Deriving structural explanations for protein stability through temperature sensitive mutants

Charles Harris

June 2020

**Abstract**

Protein stability engineering will prove invaluable in emerging fields in biotechnology and biopharmaceuticals. Additionally, many disease causing mutations are as a result of protein instability. Despite this, we have been able to determine few universal rules to explain protein stability, rather, we only have a long list of highly interdependent factors. Here we use commonly available algorithms for predicting changes in protein stability (FoldX) and changes in structural features (Missense3D), as well as experimental data, to try and provide structural explanations for which missense mutations cause protein thermal instability. Our results confirm the understanding that protein folding is largely governed by thermodynamic We also evaluate the accuracy of these programs and discuss what is needed for the advancement of the field.

**Introduction**

Understanding the structural features that give proteins their stability is important for the development and optimisation of new biotechnological applications \citep{liu2019biotech} and biopharmaceuticals \citep{frokjaer2005proteintherapeutics, demarest2008antibody, sun2016vaccine, madan2018vaccine2, rickerby2020labgen}. Undoubtedly, it will also help expand our comprehension of the mechanisms underpinning protein folding and protein interactions, the failure of which are responsible for a great number of diseases \citep{casadio2011disease+stability1,yates2014disease+stability3,peng2016disease+stability2}. Having these well understood explanations will also be essential if we are to move towards \textit{de novo} protein design \citep{huang2016denovo, baker2019denovo2}.

A new generation of recombinantly expressed, engineered protein therapeutics \citep{frokjaer2005proteintherapeutics} and monoclonal antibodies \citep{demarest2008antibody} have highlighted the issues around long-term storage and efficient delivery. Lapses in safely standards can not only reduce efficacy, but also cause unforeseen immunogenic side effects \citep{krishna2016immunogenicity}. Additionally, most current vaccines are very thermally sensitive and need to be stored in a continuous "cold chain", presenting significant logistical challenges \citep{sun2016vaccine}, especially in the developing world \citep{madan2018vaccine2}. Whilst techniques have been developed to monitor the structural integrity of such therapeutics, both before and after biomanufacturing \citep{bhirde2018therapeuticscreening}, an all together better approach, both clinically and economically, is to engineer future therapeutics to be more thermally resistant.

Most \textit{de novo} protein design strategies so far start with idealised structures with very high stability but no or little function. However, the addition of functional and binding sites to these structures usually is detrimental to stability, meaning a more nuanced understanding of protein stability will be required if we are to be able too add residues that are essential to function but do not lend themselves to stability (e.g. addition of hydrophobic residues on protein binding surfaces).

???Add paragraph on how protein stability relates to disease???

Considerable effort has been put towards the furthering of our understanding of these effects \citep{pucci2017proteinstability, feller2013psychrophilic, razvi2006lessonsstability}, and much has been learnt from the study of thermophilic proteins. Work has also been done into how variation at the genome level impact protein structure and function \citep{bhattacharya2017impactofgeneticvariation}. Unfortunately, there seems to be no definitive and universal rules but rather a complex, interconnected network of factors, the high dimensionality of which makes it difficult to paint any clear picture. Furthermore, any rigid rules observed in one protein family rarely apply to other areas.

In this report we use experimental and computationally predicted data to try and elucidate the importance certain structural features have with regards to protein thermal stability. We use the commonly available computational tools Missense3D \citep{ittisoponpisan2019Missense3D} and FoldX \citep{schymkowitz2005foldx}. While Missense3D is designed as a predictor of disease causing variants, we primarily use it for its ability to predict structural changes that occur upon a missense mutation. We also address some of the issues faced in the field with regards to data bias and abundance as well as briefly discuss new areas that might provide more insights.

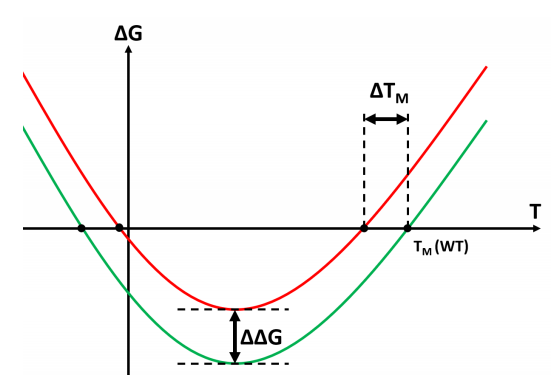
**Theoretical review of protein stability**

The thermodynamic stability of the protein folding process is easily characterised by the changes in Gibbs free energy (\textDelta G). The exact value of \textDelta G is the sum of all the entropic (e.g. hydrophobic effects) and enthalpic (e.g hydrogen bonds and van der Waals interactions) terms. Assuming protein folding to be a perfect two-state transition, the temperature dependent change in free energy upon folding can be written as:

\[\Delta G(T) = G\_{(Folded)}(T) - G\_{(Unfolded)}(T) \]

Thus, any missense mutation that causes a change in the thermodynamic stability of a protein can be described by its change on the free energy difference (Figure \ref{fig:theory}) such as:

\[ \Delta\Delta G(T) = \Delta G\_{(Mutant)}(T) - \Delta G\_{(Wild\:type)}(T) \]

Protein stability is also described sometimes by its melting temperature ($T\_M$), the point at which \textDelta G equals zero. Accordingly, changes in protein stability can be described in terms of its effect on melting temperature (\textDelta$T\_M$). The complex relationship between folding \textDelta G and other thermodynamic descriptors has been covered elsewhere in the literature \citep{becktel1987proteinstabilitycurves, prabhu2005heatcapacity, miyazawa2017foldability}, but these go beyond the level of detail needed for this investigation.

Charlieharris001@gmail.com

Figure 1: Green line: thermodynamic properties of a hypothetical wild type protein. Red line: thermodynamic properties of a simple hypothetical destabilising mutant of the original wild type. Note the negative \textDelta$T\_M$ and positive \textDelta G in this destabilising example.}

In reality, proteins do not fold by transitioning between two states, rather partially organised structures gradually organise into the native state by moving though a rugged "funnel-like" energy landscape \citep{onuchic2004landscape1}. Sometimes proteins incorrectly fold when they are forced into an undesirable local minima, the probability of which increases with topological complexity that is stabilised by long-range interactions and domains that are separate in the native structure interacting during folding \citep{hartl2009landscape2}. While we do not address this point here, this means the stability of natural and engineered proteins should not only be judged by their folding \textDelta G but also the propensity to adapt the native state. Thus, strategies known as "negative-design principles" have been developed that aim to destabilise only the misfolded states to encourage the native state to form \citep{goldenzweig2018dynamics2, schafer2014negativedesign}.

???Mention other external factors affecting stability such as pH???

**Methods**

The computational workflow used in this project is outlined in Figure \ref{fig:workflow}. We use two experimentally determined datasets (to study data bias between the two) and two programs, FoldX \citep{delgado2019foldx5} and Missense3D \citep{ittisoponpisan2019Missense3D}. Note that when we refer to the "HoTMuSiC" \citep{pucci2016HoTMuSiC} (n=1,626) and "Missense3D" (n=10,229) datasets, we are referring to the experimentally determined datasets used to train those respective algorithms and not any predictions made by those algorithms.

