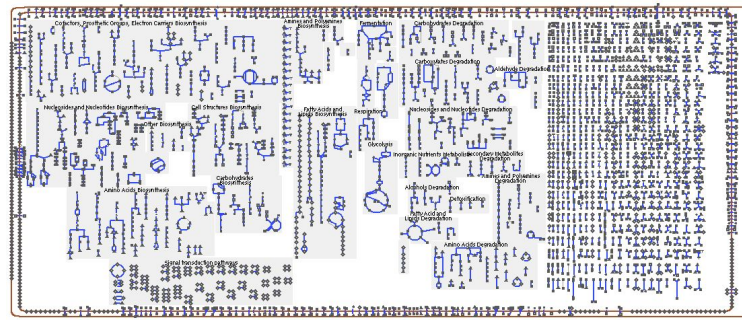




Introduction to BioCyc

Dylan Beesley
SRI International

ecocyc.org
biocyc.org
metacyc.org



**Following the wrong directions
costs time**



**Following the wrong
data costs much more**



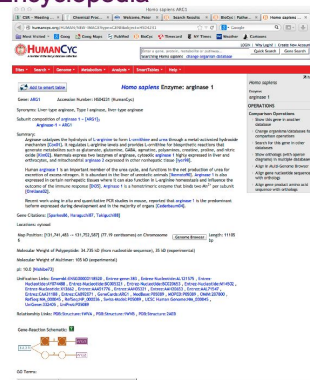
What is BioCyc and Why Use It?

- Established in 1997
- Web portal that integrates genome and metabolic pathway data with powerful bioinformatics tools
- Easy to use
- Integration of **extensive, high-quality** data – one-stop shop
 - Computational inferences
 - Imports from other databases
 - Extensive literature-based curation
- Large assortment of bioinformatics tools
- Even computationally generated databases utilize a curated baseline layer of data, MetaCyc

BioCyc Web Portal and Database Collection

- Encyclopedic reference source
 - 20,000 total databases
 - 90 curated databases, contents derived from 153,000 publications
 - Humans, pathogens, model organisms, cyanobacteria
 - <https://biocyc.org/biocyc-pgdb-list.shtml>
- Use cases:
 - Genome exploration
 - Multi-omics data analysis
 - Sequence analysis
 - Comparative analysis
 - Synthetic biology/Chassis Design
 - Microbiome studies
 - Environmental metagenomics

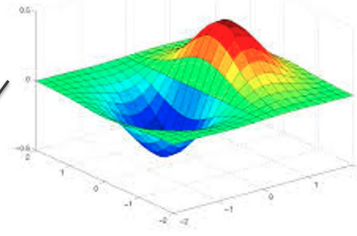
Encyclopedia



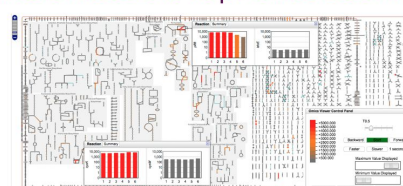
Queryable Database



Metabolic Model



Zoomable Metabolic Map / Omics Data Analysis



BioCyc Curation Team

- Performs literature searches for new gene functions, pathways
- Author mini-reviews, insert citations and evidence codes
- Updates database fields
 - Gene names, synonyms, protein functions
 - Enzyme activators, inhibitors, cofactors
 - Protein features, complexes
- Reviews and extends enzyme/transporter reaction associations
- Reviews predicted metabolic pathways
- Regulation: TF binding sites, promoters, regulation of transcription initiation and other processes

How To Quantify the Value of Curation

- **How many protein functions per genome are corrected by curators?**
- We chose 10 Random BioCyc microbial genomes
- For each genome:
 - Find all proteins whose names changed between start and end of curation
 - Randomly chose 200 curated proteins from each genome (2,000 proteins reviewed)
 - Manually score each protein name-change as either:
 - Functional change example: CapA family protein -> poly-gamma-D-glutamate transporter
 - Stylistic change example: ferrous iron transport protein B □ Fe²⁺ transporter FeoB

On average, 13.0% of protein functions per genome updated during curation

How do We Quantify the Value of Curation?

