U.S. DEPARTMENT OF ENERGY
Office of ENERGY EFFICIENCY & RENEWABLE ENERGY
SAND2020-2841C

RetSynth aids in the identification of model platform organisms for the biochemical synthesis of MCCI and SI fuels

Bernard Nguyen, Ryan Davis, Anthe George, Corey M. Hudson - Sandia National Laboratories

Introduction

Scientists use retrosynthetic analysis (RSA) to quickly identify (in-silico) potential reaction pathways for a desired target molecule. In practice, an organism may require genetic engineering to optimize biological synthesis of necessary precursors. Financially, it may actually be more cost-effective to distill a different precursor at the cost of an additional chemical reaction. Chemical reactions may then require process optimization to properly scale up. Finally, after all is said and done, environmental screening may rule out the molecule or a completely novel compound may end up having more desirable fuel properties. Rather than the traditional target-first approach to RSA, we propose a platform-first approach focused on identifying a versatile platform organism. RetSynth[1], along with its flux balance analysis (FBA), gene compatibility (GC), and separation properties (SP) toolkits, enables us to identify organisms capable of producing high yields of the chemical precursors for many high-value fuel candidates.

Figure 1: Identified and theoretically possible metabolic pathways for *E. coli* (nodes=metabolites, edges=reactions), including enzymatic reactions from PATRIC[3], MetaCyc[4], ATLAS[5], and MINE[6].

