# Using AlphaFold to predict the impact of single mutations on protein stability and function

Marina A. Pak[1], Karina A. Markhieva[2,*], Mariia S. Novikova[3,*], Dmitry S. Petrov[4,*], Ilya S. Vorobyev[1], Ekaterina S. Maksimova[5], Fyodor A. Kondrashov[5], and Dmitry N. Ivankov[1,†]

[1]Center of Life Sciences, Skolkovo Institute of Science and Technology, Moscow, Russia
[2]Peoples' Friendship University of Russia (RUDN University), Moscow, Russia
[3]Armand Hammer United World College of the American West, New Mexico, USA
[4]Specialized Educational and Scientific Center of UrFU (SUNC UrFU), Ekaterinburg, Russia
[5]Institute of Science and Technology Austria, Maria Gugging, Austria
[*]Equal contribution
[†]Corresponding author

## Abstract

AlphaFold changed the field of structural biology by achieving three-dimensional (3D) structure prediction from protein sequence at experimental quality. The astounding success even led to claims that the protein folding problem is "solved". However, protein folding problem is more than just structure prediction from sequence. Presently, it is unknown if the AlphaFold-triggered revolution could help to solve other problems related to protein folding. Here we assay the ability of AlphaFold to predict the impact of single mutations on protein stability ($\Delta\Delta G$) and function. To study the question we extracted metrics from AlphaFold predictions before and after single mutation in a protein and correlated the predicted change with the experimentally known $\Delta\Delta G$ values. Additionally, we correlated the AlphaFold predictions on the impact of a single mutation on structure with a large scale dataset of single mutations in GFP with the experimentally assayed levels of fluorescence. We found a very weak or no correlation between AlphaFold output metrics and change of protein stability or fluorescence. Our results imply that AlphaFold cannot be immediately applied to other problems or applications in protein folding.

**Abbreviations:** 3D, three-dimensional; CASP, critical assessment of protein structure prediction; PDB, protein data bank; NMR, nuclear magnetic resonance; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; PCC, Pearson correlation coefficient; LDDT, local distance difference test; pLDDT, per-residue local distance difference test; GFP, green fluorescence protein; BLAST, basic local alignment search tool

## 1 Introduction

AlphaFold is widely claimed to have revolutized protein 3D structure prediction from protein sequence, a 50-years long-standing challenge of protein physics and structural bioinformatics [1]. The fourteenth round of CASP, a blind competition on protein 3D structure prediction [2], demonstrated that AlphaFold, a newcomer to the field, significantly outperforms all other methods. Crucially, AlphaFold models showed an accuracy of their predicted structures that was comparable to structures solved by experimental methods, like X-ray crystallography, NMR, and Cryo-EM [3].

'It will change everything', said Andrei Lupas in an interview to Nature [3]. One of the primary changes may be that AlphaFold may also solve other problems related to protein folding. These

problems include the prediction of various protein interactions, such as protein-protein, protein-ligand and protein-DNA/RNA, and the prediction of the impact of mutations on protein stability. AlphaFold proved to be useful for experimental determination of protein structures with molecular replacement phasing [4, 5] and already facilitated elucidation of SARS-Cov2 protein structures [6, 7]. Furthermore, AlphaFold in collaboration with EMBL-EBI launched a global initiative on constructing the structure models for the whole protein sequence space [8]. The database of freely available structures of all human proteins, and other 20 key organisms yet to be determined, is attributed to "revolutionize the life sciences" [3]. Furthermore, AlphaFold is expected to bring new insights into our understanding of the structural organization of proteins, boost the development of new drugs and vaccines [9]. Researchers in the field are already actively testing AlphaFold performance in various bioinformatics tasks, for instance, in peptide-protein docking [10, 11].

