AlphaFold2 can predict single-mutation effects on structure and phenotype

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] *Department 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 (AF) is a promising tool, but is it accurate enough to predict single mutation effects? Here, we report that a measure for localized structural deformation between protein pairs differing by only 1-3 mutations is correlated across 4,645 experimental and AF-predicted structures. Furthermore, analysis of ∼10,000 proteins shows that the local structural change correlates with various phenotypic changes. These findings suggest that AF can predict the magnitude of single-mutation effects in many proteins, and we propose a method to identify those proteins for which AF is predictive.

Alteration of one or few amino acid residues can affect structure [1–3] and function [4] of a protein and, in extreme cases, be the difference between health and disease [5, 6]. Understanding structural consequences of point mutations is important for drug design [7, 8] and could also accelerate optimization of enzymatic function via directed evolution [9, 10]. In these and other applications, theoretical models [11] could be of immense help, provided they are sufficiently accurate. In this context, AlphaFold2 [12] has recently made breakthroughs in predicting global protein structure from sequence with unprecedented precision. Notwithstanding, it is not yet known whether AF is sensitive enough to detect small, local effects of single mutations. Even if AF 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 [13–15]. Furthermore, as accuracy improves, evaluating the quality of predictions becomes increasingly complicated by the inherent noise in experimental measurements [15–22]. So far, no study has evaluated whether AF can accurately measure structural changes due to single mutations, and there are conflicting reports as to whether AF can predict the effect of a mutation on protein stability [23–26]. Furthermore, recent evidence suggests that AF learns the energy functional underlying folding, raising the question of whether the inferred functional is sensitive enough to discern the subtle physical changes due to a single mutation [27]. We aim to resolve this issue by comparing AF predictions with extensive data on protein structure and function.

We examine AF predictions in light of structural data from a curated, non-redundant set of proteins from the Protein Data Bank (PDB) [28], and phenotype data from high-throughput experiments [29–31]. We find that AF can detect the effect of a mutation on structure by identifying local deformations between protein pairs differing by 1-3 mutations: the physical measure for such local structural changes (termed LDD) is correlated in experimental (PDB) and AF-predicted pairs. A statistical model shows that the observed correlation is about as high as expected, given the inherent measurement noise. Furthermore, analysis of ∼10,000 proteins whose function was probed in three high-throughput studies shows significant correlations between local structural change (LDD) in AF-predicted structures, and three categories of phenotype (fluorescence, folding, catalysis) across three experimental data sets [29–31]. These sets of correlations suggest that AF can predict the magnitude of single-mutation effects in many proteins, and we propose a method to assess confidence in AF predictions of mutational effects, so that users can distinguish between structural changes that are caused by mutation, as opposed to natural protein fluctuations or other variance in AF predictions: information from multiple structures is used to create a fluctuation-mutation structure profile (FMS profile), which allows us to identify proteins, and individual mutations, that are more likely to have accurate predictions. These results and the proposed recipe indicate that AF can be used to predict physicochemical effects of missense mutations, undamming vast potential in the field of protein design and 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 SM) per residue for, respectively, experimental and AF-predicted pairs of structures (Fig. 1A). LDD is calculated by obtaining a vector of distances between neighbouring C_α positions in each structure (*e.g.*, wild-type and mutant), and taking the Euclidean norm of the difference between the two vectors. This measure provides a robust estimate of the magnitude of local strain, which is less sensitive to noise than the shear tensor [32, 33]. We observe that LDD is highest at, and decays away from the mutated site (Fig. 1B). Furthermore, LDD is correlated with distance from the mutated site (Fig. 1B), and is correlated across PDB and AF structures (Fig. 1C,E). Taken together, these correlations indicate that LDD is a sensitive measure of local structural change, and that the AF is capable of predicting structural change upon mutation.

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

AF-predicted change is due to mutations.— It is possible that the LDD correlation between AF and PDB reflects in-

