

## Motivation

The COVID-19 global pandemic, caused by coronavirus SARS-CoV-2, needs no introduction to a modern audience. Since the disease's emergence, it has spurred a furious effort to understand how to combat the virus and develop robust treatments. An important avenue in this exploration is improving existing antibodies to better combat the virus. This project aims to develop a combined machine learning and computational chemistry analytic pipeline leveraging data from high throughput infectivity assays and in vivo antibody neutralization studies (Bloom et al; Robbiani et al; Baum et al). We want to predict locations on the antibody genome likely to strengthen neutralization of SARS-CoV-2 for the rational design of antibody optimization by applying a new machine learning method developed by Dullerud and Jonsson in combination with energy minimization methods through FoldX software on bound antibody-SARS-CoV-2 complexes (Fig 1). We demonstrate this pipeline using the example of anti-SARS-CoV-2 antibody C105 (PDB: 6XCM), a recently solved structure (Barnes et al.)

To achieve this, we calculated differences in Gibbs energy ( $\Delta G$ ) of mutated antibody-virus complexes and compared it to wild type stability energies, yielding the binding favorability  $\Delta\Delta G$  value. 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.

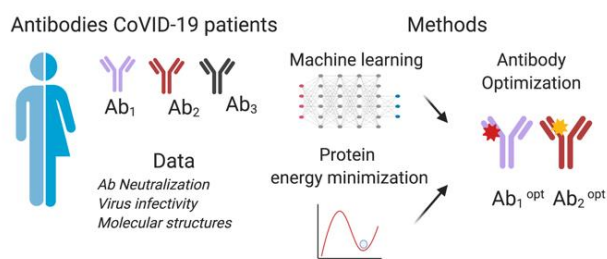


Figure 1 | Graphic overview of project. Pipeline begins with publicly available antibody data from convalescent COVID-19 patients fed into a Jonsson lab machine learning algorithm. Molecular structures including that of antibody C105 are run through FoldX energy minimization software (Barnes et al.). Calculated ddG values are combined with machine learning output to reveal genomic locations where antibodies can be optimized for enhanced neutralization.

## Progress

## Results

### *Energy minimization calculations to optimize antibody design: an example with anti-SARS-CoV-2 antibody C105*

We chose as an example to illustrate our antibody optimization pipeline the C105 antibody isolated by Caltech's Bjorkman lab (Barnes et al). C105 was derived from antibodies from a convalescent COVID-19 patient (Robbiani et al). A detailed crystal structure of the C105 structure was developed by fitting to cryo-EM density (Barnes et al). We chose this antibody because its structure was publicly available and it had already been the subject of vigorous study (Barnes and Bjorkman).

To estimate how mutations of the antibody might affect virus neutralization, we calculated differences in Gibbs free energies between the WT virus and mutations of the C105 antibody locations that were predicted by the Jonsson Lab satlasso algorithm (Dullerud and Jonsson). To do this we performed energy minimization on the structures and found several antibody genomic locations that could be amenable to optimization (Figure 2b).

We investigated two different avenues to improve the antibody—locations where the wildtype had a favorable IC<sub>50</sub> coefficient and ddG was greater than 0 or mutated proteins with ddG less than 0. We also focused specifically on T28 as an antibody location to optimize and illustrate the pipeline, and generated new structures using FoldX's BuildModel function and YASARA protein visualization software to assess the distance between the T28 C105 heavy chain mutations and the virus.