Guided by the expected immediate impact of AlphaFold for the solution of a wide range of problems in structural bioinformatics, we explored the capacity of AlphaFold predictions to serve as a proxy for the impact of mutations on protein stability change ($\Delta\Delta G$). Although AlphaFold provides a disclaimer that it "has not been validated for predicting the effect of mutations" (https://alphafold.ebi.ac.uk/faq), the expectations of AlphaFold are so high that we judged it prudent to check how well AlphaFold predictions could work for estimation of $\Delta\Delta G$ values. We found that the difference between pLDDT scores, the only local AlphaFold prediction metric reported in the output PDB file, had a very weak correlation with experimentally determined $\Delta\Delta G$ values (Pearson correlation coefficient, PCC = 0.17). The difference in the global AlphaFold metric - the pLDDT averaged for all residues - shows no correlation, both isolated and in combination with the mutated residue's pLDDT score. Similarly, the same AlphaFold metrics had a very weak correlation with the impact of single mutations on protein function, fluorescence, of GFP. Recent results show that the use of AlphaFold models instead of template structures does not improve $\Delta\Delta G$ prediction by FoldX [12] [13]. Taken together, so far we do not see a use for AlphaFold to help solve the problem of predicting the impact of a mutation on protein stability. The availability of AlphaFold models allows applying more accurate 3D protein structure-based $\Delta\Delta G$ predictors rather than sequence-based $\Delta\Delta G$ predictors; the bottleneck still seems to be the accuracy of current 3D protein structure-based $\Delta\Delta G$ predictors.

## 2  Results

### 2.1  Data set of mutations

We used experimental data on protein stability changes upon single-point variations from Thermo-Mut Database [14]. After the filtering procedure (see Methods) we randomly chose 976 mutations in 90 proteins for our analysis. For the multiple linear regression analysis, the dataset was split into two sets, a training and a testing set. The split was based on BLAST [15] results, such that the mutations were assigned to the testing set if corresponding proteins had <50% sequence identity to any other protein in the entire dataset (see Methods). All of the other mutations were assigned to the training set.

### 2.2  AlphaFold prediction metrics

Along with coordinates of all heavy atoms for a protein, AlphaFold model contains "its confidence in form of a predicted lDDT-C$\alpha$ score (pLDDT) per residue" [1]. LDDT ranges from 0 to 100 and is a superposition-free metric indicating to what extent the protein model reproduces the reference structure [16]. The pLDDT scores averaged across all residues designate the overall confidence for the whole protein chain (<pLDDT>). For each mutation in the dataset, we calculated the difference in pLDDT between the wild type and mutated structures in the mutated position as well as the difference in <pLDDT> between wild type and mutant protein structure models. By checking $\Delta$pLDDT and $\Delta$<pLDDT> values as potential proxies for the change of protein stability we explored the hypothesis that the change of protein stability due to mutation is somehow reflected in the difference of AlphaFold confidence between wild type and mutant structures.

### 2.3  Correlation between $\Delta\Delta G$ and $\Delta$pLDDT values

First, we studied the relationship between the effect of mutation on protein structure stability and the difference in the accuracy of protein structure prediction by AlphaFold for the wild-type and mutant proteins. We did not observe a pronounced correlation between the mutation effect and

the difference in confidence metrics (Figure 1). The correlation coefficient is $0.17 \pm 0.03$ (p-value $< 10^{-7}$) for $\Delta$pLDDT and $0.00 \pm 0.03$ (p-value $= 0.93$) for the $\Delta$<pLDDT>.
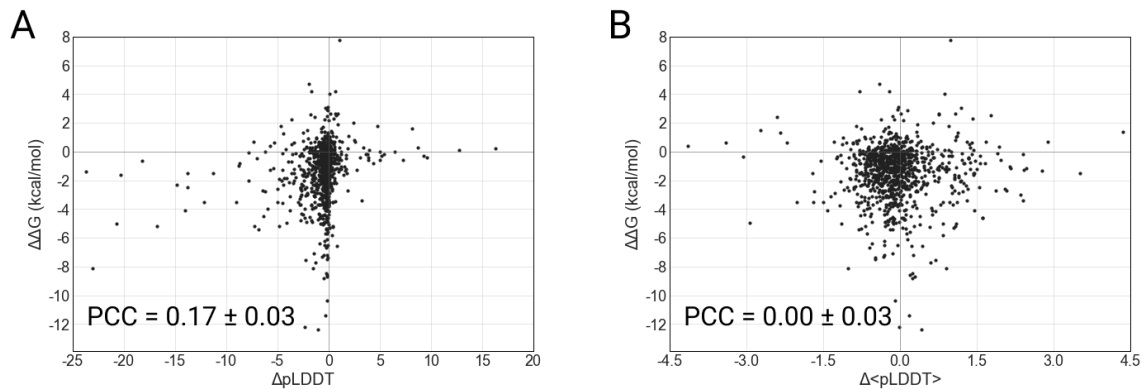


Figure 1: Correlation between the effect of mutation and change of confidence score of structure prediction for the mutated amino acid (A) and for the whole structure (B).

