Imperial College London

Final-Year Research Project Cover Sheet

Students: Please fill in and use it as the front page of your project report

Name of Student: Charles Harris

Degree Stream: Biochemistry

Date: 05/06/2020

Word Count: 6017

Project Title: Deriving structural explanations for protein stability through temperature sensitive mutants

Type of Project: DELETE AS APPROPRIATE

Practical Project Data Project (10 weeks) – DP1/DP2

I hereby declare that this Project has not been submitted, either in the same or different forms, to this or any other university for a degree. I also declare that this Project does not draw from any other work prepared under consultancy or other professional undertaking, by myself or jointly with other authors in any way other than that duly and explicitly acknowledged herewith.

Name of Supervisor: Prof Michael Sternberg

Name of Examiner 1: Dr Samraat Pawar

Name of Examiner 2: Dr Alfonso De Simone

Staff use: Cut and paste the following link into your browser for the Project Examiners forms.

https://bb.imperial.ac.uk/webapps/blackboard/content/listContentEditable.jsp?content_id=_1517936_1&course_id=_12010_1&content_id=_1772614_1

Abstract

Protein stability engineering will prove invaluable in emerging fields in biotechnology and biopharmaceuticals. Many disease causing mutations are as a result of protein instability. Despite this, we have not been able to determine rules to universally explain protein stability, rather, we only have a 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. 1088 variants across 59 proteins are analysed. Our results confirm the understanding that protein folding is largely governed by thermodynamics. However, the contributions of individual types of interactions proved hard to quantify directly. Also addressed is whether FoldX could be used to study these effects on a larger scale by using predicted data and as well as what is needed for the advancement of the field.

1 Introduction

Understanding the structural features that give proteins their stability is important for the development and optimisation of new biotechnological applications (Liu et al. 2019) and biopharmaceuticals (Frokjaer & Otzen 2005, Demarest & Glaser 2008, Sun et al. 2016, Madan et al. 2018, Rickerby et al. 2020). 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 (Casadio et al. 2011, Yates et al. 2014, Peng & Alexov 2016). Having these well understood explanations will also be essential if we are to move towards *de novo* protein design (Huang et al. 2016, Baker 2019). This report uses commonly available computational protein stability programs and datasets to try and elucidate some of the mechanisms that determine stability. Also addressed is the strengths and limitations of these programs for this kind of analysis and future areas of interest in the field.

A new generation of recombinantly expressed, engineered protein therapeutics (Frokjaer & Otzen 2005) and monoclonal antibodies (Demarest & Glaser 2008) 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 (Krishna & Nadler 2016). Additionally, most current vaccines are thermally sensitive and need to be stored in a continuous "cold chain", presenting significant logistical challenges (Sun et al. 2016), especially in the developing world (Madan et al. 2018). Whilst techniques have been developed to monitor the structural integrity of such therapeutics, both before and after biomanufacturing (Bhirde et al. 2018), an all together better approach, both clinically and economically, is to engineer future therapeutics to be more thermally resistant.

Most *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 to add residues that are essential to function but do not lend themselves to stability (e.g. addition of hydrophobic residues on protein binding surfaces) (Huang et al. 2016, Baker 2019).

Protein folding is a delicate balance between having enough thermodynamic stability to maintain the structure whilst allowing for the conformational flexibility that is required for function (Goldenzweig & Fleishman 2018). For example, in the case of enzymes this manifests in the protein having the dynamics to stabilise the transition state once the substrate is bound (Kim et al. 2017, Goldenzweig & Fleishman 2018). This means a missense mutation could increase stability such that it alters the conformational sampling in such a way that it is deleterious to the original function and therefore causes disease (Chiang et al. 2016). Alternatively, missense mutations that decrease stability can cause disease either through enhanced dynamics (Kumar et al. 2017) or misfolding and aggregation of the protein (Plaza del Pino et al. 2000, Protasevich et al. 2010).

Considerable effort has been put towards the furthering of our understanding of how structural features relate to protein thermal stability (Pucci & Rooman 2017, Miyazawa 2017, Harms & Thornton 2013), and much has been learnt from the study of thermophilic protein (Razvi & Scholtz 2006, Feller 2013). Work has also been done into how variation at the genome level impacts protein structure and function (Bhattacharya et al. 2017). Unfortunately, there seem 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 (Pucci & Rooman 2017).

Given the importance protein stability has in the life sciences and the degree to which is can change upon a single missense mutation, it is not surprising there are many programs available that aim to

predict such changes in thermodynamic stability (Khan & Vihinen 2010, Potapov et al. 2009). The majority of approaches rely on statistically derived energy potentials (Schymkowitz et al. 2005, Dehouck et al. 2011, Pucci et al. 2017) but there are also some that use support vector machines (Capriotti et al. 2005) and artificial neural networks (ANNs) (Capriotti et al. 2004). The most recent advancements have come from combining statistical potentials with ANNs (Pucci et al. 2016*b*). Additionally, while they are not specifically trained for predicting stability changes upon point mutation, tools intended for protein design (Rohl et al. 2004) have proved to perform well at this task (Nisthal et al. 2019).

In recent decades the accuracy of these computational tools has meant that we can now use them to generate large synthetic datasets, that would not be practical to collect experimentally, which allow us to investigate the causes of protein stability further (Yue et al. 2005, Tokuriki et al. 2007, Pucci et al. 2020). However, recent analysis has shown that a major limitation in the algorithms developed thus far is that they are mainly trained on destabilising mutants, meaning they perform poorly when trying to predict the effects of stabilising mutations (Pucci et al. 2018, Usmanova et al. 2018, Montanucci et al. 2019). This will undoubtedly need to be addressed if we are to seriously use these tools for protein engineering. Many methods also exist that try to predict whether missense mutations are disease causing variants (Thusberg et al. 2011, Ittisoponpisan et al. 2019).

1.1 Theoretical review of protein stability

The thermodynamic stability of the protein folding process is easily characterised by the changes in Gibbs free energy (ΔG). The exact value of ΔG 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 1) 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 ΔG equals zero. Accordingly, changes in protein stability can be described in terms of its effect on melting temperature (ΔT_M). The complex relationship between folding ΔG and other thermodynamic descriptors has been covered elsewhere in the literature (Becktel & Schellman 1987, Prabhu & Sharp 2005, Miyazawa 2017), but these go beyond the level of detail needed for this investigation.

