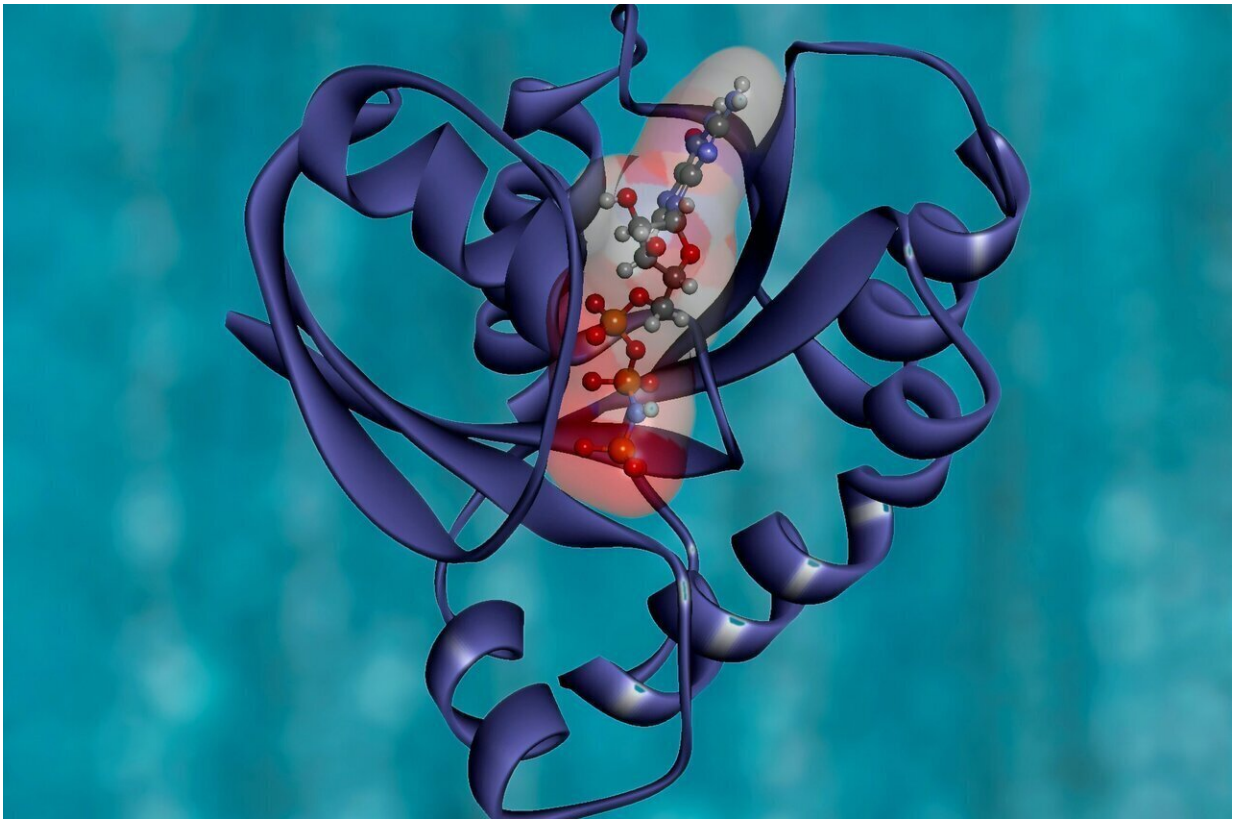


Machine learning reveals recipe for building artificial proteins

July 24 2020, by Emily Ayshford



Credit: Unsplash/CC0 Public Domain

Proteins are essential to the life of cells, carrying out complex tasks and catalyzing chemical reactions. Scientists and engineers have long sought to harness this power by designing artificial proteins that can perform

new tasks, like treat disease, capture carbon, or harvest energy, but many of the processes designed to create such proteins are slow and complex, with a high failure rate.

