AlphaFold2 can predict structural and phenotypic effects of single mutations

John M. McBride,[1,*] Konstantin Polev,[1,2] Vladimir Reinharz,[3] Bartosz A. Grzybowski,[1,4,†] and Tsvi Tlusty[1,4,‡]

[1]*Center for Soft and Living Matter, Institute for Basic Science, Ulsan 44919, South Korea*
[2]*Departments of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea*
[3]*Université du Québec à Montréal, Canada*
[4]*Departments of Physics and Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea*

AlphaFold2 (AF2) is a promising tool for structural biology, but is it sufficiently accurate to predict the effect of missense mutations? We find that structural variation between closely related (1-3 mutations) protein pairs is correlated across experimental and AF2-predicted structures ($\sim$90,000 pairs). Analysis of $\sim$10,000 predicted structures from three high-throughput studies linking sequence and phenotype (fluorescence, foldability, and catalysis) demonstrates that AF2 can predict the phenotypic effect of missense mutations.

The unprecedented success of AlphaFold2 (AF2) in predicting protein structures from sequence has the potential to revolutionize structural biology.[1] AF2 can clearly predict global structure, yet we do not know whether it is sensitive enough to detect small, local effects of single mutations. Even if AF2 achieves high accuracy, the effect of a mutation may be small compared to the inherent conformational dynamics of the protein – predicting static structures may not be particularly informative.[2–4] Furthermore, as accuracy improves, evaluating the quality of predictions becomes increasingly complicated by the inherent noise in experimental measurements.[4–11] So far, there are conflicting reports as to whether AF2 can predict the effect of a mutation on protein stability.[12–15] We aim to resolve this issue by comparing AF2 predictions with extensive data on protein structure and function.

We examine AF2 predictions in light of structural data from a curated, non-redundant set of proteins from the Protein Data Bank (PDB),[16] and phenotype data from high-throughput experiments.[17–19] We find that AF2 can detect the effect of a mutation on structure by identifying structural change between protein pairs differing by 1-3 mutations: local structural change is correlated in experimental (PDB) and AF2-predicted pairs. Furthermore, we find significant correlations between local structural changes in AF2-predicted structures, and three categories of phenotype (fluorescence, folding, catalysis) across three experimental data sets.[17–19] These results indicate that AF2 can be used to predict the effect of missense mutations, undamming vast potential in the field of protein evolution.

AF2 can predict local structural change. We illustrate our approach by analyzing wild-type (`6BDD_A`) and single-mutant (`6BDE_A`, A71G) structures of H-NOX protein from *K. algicida* (Fig. 1D). To quantify local structural change, we calculate the local distance difference (LDD, see Methods) per residue for, respectively, experimental and AF2-predicted pairs of structures (Fig. 1A). LDD is highest at, and decays away from the mutated site (Fig. 1B). LDD is correlated with distance from the mutated site, and is correlated across PDB and AF2 structures (Fig. 1C,E). Taken together, these correlations indicate that LDD is a sensitive measure of local structural change, and that the AF2 is capable of predicting structural change upon mutation.

To get better statistics on AF2's performance, we apply this approach to a large sample of protein pairs that differ by 3 or fewer mutations (see Methods). We first note that some discrepancy between experiment and prediction is expected, since experimental measurements appear to be less precise than AF2 predictions: comparisons of structures with the same sequence show much higher LDD in the PDB than in AF2 structures (SI Fig 1); LDD decays to chance levels over 2 nm in AF2, but only 1 nm in the PDB (SI Fig 1). Despite this, we find that most of the time there is a non-trivial correlation in LDD across PDB and AF2 pairs of mutated sequences (Fig. 1F).

AF2-predicted change is due to mutations. It is possible that the LDD correlation between AF2 and PDB reflects inherent dynamical heterogeneity, due to local flexibility, or large conformational change. In support of this, we find that structural change is correlated across pairs of identical sequences (Fig. 1F, for zero mutations, $M = 0$). We test whether local flexibility can account for this correlation by examining solvent accessibility (solvent-exposed regions should be more flexible);[20] we find little support for this explanation (SI Fig 2), which suggests this correlation ($M = 0$) may be due to large-scale motion.[21] However, the LDD correlation is much stronger in non-identical pairs (Fig. 1F, $M > 0$), and the correlation is strongest within 1-2 nm of the mutated site (Fig. 1G). Furthermore, we find stronger correlations for non-conservative mutations, mutations of buried residues, and structures with higher confidence score (SI Fig 3). These facts suggest that the correlation is in part due to detectable effects of mutations.