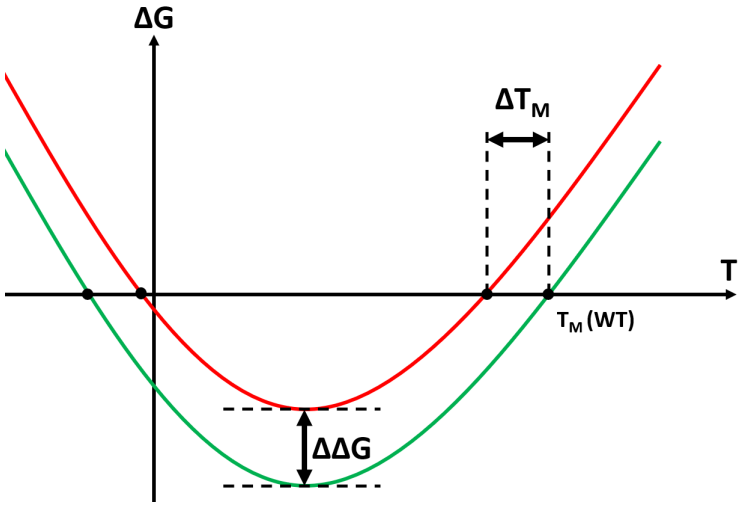

Figure 1: Green line: thermodynamic properties of a hypothetical wild type protein. Red line: thermodynamic properties of a simple hypothetical destabilising mutant of the original wild type. Note the negative ΔT_M and positive ΔG in this destabilising example.

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 (Onuchic & Wolynes 2004). 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 (Hartl & Hayer-Hartl 2009). While this point is not addressed here, this means the stability of natural and engineered proteins should not only be judged by their folding ΔG but also their propensity to adopt 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 (Goldenzweig & Fleishman 2018, Schafer et al. 2014, Berezovsky et al. 2007).

The main driving force behind protein folding has been identified as the hydrophobic effect. This is due to it being entropically more favourable for hydrophobic residues to cluster together in a protein core and exclude themselves from water (Baker 2019, Camilloni et al. 2016). Additionally, while it is temperature dependent, when compared to other interactions, the hydrophobic effect seems to be essentially constant (Folch et al. 2008). It has proved more difficult to quantify the impact of disulphide bridges due to their low occurrence, but it appears that native disulphide bridges provide increased thermal stability by decreasing the conformational degrees of freedom (and therefore entropy) of the unfolded state (Jo et al. 2016, Pecher & Arnold 2009). A lot remains to be learnt, with *de novo* protein design using disulphide bonds having produced conflicting results (Bashirova et al. 2019, Niu et al. 2016). The role of salt bridges has proved even more eluding (Pucci & Rooman 2017), but their presence is notably increased in thermophilic (Karshikoff & Ladenstein 2001) and decreased in psychrophilic organisms (Parvizpour et al. 2017). The later of which is usually explained by the need for increased conformational flexibility at lower temperatures (Struvay & Feller 2012).

Mutations near the protein surface are also known to be less destabilising on average when compared to those in the protein core. Predictions done elsewhere 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 ($\Delta\Delta$G $\approx$ 0.6 kcal/mol) while core residue mutations have a stronger destabilising mean ($\Delta\Delta$G $\approx$ 1.4 kcal/mol) and wider distribution (Tokuriki et al. 2007).

1.2 Aims

In this report, experimental and computationally predicted data is used to try and elucidate the importance certain structural features have with regards to protein thermal stability. The commonly available computational tools Missense3D (Ittisoponpisan et al. 2019) and FoldX (Schymkowitz et al. 2005) are used. While Missense3D is designed as a predictor of disease causing variants, its primarily use is for its ability to predict structural changes that occur upon a missense mutation (Ittisoponpisan et al. 2019). Additionally, the capabilities of FoldX as a predictor of changes in thermodynamic stability (Schymkowitz et al. 2005) is assessed by each structural feature that is seen in missense mutants. Also addressed are 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.

2 Methods

The computational workflow used in this project is outlined in Figure 2. Two experimentally determined datasets (to study data bias between the two) and two programs, FoldX (Delgado et al. 2019) and Missense3D (Ittisoponpisan et al. 2019) are used. Note that the "HoTMuSiC" (Pucci et al. 2016*a*) (n=1,626) and "Missense3D" (n=10,229) datasets are used to refer 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 its PositionScan function can predict folding $\Delta\Delta$G. The algorithm uses statistically derived energy potential functions trained on another subsection of the ProTherm database (Schymkowitz et al. 2005). Missense3D produces a binary prediction on whether the variant is *disease causing* and not thermodynamically unstable. This binary prediction does not feature greatly in our results but Missense3D's ability to predict 16 types of structural changes upon point mutation (referred to as "features" in the rest of the text) is used extensively. The exact list of features and how they are defined are explained in Ittisoponpisan et al. (2019). Both algorithms assume rigid mainchains but repack the sidechains. Missense3D uses the SCWRL4 (Krivov et al. 2009) algorithum for sidechain repacking (Ittisoponpisan et al. 2019) whereas FoldX performs this natively (Schymkowitz et al. 2005).

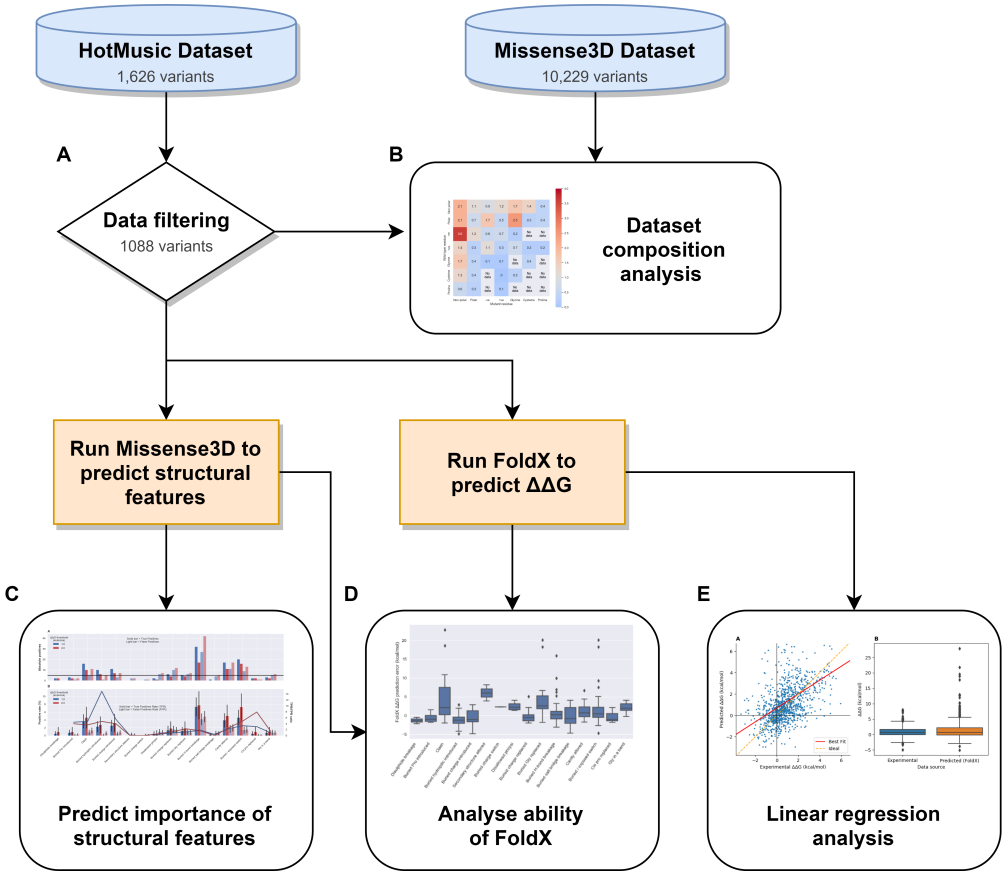

Figure 2: Flowchart demonstrating the workflow used. Note the Missense3D dataset is only used to study bias in our dataset and not to make predictions.

