AlphaFold-predicted deformation probes changes in protein stability

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*

AlphaFold2 (AF) is an excellent structure predictor, but it does not unambiguously indicate whether the predicted structure is stable. Could it be that AF contains sufficient information to predict protein stability, and one just needs a way to extract it? To address this question, we investigate single-mutation effects on protein stability using a large empirical dataset of changes in free energy of folding $\Delta\Delta G$, and calculate the deformation, measured by strain, between AF-predicted mutant and wild-type protein structures. Our initial motivation was to assess if anything could be inferred about protein stability from AF-predicted structures, but our results showed unexpectedly that 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. In particular, high strain in buried residues generally gives rise to large changes in stability. Thus, our findings indicate that AF-predicted structures encode significant information on stability, suggesting that (de-)stabilizing effects of mutations may be estimated using AF. This paves the way for the development of new structure-based algorithms that accurately predict and explain how mutations affect stability.

AlphaFold (AF) has ushered in a revolution in structural biology, yet we are still in unchartered waters.[1–3] Of particular importance for protein design and evolution is the ability to predict whether a given protein sequence will lead to a stably-folded structure.[4–6] On the one hand, AF appears to predict structures with high pLDDT – AF's measure of confidence in atomic positions – that do not fold *in vitro* or are susceptible to protease degradation,[7,8] and another study found that pLDDT changes do not correlate with stability changes stability.[9] On the other hand, some analyses suggest that AF and RoseTTTaFold can be used for predicting stability changes upon mutation.[7,10,11] This motivates us to systematically investigate the question as to whether AF-predicted structures are informative of stability changes.

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).[12] We find that a simple general measure of deformation upon mutation – effective strain – 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. 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 indeed 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.

RESULTS

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, see Methods). 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, see Methods) 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.[7] 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, orange), 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). To achieve a more accurate estimate of deformation *due to mutation*, we follow the method of McBride *et al.*[7] to calculate deformation using 'averaged' AF structures (see Methods). This drastically reduces ES in most regions (which originates chiefly from noise and fluctuations), except for regions near the mutated site (Fig. 1D-E, blue).

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 the 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,

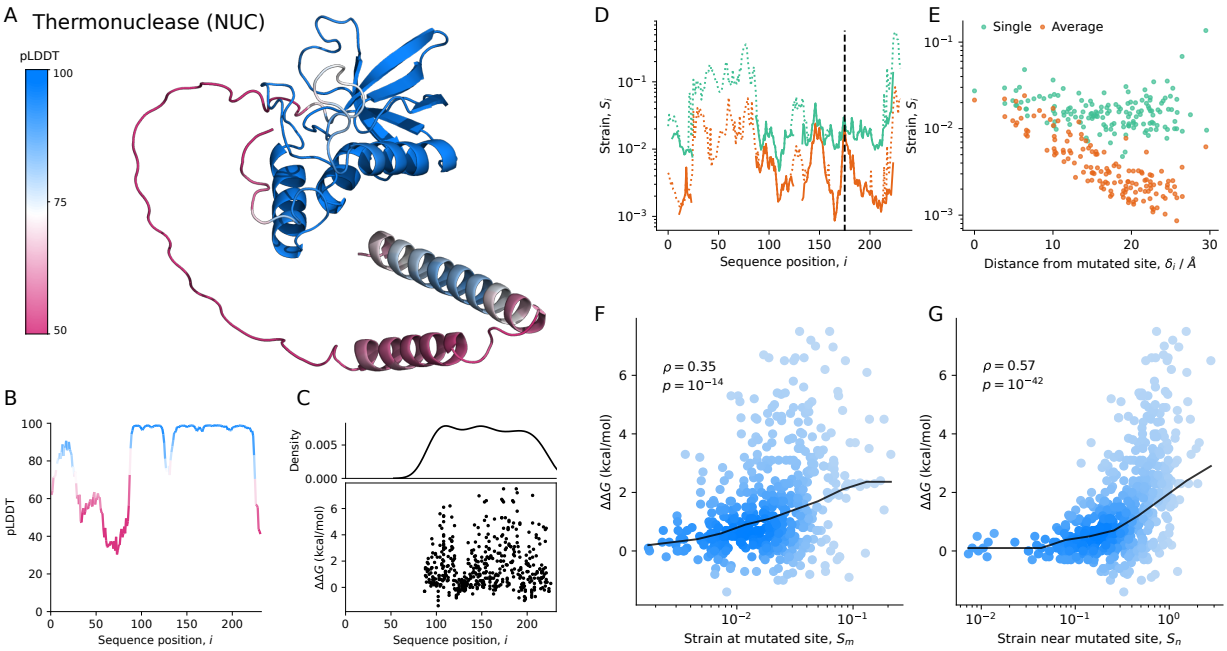

FIG. 1. **Strain calculated using AF-predicted structures correlates with $\Delta\Delta G$.** A: AF-predicted structure of thermonuclease from *staphylococcus aureus* (NUC) – the most common protein in TMDB. Residues are colored according to pLDDT (AF-predicted confidence score). B: pLDDT per residue. C: Distribution of 491 mutation sites along the sequence, and corresponding changes in stability $\Delta\Delta G$. D: Strain upon mutation (A176G, $\Delta\Delta G = 2.4\,\mathrm{kcal/mol}$) 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 and p values are shown; circles are shaded by density.

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 15 Å of each residue i. For NUC, we find that mutations in buried regions tend to have an outsize impact on stability (Fig. 2C), as expected, given the standard paradigm of buried residues having low mutation rates.[13] 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). This suggests that S_n correlates with $\Delta\Delta G$ because summing S in a neighborhood of 15 Å most probably includes the contributions of one or a few of these buried residues.

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; many regions have no mutations sampled from them, so deformation remains low no matter how buried the residues are. Further 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.

