

Article

Assessing pH-Dependent Conformational Changes in the Fusion Peptide Proximal Region of the SARS-CoV-2 Spike Glycoprotein

Darya Stepanenko ^{1,2}, Yuzhang Wang ^{1,3} and Carlos Simmerling ^{1,3,*}

¹ Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, NY 11794, USA; darya.stepanenko@stonybrook.edu (D.S.); yuzhang.wang@stonybrook.edu (Y.W.)

² Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA

³ Department of Chemistry, Stony Brook University, Stony Brook, NY 11794, USA

* Correspondence: carlos.simmerling@stonybrook.edu

Abstract: One of the entry mechanisms of the SARS-CoV-2 coronavirus into host cells involves endosomal acidification. It has been proposed that under acidic conditions, the fusion peptide proximal region (FPPR) of the SARS-CoV-2 spike glycoprotein acts as a pH-dependent switch, modulating immune response accessibility by influencing the positioning of the receptor binding domain (RBD). This would provide indirect coupling of RBD opening to the environmental pH. Here, we explored this possible pH-dependent conformational equilibrium of the FPPR within the SARS-CoV-2 spike glycoprotein. We analyzed hundreds of experimentally determined spike structures from the Protein Data Bank and carried out pH-replica exchange molecular dynamics to explore the extent to which the FPPR conformation depends on pH and the positioning of the RBD. A meta-analysis of experimental structures identified alternate conformations of the FPPR among structures in which this flexible region was resolved. However, the results did not support a correlation between the FPPR conformation and either RBD position or the reported pH of the cryo-EM experiment. We calculated pKa values for titratable side chains in the FPPR region using PDB structures, but these pKa values showed large differences between alternate PDB structures that otherwise adopt the same FPPR conformation type. This hampers the comparison of pKa values in different FPPR conformations to rationalize a pH-dependent conformational change. We supplemented these PDB-based analyses with all-atom simulations and used constant-pH replica exchange molecular dynamics to estimate pKa values in the context of flexibility and explicit water. The resulting titration curves show good reproducibility between simulations, but they also suggest that the titration curves of the different FPPR conformations are the same within the error bars. In summary, we were unable to find evidence supporting the previously published hypothesis of an FPPR pH-dependent equilibrium: neither from existing experimental data nor from constant-pH MD simulations. The study underscores the complexity of the spike system and opens avenues for further exploration into the interplay between pH and SARS-CoV-2 viral entry mechanisms.

Keywords: coronavirus; spike; membrane fusion; viral entry; pH



Citation: Stepanenko, D.; Wang, Y.; Simmerling, C. Assessing pH-Dependent Conformational Changes in the Fusion Peptide Proximal Region of the SARS-CoV-2 Spike Glycoprotein. *Viruses* **2024**, *16*, 1066. <https://doi.org/10.3390/v16071066>

Academic Editor: Rui Yin

Received: 14 May 2024

Revised: 26 June 2024

Accepted: 28 June 2024

Published: 2 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The host cell entry mechanism of coronaviruses, such as the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), remains an active area of research. For SARS-CoV-2, the entry process is believed to initiate with the interaction between the viral spike fusion glycoprotein receptor binding domain (RBD, shown in Figure 1) and the host cell receptor, angiotensin-converting enzyme 2 (ACE2). The RBD, which has open and closed states, is able to interact with ACE2 only in the open state [1,2]. Otherwise, the RBD receptor binding motif is not accessible for binding [3]. Likewise, many antibody-binding epitopes on the RBD are masked in the closed state. Numerous structures are now present in the Protein Data Bank (PDB) for the three RBD domains in a variety of open and closed combinations.

In one prior study of the SARS-CoV-2 2P (prefusion stabilized spike with K986P/V987P [4]), the authors observed only an all-closed RBD conformation at pH 4.5 and 4.0 [5]. At pH 5.5, the authors observed RBD either in the open or closed position or in an undefined position. Some other viruses, like influenza, have entry proteins that undergo structural changes in acidic pH conditions. For example, the influenza hemagglutinin (HA) fusion protein facilitates viral membrane fusion with the endosomal membrane when activated by low pH [6,7]. This led the authors to hypothesize that acidification of the environment within the endosomal pH range, which varies from 6.5 to 4.0, mediates the positioning of the RBD. It was also proposed that this positioning allows the virus to evade the immune response within the endosome. Other authors [8] have suggested that an all-closed spike structure at a reduced pH may contribute to protein stability during viral assembly, preventing premature spike activation. In both cases, the authors suggest that pH-dependent RBD positioning may arise from pH-dependent conformational changes in a portion of the spike that is not in direct contact with the RBD; this region of the spike structure is discussed next.

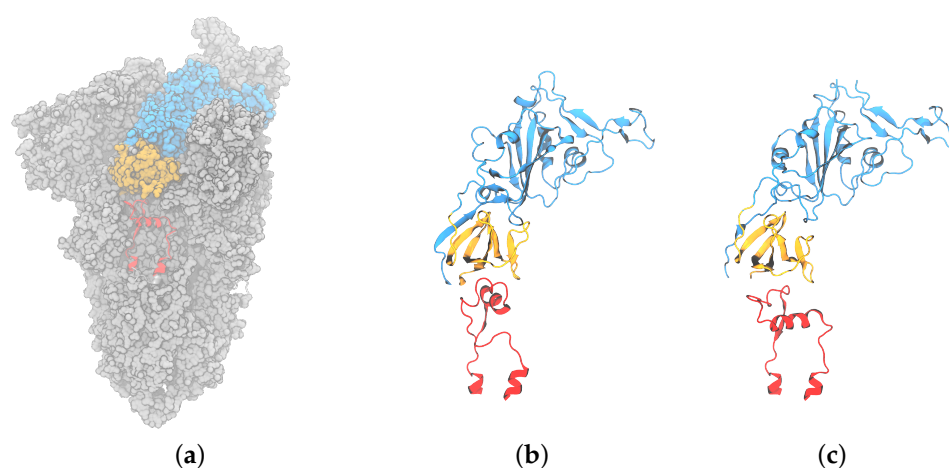


Figure 1. (a) Surface representation of spike trimeric ectodomain (PDB ID 6XM0). The receptor binding domain is colored blue, the C-terminal domain 1 is orange, and the fusion peptide proximal region is red; (b) cartoon representation of RBD, CTD1, and the compact conformation (PDB ID 6XR8 chain B [9]); (c) cartoon representation of RBD, CTD1, and the extended conformation FPPR (PDB ID 6XM0, FPPR from chain B [10]).