Since the confidence metrics for a given amino acid and whole protein are weakly correlated (PCC $= 0.20 \pm 0.03$, p-value $< 10^{-9}$) we then explored how their combination correlates with the effect of mutation. Multiple linear regression model resulted in the dependence $\Delta\Delta G = -1.08 + 0.11\cdot\Delta$pLDDT $+ 0.05\cdot\Delta$<pLDDT>. We did not obtain any pronounced correlation either for training ($0.12 \pm 0.05$, p-value $= 0.03$) or test sets ($0.20 \pm 0.04$, p-value $< 10^{-5}$).

## 2.4 Relationship of $\Delta$pLDDT and amino acid size change

We explored the outliers with high absolute values of $\Delta$pLDDT in Figure 1A. We noticed the association between $\Delta$pLDDT score and the change of the side chain size upon mutation; to check the association, we conducted the Kruskal-Wallis H-test. The $\Delta$pLDDT scores within different changes in amino acid volume were significantly different (p-value $= 0.02$). We observed that most of the outlier mutations had a destabilizing effect. We assessed the correlation between the destabilizing effect of a mutation and the decrease in the confidence score. The PCC was greater than for the whole set of mutations, but still not strong (PCC $= 0.38 \pm 0.08$, p-value $< 10^{-5}$). However, the size-changing mutations also populate the region of near-zero pLDDT values, so these mutations also poorly predict $\Delta\Delta G$ from $\Delta$pLDDT (PCC $= 0.17 \pm 0.07$, p-value $= 0.01$).

## 2.5 Correlation between GFP fluorescence and $\Delta$pLDDT values

Protein stability is intimately coupled with protein functionality. Thus, a reasonable hypothesis holds that the loss of protein functionality due to mutations in most cases results from reduced stability [17]. Therefore, along with testing correlation of AlphaFold metrics with $\Delta\Delta G$, it is reasonable to test the correlation of AlphaFold metrics with protein function. Furthermore, the change of pLDDT scores may contribute directly to protein functionality without contributing to protein stability. We checked the correlation between $\Delta$pLDDT values and the fluorescent level of 447 randomly chosen single GFP mutants from [18]. The correlation coefficient is $0.14 \pm 0.05$ (p-value $= 0.003$) for $\Delta$pLDDT and $0.16 \pm 0.05$ (p-value $< 10^{-3}$) for the $\Delta$<pLDDT> (Figure 2).

## 3 Discussion

Extraordinary success of AlphaFold in predicting protein 3D structure from protein sequence may lead to temptation to apply this tool to other questions in structural bioinformatics. Here we checked the potential of AlphaFold metrics to serve as a predictor for the impact of mutation on protein stability and function. We found a weak correlation of $0.17 \pm 0.03$ between $\Delta$pLDDT and $\Delta\Delta G$ associated with specific mutations. Although the correlation was statistically significant (p-value $< 10^{-7}$), it is so weak that it cannot be used for accurate $\Delta\Delta G$ predictions (Figure 1) and it is unclear how such predictions can be used in practical applications. Clearly, $\Delta$pLDDT would show a better correlation with $\Delta\Delta G$ if it was measured across bins of averaged $\Delta\Delta G$. Alternatively, $\Delta$pLDDT could be a separate term in a multiple linear regression model. The averaged metric $\Delta$<pLDDT> shows correlation with $\Delta\Delta G$, which is statistically indistinguishable
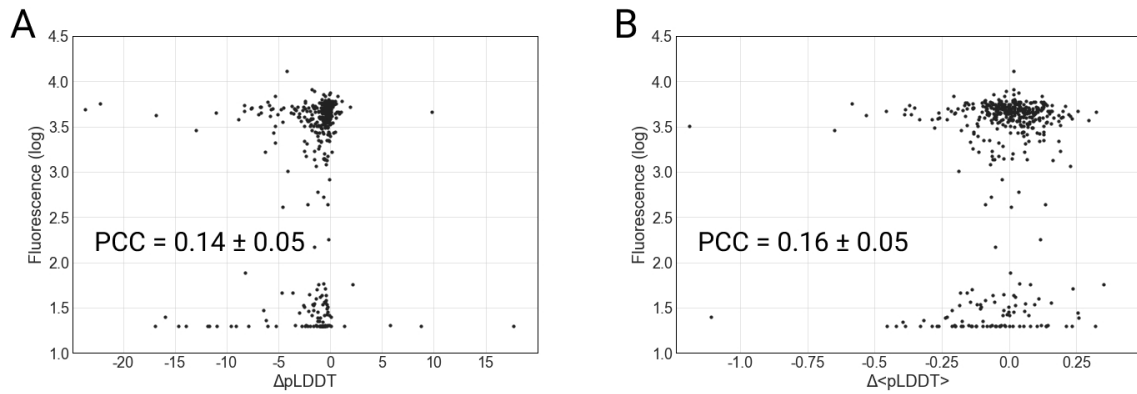
Figure 2: Correlation between the GFP fluorescence and change of confidence score of structure prediction for the mutated amino acid (A) and for the whole structure (B).

