DERIVING STRUCTURAL EXPLANATIONS FOR PROTEIN THERMAL STABILITY

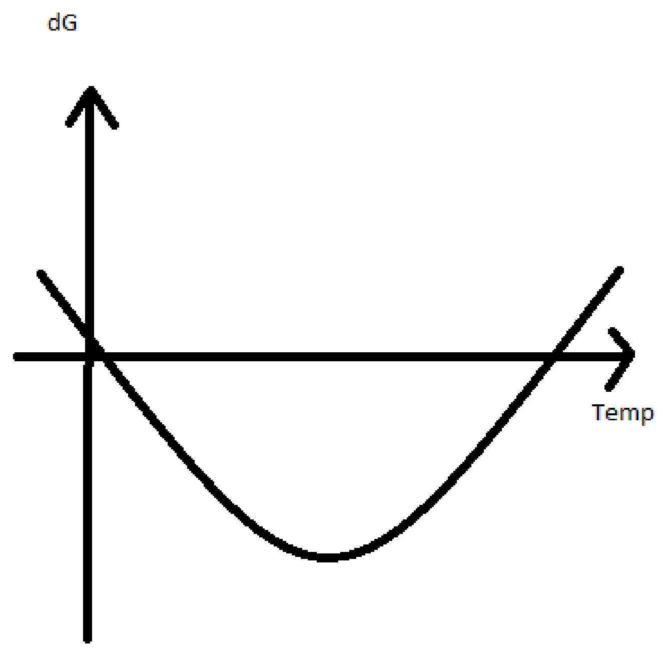
Charles Harris - Department of Life Sciences, Imperial College London

Introduction

Understanding the structural features that give proteins their stability is important for the development and optimisation of new biotechnological and pharmaceutical applications at a range of temperatures.

Unfortunately, despite the large amount of research, there is no clear set of principles that can explain protein stability totally or be used for rational *de novo* protein design. For example, the introduction of disulphide bridges in flexible regions greatly increases stability at higher temperatures but introducing them in highly structured regions has little or no effect. Whilst some features can be identified as being more important than others, any general trends that are observed vary greatly between protein families.

The difference in the Gibbs free energy of the unfolded versus the folded state (ΔG) is a quantifiable measurement of the thermodynamic stability of a given protein, with most proteins in nature having a value somewhere between (FIND SOURCE).



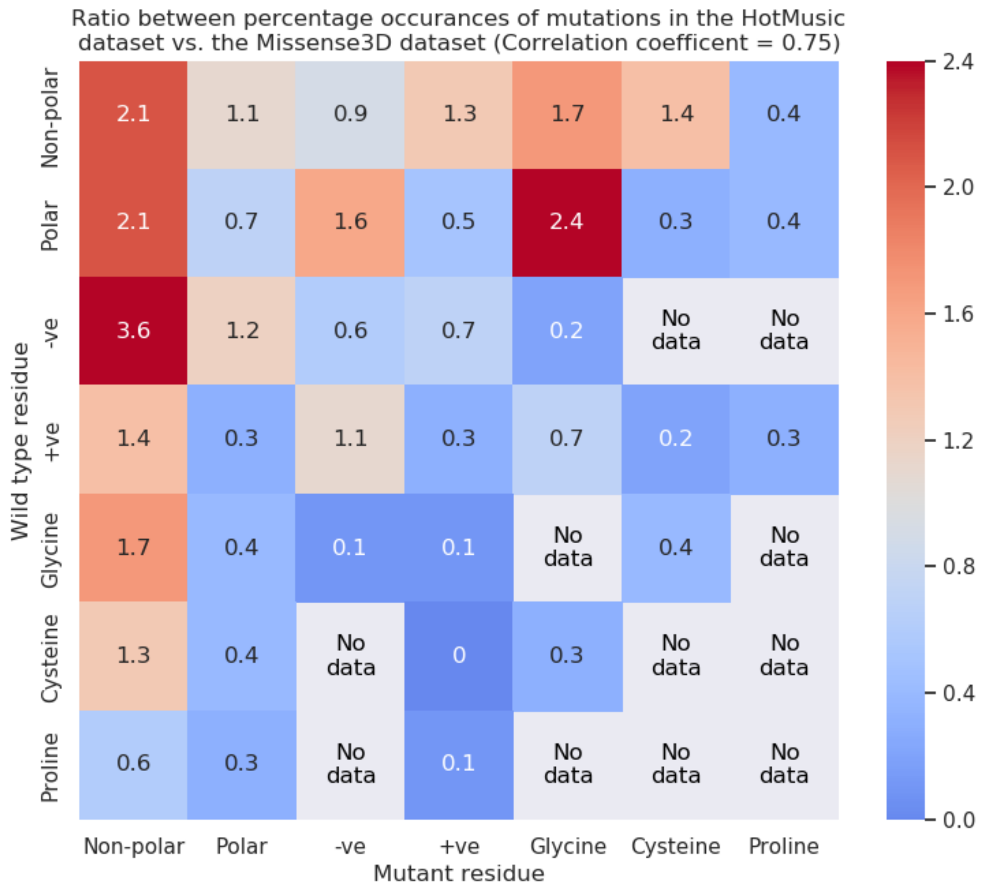
Changes in thermodynamic stability upon a point mutation can be quantified by the change in ΔG (i.e ΔΔG). Unfortunately, there is also no clear threshold value of ΔΔG for which all missense mutations become deleterious, although commonly used threshold range between 1-2 kcal mol^{-1} (check this). Compounding this with the fact that most computational thermodynamic predictive software has high inaccuracy means that we are no close to being able to accurately predict weather missense mutations will cause a disease.

