Tea Freedman-Susskind
PI Mentor: Dr. Vanessa Jonsson
PhD Mentor: Natalie Dullerud
8 July 2020
First Interim Report

Motivation

A plague of unprecedented scale in living memory, the COVID-19 pandemic has touched every aspect of life in 2020. Enormous quantities of resources have been devoted to treating it and preventing its spread, as well as to spur on a maelstrom of research and development aimed at understanding the novel SARS-CoV-2, creating more effective treatments, and preventing its deadly spread via a safe and reliable vaccine. Several studies have shown that COVID patients can produce antibodies that neutralize SARS-CoV-2 (Baum et al., Robbiani et al.). Isolating these antibodies and developing an antiviral therapeutic is a promising approach toward virus control. However, due to rapid mutations characteristic of viral pathogenesis, there exists the potential for antibody resistance. A crucial step along the path to robust treatments and useful vaccines is a thorough understanding of possible mutations of the virus and how that impacts the effectiveness of antibodies.

We know that the virus is mutating, specifically its spike (S) glycoprotein region, which mediates attachment and fusion to human cells and is a major antibody target (Korber et al.). A significant mutation (albeit one not in the following analysis) include the D614G mutation, with likely origins in Germany. This mutation occurs in the residue binding domain (RBD) of the spike protein and has quickly become the dominant form of SARS-CoV-2 in every region it appears. Mutations on the viral genome have the potential to change virus infectivity and/or antibody neutralization: a thorough characterization of their impact on infection (through ACE2 receptor binding) and antibody neutralization is essential for the development of any robust treatment or vaccine (Starr et al.) Knowing which parts of the S protein are unlikely to change enables the design of antibodies targeting these areas or combination antibody treatments that are likely to be useful despite likely mutations (Starr et al.). Understanding antibody binding favorability aids in vaccine development and improving potency (Barnes et al.).

Existing studies have used myriad experimental methods to analyze S protein mutations, such as in vitro analysis of yeast display technologies (Bloom) or structural analysis of molecular models generated with electron microscopy (Barnes et al.). Our lab's goal is to add to this knowledge using criteria of viral fitness developed in my mentor's thesis work on another virus, HIV (Jonsson). Through a combination of computational chemistry and machine learning techniques, we aim to develop analytic methods to assess viral fitness in terms of both infection and antibody neutralization in the presence of mutations. Using energy minimization tools on molecular structures, we calculate differences in binding energies caused by mutations on the viral protein/cellular receptor and viral protein/antibody and estimate their impact on infectivity and neutralization. Specifically, we calculate the change in Gibbs Energy ($\Delta\Delta G$) due to mutations and compare the difference between these $\Delta\Delta G$ values for the virus bound to the antibody and the virus bound to the human cell. Applying this analysis to the SARS-CoV-2 virus will yield useful information with potential application to analyzing neutralization potential of antibodies as well as to bolster the body of knowledge surrounding structural mutations of SARS-CoV-2.

My Role

My role on this project is to characterize changes in binding energy due to mutations on viral proteins and antibodies. To do this, we use FoldX to calculate energy data for mutations of the S glycoprotein complexed each with antibodies and the human ACE2 receptor, validate the FoldX calculations with available data, and help develop a machine learning model to identify parts of the antibody structure that could be modulated to enhance neutralization of SARS-CoV-2.

Progress

Literature search

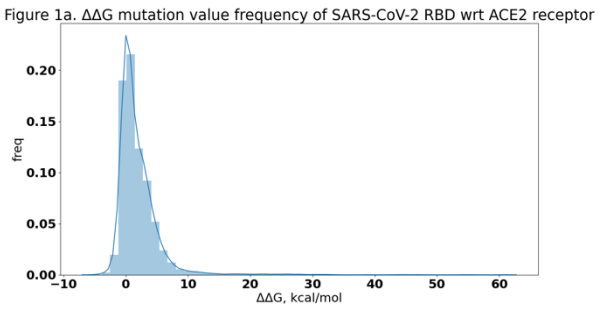
I began my project by familiarizing myself with the literature I would draw on and software I would use during the rest of the project. I read several papers on mutations on the spike glycoprotein of SARS-CoV-2, from early papers identifying commonly observed mutations to later ones pinpointing the structural effects thereof. I also experimented with and did a tutorial for the YASARA protein visualization software and explored the FoldX terminal pipeline for mutation energy analysis.