Phenotypic change correlates with AF2-predicted change. An orthogonal test of whether AF2 can predict the effect of a mutation is to study correlations between LDD and phenotypic change. This approach has the benefit of avoiding the pitfalls associated with noisy PDB measurements. However, the link between structure and function is often unknown, and likely quite complex. Therefore, a lack of a correlation between LDD and phenotype is not strong evidence that the structure is incorrect, as there maybe be a non-trivial mapping between structure and function. On the other hand, observation of correlation between LDD and phenotype is strong evidence that AF2 can predict the effect of mutations. To this end, we study three data sets from high-throughput experiments, covering three distinct phenotypes: (i) green fluorescence is measured for 2,312 GFP sequences;[17] (ii) blue and red fluorescence is measured for 8,192 sequences linking mTagBFP2 (mostly blue) and mKate2 (mostly red);[18] (iii) foldability and catalytic activity are measured for PafA.[19]

We find that structural change between wild type pro-

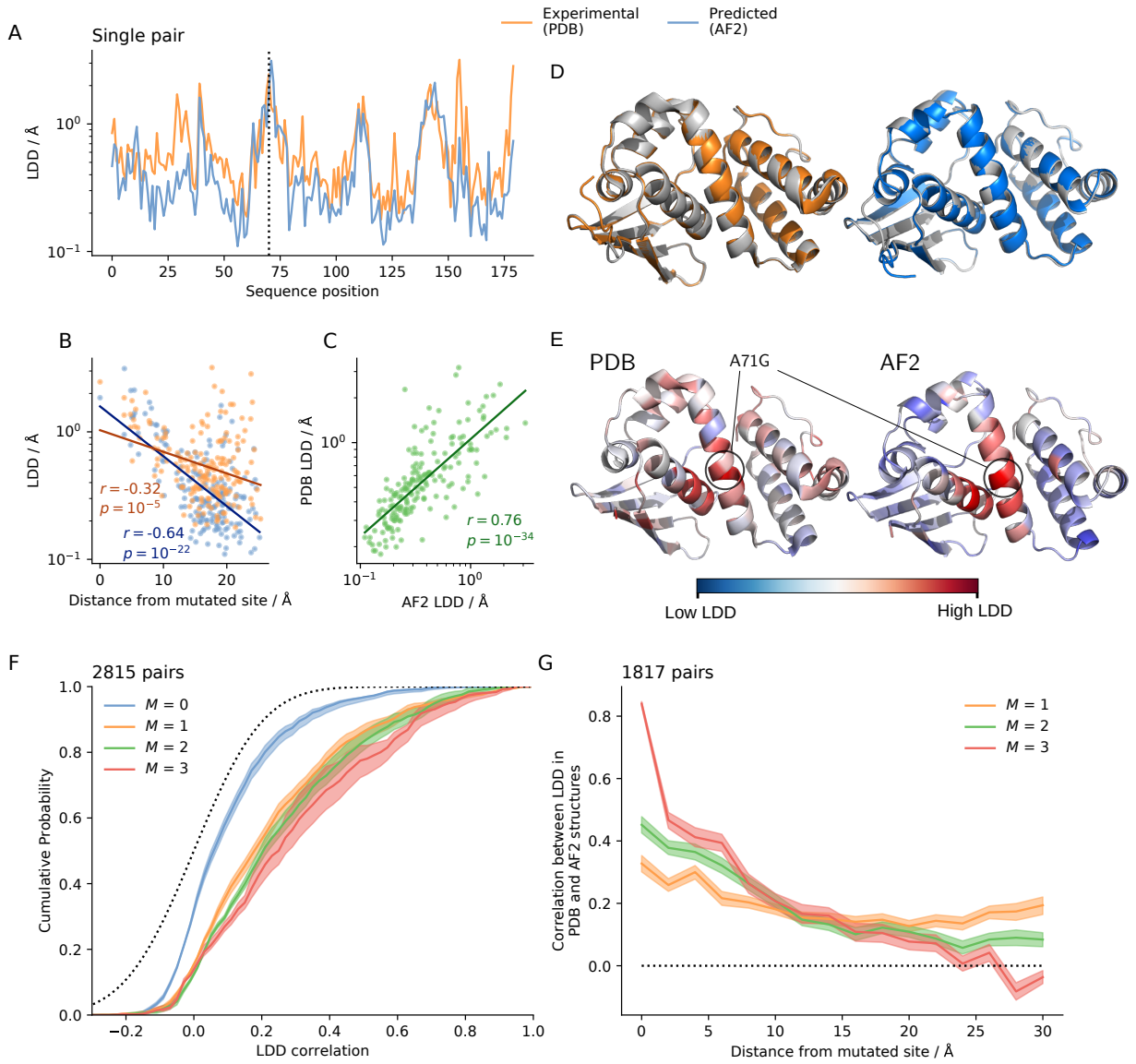

FIG. 1. A: C_α local distance difference (LDD) between wild type and mutant (A71G) H-NOX protein, for experimental (orange) and predicted (blue) structures. Dotted line indicates mutated residue. B: LDD as a function of distance from the mutated site. C: Comparison of LDD obtained from experimental and predicted structures. D: Overlaid wild type (grey, `6BDD_A`) and mutant (colour, `6BDE_A`), experimental (orange) and predicted (blue) structures. E: Wild type protein with residues coloured by LDD; location of A71G mutant is shown. F: Distribution of correlation between LDD vectors calculated for experimental and predicted structures for 2,815 protein pairs, as a function of the number of mutations. Dotted line shows a distribution with zero mean for comparison. G: LDD correlation across 1,817 pairs as a function of distance from the mutated site. Shaded region indicates 95 % confidence intervals obtained by sampling over 92,868 pairs.

