Structures of a protein that were predicted by artificial intelligence (blue) and experimentally determined (green) match almost perfectly.

IN DEPTH

STRUCTURAL BIOLOGY

‘The game has changed.’ AI triumphs at protein folding

In milestone, software predictions finally match structures calculated from experimental data

By **Robert F. Service**

Artificial intelligence (AI) has solved one of biology’s grand challenges: predicting how proteins fold from a chain of amino acids into 3D shapes that carry out life’s tasks. This week, organizers of a protein-folding competition announced the achievement by researchers at DeepMind, a U.K.-based AI company. They say the DeepMind method will have far-reaching effects, among them dramatically speeding the creation of new medications.

“What the DeepMind team has managed to achieve is fantastic and will change the future of structural biology and protein research,” says Janet Thornton, director emeritus of the European Bioinformatics Institute. “This is a 50-year-old problem,” adds John Moult, a structural biologist at the University of Maryland, Shady Grove, and co-founder of the competition, Critical Assessment of Protein Structure Prediction (CASP). “I never thought I’d see this in my lifetime.”

The body uses tens of thousands of different proteins, each a string of dozens to hundreds of amino acids. The order of the amino acids dictates how the myriad pushes and pulls between them give rise to proteins’ complex 3D shapes, which, in turn, determine how they function. Knowing those shapes helps researchers devise drugs that can lodge in proteins’ crevices. And being able to synthesize proteins with a desired structure could speed development of enzymes to make biofuels and degrade waste plastic.

For decades, researchers deciphered proteins’ structures using experimental techniques such as x-ray crystallography or cryo–electron microscopy (cryo-EM). But such methods can take years and don’t always work. Structures have been solved for only about 170,000 of the more than 200 million proteins discovered across life forms.

In the 1960s, researchers realized if they could work out all interactions within a protein’s sequence, they could predict its shape. But the amino acids in any given sequence could interact in so many different ways that the number of possible structures was astronomical. Computational scientists jumped on the problem, but progress was slow.

In 1994, Moult and colleagues launched CASP, which takes place every 2 years. Entrants get amino acid sequences for about 100 proteins whose structures are not known. Some groups compute a structure for each sequence, while others determine it experimentally. The organizers then compare the computational predictions with the lab results and give the predictions a global distance test (GDT) score. Scores above 90 on the 100-point scale are considered on par with experimental methods, Moult says.

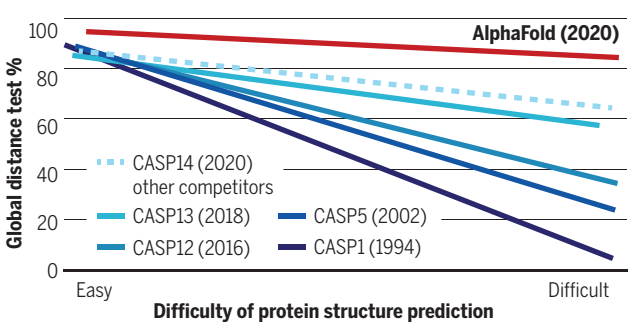
Even in 1994, predicted structures for small, simple proteins could match experimental results. But for larger, challenging proteins, computations’ GDT scores were about 20, “a complete catastrophe,” says Andrei Lupas, a CASP judge and evolutionary biologist at the Max Planck Institute for Developmental Biology. By 2016, competing groups had reached scores of about 40 for the hardest proteins, mostly by drawing insights from known structures of proteins that were closely related to the CASP targets.

When DeepMind first competed, in 2018, its algorithm, called AlphaFold, relied on this comparative strategy. But AlphaFold also incorporated a computational approach called deep learning, in which the software is trained on vast data troves—in this case, the sequences and structures of known proteins—and learns to spot patterns. DeepMind won handily, beating the competition by an average of 15% on each structure, and winning GDT scores of up to about 60 for the hardest targets.

But the predictions were still too coarse, says John Jumper, who heads AlphaFold’s development at DeepMind. “We knew how far we were from biological relevance.” So

Getting real

At the Critical Assessment of Protein Structure Prediction (CASP) competition, AlphaFold matched experimental findings on a measure of accuracy.



CREDITS: (IMAGE) DEEPMIND; (GRAPH) C. BICKEL/*SCIENCE*; (DATA) CASP

the team combined deep learning with an "attention algorithm" that mimics the way a person might assemble a jigsaw puzzle: connecting pieces in clumps—in this case clusters of amino acids—and then searching for ways to join the clumps in a larger whole. Working with a computer network built around 128 machine learning processors, they trained the algorithm on all 170,000 or so known protein structures.

And it worked. In this year's CASP, AlphaFold achieved a median GDT score of 92.4. For the most challenging proteins, AlphaFold scored a median of 87, 25 points above the next best predictions. It even excelled at solving structures of proteins that sit wedged in cell membranes, which are central to many human diseases but notoriously difficult to solve with x-ray crystallography. Venki Ramakrishnan, a structural biologist at the Medical Research Council Laboratory of Molecular Biology, calls the result "a stunning advance on the protein folding problem."

