AI-predicted protein deformation encodes energy landscape perturbation

John M. McBride[1,*] and Tsvi Tlusty[1,2,†]

[1] *Center for Soft and Living Matter, Institute for Basic Science, Ulsan 44919, South Korea*
[2] *Departments of Physics and Chemistry, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea*

AI algorithms have proven to be excellent predictors of protein structure, but whether and how much these algorithms can capture the underlying physics remains an open question. Here, we aim to test this question using the Alphafold2 (AF) algorithm: We use AF to predict the subtle structural deformation induced by single mutations, quantified by strain, and compare with experimental datasets of corresponding perturbations in folding free energy $\Delta\Delta G$. Unexpectedly, we find that physical strain alone – without any additional data or computation – correlates almost as well with $\Delta\Delta G$ as state-of-the-art energy-based and machine-learning predictors. This indicates that the AF-predicted structures alone encode fine details about the energy landscape. In particular, the structures encode significant information on stability, enough to estimate (de-)stabilizing effects of mutations, thus paving the way for the development of novel, structure-based stability predictors for protein design and evolution.

AI has ushered in a revolution in structural biology, yet we are still in uncharted waters [1, 2]. In particular, it is not clear whether AI algorithms that predict protein structure from sequence, such as AlphaFold (AF) [3] or RoseTTAFold [4], owe their unprecedented accuracy to highly sophisticated pattern recognition or these algorithms can capture some of the many-body physics underlying protein folding. Recent studies provide evidence suggesting that AF has learned an *effective* energy functional that is searched in order to accurately predict the native structure [5], even if it includes uncommon structural motifs [6].

A stringent test for whether an AI algorithm has learned the actual *physical* energy landscape would be the capacity to probe, from the predicted structure alone, changes in the thermodynamic free energy ($\Delta\Delta G$) due to single mutations. This would indicate that the predicted structure encodes fine details about the physical energy landscape. Besides this fundamental interest, such capacity may be impactful in applications: Of particular importance for protein design and sequence generation [7–10], and for protein evolution [11–13] is the ability to predict whether a given protein sequence will lead to a stably-folded structure [3, 4]. AF has been reported to predict folded structures for proteins that are not stable [14–16]. Despite this, some analyses suggest that AF and RoseTTAFold can be used for predicting stability changes upon mutation [14, 17–20]. All this motivates us to systematically investigate here the question of whether AI-predicted structures can be used to infer changes to free energy landscapes.

To this end, we study a curated subset of 2,499 measurements of stability change ($\Delta\Delta G$) due to single mutations, taken from ThermoMutDB (TMDB) [21]. Strain is a simple general measure of deformation in proteins [22–27]. We find that the deformation upon mutation – as measured by the effective strain [14] – correlates well with $\Delta\Delta G$. We show it is essential to average over an ensemble of multiple AF-predicted structures to get precise estimates of strain due to mutation, and that most of the relevant information is gleaned from the residues within 15 Å of the mutated residue. Our initial motivation was to examine whether the AF structures encode any information about stability. Surprisingly, we found that correlations between strain and $\Delta\Delta G$ compare well against those obtained using state-of-the-art $\Delta\Delta G$ predictors, suggesting that AF predictions are highly informative of stability changes. We propose that new energy-based force fields can be developed that may provide a mechanistic understanding of the effects of mutations on stability. Since stability measurements are easier than structure determination, such a development could settle the question of whether AI algorithms have truly learned the physics of protein folding by testing on massive stability datasets [5, 6, 28].

Strain correlates with $\Delta\Delta G$ in Thermonuclease.— We first examine thermonuclease (NUC, Uniprot ID, `P00644`; *staphylococbus aureus*) – the protein that has the highest number of mutants in TMDB (491 after applying controls, Appendix A). NUC consists of a folded region (starting around K88) and an extended disordered region near the N-terminus (Fig. 1A), indicated by the low pLDDT (AF-predicted confidence score) values (Fig. 1A-B). The NUC mutants are all sampled from the folded domain (Fig. 1C). Note that we define $\Delta\Delta G$ such that an increase in ΔG upon mutation is *destabilizing*.

To measure deformation, we calculate effective strain (ES; Appendix B) per residue, S_i, between wild-type (WT) and mutant structures predicted by AF (Fig. 1D). Disordered residues always show high ES (Fig. 1D, dotted lines) due to prediction noise, regardless of mutations [14]. We therefore exclude residues whose pLDDT < 70 from the ES calculations (Fig. 1D, solid lines). Even without disordered residues, if we only compare two static structures (Fig. 1D-E, teal), we still see residual ES in regions far from the mutated site. This occurs since regions in proteins with high flexibility tend to have high variability across repeat AF predictions (Supplementary Material [SM] Fig. 1)[29]. To achieve a more accurate estimate of deformation *due to mutation*, we calculate deformation using 'averaged' AF structures (Appendix B) [14]. This drastically reduces ES in most regions (which originates chiefly from noise and fluctuations), except for regions near the mutated site (Fig. 1D-E, orange).

