Identifying protein sites contributing to vaccine escape via statistical comparisons of short-term molecular dynamics simulations

Madhusudan Rajendran,[1]* Maureen C. Ferran,[1] and Gregory A. Babbitt[1,]**

[1]Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA.

*Correspondence: madhusudan.rajendran@gmail.com

**Correspondence: gabsbi@rit.edu

Abstract

The identification of viral mutations that confer escape from antibodies is crucial for understanding the interplay between immunity and viral evolution. We describe a molecular dynamic (MD) based approach that scales well to a desktop computer with a high-end modern graphics processor and enables the user to identify protein sites that are prone to vaccine escape in a viral antigen. We first implemented our MD pipeline to employ site-wise calculation of Kullback-Leibler divergence in atom fluctuation over replicate sets of short-term MD production runs to compare influenza hemagglutinin's rapid motions in the presence and absence of three well-known neutralizing antibodies. Using this simple comparative method applied to motions of viral proteins, we successfully identified *in silico* all previously empirically confirmed sites of escape in hemagglutinin, predetermined via selection experiments and neutralization assays. After this validation of our computational approach, we identified potential hot spot residues in the receptor binding domain (RBD) of the SARS-CoV-2 virus in the presence of COVOX-222 and S2H97 antibodies. We identified sites in the antigen-antibody interface with strong dampening of fluctuation that may indicate potential antibody escape due to single mutations. Many of these sites were found to match known sites of mutations in SARS-CoV-2 variants of concern. The determination of single sites with large effect on antigen-antibody binding interfaces is crucial to discriminating neutral variants from potential escape variants. In summary, we provide a cheap, fast, and accurate *in silico* method to identify and quantify potential hot spots of functional evolution in antibody binding footprints.

Introduction

In current attempts to prevent the spread of the Coronavirus disease 2019 (COVID-19) pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), clinicians and scientists have focused their efforts on the development of vaccines that are hoped to induce broad, long-lasting, neutralizing antibodies. However, the selective pressures imposed by the presence of the neutralizing antibodies in the host population can also drive the evolution of viruses towards adaptations that allow them to escape neutralization. For example, in the case of influenza viruses, immunity is provided by antibodies that target the hemagglutinin (HA), responsible for viral attachment and viral fusion to the host cell. However, these antibodies drive selection for amino acid substitutions in the HA, causing the influenza virus to rapidly evolve every year (1). As a result, humans are repeatedly re-infected over their lifetime. Similarly, the antibody selection by the host immune system can also

drive the emergence of new SARS-CoV-2 variants. Therefore, when developing vaccines that elicit antibodies against a broad range of strains, research efforts should also be aimed at identifying potential mutations that can facilitate viral escape from the neutralization effects of specific antibodies (2, 3).

The traditional approach to identifying these mutations is by passaging the virus in the presence of antibodies in a directed selection experiment, followed by validation of the variants that arise with neutralization assays. For example, in influenza viruses, escape mutant selection using a panel of monoclonal antibodies (mAbs) was used to identify the five major antigenic regions, Sa, Sb, Ca1, Ca2, and Cb (4-6). However, a significant drawback of this approach is that the directed selection typically only favors one of the many potential mutations that can escape a given antibody. Additionally, there is often a strong bias towards single amino acid replacements common in the viral stock, preventing a complete picture of the potentially multiple ways a virus can escape an antibody. Another approach is to test antibody binding to a panel of viral variants. In a typical 500 residue viral protein, there are about 10^4 potential single amino acid mutants (7). Creating all individual mutants and then testing the mutants against the antibodies is an impossible task, thus causing researchers to limit themselves to exploring only a small portion of protein space (e.g., examining only mutations to alanine) (8, 9). Such studies often use less cumbersome assays but still cannot give a complete picture of the mutational spectra that can allow a virus to escape neutralization by a given antibody (10).

Functional evolutionary studies of viral vaccine escape are often supplemented with protein structural determination via x-ray crystallography or cryogenic electron microscopy. While structural biology can provide a static image of how an antibody physically contacts a viral protein, it cannot provide complete information regarding which amino acid sites are more prone to single replacement mutations (i.e. hot-spot sites or residues). Prediction of hot-spot residues is crucial as these sites on the antibody-antigen complex have a strong propensity to disrupt binding interactions within the antibody-antigen interface (11, 12). Recently, it has been demonstrated, via site-directed selection experiments and neutralization assays, that these hot-spot regions share a common biophysical feature. They all tend to harbor single amino acid sites that have significant large affects upon binding interactions in the antibody-antigen interface (13). Given this common feature of hot-spot residues, we hypothesize that some simple analyses of computer-based dynamics simulations of the antibody-antigen interface might help predict potential vaccine escape mutations before they happen, allowing for important functional context to real-time sequence-based surveillance of current and future pandemics.

Here we utilize a relatively simple method of comparative statistical analysis of molecular dynamic (MD) simulations developed by our lab (DROIDS - Detecting Relative Outlier Impacts in Dynamic Simulation or DROIDS 4.0) that employs a site-wise Kullback-Leibler (KL) divergence metric and a multiple test corrected two sample Kolmogorov-Smirnov (KS) test to successfully validate previously known sites of antibody escape in the influenza HA (13-17). We then utilized our site-wise comparative MD approach to identify potential sites prone to antibody escape in the spike protein of SARS-CoV-2. In summary, we present a method to accurately identify hot-spot residues that are prone to single point mutations with large functional effects upon the antibody-antigen binding interface and thus are likely pre-adapted to allow for vaccine escape. Identifying such residues *in silico* will be essential for pre-screening the antigenic consequences of viral genetic variations and designing better vaccines that induce long-term and broadly neutralizing antibodies against viral pathogens.

Methods

PDB structure, glycosylation, and model preparation

The protein structures used for the primary models for analyzing the molecular dynamics of antibody interactions with HA and the SARS-CoV-2 receptor binding domain (RBD) are listed in Figure 1A. Any crystallographic reflections were removed along with any other small molecules used in crystallization. Hydrogens were added, and crystallographic waters were removed using pdb4amber (AmberTools18)(18). Glycosylation was deleted using the swapaa function in UCSF Chimera 1.14 (19). Predicted glycosylation were rebuilt for the Amber force field using the glycoprotein builder on the glycam.org webserver and the GLYCAM06j-1 force field (20, 21). Any missing loop structure in the files was inferred via homology modeling using the 'refine loop' command to Modeler within UCSF Chimera (22, 23). The globular head domain of the influenza HA is stable in its monomeric form. However, the stalk domain needs to be in a trimer to be stable (24). Therefore, MD simulations of PDB 4GMS was performed using trimmed (residues 57-270) monomeric form of the head domain. For MD simulations of the 3ZTJ and 4HLZ, the trimeric form of the stalk domain was used. The antibodies were trimmed, and only the heavy and light chains of the fragment antigen-binding (Fab) were used.

