

# SARS-CoV-2 Spike Protein Mutations and Escape from Antibodies: A Computational Model of Epitope Loss in Variants of Concern

Alice Triveri, Stefano A. Serapian, Filippo Marchetti, Filippo Doria, Silvia Pavoni, Fabrizio Cinquini, Elisabetta Moroni, Andrea Rasola, Francesco Frigerio, and Giorgio Colombo\*



Cite This: *J. Chem. Inf. Model.* 2021, 61, 4687–4700



Read Online

ACCESS |



Metrics & More

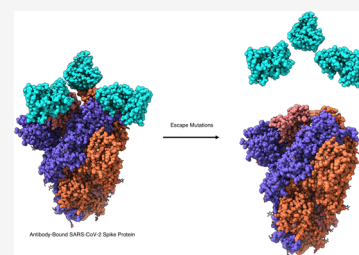


Article Recommendations



Supporting Information

**ABSTRACT:** The SARS-CoV-2 spike (S) protein is exposed on the viral surface and is the first point of contact between the virus and the host. For these reasons it represents the prime target for Covid-19 vaccines. In recent months, variants of this protein have started to emerge. Their ability to reduce or evade recognition by S-targeting antibodies poses a threat to immunological treatments and raises concerns for their consequences on vaccine efficacy. To develop a model able to predict the potential impact of S-protein mutations on antibody binding sites, we performed unbiased multi-microsecond molecular dynamics of several glycosylated S-protein variants and applied a straightforward structure-dynamics-energy based strategy to predict potential changes in immunogenic regions on each variant. We recover known epitopes on the reference D614G sequence. By comparing our results, obtained on isolated S-proteins in solution, to recently published data on antibody binding and reactivity in new S variants, we directly show that modifications in the S-protein consistently translate into the loss of potentially immunoreactive regions. Our findings can thus be qualitatively reconnected to the experimentally characterized decreased ability of some of the Abs elicited against the dominant S-sequence to recognize variants. While based on the study of SARS-CoV-2 spike variants, our computational epitope-prediction strategy is portable and could be applied to study immunoreactivity in mutants of proteins of interest whose structures have been characterized, helping the development/selection of vaccines and antibodies able to control emerging variants.



## INTRODUCTION

Protein sequences evolve as a result of selective pressure to optimize function, create improved phenotypes, and introduce new advantageous traits. In pathogens like bacteria and viruses, sequences evolve via modifications such as point mutations, recombination and deletions/insertions to induce higher infectivity, more efficient replication, and ultimately escape from the host immune systems.<sup>1–7</sup>

The SARS-CoV-2 virus, the etiological agent of Covid-19, is no exception to these general rules. The spread of the virus to more than 200 million people worldwide, combined with the pressure determined by the reactions of immunocompetent populations, led to the emergence of “variants of concern”. In this context, attention has been focused on the SARS-CoV-2 spike protein (S protein), the large, heavily glycosylated class I trimeric fusion protein which mediates host cell recognition, binding and entry. Because it represents the first point of contact with the host, and given its crucial role in viral pathogenesis,<sup>5,6,8–10</sup> the S protein has been the basis for the design of currently used vaccines effective at reducing viral spread, hospitalization and mortality rates.<sup>11–16</sup>

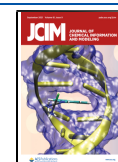
While for almost one year the only notable mutation in S has been the D614G (Asp<sup>614</sup> → Gly), which increases affinity for the cell receptor ACE2 and has immediately become dominant, novel S protein variants reported of late may pose new potential challenges for efficacy of vaccination, antibody-based therapies

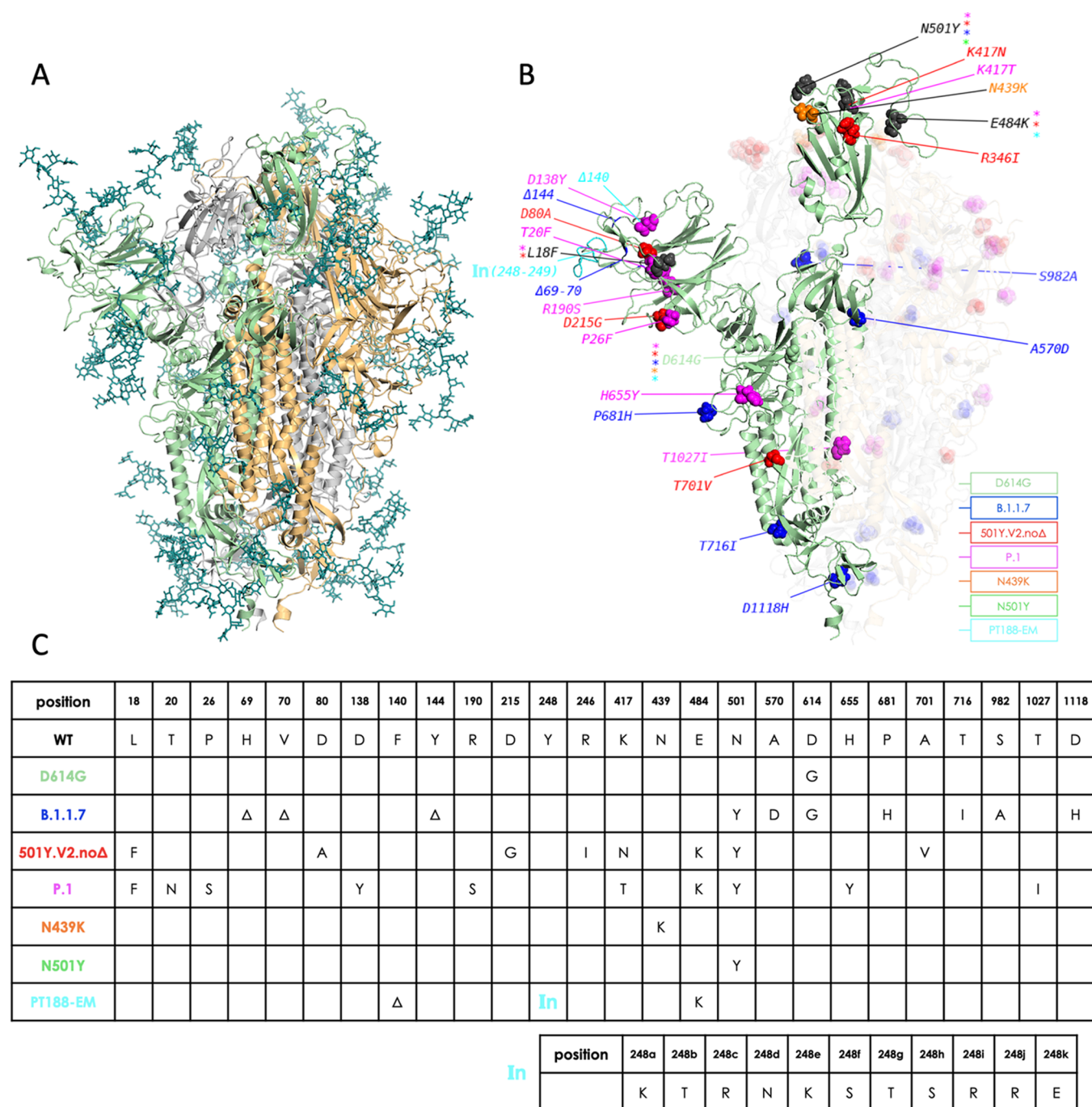
and viral diffusion control. Three notable examples of such evolved S proteins, which correspond to major circulating variants, are B.1.1.7 (the so-called UK or  $\alpha$  variant), 501Y.V2/B.1.351 (the South African or  $\beta$  variant), and B.1.1.28 (P.1, the Brazilian or  $\gamma$  variant). All such sequences contain various mutations due to nonsynonymous nucleotide changes in the receptor-binding domain (RBD), including E484K, N501Y, and/or K417N.<sup>10</sup> In B.1.1.7 and B.1.351, deletions are also present in the N-terminal domains (NTD) (Figure 1).

Several studies showed how some of these circulating variants may have reduced sensitivity to neutralizing antibodies targeting the RBD or to the NTD.<sup>10,17–19</sup> In this context, polyclonal antibodies contained in convalescent plasma (CP) from individuals infected with the D614G-containing SARS-CoV-2, showed reduced potency in neutralizing 501Y.V2/B.1.351 virus isolates.<sup>20,21</sup> Furthermore, antibodies elicited after vaccine treatment showed reduced neutralization of pseudoviruses bearing the mutations of the P.1 and 501Y.V2/B.1.351 variants.<sup>22</sup> The same was observed for pseudoviruses with

Received: July 19, 2021

Published: September 1, 2021





**Figure 1.** Overview of simulated variants (definitions in main text). (A) The full-length, fully glycosylated trimeric structure corresponding to pdb code 6VSB. Protomer A (RBD “up”): secondary structure in green; protomers B and C (RBD “down”): grey and sand, respectively. Glycans’ C, N, and O atoms rendered as teal sticks. (B) Positions and nature of mutations highlighted on protomer A of different variants. Mutant residues’ heavy atoms are rendered as spheres; a different color is assigned to each variant, as indicated in the legend. Mutations common to more than one variant are rendered and/or labeled in black, with colored asterisks denoting variants carrying the mutation. The insertion in the PT188-EM variant (cyan) is denoted by “In(248–249)”. Protomers B and C are also shown with their respective mutations, but rendered with increased transparency for clarity; glycans are omitted; (C) synopsis of mutations on the different variants simulated in this work, including the 11-residue insertion in the PT188-EM variant.

variations in S mimicking those of the B.1.1.7 lineage.<sup>22,23</sup> Yet, fortunately, it was shown that vaccine-generated antibody titers were sufficient to neutralize B.1.1.7 in sera from 40 BNT162b2-vaccinated individuals.<sup>24</sup> In this context, it is encouraging to note that new studies are reporting high levels of efficacy against severe forms of Covid-19 also in countries where these variants have become dominant.<sup>25–28</sup>

A crucial question for understanding the impact of S-protein evolution on the development of monoclonal antibody (mAb)-based and vaccine-based therapies, is whether we can develop a simple model to rationalize, and eventually predict, the effect of variations on the structural properties of S that ultimately underpin antibody recognition. Fundamentally, comparison across S-proteins mutants can help us understand the molecular basis of the protein’s evolvability, furthering our grasp of the

Table 1. PDB IDs of the S–Ab Complexes Used to Compare Epitope Predictions<sup>a</sup>