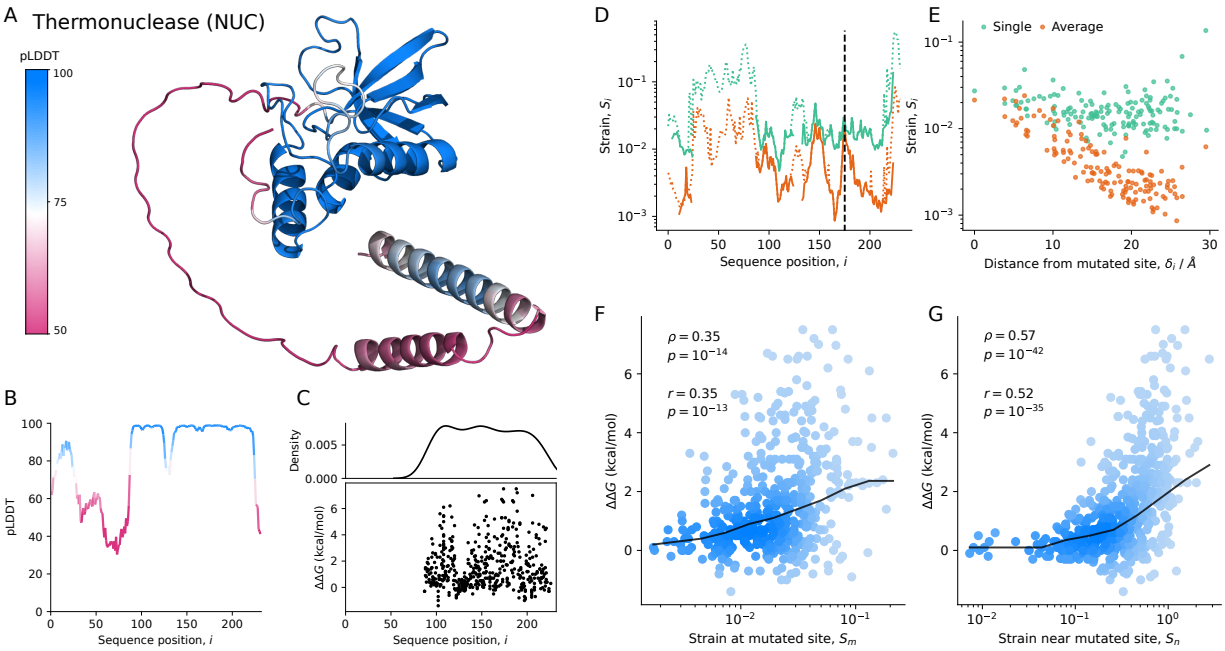

FIG. 1. **Strain calculated using AF-predicted structures correlates with $\Delta\Delta G$.** A: AF-predicted structure of thermonuclease from *staphylococcus aureus* (NUC) – the most common protein in TMDB. Residues are colored according to pLDDT (AF-predicted confidence score). B: pLDDT per residue. C: Distribution of 491 mutation sites along the sequence, and corresponding changes in stability $\Delta\Delta G$. D: Strain upon mutation (A176G, $\Delta\Delta G = 2.4$ kcal/mol) per residue; the mutated site is indicated by the black dashed line. The strain calculation either includes (dotted line) or excludes (solid line) residues with pLDDT < 70, and uses either a single pair of structures (green) or pairs of averaged structures (orange). E: Strain as a function of distance from the mutated site, δ_i, for single pairs and ensemble-averaged pairs of structures. F-G: Empirical $\Delta\Delta G$ vs. strain at the mutated site, S_m (F), and the sum of strain over all residues within 15 Å of the mutated site, S_n (G). Solid line shows the median; Spearman's ρ, Pearson's r, and corresponding p values are shown; circles are shaded by density.

Using this more precise measure of deformation due to mutation, we find a significant correlation (Spearman's $\rho = 0.35$) between strain at the mutated site S_m and change in stability $\Delta\Delta G$ (Fig. 1F). This correlation is even higher (Spearman's $\rho = 0.57$) when calculating the sum of strain over all residues within a spherical neighborhood of radius $\gamma = 15$ Å around the mutated site, S_n (Fig. 1G). See SM Fig. 2 for similar analyses without excluding low pLDDT residues and without using average structures. Our rationale for choosing $\gamma = 15$ Å will become clear in a following section. For now, we highlight that, in this particular example (NUC) it appears that AF-predicted deformation correlates quite well with empirical measurements of changes in stability.

Strain correlates with $\Delta\Delta G$ within protein families.— We expand our analysis of NUC to more protein families, again focusing on the families that have the highest coverage in TMDB. For the second-to-fifth most common proteins, we find correlations between stability change and local deformation ranging from $0.39 \leq \rho \leq 0.61$ (Fig. 2A). Extending this analysis to the 40 most common proteins reveals that there are insufficient samples to show significant strain-stability correlations in most cases. Nevertheless, in the 16 correlations that are statistically significant, the average Spearman's ρ is 0.54 (Fig. 2B) with an overall range 0.29-0.78. We see that for all cases with sufficient data, there appears to be a consistent correlation between strain and changes in stability.