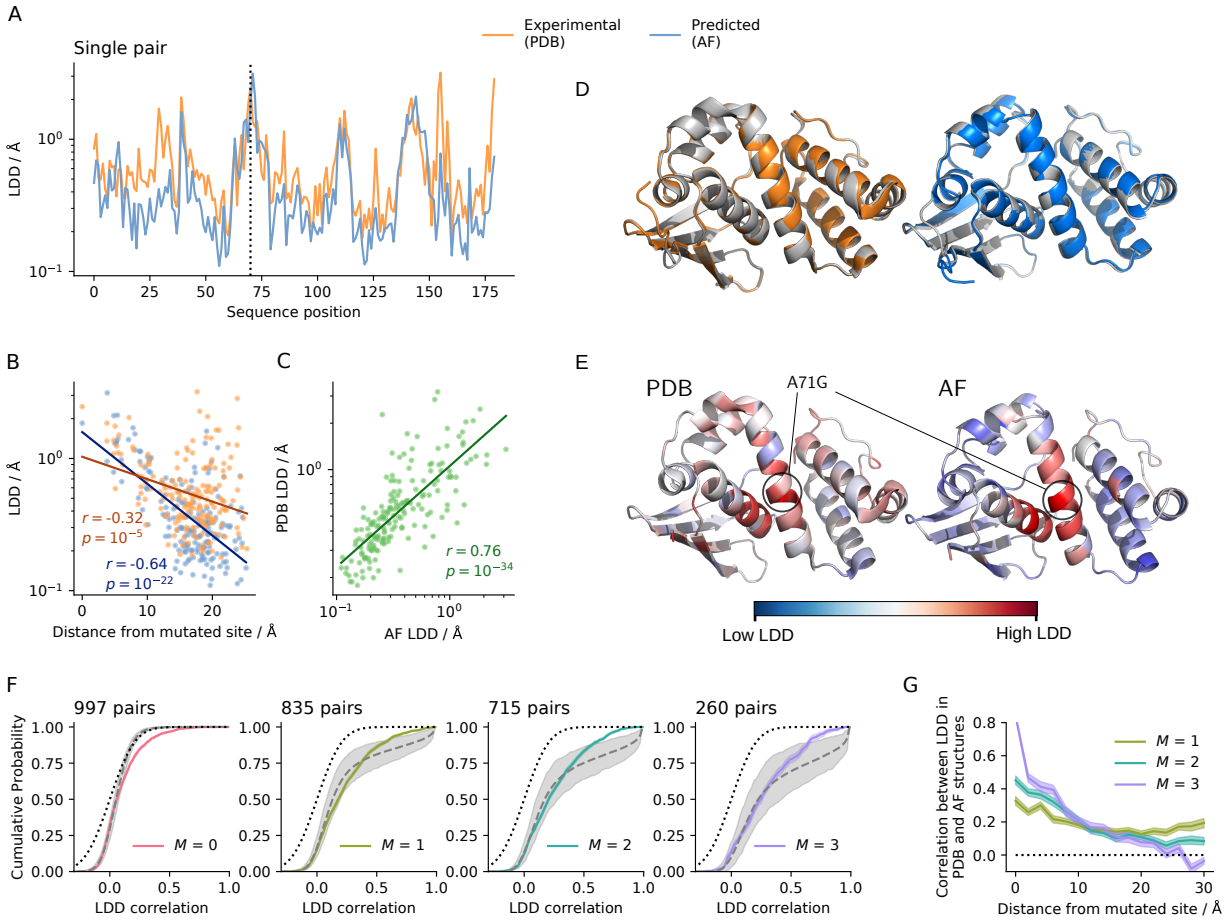

FIG. 1. A: C_α local distance difference (LDD) between wild type and mutant (A71G) H-NOX protein, for experimental (orange) and predicted (blue) structures. Dotted line indicates mutated residue. B: LDD as a function of distance from the mutated site. C: Comparison of LDD obtained from experimental and 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: Distributions of correlation between LDD vectors calculated for experimental and predicted structures (solid lines) for 2,815 protein pairs, as a function of the number of mutations, M. Dashed grey line shows the mean distribution for a statistical model of the expected correlation given some measurement error and mutation effects, both calibrated using experimental data; shaded grey area indicates the 95 % confidence intervals. 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. Colored shaded regions indicate 95 % confidence intervals obtained by sampling over 92,868 pairs.

herent dynamical heterogeneity, due to local flexibility, or large conformational change. In support of this, we find that structural change is weakly correlated across pairs of identical sequences (Fig. 1F, for zero mutations, $M = 0$). We therefore test whether local flexibility can account for this correlation by examining solvent accessibility (solvent-exposed regions should be more flexible) [34]; we find little support for this explanation (SM Fig. 2), which suggests this correlation ($M = 0$) may be due to flexibility that is not captured by solvent accessibility [35].

In contrast, for non-identical pairs ($M > 0$) the LDD correlations are much stronger (Fig. 1F), and correlations are strongest within 1-2 nm of the mutated site (Fig. 1G). These facts suggest that the LDD correlation is in part due to detectable effects of mutations.

Low LDD correlations are expected due to experimental noise.— We note that many protein pairs show weak or no correlation (*e.g.*, $r = 0.24$ on average for $M = 1$) between PDB and AF LDD vectors. One explanation for the relatively low LDD correlations is that AF predictions are inaccurate, while an alternative explanation is that the experimental data is too noisy (*i.e.*, the effect of the mutation is small compared to experimental measurement error). To test the latter explanation, we create a statistical model of both measurement error and the effect of a mutation (see SM), and calibrate two versions: one using PDB data, the other using AF data. We then simulate the distribution of LDD correlations that one should expect given: the set of PDB structures tested, the calibrated experimental noise, and mutational effect size. The results of this simple model (Fig. 1F, dotted line) show that the empirical correlations are roughly what one should expect given the high experimental noise in PDB structures relative to the magnitude of the effect of a mutation. Furthermore, we see that increasing the precision of the AF statistical model (reducing noise to zero) does not result in higher LDD correlations. All this indicates that the simulated LDD correlation is limited by the reliability of the PDB data (SM Fig. 3). In other words, it appears that AF is almost as accurate as possible, given the available experimental data and its noisiness.

