

Machine learning takes on synthetic biology: algorithms can bioengineer cells for you

25 September 2020



Berkeley Lab scientists Tijana Radivojevic (left) and Hector Garcia Martin working on mechanistic and statistical modeling, data visualizations, and metabolic maps at the Agile BioFoundry last year. Credit: Thor Swift/Berkeley Lab

If you've eaten vegan burgers that taste like meat or used synthetic collagen in your beauty routine—both products that are "grown" in the lab—then you've benefited from synthetic biology. It's a field rife with potential, as it allows scientists to design biological systems to specification, such as engineering a microbe to produce a cancer-fighting agent. Yet conventional methods of bioengineering are slow and laborious, with trial and error being the main approach.

Now scientists at the Department of Energy's Lawrence Berkeley National Laboratory (Berkeley Lab) have developed a new tool that adapts machine learning algorithms to the needs of synthetic biology to guide development systematically. The innovation means scientists will not have to spend years developing a meticulous understanding of each part of a cell and what it

does in order to manipulate it; instead, with a limited set of training data, the algorithms are able to predict how changes in a cell's DNA or biochemistry will affect its behavior, then make recommendations for the next engineering cycle along with probabilistic predictions for attaining the desired goal.

"The possibilities are revolutionary," said Hector Garcia Martin, a researcher in Berkeley Lab's Biological Systems and Engineering (BSE) Division who led the research. "Right now, bioengineering is a very slow process. It took 150 person-years to create the anti-malarial drug, artemisinin. If you're able to create new cells to specification in a couple weeks or months instead of years, you could really revolutionize what you can do with bioengineering."

Working with BSE data scientist Tijana Radivojevic and an international group of researchers, the team developed and demonstrated a patent-pending algorithm called the Automated Recommendation Tool (ART), described in a pair of papers recently published in the journal *Nature Communications*. Machine learning allows computers to make predictions after "learning" from substantial amounts of available "training" data.

In "ART: A machine learning Automated Recommendation Tool for synthetic biology," led by Radivojevic, the researchers presented the algorithm, which is tailored to the particularities of the synthetic biology field: small training data sets, the need to quantify uncertainty, and recursive cycles. The tool's capabilities were demonstrated with simulated and historical data from previous metabolic engineering projects, such as improving the production of renewable biofuels.

In "Combining mechanistic and machine learning models for predictive engineering and optimization

of tryptophan metabolism," the team used ART to guide the metabolic engineering process to increase the production of tryptophan, an amino acid with various uses, by a species of yeast called *Saccharomyces cerevisiae*, or baker's yeast. The project was led by Jie Zhang and Soren Petersen of the Novo Nordisk Foundation Center for Biosustainability at the Technical University of Denmark, in collaboration with scientists at Berkeley Lab and Teselagen, a San Francisco-based startup company.

To conduct the experiment, they selected five genes, each controlled by different gene promoters and other mechanisms within the cell and representing, in total, nearly 8,000 potential combinations of biological pathways. The researchers in Denmark then obtained experimental data on 250 of those pathways, representing just 3% of all possible combinations, and that data were used to train the algorithm. In other words, ART learned what output (amino acid production) is associated with what input (gene expression).

Then, using statistical inference, the tool was able to extrapolate how each of the remaining 7,000-plus combinations would affect tryptophan production. The design it ultimately recommended increased tryptophan production by 106% over the state-of-the-art reference strain and by 17% over the best designs used for training the model.

"This is a clear demonstration that bioengineering led by machine learning is feasible, and disruptive if scalable. We did it for five genes, but we believe it could be done for the full genome," said Garcia Martin, who is a member of the Agile BioFoundry and also the Director of the Quantitative Metabolic Modeling team at the Joint BioEnergy Institute (JBEI), a DOE Bioenergy Research Center; both supported a portion of this work. "This is just the beginning. With this, we've shown that there's an alternative way of doing metabolic engineering. Algorithms can automatically perform the routine parts of research while you devote your time to the more creative parts of the scientific endeavor: deciding on the important questions, designing the experiments, and consolidating the obtained knowledge."

More data needed

The researchers say they were surprised by how little data was needed to obtain results. Yet to truly realize synthetic biology's potential, they say the algorithms will need to be trained with much more data. Garcia Martin describes synthetic biology as being only in its infancy—the equivalent of where the Industrial Revolution was in the 1790s. "It's only by investing in automation and high-throughput technologies that you'll be able to leverage the data needed to really revolutionize bioengineering," he said.

Radivojevic added: "We provided the methodology and a demonstration on a small dataset; potential applications might be revolutionary given access to large amounts of data."

The unique capabilities of national labs

Besides the dearth of experimental data, Garcia Martin says the other limitation is human capital—or machine learning experts. Given the explosion of data in our world today, many fields and companies are competing for a limited number of experts in machine learning and artificial intelligence.

Garcia Martin notes that knowledge of biology is not an absolute prerequisite, if surrounded by the team environment provided by the national labs. Radivojevic, for example, has a doctorate in applied mathematics and no background in biology. "In two years here, she was able to productively collaborate with our multidisciplinary team of biologists, engineers, and computer scientists and make a difference in the synthetic biology field," he said. "In the traditional ways of doing metabolic engineering, she would have had to spend five or six years just learning the needed biological knowledge before even starting her own independent experiments."

"The national labs provide the environment where specialization and standardization can prosper and combine in the large multidisciplinary teams that are their hallmark," Garcia Martin said.

Synthetic biology has the potential to make significant impacts in almost every sector: food,

medicine, agriculture, climate, energy, and materials. The global [synthetic biology](#) market is currently estimated at around \$4 billion and has been forecast to grow to more than \$20 billion by 2025, according to various market reports.

"If we could automate metabolic engineering, we could strive for more audacious goals. We could engineer microbiomes for therapeutic or bioremediation purposes. We could engineer microbiomes in our gut to produce drugs to treat autism, for example, or microbiomes in the environment that convert waste to biofuels," Garcia Martin said. "The combination of [machine learning](#) and CRISPR-based gene editing enables much more efficient convergence to desired specifications."

More information: *Nature Communications* (2020). [DOI: 10.1038/s41467-020-18008-4](https://doi.org/10.1038/s41467-020-18008-4)

Provided by Lawrence Berkeley National Laboratory

APA citation: Machine learning takes on synthetic biology: algorithms can bioengineer cells for you (2020, September 25) retrieved 5 December 2020 from <https://phys.org/news/2020-09-machine-synthetic-biology-algorithms-bioengineer.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.