Determinants of strain-stability correlation.— To better understand why strain is correlated with changes in stability, we examine the correlations between strain at individual residues S_i (not necessarily the mutated ones) and $\Delta\Delta G$, and compare this with the number of neighbors within 10 Å of each residue i. For NUC, we find that mutations in buried regions (those with many neighbors) tend to have an outsize impact on stability (Fig. 2C), as expected, given the standard paradigm of buried residues having low mutation rates [30]. In general one expects mutations of buried residues to affect more bonds and therefore inflict larger stability changes. Indeed, across the five most common proteins in the TMDB, we see a clear trend whereby the residues with the highest strain-stability correlations are amongst the most buried within that protein (Fig. 2D).

Evidently, when only a few points are sampled, the resulting correlation is not particularly informative. Also, in a large protein, more points are needed to achieve sufficient sampling. This is demonstrated most clearly in IgG-binding protein G (IgGb, `P06654`): this protein has a length $L = 448$, so even though we have 225 samples, many buried residues do not correlate with stability changes (Fig. 2D). This is because many regions have no mutations sampled

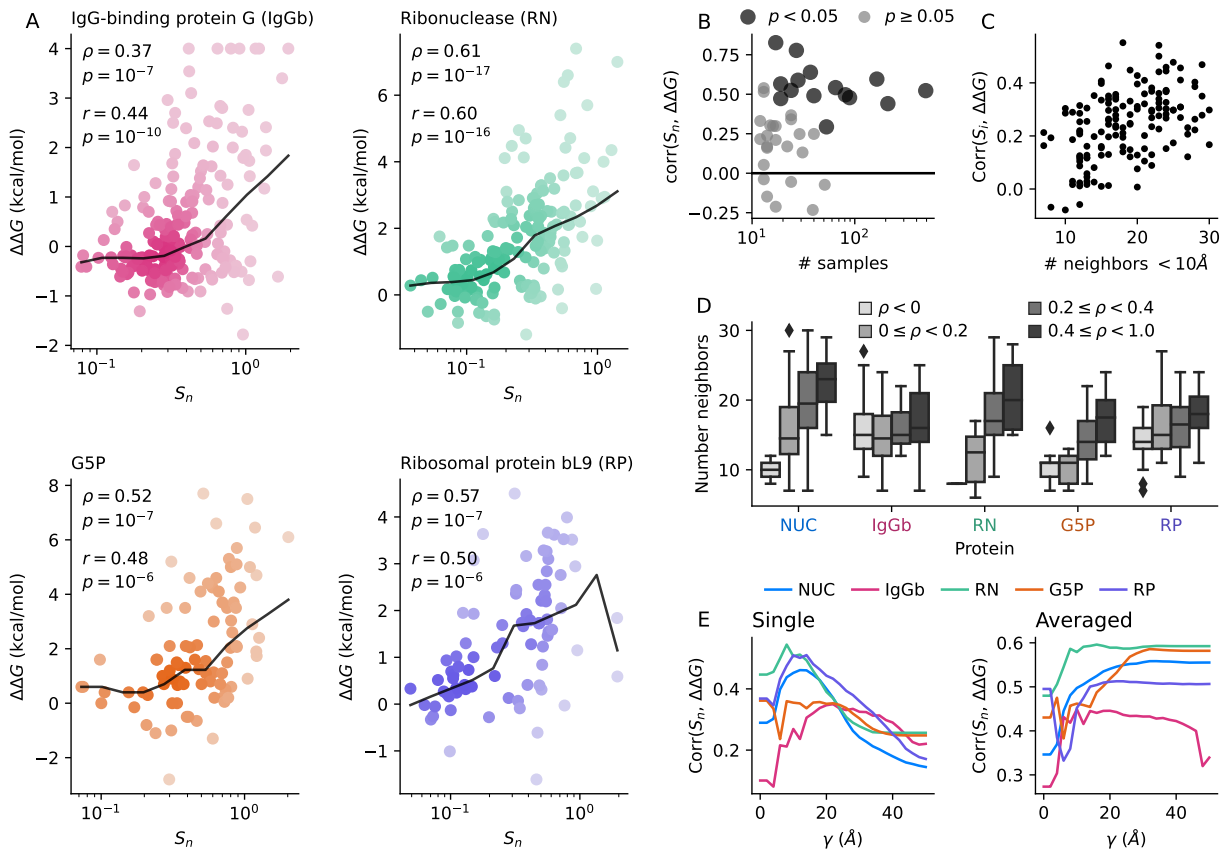

FIG. 2. **When and why does S_{n} correlate with $\Delta\Delta G$?** A: Change in stability $\Delta\Delta G$ against strain near mutated residue S_{n} for four proteins (the second to fifth most common in TMDB; the most common one, NUC, is shown in Fig. 1); Spearman's ρ, Pearson's r, and corresponding p values are shown; black line is the median; circles are shaded by density. B: $S_{\mathrm{n}}-\Delta\Delta G$ correlation for each of the 40 most common proteins in TMDB as a function of the number of samples; results with $p < 0.05$ are shown by dark circles. C: $S_i-\Delta\Delta G$ (Pearson's) correlation for each residue i as a function of the number of neighbors (for NUC only). D: Distributions of numbers of neighbors, grouped by $S_i-\Delta\Delta G$ Spearman's correlation, for the five most common proteins. E: $S_{\mathrm{n}}-\Delta\Delta G$ (Pearson's) correlation for each residue i as a function of the neighborhood threshold value γ for calculating S_{n}, for both pairs of single structures and averaged structures, for five proteins.

from them, so deformation remains low no matter how buried the residues are. Another complication arises from the abundance of disordered regions (SM Fig. 3) in which AF appears to have little capacity to predict mutation effects, while well-folded regions are small. As a result, for IgGB the link between number of neighbors and effect on stability is weak. This case highlights the need for a nuanced approach to understanding the relationship between strain and changes in stability.

Range of informative residues.— We find that the $S_{\mathrm{n}}-\Delta\Delta G$ correlation increases with γ, the radius of neighborhood used to calculate S_{n}, up to about $10 \leq \gamma \leq 20$ Å, depending on the protein family (Fig. 2E). This optimal length scale of $\sim$15 Å, could be the outcome of two possible effects:

First, we note that the residues that are most informative of stability changes are the most buried ones (Fig. 2C), and not necessarily the mutated residues. Hence, the optimal γ needs to be large enough to include some of these informative, buried residues in the S_{n} calculation. In TMDB, for example (SM Fig. 4), the average distance between buried and mutated residues is 10 Å. This length scale can be rationalized by a simple geometric argument: the average distance to the center of a sphere (the most buried part) of diameter d is $r_0 = (3/8)d$; the average volume of an amino acid is $\sim$144 Å^3 [31], so in a folded, approximately spherical domain of $\sim$100 amino acids, $r_0 \sim 11.3$ Å.

Second, when we compare pairs of single AF-predicted structures, which exhibit significant noise unrelated to mutations (SM Fig. 1), strain-stability correlations reach a maximum as a function of γ and then decrease; this decrease is not observed for averaged structures (Fig. 2E). This is consistent with reports that AF strain predictions are usually indistinguishable from noise after about 15 Å, while this range could be extended to 25 Å (and noise reduced by a factor of 4) by averaging over structures [14, 22].

In summary, we propose that the reason for the maximum information from strain around a range of 15 Å is due to both the need to increase γ to include buried residues and also because long-range mutation effects may be masked by prediction noise. In this context, we note that conformational dynamics of proteins were

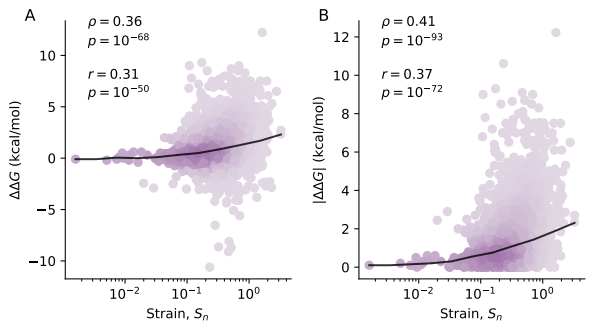

FIG. 3. $S_\mathrm{n}-\Delta\Delta G$ **correlations across proteins.** Strain near mutated site S_n vs. $\Delta\Delta G$ (A) and magnitude of stability change $|\Delta\Delta G|$ (B) for all mutations in our reduced TMDB sample of 2,499 unique mutants. , Pearson's r, and corresponding p values are shown.

reported to exhibit correlations with similar and even longer ranges, as demonstrated by normal mode analysis [22, 32, 33]. However, one needs to remember that displacement and strain – the tensorial derivative of the displacement – may show different correlation lengths [22].

Strain correlates with $\Delta\Delta G$ across protein families.— We find a moderate correlation ($\rho = 0.36$) between S_n and $\Delta\Delta G$ when comparing all the measurements together. This correlation is expected to be limited due to the inability of strain to differentiate between stabilizing and destabilizing mutations. Since Eq.(1) measures the absolute relative change in distances, $S \geq 0$ by definition, and thus the strain is invariant to reversing the reference and target structures. Indeed, we find a somewhat higher correlation ($\rho = 0.41$) between S_n and the magnitude of stability changes, $|\Delta\Delta G|$. This indicates that more generally, large structural changes lead to large stability changes, independent of the sign of the change.

We do not expect a simple mapping between S_n and $\Delta\Delta G$, given the complexity of protein structures and intramolecular interactions. We obviate protein size effects to an extent by only looking at mutation effects within $\gamma = 15$ Å, but there are other protein-specific factors – such as the degree of disorder, protein shape, flexibility, and amino acid packing – that may alter the relationship between strain and $\Delta\Delta G$ for different proteins. Nonetheless, the strain-stability correlations shown here indicate that strain due to mutation contains considerable information about stability changes that may be leveraged in subsequent development of predictors of $\Delta\Delta G$.

