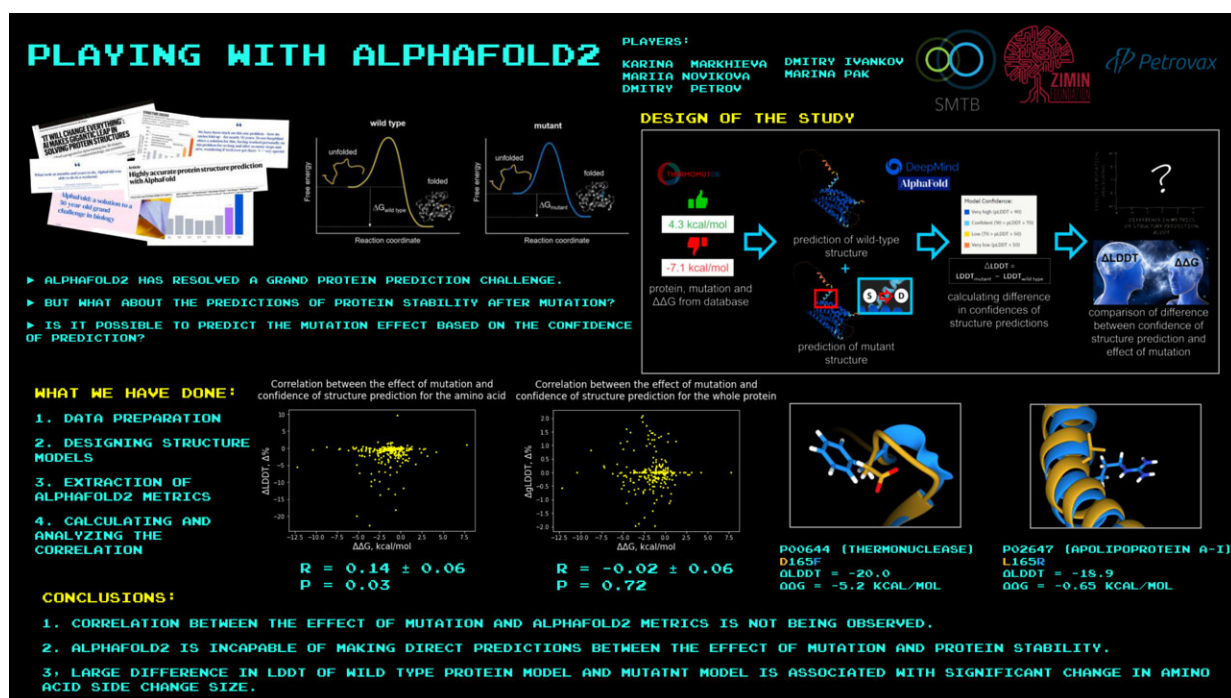


AlphaFault: High schoolers give fabled AI a problem it can't crack

April 7 2023



Poster of the project Playing With AlphaFold2 at the School of Molecular and Theoretical Biology held by Skoltech online in 2021. Credit: Dmitry Ivankov/Skoltech

A bioinformatics boot camp for high schoolers at Skoltech turned into a venue for the latest chapter in the ongoing contest between humans and artificial intelligence in science. Having earlier resolved a key 50-year-old problem of structural bioinformatics, the breakthrough AI program

AlphaFold proved inapplicable to another challenge researchers in this field are faced with.

This finding is reported in a *PLOS ONE* study, whose authors refute the claims by some AlphaFold enthusiasts that DeepMind's AI has mastered the ultimate protein physics and is the be-all and end-all of structural bioinformatics.

Structural bioinformatics is a branch of science that explores the structures of proteins, RNA, DNA and their interactions with other molecules. The findings supply the basis for drug discovery and the creation of proteins with exciting properties, such as the catalysts of reactions not seen in the natural world.

Historically, the central problem of structural bioinformatics was predicting protein structures. That is, given an arbitrary sequence of amino acids that comprise a protein, how do you reliably compute what 3D shape that protein will assume in the body—and therefore how it will function.

After 50 years, the problem was resolved by AlphaFold, an [artificial intelligence](#) program created by Google's DeepMind, whose predecessors earlier made headlines by achieving superhuman performance in chess, the game of Go, and the video game StarCraft II.