Molecular dynamics simulation protocols

All comparative molecular dynamics analysis via our DROIDS pipeline was based upon 100 replicate sets of 1 nanosecond accelerated MD runs (i.e., 100x1ns MD runs in each comparative state, e.g., unbound vs. bound). MD simulations were conducted using the particle mesh Ewald method via accelerated MD (pmemd.cuda) in Amber18 and the ff14SB protein and GLYCAM06j-1 force fields and implemented on two RTX 2080 Ti graphics processor units controlled via Linux Mint 19 operating system (21, 25-30). These replicate sets were preceded by energy minimization, 300 picoseconds of heating to 300K, and ten nanoseconds of equilibration, followed by random equilibration intervals for each replicate ranging from 0-0.5 nanoseconds. All protein systems were prepared via tLeAP (Ambertools18) and explicitly solvated, and charge neutralized with Na+ and Cl- ions in a Tip3P octahedral water box set to 12 nm beyond the surface of each protein with periodic boundaries (18, 31). All simulations were regulated using the Anderson thermostat at 300K and one atmospheric pressure (32). Root mean square atom fluctuations and atom correlations were conducted in CPPTRAJ using the atomicfluct and atomicorr commands (33).

Comparative protein dynamics analyses with DROIDS 4.5 and statistical analyses

Comparative signatures of dampened atom fluctuation during antibody binding were presented as protein site-wise divergence in atom fluctuation in the antibody-bound versus unbound states for each viral target protein. Divergences were calculated using the signed symmetric Kullback-Leibler (KL) divergence calculation in DROIDS 4.5. Significance tests and p-values for these site-wise differences were calculated in DROIDS 4.5 using two-sample Kolmogorov-Smirnov tests with the Benjamini-Hochberg multiple test correction in DROIDS 4.5. The mathematical details of DROIDS 4.0 site-wise comparative protein dynamics analysis were published previously by our group and can be found here (14, 15). This code is available at our GitHub web landing: https://gbabbitt.github.io/DROIDS-4.0-comparative-protein-dynamics/, and the latest release (version 4.5) is available at our GitHub repository https://github.com/gbabbitt/DROIDS-4.0-comparative-protein-dynamics.

Results and Discussion

Using our comparative MD analysis pipeline (DROIDS 4.5), we successfully identified and validated previously known antibody escape sites in the head and stalk domains of the influenza HA. We then expanded the scope of our study to identify potential mutational sites prone to antibody escape in the spike protein of SARS-CoV-2.

Initially, we applied a Mann-Whitney U-test to compare KL divergence values, generated using MD simulations, to amino acid residues with significant average fraction of variants surviving and non-significant average fraction of variants surviving in the presence of S139/1, (Figure 1C) (13). In our comparative analyses of MD simulations, we found significantly higher KL divergence in the average fraction of variants surviving during directed selection ($p < 0.001$), suggesting that our KL divergence metric might prove as a useful quantitative measure in discriminating sites with mutations leading to vaccine escape. We then applied our approach to anti-HA antibodies with a wide range of breadth against different influenza strains. S139/1 is a one such monoclonal antibody known to neutralize both group 1(H1, H2, H5, H9, and H13) and group 2 (H3) influenza viruses (Figure 1B) (34). Crystallization studies have revealed that the antibody targets highly conserved residues in the RBD of the head domain (35) (Figure 1B). Comparative MD simulations of HA bound to S139/1 and unbound have revealed strong dampening of atom fluctuations occurring at sites K156, G158, and N193 of the HA, with -KL divergence values -4.65 ($D = 0.810$, $p < 0.001$), -5.81 ($D = 0.895$, $p < 0.001$), and -4.81 ($D = 0.807$, $p < 0.001$), respectively (Figure 1A and 2A). Directed selection using mAb S139/1 revealed that these three residues are sites of strong escape (Figure 2B) (13). Furthermore, we observed that sites with significant average fraction of variants surviving (K156, G158, and N193) have significantly higher KL divergence values than the other amino acid positions (Figure 1C). This suggests a strong relationship between our KL divergence values and the results obtained from the directed selection experiments (13). As expected, the three amino acid residues with the highest negative KL divergence in our comparison of atom fluctuations from MD simulations also fall directly in the empirically determined physical binding footprint of the antibody. And they are the same three sites where previous works have selected escape mutants in the H1, H2, and H3 HAs (34, 35). Residue 156 is a part of the 150 loop of the influenza HA. Lys156 forms electrostatic interactions by inserting itself into the acidic pocket in the Fab formed by residues Glu^{H35} and Glu^{H50}. This interaction's importance can be seen in the case of A/Japan/305/1957 (Jap57) H2 HA. Jap57 has glutamate, in place of lysine, at position 156. As a result of this electrostatic repulsion contributed by Glu156, S139/1 does not bind to Jap57 H2. Another 150 loop binding determinant is residue 158, which has the largest dampening of atomic fluctuations. Gly158 is closely stacked on the light chain complementarity determining region 3 (LCDR3) of S139/1 and is further stabilized by a main chain hydrogen bond between S159 and Asn^{L92}. Any mutations at these positions will cause clashes at the antibody interface. S139/1 also recognizes members of the 190 helix. In the residues of 190 helix, Asn193 plays a significant role in antibody recognition and is buried by heavy chain complementarity determining regions (HCDR) 1-3 of S139/1 (34). Furthermore, it should be noted that site 193 was known to interact with the host receptor molecule (sialic acid moiety), suggesting the contribution of this residue to receptor binding of HA (36, 37). Therefore, isolates with mutated residues at site 193 have reduced viral fitness due to their inability to bind to the host receptors (38). Selective pressure created by S139/1 on residue 193 has other secondary effects on the virus, such as the reduced binding activity of the influenza virus.

Lastly, with mAb S139/1, atomic fluctuation dampening is also seen in residues 133-134 of the influenza HA (Figure 2A). These residues are part of the 130 loop of the HA, and make a significant portion of the RBD of the influenza virus. Insertion at position 133a, between residues 133 and 134, creates a change in the 130 loop structure, causing a bulge in the insertion site, which then clashes into mAb S139/1. Furthermore, the side chain of 133a is orientated in space where the antibody would bind. Lastly, in addition to atomic fluctuation dampening at residues 133-134 and at residues 156/158, we also see dampening at 145-147 (Figure 2A). Upon further examination, there lie no epitopes of S139/1 between residues 145-147. Therefore, we believe that this dampening is an artifact caused purely by the residue's location between two epitope sites (130 loop and 150 loop) of S139/1. In summary, the three sites determined to have the largest physical effect upon antibody binding in our simulations (i.e. showing the largest negative KL divergence) are exactly the same three sites most likely to evolve antibody escape under directed selection experiments.

