Check for updates

Up State of the SARS-COV-2 Spike Homotrimer Favors an Increased Virulence for New Variants

Carolina Corrêa Giron [1,2], Aatto Laaksonen [3,4,5,6] and Fernando Luís Barroso da Silva [1,7]*

[1] *Departamento de Ciências Biomoleculares, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil,* [2] *Hospital de Clínicas, Universidade Federal do Triângulo Mineiro, Uberaba, Brazil,* [3] *Arrhenius Laboratory, Department of Materials and Environmental Chemistry, Stockholm University, Stockholm, Sweden,* [4] *State Key Laboratory of Materials-Oriented and Chemical Engineering, Nanjing Tech University, Nanjing, China,* [5] *Centre of Advanced Research in Bionanoconjugates and Biopolymers, Petru Poni Institute of Macromolecular Chemistry, Iasi, Romania,* [6] *Division of Energy Science, Department of Engineering Sciences and Mathematics, Luleå University of Technology, Luleå, Sweden,* [7] *Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, United States*

The COVID-19 pandemic has spread worldwide. However, as soon as the first vaccines—the only scientifically verified and efficient therapeutic option thus far—were released, mutations combined into variants of SARS-CoV-2 that are more transmissible and virulent emerged, raising doubts about their efficiency. This study aims to explain possible molecular mechanisms responsible for the increased transmissibility and the increased rate of hospitalizations related to the new variants. A combination of theoretical methods was employed. Constant-pH Monte Carlo simulations were carried out to quantify the stability of several spike trimeric structures at different conformational states and the free energy of interactions between the receptor-binding domain (RBD) and angiotensin-converting enzyme II (ACE2) for the most worrying variants. Electrostatic epitopes were mapped using the PROCEEDpKa method. These analyses showed that the increased virulence is more likely to be due to the improved stability to the S trimer in the opened state, in which the virus can interact with the cellular receptor, ACE2, rather than due to alterations in the complexation RBD-ACE2, since the difference observed in the free energy values was small (although more attractive in general). Conversely, the South African/Beta variant (B.1.351), compared with the SARS-CoV-2 wild type (wt), is much more stable in the opened state with one or two RBDs in the up position than in the closed state with three RBDs in the down position favoring the infection. Such results contribute to understanding the natural history of disease and indicate possible strategies for developing new therapeutic molecules and adjusting the vaccine doses for higher B-cell antibody production.

Keywords: SARS-CoV-2, mutations, conformational states, coronavirus, electrostatic interactions, epitopes, binding affinity, protein–protein interactions

INTRODUCTION

Severe acute respiratory syndrome corona virus-2 (SARS-CoV-2), the new beta coronavirus responsible for the COVID-19 pandemic, infected more than 100 million people and killed more than two million, approximately 110 times more than the 2009 H1N1 pandemic officially killed within 1 year [1, 2]. Even though the numbers are alarming, the data is not being collected evenly

GRAPHICAL ABSTRACT | Effects of the conformational states of the SARS-CoV-2 spike protein variants on their electrostatic stability and their impact on virulence.
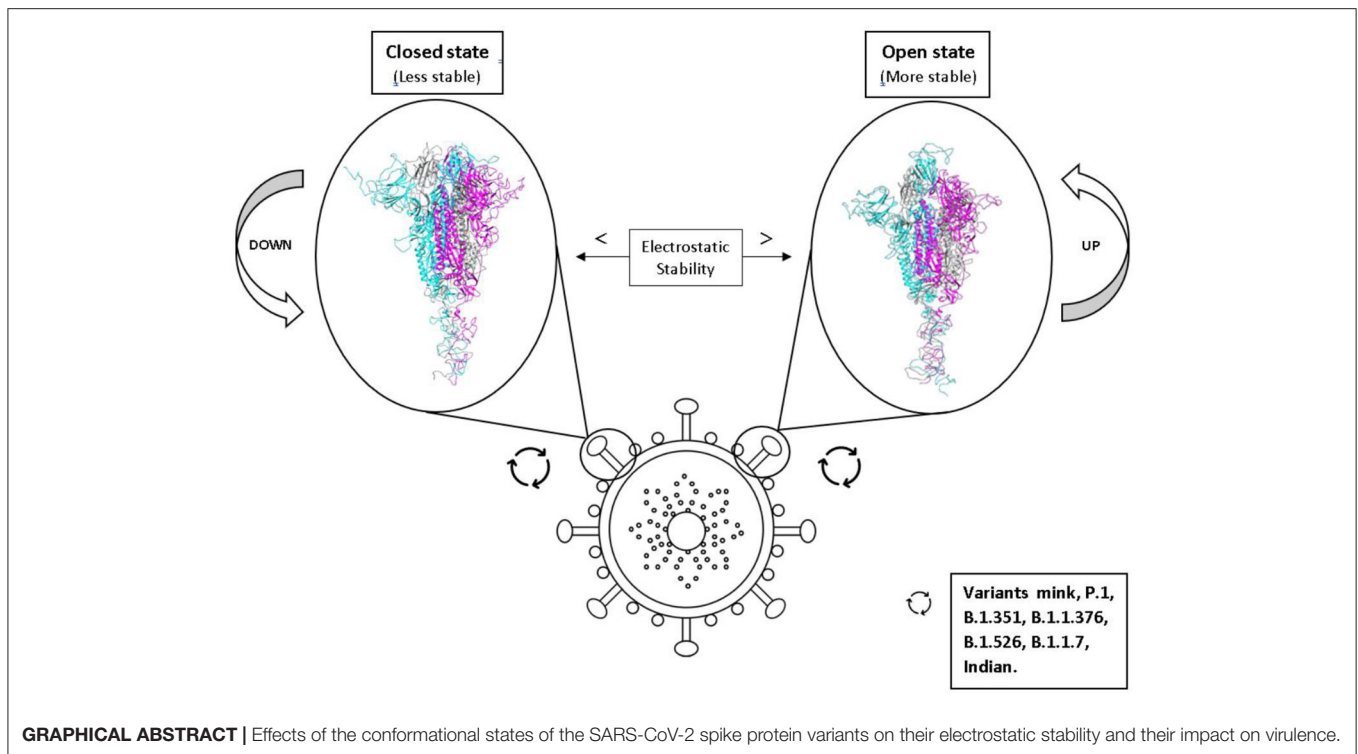
due to different coping strategies adopted worldwide. Moreover, the test results are not blindly reliable. For example, the RT-PCR test, considered trustworthy, was found to have high false-negative rates due to insufficient cellular material and different viral load kinetics of SARS-CoV-2, depending on the patient and sampling timing (3–5). The numbers of the COVID-19 deaths, as a result, are underestimated and often far from reality. The 2009 H1N1 pandemic mortality, for example, was 15 times higher than the number shown by laboratory-confirmed results (2).

Both SARS-CoV-2 and the other coronavirus SARS-CoV-1, responsible for the 2003 Severe Acute Respiratory Syndrome pandemic in China, use a very similar mechanism to promote the fusion of viral and cellular membranes to infect the human cell (6). They both interact with the Angiotensin Converting Enzyme II (ACE2) receptor through the Spike (S) glycoprotein receptor-binding domain, which shares 74% of sequence identity—see Jaimes et al. (7)—with a nearly identical binding conformation and similar affinities to promote the cell entry (6, 8). Virus entry is also related to the presence of N-linked glycans around the post-fusion spike protein structure, which probably act in a protective role against host immune responses (9). Although the virus already has some natural immune escape mechanisms, mutations can increase this escape, meaning that people who have already been infected could remain susceptible to reinfection, such as what happened in Manaus, Brazil (10). All viruses mutate as they replicate numerous times. Although coronaviruses have a relatively efficient proofreading mechanism (11), as many mutations have already been reported. For instance, the "Global Initiative on Sharing All Influenza Data"

(GISAID) (12) repository has more than 804 k (17 March 2021) genome assemblies available in their "official hCoV-19 Reference Sequence." They adopted the high-quality genome sequence "hCoV-19/Wuhan/WIV04/2019," isolated from a clinical sample at the Wuhan Jinyintan Hospital in Hubei Province on December 30, 2019, as the official reference sequence.

Typically, point mutations with a defective genome do not represent any health issue. However, mutations driven by an adaptive evolution are an advantage for the virus, and as such, they represent a possible increased risk to human health, raising doubts about the efficiency of both the antibody therapies and the developed vaccines. The major concern arises from the knowledge that small differences in genetic material can substantially alter the properties of the viral proteins and offer extra advantages to viruses, such as higher ability to transmit or the capacity to escape the control of antibodies (whether produced due to a previous infection, received *via* intravenous administration or stimulated by vaccination) (13). Due to the rising fear of possible consequences that new variants may bring to the outcome of the pandemic, local outbreaks have been studied, and some of the variants responsible for them are being classified as variants of concern (VOCs) by some health organizations from around the world, such as Centers for Disease Control and Prevention (CDC) and European Centre for Disease Prevention and Control (ECDC) (14, 15).

To date, there are currently three principal VOCs considered: (a) 501Y.V1, also called VOC 202012/01 and B.1.1.7, (b) B.1.351, also called 501Y.V2, and (c) P.1 (15). Recently, the World Health Organization (WHO) labeled them as Alpha, Beta and Gamma,

respectively. All these three variants are of concern because of the specific mutations that increased the transmissibility of the virus and its harmful effects on society. 501Y.V1, a variant detected in the United Kingdom and approximately 50% more transmissible, carries eight mutations on the Spike glycoprotein homotrimer. However, three of them are particularly worrisome—N501Y, meaning that residue number 501 had an asparagine replaced by a tyrosine, P681H, and the deletion of residues 69 (histidine) and 70 (valine) (16–19).

B.1.351, on the other hand, was first detected in South Africa and had, apart from N501Y (included in the three VOCs), the mutations K417T and E484K at the receptor-binding domain (RBD), the same key mutations present in the P.1 lineage, detected in Brazil (15, 20). Even though the impact of these mutations on the course of the pandemic is still partially unknown, different authors have elucidated some key aspects of the mutations mentioned. N501Y, for example, is located in the RBD and may increase ACE2 binding and transmissibility, while the infectivity rises by the deletion 69/70 (21, 22). The other two mutations found in both B.1.351 and P.1, E484K and K417N, have a crucial role in the viral escape, preventing the neutralization by some antibodies (10, 20). Although individually these mutations might generate changes in the trimeric properties, it is known that the combination of E484K, K417N, and N501Y mutations cause a greater conformational change in the RBD than N501Y or E484K alone (23). Another worrisome modification in B.1.351 and P.1 is the mutations D614G, once 614G variant, and ORF1ab 4715L, which were proven to be related to higher fatality rates (24).

The main clinical aspects of COVID-19, unlike efficient treatment strategies, are well documented and known in the literature, at least for the wild-type virus. The illness usually starts with nonspecific symptoms, such as fever, persistent cough, and fatigue (25). However, loss of smell, taste, and sense, delirium, skipped meals, and gastrointestinal symptoms, like abdominal pain and diarrhea, can also be considered to identify individuals infected with SARS-CoV-2 (26). In a more advanced stage, constituting the severe form of the disease, shortness of breath can also be observed, which usually leads to hospitalization. However, the clinical aspects of COVID-19 caused by the new variants are still being elucidated.

Experimental data suggest that P.1 (or Gamma) and B.1.351 (or Beta) are partially or even totally resistant to antibodies developed for the treatment of COVID-19 and are inefficiently inhibited by the serum from convalescent individuals (27). Moreover, based on analysis from the "New and Emerging Respiratory Virus Threats Advisory Group" (28), from the United Kingdom, the infection caused by the B.1.1.7 variant is associated with a higher rate of hospitalization and death when compared to the wild type of SARS-CoV-2 (SARS-CoV-2 wt) (28). All these results suggest that the increased transmissibility and a potential antigenic escape are the reason why there is a resurgence of cases and why individuals, previously infected with the wild type, are only partially protected against the VOC (27, 29).

Indeed, among several possible interdependent mechanisms that can increase viral transmissivity (e.g., increased viral shedding, the longer interval of contagiousness, people mobility, wearing masks), a couple of them are directly related to mutations: (a) an increased binding affinity between the viral adhesins with specific cell receptors (in this case, the RBD-ACE2 complexation, that defines the virus virulence and possible increased infectivity), (b) an increased environmental stability of the viral proteins and the virion particle, (c) a higher availability of human cell receptors by either direct genetic conditions [e.g., the concentration of receptors ACE2 tends to be more elevated in men (30)] or due to comorbidities [e.g., a lower stomachal pH in patients with Barrett's esophagus induces a higher expression of ACE2 (31)], (d) an increased interaction with co-receptors, and (e) immune evasion (32–35). Each of these individual factors and also their combinations have been discussed in the literature. For instance, there are studies investigating other possible interactions for the spike protein to enter the cell (36, 37), genetic factors affecting the viral virulence [e.g., the ACE2 polymorphisms can also influence the virus entry in the host cell and individual susceptibility (38)], and the immune evasion (39–41).

The interaction between viral proteins and cell receptors can vary from virus to virus and for different mutations (33, 42). The spike RBD comparisons for its binding affinity with ACE2 for both SARS-CoV-1 and SARS-CoV-2 (8, 43–45) have been reported. However, the data is still lacking for the new variants of SARS-CoV-2, quantifying their impact on virulence. Starr and co-authors partially addressed this question in their landmark experimental research work where they performed a deep mutational scanning of SARS-CoV-2 RBD (46). They provided information about the effect of single mutations on the binding, stability, and expression (46). However, nothing is known about double or triple substitutions. Moreover, the spike trimer must undergo a complex mechanism to make the RBD available for a proper binding with ACE2. Cryo-electron microscopy (CryoEM) structures of the SARS-CoV-1 spike trimer revealed that at least one chain of the homotrimer has to be at the "up" position ("open" conformation) as a prerequisite conformational state for the RBD-ACE2 interaction (47). At least a two-step "expose–dock-like" mechanism is needed first to allow the whole homotrimeric structure to perform all the conformational adjustments (first step of this complex process) before steric clashes (seen for the Spike "close" conformation) are removed to allow the complexation RBD-ACE2 to happen (second step) (43, 47). Up to three receptors, ACE2 can be bound one by one (48). Since several mutations present in the new variants occurred at the spike protein outside the RBD region (e.g., for the 501Y.V2 variant, L18F, D80A, D215G, R246I, D614G, and A701V), it is expected that they have an impact on the "up" and "down" mechanism of the exposing step. Also, the number of titratable amino acids involved in these mutations suggests an electrostatic dependence. Nevertheless, the pH effects are somehow contradictory or, at least, not fully understood. On one side, pH does not seem to be particularly important to trigger the conformational changes from the "down" to the "up" state (49). For example, no significant conformational changes were observed in CryoEM at lower pH (pH 5.6) in comparison with neutral-pH (pH 7.2) for SARS-CoV-1 (48). On

another side, Zhou et al. (50) concluded that an immune evasion could be facilitated for SARS-CoV-2 by the low pH "down" confirmation because of a pH-dependent refolded region located at the spike–interdomain interface, consisting of residues 824–858, that exhibited structural modifications and RBD-mediated positioning of the trimer apex (50).