teins and mutants in AF2-predicted structures correlates with phenotypic change for all phenotypes (Fig. 2). We consistently find negative correlations between mean LDD and fluorescence, across a range of mutation number M (Fig. 2A-B). Far away from the wild type ($M \gg 0$), the correlations disappear, and this is not simply due to undersampling (Fig. 2C). We find weak, yet significant correlations between mean LDD and foldability (Fig. 2D), and between mean LDD at the active site and catalytic activity (Fig. 2E). We do not find consistent correlations with RMSD, indicating that local change is more appropriate for measuring mutational effects (SI Fig 4). In some cases, performance is heavily dependent on which pre-trained model (see Methods) is used: surprisingly, we found that using the highest ranked (by pLDDT; see Methods) models resulted in worse performance for phenotypic change (SI Fig 4), but slightly better performance for structural change (SI Fig 5). Taken together, these results provide convincing evidence that AF2 can predict the effect of a single mutation.

LDD predicts phenotypic change for wild type proteins. It is quite unexpected that LDD should be a good predictor of phenotypic change, even if AF2 can accurately predict structure. We suspect that the correlation is strong because the structures are always compared to the wild-type proteins, where structure is optimized for function through evolution – any deviation from this optimal structure is likely to diminish protein function. To test this, we calculate the correlation between LDD and phenotype change across all possible pairs of proteins. We find that the correlation is much weaker, or completely disappeared (SI Fig