On average, 13.0% of protein functions per genome corrected during curation

Plus enz/rxn associations, pathway correction, mini-review summaries

Recently published:

Caspi R, Wilson-Mortier B, Moore LR, Karp JN, Beesley D, Paulsen IT and Karp PD (2026) Assessment of the impact of manual curation in BioCyc. *Front. Bioinform.* 6:1887950. doi: 10.3389/fbinf.2026.1887950

	A	B	D	C	E
Organism Database	Protein-Coding Genes	Proteins with Name Changes	Estimated Proteins with Function Changes	Percent Functional Changes from Sampled Proteins	Estimated Percent of Proteins Corrected
Acinetobacter baumannii ATCC 17978	3636	733	370.2	50.50%	10.2%
Chlamydia trachomatis D/UW-3/CX	889	110	68.2	62.04%	7.6%
Enterococcus faecalis V583	3193	596	369.5	62.00%	11.6%
Helicobacter pylori 26695	1445	524	138.9	26.50%	9.6%
Lactiplantibacillus plantarum WCFS1	3014	960	607.8	63.32%	20.1%
Neisseria gonorrhoeae FA 1090	1934	354	231.9	65.50%	12.0%
Staphylococcus aureus NCTC 8325	2825	1020	642.6	63.00%	22.8%
Streptococcus pneumoniae D39V	1991	608	106.4	17.50%	5.3%
Synechocystis sp. PCC 6803	3630	881	577.1	65.50%	15.9%
Vibrio cholerae O1 biovar El Tor str. N16961	3569	967	517.3	53.50%	14.5%
Sum or Average*	26126	6753	3575	52.9%	13.0%

Curated Databases

Model & Reference Organisms

- *Escherichia coli* K-12 substr. MG1655
- *Homo sapiens*
- *Arabidopsis thaliana*
- *Saccharomyces cerevisiae*
- *Caulobacter crescentus* NA1000
- *Caulobacter vibrioides* CB15
- *Drosophila melanogaster*
- *Mus musculus*
- *Bacillus subtilis* subtilis 168
- *Nostoc/Anabaena* sp. PCC 7120
- *Streptomyces coelicolor* A32

Human Microbiome & Commensals

- *Bifidobacterium longum* longum JCM 1217
- *Collinsella aerofaciens* ATCC 25986
- *Faecalibacterium prausnitzii* A2-165
- *Lactobacillus plantarum* WCFS1
- *Lactobacillus rhamnosus* GG
- *Methanobrevibacter smithii* ATCC 35061
- *Porphyromonas gingivalis* W50
- *Prevotella copri* DSM 18205
- *[Clostridium] symbiosum* WAL-14163
- *[Eubacterium rectale]* ATCC 33656

Synthetic Biology & Biotechnology

- *Escherichia coli* B str. REL606
- *Escherichia coli* K-12 substr. W3110
- *Corynebacterium glutamicum* ATCC 13032
- *Cupriavidus necator* H16
- *Clostridium saccharoperbutylacetonicum* N1-4HMT
- *Lactococcus lactis* lactis IL1403
- *Pseudomonas putida* KT2440
- *Synechococcus elongatus* PCC 7942
- *Synechocystis* sp. PCC 6803 substr. Kazusa
- *Vibrio natriegens* ATCC 14048

Environmental & Biogeochemical Research

- *Alteromonas macleodii* ATCC 27126
- *Aurantimonas manganoxydans* SI85-9A1
- *Desulfovibrio piger* ATCC 29098
- *Methanocaldococcus jannaschii* DSM 2661
- *Prochlorococcus marinus* marinus CCMP1375
- *Prochlorococcus marinus* pastoris CCMP1986
- *Sulfurospirillum multivorans*
- *Synechococcus* sp. WH 8102
- *Thalassiosira pseudonana* CCMP1335