How many residues are affected by mutations in a way that impacts stability?— When we compare pairs of single AF-predicted structures, which exhibit significant noise unrelated to mutations (SM Fig. 1), we find maximal strain-stability correlations when counting neighbors within $10 \leq \gamma \leq 20$ Å (Fig. 2E). For residues further than 20 Å, we can consider strain to have little information pertaining $\Delta\Delta G$, and their inclusion decreases the correlation due to noise. Likewise, when we compare pairs of averaged structures, we find that correlations increase up to about $10 \leq \gamma \leq 20$ Å and then plateau (Fig. 2E). The correlations of the averaged pairs do not decrease significantly with increasing γ since averaging filters out much of the noise. This demonstrates that even after increasing the precision of the prediction of mutation effects, there tends to be little additional information beyond 20 Å; although this does not preclude the possibility of long-ranged mutation effects, as in allosteric communication. The reason why correlations continue to increase with γ up to $\sim$20 Å is perhaps due to the typical size of folded domains; we find that across

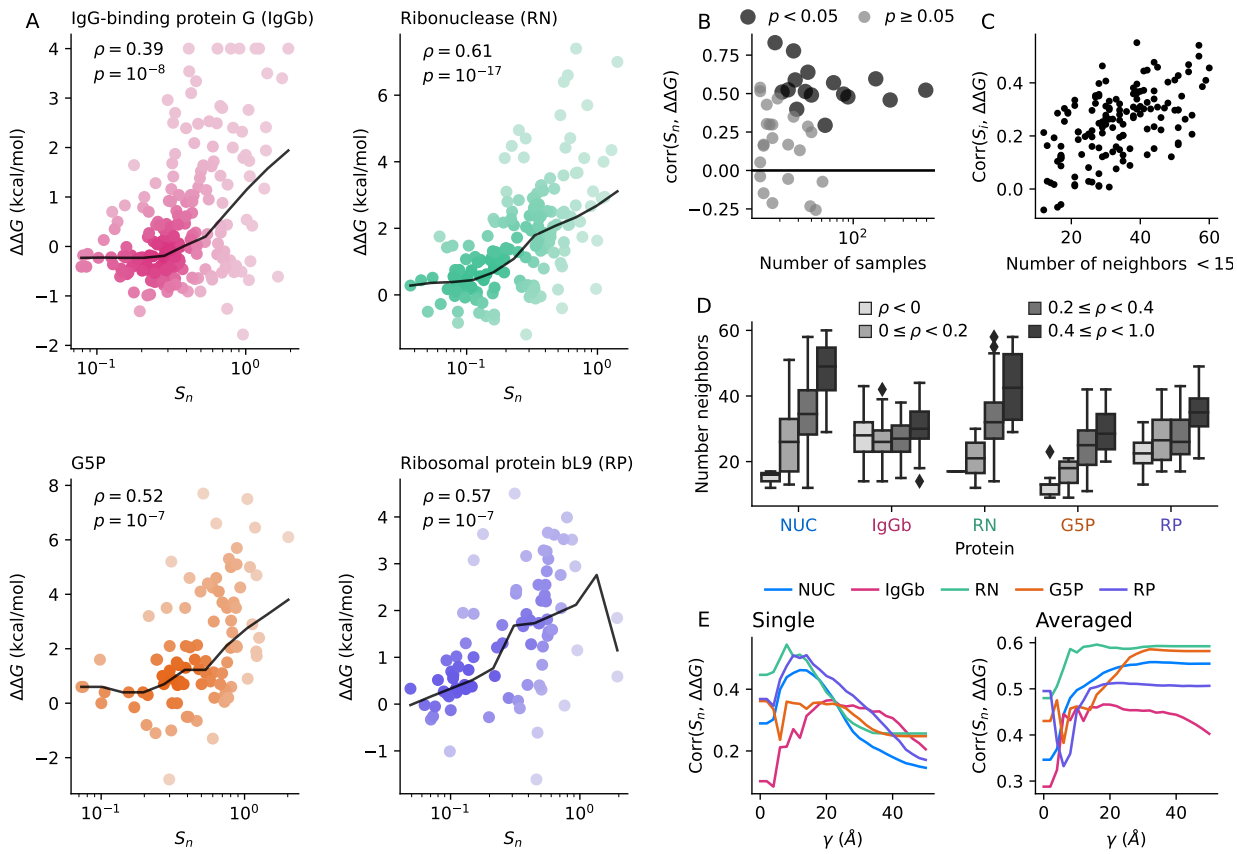

FIG. 2. **When and why does S_n correlate with $\Delta\Delta G$?** A: Change in stability $\Delta\Delta G$ against strain near mutated residue S_n for four proteins (the second to fifth most common in TMDB; the most common one, NUC, is shown in Fig. 1); Spearman's ρ and p values are shown; black line is the median; circles are shaded by density. B: S_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–$\Delta\Delta G$ 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–$\Delta\Delta G$ correlation, for the five most common proteins. E: S_n–$\Delta\Delta G$ 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.

all proteins studied here, the most buried (and hence key to stability) residue is on average 15 Å from any residue (SM Fig. 4). Thus, we need to consider up to 15 Å to find the most informative residues on average, and beyond this distance, the mutation effects are too weak to add extra information; on the contrary, they mainly add noise.

Strain correlates with $\Delta\Delta G$ across protein families.— We find a moderate correlation ($\rho = 0.32$) 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 strain is invariant to reversing the reference and target structures. Indeed, we find a somewhat higher correlation ($\rho = 0.40$) 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

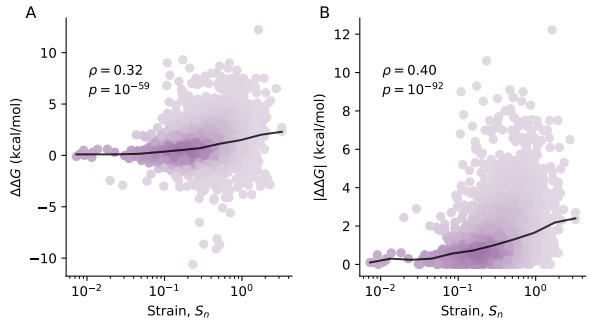

FIG. 3. **S_n–$\Delta\Delta G$ correlations across proteins.** Strain near mutated site S_n vs. $\Delta\Delta G$ (A) and magnitude of stability change $|\Delta\Delta G|$ (B) for all mutations in our reduced TMDB sample of 2,499 unique mutants. Spearman's ρ and p value are shown.

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 ΔΔG almost as well as tailored ΔΔG predictors.— To put the strain-stability correlations in context, we compare them with two state-of-the-art ΔΔG predictors, DDMut and FoldX (see Methods).[14,15] 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_{n}–$|\Delta\Delta G|$ correlations are almost as high as correlations obtained using FoldX on average, and not far behind 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. 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 ($\rho = 0.2$). 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. 5). 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. 6), 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.