Following a previous study on the interactions of the RBDs of SARS-CoV-1 and SARS-CoV-2 spike proteins and the human cell receptor ACE2 (43), we investigated the interactions of the RBD of the new most worrying variants (at present) and other mutations with ACE2. Using constant-pH biophysical simulation methods, the binding affinities between the RBDs of these variants with ACE2 were quantified at different pH regimes and the electrostatic epitopes (EEs) of each case mapped and compared. For this analysis, the spike RBD was assumed to be ready for the interaction at the proper conformational state (i.e., only the second step of the "expose–dock-like" mechanism was studied). Another important aspect explored in this study was the electrostatic stability of the different possible conformational states of the trimer [all S chains at the "down" state (DDD), one chain at the "up" state, and two others at the "down" state (UDD), and two chains at the "up" state and one "down" state (DUU)] for some variants at all pH regimes. Such combined data can provide answers at the molecular level for the increased transmissivity of the new variants observed in clinical practice, explaining the faster spread of them, the increased number of hospitalizations, and a tendency to affect younger patients.

MATERIALS AND METHODS

The field of virology is widely explored and enriched by computational approaches. It involves from bioinformatics tools and machine learning methods to biophysical simulations to understand the various aspects of viral functioning, including its immunology, pathogenesis, structural and molecular biology of the virus proteins (51–53). Complementing experimental studies, computational tools allow, for example, the understanding of molecular mechanisms of the virus, such as the comprehension of capsid proteins assembly (assembly intermediates are still difficult to be obtained by lab experiments), the quantification of the biomolecular interactions of the host–pathogen system, the prediction of conventional epitopes and EEs, the understanding of increased virulence for different strains, and the development of specific molecular binders for diagnosis, treatment, and prevention (43, 54–60).

Biophysical simulations of virus systems, based on computational molecular simulation methods such as Monte Carlo (MC) (61, 62) and classical Molecular Dynamics (MD) (61, 63) can take advantage of their long-recorded success for probing the thermodynamic, dynamic, and interactive properties of biomolecules in pharmaceuticals [see (64, 65) for reviews]. Here, a fast constant-pH MC scheme (66, 67) is applied to identify important residues for host-pathogen interactions and to clarify the intermolecular interactions involving the RBD of S proteins of SARS-CoV-1, 2 and the South African variant (B.1.351) and the electrostatic stabilities (68) of the spike

homotrimers for the DDD, DUU and UDD conformational states at all solution pHs. Additionally, other recent variants like the Brazilian P.1 (alias of B.1.1.28.1), the Californian (B.1.427/B.1.429), the New York (B.1.526), the Indian double mutations E484Q and L452R (B.1.617), and the SARS-CoV-2 mink-associated variant strain (Y453F) were included in our analyses. Due to the emergence of more than one nomenclature for these variants of the SARS-CoV-2 virus, the WHO provided quite recently a new naming system based on the Greek alphabet: Alpha is used to refer to B.1.1.7 (UK), Beta to B.1.351 (South African), Gamma to P.1 (Brazilian), Kappa to B.1.617.1 (the Indian double mutant), Epsilon to B.1.427/B.1.429 (Californian) and Iota to B.1.526 (New York).

Molecular Systems and Their Structural Modeling

In the present study, several molecular systems were investigated by employing the SARS-CoV-1, 2 and the new variants of the S RBD proteins with ACE2 (simulation set 1). Also, the electrostatic stability of the whole spike homotrimeric protein at different conformational states was studied (simulation set 2). A scheme of these two simulation sets is given in **Figure 1**. The three-dimensional coordinates of these macromolecules necessary for all these simulations were obtained from different sources:

(a) The SARS-CoV-1 S RBD wt protein (RBD1wt): It was extracted from the RCSB Protein Data Bank (PDB) (69), where it was deposited with the PDB id 2AJF (chain E, resolution 2.9 Å, pH 7.5) and found complexed with ACE2 (chain A) – see **Figure 2**.

(b) The SARS-CoV-2 S RBD wt protein (RBD2wt): For the sake of comparison with a previous study (43), we used the coordinates obtained from the comparative modeling of the three-dimensional protein structure built up at the SWISS-MODEL workspace (YP_009724390.1) based on the NCBI reference sequence NC_045512 (70). See Reference (43) for the details. The root-mean-square deviation (RMSD) of atomic positions between this modeled structure for the RBD of SARS-CoV-2 S wt protein and the available one for SARS-CoV-1 (PDB id 2AJF) is 0.64 Å. A comparison with experimental SARS-CoV-2 S RBD structures deposited after this theoretical model was proposed [e.g., PDB ids 6W41 (resolution 3.08 Å, pH 4.6) and 6YM0 (resolution 4.36 Å, pH 8)], which revealed similar RMSDs (0.51–0.56 Å) to this one (0.64Å). This is an important feature showing that the sequence-structure relationship is robust enough to handle even mutations while still preserving the overall fold of the molecules. A previous study have also indicated this feature (71). Additional runs were performed with RBDs extracted from the PBD ids 6VSB (prefusion open state with one RBD at the up position) and 6VXX (close state). These are CryoEM coordinates obtained with a resolution of 3.46 and 2.8 Å, respectively. We shall refer to them as RBD2$_{wt'}$(6VSB) and RBD2$_{wt'}$(6VXX).

(c) The SARS-CoV-2 S RBD variants (RBD2$_{variant}$): The coordinates for all studied variants were modeled by the simple replacement of amino acids from the SARS-CoV-2 S RBD wt structure followed by an energy minimization using "UCSF Chimera 1.14" (72). Rotamers using the Dunbrack 2010 library
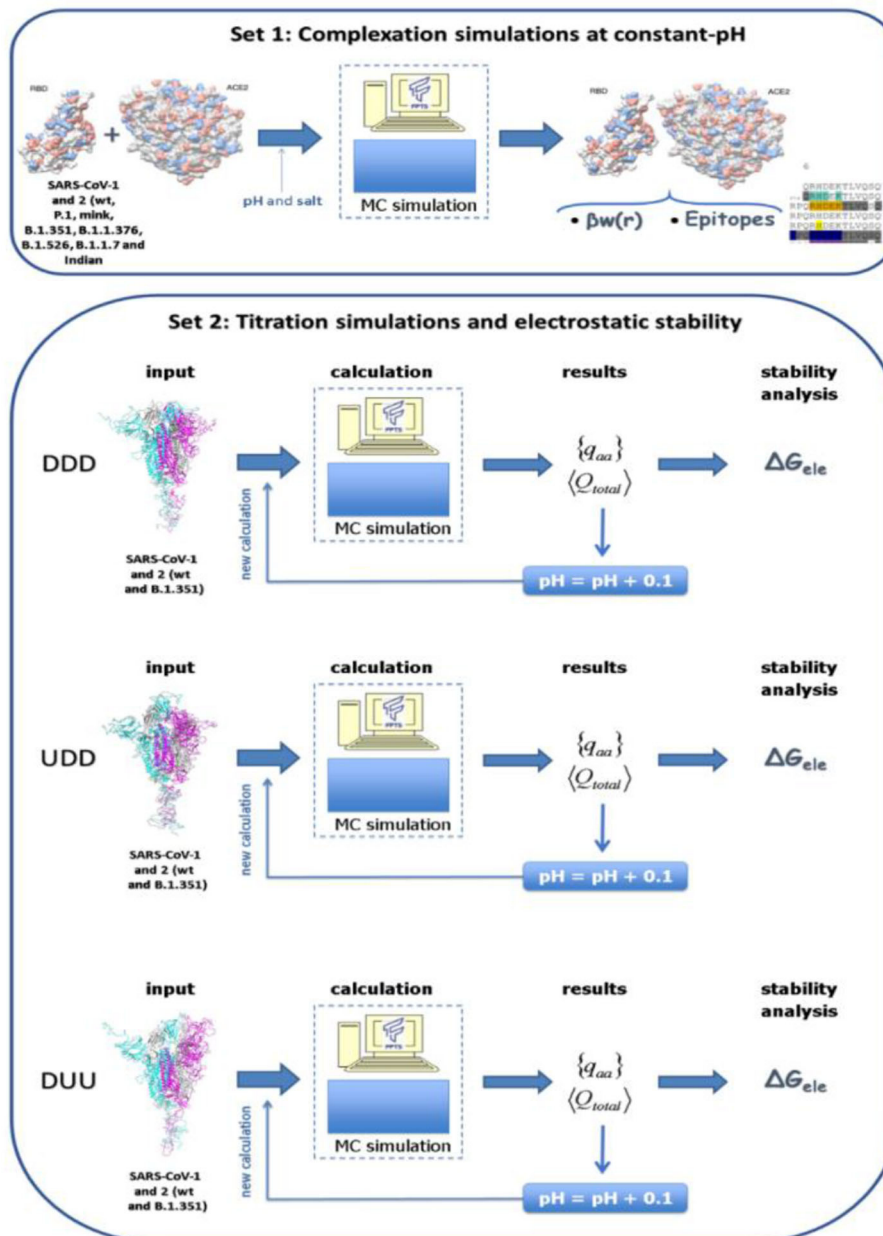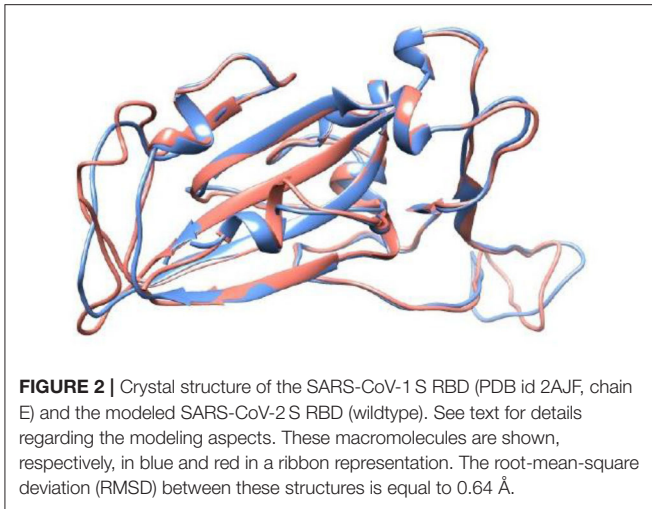
FIGURE 1 | Schematic representation for the two simulation sets of this study. Simulation set 1 represents complexation simulations at constant-pH where the interaction between the S receptor-binding domain (RBD) protein of SARS-CoV-1, 2 and the the new variants (B.1.351, P.1, and the mink variant) with the human receptor angiotensin-converting enzyme II (ACE2) were investigated. Simulation set 2 represents simulations used to estimate the electrostatic stability of whole spike homotrimeric proteins at different conformational states (DDD, UDD, and DUU, respectively) for all solution pHs. Spike wt proteins from SARS-CoV-1, 2 and its B.1.351 (Beta WHO new label) were investigated.

with the highest probabilities were selected for each case. The minimization was performed with default parameters also considering the H-bonds. Replaced amino acids are a) N501Y, K417N, and E484K (20, 46, 73, 74) for the South African (SA) variant (RBD2$_{SA}$), b) K417T, E484K, N501Y (75–78) for the Brazilian (BR) P.1 (RBD2$_{BR}$), c) Y453F (79) for the "mink" (RBD2$_m$) strain, d) N501Y (41, 75) for the UK B.1.1.7 variant (RBD2$_{UK}$), e) E484K and N501Y (80) for the New York (NY)

B.1.526 variant (RBD2$_{NY}$), and f) L452R (81, 82) for the Californian (CA) B.1.427/B.1.429 (RBD2$_{Ca}$) strain. Note that both the SA/B.1.351 (or Beta), the Brazilian P.1 (Gamma), and the NY/B.1.526 (Iota) variants share two common mutations (E484K and N501Y). The third mutation at the RBD for the BR and SA variants is absent in the NY variant and occurs at K417 that is replaced by N and T, respectively, for the BR and SA variants, and does not alter the electrostatic properties of

FIGURE 2 | Crystal structure of the SARS-CoV-1 S RBD (PDB id 2AJF, chain E) and the modeled SARS-CoV-2 S RBD (wildtype). See text for details regarding the modeling aspects. These macromolecules are shown, respectively, in blue and red in a ribbon representation. The root-mean-square deviation (RMSD) between these structures is equal to 0.64 Å.

the RBD. Both ASN and THR have the same physical–chemical characteristics being polar with uncharged side chains, indicating that these mutations are a natural adaptation that helps the virus. The E484 was also quite recently found in a double mutation (E484Q and L452R) in India (RBD2$_I$).

(d) The SARS-CoV-2 S homotrimer wt protein (Strimer2$_{wt}$): Structural data are now abundant for the spike proteins (or their parts), providing us with a rich level of information never seen before. To date, 151 experimental structures are available at the PDB with a diversity of resolution, conformational states, and bound forms (apo and holo with different partners). They were solved by either x-ray crystallography or CryoEM. A good compilation of these available structures can be found at "Universal Protein Resource" (UniProt) under the id P0DTC2 [83]. Several computer simulations were also performed, expanding even more information extracted from these experimental structures [84–86]. Despite this amazing work in such a short time, new variants are also surging in a highly competitive time frame, and key physical chemistry parameters as the solution pH have not been explored in full detail yet. From this ample source of available structural coordinates, we decided to use the configurations extracted from the MD trajectories generated at Poma's lab [85] as input structures for our present calculations. This research group has characterized the structural and energetic differences between the DDD, DUU, and UDD conformations of the SARS-CoV-2 spike wt trimer at pH 7. The advantages of this choice are: i) standardization of all physical–chemical conditions used to obtain the structures at the three conformational states (different experimental structures were solved at distinct conditions) in the absence of any additional chemical (i.e., in a genuine electrolyte solution), ii) all structures from these MD trajectories were complete (no missing amino acids) and already minimized while the experimental structures have missing residues (in fact, the CryoEM structures contain several missing amino acids particularly at the RBD), iii) this set of configurations includes thermal, structural fluctuations, iv) configurations extracted from

different simulation replicas allow us to estimate the standard-deviations (SDs) in our measurements. All these issues avoid the introduction of additional artifacts in the calculations. We also performed simulations with the experimental structure given by the PDB id 7A94 (CryoEM SARS-CoV-2 Spike homotrimer wt with one ACE2 bound, resolution 3.9 Å, pH 8). This configuration (at the UDD conformational), the most populated conformational state experimentally observed among the bound ones [48], allows us to test an experimental structure and see how the binding of ACE2 changes the stability of the complex S-ACE2.

