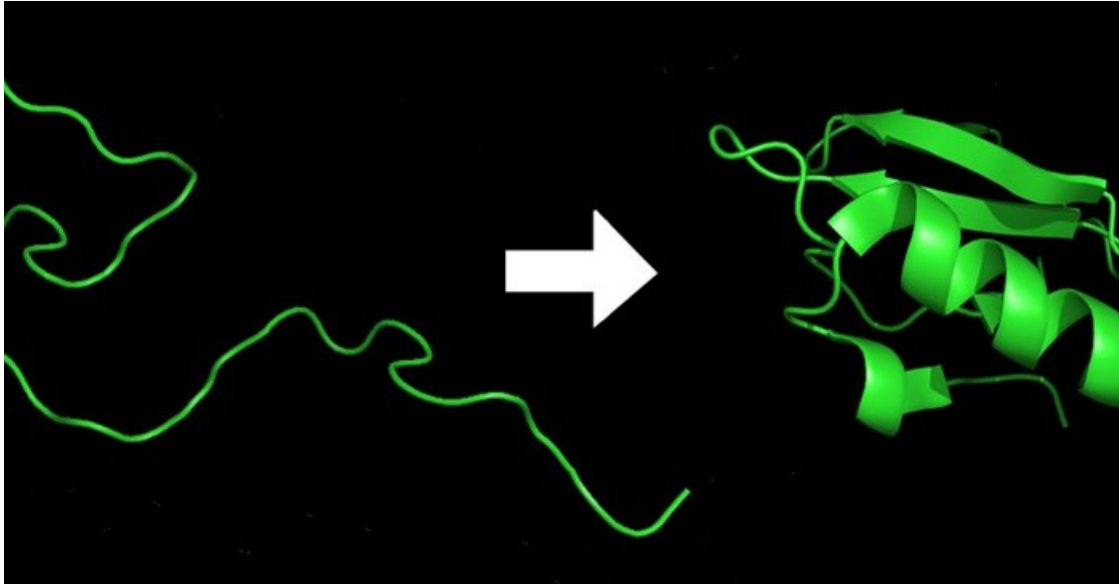


AI makes huge progress predicting how proteins fold

By [Marc Zimmer](#)

7 Dec 2020

Solving what biologists call "the protein-folding problem" is a big deal. Proteins are the workhorses of cells and are present in all living organisms. They are made up of long chains of amino acids and are vital for the structure of cells and communication between them as well as regulating all of the chemistry in the body.



A simple chain of amino acids folds into a complex three-dimensional structure.

This week, the Google-owned artificial intelligence company [DeepMind](#) demonstrated a deep-learning program called [AlphaFold2](#), which experts are calling a [breakthrough](#) toward solving the grand challenge of [protein folding](#).

Proteins are long chains of amino acids linked together like beads on a string. But for a protein to do its job in the cell, it must "fold" – a process of twisting and bending that transforms the molecule into a complex three-dimensional structure that can interact with its target in the cell. If the folding is disrupted, then the protein won't form the correct shape – and it won't be able to perform its job inside the body. This can lead to disease – as is the case in a common disease like Alzheimer's, and rare ones like cystic fibrosis.