Phenotypic change correlates with AF-predicted structural change.— An orthogonal test of whether AF can pre-

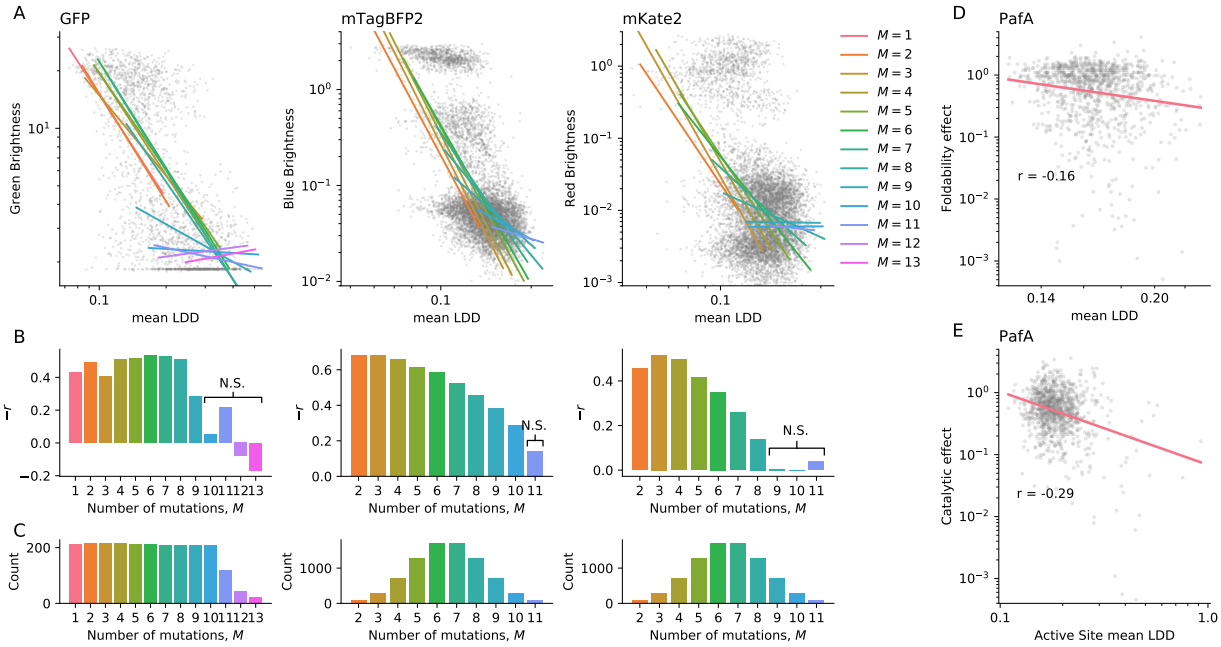

FIG. 2. A: Correlations between AF-predicted structural change, estimated by the mean LDD, and measured change in fluorescence for three proteins (GFP, mTagBFP2, mKate2), for different values of number of mutations, M, from the wild type. B: Correlation (Pearson's r) as a function of M. C: Number of 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.).

dict the effect of a mutation is to study correlations between the LDD structural measure 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 AF can be predictive in estimating 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 [29]; (ii) blue and red fluorescence is measured for 8,192 sequences linking mTagBFP2 (mostly blue) and mKate2 (mostly red) [30]; (iii) foldability and catalytic activity are measured for PafA [31].

We find that structural change between wild-type proteins and mutants in AF-predicted structures correlates with phenotypic change for all phenotypes (Fig. 2). We consistently find negative correlations between mean LDD (*i.e.*averaged over all residues) 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 statistically significant correlations between mean LDD and foldability (Fig. 2D), and between mean LDD at the active site (*i.e.*, averaged over active site residues) and catalytic activity (Fig. 2E). In contrast, we do not find consistent correlations with RMSD, a standard estimate of AF accuracy,[12] indicating that local deformation is more appropriate for measuring mutational effects (SM Fig. 4). In some cases, performance is heavily dependent on which pre-trained model (see SM) is used: surprisingly, we found that using the highest ranked (by pLDDT; see SM) models resulted in worse performance for phenotypic change (SM Fig. 4), but slightly better performance for structural change (SM Fig. 5). Taken together, these results provide evidence that AF can be used to predict the structural effect of a single mutation. For example, a mutation in GFP that causes a change such that the mean LDD is greater than 0.2, will most likely result in dimmed fluorescence.