FoldX treats protein stability as a regression problem and it's PositionScan function can predict folding \textDelta\textDelta G. The algorithm uses statistically derived energy potential functions trained on another subsection of the ProTherm database \citep{schymkowitz2005foldx}. Missense3D produces a binary prediction on weather the variant is \textit{disease casuing} and not thermodynamically unstable. This binary prediction does not feature greatly in our results but we do use extensively Missense3D's ability to predict 16 types of structural changes upon point mutation (referred to as "features" in the rest of the text). The exact list of features and how they are defined are explained in \cite{ittisoponpisan2019Missense3D}. Both algorithms assume rigid mainchains but repack the sidechains. Missense3D uses the SCWRL4 \citep{krivov2009SCWRL4} algorithum for sidechain repacking \citep{ittisoponpisan2019Missense3D} whereas FoldX performs this natively \citep{schymkowitz2005foldx}.

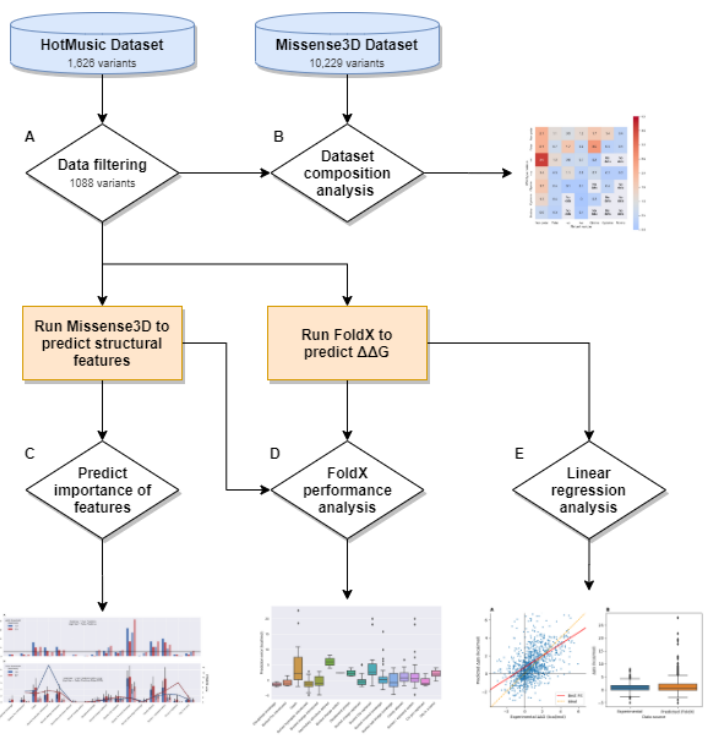


Figure 2: Flowchart demonstrating the workflow used. Note we only use the Missense3D dataset to study bias in our dataset and not to make predictions.

**Construction and analysis of dataset**

The HoTMuSiC dataset, more specifically T1626 \citep{pucci2016HoTMuSiCdataset}, is mostly a subsection of the widely used ProTherm database \citep{kumar2006protherm} but also contains a number of variants screened from the literature. All variants were manually checked in the original literature to remove errors and a criteria was used to ensure only high quality structures and variants were used. A brief summary of the criteria used is listed below, a full explanation is provided in \cite{pucci2016HoTMuSiCdataset}.

* Only mutations mapped onto x-ray structures with a resolution of 2.5 Å or greater were kept.
* Only mutations for monomeric proteins were kept. This was important to our study as it ensures that any changes in stability observed was a result of changing folding dynamics and not of quaternary structure.
* Only proteins described in the literature as undergoing a two-state folding transition were kept.

For the purposes of our study we performed some additional filtering of the dataset (Figure \ref{fig:workflow}A):

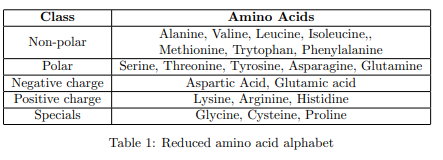
* Any mutation not containing an experimentally measured \textDelta\textDelta G value was removed (HoTMuSiC was trained only to predict $T\_m$).
* Double mutants were removed. This included single mutants of already once mutated structures as the single-mutated structures could not be sourced in a high-throughput manner.
* Variants with unconventional nomenclature were removed as these often could be be interpreted by the programs.
* One variant was removed due to a mismatch between the wild-type residue listed in the dataset and the wild-type structure.

Following this, we were left with 1088 variants.

---

Protein mutation databases, such as HoTMuSiC, are compiled from a diverse range of protein stability and mutagenisis studies and have a bias towards the mutations that were of interest for those particular studies. Therefore, the frequency at which certain mutations happen in our data set is not necessarily representative of the frequency to which they are found in nature. In order to demonstrate this we compared the frequency distribution of variants in our dataset to the Missense3D dataset which, being made of naturally occurring disease-causing and neutral variants \citep{ittisoponpisan2019Missense3D} can be thought of as being a relatively accurate representation of the frequency at which different missense variants are observed in nature. For comparison, we calculated the ratio between the percentage abundances for every type of missense mutation in both datasets (Figure \ref{fig:workflow}B). Therefore, a variant type with a ratio $>$1 is considered overepresented and $<$1 is considered underrepresented.

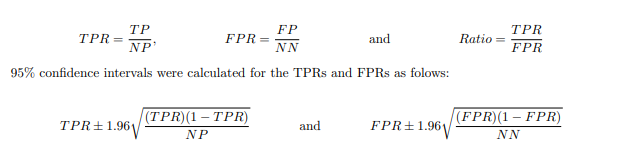
The distribution of mutations were described using a reduced amino acid alphabet (Table \ref{table:alphabet}) for easier visualisation. Additionally, we would predict that the most deleterious mutations would occur in variants that move between these classes. Glycine and proline were each allocated to their own special class due to their altered Ramanchandrian space as well as cysteine for its ability to form disulfide bridges.



**Analysis of the importance of structural features for thermal stability**

Changes in structural features was predicted using a batch version of the Missense3D program. The exact way in which theses structural changes are calculated are described in \cite{ittisoponpisan2019Missense3D}. Analysis of the degree to which these predicted features contribute towards protein stability was conducted in a simular way to as was done in \cite{ittisoponpisan2019Missense3D}, except here we look at the effect assuming different \textDelta \textDelta G thresholds has on the predicted importance of such predictions (Figure \ref{fig:workflow}C). Note these thresholds are not intended as exact cutoffs for which a protein misfolds or causes a disease, but rather to give an idea towards how deleterious each structural change is.

At each \textDelta \textDelta G threshold, we calculated the number of Total Positives (NP) and Total Negatives (NN) in our dataset based on whether the experimentally determined \textDelta \textDelta G values exceeded that particular threshold. Then, for every positively predicted feature in each variant, we calculated whether it was a True Positve (TP) or False Positve (FP). To account for the changing balance between the predicted stable and unstable variants in the dataset as we varied the \textDelta \textDelta G threshold, we also calculated the True Positives Rates (TPR) and False Positve Rates (FPR). The importance of each feature with regards to its impact on protein instability was estimated by taking the TPR/FPR ratio. All were calculated as follows:



Features with only a small number of observations are not likely to give statically significant results. Therefore, we've used a threshold of 5 TPs and FPs needing to be observed for our results to be significant.

**Evaluation of stability prediction tools**

We evaluated the ability of FoldX to predict protein stability it two ways. Firstly, we performed simple linear regression analysis (determined by least squares approach) between the FoldX predicted and experimentaly determined \textDelta\textDelta G values (Figure \ref{fig:workflow}E). Pearson correlation coefficients were also calculated. Secondly, we also wanted to investigate which kind of structural alterations FoldX does and doesn't handle well. To do this we plotted boxplots of the error between the predicted and experimental \textDelta\textDelta G values ($Error = \Delta \Delta G\_{pred} - \Delta \Delta G\_{exp}$) for every Missense3D predicted feature (Figure \ref{fig:workflow}D).

