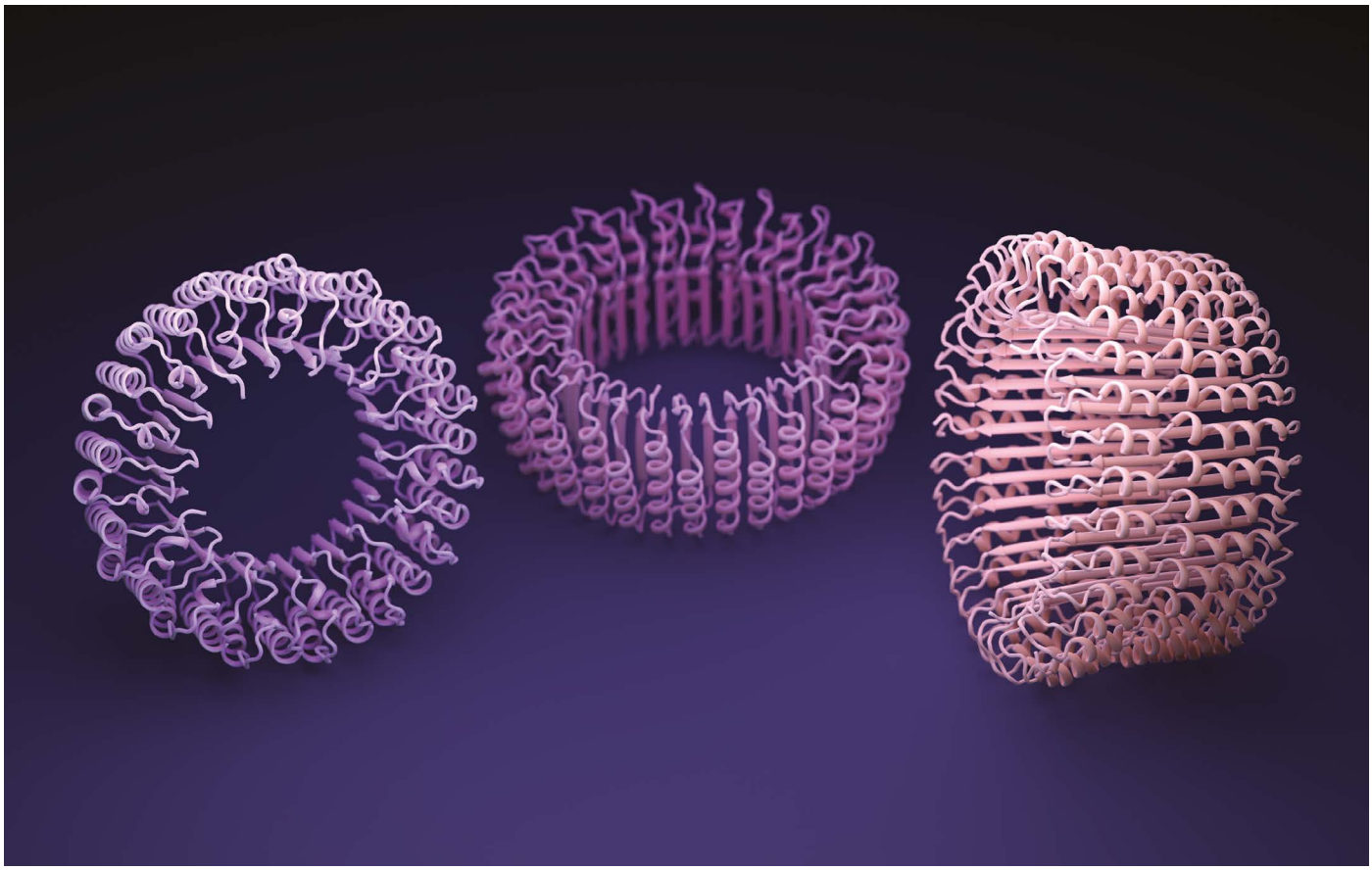
News in focus



IAN C. HAYDON/UW INST. PROTEIN DESIGN

Artificial-intelligence tools are helping scientists to come up with proteins that are shaped unlike anything in nature.

