Tea Freedman-Susskind

PI Mentor: Dr. Vanessa Jonsson

PhD Mentor: Natalie Dullerud

X July 2020

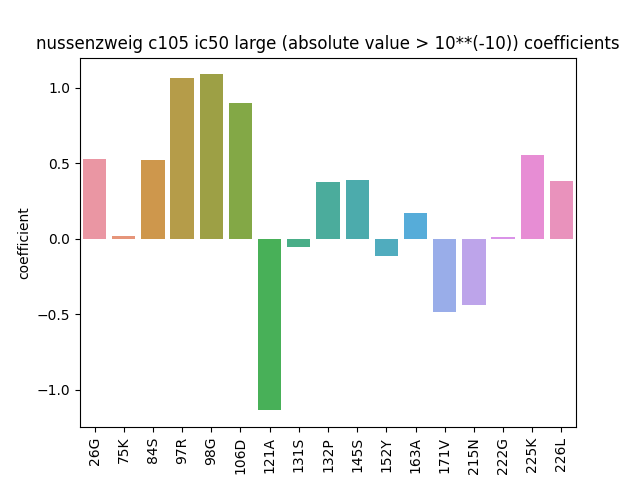
Second Interim Report

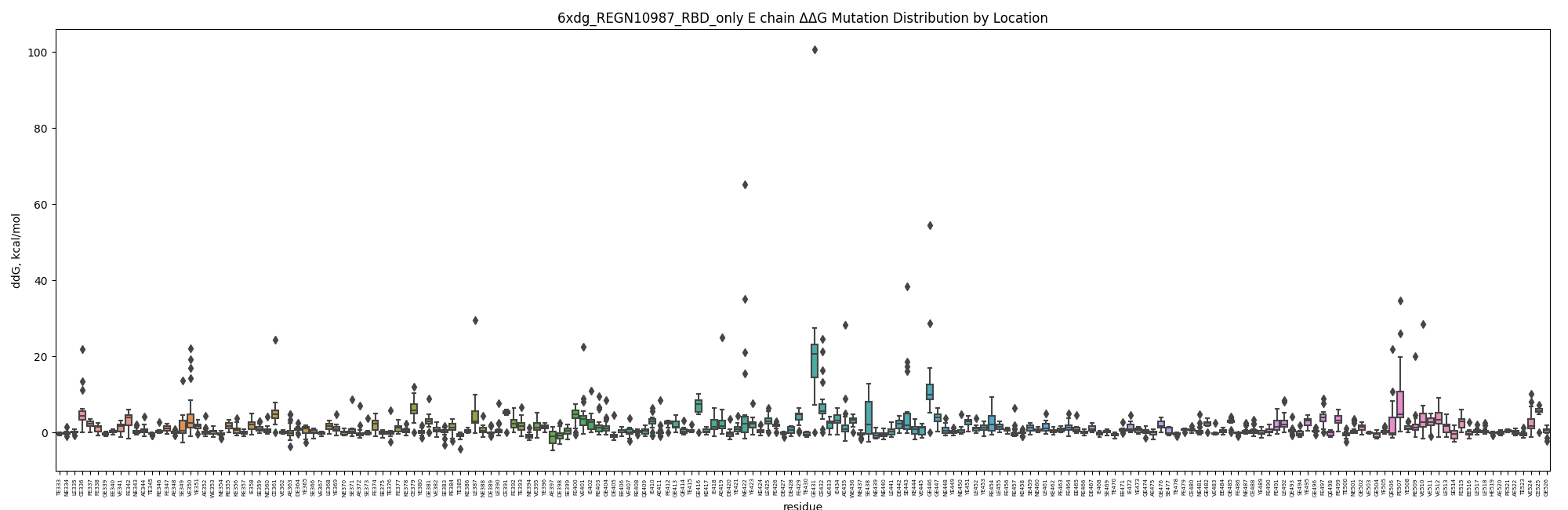
**Progress**

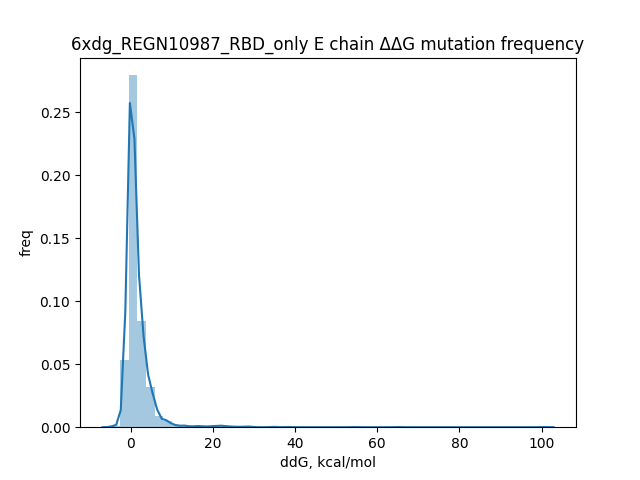
I tried to compare my calculated ΔΔG data with yeast-derived in vivo data to demonstrate its accuracy, but instead this just revealed underlying issues with how I was calculating ΔΔG—namely, that I was using a function that output ΔG instead of ΔΔG. I ran the function which produces ΔΔG data on old and new structures, all of which are SARS-CoV-2 spike protein fragments including its residue binding domain (RBD) bound to a variety of antibodies as well as the human ACE2 receptor. I used this analysis to identify aberrant mutations which either greatly enhance or detract from the bound complex’s stability. In addition, I furthered some analysis my co-mentor had undertaken of RBD mutations’ effects on the minimum concentration of that antibody necessary in vitro to neutralize half of spike proteins, the IC50 value, by plotting mutations of interest in a bar graph and analyzing them within the complex’s structure.

**Methods**

I have continued to write scripts in Python to analyze data generated through FoldX, a Terminal-based program that simulates mutations and expected properties thereof using an energy minimization algorithm. In pairing simulated data with in vivo data from experiments in yeast, it became clear that there was a problem with my data, which proved to be the use of ΔG values from FoldX’s BuildModel function instead of ΔΔG values, which could be had from the PositionScan function (Starr et al). I proceeded to run all the structures I had previously collected ΔG data from through PositionScan to get ΔΔG, and reanalyzed the structures based on the now correct values. I used this opportunity to clean up my Python analysis scripts, and automate mutation running. In addition, while troubleshooting an especially slow mutation run, I discovered that PositionScan had been repairing the original protein file with each mutation, and so added in a repair command prior to