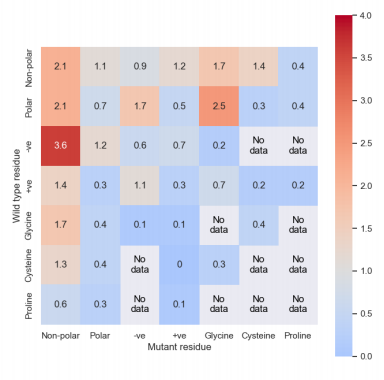
Comparison of the two algorithms was attempted via a ROC curve and other means. However, as Missense3D is trained a predictor of disease and not thermal instability \citep{ittisoponpisan2019Missense3D}, we found it performed rather poorly when compared to FoldX.

**Results**

**Nature of the dataset**

The matrix showing the composition of the dataset we used is shown in Figure \ref{fig:heatmap}. In total there are 49 types on variants given our reduced amino acid alphabet. Variants marked with "No data" are ones where the observed frequency in either (or both) datasets was zero.

Generally, most variants were highly underrepresented. A notable exception to this was residues being mutated to non-polar residues, which were generally overprepresented (excluding proline to non-polar). Negative to non-polar mutants were observed with the highest ratio of 3.6. This high occurrence of mutant non-polar residues is mostly due to alanine being the mutant residue often used, accounting for 288 (26\%) of all variants in the dataset. Mutation of polar and non-polar residues to glycine were also rather underrepresented with ratios of 2.6 and 1.7 respectively. The most underrepresented amino acid was proline, always having a ratio $<$ 1 when either observed as a wild type or mutant residue. Removal of cystienes were also highly underrepresented, with the exception of cystine to non-polar residues, thus we only observed a small number of disulphide breakages were observed.



Bias found for certain missense mutations in our dataset compared to that found in the Missense3D dataset. Red values are over-represented variants while blue are under-represented.

**A picture containing indoor, sitting, long, man

Description automatically generatedImportance of structural features for protein stability**

(A) Absolute number of positives observed for feature feature at various \textDelta\textDelta G thresholds. (B) TPRs, FPRs and TPR/FPR ratios with respect to \textDelta\textDelta G thresholds. Error bars denote 95\% confidence intervals.

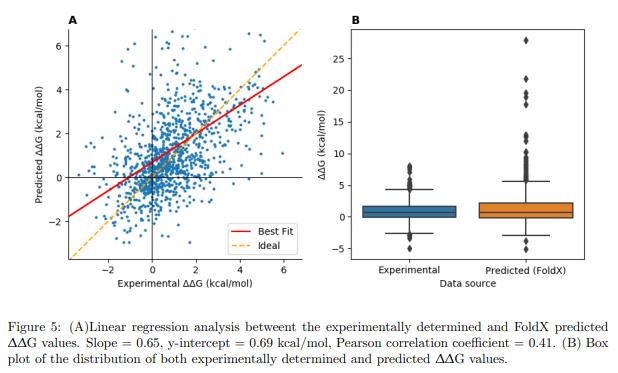
Figure \ref{fig:features} shows the structural changes predicted by Missense3D and their corresponding impact on protein stability. We have shown the absolute number of positives and the positive rates (Figure \ref{fig:features}A) to make it clearer when changes in the both the positive rates and their ratios (Figure \ref{fig:features}B) are a result in changes in the number of observations or the balance between deleterious and neutral variants in the dataset.

The most observed feature was the breakage of buried hydrogen bonds. With 32 and 17 true positives and 27 and 42 False Positives observed at 1.0 and 2.0 kcal/mol \textDelta\textDelta G thresholds respectively. Steric clashes, cavity alterations and residues being switched between the buried and exposed state were also highly observed with $\geq$ 5 true and false positives being observed for all features at both thresholds. Only two disulphide bond breakages were observed and these were both done by mutating either cystine in the same exact bond (Cys32 and Cys 35 in 2TRX). Furthermore, both of these breakages were predicted to be deleterious at both thresholds so we were unable to calculate the TPR/FPR ratio (referred to from now on simply as "the ratio") as the FPR was always zero. Many other features, such as buried proline being introduced, secondary structure alterations, buried charged switching, buried salt bridge breakage and cis-proline replacement, are observed in such little frequency that no reliable conclusions have be drawn from their ratio.

The most deleterious mutation, according to their ratios, is the introduction of buried hydrophilic residues at a \textDelta\textDelta G threshold of 1 kcal/mol with a ratio of 15.0. However,this high value is the result of only 1 false positives being observed. Furthermore, the importance is reduced to a ratio of 4.8 if a 2.0 kcal/mol \textDelta\textDelta G threshold is assumed. The feature with the second highest ratio is the replacement of cis-proline assuming a 1.0 kcal/mol threshold with a value f 7.9. However, this feature is only observed 3 times within our dataset, so there is little to no statistical significance in this observation. Residues switching between the buried and exposed state proved also to be highly deleterious with ratios of 3.3 and 4.9 for 1.0 and 2.0 kcal/cal thresholds respectively. Both of theses observations are also highly significant.

Many features, such as disallowed phi and psi angles, buried charge replacement, buried hydrogen bond breakage and the replacement of glycines in both buried positions and tight bends, had ratios $\approx$ 1. However, the small number of observations means the confidence is very low.

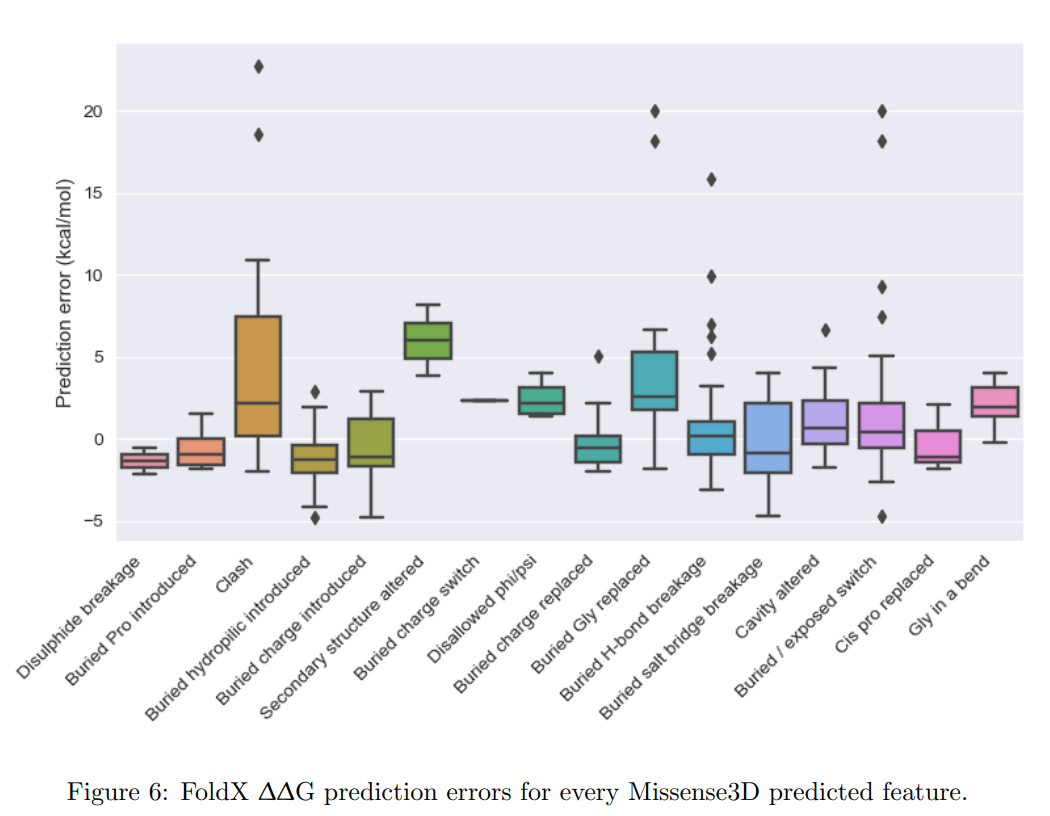
**Accuracy of protein stability prediction programs**