We also chose two broad antibodies that target the stalk domain of the HA (FI6v3 and C179) to further validate our computational approach to identify potential sites of escape. FI6v3 was first isolated by high-throughput screening of immortalized antibody-secreting cells and was found to bind to both group 1 (H1, H2, H5, H6, H8, H9, and H13) and group 2 (H3, H4, H7, and H10) viruses (Figure 1B) (39). Antibody C179 was first isolated from a mouse that had been immunized with the H2N2 virus and was later found to cross-neutralize H1, H2, H5, H6, and H9 subtypes (Figure 1B) (40). Both FI6v3 and C179 have epitopes that lie in the stalk domain and are known to interfere with membrane fusion (Figure 1B). Our MD simulations do not reveal dampening of atomic fluctuations in the stalk domain in the presence of FI6v3 or C179. When we look at the difference in atomic fluctuations of the stalk antibodies on the same scale as S139/1, we only see a few sites with strong dampening (Figures 2C and 2D).

Antibody selection revealed very similar results as our MD simulation, with only few sites in the stalk domain with slightly increased fraction of variants surviving mAbs C179 and (13). As a result, the authors concluded that the stalk domain is less capable of escaping antibodies by single mutations (13). In the presence of FI6v3, both directed selection and MD simulation picked a small bump at site 47 (KL = -2.41, D = 0.647, $p < 0.001$) that is of importance (Figures 1A and 2C) (13). Mutational studies have demonstrated that the introduction of E47R in the stalk domain has increased the resistance to FI6v3 (41). Like other studies, we also conclude that the HA stalk is intolerant of mutations, and confirmed by us here due to the lack of sites with dampened fluctuations. The absence of sites with large divergences in atomic fluctuations in our comparative MD simulations would seem to confirm that the anti-stalk antibodies studied here readily target sites with high mutational tolerance as suggested by other site directed experiments (42). Another possible explanation could be that the binding energetics at protein-protein interfaces can be asymmetrically distributed across all sites, thus preventing us from identifying the mutational tolerant sites (11, 12).

We implemented our computational approach to identify potential hot spot residues in the RBD of the spike protein of SARS-CoV-2 in the presence of two recently published antibodies against the virus: COVOX-222 and S2H97 (Figure 3A). COVOX-222 is known to bind to different residues than S2H97 and is known to neutralize strains P.1 (Gamma) from Brazil, B.1.351 (Beta) from South Africa, and B.1.1.7 (Alpha) from the United Kingdom (43). Termed "super-antibody," S2H97 is known to bind with high affinity across all sarbecovirus clades and prophylactically protects hamsters from viral challenge (44, 45). In the presence of COVOX-222, we see the most dampening of atomic fluctuations at residues G476 (KL = -5.50, D = 0.888, $p < 0.001$), F486 (KL = -4.52, D = 0.800, $p < 0.001$), and N501 (KL = -4.55, D = 0.708,

$p < 0.001$) of the RBD (Figure 1A and 3B). We see modest amount of dampening at residues D405 (KL = -2.57, D = 0.572, $p < 0.001$), K417 (KL = -3.27, D = 0.678, $p < 0.001$), and K458 (KL = -3.49, D = 0.703, $p < 0.001$) of the RBD (Figure 1A and 3B).

It should be noted that the residues picked up by our computational method are part of the epitopes of COVOX-222. Of all the residues identified by the comparison of MD simulation, the residues with the greatest dampening of atomic fluctuation have a higher chance of being classified as hot-spot residues. They are more prone to mutate, thus allowing the virus to escape COVOX-222. Residue 417, one of the residues with moderate dampening of atomic fluctuation, makes a weak salt-bridge interaction with the HCDR3 residue E99 of COVOX-222 (43). And residue N501, one with residues with the most dampening of atomic fluctuations, is known to interact with LCDR1 residue P30 of the antibody via a stacking interactions (43, 46). In the case of the mAb S2H97, we see a moderate dampening of atomic fluctuations at sites D428 (KL = -3.81, D = 0.692, $p < 0.001$) and K462 (KL = -3.37, D = 0.723, $p < 0.001$) . Sites N394 (KL = -4.25, D = 0.760, $p < 0.001$), E484 (KL = -5.20, D = 0.837, $p < 0.001$), L518 (KL = -6.56, D = 0.945, $p < 0.001$) have the most dampening of atomic fluctuations (Figure 1A and 3C). These five sites all fall in the epitope of S2H97 (44). Like mAb COVOX-222, we predict that the sites with the most dampening of atomic fluctuations (i.e. negative KL divergence) are more prone to functionally evolve under the selection pressure of the vaccine, thus allowing the virus to potentially escape the binding of S2H97.

Many of the residues identified as potential sites of escape, either in the presence of COVOX-222 or S2H97, overlap with mutations seen in the variants of concern (VOC). At present, there are mainly four kinds of VOC: Alpha, Beta, Gamma, and Delta In the RBD of the spike protein, alpha harbors an N501Y mutation, beta, and gamma harbor K417N, E484K and N501Y mutations, and delta harbors L452R and T478K mutations. Here we employed comparative MD simulation of the SARS-CoV-2 RBD bound to human angiotensin-converting enzyme 2(ACE2) (PDB 6VW1 trimmed to only include the viral RBD) modeled with and without the recently arising VOC mutations in the alpha to delta variant strains. Three of the VOC mutations (417, 484, and 501) were identified as potential escape sites using our approach (Figure 1A, 3B, and 3C). In addition to the three sites, we confirmed (i.e., 417, 484, and 501), we also saw significant differences in the binding atomic fluctuations at the positions 452 (KL = -0.206, D = 0.155, $p < 0.001$) and 478 (KL = -0.129, D = 0.190, $p < 0.001$) of RBD (Figures 1A, 3D and 3E). Investigation with pseudoviruses possessing RBD mutations harbored by VOC demonstrated that plasma neutralizing activity of vaccinated individuals showed a one-fold to three-fold significant decreases against E484K, N501Y, or the K417N + E484K + N501Y triple mutant (48). These results confirm what we have observed using our comparative MD analysis (Figures 1A and 3E). Furthermore, evidence from clinical trials on the impact of VOC on vaccine efficacy confirms what we have observed as well. For example, the ChAdOx1 nCoV-19 vaccine and the single-dose JNJ-78436735 (Johnson & Johnson/Janssen) vaccine have reduced vaccine efficacy by 10.4% and 57%, respectively, against the B.1.351 variant, which contains K417N, E484K, and N501Y mutations (49-51). Therefore, when evaluating vaccine efficacy and the effect of neutralizing antibodies against the SARS-CoV-2 virus, the focus should be given to amino acid positions that are prone to mutate and escape the antibody, as these positions might cause the emergence of new VOCs.

