173.   Metabolic RouteSearch in Pathway Databases and Expansion of MetaCyc

Mario Latendresse,[1] Ron Caspi,[1] **Peter D. Karp[1]**

[1]Bioinformatics Research Group, SRI International, Menlo Park, CA.

Project Goals:
- **Extend Pathway Tools software to support pathway design**
- **Extend the MetaCyc pathway database to support metabolic engineering and pathway analysis of bioenergy-related organisms**

The RouteSearch Software for Metabolic Engineering
A key computational problem in metabolic engineering is finding efficient metabolic routes from a source to a target compound in genome---scale reaction networks, potentially considering the addition of new reactions. Efficiency can be based on many factors, such as route lengths, atoms conserved, and the number of new reactions added to the route. Fast algorithms are needed to systematically search these large genome-scale reaction networks.

We present the algorithm used in the new RouteSearch tool within the Pathway Tools software. This algorithm is based on a general Branch-and-Bound search and involves constructing a network of atom mappings to facilitate efficient searching. As far as we know, it is the first published algorithm that finds guaranteed optimal routes where atom conservation is part of the optimality criteria. RouteSearch includes a graphical user interface that speeds user understanding of its search results. We evaluated the algorithm on five example metabolic-engineering problems from the literature; for one problem the published solution was equivalent to the optimal route found by RouteSearch; for the remaining four problems, RouteSearch found the published solution as one of its best-scored solutions. These problems were each solved in less than five seconds of computational time.

In addition, RouteSearch can be used to search for reaction paths between metabolites in an organism's metabolic network without the addition of new engineered reactions.

RouteSearch is accessible at BioCyc.org by using the menu command Metabolism ---> Route Search, and by downloading Pathway Tools. Download from: http://biocyc.org/download.shtml

Expansion of the MetaCyc Pathway Database
The goal of the curation part of this project was to enhance the bioenergy-related content of the MetaCyc database in several respects. All in all, we have created on the order of 50 new bioenergy pathways.

Lignocellulose compounds. We significantly enhanced this category of chemical compounds in MetaCyc, with a focus on natural cellulosic and hemicellulosic polymers [e.g. a $(1{\rightarrow}3)$-β-D-xylan]. We also curated many of the compounds that are routinely used by researchers in the field for characterization of these enzymes (for example, Avicel, p-nitrophenyl-derivatives).

Lignocellulose-degrading enzymes and pathways. We created many pathways that describe the degradation of lignocellulosic compounds, along with the enzymes that are involved. A few of the polymers included in these pathways are cellulose, rhamnogalacturonan, (1,3)-β-D-xylan, (1,4)-β-D-xylan, L-arabinan, xyloglucan, L-arabinan, and glucuronoarabinoxylan.

New display for Glycan degradation pathways. The traditional pathway layout, which shows successive ordered reactions, is often not suitable for describing glycan degradation pathways, where a number of enzymes act at the same time at different locations of the polymer. To help illustrate this type of degradation pathways we implemented a new display tool by integrating the GylcanBuilder software into Pathway Tools. Using this tool, polymer structures are shown using symbols for the glycan monomers, and the location of sites cleaved by different enzymes is shown using color-coded arrows pointing to the bonds within the polymer structure that are cleaved.

Biofuel production. Under this category we created pathways for the biosynthesis of bioenergy-relevant chemical products, such as alcohols, isoprenoids, alkanes, algal oil compounds, etc.

Hydrogen-production. We created eight different hydrogen production pathways, describing the process in different organisms, encompassing all known types of hydrogenase enzymes.

Engineered pathways. Since many of the important bioenergy-related pathways are engineered pathways that combine enzymes from multiple organisms within a single host, we added to MetaCyc the ability to describe such pathways, and curated a number of relevant engineered pathways.

References

[1] The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nuc. Acids Res. Jan 2014 42:D459-71.