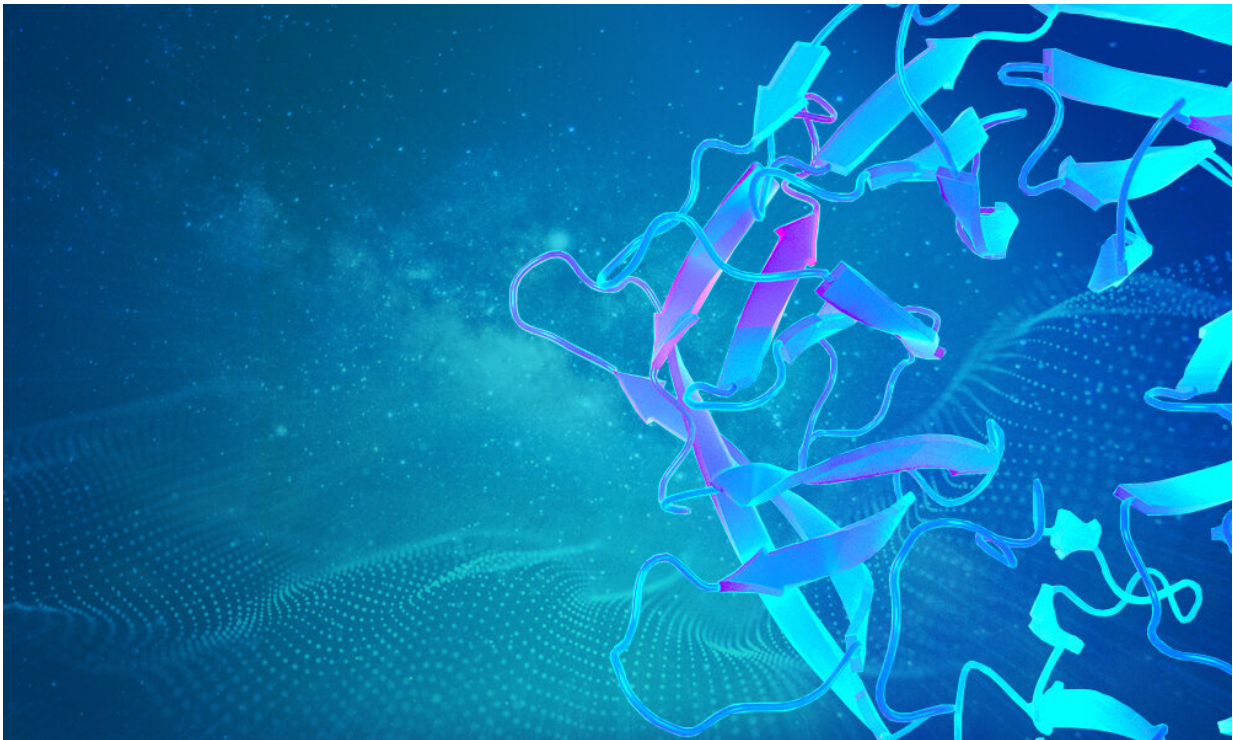


AlphaFold predicts structure of almost every catalogued protein known to science

July 28 2022, by Oana Stroe



Credit: Karen Arnott/EMBL-EBI

DeepMind and EMBL's European Bioinformatics Institute (EMBL-EBI) have made AI-powered predictions of the three-dimensional structures of nearly all cataloged proteins known to science freely and openly available to the scientific community, via the AlphaFold Protein Structure Database.

The two organizations hope the expanded database will continue to increase our understanding of biology, aiding countless more scientists in their work as they look to tackle global challenges.

The database is being expanded by approximately 200 times, from nearly 1 million protein structures to over 200 million, covering almost every organism on Earth that has had its genome sequenced. The expansion of the database includes predicted structures for a wide range of species, including plants, bacteria, animals, and other organisms, opening up new avenues of research across the life sciences that will have an impact on global challenges, including sustainability, food insecurity, and neglected diseases.

Now, almost every [protein sequence](#) on the [UniProt](#) protein database will come with a predicted structure. This release will also open up new research avenues, such as supporting bioinformatics and computational work by allowing researchers to potentially spot patterns and trends in the database.

"AlphaFold now offers a 3D view of the protein universe," said Edith Heard, Director General of EMBL. "The popularity and growth of the AlphaFold Database is testament to the success of the collaboration between DeepMind and EMBL. It shows us a glimpse of the power of multidisciplinary science."

"We've been amazed by the rate at which AlphaFold has already become an essential tool for hundreds of thousands of scientists in labs and universities across the world," said Demis Hassabis, Founder and CEO of DeepMind. "From fighting disease to tackling plastic pollution, AlphaFold has already enabled incredible impact on some of our biggest global challenges. Our hope is that this expanded database will aid countless more scientists in their important work and open up completely new avenues of scientific discovery."

An essential tool for scientists

DeepMind and EMBL-EBI launched the AlphaFold database in July 2021, with more than 350,000 protein structure predictions, including the entire human proteome. Subsequent updates saw the addition of UniProtKB/SwissProt and 27 new proteomes, 17 of which represent neglected [tropical diseases](#) that continue to devastate the lives of more than 1 billion people globally.

In just over a year, more than 1,000 [scientific papers](#) have cited the database and over 500,000 researchers from over 190 countries have accessed the AlphaFold Database to view over two million structures.

The team has also seen researchers building on AlphaFold to create and adapt tools such as [Foldseek](#) and [Dali](#) which allow users to search for entries similar to a given protein. Others have adopted the core machine learning ideas behind AlphaFold, forming the backbone of a slate of new algorithms in this space, or applying them to areas such as RNA structure prediction or in developing new models for designing proteins.

Impact and future of AlphaFold and the database

AlphaFold has also shown impact in areas such as improving our ability to fight [plastic pollution](#), gain insight into Parkinson's disease, increase the [health of honey bees](#), understand how ice forms, tackle neglected diseases such as Chagas disease and Leishmaniasis, and [explore human evolution](#).

"We released AlphaFold in the hopes that other teams could learn from and build on the advances we made, and it has been exciting to see that happen so quickly. Many other AI [research organizations](#) have now entered the field and are building on AlphaFold's advances to create

further breakthroughs. This is truly a new era in [structural biology](#), and AI-based methods are going to drive incredible progress," said John Jumper, Research Scientist and AlphaFold Lead at DeepMind.

"AlphaFold has sent ripples through the molecular biology community. In the past year alone, there have been over a thousand [scientific articles](#) on a broad range of research topics which use AlphaFold structures; I have never seen anything like it," said Sameer Velankar, Team Leader at EMBL-EBI's Protein Data Bank in Europe. "And this is just the impact of one million predictions; imagine the impact of having over 200 million [protein](#) structure predictions openly accessible in the AlphaFold Database."

DeepMind and EMBL-EBI will continue to refresh the [database](#) periodically, with the aim of improving features and functionality in response to user feedback. Access to structures will continue to be fully open, under a CC-BY 4.0 license, and bulk downloads will be made available via [Google Cloud Public Datasets](#).

More information: Database: alphafold.ebi.ac.uk/

Provided by European Molecular Biology Laboratory

Citation: AlphaFold predicts structure of almost every catalogued protein known to science (2022, July 28) retrieved 5 May 2024 from <https://phys.org/news/2022-07-alphafold-catalogued-protein-science.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.