2.1 Construction and analysis of dataset

The HoTMuSiC dataset, more specifically T1626 (Pucci et al. 2016*a*), is mostly a subsection of the widely used ProTherm database (Kumar et al. 2006) but also contains a number of variants screened from the literature. All variants were manually checked by the authors in the original literature to remove errors and criteria were used to ensure only high quality structures and variants were included. A summary of the criteria used is listed below, a full explanation is provided in Pucci et al. (2016*a*).

- 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, additional filtering of the dataset was performed (Figure 2A):

- Any mutation not containing an experimentally measured ΔΔG value was removed (HoTMuSiC was trained only to predict T_m (Pucci et al. 2016*b*)).
- 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 not 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, our dataset contains 1088 variants across 59 protein structures.

Protein mutation databases, such as HoTMuSiC, are compiled from a diverse range of protein stability and mutagenesis 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 the frequency distribution of variants in our dataset was compared to the Missense3D dataset which, being made of naturally occurring disease-causing and neutral variants (Ittisoponpisan et al. 2019), can be thought of as being a relatively accurate representation of the frequency at which different missense variants are observed in nature. For comparison, the ratio between the percentage abundances for every type of missense mutation in both datasets was calculated (Figure 2B). Therefore, a variant type with a ratio >1 is considered over-represented and <1 is considered underrepresented.

The distribution of mutations were described using a reduced amino acid alphabet (Table 1) to increase the number of observations in each category. Additionally, it is predicted 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 Ramachandran space as well as cysteine for its ability to form disulfide bridges.

Class	**Amino Acids**
Non-polar	Alanine, Valine, Leucine, Isoleucine,, Methionine, Trytophan, Phenylalanine
Polar	Serine, Threonine, Tyrosine, Asparagine, Glutamine
Negative charge	Aspartic Acid, Glutamic acid
Positive charge	Lysine, Arginine, Histidine
Specials	Glycine, Cysteine, Proline

Table 1: Reduced amino acid alphabet.

2.2 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 Ittisoponpisan et al. (2019). Analysis of the degree to which these predicted features contribute towards protein stability was conducted in a similar way to as was done in Ittisoponpisan et al. (2019), except here the study considered the effect assuming different ΔΔG thresholds has on the predicted importance of such predictions (Figure 2C). 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 ΔΔG threshold, the number of Total Positives (NP) and Total Negatives (NN) in our dataset was calculated based on whether the experimentally determined ΔΔG values exceeded that particular threshold. Then, for every positively predicted feature in each variant, it was 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 the ΔΔG threshold was varied, the True Positives Rates (TPR) and False Positve Rates (FPR) was also calculated. 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:

$$TPR = \frac{TP}{NP}, \qquad FPR = \frac{FP}{NN} \qquad \text{and} \qquad Ratio = \frac{TPR}{FPR}$$

95% confidence intervals were calculated for the TPRs and FPRs as folows:

$$TPR \pm 1.96\sqrt{\frac{(TPR)(1-TPR)}{NP}} \qquad \text{and} \qquad FPR \pm 1.96\sqrt{\frac{(FPR)(1-FPR)}{NN}}$$

Features with only a small number of observations are not likely to give statically significant results. In addition to using the confidence intervals, a threshold of 5 TPs and FPs needing to be observed for feature observation to be significant is also taken into account.

2.3 Evaluation of stability prediction tools

The ability of FoldX to predict protein stability was evaluated in two ways. Firstly, simple linear regression analysis was performed (determined by least squares approach) between the FoldX predicted and experimentaly determined ΔΔG values (Figure 2E). Pearson correlation coefficients were also calculated. Secondly, which kind of structural alterations (as predicted by Missense3D) FoldX does and doesn't handle well was investigated. To do this, boxplots of the error between the predicted and experimental ΔΔG values ($Error = \Delta\Delta G_{pred} - \Delta\Delta G_{exp}$) were constructed for every Missense3D predicted feature (Figure 2D).

3 Results

3.1 Nature of the dataset

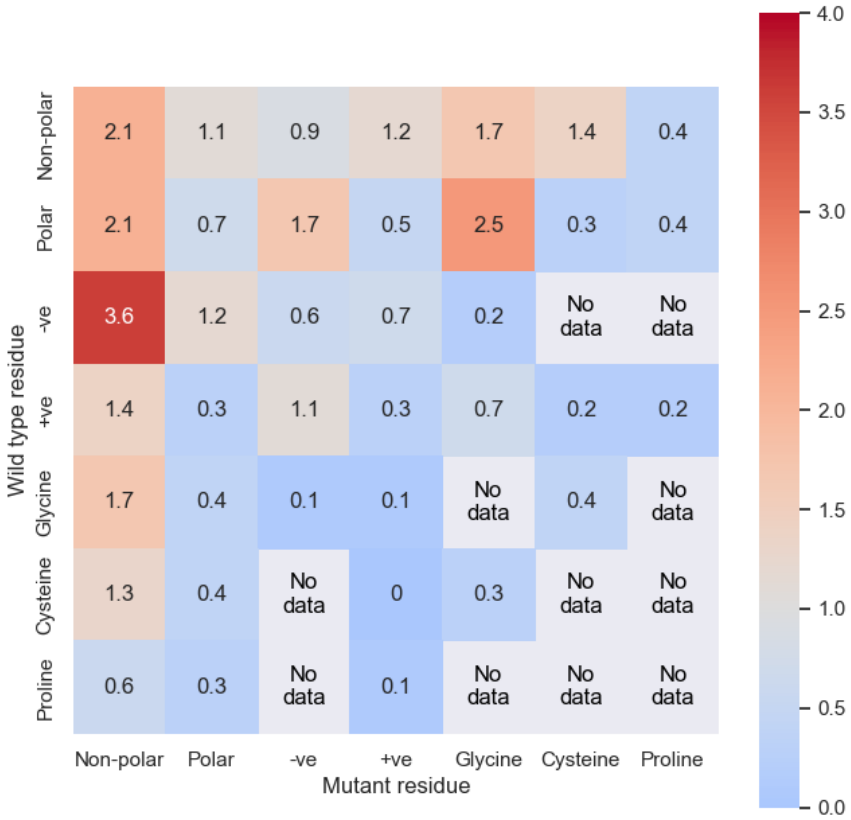

Figure 3: Bias found for certain missense mutations in our dataset compared to that found in the Missense3D dataset. Red values are over-represented variants while blue are under-represented.

The matrix showing the composition of the dataset used is shown in Figure 3. 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 in our dataset. A notable exception to this was residues being mutated to non-polar residues, which were generally over-represented (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 overrepresented 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 only a small number of disulphide breakages were observed (n=2).