Fusion peptide proximal region: The spike structure is shown in Figure 1. Directly under the RBD domain lies the C-terminal domain 1 (CTD1); under CTD1, there is a short region that is unresolved in the majority of the structures deposited in the Protein Data Bank (PDB). This specific region, located between residues 824 and 858 with a disulfide bond between C840 and C851, is denoted as the **fusion peptide proximal region (FPPR)**. FPPR density was first resolved in the all-down-RBD configuration of the wild-type (WT) SARS-CoV-2 spike (PDB 6XR8) [11]. In that structure, the FPPR residues tightly pack together under the CTD1 and RBD of the protomer located in the clockwise direction from the FPPR, as illustrated in Figures 1b and 2a. Hereafter, this conformation will be referred to as the **compact FPPR**.

Later, a different ordered FPPR conformation was reported (6XM0, using SARS-CoV-2 with the 2P mutation along with GSAS instead of RRAR at the S1/S2 furin cleavage site) and was also observed under a closed RBD, as shown in Figures 1c and 2b [5]. Here, this conformation will be referred to as the **extended FPPR**. The authors also reported a partially resolved, extended-conformation FPPR under an open RBD.

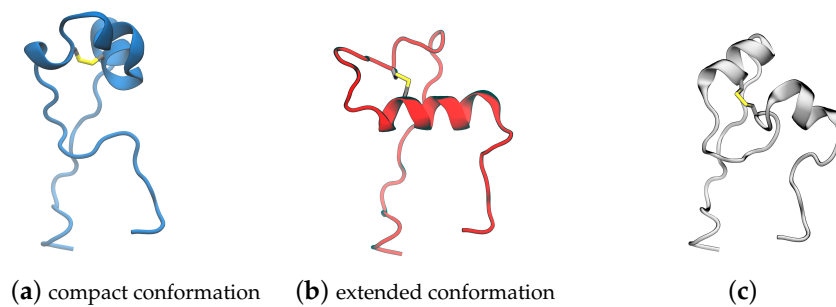


Figure 2. Representative structures for three conformational clusters of the FPPR obtained from analyzing spike structures in the PDB: (a) 7E7B (chain B), (b) 6XM0 (chain B), and (c) 7LS9 (chain A).

When the FPPR was initially resolved under a closed RBD [11], it was suggested that a structured FPPR (rather than the disordered FPPR as in all prior structures) might assist with maintaining the RBD in a downward position. The RBD and FPPR do not make direct contact but are separated by the CTD1 domain. The authors observed that the CTD1 domain shifts downward when the RBD opens, which can cause a steric clash between CTD1 and a compact FPPR. Subsequently, upon the observation of both compact and extended conformations under closed RBDs, along with exclusively an extended FPPR (partially resolved) under open RBDs, it was proposed that the conformation adopted by the FPPR could influence the ability of the RBD to open: with the compact conformation aiding with keeping the RBD closed and only the extended or disordered conformation allowing RBD opening [5].

The FPPR and surrounding region contain 13 acidic amino acids, including D614; the D614G mutation has largely replaced the original SARS-CoV-2 strain. These are partially compensated for by nine nearby Arg or Lys peptides. This high density of negative charge could lead to upward pKa shifts of these acidic side chains, which could possibly differ between alternate FPPR conformations in which the relative positions of the acidic groups vary. Some evidence exists to support a hypothesis that the FPPR conformation and pH are coupled. One cryo-EM spike study [5] observed the compact FPPR conformation at pH values 4.0 and 5.5 but only the extended conformation at pH 5.5 [5]. Based on this observation, the authors suggested that the FPPR serves as a pH-sensitive switch, with only the compact conformation being present under acidic conditions [5].

Taken together, the two hypotheses (RBD position depends on the FPPR, and the FPPR is influenced by pH) imply that the RBD would be largely closed in an acidic pH environment, shedding bound ACE2 or antibodies [5] and stabilizing the nascent spike during viral assembly [8].

Impact of acidic pH on spike behavior: There is additional evidence suggesting that the spike experiences conformational changes when exposed to acidic pH. It has been reported that acidic pH helps restore the spike protein after conformational changes that occur due to prolonged spike storage [12]. Following the storage of the spike protein (SARS-CoV-2 2P variant) for 8 days at 4 °C in PBS (pH 7.4), notable changes in conformational states were detected by negative-stain electron microscopy. The data indicate a reduction in the fraction of well-folded spike particles from 89% to 47% [12]. Additionally, a decrease in the melting temperature was observed by differential scanning calorimetry (DSC). Upon 10 min exposure to buffers with pH 5.5 or 4.0, the DSC melting profile was restored and a fully folded conformation was regained [12].

Others have reported changes in spike stability with pH. After a disulfide-stabilized spike (S-R/x3, with a disulfide linking the RBD and central helix) was stored at 4° for 40 days at pH 7.4, most of the the spike stayed folded, but the three RBDs moved slightly apart and the FPPR became disordered, as observed by cryo-EM [8]. Overnight incubation at pH 5.0 partially refolded the FPPR and brought the three RBDs closer.

In the present study, the hypothesis that the FPPR serves as a pH-sensitive switch was explored through a meta-analysis of 700 experimental spike structures deposited in the PDB and atomistic pH-replica exchange molecular dynamics simulations.

The analysis of PDB structures included extracting the RBD position, FPPR conformation, and reported pH. This revealed that both FPPR conformations have been observed in spike structures that were determined at pH values ranging from 4 to 8; likewise, both conformations have been observed under open and closed RBD domains, and both are present simultaneously in some spike structures. The lack of correlation between the FPPR conformation and either the pH or RBD position is supported by our constant-pH simulations. We quantified FPPR dynamics at different pH values and estimated acidic side chain pKa values. The simulation results indicate that the net protonation of the titratable residues in the FPPR and surrounding amino acids is determined largely by the pH rather than the FPPR conformation. These findings challenge the definition of the FPPR as a pH-sensitive switch, suggesting that both FPPR conformations may exist at neutral and acidic pH values, and factors other than pH may be responsible for the observation of alternate FPPR conformations. Consequently, the role of the FPPR in spike protein dynamics and function remains unclear. We also propose that the extensive use of artificially stabilized spike constructs for structure determination may obscure the true biological behavior.

2. Materials and Methods

All amino acid position numbers reported here are the same as in the full-length spike protein sequence.

A total of 700 SARS-CoV-2 full spike protein (Uniprot ID P0DTC2) cryo-EM structures with resolutions of 4 Å or better (see List S1) were retrieved from the PDB. The FPPR was identified using the sequence NKVTLADAGFI[KM]QYGDCLGD[IM]A[AY]RDLICAQKF[NK]GLTVLPPLLTDEMI, where residues in [] indicate variants in the sequence. Next, structures were filtered to retain only those in which all alpha-carbon coordinates are present for the FPPR. Individual structures were categorized using two different labels: by the pH of the experiment as reported in the PDB file or by the open/closed position of the RBD above that FPPR. The RBD position was determined using the SpikeScape tool [13].

Clustering FPPR conformations from the PDB: First, FPPR coordinates were extracted from the full PDB files obtained using the filters described above. These FPPR conformations were clustered into different groups using the coordinates of the alpha carbon atoms. The hierarchical agglomerative (bottom-up) method as implemented in the cptraj [14] program was used with a minimum distance between clusters of 4 Å.

