



OPEN

The effect of mutations on binding interactions between the SARS-CoV-2 receptor binding domain and neutralizing antibodies B38 and CB6

Jonathan E. Barnes¹, Peik K. Lund-Andersen^{1,2}, Jagdish Suresh Patel^{1,2}✉ & F. Marty Ytreberg^{1,3}✉

SARS-CoV-2 is the pathogen responsible for COVID-19 that has claimed over six million lives as of July 2022. The severity of COVID-19 motivates a need to understand how it could evolve to escape potential treatments and to find ways to strengthen existing treatments. Here, we used the molecular modeling methods MD + FoldX and PyRosetta to study the SARS-CoV-2 spike receptor binding domain (S-RBD) bound to two neutralizing antibodies, B38 and CB6 and generated lists of antibody escape and antibody strengthening mutations. Our resulting watchlist contains potential antibody escape mutations against B38/CB6 and consists of 211/186 mutations across 35/22 S-RBD sites. Some of these mutations have been identified in previous studies as being significant in human populations (e.g., N501Y). The list of potential antibody strengthening mutations that are predicted to improve binding of B38/CB6 to S-RBD consists of 116/45 mutations across 29/13 sites. These mutations could be used to improve the therapeutic value of these antibodies.

SARS-CoV-2, the viral pathogen behind the COVID-19 pandemic, has battered the world population medically, economically, and socially. Infected individuals can suffer from a range of symptoms, from moderate cold-like, to severe pneumonia-like¹. Furthermore, those infected can suffer long-term effects termed “Long-Haul COVID”, the mechanisms and extent of which is not fully understood². Due to the severity and high mortality rate, vaccines were rapidly developed and deployed worldwide³. As the pandemic has progressed, novel variants such as Omicron^{4,5} appeared with the ability to escape protection from antibodies and potentially vaccines^{6,7}.

The primary mechanism of SARS-CoV-2 cell entry is via the receptor binding domain of the Spike glycoprotein (S-RBD) that is located on the surface of the virus⁸. The S-RBD has been shown to be a successful target for antibody therapies. Dejnirattisai et al. tested monoclonal antibodies (mAbs) from patients with verified SARS-CoV-2 infections, finding that 19 of the mAbs that targeted S-RBD were potent neutralizers compared to only one that targeted the N-terminal domain⁹. Piccoli et al. explored SARS-CoV-2 immune sera, finding that 90% of the antibody neutralization was from those targeting the S-RBD¹⁰. For these reasons, there are a number of antibodies that have been identified and developed to target the S-RBD¹¹. Of particular importance to this study are the B38 and CB6 antibodies. B38 was one of the earliest protective antibodies discovered¹² and is currently a basis for engineering new antibodies¹³. CB6 (also known as LY-CoV016, JS016, or Etesevimab) was designed in 2020 and has been used in combination with LY-CoV555 as a treatment for patients with mild to moderate COVID-19 disease^{14,15}. These antibodies share similar epitopes suggesting that single mutations could allow escape from both B38 and CB6. When compared, B38 has five unique contacts with S-RBD (Y453, F490, G496, Q498, V503) and CB6 has three (R405, R408, G504), with the remaining 30 contacts shared between them (85% and 91% of the epitope is shared for B38 and CB6 respectively)^{12,14}.

Since it has become routine practice to monitor genetic variation during outbreaks and pandemics, an aspirational goal in the fight against COVID-19 is to predict when viral evolution is outpacing the effectiveness of protective antibodies and exploring ways to retain or improve their efficacy. Antibody escape can occur for many

¹Institute for Modeling Collaboration and Innovation, University of Idaho, Moscow, ID 83843, USA. ²Department of Biological Sciences, University of Idaho, Moscow, ID 83843, USA. ³Department of Physics, University of Idaho, Moscow, ID 83843, USA. ✉email: thejagdishpatel@gmail.com; ytreberg@uidaho.edu

possible reasons. For example, if a mutation significantly reduces affinity to an antibody, then it is likely to result in escape from that antibody regardless of other possible mutation effects. However, a mutation can also affect binding rates, folding efficiency, or downstream characteristics that could lead to escape without a change in affinity. Experimental work has been done to map mutations in the S-RBD that show antibody escape. Bloom and collaborators generated a complete map of mutations that escape binding with 10 monoclonal antibodies (not including B38 or CB6)¹⁶. Important for our study, two other studies from the Bloom lab evaluated escape mutations for CB6; one included just CB6¹⁷ and another considered CB6 and its cocktail with Ly-CoV555¹⁸. These studies used a deep mutational scanning technique to ascertain the effects of mutations in the S-RBD and binding with ACE2¹⁹ and led them to develop software that predict antibody escape mutations²⁰.

Computationally, Teng et al. used FoldX to perform deep mutational scanning to predict stability and determine the effects on binding affinity between the S-RBD and ACE2²¹. Tsai et al. also used FoldX, MutaBind2, and mCSM-PPI2 to generate heatmaps for possible antibody escape for five antibodies bound to the S-RBD²². These five antibodies all bind to the S-RBD, a few of which bind to similar parts of the S-RBD as B38 and CB6. Sharma et al. used ProAffiMuSeq, mCSM, CUPSAT, and FoldX to determine relevant structural aspects for binding affinity between the S-RBD and antibodies, noting attributes like changes to Tyrosine (Y) as significant²³. Furthermore, there have been computational studies that investigated known variants bound to B38 and CB6 and their potential escape behavior. Ray et al. used protein-graph-connectivity networks to investigate the changes in dynamics with known variants and their interaction with mAbs B38 and BD233²⁴. Hendy et al. used molecular dynamics simulations to predict the binding affinity between antibodies (including B38) with the RBD considering three mutation sets matching those found in variants B1.1.7, B1.1.28, and B1.351²⁵. They also explored potential mechanisms of escape. Miller et al. explored the effects of mutations specific to the recent Omicron variant using amino acid interaction networks²⁶. They tested binding with a few antibodies including CB6. Laurinini et al. performed alanine scanning to investigate potential escape mutations for antibodies including CB6 finding 90% agreement to experimental results²⁷. While such studies are informative at identifying sites that are significant for binding, they do not provide information specific to individual mutations. We seek to build on these previous studies to provide a comprehensive list of escape mutations for B38 and CB6 by predicting binding affinity changes due to all possible amino acid changes.

In this work, we use an approach previously developed in our group to generate watchlists of mutations for ebolavirus^{28,29} using a fast empirical method, FoldX³⁰, combined with rigorous molecular dynamics (MD) simulations (termed “MD + FoldX”). We also generated results using PyRosetta software. We applied these methodologies to the S-RBD in a complex with two antibodies, B38 and CB6 to predict a large number of possible mutations of S-RBD that can disrupt antibody binding and allow escape. The appearance of mutations in this watchlist in an emerging variant will be evidence that treatments using these antibodies may be less effective. Additionally, we identified mutations in the antibodies that are predicted to improve their binding to the S-RBD. These mutations may strengthen the therapeutic effect of the antibody against COVID-19.

