Predictions of the SARS-CoV-2 Omicron Variant (B.1.1.529) Spike Protein Receptor-Binding Domain Structure and Neutralizing Antibody Interactions

Colby T. Ford[1, 2, ✉], **Denis Jacob Machado**[1], **and Daniel A. Janies**[1]

[1]Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA
[2]School of Data Science, University of North Carolina at Charlotte, Charlotte, NC USA

The genome of the SARS-CoV-2 Omicron variant (B.1.1.529) was released on November 22, 2021, which has caused a flurry of media attention due the large number of mutations it contains. These raw data have spurred questions around vaccine efficacy. Given that neither the structural information nor the experimentally-derived antibody interaction of this variant are available, we have turned to predictive computational methods to model the mutated structure of the spike protein's receptor binding domain and posit potential changes to vaccine efficacy. In this study, we predict some structural changes in the receptor-binding domain that may reduce antibody interaction without completely evading existing neutralizing antibodies (and therefore current vaccines).

SARS-CoV-2 | COVID-19 | Protein Structure Analysis | Antibodies
Bioinformatics | Computational Biology | Machine Learning

Correspondence: ***colby.ford@uncc.edu***

Introduction

A team of researchers from the Botswana-Harvard HIV Reference Laboratory submitted a new SARS-CoV-2 genome sequence to GISAID on November 22, 2021 (GISAID accession. EPI_ISL_6752027). The specimen was taken from a living 59-year-old male from Gaborone, Botswana using a nasopharyngeal swab and was sequenced using a Nanopore MinION device.

This sample's genome contains 60 mutations from the Wuhan-derived reference genome (GenBank accession no. NC_045512.2) (1), 37 of which are in the Spike (S) protein. This variant was given the identifier B.1.1.529 by PANGO lineages (2). On November 26, 2021, the WHO has designated B.1.1.529 as a Variant of Concern (VOC), named Omicron (3).

The emergence of new SARS-CoV-2 variants is expected. Therefore, scientists have advocated for close international monitoring to determine the need for vaccination boosters and redesign (4). Hence, the identification of the omicron variant is not surprising. What is surprising is the number of mutations that the omicron variant accumulated compared to the first sequenced genome of SARS-CoV-2.

Different authors have warned that limited SARS-CoV-2 sampling and sequencing from positive cases, especially from asymptomatic and symptomatic cases that did not require hospitalization, would make it challenging to identify new mutations in the virus. For example, Brito et al. (5) analyzed the spatiotemporal heterogeneity in each country's SARS-CoV-2 genomic surveillance efforts using metadata submitted to GISAID until May 30, 2021. These authors calculate that sequencing capacity should be at least 0.5% of cases per week when incidence is more than 100 positive cases every 100,000 people. Unfortunately, most countries are not reaching this sequence threshold.

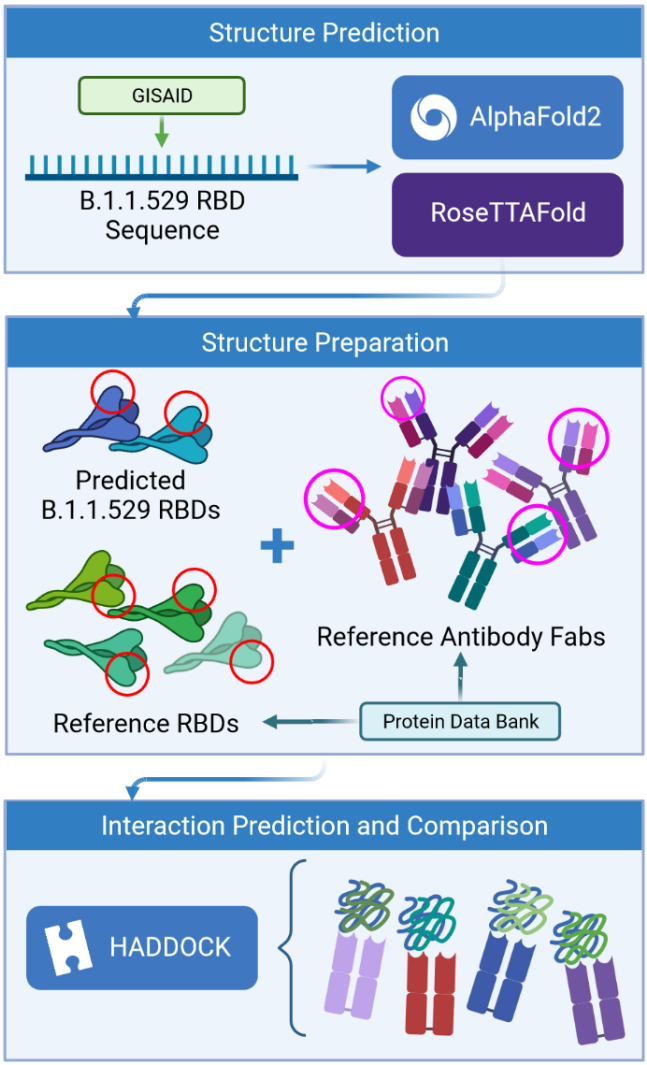

Fig. 1. Process flow of the prediction analysis steps.

While sampling bias can explain why we may miss new mutations and fail to identify new variants of low prevalence, the emergence of new variants is due to factors that favor

the transmission of SARS-CoV-2, including low vaccination rates in some regions, especially in low and middle-income countries (LMICs). Therefore, disparities in vaccination rates combined with sampling bias explain why scientists may continue to be surprised by the mutations in new SARS-CoV-2 variants.

There are many questions regarding genomic epidemiology and the lessons we can learn from the COVID-19 pandemic. Those questions are beyond the scope of this manuscript, and we addressed them elsewhere (6). While the origin and evolution of the Omicron variant are still open questions, here we focus on the potential implications of the mutations observed in this variant.

This seemingly hyper-mutated variant is of public health concern with unanswered questions surrounding vaccine protection (from available vaccines), the possibility of reinfection, transmissibility, and pathogenicity.

Regarding vaccine efficacy, we must look at the receptor-binding domain (RBD), part of the S1 subunit, of the spike protein as this is the binding site for neutralizing antibodies. This domain exists between positions 319 and 541 of the spike protein. Omicron contains 15 mutations in the RBD, none of which are deletions or insertions. In contrast, the Delta variant contains 7 mutations across the entire spike protein, only 2 of which are in the RBD.