Strain correlates with $\Delta\Delta G$ almost as well as tailored $\Delta\Delta G$ predictors.— To put the strain-stability correlations in context, we compare them with two state-of-the-art $\Delta\Delta G$ predictors, DDMut and FoldX (Appendix C) [34, 35]. FoldX predicts ΔG from structure using empirical energy-based potentials; it is used to calculate $\Delta\Delta G$ by first a generating structure for the mutant based on a reference WT structure, which enables calculation of ΔG for both WT and mutant structures. DDMut uses a neural network to predict $\Delta\Delta G$, using a reference structure and a mutation as input.

We find that $S_\mathrm{n}-|\Delta\Delta G|$ correlations are almost as high as correlations obtained using FoldX on average, but lags behind the more recent algorithm, DDMut (Fig. 4). We were genuinely surprised by this performance since strain is a *simple* general measure of deformation and not designed specifically for stability, in contrast to FoldX and DDMut. However there is clearly room for improvement: simply by counting the number of neighbors at the mutation site we see a correlation of $\rho = 0.30$ (Fig. 4D); the mean absolute error obtained from a linear fit is 1.2 kcal/mol for both S_n and DDMut (SM Fig. 5). We note that our aim here is not to use strain to predict $\Delta\Delta G$, but rather to see whether the strain predicted by AF is informative about stability changes. But given the surprisingly high correlation, we speculate that AF-predicted structures can be leveraged to produce even better $\Delta\Delta G$ predictors.

We expected that since FoldX calculates ΔG for a structure, it would give us a more accurate estimate of $\Delta\Delta G$ than strain if we use it on AF-predicted structures. Surprisingly, we find that this method (AF-FoldX) performs worse than strain in this case ($\rho = 0.26$). Likewise, we find that calculating strain using the structures generated by FoldX results in a lower correlation with $\Delta\Delta G$ ($\rho = 0.30$, SM Fig. 6). We considered also the possibility that AF-predicted structures are not as accurate as FoldX-generated structures. However, we find that strain calculated using FoldX structures is not correlated with distance from the mutated site (SM Fig. 7), indicating that FoldX is less accurate than AF in predicting the effect of mutations on structure. These results suggest that there is a promising path for generation of new energy-based methods for $\Delta\Delta G$ prediction using AF-predicted structures.

Can AF be used to predict stability changes?— We emphasize that our aim here is not to outright develop a $\Delta\Delta G$ predictor but rather to investigate whether AF predictions are informative of stability changes. We have found that a general measure of deformation, strain, correlates quite well with $\Delta\Delta G$. Although it was not designed to be a $\Delta\Delta G$-predictor, S_n appears to be almost as good at predicting the magnitude of stability changes as state-of-the-art $\Delta\Delta G$ predictors. Of course, this needs to be tested on a larger set of measurements, and more structures are needed for experimental validation of the relationship between strain and stability (SM Sec. 1C). We also note that higher correlations have been observed for these predictors on different datasets, so this analysis should be repeated on larger sets of $\Delta\Delta G$ measurements [34, 36]. Yet, within these limitations, it seems clear that there is sufficient information in AF-predicted structures to make $\Delta\Delta G$ predictions. It stands to reason that new physics-based models can be developed to achieve even better predictions, off the back of AF-predicted structures. Alternatively,

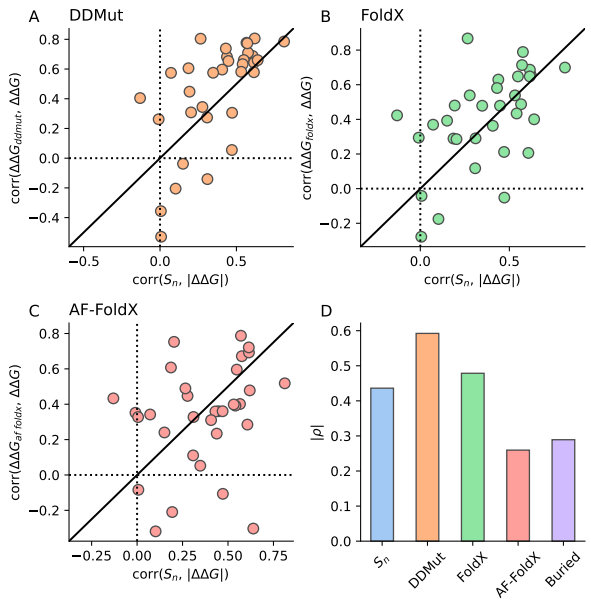

FIG. 4. **Comparison with state-of-the-art algorithms.** A-C: Correlation (Spearman's ρ; linear correlations are shown in SM Fig. 8) between strain S_n and magnitude of stability change $|\Delta\Delta G|$, vs. correlation between $\Delta\Delta G$ and predictions of DDMut (A), FoldX (B), and FoldX using AF-predicted structures (C); separate points are shown for each of the 40 most-common proteins. D: Correlations (as above) for the full sample of 2,499 unique mutants. Correlation between number of neighbors within 10 Å of the mutated site is shown as "Buried".

these structures could be used to reparameterize existing force fields such as FoldX. The recent explosion of high-throughput measurements [28, 37] will certainly lead to more machine-learning and sequence-based approaches. Nevertheless, we feel that physics-based methods are essential to offer a detailed view into the mechanistic effects of mutations on stability. The breakthrough by AF in structure prediction might offer the key to this future.

