bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

AlphaFold2 can predict single-mutation effects

John M. McBride,[1,*] Konstantin Polev,[1,2] Amirbek Abdirasulov,[3]
Vladimir Reinharz,[4] Bartosz A. Grzybowski,[1,5,†] and Tsvi Tlusty[1,5,‡]

[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]*Department of Computer Science and Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea*
[4]*Université du Québec à Montréal, Canada*
[5]*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 the localized structural deformation between protein pairs differing by only 1-3 mutations – as measured by the effective strain – is correlated across 3,901 experimental and AF-predicted structures. Furthermore, analysis of ∼11,000 proteins shows that the local structural change correlates with various phenotypic changes. These findings suggest that AF can predict the range and magnitude of single-mutation effects on average, and we propose a method to improve precision of AF predictions and to indicate when predictions are unreliable.

Alteration of one or few amino acid residues can affect structure [1–3] and function [4, 5] of a protein and, in extreme cases, be the difference between health and disease [6, 7]. Understanding structural consequences of point mutations is important for drug design [8, 9] and could also accelerate optimization of enzymatic function via directed evolution [10, 11]. In these and other applications, theoretical models [12] could be of immense help, provided they are sufficiently accurate. In this context, AlphaFold2 [13] 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 [14–16]. Furthermore, as accuracy improves, evaluating the quality of predictions becomes increasingly complicated by the inherent noise in experimental measurements [16–23]. 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 [24–28]. 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 [29]. 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 set of proteins from the Protein Data Bank (PDB) [30], and phenotype data from high-throughput experiments [31–33]. 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 deformation is probed by the effective strain (ES) measure. We show that ES computed between a pair of PDB structures is correlated with the ES computed for the corresponding pair of structures predicted by AF. Furthermore, analysis of ∼11,000 proteins whose function was probed in three high-throughput studies shows significant correlations between AF-predicted ES and three categories of phenotype (fluorescence, folding, catalysis) across three experimental data sets [31–33]. These sets of correlations suggest that AF can predict the range and magnitude of single-mutation effects. We provide new tools (github.com/mirabdi/PDAnalysis) for computing deformation in proteins, and a methodology for increasing the precision of AlphaFold predictions of mutation effects. Altogether, these results indicate that AF can be used to predict physicochemical effects of missense mutations, undamming vast potential in the field of protein design and evolution.

AF can predict local structural change.— We illustrate our approach by analyzing wild-type (WT; `6BDD_A`) and single-mutant (`6BDE_A`, A71G) structures of H-NOX protein from *K. algicida* (Fig. 1D) [34]. To quantify local deformation, we calculate the effective strain (ES) per residue S_i (See App. A) for, respectively, experimental and AF-predicted pairs of structures (Fig. 1A). The ES is the mean relative change in distance from C_α of residue i to neighboring C_α positions within a range of 13 Å. ES provides a robust estimate of the magnitude of local strain, which accounts also for non-affine deformation in addition to affine deformation [35–39]. Like the frame-aligned-point-error (FAPE) measure used in training AF [13], ES is invariant to alignment. In H-NOX, we observe that the S_i is highest at, and decays away from the mutated site, showing a correlation with the distance from the mutated site (Fig. 1B). We find that S_i is correlated across PDB and AF structures (Fig. 1C,E). Taken together, these correlations suggest that S_i is a sensitive measure of local structural change, and that AF is capable of predicting such structural change upon mutation.

Experimental measurement variability limits evaluation.— Before exploring AF predictions in more detail, we first examine variation within experimental structures by comparing repeat measurements of the same protein. In Fig. 1F we show the distribution of S_i calculated for all residues in all pairs (Supplemental Material (SM) Sec. 1A [40]) of protein structures with identical sequences (number of mutations, $M = 0$); we excluded pairs where the crystallographic group differed (SM Sec. 1B [40]). Protein structures vary considerably between repeat measurements (average ES is $\langle S_i \rangle = 0.018$, and the average Root Mean Square Deviation is RMSD = 0.24 Å). In comparison, differences between repeat predictions of AF are much lower ($\Delta S_i = 0.005$,

