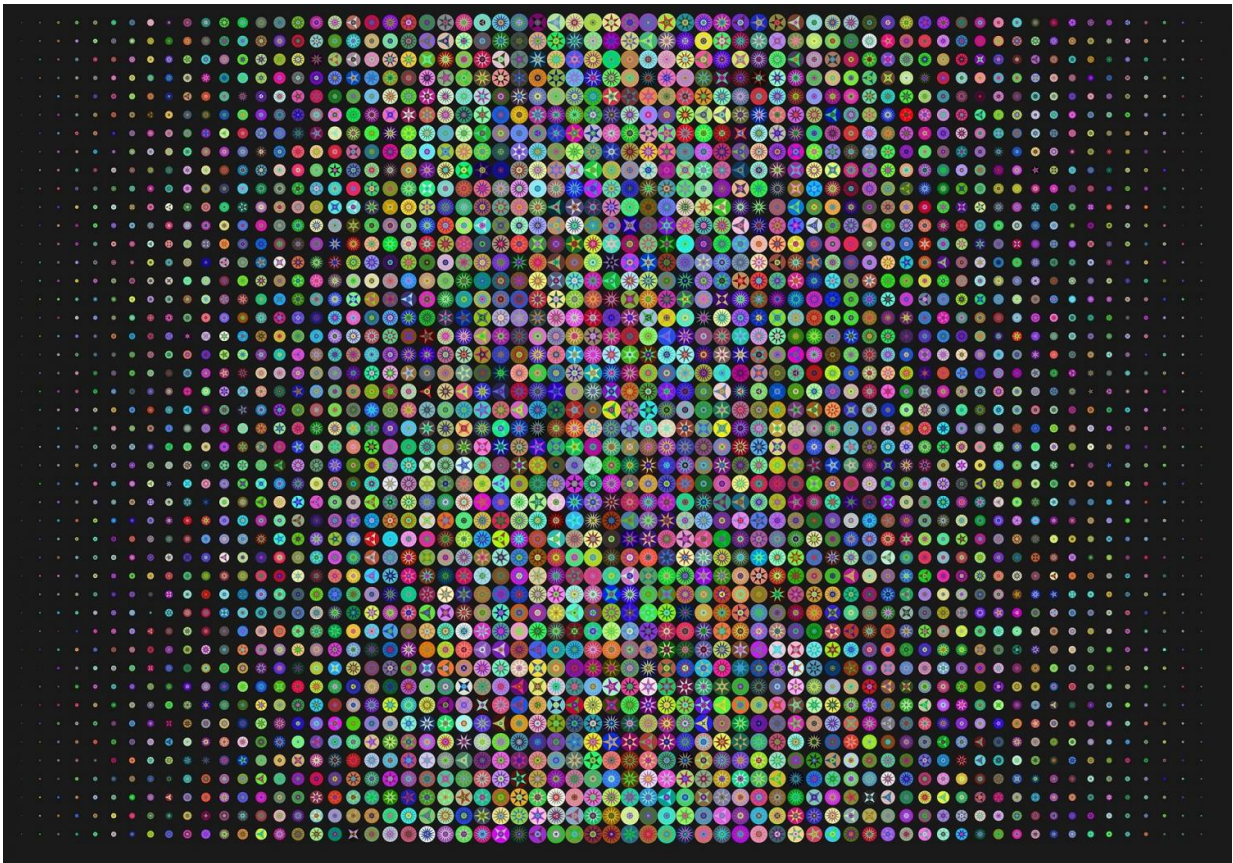


New antibiotics are desperately needed—machine learning could help

February 21 2019, by Hanae Armitage



Credit: CC0 Public Domain

Researchers at Stanford have created an algorithm that, guided by previous research, lays out the DNA sequences most likely to align with antimicrobial properties.

As the threat of [antibiotic resistance](#) looms, microbiologists aren't the only ones thinking up new solutions. James Zou, Ph.D., assistant professor of biomedical data science at Stanford, has applied machine learning to create an [algorithm](#) that generates thousands of entirely new virtual DNA sequences with the intent of one day creating antimicrobial proteins.

The algorithm, called Feedback GAN, essentially acts as a mass producer of different DNA snippets. And while these sequence attempts are somewhat random, the algorithm isn't working blindly. It's basing the new possible peptides, or small groups of amino acids, on previous research that lays out the DNA sequences most likely to align with [antimicrobial properties](#).

For now, these templates, which don't exist in nature, are theoretical, generated on a computer. But in the face of rising concerns about microbe resistance, Zou said it's critical to think about solutions that don't already exist.

"We chose to pursue antimicrobial proteins because it's a very important, high-impact problem that's also a relatively tractable problem for the algorithm," Zou said. "There are existing tools that we incorporate into our system that evaluate if a new sequence is likely to have the properties of a successful antimicrobial [protein](#)."

Feedback GAN builds on that, working to incorporate just the right balance of random chance and precision.

A paper describing the algorithm was published online Feb. 11 in *Nature Machine Learning*. Anvita Gupta, a student in computer science, is the first author; Zou is the senior author.

Self-refining

Gupta and Zou's algorithm doesn't just churn out new combinations of DNA. It also actively refines itself, learning what works and what doesn't through a feedback loop: After the algorithm spits out a wide range of DNA sequences, it runs a trial-and-error learning process that sifts through the peptide suggestions. Based on their resemblance to other known antimicrobial peptides, the "good" ones get fed back into the algorithm to inform future DNA sequences generated from the code, and to get refined themselves.

"There's a built-in arbiter and, by having this [feedback loop](#), the system learns to model newly generated sequences after those that are deemed likely to have antimicrobial properties," Zou said. "So the idea is both individual peptide sequences and the generation of the sequences get better and better."

Zou has also considered another core component of hypothetical proteins: protein folding. Proteins contort into very specific structures linked to their functions. An algorithm could create the perfect sequence, but unless it can fold up, it's useless—like the cogs of a clock strewn on a table.

Zou can tweak the algorithm so that instead of analyzing a propensity for antimicrobial properties, it determines the likelihood of correct folding.

"We can actually do these two things in parallel where we look at antimicrobial properties of one sequence and folding likelihood of another," said Zou. "We run both so that we're optimizing either the antimicrobial properties or its ability to fold."

Next, Zou hopes to merge the two variations of the algorithm to create peptide sequences that are optimized for both their microbe-killing abilities and their ability to fold into a genuine protein.

More information: Anvita Gupta et al. Feedback GAN for DNA optimizes protein functions, *Nature Machine Intelligence* (2019). [DOI: 10.1038/s42256-019-0017-4](https://doi.org/10.1038/s42256-019-0017-4)

Provided by Stanford University Medical Center

Citation: New antibiotics are desperately needed—machine learning could help (2019, February 21) retrieved 25 April 2024 from <https://techxplore.com/news/2019-02-antibiotics-desperately-neededmachine.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.