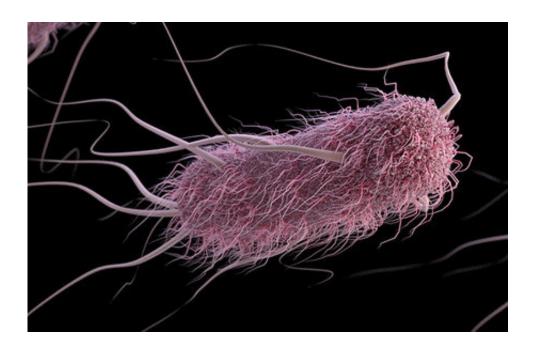# Machine-learning software predicts behavior of bacteria

April 19 2018, by Emily Velasco

An artist's depiction of a bacterial cell. Credit: Centers for Disease Control and Prevention/James Archer

In a first for machine-learning algorithms, a new piece of software developed at Caltech can predict behavior of bacteria by reading the content of a gene. The breakthrough could have significant implications for our understanding of bacterial biochemistry and for the development of new medications.

One thrust of modern pharmacology is focused on alleviating ailments

by developing drugs that target specific proteins that reside in the membranes of our bodies' cells. These proteins, known as integral membrane proteins (IMP), act as receptors or "gates" that allow materials into and out of cells. Examples of IMPs are G-protein-coupled receptors, which relay information to a cell about its environment, and ion channels, which control the interior environment of a cell by acting as gatekeepers that selectively allow ions to pass in and out of the cell. IMPs are the targets of nearly 50 percent of all drugs on the market. Unfortunately, many IMPs are poorly understood.

"These are very important molecules our body makes that we just don't know enough about," says Bil Clemons, a professor of biochemistry at Caltech.

In order to gain a more complete understanding of an IMP, researchers need to generate large amounts of it for purification and detailed study. Typically, that's done by inserting the DNA for that protein into bacteria; the protein is then produced as a matter of course as the bacteria grows and multiplies. The problem is that not all bacteria are willing to cooperate and make only measly amounts of protein. Only a few bacteria end up making enough of the proteins to be useful, and, until now, there has been no way for researchers to know if a bacterium they're working with will be a hit or a dud.

"One of the major limitations in studying membrane proteins is the lack of ability to express them in reasonable amounts," Clemons says. "We use these bacteria as factories to make things for us, but it's hit or miss ... mostly miss. Anecdotally, it's been about 10 percent successful."

All the trial and error involved in getting bacteria to cooperate wastes researchers' time and resources. Clemons wondered if it would be possible to use computers to predict how bacteria will react when asked to create a protein they normally don't produce.

"We presumed bacterial cells were doing some quantitative reading of the DNA to determine how much of these proteins to make," he says. "We wanted to know if we could use computational tools to increase the success rate of finding bacteria that express proteins in useful amounts to help us characterize molecules important to medicine."

Clemons and his graduate student, Shyam Saladi, created that tool—a machine-learning software they've dubbed IMProve—that compares bacterial DNA with data about how much protein the bacteria produces. They then used a dataset for IMProve that cultured many samples of bacteria to see how well they produced the desired membrane proteins. The researchers trained IMProve by feeding those results and the genetic codes the bacteria rely on for expressing the proteins into IMProve so it could learn which DNA sequences were going to result in high protein production.

Once the software was trained, the researchers found that it predicted bacterial behavior so well they were able to double their rate of successfully picking bacteria that would express IMPs in large quantities.

"It surprised us because there was no guarantee that this approach was going to work," Clemons says. "Cells are extremely complex, and you're asking a relatively simple statistical model to predict what a cell is going to do. From that perspective, it was pretty shocking."

But, Clemons adds that, maybe their results aren't so surprising in hindsight.

"This underlines the idea that cells are just computers, and they're just computing things," he says.

The paper, titled "A statistical model for improved membrane protein expression using sequence-derived features," appears in the March 30

issue of the *Journal of Biological Chemistry*.

  More information: IMProve is available to other researchers as a web tool by visiting clemonslab.caltech.edu/improve.html

A statistical model for improved membrane protein expression using sequence-derived features. *JBC*. DOI: 10.1074/jbc.RA117.001052

Provided by California Institute of Technology

Citation: Machine-learning software predicts behavior of bacteria (2018, April 19) retrieved 28 April 2024 from https://phys.org/news/2018-04-machine-learning-software-behavior-bacteria.html