This milestone achievement led to speculations that the neural network must have somehow internalized the underlying physics of proteins and should work beyond the task it was designed for. Some people, even in the structural bioinformatics community, expected that the AI would soon give the definitive answers to that discipline's remaining questions and consign it to the history of science.

"We decided to settle this and put AlphaFold to work on another central

task of structural bioinformatics: predicting the impact of single mutations on protein stability. That means you choose a certain known protein and introduce exactly one mutation, the smallest change possible. And you want to know whether the resulting mutant is more stable or less stable and to what extent. AlphaFold was clearly unable to do this, as evidenced by its predictions contradicting the known experimental findings," the study's principal investigator, Assistant Professor Dmitry Ivankov of Skoltech Bio, said.

Asked about the role of the high school students taking part in the project, the researcher said they were involved in mutation data processing, writing scripts for handling prediction results, visualizing the structures specified by AlphaFold, and basically fooling around with the online version of the AI.

Ivankov emphasized that AlphaFold's creators never actually claimed that the AI was applicable to other tasks besides predicting protein structures based on their amino acid sequences. "But some machine learning enthusiasts were quick to prophesy the end of structural bioinformatics. So we thought it a good idea to go ahead and check, and we now know it cannot predict the effect of single mutations," Ivankov added.

On a practical level, predicting how single mutations affect protein stability is useful for sifting through the many possible mutations to determine which ones might be useful. This comes in handy, for example, if you want to make a protein additive for [laundry detergents](#) resistant to higher temperatures so it could break down the fats, starch, fibers, or other proteins in hotter water. Also, sweet proteins are known that could someday be used in place of sugar, provided they can withstand the heat of a cup of coffee or tea.

On a more fundamental level, the findings of the study show that the

artificial intelligence of today is no cure-all, and while it might be wildly successful in solving one problem, others remain, including a dozen or so major challenges in structural bioinformatics. Among them are predicting the structures of complexes made up of proteins and either small molecules or DNA or RNA, determining how mutations affect the binding energy of proteins with other molecules, and designing proteins with [amino acid sequences](#) that endow them with desired properties, such as the ability to catalyze otherwise impossible reactions, serving as an element of a tiny "molecular factory."

Besides issuing a reminder that even in the wake of AlphaFold, scientists in their field have one or two things to do, the authors of the study in *PLOS ONE* examine the contention that the AI program's success stems from its "having learned physics," as opposed to just internalizing the totality of the protein structures known to humanity and cleverly manipulating them. Apparently this is not the case, because knowing the physics involved, it should be relatively easy to compare two very similar but not identical structures in terms of their stability, but it is precisely the task AlphaFold did not accomplish.

This point is supported by two previously voiced reservations regarding the AI's "knowledge" of physics. First, AlphaFold predicts some structures with side groups dangling in a way that suggests a zinc ion to be bound to them. However, the program's input is limited to the protein's amino acid sequence, so the only reason why the "invisible zinc" is there is that the AI was trained on analogous protein structures bound to this ion. Without the zinc, the predicted side group orientation contradicts physics.

Second, AlphaFold can predict a solitary protein structure that looks sort of like a spiral and is indeed accurate—provided that it is interlaced with two other such chains. Without them, the prediction is physically unsound. So rather than rely on physics, the program must be simply

reproducing a shape it isolated from a compound structure.

"Interestingly, this research grew out of a 'playful' project featuring the participants of the School of Molecular and Theoretical Biology. We called it 'Games With AlphaFold.' The moment AlphaFold became openly accessible, our lab installed it on the Zhores supercomputer. One of the games involved comparing the known mutation effects with what AlphaFold predicts for the original and the mutant proteins. This led to a study, in which high schoolers got the chance to simultaneously experience a supercomputer and [advanced artificial intelligence](#)," the study's lead author, Skoltech Ph.D. student Marina Pak, said.

More information: Marina A. Pak et al, Using AlphaFold to predict the impact of single mutations on protein stability and function, *PLOS ONE* (2023). [DOI: 10.1371/journal.pone.0282689](https://doi.org/10.1371/journal.pone.0282689)

Provided by Skolkovo Institute of Science and Technology

Citation: AlphaFault: High schoolers give fabled AI a problem it can't crack (2023, April 7) retrieved 21 June 2024 from <https://phys.org/news/2023-04-alphafault-high-schoolers-fabled-ai.html>

<p>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.</p>
--