LDD correlates with phenotypic change for wild-type proteins.— It is quite unexpected that LDD should be a good predictor of phenotypic change, even if AF can accurately predict structure. We suspect that the correlation is strong because the structures are always compared to the wild-type proteins, where the structure is optimized for function through evolution – any deviation from this optimal structure is likely to diminish protein function. To test this hypothesis, we calculate the correlation between LDD and phenotype change across all possible pairs of proteins (*i.e.*, comparing mutants with mutants). We find that the correlation is much weaker, or completely disappeared (SM Fig. 6) for the divergent fluorescent proteins, but unchanged for PafA sequences (which are all within $M \leq 2$ of WT). 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 phenotype landscapes, a more sophisticated mapping from structure to function is needed.

Assessing AF predictions using fluctuation-mutation

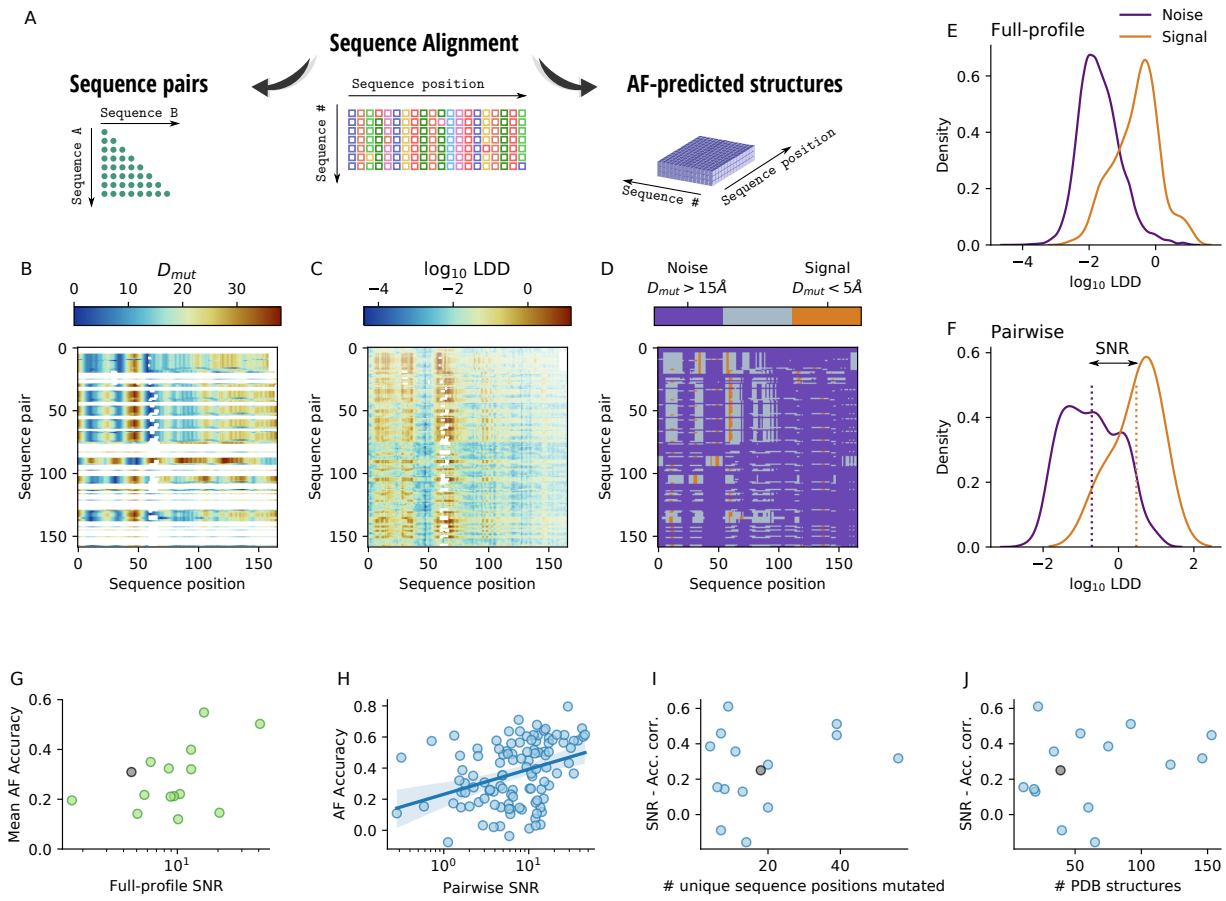

FIG. 3. **Assessing AF predictions using fluctuation-mutation structure (FMS) profiles.** A-D: We illustrate how to create an FMS profile using human H-Ras protein as an example. We first align a set of sequences of same length, use AF to predict structures, and arrange them into pairs. For each sequence pair and position we calculate LDD (B), and D_{mut}, the average distance to a mutated residue (C); blank regions indicate either disorder (pLDDT < 70) or pairs of structures from identical sequences. We assign each point to be either Signal ($D_{\text{mut}} < 5$ Å) or Noise ($D_{\text{mut}} > 15$ Å). E-F: Signal and Noise LDD distributions for (E) the full profile and (F) a single pair of sequences (H-Ras WT, `3L8Z_A` and mutant A59G, `1LF0_A`). Dotted lines indicate means; signal-to-noise ratio (SNR) is the ratio between the means. G: Mean AF Accuracy (LDD correlation) as a function of whole-protein SNR, for 15 protein families; black circle indicates H-Ras. H: AF Accuracy as a function of pairwise SNR for all H-Ras pairs. I-J: Correlation between pairwise SNR and AF Accuracy for 15 families as a function of (I) the number of unique sequence positions mutated in the structure profile, and (J) the number of unique structures used in the profile; black circle indicates H-Ras.

structure profiles.— The significant correlations between structural and phenotypic change indicate that AF *can* predict the effect of missense mutations. By this, we mean that the magnitude of change between AF-predicted structures is often, but not always, a good measure of the effect of a mutation. Naturally, we next wish to know *when* we can trust AF predictions. In general, we expect that if a mutation has a large effect, this should be easier to reliably measure, and effects should be difficult to measure in flexible regions. However, we find only weak correlations ($|r| < 0.1$, SM Fig. 7) between plausible indicators of mutation effect size (surface accessibility, Blosum score, fraction of disordered residues, and mean pLDDT). Instead, we propose that AF itself can be used to generate a more meaningful predictor of accuracy, using what we term *fluctuation-mutation structure profiles* (FMS profiles).