Methods

We employed the MD + FoldX approach developed in our group to build a watchlist of antibody escape mutations^{28,29,31}. Following this approach, we first carried out short MD simulations using the 3D structures of the antibody-S-RBD complexes (for both B38 and CB6) as inputs to sample conformational flexibility. The snapshots from the MD simulations were then used as inputs for FoldX to perform mutational scanning and generate averaged predictions for relative binding ($\Delta\Delta G_{\text{bind}}$) and folding ($\Delta\Delta G_{\text{fold}}$) free energy changes due to all possible mutations at sites near the binding interface. To overcome the limitation of FoldX software and to improve the overall escape mutation prediction accuracy, we also used the PyRosetta program. We used the built-in conformational sampling procedure in PyRosetta, and its knowledge-based scoring function to estimate $\Delta\Delta G_{\text{bind}}$ and $\Delta\Delta G_{\text{fold}}$ values. Both MD + FoldX and PyRosetta approaches were also used to scan all possible mutations of B38 and CB6 to identify mutations that may improve their binding to the S-RBD. Mutation sites were chosen to be those within 10 Å of the binding interface based on the energy minimized structures. This is reasonable since both FoldX and PyRosetta are expected to only capture local effects. The mutation sites for S-RBD, for example, were determined by selecting α -carbons on the S-RBD that were within 10 Å of any other atom belonging to either the heavy or light antibody chains.

System determination and preparation. To generate a watchlist of potential antibody escape mutations for the S-RBD and identify mutations that could improve antibody binding affinity, we selected two S-RBD-antibody complexes: B38 and CB6. These complexes have Protein Data Bank³² identifiers of 7BZ5 and 7C01 corresponding to antibodies B38 and CB6, respectively. Figure 1 shows the epitopes of B38 and CB6 bound to the S-RBD, highlighting the overlapping regions.

Structures were prepared and solvated using the OpenMM-setup package. The AMBER14 forcefield was used to generate protein topology files along with the default TIP3P-FB water model. Only the protein chains were retained with all heterogens and water removed. For the B38 complex, the missing HIS at site 519 in chain A was added since it is an internal loop; other missing residues were not added since they are part of the termini and located away from the binding interface. The CB6 complex has missing residues located at the termini only, so none were modeled. Heavy atoms were added for the B38 complex (LYS528, SER217, CYS215) and the CB6 complex (PRO527, GLU218, GLU215). The following protocol was then applied for all simulations. Hydrogens were added corresponding to a pH of 7.0. A rhombic dodecahedral water box was used with a padding distance of 1.2 nm. Sodium and chloride atoms were added to neutralize the net charge of the systems with an ionic strength of 0.15 M. MD simulations were then carried out using OpenMM³³ with the following protocol. The systems were first minimized for 2000 steps using *minimizeEnergy*. Next, we performed NVT annealing with

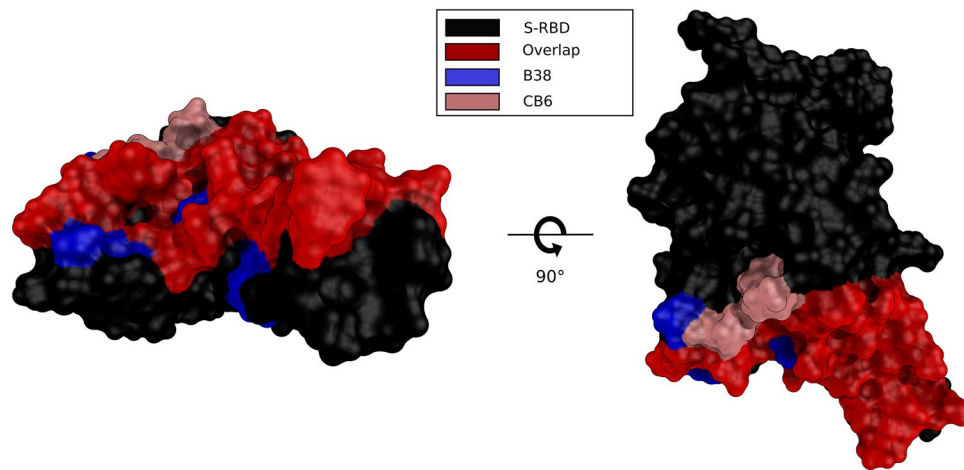


Figure 1. Comparison of epitopes for antibodies B38 and CB6 on S-RBD. Red (overlapping epitope), Blue (B38) and Salmon (CB6) colors are used to compare epitope regions on the 3-D structure of the S-RBD (black) shown in surface representation. Two viewpoints of the S-RBD structure are shown to provide full context.

position restraints for a total of 1 ns starting at 10 K and stopping at 310 K with a ΔT of 3 K and 5000 steps at each temperature. NPT equilibration was then performed for 1 ns at 310 K, with a pressure of 1 bar maintained via the Monte Carlo barostat with an interval of 25 steps. For all simulations the Langevin integrator was used and the time step set to two fs. The only constraints were hydrogen bonds. Production simulations were then run for 100 ns with the same parameters except for the barostat interval adjusted to every 100 steps, with snapshots taken every 1 ns for a total of 100 snapshots to be used for analysis. For binding affinity calculations, we then performed a total of two simulations (S-RBD-B38 and S-RBD-CB6 complexes). For folding stability calculations, we performed five simulations (S-RBD, B38 H, B38 L, CB6 H, and CB6 L) where H and L denote the heavy and light chains for each antibody, respectively.

Mutation scanning. Following the protocols used for the Ebola watchlist and watchlist expansion studies^{28,29,31}, FoldX was used to estimate the folding stability of S-RBD and the binding affinity between S-RBD and the two antibodies using 100 MD snapshots from each simulation as inputs. Each snapshot was first prepared by running the *RepairPDB* command six times in succession to ensure convergence of the potential energy. The *BuildModel* command was then used to generate mutations to all possible 19 amino acids at all sites within 10 Å of the binding interface. For completeness, we considered the union of sites between the S-RBD-B38 complex and the S-RBD-CB6 complex consisting of 75 sites for the S-RBD, 51 sites for antibody B38, and 55 sites for antibody CB6. Mutations were introduced on the S-RBD for predicting antibody escape mutations (1425 total mutations) and on both the antibodies to identify antibody strengthening mutations (1349 mutations for B38, 1407 for CB6). *AnalyseComplex* was then used to predict ΔG_{bind} values for the wildtype and mutants and the difference was calculated to obtain $\Delta\Delta G_{\text{bind}}$ values. This process was repeated for the 100 snapshots, and then averaged to generate the final predicted $\Delta\Delta G_{\text{bind}}$. Similarly, to predict $\Delta\Delta G_{\text{fold}}$ values, the MD snapshots for each individual protein chain were analyzed using the *Stability* command and averaged. In addition to MD + FoldX, PyRosetta-4³⁴ was used to estimate the differences in binding and folding stability scores between mutant and wildtype structures. The ref2015 forcefield³⁵, a knowledge-based energy function, was used for all PyRosetta calculations and procedures. The *flexddG* protocol³⁶ implemented in PyRosetta was used to estimate $\Delta\Delta G_{\text{bind}}$ and $\Delta\Delta G_{\text{fold}}$ values and is briefly described here. The experimental structures were used as inputs and they were first relaxed with PyRosetta's *FastRelax* protocol³⁷. Here, we used the experimental structures (not the MD snapshots) since protocols for PyRosetta already generate conformational ensembles, in contrast to FoldX protocols that use a single structure and hence benefits from additional conformational sampling. A conformational ensemble of 50 structures was then generated using PyRosetta's *backrub* protocol³⁸ over 50 iterations, to sample backbone torsions within 8 Å of the mutation site. For each structure, all possible mutations for a given residue were then introduced, followed by global repacking of side chains using the 2010 Dunbrack rotamer library³⁹. The structures then underwent energy minimization using the *lbfgs_armijo_nonmonotone* algorithm with a tolerance of $1e-5$ and a convergence threshold of 0.5 kcal/mol. The wildtype structures were also refined with the sidechain packing and energy minimization using the same parameters. To calculate binding ΔG values, we computed the total energy of the bound complex and subtracted the energies of the separated antibody and antigen structures then averaged the results. For both binding and folding stability, the differences between the mutated and wildtype structure energies were taken to obtain $\Delta\Delta G_{\text{bind}}$ and $\Delta\Delta G_{\text{fold}}$, respectively. These $\Delta\Delta G_{\text{bind}}$ and $\Delta\Delta G_{\text{fold}}$ values were then averaged across the 50 conformations to get the final values. To our knowledge PyRosetta has not been as thoroughly tested as MD + FoldX for antibody-antigen systems, hence we ran a test using the antibody-antigen benchmarking systems from Gonzales et al. to compare its accuracy for this purpose. We found that for two of the test systems, PyRosetta performed very similarly to FoldX (within 2% or better).

