Determining the Underlying Distributions of Change in Free Energy Change for Pathogenic and Benign Protein Mutations

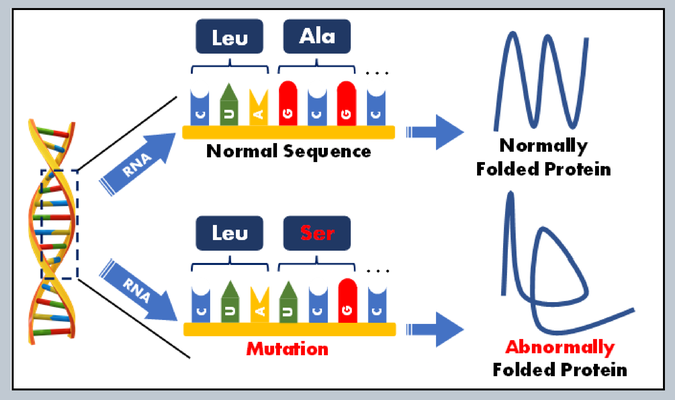
Jorden Thompson[1] (Honors Capstone), Jaie Woodard[1,2], Sriram Chandrasekaran[1,2]

1. Department of Biomedical Engineering, University of Michigan; 2. Department of Computational Medicine and Bioinformatics, University of Michigan

ENGINEERING HONORS PROGRAM
UNIVERSITY OF MICHIGAN

Introduction

A protein mutation can be pathogenic or benign and can affect protein folding or stability. A schematic that shows how the fold can be affected is shown below.



When a protein folds it experiences a change in free energy. A mutation can cause the change in free energy to change, known as a ΔΔG. This can be computationally predicted. A positive ΔΔG is destabilizing, negative is stabilizing. Our goal was to determine if there is a difference between the distributions of pathogenic and benign mutations. Information gathered could one day help predict the pathogenicity of an arbitrary mutation.

Data Analyzed

- ADDRESS from the Yang Zhang Lab contains both benign and pathogenic mutations and the corresponding ΔΔGs for a variety of genes.
- Literature search using PubMed, Google Scholar, and UniProt to identify pathogenic mutations seen in patients, corresponding benign mutations from gnomAD.
- Protein datafiles (PDBs) were obtained from Recon3D, UniProt/AlphaFold, and RCSB.

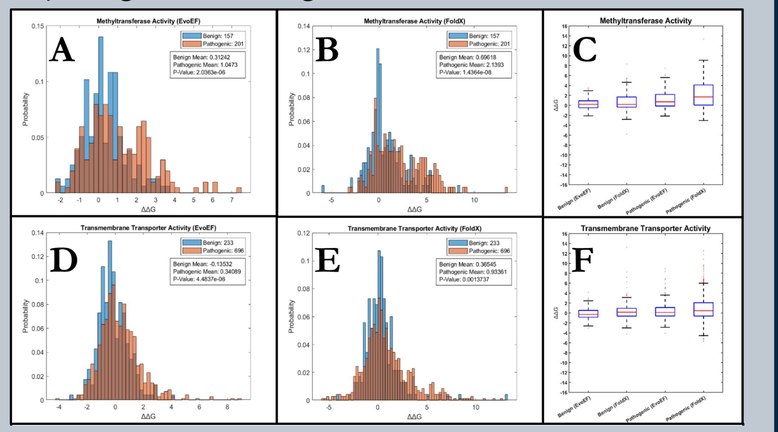
Methods

First, the ADDRESS database genes were organized by GOnet function and analyzed. Next, we obtained mutations from literature and gnomAD, found appropriate PDBs, aligned the PDBs, and simulated the mutations using EvoEF. The resulting distributions for both data sets were analyzed using MatLab by creating histograms, gaussian fits, and statistical tests.