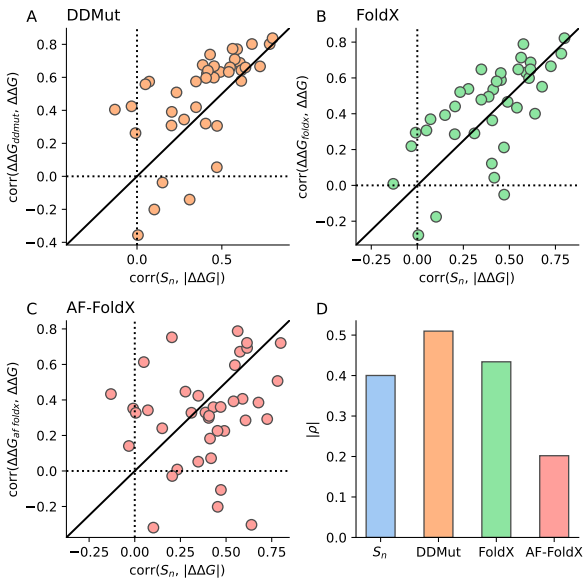

FIG. 4. **Comparison with state-of-the-art algorithms.** A-C: Correlation (Spearman's ρ) between strain S_{n} and magnitude of stability change $|\Delta\Delta G|$, vs. correlation between $\Delta\Delta G$ and predictions of DDMut (A), FoldX (B), and FoldX using AF-predicted structures (C); separate points are shown for each of the 40 most-common proteins. D: Correlations (as above) for the full sample of 2,499 unique mutants.

DISCUSSION

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.[6,14] 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, these structures could be used to reparameterize existing force fields such as FoldX. The recent explosion of high-throughput measurements[16,17] 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.

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. 7).[7] 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. 8).

CONCLUSION

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.

METHODS

Stability Data— We select all proteins from ThermoMutDB (TMDB)[12] with sequence length, $50 \leq L \leq 500$, and measurements for single mutants made within $293 \leq T \leq 313\,\mathrm{K}$ and $5 \leq \mathrm{pH} \leq 8$; 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),[18] Uniprot,[19] 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. We average $\Delta\Delta G$ when multiple measurements are available to get a representative value for each mutation. 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.

Structure Analysis— We use the ColabFold implementation of AF to predict protein structures.[20] 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 McBride *et al.*[7]. We calculate effective strain, ES,[7] 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 following McBride *et al.*[7], as

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

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.

$\Delta\Delta G$ predictors— We use two state-of-the-art methods to predict $\Delta\Delta G$: FoldX 5,[21] and DDMut.[14] 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.

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

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

[1] 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).

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

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

[4] 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.

[5] T. Sanavia, G. Birolo, L. Montanucci, P. Turina, E. Capriotti, and P. Fariselli, Limitations and challenges in protein stability prediction upon genome variations: towards future applications in precision medicine, Computational and Structural Biotechnology Journal **18**, 1968 (2020).

[6] 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).

[7] J. M. McBride, K. Polev, A. Abdirasulov, V. Reinharz, B. A. Grzybowski, and T. Tlusty, Alphafold2 can predict single-mutation effects, Physics Review Letters (in press).

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

[9] 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).

[10] 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).

[11] 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).

[12] 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).

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

[14] 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).

[15] 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).

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

[17] 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).

[18] 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).

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

[20] 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).

[21] 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).

Supplemental Material for
"AlphaFold-predicted deformation from wild-type protein correlates with changes in stability"

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*

Contents

1. PDB Structure Data 1
   A. How many structures are needed to average? 1
   B. pLDDT is less informative of stability than strain. 1
   C. Insufficient PDB data to measure empirical strain-stability correlation. 1
   D. Construction of a reduced ThermoMutDB set 1

1. PDB Structure Data

A. How many structures are needed to average?

Since we found that correlations between S_n and $\Delta\Delta G$ are higher when using averaged AF-predicted structures, a pertinent practical question is, how many structures should one use? We looked at the correlations (both $\Delta\Delta G$ and $|\Delta\Delta G|$) as a function of the number of structures used to create average structures, finding that most of the improvement is seen when using only five structures instead of one (Fig. 8). We suggest that using 20 structures is sufficient to achieve good performance.

B. pLDDT is less informative of stability than strain.

Since AF's predicted confidence measure, pLDDT, is a good indicator of disorder, we also investigated whether changes in pLDDT are informative of changes in stability. We compared changes in pLDDT at the mutated site m, changes in the averaage pLDDT (across the entire protein), for both single and averaged structures, against both $\Delta\Delta G$ and $|\Delta\Delta G|$. We find poor correlations between changes in pLDDT and $\Delta\Delta G$ (Fig. 7), in agreement with Pak *et al.* (2023).

C. Insufficient PDB data to measure empirical strain-stability correlation.

We sought to repeat our analysis using experimental structures from the Protein Data Bank (PDB). We only found ~ 100 entries which have both wild-type (WT) and mutant experimental structures. We removed any NMR structures, and only considered pairs with the same oligomeric state and same bound substrates. This left us with 36 pairs of PDB structures, which is too few to study statistical correlations. It is possible that there are sufficient $\Delta\Delta G$ measurements across other data sets, but the limit of unique mutations is determined by the number of possible structure pairs that differ by a single amini acid substitution. In a previous study (McBride *et al.*, in press) we only found $\sim$1,000 of these, so there is some hope that this analysis can be performed if the other datasets have enough $\Delta\Delta G$ measurements.

D. Construction of a reduced ThermoMutDB set

Our original approach was to use the ThermoMutDB subset provided by Pak *et al.* (2023), however we decided to work exclusively with the original ThermoMutDB data, for two of reasons. We found that the "TMDB_ID" values do not always match with the original ThermoMutDB ID numbers. We also found a case where a mutant (P00644, L89V) was assigned an incorrect mutation position; in this case, both the PDB SEQRES and Uniprot sequences have a lysine at position 89, but the correct mutation code should be L171V.

We found several issues in the ThermoMutDB that led to exclusion or correction of entries. We found that for some entries, the reference protein for which the mutation was defined (and $\Delta\Delta G$ was calculated) was not the WT sequence found in Uniprot; (Consortium, 2018; Isom *et al.*, 2008, 2010; Merkel *et al.*, 1999) we excluded these. We found duplications of entries, where measurements from one paper were referenced in several other papers, yet both were included; (Robinson *et al.*, 2018) we excluded these entries. We found that one measurement (Y27A) in one source, (Shortle *et al.*, 1990) and all measurements in another source (Carra and Privalov, 1995) were included with the incorrect sign; we corrected these.