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

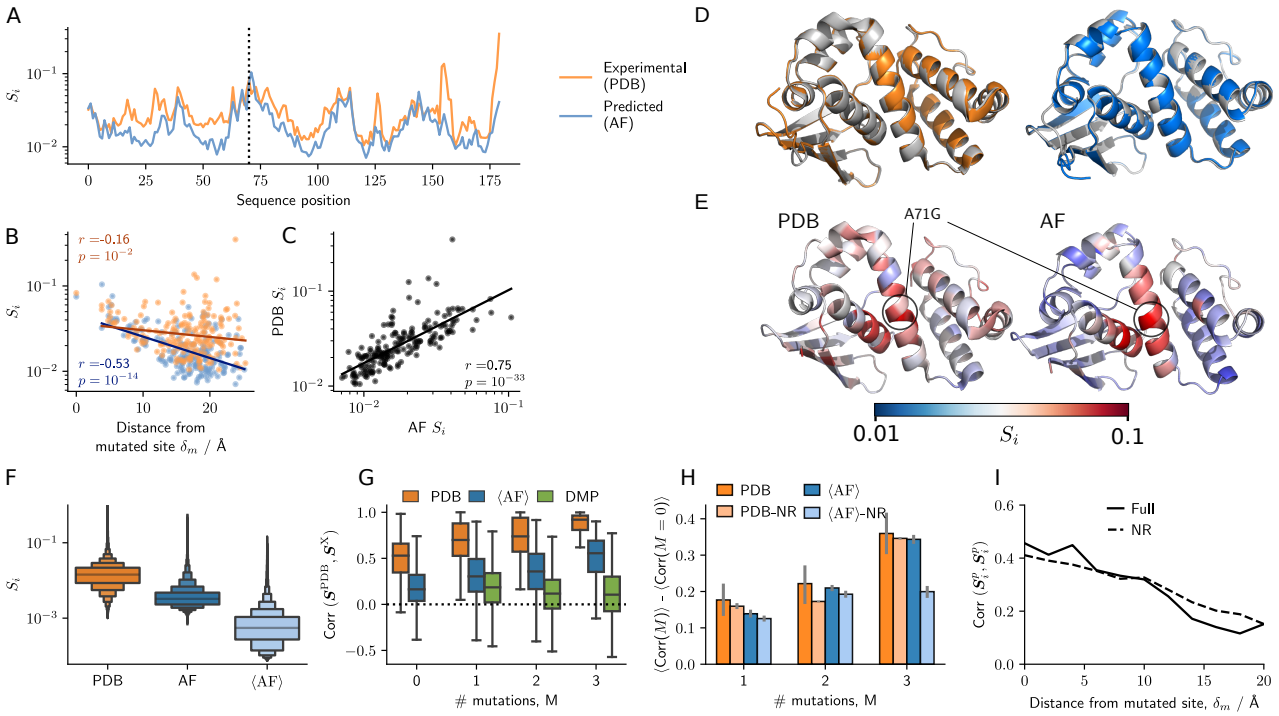

FIG. 1. A: Local deformation per residue measured by effective strain, S_i, between wild-type (WT) and mutant (A71G) H-NOX protein, for experimental (orange) and AF-predicted (blue) structures. Dotted line indicates the mutated residue. B: S_i vs distance from the nearest mutated site, δ_m. C: Comparison of S_i obtained from experimental and predicted structures. D: Overlaid WT (grey, `6BDD_A`) and mutant (colour, `6BDE_A`), experimental (orange) and predicted (blue) structures. E: Wild type protein with residues coloured by S_i; location of A71G mutation is shown. F: Distribution of S_i between matched pairs of structures with the same sequence ($M = 0$), for PDB, AF, and averaged AF ($\langle$AF$\rangle$) structures. G: Distribution of correlation between PDB strain fields and equivalent fields from PDB, AF and DMPfold, shown for different numbers of mutations, M. H: Residual correlation that is due to mutations, shown for the full dataset and a non-redundant version (NR); whiskers show bootstrapped 95 % confidence intervals. I: Correlation between PDB and $\langle$AF$\rangle$ strain fields, $\mathbf{S}_i^p$, across all pairs p and residues i that are within a distance δ_m from a mutated site, shown for the full dataset and a non-redundant version (NR).