Directed selection experiments followed by confirmation through neutralization assays are the classical approach for identifying the location of the hot-spot residues. This involves passaging the virus in the presence of an increasing antibody concentration and then sequencing the virus, which takes over

several months. However, each selection experiment only identifies one of many mutations that escape an antibody and must be conducted in a high-throughput manner via deep mutational scanning (60). Compared to the classical method, our simple computational approach can identify the locations of sites prone to vaccine escape in a matter of days. Additionally it identifies positions of all amino acids that could mutate to escape the antibody in a single *in silico* experiment. And compared to other computational tools, our MD-based approach does not require large and potentially biased machine learning training data sets. Furthermore, the *de novo* prediction is based on a given experimental structure, which enables an unprecedented close synergy between our computational approach and existing laboratory methods for identifying potential routes of viral evolution leading to enhanced transmissibility and vaccine escape.

Additionally, some of the other tools used in epitope prediction involve sequence-based methods, machine-learning methods, and structure-based methods. The epitope surface accessibility to antibody binding is generally used by the sequence-based methods (52). The availability of antigen sequence is crucial for the sequence-based method; however, the predicted epitope residues are not grouped into the corresponding epitopes (53). The machine learning (ML) based epitope prediction methods include several steps: a collection of datasets with clean and comprehensive data, extraction of antigen features of the sequence (e.g., physicochemical properties, evolutionary information, amino acid composition), and training the model using ML algorithms (54). Some of the commonly used ML tools for epitope prediction are ABCPred (uses artificial neural network), COBEpro (uses support vector machine), EPSVR (uses support vector regression method), and BepiPred (based on random forest algorithm) (52, 55-57). And lastly, structure-based methods identify epitopes by antigen structure and epitope-related propensity scales, including specific physicochemical properties and geometric attributes (58, 59). As the evolution of antibody-antigen interactions seem largely driven by single-site mutations with large functional effects and because antibody-antigen binding has quite obvious biophysical impacts on molecular dynamics, as would the formation of any protein complex, we offer a simple and effective alternative computational method to studying vaccine escape *in silico* using readily available simulation software combined with a classic and easily interpretable approach to the statistical comparison of distributions of local atom fluctuations on virus proteins in the presence/absence of antibodies.

In summary, we have computationally identified hot-spot residues known to mutate in a viral antigen in the presence of a neutralizing antibody. We first validated our approach using the influenza spike protein and three well-characterized antibodies against the influenza HA (13). We then implemented our approach to identify sites in the RBD of SARS-CoV-2 that are known to mutate in the presence of neutralizing mAbs S2H97 and COVOX-222 (43, 44). We further identified that residues known to mutate in the presence of the antibody overlap with the mutations seen in the VOC. While the method described here is not a complete substitute for laboratory-based methods we believe it can greatly complement these methods by allowing time and cost-saving through the computational prescreening of the underlying biophysics that may drive outcomes of many potential lab experiments. Determining which viral mutations escape from antibodies will be crucial for designing future therapeutics and vaccines and assessing future antigenic implications of ongoing viral evolution.

Author contributions

M.R wrote the manuscript, designed, performed the experiments, and analyzed the data. G.A.B revised and edited the manuscript, designed the experiments, and analyzed the data. M.C.F edited the manuscript and analyzed the data.

Acknowledgements

We acknowledge Dr. Ernest Fokoue (School of Mathematical Sciences, Rochester Institute of Technology) for his consultation on the prior development of our method. We also acknowledge Dr. Feng Cui and Dr. André O. Hudson (Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology) for helpful comments regarding this work.

Declaration of interests

The authors declare no competing interests.

References

1. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. 2004. Mapping the antigenic and genetic evolution of influenza virus. Science 305:371-6.
2. Luksza M, Lässig M. 2014. A predictive fitness model for influenza. Nature 507:57-61.
3. Krammer F, Palese P. 2015. Advances in the development of influenza virus vaccines. Nat Rev Drug Discov 14:167-82.
4. Gerhard W, Yewdell J, Frankel ME, Webster R. 1981. Antigenic structure of influenza virus haemagglutinin defined by hybridoma antibodies. Nature 290:713-7.
5. Caton AJ, Brownlee GG, Yewdell JW, Gerhard W. 1982. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). Cell 31:417-27.
6. Webster RG, Laver WG. 1980. Determination of the number of nonoverlapping antigenic areas on Hong Kong (H3N2) influenza virus hemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance. Virology 104:139-48.
7. Haddox HK, Dingens AS, Bloom JD. 2016. Experimental Estimation of the Effects of All Amino-Acid Mutations to HIV's Envelope Protein on Viral Replication in Cell Culture. PLoS Pathog 12:e1006114.
8. Pierce BG, Keck ZY, Lau P, Fauvelle C, Gowthaman R, Baumert TF, Fuerst TR, Mariuzza RA, Foung SKH. 2016. Global mapping of antibody recognition of the hepatitis C virus E2 glycoprotein: Implications for vaccine design. Proc Natl Acad Sci U S A 113:E6946-E6954.
9. Walker LM, Phogat SK, Chan-Hui PY, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR, Investigators PGP. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 326:285-9.
10. Polonis VR, Brown BK, Rosa Borges A, Zolla-Pazner S, Dimitrov DS, Zhang MY, Barnett SW, Ruprecht RM, Scarlatti G, Fenyö EM, Montefiori DC, McCutchan FE, Michael NL. 2008. Recent advances in the characterization of HIV-1 neutralization assays for standardized evaluation of the antibody response to infection and vaccination. Virology 375:315-20.
11. Jin L, Fendly BM, Wells JA. 1992. High resolution functional analysis of antibody-antigen interactions. J Mol Biol 226:851-65.

12. Dall'Acqua W, Goldman ER, Lin W, Teng C, Tsuchiya D, Li H, Ysern X, Braden BC, Li Y, Smith-Gill SJ, Mariuzza RA. 1998. A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. Biochemistry 37:7981-91.
13. Doud MB, Lee JM, Bloom JD. 2018. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. Nat Commun 9:1386.
14. Babbitt GA, Mortensen JS, Coppola EE, Adams LE, Liao JK. 2018. DROIDS 1.20: A GUI-Based Pipeline for GPU-Accelerated Comparative Protein Dynamics. Biophys J 114:1009-1017.
15. Babbitt GA, Fokoue EP, Evans JR, Diller KI, Adams LE. 2020. DROIDS 3.0-Detecting Genetic and Drug Class Variant Impact on Conserved Protein Binding Dynamics. Biophys J 118:541-551.
16. Babbitt GA, Lynch ML, McCoy M, Fokoue EP, Hudson AO. 2020. Function and evolution of B-Raf loop dynamics relevant to cancer recurrence under drug inhibition. J Biomol Struct Dyn:1-16.
17. Rynkiewicz P, Lynch ML, Cui F, Hudson AO, Babbitt GA. 2021. Functional binding dynamics relevant to the evolution of zoonotic spillovers in endemic and emergent. J Biomol Struct Dyn:1-19.
18. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. 2005. The Amber biomolecular simulation programs. J Comput Chem 26:1668-88.
19. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605-12.
20. Woods-Group. 2005-2021. GLYCAM Web, Complex Carbohydrate Research Center, University of Georgia, Athens, GA. http://legacy.glycam.org.
21. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ. 2008. GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 29:622-55.
22. Fiser A, Do RK, Sali A. 2000. Modeling of loops in protein structures. Protein Sci 9:1753-73.
23. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779-815.
24. Aartse A, Eggink D, Claireaux M, van Leeuwen S, Mooij P, Bogers WM, Sanders RW, Koopman G, van Gils MJ. 2021. Influenza A Virus Hemagglutinin Trimer, Head and Stem Proteins Identify and Quantify Different Hemagglutinin-Specific B Cell Subsets in Humans. Vaccines (Basel) 9.
25. Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. 98(10089-10092.
26. Ewald PP. 1921. Die Berechnung optischer und elektrostatischer Gitterpotentiale. Annalen der Physik 369:253-287.
27. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. 2012. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. J Chem Theory Comput 8:1542-1555.
28. Pierce LC, Salomon-Ferrer R, Augusto F de Oliveira C, McCammon JA, Walker RC. 2012. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. J Chem Theory Comput 8:2997-3002.
29. Salomon-Ferrer R, Götz AW, Poole D, Le Grand S, Walker RC. 2013. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. J Chem Theory Comput 9:3878-88.
30. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J Chem Theory Comput 11:3696-713.
31. Wang J, Wang W, Kollman PA, Case DA. 2006. Automatic atom type and bond type perception in molecular mechanical calculations. J Mol Graph Model 25:247-60.