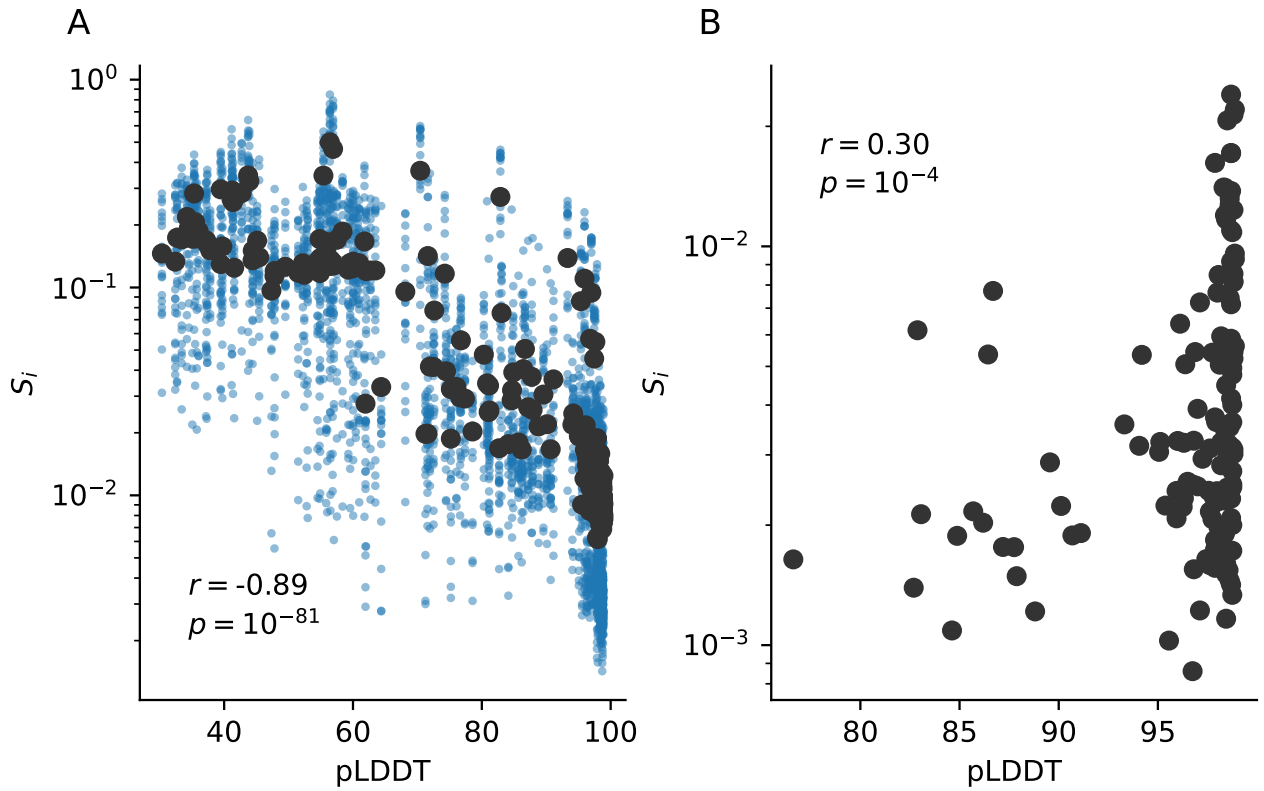

FIG. 1 **AF predictions are more variable in regions with low pLDDT.** Strain per residue S_i vs. pLDDT per residue for: (A) Strain calculated between repeat AF predictions of the same protein sequence; blue dots are shown for different pairs of single structures; black circles indicate means across multiple pairs. (B) Strain calculated for pairs of averaged structures, whose sequences differ by a single mutation; residues with pLDDT < 70 are excluded. Pearson's correlation coefficient and p value are shown.

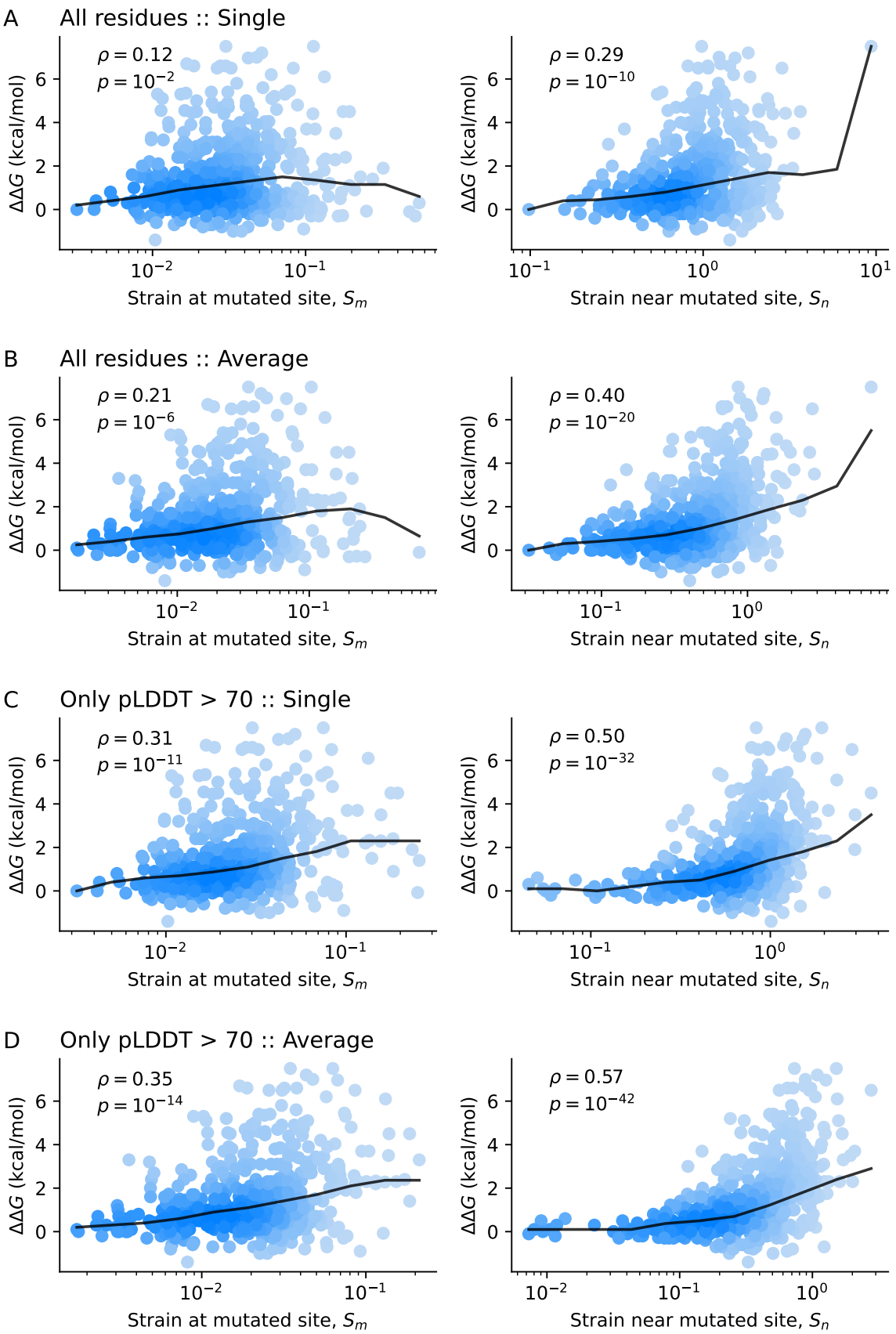

FIG. 2 **Improving precision of mutation effect prediction.** Strain at the mutated site m, S_m (left), and near the mutated site, S_n (right), for four different methods of calculating strain: (A) All residues and pairs of single structures. (B) All residues and pairs of averaged structures. (C) Only residues with pLDDT > 70 and pairs of single structures. (D) Only residues with pLDDT > 70 and pairs of averaged structures. Spearman's ρ and p value are shown.