Strain-energy relation.— Our predictions allow us to deduce an effective strain-energy relation by modeling the protein as an elastic spring network (as detailed in SM Section 2) [23, 25]. Within this framework, mutation can be seen as introducing a point defect that perturbs the network [26], thereby inducing energy change $\Delta\Delta G$, which includes contributions from entropy, changing topology (breaking and forming bonds), and energy release in pre-stressed frustrated bonds [38]. The energy change is quadratic in the strain, with average *effective* spring constants K ranging between 1-10 kcal/mol/Å^2 (SM Fig. 9). These typical values of K [27] estimate the average curvature of the high-dimensional energy landscape [39]. K varies widely in and between protein families, reflecting the anisotropic geometry and heterogeneity of the landscapes.

Advice for using AF to predict mutation effects.— A previous study examined the same ThermoMutDB dataset, yet they concluded that AF cannot be used to predict stability. This is due to using changes in pLDDT to measure mutation effects, which does not appear to be reliable for this purpose (SM Fig. 10) [14]. We recommend using strain as a more robust measure of the effect of mutations on structure, particularly when using averaged structures. We recommend using about 10 to 20 (*i.e.*, 5 models, 2-4 repeats) structures to get averages, as there are diminishing returns on performance gains (SM Fig. 11). Finally, we tested the strain predicted by the recently released AlphaFold3 [40] in three TMDB mutants by comparing with AF2 predictions (SM Fig. 12-13), finding no major differences, except in loops and disordered regions.

We have studied the correspondence between AlphaFold (AF) predictions of mutation effects on structure (measured using strain) and changes in stability. We find that strain correlates well with $\Delta\Delta G$, almost as well as state-of-the-art $\Delta\Delta G$ predictors. Altogether, our findings suggest that new algorithms can be developed to extract more information from AF structures to produce accurate physics-based models of stability change upon mutation. Furthermore, quantitative mapping of the energy landscape may pave the way for more realistic modeling of functional conformational dynamics of protein machines, beyond coarse-grained elastic-network models [22, 41].

Code used to filter the ThermoMutDB can be found at github.com/jomimc/AF2_Stability_PRL_2024. We used the PDAnalysis python package to calculate strain. All AF2-predicted PDB structures are available at [42], compressed using FoldComp [43].

This work was supported by the Institute for Basic Science, Project Code IBS-R020-D1.

Appendix A: Stability Data.— We select all proteins from ThermoMutDB (TMDB) [21] with sequence length, $50 \leq L \leq 500$, and measurements for single mutants made within $293 \leq T \leq 313$ K and $5 \leq$ pH ≤ 8; this amounts to 5,078 out of 13,337 measurements. One problem we found with TMDB is that the indices in mutation codes can be associated with either Protein Data Bank (PDB) [44], Uniprot [45], or "unsigned", yet we wanted to match mutations to Uniprot sequences. Hence, if indices were matched to PDB indices that differed from Uniprot indices, we used a custom script to convert the mutation codes to the correct Uniprot indices. This script only works when the mutation code index refers to the "_atom_site.label_seq_id" entry of the mmCIF file; we excluded many cases where the TMDB mutation indices refer to idiosyncratic indexing (*i.e.*, not starting from one) in the "_atom_site.auth_seq_id" entries in the mmCIF files. Out of caution, if no correct matches were found for a protein we excluded all TMDB entries that were related to this protein (*i.e.*, if they have the same value in the 'PDB_wild" column). We excluded "unsigned", since these were ambiguous, and we discovered that some of these were labelled with incorrect Uniprot accession ids. After this procedure we are left with 3,236 measurements. When multiple $\Delta\Delta G$ measurements are available we

report the average $\Delta\Delta G$ across all measurements. We leave out a mutant if the standard deviation of the $\Delta\Delta G$ measurements is higher than 1 kT; this occurred in only about 2 % of cases. We examined why some of these cases had such high variance, and found occasional errors where the $\Delta\Delta G$ sign differed from the reference it was taken from. A full list of errors and corrections can be found in SM Section 1D. The final set of $\Delta\Delta G$ measurements includes 2,499 unique mutants. We study correlations within individual wild-type (WT) proteins (and their single mutants), and across the full set of measurements.