(e) The SARS-CoV-2 S homotrimer SA variant (Strimer2$_{SA}$): By assuming valid the robust sequence-structure relationship to preserve the overall fold of the molecules upon point mutations (as mentioned above), the coordinates for all studied variants were modeled as described above for item (c) by the simple replacement of amino acids from the SARS-CoV-2 S wt trimer structures at the corresponding three different conformational states. Replaced amino acids are for the SA B.1.351 variants, D614G, N501Y, K417N, E484K, L18F, D80A, D215G, R246I, and A701V [20, 46, 73, 74].

(f) The SARS-CoV-1 S homotrimer wt protein (Strimer1wt): As done for the SARS-CoV-2 B.1.351 variant, deletions and amino acids substitutions were performed in the SARS-CoV-2 wt homotrimer following the sequence of the SARS-CoV-1 S protein as given by Uniprot id P59594 [83]. This introduces an additional approximation in outcomes whose effect is assumed to be small due to the high identity (at the sequence level) between them and the strong sequence-structure relationship mentioned above. Additional calculations were performed with the experimental structures given by PDB ids 6ACC (DDD conformational state) and 6ACD (UDD conformational state).

As indicated above between parenthesis, we shall refer to these systems (a)–(f) as RBD1wt, RBD2wt (and RBD2wt'), RBD2$_{variant}$ (variant = SA, BR, CA, NY, UK, I or m), Strimer2$_{wt}$, Strimer2$_{SA}$, and Strimer1$_{wt}$, respectively. Calculations performed with a different input structure will have their PDB ids included as a subscript (e.g., Strimer1$_{wt(6ACC)}$ for the SARS-CoV-1 spike wt protein using the PDB id 6ACC). The human receptor ACE2 needed for all simulations of set 1 was obtained from the PDB id 2AJF (chain A) as demonstrated in a previous study [43].

All PDB files were edited before the calculations. Missing regions in these proteins were built up using the "UCSF Chimera 1.14" interface [72] of the program "Modeler" with default parameters [87]. Water molecules and hetero atoms were completely removed from all used files. Glycosylation sites were not included (they are also often truncated in the experiments) due to the incompatibility of highly flexible molecules with a rigid protein model and all the uncertainties and arbitrariness involved in it [88]. Cysteines involved in sulfur bridges were fixed with GLYCAM web tools [89]. The "UCSF Chimera 1.14" package [72] was employed for all molecular visualizations and representations. Some images were generated by CoV3D [90]. When appropriate, it is indicated in the captions of the figures. For some analysis, it was necessary to determine structural interfaces. This was done with the online server "PDBePISa" [91] with default options.

The linear sequences of the SARS-CoV-1 and 2 S proteins are available in UniProt with the ids P59594 and P0DTC2 for SARS-CoV-1 and 2 (wt and the B.1.351 variant), respectively. The S1 subunit that binds the virion to the cell membrane receptor is the cleaved chain between residues 14 and 667 for SARS-CoV-1, and between residues13 and 685 for SARS-CoV-2. To our knowledge, there is no information yet of if the mutations have affected the cleaved region for SARS-CoV-2. Alignments of pairwise sequences were obtained by the EMBOSS Needle server (92) with default settings. They are shown in **Supplementary Figures 1–3**. The identity and similarity between SARS-CoV-1 and 2 wt are 64.2 and 78.6%, respectively. The RBD corresponds to positions 306–527 and 319–541 for SARS-CoV-1 and 2, respectively. Both identity (I) and similarity (S) are higher for this specific structural region ($I = 73.1\%$ and $S = 82.1\%$). On the other hand, the identity and similarity between SARS-CoV-1 and the B.1.351 variant are 68.2 and 75.5%, respectively. The identity and similarity between SARS-CoV-2 wt and the B.1.351 variant are 98.4 and 99.0%, respectively. This comparison shows that the B.1.351 variant of SARS-CoV-2 has an identity for the RBD slightly closer to SARS-CoV-1 than SARS-CoV-2 wt (68.2% against 64.2%).

Molecular Simulations

Despite the wide availability of molecular simulation models at different scales (64, 93–96), the so-called coarse-grained (CG) models are the most cost-effective ones, due to (i) the large number of atoms (implicitly) involved in each of the studied systems (e.g. the SARS-CoV-2 S wt homotrimer has 3,363 amino acids), (ii) the electrostatic coupling between a large number of titratable groups in these macromolecular structures (e.g., 246 groups per chain of the SARS-CoV-2 wt homotrimer), (iii) the need to repeat the calculations both at several different physical–chemical conditions (140 different pH conditions per system), for different protein conformations (e.g., 10 input structures for each conformational state of the homotrimer) and several viral protein systems (more than 7 systems as described below, see item 2.1), (iv) the estimation of the free energy of interactions based on a histogram method where the statistics on each histogram bin requires longer simulation runs for a proper sampling, and (v) the number of simulation replicates to guarantee their numerical convergence. These simplified computer models allow a lower computational cost to explore the main physical characteristics of a system with a small number of parameters (54, 97, 98). Though it may not be so explicit, all these calculations still require extensive computational resources.

A simple CG for protein–protein interactions has been devised using a rigid body description of the macromolecules and successfully applied to study several different biomolecular systems (54, 97–100). The main feature of this model is the inclusion of a fast and accurate description of the pH effects using the fast proton titration scheme—FPTS (66, 67, 96, 101). Ideally, the coupling between the proton equilibria and conformational changes should be described by a constant-pH molecular dynamics (CpH MD) scheme. Nevertheless, convergence, especially of the electrostatic properties, is still a critical issue of such methods for a single and small protein (102–105). The CPU costs are prohibitive to study all the systems and

conditions mentioned above, even for fast CpH MD approaches as our OPEP6 (105). Conversely, an interesting observation is that key dynamical properties significantly relate to electrostatic property variations, as recently demonstrated for flaviviruses (42). Therefore, we followed the constant-pH MC (CpH MC) strategy as used before for host-pathogen interactions (43, 59). In such a rigid model, internal degrees of freedom are not considered, which implies that our electrostatic model cannot assess the transition features of any reorganized structures of dynamic loops. For radial averaged properties, such as free energy of interactions, it has virtually no significant effect. Curiously, no significant conformational changes induced by pH have been seen in the spike protein in the experimental research carried out by Song and co-authors (49). Other models are also available in the literature. Each one has its characteristics, pros, and cons. For instance, Yu et al. are currently developing a SARS-CoV-2 virion CG model (solved by MD) that provides a full description of the virus due to its ability to adapt and incorporate new details of the molecules as they are released. However, pH effects are not fully included in their model, while it is done in our present work.

For the present CpH CG model, each group of atoms defining an amino acid is converted into a single charged Lennard-Jones (LJ) sphere of valence z_i (a function of the pH of the solution) and radius (R_i)—the values for each class of amino acid is taken from the study of Persson et al. (99). The centers of mass of the spheres created are used to arrange them accordingly with their experimental three-dimensional structures (as specified above). In order to obtain the valences and allow the variation of the amino acids depending on the pH during the simulation, FPTS was employed, whose physicochemical basis and explanation can be found in previous publications (66, 67, 101, 102).

For all simulations of set 1 (i.e., for the complexation RBD-ACE2—see **Figure 1**), two proteins (here, the RBD and the ACE2) are placed in an open cylinder simulation box to allow for forward and backward translations in one axis combined with rotational movements in any direction. As for the simulation data, the static dielectric constant of the medium (ε_s) was 78.7 to mimic an aqueous solution (assuming a temperature of 298 K), the radius (r_{cyl}) of the simulation cell used was 150 Å, and height (l_{cyl}) was 200 Å. Furthermore, the salt particles and the added counter-ions were represented using an electrostatic screening term as follows: for two ionizable amino acids i and j, the screening is given by [$\exp(-\kappa r_{ij})$], where κ is the modified inverse Debye length, and r_{ij} is the distance between particles (54, 97, 98, 102). For simulations of set 2 (i.e., the stability of different homotrimers), only one macromolecule was included in the simulation cell.

The electrostatic interactions [$u_{el}(r_{ij})$] between any two ionizable amino acids of valences z_i and z_j are given by

$$u_{el} = \frac{z_i z_j e^2}{4\pi \epsilon_0 \epsilon r_{ij}} exp\left(-\kappa r_{ij}\right), \tag{1}$$

where e (the elementary charge) is $1.602 \times 10-10$ C and ε_0 is the dielectric constant of the vacuum ($\varepsilon_0 = 8.854 \times 10^{-12}$ C^2/Nm^2). Except for the ionizable amino acids charges—which were defined by the FPTS (66, 67)—all the others were fixed

neutral and kept constant during all simulation runs. Further details can be found on Poveda-Cuevas et al. (54), Barroso da Silva et al. (97), Delboni and Barroso da Silva (98), and Mendonça et al. (100).

Hydrophobic effect, van der Waals interactions, and excluded volume repulsion can also affect protein-protein interactions (54, 98, 99, 106). A simple way to at least incorporate the main contributions of these interactions is using an LJ term [$u_{vdw}(r_{ij})$] between the amino acids (54, 98, 101). For any two amino acids (either charged or not), the calculation for the LJ term is given by

$$u_{vdw} = 4\varepsilon_{LJ}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \qquad (2)$$

where ε_{LJ} regulates the intensity of the attractive forces in the system (54, 97, 98) and σ_{ij} ($= R_i + R_j$)/2 is the LJ diameter for a pair of residues, represented by i and j, when they are in contact. The choice of ε_{LJ} is somehow arbitrary, although it has been used in many works a universal value of .124 kJ/mol (97, 98, 101, 107), corresponding to a Hamaker constant of ca. 9k_BT (where k_B = 1,380 × 10^{-23} m^2 kg s^{-2}K^{-1} is the Boltzmann constant, and T the temperature, in Kelvin) for amino acid pairs (98, 99, 108). The direct impact is that the calculated magnitude of the free energy of interactions can be either under or overestimated, as we discussed before (43). For instance, the absolute numbers might be shifted when compared with results obtained using other force field descriptions. In principle, when experimental second virial coefficients are known for the same physical–chemical conditions and system, proper calibration of ε_{LJ} can be obtained for this very specific situation. This is not the case for these spike proteins. Alternatively, results can be interpreted in relative terms, comparing the measurements between similar systems and conditions. This solves the possible problem of ambiguity in the interpretation of the obtained data.

The size of the amino acid beads also affects the LJ contributions. For the sake of consistency with the adopted ε_{LJ} value, all R_i's were taken from Persson et al. (99). Each amino acid has a specific value of R_i (e.g., R_{TYR} = 4.1 Å, R_{GLU} = 3.8 Å. For this pair, 2σ_{ij} = R_{TYR}+R_{GLU} = 7.9 Å) which allows the description of mostly macromolecular hydrophobic moments (109). Therefore, the simulations should correctly generate the docking orientation at short separation distances (54).

By bringing equations 1 and 2 together, one can obtain the total interaction energy of the system (whether charged or neutral) for a given configuration [$U(\{r_k\})$]:

$$U(\{r_k\}) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(u^{el}\left(r_{ij}\right) + u^{vdw}\left(r_{ij}\right)\right), \qquad (3)$$

where $\{r_k\}$ is the position of the amino acids and N is the total number.

The results were obtained using the Metropolis MC sampling performed with physiological ionic strength (NaCl at 150 mM) at different pH regimes. Calculations were performed with the Faunus biomolecular simulation package (110), where the FPTS is implemented (66). For the complexation study described in the

simulation set 1, following a previous study (43), the pHs 7.0 and 4.6 were chosen, respectively, due to the need to understand the behavior of the system in physiologic pH level conditions and the acidic pH of the endosomal environment. This is a reasonable choice to keep the general features of the present study even though the precise value of the pH in human cells can be slightly different from these numbers (111). For all simulations from set 2, pH was varied from 0 to 14 with an increment of 0.1 units of pH to explore all possible pH conditions. This pH range was also used for the epitopes mapping by the PROCEEDpKa method (54)—see below. As indicated by **Figure 1**, the main outcomes of these CpH simulations were the averaged total protein charge (<Q_{total}>), the averaged partial charge of each titratable group ({<q_{aa}>}), and the averaged protein dipole moment (< μ >). These quantities are also conveniently expressed in units of the elementary charge: the averaged total protein charge number (Z = <Q_{total}>/e), the averaged valence of each titratable group (z_{aa} = <q_{aa}>/e), and the averaged protein dipole number moment (< μ_o > = $|\sum z_i r_i|$). Free energy of interactions [$\beta w(r)$] were estimated from radial distribution functions [g(r)] between the centers of the proteins [$\beta w(r)$ = –ln g(r), where β = 1/k_BT]. r is the separation distance between the centers.

After preparing the molecular systems for the simulations as described above, equilibration and production runs were performed. Even invoking all the approximations in the CG model, these simulation runs demanded high computational resources due to (i) the large number of titratable groups involved in the system with strong electrostatic coupling; (ii) the free energy barriers of the systems; (iii) the need to fill all histogram bins used for the calculation of g(r) during sampling; and (iv) longer runs to the decrease statistical noises in the $\beta w(r)$ data (54, 97, 98). All simulations from set 1 required at least 3.0 10^9 MC steps at the production phase. Simulations from set 2 could be well performed with 10^8 MC steps at the production phase. SDs were estimated by the use of at least three replicates per simulated system. Some systems were further explored with additional replicates.

Electrostatic Epitopes Determined by the PROCEEDpKa Method

Electrostatic properties are well known to be of great importance in biomolecular interactions. Indeed, they strongly depend on the spatial distribution of intramolecular and intermolecular charges, environmental conditions, such as pH or salt concentration that can vary significantly in different cellular compartments (54, 67). Due to its intrinsic long-range characteristics and the electrostatic coupling between ionizable residues, key amino acids responsible for the host-pathogen interaction can include groups outside the classical view of the epitope–paratope interface. Such broader definition, including inner titratable residues that can also take part in the interplay of interactions, has been called "electrostatic epitopes" (EEs). They can be efficiently mapped by a computational strategy called "PROCEEDpKa" (PRediction Of electrostatic Epitopes basedED on *pKa* shifts) (54). This allows the identification of all ionizable residues of macromolecules that do drive the biomolecular interactions. The difference between the numbers of classical (based on the "key and lock" view) and

electrostatic epitopes will be higher for systems with a stronger electrostatic coupling between superficial ionizable residues with the for a given solution pH, protein system and conformational stateinner ones. Among other advantages [see Poveda-Cuevas et al. (54)], "PROCEEDpKa" includes the pH and ionic strength dependence that can dramatically affect the complexation process of the host–pathogen interactions. This is particularly important for the antibody–antigen interface with a peculiar electrostatic pattern richer in titratable amino acids (54).