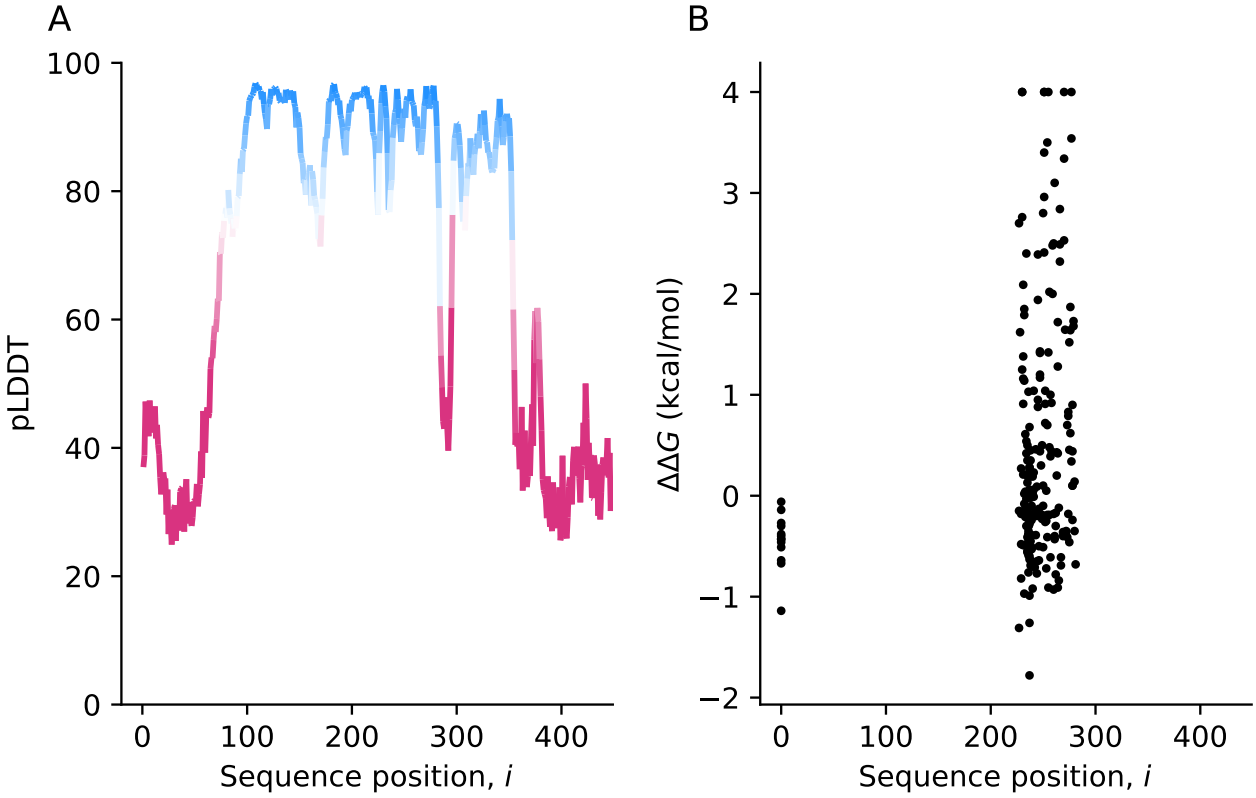

FIG. 3 **Biased sampling of IgG-binding protein G (IgGb).** A: pLDDT vs sequence position i for IgGb. B: Distribution of mutations in the ThermoMutDB for IgGb along the sequence, and the corresponding $\Delta\Delta G$ values.