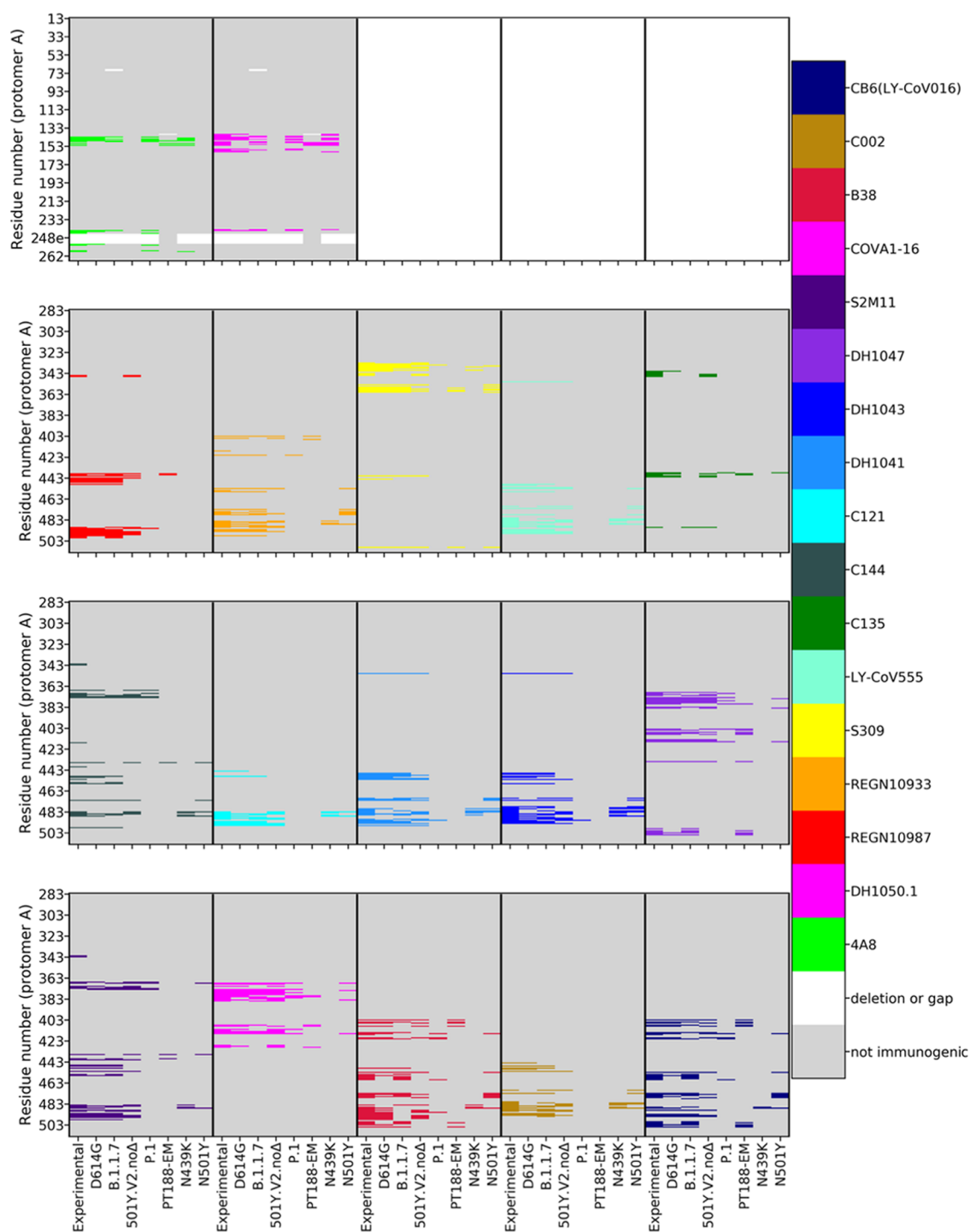
Antibody	PDB ID	D614G	B.1.1.7	501Y.V2.noΔ	P.1	N439K	N501Y	PT188-EM
REGN10987	6XDG	<a href="https://doi.org/10.1126/science.abd0827">10.1126/science.abd0827</a>	<a href="https://doi.org/10.1101/2021.02.18.431897">https://doi.org/10.1101/2021.02.18.431897</a>	<a href="https://doi.org/10.1101/2021.02.18.431897">https://doi.org/10.1101/2021.02.18.431897</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.03.01.433466">https://doi.org/10.1101/2021.03.01.433466</a>	<a href="https://doi.org/10.1016/j.cell.2021.01.037">https://doi.org/10.1016/j.cell.2021.01.037</a>	<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	
REGN10933	6XDG	<a href="https://doi.org/10.1126/science.abd0827">10.1126/science.abd0827</a>	<a href="https://doi.org/10.1101/2021.02.18.431897">https://doi.org/10.1101/2021.02.18.431897</a>	<a href="https://doi.org/10.1101/2021.02.18.431897">https://doi.org/10.1101/2021.02.18.431897</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.03.01.433466">https://doi.org/10.1101/2021.03.01.433466</a>	<a href="https://doi.org/10.1016/j.cell.2021.01.037">https://doi.org/10.1016/j.cell.2021.01.037</a>	<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	
LY-CoV555	7L3N	<a href="https://doi.org/10.1101/2020.09.30.318972">10.1101/2020.09.30.318972</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a> 501Y.V2.Δ		<a href="https://doi.org/10.1016/j.cell.2021.01.037">https://doi.org/10.1016/j.cell.2021.01.037</a>	<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	
S309	7JX3/6WPT	<a href="https://doi.org/10.1038/s41586-020-2349-y">10.1038/s41586-020-2349-y</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.03.01.433466">https://doi.org/10.1101/2021.03.01.433466</a>	<a href="https://doi.org/10.1016/j.cell.2021.01.037">https://doi.org/10.1016/j.cell.2021.01.037</a>	<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	
C135	7K8Z	<a href="https://doi.org/10.1038/s41586-020-2852-1">10.1038/s41586-020-2852-1</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a> 501Y.V2.Δ		<a href="https://doi.org/10.1101/2020.07.21.214759">https://doi.org/10.1101/2020.07.21.214759</a>		
C144	7K90	<a href="https://doi.org/10.1038/s41586-020-2852-1">10.1038/s41586-020-2852-1</a>		<a href="https://doi.org/10.1101/2021.01.27.428478">https://doi.org/10.1101/2021.01.27.428478</a> 501Y.V2.Δ	<a href="https://doi.org/10.7554/eLife.61312">https://doi.org/10.7554/eLife.61312</a>	<a href="https://doi.org/10.1101/2020.07.21.214759">https://doi.org/10.1101/2020.07.21.214759</a>		
C121	7K90	<a href="https://doi.org/10.1038/s41586-020-2852-1">10.1038/s41586-020-2852-1</a>		<a href="https://doi.org/10.1101/2021.01.27.428478">https://doi.org/10.1101/2021.01.27.428478</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.03.01.433466">https://doi.org/10.1101/2021.03.01.433466</a>	<a href="https://doi.org/10.1101/2020.07.21.214759">https://doi.org/10.1101/2020.07.21.214759</a>		
4A8	7C2L	<a href="https://doi.org/10.1126/science.abc6952">10.1126/science.abc6952</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a> 501Y.V2.Δ	<a href="https://doi.org/10.1016/j.immuni.2021.06.003">https://doi.org/10.1016/j.immuni.2021.06.003</a>			<a href="https://doi.org/10.1101/2020.10.07.328302">https://doi.org/10.1101/2020.10.07.328302</a>
DH1041	7LAA	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2021.03.11.435037">https://doi.org/10.1101/2021.03.11.435037</a>	<a href="https://doi.org/10.1016/j.chom.2021.03.002">https://doi.org/10.1016/j.chom.2021.03.002</a>			
DH1043	7LJR	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2021.03.11.435037">https://doi.org/10.1101/2021.03.11.435037</a>	<a href="https://doi.org/10.1016/j.chom.2021.03.002">https://doi.org/10.1016/j.chom.2021.03.002</a>			
DH1047	7LD1	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2021.03.11.435037">https://doi.org/10.1101/2021.03.11.435037</a>	<a href="https://doi.org/10.1016/j.chom.2021.03.002">https://doi.org/10.1016/j.chom.2021.03.002</a>			
DH1050.1	7LCN	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2020.12.31.424729">10.1101/2020.12.31.424729</a>	<a href="https://doi.org/10.1101/2021.03.11.435037">https://doi.org/10.1101/2021.03.11.435037</a>	<a href="https://doi.org/10.1016/j.chom.2021.03.002">https://doi.org/10.1016/j.chom.2021.03.002</a>			
S2M11	7K43	<a href="https://doi.org/10.1126/science.abc3354">10.1126/science.abc3354</a>		<a href="https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf">https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf</a>	<a href="https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf">https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf</a>	<a href="https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf">https://www.citid.cam.ac.uk/wp-content/uploads/2021/02/POST-SUBMISSION_vaccin e-DCv2-2.pdf</a>		
COVA1-16	7JMX	<a href="https://doi.org/10.1101/2020.08.02.233536">10.1101/2020.08.02.233536</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.05.26.21257441">https://doi.org/10.1101/2021.05.26.21257441</a>		<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	
B38	7B25	<a href="https://doi.org/10.1126/science.abc2241">10.1126/science.abc2241</a>	<a href="https://doi.org/10.1016/j.chom.2021.03.002">10.1016/j.chom.2021.03.002</a>				<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	
C002	7K8T	<a href="https://doi.org/10.1038/s41586-020-2852-1">10.1038/s41586-020-2852-1</a>						
CB6	7C01	<a href="https://doi.org/10.1038/s41586-020-2381-y">10.1038/s41586-020-2381-y</a>	<a href="https://doi.org/10.1038/s41586-021-03398-2">https://doi.org/10.1038/s41586-021-03398-2</a>	<a href="https://doi.org/10.1016/j.xcrn.2021.100255">https://doi.org/10.1016/j.xcrn.2021.100255</a> 501Y.V2.Δ	<a href="https://doi.org/10.1101/2021.03.01.433466">https://doi.org/10.1101/2021.03.01.433466</a>		<a href="https://doi.org/10.1002/1873-3468.14076">https://doi.org/10.1002/1873-3468.14076</a>	

<sup>a</sup>For each Ab considered in this work (leftmost column), we report: PDB IDs of S–Ab Cryo-EM complexes used as experimental reference for our MLCE epitope predictions; and, where available, experimental studies reporting either that Ab's gain (yellow) or loss/absence of activity (blue) towards a particular variant. White cells indicate that experimental data is unavailable. \* denotes experimental studies carried out on the 501Y.V2.noΔ S variant but with the Δ241–243 deletion.

relationships between sequence, structure and (immuno)-recognition. From the practical point of view, this knowledge could in principle be harnessed to design and engineer improved S-based antigens or multicomponent domain/peptide combinations, focusing for instance on those antibody binding regions,

known as epitopes, that are predicted to be conserved in multiple variants.

Here, we apply a straightforward structure-dynamics-energy strategy to predict potentially immunogenic regions in



**Figure 2.** Mapping epitopes on each variant. Epitope mapping on S protomer A for the 2 NTD-targeting antibodies (two top left panels; cf. numbering on Y-axis) and the 15 RBD-targeting antibodies (bottom three rows) considered in this study. In each panel, using a distinct color for each antibody (right palette), the experimentally (Cryo-EM or X-ray) detected residues that belong to an epitope (labeled “Experimental” on each panel’s X-axis) are compared to epitopes predicted in silico on each of the seven variants considered. Predicted immunogenic residues are colored according to the Ab they would be targeted by. Non-immunogenic residues are shown in gray; gaps/insertions in white. The figure shows how the extent of the epitopes for the different antibodies varies in the distinct mutants.

representative 3D conformations of several variants of the full-length glycosylated trimeric S protein (Figure 1).

The selected S proteins represent some of the major variants of concern circulating at the time of setting up simulation. In this respect, the African variant we simulate, which is named 501Y.V2.no $\Delta$ , corresponds to the S lineage originally discovered in South Africa in late November 2020 by Tegally et al.<sup>29</sup> This S

variant features the additional mutations L18F (in common with P.1) and R246I but does not feature the  $\Delta$ 241–243 deletion, whose existence was still debated when the authors released their study in January 2021.<sup>29</sup> This variant has subsequently been referred to in several papers as B.1.351.<sup>20,21</sup> The list of studied proteins is further enriched by a laboratory-evolved escape S-variant, obtained by Rappuoli and coworkers by co-



incubating the SARS-CoV-2 virus with a highly neutralizing plasma from a Covid-19 convalescent patient. Interestingly, after several passages this strategy generated a variant completely resistant to plasma neutralization. This “artificial” variant is labeled here as the PT188-EM variant.<sup>21</sup>

Conformations are extracted from independent atomistic molecular dynamics (MD) simulations totaling 4  $\mu$ s for each mutant. Our approach to the detection of epitopes on S, that is, its antibody-binding protein regions, is based on the concept that such sites should continuously evolve to escape immune recognition by the host without impairing the native protein structure required for viral function and survival. We previously showed—and experimentally confirmed—that these regions coincide with substructures that are not involved in major stabilizing intramolecular interactions with core protein residues that are important for its folding into a functional 3D structure.<sup>30</sup> In other words, antigen–antibody (Ab)-interacting regions show minimal energetic coupling with the rest of the protein, which in turn should favor accumulation of escape mutations while preserving the antigen’s 3D structure. Furthermore, minimal intramolecular coupling provides epitopes with greater conformational freedom to adapt to and be recognized by a binding partner. Actual binding to an external partner such as an Ab is expected to occur if favorable intermolecular interactions determine a lower free energy for the bound state than for the unbound state.<sup>30–33</sup>

These concepts are analyzable by the MLCE (matrix of low coupling energy) approach<sup>30,34</sup> (see also [Materials and Methods](#)). Starting from the characterization of the energy of pairwise interactions between all aminoacids and monosaccharides, and filtering the resulting interaction map with structural information extracted from the same protein’s inter-residue contact map, MLCE identifies groups of spatially contiguous residues with poor energetic coupling to the rest of the protein as potential immunogenic regions. At the same time, groups of residues with high energetic coupling are identified as stabilization centers.