32. Andersen HC. 1980. Molecular dynamics simulations at constant pressure and/or temperature. The Journal of Chemical Physics. 72(4):2384-2393.
33. Roe DR, Cheatham TE. 2013. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J Chem Theory Comput 9:3084-95.
34. Yoshida R, Igarashi M, Ozaki H, Kishida N, Tomabechi D, Kida H, Ito K, Takada A. 2009. Cross-protective potential of a novel monoclonal antibody directed against antigenic site B of the hemagglutinin of influenza A viruses. PLoS Pathog 5:e1000350.
35. Lee PS, Yoshida R, Ekiert DC, Sakai N, Suzuki Y, Takada A, Wilson IA. 2012. Heterosubtypic antibody recognition of the influenza virus hemagglutinin receptor binding site enhanced by avidity. Proc Natl Acad Sci U S A 109:17040-5.
36. Eisen MB, Sabesan S, Skehel JJ, Wiley DC. 1997. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. Virology 232:19-31.
37. Fleury D, Barrère B, Bizebard T, Daniels RS, Skehel JJ, Knossow M. 1999. A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. Nat Struct Biol 6:530-4.
38. Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y. 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol 74:8502-12.
39. Corti D, Voss J, Gamblin SJ, Codoni G, Macagno A, Jarrossay D, Vachieri SG, Pinna D, Minola A, Vanzetta F, Silacci C, Fernandez-Rodriguez BM, Agatic G, Bianchi S, Giacchetto-Sasselli I, Calder L, Sallusto F, Collins P, Haire LF, Temperton N, Langedijk JP, Skehel JJ, Lanzavecchia A. 2011. A neutralizing antibody selected from plasma cells that binds to group 1 and group 2 influenza A hemagglutinins. Science 333:850-6.
40. Smirnov YA, Lipatov AS, Gitelman AK, Okuno Y, Van Beek R, Osterhaus AD, Claas EC. 1999. An epitope shared by the hemagglutinins of H1, H2, H5, and H6 subtypes of influenza A virus. Acta Virol 43:237-44.
41. Roubidoux EK, Carreño JM, McMahon M, Jiang K, van Bakel H, Wilson P, Krammer F. 2021. Mutations in the Hemagglutinin Stalk Domain Do Not Permit Escape from a Protective, Stalk-Based Vaccine-Induced Immune Response in the Mouse Model. mBio 12.
42. Heaton NS, Sachs D, Chen CJ, Hai R, Palese P. 2013. Genome-wide mutagenesis of influenza virus reveals unique plasticity of the hemagglutinin and NS1 proteins. Proc Natl Acad Sci U S A 110:20248-53.
43. Dejnirattisai W, Zhou D, Supasa P, Liu C, Mentzer AJ, Ginn HM, Zhao Y, Duyvesteyn HME, Tuekprakhon A, Nutalai R, Wang B, López-Camacho C, Slon-Campos J, Walter TS, Skelly D, Costa Clemens SA, Naveca FG, Nascimento V, Nascimento F, Fernandes da Costa C, Resende PC, Pauvolid-Correa A, Siqueira MM, Dold C, Levin R, Dong T, Pollard AJ, Knight JC, Crook D, Lambe T, Clutterbuck E, Bibi S, Flaxman A, Bittaye M, Belij-Rammerstorfer S, Gilbert SC, Carroll MW, Klenerman P, Barnes E, Dunachie SJ, Paterson NG, Williams MA, Hall DR, Hulswit RJG, Bowden TA, Fry EE, Mongkolsapaya J, Ren J, Stuart DI, Screaton GR. 2021. Antibody evasion by the P.1 strain of SARS-CoV-2. Cell 184:2939-2954.e9.
44. Starr TN, Czudnochowski N, Liu Z, Zatta F, Park YJ, Addetia A, Pinto D, Beltramello M, Hernandez P, Greaney AJ, Marzi R, Glass WG, Zhang I, Dingens AS, Bowen JE, Tortorici MA, Walls AC, Wojcechowskyj JA, De Marco A, Rosen LE, Zhou J, Montiel-Ruiz M, Kaiser H, Dillen JR, Tucker H, Bassi J, Silacci-Fregni C, Housley MP, di Iulio J, Lombardo G, Agostini M, Sprugasci N, Culap K, Jaconi S, Meury M, Dellota E, Abdelnabi R, Foo SC, Cameroni E, Stumpf S, Croll TI, Nix JC, Havenar-Daughton C, Piccoli L, Benigni F, Neyts J, Telenti A, Lempp FA, Pizzuto MS, Chodera JD,