3.2 Importance of structural features for protein stability

Figure 4 shows the structural changes predicted by Missense3D and their corresponding impact on protein stability. Shown are the absolute number of positives and the positive rates (Figure 4A) to make it clearer when changes in the both the positive rates and their ratios (Figure 4B) are a result in changes in the number of observations or the balance between deleterious and neutral variants in the dataset. The full results is available in Supplementary information 1.

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 $\Delta\Delta$G thresholds respectively. Steric clashes, cavity alterations and residues being switched between the buried and exposed state were also highly observed with ≥ 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 the TPR/FPR ratio (referred to from now on simply as "the ratio") was unable to be calculated 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 with 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 $\Delta\Delta$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 $\Delta\Delta$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 of 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.

3.3 Accuracy of protein stability prediction using FoldX

Linear regression analysis between the experimentally determined and FoldX predicted dataset (Figure 5A) 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, the axis boundaries in Figure 5A were intentionally restricted. However, there are many variants found outside these bounds, especially many with exceptionally high FoldX predicted $\Delta\Delta$G values (Figure 5B). There was no consensus feature predicted by Missense3D that correlated well to all variants with these high $\Delta\Delta$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 $\Delta\Delta$G greater than 10.0 kcal/mol.

Figure 6 shows the error between the FoldX $\Delta\Delta$G prediction and experimental value 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). The degree to which each feature over- or under-predicted the $\Delta\Delta$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.

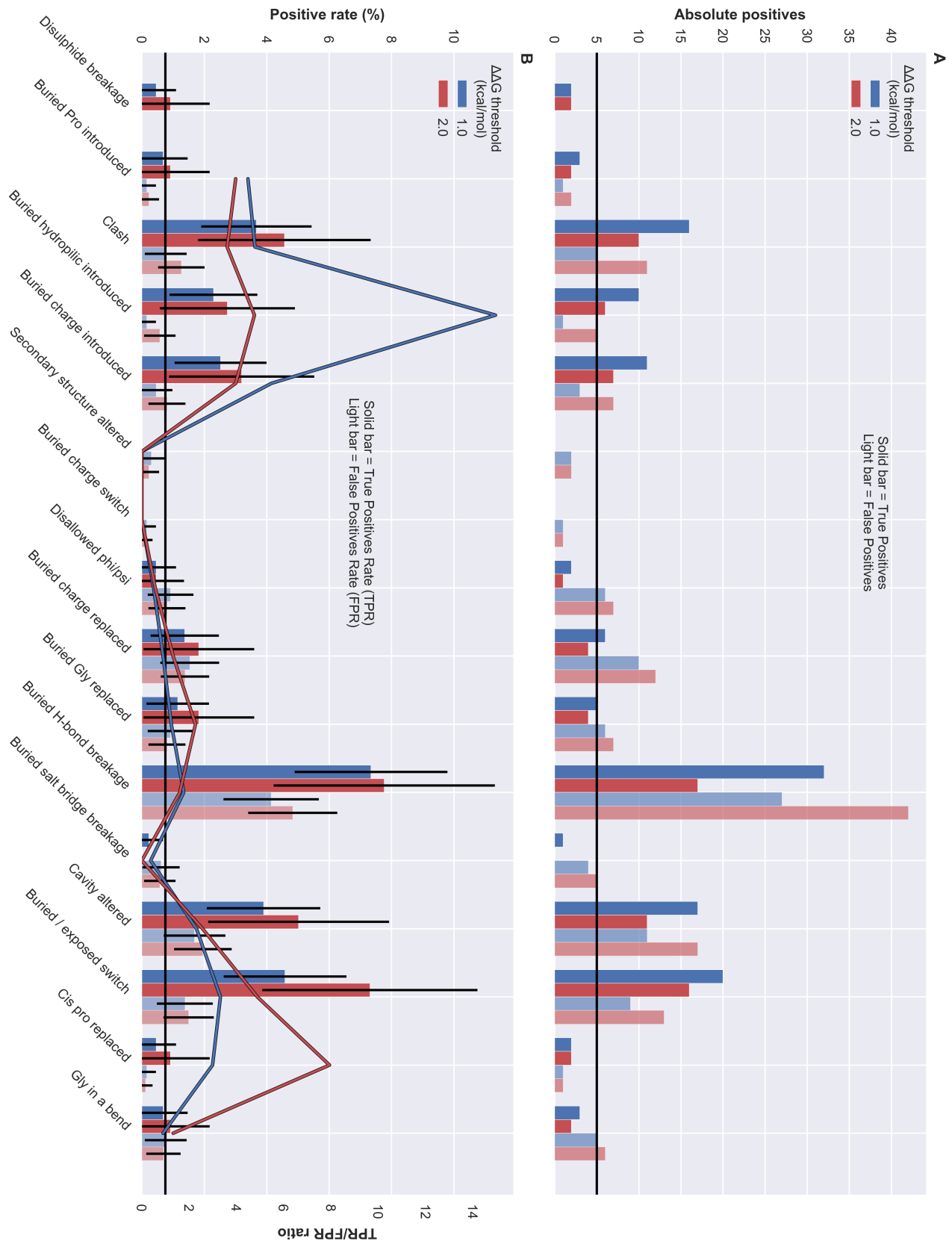

Figure 4: (A) Absolute number of positives observed for each feature at various ΔΔG thresholds. Black line indicates positive rate = 5. (B) TPRs, FPRs and TPR/FPR ratios with respect to ΔΔG thresholds. Error bars denote 95% confidence intervals. Black line indicates TPR/FPR ratio = 1.

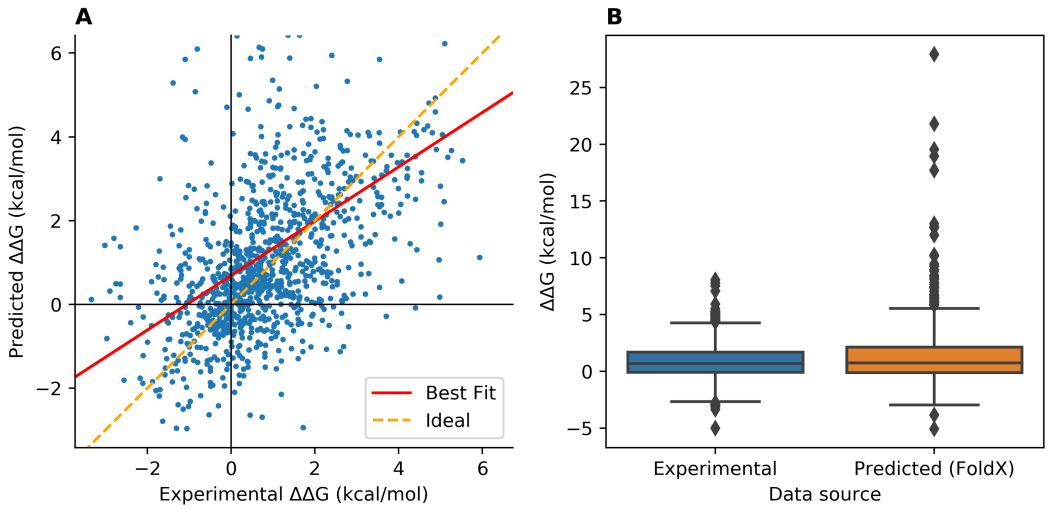

Figure 5: (A) Linear regression analysis between the experimentally determined and FoldX predicted ΔΔG values. Slope = 0.65, y-intercept = 0.69 kcal/mol, Pearson correlation coefficient = 0.41. (B) Box plot of the distribution of both experimentally determined and predicted ΔΔG values.