Upon comparing our results to recently reported characterization of Ab binding and reactivity, the analysis we report consistently shows that mutations, deletions, and/or insertions in S variants determine a reorganization of internal interactions leading to the loss of potentially immunoreactive regions on the surface. Encouragingly, these findings can be qualitatively reconnected to the decreased ability of some of the Abs elicited against the dominant S-sequence to recognize variants.

## RESULTS

To characterize the effects of mutations, deletions, and insertions on the definition of potential Ab-binding substructures in S variants, we apply a combination of the energy decomposition (ED) and MLCE methods<sup>30,34,35</sup> to representative structures extracted from long timescale MD simulations of the S protein variants reported in [Figure 1](#).

Briefly, we first run 4 independent 1  $\mu$ s long all-atom MD simulations of each variant of the full-length fully glycosylated S protein in solution ([Figure 1](#)) (each built from PDB ID: 6VSB<sup>11</sup>). Next, for each variant, we concatenate individual trajectories into one a single 4  $\mu$ s metatrajectory. Cluster analysis on each variant’s metatrajectory is then conducted to identify the 3 most representative conformations. These are then used to compute nonbonded pairwise potential energy terms (van der Waals, electrostatic interactions, solvent effects) obtaining, for a given variant with N aminoacid and monosaccharide residues, a

symmetric  $N \times N$  inter-residue interaction matrix. The three matrices extracted from a variant’s trajectory are then weighted and averaged to yield an average nonbonded interaction matrix,  $M_{ij}$ . Upon eigenvalue decomposition of  $M_{ij}$ , eigenvectors associated with the most negative eigenvalues can help build a simplified version of  $M_{ij}$  that only highlights series of residues with high- and low-intensity couplings. The former represent residues acting as folding hotspots and responsible stabilizing the protein’s 3D structure; the latter represent residue pairs with weak energetic coupling to the rest of the protein, whose mutation is expected not to impact S’ structure and thus function. In this framework, once information contained in the simplified energy map is combined with information contained in the protein’s residue–residue contact map, it permits to “filter out” clusters of residues whose energetic coupling to the rest of the structure is weak and that are spatially contiguous. Such localized networks of low-intensity couplings, located in proximity of the protein surface represent potential interaction Ab-interaction regions, or epitopes. This approach has been previously experimentally validated in a number of applications.<sup>31,32,36–43</sup>

The reference S structure we use here is the dominant D614G variant. We analyze the results of epitope predictions we obtain on isolated S variants by comparing them against selected spike-antibody complexes. To this end, we collected publicly available X-ray or Cryo-EM structural data of complexes between S and various Abs, reported in [Tables 1](#) and [S1](#). Epitopes in experimental structures are defined as the sets of S protein residues within 5 Å of any Ab residue (see [Supporting Information Table S2](#)). The experimental epitopes thus derived are used as the reference against which to compare epitopes predicted in silico.

[Figure 2](#) schematically reports the sequences of the RBD and NTD for each variant studied (sequences on the Y-axis, variant on the X-axis). The different colors point out the residues of a certain variant that are predicted to be part of an epitope for a certain characterized antibody.

Importantly, for the reference variant predicted epitopes largely overlap with experimentally identified regions. In particular, epitopes are correctly predicted for Abs targeting both the RBD and the NTD of the protein ([Table S2, Figure 2](#)).

In the remaining variants of concern, a diverse landscape of epitopes emerges. A number of residues/regions that are predicted immunogenic in the reference S-protein disappear in the variants. Overall, this is observed for all the Abs considered.

In this framework, after running an epitope prediction on each variant we monitor epitope conservation across variants through a conservation ratio: the number of residues in each predicted epitope for a given variant is divided by the number of residues in the corresponding experimental epitope in the reference S structure, which is defined based on the 5 Å threshold from its respective Ab, as discussed above. We define epitope loss when the conservation ratio is lower than 0.5; otherwise the epitope is considered to be conserved. In [Table 2](#) and [Figure 3](#) we report such conservation ratios for each D614G S epitope on each simulated variant, and confront them with available experimental data (at the time of writing) on the variant’s reactivity towards the Ab that would be expected to bind to that particular epitope. Each cell in the table is color-coded according to the experimentally measured activity of the corresponding Ab on one of the given variants. If the Ab remains active, the cell is yellow. If the Ab has lost activity against that variant, the cell is

Table 2. Epitope Predictions on Each Variant and Epitope Conservation Ratio<sup>a</sup>

Antibodies	D614G	B.1.1.7	501Y.V2.noΔ	P.1	N439K	N501Y	PT188-EM
REGN10987	0.76	0.86	0.52	0.05	0.00	0.00	0.1
REGN10933	0.65	0.78	0.39	0.04	0.13	0.30	0.09
LY-CoV555	0.67	0.83	0.84	0.00	0.21	0.29	0.00
S309	0.77	0.77	0.59	0.05	0.09	0.32	0.18
C135	0.42	0.50	0.58	0.08	0.00	0.08	0.17
C144	0.72	0.31	0.34	0.14	0.14	0.14	0.03
C121	0.77	0.85	0.82	0.00	0.31	0.15	0.00
4A8	0.50	0.31	0.00	0.56	0.44	0.00	0.19
DH1041	0.68	0.59	0.59	0.05	0.23	0.27	0.00
DH1043	0.62	0.81	0.42	0.04	0.27	0.31	0.00
DH1047	0.79	0.71	0.64	0.21	0.00	0.11	0.39
DH1050.1	0.59	0.53	0.00	0.59	0.65	0.00	0.24
S2M11	0.69	0.76	0.55	0.14	0.10	0.10	0.07
COVA1-16	0.76	0.60	0.88	0.32	0.00	0.16	0.20
B38	0.63	0.73	0.41	0.12	0.07	0.20	0.17
C002	0.60	0.70	0.40	0.00	0.25	0.25	0.00
CB6 (LY-CoV016)	0.62	0.74	0.35	0.12	0.06	0.24	0.26

<sup>a</sup>Each cell reports an epitope conservation ratio for each S variant–Ab combination, relating in silico predictions to experimental epitopes from experimental Cryo-EM and/or crystal structures. Conservation ratios lower than 0.5 indicate epitope loss; otherwise an epitope is considered to be conserved. Each cell in the table is color-coded according to the experimentally measured activity of the corresponding Ab on the respective variant. If the Ab remains active, the cell is yellow. If the Ab has lost activity against that variant, the cell is blue. If experimental data is unavailable for a particular Ab on a particular variant, the cell is white. Disagreement between predictions and experiment (i.e., blue and conservation ratio >0.5 or yellow and conservation ratio <0.5) is indicated by thick borders and dotted-line diagonal.

blue. If experimental data is unavailable for a particular Ab on a particular variant, the cell is white.

Analysis of Table 2 clearly shows that the vast majority of blue cells, indicative of a loss of Ab reactivity, contain ratios lower than 0.5. This is an important validation of our prediction: whenever a variant's predicted epitope residues—that is according to MLCE, contiguous residues uncoupled from the S protein core—shrink in number compared to D614G S, it is very likely that experimental data will also confirm that variant evades Abs binding to the shrunk or lost epitopes. On the other hand, the overwhelming majority of cases for which Abs retain activity against a variant (yellow cells) are also confirmed by our prediction to retain their respective epitopes (conservation ratio

> 0.5) with respect to D614G S. Disagreement between our predictions and experiment only occurs in a minority of cases: corresponding cells are marked by thicker borders.

Analysis of B.1.1.7 (U.K.) and 501Y.V2.noΔ South Africa; (late November 2020) immediately shows that a large portion of predicted epitopes in the RBD are conserved compared to the reference D614G. Interestingly, however, we also observe a dramatic drop in the number of NTD residues predicted as epitopes for the 501Y.V2.noΔ. Epitope loss in the NTD, which was deemed to host a super-antigenic hotspot<sup>44</sup> can help explain the ability for immune evasiveness observed for these two variants. In B.1.1.7, the NTD epitope is largely conserved consistent with the conservation of activity of Abs targeting this region against the variant (Table 2, Figures 2 and 3).

Importantly, conservation of a dominant part of the epitopes in the RBD still endows the two variants with reactivity against Abs directed to this domain, which may help explain the observed effectiveness of some convalescent plasma treatments and vaccines.<sup>13,45</sup>

Calculations on the Brazilian variant correctly indicate loss of immunoreactivity of several Abs as well as conserved reactivity of Abs 4A8 and S2M11. This variant is the only one for which our predictions of epitopes binding Abs of the DH family generally disagree with experimental data.

