DERIVING STRUCTURAL EXPLANATIONS FOR PROTEIN THERMAL STABILITY

Charles Harris, Prof Michael Sternberg, Dr Suhail Islam - 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, there are currently no such universal explanations.

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. As long as ΔG is negative, protein folding will be spontaneous [1].

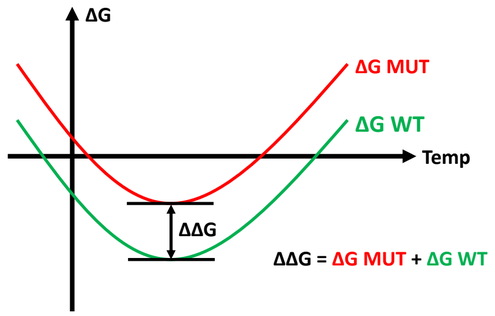

Figure 1: A thermodynamic representation of a deletions mutation.

Figure 1 shows that changes in protein thermodynamic stability caused by a missense mutation can be quantified by the change in ΔG of folding (ΔΔG). Unfortunately, there is also no clear threshold value of ΔΔG at which all missense mutations become deleterious or destabilising, although commonly used values range between 1-2 kcal mol^{-1} [2,3]. Adding to this most computational protein stability predictive software has high inaccuracy meaning that we are not close to being able to reliably predict whether missense mutations will cause a disease.

Aims

- To quantify the importance of certain structural features with regards to protein thermal stability.
- To investigate systematic biases found in protein stability datasets.
- To assess the accuracy of stability prediction programs found in the literature.

Methods

Datasets

So far all the analysis carried was done using the HotMusic dataset. The dataset from Missense3D can be thought of as being representative of all missense mutations found in nature. Figure 2 graphically shows the differences in the representation of mutations between the HotMusic and Missense3D datasets by way of the ratios between percentage abundance.

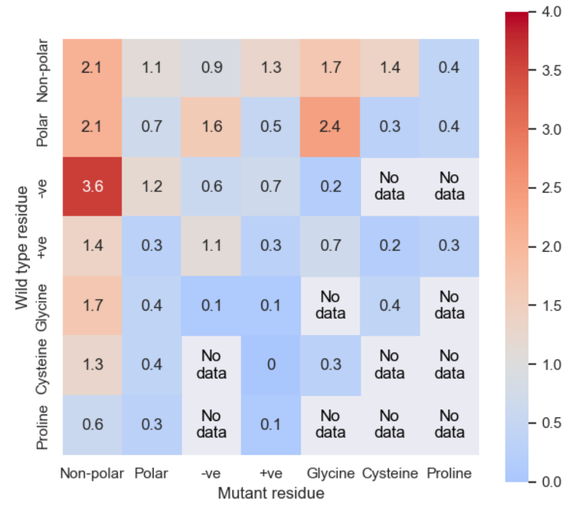

Figure 2: Heatmap representation of missense mutations in the HotMusic training data set.

While there is a correlation coefficient of 0.75 between the two, there are many types of mutations both over and underrepresented in the HotMusic dataset. More data may be required in order to provide insights into the nature of all missense mutations.

Data analysis

The importance of structural changes predicted using Missense3D were assessed by way of True Positive Rate (TPR)/False Positive Rate (FPR) ratios (as done in Ittisoponpisan et al), at a range of ΔΔG thresholds (Figure 3). Evaluation of the popular protein stability prediction programs Missense3D [4] and FoldX [5] was also conducted by way of a Receiver Operating Characteristic (ROC) curve (Figure 4).