The concept of FMS profiles (Fig. 3A) is based on the assumption that there is a non-trivial mapping between a mutation (defined by position and amino acid identity) and the magnitude of its effect, which can be estimated using AF (see SM). To this end, we take a set of protein sequences, generate structures using AF, and compare pairs of structures. We calculate matrices for LDD (Fig. 3B) and D_{mut}, the minimum distance to a mutated position (Fig. 3C). We then group LDD values into signal ($D_{\text{mut}} < 5$ Å) or noise ($D_{\text{mut}} > 15$ Å) categories (Fig. 3D), based on our finding that LDD is greatest within 5 Å of a mutated site, and LDD appears to be independent of mutations at residues further than 15 Å from a mutated site (SM Fig. 8). Finally, using the signal and noise LDD distributions (Fig. 3E-F), we calculate the overall signal-to-noise ratio (SNR) as the ratio of the geometric means of the distributions.

Using this approach, we can calculate SNR for the full profile (Fig. 3E), which estimates the ability of AF to predict mutational effects for a given protein; or we can calculate SNR for a pair of sequences ($M \geq 1$) to estimate the effect of a specific mutation (Fig. 3F). As an example (Fig. 3A-F), we created a profile for GPTase H-Ras (Uniprot acc. `P01112`), using only the set of 159 sequence pairs for which structures exist in the PDB (34 in total) that conform to our selection criteria (see SM).

To evaluate the utility of FMS profiles, we examine how

accuracy varies with SNR. To aid legibility, we hereafter report LDD correlation (Fig. 1F) as AF accuracy; although, we remind the reader that this measure of accuracy is not expected to reach a value of one, due to inherent protein fluctuations. We generate profiles for all (15) protein families for which we have sufficient data (see SM); we limit profiles to the same sets of pairs that we have empirical data for. We find that mean AF accuracy is positively correlated with full-profile SNR (Fig. 3F, $r = 0.46$, $p = 0.09$), although statistical power is low due to a lack of empirical data. We also looked at whether full-profile SNR could explain why some AF models produced better correlations with phenotypic change (SM Fig. 4); the models with high correlations had SNR > 3, while the two PafA models that were useless had SNR < 1.2 (SM Fig. 9). Within H-Ras structures, we find that pairwise SNR is positively correlated with AF accuracy ($r = 0.36$, $p < 0.005$). Across all 15 families, the correlation between SNR and AF accuracy is typically positive, but it is sometimes low or negative when there are insufficient mutation positions (Fig. 3I, SM Fig. 10) or structures sampled (Fig. 3J). Although the lack of empirical data hinders validation, it seems that if a profile samples enough sequences, and sequence positions, SNR can be used to assess the accuracy of AF in predicting single-mutation effects.