RMSD =0.11 Å). For example, the experimental RMSD between WT and mutant H-NOX is 1.6 Å, while the AF-predicted RMSD is 0.3 Å. We can refine AF predictions further by making multiple repeat predictions and averaging over the local neighborhoods ($\langle$AF$\rangle$ in Fig. 1F, App. B), which results in even lower differences ($\Delta S_i = 0.001$). We find that averaging decreases deformation away from mutated residues, while preserving deformation in mutated areas (SM Sec. 6 [40]), thus we henceforth report results for averaged structures, except where noted. The variation between experimental measurements might mask the deformation due to mutation, and therefore limits our ability to evaluate AF predictions.

Mutation effects are measurable in PDB structures.— To quantify how well we can measure mutation effects from PDB structures, we compare deformation between two matching pairs of PDB structures with identical ($M = 0$) and non-identical ($M > 0$) sequences (SM Sec. 5B [40]) of length L (number of residues). For each pair, we calculate the strain fields, $\mathbf{S} = (S_1, \ldots, S_L)$, which record ES values for all residues, and we calculate Pearson's correlation coefficient r as in Fig. 1C. We find that even among protein structures with identical sequences, strain fields are highly correlated (Fig. 1G). This occurs because the magnitude of positional fluctuations depends on local flexibility; more flexible regions exhibit higher strain in repeat measurements (App. B). Thus, a portion of the $\mathbf{S}$ correlation in Fig. 1C is due to effects other than mutation. Despite this, we find that correlations are much higher when comparing pairs of structures that differ by one or more mutations ($M > 0$), and correlations increase with M (Fig. 1G). Thus, the strength of PDB-PDB deformation correlations is partly due to differences in local flexibility, and partly also due mutations.

Mutation effects are correlated across PDB and AF structures.— To evaluate the performance of AF in predicting mutation effects, we calculate correlations between PDB and AF-predicted strain fields, $\mathbf{S}^\mathrm{PDB}$ and $\mathbf{S}^\mathrm{AF}$, calculated for all matched pairs of proteins (SM Sec. 5B [40]). The PDB-$\langle$AF$\rangle$ correlations between pairs of structures with identical sequences ($M = 0$) are lower than PDB-PDB correlations (Fig. 1G), as are the correlations for non-identical sequences ($M > 0$). Nonetheless, the correlations are significant and they increase with M. To put this result in context, the PDB-AF correlations are considerably higher than correlations obtained by using another algorithm to predict protein structure (DMPfold2) [41]. To compare the degree of correlation that is due to mutation effects, we plot the mean correlation for non-identical sequences $\langle \mathrm{Corr}\big(M \in \{1,2,3\}\big)\rangle$ subtracted from the mean correlation that can be attributed to fluctuations, $\langle \mathrm{Corr}\big(M = 0\big)\rangle$. Fig. 1H shows that the degree of correlation due to mutations is as high for AF-PDB comparisons as it is for PDB-PDB comparisons. Since many protein families are