The pKa values from static PDB structures: The pKa values were calculated using PROPKA3.5.1 with the option 'Display alternative pKa values due to coupling of titratable groups' enabled. [15,16].

Building the model system for MD simulations: Two independent systems were built using experimental spike structures with compact and extended FPPRs (6XLU and 6XM0). These were selected since they do not have any missing coordinates in or within 20 Å of the FPPR. Also, they were determined by the same lab [5], minimizing the possible impact of inconsistent structure determination protocols on our comparisons. The specific regions were: chain B, residues 35–54, 273–317, 726–783, 818–872 (the FPPR), 940–1023, and 1054–1061 and chain A, residues 272–330, 530–604, 607–656, and 661–673). To extract the coordinates of this region, the residues were selected in Pymol [17] and then saved to a new PDB file. An example obtained using spike PDB ID 6XM0 is shown in Figure 3.

The 20 Å cutoff for the model system was selected based on the assumption that residues farther than 20 Å from the FPPR do not significantly contribute to the electrostatics of the titratable residues. The assumption was tested via calculations using PROPKA3 that compared pKa values obtained on the full spike as well as the smaller fragment obtained from the same PDB file. All pKa values of titratable residues in a fragment were within 0.5 units of those obtained from the complete spike structure, as shown in Table S1. This test provides an estimate of the pKa uncertainty introduced by using the smaller model.

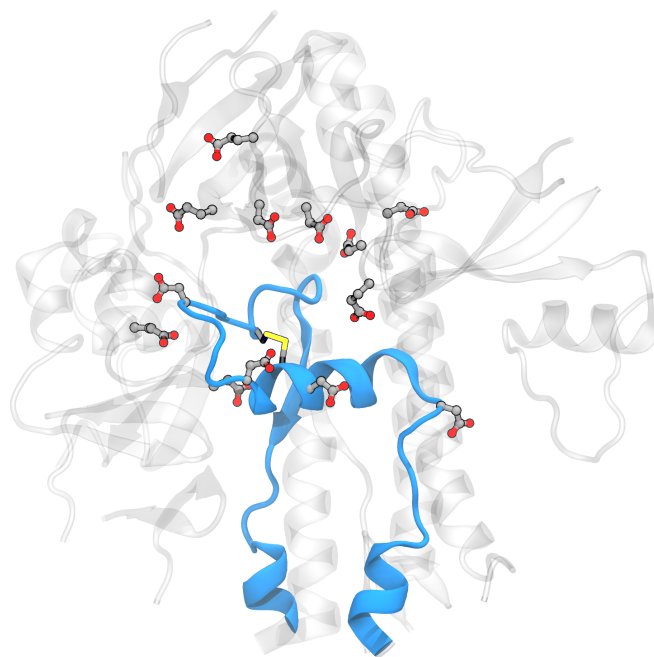


Figure 3. Model system for the spike FPPR region showing the protein backbone in blue for the extended-conformation FPPR and gray for the surrounding region. Asp, Glu, and Cys are shown in licorice (PDB ID 6XM0, FPPR from chain B).

SARS-CoV-2 is extensively N-glycosylated, which is necessary for host cell invasion and immune evasion [18,19]. This glycosylation is also predicted to modulate spike binding and enhance its open conformation, with each spike monomer bearing up to 22 glycosylation sites, totaling 66 across three monomers [20]. The composition and abundance of glycans varies significantly based on factors such as cell type, pH, CO₂ levels, oxygen levels, temperature, nutrient composition, and cellular membrane health, particularly in the ER and Golgi. However, the glycans are highly flexible, as evidenced by their lack of resolution in experimental structures. Converged sampling of these flexible ensembles remains intractable using current computer hardware; thus, these flexible regions were omitted from the simulations reported here.

Molecular dynamics general protocol: The ff14SB [21] and TIP3P [22] force fields were used for the protein and water, respectively. Unless specified, all simulations used the default settings in Amber v20 [23] with a 2 fs time step at constant temperature and volume, an 8 Å cutoff with particle mesh Ewald [24] for long range electrostatics, a Langevin thermostat with a collision frequency of 5 ps⁻¹, and SHAKE to constrain bonds involving hydrogen. Simulations were performed using the pmemd.CUDA modules of Amber v20 [23].

For both systems, hydrogen atoms and water were added using the Amber *tleap* program with a truncated octahedral periodic box with a minimum distance of the solute to the box edge of ~15 Å; the specific values varied in order to include the same number of water molecules in both systems (27,734 molecules). A total of 11 Na⁺ ions were added to neutralize the net charge of the system. The following residues were selected as titratable: Glu281, Asp830, Asp839, Asp843, Asp848, Glu554, Asp568, Asp571, Asp574, Glu583, Asp586, Asp614, and Glu619, as shown in Figure 4. The final system consisted of 90,405 atoms.

Both systems were relaxed with an eight-step protocol. Firstly, water and hydrogen atom positions were minimized for 1000 steps using the steepest descent and then for an additional 9000 steps with a conjugate gradient, while the rest of the system was restrained with 1 kcal/(mol · Å²) Cartesian positional restraints.

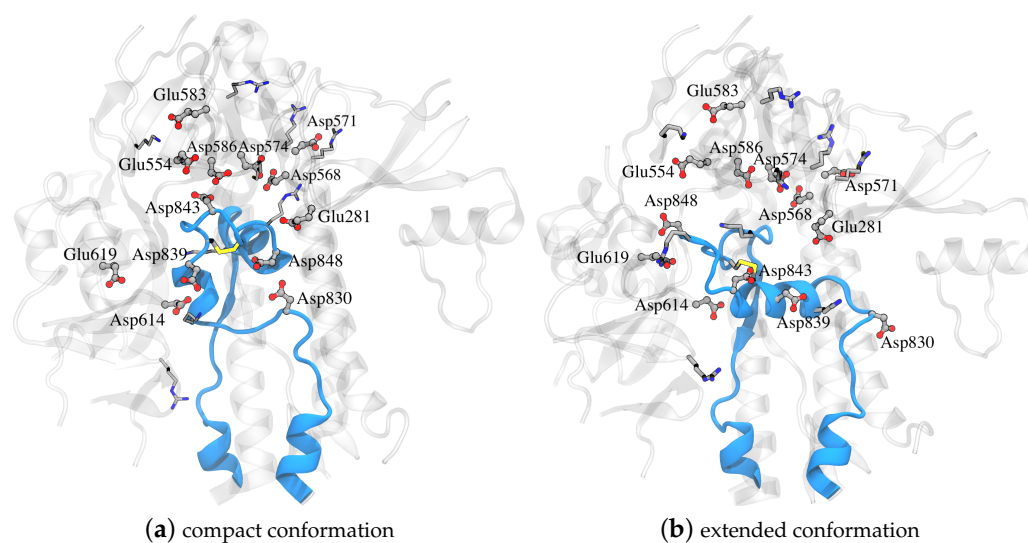


Figure 4. Part of the spike glycoprotein with the FPPR colored blue in (a) compact and (b) extended conformations with the rest of the protein around it colored gray and with Asp, Glu, Lys, and Arg in licorice (PDB ID 6XLU, 6XM0, FPPR— chain B).