Linear regression analysis between the experimentally determined and FoldX predicted dataset (Figure \ref{fig:linear}A) found the line of best fit to have a slope of 0.65 and a y-intercept of 0.69 kcal/mol. Overall the dataset had a Pearson correlation coefficient of 0.41. To aid in easily identifying the trend, we have intentionally restricted the axis boundaries in Figure \ref{fig:linear}A. However, there are many variants found outside these bounds, especially many with exceptionally high FoldX predicted \textDelta \textDelta G values (Figure \ref{fig:linear}B). There was no consensus feature predicted by Missense3D that correlated well to all variants with these high \textDelta \textDelta G predictions (apart from disease causing). The highest frequency feature in this subset was the prediction of a steric clash, appearing in 4/11 of the variants with a predicted \textDelta \textDelta G greater than 10.0 kcal/mol.

Figure \ref{fig:errors} shows the FoldX \textDelta \textDelta G prediction error for every variant by structural feature as predicted by Missense3D (note that each variant can be plotted multiple times depending on how many features it was predicted to have by Missense3D). The degree to which each feature over or underpredicted the \textDelta \textDelta G value and the amount of outliers varies considerably, although many features were consistently predicted as being more or less stabilising than they actually were.

Steric clashes, buried glycine replacement, buried hydrogen bond breakage and the switching of buried and exposed residues all had significant (mostly positive) outliers. Variants with altered secondary structure were the most inaccurate on average with a mean of 5.9 kcal/mol above the true value, but this was with only 2 observations.



**Discussion** – Subheadings here will either be improved or removed – just there to help with righting at the moment

**Discussion of the results**

The TPR/FPR ratios (referred to as the "ratio" from now on) of Figure \ref{fig:features}B not only indicate the importance of the feature, but also gives an idea of how much every feature contributes towards protein stability. For example, with an assumed \textDelta\textDelta G threshold of 1.0 kcal/mol, the introduction of buried hydrophilic residues has a ratio of 15.0 but this decreases to a ratio of 4.8 assuming a 2.0 kcal/mol threshold. This larger impact ratio at low thresholds would suggest that disturbing the hydrophobic core has a disproportionately large impact on marginally stable proteins. Additionally, the fact that our most damaging predicted feature is related to the hydrophobic effect follows the general observation that protein folding is largely determined by thermodynamics \citep{baker2019denovo2}. In this case, the entropic cost of introducing a hydrophilic residue into the hydrophobic core easily outweighs any enthalpic contributions from the stabilising intramolecular interactions \citep{goldenzweig2018dynamics2}.

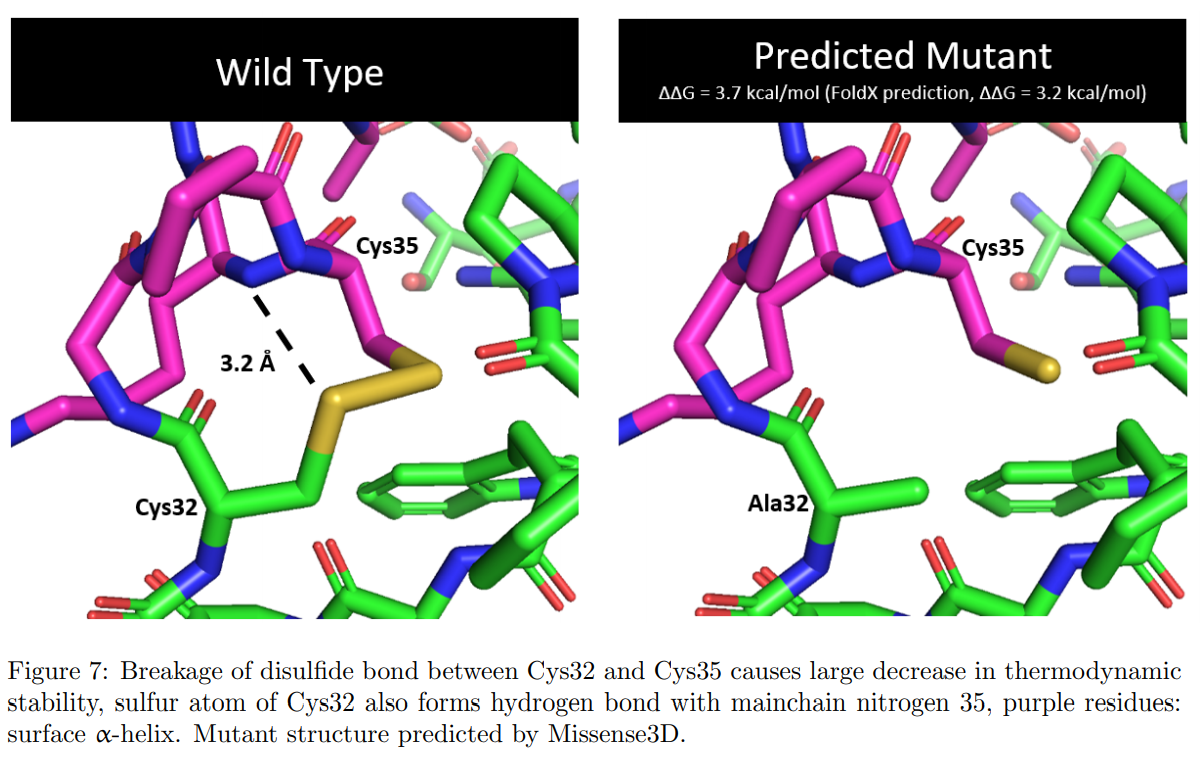
An inverse effect can be observed with the replacement of cis-proline residues, with ratios of 3.0 and 7.9 at 1.0 and 2.0 kcal/mol thresholds respectively. However, given that this is not due to changes in the number of true and false positives observed, but rather the balance between neutral and unstable variants, we cannot say this with any confidence (especially as the feature is observed only 3 times in the whole dataset). However, it appears the replacement of cis-proline can either inhibit or assist folding depending on the kink angle the mainchain is expected to take in the native state \citep{nathaniel2003proline2, williams2004proline3, yohannan2004proline1}.

Conversely, some features displayed FPRs greater than their TPRs, meaning they are a poor predictor of instability. For example, the occurrence of disallowed phi/psi angles in the mainchain of the predicted mutant structure had the second lowest ratio score (0.5-0.6). This shows that there are perhaps variants that can have a significant impact on the conformation of the mainchain without causing significant changes in protein stability with regards to folding \textDelta G. However, this doesn't mean that the conformational changes in these mutants couldn't have a disease-causing effect by altering enzyme activity or protein-protein interactions. It is not surprising Missense3D predicts these disallowed torsion angles given the fact that it assumes perfectly rigid mainchain conformations \citep{ittisoponpisan2019Missense3D}. Not surprisingly, FoldX performed poorly on the 8 variants identified to have this feature, with all being predicted has having a larger \textDelta\textDelta G than reality by 2.4 kcal/mol on average ($\pm$ 1.4 kcal/mol for whole dataset). Given our approach, we are not able to tell whether the variants adopt stable alternate conformations or completely denature. In order to investigate these mutants with that likely have an altered backbone we would suggest using an algorithm capable of modelling backbone flexibility such as Rosetta \citep{rohl2004rosetta}.