One of the test systems performed significantly better, and we found poor performance for the other half of the systems, primarily driven by a handful of outliers. The full details are shown in Supplementary Figure S1.

Results and discussion

The goal of this study is to develop both a watchlist of potential antibody escape mutations against two antibodies that interact with the S-RBD of SARS-CoV-2, and a list of possible antibody strengthening mutations. Watchlist mutations are amino acid changes in S-RBD that either MD + FoldX or PyRosetta predict to disrupt its binding to the B38 or CB6 antibodies, hence potentially reducing therapeutic efficacy of these antibodies. Similarly, antibody strengthening mutations are mutations in the antibodies that are predicted to strengthen their binding interaction to S-RBD, hence potentially increasing the therapeutic efficacy of the antibodies.

For a mutation to be part of our watchlist, i.e., be a potential escape mutation, we used the cutoff of $\Delta\Delta G_{\text{bind}} > 2.0$ kcal/mol from our previous studies to classify binding disruption^{28,29,31}. For antibody strengthening, our cutoff of $\Delta\Delta G_{\text{bind}} < -0.5$ kcal/mol was chosen to be consistent with the standard deviation determined during the validation study of FoldX that is reported as 0.46 kcal/mol³⁰. For folding, we used a cutoff of -3.0 kcal/mol $< \Delta\Delta G_{\text{fold}} < 3.0$ kcal/mol to define when a protein can still fold. All cutoffs were purposely chosen to be inclusive rather than conservative given the uncertainty of the calculations. Finally, note that the *REF15* energy function we used in PyRosetta provides results expressed in kcal/mol not the previously-used Rosetta Energy Units³⁵. This allows us to use the same cutoffs for both FoldX and PyRosetta predictions. We note that FoldX and PyRosetta use different strategies to estimate affinities and so we do not expect full agreement. FoldX uses a semi-empirical forcefield and relaxes the input structures using a proprietary algorithm termed *repairPDB*. PyRosetta performs Monte Carlo minimization of the input structure, generates a conformational ensemble and then uses a physics-based scoring function for binding affinity predictions of the ensemble. Our union dataset includes results from both methods to minimize the risk of leaving out possible escape mutations.

The heat maps in Fig. 2 show S-RBD-antibody binding affinity predictions from MD + FoldX and PyRosetta due to mutations in S-RBD (full results are available in Supplemental Dataset S1). Each method predicts possible escape mutations at specific sites. For B38, MD + FoldX predicts a total of 73 potential escape mutations at eight sites in S-RBD. More than 50% of these belong to sites R403, A475, N487, N501, and Y505. For CB6, MD + FoldX predicts 16 escape mutations at three sites, with site G476 containing the most. PyRosetta predicts a much larger escape mutation dataset with 165 and 187 escape mutations for B38 and CB6, respectively. We note that for some sites in B38 (e.g., D420 and Y421), MD + FoldX predicts destabilizing mutations but none that exceed our cutoff ($\Delta\Delta G_{\text{bind}} > 2.0$ kcal/mol) to qualify as antibody escape mutations. By contrast, PyRosetta predicts 16 antibody escape mutations for sites D420 and Y421. Similarly, this can also be seen in the case of CB6 for sites L455 and F456.

We believe that the most valuable results in this study are the union of the MD + FoldX and PyRosetta datasets; the union dataset limits possible false exclusion of mutations on a watch list. To motivate our reasoning behind using the union dataset, consider that Fig. 2 shows that MD + FoldX predicts N501Y as a possible escape mutation for B38, however PyRosetta does not predict it for either antibody. N501Y is a known mutation present in a number of variants of concern, both on its own and in combination with others^{41–44}. Additionally, it has been shown to be key in epistatic interactions associated with increased infectivity⁴⁵. We note that PyRosetta still predicts a mutation at site 501, but only N501W for B38.

The union dataset (i.e., includes both MD + FoldX and PyRosetta) is shown in Table 1 and Fig. 3. This consists of 211 predicted escape mutations ($\Delta\Delta G_{\text{bind}} > 2.0$ kcal/mol) across 35 sites for B38 and 186 mutations across 22 sites for CB6. There are 178 mutations within this combined watchlist that are predicted to not fold properly, however, we still include them in an effort to be as inclusive as possible. We also note that if a mutation does indeed significantly disrupt the protein fold, then it will not appear in a natural population. The sites with the highest concentration of potential escape mutations for B38 are R403, D420, Y421, L455, A475, G476, N487, G502, and Y505 (contain 51% of the predicted escape mutations). For CB6 these are V445, G447, L455, F456, Y473, A475, G476, Q493, Q498, and Y505 (contain 85% of the predicted escape mutations). We predict 11 sites to harbor antibody escape mutations for both antibodies, all within the shared epitope region. Table 2 shows the frequency of the mutated amino acids predicted to allow antibody escape. We see that the most common mutations for both B38 and CB6 are to tryptophan (W) – a large side chain – intuitively likely to disrupt function, especially if the side chain is small in the wildtype protein. Furthermore, we see that smaller side chains like alanine and glycine appear less frequently in the table.