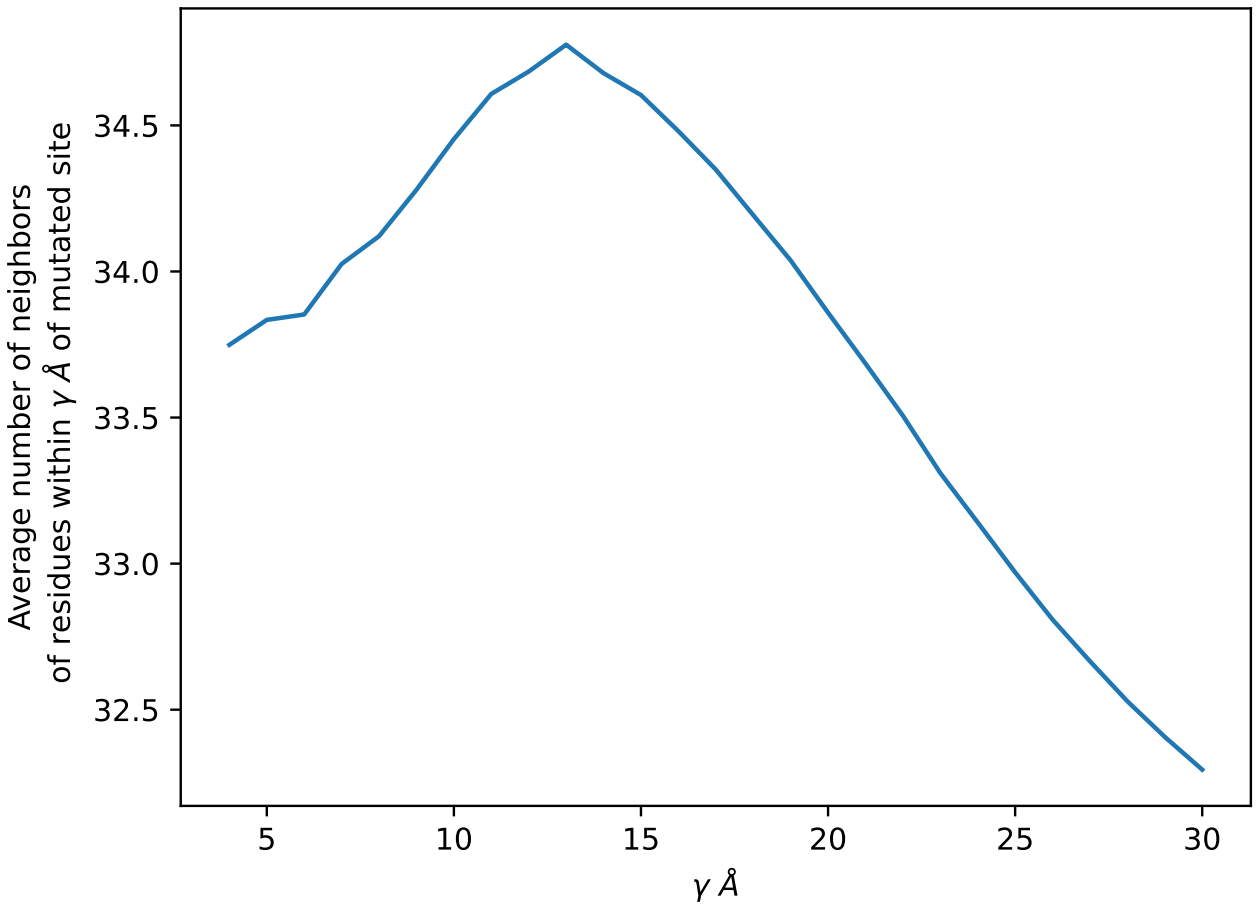

FIG. 4 **Buried residues tend to be within 15 Å of other resiudes.** For each residue i, its neighbors are defined as the set of residues whose C_α distances are within 13 Å. Plot shows average number of neighbors of residues within γ of the mutated site m, as a function of γ. Average is calculated for all unique proteins in our reduced ThermoMutDB sample.

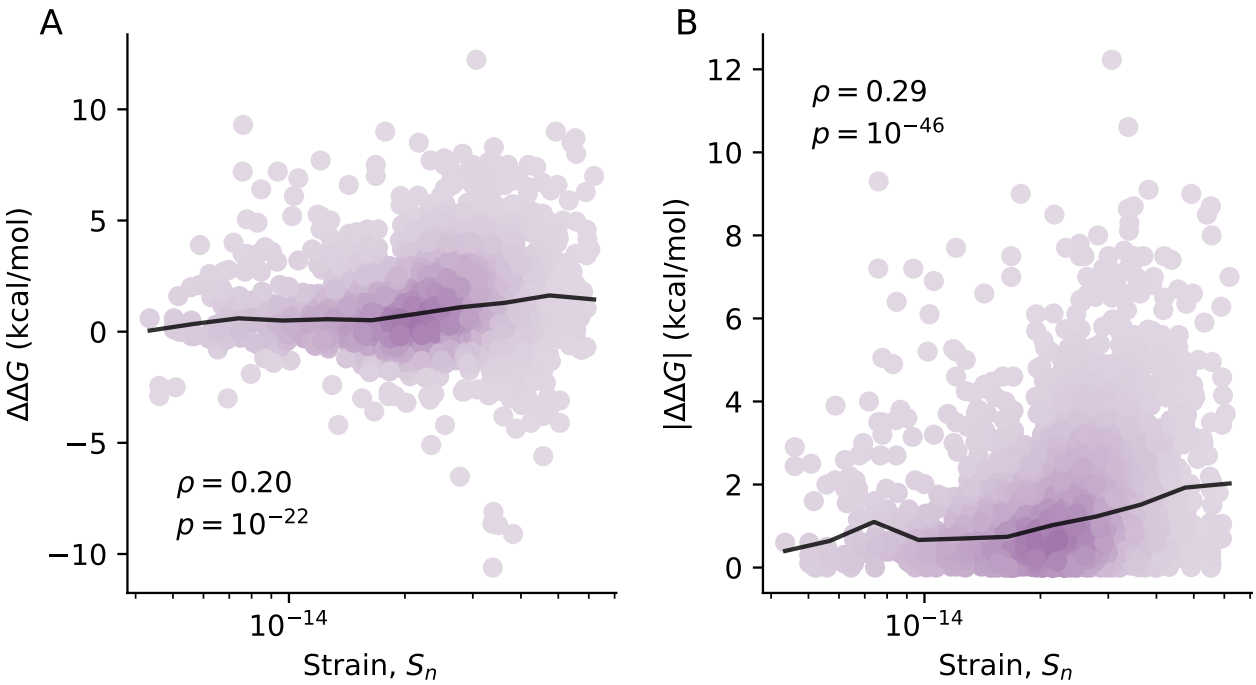

FIG. 5 S_n–$\Delta\Delta G$ **correlations calculated using FoldX structures.** Strain near mutated site S_n vs. $\Delta\Delta G$ (A) and magnitude of stability change $|\Delta\Delta G|$ for all mutations in our reduced ThermoMutDB sample of 2,499 unique mutants. Spearman's ρ and p value are shown.

