拡張代謝パスウェイデータベース「M-path」の構築

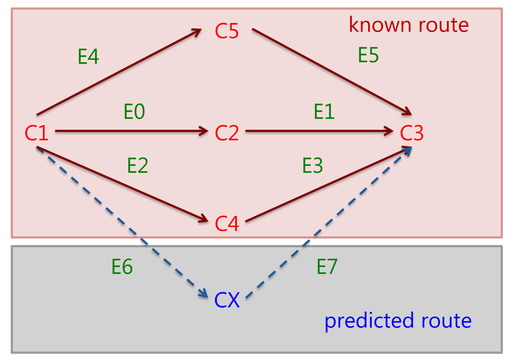
小川哲平[1], 牧口大旭[1], 荒木通啓[2,3]

[1]三井情報株式会社
[2]京都大学大学院 医学研究科
[3]神戸大学大学院 科学技術イノベーション研究科

Abstract

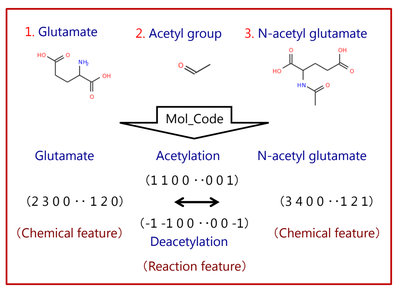
Recent developments in synthetic biology and metabolic engineering have led to the construction of synthetic metabolic pathways for efficient production of various natural and non-natural chemicals. We introduce the M-path database and viewer to explore latent synthetic metabolic pathways of putative compounds and reactions. In the current version of M-path, the set of reaction is increased up to 20,000 by using a manually curated BRENDA database and compound data from PubChem are regularly updated, which will increase a chance to find unknown metabolic pathways with diverse chemical and enzymatic information.

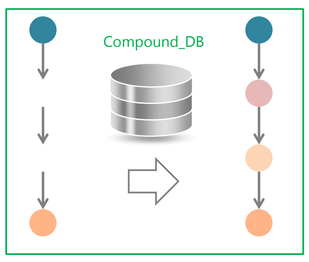
Concept

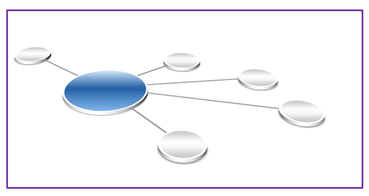


The concept of M-path is predicting "unknown" route (ex. C1 -> CX -> C3) from start compound (C1) to last compound (C3) based on "known" enzyme reactions (E6 and E7).
It can also be possible to predict the synthetic route that has not been found to synthesize yet.

Method & Result







Calculate pathway feature

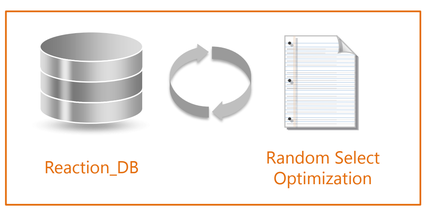
Solve linear problem.

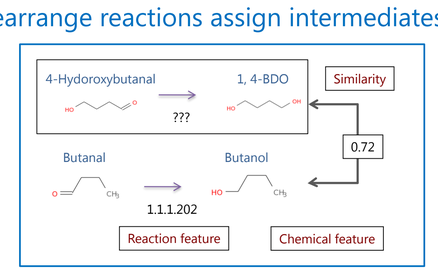
Output pathway candidates

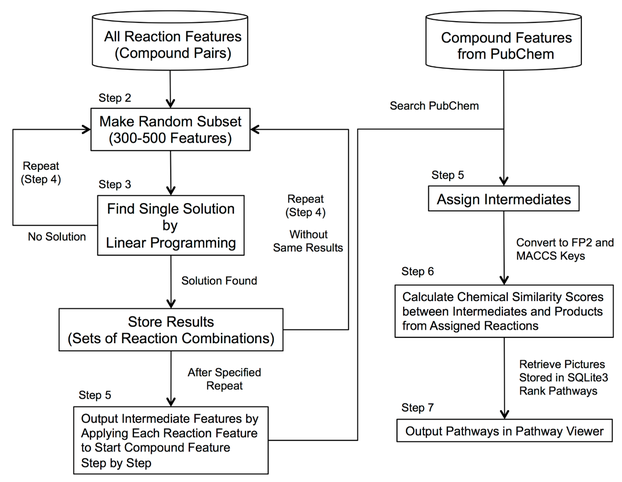
Step 1 **Step 2** **Step 3** **Step 4** **Step 5**