Electrostatic Stability

The electrostatic stability of the different spike homotrimers (Strimer1$_{wt}$, Strimer2$_{wt}$, and Strimer2$_{SA}$) at three distinct conformational states (DDD, UDD, and DUU) as a function of solution pH were estimated using the electrostatic free energy (ΔG_{elec}). This physical quantity was calculated in terms of Coulombic contributions from the individual titratable groups for a given protein structure in a specific conformation and physical–chemical conditions (68). As proposed by Ibarra-Molero and coauthors (112), ΔG_{elec} (in kJ/mol) can be given as

$$G_{elec} \approx -\frac{1}{2}\frac{1389}{78.7}\frac{z_i z_j}{r_{ij}} exp\left(\frac{-r_{ij}}{\kappa}\right), \qquad (4)$$

where r_{ij} is the separation distance between the ionizable sites i and j, as defined by the spike homotrimer conformation, all charge numbers, z_k, are the averaged ones obtained from the titration studies, i.e., $z_k = <q_{aa}>/e$ for a given pH, protein system and conformational state. All $z\kappa$ values as a function of the solution pH were obtained from the FPTS calculations. κ was fixed at 7.86 Å to correspond to an electrolyte solution at 1:1 ratio with 150 mM NaCl.
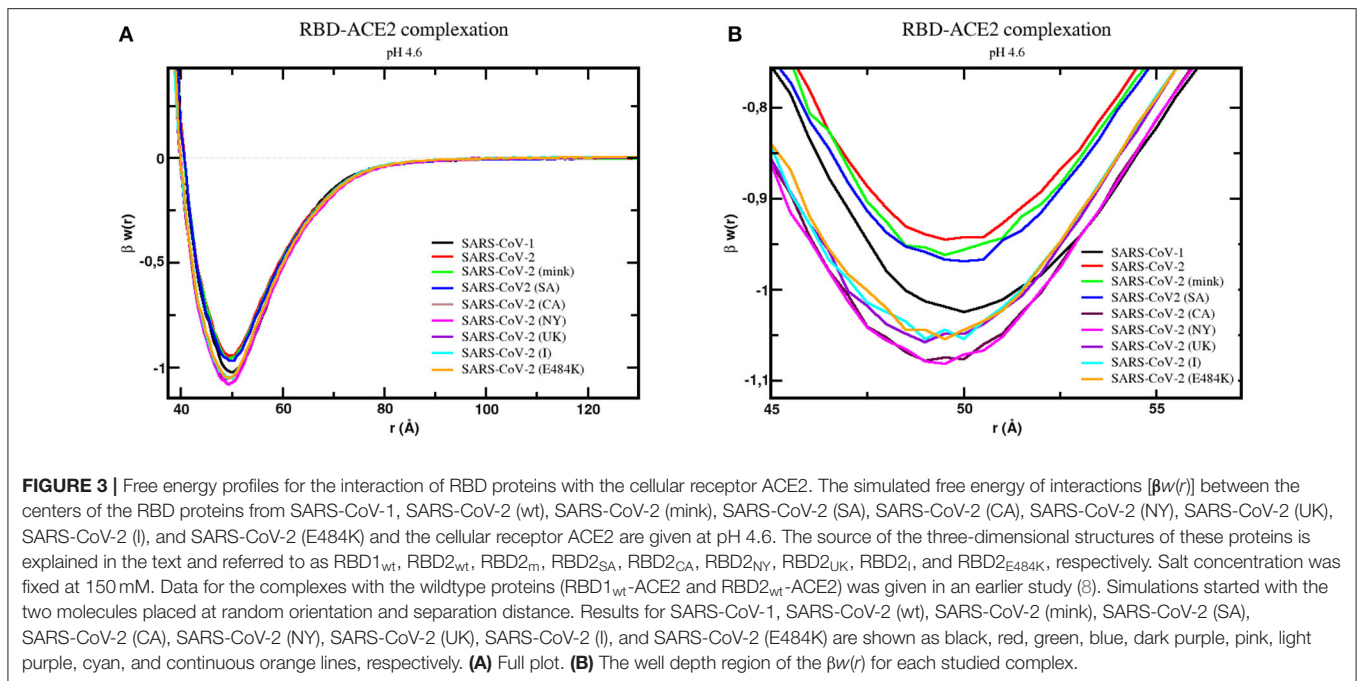
RESULTS

Free Energy of Interactions of SARS Spike RBD Proteins and ACE2

An essential step in cell invasion by a virus is the interaction with a cell receptor. The receptor used by SARS-CoV-1 has been known since 2003: the angiotensin-converting enzyme II (ACE2) (113). Because of the great similarity between the viral proteins, the same human receptor was promptly proposed to be used by the virus responsible for the COVID-19 pandemic too (114). This has been repeatedly confirmed by different studies (6, 38, 43, 114, 115). Modifications in this molecular region (which can be visualized in **Supplementary Figure 4**), either by mutations in the Spike protein or by alterations in the receptor itself, can change the free energy of the interaction and result in higher or lower affinity. This is easily seen for some mutations. For example, the mutation E484K presented at least in three new variants (B1.1.351, P.1, and B.1.526) has an acid residue (GLU) replaced by a basic one (LYS). This substitution changes the physical–chemical nature of the residue 484 and implies a complete inversion of its electrostatic interactions. GLU has its electrical charge (in elementary units) varying from −1 (when fully deprotonated) to 0 (when fully protonated), while LYS has a zero charge number when deprotonated and is positively charged

(+1) when protonated. All neighboring ionizable groups might also be affected by this inversion. This kind of evolutive viral signature suggests that the virus uses electrostatic properties to increase its virulence. When comparing the main simulated physical-chemical quantities of the RBD proteins, we can see the effect of the substitutions of the amino acids both at the protein charge number level (varying from +2.1 to +4.1, see **Supplementary Table 1**) and the dipole number moment level (varying from 31 to 90, see **Supplementary Table 1**). These results given between parenthesis were obtained by the FPTS at pH 7.0 and are summarized in **Supplementary Table 1** for all studied variants. Data for pH 4.6 is included in this table too. Being the ACE2 receptor negatively charged at pH 7, the attractive charge–charge interaction will be stronger for most variants favoring the RBD-ACE2 complexation. The differences in the dipole moments reveal that the binding orientation can also be altered for some systems.

This preliminary and simplest physical–chemical analysis was investigated with more quantitative details of the complexation process. As a continuation of a previous study on the molecular interactions of SARS-CoV-1 and 2 wt S RBDs done at the beginning of the pandemic (43), we investigated in the present work the binding association of the S RBD proteins new variants to ACE2. The RBD was assumed to be exposed and ready for the docking phase [i.e., the RBDs were out of the homotrimeric S protein, assuming that the missed parts of the whole trimer do not interfere with the binding as suggested by the available crystallographic data (43)]. The mink, SA (B.1.351), the BR (P.1), the UK (B.1.1.7), CA (B.1.427/B.1.429), the NY (B.1.526), the India (B.1.617) RBDs complexations with ACE2 were studied using our CpH MC protein–protein simulations at pH 4.6 and 7.0. The wild type with a single E484K mutation was also included in this study. Free energy of interactions [$\beta w(r)$] calculated using the potential of mean forces sampled during the simulation runs for these systems are shown in **Figure 3** at physiological salt concentration. The estimated maximum SDs obtained by comparing results from at least three replicates are .01 for $\beta w(r)$.

Simulations confirmed the binding of all RBDs to ACE2, as can be seen by the negative values of $\beta w(r)$ around 50Å. In comparison with the SARS-CoV-2 S RBD wt protein, the SARS-CoV-1 S RBD protein has the strongest tendency to bind to ACE2 in our simulations, as it was previously reported in some theoretical and experimental studies (43, 116). However, this is not a complete consensus in the literature (6, 116–120). In one of the first experimental comparisons, Tian and co-workers provided quite similar binding affinities (K_D = 15.2 nM for SARS-CoV-1 S RBD and K_D = 15.0 nM for SARS-CoV-2 S RBD, where K_D is the equilibrium dissociation constant), showing that SARS-CoV-2 S RBD-ACE2 had a slightly stronger affinity (117). Conversely, Brielle and others compiled a set of experimental measurements suggesting the higher binding affinity of SARS-CoV-1 S RBD–ACE2 (KD ∼1.5–10.0 nM) (K_D ∼1.5–10.0 nM) comparable to the binding affinity of SARS-CoV-2 S RBD–CACE2 (K_D = ∼1.2–14.7 nM)—see Brielle et al. (116). This discrepancy can have different sources (e.g., use of other structural coordinates, molecular dynamical runs not long enough to sample larger structural fluctuations,

FIGURE 3 | Free energy profiles for the interaction of RBD proteins with the cellular receptor ACE2. The simulated free energy of interactions [$\beta w(r)$] between the centers of the RBD proteins from SARS-CoV-1, SARS-CoV-2 (wt), SARS-CoV-2 (mink), SARS-CoV-2 (SA), SARS-CoV-2 (CA), SARS-CoV-2 (NY), SARS-CoV-2 (UK), SARS-CoV-2 (I), and SARS-CoV-2 (E484K) and the cellular receptor ACE2 are given at pH 4.6. The source of the three-dimensional structures of these proteins is explained in the text and referred to as RBD1$_{wt}$, RBD2$_{wt}$, RBD2$_m$, RBD2$_{SA}$, RBD2$_{CA}$, RBD2$_{NY}$, RBD2$_{UK}$, RBD2$_I$, and RBD2$_{E484K}$, respectively. Salt concentration was fixed at 150 mM. Data for the complexes with the wildtype proteins (RBD1$_{wt}$-ACE2 and RBD2$_{wt}$-ACE2) was given in an earlier study (8). Simulations started with the two molecules placed at random orientation and separation distance. Results for SARS-CoV-1, SARS-CoV-2 (wt), SARS-CoV-2 (mink), SARS-CoV-2 (SA), SARS-CoV-2 (CA), SARS-CoV-2 (NY), SARS-CoV-2 (UK), SARS-CoV-2 (I), and SARS-CoV-2 (E484K) are shown as black, red, green, blue, dark purple, pink, light purple, cyan, and continuous orange lines, respectively. **(A)** Full plot. **(B)** The well depth region of the $\beta w(r)$ for each studied complex.

glycosylation, presence of other monomers of the ACE2, use of the homotrimer instead of a single RBD, constant-charge vs. constant-pH simulations, different experimental assays, other physical–chemical conditions, etc.) (118). Calculations performed with RBD coordinates extracted from CryoEM structures will suffer from the lack of coordinates for this critically important region. For instance, the prefusion S homotrimer with a single RBD "up" as given by the PDB id 6SVB (chain A) has several missing amino acids at the RBD (INT….KVGGN….LFRKSNLKPFERDISTEIYQAGSTPC NGVEGFNCYF….NG….).

Although these missing amino acids can be fully reconstructed by homology modeling, it becomes difficult to precisely identify if the final model provides more reliable coordinates for the complexation studies than the one built-up using the SARS-CoV-1 RBD as the template (with fewer unknown coordinates). Testing RBDs extracted from the CryoEM structures (completed with homology modeling) can result in binding affinities equivalent to what was measured for SARS-CoV-1 or slightly superior—see **Supplementary Figure 5**. As expected and reported before (97, 102), there is a dependence on the used structural coordinates for the protein–protein calculations. A different set of coordinates for ACE2 could also affect the possible structural rearrangements during the binding. Constant-charge MD simulations would not be a definitive answer, as indicated by a previous study that also predicted a higher affinity of SARS-CoV-1 RBD–ACE2 binding (116). An ideal solution would be applying CpH MC simulations to all variants but this approach is still nonfeasible at present. Instead, we opted to use as the input structure the RBD modeled as before (43) and focus on the aspects given by the effects of pH and the different substitutions of the amino acids, which is the main aim

of the present study. It is a consensus from different studies that electrostatic interactions drive the RBD-ACE2 complexation in both cases (43, 118), which means that CpH models (as used in this work), even with other approximations, should better capture the main physics of the system.

Moreover, despite the good precision in the protein–protein simulations [0.01 units of $\beta w(r)$ as mentioned above], the differences between the potential depths for each protein sequence are small. Considering all the intrinsic approximations assumed in this study (including possible fluctuations due to structural dynamics), we think it is safer to conclude that these outcomes indicate tendencies given by the differences in the linear sequences. Therefore, this analysis showed a slight tendency toward a higher affinity for ACE2 by all studied new variants. The highest affinity was found for the RBD2$_{NY}$. In a crescent order, we have RBD2$_{wt}$ < RBD2$_{mink}$ < RBD2$_{SA}$ and RBD2$_{BR}$ < RBD2$_I$ < RBD2$_{UK}$ < RBD2$_{CA}$ < RBD2$_{NY}$. The BR variant P.1 (data not included in **Figure 3** because it was on the top of the plot of the SA case) behaves identically to the SA variant (B.1.351) due to the presence of the same key mutations in both of them—E484K and N501Y—at the receptor-binding motif. The mutations K417N/T do not make any difference for these two variants in terms of their binding affinities, at least captured by our CpH CG model. These substitutions on K417 might work to inhibit the ACE2-affinity (46). The effect of the absence of the mutation K417 in the NY variant can also be seen. This enhances the RBD-ACE2 affinity, especially when compared with the two other variants (SA and BR) that share the same substitutions but have either K417N or K417T to decrease it. The key mutations at the RBD region for the different variants included in this study are illustrated in **Supplementary Figure 7**. The increased affinity between RBD2$_{UK}$ and ACE2 has recently been seen in

TABLE 1 | Estimated binding affinities for RBD–ACE2 interactions.