et al. 2021. SARS-CoV-2 RBD antibodies that maximize breadth and resistance to escape. Nature 597:97-102.
45. Kwon D. 2021. This 'super antibody' for COVID fights off multiple coronaviruses, *on* Nature. **https://www.nature.com/articles/d41586-021-01917-9#correction-0**. Accessed November 19.
46. Shrestha LB, Tedla N, Bull RA. 2021. Broadly-Neutralizing Antibodies Against Emerging SARS-CoV-2 Variants. Front Immunol 12:752003.
47. CDC. 2021. SARS-CoV-2 Variant Classifications and Definitions, *on* Centers for Disease Control and Prevention. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html. Accessed November 20.
48. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Da Silva J, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatziioannou T, Bieniasz PD, Nussenzweig MC. 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature 592:616-622.
49. Mahase E. 2021. Covid-19: Where are we on vaccines and variants? BMJ 372:n597.
50. Emary KRW, Golubchik T, Aley PK, Ariani CV, Angus B, Bibi S, Blane B, Bonsall D, Cicconi P, Charlton S, Clutterbuck EA, Collins AM, Cox T, Darton TC, Dold C, Douglas AD, Duncan CJA, Ewer KJ, Flaxman AL, Faust SN, Ferreira DM, Feng S, Finn A, Folegatti PM, Fuskova M, Galiza E, Goodman AL, Green CM, Green CA, Greenland M, Hallis B, Heath PT, Hay J, Hill HC, Jenkin D, Kerridge S, Lazarus R, Libri V, Lillie PJ, Ludden C, Marchevsky NG, Minassian AM, McGregor AC, Mujadidi YF, Phillips DJ, Plested E, Pollock KM, Robinson H, Smith A, Song R, et al. 2021. Efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine against SARS-CoV-2 variant of concern 202012/01 (B.1.1.7): an exploratory analysis of a randomised controlled trial. Lancet 397:1351-1362.
51. Madhi SA, Baillie V, Cutland CL, Voysey M, Koen AL, Fairlie L, Padayachee SD, Dheda K, Barnabas SL, Bhorat QE, Briner C, Kwatra G, Ahmed K, Aley P, Bhikha S, Bhiman JN, Bhorat AE, du Plessis J, Esmail A, Groenewald M, Horne E, Hwa SH, Jose A, Lambe T, Laubscher M, Malahleha M, Masenya M, Masilela M, McKenzie S, Molapo K, Moultrie A, Oelofse S, Patel F, Pillay S, Rhead S, Rodel H, Rossouw L, Taoushanis C, Tegally H, Thombrayil A, van Eck S, Wibmer CK, Durham NM, Kelly EJ, Villafana TL, Gilbert S, Pollard AJ, de Oliveira T, Moore PL, Sigal A, et al. 2021. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. N Engl J Med 384:1885-1898.
52. Saha S, Raghava GP. 2006. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65:40-8.
53. Yao B, Zheng D, Liang S, Zhang C. 2013. Conformational B-cell epitope prediction on antigen protein structures: a review of current algorithms and comparison with common binding site prediction methods. PLoS One 8:e62249.
54. Sun P, Ju H, Liu Z, Ning Q, Zhang J, Zhao X, Huang Y, Ma Z, Li Y. 2013. Bioinformatics resources and tools for conformational B-cell epitope prediction. Comput Math Methods Med 2013:943636.
55. Sweredoski MJ, Baldi P. 2009. COBEpro: a novel system for predicting continuous B-cell epitopes. Protein Eng Des Sel 22:113-20.
56. Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. 2010. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. BMC Bioinformatics 11:381.
57. Jespersen MC, Peters B, Nielsen M, Marcatili P. 2017. BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes. Nucleic Acids Res 45:W24-W29.

58. Saha S, Raghava GPS. 2004. BcePred: Prediction of Continuous B-Cell Epitopes in Antigenic Sequences Using Physico-chemical Properties, p 197-204. *In* Nicosia G, Cutello V, Bentley P, Timmis J (ed), Artificial Immune Systems. Springer, Italy.
59. Negi SS, Braun W. 2009. Automated detection of conformational epitopes using phage display Peptide sequences. Bioinform Biol Insights 3:71-81.
60. Leon PE, Wohlbold TJ, He W, Bailey MJ, Henry CJ, Wilson PC, Krammer F, Tan GS. 2017. Generation of Escape Variants of Neutralizing Influenza Virus Monoclonal Antibodies. J Vis Exp.
61. Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ. 2004. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science 303:1838-42.

Figure Legends

Figure 1: Description of PDB files and antibodies used in this study, and the comparison between KL divergence and average fraction of variants surviving. (A) Table summarizing the protein structure used for primary models for analyzing the molecular dynamics of antibody interactions with the influenza HA and SARS-CoV-2 RBD, and ACE2 interactions with wildtype and mutated SARS-CoV-2 RBD (includes VOC mutations). The table includes amino acid positions and the corresponding -KL divergence value denoting atomic fluctuations dampening for antibody-antigen/protein 1-protein 2 MD simulations. D value and the level of significance for the corresponding amino acid position are also given. (B) S139/1 (green), FI6v3 (blue), and C179 (red) epitopes are mapped onto the HA trimer, shown in grey (PDB 1RVX). Overlapping epitopes between FI6v3 and C179 are shown in cyan. Next to the HA trimer is a phylogenetic tree of HA subtypes. Circles denote reported antibody binding or neutralization against that subtype (C) Box and whisker plots showing the KL divergence values between amino acid sites that had significant average fraction of variants surviving and non-significant average fraction of variants in the presence of monoclonal antibody S139/1 (13). Only sites K156, G158, and N193 of the HA head domain had significant average fraction surviving viral particles.

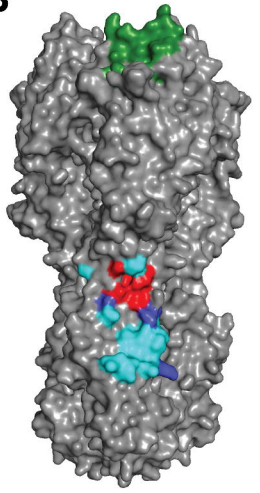
Figure 2: Epitope prediction and validation of hot-spot residues of the influenza hemagglutinin (HA) in the presence of neutralizing antibodies. Sequence positional plotting of dampening of atom motion on the influenza HA head domain by (A) S139/1 and on the HA stalk domain by (C) FI6v3 and (D) C179. The sequence profile of the -KL divergence between S139/1 and the head domain produces strong negative peaks in (A) at K156, G158, and N193. A modest negative peak is observed in the stalk domain in (C) at E47 in the presence of FI6v3. HA1 numbering is used to identify the amino acid positions in (A). HA2 numbering is used to identify the amino acid position in (C) and (D). The grey bar on top of the -KL divergence plots denotes the HA amino acid backbone with the location of (A) S139/1 epitopes shown in green, (C) FI6v3 epitopes in blue, and (D) C179 epitopes in red. (B) Logo plots of S139/1 show the amino acid position that has the largest effect. Letter heights are proportional to the excess fraction of virions with that mutation that survive the antibody, as indicated by scale bars. The logo plot was prepared using deep mutational scanning data from Doud, M.B. *et al.* (13).

