BIOINFORMATICS APPLICATIONS NOTE Vol. 27 no. 11 2011, pages 1581–1582
doi:10.1093/bioinformatics/btr160

Systems biology

Advance Access publication March 30, 2011

MRSD: a web server for Metabolic Route Search and Design

Deguo Xia[1,†], Haoran Zheng[1,2,*,†], Zhiqiang Liu[1], Guisheng Li[1], Jinlong Li[1], Jiong Hong[2] and Kai Zhao[1]

[1]School of Computer Science and Technology and [2]Department of Systems Biology, University of Science and Technology of China, Hefei 230026, PR China

Associate Editor: Trey Ideker

ABSTRACT

Summary: We present a tool called MRSD (Metabolic Route Search and Design) to search and design routes based on the weighted compound transform diagraph. The search submodule returns routes between a source and product compound within seconds in the network of one or multiple organisms based on data from KEGG. The design submodule designs a route from an appointed compound in an interactive mode. The two complementary functions, Metabolic Route Search and Design, can be broadly used in biosynthesis, bio-pharmaceuticals and the other related fields.

Availability: bioinfo.ustc.edu.cn/softwares/MRSD/

Contact: hrzheng@ustc.edu.cn

Supplementary information: Supplementary data are available at the *Bioinformatics* online.

Received and revised on December 15, 2010; accepted on February 27, 2011

1 INTRODUCTION

With the advent of post-genomic era and increasing use of computational techniques, great development has taken place in computational prediction in metabolic pathways which has numerous applications in systems biology and synthetic biology, for example, the design of tracer or knockout experiments (Blum and Kohlbacher, 2008). The primary method adopted by current tools for studying the metabolic pathways is to use the traditional algorithms of paths searching to compute the metabolic routes. Related works include Rahnuma (Mithani *et al*., 2009), MetaRoute (Blum and Kohlbacher, 2008), Pathway Hunter Tool (Rahman *et al*., 2005), FMM (Chou *et al*., 2009) and so on. Unlike previous methods, our application takes into account the relative conservation of reactions that can be measured by the reaction's frequency which is equals to the number of organisms that the reaction exists. Also, these tools either have no intermediates constraint or do not give a global and visual view of the search result. Furthermore, the user may want to design a route manually when there is no route to fulfil his needs in the search results; unfortunately, there are currently no tools provide such function, including the tools mentioned above.

Here, we described a tool called MRSD (Metabolic Route Search and Design) for searching and designing the metabolic routes based on the different organisms' metabolic networks in KEGG. The search function enables users to perform route search in a weighted compound transforming graph in which the weight of the edge is a relevance function about the reaction's frequency (Supplementary Material), in single organisms or groups of organisms with several constraints (such as using the intermediate metabolites or not). The new metabolic routes in the search results can be broadly used in synthetic biology. The design function can be employed when there is no metabolic route that satisfied the users' need in the results computed by search function, or in another situation that the user only knows or only cares for the metabolic route from a specific source or end up with a specific product compound. The design function, which produces user's interested routes in an interactive mode, can be used in biopharmaceuticals and other related fields.

2 METHODS

The MRSD functions based on the compound transforming graph which is a weighted digraph representing a metabolic network. The vertices in the graph represent the compounds, and the edges are the reactions connecting the substrates and products of the reactions. The head vertex acts as substrate in the reaction represented by the edge and the tail vertex acts as product. In this graph, one edge may represent various reactions because different reactions may have the same subset of reactants and products. In the graph of an individual organism, the weight of each edge is equal to the sum of frequencies of reactions that produce the corresponding product with the reactant. For a combined graph of various organisms, the weight of each edge is the sum frequencies of that in these organisms without regard to the duplicate of the reactions which may exits in more than one organism. The reason for this strategy is that we consider these reactions which exist in most of organisms have higher priorities than the others which only appear in few organisms.

MRSD contains two major functions: MRS (Metabolic Route Search) and MRD (Metabolic Route Design). On one hand, if you want to produce a known product from a known source, then the MRS can be employed. On the other hand, if only one metabolite is known and you want to know how to produce this metabolite or which metabolites can be produced by this one, then the MRD can be used in this situation. In addition, if the result routes of MRS do not fulfil your needs exactly, you can use MRD to design a route to introduce some minority modifies on the routes provided by MRS, which means that you can optimize the route by MRD according to the routes computed by MRS (Supplementary Material).

MRS computes pathways between the source metabolite and the product metabolite by performing the k-shortest algorithm (Eppstein, 1998) on the weighted compound transform graph. MRS finds these shortest routes with maximum weight. Under these constraints, we convert the weighted graph to a probability weighted graph in which the edge with higher weight gets lower value in order to satisfied Eppstein's algorithm (Supplementary Material). The result routes are a connected sequence of distinct reactions in which the product of one reaction acts as a substrate in the next reaction.

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

MRD designs a route in an interactive mode according to the assigned metabolite and the users' choices. All the corresponding candidate metabolites are provided by MRD based on the user-preferred compound in each step. The computation of the candidates can be introduced to find the neighbour nodes of the current compound in the compound transfer graph.

