DOI 10.18699/SBB-2022-306

Software package for retrosynthesis-based prediction of metabolic pathways

Baboshin M.[1]*, Shupletsov M.[1, 2], Golubeva L.[1]

[1] *Joint Stock Company "Ajinomoto-Genetika Research Institute", Moscow, Russia*

[2] *Moscow State University, Moscow, Russia*

* *Mikhail_Baboshin@agri.ru*

Key words: retrosynthesis, bioretrosynthesis, metabolic pathways

Motivation and Aim: Finding of metabolic pathways for synthesis of a target compound is a common task in metabolic engineering. Retrosynthesis became a popular approach for seeking pathways in recent years. The method is backward trace from target compound to available metabolites, that models enzyme promiscuity and prediction of hypothetical reactions absent in databases by application of so-called generalized reaction rules [1]. While several retrosynthesis tools were reported [2], the software have shortcomings, such as limited availability, instability, inflexibility, solution loss and proposal of false solutions, lack of appropriate pathway enumeration or thermodynamic analysis. The above shortcomings of current software hinder the practical application of the method. Therefore, more efficient computational tool is required for implementation of retrosynthesis in metabolic engineering. The aims of this work were to develop and test the software package capable to support not only finding of biosynthetic pathway, but their preliminary characterization to simplify the choice of more attractive one, and to create an updated database of generalized reaction rules that takes into account new data on biochemical reactions in living organisms.

Results: Given that the database of generalized reaction rules, which is the basis of the retrobiosynthetic approach, depends on completeness of biochemical knowledge, we generated in-house database of reaction rules. Validation of the reaction rules was performed by reproduction of reference reactions (i.e. reactions used for rule generation) and by comparison of global metabolic space with that generated using reaction rules from the database RetroRules [1]. The newly constructed reaction rules reproduce 97 % of their reference reactions. The metabolic spaces generated using the new rules and RetroRules constructed on the same set of reference reactions were close to each other (about 83 % of overlapping), which indicates almost the same quality of reaction rule construction. The entire set of our in-house reaction rules resulted in a larger metabolic space than RetroRules, probably, since the latter were constructed using a less complete data on biochemical reactions.

The software package is organized as a workflow consisting of both in-house components and external software/data and includes following steps: 1) net generation; 2) net processing; 3) pathway enumeration; and 4) pathway ranking.

Net generation means generation of metabolic net containing desired pathways. Restriction of net growth is required to prevent combinatorial explosion. We use constraining by metabolite database as the primary option, but other types of constraints are also possible.

Net processing includes construction of reaction equations, removal of the reactions that do not connect metabolites of the chassis strain with the target compound, and deletion

of inappropriate reactions (stoichiometrically and thermodynamically inconsistent reactions are removed, some other filters may be optionally applied).
Pathway enumeration, i.e. composing of pathways from the list of reactions, is performed by elementary mode analysis. The obtained pathways are added one-by-one to the model of the chassis strain to check possibility of steady-state flux and calculate theoretical yields by flux balance analysis. Pathway ranking is sorting of pathways by fitness criteria to select a pathway for implementation. First, pathways are sorted by number of hypothetical reactions. Then, in each group of pathways with equal number of hypothetical reactions, pathways are sorted by theoretical yield of target product.
To test quality of metabolic net generation, we apply our program to 20 target products (2-amino-1,3-propanediol, 3-methylbutanol, 1,4-butanediol, 2,5-dihydroxybenzoate, benzyl alcohol, β-carotene, muconate, glutarate, lycopene, mesaconate, naringenin, N-methylpyrrolinium, 4-hydroxystyrene, piceatannol, pinocembrin, protopanaxadiol, styrene, terephthalate, vanillin, violacein) with pathways reported in literature that were used earlier for testing of RetroPath2 [2]. We were able to find reference pathway in all 20 cases, compared with 13 out of 20 cases reported for RetroPath2. The result shows that simple constraining of net growth by metabolites may be more efficient than sophisticated constraining methods, such as that implemented in RetroPath2.
All four steps of the workflow were tested on 8 target products. Reference pathway was present in the final list of pathways in all cases, but only in 4 out of 8 cases it matched with the first rank pathway. This, probably, indicates that suboptimal synthetic pathways are often implemented in producer strains.
Conclusion: A software package for retrosynthesis-based prediction of metabolic pathways was created; testing on 20 target products confirmed its applicability for finding the pathways with promising properties. This software, together with an updated database of generalized reaction rules, is an efficient tool for *in silico* prediction of metabolic pathways.

References

1. Duigou T. et al. RetroRules: a database of reaction rules for engineering biology. *Nucleic Acids Res*. 2018;47:D1229-D1235.
2. Koch M. et al. Reinforcement learning for bioretrosynthesis. *ACS Synth Biol*. 2020;9:157-168.