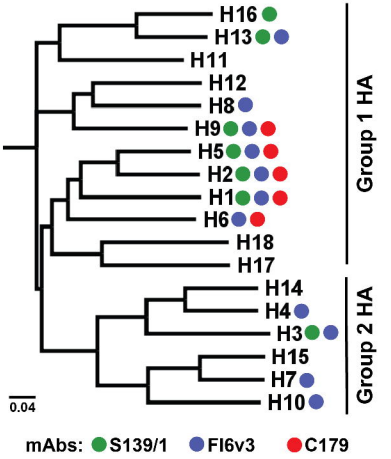
Figure 3: Identification of hot-spot residues that escape neutralizing antibodies and the importance of the variants of concern mutations. (A) Crystal structure of COVOX-222 (green, PDB 7OR9) and S2H97(cyan, PDB 7M7W) superimposed onto the structure of RBD of SARS-CoV-2 (grey, PDB 7M7W) (43, 44). COVOX-222 epitopes with the greatest KL divergence dampening are highlighted in pink, and S2H97 epitopes with the greatest KL divergence dampening are highlighted in olive. Sequence positional plotting of dampening of atom motion on the RBD of SARS-CoV-2 by (B) COVOX-222 and (C) S2H97. Amino acid positions with moderate to modest dampening of atomic fluctuation are identified in (B) and (C). The grey bar on the top of the KL divergence plots denotes the RBD domain amino acid backbone with the location of COVOX-222 epitopes shown in pink and the S2H97 epitopes shown in olive. (D) The site-wise KL divergence profiles showing the dampening of atom motion between the wildtype and mutated RBD of the spike protein of SARS-CoV-2 in the presence of human angiotensin-converting enzyme 2 (ACE2). The wildtype RBD was computationally mutated to include mutations from the variants of concern (VOC). (E) Multiple test-corrected two sample KS tests of significance for the impact of the mutations are also shown. Amino acid positions that correspond to the VOC mutations are highlighted and show higher level significance than other positions. Mutations that correspond to the four VOC in the RBD of spike protein are noted on top of (E).

A

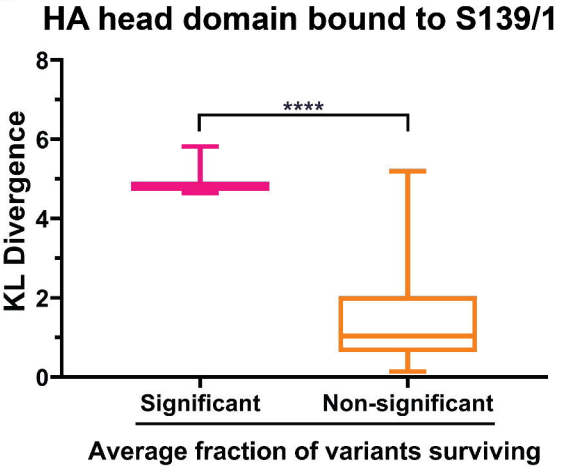
PDB ID	Antigen	Antibody	AA and position	-KL divergence	D value	p value
4GMS	A/Victoria/3/1975 H3 Head domain	S139/1	K156	-4.65	0.810	p < 0.001
			G158	-5.81	0.895	p < 0.001
			N193	-4.81	0.807	p < 0.001
3ZTJ	A/California/4/2009 H1 Stalk domain	FI6v3	E47	-2.41	0.647	p < 0.001
4HLZ	A/Japan/305/1957 H2 Stalk domain	C179	-	-		
7OR9	SARS-CoV-2 RBD	COVOX-222	D405	-2.57	0.572	p < 0.001
			K417	-3.27	0.678	p < 0.001
			K458	-3.49	0.703	p < 0.001
			G476	-5.50	0.888	p < 0.001
			F486	-4.52	0.800	p < 0.001
			N501	-4.55	0.708	p < 0.001
7M7W	SARS-CoV-2 RBD	S2H97	N394	-4.25	0.760	p < 0.001
			D428	-3.81	0.692	p < 0.001
			K462	-3.37	0.723	p < 0.001
			E484	-5.20	0.837	p < 0.001
			L518	-6.56	0.945	p < 0.001
PDB ID	**Protein 1**	**Protein 2**	**AA and position**	**-KL divergence**	**D value**	**p value**
6VW1	SARS-CoV-2 RBD mutated to include VOC mutations	ACE2	K417N	-0.053	0.140	p < 0.001
			L452R	-0.206	0.155	p < 0.001
			T478K	-0.129	0.190	p < 0.001
			E484K	-0.204	0.165	p < 0.001
			N501Y	-0.208	0.130	p < 0.05

