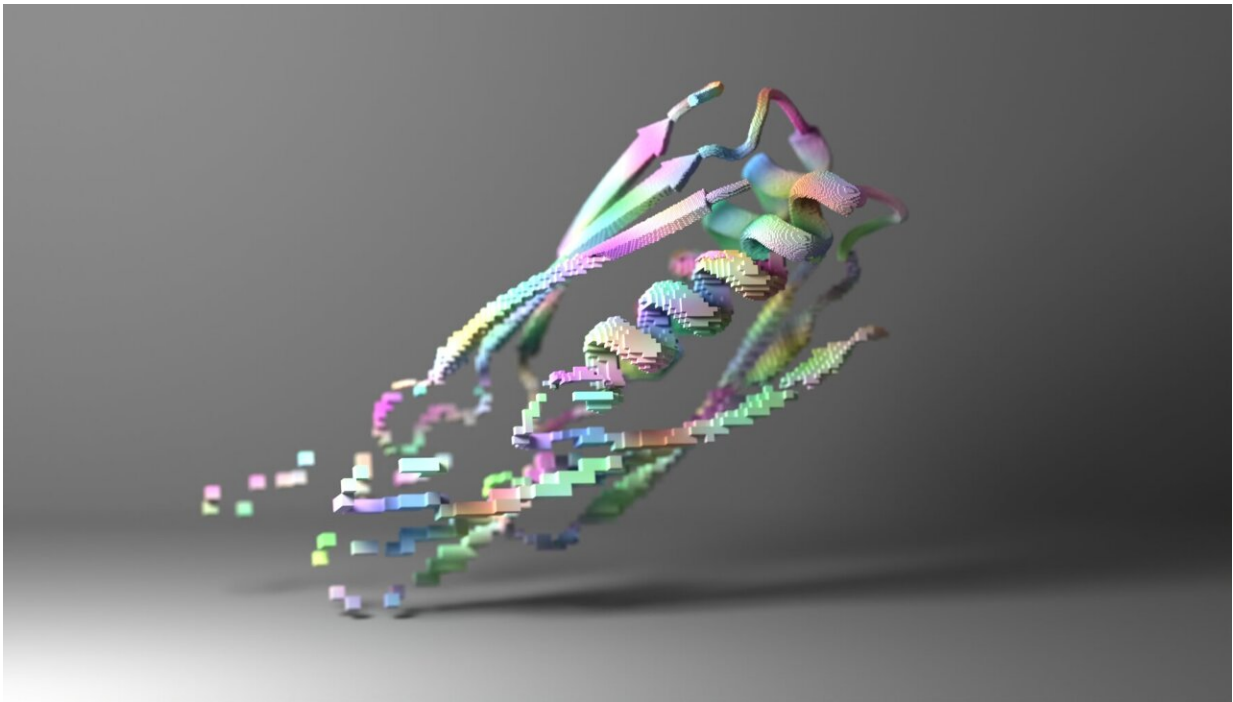


Beyond AlphaFold: AI excels at creating new proteins

September 15 2022, by Ian Haydon



Proteins designed with an ultra-rapid software tool called ProteinMPNN were much more likely to fold up as intended. Credit: Ian Haydon, UW Medicine Institute for Protein Design

Over the past two years, machine learning has revolutionized protein structure prediction. Now, three papers in *Science* describe a similar revolution in protein design.

