Scylax™: Automated Biochemical Pathway Design

Kovas Palunas, Rudesh Toofanny, Michal Galdzicki, Phillip Yen, Yih-En Andrew Ban, and Alexandre Zanghellini*

Arzeda, 3421 Thorndyke Ave W, Seattle, WA 98119; *Corresponding author: alex.zanghellini@arzeda.com

Arzeda.

Abstract

To maximize the utility of Arzeda's enzyme design platform, we have developed Scylax™, a biochemical pathway design engine. Based on databases of known biochemical reactions and predicted, designable enzyme functionalities, Scylax™ exhaustively finds all pathways that could exist to a given target molecule. Scylax™ then ranks and filters its results with the aid of reaction free energy estimates, atom flow heuristics, and predicted yields, all assisted by pathway clustering.

For many target molecules, natural pathways are suboptimal or may be non-existant. We have found that Scylax™, with the rest of Arzeda's enzyme design and testing pipeline, is able to find efficient pathways that can produce these targets biologically in an economically competitive way.

Databases

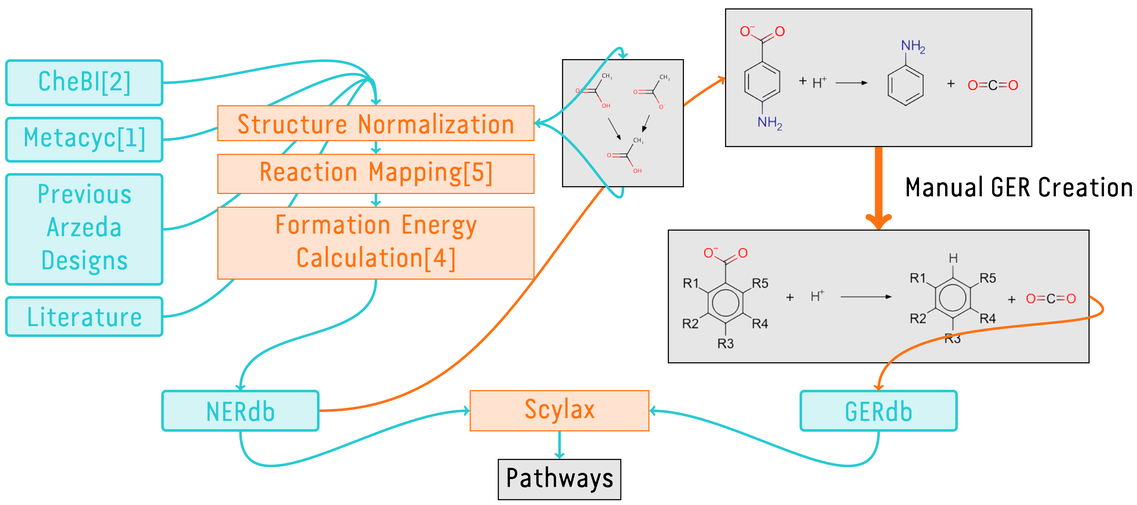

Figure 1: Database construction process. Natural Enzyme Reaction (NER) data is sourced from public and internal data sources and deposited into an internal database (NERdb). Generalized Enzyme Reactions (GERs) are then manually constructed using NER data as inspiration (and are stored in the GERdb). These reactions predict what kinds of chemistry could be done by designed enzymes, allowing for designer pathway discovery.

Sample Result

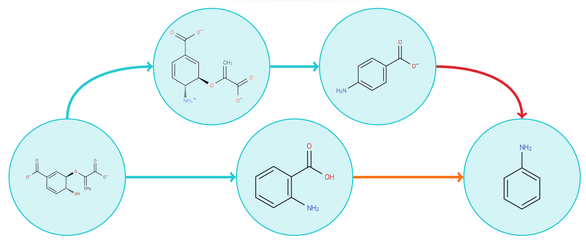

Figure 2: Sample pathways found by Scylax™ through unknown sequence and generalized reactions.

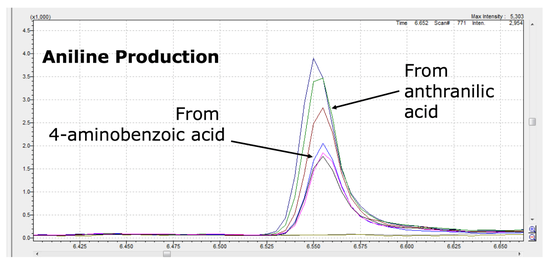

Figure 3: GCMS trace showing production of aniline using designed final step enzymes.