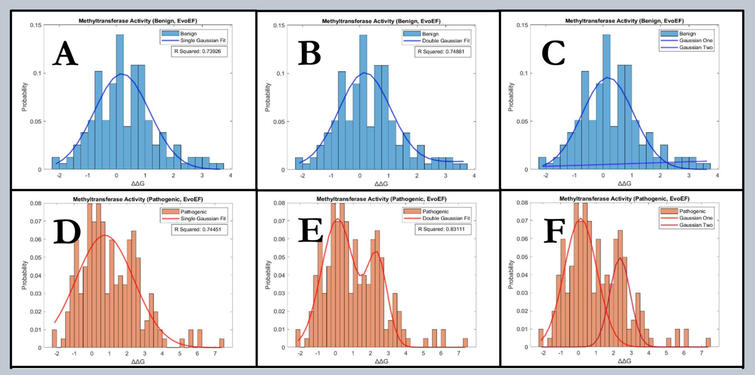
Results – ADDRESS

Out of 37 gene function categories, compared to benign distributions:

- 35 had more positive pathogenic distributions
- 36 had pathogenic distributions with greater standard deviation
- 33 had pathogenic distributions with greater range
- 31 rejected the null hypothesis (p-value < 0.05) between pathogenic and benign distributions



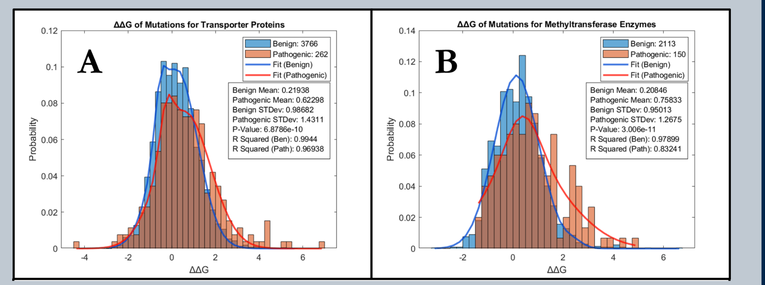
- 26 had pathogenic distributions that were better described by bimodal distributions compared to benign.



Results – Literature

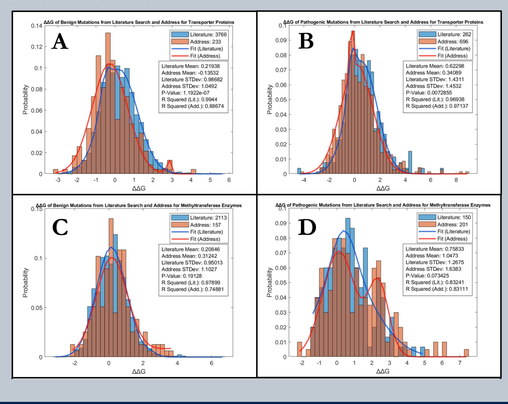
Analyzed transporter proteins and methyltransferase enzymes

- Pathogenic distributions were more positive
- Pathogenic distributions had greater standard deviations
- A t-test between pathogenic and benign distributions rejected the null hypothesis (p-value < 0.05)
- Pathogenic distributions <u>appear</u> to be better described by bimodal distributions visually, but could not confirm



We also compared the literature search distributions to their corresponding ADDRESS distributions

- Methyltransferase activity appeared to match well, accepted the null hypothesis (p-value > 0.05)
- Transporter proteins statistically did not match, visually shared characteristics



Conclusion/Future Steps

Pathogenic distributions appear to be more positive (destabilizing), have greater range and standard deviation (potentially more outliers), and be better described by a bimodal distribution. We believe this may be because pathogenic mutations cause diseases in multiple ways, such as affecting protein function or just by destabilizing the protein.

Moving forward, we can analyze more data and try to determine why pathogenic mutations appear bimodal. This would involve identifying subsets of pathogenic distributions and analyzing how they contribute to the overall distribution. We can also try different fitting techniques, different histogram settings, and different computational estimation techniques. We should also try simulating multiple mutations at a time since disease is typically caused by combinatory effects of multiple mutations.