Binding energy calculations

After these initial steps, I began a phase of calculating the mutation data by writing some Python code to process the protein files into a format indicating which mutations to perform able to be passed into a FoldX command. I ran these mutation files through FoldX to calculate the ΔΔG of these mutations. I am currently using the seaborn Python package coupled with various data frames to produce graphs to visualize the data and inform the next steps of the project.

Results

In order to analyze the effect of mutating the S protein on cell infection or neutralization by antibodies, we generated all possible amino acid point mutations (methods, code availability), on two molecular structures publicly available in the RCSB protein data bank. Based on the protein sequence, we generated all possible amino acid point mutations and formatted it into a format acceptable as an input to FoldX, then used FoldX to calculate ΔΔG for two different structures, one of them the RBD complexed with the ACE2 receptor, PDB 6VW1, and the other being the RBD complexed with an Antibody, PDB 7BZ5 (Shang et al., Wu et al.).



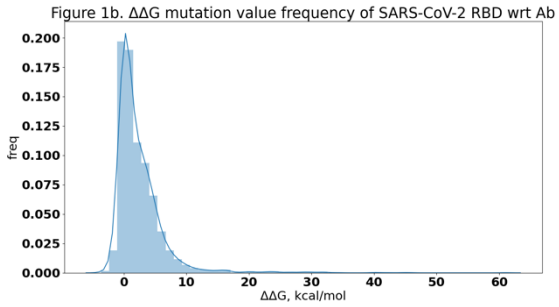


Figure 1 | Distribution of Gibbs binding energies based on FoldX calculations. A. Differences in Gibbs energies due to mutations of RBD for virus bound to ACE2 receptor (PDB: 6vw1). B. Differences due to mutations of RBD for virus bound to B38 antibody (PDB: 7BZ5).

We calculated the distribution of Gibbs energy differences based on mutations of RBD bound with ACE2 receptor and RBD bound to a B38 neutralizing antibody (fig 1ab). We found that most mutations had a ΔΔG over 0, meaning they are deleterious to binding.

To characterize regions of the viral genome that could potentially cause higher infectivity or lower antibody neutralization, we used box plots to visualize differences in Gibbs energies at every genome location of RBD bound with ACE2 receptor (fig 2abcd) and RBD bound to antibody B38 (fig 3ab). Preliminary analysis shows two notable locations. In the complex of two chains in the S trimer and the ACE2 receptor, K458 emerges with a significantly higher minimum, mean, and maximum ΔΔG of its mutations than any other location. The wild type residue is highlighted in orange in Figure 4a. We find a similarly large outlier in the S protein complexed with the B38 neutralizing antibody in the location G502, wildtype highlighted in orange in Figure 4b. A drastically high ΔΔG value means that any mutation at this location will be very deleterious to binding, which means that a mutation on the S protein at G502 would negatively impact the utility of the B38 neutralizing antibody and a mutation on the S protein at K458 would cripple the ability of SARS-CoV-2 to infect via the ACE2 receptor.

Conclusion

Even the very cursory initial analysis has revealed a wide range of changes in binding stability due to point mutations. It has also shown that most mutations are detrimental to binding for both the spike complexed with a B38 neutralizing antibody and with an ACE2 receptor.

Future Work

We will continue to analyze these initial findings. Next steps involve comparing these theoretical values to available experimental data, namely, data from Starr et al. and Baum et al.

Methods

Protein Data Availability

SARS-CoV-2 RBD structure complexed with the human ACE2 receptor from Shang et al (PDB 6vw1, https://www.rcsb.org/structure/6vw1) and the SARS-CoV-2 RBD structure complexed with an Antibody from Wu et al (PDB 7bz5, https://www.rcsb.org/structure/7BZ5) were used for this study.

FoldX calculations

FoldX is a protein energy minimization tool run through the Linux terminal. FoldX can take in a Protein Data Bank (PDB) file as well as a list of desired mutations and calculate the new ΔΔG of binding.

Software

Original software developed for this project is written in Python. It can read in a PDB file and output a FoldX-friendly formatted list of desired mutations to calculate ΔΔG for, then uses returned data to create graphical representations of Gibbs energy distributions. Software will be posted to GitHub and will be publicly available upon project completion.