Discussion.— We have shown that AF is capable of predicting structures with sufficient accuracy and that it can pick up changes as small as those resulting from a single missense mutation. Direct validation of predicted mutational effects on structure is limited by the accuracy of empirical structures, and further hindered by the lack of sequence pairs that are suitable for comparison. Likewise, predicting phenotypic change from structure alone ought to be challenging, to say the least. Despite these steep hurdles, we have shown, using the local distance difference (LDD) metric, that differences between AF-predicted structures do correlate with both structural (Fig. 1) and phenotypic changes (Fig. 2) in empirical data.

Since AF predictions are not always in agreement with empirical data, we recommend that users assess prediction accuracy by using a recipe based on FMS profiles: Full-profile SNR can be used to identify whether AF is likely to make reliable predictions for a given protein family (Fig. 3G), and can be used to distinguish good- from bad-performing models where pLDDT fails (SM Fig. 4,9). A pairwise SNR over 10 indicates that a mutations is likely to have large effects that is reliably predicted (Fig. 3H, SM Fig. 10); low values of SNR indicate that the effect on structure is small compared to the inherent flexibility of a particular region in the protein. This method may eventually be extended to work with the effect of mutations on conformational diversity, through sampling of the multiple sequence alignment [35]; however, for now there is scant empirical data to test such a method. Indeed, the difficulties we faced in assembling sufficient data highlight that the age of experimental protein structure identification is far from over [36], despite the success of AF and RoseTTAFold [12, 37].

We need to emphasise that AF is only trained to predict structures of stable proteins, and we make no claims about whether the proteins will indeed fold into the predicted structure. Given the marginal stability of most proteins, mutations may easily destabilize a protein so that its melting temperature falls below room temperature. The process of protein folding is carefully tuned *in vivo* for folding on the ribosome, and through interactions with chaperones, and mutations that do not change structure may retard folding through other means. To see whether the AF confidence score (pLDDT) is predictive of whether a protein will fold or not, we studied a set of 147 WW-domain-like sequences, of which 40 were found to fold *in vitro*. Although more sophisticated methods may perform better, the mean pLDDT by itself proved insufficient to sort folding from non-folding proteins (SM Fig. 11). Now that one question – what structure will a protein likely fold into? – has been seemingly solved, at least partially, it is crucial to next answer the question of *whether* a protein will spontaneously fold.

Placing the current results in a broader context, we note that the evidence in support of AF'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. [23] We see no clear *a priori* reason that pLDDT should correlate with either stability or fluorescence. A low pLDDT score merely indicates that AF 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 [24], but the supporting evidence includes only three pairs of structures. Two studies suggested that AF can be used to predict stability, however, it is difficult to evaluate the unique contribution of AF since it was used in combination with additional machine learning methods to predict stability [25, 26].

In summary, we showed here that AF predictions of local structural change can be used to study missense mutations in proteins. These analyses suggest that AF can, indeed, be a powerful tool, if used in the right context and backed up by appropriate analyses. Using AF, we can bridge the gap between sequence and function in high-throughput deep-mutational scan experiments, guide directed evolution studies [9], and design drugs *in silico* [10]. For example, on a smaller scale AF can be used to screen potential mutants, and in costly experiments where the number of mutations is limited, one can select for mutations with strong or weak effects in desired regions of the protein. Overall, it appears that AF provides a step change in our ability to study, and guide protein evolution.

We thank Jacques Rougemont and Jean-Pierre Eckmann for discussions. This work was supported by the Institute for Basic Science, Project Code IBS-R020-D1.

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

[1] N. Tokuriki and D. S. Tawfik, Stability effects of mutations

and protein evolvability, Curr. Opin. Struc. Biol. **19**, 596 (2009).

[2] M. Lorch, J. M. Mason, R. B. Sessions, and A. R. Clarke, Effects of mutations on the thermodynamics of a protein folding reaction: Implications for the mechanism of formation of the intermediate and transition states, Biochemistry-us. **39**, 3480 (2000).

[3] M. Zhang, D. A. Case, and J. W. Peng, Propagated perturbations from a peripheral mutation show interactions supporting ww domain thermostability, Structure **26**, 1474 (2018).

[4] G. Yang, N. Hong, F. Baier, C. J. Jackson, and N. Tokuriki, Conformational tinkering drives evolution of a promiscuous activity through indirect mutational effects, Biochemistry-us. **55**, 4583 (2016).

[5] N. Sahni, S. Yi, M. Taipale, J. Fuxman Bass, J. Coulombe-Huntington, F. Yang, J. Peng, J. Weile, G. Karras, Y. Wang, I. Kovács, A. Kamburov, I. Krykbaeva, M. Lam, G. Tucker, V. Khurana, A. Sharma, Y.-Y. Liu, N. Yachie, Q. Zhong, Y. Shen, A. Palagi, A. San-Miguel, C. Fan, D. Balcha, A. Dricot, D. Jordan, J. Walsh, A. Shah, X. Yang, A. Stoyanova, A. Leighton, M. Calderwood, Y. Jacob, M. Cusick, K. Salehi-Ashtiani, L. Whitesell, S. Sunyaev, B. Berger, A.-L. Barabási, B. Charloteaux, D. Hill, T. Hao, F. Roth, Y. Xia, A. Walhout, S. Lindquist, and M. Vidal, Widespread macromolecular interaction perturbations in human genetic disorders, Cell **161**, 647 (2015).