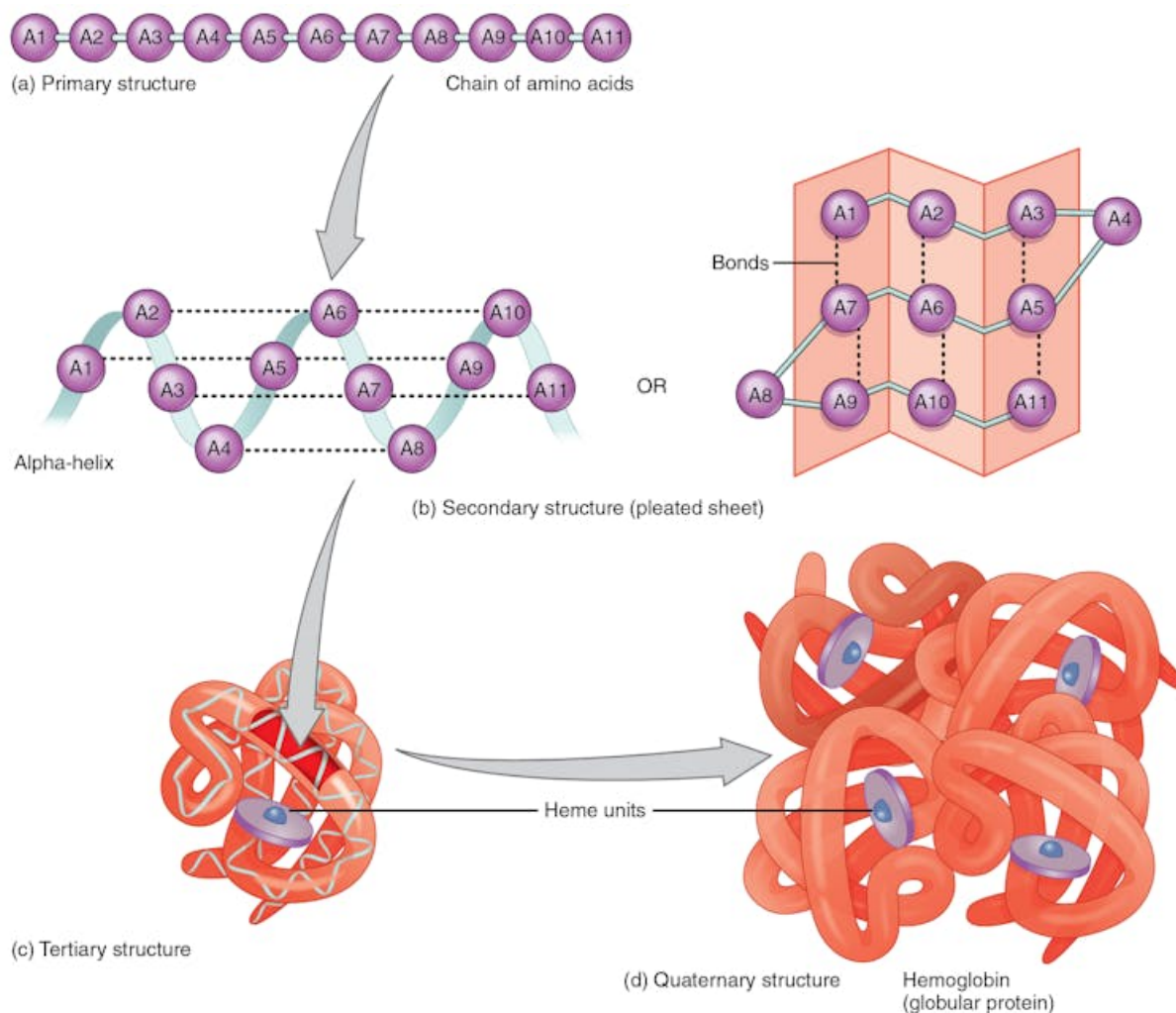
Deep learning is a computational technique that uses the often hidden information contained in vast datasets to solve questions of interest. It's been used widely in fields such as games, speech and voice recognition, autonomous cars, science and medicine.

I believe that tools like AlphaFold2 will help scientists to design new types of proteins, ones that may, for example, help break down plastics and fight future viral pandemics and disease.

[I am a computational chemist](#) and author of the book [The State of Science](#). My students and I study the structure and properties of [fluorescent proteins](#) using protein-folding computer programs based on classical physics.

After decades of study by thousands of research groups, these protein-folding prediction programs are very good at calculating structural changes that occur when we make small alterations to known molecules.

But they haven't adequately managed to predict how proteins fold from scratch. Before deep learning came along, the protein-folding problem seemed impossibly hard, and it seemed poised to frustrate computational chemists for many decades to come.

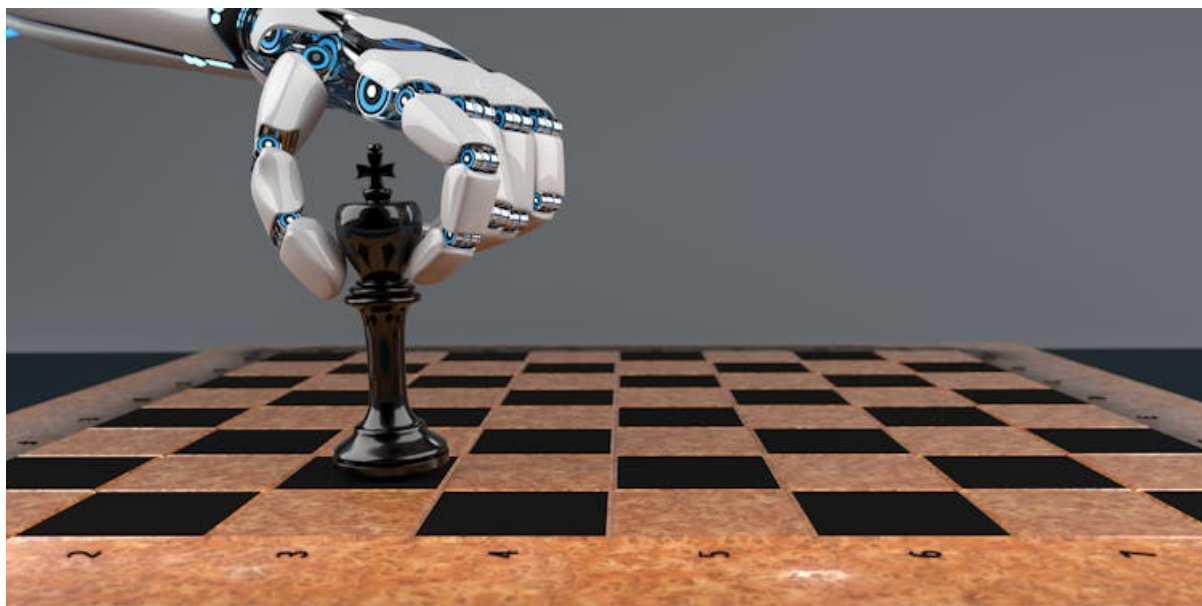


Protein folding

The sequence of the amino acids – which is encoded in DNA – defines the protein's 3D shape. The shape determines its function. If the structure of the protein changes, it is unable to perform its function. Correctly predicting protein folds based on the amino acid sequence could revolutionize drug design, and explain the causes of new and old diseases.