RBD structure	Mutation(s)	Predicted binding tendencies (pH 4.6)	Experimental relative binding affinity for each single mutation[*]
RBD2$_{mink}$	Y453F	0.01	0.25
RBD2$_{SA}$	K417N	0.02	−0.45
	E484K		0.06
	N501Y		0.24
RBD2$_{BR}$	K417T	0.02	−0.26
	E484K		0.06
	N501Y		0.24
RBD2$_{CA}$	L452R	0.12	0.02
RBD2$_{NY}$	E484K	0.13	0.06
	N501Y		0.24
RBD2$_{UK}$	N501Y	0.11	0.24
RBD2$_I$	L452R	0.10	0.02
	E484Q		0.03
RBD2$_{E484K}$	E484K	0.10	0.06

The theoretical predicted binding tendencies were calculated taken the differences between the minima for βw(r) of the wild type with the simulated value for a given variant at pH 4.6. [*] Data from Starr et al. (46). See text for other details.

another computational approach (121). This result also revealed that the binding affinity of these new variants is approaching and enhancing the affinity measured for SARS-CoV-1 that was more virulent.

Most viruses use a drop in pH to trigger their host penetration (122), including some coronaviruses (123). For SARS-CoV-2, as we mentioned above, the pH effects are somehow contradictory. To further study the effects of pH in the RBDs-ACE2 complexation, calculations were also performed at pH 7 for all the variants. Data for the SARS-CoV-2 wt was obtained before in a previous study (43). As shown in **Supplementary Figures 6A,B**, the increased pH slightly favors the attraction for all of them. The same relative behavior between the strains seen at pH 4.6 is essentially reproduced at the physiological pH conditions. In the two studied pH regimes, the NY variant has the strongest binding affinity. Recent experimental data in the preprint format confirms the same trend for the wild-type virus (50). Surprisingly, the low pH does not seem essential for the viral cell invasion for the variants. Indeed, it was suggested before that an acid regime did not appear to act as a direct trigger of this entry process for SARS-CoV-1 (124). However, the low pH could be necessary for an additional player in the process, i.e., acidic protease involvement and membrane-fusion activity (125). pH not being so critical for the RBD-ACE2 complexation offers the SARS viruses and their variants an opportunity to easily and rapidly infect the human cells. This characteristic might also contribute to increasing their infectivity, as noted before (43).

Table 1 summarizes the main theoretical results together with experimental data for single mutations. For the sake of clarity, mutations are also listed in this table. Although the experimentally available data was not obtained exactly under the same conditions, it can be seen that they are qualitatively similar to the theoretical data. The presence of more than one mutation

does not seem additive. For this set of studied cases, both N501Y and Y453F are mutations that experimentally resulted in the highest RBD-ACE2 affinities (46). The theoretical predictions confirm that the N501Y mutation increases this affinity. The difference between this strain and the NY and CA variants seen in our calculations is within the estimated errors (0.01). Conversely, the Y453F mutation is equivalent to the wild type in the theoretical predictions when the estimated errors are considered. Despite quantitative discrepancies, it is clear that the new variants tend to be more virulent than SARS-CoV-2 wt due to the molecular properties of their more evolutionary adapted RBDs.

Estimated Antigenic Regions by the PROCEEDpKa Method

In this timely research field that is evolving so fast with new variants frequently appearing at an increased rate, it becomes challenging to catch up with such "storms" of mutations and run calculations for all possible new cases. For this reason, we selected the SA variant (B.1.351 or Beta) to be investigated in more detail during the last phases of the present study. An important issue now is understanding what amino acids of its RBD are the relevant ones for the complexation mechanism. On the one hand, these epitopes can be used to design peptides for vaccines. On the other hand, they contribute to the characterization of the binding modes indicating interesting potential targets for specific therapeutic binders.

To determine the EEs of the S RBD proteins of SARS-CoV-1, SARS-CoV-2 wt, and the SA variant (B.1.351) of SARS-CoV-2 for the ACE2 complex, the PROCEEDpKa method (54) was used. This method uses pKa shifts to identify the key amino acids responsible for a host–pathogen association. It is rooted in the physical–chemical fact that the presence of an electric charge (a fixed charge or the instant charge of another ionizable residue at a given protonation state) can perturb the acid–base equilibrium of a titratable group. Consequently, identifying the pKa shifts is a practical means to probe intermolecular interactions as demonstrated in an earlier study (120). The pKas were measured from the theoretical titrations for the isolated RBDs and during computer simulations of a protein–protein complexation by the CpH MC scheme (simulations set 1).

The main questions to be addressed here are to determine whether the mentioned proteins share a common binding region when interacting with ACE2 and whether there is a change in the number of amino acids involved in it (indicating a more or a less specific association). The obtained EEs during the simulations were mapped at the sequence level to allow a direct comparison. These results are shown in **Figure 4**. In this figure, the primary sequences of the RBDs of SARS-CoV-1, SARS-CoV-2 wt, and 2' (representing the SA variant) are superimposed. Amino acids identified as EE as classified by the PROCEEDpKa method are shown in blue. Although the general patterns observed for these three viral proteins are relatively similar, some differences can be seen in the number of perturbed amino acids and their location. These differences show how the electrostatic theoretical method is sensitive to the different sequences, as seen previously (54).

FIGURE 4 | Electrostatic epitopes. Primary sequences of the SARS-CoV-1 S RBD, the SARS-CoV-2 S wt RBD, and the SARS-CoV-2′ — representing the South African variant (B.1.351)—with the interface with ACE2 (shown in blue). Data obtained using the threshold |ΔpKa| > 0.01. Calculations for the RBD of SARS-CoV-2′ were performed with the structure RBD2$_{SA}$. Gaps are represented by "–." The numbers next to the chains are used to guide the identification of the amino acid sequence numbers of the RBD.

Indeed, it was noted before that SARS-CoV-1 and 2 (wt) are antigenically different (120). The number of ionizable residues involved in the intermolecular interactions between SARS-CoV-1 S RBD and ACE2, SARS-CoV-2 S wt RBD and ACE2, and SARS-CoV-2' (RBD2$_{SA}$) and ACE2 pairs increased from 30 to 40 and from 40 to 43, respectively, with a high number of common EEs. Qualitatively, the theoretically predicted antigenic patterns for SARS-CoV-1 and 2 (wt) are similar to another work published later, reporting 17 and 21 epitopes, respectively (120). Other theoretical methods also support this behavior (116). The quantitative differences are because of the inclusion of more internalized amino acids in the structures that are electrostatically coupled with the superficial ones, and, for this reason, classified as EEs. They can have an important role in the complexation together with the superficial residues (54).

Together with the amplified tendency for stronger complexation (see above), these results suggest that, as the virus SARS-CoV-2 evolved, the binding to the ACE2 receptor, which occurs with an increasing number of involved residues, becomes more specific. It is different from what is seen for the wildtype, where the number of EEs in SARS-CoV-2 was larger than that in SARS-CoV-1, but the affinity was slightly weaker (a kind of "key-loose door lock cylinder" interaction). The rise in specificity contributes to higher virulence. This behavior might also affect how antibodies should block the RBD. If an antibody does not cover all the surface area given these EEs, there is a

chance that the RBD can still come closer to the ACE2 and probably will allow the next steps of the infection to continue. This may reduce vaccine effectiveness.

The "Up" State as a Requirement for Efficient Binding

The spike protein hides several "tricks" *via* the changes in its conformational states. Wrapp et al. (119) revealed the movement of the RBD between the up/open and down/closed conformational states for SARS-CoV-2. This study was followed by many others that used experimental techniques to provide rich structural information regarding the interplay of the conformational transitions of the spike homotrimer (48, 119, 126). At least one chain of the homotrimer has to be in the "up" conformational state to allow the binding of the RBD with the receptor ACE2.

From a structural point of view, it is clear that most of the EEs involved in the RBD-ACE2 complexation should be more exposed when the homotrimer is at the "up" conformational state, as it is not only the steric clashes that are important to be removed for their binding. The key amino acids should be available for effective interaction between these molecules. For this reason, we analyzed the predicted EE for the RBD out of the trimer and mapped them on the three-dimensional macromolecular structures of the trimers. As can be seen in **Table 2** for the SARS-CoV-2 wt RBD, the

TABLE 2 | Location of the epitopes at different interfaces for SARS-CoV-2 wt S homotrimer.

PDB id	6VSB (UDD)			6VXX (DDD)		
Chain	A (*up*)	B (*down*)	C (*down*)	A (*down*)	B (*down*)	C (*down*)
NEE	**32**	28	30	20	24	**25**
NHE	8	12	10	20	16	15

The number of exposed epitopes (NEE) and hidden epitopes (NHE) for interactions with ACE2 for two S homotrimer conformational states, one closed (DDD) and one open (UDD). This analysis used the PDB ids 6VXX and 6VSB, respectively. The state of each protein chain (D or U) is indicated in parentheses, where D corresponds to the closed (or "down") conformational state and U to the open (or "up") conformational state. Bold fonts highlight the highest values.

"down" conformational state (DDD) reduces the number of exposed (electrostatic) epitopes (NEE). The comparison of two experimental structures at different conformational states (DDD and UDD) indicates that there are 32 EEs when one chain is at the "up" conformation (UDD state) while 25 EEs are seen at the other conformational state. Interestingly, there is an increase in the NEE for the chains at the "down" state when at least one is at the "up" conformational state. This analysis was done using the detailed data for the mapping of EEs for the SARS-CoV-2 wt RBD at different conformational states of the spike homotrimer at the amino acid level given in **Supplementary Table 2**.

Electrostatic Stability

Another molecular factor that can influence the transmissivity and virulence of SARS-CoV-2 through phenotypic changes is the increased stability of the spike homotrimers in the open state, exposing the EEs of the RBD and, consequently, allowing the interaction between the RBD and ACE2 (42, 43, 54–56, 58, 59). Viruses often undergo mutations that improve either their binding affinities (as discussed above) or their protein stabilities. An improvement in the affinities can reduce the stability, which drives multiple mutations to keep or even increase virulence. In principle, pH can have an important influence on this process too.

Therefore, we investigated (with set 2 of the simulations) the electrostatic stability of the S trimeric structure of SARS-CoV-1, 2 wt and the SA variant (Beta) as a function of pH for three different conformational states (namely, DDD, DUU, and UDD). The aim was to identify which state has the greatest impact (favoring the upstream confirmation of the Spike glycoprotein), thus providing key information for developing broader spectrum-coping strategies against COVID-19.

Different analyses can be made by exploring the stability of each protein sequence at different conformational states and the comparison between them. Starting with the sequence from SARS-CoV-1, its stability is greater (the values are more negative) over almost the entire extent of the relevant pH regimes when compared with SARS-CoV-2 wt, as seen in **Figure 5**. The maximum estimated standard deviations on ΔG_{elec} for DDD, DUU, and UDD are, respectively, 7, 14, and 10 kJ/mol for all studied pH. They were calculated based on the ten CpH
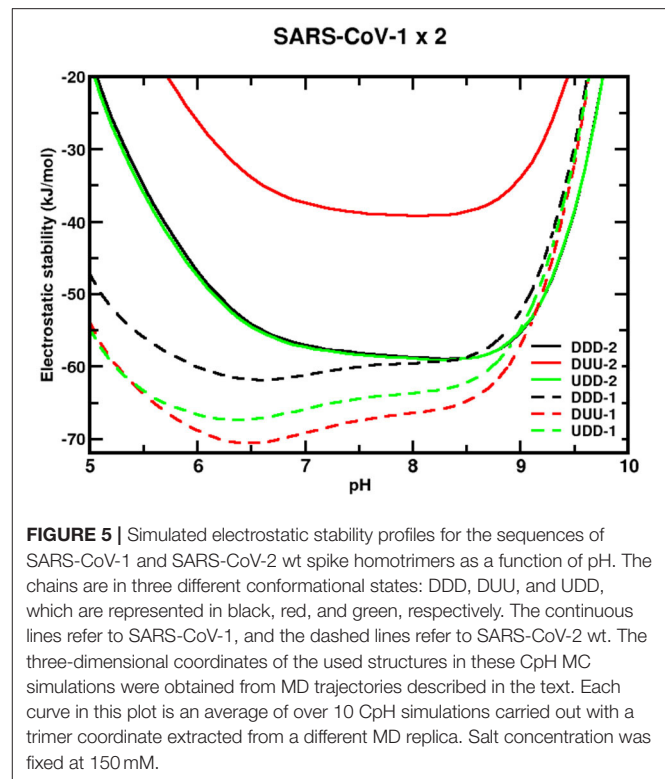

FIGURE 5 | Simulated electrostatic stability profiles for the sequences of SARS-CoV-1 and SARS-CoV-2 wt spike homotrimers as a function of pH. The chains are in three different conformational states: DDD, DUU, and UDD, which are represented in black, red, and green, respectively. The continuous lines refer to SARS-CoV-1, and the dashed lines refer to SARS-CoV-2 wt. The three-dimensional coordinates of the used structures in these CpH MC simulations were obtained from MD trajectories described in the text. Each curve in this plot is an average of over 10 CpH simulations carried out with a trimer coordinate extracted from a different MD replica. Salt concentration was fixed at 150 mM.

simulations carried out with different coordinates of the trimer at the same conformational state, which lets the calculations to include some of the thermal fluctuations of the homotrimer structures. From pH 4.3 to 9.8, ΔG_{elec} is always negative for these three conformational states. The most stable conformational state for this sequence (SARS-CoV-1) is DUU, followed by UDD and DDD. Nevertheless, considering the differences between them with the estimated SDs, it does not allow us to identify the most stable one. For instance, at pH 7, ΔG_{elec} is equal to −61 (4), −69 (8), and −66 (9) kJ/mol, respectively, for DDD, DUU, and UDD. It was somehow unexpected to find that all the three conformational states would have statistically equivalent probabilities. However, there are twice more chances for the virus to be in an open state because two states (DUU and UDD) offer this possibility. It infers that SARS-CoV-1 tends to be more often ready to enter the human cell with at least one chain at the open state when its RBD is available to bind to the receptor ACE2 with great affinity (see discussion above). Also, it is tempting to make a connection between this molecular behavior with clinical observations. A higher probability of having the homotrimer at the "up" position should result in more symptomatic patients.

pH has virtually a minor effect on the stability of this viral sequence for the most important biological regimes (pH ∼4–7). The pH values where these states for SARS-CoV-1 are more stable are 6.6 (−62 kJ/mol), 6.5 (−71 kJ/mol), and 6.4 (−67 kJ/mol), respectively, for DDD, DUU, and UDD. From pH 4–7, the differences in ΔG_{elec} [$\Delta\Delta G_{elec} = \Delta\Delta G_{elec}$(pH 7) − $\Delta\Delta G_{elec}$(pH 4)] for them is −103, −102, and −102 kJ/mol. The transitions from the close state (DDD) to the open states (UDD and DUU)

require around 9 kJ/mol (depending on the final configurational state) at pH 4, whose value has the same order of magnitude as the estimated error. There is another factor that can help to trigger the "chameleon" behavior of the homotrimer. If experimentally solved CryoEM structures for the sequence given by SARS-CoV-1 [PDB ids 6ACC (DDD) and 6ACD (DUU)] are used in the calculations, the difference is relatively larger (19 kJ/mol), favoring the DDD state, although the unique structures for each state do not permit to estimate the errors as done for the MD coordinates ($\Delta G_{elec} = -100$ kJ/mol, for DDD, and $\Delta G_{elec} = -71$ kJ/mol, for DUU).