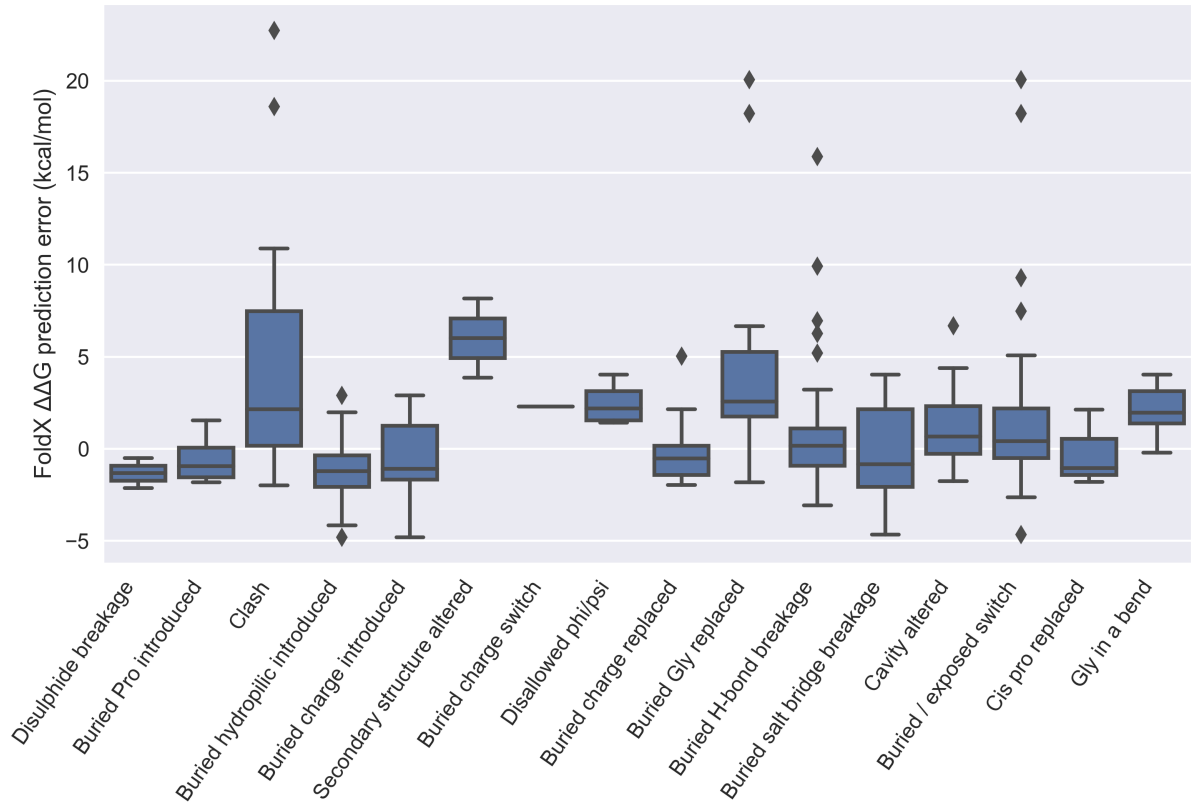

Figure 6: FoldX ΔΔG prediction errors for every Missense3D predicted feature.

4 Discussion

4.1 Importance of structural features on protein stability

The TPR/FPR ratios of Figure 4B not only indicate the importance of the features, but by using the ΔΔG thresholds we get a quantitative idea towards how much each feature contributes to protein

stability. For example, with an assumed ΔΔG 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 (Baker 2019). 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 (Goldenzweig & Fleishman 2018).

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, this cannot be said with any confidence (especially as the feature is observed only 3 times in the whole dataset). Either way, 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 (Nathaniel et al. 2003, Williams et al. 2004, Yohannan et al. 2004).

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 ΔG. 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 (Ittisoponpisan et al. 2019). Not surprisingly, FoldX performed poorly on the 8 variants identified to have this feature, with all being predicted has having a larger ΔΔG than reality by 2.4 kcal/mol on average (± 1.4 kcal/mol for whole dataset). Given our approach, it is not possible 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, it is suggested that an algorithm capable of modelling backbone flexibility such as Rosetta (Rohl et al. 2004) should be used.

The breakage of buried salt bridges had the lowest ratio of 0.4 for 1.0 kcal/mol ΔΔG and no observed true positives for 2.0 kcal/mol. This feature was observed only 5 times with a mean ΔΔG 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) (Anderson et al. 1990), 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 (Pace et al. 2014). 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 it can be assumed that most of the new charges do not 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, it can be seen 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 (Bellissent-Funel et al. 2016), 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. It is a possibility the low abundance of disulfide breakages is a result of the experimental data being taken from protein engineering studies (Montanucci et al. 2019), where the breakage of disulfide bonds is rarely attempted due to the 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. However, Missense3D did identify the breakage of a disulfide bond found in thioredoxin in *E. coli* (PDB: 2TRX) upon point mutagenesis of Cys32 to Ala32 (Figure 7). A disulfide bond in a highly structured region like this (in this case holding a α-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, due to being part of the active site (Katti et al. 1990), Missense3D