The breakage of buried salt bridges had the lowest ratio of 0.4 for 1.0 kcal/mol \textDelta\textDelta G and no observed true positives for 2.0 kcal/mol. This feature was observed only 5 times with a mean \textDelta\textDelta G of -0.1 kcal/mol (with the highest value being 1.72 kcal/mol), meaning some of our variants were stabilised by the removal of the salt bridge. While the salt bridge interaction itself can be very stabilising (3-5 kcal/mol) \citep{anderson1990salt}, buried salt bridges are rarely observed as the stabilising ionic interactions are often not enough to compensate for the entropic cost of moving two charged residues away from water to a hydrophobic protein core \citep{pace2014salt2}. This is supported by the high ratio for introduction of buried charged residues (5.5 and 4.0 for 1.0 and 2.0 kcal/mol thresholds respectively), where we can assume that most of the new charges do on form a ionic interaction in the hydrophobic core, leaving only the thermodynamic penalty. Alternatively, the destabilising effects could also be due to repulsive coulombic forces from like-charges being introduced into an existing charge network. Counter to all this evidence we see that the removal of buried charges had roughly neutral effects on stability, with ratios of 0.9 and 1.3 for 1.0 and 2.0 kcal/mol thresholds respectively. This could be due the wild type charged residues being hydrated by internal water pockets \citep{bellissent2016water}, meaning their presence or absence has a relatively neutral effect on proteins stability. Thus, our results suggest that natural selection causes a bias for buried charges to only be introduced into certain environments when compared to random mutagenesis. However, the confidence intervals of the TPRs and FPRs for the removal of buried charges are such that the true ratios could be substantially different either way.

Missense3D predicted only two disulfide breakages, both of which were for either side of the same bond. We suggest the low abundance of disulfide breakages is a result of the experimental data being taken from protein engineerning studies, where the breakage of disulfide bonds is rarely attempted due to the likely deleterious effect it would likely have on protein stability. This low sample size makes it difficult to draw any meaningful and statistically significant conclusions about the importance of disulphide bonds.

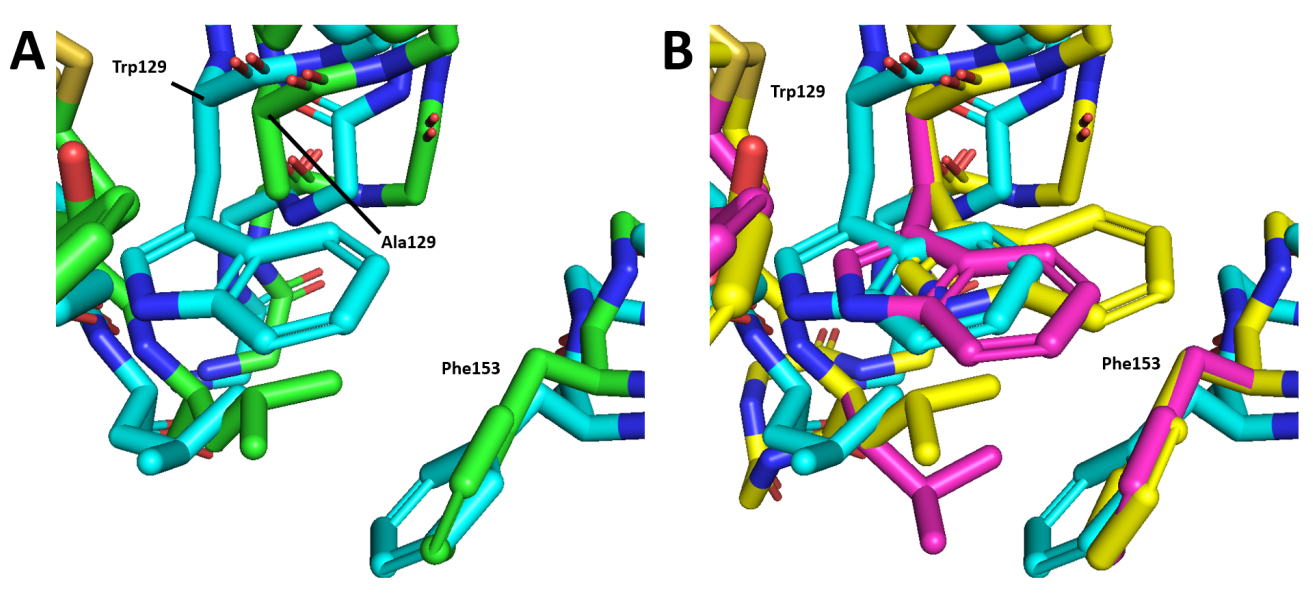
Missense3D identified the breakage of a disulfide bond found in thioredoxin in \textit{E. coli} (PDB: 2TRX) upon point mutagenasis of Cys32 to Ala32 (Figure \ref{fig:ss-bond}). A disulfide bond in a highly structured region like this (in this case holding a \textalpha-helix close to the surface) is likely to make strong contributions to protein thermal stability. It was also found in the literature that the sulfur atom of the Cys32 forms a hydrogen bond with the mainchain nitrogen of the Cys35 residue. As it has partial solvent accessibility \citep{katti1990disulfidebond}, Missense3D has not predicted it as a buried hydrogen bond breakage. Given both of these stabilising effects it is not surprising that the removal of the cystine at position 32 had a detrimental effect on protein stability, with an experimentally measured \textDelta\textDelta G value of 3.7 kcal/mol. Interestingly, mutagenisis of the Cys35 to Ala35 only had a \textDelta\textDelta G value of 3.1 kcal/mol, suggesting that the remaining Cys32 residue was still able to form stabilising interactions with the mainchain nitrogen.



**FoldX ability**

Our results have demonstrated that the predictive ability of FoldX varies significantly depending on the scenario. FoldX especially struggles with anomalies where the predicted \textDelta\textDelta G is significantly higher than the true value (Figure \ref{fig:linear}B) and most of these tended to be in the same kind of variants (Figure \ref{fig:errors}). In the case of hydrogen bond breakage and switching buried exposed residues, we propose these large inaccuracies could be the result of the small and biased dataset used to train FoldX \citep{nisthal2019domain-wide} and the use of statistical potentials \citep{schymkowitz2005foldx} that do not result in versatile energy equations. It is not a surprise then that the latest improvements in the field have backed up statistical potentials with a machine learning approach \citep{pucci2016hotmusic}. In the case of glycine replacement and steric clashes we expect the high \textDelta\textDelta G predictions are a result of extremely unfavourable energy calculations that arise from the restriction on conformational rearrangement due to the rigid mainchain assumptions \citep{schymkowitz2005foldx}.

An example of this can been seen when introducing a buried tryptophan residue at position 129 in place of alanine in the T4 lysozyme. When comparing the experimentally determined pseudo-wild-type and mutant structures (Figure \ref{fig:clash}A) the tryptophan residue adopts a pose within a pre-existing cavity. Nevertheless, both the backbone alpha-helices with residues 129 and 153 need to be shifted away from the cavity by 0.9 \AA each in the mutant native state with an experimentally measured \textDelta\textDelta G of 3.2 kcal/mol \citep{liu2000clash}. Both the Missense3D and FoldX predicted structures struggled significantly by not being able to move the backbone and the resultant structures have Trp129 clashing with Phe153 (Figure \ref{fig:clash}B). This was especially true in the FoldX predicted strucuture, with a predicted folding \textDelta\textDelta G of 21.8 kcal/mol, an absolute error of 18.7 kcal/mol.