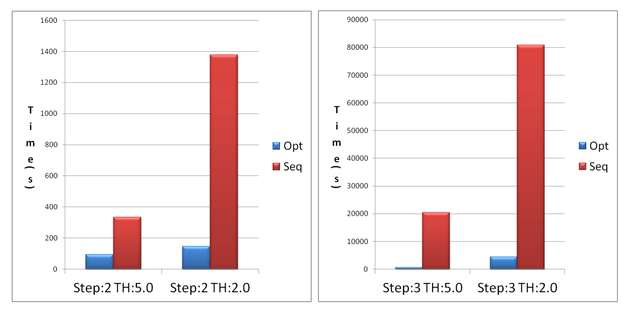
Make random subset from reaction feature

Rearrange reactions assign intermediates







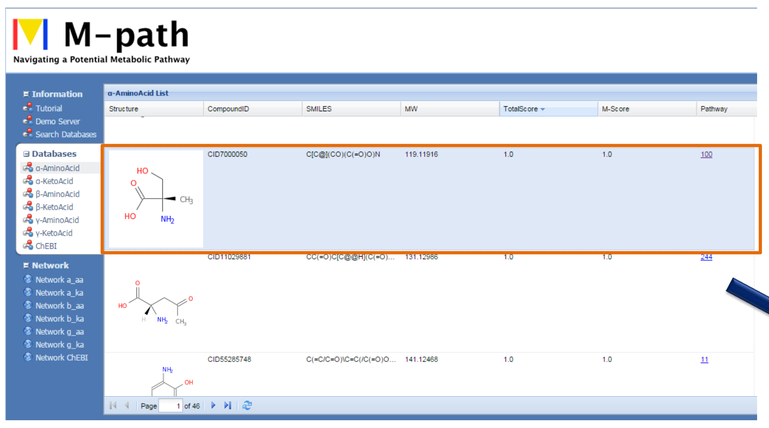


Methodology of M-path making:
M-path makes use of chemical structure information and enzymatic reaction rules, and is performed as shown scheme.

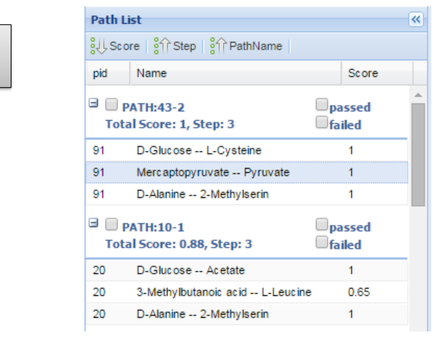
Performance of solving pathway:
Seq : extending pathways step by step until reaching target.
Opt : solving pathways by using the solution to a linear programming problem.

M-path

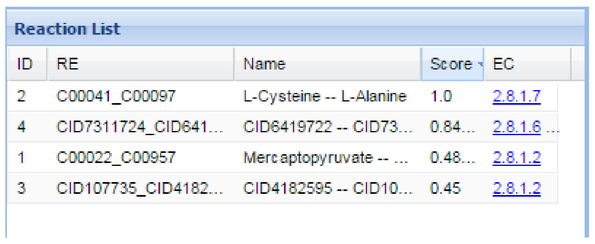
Sample reaction: D-glucose -> L-2-Methylserine



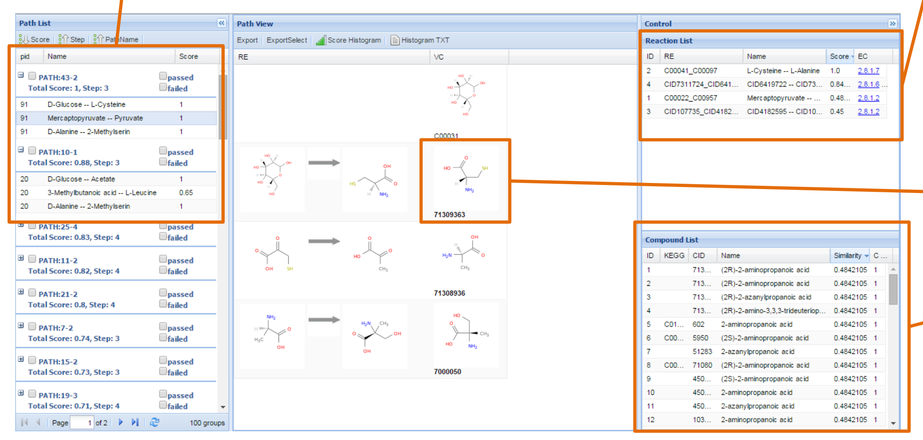
Compound viewer:
Compound viewer shows chemical feature of each compounds and the score of the pathway candidate with highest score.