Given that an experimentally-derived structure of the Omicron spike protein is not yet available, we must derive a predicted structure from its sequence *in silico*. Then, we can use available neutralizing antibody structures to computationally model the interaction between Omicron and the paratopes of the antibodies, thus allowing us to compare potential affinity changes due to the mutations and posit their effects to vaccine efficacy.

Methods

Sequence Comparison among VBMs and VOCs. We downloaded the reference genome of SARS-CoV-2 (Wuhan-Hu-1, NCBI's RefSeq accession no. NC_045512.2) as well as the first 100 complete genome sequences ($\geq$29,000 bp) of each Variant of Concern (VOC) and Variant Being Monitored (VBM). The total number of input sequences was 1,301. We aligned all of these complete genomes using `MAFFT` version 7.475 (7) with the "auto" option and trimmed the alignment to remove the 5'-UTR and 3'-UTR regions. We also removed duplicated sequences or sequences with more than 5% of missing data, leaving us with 1,026 sequences.

We annotated each of the 1,026 remaining sequences using the strategy described in Machado et al. (2021) (8). Once we had all the predicted spike proteins for each of the 1,026 genomes, we aligned those sequences based on their translation with the help of `MAFFT` using the `TranslatorX` pipeline (9). We removed duplicated sequences and sequences with more than 5% of missing data. Finally, we identified the receptor binding motif on that alignment based on sequence similarity with the reference.

We then calculated the pairwise p-distances between each pair of sequences were calculated using `MEGA` version 11.0.10 (10). This distance is the proportion (p) of nucleotide sites at which two sequences being compared are different. The p-distances were calculated for the whole spike alignments (nucleotides) but also for the alignment of its receptor binding motif (RBM; position 430–522 of the spike amino acid sequence, a subset of the positions in the RBD).

This variant nucleotide sequence for the spike protein was then translated into amino acids using the standard translation table. This sequence was then trimmed to only contain the RBD of the spike protein (positions 319 to 541).

Receptor-Binding Domain Structural Prediction. Using the derived RBD amino acid sequence for Omicron, we used `AlphaFold2` and `RoseTTAFold` to create a predicted 3D protein structures. `AlphaFold2` is a neural network-based deep learning model created by Google DeepMind (11). The algorithm first searches for homologous sequences with existing structures to use as a scaffold on which to place the new sequence. `RoseTTAFold` is a similar neural network-based system by the Institute for Protein Design at the University of Washington (12).

The `AlphaFold2`-based prediction was run with the "single sequence" mode using the predicted TM-score (PTM) method. We also specified that the algorithm should run an Amber relaxation procedure to repair any structural violations in the predicted model (13). The `RoseTTAFold`-based prediction was run with the "mmseqs2" mode (by ColabFold (14)).

Both systems each resulted in a .PDB file of the predicted RBD structure for Omicron along with metrics surrounding the multiple sequence alignment coverage, predicted aligned error (PAE), and predicted confidence (pLDDT) by position (available in Supplementary Materials).

Given that this study focuses on antibodies that bind to the top of the RBD of the spike protein. Since that `AlphaFold2` and `RoseTTAFold` are template-based models that generate the predicted RBD structures using homologous sequences for which we have actual structures, we can avoid modeling the entire spike protein.

Neutralizing Antibody Interaction Simulation. Using the predicted structures of the Omicron RBD, we simulated the interaction with four available neutralizing antibody structures: C105, CC12.1, CC12.3, and CV30 (PDBs: 6XCM, 6XC2, 6XC7, and 6XE1, respectively) (15–17). We used only a single fragment antigen-binding (Fab) region of the antibody structures as the paratope location against which to dock. Each of the RBD structures from these reference files have identical sequences to the Wuhan-Hu-1 spike RBD.

Each of these neutralizing antibody structures were collected from patients who had been infected with SARS-CoV-2. Thus, these structures are serologically-derived antibodies rather than structures of therapeutic antibodies. All of them bind to the same "up" location of the S1 subunit of the spike protein (class I binders). This is a similar location to the interaction site between the human ACE2 receptor epitope. Thus, the neutralizing mechanism of these antibodies is in the prevention of SARS-CoV-2 binding to ACE2 on human cells.

We used HADDOCK version 2.4, a biomolecular modeling software that provides docking predictions for provided structures, to predict the binding affinity between the epitope of the RBD with the paratope of the neutralizing antibody structures (18). This takes in two or more .PDB files as inputs and outputs multiple predicted protein complexes in .PDB format along with docking metrics.

We first renumbered the residues according to HADDOCK's requirements and then specified the interacting residues between the RBD structure and the Fab. Specifically, ensure there are not overlapping residue IDs between the chains of a .PDB file and then specify the residues that are assumed to interact between the structures. This analysis was performed on the antibody-RBD structure pairs shown in Table 1.

Antibody Fab	Analysis Type	RBD Source
C105 (6XCM, chains N and S)	Reference	6XCM (chain B)
	Prediction	B.1.1.529 (from AlphaFold2)
	Prediction	B.1.1.529 (from RoseTTAFold)
CC12.1 (6XC2, chains H and L)	Reference	6XC2 (chain A)
	Prediction	B.1.1.529 (from AlphaFold2)
	Prediction	B.1.1.529 (from RoseTTAFold)
CC12.3 (6XC7, chains C and D)	Reference	6XC7 (chain A)
	Prediction	B.1.1.529 (from AlphaFold2)
	Prediction	B.1.1.529 (from RoseTTAFold)
CV30 (6XE1, chains H and L)	Reference	6XE1 (chain B)
	Prediction	B.1.1.529 (from AlphaFold2)
	Prediction	B.1.1.529 (from RoseTTAFold)

Table 1. List of analyses performed, comparing reference and predicted RBD structures in complex with reference Fab structures.

The assessment of these interactions was measured by multiple biophysical factors including van der Waals energy, electrostatic energy, desolvation energy, and restraints violation energy, which were collectively used to derive a HADDOCK score to quantify changes in protein-protein interaction resulting from mutations in the RBD. Further, interfacing residues between the RBD and Fab structures were determined by identifying residues above the 1.0 $Å^2$ difference in surface area cutoff between the chains of the RBD and the Fab using the *InterfaceResidues*[1] functionality in PyMol version 2.4.1 (19).