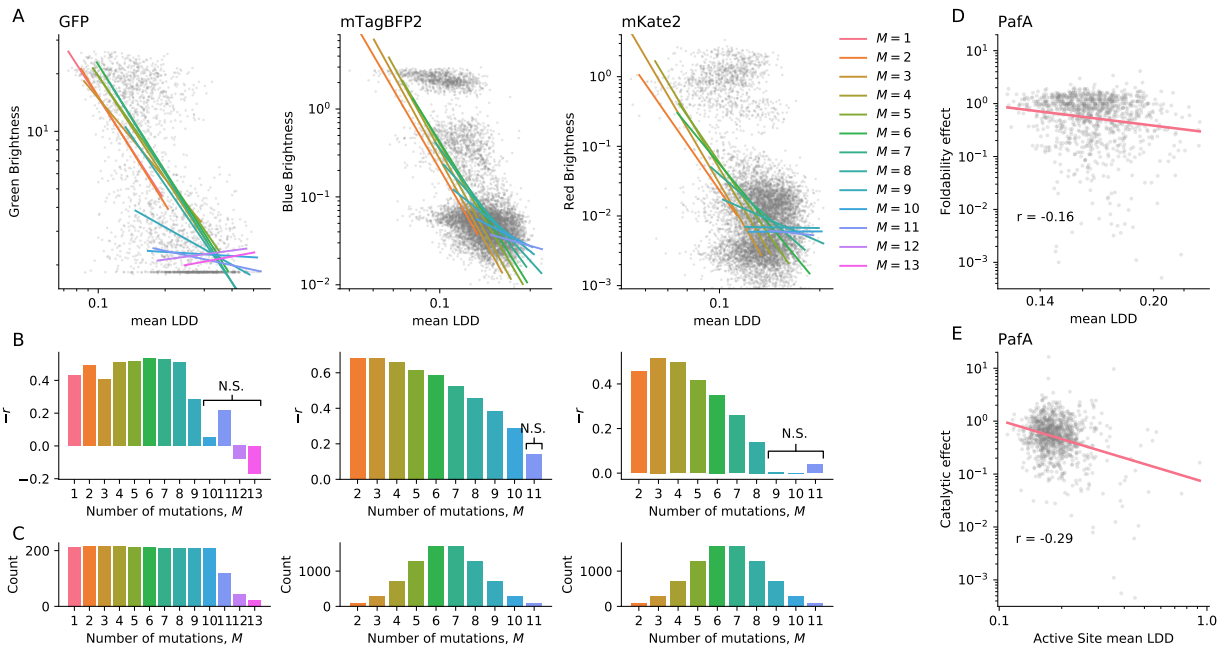

FIG. 2. A: Correlations between AF2-predicted structural change (mean LDD) and measured change in fluorescence for three proteins (GFP, mTagBFP2, mKate2), for different values of number of mutations, M, from the wild type. B: Correlation (Pearson's r) as a function of M. C: Number of sequences for each value of M. D: Correlation between mean LDD and foldability effect for PafA. E: Correlation between mean LDD at the active site, and catalytic effect for PafA. Correlations are statistically significant ($p < 0.005$), except where noted (N.S.).

6) for the divergent fluorescent proteins, but unchanged for PafA sequences (which have few mutations). This is consistent with the finding that correlations decrease as sequences deviate from the wild type (Fig. 2B). Thus we conclude that LDD is a good predictor of phenotypic change from native protein sequences. For studying phenotypic change away from optima in the phenotype landscape, a more sophisticated mapping from structure to function is needed.

Discussion. We have shown that AF2 can be used to predict the effect of a single mutation on structure and phenotype. We discuss why this prediction is not always correct, and offer recommendations for appropriate use of AF2. In general, the question of whether it is possible to measure the effect of a single mutation by comparing static structures depends on how large the effect is compared to the noise in the measurement. In this case, noise is the product of both measurement error and protein flexibility. One can use AF2's predicted confidence measure (pLDDT) as a proxy for accuracy (SI Fig 3), but additional information is necessary to understand the effect of protein flexibility. Effects will be greater for non-conservative mutations, and mutations in the buried hydrophobic core (SI Fig 3). Conversely, we would not have confidence in predictions of the effect of conservative substitutions made in flexible loops. To get maximal information from AF2 one should use all five models (SI Fig 4), and enhance conformational diversity through sampling of the multiple sequence alignment.[21] AF2 should be used in conjunction with other sources of information so as to be better equipped to assess the quality of predictions. Future validation studies should explore how predicted mutational effects depend on protein length, and availability of structural templates and homologous sequences.

