Bioenergy Curation in the MetaCyc Database of Metabolic Pathways

Ron Caspi* (caspi@ai.sri.com), Deepika Weerasinghe, and **Peter Karp**

Bioinformatics Research Group, SRI International, Menlo Park, CA

http://metacyc.org
http://biocyc.org

Project Goals: Two goals of our project are (1) To expand the coverage of bioenergy-related metabolic information in the MetaCyc database and (2) To generate within the BioCyc database collection organism-specific PGDBs for all energy-relevant organisms sequenced by JGI.

MetaCyc (metacyc.org) is a literature-curated database containing more than 1,800 metabolic pathways, collected from a wide variety of organisms, with an emphasis on microorganisms and plants. The goal of MetaCyc is to contain a representative sample of every experimentally elucidated pathway, thereby cataloging the universe of known pathways and enzymes. MetaCyc contains rich, detailed, and high quality data, including minireview summaries with extensive literature citations, enzyme information, evidence codes, and links to other databases.

In addition to its role as an encyclopedic resource for metabolic pathways, MetaCyc also serves as a reference database for the Pathway Tools software, which predicts the metabolic pathway complement of an organism from its annotated genome, creating a Pathway/Genome Database (PGDB) for that organism.

Since only pathways that already exist in MetaCyc can be inferred by Pathway Tools in organism-specific databases, the pathway content of MetaCyc strongly influences the metabolic networks predicted by Pathway Tools. Two goals of our project are (1) To expand the coverage of bioenergy-related metabolic information in MetaCyc, and (2) To generate within our BioCyc database collection organism-specific PGDBs for all energy-relevant organisms sequenced by JGI.

During the past year we have done the following:

- Greatly expanded the coverage in MetaCyc of compounds found in cellulosic biomass, with an emphasis on natural cellulosic and hemicellulosic polymers.

- Created 8 pathways for the degradation of important biomass compounds such as celluloses, rhamnogalacturonans, xylans, arabinans, xyloglucans, and carrageenans.

- Curated many enzymes that are involved in cellulosic biomass degradation.

- Curated eight naturally occurring and/or bioengineered hydrogen production pathways and related enzymes.

- Curated ten naturally-occurring and/or engineered biosynthetic pathways for potential biofuel compounds, including 1-butanol, 3-methyl butanol, isopropanol, all-trans-farnesol, long chain fatty acid esters, the algal lipid diacylglyceryl-N,N,N-trimethylhomoserine, the algal fatty acids docosahexanoate, arachidonate , and eicosapentaenoate, and the algal triterpinoid botryococcene.

In the remainder of the second year of this project we will add a new type of pathway diagram that will facilitate the presentation of enzymatic degradation of complex polymers, and will use this new tool to create many pathway diagrams for lignocellulose-degradation pathways.

References

Caspi et al (2012) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. Database Issue[Epub ahead of print] PMID 22102576.

Funding

MetaCyc and BioCyc are funded by grant GM80746 from the NIH National Institute of General Medical Sciences. Bioenergy curation is funded by grant DE-SC0004878 from the Department of Energy.