Many datasets exist which contain information about how certain missense mutations affect protein stability. Two are considered here, the training datasets for Missense3D and HotMusic. It is worth making very clear that while they both contain missense mutations, they both measure stability in different ways. Missense3D contains phylotypic outcomes (i.e. deleterious or neutral) and HotMusic contains changes in thermodynamic stability (ΔΔG).

Here we look at...

Methods

As there are 20 amino acids there are 400 possible missense mutations so visual comparison of all of these in not feasible. Therefore, a reduced amino acids alphabet consisting of 7 categories were used instead for easy comparison of missense mutations. The alphabet was constructed as follows: Non-polar are A, V, L, I, M, W and F; Polar are S, T,Y, N and Q; Negative are D and E; Positive are K, R and H. Glycine and proline both have their own category for their altered allowed Ramachandran space as well as cysteine for its ability to form disulphide bridges.

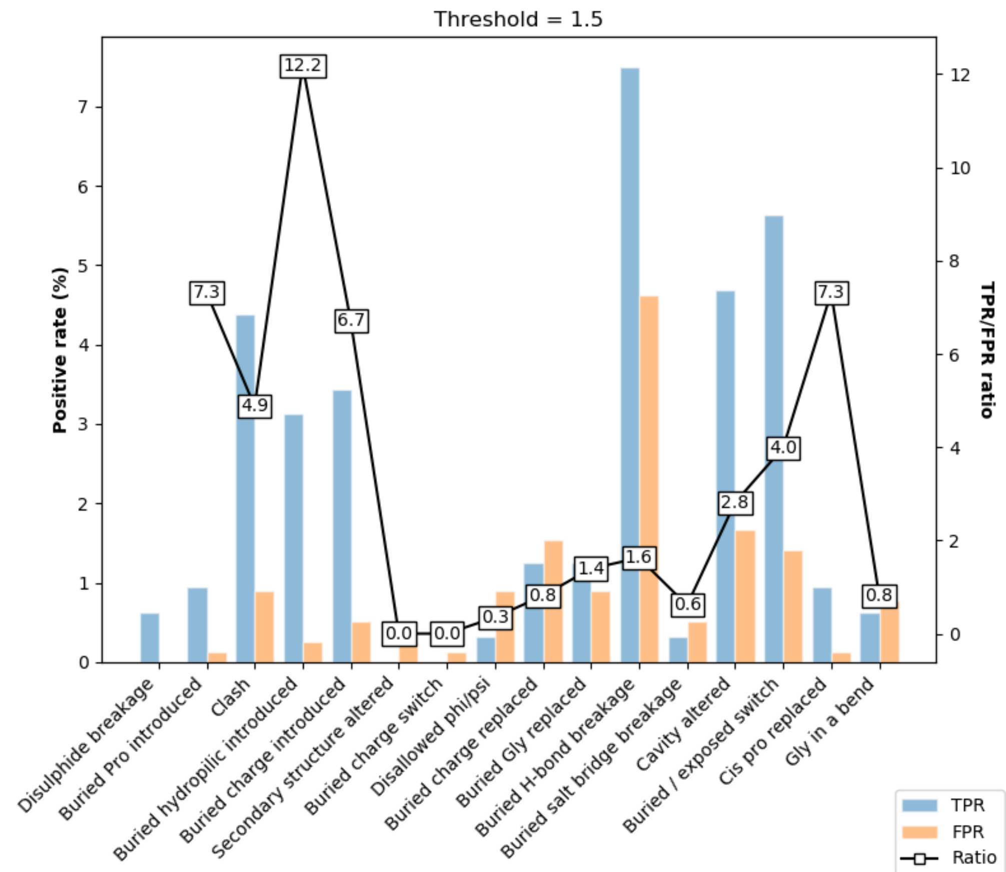


As it is taken from ... the Missense3D dataset can be thought to be representative of missense mutations found in nature. Therefore, the reduced amino acid alphabet was used to compare the differences between the HotMusic data and the Missense3D dataset. Figure 2 shows the ratio of the percentage abundance of certain missense mutations.