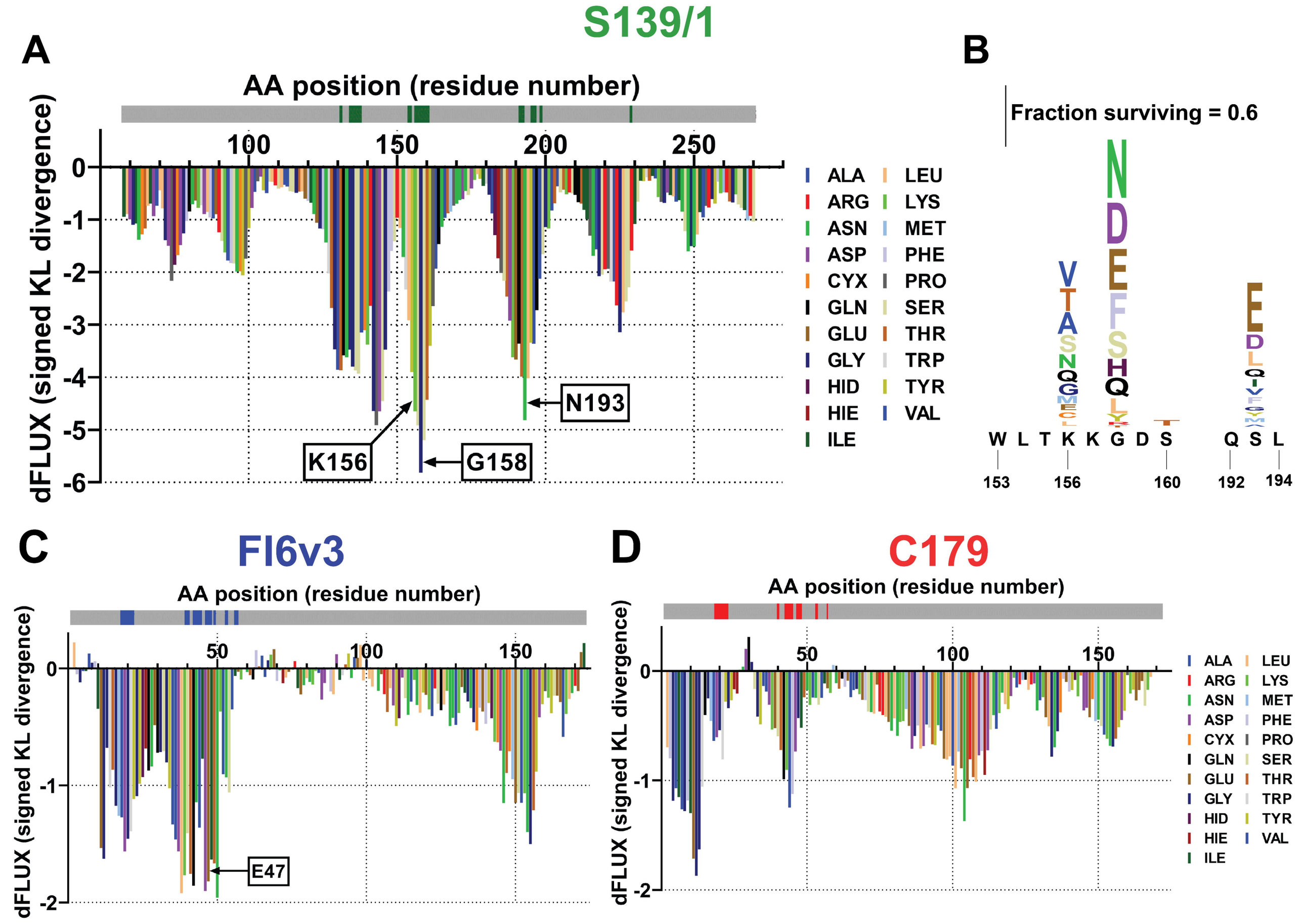
B

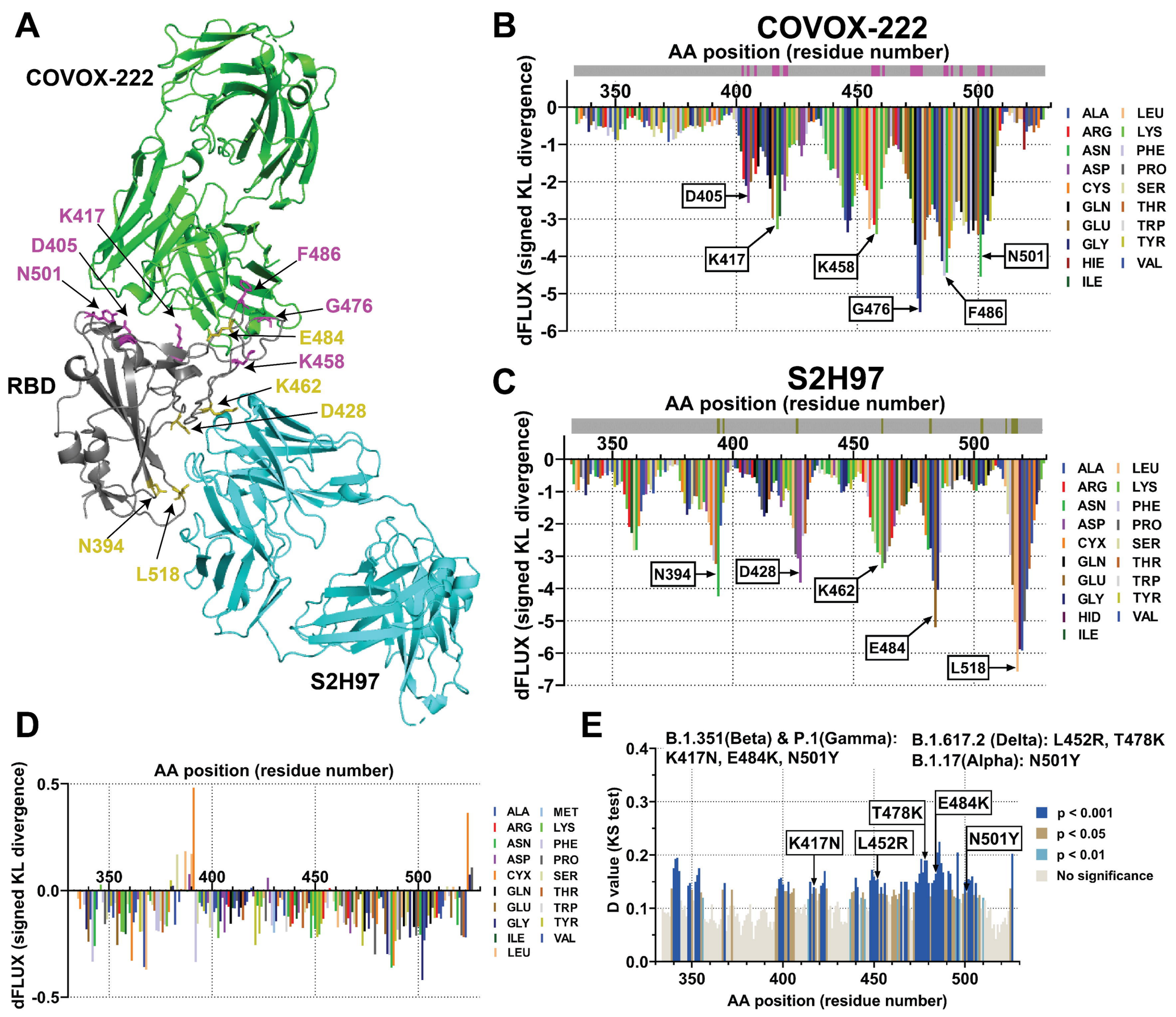




C




S139/1
A
AA position (residue number)
dFLUX (signed KL divergence)
100
150
200
250
0
-1
-2
-3
-4
-5
-6
K156
G158
N193
ALA
ARG
ASN
ASP
CYX
GLN
GLU
GLY
HID
HIE
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL
B
Fraction surviving = 0.6
W L T K K G D S Q S L
153
156
160
192
194
C
FI6v3
AA position (residue number)
dFLUX (signed KL divergence)
50
100
150
0
-1
-2
E47
D
C179
AA position (residue number)
dFLUX (signed KL divergence)
50
100
150
0
-1
-2
ALA
ARG
ASN
ASP
CYX
GLN
GLU
GLY
HID
HIE
ILE
LEU
LYS
MET
PHE
PRO
SER
THR
TRP
TYR
VAL


A
COVOX-222
K417
D405
N501
F486
G476
E484
K458
K462
D428
RBD
N394
L518
S2H97
B
COVOX-222
AA position (residue number)
dFLUX (signed KL divergence)
D405
K417
K458
G476
F486
N501
ALA
ARG
ASN
ASP
CYS
GLN
GLU
GLY
HIE
ILE
LEU
LYS
PHE
PRO
SER
THR
TRP
TYR
VAL
C
S2H97
AA position (residue number)
dFLUX (signed KL divergence)
N394
D428
K462
E484
L518
ALA
ARG
ASN
ASP
CYX
GLN
GLU
GLY
HID
ILE
LEU
LYS
PHE
PRO
SER
THR
TRP
TYR
VAL
D
AA position (residue number)
dFLUX (signed KL divergence)
ALA
ARG
ASN
ASP
CYX
GLN
GLU
GLY
ILE
LEU
MET
LYS
PHE
PRO
SER
THR
TRP
TYR
VAL
E
B.1.351(Beta) & P.1(Gamma):
K417N, E484K, N501Y
B.1.617.2 (Delta): L452R, T478K
B.1.17(Alpha): N501Y
D value (KS test)
K417N
L452R
T478K
E484K
N501Y
p < 0.001
p < 0.05
p < 0.01
No significance
AA position (residue number)