Conversely, all conformations of the sequence given by the SARS-CoV-2 wt are most stable at a solution pH between 6.5 and 9, with the most unstable conformation being the one with two RBDs in the "up" position (i.e., DUU). The transition from DDD to DUU requires 20 kJ/mol at pH 7. Less is needed at lower pH regimes. The highest stability is observed for pHs 8.3 (-59 kJ/mol), 8.1 (-39 kJ/mol), and 8.3 (-59 kJ/mol) for DDD, DUU, and UDD, respectively. Furthermore, it can be seen that ΔG_{elec} for the whole homotrimer at the closed state (DDD) and ΔG_{elec} with only one RBD (UDD) "up" are almost identical to each other [$\Delta G_{elec(DDD)} = -57 \pm 6$ kJ/mol and $\Delta G_{elec(UDD)} = -57 \pm 9$ kJ/mol at pH 7], showing a tendency for this state transition to be done without any energetic cost. Experimentally, the UDD was the most seen ACE2-bound conformational state (48). It can be hypothesized that this is the molecular reason that facilitates infections. An equal probability for DDD and UDD states as given by this theoretical analysis reduces the number of interactions RBD-ACE2 in comparison to what could be seen for SARS-CoV-1 (∼66% for the two open states). Again, we can extrapolate our data suggesting that this can explain the molecular reasons for such a higher number of transmissions by asymptomatic people as observed for SARS-CoV-2 and longer intervals of contagiousness. In truth, the CryoEM study performed by Benton and others showed that 11% of the total trimeric structures were at the closed state (DDD) and 20% of them at the open state either with one RBD up (16%) or two (4%) (48). Also, most of the bound ACE2 cases in their study were observed for the UDD state (49 vs. 14% for DUU and 3% for UUU).

Moreira and co-workers reported 10.4 and 32.5 kcal/mol for DDD→ UDD and DDD→ DUU transitions, respectively, at pH 7 and 150 mM, using a Poisson–Boltzmann solver (85). No SDs were given. Our data agree qualitatively only with the DDD→ DUU transition. What triggers the conformational changes from one state to another (controlling the viral load in the patient) is not clear from these results. pH is not directly responsible for this transition, but it can still indirectly change the interactions with another co-factor (e.g., glycans). Our present simulations did not address this trigger mechanism. It could be that pH promotes intermediate conformational states which can only be properly described by a full CpH MD simulation, whose computational costs are still prohibitive.

We next analyzed the sequence effects of the SA variant (B.1.351), comparing it with the wildtype of SARS-CoV-2. This comparison revealed a significant improvement in the homotrimer stability for the SA variant, although the stability
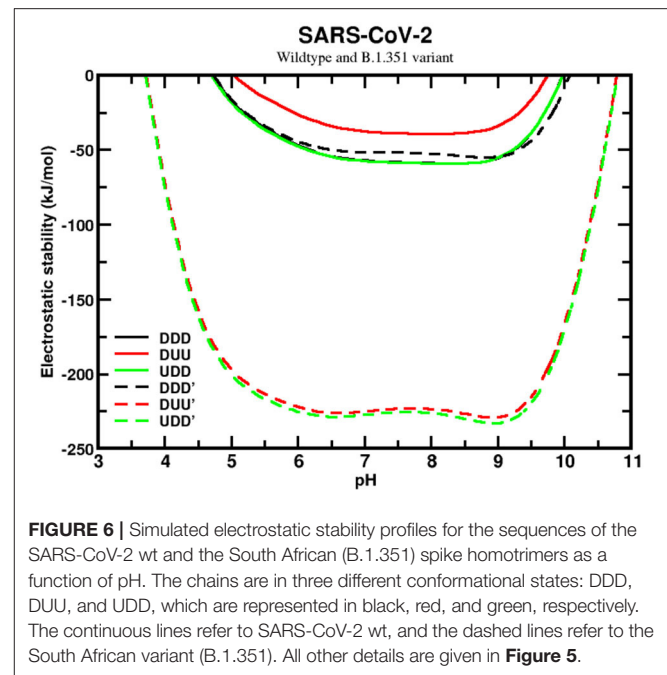

FIGURE 6 | Simulated electrostatic stability profiles for the sequences of the SARS-CoV-2 wt and the South African (B.1.351) spike homotrimers as a function of pH. The chains are in three different conformational states: DDD, DUU, and UDD, which are represented in black, red, and green, respectively. The continuous lines refer to SARS-CoV-2 wt, and the dashed lines refer to the South African variant (B.1.351). All other details are given in **Figure 5**.

curves are still qualitatively similar (see **Figure 6**). Unlike the wildtype, the SA variant presents the closed (DDD) state as the less favorable one, while the UDD and DUU forms have similar values. However, the most relevant information is the high difference of approximately 175 kJ/mol between the most stable curve of the wildtype and this variant. This shows how much more stable the SA variant (B.1.351) is in conformational states that enables it to infect the human cells easily. The mutations D80A, D215G, and R246I are probably the main ones responsible for this higher stability. These substitutions contribute to making the trimer more positively charged. Comparing the net charges of the wildtype with the SA variant, we found that they increased from +4.6 to +10.6, for DDD, from +4.7 to +10.7, for DUU, from +4.6 to +10.6, for UDD, at pH 7.

Based on this data, it appears that evolution has switched off its "chameleon" feature, dismissing the need for the DDD state due to the facility for people being contaminated. The SA variant seems to be always ready to let the RBD be prepared to form a complex, with ACE2 explaining how much more harmful to society it can be. If the hypothesis that an equal probability for DDD and UDD implies more asymptomatic cases for the wildtype (see **Figure 6**), the increased probability here for the two open states of the SA sequence should decrease the percentage of asymptomatic and increase the number of symptomatic ones. As time passes, more epidemiological data might be available to corroborate this hypothesis.

The stability of the trimer with one ACE2-bound was also investigated to assess the contribution that the receptor ACE2 could have on it. Bai and Warshel noticed that the RBD of SARS-CoV-2 had been optimized to bind stronger at distant sites (127). They also questioned whether this stronger binding could be related to conformational changes of the homotrimer. The

present analysis can partially answer it. Here, an experimental CryoEM structure (PDB id 7A94) was used for these calculations. This structure corresponds to the UDD conformational state. **Figure 7** shows that the ΔG_{elec} for the sequence given by SARS-CoV-2 wt complexed with one ACE2 molecule (holo state). In the same figure, ΔG_{elec} for the homotrimer at the *same* conformation but with the ACE2 molecule removed (apo state) is plotted. The trimer confirmation at the apo state is a bit more elongated than the ones used in the previous calculations discussed above with the structures from MD. The RMSD between this trimer structure and one from the MD trajectories is 1.3 Å. This explains why ΔG_{ele} dropped to values much smaller than what was seen above. Indeed, this suggests that approaching the receptor ACE2 modifies the trimer conformation and turns its structure into a more stable one. Again, it is another molecular mechanism to promote the infection. Although not explicitly investigated here, this could also trigger the conformational changes from the DDD to UDD (or DUU). By comparing ΔG_{elec} for the homotrimer at the halo and apo states, the effect of ACE2 in the stability is finally elucidated. As expected, ACE2 increases the stability of the trimer, as can be seen in **Figure 7**, favoring, even more, the complexation and the cell entry itself. These structures are more stable around pH 6.3 pH 6.4 ($\Delta G_{elec} = -254$ kJ/mol) for the trimer and pH 6.2 ($\Delta G_{elec} = -323$ kJ/mol) for the complex. The ACE2-bound form of the trimer favors its stability. At pH 7, the difference between the holo and apo forms is -59 kJ/mol. This value might be underestimated because we are comparing the trimer in an already modified confirmation by the previous presence of ACE2. If the initial state is the UDD obtained in the absence of ACE2 (as the ones generated at the MD trajectory), the difference is increased to -249 kJ/mol. Although the physical–chemical conditions of the two trimer structures (the CryoEM and the MD) are not the same, this difference should reflect at least the order of magnitude of the energetic gain when the trimer goes from a single molecule to a complex with ACE2.

DISCUSSION

Some molecular aspects relevant to understanding the increased transmissivity and virulence of new variants of SARS-CoV-2 are discussed. This study involved complementary pieces of information: (1) the theoretical estimation of the binding affinities between the RBD of the Spike proteins from different mutants with the cellular receptor, ACE2, (2) the mapping of the electrostatic epitopes, and (3) the electrostatic stability of different conformational states of the whole spike homotrimer. In particular, we identified, for the studied strains, the probability of having the trimer at the open state since the interaction with ACE2 only occurs if the epitopes are available for binding. Even though unknown factors may connect these mechanisms, we could identify some key aspects of the natural evolution of this virus.

All studied variants showed a tendency for a higher RBD-ACE2 binding affinity when compared with the wild-type version of SARS-CoV-2. Once the analysis of the complexion between ACE2 and the RBD from the SA variant (B.1.351 or Beta) showed minimal increases in the free energy of interactions (despite an increase in the specificity with more amino acids
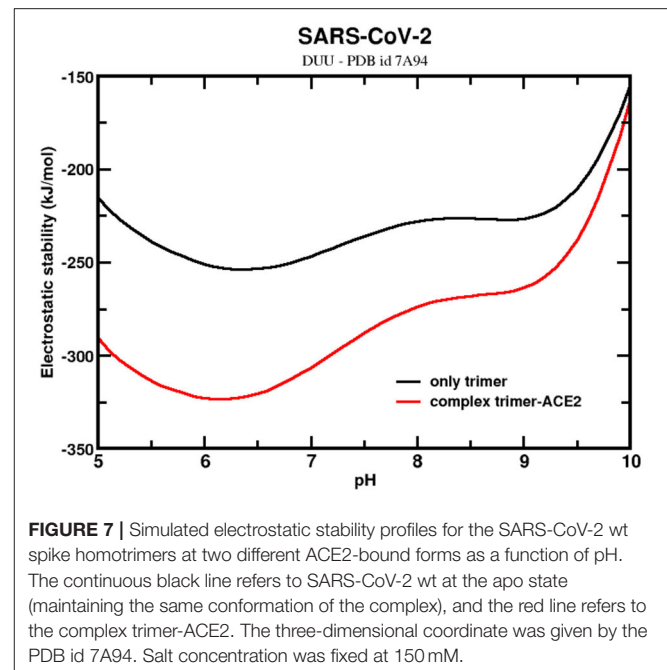

FIGURE 7 | Simulated electrostatic stability profiles for the SARS-CoV-2 wt spike homotrimers at two different ACE2-bound forms as a function of pH. The continuous black line refers to SARS-CoV-2 wt at the apo state (maintaining the same conformation of the complex), and the red line refers to the complex trimer-ACE2. The three-dimensional coordinate was given by the PDB id 7A94. Salt concentration was fixed at 150 mM.

that are important for this process), we investigated another aspect that is often used by viruses upon mutations: the stability. Comparing three conformational states (DDD, UDD, and DUU) for SARS-CoV-1, SARS-CoV-2 wt, and its SA variant, we could see differences in the homotrimer probabilities at these states. From SARS-CoV-1 to SARS-CoV-2 wt, there is a slightly smaller chance for the SARS-CoV-2 trimer to be available for binding with ACE2. The increase in DDD probability was also observed and interpreted as a possible explanation for a higher number of asymptomatic patients and longer intervals of contagiousness. Conversely, the SA variant does favor the open state of the trimer. At pH 7, the stability of the trimer is ca. 6 times higher for the SA variant in comparison with SARS-CoV-2 S wt ($\Delta G_{elec}[wt] = -72 \pm 4$ kJ/mol and $\Delta G_{elec}[SA] = -225 \pm 4$ kJ/mol for DUU, and $\Delta G_{elec}[wt] = -75 \pm 5$ kJ/mol and $\Delta G_{elec}[SA] = -227 \pm 5$ kJ/mol for UDD). The main conclusions are, schematically, summarized in **Supplementary Figure 8**. The synergistic effect between the tendency to increase the affinity with ACE2 and the availability of RBD up for binding promotes higher transmissivity and virulence for the SA variant. This might be even more substantial than could be estimated here due to all the approximations assumed. For instance, we did not touch on other effects that can be affected by pH and amplify more the virulence [e.g., the contributions from the glycans; the interaction of the viral proteins with other cellular receptors, such as TMPRSS2, which cleaves protein S at two sites, and NEUROPILIN-1, an alternative host factor for SARS-CoV-2 (128, 129)].

Viruses are simply a group of organisms that exploits their environment in order to survive. Thus, they have to balance the harm they cause and their ability to transmit itself (130). By looking historically at the human diseases caused by the Coronaviridae family, the Severe Acute Respiratory Disease from

2003, caused by SARS-CoV-1, did not cause such disastrous effects because, even though the case fatality rate was 9.6%, the disease caused such deleterious effects on the body that the infected individuals isolated themselves in their homes (15). As discussed above, disposability of spike homotrimers might interpret this as being in the open state (DUU and UDD). Transmissivity was also affected because the virus could not transmit itself to a large number of people.

SARS-CoV-2, on the other hand, developed a mechanism that enabled it to stay closed to silently spread more widely—a kind of "chameleon" feature as cited above. As a result, the number of cases raised alarmingly, and the average COVID-19 case fatality rate is about 2%−3% worldwide (131). As time went by and most of the population stopped following the security measures, the VOCs included mutations that dismissed the practical need of the "chameleon" feature (probably due to a higher probability of trimers at the DDD conformation state), which made them more virulent and dangerous. The combination of all the mechanisms listed above gave the SA variant (and possibly for others as well) a well-improved fitness from the virus perspective. Even though there are still very few studies on the matter, the case fatality rate is expected to increase.

Besides shedding some light on how the virus uses physical–chemical properties to evolve, the molecular mechanisms reported here also have direct implications for physiopathology, therapeutic strategies, and vaccines. The number of available receptors ACE2 can be easily compensated by an increased affinity and an elevated number of receptors ready to dock at the human cell. Higher RBD-ACE2 binding affinities observed for the new variants imply that younger infected individuals without comorbidities and with naturally less disposability of receptors ACE2 can have similar clinical conditions as observed for older patients contaminated by the wild-type version.

