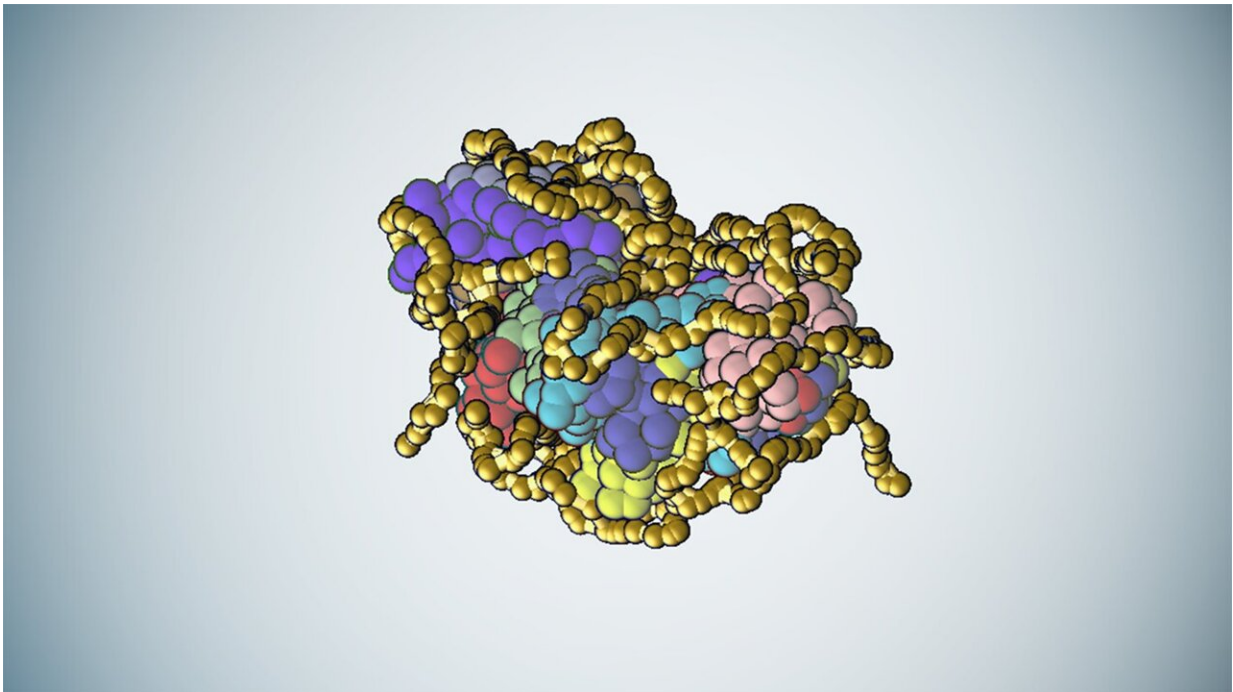


# Automation speeds the search for stable proteins

August 2 2022, by Stephanie Pappas

---



Researchers stabilized an important protein by harnessing the twin power of robotics and machine learning. They also showed that the system works for a range of proteins. Their robotics platform synthesized the reinforcement molecules more than 10 times faster than leading methods. And the machine learning model opened vast new possibilities for finding the right combination of materials, slashing months or even years from the time it takes scientists to pair proteins with their ideal supports. Image courtesy of the researchers. Credit: Michael Webb/Princeton University

Harnessing the power of robotics and machine intelligence, researchers from Princeton Engineering and Rutgers University have found a way to design stable proteins in a fraction of the time of current state of the art. The team's robotics platform speeds things up more than tenfold, and their computational approach finds solutions anywhere from weeks to years faster than what is possible by human intelligence alone.

Stabilizing proteins is a central challenge for research into drug creation, biofuel production and plastic recycling. Currently, scientists use their knowledge of chemistry to estimate which chemical compounds will pair well with proteins under different conditions. The conventional approach uses trial and error to refine results. This painstaking method can take months as scientists create and test samples of molecules, and it often leads nowhere.

