Computational Design of Bioenergy-Related Metabolic Pathways

Mario Latendresse* (latendre@ai.sri.com), Mike Travers, and **Peter Karp**

SRI International, Menlo Park, California

http://metacyc.org, http://brg.ai.sri.com/ptools

Project Goals

The general goal of our project is to support computational design of metabolic pathways for metabolic engineering applications. Our approach combines the following elements. (1) Users will specify a target metabolite, a feedstock compound, and other constraints on their design problem. (2) A pathway search algorithm will construct alternative pathways by combining reactions from the MetaCyc database. MetaCyc is a highly curated multi-organism metabolic database that currently contains more than 9500 metabolic reactions. The search algorithm will rank pathways according to multiple criteria. (3) Users will view the output of the pathway search algorithm using a graphical interface that facilitates user comprehension and evaluation of the pathways. The resulting software will be an add-on module to SRI's Pathway Tools software.

Pathway Search

We have implemented an initial version of a graph search algorithm to design novel pathways, given a source and target metabolite. This software is capable of integrating a variety of metrics and filters that affect the search. Metrics implemented to date include atom conservation, molecular similarity, avoiding certain molecules, penalties for using a reaction that is outside of the taxonomic range of the target organism, and penalties for generating or consuming certain kinds of side products/reactants. The search engine is capable of using reactions from any PGDB including MetaCyc. The results of the search can be displayed using the web-based reaction atom-mapping viewer, a new capability of Pathway Tools that can display the trajectories of individual atoms through a sequence of reactions.

Future planned work includes adding new search metrics (such as thermodynamic constraints), and an enhanced user interface that allows scientists to interactively view and control the search process.

Automated Computation of Reaction Atom Mappings

An atom mapping describes explicitly the one to one transfer of each atom from the reactants of a reaction to its products. A reaction might have several chemically valid mappings due to the symmetries of reactants and products or for other reasons. These mappings allow the computation of the flow of essential atoms from source to target metabolites in a pathway. Note that a typical reaction equation used by biologists or chemists does not describe the atom mapping, as such details are difficult to provide and might be overwhelming to the reader. When provided, atom mappings are typically described as supplementary data for each reaction.

Some previous work on computing atom mappings does not compute all possible atom mappings and is therefore incomplete. For example, the KEGG RPAIR database describes only one atom mapping per reaction. As far as we know, all computational approaches to atom mapping published so far, are combined with a post-processing step involving manual curation. That is, a scientist reviews the computed atom mappings for possible errors and appropriate corrections are applied when necessary. But the necessary corrections are not applied to the computational approach itself to avoid future computed errors.

In the approach we have taken, all the chemically valid mappings of each reaction of MetaCyc are computed. Moreover, we aim to have a computational approach that does not require manual post-processing. Technically, our approach is based on mixed integer linear programming: for each reaction, a linear program is generated from all possible valid mappings of the reaction where the objective is to minimize the sum of the costs of the bonds broken and made. We currently use bond costs that have been determined by a chemist. A linear solver solves the linear program, giving all possible optimal solutions, that is, all possible correct mappings for one reaction. This technique has been applied to almost all the reactions of MetaCyc. It is more computationally efficient than all other known techniques published so far.

The computation of correct mappings depends on the bond costs used. In the near future we intend to systematically validate these bond costs using a linear program based on a sample of correct mappings (positive examples) and incorrect mappings (negative examples). We also plan to include atom mappings in future versions of MetaCyc and of other Pathway/Genome Databases.

This work was funded by the Department of Energy under grant DE-SC0004878.