[6] V. M. Prabantu, N. Naveenkumar, and N. Srinivasan, Influence of disease-causing mutations on protein structural networks, Frontiers in Molecular Biosciences **7**, 10.3389/fmolb.2020.620554 (2021).

[7] H. Jacquier, A. Birgy, H. L. Nagard, Y. Mechulam, E. Schmitt, J. Glodt, B. Bercot, E. Petit, J. Poulain, G. Barnaud, P.-A. Gros, and O. Tenaillon, Capturing the mutational landscape of the beta-lactamase tem-1, P. Natl. Acad. Sci. Usa. **110**, 13067 (2013).

[8] A. T. Albanaz, C. H. Rodrigues, D. E. Pires, and D. B. Ascher, Combating mutations in genetic disease and drug resistance: Understanding molecular mechanisms to guide drug design, Expert Opin. Drug Dis. **12**, 553 (2017).

[9] F. H. Arnold, Design by directed evolution, Accounts Chem. Res. **31**, 125 (1998).

[10] L. Cao, B. Coventry, I. Goreshnik, B. Huang, J. S. Park, K. M. Jude, I. Marković, R. U. Kadam, K. H. G. Verschueren, K. Verstraete, S. T. R. Walsh, N. Bennett, A. Phal, A. Yang, L. Kozodoy, M. DeWitt, L. Picton, L. Miller, E. M. Strauch, N. D. DeBouver, A. Pires, A. K. Bera, S. Halabiya, B. Hammerson, W. Yang, S. Bernard, L. Stewart, I. A. Wilson, H. Ruohola-Baker, J. Schlessinger, S. Lee, S. N. Savvides, K. C. Garcia, and D. Baker, Design of protein binding proteins from target structure alone, Nature 10.1038/s41586-022-04654-9 (2022).

[11] A. Stein, D. M. Fowler, R. Hartmann-Petersen, and K. Lindorff-Larsen, Biophysical and mechanistic models for disease-causing protein variants, Trends Biochem. Sci. **44**, 575 (2019).

[12] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, Highly accurate protein structure prediction with alphafold, Nature **596**, 583 (2021).

[13] S. J. Fleishman and A. Horovitz, Extending the new generation of structure predictors to account for dynamics and allostery, J. Mol. Biol. **433**, 167007 (2021).

[14] A. S. Morgunov, K. L. Saar, M. Vendruscolo, and T. P. J. Knowles, New frontiers for machine learning in protein science, J. Mol. Biol. **433**, 167232 (2021).

[15] G. Masrati, M. Landau, N. Ben-Tal, A. Lupas, M. Kosloff, and J. Kosinski, Integrative structural biology in the era of accurate structure prediction, J. Mol. Biol. **433**, 167127 (2021).

[16] M. Andrec, D. A. Snyder, Z. Zhou, J. Young, G. T. Montelione, and R. M. Levy, 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 (2007).

[17] A. A. Rashin, A. H. L. Rashin, and R. L. Jernigan, Protein Flexibility: Coordinate Uncertainties and Interpretation of Structural Differences, Acta Crystallogr. D. **65**, 1140 (2009).

[18] B. Venkatakrishnan, M. L. Palii, M. Agbandje-McKenna, and R. McKenna, Mining the protein data bank to differentiate error from structural variation in clustered static structures: An examination of hiv protease, Viruses **4**, 348 (2012).

[19] J. K. Everett, R. Tejero, S. B. K. Murthy, T. B. Acton, J. M. Aramini, M. C. Baran, J. Benach, J. R. Cort, A. Eletsky, F. Forouhar, R. Guan, A. P. Kuzin, H. W. Lee, G. Liu, R. Mani, B. Mao, J. L. Mills, A. F. Montelione, K. Pederson, R. Powers, T. Ramelot, P. Rossi, J. Seetharaman, D. Snyder, G. V. T. Swapna, S. M. Vorobiev, Y. Wu, R. Xiao, Y. Yang, C. H. Arrowsmith, J. F. Hunt, M. A. Kennedy, J. H. Prestegard, T. Szyperski, L. Tong, and G. T. Montelione, A community resource of experimental data for nmr / x-ray crystal structure pairs, Protein Sci. **25**, 30 (2016).

[20] G. L. Hura, C. D. Hodge, D. Rosenberg, D. Guzenko, J. M. Duarte, B. Monastyrskyy, S. Grudinin, A. Kryshtafovych, J. A. Tainer, K. Fidelis, and S. E. Tsutakawa, Small angle x-ray scattering-assisted protein structure prediction in casp13 and emergence of solution structure differences, Proteins. **87**, 1298 (2019).