We then compared the metrics of the actual complexes (i.e., the real RBD structure and the Fab) versus the predicted RBD structures of Omicron (with the same Fab). This provides a baseline interaction that was then measured against the mutated interactions with each respective Fab.

[1] *InterfaceResidues* Function Documentation: https://pymolwiki.org/index.php/InterfaceResidues

Differences between HADDOCK results were assessed using Kruskal-Wallis tests and ad hoc Wilcoxon pairwise comparisons. All statistical analyses were performed using R version 4.0.4 (20).

Results

Variant Sequence Comparison. Although the Omicron RBM (spike amino acid sequence, positions 430–522) can be efficiently categorized by nine characteristic mutations (S:N440K, S:G446S, S:S447N, S:T478K, S:E484A, S:Q493R, S:G496S, S:Q298R, S:N501Y), at least two of them (S:N440K and S:G446S) may be missing from some samples classified as Omicron. Also, some Omicron RBMs contains an additional mutation at S:Y505H.

The Omicron variant is the variant more distantly related to the reference genome (SARS-CoV-2 Wuhan-Hu-1; NCBI's RefSeq accession no. NC_045512.2) in the proportion of shared nucleotides. Also, Omicron is the variant that is more distantly related to Gamma. See Figures 2 and 3.

Variant	1	2	3	4	5	6	7	8	9	10	11
1. Reference: Wuhan-Hu-1	0.00%										
2. VBM: B.1.617.3	0.19%	0.00%									
3. VBM: Beta	0.20%	0.32%	0.00%								
4. VBM: Epsilon	0.15%	0.23%	0.31%	0.00%							
5. VBM: Eta	0.23%	0.33%	0.33%	0.32%	0.00%						
6. VBM: Gamma	0.33%	0.43%	0.38%	0.41%	0.46%	0.00%					
7. VBM: Iota	0.17%	0.29%	0.27%	0.26%	0.33%	0.43%	0.00%				
8. VBM: Kappa	0.24%	0.20%	0.37%	0.28%	0.39%	0.49%	0.32%	0.00%			
9. VBM: Mu	0.22%	0.26%	0.27%	0.32%	0.33%	0.40%	0.27%	0.33%	0.00%		
10. VBM: Zeta	0.11%	0.22%	0.21%	0.21%	0.23%	0.29%	0.20%	0.27%	0.23%	0.00%	
11. VOC: Delta	0.23%	0.14%	0.39%	0.26%	0.40%	0.50%	0.34%	0.29%	0.33%	0.29%	0.00%
12. VOC: Omicron	0.81%	0.88%	0.90%	0.90%	0.88%	0.98%	0.84%	0.90%	0.79%	0.86%	0.88%

Proportion (p) 0.00% – 0.98%

Fig. 2. Distance matrix of the spike gene (using nucleotides) for 9 Variants Being Monitored (VBM) and 2 Variants of Concern (VOC). The distance is the average proportion (*p*) of nucleotide sites at which two sequences being compared are different.

Variant	1	2	3	4	5	6	7	8	9	10	11
1. Reference: Wuhan-Hu-1	0.00%										
2. VBM: B.1.617.3	2.20%	0.00%									
3. VBM: Beta	2.20%	3.33%	0.00%								
4. VBM: Epsilon	1.09%	1.09%	3.33%	0.00%							
5. VBM: Eta	1.25%	2.36%	1.25%	2.36%	0.00%						
6. VBM: Gamma	2.20%	3.33%	0.00%	3.33%	1.25%	0.00%					
7. VBM: Iota	0.82%	2.67%	2.30%	1.92%	1.34%	2.30%	0.00%				
8. VBM: Kappa	2.20%	0.00%	3.33%	1.09%	2.36%	3.33%	2.67%	0.00%			
9. VBM: Mu	2.20%	3.33%	0.00%	3.33%	1.25%	0.00%	2.30%	3.33%	0.00%		
10. VBM: Zeta	1.13%	2.28%	1.21%	2.24%	0.28%	1.21%	1.22%	2.28%	1.21%	0.00%	
11. VOC: Delta	2.12%	2.12%	4.42%	1.02%	3.42%	4.42%	2.97%	2.12%	4.42%	3.30%	0.00%
12. VOC: Omicron	11.09%	12.43%	9.78%	12.43%	11.28%	9.78%	10.76%	12.43%	9.78%	11.09%	11.27%

Proportion (p) 0.00% – 12.43%

Fig. 3. Distance matrix of the receptor binding motif (RBM) of spike gene (using amino acids) for 9 Variants Being Monitored (VBM) and 2 Variants of Concern (VOC). The distance is the average proportion (*p*) of nucleotide sites at which two sequences being compared are different.

Mutational Analysis. Comparing the RBD of Omicron to the reference genome, there are 15 mutations, all of which are single amino acid substitutions. Most of the substitutions result in a change in the residue type. (See Table 2.)

Position	Ref	Alt	Ref Type	Alt Type	Type Difference
339	G	D	Non-Polar	Negative	Yes
371	S	L	Polar	Non-Polar	Yes
373	S	P	Polar	Non-Polar	Yes
375	S	F	Polar	Non-Polar	Yes
417	K	N	Positive	Polar	Yes
440	N	K	Polar	Positive	Yes
446	G	S	Non-Polar	Polar	Yes
477	S	N	Polar	Polar	No
478	T	K	Polar	Positive	Yes
484	E	A	Negative	Non-Polar	Yes
493	Q	R	Polar	Positive	Yes
496	G	S	Non-Polar	Polar	Yes
498	Q	R	Polar	Positive	Yes
501	N	Y	Polar	Polar	No
505	Y	H	Polar	Positive	Yes

Table 2. Mutations in the receptor-binding domain (RBD) of the spike protein in the Omicron variant (B.1.1.529).

The resulting RBD structure from `AlphaFold2` predicts that there is little conformational change from the reference structure. Conversely, there is a significant conformational change in the predicted RBD structure from `RoseTTAFold`. See Figure 4.