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

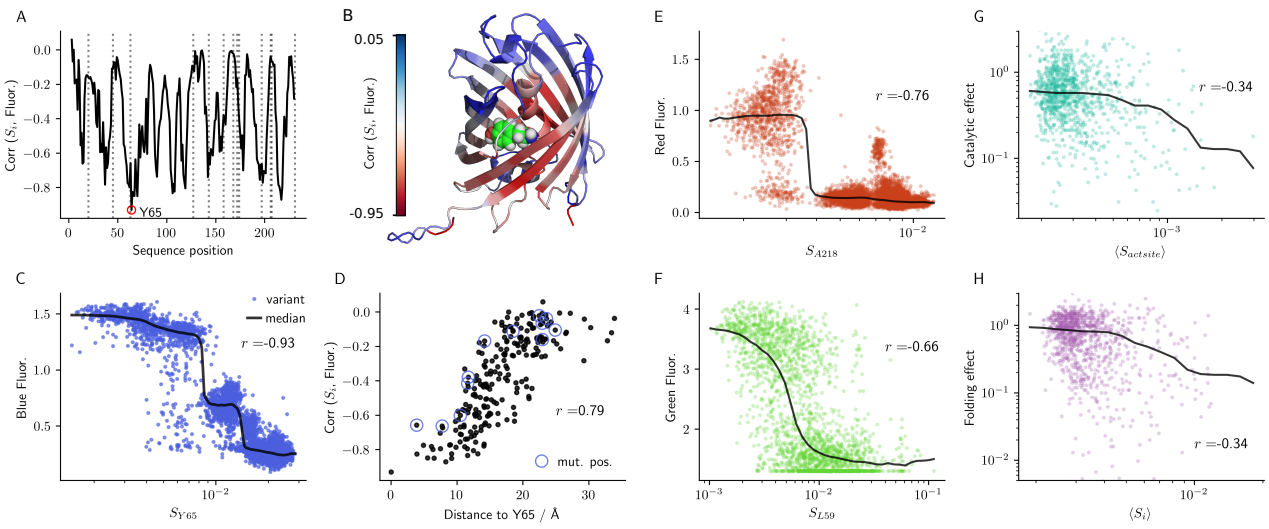

FIG. 2. A: Correlation (Pearson's r) between blue fluorescence (mTagBFP2) and AF-predicted effective strain (ES), S_i, between WT and 8,191 variants for all sequence positions i; positions of mutated residues are shown by dotted lines; chromophore site (Y65) is indicated (red circle). B: Structure of BFP, with each residue colored according to Corr(S_i, Fluor.) (A); Y65 atoms are shown as spheres. C: Strain at residue Y65 vs. fluorescence for mTagBFP2 variants. D: Fluorescence-strain correlation per residue vs. distance from residue i to Y65; mutated positions are indicated (blue circle). E-H: Correlations between: S_{A218} and red fluorescence (mKate2); S_{L59} and green fluorescence (GFP); catalytic activity and S at the active site (PafA); folding ability (fraction of active enzymes) and average strain, $\langle S_i \rangle$, of the 50 residues that correlate best with folding ability (PafA).

over-represented in the PDB, we repeat the analyses on non-redundant sets of proteins (SM Sec. 1C [40]), finding that AF-PDB correlations are still comparable to PDB-PDB correlations (NR in Fig. 1H).

AF predicts the range of mutation effects.— Fig. 1G-H shows that within matched protein pairs, deformation is correlated between PDB and AF, although the magnitude of deformation can differ (Fig. 1F). This indicates that AF is at least correctly predicting the range and the relative strength of the effect of a mutation. On average, AF predicts that mutations can produce changes in structure up to 16-18 Å (SM Sec. 7 [40]), whereas the average range in the PDB data is only14 Å due to the higher measurement variance in the PDB. This suggests that AF correctly predicts the range of a mutation's effect on structure.

AF predicts the relative magnitude of mutation effects.— It is essential to be able to predict whether a mutation will lead to a big or small effect on structure. While the previous analysis did not show this, we directly address this problem by examining whether predicted effects correlate with empirical effects *across* proteins. To do this, we group S_i values from all matched pairs p by distance from the nearest mutated residue, δ_{m} (in bins of 2 Å), to get sets of $\mathbf{S}_i^p$ for both PDB and $\langle$AF$\rangle$ pairs of structures. This allows us to compare ES magnitudes across proteins, by calculating the correlation between $\mathbf{S}_i^p$ for PDB and $\langle$AF$\rangle$. At mutated sites, the correlation is quite high, and decreases away from the mutated site as expected (Fig. 1I); this is also true for the non-redundant sample. Hence, AF is capable of distinguishing between mutations that have relatively large or small effects on structure.

Phenotypic change correlates with AF-predicted ES.— An orthogonal test of whether AF can predict the effect of a mutation is to study correlations between the effective strain (ES), S_i, and phenotypic change. This approach avoids the pitfalls associated with noisy PDB measurements, and allows us to test predictions of structures that AF was not directly trained on. However, the link between structure and function is often unknown, and likely quite complex. Therefore, a lack of a correlation between S_i 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 correlations between S_i and phenotype is strong evidence that AF can be predictive in estimating the effect of mutations. We study three data sets from high-throughput experiments, covering three distinct phenotypes (SM Sec. 2 [40]): (i) blue and red fluorescence is measured for 8,192 sequences linking mTagBFP2 (blue) and mKate2 (red) [32]; (ii) green fluorescence is measured for 2,312 GFP sequences [31]; (iii) folding (fraction of active enzymes) and catalytic (k_{cat}) effects of mutations are measured for PafA [33] (SM Sec. 2C [40]).