3 APPLICATIONS

3.1 MRS

The main application of the MRS is exploring the appointed metabolic network for all routes transforming the source metabolite to the product metabolite, which can be extensively used in synthetic biology and other related fields.

There are several settings to be defined before starting a search: the source metabolite, product metabolite and the selected networks are required; and the remaining settings including whether the search using the intermediate metabolites or not are optional. The user can specify a list of intermediate metabolites separated by vertical line. The user can perform a route search on the metabolic network of one bacterium or on the combined network of a group of bacteria. Furthermore, it is possible for the user to search all routes in the super bacteria' of the KEGG bacteria database. The 'super bacteria' is a virtual bacterium which is combined from all the bacteria in the KEGG bacteria database and contains all the reactions and compounds in this database. The user can also search the chemosynthesis route between two compounds represented by InChi string (Supplementary Material) by changing the interface type.

We perform a search on the metabolic network of the 'super-bacteria' to find a route from glucose to 1-butanol (Atsumi *et al*., 2008). This is an interesting and useful route, which can facilitate strain improvement for specificity and productivity, as its production 1-butanol offers many advantages as a substitute for gasoline because of higher energy content and higher hydrophobicity. We just reset the maximum number of routes to 30 and leave other constraints blank to use the default values. As shown in the Figure 1 (A2), the relevant routes from source compound to product compound construct a complex metabolic network. We also noticed that the highlighted route is exactly in accordance with the route designed by Atsumi. Except the route mentioned above, MRS also provides other routes that may be more interesting and useful for researchers.

3.2 MRD

Consider you only know a product (source) compound and you want to design a metabolic route that ends with (starts from) this compound; in this case you may find it useful to employ the MRD. A user can enter a query compound into the MRD by name, KEGG ID, formula, directly paste or enter an InChi string. The MRD provides all possible candidates in KEGG according to the input compound. The user can choose any of the predicted products to continue the prediction recurrently. If there are no candidates available or the user has found the route he wanted, the process ends.

As an example, we design a route to produce succinate. After we set the product compound and the design direction, the system provides all possible candidates compound which includes succinyl coenzyme A, trans-butenedioic acid, etc. Then we select succinyl coenzyme A and continue. Few steps later, we can get the designed route shown in Figure 1B, which is part of the tricarboxylic acid cycle (TCA) apparently.

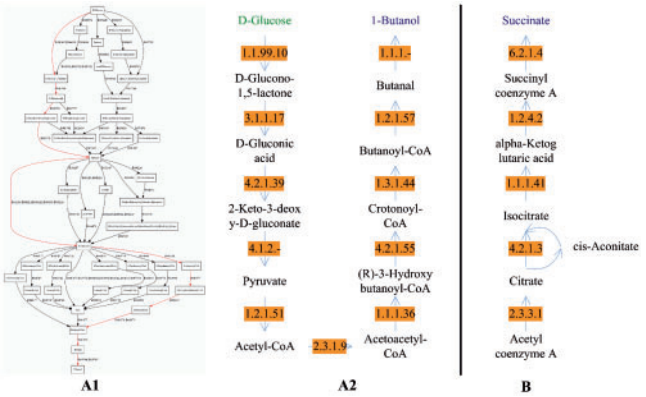

Fig. 1. (**A1**) The shortest routes between D-glucose and 1-butanol in the metabolic network of the 'super-bacteria'. (**A2**) The details of the highlighted route. (**B**) The designed route to produce succinate. Enzymatic reactions are drawn in orange.

4 CONCLUSIONS

The MRSD is a user-friendly and versatile tool that can be employed in many different applications including predicting biosynthetic or catabolic routes, biopharmaceuticals and other related fields through the separate or complementary use of MRS and MRD. MRS performs a route searching in the weighted compound transforming graph which take into account of the statistical property of the reactions, with several constraints considered to produce more convincible results. MRD is a new and useful tool to assist user in designing a route interactively. MRS and MRD can complement each other in helping researchers to continue their work conveniently.

ACKNOWLEDGEMENTS

The authors would like to thank Zhou Hong for advice about editing the article.

Funding: National Key Technologies R&D Program (2011CB910200).

Conflict of Interest: none declared.

REFERENCES

Atsumi,S. *et al.* (2008) Metabolic engineering of Escherichia coli for 1-butanol production. *Metab. Eng.*, **10**, 305–311.

Blum,T. and Kohlbacher,O. (2008) MetaRoute: fast search for relevant metabolic routes for interactive network navigation and visualization. *Bioinformatics*, **24**, 2108–2109.

Chou,C.H. *et al.* (2009) FMM: a web server for metabolic pathway reconstruction and comparative analysis. *Nucleic Acids Res.*, **37**, W129–W134.

Eppstein,D. (1998) Finding the k shortest paths. *SIAM J. Comput.*, **28**, 652–673.

Mithani,A. *et al.* (2009) Rahnuma: hypergraph-based tool for metabolic pathway prediction and network comparison. *Bioinformatics*, **25**, 1831–1832.

Rahman,S.A. *et al.* (2005) Metabolic pathway analysis web service (Pathway Hunter Tool at CUBIC). *Bioinformatics*, **21**, 1189–1193.