from zero. However, a linear combination of the two metrics, $\Delta$pLDDT and $\Delta$<pLDDT>, does not greatly improve the correlation. As for the loss-of-function prediction, the correlation with the impact of mutation on GFP fluorescence showed similar results: PCC was $0.14 \pm 0.05$ and $0.16 \pm 0.05$ for $\Delta$pLDDT and $\Delta$<pLDDT>, respectively (Figure 2).

Taken together, our data indicate that AlphaFold predictions cannot be used directly to reliably estimate the impact of mutation on protein stability or function. But why should we have expected such a correlation in the first place? Indeed, AlphaFold was not designed to predict the change of protein stability or function due to mutation. In the words of the authors "AlphaFold is not expected to produce an unfolded protein structure given a sequence containing a destabilising point mutation" (https://alphafold.ebi.ac.uk/faq). However, the only reason for a protein to fold into the distinct native structure is the stability of this structure, so the protein 3D structure and its stability are closely connected. Logically, an algorithm predicting protein 3D structure from sequence should search for the most stable 3D state under the native (or standard) conditions. If a compact structure becomes unstable (for example, due to mutation) then we might expect that the algorithm shifts its predictions toward an unfolded state. Evidence in favor of this point of view is the successful prediction of natively disordered protein regions by AlphaFold and the correlation between the decrease of pLDDT and propensity to be in a disordered region [19]. Thus, it is not unreasonable to expect a decrease in the confidence score of the mutated residue or the whole native structure.

Our results show that AlphaFold repurposing for $\Delta\Delta$G prediction did not work for the proteins we studied. However, AlphaFold may be used to help predict the impact of a mutation on protein stability or function using AlphaFold 3D models for 3D-structure based $\Delta\Delta$G predictors. Indeed, it was reported many times that 3D-based predictors perform better than 1D-based [20, 21, 22], so the availability of a pool of high-quality 3D predicted structures could be a plus. However, the performance of the resulting predictions is going to be far from perfect: the 3D-structure based $\Delta\Delta$G predictors are far from being perfect even using 3D structures from PDB [23], they show correlation of 0.59 or less in independent tests [24]. Thus, using AlphaFold models instead of PDB structures does not make $\Delta\Delta$G predictions more accurate [13], so availability of AlphaFold models is expected to show an approximately 0.59 correlation with predictions of $\Delta\Delta$G, which may be too low for many applications.

Will the newly AlphaFold-generated 3D structures for proteins with unknown 3D structures be useful for the training of the new $\Delta\Delta$G predictors? It seems unlikely, because the vast majority of the current experimental $\Delta\Delta$G values relate to proteins with known 3D structures. Thus, the newly generated structures will be useful to the same extent as the entire PDB has been across the last few decades.