Next, the system was heated to 300 K at a constant volume for 1 ns with $100 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ restraints applied to everything except hydrogen and water. The next steps involved 1 ns MD at constant pressure (1 bar) to relax the system density with the same positional restraints. Then, restraint force constants were lowered to $10 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$ for an additional 1 ns MD at constant pressure. Next, 10,000 steps of conjugate gradient minimization were performed, with restraints applied only to protein backbone atoms using a force constant of $10 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$. The next three steps of relaxation MD used 1 ns each at constant NPT, with positional restraints on protein backbone atoms with force constant of 10, 1, and then $0.1 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$. Each titratable carboxylate was deprotonated during relaxation, and no protonation state changes were attempted during these steps.

As only a subset of the spike protein was simulated, Cartesian positional restraints were maintained on all backbone atoms at the truncation boundary along with the backbone atoms of residues 272 to 291 in chain A and 292 to 317 in chain B using a force constant of $10 \text{ kcal}/(\text{mol} \cdot \text{\AA}^2)$.

The pH replica exchange molecular dynamics and estimation of pKa values: All pH-REMD simulations were performed with eight replicas that were equally spaced 1 pH unit apart over the range of 1 to 8. The structure obtained after relaxation was used as the initial coordinates and velocities for all replicas. A 2 fs time step was used, and protonation state changes were attempted every 100 MD steps. For the protonation state change attempt, the Metropolis energy was calculated using the generalized Born implicit solvent model (igb = 5 in Amber) with a 0.1 M salt concentration [25,26]. The intrinsic Born radii used the mbondi2 set [25] except for carboxylate oxygens in Asp and Glu, which were changed to 1.3 \AA to compensate for having two dummy protons present on each oxygen [27,28]. An in-house, modified version of the reference energies and a modified cpinutil.py script were used, as the ones present in Amber are intended for usage with the ff99SB force field. The reference energies were trained on model systems ACE-AS4-NME and ACE-GL4-NME using ff14SB and TIP3P force fields and were $31.578463 \text{ kcal/mol}$ for AS4 and $13.420729 \text{ kcal/mol}$ for GL4. Force field files for Glu and Asp containing dummy protons (ASH and GLH) were adapted from the versions available for ff99SB. In frmod.constph.ff14SB, AS4.ff14SB.off and GL4.ff14SB.off, the atom type of the $C\beta$ atom was changed from CT to 2C, matching the change from ff99SB to ff14SB. Those files are available on github.

During the constant-pH REMD simulation, residues Glu281, Asp830, Asp839, Asp843, Asp848, Glu554, Asp568, Asp571, Asp574, Glu583, Asp586, Asp614, and Glu619 were

allowed to change their protonation state. Attempts were performed every 100 MD steps. Each successful protonation change was followed by 100 steps of solvent relaxation [27].

At every 1000 MD steps, exchanges were attempted between the solution pH values of neighboring replicas.

All pH-REMD simulations were run for $9 \cdot 10^7$ steps or 180 ns at constant volume. These simulations were performed in triplicate to assess reproducibility.

It was confirmed that the protonation fractions of each of the 13 titratable residues in both conformations were reasonably converged within 180 ns of simulation, as illustrated in Figures S1 and S2.

As illustrated in the pH-REMD pH ladder in Figures S3 and S4, structures from different replicas sampled multiple solution pH values throughout the simulation.

The pKa values of each of the 13 residues were calculated using the Levenberg–Marquardt algorithm from `scipy.optimize.curve_fit` [29] by fitting pH-REMD protonation fraction data at different pH values to the Hill equation as shown in Equation (1).

$$f_{\text{protonation}} = \frac{1}{1 + 10^{-n(\text{pKa}-\text{pH})}} \quad (1)$$

where $f_{\text{protonation}}$ is the fraction of the total simulation that the titratable residue spent in a protonated state, and n is the Hill coefficient.

Backbone RMSDs and all atom-symmetry-corrected non-hydrogen RMSDs were calculated with `cpptraj` [14]. The clustering of pH-REMD simulations was performed using one of the two methods mentioned in the text. The first method is a hierarchical agglomerative (bottom-up) approach. It considered the symmetrical RMSD of all non-hydrogen atoms with a minimum distance between clusters of 4 Å. This was used for the PDB analysis, where the expected number of clusters was not known. The second method used the k-means clustering algorithm on CA atoms. It had a setup of two clusters, a sieve of three, and a randomized initial set of points [14]. This was used for the simulation analysis, where only two clusters were expected due to their use as initial coordinates. Visualizations were performed using VMD [30]. The Python Matplotlib library was used for plot visualization [31].

3. Results and Discussion

3.1. Experimental Structural Analysis

To test the hypothesis that the FPPR conformation is pH-dependent, all currently available SARS-CoV-2 trimeric ectodomain spike structures in the PDB (List S1) along with their reported experimental pH values were analyzed. To test the hypothesis that the compact FPPR conformation keeps the receptor binding domain (RBD) closed, the positions of the RBD domains above each FPPR were analyzed.

Out of the 700 analyzed spike structures, FPPR coordinates were present in 158 chains (see Methods for details). Structures in which the FPPR was not modeled were excluded from further analysis.

Clustering PDB files: The 158 FPPRs were clustered using the alpha carbon coordinates, and five clusters were obtained. The distribution of structures among the five clusters is shown in Table 1 and Figure 5. The largest cluster contained 136 FPPRs, all of which resemble a compact conformation (a representative structure for this cluster is PDB ID 7E7B [32]), as shown in Figure 2a. The second, smaller cluster consists of 12 FPPRs, and all structures in this cluster exhibited an extended conformation (a representative structure for the cluster is PDB ID 6XM0 [5]), as shown in Figure 2b. The third cluster includes six structures and is illustrated in Figure 2c (a representative structure is PDB ID 7LS9 [33]). The fourth and fifth clusters include three and one structures, respectively; the PDB validation report of the representative structure of the fourth cluster (PDB 7N9T) states that the modeled FPPR does not fit well into the experimental density [34–36], while the single entry for the fifth cluster (PDB 7TPL) has a missing RBD above the FPPR [37,38].

Table 1. FPPR cluster statistics: AvgDist—average RMSD between all pairs of structures in the cluster, and AvgCDist—average RMSD of this cluster to every other cluster.