(A) Experimentally determined pseudo-wild-type (Green, 1L63) and mutant (Cyan, 1QTD) structures for single mutant A129W on T4 lysozyme. Notice the large mainchain conformational changed needed to fit the tryptophan side chain in the packed protein core. (B) Experimentally determined mutant (Cyan, 1QTD), Missense3D predicted (Purple) and FoldX predicted (Yellow) structures of the A129W mutant. Both programs are unable to perform the conformational rearrangements needed to model the mutant Trp129 residue without a steric clash with the beta-carbon of Phe153.

Some features in our dataset, such as disulphide breakages, introduction of buried proline residues and cis-proline replacement, had \textDelta\textDelta G values predicted to a high degree of accuracy (Figure \ref{fig:errors}). However, the small number of observations for each of these features means we can have little confidence that this would extrapolate well in studies on a wider (possibly proteome) scale.

Predictions using FoldX showed that almost universally, the distribution of stability effects follows a bi-Gaussian distribution, with surface mutations being narrowly distributed with a mildly destabilising mean (\textDelta\textDelta G $\approx$ 0.6 kcal/mol) while core residue mutations have a stronger destabilising mean (\textDelta\textDelta G $\approx$ 1.4 kcal/mol) and wider distribution \citep{tokuriki2007surface\_vs\_core}.

While we have only used the final \textDelta\textDelta G prediction here, the output of FoldX contains energetic breakdowns of the contributions for each type of molecular interaction it models \citep{schymkowitz2005foldx}. We would suggest taking a similar approach as in Figure \ref{fig:errors} but instead correlating the prediction errors with the more detailed breakdowns to elucidate the energy functions in FoldX that prove to be most anomalous.

Analysis on another dataset similar to ours found that the SomeMin method within Rosetta performed the best with regards to predicting folding \textDelta\textDelta G, with a Pearson correction coefficient of 0.64 (compared to ours of 0.41). Notably, the best results were achieved with only allowing partial backbone flexibility, whilst unconstrained flexibility produced less clashes at the expense of accuracy \citep{nisthal2019domain-wide}. The other methods assessed by the same author, PoPMuSiC \citep{dehouck2011popmusic} and FoldX \citep{schymkowitz2005foldx}, performed less well with r = 0.56 and 0.51 respectively \citep{nisthal2019domain-wide}, suggesting FoldX is not the ideal tool for trying to do this kind of study computationally. The disparity in performance can likely be attributed to the rigid backbone assumptions and the small, biased dataset (n=339) \citep{schymkowitz2005foldx}).

**Data limitations**

A major limitation of our approach is that the mutatgenesis studies used to construct our dataset are often trying to engineer increased thermal stability and will actively avoid conducting experiments that they think will do the opposite. This results in making it very difficult to observe many structural changes that are likely to be the most deleterious to protein folding (such as disulphide breakages), let alone those results being statistically significant. Additionally, variants collected from older random mutagensis studies suffered from bias towards certain nucleotide (and therefore amino acid) substitutions due to the low-fidelity DNA polymerases of the time \citep{vanhercke2005randommutatgen}. We suggest taking a similar approach but instead looking for features which are likely to cause increased protein stability (e.g. disulphide and salt bridge formation and buried hydrophobic residues being introduced).

Construction of a data set larger than that of HoTMuSiC was attempted by trying to recreate a similar dataset from ProTherm but using a loosened criteria for variants. Notably, we removed the need for variants not to cause a large change in thermal stability \citep{pucci2016HoTMuSiCdataset}, as we hoped this would capture greater changes in protein structure. While this did result in having roughly 50 \% more variants, the accuracy of the prediction tools deteriorated significantly (r = $\approx$ 0.2). This could be due to both FoldX and Missense3D assuming that mainchain conformations are unchanged and only side chains are repacked in all their predicted variants \citep{delgado2019foldx5, ittisoponpisan2019Missense3D}. Which would mean that adding these variants with the high degree of structural change could not be accurately predicted by either program. This is also a refection of the high-quality manual curation that was done in the construction of the HoTMuSiC data set.

**Better techniques**

For every missense mutation within a protein there are many more local environmental factors that can be taken into account such as ASA and secondary structure. Unfortunately, the highly dimensional nature of our data makes it difficult to convey all the information that we have available to explain changes in proteins stability in one figure. For example, mutated residues with a high ASA and residues found in disordered regions (instead of regular secondary structures) are likely to have a less detrimental effect on protein stability. Additionally, the combination of binary and scalar information does not lend itself to traditional forms of dimensionality reduction (Multiple Component Analysis \citep{greenacre2006MCA} was attempted but with poor results). Although we do think this could be explored more. For this reason, the data (including information not presented in figures) and code have been made available for further investigation.

We have been clear throughout that we are not answering the question of what features increase stability, rather which cause instability. However, we have shown that already existing computational methods are accurate predictors of both changes in protein thermal stability and structural changes upon mutation. Therefore, we propose that future work using similar tools could produce useful insights into how structural features in a protein can be changed in order to increase stability, as well as further our understanding of protein engineering as a whole. At the very least, such work would certainly be invaluable in guiding experimental work in the field.

Here, we have only assessed 16 structural features with out approach but we believe more insights could be obtained with a greater range of structural observations. For example, $\pi$-$\pi$ and cation-$\pi$ interactions were not studied here but have a higher occurrence in thermophilic proteins \citep{makwana2015aromatic1, gromiha2002aromatic2}. One investigation showed that stabilisation of the dimer interface in insulin was achieved by the addition of a inter-chain $\pi$-$\pi$ interaction, causing a 150-fold increase in \textit{in vitro} lifetime and a structure virtually identical to that of the wild-type \citep{rege2018insulin}. Furthermore, it has been reported that cation-$\pi$ interactions become more stabilising as the protein approaches its melting temperature \citep{prajapati2006cation-pi1} so there is certainly interesting dynamics that could be investigated for biopharmaceutical and high-temperature industrial applications.

**Conclusion**

In this work we have tried to elucidate the some of the structural causes for protein instability. As expected, there seem to be no universal explanations and instead we have tried to present our results in a informative way that gives an appreciation for which factors tend to influence stability the most and an idea of where the field is heading. Although our results have shown that protein folding and stability seems to be governed much more by thermodynamics rather than energetics.

While the statistical significance of our work struggles from small data size, we have demonstrated that existing computational programs are good investigative tools for studying the mechanisms of protein stability. However, we would be cautious in advising to use FoldX to generate large artifical datasets as we have shown its results to be highly anomalous in certain scenarios . Further work could also be done into whether predicted three-dimensional protein structures can be used to analyse the stabilising or destabilising effects of missense mutations on a larger proportion of the proteome, similar to how \cite{ittisoponpisan2019Missense3D} did with disease-causing variants.

Another aspect we have not investigated is that of protein dynamics. It is often not a correct assumption that proteins are found to be static structures. Rather, they are highly dynamical molecules with a range of flexibility depending on the environmental conditions and regularly move between local minima in the high dimensional energy landscape. The activity of enzymes is determined to a large extent by the conformational fluctuations that the protein undergoes. In many cases these large fluctuations are necessary to stabilise the transition state on an enzyme substrate \citep{kim2017dynamics1, goldenzweig2018dynamics2}. --FINISH PARAGRAPH

Note that while we have critiqued FoldX and Missense3D for its rigid mainchain assumptions \citep{schymkowitz2005foldx, ittisoponpisan2019Missense3D}, modeling the mainchain flexibly increases the number of degrees of freedom significantly and therefore computational cost also. In many cases these costs are not advantageous due to the simplicity of the problem or size of the analysis, meaning the choice of program is highly situational. This is reflected in the wide variety of algorithms available for different goals \citep{rohl2004rosetta, pucci2016hotmusic, schymkowitz2005foldx}.

