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

Background

- Existing retrosynthesis tools generally traverse production routes from a starting to a target metabolite using either known enzymes or retrosynthetic rules but not both.
- novoStoic allows for seamlessly blending known reactions with *de novo* steps to construct balanced and thermodynamically feasible pathways that maximize performance objectives such product yield or conversion profit margin.

Approach

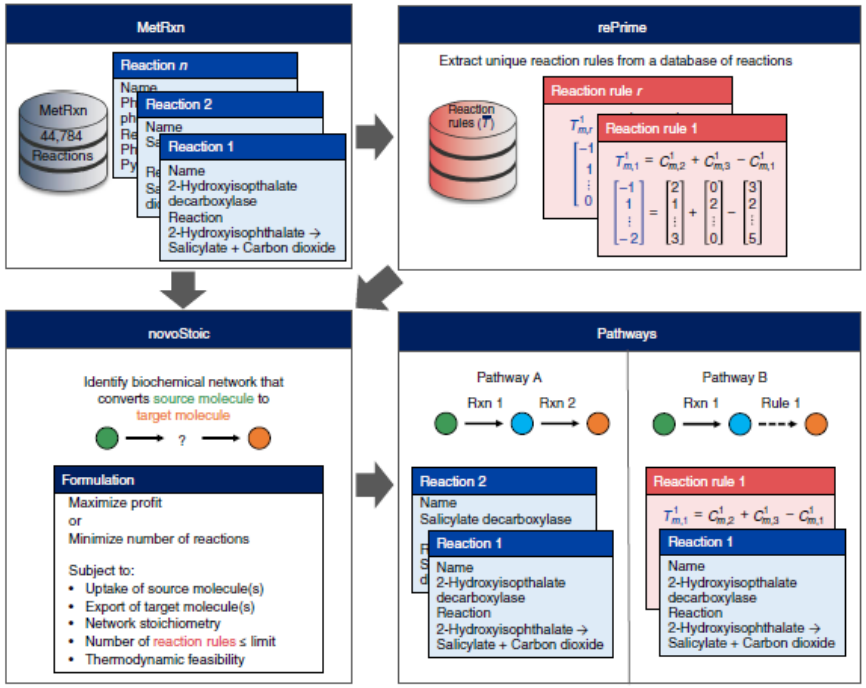
- rePrime uses a prime factoring algorithm to extract reaction rules at different levels of detail by analyzing known reactions catalogued in databases. Reaction rules are codified as operators between the molecular groups undergoing bond changes.
- novoStoic uses mixed-integer linear (MILP) optimization to design cofactor balanced pathways with a pre-specified number of novel steps from starting metabolites to products.

Outcome

- A computational tool novoStoic available for download at www.maranasgroup.com for the design of bioconversion pathways from source to target metabolites.

Significance

- novoStoic can be used for the exhaustive exploration of all possible ways of converting lignocellulosic feedstocks to advanced biofuel molecules. By linking novel steps to catalogued reactions with similar chemistry, enzymes that have to be engineered for altered substrate specificity are identified.
- novoStoic can also prospect the most efficient ways of lignin valorization by funneling one or more lignin monomers towards a single aromatic precursor with the highest value and assess the impact on lignin composition alterations on product yield.



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, R^{λ} (red boxes) at moiety size λ. The reaction rules are derived from the molecular signature C_{mi}^{λ} of each participating metabolites and are captured by T_{mr}^{λ}. 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. Other criteria including the number of reaction rules and thermodynamic feasibility of the pathway can be flexibly incorporated as constraints. By controlling the number of reaction rules, two possible pathways designed by novoStoic are shown in the bottom right panel. Pathway A uses two known reactions (blue boxes) that present in the MetRxn database, whereas pathway B uses a combination of a known reaction (blue box) and a reaction rule (red box) to perform the same conversion.

Kumar, et al., "Pathway design using de novo steps through uncharted biochemical spaces", *Nature Communications*, 2018, 9(1), 184. doi: 10.1038/s41467-017-02362

THE CENTER FOR BIOENERGY INNOVATION