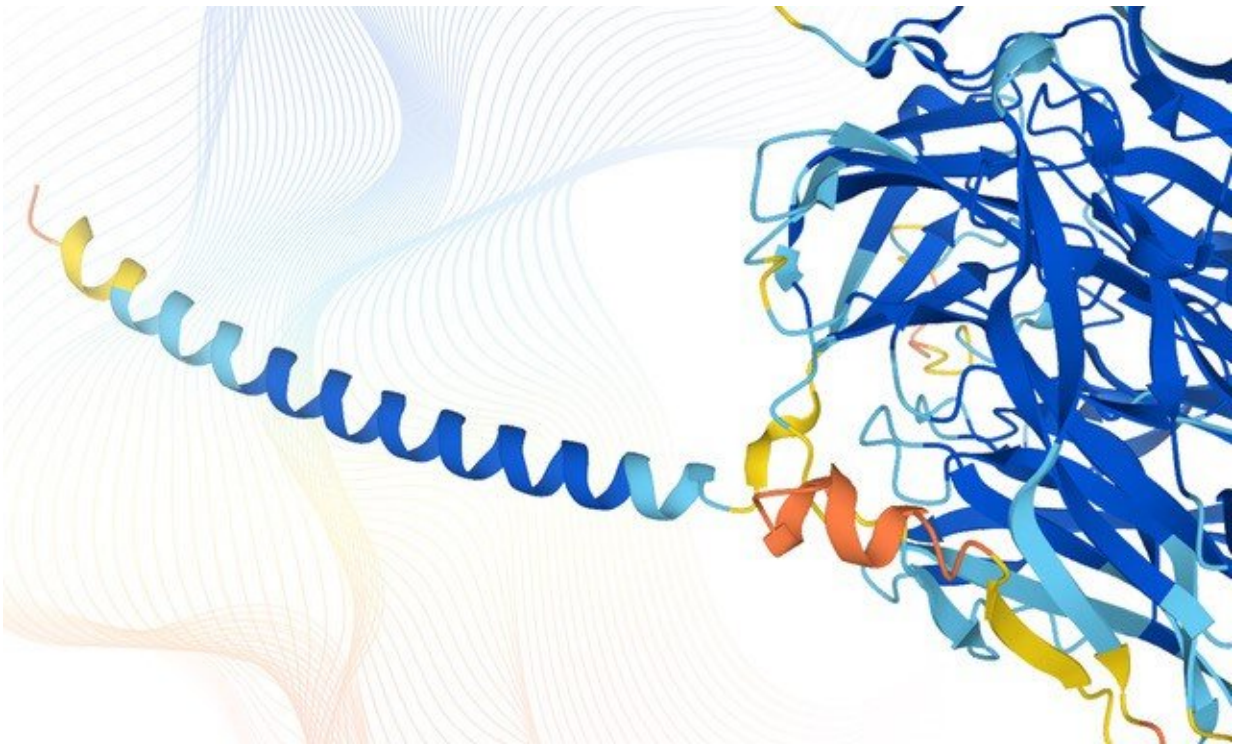


# DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins

July 22 2021

---



Protein structures to represent the data obtained via AlphaFold. Credit: Karen Arnott/EMBL-EBI

DeepMind today announced its partnership with the European Molecular Biology Laboratory (EMBL), Europe's flagship laboratory for the life

sciences, to make the most complete and accurate database yet of predicted protein structure models for the human proteome. This will cover all ~20,000 proteins expressed by the human genome, and the data will be freely and openly available to the scientific community. The database and artificial intelligence system provide structural biologists with powerful new tools for examining a protein's three-dimensional structure, and offer a treasure trove of data that could unlock future advances and herald a new era for AI-enabled biology.

AlphaFold's recognition in December 2020 by the organizers of the Critical Assessment of protein Structure Prediction (CASP) benchmark as a solution to the 50-year-old grand challenge of protein structure prediction was a stunning breakthrough for the field. The AlphaFold Protein Structure Database builds on this innovation and the discoveries of generations of scientists, from the early pioneers of protein imaging and crystallography, to the thousands of prediction specialists and structural biologists who've spent years experimenting with proteins since. The database dramatically expands the accumulated knowledge of protein structures, more than doubling the number of high-accuracy human protein structures available to researchers. Advancing the understanding of these building blocks of life, which underpin every biological process in every living thing, will help enable researchers across a huge variety of fields to accelerate their work.