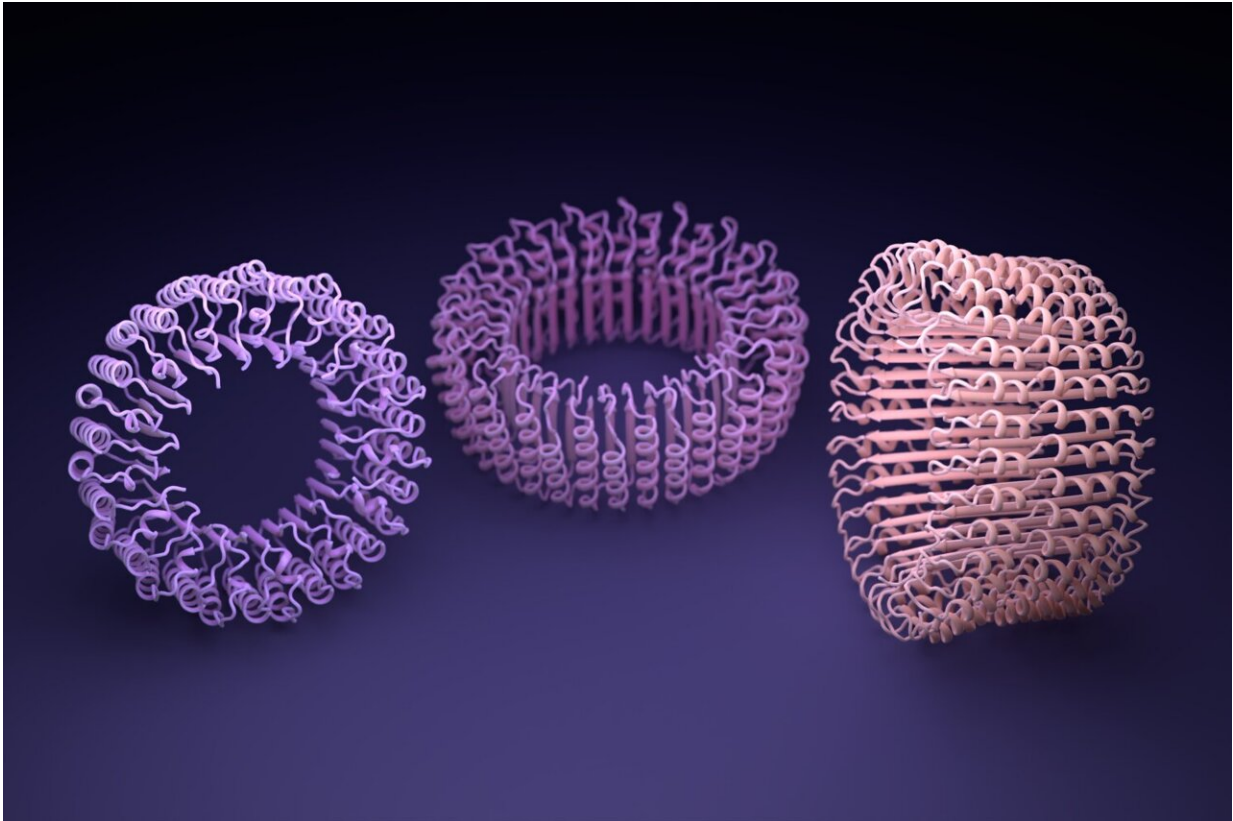
In the new papers, biologists at the University of Washington School of Medicine show that machine learning can be used to create [protein molecules](#) much more accurately and quickly than previously possible. The scientists hope this advance will lead to many new vaccines, treatments, tools for [carbon capture](#), and sustainable biomaterials.

"Proteins are fundamental across biology, but we know that all the proteins found in every plant, animal, and microbe make up far less than one percent of what is possible. With these new software tools, researchers should be able to find solutions to long-standing challenges in medicine, energy, and technology," said senior author David Baker, professor of biochemistry at the University of Washington School of Medicine and recipient of a [2021 Breakthrough Prize in Life Sciences](#).

Proteins are often referred to as the "building blocks of life" because they are essential for the structure and function of all living things. They are involved in virtually every process that takes place inside cells, including growth, division, and repair. Proteins are made up of long chains of chemicals called amino acids. The sequence of [amino acids](#) in a [protein](#) determines its three-dimensional shape. This intricate shape is crucial for the protein to function.

Recently, powerful machine learning algorithms including [AlphaFold](#) and [RoseTTAFold](#) have been trained to predict the detailed shapes of natural proteins based solely on their amino acid sequences. Machine learning is a type of artificial intelligence that allows computers to learn from data without being explicitly programmed. Machine learning can be used to model complex scientific problems that are too difficult for humans to understand.

To go beyond the proteins found in nature, Baker's [team members](#) broke down the challenge of [protein design](#) into three parts and used new software solutions for each.



Artificial intelligence hallucinated these symmetric protein assemblies, in a way similar to other A.I. generative tools that produce output based on simple prompts. Credit: Ian Haydon, UW Medicine Institute for Protein Design

First, a new protein shape must be generated. In a [paper](#) published July 21 in the journal *Science*, the team showed that artificial intelligence can generate new protein shapes in two ways. The first, dubbed "hallucination," is akin to DALL-E or other generative A.I. tools that produce output based on simple prompts. The second, dubbed "inpainting," is analogous to the autocomplete feature found in modern search bars.

