

Metabolic Pathway Discovery via Meta-graph Construction and Search

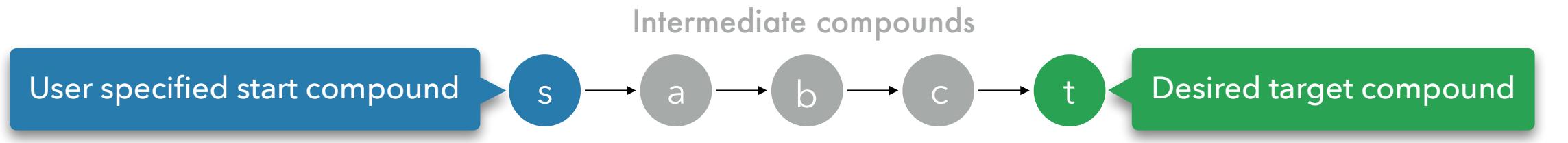
Prudhvi Boyapalli, Sarah Kim, Mark Moll, Lydia Kavraki

Department of Computer Science, Rice University

Metabolic Pathfinding

Metabolic pathways are a series of chemical reactions, and the intermediate compounds produced by those reactions, which connect two compounds.

The metabolic pathfinding problem is to find all biologically feasible pathways from a given start compound to a given target compound. To ensure that the discovered pathways are biologically significant, a threshold is enforced so that at least a minimum number of carbon atoms from the start compound are preserved in the target compound.