Last week, the methodology behind the latest highly innovative version of AlphaFold, the sophisticated AI system announced last December that powers these structure predictions, and its open source code were published in *Nature*. Today's announcement coincides with a second *Nature* paper that provides the fullest picture of proteins that make up the [human proteome](#), and the release of 20 additional organisms that are important for [biological research](#).

"Our goal at DeepMind has always been to build AI and then use it as a

tool to help accelerate the pace of scientific discovery itself, thereby advancing our understanding of the world around us," said DeepMind Founder and CEO Demis Hassabis, Ph.D. "We used AlphaFold to generate the most complete and accurate picture of the human proteome. We believe this represents the most significant contribution AI has made to advancing scientific knowledge to date, and is a great illustration of the sorts of benefits AI can bring to society."

## **AlphaFold is already helping scientists to accelerate discovery**

The ability to predict a protein's shape computationally from its amino acid sequence—rather than determining it experimentally through years of painstaking, laborious and often costly techniques—is already helping scientists to achieve in months what previously took years.

"The AlphaFold database is a perfect example of the virtuous circle of open science," said EMBL Director General Edith Heard. "AlphaFold was trained using data from public resources built by the [scientific community](#) so it makes sense for its predictions to be public. Sharing AlphaFold predictions openly and freely will empower researchers everywhere to gain new insights and drive discovery. I believe that AlphaFold is truly a revolution for the [life sciences](#), just as genomics was several decades ago and I am very proud that EMBL has been able to help DeepMind in enabling open access to this remarkable resource."

AlphaFold is already being used by partners such as the Drugs for Neglected Diseases Initiative (DNDi), which has advanced their research into life-saving cures for diseases that disproportionately affect the poorer parts of the world, and the Centre for Enzyme Innovation (CEI) is using AlphaFold to help engineer faster enzymes for recycling some of our most polluting [single-use plastics](#). For those scientists who rely on

experimental protein [structure](#) determination, AlphaFold's predictions have helped accelerate their research. For example, a team at the University of Colorado Boulder is finding promise in using AlphaFold predictions to study antibiotic resistance, while a group at the University of California San Francisco has used them to increase their understanding of SARS-CoV-2 biology.

## **The AlphaFold Protein Structure Database**

The AlphaFold Protein Structure Database builds on many contributions from the international scientific community, as well as AlphaFold's sophisticated algorithmic innovations and EMBL-EBI's decades of experience in sharing the world's biological data. DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) are providing access to AlphaFold's predictions so that others can use the system as a tool to enable and accelerate research and open up completely new avenues of scientific discovery.

"This will be one of the most important datasets since the mapping of the Human Genome," said EMBL Deputy Director General, and EMBL-EBI Director Ewan Birney. "Making AlphaFold predictions accessible to the international scientific community opens up so many new research avenues, from neglected diseases to new enzymes for biotechnology and everything in between. This is a great new scientific tool, which complements existing technologies, and will allow us to push the boundaries of our understanding of the world."

In addition to the human proteome, the database launches with ~350,000 structures including 20 biologically-significant organisms such as E.coli, fruit fly, mouse, zebrafish, malaria parasite and tuberculosis bacteria. Research into these organisms has been the subject of countless research papers and numerous major breakthroughs. These structures will enable researchers across a huge variety of fields—from neuroscience to

medicine—to accelerate their work.

## The future of AlphaFold

The database and system will be periodically updated as we continue to invest in future improvements to AlphaFold, and over the coming months we plan to vastly expand the coverage to almost every sequenced [protein](#) known to science—over 100 million structures covering most of the UniProt reference database.

**More information:** Highly accurate protein structure prediction for the human proteome, *Nature* (2021). DOI: [10.1038/s41586-021-03828-1](https://doi.org/10.1038/s41586-021-03828-1), [www.nature.com/articles/s41586-021-03828-1](https://www.nature.com/articles/s41586-021-03828-1)

Provided by European Molecular Biology Laboratory

Citation: DeepMind and EMBL release the most complete database of predicted 3D structures of human proteins (2021, July 22) retrieved 2 May 2024 from <https://phys.org/news/2021-07-deepmind-embl-database-3d-human.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.