There are currently no large-scale experimental studies evaluating antibody escape for B38. For CB6, however, there are experimental results that we can use to compare with our predictions. Starr et al. used deep mutational scanning to build a map of mutations in S-RBD and measured how these mutations modify the ability of S-RBD to bind CB6¹⁷. They measured the escape fraction for nearly all possible mutants in the S-RBD. Mutations that were excluded were filtered out for having low sequencing counts, for low binding with ACE2 (based on the

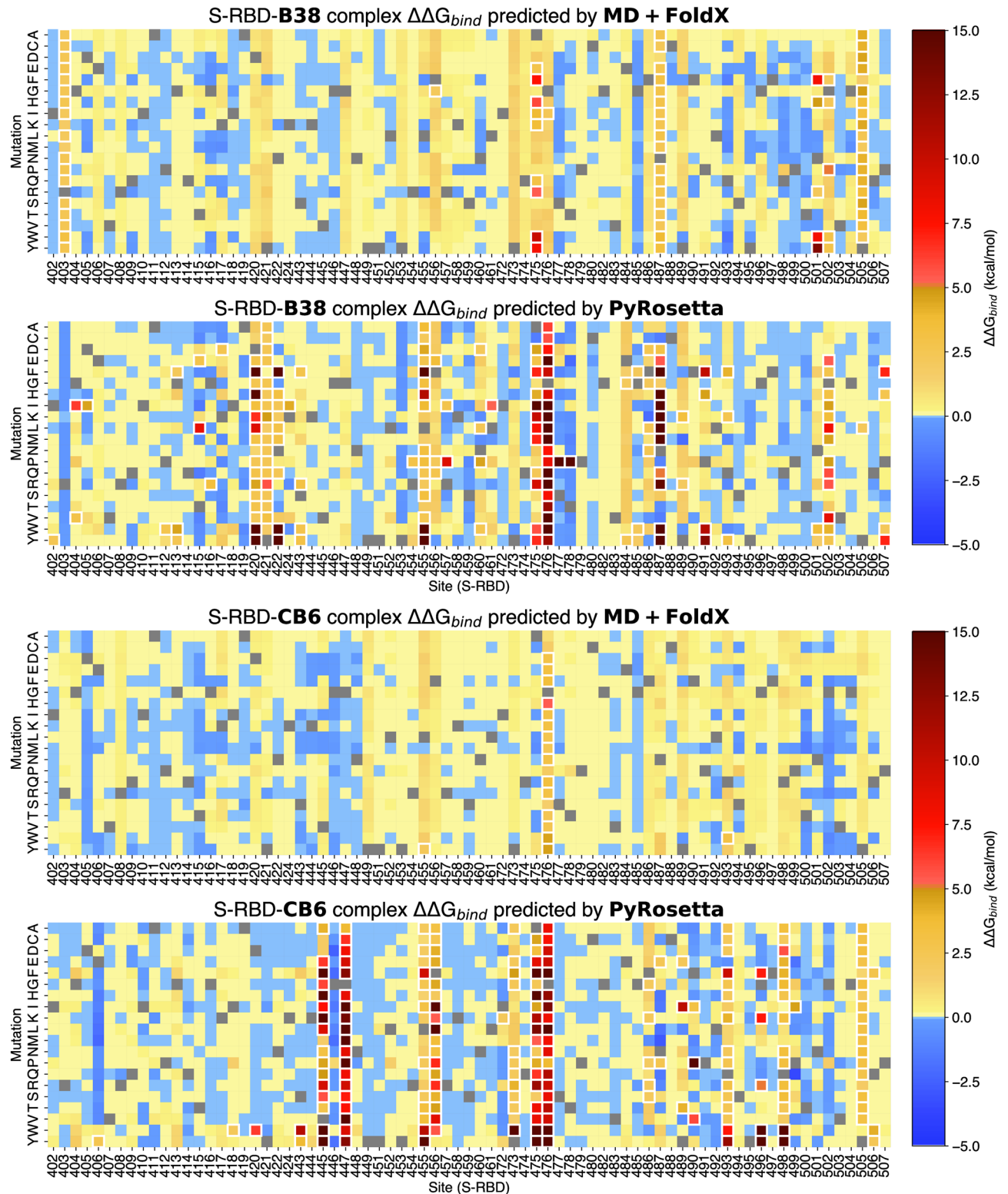


Figure 2. Heatmaps for S-RBD mutations showing PyRosetta and MD + FoldX predicted $\Delta\Delta G_{bind}$ values for all possible 19 mutations at each site on the epitopes of B38 (top) and CB6 (bottom) antibodies. Blue indicates mutations that lead to stronger binding ($\Delta\Delta G_{bind} < 0$), yellow and red indicates binding disruption ($\Delta\Delta G_{bind} > 0$). The white highlighted data points are cases where the $\Delta\Delta G_{bind}$ values are above the cutoff, hence are predicted to be antibody escape mutations ($\Delta\Delta G_{bind} > 2$ kcal/mol). This figure was generated using the Python programming language v3.10.8 (Python software foundation, <https://python.org>) and the matplotlib plotting module v3.5.2⁴⁰.

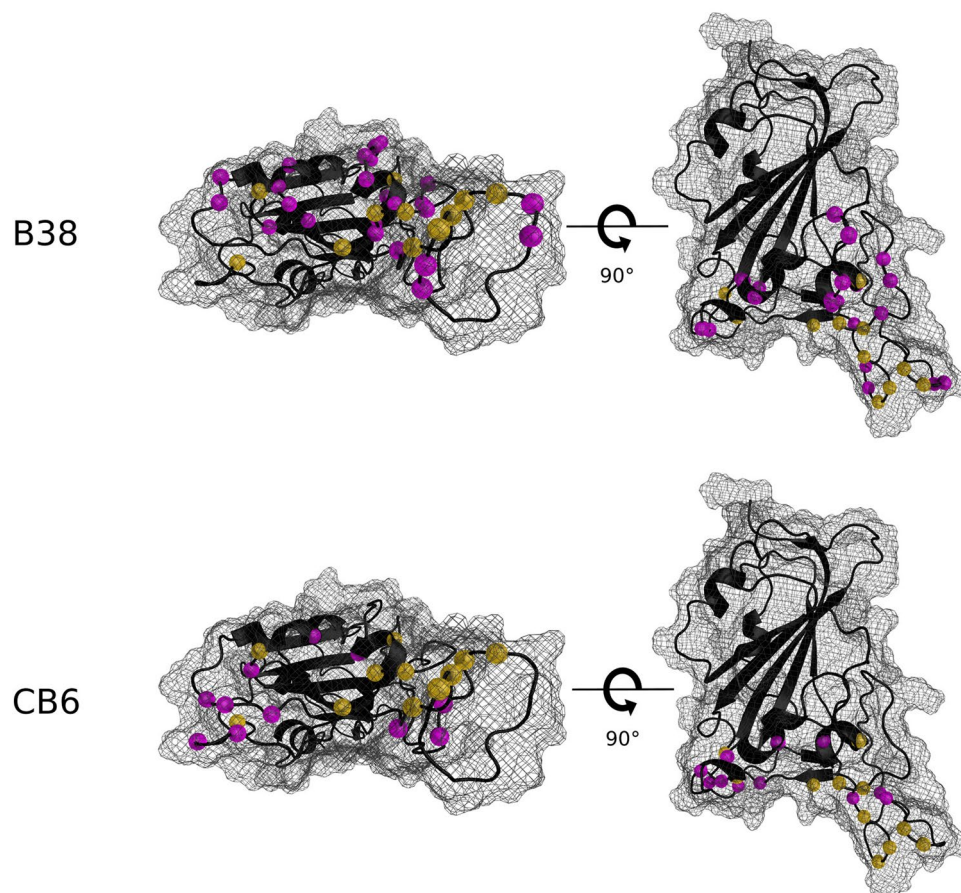


Figure 3. The S-RBD amino acid sites predicted to contain antibody escape mutations (magenta) by MD + FoldX or PyRosetta (i.e., all sites listed in Table 1). The S-RBD sites that are predicted to harbor antibody escape mutations for both B38 and CB6 are shown in gold. For both B38 (top) and CB6 (bottom) two views are shown to provide context. The S-RBD is indicated in black, with the surface shown as a mesh.