All proteins with the same sequence of amino acid building blocks fold into the same three-dimensional form, which optimises the interactions between the amino acids. They do this within milliseconds, although they have an astronomical number of possible configurations available to them – [about 10 to the power of 300](#). This massive number is what makes it hard to predict how a protein folds even when scientists know the full sequence of amino acids that go into making it. Previously predicting the structure of protein from the amino acid sequence was impossible. Protein structures were experimentally determined, a time-consuming and expensive endeavour.

Once researchers can better predict how proteins fold, they'll be able to better understand how cells function and how misfolded proteins cause disease. Better protein prediction tools will also help us design drugs that can target a particular topological region of a protein where chemical reactions take place.



AlphaFold is born from deep-learning chess, Go and poker games

The success of DeepMind's protein-folding prediction program, called [AlphaFold](#), is not unexpected. Other deep-learning programs written by [DeepMind](#) have demolished the world's best chess, Go and poker players.

In 2016 [Stockfish-8](#), an open-source chess engine, was the world's computer chess champion. It evaluated 70 million chess positions per second and had centuries of accumulated human chess strategies and decades of computer experience to draw upon. It played efficiently and brutally, mercilessly beating all its human challengers without an ounce of finesse. Enter deep learning.

On Dec. 7, 2017, Google's deep-learning chess program [AlphaZero](#) thrashed Stockfish-8. The chess engines played 100 games, with AlphaZero winning 28 and tying 72. It didn't lose a single game. AlphaZero did only 80,000 calculations per second, as opposed to Stockfish-8's 70-million calculations, and it took just four hours to learn chess from scratch by playing against itself a few million times and optimizing its neural networks as it learned from its experience.

[AlphaZero](#) didn't learn anything from humans or chess games played by humans. It taught itself and, in the process, derived strategies never seen before. In a [commentary](#) in Science magazine, former world chess champion Garry Kasparov wrote that by learning from playing itself, AlphaZero developed strategies that "reflect the truth" of chess rather than reflecting "the priorities and prejudices" of the programmers. "It's the embodiment of the cliché 'work smarter, not harder.'"

Casp – the Olympics for molecular modellers

Every two years, the world's top computational chemists test the abilities of their programs to predict the folding of proteins and compete in the [Critical Assessment of Structure Prediction](#) (Casp) competition.

In the competition, teams are given the linear sequence of amino acids for about 100 proteins for which the 3D shape is known but hasn't yet been published; they then have to compute how these sequences would fold. In 2018 AlphaFold, the deep-learning rookie at the competition, beat all the traditional programs – but barely.

Two years later, on Monday, it was announced that Alphafold2 had won the 2020 competition by a healthy margin. It whipped its competitors, and its predictions were comparable to the existing experimental results determined through gold standard techniques like X-ray diffraction crystallography and cryo-electron microscopy. Soon I expect AlphaFold2 and its progeny will be the methods of choice to determine protein structures before resorting to experimental techniques that require painstaking, laborious work on expensive instrumentation.

One of the reasons for AlphaFold2's success is that it could use the [Protein Database](#), which has over 170,000 experimentally determined 3D structures, to train itself to calculate the correctly folded structures of proteins.

The potential impact of AlphaFold can be appreciated if one compares the number of all published protein structures – approximately 170,000 – with the 180 million DNA and protein sequences deposited in the [Universal Protein Database](#). AlphaFold will help us sort through treasure troves of DNA sequences hunting for new proteins with unique structures and [functions](#).

Has AlphaFold made me, a molecular modeller, redundant?

As with the chess and Go programs – AlphaZero and AlphaGo – we don't exactly know what the AlphaFold2 algorithm is doing and why it uses certain correlations, but we do know that it works.

Besides helping us predict the structures of important proteins, understanding AlphaFold's "thinking" will also help us gain new insights into the mechanism of protein folding.

[Deep knowledge, daily. [Sign up for The Conversation's newsletter.](#)]

One of the most common fears expressed about AI is that it will lead to large-scale unemployment. AlphaFold still has a significant way to go before it can consistently and successfully predict protein folding.

However, once it has matured and the program can simulate protein folding, computational chemists will be integrally involved in improving the programs, trying to understand the underlying correlations used, and applying the program to solve important problems such as the protein misfolding associated with many diseases such as Alzheimer's, Parkinson's, cystic fibrosis and Huntington's disease.

AlphaFold and its offspring will certainly change the way computational chemists work, but it won't make them redundant. Other areas won't be as fortunate. In the past robots were able to replace humans doing manual labour; with AI, our cognitive skills are also being challenged.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

ABOUT THE AUTHOR

Marc Zimmer, *Connecticut College*

For more, visit: <https://www.bizcommunity.com>