We find significant correlations between phenotype and AF-predicted ES (compared to WT) for all phenotypes (Fig. 2). It is possible to predict blue, red and green fluorescence (Pearson's $r = -0.93$, $r = -0.76$, $r = -0.67$) by measuring the ES at residues Y65, A218 and L59, respectively (Fig. 2C,E,F). There are many other residues at which deformation measured by ES is predictive of fluorescence (Fig. 2A-B), and these residues are found to be closer to residue Y65 (Fig. 2D, Y65 covalently binds to a chromophore); this is despite no mutations to Y65, which suggests that AF can predict allosteric effects. We also find weaker, yet significant correlations between ES and the empirical effects of mutations on folding and catalytic activity (Fig. 2G-H). For catalytic activity, we measure mean deformation at the active site; for the folding effect, we measure mean ES between the 50 residues that correlate best with the folding effect (SM Sec. 9 [40]).

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In contrast, we do not find consistent correlations with RMSD, a standard estimate of AF accuracy [13], indicating that local deformation, as measured by the ES, is more appropriate for measuring mutational effects (SM Sec. 10). In some cases, performance is heavily dependent on which pre-trained model (SM Fig. 4 [40]) is used: surprisingly, we found that using the highest ranked (by pLDDT; see SM Sec. 3 & 11 [40]) models resulted in worse performance for phenotypic change (SM Fig. 4 [40]), and performance for structural change was close to average (SM Sec. 12 [40]). Taken together, these results provide evidence that AF can be used to predict the structural effect of a single mutation.

ES correlates with phenotypic change for wild-type proteins.— It is quite unexpected that ES, S_i, 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 (WT) proteins, where the structure is adapted for function through evolution – any deviation from this optimal structure is likely to diminish protein function. We find that high correlations are only found within $M \leq 8$ mutations from the WT, and phenotype-ES correlations are much weaker between non-WT pairs (SM Sec. 9 [40]). Thus we conclude that S_i is a good predictor of phenotypic change from native protein sequences. For studying phenotypic change away from optima in phenotype landscapes, another mapping from structure to function is needed.

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 (Fig. 1F), and further hindered by the lack of sequence pairs that are suitable for comparison (SM Sec. 1 [40]). Likewise, predicting phenotypic change from structure alone ought to be challenging, to say the least. Despite these steep hurdles, we have shown, using effective strain (ES) as a measure of deformation, that differences between AF-predicted structures do correlate with both structural (Fig. 1) and phenotypic changes (Fig. 2) in empirical data. Examining individual pairs of PDB structures, mutation effects are masked by fluctuations, but this inherent noise is filtered by analyzing the statistics of many pairs, demonstrating that AF is accurate. The difficulties in assembling sufficient data for validation highlight that the age of experimental protein structure identification is far from over [42], despite the success of AF and RoseTTAFold [13, 43]. Our methodology for evaluating mutation effects using deformation can be used in future empirical evaluation of mutation effects.

Advice for using AF to study mutations.— We find higher correlations between AF and PDB when mutations are in less flexible regions of proteins, and when mutations have large effects (App. C). One can quickly estimate flexibility using pLDDT (AF's confidence in a residue's predicted position, or a proxy measure of rigidity; SM Sec. 11 [40]), but it is more useful to measure the variance of AF predictions by predicting multiple structures (App. B, SM Sec. 6 [40]). Depending on the flexibility, and mutation effect size, one can achieve much more reliable estimates of mutation effects by averaging across many repeat structures. We advise against using templates in predictions (used by default in AF models 1 and 2), since this appears to offer at best negligible increases in accuracy, and we found one example where including templates made the predictions much worse (SM Sec. 12 [40]). We recommend using effective strain as a measure of local deformation, rather than using RMSD or pLDDT. We provide code for calculating deformation, producing average structures, and calculating repeat-prediction variance at github.com/mirabdi/PDAnalysis.

AF predicts structure, not folding.— 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 mechanisms [44]. To see whether 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, mean pLDDT by itself proved insufficient to sort folding from non-folding proteins (SM Sec. 11B [40]). 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.