binding score of RaTG13 to human ACE2, lowest known affinity capable of cell entry), or for being a variant with poor RBD expression. Of the 186 mutations in our watchlist, 48 are excluded from the Starr et al. study (see Supplemental Table S1). Of those excluded, we predict 39 to fold ($-3.0 \text{ kcal/mol} < \Delta\Delta G_{\text{fold}} < 3.0 \text{ kcal/mol}$). The binding results from Starr et al. were provided in escape fraction, a metric that ranges from 0 (no observed escape, antibody always bound) to 1 (a decrease in binding frequency by $100\times$ or more). While we cannot directly compare escape fraction to $\Delta\Delta G_{\text{bind}}$, we can safely assume that a mutation with a large (actual, not predicted) $\Delta\Delta G_{\text{bind}}$ has a higher chance to escape and hence a larger escape fraction. Note that the converse is not true, that is, high escape fraction does not imply large $\Delta\Delta G_{\text{bind}}$ since mutations can escape antibody binding for a variety of reasons other than reduced affinity (e.g., modified binding rates, misfolding, etc.). Comparing our union CB6 dataset with Starr et al. we find that 20% (38) of our predicted escape mutations correspond to an escape fraction of 0.5 or greater with the majority falling below 0.1. We next looked more closely at the subset of mutations that are in agreement between our study and Starr et al., that is mutations we predict to escape that they determined to have an escape fraction greater than 0.5. For PyRosetta, we found that all 38 mutations have $\Delta\Delta G_{\text{bind}}$ values above the cutoff. In addition, the mutation with the largest predicted $\Delta\Delta G_{\text{bind}}$ of 44.52 kcal/mol corresponded to an escape fraction of 0.98. For FoldX, only four of the 38 mutations are above the cutoff, and the largest $\Delta\Delta G_{\text{bind}}$

S-RBD-B38		S-RBD-CB6	
Site	Predicted Escape Mutants	Site	Predicted Escape Mutants
I402	Y	E406	Y
R403	A,C,D,E,G,H,I,L,N,P,Q,S,T,V,W,Y	I418	W
G404	I,V	D420	W
D405	I	S443	W,Y
P412	W	V445	A,E,F,G,H,I,K,L,N,P,Q,R,S,T,W,Y
G413	F,W,Y	G447	A,C,D,E,F,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
T415	E,L	L455	A,D,F,G,I,P,S,T,V,W,Y
G416	R	F456	A,C,D,E,G,H,I,K,L,M,N,P,Q,R,S,T,V,W
K417	D	I472	K
D420	E,F,G,H,I,K,L,M,P,Q,R,S,T,V,W,Y	Y473	A,E,F,G,N,P,Q,R,S,T,W
Y421	A,C,D,E,G,H,I,K,M,N,P,Q,R,S,T,V	A475	C,D,E,F,G,H,I,K,L,M,N,P,Q,R,T,V,W,Y
N422	F,H,I,L,M,Q,W,Y	G476	A,C,D,E,F,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y
K424	I	F486	G,K,P,R,S
S443	F,R,W,Y	N487	E,P
R454	P	Y489	E,I,T
L455	A,C,D,E,F,G,H,I,N,P,Q,R,S,V,W,Y	F490	I,P,V
F456	E,G,P,Q	Q493	A,C,D,F,G,H,K,N,P,R,S,T,V,W,Y
R457	I,P	G496	F,K,R,W,Y
N460	D,L,P,W	Q498	A,C,D,E,F,G,H,K,N,P,R,S,T,W,Y
L461	I	P499	I
A475	D,E,F,H,I,K,L,M,Q,R,W,Y	Y505	A,C,D,E,G,H,I,K,L,M,N,P,Q,S,T,V
G476	A,C,D,E,F,H,I,K,L,M,N,P,Q,R,S,T,V,W,Y	Q506	F,W,Y
S477	P		
T478	P		Blue: MD+FoldX
E484	G,Y		Black: PyRosetta
G485	F,W		Red: Both MD+FoldX and PyRosetta
F486	D,G,L,N		Yellow: Site predicted for both antibodies
N487	A,C,D,E,F,G,H,I,K,L,M,P,Q,R,S,T,V,W,Y	Key:	Bold: Predicted to Fold
Y489	K,R		
P491	F,H,W,Y		
Q493	F,K,Y		
N501	F,H,R,W,Y		
G502	E,F,H,I,K,L,M,P,Q,R,V,W,Y		
Y505	A,C,D,E,G,H,I,K,L,M,N,P,Q,R,S,T,V,W		
P507	F,H,W,Y		

Table 1. Union dataset of predicted escape mutations for B38 and CB6 antibodies based on both MD + FoldX and PyRosetta simulations. The left and right tables show results for B38 and CB6, respectively. For each subtable, the left column is the S-RBD site and the right contains the mutations predicted by either MD + FoldX and/or PyRosetta to lead to antibody escape. Predicted escape mutations are colored based on whether both methods predict escape (red), only MD + FoldX predicts escape (blue), or only PyRosetta predicts escape (black). The yellow highlighted sites are those predicted to contain antibody escape mutations for both antibodies. Bold font denotes mutations that are predicted to allow folding.

is 3.23 kcal/mol corresponding to an escape fraction of 0.871. We note that there are two additional mutations in this subset, both at site L455, where FoldX predicted $\Delta\Delta G_{\text{bind}}$ values just below our cutoff (1.34 and 1.59 kcal/mol) that would have been excluded from our watchlist if we had only considered FoldX results. Finally, if we compare our results with Starr et al. at the site level, we see about 50% agreement (Supplemental Table S1). In some of these sites 50% of mutations are predicted to escape: L455, F456, Y473, A475, and G476. To summarize this comparison, we believe that the discrepancies between our results and the Starr et al. data highlight the need for more accurate modeling approaches and techniques.

We have also identified sites and mutations relevant to recent variants of concern. Predominant variants of concern include the UK variant (B.1.1.7; N501Y), South African variant (B.1.351; N501Y, E484K, K417N, L18F, A701V), Brazilians (P1/P2; N501Y, E484K, K417T, D614G, V1176F), Californians (B.1.429/7; L452R, W152C, S13I), and Indian (Delta) (B.1.617.1, B.1.617.2/B.1.617.3; E484Q, L452R, P681R)⁴⁶. N501Y is common in these variants of concern and is consistent with our findings, is a known antibody escape mutation, and has been shown to reduce efficacy of B38⁴⁷. Our methods indicate E484 is also a site of significance; present in the South African, Brazilian, and Delta variants. The most recent variant, Omicron, has mysterious origins as it is a departure from all previously determined lineages⁴⁸. Omicron consists of a large number of mutations (62 vs. 45 in other variants on average)⁴⁹. It has also been noted to escape vaccines and other antibody treatments^{6,50}. It consists of a large number of mutations in the S-RBD: G339D, S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y, Y505H⁵¹. Of these, we identified N501Y and Y505H in B38 and Q493R, Q498R, and Y505H as CB6 escape mutations. While they were not the known mutations of significance, we did identify K417D, S477P, and T478P. We cannot expect total overlap with our findings since the Omicron variant contains a large number of mutations likely working in concert, and with likely epistatic effects⁴⁵.