[21] M. L. Lynch, M. F. Dudek, and S. E. J. Bowman, A searchable database of crystallization cocktails in the pdb: Analyzing the chemical condition space, Patterns **1**, 100024 (2020).

[22] N. J. Fowler, A. Sljoka, and M. P. Williamson, The accuracy of nmr protein structures in the protein data bank, Structure **29**, 1430 (2021).

[23] M. A. Pak, K. A. Markhieva, M. S. Novikova, D. S. Petrov, I. S. Vorobyev, E. S. Maksimova, F. A. Kondrashov, and D. N. Ivankov, Using alphafold to predict the impact of single mutations on protein stability and function, bioRxiv 10.1101/2021.09.19.460937 (2021).

[24] G. R. Buel and K. J. Walters, Can alphafold2 predict the impact of missense mutations on structure?, Nature Structural & Molecular Biology **29**, 1 (2022).

[25] M. Akdel, D. E. V. Pires, E. P. Pardo, J. Jänes, A. O. Zalevsky, B. Mészáros, P. Bryant, L. L. Good, R. A. Laskowski, G. Pozzati, A. Shenoy, W. Zhu, P. Kundrotas, V. R. Serra, C. H. M. Rodrigues, A. S. Dunham, D. Burke, N. Borkakoti, S. Velankar, A. Frost, J. Basquin, K. Lindorff-Larsen, A. Bateman, A. V. Kajava, A. Valencia, S. Ovchinnikov, J. Durairaj, D. B. Ascher, J. M. Thornton, N. E. Davey, A. Stein, A. Elofsson, T. I. Croll, and P. Beltrao, A structural biology community assessment of alphafold2 applications, Nature Structural & Molecular Biology **29**, 1056 (2022).

[26] Y. Zhang, P. Li, F. Pan, H. Liu, P. Hong, X. Liu, and J. Zhang, Applications of alphafold beyond protein structure prediction, bioRxiv 10.1101/2021.11.03.467194 (2021).

[27] J. P. Roney and S. Ovchinnikov, State-of-the-art estimation of protein model accuracy using alphafold, Phys. Rev. Lett. **129**, 238101 (2022).

[28] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, The Protein Data Bank, Nucleic Acids Res. **28**, 235 (2000).

[29] K. S. Sarkisyan, D. A. Bolotin, M. V. Meer, D. R. Usmanova, A. S. Mishin, G. V. Sharonov, D. N. Ivankov, N. G. Bozhanova, M. S. Baranov, O. Soylemez, N. S. Bogatyreva, P. K. Vlasov, E. S. Egorov, M. D. Logacheva, A. S. Kon-

drashov, D. M. Chudakov, E. V. Putintseva, I. Z. Mamedov, D. S. Tawfik, K. A. Lukyanov, and F. A. Kondrashov, Local fitness landscape of the green fluorescent protein, Nature **533**, 397 (2016).

[30] F. J. Poelwijk, M. Socolich, and R. Ranganathan, Learning the pattern of epistasis linking genotype and phenotype in a protein, Nat. Commun. **10**, 4213 (2019).

[31] C. J. Markin, D. A. Mokhtari, F. Sunden, M. J. Appel, E. Akiva, S. A. Longwell, C. Sabatti, D. Herschlag, and P. M. Fordyce, Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics, Science **373**, eabf8761 (2021).

[32] M. R. Mitchell, T. Tlusty, and S. Leibler, Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings, Proceedings of the National Academy of Sciences **113**, E5847 (2016), https://www.pnas.org/doi/pdf/10.1073/pnas.1609462113.

[33] J. P. Eckmann, J. Rougemont, and T. Tlusty, Colloquium: Proteins: The physics of amorphous evolving matter, Rev. Mod. Phys. **91**, 031001 (2019).

[34] H. Zhang, T. Zhang, K. Chen, S. Shen, J. Ruan, and L. Kurgan, On the relation between residue flexibility and local solvent accessibility in proteins, Proteins. **76**, 617 (2009).

[35] R. A. Stein and H. S. Mchaourab, Speach_af: Sampling protein ensembles and conformational heterogeneity with alphafold2, Plos Comput. Biol. **18**, 1 (2022).

[36] T. C. Terwilliger, D. Liebschner, T. I. Croll, C. J. Williams, A. J. McCoy, B. K. Poon, P. V. Afonine, R. D. Oeffner, J. S. Richardson, R. J. Read, and P. D. Adams, Alphafold predictions: Great hypotheses but no match for experiment, bioRxiv 10.1101/2022.11.21.517405 (2022).

[37] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, and D. Baker, Accurate prediction of protein structures and interactions using a three-track neural network, Science **373**, 871 (2021).