Local deformation should be used to measure mutation effects.— 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. Some studies suggest that AF and RoseTTTaFold can be indirectly used to predict phenotype, but not by comparing structures [26–28]. Two studies have reported negative results [24, 25], which we attribute primarily to their use of pLDDT and RMSD – measures much less precise of mutational effects compared to strain (SM Sec. 10-11 [40]). In one study, the authors found only weak correlations between pLDDT and fluorescence using the same GFP dataset used here. Although we do not expect pLDDT to strongly correlate with fluorescence, we do find higher correlations than those reported in [24] by examining allosteric effects (SM Sec. 11A [40]). In another analysis [25], the authors appear to assume that structure-disrupting mutations should result in a large change in predicted structure or pLDDT [25]. We first note that this paper only studied three proteins, limiting our ability to draw general conclusions. We also see that the deformation due to mutations in one of these proteins is higher than 96 % of mutation effects in our PDB sample (SM Sec. 13 [40]); it is possible that such large deformation is predictive of destabilization, and testing this is a promising future direction. Ultimately, we think the present study has demonstrated that deformation (measured by ES) is a more robust measure of structural change upon mutation.

Limitations.— Our structural analysis is limited to showing statistical correlations, and more precise experimental measurements are needed to validate the prediction accu-

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

racy of single proteins. Likewise, we are limited to evaluating structural change in the actual training data, but a less biased evaluation may become possible as more mutation effects are empirically determined. Further work is needed to more extensively examine the effects of MSA coverage and depth on mutation prediction accuracy. As for the phenotypic effect, we analyzed two protein folds and three phenotypes, a this analysis ought to be replicated on a greater variety of proteins and phenotypes.

In summary, we showed here that AF predictions of local structural change, probed by strain [35–38], 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 [10], and design drugs *in silico* [11]. 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 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, Jean-Pierre Eckmann, Martin Steinegger and Milot Mirdita for discussions. We thank Jacque Rougemont for providing code to calculate shear. This work was supported by the Institute for Basic Science (IBS-R020-D1).

Appendix A: Calculating local deformation.— As a measure of local deformation, we compute the effective strain (ES), S_i. ES is simply the mean relative change of the interparticle distances around a given residue and is partially correlated with shear strain (SM Sec. 4-5 [40]). To calculate S_i per residue i, we first define a neighborhood N_i that includes the $n_i = |N_i|$ residues $j \in N_i$ whose C_α positions $\mathbf{r}_j$ are within 13 Å of $\mathbf{r}_i$, the C_α position of residue i (in both reference and target structures). We obtain a $3 \times n_i$ *neighborhood* tensor $\mathbf{D}_i$ whose n_i rows are the distance vectors, $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. We calculate, respectively, $\mathbf{D}_i$ and $\mathbf{D}'_i$ for the two structures we are comparing (*e.g.*, WT and mutant), and rotate $\mathbf{D}'_i$ to maximize overlap between the tensors. The ES is the average over the n_i neighbors of the relative change in the distance vectors,

$$S_i = \left\langle \frac{|\Delta \mathbf{r}_{ij}|}{|\mathbf{r}_{ij}|} \right\rangle = \frac{1}{n_i} \sum_{j \in N_i} \frac{|\mathbf{r}_{ij} - \mathbf{r}'_{ij}|}{|\mathbf{r}_{ij}|} \,. \tag{1}$$

We have evaluated several other local metrics, similar in nature to ES, finding that the conclusions are not very sensitive to the specific choice of metric or neighborhood cutoff (SM Sec. 4-5 [40]). We only include AF-predicted residues in strain calculations if pLDDT > 70, and treat them as disordered otherwise.

Appendix B: Averaging local neighborhoods increases accuracy.— Since AF predictions are stochastic, repeat predictions vary. We find that deformation between repeat predictions of the same protein leads to non-negligible ES (Fig. 1G). The ES is higher in flexible regions, which is indicated by higher B-factor, solvent accessibility (RSA), and lower pLDDT (Fig. 3). It is possible to obtain more reliable estimates of mutation effects by averaging across local neighborhoods, $\mathbf{D}_i$, in repeat predictions (SM Sec. 6 [40]). Our average structures ($\langle$AF$\rangle$) are typically averaged over all 5 AF models, with one set of predictions from DeepMind's AF implementation, and five sets of predictions from ColabFold's AF implementation [45]. Averaging typically increases deformation-phenotype correlations (SM Sec. 6B [40]). One exception is the mTagBFP2/mKate2 dataset, where DeepMind's implementation of AF produces a better correlation than the average; we find that this is due to the ColabFold implementation performing poorly on this specific protein (SM Sec. 6B [40]). We see little increase in PDB-AF structure correlations (SM Sec. 6A [40]), and we attribute this to limitations of higher repeat-measurement variability in PDB structures.