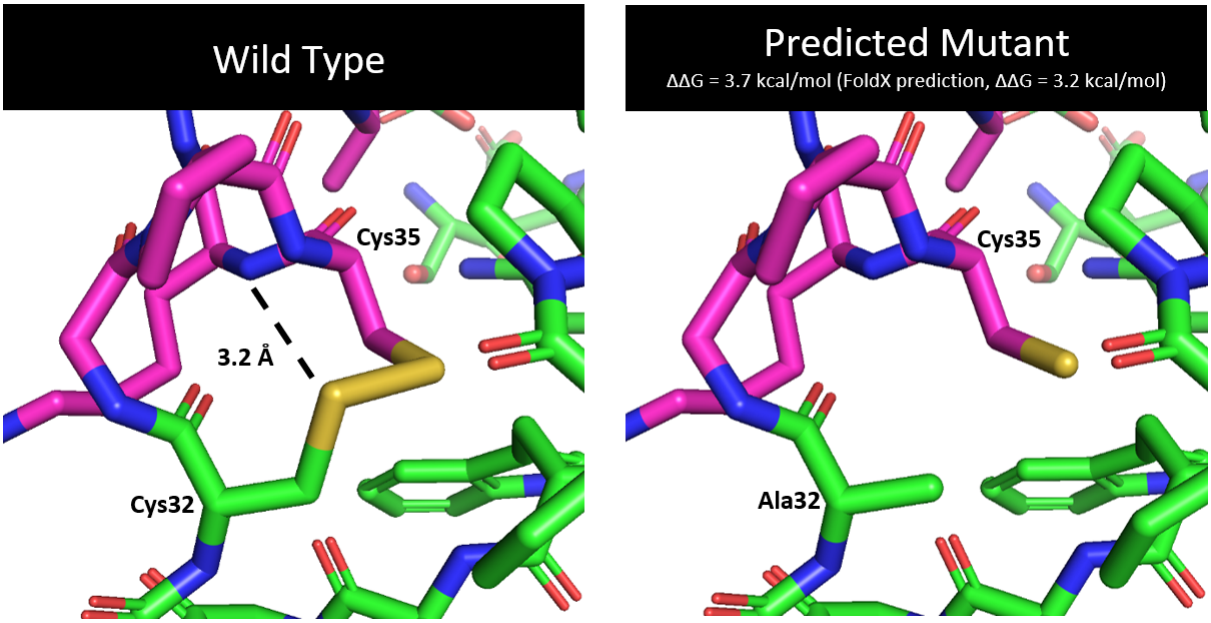

Figure 7: Breakage of disulfide bond between Cys32 and Cys35 causes large decrease in thermodynamic stability, sulfur atom of Cys32 also forms hydrogen bond with mainchain nitrogen 35, purple residues: surface α-helix. Mutant structure predicted by Missense3D.

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 $\Delta\Delta G$ value of 3.7 kcal/mol. Interestingly, mutagenesis of the Cys35 to Ala35 only had a $\Delta\Delta G$ value of 3.1 kcal/mol, suggesting that the remaining Cys32 residue was still able to form stabilising interactions with the mainchain nitrogen.

4.2 Thermodynamic prediction ability of FoldX

Our results have demonstrated that the predictive ability of FoldX varies significantly depending on the scenario. FoldX especially struggles with anomalies where the predicted $\Delta\Delta G$ is significantly higher than the true value (Figure 5B) and most of these tended to be in the same kind of variants (Figure 6). In the case of hydrogen bond breakage and switching buried exposed residues, these large inaccuracies could be the result of the small and biased dataset used to train FoldX (Nisthal et al. 2019) and the use of statistical potentials (Schymkowitz et al. 2005) 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 (Pucci et al. 2016*b*). In the case of glycine replacement and steric clashes, it can be expected that the high $\Delta\Delta G$ predictions are a result of extremely unfavourable energy calculations that arise from the restriction on conformational rearrangement due to the rigid mainchain assumptions (Schymkowitz et al. 2005).

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 8A) 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 $\mathring{A}$ each in the mutant native state with an experimentally measured $\Delta\Delta G$ of 3.2 kcal/mol (Liu et al. 2000). 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 8B). This was especially true in the FoldX predicted strucuture, with a predicted folding $\Delta\Delta G$ of 21.8 kcal/mol, an absolute error of 18.7 kcal/mol.

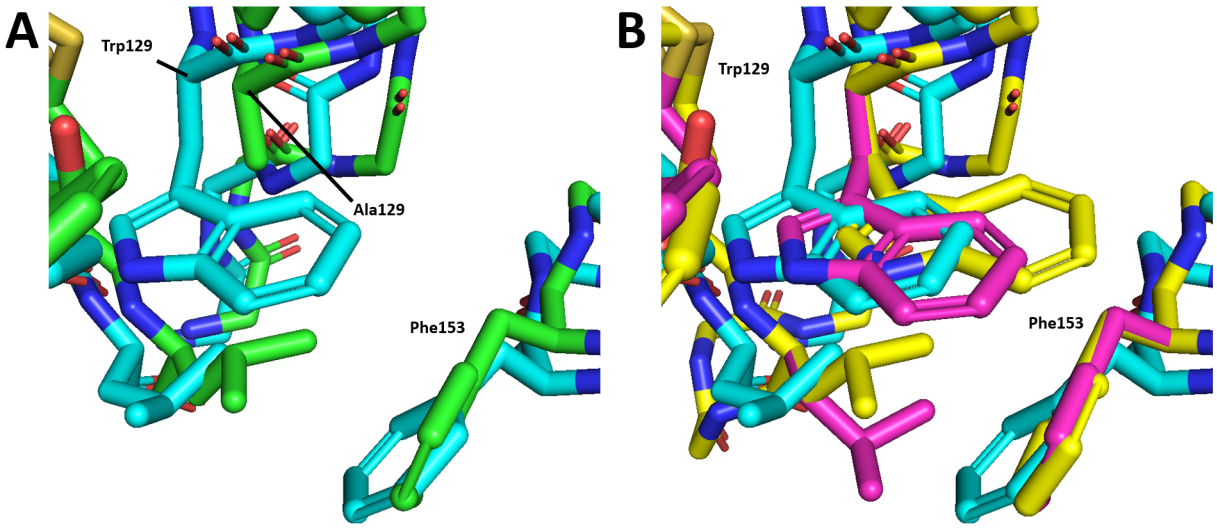

Figure 8: (A) Experimentally determined pseudo-wild-type (Green, 1L63) and mutant (Cyan, 1QTD) structures for single mutant A129W on T4 lysozyme. Notice the large mainchain conformational changed needed to fit the tryptophan side chain in the packed protein core. (B) Experimentally determined mutant (Cyan, 1QTD), Missense3D predicted (Magenta) 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 ΔΔG values predicted to a high degree of accuracy (Figure 6). However, the small number of observations for each of these features means that there is little confidence that this would extrapolate well in studies on a wider (possibly proteome) scale.

While only the final ΔΔG prediction is considered here, the output of FoldX contains energetic breakdowns of the contributions for each type of molecular interaction it models (Schymkowitz et al. 2005). Taking a similar approach as in Figure 6 but instead correlating the prediction errors with the more detailed breakdowns could elucidate the energy functions in FoldX that prove to be most anomalous. The prediction errors by mutant residue type or wild type-mutant pair (similar to the classes seen in Figure 3) could also be investigated to find which mutations FoldX performs most poorly on.

Analysis on another dataset similar to ours found that the SomeMin method within Rosetta performed the best with regards to predicting folding ΔΔG, with a Pearson correction coefficient of 0.64 (compared to the result of using FoldX on our dataset 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 (Nisthal et al. 2019). The other methods assessed by the same author, PoPMuSiC (Dehouck et al. 2011) and FoldX (Schymkowitz et al. 2005), performed less well with r = 0.56 and 0.51 respectively (Nisthal et al. 2019), 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) used to train FoldX (Schymkowitz et al. 2005).

4.3 Data limitations

As demonstrated by Figure 3, a major limitation of our approach is that the mutagenesis studies used to construct our dataset are often trying to engineer increased thermal stability and will actively avoid conducting experiments that they think will be detrimental (Pucci et al. 2018, Usmanova et al. 2018, Montanucci et al. 2019). This makes 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 mutagenesis studies suffered from bias towards certain nucleotide (and therefore amino acid) substitutions due to the low-fidelity DNA polymerases of the time (Vanhercke et al. 2005).