Curated Databases

Infectious Disease & Pathogens

- *Acinetobacter baumannii* AB5075
- *Acinetobacter baumannii* ATCC 17978
- *Bacillus anthracis* Ames
- *Brucella abortus* 2308
- *Brucella ovis* ATCC 25840
- *Chlamydia trachomatis* D/UW-3/CX
- *Clostridioides difficile* 630
- *Clostridioides difficile* R20291
- *Enterococcus faecalis* V583
- *Enterococcus faecalis* OG1RF
- *Escherichia coli* CFT073
- *Escherichia coli* O157:H7 str. EDL933
- *Helicobacter pylori* 26695
- *Klebsiella pneumoniae* NTUH-K2044
- *Listeria monocytogenes* 10403S
- *Mycobacterium tuberculosis* CDC1551
- *Mycobacterium tuberculosis* H37Rv
- *Neisseria gonorrhoeae* FA 1090
- *Salmonella enterica* enterica serovar Typhimurium str. 14028S
- *Salmonella enterica* enterica serovar Typhimurium str. LT2
- *Salmonella enterica* enterica serovar Typhimurium str. SL1344
- *Shigella flexneri* 2a str. 2457T
- *Staphylococcus aureus* aureus NCTC 8325
- *Staphylococcus epidermidis* RP62A
- *Streptococcus pneumoniae* D39V
- *Vibrio cholerae* O1 biovar El Tor str. N16961

Parasites & Disease Vectors

- *Anopheles gambiae*
- *Leishmania major* strain Friedlin
- *Plasmodium falciparum* 3D7
- *Schistosoma mansoni*
- *Trypanosoma brucei*

Plant & Host-Associated Symbiosis

- *Agrobacterium fabrum* C58
- *Candidatus Cardinium hertigii*
- *Candidatus Evansia muelleri*
- *Candidatus Portiera aleyrodidarum* BT-QVLC
- *Sinorhizobium meliloti* 1021