New and high-throughput experimental pipelines, such as deep mutational screening \citep{fowler2014dms}, could in the future produce higher quality, domain and proteome-wide datasets with more consistent experimental conditions than are seen across compiled databases such as ProTherm. One such example of this is presented in \cite{nisthal2019domain-wide} where they produce and experimentally measure the folding \textDelta\textDelta G for nearly every possible single mutant in the 56-residue \textbeta 1 domain of the streptococcal protein G (n=935). The key being that this data was collected on the scale of weeks instead of years with a roughly 20-fold increase in speed \citep{nisthal2019domain-wide}. We have not performed any analysis on this dataset here as we think our approach would suffer from the multiple testing problem by mutating every residue on a single small protein. However, these new methods will produce the data needed for more insights into protein stability and to train better prediction algorithms.

Despite all the work that have been done to try and explain protein stability, a lot remains unanswered. Listed below are some possible future areas that could lead to more valuable conclusions (some have been more explored than others).

* More accurate and faster prediction methods of protein thermal stability}, or the change in stability upon mutation, are essential for the development of new biotechnological and biopharmeutical applications \citep{pucci2016HoTMuSiC}.
* High thought-put experimental determination of protein stability} is going to pave the way for large, representative and high-quality datasets that will provide greater understanding of protein stability and improve the accuracy of prediction algorithums \citep{fowler2014dms, nisthal2019domain-wide}.
* Membrane proteins} can maintain very high stability and function whilst in a completely different environment to that of globular proteins, a lot remains to be learnt from their analysis \citep{hayashi2020membrane3,mbaye2019membrane1, yasuda2016membrane2}.
* The eukaryotic cell cycle} appears to modulate protein stability at the proteome-level within single cells through changes in protein abundance and post-translational modifications \citep{becher2018cellcyle}.
* Molecular dynamics} will be essential to discover new information that cannot be explained from static structures as well as designing proteins that have a folding pathway that is easily guided to the native state \citep{childers2017md1}
* Chaperones} help the cell respond to temperature fluctuations \citep{zhang2015chaperones, jia2016intelligent} and accelerates protein evolution by providing a thermodynamic buffer which facilitates the destabilising mutations needed to acquire new functions before compensatory mutations restore stability \citep{tokuriki2009chaperonin}.
* Domain- and proteome-wide analysis} \citep{nisthal2019domain-wide}.
* Terahertz time-domain spectroscopy} can be used to measure the changes in water dynamics around a protein molecule at the picosecond scale \citep{aoki2016terahertz}.

