

Testing the limits of AlphaFold2's accuracy in predicting protein structure

November 22 2023

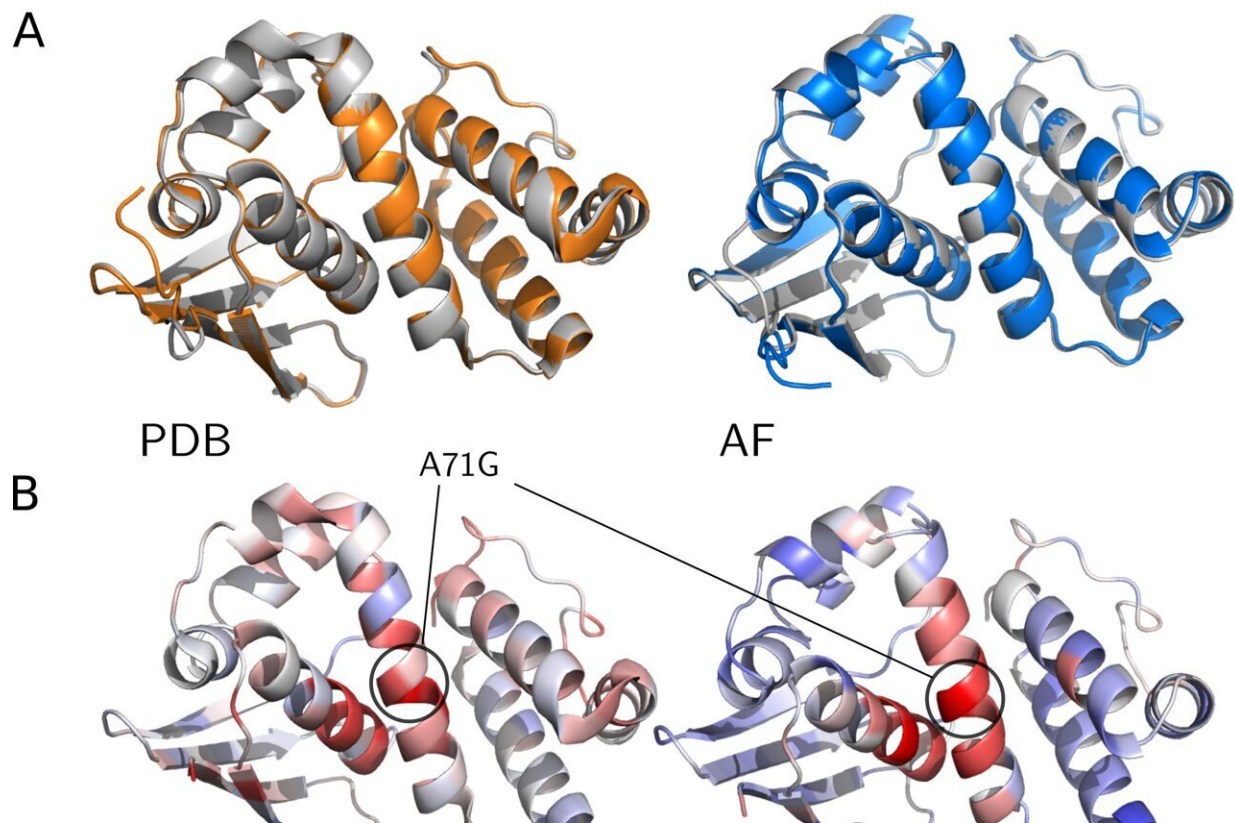


Figure 1. A: Overlaid wild-type (grey) and mutant (color), experimental (orange), and predicted (blue) structures of H-NOX protein. B: Wild-type protein with residues colored by strain (a measure of structural deformation), S_i ; the location of the mutation (residue 71 is mutated from alanine (A) to glycine (G)) is indicated. C: Strain, S_i , per residue along the protein sequence for both experimental and predicted structures; mutation location is indicated with the dotted line. Credit: *Physical Review Letters* (2023). DOI: 10.1103/PhysRevLett.131.218401

Proteins, the workhorses of biology, are encoded by DNA sequences and are responsible for vital functions within cells. Since the first experimental measurement of a protein structure was made by John Kendrew in the 1950s, protein's ability to fold into complex three-dimensional structures has long been a subject of scientific fascination and importance. However, determining these structures experimentally has remained a formidable challenge for decades.

IBS researcher John M. McBride said, "Sequencing DNA is a far simpler process than analyzing [protein](#) structures. For example, let's compare the progress of research in DNA and proteins. So far, we have sequenced hundreds of millions of DNA sequences, while on the other hand, we have managed to characterize only several hundred thousand protein structures."

Hence, a central focus of computational biology has been predicting protein structures from their sequences. Understanding a protein's structure is pivotal for deciphering its functions, delving into diseases, unraveling aging, and engineering proteins for various technological applications.

Google DeepMind extended the application of artificial intelligence into the biophysics domain. The company's AlphaFold2 represents the latest milestone in tackling the problem of protein structure prediction, bridging the gap between computational predictions and experimental accuracy. This achievement is substantial enough for some to declare the problem of protein structure prediction as "solved."

But the question is: How accurate is it?

Despite AlphaFold2's success in predicting protein structures, questions

remain regarding the limits of its accuracy. A fundamental concern arises when attempting to predict the effects of the tiniest changes in a protein—for example, in single point mutation where a single amino acid is substituted with another of differing chemical properties. Achieving accuracy at this level is essential for studying diseases and evolution.