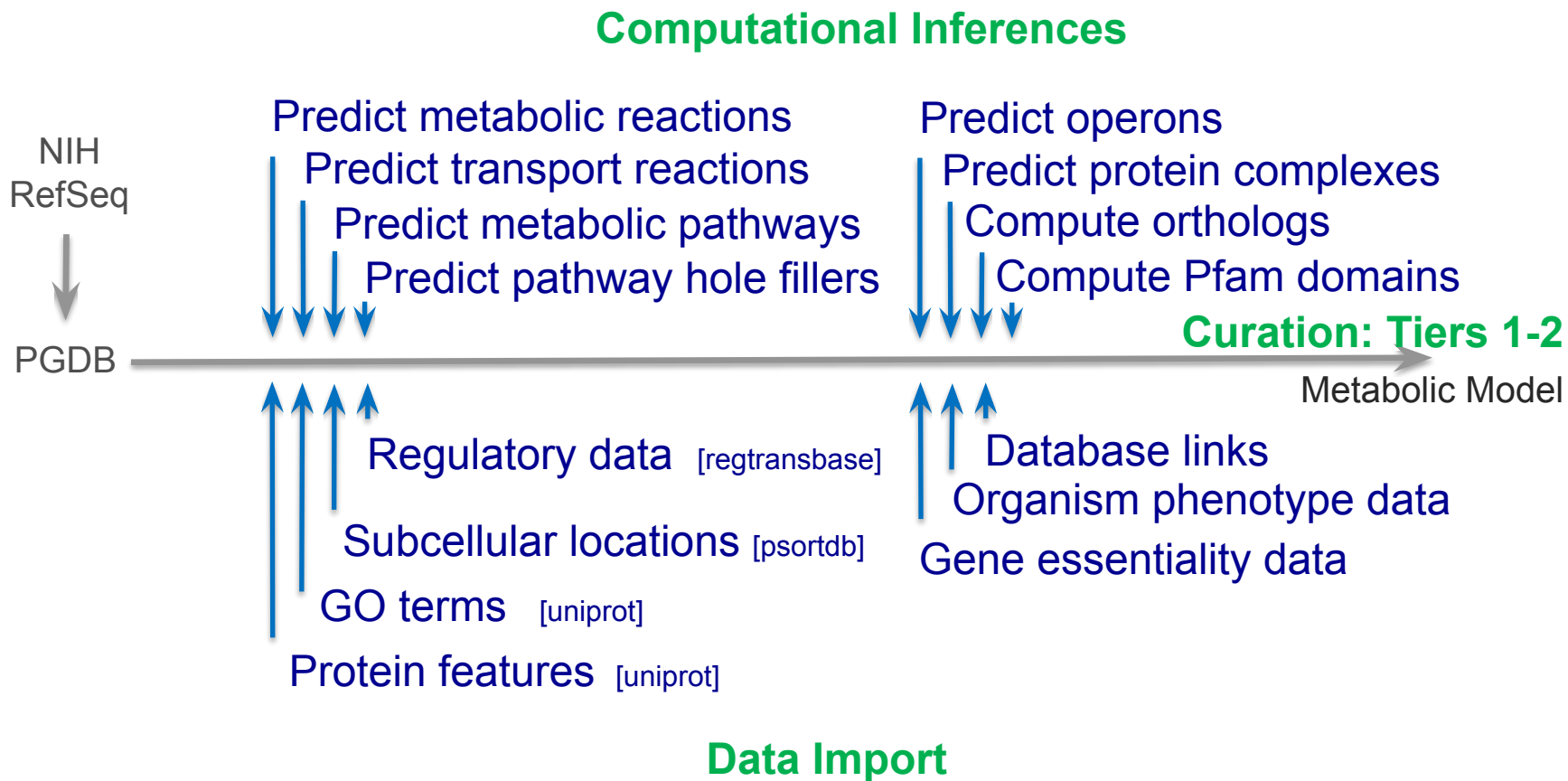
Synthetic Biology Applications

- **Gene Essentiality Data** for key organisms can assist in designing knock-out's
 - Gene essentiality provides experimental evidence of the impact of knock-outs on the organism under various experimental conditions
- The **Metabolic Route Search** tool can be utilized to identify potential biosynthetic pathways between a starting compound and a target compound
 - Search within an organism using its native enzymes and pathways
 - Search across the entire BioCyc database, allowing incorporation of non-native reactions for pathway engineering
- **Multi-Omics Integration** allows users to upload gene expression, metabolomics, proteomics, or other large data sets and paint them on to the metabolic network diagram

MetaCyc: Metabolic Encyclopedia

- Experimentally determined metabolic pathways, reactions, enzymes, and compounds
- Metabolism from all domains of life
- Literature-based DB with extensive references and commentary
 - Derived from 75,000 publications
 - 3,200 metabolic pathways; 19,700 reactions; 20,000 metabolites
- Extensive information on pathway variants, enzyme properties
 - 8 variants of the TCA cycle
 - 6 variants of L-lysine biosynthesis

Creation of BioCyc Databases (PGDBs)



BioCyc Software Tools

- **Genome Informatics**

- Gene/protein/RNA searches, gene information page
- Circular and linear genome browsers
- BLAST search
- Sequence pattern search
- Multiple sequence alignment
- Sequence retrieval

- **Pathway Informatics**

- Reaction/metabolite/pathway searches and info pages
- Infer metabolic network from genome
- Zoomable metabolic network diagrams
- Metabolic network explorer
- Metabolic route search for single organisms and communities
- Pathway-based inference of nutrient requirements
- Metabolic modeling using flux-balance analysis
 - FBA, dFBA, FVA, gap filling, blocked metabolites

- **Regulatory Informatics**

- Capture/visualize promoters, TF binding sites
- Visualize regulon of transcription factor
- Regulatory network viewer
- Operon inference

- **Microbiome Informatics**

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

- **Comparative Analysis**

- Search across PGDBs for genes, proteins, pathways, metabolites
- Comparative genome browser
- Compare pathway, reaction, and metabolite complements

- **Transcriptomics and Metabolomics Data Analysis**

- Enrichment analysis
- Paint transcriptomics data onto pathways, zoomable metabolic network
- Omics Dashboard
- Sort pathways by pathway activation score
- Pathway covering analysis

- **Advanced Data Manipulation**

- SmartTables
- Advanced Query Interface
- Interactive editing tools
- APIs: Python, Perl, Java, Lisp, R, web services
- Import/Export: SBML, GenBank, GFF3, BioPAX, tab-delimited files