**References**

Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990), ‘ph-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of t4 lysozyme’, Biochemistry 29(9), 2403–2408. 13 Aoki, K., Shiraki, K. & Hattori, T. (2016), ‘Salt effects on the picosecond dynamics of lysozyme hydration water investigated by terahertz time-domain spectroscopy and an insight into the hofmeister series for protein stability and solubility’, Physical Chemistry Chemical Physics 18(22), 15060–15069. Baker, D. (2019), ‘What has de novo protein design taught us about protein folding and biophysics?’, Protein Science 28(4), 678–683. Becher, I., Andr´es-Pons, A., Romanov, N., Stein, F., Schramm, M., Baudin, F., Helm, D., Kurzawa, N., Mateus, A., Mackmull, M.-T. et al. (2018), ‘Pervasive protein thermal stability variation during the cell cycle’, Cell 173(6), 1495–1507. Becktel, W. J. & Schellman, J. A. (1987), ‘Protein stability curves’, Biopolymers: Original Research on Biomolecules 26(11), 1859–1877. Bellissent-Funel, M.-C., Hassanali, A., Havenith, M., Henchman, R., Pohl, P., Sterpone, F., van der Spoel, D., Xu, Y. & Garcia, A. E. (2016), ‘Water determines the structure and dynamics of proteins’, Chemical reviews 116(13), 7673–7697. Bhattacharya, R., Rose, P. W., Burley, S. K. & Prli´c, A. (2017), ‘Impact of genetic variation on three dimensional structure and function of proteins’, PloS one 12(3). Bhirde, A. A., Chiang, M.-J., Venna, R., Beaucage, S. & Brorson, K. (2018), ‘High-throughput in-use and stress size stability screening of protein therapeutics using algorithm-driven dynamic light scattering’, Journal of pharmaceutical sciences 107(8), 2055–2062. Casadio, R., Vassura, M., Tiwari, S., Fariselli, P. & Luigi Martelli, P. (2011), ‘Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome’, Human Mutation 32(10), 1161–1170. Childers, M. C. & Daggett, V. (2017), ‘Insights from molecular dynamics simulations for computational protein design’, Molecular systems design & engineering 2(1), 9–33. Dehouck, Y., Kwasigroch, J. M., Gilis, D. & Rooman, M. (2011), ‘Popmusic 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality’, BMC bioinformatics 12(1), 151. Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. (2019), ‘Foldx 5.0: working with rna, small molecules and a new graphical interface’, Bioinformatics 35(20), 4168–4169. Demarest, S. J. & Glaser, S. M. (2008), ‘Antibody therapeutics, antibody engineering, and the merits of protein stability.’, Current opinion in drug discovery & development 11(5), 675–687. Feller, G. (2013), ‘Psychrophilic enzymes: from folding to function and biotechnology’, Scientifica 2013. Fowler, D. M. & Fields, S. (2014), ‘Deep mutational scanning: a new style of protein science’, Nature methods 11(8), 801. Frokjaer, S. & Otzen, D. E. (2005), ‘Protein drug stability: a formulation challenge’, Nature reviews drug discovery 4(4), 298–306. Goldenzweig, A. & Fleishman, S. J. (2018), ‘Principles of protein stability and their application in computational design’, Annual review of biochemistry 87, 105–129. Greenacre, M. & Blasius, J. (2006), Multiple correspondence analysis and related methods, CRC press. Gromiha, M. M., Thomas, S. & Santhosh, C. (2002), ‘Role of cation-π interactions to the stability of thermophilic proteins’, Preparative Biochemistry and Biotechnology 32(4), 355–362. Hartl, F. U. & Hayer-Hartl, M. (2009), ‘Converging concepts of protein folding in vitro and in vivo’, Nature structural & molecular biology 16(6), 574. Hayashi, T., Yasuda, S., Suzuki, K., Akiyama, T., Kanehara, K., Kojima, K., Tanabe, M., Kato, R., Senda, T., Sudo, Y. et al. (2020), ‘How does a microbial rhodopsin rxr realize its exceptionally high thermostability with the proton-pumping function being retained?’, The Journal of Physical Chemistry B 124(6), 990–1000. 14 Huang, P.-S., Boyken, S. E. & Baker, D. (2016), ‘The coming of age of de novo protein design’, Nature 537(7620), 320–327. Ittisoponpisan, S., Islam, S. A., Khanna, T., Alhuzimi, E., David, A. & Sternberg, M. J. (2019), ‘Can predicted protein 3d structures provide reliable insights into whether missense variants are disease associated?’, Journal of molecular biology 431(11), 2197–2212. Jia, H., Sun, X., Sun, H., Li, C., Wang, Y., Feng, X. & Li, C. (2016), ‘Intelligent microbial heat-regulating engine (imhere) for improved thermo-robustness and efficiency of bioconversion’, ACS synthetic biology 5(4), 312–320. Katti, S. K., LeMaster, D. M. & Eklund, H. (1990), ‘Crystal structure of thioredoxin from escherichia coli at 1.68 ˚a resolution’, Journal of molecular biology 212(1), 167–184. Kim, T. H., Mehrabi, P., Ren, Z., Sljoka, A., Ing, C., Bezginov, A., Ye, L., Pom`es, R., Prosser, R. S. & Pai, E. F. (2017), ‘The role of dimer asymmetry and protomer dynamics in enzyme catalysis’, Science 355(6322), eaag2355. Krishna, M. & Nadler, S. G. (2016), ‘Immunogenicity to biotherapeutics–the role of anti-drug immune complexes’, Frontiers in immunology 7, 21. Krivov, G. G., Shapovalov, M. V. & Dunbrack Jr, R. L. (2009), ‘Improved prediction of protein side-chain conformations with scwrl4’, Proteins: Structure, Function, and Bioinformatics 77(4), 778–795. Kumar, M. S., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H. & Sarai, A. (2006), ‘Protherm and pronit: thermodynamic databases for proteins and protein–nucleic acid interactions’, Nucleic acids research 34(suppl 1), D204–D206. Liu, Q., Xun, G. & Feng, Y. (2019), ‘The state-of-the-art strategies of protein engineering for enzyme stabilization’, Biotechnology advances 37(4), 530–537. Liu, R., Baase, W. A. & Matthews, B. W. (2000), ‘The introduction of strain and its effects on the structure and stability of t4 lysozyme’, Journal of molecular biology 295(1), 127–145. Madan, M., Sikriwal, D., Sharma, G., Shukla, N., Mandyal, A. K., Kale, S. & Gill, D. (2018), ‘Rational design of heat stable lyophilized rotavirus vaccine formulations’, Human vaccines & immunotherapeutics 14(9), 2132–2141. Makwana, K. M. & Mahalakshmi, R. (2015), ‘Implications of aromatic–aromatic interactions: From protein structures to peptide models’, Protein Science 24(12), 1920–1933. Mbaye, M. N., Hou, Q., Basu, S., Teheux, F., Pucci, F. & Rooman, M. (2019), ‘A comprehensive computational study of amino acid interactions in membrane proteins’, Scientific reports 9(1), 1–14. Miyazawa, S. (2017), ‘Selection originating from protein stability/foldability: Relationships between protein folding free energy, sequence ensemble, and fitness’, Journal of theoretical biology 433, 21–38. Nathaniel, C., Wallace, L. A., Burke, J. & Dirr, H. W. (2003), ‘The role of an evolutionarily conserved cisproline in the thioredoxin-like domain of human class alpha glutathione transferase a1-1’, Biochemical Journal 372(1), 241–246. Nisthal, A., Wang, C. Y., Ary, M. L. & Mayo, S. L. (2019), ‘Protein stability engineering insights revealed by domain-wide comprehensive mutagenesis’, Proceedings of the National Academy of Sciences 116(33), 16367–16377. Onuchic, J. N. & Wolynes, P. G. (2004), ‘Theory of protein folding’, Current opinion in structural biology 14(1), 70–75. Pace, C. N., Scholtz, J. M. & Grimsley, G. R. (2014), ‘Forces stabilizing proteins’, FEBS letters 588(14), 2177–2184. Peng, Y. & Alexov, E. (2016), ‘Investigating the linkage between disease-causing amino acid variants and their effect on protein stability and binding’, Proteins: Structure, Function, and Bioinformatics 84(2), 232–239. 15 Prabhu, N. V. & Sharp, K. A. (2005), ‘Heat capacity in proteins’, Annu. Rev. Phys. Chem. 56, 521–548. Prajapati, R. S., Sirajuddin, M., Durani, V., Sreeramulu, S. & Varadarajan, R. (2006), ‘Contribution of cation- π interactions to protein stability’, Biochemistry 45(50), 15000–15010. Pucci, F., Bourgeas, R. & Rooman, M. (2016a), ‘High-quality thermodynamic data on the stability changes of proteins upon single-site mutations’, Journal of physical and chemical reference data 45(2), 023104. Pucci, F., Bourgeas, R. & Rooman, M. (2016b), ‘Predicting protein thermal stability changes upon point mutations using statistical potentials: Introducing hotmusic’, Scientific reports 6, 23257. Pucci, F. & Rooman, M. (2017), ‘Physical and molecular bases of protein thermal stability and cold adaptation’, Current opinion in structural biology 42, 117–128. Razvi, A. & Scholtz, J. M. (2006), ‘Lessons in stability from thermophilic proteins’, Protein Science 15(7), 1569–1578. Rege, N. K., Wickramasinghe, N. P., Tustan, A. N., Phillips, N. F., Yee, V. C., Ismail-Beigi, F. & Weiss, M. A. (2018), ‘Structure-based stabilization of insulin as a therapeutic protein assembly via enhanced aromatic–aromatic interactions’, Journal of Biological Chemistry 293(28), 10895–10910. Rickerby, H. F., Putintseva, K. & Cozens, C. (2020), ‘Machine learning-driven protein engineering: a case study in computational drug discovery’, Engineering Biology 4(1), 7–9. Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004), Protein structure prediction using rosetta, in ‘Methods in enzymology’, Vol. 383, Elsevier, pp. 66–93. Schafer, N. P., Kim, B. L., Zheng, W. & Wolynes, P. G. (2014), ‘Learning to fold proteins using energy landscape theory’, Israel journal of chemistry 54(8-9), 1311–1337. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. & Serrano, L. (2005), ‘The foldx web server: an online force field’, Nucleic acids research 33(suppl 2), W382–W388. Sun, T., Han, H., Hudalla, G. A., Wen, Y., Pompano, R. R. & Collier, J. H. (2016), ‘Thermal stability of self-assembled peptide vaccine materials’, Acta biomaterialia 30, 62–71. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. (2007), ‘The stability effects of protein mutations appear to be universally distributed’, Journal of molecular biology 369(5), 1318– 1332. Tokuriki, N. & Tawfik, D. S. (2009), ‘Chaperonin overexpression promotes genetic variation and enzyme evolution’, Nature 459(7247), 668–673. Vanhercke, T., Ampe, C., Tirry, L. & Denolf, P. (2005), ‘Reducing mutational bias in random protein libraries’, Analytical biochemistry 339(1), 9–14. Williams, A. D., Portelius, E., Kheterpal, I., Guo, J.-t., Cook, K. D., Xu, Y. & Wetzel, R. (2004), ‘Mapping aβ amyloid fibril secondary structure using scanning proline mutagenesis’, Journal of molecular biology 335(3), 833–842. Yasuda, S., Kajiwara, Y., Takamuku, Y., Suzuki, N., Murata, T. & Kinoshita, M. (2016), ‘Identification of thermostabilizing mutations for membrane proteins: rapid method based on statistical thermodynamics’, The Journal of Physical Chemistry B 120(16), 3833–3843. Yates, C. M., Filippis, I., Kelley, L. A. & Sternberg, M. J. (2014), ‘Suspect: enhanced prediction of single amino acid variant (sav) phenotype using network features’, Journal of molecular biology 426(14), 2692–2701. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P. & Bowie, J. U. (2004), ‘The evolution of transmembrane helix kinks and the structural diversity of g protein-coupled receptors’, Proceedings of the National Academy of Sciences 101(4), 959–963. Zhang, K., Ezemaduka, A. N., Wang, Z., Hu, H., Shi, X., Liu, C., Lu, X., Fu, X., Chang, Z. & Yin, C.-C. (2015), ‘A novel mechanism for small heat shock proteins to function as molecular chaperones’, Scientific reports 5, 8811. 1