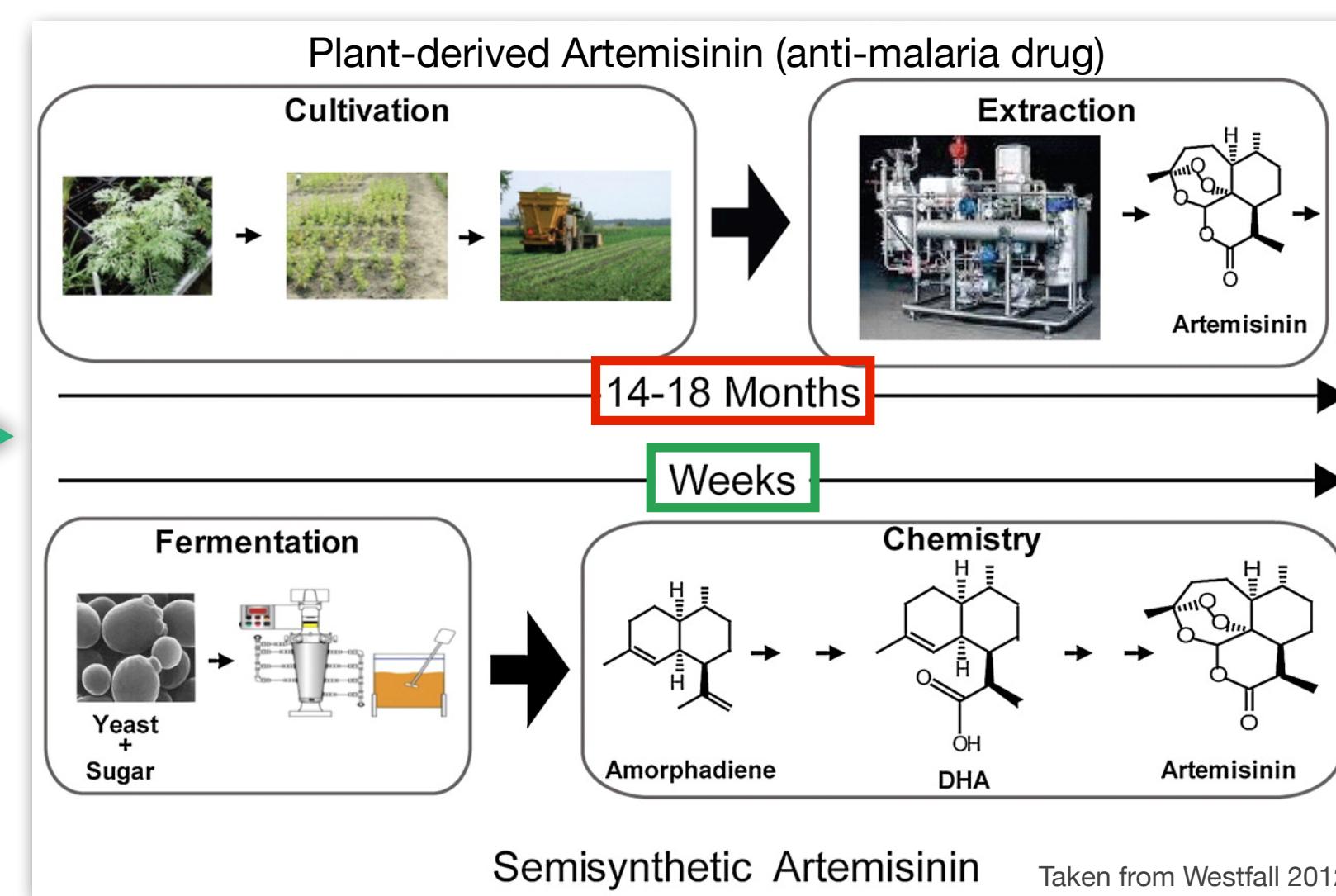


Motivation

- Finding novel metabolic pathways, possibly spanning multiple species, has important applications in fields such as metabolic engineering, metabolic network analysis, and metabolic network reconstruction [1].
- The ability to discover metabolic pathways computationally, as opposed to experimentally, can significantly speed up the development process for drug / biofuel synthesis

Example: Impact on Metabolic Engineering

Conventional production of Artemisinin can take between 14-18 months. However, using a synthetic process to produce the same compound takes a matter of weeks. Thus, the ability to find metabolic pathways between compounds, such as Amorphadiene and Artemisinin, which can then be utilized to provide alternative methods of production, can significantly increase the producibility of a desired compound.



Contributions

I present an algorithm for finding biologically important pathways in metabolic networks through the search of a graph containing atom mapping data from metabolic pathway databases such as KEGG, MetaCyc, and BRENDA. While prior