running the PositionScan. I also added to the library of graphs generated using the Seaborn Python library for each protein file, creating a box plot and frequency plot for each chain in the protein structures and implemented automatic saving. I used some of these graphs to analyze data generated by my co-mentor on mutational improvements to and detractions of IC50 neutralization data for the C105 antibody, and explored its location in the structure using YASARA (cite 6xcm).

  
Figure 1|Bar plot of large values of coefficients relating impact of mutations to the C105 antibody to neutralization potency.

  
Figure 2|Box plot of Δ Δ G of mutations at each location of the RBD complexed with a REGN10987 antibody.

  
Figure 3|Frequency plot of Δ Δ G of mutations of the RBD complexed with a REGN10987 antibody.

**Challenges**

The speed that mutation analysis runs at has proved a major limiting factor in my research. I alleviated part of this problem by repairing protein files prior to running the analysis, streamlining the process because a file repair job is run automatically with each mutation for an unrepaired file. I also set the FoldX task to the highest priority on my machine.

Another challenge has been fitting my calculated ΔΔG values with data from in vivo experiments. Investigating this problem led me to realize that I had been using the wrong values (ΔG) and the wrong FoldX function for my purposes. Additionally, I had been pairing data from the wrong protein file—the spike protein’s residue binding domain (RBD) complexed with a B38 antibody—instead of the file with a comparable complex, that of the RBD and the human ACE2 receptor.

**Future Work**

Before the end of the SURF program, I would like to finish running FoldX analysis to garner ΔΔG mutation data for all relevant complexes. I would like to pair this data with relevant in vivo data to demonstrate the technique’s effectiveness, and pick out and explore mutations of interests. I also hope to identify binding regions between antibodies and the spike protein’s RBD through analysis of distance between the molecules, and pair this distance analysis with ΔΔG analysis.