Using the same MD + FoldX and PyRosetta protocols, we also built a list of B38 and CB6 antibody mutations that could strengthen their interaction with the S-RBD (Fig. 4, Table 3). The figure shows a heatmap of antibody mutations and their predicted effect on binding S-RBD (full results are available in Supplemental Dataset S2). The table shows predicted antibody strengthening mutations based on our cutoff ($\Delta\Delta G_{\text{bind}} < -0.5$ kcal/mol) with those predicted to allow folding in bold (-3.0 kcal/mol $< \Delta\Delta G_{\text{fold}} < 3.0$ kcal/mol). We found 116 mutations across 29 sites and 45 mutations across 13 sites for B38 and CB6 respectively. There are currently no experimental studies detailing mutations in these antibodies and their effects on binding to the S-RBD, but B38 is currently being used to engineer improved antibodies¹³. We also note that computational studies such as ours have been used in the past to design and improve on existing antibodies for other antigens by performing single mutation mutagenesis and evaluating the effects on binding affinity⁵². We hope these results can serve as a reference for further analysis and future antibody design.

Our choice of using MD + FoldX was motivated by our previous studies^{28,29}. MD + FoldX was compared to eight other methods and shown to have the best accuracy for antibody-antigen complexes³¹. Another recent study experimentally validated this approach, demonstrating that six of eight escape mutations predicted by MD + FoldX showed decreased neutralization in respiratory syncytial virus⁵³. To provide a point of comparison and include mutations that might be “missed” by MD + FoldX, we also used PyRosetta. A limitation of our study, and indeed all studies that consider only single point mutations, is the possibility of epistasis. Computational methods that may be capable of predicting epistatic effects are rigorous and computationally expensive. Fast methods such as FoldX and PyRosetta are trained or referenced by datasets consisting of single point mutations, hence tend to predict additivity in the case of two or more mutations. It is already understood that epistasis is present in SARS-CoV-2⁵⁴. Furthermore, emerging variants of concern are frequently due to multiple mutations, such as Omicron with 62 mutations throughout the genome⁴⁹. While well beyond the scope of the current report, it would be beneficial to future studies to develop rapid methods that are capable of accurately predicting the effects due to multiple mutations.

Our choices of cutoffs have a significant effect on results. If we consider sites and mutations from experimental data that have a strong antibody escape effect (like E484K in the Chakraborty study)⁴⁷ we find the mutation does have a positive $\Delta\Delta G_{\text{bind}}$ for both FoldX (0.22 kcal/mol, well below cutoff and within FoldX error) and PyRosetta (1.81 kcal/mol). If our cutoff were 1.5 kcal/mol instead this mutation would have been included. However, this cutoff choice would also include a large set of other mutations that may or may not exhibit escape. While any choice of cutoff is inherently arbitrary, we have attempted to choose reasonable values based on previous

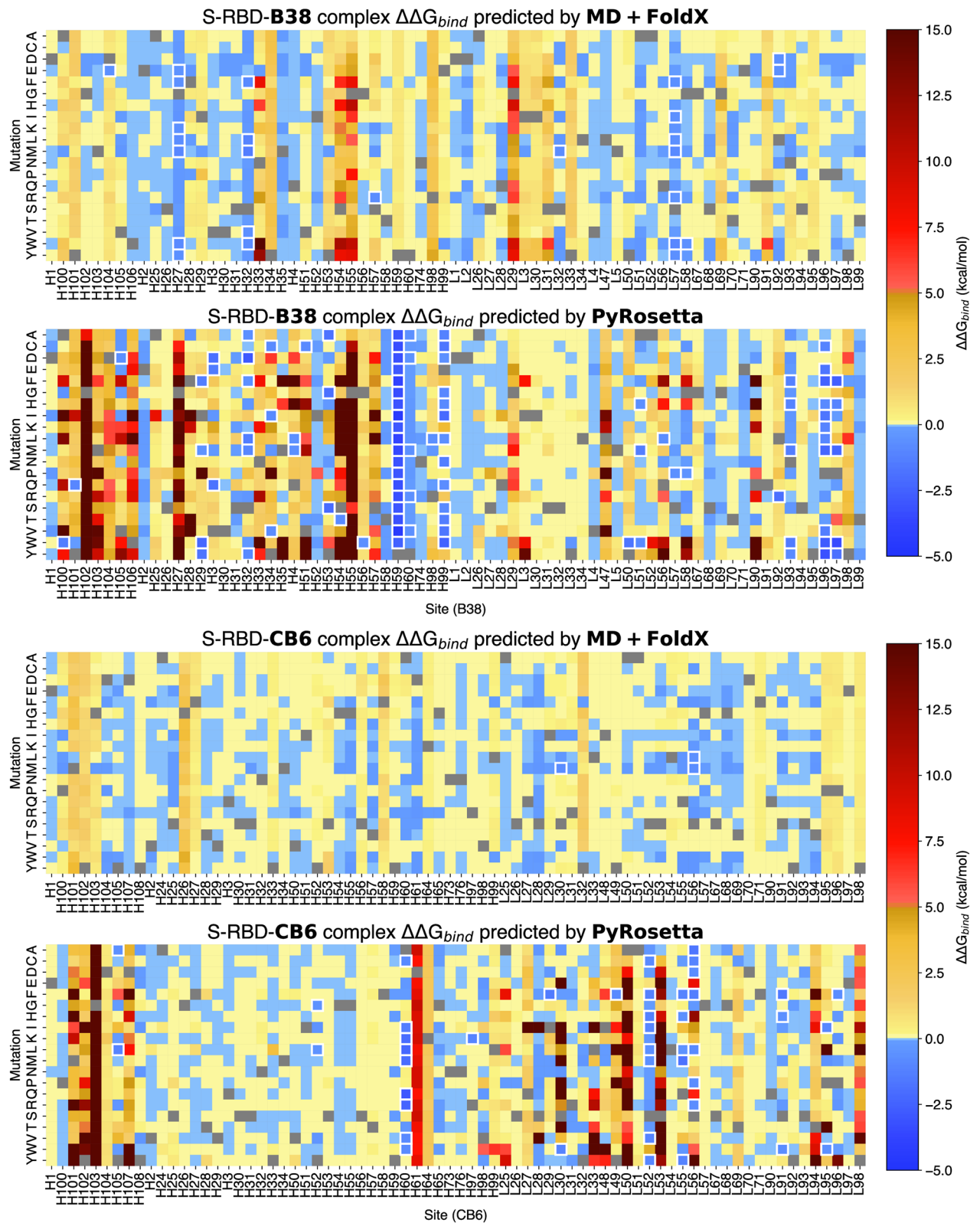


Figure 4. Heatmaps showing how mutations of B38 (top) and CB6 (bottom) modify binding to S-RBD. For each antibody, both MD + FoldX predictions (top) and PyRosetta predictions (bottom) are shown. Blue indicates stronger binding ($\Delta\Delta G_{bind} < 0$), yellow and red indicates binding disruption ($\Delta\Delta G_{bind} > 0$). Mutations predicted to be antibody strengthening ($\Delta\Delta G_{bind} < -0.5$ kcal/mol) are highlighted in white. This figure was generated using the Python programming language v3.10.8 (Python software foundation, <https://python.org>) and the matplotlib plotting module v3.5.2⁴⁰.