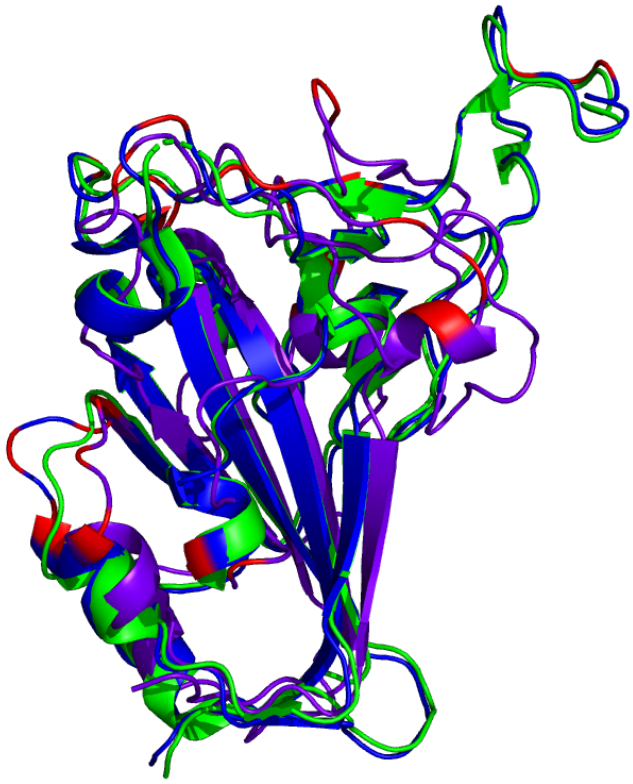

Fig. 4. Comparison of reference RBD structure (PDB: 6XC2, shown in green) and the predicted Omicron (B.1.1.529) RBD structures (`AlphaFold2` shown in blue, `RoseTTAFold` shown in purple). Mutated residues are highlighted in red.

There are multiple mutated residues (shown in red in Figures 5 and 6) in positions that may affect the ability of a neutralizing antibody to sufficiently bind. Some of these mutated residues change to much longer side-chained or differently-charged amino acids. For example, there are two "to lysine" mutations: N440K and T478K (i.e., from polar, smaller side chain residues to a positive-charged, longer side chain residue). These types of changes may have an effect on the binding affinity between the RBD and an antibody, either by changing the surface charge on the protein or by inhibiting a tighter antibody interaction.

Fig. 5. Possible inhibitory mutated RBD residues in the `AlphaFold2` Omicron (B.1.1.529) (structure shown in blue with mutated residues of interest shown in red) superimposed on a reference RBD structure. Note: the reference RBD structure (PDB: 6XC2) is shown in green with equivalent position residues highlighted in yellow. CC12.1 antibody Fab (from PDB 6XC2) is shown in magenta / pink.

Superimposing the predicted RBDs on the reference RBD with the CC12.1 antibody in place, shown in Figures 5 and 6, shows that mutations Q493R, Q498R, and N501Y in the `AlphaFold2` structure and the E484A mutation in the `RoseTTAFold` structure may affect the binding position of the antibody. These side chains clash with particular residues of the Fab, which may cause a less effective and more distant binding. Longer/larger side chains may increase the distance between the Fab paratope of the antibody and the epitope of the RBD.

Fig. 6. Possible inhibitory mutated RBD residues in the RoseTTAFold Omicron (B.1.1.529) (structure shown in purple with mutated residues of interest shown in red) superimposed on a reference RBD structure. Note: the reference RBD structure (PDB: 6XC2) is shown in green with equivalent position residues highlighted in yellow. CC12.1 antibody Fab (from PDB 6XC2) is shown in magenta / pink.

Antibody Binding Analyses. The results of all four antibody docking exercises with the predicted AlphaFold2 RBD structure show that the Fab of the respective neutralizing antibodies continue to bind to the RBD of Omicron, though not as well as the reference interaction. Note that there is a consistent decrease (increase in value) in the electrostatic energy and an increase in restraints violation energy between the binding from the reference RBDs and the predicted RBD of Omicron. The HADDOCK score is worse (higher) across the board and it appears that the interaction of the Omicron RBD with the antibodies are more distant, as shown by the buried surface area changes below.

When looking at the antibody docking exercises with the predicted RoseTTAFold RBD structure, there is agreement with the AlphaFold2 results in that there still seems to be interaction with the neutralizing antibodies. However, the reduction in binding affinity is much more severe given the conformational changes seen only in the RoseTTAFold structure.

Note: All values in Tables 3, 4, 5, and 6 below represent the best docking predictions from HADDOCK. Also, values in parentheses represent the percentage difference between the given metric for the predicted structure and the reference structure.

C105 Antibody Binding. Resulting binding metrics from the C105 HADDOCK docking analysis are shown in Table 3. This interaction shows that there is a $\sim 4\%$ reduction in the electrostatic energy and an $\sim 5\%$ decrease in buried surface area comparing between the 6XCM RBD and the predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is $\sim 16\%$ increase in the electrostatic energy but with a $\sim 36\%$ decrease in buried surface area.

Metric	6XCM RBD w/ C105	Predicted B.1.1.529 RBD w/ C105	
		AlphaFold	RoseTTAFold
van der Waals energy	-85	-71.8 (16%)	-44.2 (48%)
Electrostatic energy	-280.9	-268.9 (4%)	-324.5 (-16%)
Desolvation energy	-18.6	-17.1 (8%)	-3.2 (83%)
Restraints violation energy	154.1	158.8 (3%)	338.6 (120%)
HADDOCK score	-144.4	-126.8 (12%)	-78.4 (46%)
Buried Surface Area	2417.8	2299.6 (-5%)	1555.1 (-36%)

Table 3. HADDOCK metrics for the C105 docking prediction, comparing the 6XCM RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Fig. 7. HADDOCK docking prediction using C105 (shown in magenta / pink), comparing the 6XCM RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (AlphaFold2 shown in blue, RoseTTAFold shown in purple).

CC12.1 Antibody Binding. Resulting binding metrics from the CC12.1 HADDOCK docking analysis are shown in Table 4. This interaction shows that there is a $\sim 17\%$ reduction in the electrostatic energy and an $\sim 7\%$ decrease in buried surface area comparing between the 6XC2 RBD and predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is $\sim 24\%$ reduction in the electrostatic energy but and a $\sim 43\%$ decrease in buried surface area.