Pathway Enumeration

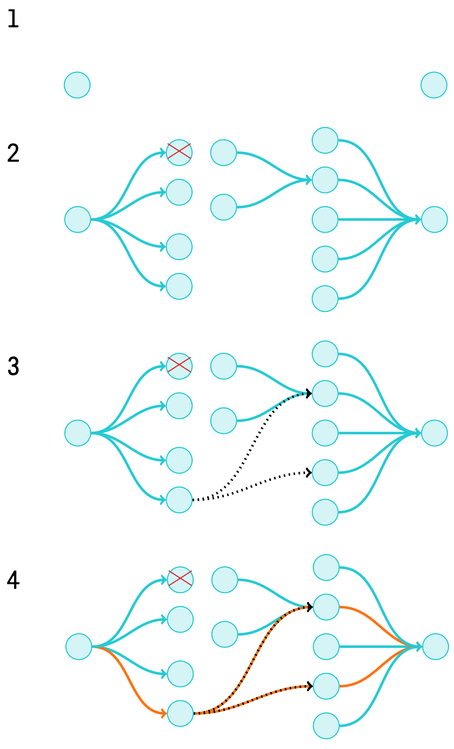

Figure 4: Pathway enumeration process: start/target compound selection (1), filtered enumeration (2), merging (3), final pathway identification (4).

Atom Tracing and Yield

Given atom mapping data, Scylax™ identifies how atoms flow across reactions and pathways:

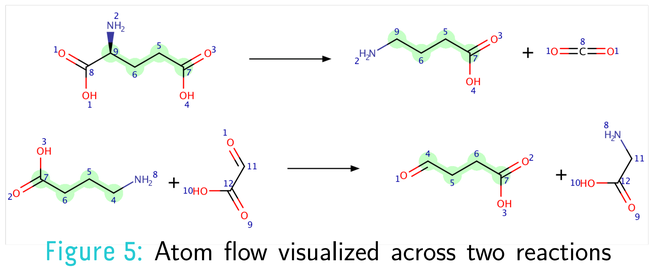

Figure 5: Atom flow visualized across two reactions of a pathway.

- Compounds store canonical atom mappings
- Reactions store canon-to-canon atom mapping information
- Pathways combine reaction mappings to tell which atoms went where
- Ultimate pathway atom transfer used to calculate mass/molar yield
- Yield can be adjusted by metabolic costs for any coreactants/byproducts

All mappings taken from source data, and additionally automapped with Reaction Decoder Tool[5]. All GERs are manually mapped.

Free Energy Estimation

Using precomputed standard compound formation energies ($\Delta_f G'^{\circ}$), concentration (C) dependent free energy changes for all reactions ($\Delta_r G'$) in Scylax™ can be computed as follows:

$$\Delta_r G' = \sum_{p=0}^{\text{prods}} \Delta_f G'^{\circ}_p - \sum_{r=0}^{\text{reacts}} \Delta_f G'^{\circ}_r + RT \ln\left(\frac{\prod_{p=0}^{\text{prods}} C_p}{\prod_{r=0}^{\text{reacts}} C_r}\right),$$

This equation gives a numerical value for reaction driving force, which is used to quantify a pathway's ability to carry flux efficiently.

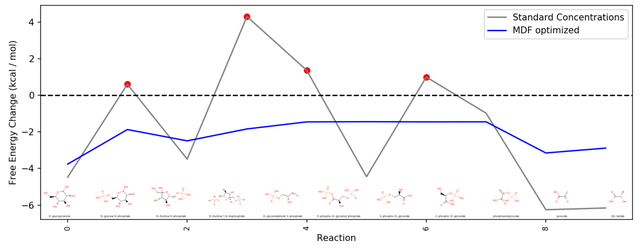

Figure 6: Reaction free energies can be connected to create a pathway free energy profile. Using defined concentration variations, we can ask if there is a set of concentrations that make all pathway reactions feasible ($\Delta_r G' < 0$) using MDF analysis[3].

Pathway Clustering

Many distinct result pathways can be very similar! Solution:

MKFF-VFALILA
MKF-AVFALKLA

$\longleftrightarrow$

$R_1\ R_2\ R_3\ R_4\ R_5$
$R_1\ R_2\ R_3\ R_4\ R_6$
$R_1\ R_8\ R_3\ R_4\ R_7$

1. Cast pathways as sequences of reactions, compounds, etc.
2. Find distance between all pathways using sequence alignment
3. Identify groups of pathways with similar sequences, and outliers!

Acknowledgements

This work was funded by the DARPA 1000 Molecules program (Scylax™ and protein design improvement, application to aniline). The original work for pathway design was funded by NSF SBIR Phase I and II #1456372.

[1] Caspi, R. et. al. (2014) The MetaCyc Database of metabolic pathways ...

[2] Hastings J, et. al. (2016). ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res.

[3] Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, et al. (2014) Pathway Thermodynamics Highlights Kinetic Obstacles in Central Metabolism. PLOS Computational Biology 10(2): e1003483. https://doi.org/10.1371/journal.pcbi.1003483

[4] Noor E, Haraldsdóttir HS, Milo R, Fleming RMT (2013) Consistent Estimation of Gibbs Energy Using Component Contributions. PLOS Computational Biology 9(7): e1003098. https://doi.org/10.1371/journal.pcbi.1003098

[5] Rahman SA, Torrance G, Baldacci L, Martínez Cuesta S, Fenninger F, Gopal N, et al. Reaction Decoder Tool (RDT): extracting features from chemical reactions. Bioinformatics. 2016;32:2065-6.