Challenges

This project will face several challenges. Fundamentally, for many reasons, data availability might be limited. Additionally, FoldX only accurately predicts single mutations so analysis is limited to single mutations and not multiple coexisting mutations, which are known to occur (Korber et al.). On a personal level, I suspect that my lack of biology knowledge might pose a problem as the project progresses.

References

Barnes, Christopher O., et al. “Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies.” *Cell*, 2020, doi:10.1016/j.cell.2020.06.025.

Baum, Alina, et al. “Antibody Cocktail to SARS-CoV-2 Spike Protein Prevents Rapid Mutational Escape Seen with Individual Antibodies.” *Science*, 2020, doi:10.1126/science.abd0831.

Jonsson, Vanessa D. “Robust Control of Evolutionary Dynamics.” *California Institute of Technology*, 2016.

Korber, B, et al. “Spike Mutation Pipeline Reveals the Emergence of a More Transmissible Form of SARS-CoV-2.” 2020, doi:10.1101/2020.04.29.069054.

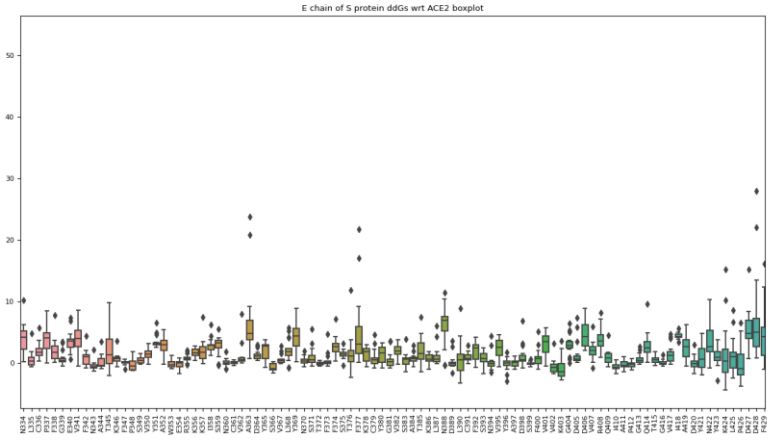
Robbiani, Davide F., et al. “Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals.” *Nature*, 18 June 2020, doi:10.1038/s41586-020-2456-9.

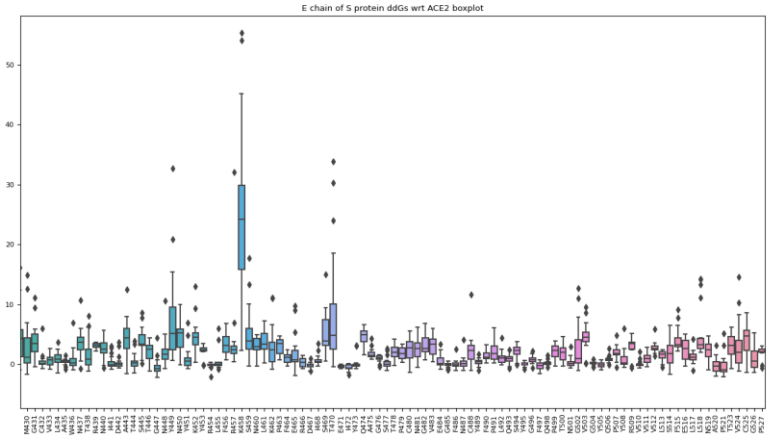
Shang, J., et al. “Structure of SARS-CoV-2 Chimeric Receptor-Binding Domain Complexed with Its Receptor Human ACE2.” 4 Mar. 2020, doi:10.2210/pdb6vw1/pdb.

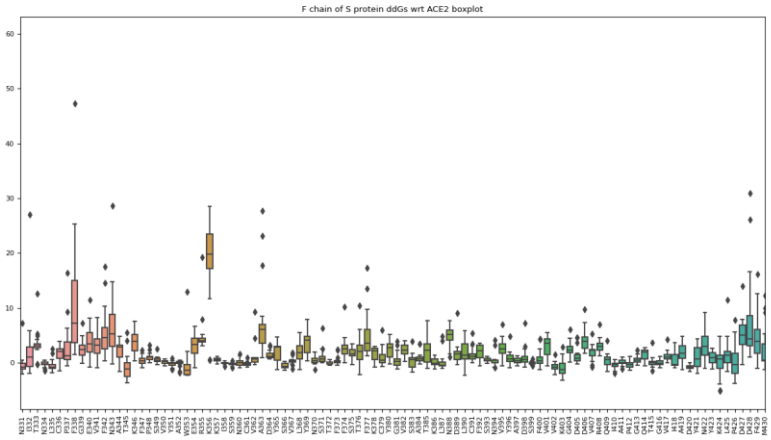
Starr, Tyler N., et al. “Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding.” 2020, doi:10.1101/2020.06.17.157982.