Metric	6XC2 RBD w/ CC12.1	Predicted B.1.1.529 RBD w/ CC12.1	
		AlphaFold	RoseTTAFold
van der Waals energy	-106.9	-90.7 (15%)	-27.5 (74%)
Electrostatic energy	-342.6	-284.9 (17%)	-259.2 (24%)
Desolvation energy	-37.5	-28.8 (23%)	-12.4 (67%)
Restraints violation energy	143.2	152.6 (7%)	201.4 (41%)
HADDOCK score	-198.6	-163.3 (18%)	-71.6 (64%)
Buried Surface Area	2778.5	2584.3 (-7%)	1578.4 (-43%)

Table 4. HADDOCK metrics for the CC12.1 docking prediction, comparing the 6XC2 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

CC12.3 Antibody Binding. Resulting binding metrics from the CC12.3 HADDOCK docking analysis are shown in Table 4. This interaction shows that there is a $\sim 22\%$ reduction in the electrostatic energy and an $\sim 3\%$ decrease in buried surface area comparing between the 6XC7 RBD and predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is $\sim 24\%$ reduction in the electrostatic energy and a $\sim 43\%$ decrease in buried surface area.

Metric	6XC7 RBD w/ CC12.3	Predicted B.1.1.529 RBD w/ CC12.3	
		AlphaFold	RoseTTAFold
van der Waals energy	-86.1	-93.1 (-8%)	-48.1 (44%)
Electrostatic energy	-248.1	-194.1 (22%)	-189.2 (24%)
Desolvation energy	-41.7	-37.9 (9%)	-23.5 (44%)
Restraints violation energy	159.1	104.7 (-34%)	221.4 (39%)
HADDOCK score	-161.5	-159.4 (1%)	-87.4 (46%)
Buried Surface Area	2489.8	2416.7 (-3%)	1420.6 (-43%)

Table 5. HADDOCK metrics for the CC12.3 docking prediction, comparing the 6XC7 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

Fig. 8. HADDOCK docking prediction using CC12.1 (shown in magenta / pink), comparing the 6XC2 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (AlphaFold2 shown in blue, RoseTTAFold shown in purple).

Fig. 9. HADDOCK docking prediction using CC12.3 (shown in magenta / pink), comparing the 6XC7 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (AlphaFold2 shown in blue, RoseTTAFold shown in purple).

CV30 Antibody Binding. Resulting binding metrics from the CV30 HADDOCK docking analysis are shown in Table 6. This interaction shows that there is a $\sim 52\%$ reduction in the electrostatic energy and an $\sim 20\%$ decrease in buried surface area comparing between the 6XE1 RBD and predicted RBD of Omicron from AlphaFold2. From RoseTTAFold, the interaction shows that there is $\sim 41\%$ reduction in the electrostatic energy and a $\sim 35\%$ decrease in buried surface area.

Metric	6XE1 RBD w/ CV30	Predicted B.1.1.529 RBD w/ CV30	
		AlphaFold	RoseTTAFold
van der Waals energy	-86.1	-61.7 (28%)	-41.1 (52%)
Electrostatic energy	-354.4	-171.8 (52%)	-207.9 (41%)
Desolvation energy	-13.8	-19.5 (-41%)	-10.2 (26%)
Restraints violation energy	147.6	159.4 (8%)	286.0 (93%)
HADDOCK score	-156.0	-99.6 (36%)	-43.9 (71%)
Buried Surface Area	2479.0	1992.6 (-20%)	1600.2 (-35%)

Table 6. HADDOCK metrics for the CV30 docking prediction, comparing the 6XE1 RBD vs. the Omicron (B.1.1.529) predicted RBD structures.

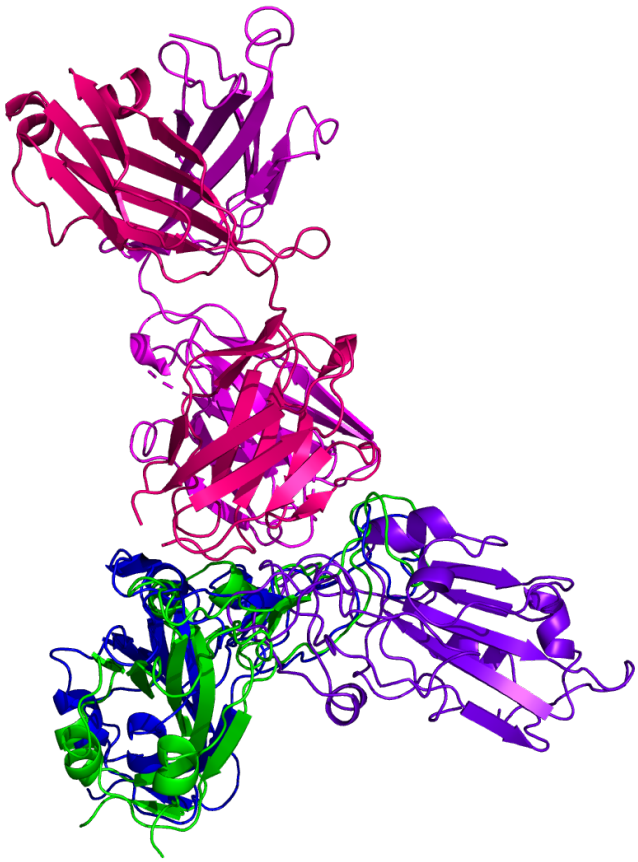

Fig. 10. HADDOCK docking prediction using CV30 (shown in magenta / pink), comparing the 6XE1 RBD (shown in shown in green) vs. the Omicron (B.1.1.529) predicted RBD structures (AlphaFold2 shown in blue, RoseTTAFold shown in purple).

Antibody Interaction Comparison. All of the interaction predictions among the four antibodies tested in this study (C105, CC12.1, CC12.3, and CV30) agree that there is a decrease in binding affinity when comparing the respective RBD interactions with the Omicron RBD interactions. Across all of the docking predictions using the AlphaFold2 RBD structure, we see a drop in electrostatic interaction (increase in the electrostatic energy value) ranging from $\sim 4\%$ to $\sim 52\%$ and a consistent decrease in buried surface area (increase in distance) of the RBD and the antibody Fab. In addition, we see a variable increase (worsening) in the HADDOCK score, indicating that all of the Omicron RBD structures have a lower binding affinity when compared to their respective reference RBD structures as a benchmark. See Figure 11.

