# Towards Effective Metabolic Engineering Using Multilevel and Probabilistic Modeling

**Tijana Radivojević*, Zak Costello† and Hector Garcia Martin‡**

*DOE Agile BioFoundry, USA
Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, USA
5885 Hollis St., 94608 Emeryville, CA, USA
e-mail: tradivojevic@lbl.gov

†Biofuels and Bioproducts Division, DOE Joint BioEnergy Institute, USA
DOE Agile BioFoundry, USA
Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, USA
5885 Hollis St., 94608 Emeryville, CA, USA
e-mail: zak.costello@lbl.gov

‡Biofuels and Bioproducts Division, DOE Joint BioEnergy Institute, USA
DOE Agile BioFoundry, USA
Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, USA
Basque Center for Applied Mathematics, Spain
5885 Hollis St., 94608 Emeryville, CA, USA
e-mail: hgmartin@lbl.gov

## ABSTRACT

Biology has changed radically in the last two decades, transitioning from a descriptive science into a design science. Synthetic biology allows us to bioengineer cells to synthesize valuable molecules such as renewable biofuels or antimalarial drugs. However, traditional synthetic biology approaches involve ad-hoc non systematic engineering practices, which lead to long development times. Here, we present the Automatic Recommendation Tool (ART), a tool that leverages machine learning, probabilistic and multilevel modeling techniques to guide metabolic engineering in a systematic fashion, without the need for a full mechanistic understanding of the biological system. ART uses a collection of experimental data (e.g. targeted proteomics, growth media components' levels) and creates a probabilistic predictive model for the target variable. A full probabilistic characterization of the production levels enables a principled way of testing hypothetical scenarios in silico. Furthermore, 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. We demonstrate the capabilities of ART on simulated and real data sets and discuss possible difficulties in achieving satisfactory predictive power.