The neutralization of the new variants would require an elevated number of binders. For instance, it has to be investigated if higher doses of the available vaccines could boost the immune system to produce enough concentration of antibodies to neutralize such a high number of RBDs that are ready to interact with ACE2. Considering the increased stability of the B.1.351 (or Beta) variant in its infectious form, the need to adopt prompt and effective measures to contain the advance of the pandemic must be reinforced. Even though one is supposed to have protection on some level with vaccination against the early variants, the necessary concentration of antibodies for the new variants might not be achieved so fast by the human body. Social distance, masks, and rapid and global mass vaccination should be adopted to prevent transmission of the virus since it only tends to accumulate mutations to be more transmittable and more prepared to evade the immune strategies of the body (40). It is worth noting that the strongest ACE2-affinity-enhancing

mutations have not yet been selected in current variants. As the virus evolves, it can still find other substitutions to improve its destructive power to spread and thrive (from his perspective), increasing our survival risk (46).

AUTHOR'S NOTE

The following note added in Proof was submitted by the authors after this paper was accepted. After acceptance, we were informed by an experimental work (https://doi.org/10.1016/j.chom.2021.04.007) which also suggests the relevance of the up state for one of the variants.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CG: visualization, investigation, and writing–original draft. FS: conceptualization, investigation, methodology, software, data curation, and writing–review and editing, and supervision. AL: data curation, methodology, and writing–review and editing, and supervision. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmedt.2021.694347/full#supplementary-material

REFERENCES

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* (2020) 20:5. doi: 10.1016/S1473-3099(20)30120-1

2. Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect. Dis.* (2012) 12:9. doi: 10.1016/S1473-3099(12)70121-4

3. Li Y, Yao L, Li J, Chen L, Song Y, Cai Z, et al. Stability issues of RT-PCR testing of SARS-CoV-2 for hospitalized patients clinically diagnosed with COVID-19. *J Med Virol.* (2020) 92:7. doi: 10.1002/jmv.25786

4. Xie X, Zhong Z, Zhao W, Zheng C, Wang F, Liu J. Chest CT for Typical Coronavirus Disease 2019 (COVID-19) Pneumonia: relationship to negative rt-pcr testing. *Radiology.* (2020) 296:2. doi: 10.1148/radiol.2020200343

5. Tahamtan A, Ardebili A. Real-time RT-PCR in COVID-19 detection: issues affecting the results. *Expert Rev Mol Diagn.* (2020) 20:5. doi: 10.1080/14737159.2020.1757437

6. Walls AC, Park Y, Tortorici MA, Wall A, Mcguire AT, Veesler D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* (2020) 181:2. doi: 10.1016/j.cell.2020.11.032

7. Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. *J. Mol. Biol.* (2020) 432:10. doi: 10.1016/j.jmb.2020.04.009

8. Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science.* (2005) 309:5742. doi: 10.1126/science.1116480

9. Cai Y, Zhang J, Xaio T, Peng H, Sterling S, Walsh R, et al. Distinct conformational states of SARS-CoV-2 spike protein. *Science.* (2020) 369:6511. doi: 10.1126/science.abd4251

10. Kupferschmidt K. New mutations raise specter of 'immune escape'. *Science.* (2021) 371:6527. doi: 10.1126/science.371.6527.329

11. Robson F, Khan KS, Le TK, Paris C, Demirbag S, Barfuss P, et al. Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell.* (2020) 79:5. doi: 10.1016/j.molcel.2020.07.027

12. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall.* (2017) 1:1. doi: 10.1002/gch2.1018

13. Yao H, Lu X, Chen Q, Xu K, Chen Y, Cheng M, et al. Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity *in vitro* and with clinical implications *in vivo*. *Cell Discov.* (2020) 6:76. doi: 10.1038/s41421-020-00226-1

14. CDC. *Update: Severe Acute Respiratory Syndrome-Worldwide and United States.* (2003). Available online at: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5228a4.htm (accessed March 7, 2021).

15. European Centre for Disease Prevention and Control. *Risk Related to Spread of New SARS-CoV-2 Variants of Concern in the EU/EEA, First Update.* (2021) Available online at: https://www.ecdc.europa.eu/en/publications-data/covid-19-risk-assessment-spread-new-variants-concern-eueea-first-update (accessed March 7, 2021).

16. Davies NG, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, Pearson CAB, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science.* (2021) 372:6538. doi: 10.1126/science.abg3055

17. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. *Cell.* (2021) 184:5. doi: 10.1016/j.cell.2021.01.044

18. Coutinho RM, Marquitti FMD, Ferreira LS, Borges ME, Paixão da Silva RL, Canton O et al. Model-based estimation of transmissibility and reinfection of SARS-CoV-2 P.1 variant. *medRxiv [Preprint].* (2021) doi: 10.1101/2021.03.03.21252706

19. Capozzi L, Bianco A, Del Sambro L, Simone D, Lippolis A, Notarnicola M. et al. Genomic surveillance of circulating SARS-CoV-2 in South East Italy: a one-year retrospective genetic study. *Viruses .* (2021) 13:731. doi: 10.3390/v13050731

20. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Lambson BE, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat Med.* (2021). 27:622–5. doi: 10.1101/2021.01.18.427166

21. Teruel N, Mailhot O, Najmanovich RJ. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. *BioRxiv [Preprint].* (2020). Available online at: https://www.biorxiv.org/content/10.1101/2020.12.16.423118v3 (accessed March 9, 2021).

22. Kemp SA, Collier DA, Datir RP, Ferreira IATM, Gayed S, Jahun A, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature.* (2021) 592:277–82. doi: 10.1038/s41586-021-03291-y

23. Nelson G, Buzko O, Spilman P, Niazi K, Rabizadeh S, Soon-Shiong P. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. *BioRxiv [Preprint].* (2021) Available at: https://www.biorxiv.org/content/10.1101/2021.01.13.426558v1 (accessed March 9, 2021).

24. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *J Hum Genet.* (2020) 65:1075–82. doi: 10.1038/s10038-020-0808-9

25. Kolifarhood G, Aghaali M, Mozafar Saadati H, Taherpour N, Rahimi S, Izadi N, et al. Epidemiological and clinical aspects of COVID-19; a narrative review. *Arch Acad Emerg Med.* (2020) 8:1.

26. Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med.* (2020) 26:1037–40. doi: 10.1038/s41591-020-0916-2

27. Hoffman M, Arora P, Groß R, Seidel A, Hörnich B, Hahn A, et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. *Cell.* (2021) 184:9. doi: 10.1016/j.cell.2021.03.036

28. Public Health England. *NERVTAG: Update Note on B.1.1.7 Severity, 11 February 2021.* Available online at: https://www.gov.uk/government/publications/nervtag-update-note-on-b117-severity-11-february-2021 (accessed March 7, 2021).

29. Sabino EC, Buss LF, Carvalho MPS, A Prete Jr C, Crispim MAEC, A Fraiji N, et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet.* (2021) 397:10273. doi: 10.1016/S0140-6736(21)00183-5

30. Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nat Commun.* (2020) 11:6317. doi: 10.1038/s41467-020-19741-6

31. Jimenez L, Codo AC, Sampaio VSS, Oliveira ERA, Ferreira LKK, Davanzo GG, et al. The influence of pH on SARS-CoV-2 infection and COVID-19 severity. *medXriv [Preprint].* (2020). Available at: https://www.medrxiv.org/content/10.1101/2020.09.10.20179135v1 (accessed March 8, 2021).

32. Bhella D. The role of cellular adhesion molecules in virus attachment and entry. *Philos Trans R Soc B: Biol Sci.* (2015) 370:1661. doi: 10.1098/rstb.2014.0035

33. Maginnis MS. Virus–Receptor Interactions: the key to cellular invasion. *J Mol Biol.* (2018) 430:17. doi: 10.1016/j.jmb.2018.06.024

34. Li W, Su Y-Y, Zhi S-S, Huang J, Zhuang C-L, Bai W-Z, et al. Virus shedding dynamics in asymptomatic and mildly symptomatic patients infected with SARS-CoV-2. *CMI.* (2020) 26:11. doi: 10.1016/j.cmi.2020.07.008

35. Katze M, Korth M, Law GL, Nathanson N. *Viral Pathogenesis.* 3rd ed. London: Academic Press. (2016).

36. Sigrist CJ, Bridge A, Mercier P. A potential role for integrins in host cell entry by SARS-CoV-2. *Antiviral Res.* (2020) 177. doi: 10.1016/j.antiviral.2020.104759

37. Bò L, Miotto M, Rienzo L, Milanetti E, Ruocco G. Exploring the association between sialic acid and SARS-CoV-2 spike protein through a molecular dynamics-based approach. *Front. Med. Technol.* (2021) 2:614652. doi: 10.3389/fmedt.2020.614652

38. Gao S, Zhang L. ACE2 partially dictates the host range and tropism of SARS-CoV-2. *Comput Struct Biotechnol J.* (2020) 18: 4040–7. doi: 10.1016/j.csbj.2020.11.032

39. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature.* (2021) 593:130–35. doi: 10.1101/2021.01.25.428137

40. Altmann DM, Boyton RJ, Beale R. Immunity to SARS-CoV-2 variants of concern. *Science.* (2021) 371:6534. doi: 10.1126/science.abg7404

41. Starr TN, Greaney AJ, Addetia A, Hannon WW, Choudhary MC, Dingens AS, et al. Prospective mapping of viral mutations that escape antibodies used to treat COVID-19. *Science.* (2021) 371:6531. doi: 10.1126/science.abf9302

42. Poveda-Cuevas AS, Barroso da Silva FL, Etchebest C. How the strain origin of Zika virus NS1 protein impacts its dynamics and implications to their differential virulence. *J. Chem. Inf. Model.* (2021) 61:3. doi: 10.1021/acs.jcim.0c01377

43. Giron CC, Barroso da Silva FL, Laaksonen A. On the interactions of the receptor-binding domain of SARS-CoV-1 and SARS-CoV-2 spike proteins with monoclonal antibodies and the receptor ACE2. *Virus Res.* (2020) 285:198021. doi: 10.1016/j.virusres.2020.198021

44. Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, et al. Structural basis of receptor recognition by SARS-CoV-2. *Nature.* (2020) 581. doi: 10.1038/s41586-020-2179-y

45. Khatri I, Staal FJT & van Dongen, JJM. Blocking of the High-Affinity Interaction-Synapse Between SARS-CoV-2 Spike and Human ACE2 Proteins Likely Requires Multiple High-Affinity Antibodies: An Immune Perspective. *Front. Immunol.* (2020) 11:570018. doi: 10.3389/fimmu.2020.570018

46. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Bloom JD, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell.* (2020) 182:5. doi: 10.1016/j.cell.2020.08.012

47. Gui M, Song W, Zhou H, Xu J, Chen S, Xiang Y, et al. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell Res.* (2016) 27:1. doi: 10.1038/cr.2016.152

48. Benton DJ, Wrobel AG, Xu P, Roustan C, Martin SR, Rosenthal PB, et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature.* (2020) 588:7837. doi: 10.1038/s41586-020-2772-0

49. Song W, Gui M, Wang X, Xiang Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathogens.* (2018) 14:e1007236. doi: 10.1371/journal.ppat.1007236

50. Zhou T, Tsybovsky Y, Olia A, Gorman J, Rapp M, Cerutti G, et al. Cryo-EM structures of SARS-CoV-2 spike without and with ACE2 reveal a pH-dependent switch to mediate endosomal positioning of receptor-binding domains. *Cell Host Microbe.* (2020) 28:6. doi: 10.1016/j.chom.2020.11.004

51. Ibrahim B, McMahon DP, Hufsky F, Beer M, Deng L, le Mercier P, et al. A new era of virus bioinformatics. *Virus Res.* (2018) 251. doi: 10.1016/j.virusres.2018.05.009

52. Sharma D, Priyadarshini P, Vrati S. Unraveling the web of viroinformatics: computational tools and databases in virus research. *J. Virol.* (2015) 89:3. doi: 10.1128/JVI.02027-14

53. Sato H, Yokoyama M, Toh H. Genomics and computational science for virus research. *Front. Microbiol.* (2013) 4:42. doi: 10.3389/978-2-88919-126-0

54. Poveda-Cuevas AS, Etchebest C, Barroso da Silva FL. Identification of Electrostatic Epitopes in Flavivirus by Computer Simulations: The PROCEEDpKa Method. *J. Chem. Inf. Model.* (2020) 60:2. doi: 10.1021/acs.jcim.9b00895

55. Keretsu S, Bhujbal SP, Cho SJ. Rational approach toward COVID-19 main protease inhibitors via molecular docking, molecular dynamics simulation and free energy calculation. *Sci Rep.* (2020) 10:17716. doi: 10.1038/s41598-020-74468-0

56. Francés-Monerris A, Hognon C, Miclot T, García-Iriepa C, Iriepa I, Terenzi A. Molecular basis of SARS-CoV-2 infection and rational design of potential antiviral agents: modeling and simulation approaches. *J. Proteome Res.* (2020) 19:11. doi: 10.1021/acs.jproteome.0c00779

57. Gao K, Wang R, Chen J, Cheng L, Frishcosy J, Huzumi Y, et al. Methodology-centered review of molecular modeling, simulation, and prediction of SARS-CoV-2. *arXiv [Preprint].* (2021) Available at: https://arxiv.org/abs/2102.00971 (accessed March 19, 2021).

58. Lizbeth RG, Jazmín GM, José CB, Marlet MA. Immunoinformatics study to search epitopes of spike glycoprotein from SARS-CoV-2 as potential vaccine. *J Biomol Struct Dyn.* (2020) 1–5. doi: 10.1080/07391102.2020.1780944

59. Poveda-Cuevas AS, Etchebest C, Barroso da Silva FL. Insights into the ZIKV NS1 virology from different strains through a fine analysis of physicochemical properties. *Acs Omega.* (2018) 3:11. doi: 10.1021/acsomega.8b02081

60. Mondal D, Warshel A. Exploring the mechanism of covalent inhibition: simulating the binding free energy of α-ketoamide inhibitors of the main protease of sars-cov-2. *Biochemistry.* (2020) 59:48. doi: 10.1021/acs.biochem.0c00782

61. Frenkel D, Frenkel D, Smit B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd Edn. San Diego, CA: Academic Press. (2001).

62. Binder K. (ed.). *Monte Carlo Methods in. Statistical Physics.* Springer-Verlag. Böttcher-Friebertshäuser E, Garten W, Klenk HD. (eds.), (2018). *Activation of Viruses by Host Proteases.* Cham: Springer International Publishing.