SCIENTISTS ARE USING AI TO DREAM UP REVOLUTIONARY NEW PROTEINS

Huge advances in artificial intelligence mean researchers can design completely original molecules in seconds instead of months.

By Ewen Callaway

In June, South Korean regulators authorized the first-ever medicine, a COVID-19 vaccine, to be made from a novel protein designed by humans. The vaccine is based on a spherical protein 'nanoparticle' that was created by researchers nearly a decade ago, through a labour-intensive trial-and-error-process[1].

Now, thanks to gargantuan advances in artificial intelligence (AI), a team led by David Baker, a biochemist at the University of Washington in Seattle, reports in *Science*[2,3] that it can design such molecules in seconds instead of months.

Such efforts are a part of a scientific sea change, as AI tools such as DeepMind's protein-structure-prediction software AlphaFold are embraced by life scientists.

"Since AlphaFold, there's been a shift in the way we work with protein design," says Noelia Ferruz, a computational biologist at the University of Girona, Spain. "We are witnessing very exciting times."

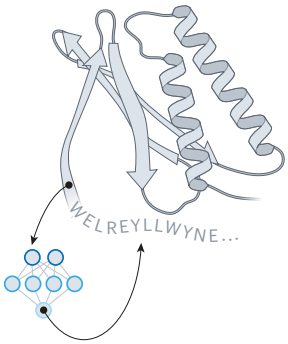
Baker's laboratory has spent the past three decades making new proteins. Software called Rosetta, which his lab started developing in the 1990s, splits the process into steps. Initially, researchers had conceived a shape for a novel protein — often by cobbling together bits of other proteins — and the software had deduced a sequence of amino acids that corresponded to this shape.

But these 'first draft' proteins rarely folded into the desired shape when made in the lab, and instead ended up stuck in different conformations. So another step was needed to ensure that the protein sequence folded only into a single desired structure. This step, which involved simulating all the ways in which sequences might fold, was computationally expensive,

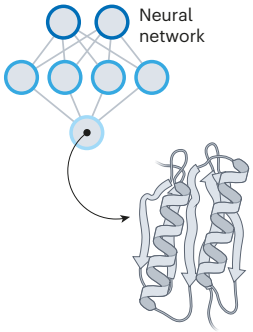
HOW TO DESIGN A PROTEIN

Researchers have developed dozens of artificial-intelligence (AI) protein-design tools in recent years, using several approaches. Here are four techniques that design new protein structures or sequences using AI. In practice, researchers can mix and match methods to come up with a working final design.

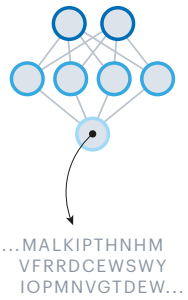
Fixed-backbone design
Given a predetermined protein structure, an AI network determines an amino-acid sequence for that protein.



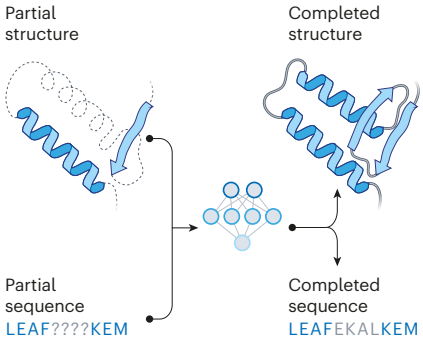
Structure generation
A neural network trained on protein structures can generate totally novel protein structures, but often with limited control of the output.



Sequence generation
Using language models, a neural network learns to 'speak' protein. These networks can be fine-tuned to generate novel sequences resembling members of specific protein families.



Sequence and structure design
Using one approach called inpainting, researchers input a structure or sequence that they want included in a protein, and the AI network fills in the rest.



NIK SPENCER/*NATURE*; SOURCE: ADAPTED FROM REF 5 AND 6

says Sergey Ovchinnikov, an evolutionary biologist at Harvard University in Cambridge, Massachusetts, who used to work in Baker's lab. "You would literally have, like, 10,000 computers running for weeks doing this."