Cluster	Structures #	AvgDist	AvgCDist
1	136	1.9	9.4
2	12	1.7	8.0
3	6	0.6	8.2
4	3	0.0	10.2
5	1	0.0	9.8

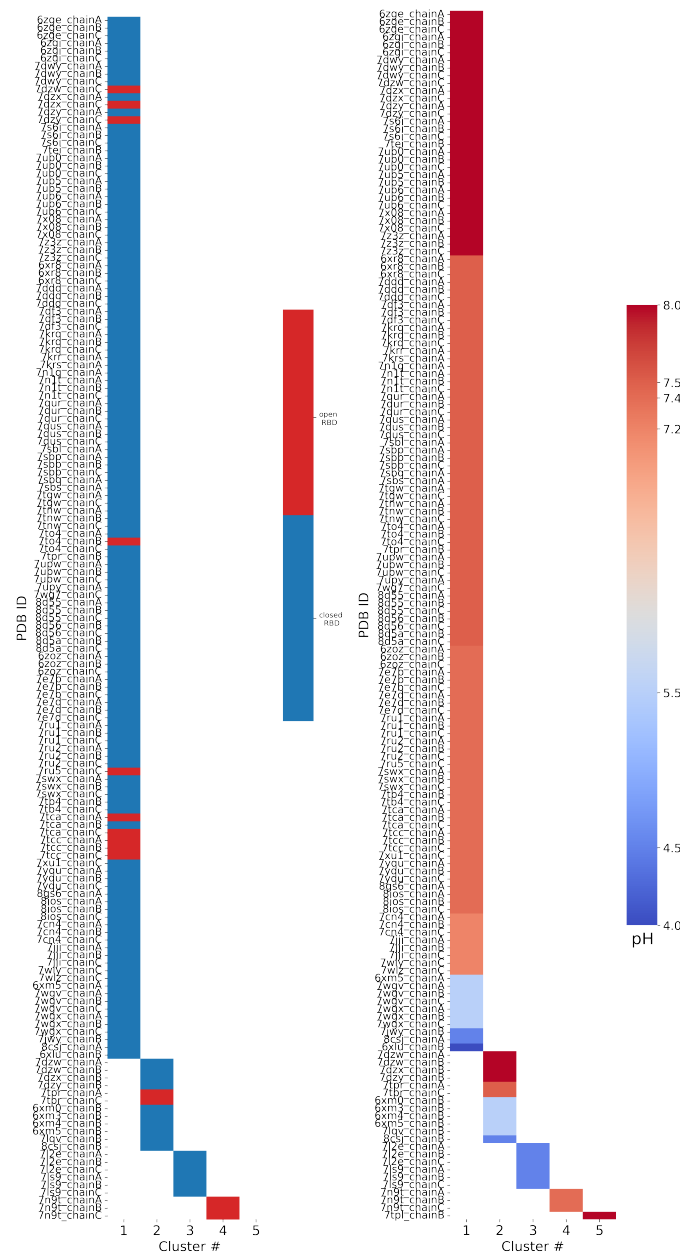


Figure 5. Results of FPPR cluster analysis applied to spike PDB structures. The PDB code and chain ID are given next to a colored box that is placed in the column according to the FPPR cluster number. The box color indicates the position of the RBD above the FPPR (left) or the pH of the experiment (right).

RBD–FPPR conformation coupling: We analyzed spike RBD positioning using the SpikeScape [13] tool. The majority of structures with a resolved FPPR have it positioned under the CTD1 connected to a closed RBD (142 of 158), as shown in Figure 5 left. Among these closed-RBD structures, 126 have a compact FPPR, 10 have an extended FPPR, and 6 have a conformation from the third cluster.

Among the 15 open-RBD structures with a resolved FPPR, 10 have a compact FPPR, 2 have an extended FPPR, and 3 exhibit the FPPR conformation from the fourth cluster. The remaining FPPR structure has an unresolved RBD. Notably, three of the ten structures featuring the compact conformation modeled under an open RBD (7DZW [39], 7DZX [40], and 7DZY [41,42]) exhibit steric clashes with the CTD1 region directly above the FPPR, as observed in the 3D structure (Figure S5) and PDB validation report.

With a closed RBD, the compact FPPR is more prevalent than the extended FPPR (126:10); under an open RBD, the preference for a compact FPPR relaxes somewhat (7:2 after removing those with clashes). An alternate view is to focus on the FPPR; above a compact FPPR, the RBD closed:open ratio (126:7) is somewhat higher than for the extended FPPR (10:2).

Although a compact FPPR appears more frequently than an extended FPPR when under a closed RBD, it remains that the compact FPPR conformation has been reported under an open RBD, casting doubt on the hypothesis that a compact FPPR conformation locks the RBD in a closed position. Moreover, among the open-RBD structures, the FPPR is more frequently modeled in the compact rather than the extended conformation. However, the scarcity of structures with a fully resolved FPPR positioned under an open RBD complicates drawing reliable conclusions regarding a relationship between the FPPR conformation and RBD position. We therefore shift focus to the role of the pH.

The pH–FPPR conformation coupling: To explore the relationship between FPPR and pH, the reported pH during structure determination was extracted from the PDB files. As shown in Figure 5 right, 136 of 158 structures report a neutral pH ranging from 7.2 to 8.0, while only 22 structures were resolved at pH values of 4.0 to 5.5. However, the two most populated FPPR clusters, representing the compact and extended conformations, are both present across the entire pH range from 4.0 to 8.0.

Interestingly, there are cases at both neutral and acidic pH values where the compact and extended FPPR conformations are present within the same spike structure (in different chains), despite the RBDs above these FPPRs all adopting the same open or closed position. Examples include 7DZX (pH 8), 7DZY (pH 8), 8CSJ [35] (pH 4.5), and 6XM5 [5] (pH 5.5). The sampling of alternate FPPR conformations in the same spike structure suggests a small free energy difference between FPPR conformations, even at different pH values. These observations undermine the hypothesis that the compact conformation is strongly favored at acidic pH. However, as with the relatively rare open-RBD structures with an ordered FPPR, the scarcity of spike structures determined at acidic pH makes it difficult to draw reliable conclusions.

In summary, analyzing all spike structures from the PDB yields results inconsistent with strong coupling between the FPPR conformation and either the experimental pH or RBD position, with both FPPR conformations observed in both neutral and acidic environments as well as under both open and closed RBDs (sometimes in the same experimental structure). However, the reliability of the analysis is reduced by the scarcity of structures determined at acidic pH or with the extended FPPR along with issues related to poor-quality regions of these cryo-EM structure models. Given these uncertainties, we explored avenues to supplement the PDB-based analysis.

Calculating pKa values for titratable side chains using PDB structures: If the equilibrium ratio of two protein conformations shows pH dependence, the titration curves of these two conformations should differ [43]. Comparison of calculated pKa values for different conformations has been used to rationalize pH-dependent conformational changes [44–48]. In the case of the FPPR, if the equilibrium preference between compact and extended FPPRs has pH dependence, there should be at least one residue with different pKa values in those

two conformations. The reference pKa values of Asp and Glu amino acids in water are 3.8 and 4.5, respectively. However, the pKa values in a protein could be quite different from these due to interactions with the local environment (e.g., formation of a salt bridge leads to a lower pKa). To test if the titration curves of compact and extended FPPRs are different, we calculated the pKa values of titratable residues in the FPPR region for both conformations.