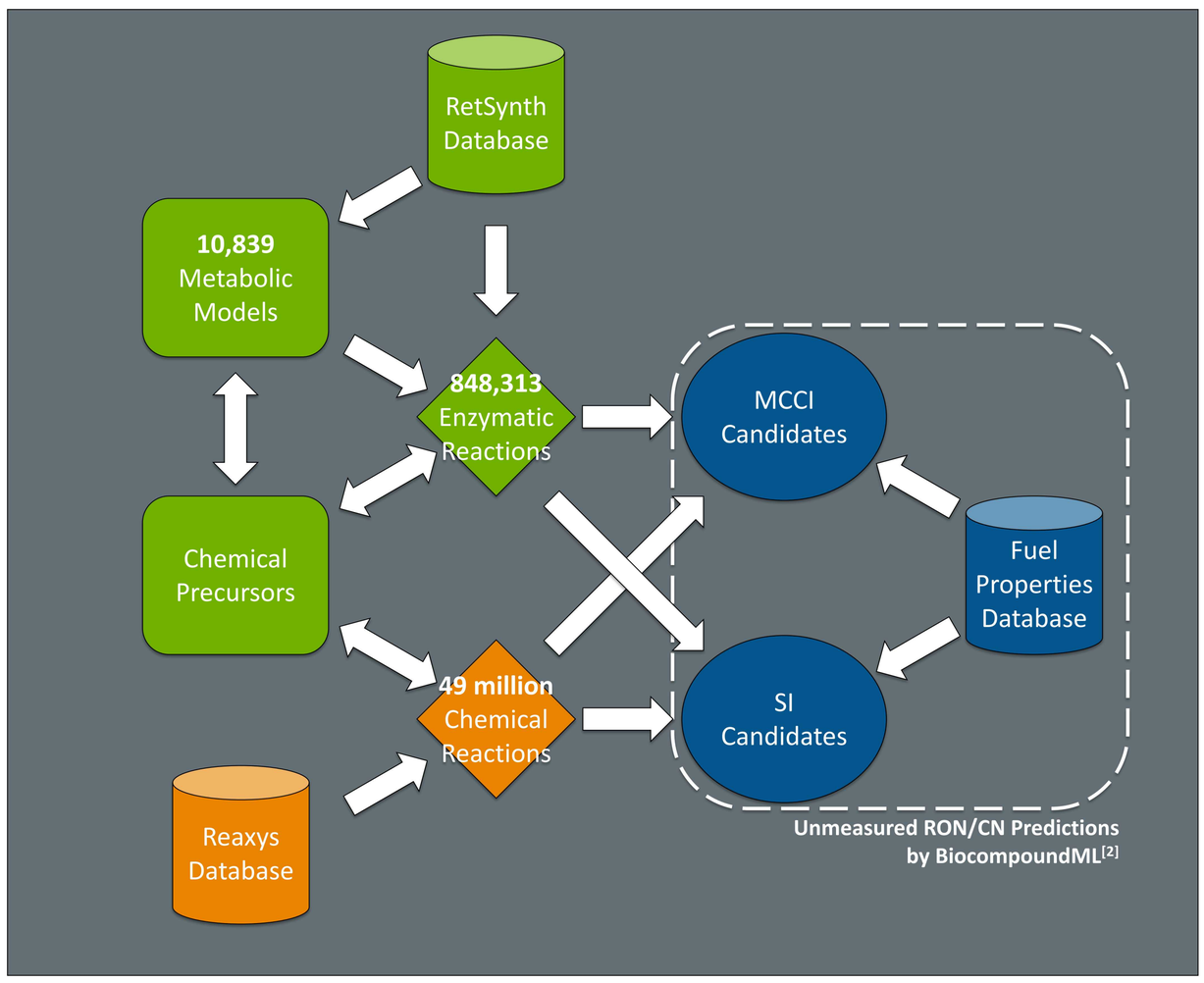

Figure 3: Workflow to identify organisms best suited to biochemically produce MCCI and SI fuels.

Methods

The RetSynth database is a centralized and organized resource containing metabolic models, enzymatic reactions, and chemical reactions from PATRIC, MetaCyc, ATLAS, MINE, KBase, SPRESI, and Reaxys. The potential metabolic pathways generated by aggregating all of this information are often vast and complex (Figure 1). In order to find the organism(s) best suited for biochemical synthesis of MCCI and SI fuel candidates, we must first narrow down the list of potential organisms.

The first step was to identify our target molecules. Of the 420 pure substances, we screened out 62 SI and 30 MCCI candidates based on the criteria in Table 1, filling missing measurements with chemical and fuel property predictions from ChemDraw and BiocompoundML respectively. We then identified the chemical precursors for all of the targets and screened for organisms in the RetSynth database that had the most potential pathways to produce the MCCI/SI targets.

Work is currently being done to take these organisms and identify the best, most versatile candidate(s) (using the FBA, GC, and SP modules) for biochemical production of MCCI and SI fuels.

Engine Type	Boiling Point (°C)	Flash Point (°C)	Melting Point (°C)	Solubility in H_2O (g/L)	Fuel Properties
SI	20 < BP < 165	N/A	< -10	< 20	RON ≥ 98
MCCI	< 338	> 52	< 0		CN ≥ 40

Table 1: SI and MCCI fuel screening criteria as determined by McCormick 2017[7] and the Co-Optima Tier 1 criteria respectively.

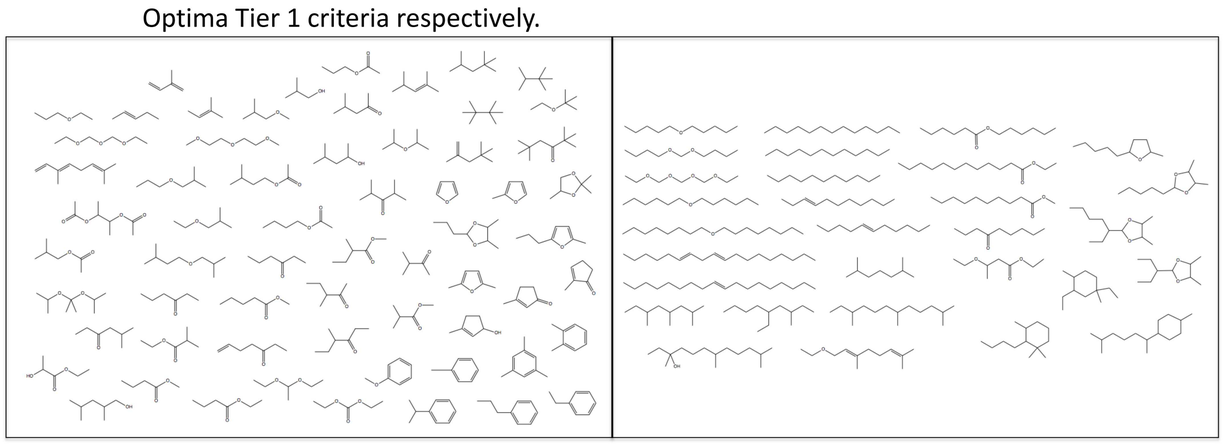

Figure 2: SI (left) and MCCI (right) fuel targets based on the screening criteria listed in Table 1.