The deep learning approach demonstrated by AlphaFold may be an inspiring example to develop a deep learning $\Delta\Delta$G predictor. However, we see the dramatic difference between the situations with 3D structure prediction and $\Delta\Delta$G prediction that may impede this development. The difference is in the amount of available data. For protein structure prediction AlphaFold used PDB with >180,000 files, and each file contained a wealth of information. In contrast to PDB, the number of experimentally measured $\Delta\Delta$G values are of the order of 10,000 and these are just numbers without accompanying extra data. To make a rough comparison of information in bits, PDB structures occupy 100 Gb, while all the known experimentally $\Delta\Delta$G values occupy  10 kb. Neural networks are very sensitive to the amount of information in the training set so the ability

of deep learning to tackle the $\Delta\Delta G$ prediction task at present looks hindered mostly by the lack of experimental data.

Overall, we explored the capacity of direct prediction of $\Delta\Delta G$ by all of the reported AlphaFold metrics: (i) the difference in the pLDDT score before and after mutation in the mutated position, (ii) the difference in the averaged pLDDT score across all positions before and after mutation. We found that the correlation was weak or absent, and, therefore, AlphaFold predictions are unlikely to be useful for $\Delta\Delta G$ predictions. Taken together with our recent result that AlphaFold models are not better for $\Delta\Delta G$ predictions by FoldX than best templates [13], we see no straightforward way to use AlphaFold advances for solving the task of prediction of $\Delta\Delta G$ upon mutation. The task of $\Delta\Delta G$ prediction should be solved separately and it will face the problem of limited amount of data for training neural networks.

# 4    Methods

## 4.1    Dataset of experimental mutations

The data on experimentally measured effects of mutations on protein stability were taken from ThermomutDB [14] (version 1.3). From 13,337 mutations in the database we extracted single-point mutations with data on $\Delta\Delta G$.

The filtered dataset resulted in 5712 mutations in 286 proteins. We have done the analysis for randomly chosen 976 mutations in 90 proteins.

## 4.2    Dataset of GFP mutants fluorescence

We took data on fluorescence levels of GFP mutants from [18]. From the original dataset we randomly extracted 447 single mutants for our analysis.

## 4.3    Protein structure modeling with AlphaFold

The wild type protein structures were retrieved from the AlphaFold Protein Structure Database (AlphaFold DB) [8] by their Uniprot accession code. The structures of original proteins that were absent in the AlphaFold DB as well as structures of mutant proteins were modeled by the standalone version of AlphaFold [1] using the fasta file with Uniprot sequence of a protein as the only input in the '–fasta_paths' flag.

## 4.4    Prediction metrics

The per-residue local distance difference test (pLDDT) confidence scores for the protein structure models downloaded from the AlphaFold DB were retrieved from the B-factor field of the coordinate section of the pdb file. The pLDDT confidence scores for the protein structure models that we predicted by standalone AlphaFold were extracted from the pickle file for the best ranked model from "plddt" array.

## 4.5    Sequence identity of proteins within the dataset of mutations

To identify the sequence identities of the proteins in the dataset of mutations we performed protein BLAST [15] search of protein sequences against themselves.

We divided the dataset into training and test sets for linear regression model based on the arbitrary sequence identity threshold of 50%. Mutations in proteins above the threshold comprised the training set, and the rest of mutations were used as the test set. The training and test sets resulted in 364 mutations in 60 proteins and 612 mutations in 30 proteins.

## 4.6    Linear regression analysis

Multiple linear regression fit with two parameters was performed using the linear_model module of Sklearn library with default parameters.

# 5   Acknowledgments

# References

1. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., . . . Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589. doi:10.1038/s41586-021-03819-2

2. Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., & Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins: Structure, Function, and Bioinformatics, 87(12), 1011–1020. doi:10.1002/prot.25823

3. Callaway, E. (2020). "It will change everything": DeepMind's AI makes gigantic leap in solving protein structures. Nature, 588(7837), 203–204. doi:10.1038/d41586-020-03348-4

4. Millán, C., Keegan, R. M., Pereira, J., Sammito, M. D., Simpkin, A. J., McCoy, A. J., . . . Read, R. J. (2021). Assessing the utility of CASP14 models for molecular replacement. doi:10.1101/2021.06.21.449228

5. Hegedűs, T., Geisler, M., Lukács, G., & Farkas, B. (2021). AlphaFold2 transmembrane protein structure prediction shines. doi:10.1101/2021.08.21.457196