The pKa values of titratable amino acids in the FPPR have been calculated previously [5]. It was reported that the pKa values of Asp830, Asp843, Asp574, Asp586, and Asp614 in the compact conformation were higher than those in the extended conformation, resulting in increased protonation of the FPPR region at pH 4.0. This result supports the hypothesis of pH-dependent FPPR conformations. However, the pKa calculations in that study were performed on static structures. Given the sensitivity of pKa values to local electrostatic interactions, it is possible that pKa values obtained using cryo-EM models may have larger uncertainties than those estimated using high-resolution crystal structures.

To assess the reproducibility of pKa calculations for the spike, pKa calculations were conducted here for multiple structures exhibiting each FPPR conformation. Specifically, two structures with the compact FPPR conformation and two structures with the extended FPPR conformation were selected, with the expectation that the differences in pKa values between spike structures exhibiting different FPPR conformations would be larger than those obtained from pairs of structures adopting the same FPPR conformation. For the extended conformation, coordinates from PDB entries 6XM0 [5] (determined at pH 5.5) and 7LQV [49] (pH 5.5) were used, while for the compact conformation, 6XLU [5] (pH 4) and 7WGX [50] (pH 5.5) were used.

To measure the pKa values of titratable residues, the widely used software PROPKA3.5.1 [15,16] was utilized. This software uses a physics-based model with empirical adjustments to calculate the pKa values for a static structure. The resulting pKa values are presented in Table 2. For the extended FPPR conformation, pKa values from different spike structures agree within 1 pH unit. In contrast, the pKa values in the compact conformation for Glu583, Asp614, and Asp843 display differences of more than 1.5 units between two structures adopting the same FPPR conformation. Such discrepancies might be attributed to the local differences in structures. For example, in structure 6XLU, Asp843 is near Asp586 (see Figure S6), leading to stronger preference for the protonated acid and an increased pKa of 6.0. Conversely, in 7WGX where the carboxyl group is solvent-exposed, the pKa returns to 3.8.

Table 2. The pKa values from PROPKA3 for 13 titratable residues in example PDB structures adopting different FPPR conformations. Residues with a difference of 1.5 units or more for pKa values within the same conformation are colored red.

Residue	FPPR Chain	Extended Conformation		Compact Conformation	
		6XM0	7LQV	6XLU	7WGX
Glu 281	B	3.6	4.0	4.1	4.0
Glu 554	A	4.2	3.9	4.6	3.9
Asp 568	A	4.7	4.9	6.1	5.9
Asp 571	A	4.3	4.9	4.8	5.1
Asp 574	A	6.0	6.9	5.4	5.5
Glu 583	A	4.9	5.3	4.9	3.4
Asp 586	A	4.7	5.0	5.0	4.9
Asp 614	A	5.5	5.8	7.2	4.9
Glu 619	A	4.9	4.9	4.2	3.6
Asp 830	B	4.2	4.6	4.5	4.1
Asp 839	B	3.9	3.8	3.6	4.5
Asp 843	B	4.2	4.3	6.0	3.8
Asp 848	B	4.4	4.5	4.9	3.8

This analysis highlights how the specific choice of spike model from the PDB significantly affects pKa values even when the backbone conformation is highly similar, making the quantitative comparison of pKa values between different conformations unreliable.

3.2. Constant-pH REMD Simulations

As demonstrated above, pKa calculations from static structures are sensitive to the details of the input experimental model. In contrast, molecular dynamics simulations are widely used to explore flexibility in biomolecular systems by sampling a dynamic ensemble in the context of an all-atom model with explicit water [51]. We therefore explored whether atomistic molecular dynamics simulations could provide estimates of pKa values for acidic side chains in the FPPR region. MD simulations can sample local flexibility, which we hypothesized would lead to more precise pKa values. We also assess whether simulations are capable of directly sampling the interconversion between FPPR conformations or show a preference for compact vs. extended FPPRs under different pH conditions.

Traditional MD simulations employ a fixed protonation state for each titratable group that is set prior to the simulation (typically using the amino acid reference pKas or, less often, predictions from a static pKa calculation). Thus, protonation states during these simulations will not be conformation-dependent.

A recent addition to the MD toolbox is constant-pH molecular dynamics (cpHMD), which adds the ability to maintain a constant system pH rather than a protonation state. CpHMD has been used to estimate pKa values with explicit inclusion of conformational dynamics and to explore the coupling of a protonation state and a conformational equilibrium [46,52,53]. In the discrete protonation variant of cpHMD that is used here, protonation state changes are attempted at intervals during the simulation, with the probability of accepting the change depending on the desired pH, the reference energy for the titratable amino acid, and the electrostatic energy of the system before and after the trial addition or removal of the proton [26]. Performing cpHMD simulations over a range of pH values provides titration curves for each group, from which pKa values can be estimated. A more advanced version of the method, pH replica exchange MD (pH-REMD) [54–56], allows simulations at different pH values to exchange conformations, leading to increased conformational sampling and better convergence. During pH-REMD, a series of coupled simulations (“replicas”) at different pH values are performed. At intervals, replicas can swap solution pH values, with the probability of accepting the swap based on the pH values and number of protons in each conformation. We used pH-REMD here to estimate pKa values for the FPPR region.

Achieving proper statistical significance requires extensive sampling in both structural and protonation spaces. However, the size of the trimeric spike in explicit water exceeds 1 million atoms [57], for which convergence of protonation states and conformational ensembles becomes intractable with current computational resources. Therefore, we opted to focus on a smaller model of the spike by including a single FPPR and any portion of the spike within 20 Å (see Methods).

Two simulation systems were built in an explicit solvent, including the compact and extended FPPR. Thirteen amino acids were designated as titratable; these matched the ones examined for the static pKa calculations described above in Table 2. Eight equally spaced replicas, spanning pH values from 1 to 8 in unit increments, were generated. Although the pH of biological interest ranges from 4 to 7, the range from 1 to 8 was selected as a ladder for better sampling and titration curve fitting. Then, pH-REMD simulations of 180 ns were generated using an initial structure with either the compact (PDB ID 6XLU [5]) or extended (6XM0 [5]) FPPR conformation. For each conformation, pH-REMD simulations were repeated in triplicate to assess the reproducibility of the calculated pKa values.