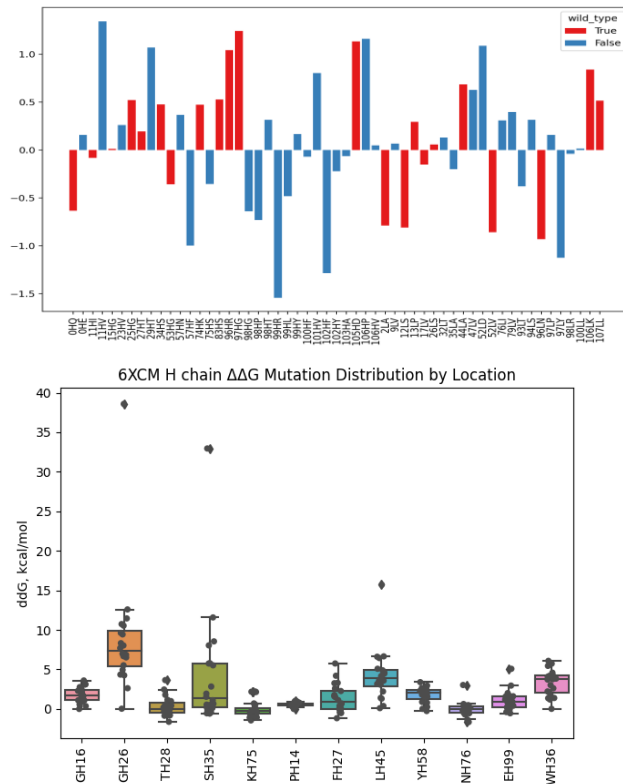


Figure 2 | Analysis of C105 antibody through each pipeline lens: machine learning and computational chemistry. a. Bar plot of coefficients relating impact of mutations to the C105 antibody to neutralization potency solved by satlasso regression (Dullerud and Jonsson). Image credit: Dullerud. b. Box and strip plot of ddG energies calculated with FoldX for selected mutations of the heavy and light chains of C105 antibody defined in PDB structure 6XCM.

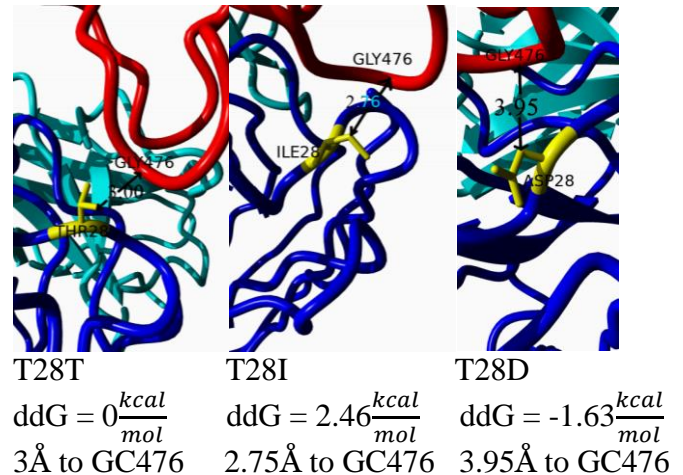


Figure 3 | Heavy chain structures. Molecular structures of residue 28 (highlighted in yellow) on the C105 heavy chain (dark blue) complexed with C105 light chain (light blue) with distances to closest residue on SARS-CoV-2 RBD, GLY476, labeled. Distances calculated using YASARA protein visualization software. a. Wildtype THR28, distance of 3Å to RBD. b. T28I mutation, distance of 2.76Å to RBD. c. T28D mutation, distance of 3.95 Å to RBD.

## 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>), the SARS-CoV-2 RBD structure complexed with a B38 antibody from Wu et al (PDB 7bz5, <https://www.rcsb.org/structure/7bz5>) were used for this study, and the SARS-CoV-2 spike protein complexed with a C105 antibody (PDB 6xcm, <https://www.rcsb.org/structure/6xcm>).

### FoldX Calculations

FoldX 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  $\Delta\Delta G$  of binding.

### Software

Original software developed for this project is written in Python. It can be run from the terminal, read in a PDB file, run a repair on the structure, and run a PositionScan FoldX function to calculate  $\Delta\Delta G$  for every desired mutation. More software uses this returned data to create graphical representations of Gibbs energy distributions. Additionally, Python software was developed to read in, analyze, and graph IC50 neutralization data, pairs it with  $\Delta\Delta G$  data, and selects mutations of interest to output to a FoldX-friendly formatted list to create structures for. Software will be posted to GitHub and will be publicly available upon project completion.

## 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  $\Delta\Delta G$  values with data from in vivo experiments. Investigating this problem led me to realize that I had been using the wrong values ( $\Delta G$ ) 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  $\Delta\Delta G$  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  $\Delta\Delta G$  analysis.

### **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.
- Barnes, C.O., and P.J. Bjorkman. “Structure of the SARS-CoV-2 Spike Glycoprotein in Complex with the C105 Neutralizing Antibody Fab Fragment (State 1).” *Cell*, 1 July 2020, doi:10.2210/pdb6xcm/pdb.
- 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-CoV2.” 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.