In the new system, engineers use a machine learning [model](#) to identify [chemical compounds](#) most likely to stabilize desired proteins. The model helps narrow hundreds of thousands of possibilities to a few likely candidates. A robotic assembly platform produces samples of the molecules for evaluation. Combining the robotic platform with the [machine-learning model](#) turns out results in as little as a few days.

This twin-turbo approach offers an additional advantage: because of its ability to churn through vast amounts of data, the machine-learning model often recommends candidate molecules that would not have occurred to scientists.

"In terms of the increase in what we can look for, it's pretty much unbounded," said Michael Webb, assistant professor of chemical and biological engineering at Princeton and one of the study's two senior authors. "The utilization of machine learning to direct our search accelerates discovery by an amount which is hard to quantify but is very important. It's possible that you could be spinning your wheels for a very

long time if you continued to rely on systematic search or trial-and-error."

Led by Webb and Adam Gormley, assistant professor of biomedical engineering at Rutgers, the researchers published their findings in the journal *Advanced Materials*.

In developing their system, the team turned to three proteins with distinctive properties, including a protein found in horseradish that is used widely in hospitals and water treatment plants.

"If we could solve the problem for these three, then theoretically we could extend the same procedure to all sorts of enzymes," said Roshan Patel, a graduate student in Webb's lab and one of the first authors of the new paper.

While proteins perform all kinds of amazing feats in nature, they tend to be picky about their working conditions. Changes in temperature or exposure to solvents can stop them in their tracks. To harden proteins for use outside their native environments, scientists often reinforce them with specialized supporting materials—like rebar in concrete—making these fragile structures more robust. That's a key step for enabling a vast body of biomedical, environmental and other industrial technologies.

But finding the perfect match between a protein and its support molecule means optimizing an astronomical number of choices. Conventional methods are slow and largely unsystematic—think trial and error—which means most of the possible solutions go unexplored.

With the horseradish protein, the researchers started by making 500 different support molecules based on the traditional, intuitive approach. Each support had some potential to steel the protein against harsh industrial conditions, but the researchers didn't know much more than

that. They then tested each of the 500 molecules as a support, gathering real data on its performance, and simultaneously tasked the computer model with making predictions about what they would find. Comparing the predictions against the findings allowed them to improve the computer model through a process of positive reinforcement, called reinforcement learning.

With the newly trained computer model, the researchers expanded their search to more than half a million possible support molecules. Each molecule represented a different recipe pieced together from thousands of ingredients in various configurations. They ran the data through the model four times, each time looking for two things: molecules that would out-perform the rest of the field, or molecules that held some interesting quality that might make the algorithm even more sophisticated.

"On the fifth run," Webb said, "we took the cuffs off. We said, OK, give us the best 24 [molecules] you can find."

Compared to the molecules they identified using intuition-based methods, the new machine-assisted approach found support molecules that worked more than five times better for the horseradish protein. When working with lipase, a protein that breaks down fat in the body, the results were far more dramatic. The new system found a support molecule that improved performance by roughly 50 times compared to the initial choices, even pushing the [protein](#) to work better outside its native environment than it would in its natural state.

"There are a lot of things you can manipulate about [these molecules], including the chemistry of their underlying units, their size, their architecture, their sequence," Webb said. "All of those things can impact the properties in a way you might exploit" for a useful application.

Webb said they could streamline the process and speed it further by integrating the [machine learning model](#) with the physical robotics system on site. Much of the initial work was done by sending data back and forth between the two labs.

He also pointed to specific applications the team was starting to work on, where finding [molecules](#) to stabilize proteins could lead to transformative solutions: a new way to recycle hard-to-break plastics and a non-invasive treatment for spinal cord injuries.

"There is an opportunity to do a follow-up and figure out more precisely why these things are working and the conditions in which they do," Webb said.

**More information:** Matthew J. Tamasi et al, Machine Learning on a Robotic Platform for the Design of Polymer–Protein Hybrids, *Advanced Materials* (2022). [DOI: 10.1002/adma.202201809](https://doi.org/10.1002/adma.202201809)

Provided by Princeton University

Citation: Automation speeds the search for stable proteins (2022, August 2) retrieved 8 August 2024 from <https://techxplore.com/news/2022-08-automation-stable-proteins.html>

<p>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.</p>
--