Placing the current results in a broader context, we note that the evidence in support of AF2's capacity to predict the effect of a mutation has so far been mixed. In one study, the authors found no correlation between pLDDT and either stability or fluorescence. [12] We see no clear *a priori* reason that pLDDT should correlate with either stability or fluorescence. A low pLDDT score indicates that AF2 is not confident in the accuracy of a residue, which can be due to inaccuracy, or protein flexibility/disorder – a flexible protein can still be fluorescent or stable. In another analysis, the authors appear to assume that structure-disrupting mutations should result in a large change in predicted structure or pLDDT.[13] This is plausible, but it is also possible that structure-disrupting mutations might alter the kinetic or thermal stability, with only minor structural changes to the native state (which may fold correctly at lower temperature). It is premature to judge the performance of AF2 in this regard by only examining three pairs of structures. Two studies suggested that AF2 can be used to predict stability, however it is difficult to evaluate the unique contribution of AF2 since it was used in combination with additional machine learning methods to predict stability.[14,15]

In summary, we showed here that AF2 predictions of local structural change (without extra machine-learning algorithms), can be used to study missense mutations in proteins. These analyses suggest that AF2 can, indeed, be a powerful tool, if used in the right context, with appropriate analyses. Using AF2, we can bridge the gap between sequence and function in deep-mutational scan experiments, guide directed evolution studies,[22] and design drugs *in silico*.[23] On a smaller scale, AF2 can be used to screen potential mutants in costly experiments where

the number of mutations is limited. Overall, it appears that AF2 provides a step change in our ability to study evolution.

Methods

Structure Data. We curate a set of structures from the PDB to study the effect of mutations on protein structure. We select all proteins from the PDB that have equal sequence length, for which there are multiple structures whose sequences differ by no more than 3 mutations; we include proteins with identical sequences as a control group. We only include proteins of length $50 \leq L \leq 500$. We exclude protein complexes, or proteins bound to RNA or DNA. We only match pairs of proteins if they are bound to the exact same types of small molecules. We exclude NMR structures for simplicity, avoiding the need to determine additional cutoffs to infer disorder, or to choose a representative structural model. We only consider pairs that were prepared at a similar pH (within ± 0.5). We exclude pairs of structures that are almost identical (*e.g.*, from time-resolved crystallography experiments); *i.e.*, pairs with a RMSD smaller than 0.001 Å. This leaves us with 4,653 PDB structures, and 92,868 pairs. To create a non-redundant sample, we cluster protein sequences using CD-hit,[24] with a 90 % sequence identity threshold. From the total pool of pairs we create sub-samples with no more than 10 examples per group, for each value of M, the mutation number (in total, 2,815 pairs); we run analyses with 1,000 samples to estimate sampling error. We predict structures using AF2 with default template cutoff date (14 May 2020) and reduced genomic database. We run AF2 using all five pre-trained models: we show results for the models with the highest pLDDT (AF2-predicted measure of residue-level confidence) in Fig. 1, and compare other models in SI Fig 5; we show results in Fig. 2 for single, best-performing model for each separate phenotype, and the other models in SI Fig 4.

Structure analysis. Metrics of structural change are often differentiated by whether they provide absolute measures or scores, and by whether they measure global or local differences: *e.g.*, root-mean-square-deviation (RSMD) is an absolute, global measure. For our purpose, we want an absolute, local measure. A common metric of local change is the local distance difference test (LDDT),[25] but this is designed to score the similarity of experimental and predicted structures for a single protein sequence; it is bounded by 0 and 1 so as to ignore large effects, and uses a total of five cutoff parameters. Instead, we design a simplified LDDT variant, the local distance difference (LDD), which requires only one cutoff, and measures perturbations on an absolute scale. For each residue we calculate a distance vector of the scalar distances between neighbouring C_α atoms, for each structure; atoms are neighbours if they are within a radius of 11 Å in both structures. The LDD is then the Euclidean distance between these two distance vectors.

Author Contributions. JM, KP, VR and TT designed research. JM and KP produced data. JM analysed data. BG and TT supervised the project. JM wrote the paper. JM, KP, VR, BG and TT revised the paper.

