway Tools

The MetaFlux module generates a flux balance analysis (FBA) model automatically from a PGDB, enabling quantitative modeling of steady-state metabolic fluxes. Combine models for multiple organisms to model organism communities.

By combining pathway databases with FBA, MetaFlux achieves close coupling of the FBA model to genome and metabolite data and high accessibility of the FBA model via the query and visualization features of Pathway Tools. MetaFlux speeds comprehension of simulation results by painting computed fluxes on metabolic-map diagrams and on individual pathways, and MetaFlux will plot metabolite concentrations and organism biomass changes. Dynamic FBA enables temporal simulations.

Development of FBA models is accelerated by a multiple gap-filling tool that postulates additional reactions to add to an FBA model to complete it, and that identifies what subset of biomass components can be produced by the current model. Modeling of gene knock-outs is supported.

## Pathway Tools Benefits

Bring BioCyc into a secure proprietary environment.  
Extract more information from proprietary genomes via

- Computational inferences such as pathway prediction and operon prediction
- Search, visualization, and comparative tools
- Omics data analysis tools

## Technical Specifications/Configurations

1. Pathway/Genome Navigator bundled with BioCyc Databases provides search, visualization, and analysis of BioCyc databases. The same binary application can run as both a desktop application and as a Web server within an organization's intranet.

- 1) Base configuration: Pathway/Genome Navigator plus as many as 500 user-selected BioCyc databases
  - Platforms supported: Linux, Windows (Web mode not available), Macintosh
  - Hardware: 2.4 GHz processor, 16 GB RAM, 5 GB disk
  - For more than 500 databases requires Linux with 2.4 GHz processor, 128 GB RAM, 3 TB disk

2. These software modules may be added to configuration 1:

- Editors module: enables editing of PGDBs
- PathoLogic module: enables creation of new Pathway/Genome Databases from annotated genomes
- PathoLogic requires the EcoCyc and MetaCyc PGDBs.
- MetaFlux module: Enables creation of quantitative metabolic flux models

## About SRI's Bioinformatics Research Group

SRI International, an independent research institute, is a key player in the field of computational biology. SRI's Bioinformatics Research Group, which produces BioCyc and Pathway Tools, is a leader in the development of database content and software tools for bioinformatics.

## References

- [1] The EcoCyc Database. EcoSal Plus. 2018. <http://www.asmscience.org/content/journal/ecosalplus/10.1128/ecosalplus.ESP-0006-2018>
  - [2] MetaCyc. Nucleic Acids Research 2020. doi: 10.1093/nar/gkz862
  - [3] Pathway Tools version 24.0: Integrated Software for Pathway/Genome Informatics and Systems Biology, 2020. <http://arxiv.org/abs/1510.03964v4>
  - [4] The Pathway Tools Software and Its Role in Anti-Microbial Drug Discovery, Microbial Genomics and Drug Discovery, T.J. Dougherty and S.J. Projan eds., Marcel Dekker Inc., New York, 2003
- Additional publications: <http://biocyc.org/publications.shtml>

## For More Information

Downloadable Pathway Tools Software

- Freely available to academics and government laboratories for research purposes; see <http://biocyc.org/download.shtml>
- For commercial use click here <https://share.hsforms.com/13fwJvezjSDyUXI8BRPpkQ3ipje> or contact [ptools-commercials@ai.sri.com](mailto:ptools-commercials@ai.sri.com).

BioCyc.org Website Subscriptions

- Subscriptions provide access to more than 20,000 organism databases and extensive online software tools.
- The EcoCyc database is free to all users.
- Subscriptions are available for institutions, labs, and individuals.
- Contact [biocyc-sales@sri.com](mailto:biocyc-sales@sri.com).



333 Ravenswood Avenue  
Menlo Park, CA 94025-3493  
650.859.2000  
[www.sri.com](http://www.sri.com)

SRI and SRI International are registered trademarks of SRI International.  
All other trademarks are the property of their respective owners. ©2024. All rights reserved. 06/24