Wu, Y., et al. “Structure of COVID-19 Virus Spike Receptor-Binding Domain Complexed with a Neutralizing Antibody.” 2020, doi:10.2210/pdb7bz5/pdb.

Additional figures


E chain of S protein ddGs wrt ACE2 boxplot


E chain of S protein ddGs wrt ACE2 boxplot


F chain of S protein ddGs wrt ACE2 boxplot

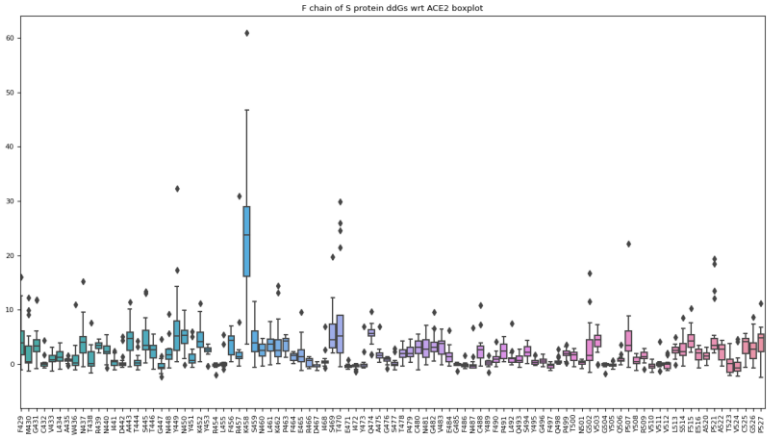


Figure 2 | Range of Gibbs energies at each location on two chain fragments of the S protein RBD complexed with the human ACE2 receptor (PDB: 6vw1). AB. Range of Gibbs energies at each location for chain E. CD. Range of Gibbs energies at each location for chain F.