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, the need for variants not to cause a large change in thermal stability (Pucci et al. 2016*b*) was removed, as it was 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 (Delgado et al. 2019, Ittisoponpisan et al. 2019). This 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.

4.4 Limitations of our approach

For every missense mutation within a protein there are many more local environmental factors that can be taken into account such as solvent accessibility and secondary structure. Unfortunately, the highly dimensional nature of our data makes it difficult to convey all the information that is available to explain changes in proteins stability in one figure. For example, mutated residues with a high solvent accessibility and found in disordered regions (instead of regular secondary structures) are likely to have a less detrimental effect on protein stability (Pucci & Rooman 2017, Tokuriki et al. 2007). Additionally, the combination of binary and scalar information does not lend itself to traditional forms of dimensionality reduction (Multiple Component Analysis (Greenacre & Blasius 2006) was attempted but with poor results), although this could be explored more.

Here, only 16 structural features were assessed with our approach but more insights could be obtained by observing a greater range of structural changes. For example, π-π and cation-π interactions were not studied here but have a higher occurrence in thermophilic proteins (Makwana & Mahalakshmi 2015, Gromiha et al. 2002). One investigation showed that stabilisation of the dimer interface in insulin was achieved by the addition of a inter-chain π-π interaction, causing a 150-fold increase in *in vitro* lifetime and a structure virtually identical to that of the wild-type (Rege et al. 2018). Furthermore, it has been reported that cation-π interactions become more stabilising as the protein approaches its melting temperature (Prajapati et al. 2006) so there is certainly interesting dynamics that could be investigated for biopharmaceutical and high-temperature industrial applications.

5 Conclusion

In this work we have tried to elucidate the some of the structural causes for protein instability. While the data limitations have proved it is difficult to reliably quantify the importance of certain features commonly associated with high stability (such as disulphide bonds and buried salt bridges), it has been indirectly demonstrated that protein folding is largely governed thermodynamics over 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, caution should be taken in using FoldX to generate large artifical $\Delta\Delta$G datasets as we have shown its results to be highly anomalous in certain scenarios.

Note that while FoldX and Missense3D have been critiqued for their rigid mainchain assumptions (Schymkowitz et al. 2005, Ittisoponpisan et al. 2019), 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 (Rohl et al. 2004, Pucci et al. 2016*b*, Schymkowitz et al. 2005). It is worth noting that, even with low sequence identity homology models, Missense3D has performed well as a predictor of disease causing mutants (Ittisoponpisan et al. 2019). This means that a significantly larger proportion of the proteome (and structural features) could be analysed if our approach is found to work well on predicted 3D coordinates.

New and high-throughput experimental pipelines, such as deep mutational screening (Fowler & Fields 2014), 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 Nisthal et al. (2019) where they produce and experimentally measure the folding $\Delta\Delta$G for nearly every possible single mutant in the 56-residue β1 domain of the streptococcal protein G (n=935). The key here was that this data was collected over of weeks instead of years with a roughly 20-fold increase in speed (Nisthal et al. 2019). Analysis on this dataset was not performed here as it is likely 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 has 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 (Pucci et al. 2016*b*).
- **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 (Fowler & Fields 2014, Nisthal et al. 2019).
- **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 (Hayashi et al. 2020, Mbaye et al. 2019, Yasuda et al. 2016).
- **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 (Becher et al. 2018).
- **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 (Childers & Daggett 2017).
- **Chaperones** help the cell respond to temperature fluctuations (Zhang et al. 2015, Jia et al. 2016) and accelerates protein evolution by providing a thermodynamic buffer which facilitates the destabilising mutations needed to acquire new functions before compensatory mutations restore stability (Tokuriki & Tawfik 2009).
- **Terahertz time-domain spectroscopy** can be used to measure the changes in water dynamics around a protein molecule at the picosecond scale (Aoki et al. 2016).

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.

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.

Bashirova, A., Pramanik, S., Volkov, P., Rozhkova, A., Nemashkalov, V., Zorov, I., Gusakov, A., Sinitsyn, A., Schwaneberg, U. & Davari, M. D. (2019), 'Disulfide bond engineering of an endoglucanase from penicillium verruculosum to improve its thermostability', *International journal of molecular sciences* **20**(7), 1602.

Becher, I., Andrés-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.

Berezovsky, I. N., Zeldovich, K. B. & Shakhnovich, E. I. (2007), 'Positive and negative design in stability and thermal adaptation of natural proteins', *PLoS computational biology* **3**(3).

Bhattacharya, R., Rose, P. W., Burley, S. K. & Prlić, 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.

Camilloni, C., Bonetti, D., Morrone, A., Giri, R., Dobson, C. M., Brunori, M., Gianni, S. & Vendruscolo, M. (2016), 'Towards a structural biology of the hydrophobic effect in protein folding', *Scientific reports* **6**(1), 1–9.

Capriotti, E., Fariselli, P., Calabrese, R. & Casadio, R. (2005), 'Predicting protein stability changes from sequences using support vector machines', *Bioinformatics* **21**(suppl_2), ii54–ii58.

Capriotti, E., Fariselli, P. & Casadio, R. (2004), 'A neural-network-based method for predicting protein stability changes upon single point mutations', *Bioinformatics* **20**(suppl_1), i63–i68.

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.

Chiang, C.-H., Grauffel, C., Wu, L.-S., Kuo, P.-H., Doudeva, L. G., Lim, C., Shen, C.-K. J. & Yuan, H. S. (2016), 'Structural analysis of disease-related tdp-43 d169g mutation: linking enhanced stability and caspase cleavage efficiency to protein accumulation', *Scientific reports* **6**(1), 1–14.

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**.

Folch, B., Rooman, M. & Dehouck, Y. (2008), 'Thermostability of salt bridges versus hydrophobic interactions in proteins probed by statistical potentials', *Journal of chemical information and modeling* **48**(1), 119–127.

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.

Harms, M. J. & Thornton, J. W. (2013), 'Evolutionary biochemistry: revealing the historical and physical causes of protein properties', *Nature Reviews Genetics* **14**(8), 559–571.

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.

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.

Jo, B. H., Park, T. Y., Park, H. J., Yeon, Y. J., Yoo, Y. J. & Cha, H. J. (2016), 'Engineering de novo disulfide bond in bacterial α-type carbonic anhydrase for thermostable carbon sequestration', *Scientific reports* **6**(1), 1–9.

Karshikoff, A. & Ladenstein, R. (2001), 'Ion pairs and the thermotolerance of proteins from hyperthermophiles: a 'traffic rule'for hot roads', *Trends in biochemical sciences* **26**(9), 550–557.

Katti, S. K., LeMaster, D. M. & Eklund, H. (1990), 'Crystal structure of thioredoxin from escherichia coli at 1.68 å resolution', *Journal of molecular biology* **212**(1), 167–184.