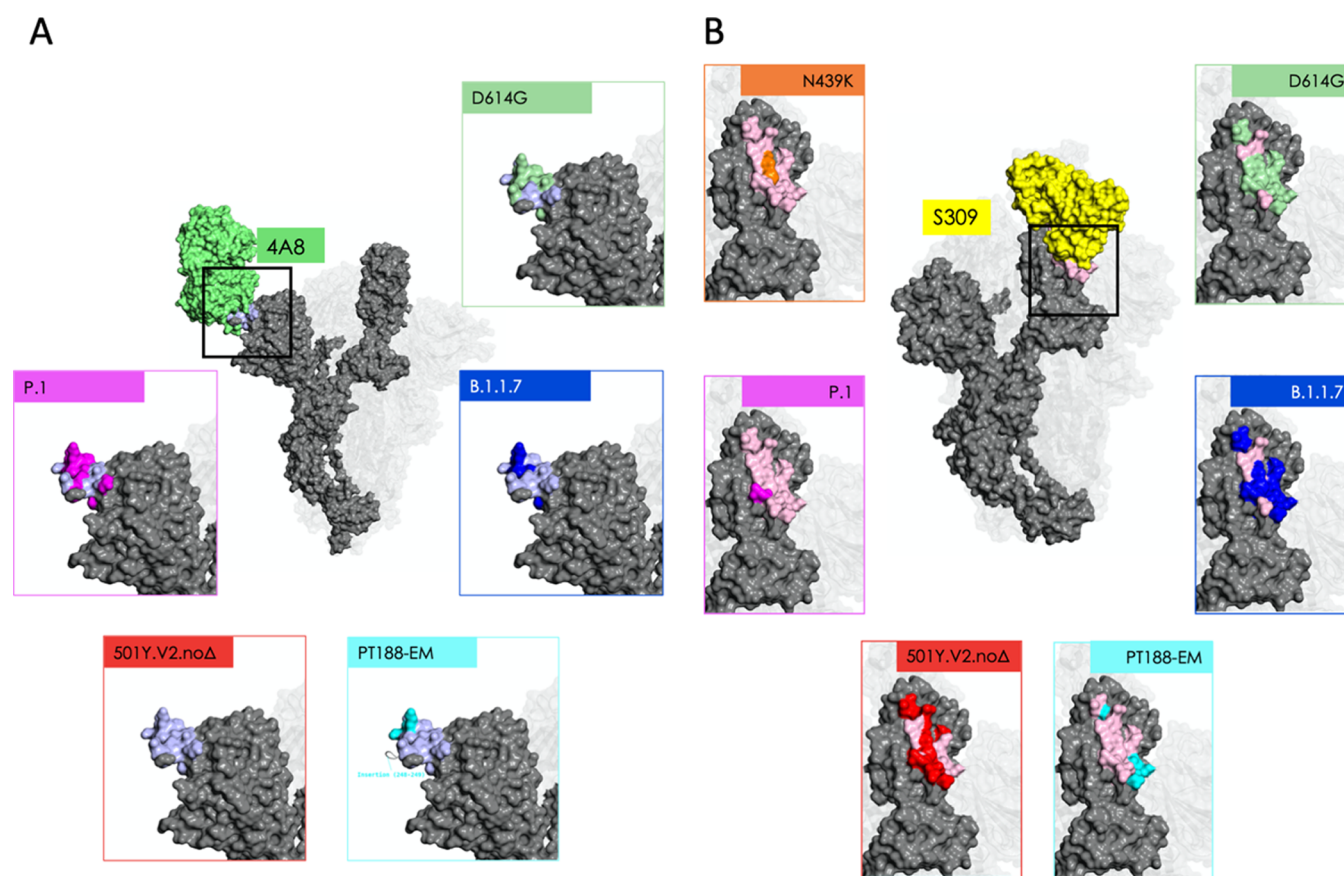
Finally, it is important to note that the “artificial” PT188-EM variant, evolved in the lab under the pressure of convalescent serum to evade Ab-effects, appears to have lost a very large number of protein epitopes (see Table 2, Figures 2 and 3). In particular, the insertion at residues 248 modifies the conformational properties of the region otherwise recognized by Ab 4A8. As a consequence, the epitope to this antibody disappears from the predictions on the PT188-EM variant.<sup>21</sup> Interestingly, in this case, the carbohydrate motifs coating the protein appear to host most of the uncoupled regions (117 carbohydrate moieties in the PT188-EM variant vs 90 in the reference S-protein), pointing to a role of the glycan shield in protecting the protein from immune recognition, besides playing a key part in modulating interactions for ACE2 recognition and cell-entry.<sup>46–52</sup> (see Figure 4).

Importantly, mutants N439K, is correctly predicted as an escape variant from all Abs for which experimental data proved lower efficacy.

## DISCUSSION

In this work, we analyzed full-length models of 7 trimeric glycosylated SARS-CoV-2 S protein variants, derived from the prefusion conformation of the Cryo-EM structure 6VSB,<sup>11</sup> in which the receptor binding domain of chain A (RBD-A) is in an “up” conformation, exposed to interaction with host cell receptors and potential targeting by Abs. The data from our energetic analyses can be aptly integrated in the characterization of the properties of S and other SARS-CoV-2 proteins from long scale simulations, such as those recently presented by Zimmerman et al.,<sup>53</sup> Casalino et al.,<sup>50</sup> Spinello et al.,<sup>54,55</sup> Oliveira et al.,<sup>56</sup> Shoemark et al.,<sup>57</sup> Wang et al.,<sup>58</sup> and Fallon.<sup>59</sup>

Our MLCE analysis of the full-length trimers correctly identifies a number of epitopes in the RBD that have been previously experimentally characterized. RBD is in fact targeted by the largest fraction of neutralizing antibodies. MLCE also identifies regions in the NTD, which are known to be targeted by different Abs, some of which potentially neutralize SARS-CoV-2<sup>19</sup> and highlights putative immunoreactive substructures at the end



**Figure 3.** Mutations modify epitope identity. Central images in panels (A,B) depict the Cryo-EM structure of the antigen-binding fragments of two representative Abs bound to protomer A: 4A8 (panel A; Ab in green; experimental epitope in light blue); and S309 (panel B; Ab in yellow; experimental epitopes in light pink). Insets in each panel contrast the extent of the experimental epitope with epitopic residues predicted by the MLCE method (see main text) for five (panel A) or six (panel B) of the variants considered in this work: these residues are rendered using the same color code used for variants in Figure 1; residues in the experimental epitope not predicted by MLCE are rendered as in the central image (panel A: light blue; panel B: light pink). Other residues on the S protein (not comprised in the experimental epitope) are rendered in gray. Glycans are omitted for clarity; positions of protomers B and C are shown for reference.

of the S2 domain, where sugar-engaging Abs have recently been characterized (see Table S2, Figure 4).

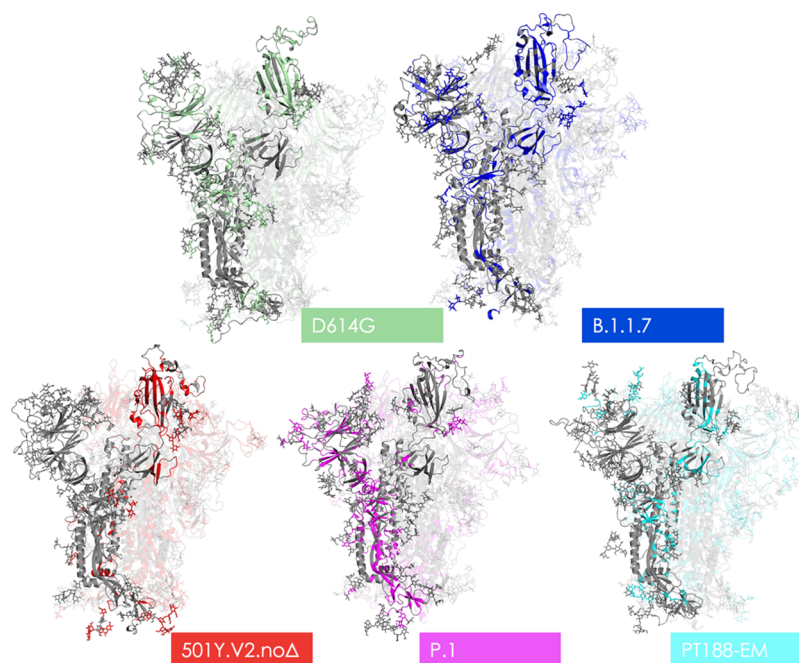
As a caveat, it is worth pointing out here the fact that antibodies can bind to conformations of the spike protein that are different from the ones sampled here. Indeed, work by Casalino et al.,<sup>50</sup> Zimmerman et al.,<sup>53</sup> and by Fallon et al.<sup>59</sup> show that the protein can undergo dramatic structural changes. In our simulations, despite running 4 ms of all-atom MD simulation for each system, we could not observe such changes, if not in their initial stages. To be consistent in our comparative among the different species, we in fact decided to use the same protocol on every system and benchmarked the data obtained against available experiments. This may indeed partly limit the exploration of the conformational space available to this flexible protein, in turn somewhat limiting the prediction of immunogenic regions. We hypothesize that this is the reason behind the limited success we have with the Brazilian variant. The mutations, insertions, deletions in this sequence can expectedly favor the exploration of structures that are different from the ones we are considering here. We notice, however, that a significant number of experimental immunoreactivity data are correctly captured by our approach even on the Brazilian variant, supporting the validity of MLCE in this context.

Energy-based epitope prediction through the MLCE approach reveals a common theme across variants: the number

and surface exposure of potentially immunoreactive regions decrease in S protein mutants compared to the reference D641G. In particular, the number of residues defining the epitope located in the long RBD loop (residues 417–503, recognized by many protective Abs) is much lower in mutants 501Y.V2.noΔ, B.1.1.28, and N439K (see Figures 2 and 3, Table S2). Interestingly, in the case of B.1.1.7, which shows limited evasion, the loop is largely active in terms of immunoreactivity. In contrast, in the evading variant PT188-EM the entire loop disappears from the list of potential Ab-targets.

Potentially important contributions to the perturbation of epitopes' physico-chemical properties may be related to charge variations. Two striking examples are the loss or reduction of epitopes determined by the N439K and E484K mutations. Both cases involve residues that are part of epitopes of a large number of antibodies and after these mutations the antibodies completely or partially lose their efficacies. In the case of the mutation N439K, it has been reported<sup>60</sup> that this variant maintains fitness while evading antibodies immunity. In fact, N439K RBD forms a new interaction with the human ACE2 receptor (hACE2) and has enhanced affinity for hACE2. The salt bridge at the RBD-hACE2 interface (RBD N439K/hACE2 E329) plausibly adds a strong interaction at the binding interface during viral cell entry. On the other hand, the N to K mutation determines stronger intra-spike protein interactions which





**Figure 4.** Structural representations of epitopes on different variants. The various structures depicted show the 3D structure of protomer A in gray. Residues rendered in the color assigned to their respective variant in Figure 2 mark the locations of all predicted epitopes; areas in gray represent non-immunogenic regions. Glycan heavy atoms are rendered as sticks.

dramatically decrease the decoupling of this region from the core, making it substantially less prone to interaction with Abs.

The E484K mutation is of particular concern due to its location within nAb epitopes, and it has been shown to reduce or eliminate binding to many potent RBD-directed nAbs.<sup>61</sup>

Experimental characterization of Abs targeting the NTD revealed a site recognized by most Abs, located between the N3 and N5 loops of the domain. This epitope was correctly predicted in our previous work.<sup>43</sup> Specifically, Lys147 and Arg246, known to be important in stabilizing interactions with the complementarity-determining regions of different Abs are correctly predicted as epitope elements.

On the other hand, sequence mutations in SARS-CoV-2 variants lead to the N3 and N5 NTD loops disappearing from the ensemble of Ab-binding substructures. This is observed computationally and is corroborated by recent experimental data by Veesler and coworkers.<sup>19,44</sup> Interestingly, these epitopes largely coincide with the regions where alanine substitutions reduced affinity for antibodies 4A8, CM17, and CM25 (see ref 62). The impact of epitope loss in these regions is also confirmed by the observation that an engineered N3–N5 double mutant and native  $\beta$  variant<sup>29</sup> both evade neutralization by mAbs CM25 and 4A8.

Interestingly, our approach correctly captures the epitopes for Abs, such as C121 and C144, that are known to engage different RBDs.<sup>63</sup> Antibody C121, for instance, can bind to an RBD in the down conformation and to an adjacent RBD in the up conformation.<sup>63</sup> In the structural paper, the epitope is reported to entail only residues in protomer A with the RBD in the up conformation. Contacts with the nearby RBD in the down conformation are made by Ab residues that are outside the complementary determining region. In this respect, our approach can correctly predict potential immunoreactive sequences even for Abs that would end up binding across different domains. MLCE in fact only aims to predict substructures on the antigen that can potentially be complexed

by one or more Fabs. Focusing only on the antigen, MLCE would not be able to predict whether different epitopes are targeted by the same or distinct Abs at the same time.

Finally, our strategy correctly predicts the loss of most epitopes in the lab-evolved escape variant described by Andreano et al.<sup>21</sup> (see Figure 4, Table S2).

We propose a model for the study of Ab-reactivity of SARS-CoV-2 S protein variants that integrates sequence and structural information and incorporates dynamics and energetics into the analysis of the variation/loss of epitopes. Mutations in S variants determine the loss of epitopes and as a consequence can confer escape from antibodies. Upon sequence variation, the protein shifts to states characterized by different intramolecular interactions compared to the initial D614G structure; this transition decreases the number of energetically uncoupled substructures available for engaging interactors such as Abs. Unique to this model is the observation that mutations, insertions, and deletions exhibiting different immunoreactivity experimentally are consistently captured by the energy based decomposition of structures extracted from unbiased classical MD simulations of the glycosylated S protein isolated in solution, without any input of prior information on Ab-binding propensities. Although qualitative in nature and focused on the study of S variants of concern, our approach is general and immediately portable to other targets to provide physicochemical information on the determinants of Abs recognition.