Similarly, we see a more extreme reduction in the binding affinity of the Fab structures and the RBD structure predicted by RoseTTAFold. We see a drop in electrostatic interaction (increase in the electrostatic energy value) ranging from $\sim 24\%$ to $\sim 41\%$ in three of the four interactions (with an odd $\sim 16\%$ increase in the C105 interaction) and a consistent $\sim 35\%$ to $\sim 43\%$ decrease in buried surface area (increase in distance) of the RBD and the antibody Fab.

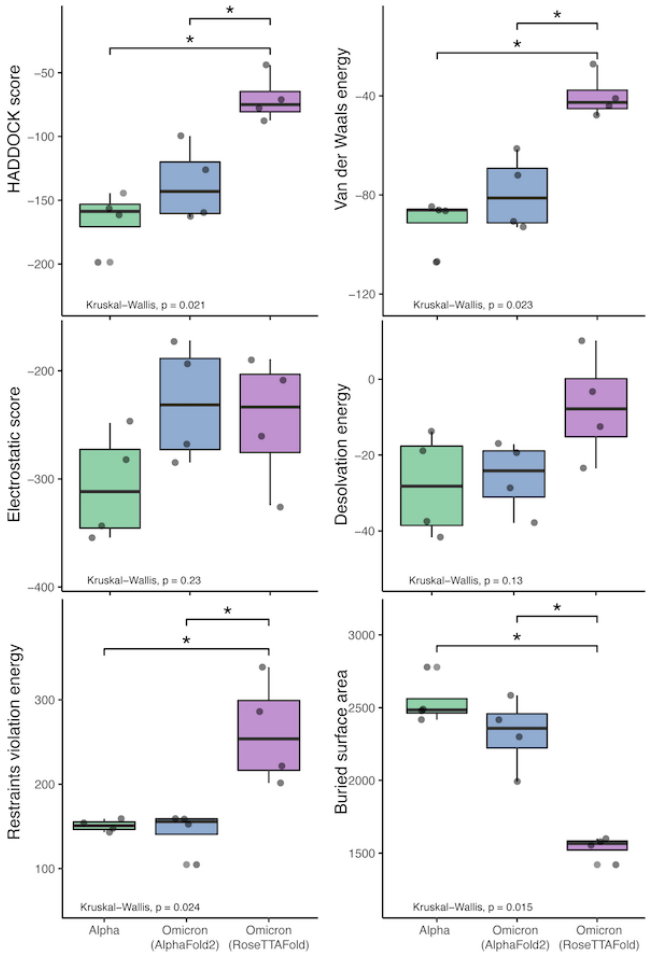

Fig. 11. HADDOCK results comparison between the reference RBD structures and the predicted Omicron (B.1.1.529) RBD structures. The bars indicate results that are significantly different and the "*" indicates Wilcoxon Rank-Sum's p-value ≤ 0.005).

We performed Kruskal-Wallis tests with each of the HADDOCK metrics, which show that there is a statistically-significant difference between the three sets of binding experiments (Alpha and the two predicted Omicron sets). See Figure 11. Interestingly, performing the Wilcoxon Rank-Sum tests on these metrics to compare the differences between the predictions and reference results shows that there is no statistically-significant difference at the $\alpha = 0.05$ level between the AlphaFold2 RBD structure binding and the reference structures. However, when comparing the RoseTTAFold RBD structure to the reference results, differences in the restraints violation energy, van der Waals energy, HADDOCK score, and buried surface area are statistically significant at the $\alpha = 0.05$ level and the differences in desolvation energy is statistically significant at the $\alpha = 0.10$ level. Also, the same metrics are significantly different between the AlphaFold2 and RoseTTAFold, further showing that these predicted structures are quite different from one another.

Fab-RBD Interfacing Residues. Furthermore, when comparing residues that are interfaced between the Fab and RBD, there is agreement in that particular residues in Omicron are no longer interfacing with the antibodies analyzed in this study. In particular, residues 448N, 484A, and 494S may not interface with the Fab structure as they are in the reference RBD-Fab complexes. However, the aforementioned N501Y and S477N mutations (along with a variety of other mutations) do not appear to affect the interfacing of the residues at these positions.

This implies that there are certain positions that are more sensitive to mutations in that substitutions at these loci are more likely to affect the interface of the RBD with the antibody's Fab (denoted by a △ symbol in Table 7).

In contrast, there are other positions that have been substituted between the reference RBDs and the predicted Omicron RBDs that continue to interface with most of the Fab structures (denoted by a ♦ symbol in Table 7).

Finally, there are some residues that remain unchanged in the Omicron variant RBD structure, yet we see changes in the interfacing at these loci (denoted by a ○ symbol in Table 7). This suggests that there are other mutated residues around these stable positions that may be affecting their ability to interface.

Many of these findings concur with the findings in Sharma et al. (21). Specifically, we also continue to see a reliance on residues at positions K417, S477, Q493, Q498, N501, and Y505, which are stated to increase binding affinity. Interestingly, the interfacing residues that are non-mutated N448 and S494 (denoted with a ○ symbol) are rarely, if ever, listed as interfacing residues in Table 3 in Sharma et al. (21). This further supports the notion that there are residue positions that are important for interfacing with neutralizing antibodies, but that the mutations seen here in the Omicron RBD may not seriously affect this RBD-antibody interface.

Interface Position		**417** (K▶N)	**448** (N)	**477** (S▶N)	**484** (E▶A)	**493** (Q▶R)	**494** (S)	**498** (Q▶R)	**501** (N▶Y)	**505** (Y▶H)
6XC2 vs. B.1.1.529	Reference	K	N	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	-	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
6XC7 vs. B.1.1.529	Reference	K	N	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	-	R	Y	H
	RoseTTAFold Prediction	-	-	N	-	R	-	R	Y	-
6XCM vs. B.1.1.529	Reference	K	-	S	E	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	N	-	R	S	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
6XE1 vs. B.1.1.529	Reference	K	-	S	-	Q	S	Q	N	Y
	AlphaFold2 Prediction	N	-	-	A	R	S	R	Y	H
	RoseTTAFold Prediction	-	N	N	A	-	-	-	-	H
Legend: ♦ - Mutated positions with few interface changes, ○ - Non-mutated positions with interface changes, △ - Mutated positions with interface changes		△	○	♦	△	△	○	△	△	♦

Table 7. Interfacing residue changes of interest between the Fab paratope and the RBD structures. (Note that a '-' means that the residue at this position no longer interfaces with the Fab structure.)

