ART: A machine learning Automated Recommendation Tool for synthetic biology

Tijana Radivojevic,[1,2,3], Zak Costello,[1,2,3] Kenneth Workman,[1,3,4] William Morrell,[1,5] Mark Forrer,[1,2,5] Somtirtha Roy,[1,3] Nahan Hillson,[1,2,3] **Jay Keasling (JBEI PI),** [2,3,4,6,7,8] and Hector Garcia Martin[1,2,3,9,10*] (hgmartin@lbl.gov)

[1]DOE Agile BioFoundry, Emeryville, CA 94608, USA. [2]Biofuels and Bioproducts Division, DOE Joint BioEnergy Institute, Emeryville, CA 94608, USA. [3]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Department of Bioengineering, University of California, Berkeley, CA 94720, USA., [5]Biotechnology and Bioengineering and Biomass Science and Conversion Department, Sandia National Laboratories, Livermore, California 94550, United States, [6]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kgs., Lyngby, Denmark. [7]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA, USA, [8]Center for Synthetic Biochemistry, Institute for Synthetic Biology, Shenzhen Institutes of Advanced Technologies, Shenzhen, China, [9]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [10]BCAM, Basque Center for Applied Mathematics, Bilbao 48009, Spain.

http://jbei.org

Project Goals: To develop machine learning methods to effectively guide bioengineering

Synthetic biology allows us to bioengineer cells to synthesize novel valuable molecules such as renewable biofuels or anticancer drugs. However, traditional synthetic biology approaches involve ad-hoc engineering practices, which lead to long development times. Here, we present the Automated Recommendation Tool (ART[1]), a tool that leverages machine learning and probabilistic modeling techniques to guide synthetic biology in a systematic fashion, without the need for a full mechanistic understanding of the biological system. Using sampling-based optimization, ART provides a set of recommended strains to be built in the next engineering cycle, alongside probabilistic predictions of their production levels[2]. We demonstrate the capabilities of ART on simulated data sets, as well as experimental data from real metabolic engineering projects producing renewable biofuels, hoppy flavored beer without hops, fatty acids, and tryptophan. We also discuss the limitations of this approach, and the practical consequences of the underlying assumptions failing. A fundamental part of using machine learning in synthetic biology involves the availability of large amounts of high-quality training data. We show how to use the Experiment Data Depot (EDD) to store, visualize and export data in a standardized fashion.

References

1. Radivojević, Tijana, et al. "A machine learning Automated Recommendation Tool for synthetic biology." Nature communications 11.1 (2020): 1-14.
2. Zhang, Jie, et al. "Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism." Nature communications 11.1 (2020): 1-13.

This work was part of the Agile BioFoundry (http://agilebiofoundry.org) and the DOE Joint BioEnergy Institute (http://www.jbei.org), supported by the U. S. Department of Energy, Energy Efficiency and Renewable Energy, Bioenergy Technologies Office, and the Office of Science, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the

U.S. Department of Energy. NJH was also supported by the DOE Joint Genome Institute (https://jgi.doe.gov) by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DEAC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. This research is also supported by the Basque Government through the BERC 2014-2017 program and by Spanish Ministry of Economy and Competitiveness MINECO: BCAM Severo Ochoa excellence accreditation SEV-2013-0323.