Since one of the fundamental goals of structural vaccinology is the identification and design of structures with optimized properties for immunoreactivity, development and validation of computational methods that help identify conserved versus non-conserved epitope regions in different variants independently of whether structures of related protein–antibody complexes are available may hold great potential. In the case we have presented here, one may consider designing chimeras or multicomponent systems (peptide- or domain-based) presenting all (or most of)



the conserved sequences that are predicted to be potentially Ab-reactive.

Furthermore, our results suggest that approaches like the one we presented here may be used prospectively as an aid in the analysis and characterization of emerging variants.

Though targeted experiments and design of mutants with tailored reactivities based on MLCE analysis are required to further validate these ideas and precisely define their progression to real-world applicability, our findings provide a new basis to understand how mutations could directly result in escape from immunorecognition.

## MATERIALS AND METHODS

**Preparation of Spike Protein Variants.** Fully glycosylated S protein variants simulated in this work were variously derived from simulations described by Grant et al.<sup>47</sup> based on the Cryo-EM structure of the WT S protein at PDB entry 6VSB,<sup>11</sup> wherein one RBD is in the “up” conformation and the other two are “down”. All mutations, including the “reference” D614G, are introduced using the “mutations wizard” in the PyMOL molecular modeling package (Schrodinger LLC): rotamers of non-glycine side chains are chosen from the first suggested option for S protomer A, and then, where possible, we have sought to adopt the same rotamers for protomers B and C. Histidine tautomers and disulfide bridges are retained as in our reference simulations. In B.1.1.7 variant S protomers, mutant histidines 681 and 1118 are introduced with protonation at N $\epsilon$ 2, and mutant aspartate 570 side chains are left unprotonated. Mutant lysine 484 sidechains (B.1.1.28 variant; E484K variant) are left protonated.

Consistent with our reference simulations,<sup>43,47</sup> all three protomers are modeled without gaps, from Ala27 in the NTD to Asp1146 just downstream of heptapeptide repeat 1 (HR1);  $-\text{NH}_3^+$  and  $-\text{COO}^-$  caps are added, respectively, at N- and C-termini of each protomer.

In the case of the B.1.1.7 variant, gaps left by deletions in all three protomers are replaced with artificially long C–N bonds; systems are then allowed to relax with a 400-step preminimization cycle in vacuo (200 steepest-descent + 200 conjugate gradient), using the AMBER platform’s sander utility (version 18),<sup>64</sup> in which harmonic positional restraints ( $k = 5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ) are applied to all atoms except those in the five residues on either side of the gap. Distortions and clashes introduced with the glycosylated Ser13-Pro26 fragment are resolved using a similar approach.

The artificial PT188-EM was modeled following the methods described in ref 21.

**MD Simulation Details.** After preparation, glycosylated S protein structures are solvated in a cuboidal box of TIP3P water molecules using AMBER’s tleap tool; where necessary,  $\text{Na}^+$  or  $\text{Cl}^-$  ions are added accordingly to neutralize the charge. N-glycosylated asparagines and oligosaccharides are treated using the GLYCAM\_06j forcefield,<sup>65</sup> whereas ions are modeled with parameters by Joung and Cheatham.<sup>66</sup> To all other (protein) atoms, we apply the ff14SB forcefield.<sup>67</sup> Starting structures and topologies for all simulated variants are electronically provided as Supporting Information.

On each glycosylated S protein variant, we conduct 4 independently replicated atomistic MD simulations, using the AMBER package (version 18): each replica consists of two 300-step rounds of minimization, 2.069 ns preproduction, and 1  $\mu\text{s}$  production. The sander MD engine<sup>64</sup> is used into the earlier

stages of preproduction; thereafter, we switch to the GPU-accelerated pmemd.cuda.<sup>64</sup>

**Details on MD Production.** The 1  $\mu\text{s}$  production stage is carried out in the  $NpT$  ensemble ( $T = 300 \text{ K}$ ;  $p = 1 \text{ atm}$ ) using a 2 fs time step; a cutoff of 8.0  $\text{\AA}$  is applied for the calculation of Lennard-Jones and Coulomb interactions alike. Coulomb interactions beyond this limit are computed using the particle mesh Ewald method.<sup>68</sup> All bonds containing hydrogen are restrained using the SHAKE algorithm.<sup>69</sup> Constant pressure is enforced via Berendsen’s barostat<sup>70</sup> with a 1 ps relaxation time, whereas temperature is stabilized by Langevin’s thermostat<sup>71</sup> with a 5  $\text{ps}^{-1}$  collision frequency.

**Details on MD Preproduction.** Prior to the production stage, every independent MD replica for every S variant goes through a series of preproduction steps, namely: minimization, solvent equilibration, system heating, and equilibration. The first two are conducted using the sander utility, after which the GPU-accelerated pmemd.cuda is invoked instead.

Minimization takes place in two 300-step rounds, the first 10 of which use the steepest-descent algorithm and the last 290 conjugate gradient. In the first round, we only minimize backbone  $\text{H}\alpha$  and  $\text{H1}$  hydrogens on aminoacids and monosaccharides, respectively, restraining all other atoms harmonically ( $k = 5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ). Thereafter, all atoms are released, including solvent and ions.

Solvent equilibration occurs over 9 ps with a time step of 1 fs; the ensemble is  $NVT$ , with temperatures in this case enforced by the Berendsen thermostat.<sup>70</sup> Positions of non-solvent atoms are harmonically restrained ( $k = 10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ). Solvent molecules are assigned initial random velocities to match a temperature of 25 K. Fast heating to 400 K (coupling: 0.2 ps) is performed over the first 3 ps; the solvent is then retained at 400 K for another 3 ps; and cooled back down to 25 K over the last 3 ps, more slowly (coupling: 2.0). The cutoff for determining Lennard-Jones and Coulomb interactions remains at 8.0  $\text{\AA}$  for this and all subsequent stages, as does the particle mesh Ewald method<sup>68</sup> to determine Coulomb interactions beyond this cutoff. SHAKE constraints<sup>69</sup> are not applied at this stage, but are always present thereafter.

For system heating, the time step is increased to 2 fs and, whilst continuing in the  $NVT$  ensemble, temperatures are now enforced by the Langevin thermostat<sup>71</sup> (which remains in place for all subsequent stages). With an initial collision frequency of 0.75  $\text{ps}^{-1}$ , the system is heated from 25 to 300 K over 20 ps: all atoms are free to move except aminoacids’  $\text{C}\alpha$  atoms, which are positionally restrained with  $k = 5 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ .

For equilibration, the ensemble is switched to  $NpT$  ( $p = 1 \text{ atm}$ ; Berendsen barostat coupling: 1 ps), and the system is simulated for a further 2040 ps. The thermostat’s collision frequency is kept lower than in the production stage (1  $\text{ps}^{-1}$ ). Restraints on  $\text{C}\alpha$  atoms are lifted gradually:  $k = 3.75 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  for the first 20 ps; 1.75  $\text{kcal mol}^{-1} \text{ \AA}^{-2}$  for the following 20 ps; none thereafter.

**Clustering of MD Simulations.** Following MD, each variant’s 4 replicas are concatenated into a single 4  $\mu\text{s}$  “metatrajectory”, desolvated, stripped of any ions, and aligned on backbone heavy atoms of all aminoacid residues, in all three protomers, that belong to neither the NTD nor the RBD according to domain definitions by Huang et al.<sup>72</sup> Clustering calculations are then conducted using the hierarchical agglomerative algorithm,<sup>73</sup> considering every 20th metatrajectory frame (i.e., every 50 ps), based on the root-mean-square deviation of backbone heavy atoms of aminoacid residues

composing the NTD and the RBD in all three protomers. Values of  $\epsilon$  are chosen so that they provide the best compromise between maximizing cluster homogeneity, based on silhouette score, and ensuring at least 60–80% of the metatrjectory is covered by the three most populated clusters: this usually means  $\epsilon = 9–12$ .

All of the steps discussed in the previous paragraph are conducted using AMBER's postprocessing utility *cpptraj*.

**MLCE Method.** Potential epitopes on each S variant are predicted using the MLCE method (of which we also provide a more detailed account in our previous work).<sup>43</sup> The procedure is automatically carried out by our own in-house code (<https://github.com/colombolab/MLCE>) which we have now rewritten to rely on the computationally more efficient MMPBSA.py utility<sup>74</sup> instead of *mm\_pbsa.pl*.

To begin with, three representative S protein structures for each variant (i.e., the centroids of its three most populated clusters; see [Clustering of MD Simulations](#)) are minimized for 200 steps (10 steepest descent; 190 conjugate gradient), with the sander engine, in implicit solvent (per the modified generalized born model by Onufriev et al.).<sup>75</sup> The cutoff used for the computation of Lennard-Jones and Coulomb interactions is 12.0 Å, under nonperiodic conditions; the mobile counterion concentration is always set at 0.1 M for all variants, and the solvent-accessible surface area is calculated by employing linear combinations of pairwise overlaps.

After minimization, MMPBSA.py<sup>74</sup> is initialized and uses the MM/GBSA method<sup>76</sup> to construct a nonbonded pairwise interaction matrix  $M_s$  for each of the variant's representative structures  $s$ : otherwise put, each term  $M_{s,ij}$  in this matrix contains the van der Waals and Coulomb interactions (including 1–4 interaction terms) for a representative structure's  $i$ th and  $j$ th residues (aminoacids and monosaccharides alike). Settings for this stage are identical to minimization, apart from a 0 M implicit ion concentration.  $M_s$  matrices for a variant's representative structures  $s = 1, 2, \text{ and } 3$  are then averaged, and scaled by the fraction of that variant's trajectory represented by their parent cluster: this gives an average weighted nonbonded interaction matrix  $M$ .

Using the validated approach explained in detail by Genoni and coworkers,<sup>77</sup> our code performs eigen decomposition of  $M$ : in other words, each averaged interaction matrix element  $M_{ij}$ —that is for each individual pair formed by the  $i$ th and  $j$ th residues—is first diagonalized and re-expressed as

$$M_{ij} = \sum_{\alpha=1}^N \lambda_{\alpha} v_i^{\alpha} v_j^{\alpha} \quad (1)$$

where  $N$  is the total number of aminoacid and glycan residues,  $\lambda_{\alpha}$  is the  $\alpha$ th eigenvalue, and  $v_i^{\alpha}$  and  $v_j^{\alpha}$  are the  $i$ th and  $j$ th components of its associated vector. From this eigen decomposition, the code is programmed to (re)select a minimum number of essential eigenvectors  $N_e$  that are sufficient to “cover” interactions of the maximum number of residues in the protein. The full details of the selection are described in ref 77. This step results in an “essential folding matrix”  $M^{\text{fold}}$  constructed from  $M$ 's  $N_e$  essential eigen vectors, and whose elements are each expressed as follows

$$M_{ij}^{\text{fold}} = \sum_{\alpha=1}^{N_e} \lambda_{\alpha} v_i^{\alpha} v_j^{\alpha} \quad (2)$$

that is, no longer as a sum over all vectors (residues)  $N$ , but only over the chosen  $N_e$  essential eigenvectors.  $M^{\text{fold}}$  thus only contains information on whether interaction of a variant's  $i$ th and  $j$ th's residues is actually more or less stabilizing for the folding of one of that variant's domains<sup>77</sup> (e.g., NTD, RBD, etc.). This is in contrast to  $M$ , whose elements  $M_{ij}$  simply indicate whether a variant's  $i$ th and  $j$ th's residues attract, repel, or don't interact.