B38		CB6	
Mutation	Frequency	Mutation	Frequency
W	17	W	14
Y	16	P	13
F	15	K	11
I	15	T	11
P	14	Y	11
H	13	E	10
R	12	F	10
E	11	I	10
L	11	R	10
Q	11	S	10
D	10	A	9
G	9	G	9
K	9	N	9
V	9	D	8
M	8	H	8
S	7	V	8
A	6	C	7
C	6	Q	7
N	6	L	6
T	6	M	5

Table 2. Residue type frequencies for predicted escape mutations B38 and CB6 are shown left and right respectively. The mutated residue is indicated in the left subcolumn for each and the frequency it occurs in our escape dataset is shown on the right.

studies and the inherent error in the calculations. For example, if we chose an antibody strengthening cutoff of $\Delta\Delta G_{\text{bind}} < 0$ kcal/mol this would lead to 757 mutations across 69 sites for B38 and 773 mutations across 70 sites for CB6, of which 85%+ of the results would be within the error for FoldX. If we consider known escape variants such as K417N for CB6, our results predict a positive $\Delta\Delta G_{\text{bind}}$ (0.566 kcal/mol and 0.0046 kcal/mol for FoldX and PyRosetta, respectively) meaning our results predict possible escape, but the mutations were not included in our watchlist since these values are below our threshold.

Conclusions

In this study, we investigated how single amino acid mutations affect binding between the S-RBD of SARS-CoV-2 and two antibodies with similar epitopes, B38 and CB6. We considered both how mutations in S-RBD could lead to antibody escape (termed “watchlist mutations”), and how mutations in the antibodies could improve binding to the S-RBD (termed “antibody strengthening mutations”). Our watchlist for potential antibody escape (Table 1) from B38 consists of 211 mutations across 35 sites, and from CB6 consists of 186 mutations across 22 sites. Some of these mutations have been previously identified as significant in observed populations (e.g., N501Y). Our list of potential antibody strengthening mutations (Table 3) for B38 consists of 116 mutations across 29 sites, and for CB6 consists of 45 mutations across 13 sites. Our watchlist provides predictions for possible reductions in efficacy of B38 and CB6 antibodies in treating some strains of SARS-CoV-2. Similar methods could be used to predict the efficacy of other therapeutics or for other variants. Furthermore, our antibody strengthening mutation list could be used to potentially improve existing therapeutic antibodies. Finally, our comparison to the Starr et al. study highlights the need for more accurate computational methods for predicting binding affinity changes due to mutations.

S-RBD-B38		S-RBD-CB6	
Site	Mutations	Site	Mutations
VH3	D,E,Q	YH52	G,L
QH4	L,M	AH60	C,I,K,L,M,Q,R,V,W
GH27	E,F,K,L,M,W,Y	RH97	K
IH29	F,M,W,Y	YH105	A,L
SH32	A,D,F,L,M,V,W,Y	IL29	F
YH34	C,I,V	SL30	M,W
VH51	C	YL49	F
YH53	A,G,S	SL52	C,F,H,I,L,M,V,Y
SH54	T	QL55	F,H,L,M,Y
GH56	W	SL56	A,C,D,F,L,M,P,R
SH57	R	SL91	F,H,W
YH59	A,C,D,E,F,G,H,I,K,L,M,N,P,Q,R,S,T,V,W	PL95	I,W
YH60	A,D,M,R,V	PL96	F
RH98	L		
EH99	A,C,D,G,H,I,L,N,Q,S,T,V,W		MD+FoldX
AH100	W		PyRosetta
YH101	Q		MD+FoldX and PyRosetta
DH104	E	Key:	Predicted to fold
VH105	D		
SL32	M		
YL50	W		
AL51	H,M,W		
QL56	L		
SL57	F,I,L,M,P,R,W,Y		
GL58	P,W		
LL92	D,E,R		
NL93	F,H,M,W,Y		
PL96	C,E,F,H,I,K,L,M,V,W,Y		
PL97	F,H,M,W,Y		

Table 3. Union of the individual datasets for predicting antibody strengthening mutations from MD + FoldX and PyRosetta. The left table contains those predicted for B38, and right is for CB6. The left column in each subtable is the site and the right column shows the list of mutations predicted to strengthen the binding interaction between the S-RBD and the antibody. Predicted strengthening mutations are colored based on whether both methods predict strengthening (red), only MD + FoldX predicts strengthening (blue), or only PyRosetta predicts strengthening (black). Bold font denotes mutations that are predicted to allow folding.

Data availability

Datasets generated for this work are publicly available and provided in full as supplemental datasets S1 and S2.