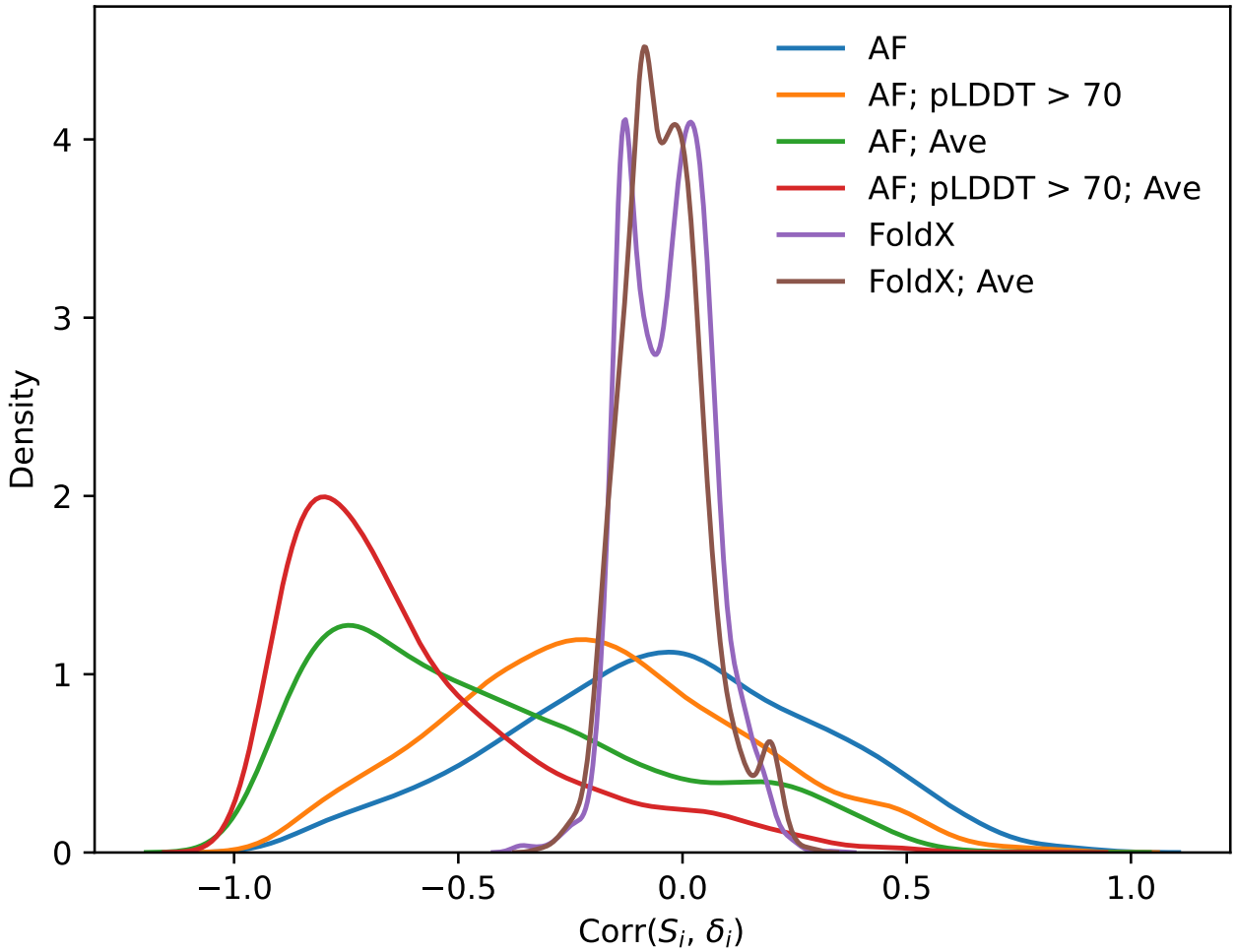

FIG. 6 **Correlation between distance from mutated site δ_i and S_i** for all proteins in our reduced ThermoMutDB sample. Different distributions are shown for shear calculated using: AF-predicted structures, all residues and pairs of single structures (AF). AF-predicted structures, only residues with pLDDT > 70 and pairs of single structures (AF; pLDDT > 70). AF-predicted structures, all residues and pairs of averaged structures (AF; Ave). AF-predicted structures, only residues with pLDDT > 70 and pairs of averaged structures (AF; pLDDT > 70; Ave). FoldX-predicted structures, all residues and pairs of single structures (FoldX). FoldX-predicted structures, all residues and pairs of averaged structures (FoldX; Ave).

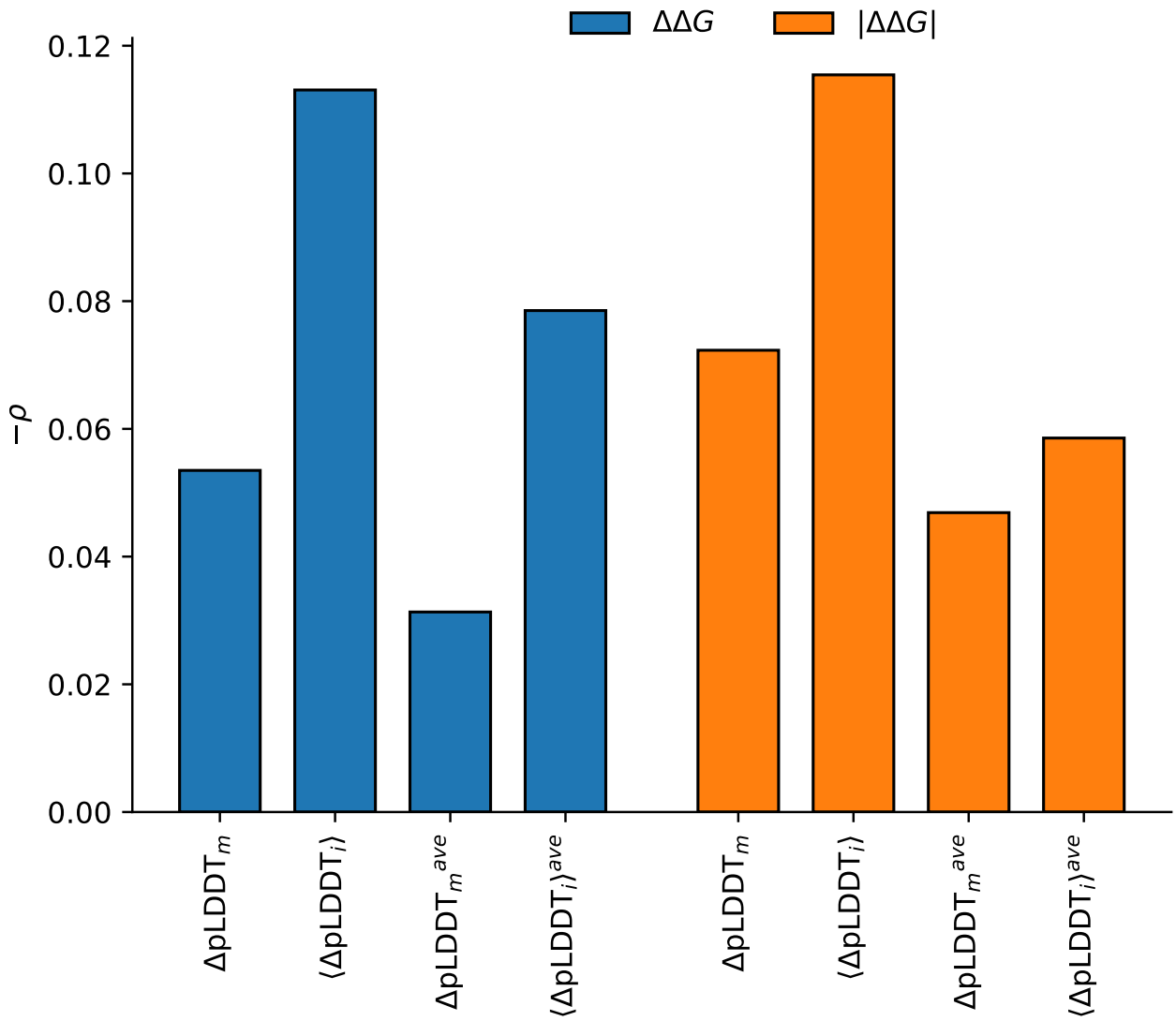

FIG. 7 **ΔpLDDT is a poor predictor of ΔΔ*G*.** Spearman's correlation between differences in pLDDT and stability. Differences in pLDDT are measured in four ways: difference in pLDDT at the mutated site ΔpLDDT_m; average difference in pLDDT across the whole protein $\langle\Delta\text{pLDDT}_i\rangle$; difference in pLDDT at the mutated site, averaged over 50 pLDDT predictions $\Delta\text{pLDDT}_m^{ave}$; average difference in pLDDT across the whole protein, averaged over 50 pLDDT predictions $\langle\Delta\text{pLDDT}_i\rangle^{ave}$;