Khan, S. & Vihinen, M. (2010), 'Performance of protein stability predictors', *Human mutation* **31**(6), 675–684.

Kim, T. H., Mehrabi, P., Ren, Z., Sljoka, A., Ing, C., Bezginov, A., Ye, L., Pomès, 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.

Kumar, R., Ariöz, C., Li, Y., Bosaeus, N., Rocha, S. & Wittung-Stafshede, P. (2017), 'Disease-causing point-mutations in metal-binding domains of wilson disease protein decrease stability and increase structural dynamics', *Biometals* **30**(1), 27–35.

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.

Montanucci, L., Savojardo, C., Martelli, P. L., Casadio, R. & Fariselli, P. (2019), 'On the biases in predictions of protein stability changes upon variations: the inps test case', *Bioinformatics* **35**(14), 2525–2527.

Nathaniel, C., Wallace, L. A., Burke, J. & Dirr, H. W. (2003), 'The role of an evolutionarily conserved cis-proline 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.

Niu, C., Zhu, L., Xu, X. & Li, Q. (2016), 'Rational design of disulfide bonds increases thermostability of a mesophilic 1, 3-1, 4-β-glucanase from bacillus terquilensis', *PLoS One* **11**(4).

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.

Parvizpour, S., Razmara, J., Shamsir, M. S., Illias, R. M. & Abdul Murad, A. M. (2017), 'The role of alternative salt bridges in cold adaptation of a novel psychrophilic laminarinase', *Journal of Biomolecular Structure and Dynamics* **35**(8), 1685–1692.

Pecher, P. & Arnold, U. (2009), 'The effect of additional disulfide bonds on the stability and folding of ribonuclease a', *Biophysical chemistry* **141**(1), 21–28.

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.

Plaza del Pino, I. M., Ibarra-Molero, B. & Sanchez-Ruiz, J. M. (2000), 'Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases', *Proteins: Structure, Function, and Bioinformatics* **40**(1), 58–70.

Potapov, V., Cohen, M. & Schreiber, G. (2009), 'Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details', *Protein engineering, design & selection* **22**(9), 553–560.

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.

Protasevich, I., Yang, Z., Wang, C., Atwell, S., Zhao, X., Emtage, S., Wetmore, D., Hunt, J. F. & Brouillette, C. G. (2010), 'Thermal unfolding studies show the disease causing f508del mutation in cftr thermodynamically destabilizes nucleotide-binding domain 1', *Protein Science* **19**(10), 1917–1931.

Pucci, F., Bernaerts, K. V., Kwasigroch, J. M. & Rooman, M. (2018), 'Quantification of biases in predictions of protein stability changes upon mutations', *Bioinformatics* **34**(21), 3659–3665.

Pucci, F., Bourgeas, R. & Rooman, M. (2016*a*), '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. (2016*b*), 'Predicting protein thermal stability changes upon point mutations using statistical potentials: Introducing hotmusic', *Scientific reports* **6**, 23257.

Pucci, F., Kwasigroch, J. M. & Rooman, M. (2017), 'Scoop: an accurate and fast predictor of protein stability curves as a function of temperature', *Bioinformatics* **33**(21), 3415–3422.

Pucci, F., Kwasigroch, J. M. & Rooman, M. (2020), Protein thermal stability engineering using hotmusic, *in* 'Structural Bioinformatics', Springer, pp. 59–73.

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.

Struvay, C. & Feller, G. (2012), 'Optimization to low temperature activity in psychrophilic enzymes', *International journal of molecular sciences* **13**(9), 11643–11665.

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.

Thusberg, J., Olatubosun, A. & Vihinen, M. (2011), 'Performance of mutation pathogenicity prediction methods on missense variants', *Human mutation* **32**(4), 358–368.

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.

Usmanova, D. R., Bogatyreva, N. S., Ariño Bernad, J., Eremina, A. A., Gorshkova, A. A., Kanevskiy, G. M., Lonishin, L. R., Meister, A. V., Yakupova, A. G., Kondrashov, F. A. et al. (2018), 'Self-consistency test reveals systematic bias in programs for prediction change of stability upon mutation', *Bioinformatics* **34**(21), 3653–3658.

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.

Yue, P., Li, Z. & Moult, J. (2005), 'Loss of protein structure stability as a major causative factor in monogenic disease', *Journal of molecular biology* **353**(2), 459–473.

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.

Supplementary information 1

ΔΔG Threshold (kcal/mol)	1.0							2.0						
	Absolute Postives	Absolute Negatives	TPR (%)	FPR (%)	TPR/TPR ratio	TPR 95% confidence interval (%)	FPR 95% confidence interval (%)	Absolute Postives	Absolute Negatives	TPR (%)	FPR (%)	TPR/TPR ratio	TPR 95% confidence interval (%)	FPR 95% confidence interval (%)
Disulphide breakage	2	0	0.46	0.00	N/A	0.63	0.00	2	0	0.91	0.00	N/A	1.26	0.00
Buried Pro introduced	3	1	0.69	0.15	4.49	0.78	0.30	2	2	0.91	0.23	3.97	1.26	0.32
Clash	16	5	3.67	0.77	4.79	1.76	0.67	10	11	4.57	1.27	3.61	2.76	0.74
Buried hydropilic introduced	10	1	2.29	0.15	14.95	1.41	0.30	6	5	2.74	0.58	4.76	2.16	0.50
Buried charge introduced	11	3	2.52	0.46	5.48	1.47	0.52	7	7	3.20	0.81	3.97	2.33	0.59
Secondary structure altered	0	2	0.00	0.31	0.00	0.00	0.42	0	2	0.00	0.23	0.00	0.00	0.32
Buried charge switch	0	1	0.00	0.15	0.00	0.00	0.30	0	1	0.00	0.12	0.00	0.00	0.23
Disallowed phi/psi	2	6	0.46	0.92	0.50	0.63	0.73	1	7	0.46	0.81	0.57	0.89	0.59
Buried charge replaced	6	10	1.38	1.53	0.90	1.09	0.94	4	12	1.83	1.38	1.32	1.77	0.78
Buried Gly replaced	5	6	1.15	0.92	1.25	1.00	0.73	4	7	1.83	0.81	2.27	1.77	0.59
Buried H-bond breakage	32	27	7.34	4.14	1.77	2.45	1.53	17	42	7.76	4.83	1.61	3.54	1.43
Buried salt bridge breakage	1	4	0.23	0.61	0.37	0.45	0.60	0	5	0.00	0.58	0.00	0.00	0.50
Cavity altered	17	11	3.90	1.69	2.31	1.82	0.99	11	17	5.02	1.96	2.57	2.89	0.92
Buried / exposed switch	20	9	4.59	1.38	3.32	1.96	0.90	16	13	7.31	1.50	4.88	3.45	0.81
Cis pro replaced	2	1	0.46	0.15	2.99	0.63	0.30	2	1	0.91	0.12	7.94	1.26	0.23
Gly in a bend	3	5	0.69	0.77	0.90	0.78	0.67	2	6	0.91	0.69	1.32	1.26	0.55