Second, to speed up the process, the team devised a new algorithm for

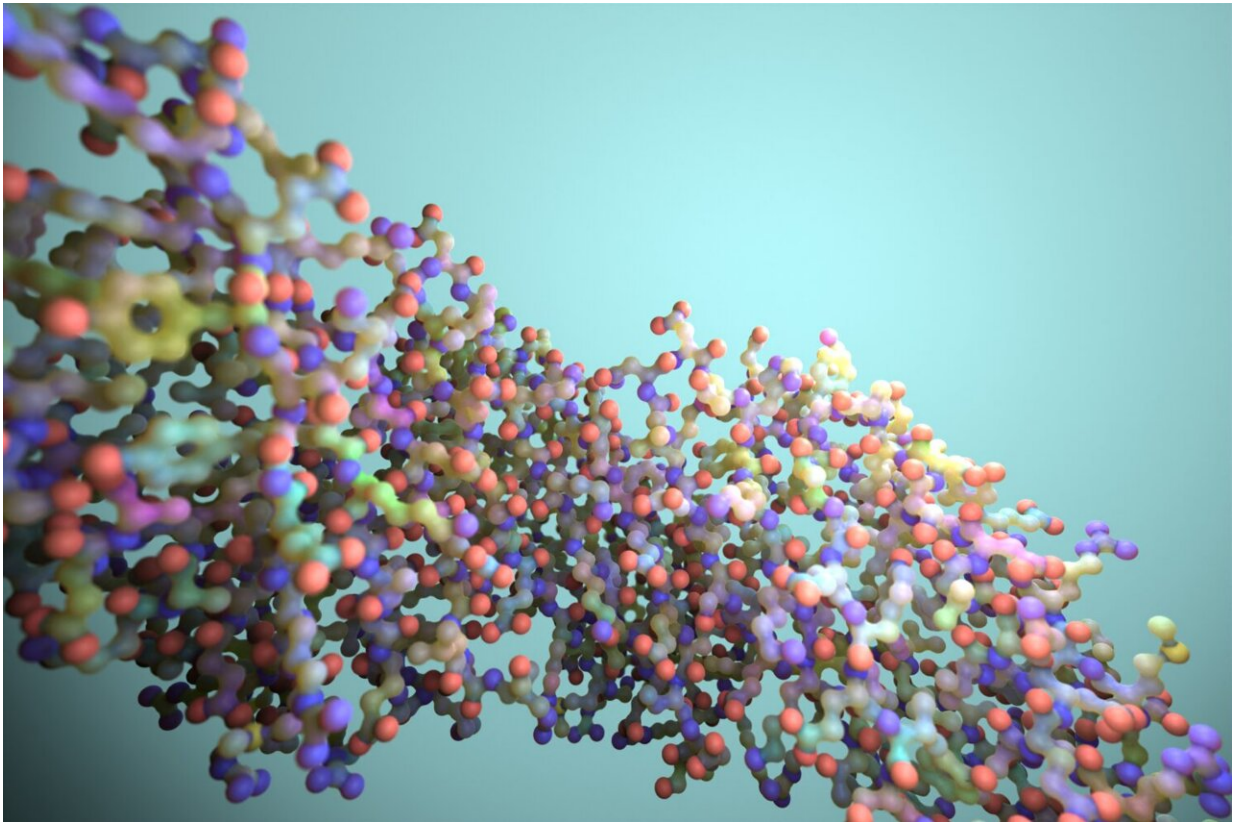
generating amino acid sequences. Described in the Sept.15 issue of *Science*, this [software tool](#), called ProteinMPNN, runs in about one second. That's more than 200 times faster than the previous best software. Its results are superior to prior tools, and the software requires no expert customization to run.

"Neural networks are easy to train if you have a ton of data, but with proteins, we don't have as many examples as we would like. We had to go in and identify which features in these molecules are the most important. It was a bit of trial and error," said project scientist Justas Dauparas, a postdoctoral fellow at the Institute for Protein Design

Third, the team used AlphaFold, a [tool](#) developed by Alphabet's DeepMind, to independently assess whether the amino acid sequences they came up with were likely to fold into the intended shapes.

"Software for predicting protein structures is part of the solution but it cannot come up with anything new on its own," explained Dauparas.

"ProteinMPNN is to protein design what AlphaFold was to protein structure prediction," added Baker.



Detail of a protein designed using a rapid tool called ProteinMPNN, another advance in the use of artificial intelligence and machine learning in protein design. Credit: Ian Haydon, UW Medicine Institute for Protein Design

In another paper appearing in *Science* Sept. 15, a team from the Baker lab confirmed that the combination of new machine learning tools could reliably generate new proteins that functioned in the laboratory.

"We found that proteins made using ProteinMPNN were much more likely to fold up as intended, and we could create very complex protein assemblies using these methods" said project scientist Basile Wicky, a postdoctoral fellow at the Institute for Protein Design.

Among the new proteins made were nanoscale rings that the researchers

believe could become parts for custom nanomachines. Electron microscopes were used to observe the rings, which have diameters roughly a billion times smaller than a poppy seed.

"This is the very beginning of [machine learning](#) in protein design. In the coming months, we will be working to improve these tools to create even more dynamic and functional proteins," said Baker.

Computer resources for this work were donated by Microsoft and Amazon Web Services.

More information: J. Dauparas et al, Robust deep learning based protein sequence design using ProteinMPNN, *Science* (2022). [DOI: 10.1126/science.add2187](https://doi.org/10.1126/science.add2187).
www.science.org/doi/10.1126/science.add2187

B. I. M. Wicky et al, Hallucinating symmetric protein assemblies, *Science* (2022). [DOI: 10.1126/science.add1964](https://doi.org/10.1126/science.add1964).
www.science.org/doi/10.1126/science.add1964

Provided by University of Washington School of Medicine

Citation: Beyond AlphaFold: AI excels at creating new proteins (2022, September 15) retrieved 27 April 2024 from <https://phys.org/news/2022-09-alphafold-ai-excels-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.