Preliminary insights into structural features

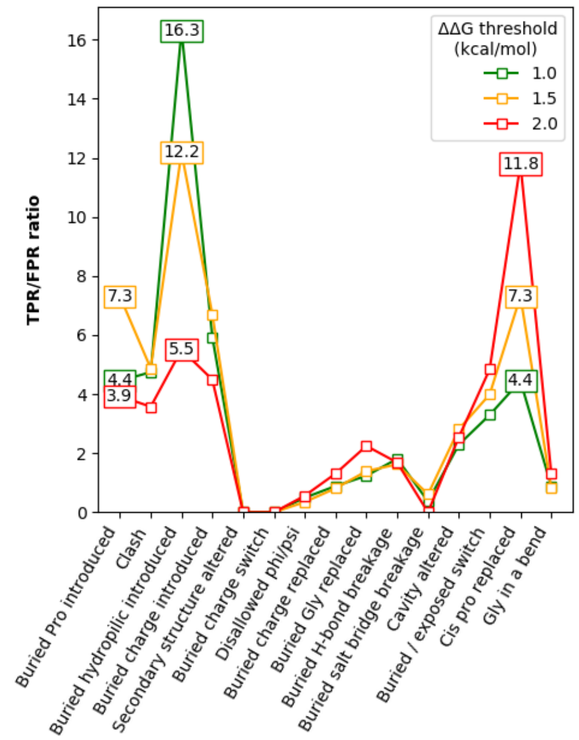

Figure 3: Importance of structural features as predicted by Missense3D.

Key findings:

- The introduction of buried hydrophobic residues seem to have the most detrimental affect on protein stability (especially at low ΔΔG thresholds).
- Removal of cis-proline is also detrimental.
- Results suggest low importance of hydrogen and ionic bonds.

Points going forward:

- The effect of disulphide bond breakage cannot be assessed here as the FPR in all cases was 0.
- Further work is needed to find the statistical significance of these results (as done in Ittisoponpisan et al).
- Both of these points can be addressed by compiling multiple protein stability datasets.

Evaluation of stability predicting programs

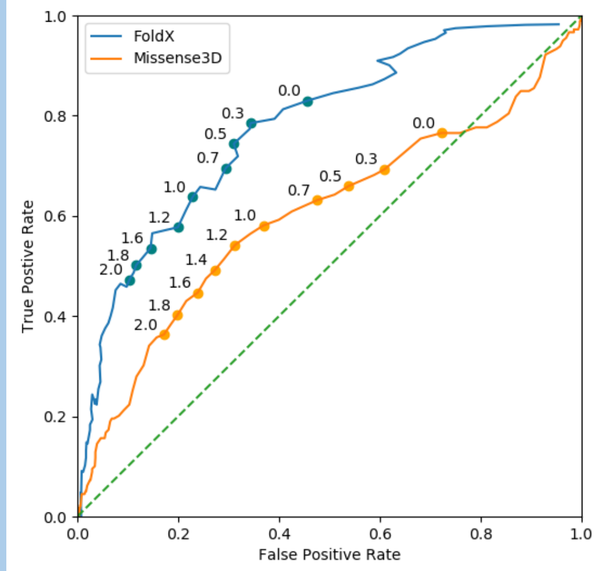

Figure 4: ROC curve evaluating current protein stability prediction programs. Plotted values are ΔΔG thresholds. Ideal location is in top left of graph.

Key points:

- FoldX performed measurably better at all ΔΔG thresholds when compared to Missense3D.
- More programs could be compared in this way if a high-throughput pipeline for their use can be identified.

Acknowledgements

The author would like to thank Dr Tarun Khanna for his help with the statistical work and in the use of the Missense3D program.

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) Bromberg, Y. and Rost, B., 2009. Correlating protein function and stability through the analysis of single amino acid substitutions. *Bmc Bioinformatics*, *10*(8), p.S8.
(3) Capriotti, E., Fariselli, P., Rossi, I. and Casadio, R., 2008. A three-state prediction of single point mutations on protein stability changes. *BMC bioinformatics*, *9*(2), p.S6.
(4) 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.
(5) Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L., 2005. The FoldX web server: an online force field. *Nucleic acids research*, *33*(suppl_2), pp.W382-W388.