All groups in this year's competition improved, Moult says. But with AlphaFold, Lupas says, "The game has changed." The organizers even worried DeepMind may have cheated somehow. So Lupas set a special challenge: a membrane protein from a species of archaea, an ancient group of microbes. For 10 years, his team had tried to get its x-ray crystal structure. "We couldn't solve it."

But AlphaFold had no trouble. It returned a detailed image of a three-part protein with two helical arms in the middle. The model enabled Lupas and his team to make sense of their x-ray data; within half an hour, they had fit their experimental results to AlphaFold's predicted structure. "It's almost perfect," Lupas says. "They could not possibly have cheated on this. I don't know how they do it."

As a condition of entering CASP, DeepMind—like all groups—agreed to reveal sufficient details about its method for other groups to re-create it. That will be a boon for experimentalists, who will be able to use structure predictions to make sense of opaque x-ray and cryo-EM data. It could also enable drug designers to work out the structure of every protein in new and dangerous pathogens like SARS-CoV-2, a key step in the hunt for molecules to block them, Moult says.

Still, AlphaFold doesn't do everything well. In CASP, it faltered on one protein, an amalgam of 52 small repeating segments, which distort each others' positions as they assemble. Jumper says the team now wants to train AlphaFold to solve such structures, as well as those of complexes of proteins that work together to carry out key functions in the cell.

Even though one grand challenge has fallen, others will undoubtedly emerge. "This isn't the end of something," Thornton says. "It's the beginning of many new things." ■

PHOTO: KEVIN SCHAFER/GETTY IMAGES

WATER POLLUTION

Why were salmon dying? The answer washed off the road

Common tire chemical implicated in coho salmon kills

By **Erik Stokstad**

For decades, something in urban streams has been killing coho salmon in the U.S. Pacific Northwest. Even after Seattle began to restore salmon habitat in the 1990s, up to 90% of the adults migrating up certain streams to spawn would suddenly die after rainstorms. Researchers suspected the killer was washing off nearby roads, but couldn't identify it. "This was a serious mystery," says Edward Kolodziej, an environmental engineer at the University of Washington's (UW's) Tacoma and Seattle campuses.

Online this week in *Science*, researchers led by Kolodziej report the primary culprit comes from a chemical widely used to protect tires from ozone, a reactive atmospheric gas. The toxicant, called 6PPD-quinone, leaches out of the particles that tires shed onto pavement. Even small doses killed coho salmon in the lab. "It's a brilliant piece of work," says Miriam Diamond, an environmental chemist at the University of Toronto. "They've done a tremendous job at sleuthing out a very challenging problem."

Particles that erode from tires wash into streams used by coho salmon.

Manufacturers annually produce some 3.1 billion tires worldwide. Tire rubber is a complex mixture of chemicals, and companies closely guard their formulations. Because tire particles are a common component of water pollution, researchers have been examining how they affect aquatic life.

After Kolodziej arrived at UW's Center for Urban Waters in 2014, he joined the effort to solve the coho salmon mystery. The group created a mixture of particles from nine tires—some bought new, others provided by two undergraduates who moonlight as mechanics—to mimic what might wash off typical highways. They found several thousand unidentified chemicals in the mixture. Postdoc Zhenyu Tian spent more than 2 years narrowing down the list, separating the molecules based on their electrical charge and other properties. By May 2019, he had narrowed the focus to about 50 unknown chemicals, and then further work revealed the chemical formula of a prime suspect. "If you're looking for an unexplained toxicant that's killing fish, we had the perfect instruments and expertise," Kolodziej recalls.

But what was it? A 2019 report from the Environmental Protection Agency on chemicals in recycled tires mentioned 6PPD, which has a similar formula. The final clue was buried in an industry report from 1983, which contained the exact formula of 6PPD-quinone, the molecule created when 6PPD reacts with ozone. The team synthesized 6PPD-quinone and found it was highly lethal to coho salmon.

Now, the team is working to understand how the chemical kills fish. Kolodziej and colleagues say other species of fish should also be evaluated for sensitivity. Because you can't buy the molecule, Kolodziej's team is making it. "My lab might even be the only place that actually has this," he says.

The researchers suspect the compound is present on busy roads everywhere. They've found it washes off pavement and into streams in Los Angeles and San Francisco, for example. The simplest solution might be for tire manufacturers to switch to an environmentally benign alternative. But Sarah Amick, vice president of environment, health, safety, and sustainability at the U.S. Tire Manufacturers Association, says it's too early to discuss alternatives. "It's important that additional research be done to validate and verify these results."

Another way to protect salmon is to filter stormwater through soil, but installing enough infiltration basins to treat road runoff before it reaches spawning streams would be very expensive, says co-author Jenifer McIntyre, an ecotoxicologist at Washington State University's Puyallup Research and Extension Center. In the meantime, Kolodziej says he "can't walk along a street without staring at all the skid marks," thinking about tire chemicals, and "wondering what's there." ■

‘The game has changed.’ AI triumphs at protein folding

Robert F. Service

Science **370** (6521), . DOI: 10.1126/science.370.6521.1144