63. Rapaport DC. *The Art of Molecular Dynamics Simulation*, 2nd edn. Cambridge University Presshttps. (2004).

64. Barroso da Silva FL, Carloni P, Cheung D, Cottone G, Donnini S, Foegeding EA, et al. Understanding and controlling food protein structure and function in foods: perspectives from experiments and computer simulations. *Annu. Rev. Food Sci. Technol.* (2020) 11:1. doi: 10.1146/annurev-food-032519-051640

65. van Gunsteren WF, Dolenc J. Thirty-five years of biomolecular simulation: development of methodology, force fields and software. *Mol. Simul.* (2012) 38:14–15. doi: 10.1080/08927022.2012.701744

66. Teixeira AA, Lund M, Barroso da Silva F L. Fast proton titration scheme for multiscale modeling of protein solutions. *J. Chem. Theory Comput.* (2010) 6: 3259–66. doi: 10.1021/ct1003093

67. Barroso da Silva FL, MacKernan D. Benchmarking a fast proton titration scheme in implicit solvent for biomolecular simulations. *J. Chem. Theory Comput.* (2017) 13:6. doi: 10.1021/acs.jctc.6b01114

68. Frigori RB, Barroso da Silva FL, Carvalho PPD, Alves NA. Occurrence of biased conformations as precursors of assembly states in fibril elongation of amyloid-β fibril variants: an in silico study. *J. Phys. Chem. B.* (2020) 124:14. doi: 10.1021/acs.jpcb.0c01360

69. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. *Nucleic Acids Res.* (2020) 28:1. doi: 10.1093/nar/28.1.235

70. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics.* (2006) 22:2. doi: 10.1093/bioinformatics/bti770

71. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet.* (2020) 395:10224. doi: 10.1016/S0140-6736(20)30251-8

72. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Ferrin TE. UCSF chimera: a visualization system for exploratory research and analysis. *J Comp Chem.* (2004) 25:13. doi: 10.1002/jcc.20084

73. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv [Preprint].* (2020). Available at: https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1.full-text#F2 (accessed March 21, 2021).

74. Tegally H, Wilkinson E, de Oliveira T, Lessels R, Giandhari J, Pillay S, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* (2021) 27. doi: 10.1038/s41586-021-03402-9

75. Rambaut A, Loman N, Pybus O, Pybus O, Barclay W, Barrett J, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological [Preprint].* (2020). Available online at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed June 08, 2020).

76. Naveca F, Nascimento V, Souza V, Corado A, Nascimento F, Resende P, et al. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein. *Virological [Preprint].* (2021). Available online at: https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585 (accessed June 08, 2021).

77. Wang P, Wang M, Yu J, Cerutti G, Nair MS, Huang Y, et al. Increased Resistance of SARS-CoV-2 Variant P.1 to Antibody Neutralization. *bioRxiv [Preprint].* (2021) Available online at: https://www.biorxiv.org/content/10.1101/2021.03.01.433466v1 (accessed March 18, 2021).

78. Mahase E. Covid-19: What new variants are emerging and how are they being investigated? *BMJ.* (2021) 158:372. doi: 10.1136/bmj.n158

79. Mallapaty S. COVID mink analysis shows mutations are not dangerous — yet. *Nature.* (2020) 587:340–1. doi: 10.1038/d41586-020-03218-z

80. Annavajhala MK, Mohri H, Zucker JE, Shen Z, Wang P, Gomez-Simmonds A, et al. A novel SARS-CoV-2 variant of concern, B.1.526, identified in New

York. *medRxiv [Preprint]*. (2021) Available online at: https://www.medrxiv.org/content/10.1101/2021.02.23.21252259v1 (accessed March 30, 2021).

81. Callaway E. Multitude of coronavirus variants found in the US — but the threat is unclear. *Nature*. (2021) 591:190. doi: 10.1038/d41586-021-00564-4

82. Deng X, Garcia-Knight MA, Khalid MM, Servellita V, Wang C, Chiu CY, et al. Transmission, infectivity, and neutralization of a spike L452R SARS-CoV-2 variant. *Cell*. (2021). doi: 10.1016/j.cell.2021.04.025

83. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. (2021) 49:D1. doi: 10.1093/nar/gkaa1100

84. Garay PG, Barrera E, Klein F, Machado MR, Soñora M, Pantano S. The SIRAH-CoV-2 Initiative: a coarse-grained simulations' dataset of the sars-cov-2 proteome. *Front. Med. Technol.* (2021) 3:644039. doi: 10.3389/fmedt.2021.644039

85. Moreira RA, Guzman HV, Boopathi S, Baker JL, Poma AB. Characterization of Structural and Energetic Differences between Conformations of the SARS-CoV-2 Spike Protein. *Materials*. (2020) 13:23. doi: 10.3390/ma13235362

86. Yu A, Pak AJ, He P, Monje-Galvan V, Casalino L, Gaieb Z, et al. A multiscale coarse-grained model of the SARS-CoV-2 virion. *Biophys. J.* (2021) 120:6. doi: 10.1016/j.bpj.2020.10.048

87. Eswar N, Webb B, Merti-Renom MA, Madhusudhan MS, Eramian D, Sehn, et al. Comparative protein structure modeling using modeller. *Curr Protocols Bioinform*. (2006) 15:1. doi: 10.1002/0471250953.bi0506s15

88. Cao W, Dong C, Kim S, Du L, Im W, Zhang X, et al. Biomechanical characterization of SARS-CoV-2 spike RBD and human ACE2 protein-protein interaction. *Biophys. J.* (2021) 120:6. doi: 10.1016/j.bpj.2021.02.007

89. Singh A, Montgomery D, Xue X, Foley BL, Woods RJ. GAG Builder: a web-tool for modeling 3D structures of glycosaminoglycans. *Glycobiology*. (2019) 29:7. doi: 10.1093/glycob/cwz027

90. Gowthaman R, Guest JD, Yin R, Adolf-Bryfogle J, Schief WR, Pierce BG. CoV3D: a database of high resolution coronavirus protein structures. *Nucleic Acids Res*. (2021) 49:D1. doi: 10.1093/nar/gkaa731

91. Krissinel E, Henrick K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* (2007) 372:3. doi: 10.1016/j.jmb.2007.05.022

92. Needleman S, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* (1970) 48:3. doi: 10.1016/0022-2836(70)90057-4

93. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. *Chem. Rev.* (2016) 116:14. doi: 10.1021/acs.chemrev.6b00163

94. Fossepre M, Laaksonen A, Lyubartsev A, Mocci F, Naômé A, Vercauteren D. Biomolecular Modeling across Spatial & Temporal Scales (2020) 2.

95. Leach AR. *Molecular Modelling–Principles and Applications*, 1st Edn. Singapore: Longman. (1996).

96. Barroso da Silva FL, Dias LG. Development of constant-pH simulation methods in implicit solvent and applications in biomolecular systems. *Biophys. Rev.* (2017) 9:699–728. doi: 10.1007/s12551-017-0311-5

97. Barroso da Silva FL, Pasquali S, Derremaux P, Dias LG. Electrostatics analysis of the mutational and pH effects of the N-terminal domain self-association of the major ampullate spidroin. *Soft Matter*. (2016) 12:25. doi: 10.1039/C6SM00860G

98. Delboni LA, Barroso da Silva FL. On the complexation of whey proteins. *Food Hydrocoll*. (2016) 55:89–99. doi: 10.1016/j.foodhyd.2015.11.010

99. Persson BA, Lund M, Forsman J, Chatterton DEW, Torbjörn AAkesson. Molecular evidence of stereo-specific lactoferrin dimers in solution. *Biophys. Chem*. (2010) 151:3. doi: 10.1016/j.bpc.2010.06.005

100. Mendonça DC, Macedo JN, Guimarães SL, Barroso da Silva FL, Cassago A, Garratt RC et al. A Revised Order of Subunits in Mammalian Septin Complexes. *Cytoskeleton*. (2019) 76:9–10. doi: 10.1002/cm.21569

101. Barroso da Silva FL, Derreumaux P, Pasquali S. Fast coarse-grained model for RNA titration. *J. Chem. Phys.* (2017) 146:3. doi: 10.1063/1.4972986

102. Barroso da Silva FL, Derreumaux P, Pasquali S. Protein-RNA complexation driven by the charge regulation mechanism. *Biochem. Biophys. Res. Commun.* (2018) 498:2. doi: 10.1016/j.bbrc.2017.07.027

103. Williams SL, Blachly PG, Mccammon JA. Measuring the successes and deficiencies of constant pH molecular dynamics: a blind prediction study. *Proteins*. (2011) 79:12. doi: 10.1002/prot.23136

104. Chen W, Huang Y, Shen J. Conformational activation of a transmembrane proton channel from constant pH molecular dynamics. *J Phys Chem Lett*. (2016) 7:19. doi: 10.1021/acs.jpclett.6b01853

105. Barroso da Silva FL, Sterpone F, Derremaux P. OPEP6: a new constant-ph molecular dynamics simulation scheme with opep coarse-grained force field. *J Chem Theory Comput*. (2019) 15:6. doi: 10.1021/acs.jctc.9b00202

106. Mezzenga R, Fischer P. The self-assembly, aggregation and phase transitions of food protein systems in one, two and three dimensions. *Rep. Prog. Phys.* (2013) 76:4. doi: 10.1088/0034-4885/76/4/046601

107. Hyltegren K, Polimeni M, Skepö M, Lund M. Integrating all-atom and coarse-grained simulations—toward understanding of IDPs at surfaces. *J. Chem. Theory Comput.* (2020) 16:3. doi: 10.1021/acs.jctc.9b01041

108. Kurut A, Persson BA, AAkesson T, Forsman J, Lund M. Anisotropic interactions in protein mixtures: self assembly and phase behavior in aqueous solution. *J. Phys. Chem. Lett.* (2012) 3:6. doi: 10.1021/jz201680m

109. Eisemberg D, Weiss RM, Terwilliger TC, Wilcox W. Hydrophobic Moments and Protein Structure. *Faraday Symp Chem Soc*. (1982) 17:109–20. doi: 10.1039/fs9821700109

110. Stenqvist B., Thuresson A., Kurut A, Vácha R, Lund M. Faunus–a flexible framework for monte carlo simulation. *Mol Simulat*. (2013) 39:14–5. doi: 10.1080/08927022.2013.828207

111. Qi T, Cao Y. In translation: FcRn across the therapeutic spectrum. *International J Mol Sci*. (2021) 22:6. doi: 10.3390/ijms22063048

112. Ibarra-Molero B, Loladze VV, Makhatadze GI, Sanchez-Ruiz JM. Thermal versus guanidine-induced unfolding of ubiquitin. AnBjo analysis in terms of the contributions from charge-charge interactions to protein stability. *Biochemistry*. (1999) 38:25. doi: 10.1021/bi9905819

113. Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*. (2003) 426:6965. doi: 10.1038/nature02145

114. Zhou P, Yang X, Wang X, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. (2020) 579: 270–3. doi: 10.1038/s41586-020-2951-z

115. Yan R, Zhang Y, Li Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. (2020) 367:6485. doi: 10.1126/science.abb2762

116. Brielle ES, Schneidman-Duhovny D, Linial M. The SARS-CoV-2 Exerts a Distinctive Strategy for Interacting with the ACE2 Human Receptor. *Viruses [Internet]. MDPI AG*. (2020) 12:5. doi: 10.3390/v12050497

117. Tian X, Li C, Huang A, Xia S, Lu S, Shi Z, et al. Potent binding of 2019 novel coronavirus spike protein by a SARS coronavirus-specific human monoclonal antibody. *Emerg. Microbes Infect.* (2020) 9:1. doi: 10.1080/22221751.2020.1729069

118. Nguyen HL, Lan PD, Thai NQ, Nissley DA, O'Brien EP, Li MS. Does SARS-CoV-2 Bind to Human ACE2 More Strongly Than Does SARS-CoV? *J. Phys. Chem. B.* (2020) 124:34. doi: 10.1021/acs.jpcb.0c04511

119. Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, Abiona O, et al. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science*. (2020) 367:6483. doi: 10.1126/science.abb2507

120. Wang Q, Zhang Y, Wu L, Niu S, Song C, Zhang Z et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. *Cell*. (2020) 181:4. doi: 10.1016/j.cell.2020.03.045

121. Ali F, Kasry A, Amin M. The new SARS-CoV-2 strain shows a stronger binding affinity to ACE2 due to N501Y mutant. *Med Drug Discov*. (2021) 10:100086. doi: 10.1016/j.medidd.2021.100086

122. Helenius A. Virus entry: what has pH got to do with it?. *Nat Cell Biol*. (2013) 15:125. doi: 10.1038/ncb2678

123. Yuan P, Yang Z, Song H, Wang K, Yang Y, Xie L et al. Three main inducers of alphacoronavirus infection of enterocytes: sialic acid, proteases, and low pH. *Intervirology*. (2018) 61:53–63. doi: 10.1159/000492424

124. Simmons G, Gosalia DN, Rennekamp AJ, Reeves JD, Diamond SL, Bates P. Inhibitors of cathepsin L prevent severe acute respiratory syndrome coronavirus entry. *Proc Natl Acad Sci USA*. (2005) 102:33. doi: 10.1073/pnas.0505577102

125. Srivastava D, Santiso E, Gubbins K, Barroso da Silva FL. Computationally mapping pKa shifts due to the presence of a polyelectrolyte chain around whey proteins. *Langmuir*. (2017) 33:42. doi: 10.1021/acs.langmuir.7b02271

126. Yuan M, Wu NC, Zhu X, Lee C-CD, So RTY, Lv H, et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science.* (2020) 386:6491. doi: 10.1126/science.abb7269

127. Bai C, Warshel A. Critical Differences between the binding features of the spike proteins of SARS-CoV-2 and SARS-CoV. *J. Phys. Chem. B.* (2020) 124:28. doi: 10.1021/acs.jpcb.0c04317

128. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell.* (2020) 181:2. doi: 10.1016/j.cell.2020.02.052

129. Daly JL, Simonetti B, Klein K, Chen K, Williamson MK, Antón-Plágaro C, et al. Neuropilin-1 is a host factor for SARS-CoV-2 infection. *Science.* (2020) 370:6518. doi: 10.1126/science.abd3072

130. Ewald PW, Sussman JB, Distler MT, Libel C, Chammas WP, Dirita VJ, et al. Evolutionary control of infectious disease: prospects for vectorborne and waterborne pathogens. *Mem. Inst. Oswaldo Cruz.* (1998) 93:5. doi: 10.1590/S0074-02761998000500002

131. Cao Y, Hiyoshi A, Montgomery S. COVID-19 case-fatality rate and demographic and socioeconomic influencers: worldwide spatial regression analysis based on country-level data. *BMJ Open.* (2020) 10:43560. doi: 10.1136/bmjopen-2020-043560