6. Gupta, M., Azumaya, C. M., Moritz, M., Pourmal, S., Diallo, A., Merz, G. E., . . . Brilot, A. F. (2021). CryoEM and AI reveal a structure of SARS-CoV-2 Nsp2, a multifunctional protein involved in key host processes. doi:10.1101/2021.05.10.443524

7. Flower, T. G., & Hurley, J. H. (2021). Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. Protein Science, 30(4), 728–734. doi:10.1002/pro.4050

8. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., . . . Hassabis, D. (2021). Highly accurate protein structure prediction for the human proteome. Nature, 596(7873), 590–596. doi:10.1038/s41586-021-03828-1

9. Higgins, M. K. (2021). Can We AlphaFold Our Way Out of the Next Pandemic? Journal of Molecular Biology, 167093. doi:10.1016/j.jmb.2021.167093

10. Ko, J., & Lee, J. (2021). Can AlphaFold2 predict protein-peptide complex structures accurately? doi:10.1101/2021.07.27.453972

11. Tsaban, T., Varga, J., Avraham, O., Ben-Aharon, Z., Khramushin, A., & Schueler-Furman, O. (2021). Harnessing protein folding neural networks for peptide-protein docking. doi:10.1101/2021.08.01.454656

12. Guerois, R., Nielsen, J. E., & Serrano, L. (2002). Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations. Journal of Molecular Biology, 320(2), 369–387. doi:10.1016/s0022-2836(02)00442-4

13. Pak, M. A., & Ivankov, D. N. (2021). Best templates outperform homology models in predicting the impact of mutations on protein stability. doi:10.1101/2021.08.26.457758

14. Xavier, J. S., Nguyen, T.-B., Karmarkar, M., Portelli, S., Rezende, P. M., Velloso, J. P. L., . . . Pires, D. E. V. (2020). ThermoMutDB: a thermodynamic database for missense mutations. Nucleic Acids Research, 49(D1), D475–D479. doi:10.1093/nar/gkaa925

15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of Molecular Biology, 215(3), 403–410. doi:10.1016/s0022-2836(05)80360-2

16. Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. Bioinformatics, 29(21), 2722–2728. doi:10.1093/bioinformatics/btt473

17. Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N., & Tawfik, D. S. (2006). Robustness–epistasis link shapes the fitness landscape of a randomly drifting protein. Nature, 444(7121), 929–932. doi:10.1038/nature05385

18. Sarkisyan, K. S., Bolotin, D. A., Meer, M. V., Usmanova, D. R., Mishin, A. S., Sharonov, G. V., . . . Kondrashov, F. A. (2016). Local fitness landscape of the green fluorescent protein. Nature, 533(7603), 397–401. doi:10.1038/nature17995

19. Ruff, K. M., & Pappu, R. V. (2021). AlphaFold and Implications for Intrinsically Disordered Proteins. Journal of Molecular Biology, 167208. doi:10.1016/j.jmb.2021.167208

20. Montanucci, L., Capriotti, E., Frank, Y., Ben-Tal, N., & Fariselli, P. (2019). DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinformatics, 20(S14). doi:10.1186/s12859-019-2923-1

21. Savojardo, C., Fariselli, P., Martelli, P. L., & Casadio, R. (2016). INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics, 32(16), 2542–2544. doi:10.1093/bioinformatics/btw192

22. Lv, X., Chen, J., Lu, Y., Chen, Z., Xiao, N., & Yang, Y. (2020). Accurately Predicting Mutation-Caused Stability Changes from Protein Sequences Using Extreme Gradient Boosting. Journal of Chemical Information and Modeling, 60(4), 2388–2395. doi:10.1021/acs.jcim.0c00064

23. Berman, H. M. (2000). The Protein Data Bank. Nucleic Acids Research, 28(1), 235–242. doi:10.1093/nar/28.1.235

24. Potapov, V., Cohen, M., & Schreiber, G. (2009). Assessing computational methods for predicting protein stability upon mutation : good on average but not in the details. Prot. Eng. Des. Sel. 22(9), 553–560. https://doi.org/10.1093/protein/gzp030

25. Zacharov, I., Arslanov, R., Gunin, M., Stefonishin, D., Bykov, A., Pavlov, S., . . . Fedorov, M. (2019). "Zhores" — Petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. Open Engineering, 9(1), 512–520. doi:10.1515/eng-2019-0059