Results and Initial Takeaways

The candidate chassis for SI fuels are *Ensifer adhaerens, Klebsiella quasipneumoniae, Klebsiella sp. KPN1705, Klebsiella variicola,* and *Shinella sp. HZN7*. The candidate chassis for MCCI fuels are *Bacillus thuringiensis, Cupriavidus basilensis, Paraburkholderia carbiensis, Paraburkholderia hospital, Paraburkholderia terrae, and Variovorax paradoxus*.

Pathways to production of MCCI compounds contain a lot of chemical reactions. As a result, there are more candidate chassis organisms capable of producing some MCCI fuels, whereas there are few SI candidate organisms capable of producing many of the SI targets. Additionally, while metabolic models suggest SI fuels can be synthesized from limited lignocellulosic media, that is not the case for MCCI fuels. Only 3 of the MCCI targets have the potential to be made biologically without any chemical processing, none of which are present in lignocellulosic models.

References

[1] Whitmore L. S, Nguyen B. , Pinar A., George A., Hudson C.M., 2019. "RetSynth: Determining all optimal and sub-optimal synthetic pathways that facilitate synthesis of target compounds in chassis organisms." BMC Bioinformatics
[2] Whitmore, Leanne S., et al. "BioCompoundML: A General Biofuel Property Screening Tool for Biological Molecules Using Random Forest Classifiers." Energy & Fuels, vol. 30, no. 10, 2016, pp. 8410-8418., doi:10.1021/acs.energyfuels.6b01952
[3] Wattam AR, Davis JJ, Assaf R, Boisvert S, Brettin T, Bun C, Conrad N, Dietrich EM, Disz T, Gabbard JL, Gerdes S, Henry CS, Kenyon RW, Machi D, Mao C, Nordberg EK, Olsen GJ, Murphy-Olson DE, Olson R, Overbeek R, Parrello B, Pusch GD, Shukla M, Vonstein V, Warren A, Xia FF, Yoo H, Stevens RL. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. Nucleic Acids Res. 2017; 45(D1):535–42.
[4] Caspi R, Billington R, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Midford PE, Ong Q, Ong WK, Paley S, Subhraveti P, Karp PD. The MetaCyc database of metabolic pathways and enzymes. Nucleic Acids Res. 2018; 46(D1):633–9.
[5] Hadadi N, Hafner J, Shajkofci A, Zisaki A, Hatzimanikatis V. Atlas of biochemistry: A repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies. ACS Synth Biol. 2016; 5(10):1155–66.
[6] Jeffryes JG, Colastani RL, Elbadawi-Sidhu M, Kind T, Niehaus TD, Broadbelt LJ, Hanson AD, Fiehn O, Tyo KE, Henry CS. MINES: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics. J Cheminform. 2015; 7:44
[7] Mccormick, Robert L., et al. "Selection Criteria and Screening of Potential Biomass-Derived Streams as Fuel Blendstocks for Advanced Spark-Ignition Engines." SAE International Journal of Fuels and Lubricants, vol. 10, no. 2, 2017, pp. 442–460., doi:10.4271/2017-01-0868.
[8] "Co-Optimization of Fuels and Engines: Fuel Properties Database: NREL." NREL.gov, www.nrel.gov/transportation/fuels-properties-database/.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. Research at Sandia National Laboratories was supported by the BioEnergy Technology Office of the Office of Energy Efficiency and Renewable Energy, U.S. Department of Energy under Award #NL00_____. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. This research was conducted as part of the Co-Optimization of Fuels & Engines (Co-Optima) project sponsored by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies and Vehicle Technologies Office.