No transitions between FPPR conformations are observed: Ideally, frequent transitions between the compact and extended FPPR conformations could be sampled to provide an estimate of the equilibrium constant for the FPPR conformational change at various pH values. However, no transitions between compact and extended conformations were sampled at any pH throughout the 180 ns pH-REMD simulation. This was confirmed by comparing the FPPR backbone RMSD during time to both experimental conformations. The initially extended FPPR simulation exhibited a low RMSD to the starting structure, with most backbone RMSD values ranging between 1 and 3 Å regardless of the pH (Figures 6 and S7). When compared to the compact FPPR, the RMSD for this initially extended simulation was much larger, ranging between 6 Å and 8 Å as shown in Figure 7. This indicates that pH-REMD does not sample the compact FPPR conformation when started in the extended conformation. The initially compact FPPR simulation exhibited more variability, with backbone RMSD values for the compact FPPR generally remaining below 4 Å but occasionally spiking up to 6 Å, as shown in Figures 6 and S7. However, no transitions to the extended FPPR were sampled, with RMSD values for the initially compact simulations remaining above 6–8 Å. In summary, simulations starting from either FPPR conformation remained relatively stable at all pH values, and no transition to the other FPPR conformation was sampled. This slow rearrangement is consistent with the trapping of alternate FPPR conformations in some cryo-EM spike structures.

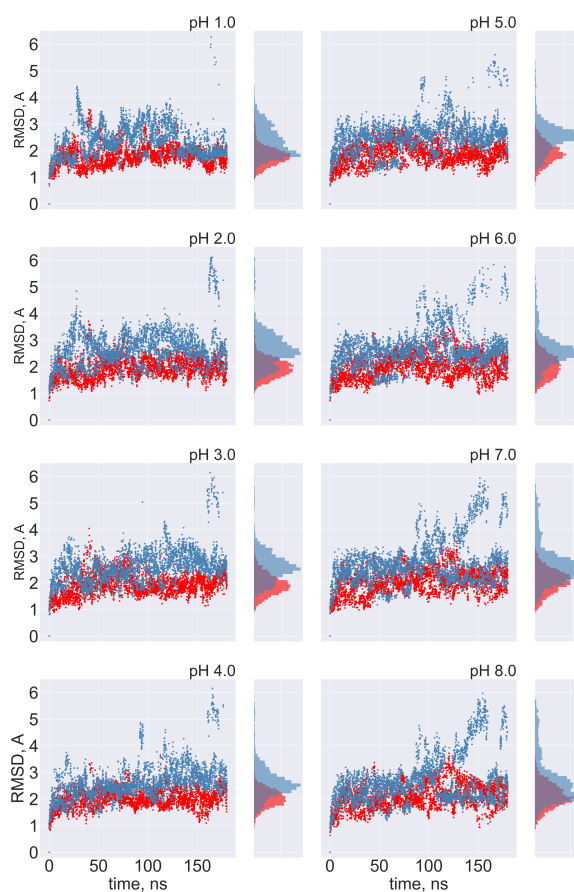


Figure 6. The FPPR remains in the initial conformation during simulations of extended and compact FPPRs. FPPR RMSD values are shown at different pH values during two pH-REMD simulations, each initiated with a different FPPR conformation. Red symbols—initially extended simulation, with backbone RMSD calculated using the experimental extended conformation; blue symbols—initially compact simulation, with backbone RMSD calculated using the experimental compact conformation. Low values indicate that the initial FPPR conformation is stable.

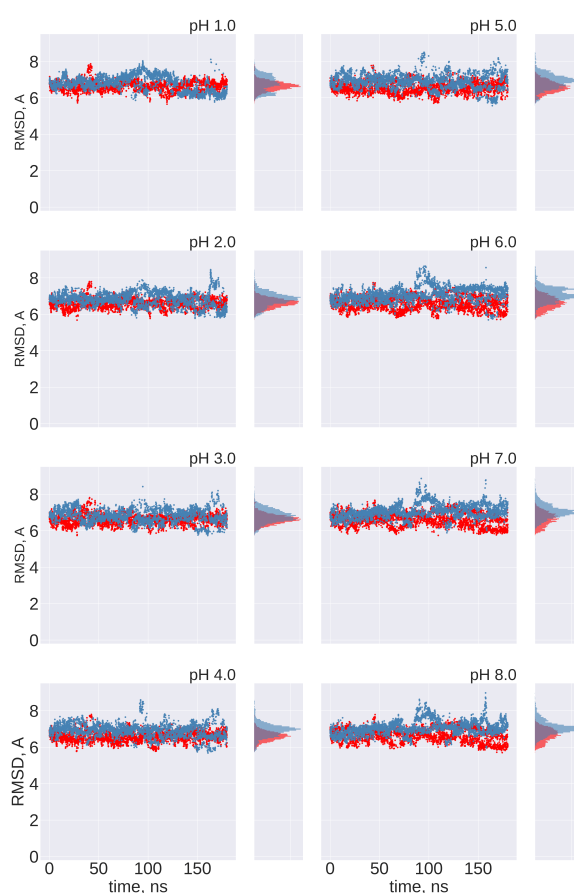


Figure 7. Cross-comparison of FPPR RMSD values as a metric for conformational transitions. FPPR RMSD values vs. time are shown during two pH-REMD simulations initiated with different FPPR conformations. In each, the RMSD is calculated using the *other* FPPR conformation as the reference structure. Red symbols—initially extended, with backbone RMSD calculated relative to the experimental compact conformation; blue symbols—initially compact, with backbone RMSD calculated relative to the experimental extended conformation. Large values at all pH ranges indicate that no simulation samples a transition to the alternate FPPR conformation.

To confirm the results of the RMSD analysis, trajectory frames from both simulations across all pH ranges were combined and clustered using a hierarchical agglomerative (bottom-up) approach using all non-hydrogen FPPR atoms (see Methods). The simulations starting from the two different FPPR conformations share no clusters, regardless of the pH (Figure S8). Frames from the simulations starting in the extended conformation fall into a single cluster, while frames from the initially compact conformation form two clusters, with a dominant cluster that is similar to the compact conformation. These observations confirm that no transition between conformations happens within 180 ns at any pH value.

Since both FPPR conformations appear to be kinetically trapped during the 180 ns pH-REMD simulations, we extended one of the three pH-REMD runs to 900 ns for both FPPR conformations. Even with the longer timescale, no transitions between conformations were sampled. This precludes the direct calculation of a pH-dependent equilibrium constant for the FPPR conformational change.

However, the lack of transitions between different FPPR conformations allows us to assign the pH-REMD titration profiles uniquely to a single FPPR conformation type. We therefore exploited the lack of conformational switching to compare the protonation behavior of the alternate FPPR conformations at each simulated pH, focusing on endosomal pH values of 4.0 to 6.5.

Calculating the pKa values for each FPPR conformation via pH-REMD: To test the hypothesis that the equilibrium distribution is pH-dependent, the titration curves of the titratable residues in the two FPPR conformations were compared. If the equilibrium distribution is indeed pH-dependent, the set of titration curves should differ [43]. Here, the most important residues are those exhibiting *increased* pKa values compared to the reference pKa, since this would lead to protonation changes when moving from extracellular to endosomal pH. Downshifted pKa values would be less likely to affect the FPPR over the biologically relevant pH range.