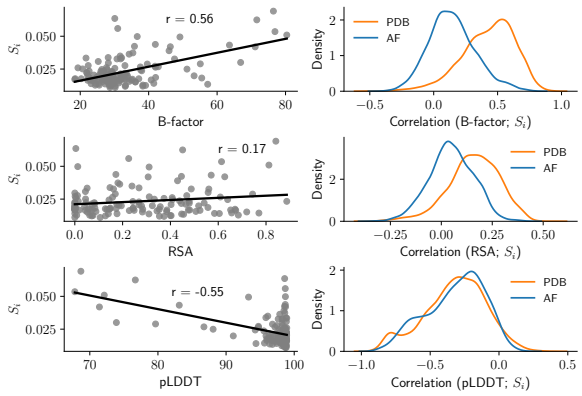

FIG. 3. **Fluctuations are greater in flexible regions.** Deformation (ES) between experimental hen lysozyme structures (`194L_A` and `6RTA_A`), S_i is correlated (Pearson's r) with B-factor, relative solvent accessibility (RSA), and pLDDT (left). Distributions (kernel density estimates) of correlations for all proteins (right).

Appendix C: When do AF predictions correlate with PDB data? — Here we assess why AF sometimes predicts mutation effects similar to those measured in experimental structures (Fig. 1G). Across all proteins, AF-PDB correlations are higher for mutant pairs of proteins in two situations (Fig. 4, SM Sec. 8 [40]): when flexibility is low (low B-factor, low RSA, high pLDDT, high ES when comparing repeat predictions $\langle S_i \rangle$); and when mutations have large effects that are easier to measure (high PDB-PDB correlation, high deformation at mutated site S_m, BLOSUM score). One might expect a negative correlation with the frequency of mutation in MSA, as more frequent mutations might have smaller effects; instead, it appears that wider MSA coverage leads to more evolutionary information that improves predictions, but this needs to be tested further. There was no significant effect due to secondary structure type or MSA size (SM Fig. 21 [40]).

* 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

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

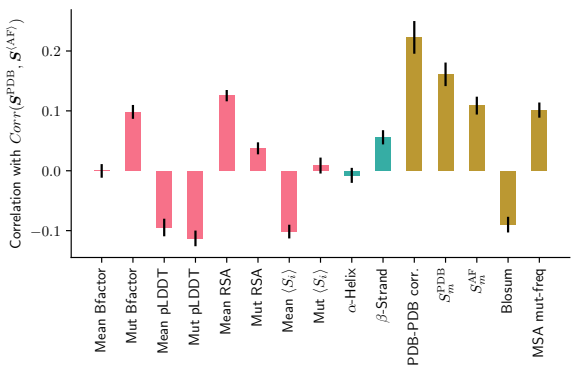

FIG. 4. **Correlations are higher if mutations have large effects in rigid regions.** A: Pearson's correlation between PDB-AF **S**-correlation and: mean and mutated residue values of flexibility (B-factor, RSA, pLDDT, $\langle S_i \rangle$); fraction of secondary structure (α-helix or β-sheet); magnitude of mutation effect (PDB-PDB **S** correlation, ES at mutated site in PDB and AF, $S_{\mathrm{m}}^{\mathrm{PDB}}$ and $S_{\mathrm{m}}^{\mathrm{AF}}$, BLOSUM score, frequency of mutation in MSA). Results are shown for the non-redundant sample; whiskers show bootstrapped standard deviations.

(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 **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 **55**, 4583 (2016).

[5] J. M. McBride, J.-P. Eckmann, and T. Tlusty, General Theory of Specific Binding: Insights from a Genetic-Mechano-Chemical Protein Model, Molecular Biology and Evolution **39**, msac217 (2022).

[6] 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).

[7] 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).

[8] 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, Proc. Natl. Acad. Sci. USA **110**, 13067 (2013).

[9] 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).

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

[11] 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).

[12] 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).

[13] 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).

[14] 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).