While there is a correlation coefficient of 0.75 between the HotMusic and Missense3D dataset, Figure 2 shows that there are many types of mutations bother over and underrepresented in the HotMusic dataset. For example, there is an abundance of residues being mutated to other non-polar residues and non-charged residues being mutated to glycine residues. Additionally, the removal of cysteines (i.e. breakage of disulphide bridges) are mostly under represented.

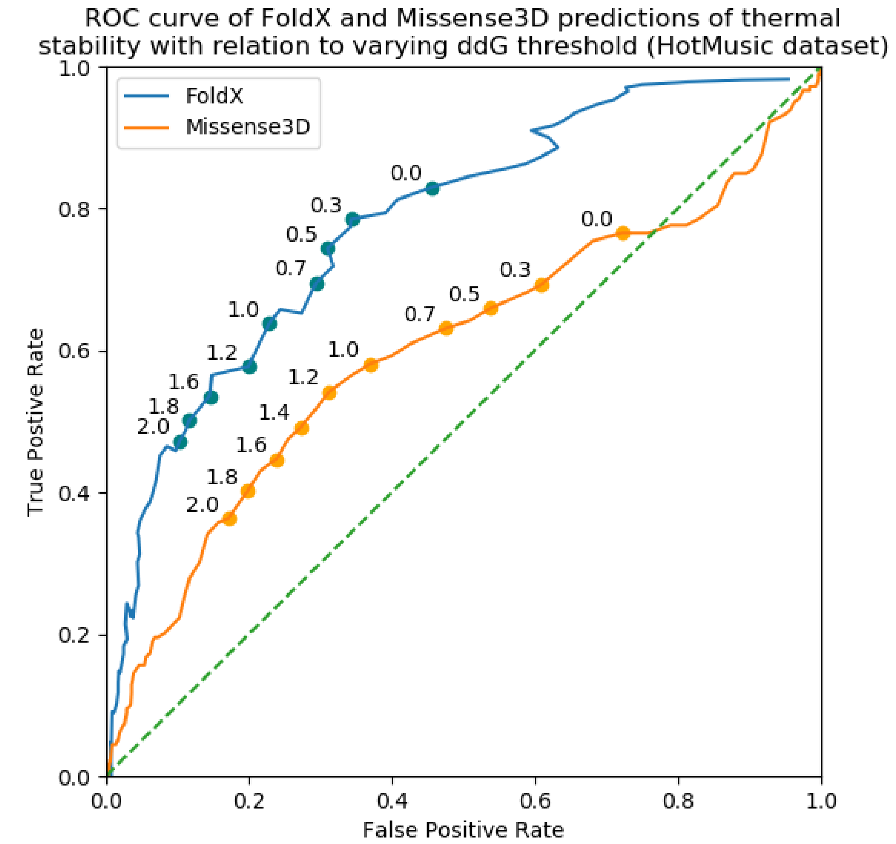
Results

The importance of structural features predicted using Missense3D were assessed in the same way as in Ittisoponpisan et al. However, to assess the importance of certain structural features with respect to varying ΔΔG thresholds, the True Positive Rate (TPR) and False Positive Rate (FPR) bars were omitted and instead multiple TPR/FPR ratios are shown one graph.



Discussion

- Mention importance of protein dynamics
- Can compare more datasets and programs
- Review current state of stability prediction programs/future areas of work



Acknowledgements

References

(1) Pucci, F. and Rooman, M., 2017. Physical and molecular bases of protein thermal stability and cold adaptation. *Current opinion in structural biology*, *42*, pp.117-128.

(2) Ittisoponpisan, S., Islam, S.A., Khanna, T., Alhuzimi, E., David, A. and 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), pp.2197-2212.