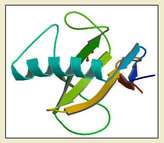
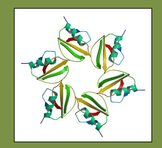
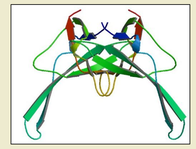
Analyzing the Range of Mutant Variability

Mary Foster, Samantha Loo, Taylor McGilton, Adam Rockett

CSCI 474 Bioinformatics, Filip Jagodzinski, 2019







Introduction

All living things contain proteins made from smaller “pieces” called amino acids. Proteins provide a plethora of roles for an organism, ranging from transporting molecules to enzymes, aiding in biochemical reactions. With proteins providing such critical functions to our existence, it is important to understand how mutations are impacting the proteins ability to carry out its intended function. Through the use of x-ray crystallography researchers have managed to obtain delta delta G values, which present the range of stabilizing or destabilizing effect a mutation has on the protein. Larger delta delta G values correspond to great stabilizing impacts, similarly very negative values coincide with significant destabilizing effects. In order to analyze a variety of mutations that occur on five selected proteins (1BNI, 1VQB, 2CI2, 1WQ5, 1RN1) delta delta G values were collected. These values can then be displayed in a plot in order to determine if there are trends or notable outliers that illustrate substantial impacts on protein stability. Looking closely at the mutant structures in comparison to the wild type provides insight into the structural and functional differences of the proteins.

Motivation

By analyzing the mutations of the proteins we’ll gain results that can help further development in medicine and research. Specifically, we’ll be able to locate trends in what types of mutations are occuring at what residues. With this information we can infer what parts of the protein will have what kinds of functions. We’ll also use the delta delta G values to determine if the mutations that occurred caused the protein to stabilize or destabilize. Combined with our inference of the protein’s functions we can then pinpoint locations where there will be more impact to the protein’s overall functions. Using this information will be helpful in designing drugs to target areas of a protein that will have the most impact while also avoiding areas of low impact or areas that would negatively impact the protein in a way that is undesired.

Methods

We gathered all of the known delta delta G values from all of the different mutations for our five proteins from the ProTherm Database using various bash scripts to extract the necessary information. The ProTherm Database provides experimentally found values that are the most accurate values in comparison to ones that may be predicted using algorithms. From there, we extracted all of the proteins that only had one mutation and their delta delta G values. Once we found all of the single mutation proteins, we looked at the average delta delta G value at each residue location to see if we could notice any trends. We looked for which locations caused the largest destabilization, as well as which locations cause the protein to become more stable. After looking at the single mutations we looked at multiple mutations by creating a script to generate a list of those mutations and ddG values so they could be plotted.

```bash
#/bin/bash

var="1"
count="1"
echo "Results" > results2.txt
thing="PDB_wild 1BNI"
while [ $var -lt 26046 ]
do
echo "...$var"
grep -A 4 -w "NO. $var" 'results wild mutation ddg.txt' > results1.txt

curr=$(grep -w "PDB_wild" results1.txt)
ddg=$(grep -w "ddG" results1.txt)
echo ${curr}
if [ "$thing" == "$curr" ]
then
    echo "$count $ddg" >> results2.txt
    count=$((count + 1))
fi
var=$((var + 1))
done
```

Results

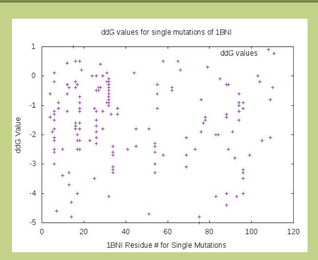


Figure 1. Delta delta G (ddG) values for the protein 1BNI displayed in a plot for each mutation. Mutations are present all throughout the residues. The majority of the mutations remained below the zero threshold.

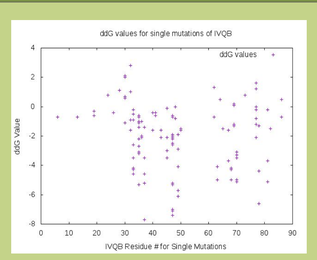


Figure 2. The ddG values for the protein 1VBQ displayed for each mutation residue number. Overall mostly spread between a ddG of -8 and 4, with the highest concentration around -2 to 0.

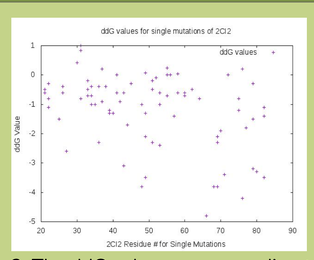


Figure 3. The ddG values corresponding to the residue number of single mutations of 2CI2. Each residue number had a small number of single mutations associated with it with ddG’s ranging from -5 to 1.

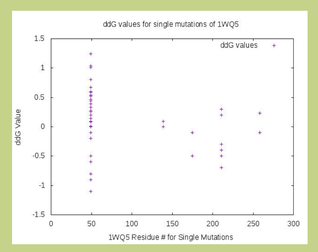


Figure 4. Plot for the protein 1WQ5 of the ddG values for the corresponding residue number at which a mutation has occurred. Mutations remained relatively localized to a couple residue numbers. None of the ddG values are extreme and has more mutations above the zero threshold.

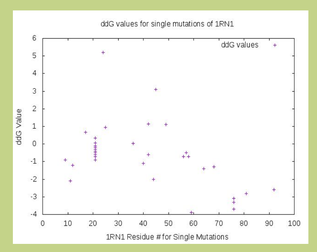


Figure 5. Plot of ddG value vs residue number for the protein 1RN1. The average ddG value for 1RN1 was just below zero and there is a cluster of mutations at the 21st residue.

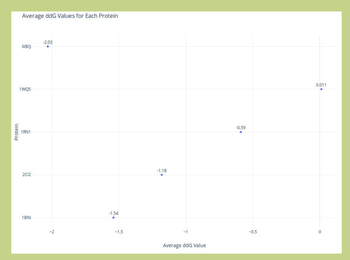


Figure 6. Plot of average delta delta G (ddG) values for single mutations among the chosen proteins. All averages are below zero except for 1WQ5.

Conclusions

Although the mutations were different for each of the proteins an overall trend found was that most mutations had a destabilizing effect on the protein. 1WQ5 was an outlier in that the average of the delta delta G values was just barely above zero, demonstrating that the average of the mutations are slightly stabilizing to the protein. In figure 1 the mutations occur all throughout the residues, signifying that these mutations may not have a critical impact on the protein’s ability to fold correctly. Protein 1VQB in figure 2 has the largest range for ddG values, many extremely destabilizing. Displayed in figure 3, the mutations are not collected in one region or contain a lot at the same residue. Referencing figure 4, mutations arose in confined regions, demonstrating that these regions may not be as essential to the protein folding as other parts. Figure 4 also displays the outlier with an average stabilizing effect of mutations. In Figure 5 the cluster of points at the 21st residue are all mutations from Alanine to different amino acids, suggesting that this residue location is not very essential to the protein, as the ddG values are close to 0. Overall, mutations among the proteins tended to be destabilizing, with mutations to IVQB being the most destabilizing as seen in figure 6. 1WQ5, which became more stable on average after mutation would be beneficial the protein’s ability to function.

Reference

Kumar, M. D., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H., & Sarai, A. (2006, January 01). ProTherm and ProNIT: Thermodynamic databases for proteins and protein-nucleic acid interactions.