Conclusion and Discussion

While *in vitro* experiments are needed to validate these predictions, the predicted results here suggest that existing neutralizing antibodies may still bind to the mutated spike protein of the Omicron variant. However, it appears that the affinity of Omicron's RBD for neutralizing antibodies is reduced compared to the reference RBD structures. The results of both `AlphaFold2` and `RoseTTAFold` suggest that antibodies elicited from previous infection will provide at least some protection against Omicron. Additionally, these results indicate that the SARS-CoV-2 Omicron variant will note completely evade vaccines based on the spike protein.

Though there are many mutations in the RBD of Omicron, the predicted structure from `AlphaFold2` suggests these mutations do not appear to be causing any large conformational change that would totally evade antibody interaction. However, we do see some amino acid substitutions to different, longer side chained residues at the binding site. This result may be due to the slightly more distant interaction with the antibody and therefore may reduce the binding affinity.

The results of `RoseTTAFold` suggest that the mutations have a different effect on the overall 3D structure of the Omicron RBD. The conformational change seen in this structure may contribute to antibody evasion or more severely reduce antibody binding affinity.

`AlphaFold2` has not been validated for predicting the effect of mutations and is not expected to produce a completely unfolded structure if the Omicron sequence contains any destabilising point mutations. This may explain the extremely similar backbone structure to the reference RBD structures.

Given that the antibody docking results between `AlphaFold2` and `RoseTTAFold` are statistically different, this further posits that this potential conformational change seen in the `RoseTTAFold` RBD may directly affect the binding of neutralizing antibodies.

Though the predicted RBD structures between `AlphaFold2` and `RoseTTAFold` are quite different, the `HADDOCK` antibody docking results seem to converge on specific positions with which the Fab will interface. S477N and Y505H may be useful positions on which to focus for future vaccine design or mutational surveillance in new variants.

Further analyses are needed using a broader range of different classes of antibodies, including therapeutic antibodies and neutralizing antibodies that bind to other locations on the spike protein. Recent preprint articles show more drastic reductions in the binding affinity of some other antibodies like CB6, a neutralizing antibody similar to ones used in this study, as well as a variety of therapeutic antibodies (22, 23). Thus, while we fail to see complete evasion of the antibodies used in this study, there are far more Omicron-antibody interactions to be understood.

Once a true structure of the Omicron RBD is determined, it will be of interest to compare the true structure to both of the predicted RBD structures from `AlphaFold2` and `RoseTTAFold`, mainly to see if a large conformational change occurs (as is seen in the `RoseTTAFold`) or if the overall backbone/3D structure is very similar to the reference structures used in this study (as is seen in the `AlphaFold2`). In addition, it will be necessary to validate the antibody interactions predicted using `HADDOCK` with true, experimentally-derived binding measurements.

Determining the actual structure of a protein is a time-consuming process. Further, quantifying protein-protein interactions (like spike-to-antibody interactions) are also experimentally difficult to perform *in vitro*. Given the public health urgency in understanding the impacts of new SARS-CoV-2 variants quickly requires that we act quicker than is possible in a lab. Thus, *in silico* predictive tools like `AlphaFold2`, `RoseTTAFold`, and `HADDOCK` are important for quickly understanding the biochemistry of variants and can help us to infer the epidemiological implications of the variant.

Supplementary Materials

All data, scripts, and results from this work are available at GitHub.com/colbyford/SARS-CoV-2_B.1.1.529_Spike-RBD_Predictions (24).

ACKNOWLEDGEMENTS

We acknowledge these entities at UNC Charlotte: The Department of Bioinformatics and Genomics, The Bioinformatics Research Center, the College of Computing and Informatics, The College of Liberal Arts and Sciences, Research and Economic Development, University Research Computing, and the Graduate School. We also thank the Belk family for support.

Also, we appreciate the initial feedback from members of the Google DeepMind team on the usage of `AlphaFold2` for point mutations. This prompted our further investigation and comparison with `RoseTTAFold` and the differing results we now show in this study.