REFERENCES

* jmmcbride@protonmail.com
† nanogrzybowski@gmail.com
‡ tsvitlusty@gmail.com

1. Jumper, J. *et al.* Highly Accurate Protein Structure Prediction With Alphafold. *Nature* **596**, 583–589 (2021).
2. Fleishman, S. J. & Horovitz, A. Extending the New Generation of Structure Predictors to Account for Dynamics and Allostery. *J. Mol. Biol.* **433**, 167007 (2021).
3. Morgunov, A. S., Saar, K. L., Vendruscolo, M. & Knowles, T. P. J. New Frontiers for Machine Learning in Protein Science. *J. Mol. Biol.* **433**, 167232 (2021).
4. Masrati, G. *et al.* Integrative Structural Biology in the Era of Accurate Structure Prediction. *J. Mol. Biol.* **433**, 167127 (2021).
5. Andrec, M. *et al.* A Large Data Set Comparison of Protein Structures Determined by Crystallography and NMR: Statistical Test for Structural Differences and the Effect of Crystal Packing. *Proteins.* **69**, 449–465 (2007).
6. Rashin, A. A., Rashin, A. H. L. & Jernigan, R. L. Protein Flexibility: Coordinate Uncertainties and Interpretation of Structural Differences. *Acta Crystallogr. D.* **65**, 1140–1161 (2009).
7. Venkatakrishnan, B., Palii, M. L., Agbandje-McKenna, M. & McKenna, R. Mining the Protein Data Bank to Differentiate Error From Structural Variation in Clustered Static Structures: An Examination of Hiv Protease. *Viruses* **4**, 348–362 (2012).
8. Everett, J. K. *et al.* A Community Resource of Experimental Data for NMR / X-Ray Crystal Structure Pairs. *Protein Sci.* **25**, 30–45 (2016).
9. Hura, G. L. *et al.* Small Angle X-Ray Scattering-Assisted Protein Structure Prediction in Casp13 and Emergence of Solution Structure Differences. *Proteins.* **87**, 1298–1314 (2019).
10. Lynch, M. L., Dudek, M. F. & Bowman, S. E. J. A Searchable Database of Crystallization Cocktails in the Pdb: Analyzing the Chemical Condition Space. *Patterns* **1**, 100024 (2020).
11. Fowler, N. J., Sljoka, A. & Williamson, M. P. The Accuracy of NMR Protein Structures in the Protein Data Bank. *Structure* **29**, 1430–1439 (2021).
12. Pak, M. A. *et al.* Using Alphafold to Predict the Impact of Single Mutations on Protein Stability and Function. *bioRxiv* (2021).
13. Buel, G. R. & Walters, K. J. Can Alphafold2 Predict the Impact of Missense Mutations on Structure? *Nature Structural & Molecular Biology* **29**, 1–2 (2022).
14. Akdel, M. *et al.* A Structural Biology Community Assessment of Alphafold 2 Applications. *bioRxiv* (2021).
15. Zhang, Y. *et al.* Applications of Alphafold Beyond Protein Structure Prediction. *bioRxiv* (2021).
16. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
17. Sarkisyan, K. S. *et al.* Local Fitness Landscape of the Green Fluorescent Protein. *Nature* **533**, 397–401 (2016).
18. Poelwijk, F. J., Socolich, M. & Ranganathan, R. Learning the Pattern of Epistasis Linking Genotype and Phenotype in a Protein. *Nat. Commun.* **10**, 4213 (2019).
19. Markin, C. J. *et al.* Revealing Enzyme Functional Architecture via High-Throughput Microfluidic Enzyme Kinetics. *Science* **373**, eabf8761 (2021).
20. Zhang, H. *et al.* On the relation between residue flexibility and local solvent accessibility in proteins. *Proteins: Structure, Function, and Bioinformatics* **76**, 617–636 (2009). https://onlinelibrary.wiley.com/doi/pdf/10.1002/prot.22375.
21. Stein, R. A. & Mchaourab, H. S. Modeling Alternate Conformations With Alphafold2 via Modification of the Multiple Sequence Alignment. *bioRxiv* (2021).
22. Arnold, F. H. Design by Directed Evolution. *Accounts of Chemical Research* **31**, 125–131 (1998).
23. Cao, L. *et al.* Design of protein binding proteins from target structure alone. *Nature* (2022).
24. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006). https://academic.oup.com/bioinformatics/article-pdf/22/13/1658/484588/btl158.pdf.
25. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. Lddt: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests. *Method. Biochem. Anal.* **29**, 2722–2728 (2013).