There is skepticism about whether AlphaFold2 can achieve such accuracy. The official AlphaFold database clearly states, "AlphaFold has not been validated for predicting the effect of [mutations](#). In particular, AlphaFold is not expected to produce an unfolded protein structure given a sequence containing a destabilizing point mutation." Additionally, several recent assessments have failed to provide evidence that AlphaFold can predict mutation effects.

Researchers from the Center for Soft and Living Matter within the Institute for Basic Science (South Korea), recently explored the limits of AlphaFold2 AI's ability to predict protein structures. The team used a two-pronged approach to provide a compelling, comprehensive demonstration that AlphaFold can indeed predict mutation effects. First, they directly validated AlphaFold predictions by comparing them with experimental structures. The researchers combined this with an indirect validation of AlphaFold predictions comparing AlphaFold-predicted mutation effects on structure to empirical measurements of protein phenotypes.

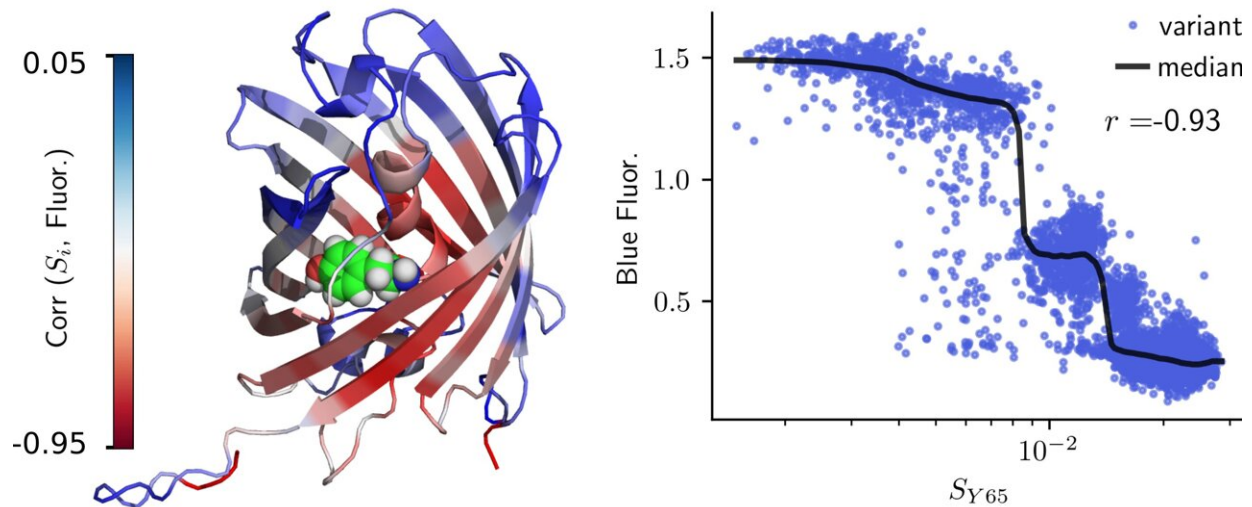


Figure 2. The structure of the blue fluorescent protein is shown, colored according to how well strain at each residue correlates with fluorescence. The atoms of the tryptophan residue (Y65) that bind to a chromophore are shown by spheres. Deformation at this residue, S_{Y65} , leads to decreases in fluorescence in a two-step manner. Credit: *Physical Review Letters* (2023). DOI: 10.1103/PhysRevLett.131.218401

However, this whole process was extremely challenging.

The first major obstacle was that there was very little data that could be used for comparison. Even though there are more than half a million structures in the public Protein Data Bank (PDB), only a small fraction of these can be used to measure mutation effects. After rigorous data selection and controlling for various factors, researchers were left with just a few thousand proteins with experimental structures involving minor amino acid changes. These data also contained lots of random noise, which made it challenging to distinguish between structural variations due to measurement error and those caused by mutations.

Despite this, the researchers showcased that mutation effects can be

statistically measured using experimental structures, providing a robust methodology for quantifying these effects. By applying this methodology, researchers demonstrated that AlphaFold's predictions are nearly as accurate as experimental measurements.

The second challenge involves the inadequacy of typical structural similarity measures to capture structural differences due to mutations. Conventional measurements, like the root-mean-square deviation (RMSD), primarily account for changes across the entire protein structure, obscuring small local effects in mutated regions. Local measurements such as the local distance difference test (LDDT) also have low resolution and are limited in their ability to capture fine differences.

In response, the research team adopted tools from physics, specifically concepts from continuum mechanics, to measure strain in proteins, a natural measure of deformation. They tested this approach on measurements of fluorescence in several fluorescent proteins. It was found that AlphaFold can accurately predict deformation at the chromophore-binding site (which is important for fluorescence), leading to accurate predictions of fluorescence in fluorescent proteins.

Two years after the release of AlphaFold2 we are still exploring the limits and the pitfalls of this fantastic new algorithm. This first successful validation of AlphaFold for predicting mutation effects paves the way for investigations into disease and drug development, leading to improvements in human health. The ability to predict mutation effects will enhance the study of evolution, looking both forward—using directed evolution to develop new enzymes—and backward—understanding the evolutionary history of life itself. The future of protein science is indeed bright.

The work is [published](#) in the journal *Physical Review Letters*.

More information: John M. McBride et al, AlphaFold2 Can Predict Single-Mutation Effects, *Physical Review Letters* (2023). [DOI: 10.1103/PhysRevLett.131.218401](https://doi.org/10.1103/PhysRevLett.131.218401)

Provided by Institute for Basic Science

Citation: Testing the limits of AlphaFold2's accuracy in predicting protein structure (2023, November 22) retrieved 2 May 2024 from <https://phys.org/news/2023-11-limits-alphafold2-accuracy-protein.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.