To obtain the final MLCE matrix, the essential folding matrix  $M^{\text{fold}}$  is subsequently Hadamard-multiplied by a pairwise residue–residue contact matrix  $C$

$$\text{MLCE} = M^{\text{fold}} \odot C \quad (3)$$

$C$ 's elements  $C_{ij}$  are either 0 or 1, depending on whether  $C\beta$  atoms in the  $i$ th and  $j$ th residues ( $C1$  in the case of monosaccharides;  $H$  in the case of glycines) fall below or above an arbitrary 6.0 Å threshold, respectively. In this way, each MLCE element  $\text{MLCE}_{ij}$  will be nonzero if and only if residues  $i$  and  $j$  are spatially contiguous and exhibit an energetic interaction that is stabilizing for the folding of a particular domain, in which case  $\text{MLCE}_{ij}$  takes the actual value of the pair's degree of stabilization.

Individual  $\text{MLCE}_{ij}$  elements are ultimately ranked from the most stabilizing (i.e., pairs that are most energetically relevant for the folding of their particular domain) to least stabilizing (i.e., pairs showing the weakest energetic coupling within their domains). Our final epitope predictions are made by isolating the top 10% weakest-interacting spatially contiguous residue pairs yielded by this ranking.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.1c00857>.

Table S1 explicitly showing all the references used in which Abs are characterized; Table S2 with the residues that are part of the experimentally determined epitopes and of the predicted epitopes on the various spike variants. The numbering of the residues refers to that reported in UNIPROT entry P0DTC2. <https://www.uniprot.org/uniprot/P0DTC2> (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

Giorgio Colombo – Department of Chemistry, University of Pavia, Pavia 27100, Italy; [orcid.org/0000-0002-1318-668X](https://orcid.org/0000-0002-1318-668X); Email: [g.colombo@unipv.it](mailto:g.colombo@unipv.it)

### Authors

Alice Triveri – Department of Chemistry, University of Pavia, Pavia 27100, Italy

Stefano A. Serapian – Department of Chemistry, University of Pavia, Pavia 27100, Italy

Filippo Marchetti – Department of Chemistry, University of Pavia, Pavia 27100, Italy

Filippo Doria – Department of Chemistry, University of Pavia, Pavia 27100, Italy

Silvia Pavoni – Department of Physical Chemistry, R&D Eni SpA, Milan 20097, Italy

Fabrizio Cinquini – Upstream & Technical Services—TECS/STES—Eni Spa, Milan 20097, Italy

Elisabetta Moroni – Istituto di Scienze e Tecnologie Chimiche “Giulio Natta”—SCITEC, Milano 20131, Italy

Andrea Rasola – Department of Biomedical Sciences, University of Padua, Padova 35131, Italy

Francesco Frigerio – Department of Physical Chemistry, R&D Eni SpA, Milan 20097, Italy

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jcim.1c00857>

## Notes

The authors declare no competing financial interest.

Requirements for existing software applies to this manuscript: software used for MD simulations is AMBER version 18 available at <http://ambermd.org/>; epitope prediction is run with the MLCE code available at: <https://github.com/colombolab/MLCE>. Requirements for data availability apply to this manuscript. For simulations: AMBER input files, topologies, and restart (rst) files for each variant are provided in **Supporting Information** as a GDRIVE directory at the link [https://drive.google.com/drive/folders/1Z9dkf6HPNsQ79Q0D2f07OGjNTmj0\\_eXF?usp=sharing](https://drive.google.com/drive/folders/1Z9dkf6HPNsQ79Q0D2f07OGjNTmj0_eXF?usp=sharing). The material for each simulation is stored in a directory named after the mutant. The structures of all variants on which epitope predictions were run are provided in the same directory in a file named spike-clusters.zip. Given their large dimensions, full trajectories are available upon request.

## ACKNOWLEDGMENTS

The authors thank Prof. Robert J. Woods and Dr. Oliver Grant (University of Georgia) for kindly providing the original atomistic molecular dynamics simulations of the fully glycosylated spike proteins. G.C. gratefully acknowledges Dr. Guido Scarabelli, Dr. Riccardo Capelli, and Dr. Claudio Peri for previous work on epitope prediction. This research was partially supported by a Grant from “Cassa di Risparmio di Padova e Rovigo (CARIPARO) PROGETTI DI RICERCA SUL COVID-19”.

## REFERENCES

- (1) Dellus-Gur, E.; Toth-Petroczy, A.; Elias, M.; Tawfik, D. S. What Makes a Protein Fold Amenable to Functional Innovation? Fold Polarity and Stability Trade-Offs. *J. Mol. Biol.* **2013**, *425*, 2609–2621.
- (2) Wellner, A.; Raitsev Gurevich, M.; Tawfik, D. S. Mechanisms of Protein Sequence Divergence and Incompatibility. *PLoS Genet.* **2013**, *9*, No. e1003665.
- (3) Tóth-Petróczy, Á.; Tawfik, D. S. The Robustness and Innovability of Protein Folds. *Curr. Opin. Struct. Biol.* **2014**, *26*, 131–138.
- (4) Yanagida, H.; Gispan, A.; Kadouri, N.; Rozen, S.; Sharon, M.; Barkai, N.; Tawfik, D. S. The Evolutionary Potential of Phenotypic Mutations. *PLoS Genet.* **2015**, *11*, No. e1005445.
- (5) Andreano, E.; Rappuoli, R. Sars-Cov-2 Escaped Natural Immunity, Raising Questions About Vaccines and Therapies. *Nat. Med.* **2021**, *27*, 759–761.
- (6) Pecetta, S.; Pizza, M.; Sala, C.; Andreano, E.; Pileri, P.; Troisi, M.; Pantano, E.; Manganaro, N.; Rappuoli, R. Antibodies, Epicenter of Sars-Cov-2 Immunology. *Cell Death Differ.* **2021**, *28*, 821–824.
- (7) Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. Protein Stability Promotes Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5869–5874.
- (8) Yuan, M.; Huang, D.; Lee, C.-C. D.; Wu, N. C.; Jackson, A. M.; Zhu, X.; Liu, H.; Peng, L.; van Gils, M. J.; Burton, D. R.; Reincke, S. M.; Prüss, H.; Kreye, J.; Nemazee, D.; Ward, A. B.; Wilson, I. A. Structural and Functional Ramifications of Antigenic Drift in Recent Sars-Cov-2 Variants. *Science* **2021**, No. eabh1139.

(9) Altmann, D. M.; Boyton, R. J.; Beale, R. Immunity to Sars-Cov-2 Variants of Concern. *Science* **2021**, *371*, 1103–1104.

(10) Harvey, W. T.; Carabelli, A. M.; Jackson, B.; Gupta, R. K.; Thomson, E. C.; Harrison, E. M.; Ludden, C.; Reeve, R.; Rambaut, A.; Peacock, S. J.; Robertson, D. L.; Consortium, C.-G. U. Sars-Cov-2 Variants, Spike Mutations and Immune Escape. *Nat. Rev. Microbiol.* **2021**, *19*, 409–424.

(11) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-Em Structure of the 2019-Ncov Spike in the Prefusion Conformation. *Science* **2020**, *367*, 1260–1263.

(12) McDonald, I.; Murray, S. M.; Reynolds, C. J.; Altmann, D. M.; Boyton, R. J. Comparative Systematic Review and Meta-Analysis of Reactogenicity, Immunogenicity and Efficacy of Vaccines against Sars-Cov-2. *npj Vaccines* **2021**, *6*, 74.

(13) Wall, E. C.; Wu, M.; Harvey, R.; Kelly, G.; Warchal, S.; Sawyer, C.; Daniels, R.; Hobson, P.; Hatipoglu, E.; Ngai, Y.; Hussain, S.; Nicod, J.; Goldstone, R.; Ambrose, K.; Hindmarsh, S.; Beale, R.; Riddell, A.; Gambin, S.; Howell, M.; Kassiotis, G.; Libri, V.; Williams, B.; Swanton, C.; Gandhi, S.; Bauer, D. L. Neutralising Antibody Activity against Sars-Cov-2 Voccs B.1.617.2 and B.1.351 by Bnt162b2 Vaccination. *Lancet* **2021**, *397*, 2331–2333.

(14) Haas, E. J.; Angulo, F. J.; McLaughlin, J. M.; Anis, E.; Singer, S. R.; Khan, F.; Brooks, N.; Smaja, M.; Mircus, G.; Pan, K.; Southern, J.; Swerdlow, D. L.; Jodar, L.; Levy, Y.; Alroy-Preis, S. Impact and Effectiveness of Mrna Bnt162b2 Vaccine against Sars-Cov-2 Infections and Covid-19 Cases, Hospitalisations, and Deaths Following a Nationwide Vaccination Campaign in Israel: An Observational Study Using National Surveillance Data. *Lancet* **2021**, *397*, 1819–1829.

(15) Hsieh, C.-L.; Goldsmith, J. A.; Schaub, J. M.; DiVenere, A. M.; Kuo, H.-C.; Javanmardi, K.; Le, K. C.; Wrapp, D.; Lee, A. G.; Liu, Y.; Chou, C.-W.; Byrne, P. O.; Hjorth, C. K.; Johnson, N. V.; Ludes-Meyers, J.; Nguyen, A. W.; Park, J.; Wang, N.; Amengor, D.; Lavinder, J. J.; Ippolito, G. C.; Maynard, J. A.; Finkelstein, I. J.; McLellan, J. S. Structure-Based Design of Prefusion-Stabilized Sars-Cov-2 Spikes. *Science* **2020**, *369*, 1501.

(16) Wang, X.; Du, Z.; Johnson, K. E.; Pasco, R. F.; Fox, S. J.; Lachmann, M.; McLellan, J. S.; Meyers, L. A. Effects of Covid-19 Vaccination Timing and Risk Prioritization on Mortality Rates, United States. *Emerging Infect. Dis.* **2021**, *27*, 1976–1979.

(17) Greaney, A. J.; Starr, T. N.; Gilchuk, P.; Zost, S. J.; Binshtein, E.; Loes, A. N.; Hilton, S. K.; Huddleston, J.; Eguia, R.; Crawford, K. H. D.; Dingens, A. S.; Nargi, R. S.; Sutton, R. E.; Suryadevara, N.; Rothlauf, P. W.; Liu, Z.; Whelan, S. P. J.; Carnahan, R. H.; Crowe, J. E.; Bloom, J. D. Complete Mapping of Mutations to the Sars-Cov-2 Spike Receptor-Binding Domain That Escape Antibody Recognition. *Cell Host Microbe* **2021**, *29*, 44–57.

(18) Starr, T. N.; Greaney, A. J.; Addetia, A.; Hannon, W. W.; Choudhary, M. C.; Dingens, A. S.; Li, J. Z.; Bloom, J. D. Prospective Mapping of Viral Mutations That Escape Antibodies Used to Treat Covid-19. *Science* **2021**, *371*, 850.

(19) McCallum, M.; Bassi, J.; Marco, A. D.; Chen, A.; Walls, A. C.; Julio, J. D.; Tortorici, M. A.; Navarro, M.-J.; Silacci-Fregni, C.; Saliba, C.; Agostini, M.; Pinto, D.; Culap, K.; Bianchi, S.; Jaconi, S.; Cameroni, E.; Bowen, J. E.; Tilles, S. W.; Pizzuto, M. S.; Guastalla, S. B.; Bona, G.; Pellanda, A. F.; Garzoni, C.; Van Voorhis, W. C.; Rosen, L. E.; Snell, G.; Telenti, A.; Virgin, H. W.; Piccoli, L.; Corti, D.; Veessler, D. Sars-Cov-2 Immune Evasion by Variant B.1.427/B.1.429. **2021**, bioRxiv:10.1101/2021.03.31.437925.

