

PRESS RELEASE

Embargo: 15.30 hrs GMT Monday 30 November 2020

**Artificial intelligence solution to a 50-year-old science challenge could ‘revolutionise’ medical research**

Inside every cell, thousands of different proteins form the machinery that keeps all living things – from humans and plants to microscopic bacteria – alive and well. Almost all diseases, including cancer, dementia and even infectious diseases such as COVID-19, are related to the way these proteins function.

Because each protein's function is directly related to its three-dimensional shape, scientists around the world have strived for half a century to find an accurate and fast method to enable them to discover the shape of any protein.

Today (Monday) researchers at the 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14) will announce that an artificial intelligence (AI) solution to the challenge has been found.

Building on the work of hundreds of researchers across the globe, an AI program called AlphaFold, created by London-based AI lab DeepMind, has proved capable of determining the shape of many proteins. It has done so to a level of accuracy comparable to that achieved with expensive and time-consuming lab experiments.

CASP14 is organised by Dr John Moult (chair), University of Maryland, USA; Dr Krzysztof Fidelis, UC Davis, USA; Dr Andriy Kryshchuk, UC Davis, USA; Dr Torsten Schwede, University of Basel and SIB Swiss Institute of Bioinformatics, Switzerland; and Dr Maya Topf, Birkbeck, University of London, UK and CSSB (HPI and UKE) Hamburg, Germany.

Dr Moult said: “Proteins are extremely complicated molecules, and their precise three-dimensional structure is key to the many roles they perform, for example the insulin that regulates sugar levels in our blood and the antibodies that help us fight infections. Even tiny rearrangements of these vital molecules can have catastrophic effects on our health, so one of the most efficient ways to understand disease and find new treatments is to study the proteins involved.

“There are tens of thousands of human proteins and many billions in other species, including bacteria and viruses, but working out the shape of just one requires expensive equipment and can take years.

“Nearly 50 years ago, Christian Anfinsen was awarded a Nobel Prize for showing that it should be possible to determine the shape of proteins based on their sequence of amino acids – the individual building blocks that make up proteins. That’s why our community of scientists have been working on the biennial CASP challenge.”

Teams taking part in the CASP challenge are given the amino acid sequences for a set of around 100 proteins. While scientists study the proteins in the lab to determine their shape experimentally, about a 100 participating CASP teams from more than 20 countries will try to do the same thing using computers. The results are assessed by independent scientists.

Dr Fidelis said: “The CASP approach has created intense collaboration between researchers working in this field of science and we have seen how it has accelerated scientific developments.

“Since we first ran the challenge back in 1994, we have seen a succession of discoveries, each solving an aspect of this problem, so that computed models of protein structures have become progressively more useful in medical research.”

During the latest round of the challenge, DeepMind’s AlphaFold program has determined the shape of around two thirds of the proteins with accuracy comparable to laboratory experiments\*. AlphaFold’s accuracy with most of the other proteins was also high, though not quite at that level.

The CASP organisers say that this success builds on achievements made in previous CASP rounds, both by the DeepMind team and other participants, and that other teams taking part in CASP14 have also produced some highly accurate structures during this round.

Dr Kryshtafovych said: “What AlphaFold has achieved is truly remarkable and today’s announcement is a win for DeepMind, but it’s also a triumph for team science. The unique and intense way we collaborate with researchers around the world through CASP, and the contributions from many teams of scientists over the years, have brought us to this breakthrough.”

He adds: “Being able to investigate the shape of proteins quickly and accurately has the potential to revolutionise life sciences. Now that the problem has been largely solved for single proteins, the way is open for development of new methods for determining the shape of protein complexes – collections of proteins that work together to form much of the machinery of life, and for other applications.”

Professor Dame Janet Thornton, Director Emeritus of EMBL’s European Bioinformatics Institute (EMBL-EBI), who is not affiliated with CASP or DeepMind, said: “One of biology’s biggest mysteries is how proteins fold to create exquisitely unique three-dimensional structures. Every living thing – from the smallest bacteria to plants, animals and humans – is defined and powered by the proteins that help it function at the molecular level.

“So far, this mystery remained unsolved, and determining a single protein structure often required years of experimental effort. It’s tremendous to see the triumph of human curiosity, endeavour and intelligence in solving this problem. A better understanding of protein structures and the ability to predict them using a computer means a better understanding of life, evolution and, of course, human health and disease.”

(ends)

**For more information contact Kerry Noble, CASP press officer.**

Email: [kerry\\_noble@hotmail.com](mailto:kerry_noble@hotmail.com)

Tel: +44 7446 869 433

### **Notes to editors**

\*AlphaFold produced models for about two-thirds of the CASP14 target proteins with global distance test scores above 90 out of 100. Above the 90-score threshold, remaining differences between the models and the experimental structures are small and of the size expected for experimental artefacts and errors, and alternative low energy local conformations. Note that these CASP targets are single proteins or domains, not protein complexes, which are a next frontier. The global distance test is a measure of how closely the shape of the protein model matches the shape from lab experiments: Zemla A, Venclovas, Moult J, Fidelis K. Processing and evaluation of predictions in CASP4. *Proteins* 2001;Suppl 5:13-21; Zemla A. LGA: A method for finding 3D similarities in protein structures. *Nucleic Acids Res* 2003;31(13):3370-3374).

Results of the CASP14 experiment will be available at [www.predictioncenter.org](http://www.predictioncenter.org) at the end of November and presented at a virtual conference from 30 November to 4 December. Results will also be published in a special issue of the journal 'PROTEINS'.

Funding: CASP operations are partially supported by a grant from the National Institutes of Health, NIH R01GM100482.