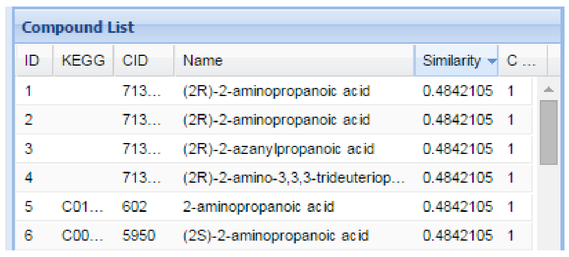


Path List:
The predicted route is shown in Path List. Each step is scored based on the similarity to the known reaction. Score 1 shows that predicted route is the same reaction as known reaction, and score < 1 shows the similar to it. Similarity is calculated based on Tanimoto score of Substrate/Product.



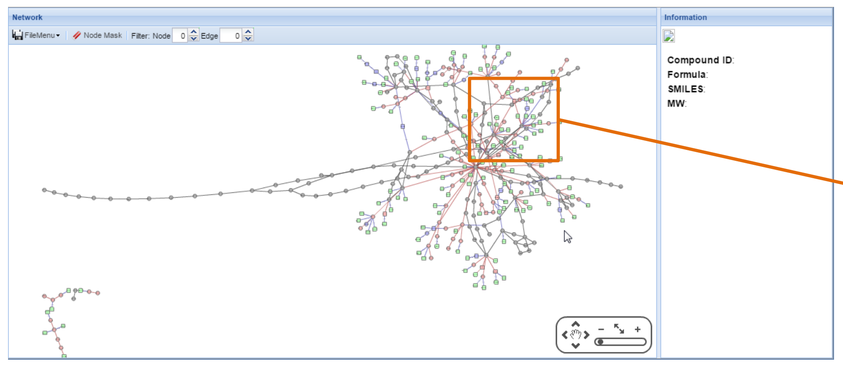
Reaction List:
The reaction which has same substrate-product pair as predicted reaction is scored 1, then similar reaction is scored based on the similarity of chemical structure of its substrate and product.



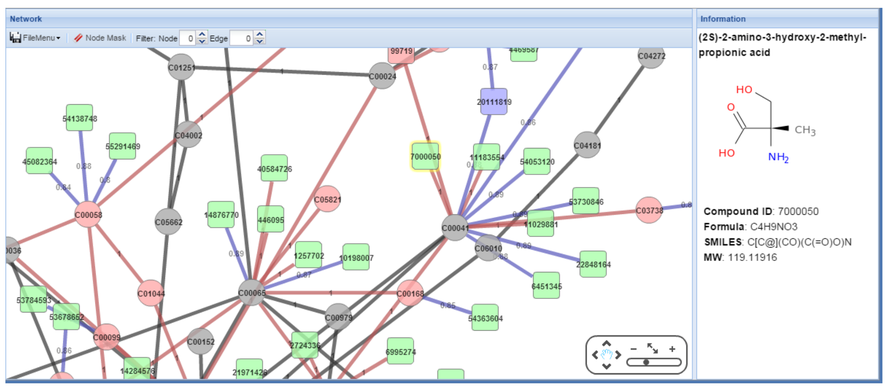


Compound List:
The predicted compound is shown in Path View and its similar compounds were shown in Compound List.

Pathway viewer:
Pathway viewer shows potential metabolic pathways including reaction data and the transition of chemical structures and reaction orders from top to bottom for each pathway.



Network viewer:
Network Viewer is an integrated network using known amino acid network data, compounds and reactions from the pathway candidates with M-score > 0.8 for all data.



Network Viewer Information:
The core L-amino acid biosynthetic network annotated by KEGG (grey). Naturally reachable (PATH_0) compounds are shown as pink circles. Blue edges are putative reaction steps predicted by M-path. The green squares show putative L-amino acids not found in KEGG, with blue squares as their intermediates such as keto acids.

References

(1) M. Araki, R. S. Cox III, H. Makiguchi, T. Ogawa,T. Taniguchi, K. Miyaoku, M. Nakatsui, K. Y. Hara and A. Kondo, *Bioinformatics* **2015**, *31*, 905-911.

(2) http://bp.scitec.kobe-u.ac.jp/m-path/aa/

(3) http://bp.scitec.kobe-u.ac.jp/m-path/db/