(20) Planas, D.; Bruel, T.; Grzelak, L.; Guivel-Benhassine, F.; Staropoli, I.; Porrot, F.; Planchais, C.; Buchrieser, J.; Rajah, M. M.; Bishop, E.; Albert, M.; Donati, F.; Prot, M.; Behillil, S.; Enouf, V.; Maquart, M.; Smati-Lafarge, M.; Varon, E.; Schortgen, F.; Yahyaoui, L.; Gonzalez, M.; De Sèze, J.; Péré, H.; Veyer, D.; Sève, A.; Simon-Lorière, E.; Fafi-Kremer, S.; Stefic, K.; Mouquet, H.; Hocqueloux, L.; van der Werf, S.; Prazuck, T.; Schwartz, O. Sensitivity of Infectious Sars-Cov-2 B.1.1.7 and B.1.351 Variants to Neutralizing Antibodies. *Nat. Med.* **2021**, *27*, 917–924.



- (21) Andreano, E.; Piccini, G.; Licastro, D.; Casalino, L.; Johnson, N. V.; Paciello, I.; Monego, S. D.; Pantano, E.; Manganaro, N.; Manenti, A.; Manna, R.; Casa, E.; Hyseni, I.; Benincasa, L.; Montomoli, E.; Amaro, R. E.; McLellan, J. S.; Rappuoli, R. Sars-Cov-2 Escape in Vitro from a Highly Neutralizing Covid-19 Convalescent Plasma. *2020*, bioRxiv:10.1101/2020.12.28.424451.
- (22) McCormick, K. D.; Jacobs, J. L.; Mellors, J. W. The Emerging Plasticity of Sars-Cov-2. *Science* **2021**, *371*, 1306.
- (23) Wang, Z.; Schmidt, F.; Weisblum, Y.; Muecksch, F.; Barnes, C. O.; Finkin, S.; Schaefer-Babajew, D.; Cipolla, M.; Gaebler, C.; Lieberman, J. A.; Oliveira, T. Y.; Yang, Z.; Abernathy, M. E.; Huey-Tubman, K. E.; Hurley, A.; Turroja, M.; West, K. A.; Gordon, K.; Millard, K. G.; Ramos, V.; Da Silva, J.; Xu, J.; Colbert, R. A.; Patel, R.; Dizon, J.; Unson-O'Brien, C.; Shimeliovich, I.; Gazumyan, A.; Caskey, M.; Bjorkman, P. J.; Casellas, R.; Hatzioannou, T.; Bieniasz, P. D.; Nussenzweig, M. C. Mrna Vaccine-Elicited Antibodies to Sars-Cov-2 and Circulating Variants. *Nature* **2021**, *592*, 616–622.
- (24) Muik, A.; Wallisch, A.-K.; Sanger, B.; Swanson, K. A.; Muhl, J.; Chen, W.; Cai, H.; Maurus, D.; Sarkar, R.; Tureci, ˆ.; Dormitzer, P. R.; ˆahin, U. Neutralization of Sars-Cov-2 Lineage B.1.1.7 Pseudovirus by Bnt162b2 Vaccine–Elicited Human Sera. *Science* **2021**, *371*, 1152.
- (25) Abu-Raddad, L. J.; Chemaitelly, H.; Butt, A. A. Effectiveness of the Bnt162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants. *N. Engl. J. Med.* **2021**, *385*, 187–189.
- (26) Tarke, A.; Sidney, J.; Methot, N.; Yu, E. D.; Zhang, Y.; Dan, J. M.; Goodwin, B.; Rubiro, P.; Sutherland, A.; Wang, E.; Frazier, A.; Ramirez, S. I.; Rawlings, S. A.; Smith, D. M.; da Silva Antunes, R.; Peters, B.; Scheuermann, R. H.; Weiskopf, D.; Crotty, S.; Grifoni, A.; Sette, A. Impact of Sars-Cov-2 Variants on the Total Cd4<sup>+</sup> and Cd8<sup>+</sup> T Cell Reactivity in Infected or Vaccinated Individuals. *Cell Rep. Med.* **2021**, *2*, 100355.
- (27) European Centre for Disease Prevention and Control. *A Assessing Sars-Cov-2 Circulation, Variants of Concern, Non-pharmaceutical Interventions and Vaccine Rollout in the Eu/Eea*; ECDC: Stockholm, 2021, 15th Update.
- (28) Madhi, S. A.; Baillie, V.; Cutland, C. L.; Voysey, M.; Koen, A. L.; Fairlie, L.; Padayachee, S. D.; Dheda, K.; Barnabas, S. L.; Bhorat, Q. E.; Briner, C.; Kwatra, G.; Ahmed, K.; Aley, P.; Bhikha, S.; Bhiman, J. N.; Bhorat, A. a. E.; du Plessis, J.; Esmail, A.; Groenewald, M.; Horne, E.; Hwa, S.-H.; 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, C. K.; Durham, N. M.; Kelly, E. J.; Villafana, T. L.; Gilbert, S.; Pollard, A. J.; de Oliveira, T.; Moore, P. L.; Sigal, A.; Izu, A. Efficacy of the Chadox1 Ncov-19 Covid-19 Vaccine against the B.1.351 Variant. *N. Engl. J. Med.* **2021**, *384*, 1885–1898.
- (29) Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E. J.; Msomi, N.; Mlisana, K.; von Gottberg, A.; Walaza, S.; Allam, M.; Ismail, A.; Mohale, T.; Glass, A. J.; Engelbrecht, S.; Van Zyl, G.; Preiser, W.; Petruccione, F.; Sigal, A.; Hardie, D.; Marais, G.; Hsiao, N.-y.; Korsman, S.; Davies, M.-A.; Tyers, L.; Mudau, I.; York, D.; Maslo, C.; Goedhals, D.; Abrahams, S.; Laguda-Akingba, O.; Alisoltani-Dehkordi, A.; Godzik, A.; Wibmer, C. K.; Sewell, B. T.; Lourenço, J.; Alcantara, L. C. J.; Kosakovsky Pond, S. L.; Weaver, S.; Martin, D.; Lessells, R. J.; Bhiman, J. N.; Williamson, C.; de Oliveira, T. Detection of a Sars-Cov-2 Variant of Concern in South Africa. *Nature* **2021**, *592*, 438–443.
- (30) Scarabelli, G.; Morra, G.; Colombo, G. Predicting Interaction Sited from the Energetics of Isolated Proteins: A New Approach to Epitope Mapping. *Biophys. J.* **2010**, *98*, 1966–1975.
- (31) Gourlay, L. J.; Peri, C.; Ferrer-Navarro, M.; Conchillo-Sole, O.; Gori, A.; Rinchai, D.; Thomas, R. J.; Champion, O. L.; Michell, S. L.; Kewcharoenwong, C.; Nithichanon, A.; Lassaux, P.; Perletti, L.; Longhi, R.; Lertmemongkolchai, G.; Titball, R. W.; Daura, X.; Colombo, G.; Bolognesi, M. Exploiting the Burkholderia Pseudomallei Acute Phase Antigen Bpsl2765 for Structure-Based Epitope Discovery/Design in Structural Vaccinology. *Chem. Biol.* **2013**, *20*, 1147–1156.
- (32) Lassaux, P.; Peri, C.; Ferrer-Navarro, M.; Gourlay, L.; Gori, A.; Conchillo-Sole, O.; Rinchai, D.; Lertmemongkolchai, G.; Longhi, R.; Daura, X.; Colombo, G.; Bolognesi, M. A Structure-Based Strategy for Epitope Discovery in Burkholderia Pseudomallei Oppa Antigen. *Structure* **2013**, *21*, 167–175.
- (33) Gourlay, L.; Peri, C.; Bolognesi, M.; Colombo, G. Structure and Computation in Immunoreagent Design: From Diagnostics to Vaccines. *Trends Biotechnol.* **2017**, *35*, 1208–1220.
- (34) Capelli, R.; Serapian, S. A.; Colombo, G. Computational Epitope Prediction and Design for Antibody Development and Detection. *Meth. Mol. Biol.* **2021**. In press.
- (35) Marchetti, F.; Capelli, R.; Rizzato, F.; Laio, A.; Colombo, G. The Subtle Trade-Off between Evolutionary and Energetic Constraints in Protein-Protein Interactions. *J. Phys. Chem. Lett.* **2019**, *10*, 1489–1497.
- (36) Soriani, M.; Petit, P.; Grifantini, R.; Petracca, R.; Gancitano, G.; Frigimelica, E.; Nardelli, F.; Garcia, C.; Spinelli, S.; Scarabelli, G.; Fiorucci, S.; Affentranger, R.; Ferrer-Navarro, M.; Zacharias, M.; Colombo, G.; Vuillard, L.; Daura, X.; Grandi, G. Exploiting Antigenic Diversity for Vaccine Design: The Chlamydia Artj Paradigm. *J. Biol. Chem.* **2010**, *285*, 30126–30138.
- (37) Gourlay, L. J.; Lassaux, P.; Thomas, R. J.; Peri, C.; Conchillo-Sole, O.; Nithichanon, A.; Ferrer-Navarro, M.; Vila, J.; Daura, X.; Lertmemongkolchai, G.; Titball, R.; Colombo, G.; Bolognesi, M. Flagellar Subunits as Targets for Structure-Based Epitope Discovery Approaches and Melioidosis Vaccine Development. *FEBS J.* **2015**, *282*, 338.
- (38) Gourlay, L. J.; Thomas, R. J.; Peri, C.; Conchillo-Sole, O.; Ferrer-Navarro, M.; Nithichanon, A.; Vila, J.; Daura, X.; Lertmemongkolchai, G.; Titball, R.; Colombo, G.; Bolognesi, M. From Crystal Structure to in Silico Epitope Discovery in the Burkholderia Pseudomallei Flagellar Hook-Associated Protein Flgk. *FEBS J.* **2015**, *282*, 1319–1333.
- (39) Nithichanon, A.; Rinchai, D.; Gori, A.; Lassaux, P.; Peri, C.; Conchillo-Sole, O.; Ferrer-Navarro, M.; Gourlay, L. J.; Nardini, M.; Vila, J.; Daura, X.; Colombo, G.; Bolognesi, M.; Lertmemongkolchai, G. Sequence- and Structure-Based Immunoreactive Epitope Discovery for Burkholderia Pseudomallei Flagellin. *PLoS Neglected Trop. Dis.* **2015**, *9*, No. e0003917.
- (40) Gori, A.; Peri, C.; Quilici, G.; Nithichanon, A.; Gaudesi, D.; Longhi, R.; Gourlay, L.; Bolognesi, M.; Lertmemongkolchai, G.; Musco, G.; Colombo, G. Flexible Vs Rigid Epitope Conformations for Diagnostic- and Vaccine-Oriented Applications: Novel Insights from the Burkholderia Pseudomallei Bpsl2765 Pa13 Epitope. *ACS Infect. Dis.* **2016**, *2*, 221–230.
- (41) Gori, A.; Sola, L.; Gagni, P.; Bruni, G.; Liprino, M.; Peri, C.; Colombo, G.; Cretich, M.; Chiari, M. Screening Complex Biological Samples with Peptide Microarrays: The Favorable Impact of Probe Orientation Via Chemoselective Immobilization Strategies on Clickable Polymeric Coatings. *Bioconjugate Chem.* **2016**, *27*, 2669–2677.
- (42) Sola, L.; Gagni, P.; D'Annessa, I.; Capelli, R.; Bertino, C.; Romanato, A.; Damin, F.; Bergamaschi, G.; Marchisio, E.; Cuzzocrea, A.; Bombaci, M.; Grifantini, R.; Chiari, M.; Colombo, G.; Gori, A.; Cretich, M. Enhancing Antibody Serodiagnosis Using a Controlled Peptide Coimmobilization Strategy. *ACS Infect. Dis.* **2018**, *4*, 998–1006.
- (43) Serapian, S. A.; Marchetti, F.; Triveri, A.; Morra, G.; Meli, M.; Moroni, E.; Sautto, G. A.; Rasola, A.; Colombo, G. The Answer Lies in the Energy: How Simple Atomistic Molecular Dynamics Simulations May Hold the Key to Epitope Prediction on the Fully Glycosylated Sars-Cov-2 Spike Protein. *J. Phys. Chem. Lett.* **2020**, *11*, 8084–8093.
- (44) McCallum, M.; De Marco, A.; Lempp, F. A.; Tortorici, M. A.; Pinto, D.; Walls, A. C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; Zepeda, S.; di Iulio, J.; Bowen, J. E.; Montiel-Ruiz, M.; Zhou, J.; Rosen, L. E.; Bianchi, S.; Guarino, B.; Fregni, C. S.; Abdelnabi, R.; Foo, S.-Y. C.; Rothlauf, P. W.; Bloyet, L.-M.; Benigni, F.; Camerini, E.; Neyts, J.; Riva, A.; Snell, G.; Telenti, A.; Whelan, S. P. J.; Virgin, H. W.; Corti, D.; Pizzuto, M. S.; Veessler, D. N-Terminal Domain Antigenic Mapping Reveals a Site of Vulnerability for Sars-Cov-2. *Cell* **2021**, *184*, 2332–2347.e16.