Bibliography

1. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H.a Tian, Y.-Y.n Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M.n Wang, J.-J.o Zheng, L. Xu, E. C. Holmes, and Y.-Z. Zhang. A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798):265–269, 2020. doi: 10.1038/s41586-020-2008-3.
2. A. Rambaut, E. C. Holmes, Á. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*, 5:1403–1407, 2020. doi: 10.1038/s41564-020-0770-5.
3. World Health Organization. *Classification of Omicron (B.1.1.529): SARS-CoV-2 Variant of Concern*, 2021. Available at: https://www.who.int/news/item/26-11-2021-classification-of-omicron-%28b.1.1.529%29-sars-cov-2-variant-of-concern (accessed 1 December 2021).
4. T. Farooqi, J. A. Malik, A. H. Mulla, T. Al Hagbani, K. Almansour, M. A. Ubaid, S. Alghamdi, and S. Anwar. An overview of SARS-CoV-2 epidemiology, mutant variants, vaccines, and management strategies. *Journal of Infection and Public Health*, 14:1299–1312, 2021. doi: 10.1016/j.jiph.2021.08.014.
5. A. F. Brito, E. Semenova, G. Dudas, G. W. Hassler, C. C. Kalinich, M. U.G. Kraemer, S. C. Hill, Danish COVID-19 Genome Consortium, E. C. Sabino, O. G. Pybus, C. Dye, S. Bhatt, S. Flaxamn, M. A. Suchard, N.n D. Grubaugh, G. Baele, and N. R. Faria. Global disparities in SARS-CoV-2 genomic surveillance. *medRxiv*, 2021. doi: 10.1101/2021.08.21.21262393.
6. D. J. Machado, R. A. White III, J. Kofsky, and D. A. Janies. Fundamentals of genomic epidemiology, lessons learned from the coronavirus disease 2019 (covid-19) pandemic, and new directions. *Antimicrobial Stewardship & Healthcare Epidemiology (2021)*, pages 1–11, 2021. doi: 10.1017/ash.2021.222.
7. K. Katoh and Daron M. Standley. Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780, Apr 2013. ISSN 1537-1719. doi: 10.1093/molbev/mst010.
8. D. Jacob Machado, R. Scott, S. Guirales, and D. A. Janies. Fundamental evolution of all orthocoronavirinae including three deadly lineages descendent from chiroptera-hosted coronaviruses: Sars-cov, mers-cov and sars-cov-2. *Cladistics*, 37(5):461–488, 2021. doi: https://doi.org/10.1111/cla.12454.
9. F. Abascal, R. Zardoya, and M. J. Telford. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, 38(suppl_2):W7–W13, 04 2010. ISSN 0305-1048. doi: 10.1093/nar/gkq291.
10. G. Stecher, K. Tamura, and S. Kumar. Molecular Evolutionary Genetics Analysis (MEGA) for macOS. *Molecular Biology and Evolution*, 37(4):1237–1239, 01 2020. ISSN 0737-4038. doi: 10.1093/molbev/msz312.
11. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2.
12. Minkyung Baek, Frank DiMaio, Ivan Anishchenko, Justas Dauparas, Sergey Ovchinnikov, Gyu Rie Lee, Jue Wang, Qian Cong, Lisa N. Kinch, R. Dustin Schaeffer, Claudia Millán, Hahnbeom Park, Carson Adams, Caleb R. Glassman, Andy DeGiovanni, Jose H. Pereira, Andria V. Rodrigues, Alberdina A. van Dijk, Ana C. Ebrecht, Diederik J. Opperman, Theo Sagmeister, Christoph Buhlheller, Tea Pavkov-Keller, Manoj K. Rathinaswamy, Udit Dalwadi, Calvin K. Yip, John E. Burke, K. Christopher Garcia, Nick V. Grishin, Paul D. Adams, Randy J. Read, and David Baker. Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, 373(6557):871–876, 2021. doi: 10.1126/science.abj8754.
13. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006. doi: 10.1002/prot.21123.
14. Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. Colabfold - making protein folding accessible to all. *bioRxiv*, 2021. doi: 10.1101/2021.08.15.456425.
15. C. O. Barnes, A. P. West Jr., K. E. Huey-Tubman, M. A.G. Hoffmann, N. G. Sharaf, P. R. Hoffman, N. Koranda, H. B. Gristick, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, S. Finkin, T. Hägglöf, A. Hurley, K. G. Millard, Y. Weisblum, F. Schmidt, T. Hatziioannou, P. D. Bieniasz, M. Caskey, D. F. Robbiani, M. C. Nussenzweig, and P. J. Bjorkman. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. *Cell*, 182(4):828–842.e16, 2020. doi: 10.1016/j.cell.2020.06.025.
16. M. Yuan, H. Liu, N. C. Wu, C.-C. D. Lee, X. Zhu, F. Zhao, D. Huang, W. Yu, Y. Hua, H. Tien, T. F. Rogers, E. Landais, D. Sok, J. G. Jardine, D. R. Burton, and I. A. Wilson. Structural basis of a shared antibody response to SARS-CoV-2. *Science*, 369(6507):1119–1123, 2020. doi: 10.1126/science.abd2321.
17. N. K. Hurlburt, E. Seydoux, Y.-H. Wan, V. V. Edara, A. B. Stuart, J. Feng, M. S. Suthar, A. T. McGuire, L. Stamatatos, and M. Pancera. Structural basis for potent neutralization of SARS-CoV-2 and role of antibody affinity maturation. *Nature Communications*, 11(1):5413, 2020. doi: 10.1038/s41467-020-19231-9.
18. G.C.P. van Zundert, J.P.G.L.M. Rodrigues, M. Trellet, C. Schmitz, P.L. Kastritis, E. Karaca, A.S.J. Melquiond, M. van Dijk, S.J. de Vries, and A.M.J.J. Bonvin. The HADDOCK2.2 web server: User-friendly integrative modeling of biomolecular complexes. *Journal of Molecular Biology*, 428(4):720–725, 2016. doi: 10.1016/j.jmb.2015.09.014.
19. Schrödinger, L. L. C. *The PyMOL Molecular Graphics System, version 2.4.1*, 2015. Available at: http://www.pymol.org/pymol (accessed 29 November 2021).
20. R Core Team. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
21. D. Sharma, P. Rawat, V. Janakiraman, and M. M. Gromiha. Elucidating important structural features for the binding affinity of spike - sars-cov-2 neutralizing antibody complexes. *Proteins: Structure, Function, and Bioinformatics*, n/a(n/a), 2021. doi: https://doi.org/10.1002/prot.26277.
22. Lihong Liu, Sho Iketani, Yicheng Guo, Jasper F-W. Chan, Maple Wang, Liyuan Liu, Yang Luo, Hin Chu, Yiming Huang, Manoj S. Nair, Jian Yu, Kenn K-H. Chik, Terrence T-T. Yuen, Chaemin Yoon, Kelvin K-W. To, Honglin Chen, Michael T. Yin, Magdalena E. Sobieszczyk, Yaoxing Huang, Harris H. Wang, Zizhang Sheng, Kwok-Yung Yuen, and David D. Ho. Striking antibody evasion manifested by the omicron variant of sars-cov-2. *bioRxiv*, 2021. doi: 10.1101/2021.12.14.472719.
23. Laura A. VanBlargan, John M. Errico, Peter J. Halfmann, Seth J. Zost, James E. Crowe, Lisa A. Purcell, Yoshihiro Kawaoka, Davide Corti, Daved H. Fremont, and Michael S. Diamond. An infectious sars-cov-2 b.1.1.529 omicron virus escapes neutralization by several therapeutic monoclonal antibodies. *bioRxiv*, 2021. doi: 10.1101/2021.12.15.472828.
24. C. T. Ford. *Predictions of the SARS-CoV-2 B.1.1.529 Variant Spike Protein Receptor Binding Domain Structure and Neutralizing Antibody Interactions*, 2021. Available at: https://zenodo.org/record/5733161#.YalzFvFKhhE (accessed 1 December 2021).