Received: 7 July 2022; Accepted: 1 November 2022

Published online: 05 November 2022

References

- Baj, J. *et al.* COVID-19: Specific and non-specific clinical manifestations and symptoms: The current state of knowledge. *J. Clin. Med.* **9**, 1753 (2020).
- Mehandru, S. & Merad, M. Pathological sequelae of long-haul COVID. *Nat. Immunol.* **23**, 194–202 (2022).
- Creech, C. B., Walker, S. C. & Samuels, R. J. SARS-CoV-2 vaccines. *JAMA* **325**, 1318–1320 (2021).
- Khandia, R. *et al.* Emergence of SARS-CoV-2 omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. *Environ. Res.* **209**, 112816 (2022).
- He, X., Hong, W., Pan, X., Lu, G. & Wei, X. SARS-CoV-2 omicron variant: Characteristics and prevention. *MedComm* **2**, 838–845 (2021).
- Planas, D. *et al.* Considerable escape of SARS-CoV-2 omicron to antibody neutralization. *Nature* **602**, 671–675 (2022).
- Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657–663 (2022).
- V'kovski, P., Kratzel, A., Steiner, S., Stalder, H. & Thiel, V. Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **19**, 155–170 (2021).
- Dejnirattisai, W. *et al.* The antigenic anatomy of SARS-CoV-2 receptor binding domain. *Cell* **184**, 2183–2200.e22 (2021).
- Piccoli, L. *et al.* Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. *Cell* **183**, 1024–1042.e21 (2020).
- Li, D., Sempowski, G. D., Saunders, K. O., Acharya, P. & Haynes, B. F. SARS-CoV-2 neutralizing antibodies for COVID-19 prevention and treatment. *Annu. Rev. Med.* **73**, 1–16 (2022).
- Wu, Y. *et al.* A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).
- Li, Z. *et al.* An engineered bispecific human monoclonal antibody against SARS-CoV-2. *Nat. Immunol.* **23**, 423–430 (2022).
- Shi, R. *et al.* A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. *Nature* **584**, 120–124 (2020).
- Gottlieb, R. L. *et al.* Effect of bamlanivimab as monotherapy or in combination with etesevimab on viral load in patients with mild to moderate COVID-19: A randomized clinical trial. *JAMA* **325**, 632–644 (2021).
- Greaney, A. J. *et al.* Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. *Cell Host Microbe* **29**, 44–57.e9 (2021).
- Starr, T. N. *et al.* Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science* **371**, 850–854 (2021).
- Starr, T. N., Greaney, A. J., Dingens, A. S. & Bloom, J. D. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep. Med.* **2**, 100255 (2021).
- Starr, T. N. *et al.* Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* **182**, 1295–1310.e20 (2020).
- Greaney, A. J., Starr, T. N. & Bloom, J. D. An antibody-escape estimator for mutations to the SARS-CoV-2 receptor-binding domain. *Virus Evolut.* **8**(1), veac021 (2022).
- Teng, S., Sobitan, A., Rhoades, R., Liu, D. & Tang, Q. Systemic effects of missense mutations on SARS-CoV-2 spike glycoprotein stability and receptor-binding affinity. *Brief. Bioinform.* **22**, 1239–1253 (2021).
- Tsai, K.-C., Lee, Y.-C. & Tseng, T.-S. Comprehensive deep mutational scanning reveals the immune-escaping hotspots of SARS-CoV-2 receptor-binding domain targeting neutralizing antibodies. *Front. Microbiol.* **12**, (2021).
- Sharma, D., Rawat, P., Janakiraman, V. & Gromiha, M. M. Elucidating important structural features for the binding affinity of spike - SARS-CoV-2 neutralizing antibody complexes. *Prot. Struct. Funct. Bioinform.* **90**, 824–834 (2022).
- Ray, D., Quijano, R. N. & Andricioaei, I. Point mutations in SARS-CoV-2 variants induce long-range dynamical perturbations in neutralizing antibodies. *Chem. Sci.* **13**, 7224–7239 (2022).
- Hendy, M., Kaufman, S. & Ponga, M. Molecular strategies for antibody binding and escape of SARS-CoV-2 and its mutations. *Sci. Rep.* **11**, 21735 (2021).
- Miller, N. L., Clark, T., Raman, R. & Sasisekharan, R. Insights on the mutational landscape of the SARS-CoV-2 omicron variant receptor-binding domain. *Cell Rep. Med.* **3**, 100527 (2022).
- Laurini, E., Marson, D., Aulic, S., Fermeglia, A. & Priel, S. Molecular rationale for SARS-CoV-2 spike circulating mutations able to escape bamlanivimab and etesevimab monoclonal antibodies. *Sci. Rep.* **11**, 20274 (2021).
- Miller, C. R. *et al.* Initiating a watch list for Ebola virus antibody escape mutations. *PeerJ* **4**, e1674 (2016).
- Patel, J. S., Quates, C. J., Johnson, E. L. & Ytreberg, F. M. Expanding the watch list for potential ebola virus antibody escape mutations. *PLoS ONE* **14**, e0211093 (2019).
- Schymkowitz, J. *et al.* The FoldX web server: An online force field. *Nucleic Acids Res.* **33**, W382–W388 (2005).
- Gonzalez, T. R., Martin, K. P., Barnes, J. E., Patel, J. S. & Ytreberg, F. M. Assessment of software methods for estimating protein-protein relative binding affinities. *PLoS ONE* **15**, e0240573 (2020).
- Berman, H. M., Westbrook, J. & Feng, Z. The protein databank. *Nucleic Acids Res.* **28**, 235–242 (2000).
- Eastman, P., Swails, J. & Chodera, J. D. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, (2017).
- Chaudhury, S., Lyskov, S. & Gray, J. J. PyRosetta: A script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* **26**, 689–691 (2010).
- Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
- Barlow, K. A. *et al.* Flex ddG: Rosetta ensemble-based estimation of changes in protein-protein binding affinity upon mutation. *J. Phys. Chem. B* **122**, 5389–5399 (2018).
- Tyka, M. D. *et al.* Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618 (2011).
- Lauck, F., Smith, C. A., Friedland, G. F., Humphris, E. L. & Kortemme, T. RosettaBackrub—a web server for flexible backbone protein structure modeling and design. *Nucleic Acids Res.* **38**, W569–W575 (2010).
- Shapovalov, M. V. & Dunbrack, R. L. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844–858 (2011).
- Caswell, T. A. *et al.* matplotlib/matplotlib: REL: v3.5.2. Doi: <https://doi.org/10.5281/zenodo.6513224>(2022).
- Khan, A. *et al.* Higher infectivity of the SARS-CoV-2 new variants is associated with K417N/T, E484K, and N501Y mutants: An insight from structural data. *J. Cell. Physiol.* **236**, 7045–7057 (2021).
- Lu, L. *et al.* The impact of spike N501Y mutation on neutralizing activity and RBD binding of SARS-CoV-2 convalescent serum. *EBioMedicine* **71**, 103544 (2021).

43. Huang, H., Zhu, Y., Niu, Z., Zhou, L. & Sun, Q. SARS-CoV-2 N501Y variants of concern and their potential transmission by mouse. *Cell Death Differ.* **28**, 2840–2842 (2021).
44. Colson, P. *et al.* Spreading of a new SARS-CoV-2 N501Y spike variant in a new lineage. *Clin. Microbiol. Infect.* **27**, 1352e1–1352.e5 (2021).
45. Starr, T. N. *et al.* Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. 2022.02.24.481899 Preprint at Doi: <https://doi.org/10.1101/2022.02.24.481899>(2022).
46. Cantón, R. *et al.* New variants of SARS-CoV-2. *Rev. Esp. Quimioter.* **34**, 419–428 (2021).
47. Chakraborty, S. E484K and N501Y SARS-CoV 2 spike mutants Increase ACE2 recognition but reduce affinity for neutralizing antibody. *Int. Immunopharmacol.* **102**, 108424 (2022).
48. Du, P., Gao, G. F. & Wang, Q. The mysterious origins of the Omicron variant of SARS-CoV-2. *Innovation* **3**, 100206 (2022).
49. Ma, W. *et al.* Genomic perspectives on the emerging SARS-CoV-2 omicron variant. *Genomics Proteomics Bioinform.* <https://doi.org/10.1016/j.gpb.2022.01.001> (2022).
50. Flemming, A. Omicron, the great escape artist. *Nat. Rev. Immunol.* **22**, 75–75 (2022).
51. Cameroni, E. *et al.* Broadly neutralizing antibodies overcome SARS-CoV-2 Omicron antigenic shift. *Nature* **602**, 664–670 (2022).
52. Lippow, S. M., Wittrup, K. D. & Tidor, B. Computational design of antibody-affinity improvement beyond in vivo maturation. *Nat. Biotechnol.* **25**, 1171–1176 (2007).
53. Beach, S. S., Hull, M. A., Ytreberg, F. M., Patel, J. S. & Miura, T. A. Molecular modeling predicts novel antibody escape mutations in the respiratory syncytial virus fusion glycoprotein. *J. Virol.* **0**, e00353–22 (2022).
54. Zeng, H.-L., Dichio, V., Rodríguez Horta, E., Thorell, K. & Aurell, E. Global analysis of more than 50,000 SARS-CoV-2 genomes reveals epistasis between eight viral genes. *Proc. Nat. Acad. Sci.* **117**, 31519–31526 (2020).

Acknowledgements

The research was supported by the Institute for Modeling Collaboration and Innovation sponsored by the National Institute of General Medical Sciences (<https://www.nigms.nih.gov>) under award number P20 GM104420 and the National Science Foundation (<https://www.nsf.gov>) EPSCoR Track-II under award number OIA1736253. Computer resources were provided in part by the Institute for Interdisciplinary Data Sciences Research Computing and Data Services sponsored by the National Institutes of Health (NIH P30 GM103324). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

J.S.P. and F.M.Y. were supervisors. J.E.B. performed the molecular dynamics simulations and FoldX estimates. P.K.L. performed the Rosetta calculations. J.E.B. prepared all figures and tables. J.E.B. and P.K.L. performed the analysis and wrote the original draft of the manuscript. All authors reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-23482-5>.

Correspondence and requests for materials should be addressed to J.S.P. or F.M.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022