

Comprehensive mutagenesis prediction using delaunay tessellation

Shaimaa Sait, Iosif Vaisman.

Abstract— Multiple amino acid substitutions often result in the modified substrate selection and the profiles' activities. The variations with only any of these substitutions are usually less discriminatory, as synergy is critical for specificity.

Clinical Relevance— Mutation in TP53 is found in more than 50% of human cancer, where this mutation affects the protein structure conformation that further impacts the function. DNA-binding Domaine in TP53 is consider where most of the mutation occurred.

I. INTRODUCTION

This paper describes a computational geometry technique based on the delaunay tessellation of protein structure, which explores a statistical potential to calculate residues environmental changes compared to the wild-type. Automute2 is a machine learning-based method using a comparison between wild-type and mutant amino acid. This computation tool predicts the protein stability using different approaches such as energy, activity, thermal change, and disease potential.

Applying a comprehensive mutagenesis on different x-ray crystallography for different structure of TP53 protein.

II. METHODS

Two different structures of Tp53 protein were retrieved from protein data bank (PDB) for comprehensive mutation prediction. The protein structures were analyzed using different computational techniques to investigate the phenotype.

The protein structure's α -carbon coordinates are constrained to the points in three-dimensional space. The points were used as vertices and classified the quadruplets of each amino acid in the protein. The tetrahedral of six edges with a size smaller than 12 Å were used as quadruplets. This algorithm is used to measure the frequency of occurrence (f_{ijkl}) for approximate 8855 quadruplets formed from the naturally occurring twenty amino acids. For the formation of quadruplet, multinominal distribution ($n=4$) is used to measure the expected rate of frequency (q_{ijkl}) A log-likelihood score calculates the event's

tendency for each quadruplet form, which is calculated by equation 1.

$$q_{ijkl} = \log (f_{ijkl}/p_{ijkl}) \quad (1)$$

III. RESULTS

During class-based polarity analysis, the H-H (Hydrophobic to Hydrophobic) polarity class had the highest frequency stability of 20.5 %. While the H-P (hydrophobic to polar) polarity class had the highest frequency thermal of 17.2 %, 19.5% in activity change, and 18.4% in disease potential. (Fig 2). While in structure-location class analysis, the B-B (buried to buried) class showed the highest 50% to 88%. In secondary-structure class analysis, the count stability and thermal, the C-C (coiled to coiled) class exhibited a high count in average 40% to 65% in all the prediction (fig 2).

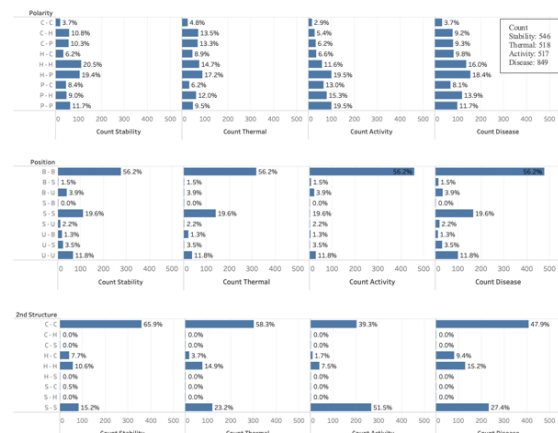


Figure 1. The frequency analysis by different modes of Automute2. (A) shows the frequency analysis of polarity class (B) shows the frequency analysis of structure location class (C) shows the frequency analysis of the secondary class

DISCUSSION & CONCLUSION

In conclusion, there is a compelling argument to be made those statistical approaches are increasingly being used in medicine to forecast disease potential and drug sensitivity.

REFERENCES

- [1] Masso, M., & Vaisman, I. I. (2010b). Knowledge-based computational mutagenesis for predicting the disease potential of human non-synonymous single nucleotide

- polymorphisms. *Journal of Theoretical Biology*, 266(4), 560–568.
- [2] Masso, M., & Vaisman, I. I. (2014). AUTO-MUTE 2.0: A Portable Framework with Enhanced Capabilities for Predicting Protein Functional Consequences upon Mutation. *Advances in Bioinformatics*, 2014, 278385.