Appendix B: Structure Analysis.— We use the ColabFold implementation of AF to predict protein structures [46]. We run all 5 models, without using templates, and using 6 recycles, and minimization using the Amber forcefield. We run 10 repeat predictions for each sequence, for each model, and create averaged structures following [14]. We calculate effective strain, ES, which measures the deformation between mutants and WT proteins, for both averaged and non-averaged structures. ES can be described simply as the average relative change of $C_\alpha - C_\alpha$ distances between neighboring residues; we define neighbors as residues whose C_α positions are within 13 Å. We calculate ES as [14],

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

where $\mathbf{r}_{ij}$ is the distance vector between C_α positions in residue i and neighbor j in a reference structure, $\mathbf{r}'_{ij}$ is the corresponding distance vector in a target structure which has been aligned to $\mathbf{r}_{ij}$, N_i is the set of neighbors, and $n_i = |N_i|$ is the number of neighbors. We refer to ES measured at the mutated site m as S_m. We also calculate the neighborhood sum of strain S_n over residues whose distance to the mutation site δ_i, is shorter than some threshold γ,

$$S_\mathrm{n} = \sum_{\delta_i \leq \gamma} S_i \ . \tag{2}$$

We mainly use $\gamma = 15$ Å. We typically only include AF-predicted residues in strain calculations if pLDDT > 70, and treat the rest as disordered, except where otherwise noted.

Appendix C: $\Delta\Delta G$ predictors.— We use two state-of-the-art methods to predict $\Delta\Delta G$: FoldX 5 [47], and DDMut [34]. We use the API for the DDMut web server to predict $\Delta\Delta G$. For FoldX, we use the BuildModel command to generate structures of the WT and mutant sequences and $\Delta\Delta G$ predictions. We use five runs as recommended, and report average values of $\Delta\Delta G$. For each algorithm, we provide the top-ranked (by pLDDT) AF-predicted WT structure as an input, along with a list of all mutants for that protein.

* jmmcbride@protonmail.com

† tsvitlusty@gmail.com

[1] J. M. Thornton, R. A. Laskowski, and N. Borkakoti, Alphafold heralds a data-driven revolution in biology and medicine, Nature Medicine **27**, 1666 (2021).

[2] S. Subramaniam and G. J. Kleywegt, A paradigm shift in structural biology, Nature Methods **19**, 20 (2022).

[3] 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).

[4] 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).

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

[6] O. Herzberg and J. Moult, More than just pattern recognition: Prediction of uncommon protein structure features by ai methods, Proceedings of the National Academy of Sciences **120**, e2221745120 (2023).

[7] A. Goldenzweig and S. J. Fleishman, Principles of protein stability and their application in computational design, Annual Review of Biochemistry **87**, 105 (2018), pMID: 29401000.

[8] B. Kuhlman and P. Bradley, Advances in protein structure prediction and design, Nature Reviews Molecular Cell Biology **20**, 681 (2019).

[9] D. Sgarbossa, U. Lupo, and A.-F. Bitbol, Generative power of a protein language model trained on multiple sequence alignments, eLife **12**, e79854 (2023).

[10] C. Malbranke, D. Bikard, S. Cocco, R. Monasson, and J. Tubiana, Machine learning for evolutionary-based and physics-inspired protein design: Current and future synergies, Current Opinion in Structural Biology **80**, 102571 (2023).

[11] N. Tokuriki and D. S. Tawfik, Stability effects of mutations and protein evolvability, Curr. Opin. Struc. Biol. **19**, 596 (2009).

[12] N. Ota, R. Kurahashi, S. Sano, and K. Takano, The direction of protein evolution is destined by the stability, Biochimie **150**, 100 (2018).

[13] E. Mauri, S. Cocco, and R. Monasson, Mutational paths with sequence-based models of proteins: From sampling to mean-field characterization, Phys. Rev. Lett. **130**, 158402 (2023).

[14] J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, and T. Tlusty, Alphafold2 can predict single-mutation effects, Phys. Rev. Lett. **131**, 218401 (2023).

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

[16] 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, PLOS ONE **18**, 1 (2023).

[17] 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).

[18] 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).

[19] S. Iqbal, F. Ge, F. Li, T. Akutsu, Y. Zheng, R. B. Gasser, D.-J. Yu, G. I. Webb, and J. Song, Prost: Alphafold2-aware sequence-based predictor to estimate protein stability changes upon missense mutations, Journal of Chemical Information and Modeling **62**, 4270 (2022), pMID: 35973091.

[20] F. Peccati, S. Alunno-Rufini, and G. Jiménez-Osés, Accurate prediction of enzyme thermostabilization with rosetta using alphafold ensembles, Journal of Chemical Information and Modeling **63**, 898 (2023), pMID: 36647575.

[21] J. S. Xavier, T.-B. Nguyen, M. Karmarkar, S. Portelli, P. M. Rezende, J. P. L. Velloso, D. B. Ascher, and D. E. V. Pires, ThermoMutDB: a thermodynamic database for missense mutations, Nucleic Acids Research **49**, D475 (2020).