work is limited to simple linear pathways or certain types of branched pathways, my algorithm builds a concise representation of all branched pathways, enabling the discovery of complex, yet biophysically feasible pathways.

The algorithm works in two phases. In the **meta-graph creation phase**, metabolic reaction data is parsed and used to build an atom mapping graph which tracks the movement of atoms from the source compound, through various reactions. The subsequent **meta-graph search phase** operates on the meta-graph and utilizes standard graph search algorithms to find both linear and branched pathways.

In addition, I've created a web-based application for visualizing and interacting with the discovered metabolic pathways.

Preliminary Results

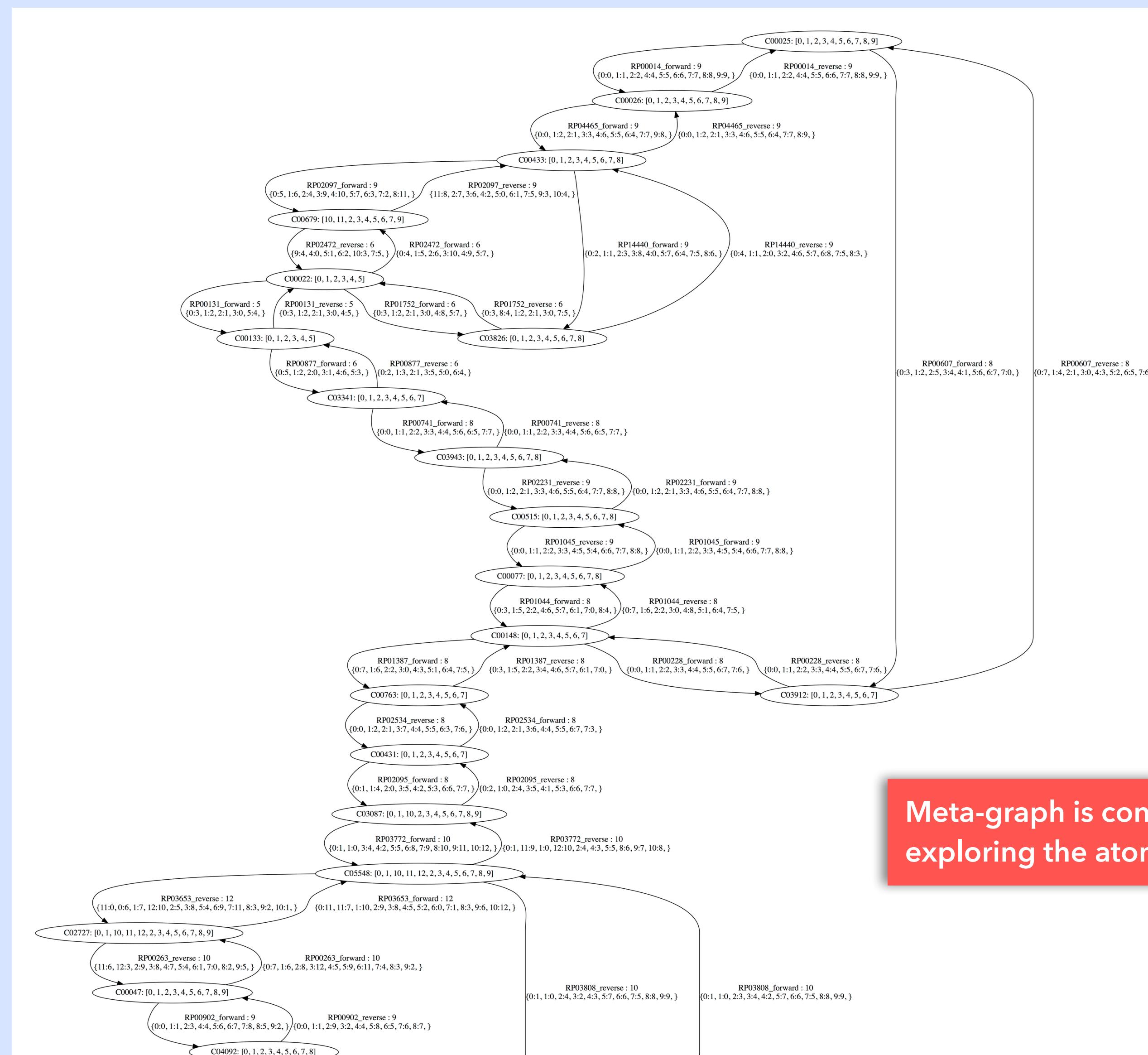
The algorithm has been tested against pathways found by existing methods, specifically the Linear Pathfinding with Atom Tracking (LPAT) algorithm [1]. When working with small subsets of the reaction data, the meta-graph construction and search method is capable of finding the same pathways as those discovered by LPAT.