In a breakthrough that could have implications across the healthcare, agriculture, and energy sectors, a team lead by researchers in the Pritzker School of Molecular Engineering (PME) at the University of Chicago has developed an [artificial intelligence](#)-led process that uses big data to design new proteins.

By developing machine-learning models that can review protein information culled from genome databases, the researchers found relatively simple design rules for building [artificial proteins](#). When the team constructed these artificial proteins in the lab, they found that they performed chemistries so well that they rivaled those found in nature.

"We have all wondered how a simple process like evolution can lead to such a high-performance material as a protein," said Rama Ranganathan, Joseph Regenstein Professor in the Department of Biochemistry and Molecular Biology, Pritzker Molecular Engineering, and the College.

"We found that genome data contains enormous amounts of information about the basic rules of protein structure and function, and now we've been able to bottle nature's rules to create proteins ourselves."

The results were published July 24 in the journal *Science*.

Using artificial intelligence to learn design rules

Proteins are made up of hundreds or thousands of [amino acids](#), and these [amino acid sequences](#) specify the protein's structure and function. But understanding just how to build these sequences to create novel proteins has been challenging. Past work has resulted in methods that can specify structure, but function has been more elusive.

What Ranganathan and his collaborators realized over the past 15 years is that genome databases—which are growing exponentially—contain enormous amounts of information about the basic rules of protein structure and function. His group developed mathematical models based on this data and then began using machine-learning methods to reveal new information about proteins' basic design rules.

For this research, they studied the chorismate mutase family of metabolic enzymes, a type of protein that is important for life in many bacteria, fungi, and plants. Using machine-learning models, the researchers were able to reveal the simple design rules behind these proteins.

The model shows that just conservation at amino acid positions and correlations in the evolution of pairs of amino acids are sufficient to predict new artificial sequences that would have the properties of the protein family.

"We generally assume that to build something, you have to first deeply understand how it works," Ranganathan said. "But if you have enough data examples, you can use deep learning methods to learn the rules of design, even as you are understanding how it works or why it's built that way."

He and his collaborators then created synthetic genes to encode for the proteins, cloned them into bacteria, and watched as the bacteria then made the synthetic proteins using their normal cellular machinery. They found that the artificial proteins had the same catalytic function as the natural chorismate mutase proteins.

A platform to understand other complex systems

Because the design rules are so relatively simple, the number of artificial

proteins that researchers could potentially create with them is extremely large.

"The constraints are much smaller than we ever imagined they would be," Ranganathan said. "There is a simplicity in nature's design rules, and we believe similar approaches could help us search for models for design in other complex systems in biology, like ecosystems or the brain."

Though artificial intelligence revealed the design rules, Ranganathan and his collaborators still don't fully understand why the models work. Next they will work to understand just how the models came to this conclusion. "There is much more work to be done," he said.

In the meantime, they also hope to use this platform to develop proteins that can address pressing societal problems, like climate change. Ranganathan and Assoc. Prof. Andrew Ferguson have spun out a company called Evozyne that will commercialize this technology with applications in energy, environment, catalysis, and agriculture. Ranganathan has worked with UChicago's Polsky Center for Entrepreneurship and Innovation to file patents and license the IP to the company.

"This system gives us a platform for rationally engineering [protein](#) molecules in a way that we always dreamed we could," he said. "Not only can it teach us the physics of how proteins work and how they evolve, it can help us find solutions for issues like carbon capture and energy harvesting. Even more generally, the studies in proteins might even help teach us how the deep neural networks behind modern machine learning actually work."

More information: An evolution-based model for designing chorismate mutase enzymes, *Science* (2020). [DOI: 10.1126/science.aba3304](https://doi.org/10.1126/science.aba3304)

Provided by University of Chicago

Citation: Machine learning reveals recipe for building artificial proteins (2020, July 24) retrieved 29 April 2024 from

<https://phys.org/news/2020-07-machine-reveals-recipe-artificial-proteins.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.