[22] J. M. McBride and T. Tlusty, The physical logic of protein machines, Journal of Statistical Mechanics: Theory and Experiment **2024**, 024001 (2024).

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

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

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

[26] S. Dutta, J.-P. Eckmann, A. Libchaber, and T. Tlusty, Green function of correlated genes in a minimal mechanical model of protein evolution, Proceedings of the National Academy of Sciences **115**, E4559 (2018).

[27] 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).

[28] K. Tsuboyama, J. Dauparas, J. Chen, E. Laine, Y. Mohseni Behbahani, J. J. Weinstein, N. M. Mangan, S. Ovchinnikov, and G. J. Rocklin, Mega-scale experimental analysis of protein folding stability in biology and design, Nature **620**, 434 (2023).

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

[30] J. Echave, S. J. Spielman, and C. O. Wilke, Causes of evolutionary rate variation among protein sites, Nature Reviews Genetics **17**, 109 (2016).

[31] S. J. Perkins, Protein volumes and hydration effects, European Journal of Biochemistry **157**, 169 (1986).

[32] Q.-Y. Tang and K. Kaneko, Long-range correlation in protein dynamics: Confirmation by structural data and normal mode analysis, PLOS Computational Biology **16**, 1 (2020).

[33] Q.-Y. Tang and K. Kaneko, Dynamics-evolution correspondence in protein structures, Phys. Rev. Lett. **127**, 098103 (2021).

[34] Y. Zhou, Q. Pan, D. E. V. Pires, C. M. Rodrigues, and D. Ascher, DDMut: predicting effects of mutations on protein stability using deep learning, Nucleic Acids Research **51**, W122 (2023).

[35] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, and L. Serrano, The FoldX web server: an online force field, Nucleic Acids Research **33**, W382 (2005).

[36] A. Broom, K. Trainor, Z. Jacobi, and E. M. Meiering, Computational modeling of protein stability: Quantitative analysis reveals solutions to pervasive problems, Structure **28**, 717 (2020).

[37] A. S. Dunham and P. Beltrao, Exploring amino acid functions in a deep mutational landscape, Molecular Systems Biology **17**, e10305 (2021).

[38] D. U. Ferreiro, E. A. Komives, and P. G. Wolynes, Frustration in biomolecules, Quarterly Reviews of Biophysics **47**, 285–363 (2014).

[39] J.-P. Eckmann and T. Tlusty, Dimensional reduction in complex living systems: Where, why, and how, BioEssays **43**, 2100062 (2021).

[40] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, and J. M. Jumper, Accurate structure prediction of biomolecular interactions with alphafold 3, Nature **630**, 493 (2024).

[41] H. Flechsig and A. S. Mikhailov, Simple mechanics of protein machines, Journal of the Royal Society Interface **16**, 20190244 (2019).

[42] J. M. McBride, Protein structure data for "ai-predicted protein deformation encodes energy landscape perturbation" (2024).

[43] H. Kim, M. Mirdita, and M. Steinegger, Foldcomp: a library and format for compressing and indexing large protein structure sets, Bioinformatics **39**, btad153 (2023), https://academic.oup.com/bioinformatics/article-pdf/39/4/btad153/49807920/btad153.pdf.

[44] 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).

[45] T. U. Consortium, Uniprot: A worldwide hub of protein knowledge, Nucleic Acids Res. **47**, D506 (2018).

[46] 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).

[47] J. Delgado, L. G. Radusky, D. Cianferoni, and L. Serrano, FoldX 5.0: working with RNA, small molecules and a new graphical interface, Bioinformatics **35**, 4168 (2019).

[48] L. D. Landau, E. M. Lifšic, E. M. Lifshitz, A. M. Kosevich, and L. P. Pitaevskii, *Theory of Elasticity: Volume 7*, Vol. 7 (Elsevier, 1986).

[49] J. S. Merkel, J. M. Sturtevant, and L. Regan, Sidechain

interactions in parallel beta sheets: the energetics of cross-strand pairings, Structure **7**, 1333 (1999).

[50] D. G. Isom, B. R. Cannon, C. A. Castañeda, A. Robinson, and B. G.-M. E., High tolerance for ionizable residues in the hydrophobic interior of proteins, Proceedings of the National Academy of Sciences **105**, 17784 (2008).

[51] D. G. Isom, C. A. Castañeda, B. R. Cannon, P. D. Velu, and B. G.-M. E., Charges in the hydrophobic interior of proteins, Proceedings of the National Academy of Sciences **107**, 16096 (2010).

[52] A. C. Robinson, J. L. Schlessman, and B. García-Moreno E, Dielectric properties of a protein probed by reversal of a buried ion pair, The Journal of Physical Chemistry B **122**, 2516 (2018), pMID: 29466010.

[53] D. Shortle, W. E. Stites, and A. K. Meeker, Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease, Biochemistry **29**, 8033 (1990), pMID: 2261461.

[54] J. H. Carra and P. L. Privalov, Energetics of denaturation and m values of staphylococcal nuclease mutants, Biochemistry **34**, 2034 (1995), pMID: 7849061.