novoStoic: Pathway design using *de novo* steps through uncharted biochemical spaces

Lin Wang[1*] (lqw5322@psu.edu), Akhil Kumar[2], Chiam Yu Ng[1], and **Costas D. Maranas**[1]

[1]Department of Chemical Engineering, The Pennsylvania State University of State, University Park, PA 16802; [2]The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA 16802

http://www.maranasgroup.com/

Project Goals: The goal of this project is to develop an optimization based pathway design tool that seamlessly integrates known reactions with reaction rules. The software tool (i.e., novoStoic [1]) allows for the incorporation of performance criteria for rank ordering pathway designs, thermodynamic feasibility constraints, limits on the number of novel reaction steps, and restrictions on the organism or class of organisms the reactions are chosen from. In addition, enzymes are flagged that will have to be re-engineered for an altered substrate specificity to enable the novel reaction steps.

Computational pathway design algorithms enumerate potential routes linking the two molecules, while often taking into consideration a multitude of criteria such as shortest route, the minimal number of non-native reactions, thermodynamic feasibility, and enzyme availability. Although most methods rely on the large number of enzymatic reactions catalogued already, there are also increasing number of tools that employ biotransformation rules derived from the existing reactions to discover novel pathways. The latter relies on the remarkable malleability of enzymes to accept a broad range of substrates as well as the potential of protein engineering for enzyme redesign. However, existing retrosynthesis tools generally traverse production routes from a source to a sink metabolite using known enzymes or *de novo* steps by graph-based methods. Generally, important considerations such as blending known transformations with putative steps, complexity of pathway topology, mass conservation, cofactor balance, thermodynamic feasibility, microbial chassis selection, and cost are largely dealt with in a posteriori fashion.

To address the problems in current computational methods, novoStoic [1] (see Figure 1) designs bioconversion routes while simultaneously considering any combination of the aforementioned design criteria. First, we augmented the MetRxn [2] (www.maranasgroup.com/metrxn) repository with a new dataset of elementally balanced reaction operators using the automated CLCA [3] based reaction rule extraction procedure termed rePrime. For each reaction, the rePrime procedure identifies and captures as reaction rules the molecular graph topological changes underpinning the substrate to product graph conversion. A reaction rule is a vector that captures the location of active reaction centers affected by the conversion of substrates to products. The reaction rules and known reactions are then operated upon a mixed integer linear programming (MILP) procedure, novoStoic then identifies a mass-balanced biochemical network that converts a source metabolite to a target while satisfying a multitude of constraints and optimizing the objective function. We demonstrate the use of novoStoic in bypassing steps in existing pathways through putative transformations, assembling complex pathways blending both known and putative steps towards biofuel and biorenewable molecules and postulating ways to biodegrade xenobiotics. novoStoic is available at Maranas group website (https://github.com/maranasgroup). Efforts are currently under way to integrate the tools within KBase.
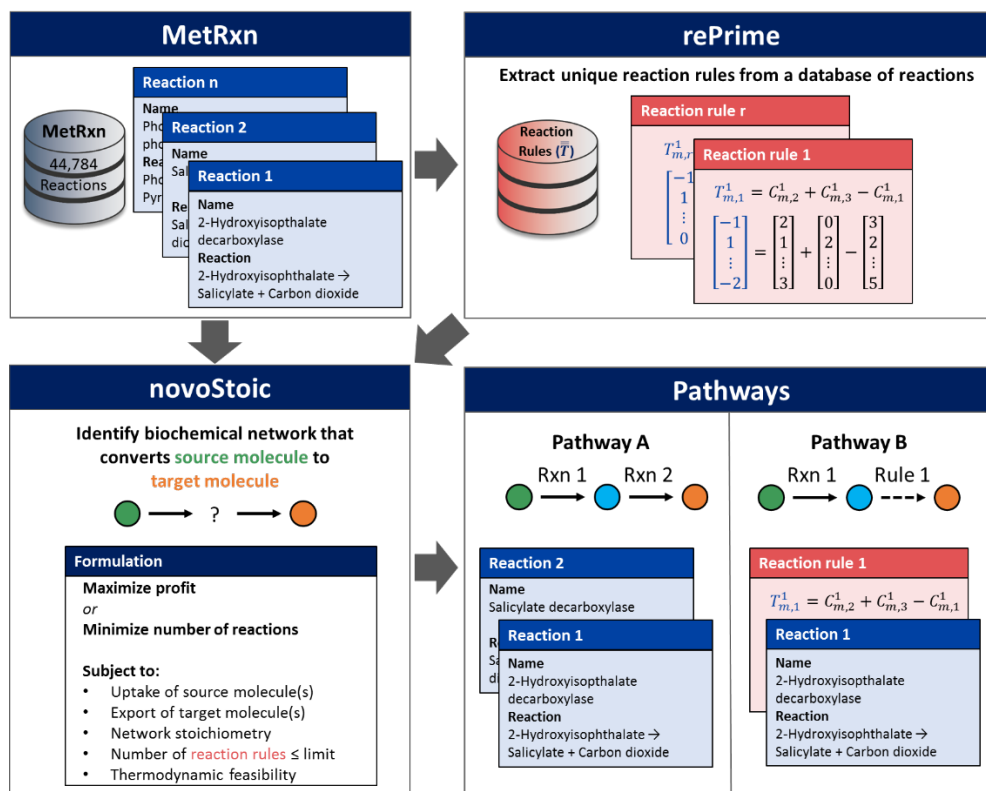
Figure 1: Schematic overview of the rePrime/novoStoic procedure. First, the rePrime procedure is used to pre-process the MetRxn database of reactions (blue boxes) to extract a unique set of reaction rules. The novoStoic procedure is then used to identify a series of intervening reactions and reaction rules that convert source molecule(s) (green circle) into target molecule(s) (orange circle) such that the profit can be maximized or the number of reactions in the pathway can be minimized.

References

[1] Kumar, A., Wang, L., Ng, C. Y., & Maranas, C. D. (2018). Pathway design using de novo steps through uncharted biochemical spaces. *Nature Communications*, *9*(1), 184.

[2] Kumar, A., Suthers, P. F., & Maranas, C. D. (2012). MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC bioinformatics*, 13(1), 6.

[3] Kumar, A., & Maranas, C. D. (2014). CLCA: maximum common molecular substructure queries within the MetRxn database. *Journal of chemical information and modeling*, 54(12), 3417-3438.