By tweaking AlphaFold and other AI programmes, that time-consuming step has become instantaneous, says Ovchinnikov. In one approach developed by Baker's team, called hallucination, researchers feed random amino-acid sequences into a structure-prediction network; this makes the structure ever-more protein-like, as judged by the network's predictions. In a 2021 paper, Baker's team created more than 100 small, 'hallucinated' proteins in the lab and found signs that about one-fifth resembled the predicted shape[4].

AlphaFold, and a similar tool developed by Baker's lab called RoseTTAFold, were trained to predict the structure of individual protein chains. But researchers soon discovered that such networks could also model assemblies of multiple interacting proteins. On this basis, Baker and his team were confident that they could hallucinate proteins that would self-assemble into nanoparticles of different shapes and sizes; these would be made up of numerous copies of a single protein and would be similar to those on which the COVID-19 vaccine is based.

But when they instructed microorganisms to make their creations in the lab, none of the 150 designs worked. "They didn't fold at all: they were just gunk at the bottom of the test tube," says Baker.

Around the same time, another researcher in the lab, machine-learning scientist Justas Dauparas, was developing a deep-learning tool to address what is known as the inverse folding problem — determining a protein sequence that corresponds to a given protein's overall shape[3]. The network, called ProteinMPNN, can act as a 'spellcheck' for designer proteins created using AlphaFold and other tools, says Ovchinnikov, by tweaking sequences while maintaining the molecules' overall shape.

When Baker and his team applied this second network to their hallucinated protein nanoparticles, they had much greater success making the molecules experimentally. The researchers checked 30 of their new proteins using cryo-electron microscopy and other experimental techniques, and 27 of them matched the AI-led designs[2]. The team's creations included giant rings with complex symmetries, unlike anything found in nature. In theory, the approach could be used to design nanoparticles corresponding to almost any symmetrical shape, says Lukas Milles, a biophysicist who co-led the effort. "It is electrifying to see what these networks can do."

"There's been a shift in the way we work with protein design. We are witnessing very exciting times."

Deep-learning revolution

Baker's isn't the only lab applying AI to protein design. In a review paper posted to the bioRxiv preprint server this month, Ferruz and her colleagues counted more than 40 AI protein-design tools that have been developed in recent years, using various approaches[5] (see 'How to design a protein').

Many of these tools, including ProteinMPNN, tackle the inverse folding problem: they specify a sequence that corresponds to a particular structure, often using approaches borrowed from image-recognition tools. Some others are based on an architecture similar to that of language neural networks such as GPT-3, which produces human-like text; but, instead, the tools are capable of producing novel protein sequences. "These networks are able to 'speak' proteins," says Ferruz.

For Baker and his colleagues, making a novel protein in the lab is the ultimate test of their methods. But not all scientists developing AI tools for protein design have easy access to experimental set-ups, notes Jinbo Xu, a computational biologist at the Toyota Technological Institute at Chicago in Illinois. Finding a lab to collaborate with can take time, so Xu is establishing his own wet lab to put his team's creations to the test.

Experiments will also be essential when it comes to designing proteins with specific tasks in mind, says Baker. In July, his team described a pair of AI methods that allow researchers to embed a specific sequence or structure in a novel protein[6]. They used these approaches to design enzymes that catalyse particular reactions; proteins capable of binding to other molecules; and a protein that could be used in a vaccine against a respiratory virus that is a leading cause of infant hospitalizations.

Last year, London-based DeepMind launched a spin-off company called Isomorphic Labs that will apply AI tools such as AlphaFold to drug discovery. DeepMind's chief executive, Demis Hassabis, says that he sees protein design as an obvious and promising application for deep-learning technology, and for AlphaFold in particular. "We're working quite a lot in the protein design space. It's pretty early days."

1. Hsia, Y. *et al. Nature* **535**, 136–139 (2016).
2. Wicky, B. I. M. *et al. Science* https://doi.org/10.1126/science.add1964 (2022).
3. Dauparas, J. *et al. Science* https://doi.org/10.1126/science.add2187 (2022).
4. Anishchenko, I. *et al. Nature* **600**, 547–552 (2021).
5. Ferruz, N. *et al.* Preprint at bioRxiv https://doi.org/10.1101/2022.08.31.505981 (2022).
6. Wang, J. *et al. Science* **377**, 387–394 (2022).