[15] 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).

[16] 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).

[17] 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).

[18] 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).

[19] 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).

[20] 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).

[21] 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).

[22] 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).

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

[24] M. A. Pak, K. A. Markhieva, M. S. Novikova, D. S. Petrov, I. S. Vorobyev, E. S. Maksimova, F. A. Kondrashov, and

bioRxiv preprint doi: https://doi.org/10.1101/2022.04.14.488301; this version posted September 3, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D. N. Ivankov, Using alphafold to predict the impact of single mutations on protein stability and function, PLOS ONE **18**, 1 (2023).

[25] G. R. Buel and K. J. Walters, Can alphafold2 predict the impact of missense mutations on structure?, Nat. Struct. Mol. Biol. **29**, 1 (2022).

[26] 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, Nat. Struct. Mol. Biol. **29**, 1056 (2022).

[27] 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).

[28] S. Mansoor, M. Baek, D. Juergens, J. L. Watson, and D. Baker, Accurate mutation effect prediction using rosettafold, bioRxiv 10.1101/2022.11.04.515218 (2022).

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

[30] 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).

[31] 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. Kondrashov, 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).

[32] 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).

[33] 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).

[34] C. W. Hespen, J. J. Bruegger, Y. Guo, and M. A. Marletta, Native alanine substitution in the glycine hinge modulates conformational flexibility of heme nitric oxide/oxygen (h-nox) sensing proteins, ACS Chem. Biol. **13**, 1631 (2018).

[35] M. R. Mitchell, T. Tlusty, and S. Leibler, Strain analysis of protein structures and low dimensionality of mechanical allosteric couplings, Proc. Natl. Acad. Sci. USA **113**, E5847 (2016).

[36] T. Tlusty, A. Libchaber, and J.-P. Eckmann, Physical model of the genotype-to-phenotype map of proteins, Phys. Rev. X **7**, 021037 (2017).

[37] S. Dutta, J.-P. Eckmann, A. Libchaber, and T. Tlusty, Green function of correlated genes in a minimal mechanical model of protein evolution, Proc. Nat. Acad. Sci. USA **115**, E4559 (2018).

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

[39] M. L. Falk and J. S. Langer, Dynamics of viscoplastic deformation in amorphous solids, Phys. Rev. E **57**, 7192 (1998).

[40] See supplemental material at url for detailed information about methods used in this letter. (2023).

[41] S. M. Kandathil, J. G. Greener, A. M. Lau, and D. T. Jones, Ultrafast end-to-end protein structure prediction enables high-throughput exploration of uncharacterized proteins, Proc. Natl. Acad. Sci. USA **119**, e2113348119 (2022).

[42] 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).

[43] 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).

[44] J. M. McBride and T. Tlusty, Slowest-first protein translation scheme: Structural asymmetry and co-translational folding, Biophysical Journal **120**, 5466 (2021).

[45] M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger, Colabfold: making protein folding accessible to all, Nature Methods **19**, 679 (2022).

[46] W. Li and A. Godzik, Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences, Method. Biochem. Anal. **22**, 1658 (2006).

[47] I. Kufareva and R. Abagyan, Methods of protein structure comparison, in *Homology Modeling: Methods and Protocols*, edited by A. J. W. Orry and R. Abagyan (Humana Press, Totowa, NJ, 2012) pp. 231–257.

[48] W. Kabsch, A Solution for the Best Rotation to Relate Two Sets of Vectors, Acta Crystallogr. A **32**, 922 (1976).

[49] Y. Zhang and J. Skolnick, Scoring function for automated assessment of protein structure template quality, Proteins. **57**, 702 (2004).

[50] A. Zemla, Lga: A Method for Finding 3d Similarities in Protein Structures, Nucleic Acids Res. **31**, 3370 (2003).

[51] V. Mariani, M. Biasini, A. Barbato, and T. Schwede, Lddt: A Local Superposition-Free Score for Comparing Protein Structures and Models Using Distance Difference Tests, Method. Biochem. Anal. **29**, 2722 (2013).

[52] J. Lubliner, *Plasticity Theory* (Courier Corporation, 2008).

[53] 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).

[54] M. Socolich, S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner, and R. Ranganathan, Evolutionary information for specifying a protein fold, Nature **437**, 512 (2005).