

# AlphaFold and beyond



**A great deal has happened in the protein structure prediction field since *Nature Methods* selected this topic as our Method of the Year 2021. Here's a quick, non-comprehensive update.**

In 2021, our choice for the [Method of the Year](#) was clear: major advances in three-dimensional protein structure prediction from amino acid sequence were expected to have a transformative impact in structural biology. In particular, deep learning-based approaches for structure prediction, exemplified by [AlphaFold](#) from DeepMind, made an astounding leap in accuracy, allowing the structures of many proteins to be predicted with high confidence. We certainly have not been disappointed with our choice; the pace of rapid new development in the structure prediction field since has even exceeded our expectations.

In July 2021, DeepMind published the results of applying AlphaFold to predict highly accurate structures for [98.5% of the human proteome](#), about 20,000 proteins in total – a feat that cemented our Method of the Year decision. One year later, DeepMind had increased this number by 10,000-fold. The [AlphaFold Protein Structure Database](#) currently contains an incredible ~200 million predicted protein structures from hundreds of thousands of species.

Other groups have tried their hand at besting AlphaFold. One preprint that has received a great deal of attention comes from authors at Meta AI, who developed a large language model-based protein structure prediction method called [ESMFold](#). Whereas AlphaFold relies on multiple sequence alignments and template structures, ESMFold requires only a single input sequence. Though ESMFold does not quite meet the accuracy of AlphaFold, according to the authors, it is an order of magnitude faster and is also uniquely able to predict accurate structures for orphan proteins with limited sequence homologs. The authors present their predictions for 617 million microbial proteins in the [ESM Metagenomic Atlas](#). Another, recently published tool called [trRosettaX-Single](#) also takes single sequences as inputs. The authors show it can predict the

structures of orphan proteins with higher accuracy than AlphaFold and also works well to predict the structures of designed proteins. In general, the protein design field stands to benefit hugely from large language models, as exemplified by the recently published method [ProGen](#), which generates novel sequences with predictable function.

AlphaFold is powerful, but it also requires powerful computing infrastructure. In particular, the process of building multiple sequence alignments requires computationally intensive searches of protein databases using homology detection methods. Earlier this year, we published [ColabFold](#), which allows users to perform homology searches 40- to 60-fold faster than AlphaFold, enabling a thousand structures to be predicted in a day using a server with one graphics processing unit.

Researchers are also using experimental information to help boost the performance of AlphaFold and other prediction tools, as exemplified by a paper we published a few months ago. Using a [hybrid, iterative procedure](#), experimental information such as a density map allows greater portions of models to be predicted with higher accuracy than using AlphaFold alone.

An exciting frontier in the protein structure prediction field is the harder challenge of accurately predicting protein function. Understanding what a protein binds to, whether another protein or proteins or a small-molecule ligand, and what that binding interaction looks like can give researchers important clues about function. A preprint from the DeepMind group presents [AlphaFold-Multimer](#), a method for predicting multichain protein complexes. When applied to a dataset of ~4,500 protein complexes, it predicted the interface of heteromeric complexes with high accuracy 26% of the time, and with 36% accuracy for homomeric complexes – a good start, but clearly a problem that needs more development to fully crack. Another group [combined AlphaFold with Monte Carlo tree search](#) to predict the structures of large protein complexes, some with high accuracy. However, the method requires that the stoichiometry of the components be known, and it did not perform well for asymmetric complexes.

In this issue, we present [AlphaFill](#), an algorithm that ‘transplants’ small-molecule ligands and ions from experimentally solved structures to protein models predicted by AlphaFold. The predictions for nearly a million models are available in the [alphafill.eu databank](#).

There has also been much interest in developing methods for accurately predicting three-dimensional RNA structures. Last year, we highlighted a [pioneering approach](#) based on geometric deep learning. More recently, several preprints have been released, one describing a tool called [DeepFoldRNA](#) and another [RoseTTAFoldNA](#), which models both RNA structures and protein–nucleic acid complexes.

Though these methods have been rightly celebrated as major advances with implications across many areas of research, we should not blindly accept prediction results as biological truth. An important, independent, [community-driven effort](#) assessed how well AlphaFold could be used to predict the effects of missense variants, ligand binding sites and model interactions, among other tasks. A recently posted preprint independently assessed the accuracy of AlphaFold predictions in comparison to density maps from recently solved crystal structures, [finding inconsistencies in domain orientation and backbone and side chain conformation](#). This led the authors to conclude that “while AlphaFold predictions are useful hypotheses about protein structures, experimental information remains essential for creating an accurate model.” In this issue, a [Comment](#) makes the case that AlphaFold and its siblings are fundamentally limited in that they predict only single structures, whereas structural distributions would better represent the dynamic, structurally heterogeneous nature of proteins.

Deep learning-based methods are here to stay, in protein structure prediction and far beyond. We are excited to see how such methods can be extended to tackle the more difficult challenges of protein function prediction and RNA structure determination. We will continue watching and highlighting how such approaches are transforming of the practice of structural biology.

Published online: 10 February 2023