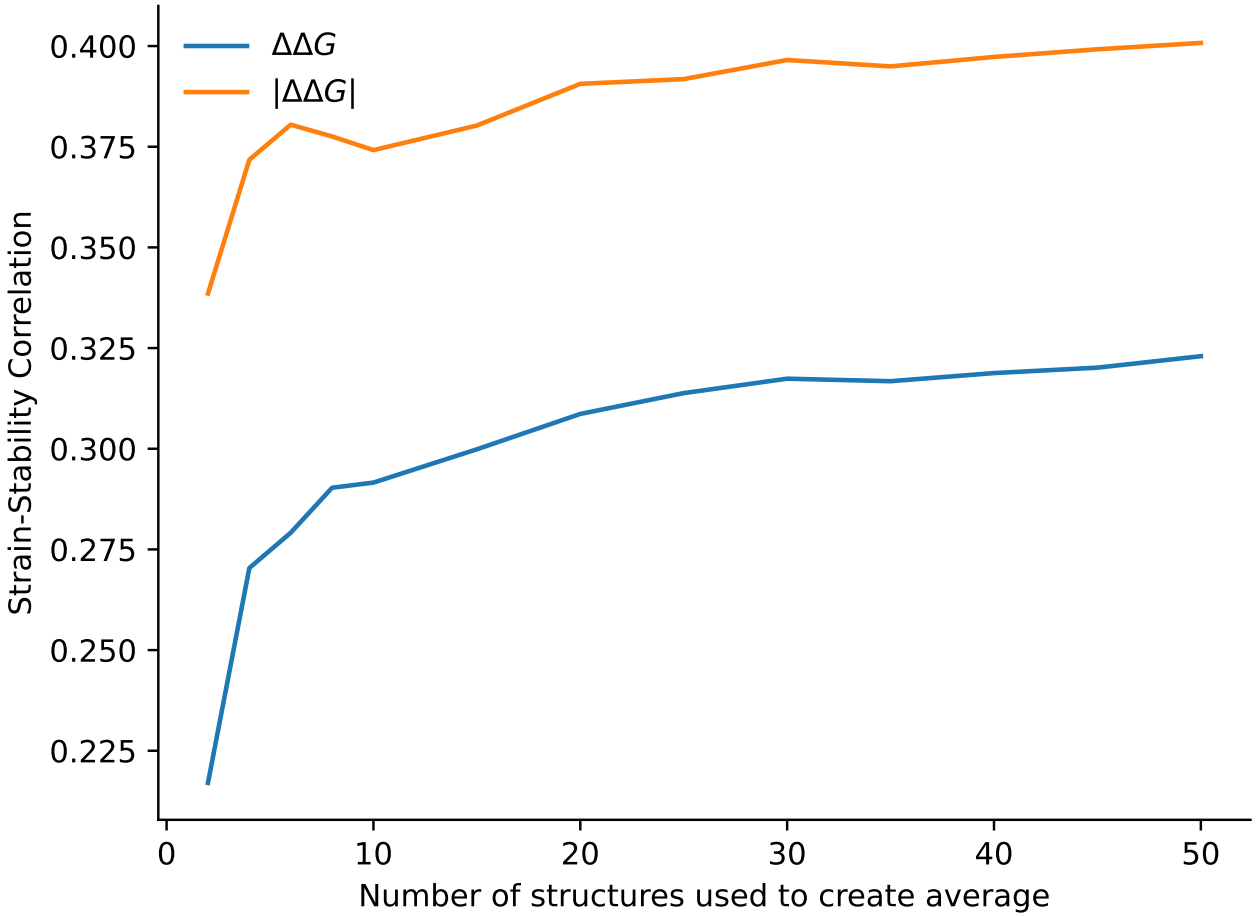

FIG. 8 **Effect of the number of structures used to create averages.** Strain-stability correlation for all mutations in our reduced ThermoMutDB sample of 2,499 unique mutants, as a function of the number of structures used to create average structures.

References

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

Carra, John H, and Peter L. Privalov (1995), "Energetics of denaturation and m values of staphylococcal nuclease mutants," Biochemistry **34** (6), 2034–2041, pMID: 7849061, https://doi.org/10.1021/bi00006a025.

Consortium, The UniProt (2018), "Uniprot: A worldwide hub of protein knowledge," Nucleic Acids Res. **47** (D1), D506–D515.

Isom, Daniel G, Brian R. Cannon, Carlos A. Castañeda, Aaron Robinson, and Bertrand García-Moreno E. (2008), "High tolerance for ionizable residues in the hydrophobic interior of proteins," Proceedings of the National Academy of Sciences **105** (46), 17784–17788, https://www.pnas.org/doi/pdf/10.1073/pnas.0805113105.

Isom, Daniel G, Carlos A. Castañeda, Brian R. Cannon, Priya D. Velu, and Bertrand García-Moreno E. (2010), "Charges in the hydrophobic interior of proteins," Proceedings of the National Academy of Sciences **107** (37), 16096–16100, https://www.pnas.org/doi/pdf/10.1073/pnas.1004213107.

McBride, John M, Konstantin Polev, Amirbek Abdirasulov, Vladimir Reinharz, Bartosz A. Grzybowski, and Tsvi Tlusty (in press), "Alphafold2 can predict single-mutation effects," Physics Review Letters.

Merkel, Jane S, Julian M. Sturtevant, and Lynne Regan (1999), "Sidechain interactions in parallel β sheets: the energetics of cross-strand pairings," Structure **7** (11), 1333–1343.

Pak, Marina A, Karina A. Markhieva, Mariia S. Novikova, Dmitry S. Petrov, Ilya S. Vorobyev, Ekaterina S. Maksimova, Fyodor A. Kondrashov, and Dmitry N. Ivankov (2023), "Using alphafold to predict the impact of single mutations on protein stability and function," PLOS ONE **18** (3), 1–9.

Robinson, Aaron C, Jamie L. Schlessman, and Bertrand García-Moreno E (2018), "Dielectric properties of a protein probed by reversal of a buried ion pair," The Journal of Physical Chemistry B **122** (9), 2516–2524, pMID: 29466010, https://doi.org/10.1021/acs.jpcb.7b12121.

Shortle, David, Wesley E. Stites, and Alan K. Meeker (1990), "Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease," Biochemistry **29** (35), 8033–8041, pMID: 2261461, https://doi.org/10.1021/bi00487a007.