- (45) Yadav, P. D.; Sapkal, G. N.; Ella, R.; Sahay, R. R.; Nyayanit, D. A.; Patil, D. Y.; Deshpande, G.; Shete, A. M.; Gupta, N.; Mohan, V. K.; Abraham, P.; Panda, S.; Bhargava, B. Neutralization against B.1.351 and B.1.617.2 with Sera of Covid-19 Recovered Cases and Vaccinees of Bbv152. **2021**, bioRxiv:10.1101/2021.06.05.447177.
- (46) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. 3d Models of Glycosylated Sars-Cov-2 Spike Protein Suggest Challenges and Opportunities for Vaccine Development. **2020**, bioRxiv:10.1101/2020.04.07.030445.
- (47) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the Sars-Cov-2 Spike Protein Glycan Shield Reveals Implications for Immune Recognition. *Sci. Rep.* **2020**, *10*, 14991.
- (48) Smith, C. C.; Entwistle, S.; Willis, C.; Vensko, S.; Beck, W.; Garness, J.; Sambade, M.; Routh, E.; Olsen, K.; Kodysh, J.; O'Donnell, T.; Haber, C.; Heiss, K.; Stadler, V.; Garrison, E.; Grant, O. C.; Woods, R. J.; Heise, M.; Vincent, B. G.; Rubinsteyn, A. Landscape and Selection of Vaccine Epitopes in Sars-Cov-2. **2020**, bioRxiv:10.1101/2020.06.04.135004.
- (49) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. Virus-Receptor Interactions of Glycosylated Sars-Cov-2 Spike and Human Ace2 Receptor. *Cell Host Microbe* **2020**, *28*, 586–601.
- (50) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the Sars-Cov-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722–1734.
- (51) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of Ace2 in the Context of Sars-Cov-2 Infection. *Biophys. J.* **2021**, *120*, 1072–1084.
- (52) Casalino, L.; Dommer, A.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L.; Simmerling, C.; Hardy, D. J.; Maia, J. D. C.; Phillips, J. C.; Kurth, T.; Stern, A.; Huang, L.; McCalpin, J.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. Ai-Driven Multiscale Simulations Illuminate Mechanisms of Sars-Cov-2 Spike Dynamics. **2020**, bioRxiv:10.1101/2020.11.19.390187.
- (53) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallamadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; Hurley, M. F. D.; Harbison, A. M.; Fogarty, C. A.; Coffland, J. E.; Fadda, E.; Voelz, V. A.; Chodera, J. D.; Bowman, G. R. Sars-Cov-2 Simulations Go Exascale to Predict Dramatic Spike Opening and Cryptic Pockets across the Proteome. *Nat. Chem.* **2021**, *13*, 651–659.
- (54) Spinello, A.; Saltalamacchia, A.; Magistrato, A. Is the Rigidity of Sars-Cov-2 Spike Receptor-Binding Motif the Hallmark for Its Enhanced Infectivity? Insights from All-Atom Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 4785–4790.
- (55) Spinello, A.; Saltalamacchia, A.; Borišek, J.; Magistrato, A. Allosteric Cross-Talk among Spike's Receptor-Binding Domain Mutations of the Sars-Cov-2 South African Variant Triggers an Effective Hijacking of Human Cell Receptor. *J. Phys. Chem. Lett.* **2021**, *12*, 5987–5993.
- (56) Oliveira, A. S. F.; Ibarra, A. A.; Bermudez, I.; Casalino, L.; Gaieb, Z.; Shoemark, D. K.; Gallagher, T.; Sessions, R. B.; Amaro, R. E.; Mulholland, A. J. A Potential Interaction between the Sars-Cov-2 Spike Protein and Nicotinic Acetylcholine Receptors. *Biophys. J.* **2021**, *120*, 983–993.
- (57) Shoemark, D. K.; Colenso, C. K.; Toelzer, C.; Gupta, K.; Sessions, R. B.; Davidson, A. D.; Berger, I.; Schaffitzel, C.; Spencer, J.; Mulholland, A. J. Molecular Simulations Suggest Vitamins, Retinoids and Steroids as Ligands of the Free Fatty Acid Pocket of the Sars-Cov-2 Spike Protein\*. *Angew. Chem., Int. Ed.* **2021**, *60*, 7098–7110.
- (58) Wang, Y.; Fallon, L.; Raguette, L.; Budhan, S.; Belfon, K.; Stepanenko, D.; Helbock, S.; Varghese, S.; Simmerling, C. Receptor Binding May Directly Activate the Fusion Machinery in Coronavirus Spike Glycoproteins. **2021**, bioRxiv:10.1101/2021.05.10.443496.
- (59) Fallon, L.; Belfon, K. A. A.; Raguette, L.; Wang, Y.; Stepanenko, D.; Cuomo, A.; Guerra, J.; Budhan, S.; Varghese, S.; Corbo, C. P.; Rizzo, R. C.; Simmerling, C. Free Energy Landscapes from Sars-Cov-2 Spike Glycoprotein Simulations Suggest That Rbd Opening Can Be Modulated Via Interactions in an Allosteric Pocket. *J. Am. Chem. Soc.* **2021**, *143*, 11349–11360.
- (60) Thomson, E. C.; Rosen, L. E.; Shepherd, J. G.; Spreafico, R.; da Silva Filipe, A.; Wojcechowskyj, J. A.; Davis, C.; Piccoli, L.; Pascall, D. J.; Dillen, J.; Lytras, S.; Czudnochowski, N.; Shah, R.; Meury, M.; Jesudason, N.; De Marco, A.; Li, K.; Bassi, J.; O'Toole, A.; Pinto, D.; Colquhoun, R. M.; Culap, K.; Jackson, B.; Zatta, F.; Rambaut, A.; Jaconi, S.; Sreenu, V. B.; Nix, J.; Zhang, L.; Jarrett, R. F.; Glass, W. G.; Beltramello, M.; Nomikou, K.; Pizzuto, M.; Tong, L.; Cameroni, E.; Croll, T. I.; Johnson, N.; Di Iulio, J.; Wickenhagen, A.; Ceschi, A.; Harbison, A. M.; Mair, D.; Ferrari, P.; Smollett, K.; Sallusto, F.; Carmichael, S.; Garzoni, C.; Nichols, J.; Galli, M.; Hughes, J.; Riva, A.; Ho, A.; Schiuma, M.; Semple, M. G.; Openshaw, P. J. M.; Fadda, E.; Baillie, J. K.; Chodera, J. D.; Rihn, N. S.; Lycett, S. J.; Virgin, H. W.; Telenti, A.; Corti, D.; Robertson, D. L.; Snell, G. Circulating Sars-Cov-2 Spike N439k Variants Maintain Fitness While Evading Antibody-Mediated Immunity. *Cell* **2021**, *184*, 1171–1187.
- (61) Gobeil, S. M.-C.; Janowska, K.; McDowell, S.; Mansouri, K.; Parks, R.; Stalls, V.; Kopp, M. F.; Manne, K.; Li, D.; Wiehe, K.; Saunders, K. O.; Edwards, R. J.; Korber, B.; Haynes, B. F.; Henderson, R.; Acharya, P. Effect of Natural Mutations of Sars-Cov-2 on Spike Structure, Conformation, and Antigenicity. *Science* **2021**, *373*, No. eabi6226.
- (62) Voss, W. N.; Hou, Y. J.; Johnson, N. V.; Delidakis, G.; Kim, J. E.; Javanmardi, K.; Horton, A. P.; Bartzoka, F.; Paresi, C. J.; Tanno, Y.; Chou, C.-W.; Abbasi, S. A.; Pickens, W.; George, K.; Boutz, D. R.; Towers, D. M.; McDaniel, J. R.; Billick, D.; Goike, J.; Rowe, L.; Batra, D.; Pohl, J.; Lee, J.; Gangappa, S.; Sambhara, S.; Gadush, M.; Wang, N.; Person, M. D.; Iverson, B. L.; Gollihar, J. D.; Dye, J. M.; Herbert, A. S.; Finkelstein, I. J.; Baric, R. S.; McLellan, J. S.; Georgiou, G.; Lavinder, J. J.; Ippolito, G. C. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science* **2021**, *372*, 1108.
- (63) Barnes, C. O.; West, A. P., Jr.; Huey-Tubman, K. E.; Hoffmann, M. A. G.; Sharaf, N. G.; Hoffman, P. R.; Koranda, N.; Gristick, H. B.; Gaebler, C.; Muecksch, F.; Lorenzi, J. C. C.; Finkin, S.; Hägglöf, T.; Hurley, A.; Millard, K. G.; Weisblum, Y.; Schmidt, F.; Hatzioannou, T.; Bieniasz, P. D.; Caskey, M.; Robbiani, D. F.; Nussenzweig, M. C.; Bjorkman, P. J. Structures of Human Antibodies Bound to Sars-Cov-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **2020**, *182*, 828.
- (64) Case, D. A.; Cerutti, D. S.; Cheatham, T. E. I.; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Greene, D.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P. L.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. M.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2018*; University of California: San Francisco, 2018.
- (65) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. Glycam06: A Generalizable Biomolecular Force Field. *J. Comput. Chem.* **2008**, *29*, 622–655.
- (66) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
- (67) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99sb. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (68) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N-Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(69) Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the Shake and Rattle Algorithms for Rigid Water Models. *J. Comp. Chem.* **1992**, *13*, 952–962.

(70) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(71) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanine-N'-Methylamide. *Biopolymers* **1992**, *32*, 523–535.

(72) Huang, Y.; Yang, C.; Xu, X.-f.; Xu, W.; Liu, S.-w. Structural and Functional Properties of Sars-Cov-2 Spike Protein: Potential Antivirus Drug Development for Covid-19. *Acta Pharmacol. Sin.* **2020**, *41*, 1141–1149.

(73) Defays, D. An Efficient Algorithm for a Complete Link Method. *Comput. J.* **1977**, *20*, 364–366.

(74) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. Mmpbsa.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314–3321.

(75) Onufriev, A.; Bashford, D.; Case, D. A. Modification of the Generalized Born Model Suitable for Macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.

(76) Genheden, S.; Ryde, U. The Mm/Pbsa and Mm/Gbsa Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10*, 449–461.

(77) Genoni, A.; Morra, G.; Colombo, G. Identification of Domains in Protein Structures from the Analysis of Intramolecular Interactions. *J. Phys. Chem. B* **2012**, *116*, 3331–3343.