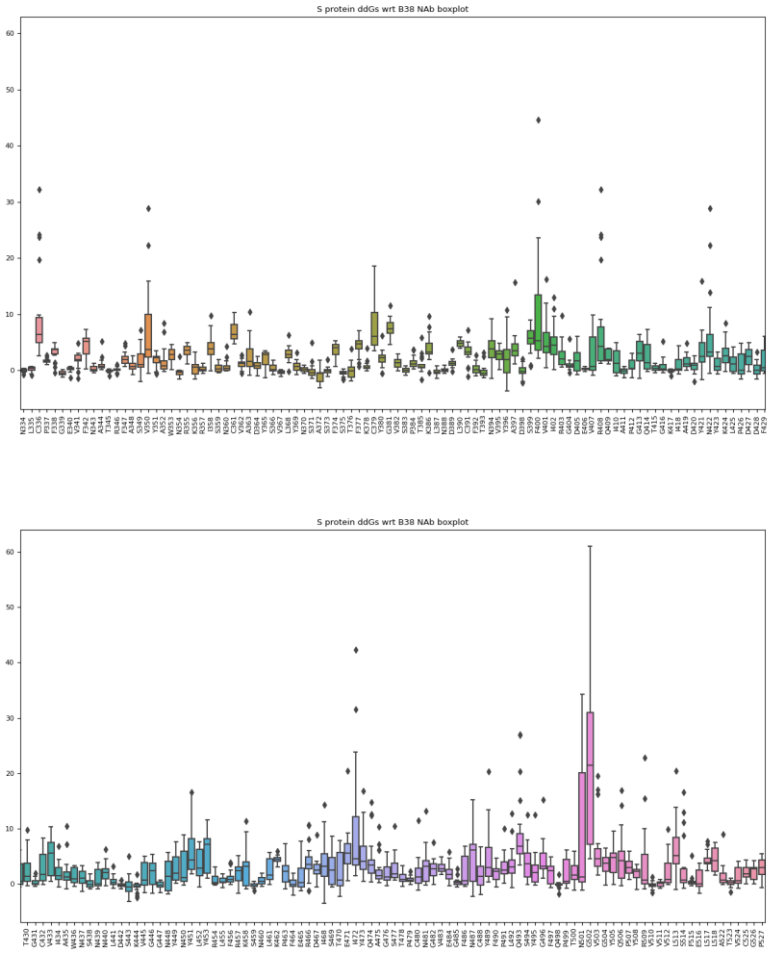


Figure 3 | Range of Gibbs energies at each location on a chain fragment of the S protein RBD complexed with the B38 neutralizing antibody (PDB: 7bz5). AB. Range of Gibbs energies at each location for the chain.

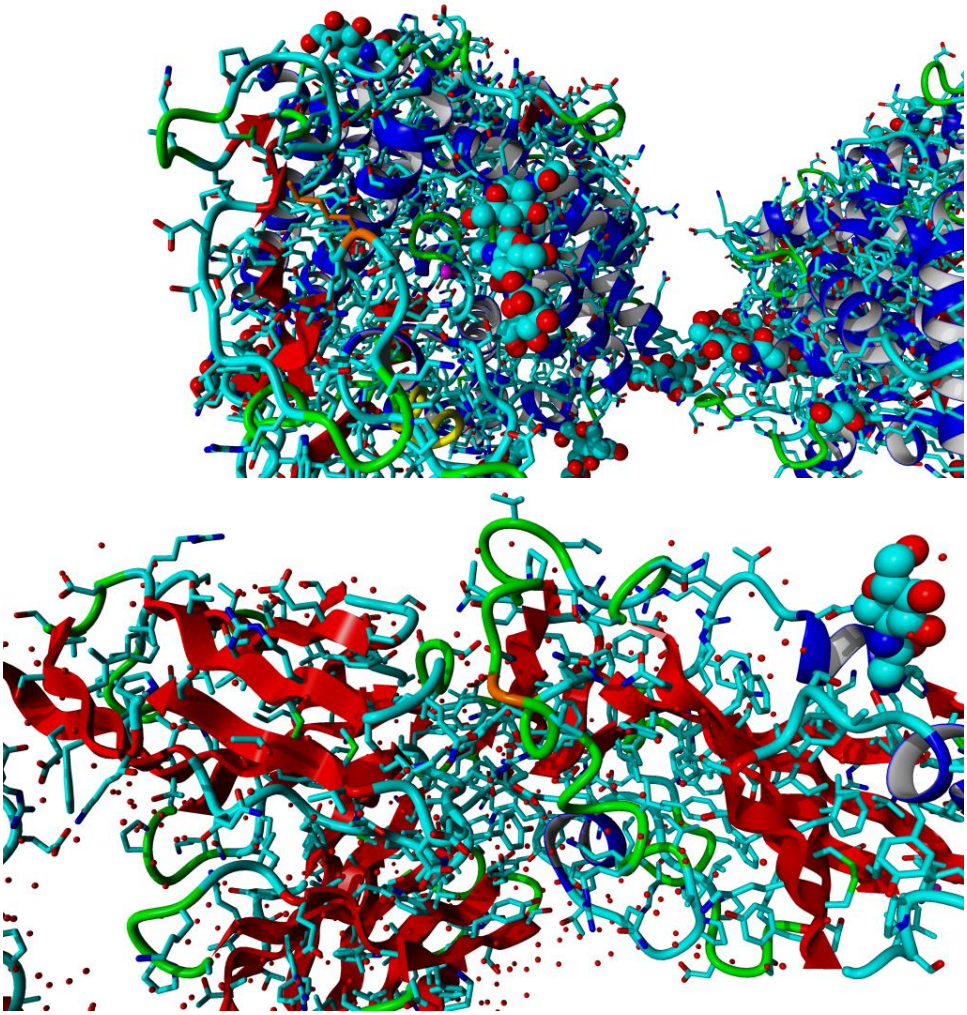


Figure 4 | Wildtype residue locations of outliers in ΔΔG of mutation highlighted in structures in orange. A. Residue K458 in RBD of SARS-CoV-2 in a complex with the human ACE2 receptor (PDB: 6vw1). B. Residue G502 in RBD of SARS-CoV-2 in a complex with the B38 neutralizing antibody (PDB: 7bz5)