Using only a small subset of the reaction data, the following paths were found in the meta-graph between the start compound Pyruvate (C00022) and the target compound L-Lysine (C00047):

- [C00022,C00079,C00433,C00026,C00025,C03912,C00148,C00763,C00431,C03087,C05548,C02727,C00047]
- [C00022,C003826,C00433,C00026,C00025,C03912,C00148,C00763,C00431,C03087,C05548,C02727,C00047]
- [C00022,C00133,C03341,C03943,C00515,C00077,C00148,C00763,C00431,C03087,C05548,C02727,C00047]

Further development of the algorithm implementation is needed before it can be run on the entire database of reaction information. Thus far, the output of the algorithm has only been checked for the existence of known linear pathways. Though the ability to find branched pathways exists by design in the presented method, extensive testing has not yet been done to compare discovered branched pathways with those found by the existing algorithms.

[1] A. P. Heath, G. N. Bennett, and L. E. Kavraki, "An Algorithm for Efficient Identification of Branched Metabolic Pathways," *Journal of Computational Biology*, vol. 18, no. 11, pp. 1575–1597, Nov. 2011.



Atom Mapping Graph

Meta-graph is constructed by exploring the atom mapping graph

Algorithm

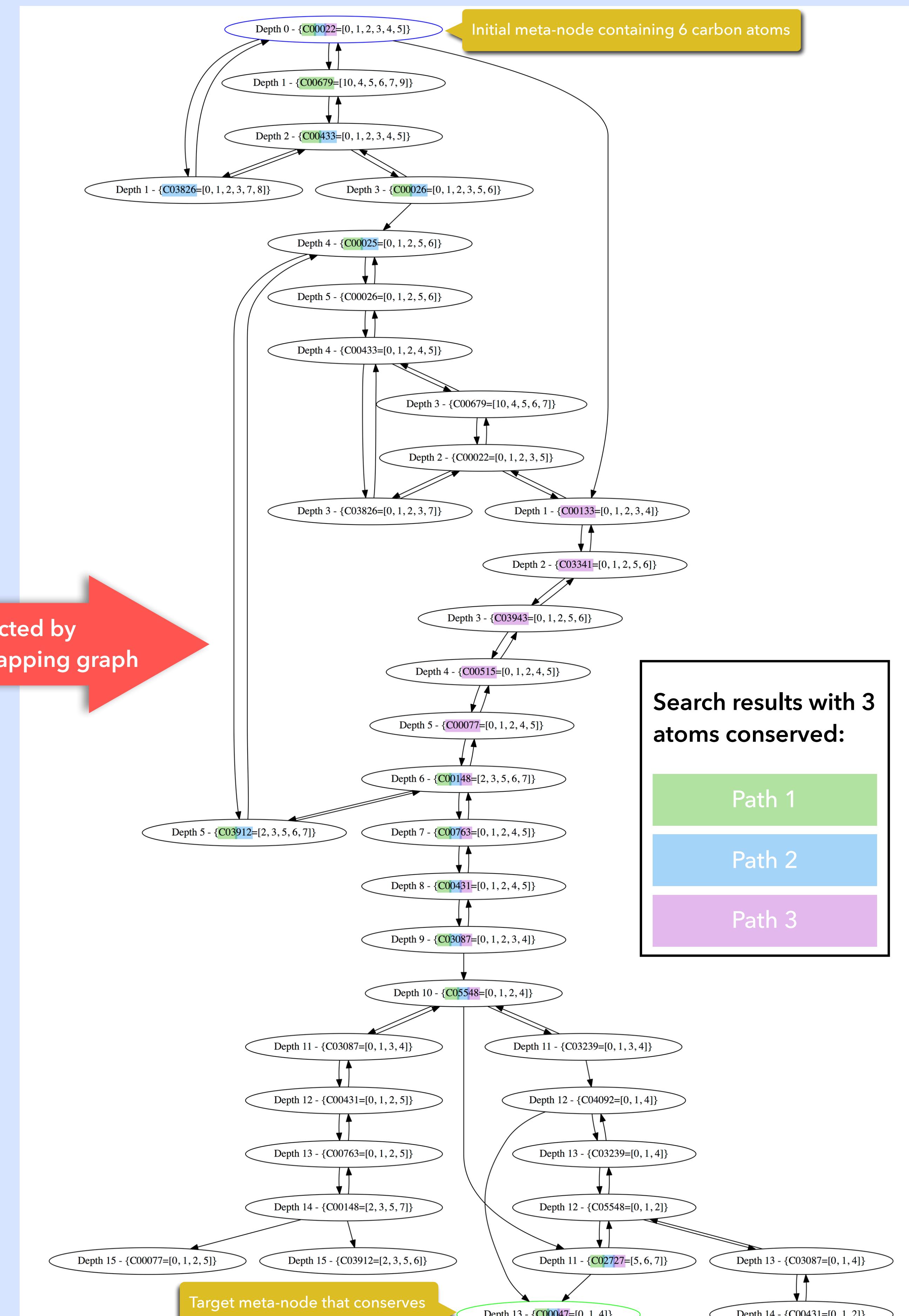
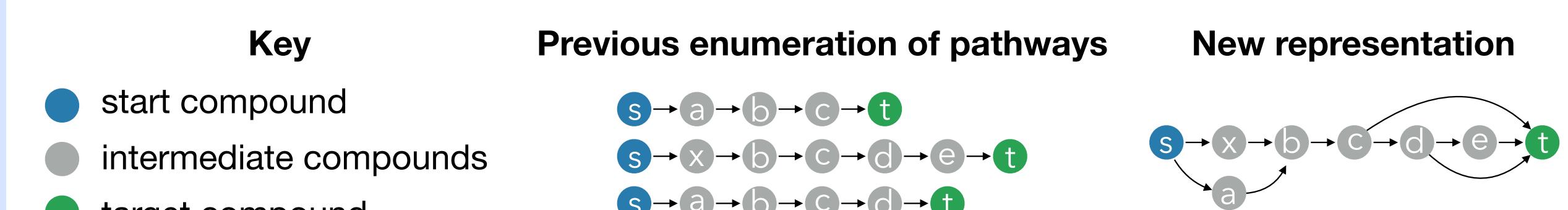
Input: Database of metabolic reaction data parsed from the Kyoto Encyclopedia of Genes and Genomes (KEGG); start compound s ; target compound t ; minimum number of carbon atoms to conserve k

Output: A list of pathways from s to t that conserve at least k carbon atoms.

- The reaction data (or a subset of it) is used to construct an atom mapping graph, am
 - The graph consists of compound nodes connected by reaction pairs, which contain the atom mapping information from the input to output compound
- The meta-graph, mg , is constructed by traversing am
 - The initial meta-node contains the start compound and the atoms which reside there
 - The meta-graph is built by searching for neighbors of existing meta-nodes. A neighboring meta-node is one which contains compounds that are reachable from compounds contained in the parent meta-node
 - The atom mapping graph is explored to find these neighbors and newly found meta-nodes must contain at least k atoms that originated from the start compound
- The meta-graph is searched for paths from the initial meta-node to any meta-nodes containing the target compound with at least k atoms

Improvements Over Prior Works

- More concise representation of pathways
 - No need to enumerate all variations of pathways between two compounds
 - Saves time and space when searching the graph for pathways
- A single algorithm and graph representation is used to discover both linear and branched metabolic pathways



Pathway Visualization