The titration curves of the majority of the 13 acidic side chains are similar between FPPR conformations, and the pKa values deviate by no more than 1 unit from the reference pKa values, as illustrated in Tables 3 and S2 as well as Figure S9.

Table 3. The pKa values of the 13 acidic side chains in the FPPR region, calculated from pH-REMD simulations. Only two have significantly different pKas in the different FPPR conformations (highlighted in red). Four residues have pKas that are shifted significantly away from the reference values, but the shifts are similar in the different conformations (indicated in italics). Error bars were obtained from independent simulations of each FPPR conformation type.

Residue	Chain	Extended	Compact
		pKa	pKa
Glu 281	B	3.8 ± 0.0	3.7 ± 0.3
<i>Glu 554</i>	A	3.0 ± 0.1	2.8 ± 0.1
<i>Asp 568</i>	A	1.4 ± 0.1	1.1 ± 0.3
Asp 571	A	3.2 ± 0.0	3.8 ± 0.2
<i>Asp 574</i>	A	<i>negative</i>	<i>negative</i>
Glu 583	A	3.6 ± 0.0	3.2 ± 0.1
<i>Asp 586</i>	A	<i>negative</i>	<i>negative</i>
Asp 614	A	3.8 ± 0.2	3.7 ± 0.4
Glu 619	A	4.7 ± 0.0	4.5 ± 0.1
Asp 830	B	4.9 ± 0.1	5.4 ± 0.5
Asp 839	B	1.6 ± 0.1	2.9 ± 0.2
Asp 843	B	4.4 ± 0.3	4.3 ± 0.5
Asp 848	B	4.5 ± 0.2	2.7 ± 0.1

The only amino acid whose pKa value exhibits an upward shift is Asp830. However, this shift occurs in both conformations, suggesting minimal contribution to a pH-dependent conformational switch. Other pKa values also are shifted downward from the reference values, but the shifts again are similar for compact and extended FPPRs. For example, Asp574 and Asp586 demonstrate significant shifts in pKa values towards the negative range in both conformations. The pKa values of Asp568 and Glu554 show a downward shift as well. All four are located above the FPPR in the CTD1 domain, as shown in Figure 4.

Only Asp848 and Asp839 exhibit distinct titration curve patterns between the compact and extended FPPRs, as shown in Figure 8. In the compact conformation, Asp848 displays a downshifted pKa of 2.7 ± 0.1, while in the extended conformation, the pKa is 4.5 ± 0.2. Asp839 is downshifted in both FPPR conformations (2.9 ± 0.2 in the compact conformation and even lower at 1.6 ± 0.1 in the extended conformation), as seen in Figure 8.

We next explored a structural basis for these calculated pKa shifts. In the compact conformation, the positioning of Asp848 at the N-terminal end of the helix (Ncap), as shown in Figure 4, may contribute to the observed lower pKa value due to interaction with the helix macrodipole [58]. In contrast, in the extended conformation, Asp848 is not situated at the N-terminal of the helix, resulting in the pKa being relatively unchanged from the reference value.

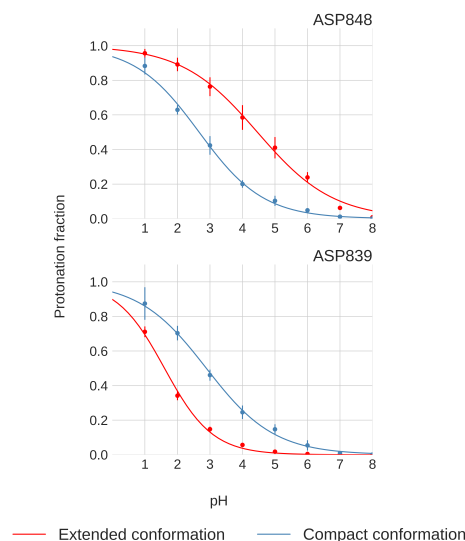


Figure 8. Titration curves of Asp848 and Asp839 from pH-REMD simulations. Error bars are calculated from independent simulations of that FPPR conformation.

For Asp839, a downward shift in pKa occurs in both conformations, with a more pronounced shift in the extended conformation. The presence of Lys835 nearby, as shown in Figure 4, may be the source of this shift, indicating a potential salt bridge formation. The reason for the magnified pKa shift in the extended conformation is less clear; however, since the pKa value is well below 4 and thus protonation is unlikely to play a biological role, we did not investigate further.

The presence of differences between individual pKa values across conformations suggests that the equilibrium distribution between compact and extended conformations might depend on the pH. Despite these pH-dependent variations in specific residues, however, the total protonation of the 13 acidic side chains remains consistent across the two different conformations over the pH range of 4.0 to 6.5, as shown in Figure 9 and Table S3. This finding suggests that the total protonation for both FPPR conformations varies with pH but is not influenced by conformation. This observation opposes the initial hypothesis proposing a pH-dependent equilibrium between FPPR conformations. MD simulations in the absence of experimental data should be treated with caution; however, these conclusions from our pH-REMD simulations are consistent with our PDB-based analysis of experimental FPPR conformations at different pH values.

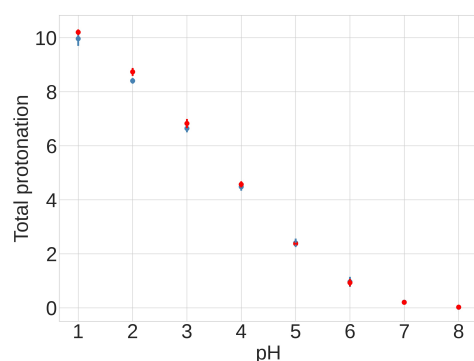


Figure 9. Results from pH-REMD simulations showing the relationship between the total protonation of titratable groups and pH. The Y-axis shows the average number of protonated acidic groups, with a maximum of 13 (total number of Asp and Glu titrated): red—extended FPPR conformation; blue—compact FPPR conformation. Error bars are from independent simulations of the same FPPR conformation type.

Forced competition between FPPR conformations at different pH values: The pH-REMD simulations described above involved initiation of all replicas in the same FPPR conformation (in the independent runs, either all extended or all compact). No transitions to the other conformation were sampled, which permitted calculation of the pKa curves for the initial conformation as shown in Figure 8. However, inclusion of a single FPPR conformation type precluded observation of a preference for one conformation over another at a given pH. To overcome this obstacle, we initiated pH-REMD simulations in which both FPPR conformations were represented in the initial coordinates for a single pH-REMD run. After evolving the simulation for a period of time, the ratio of the two conformations is calculated at each pH. If one of the FPPR conformations is preferred at a higher (or lower) pH than the other, we expect the REMD exchange process to lead to uneven sorting of conformations across the pH ladder and a pH-dependent conformational preference to be apparent.

We initiated a 180 ns pH-REMD simulation with eight replicas as before, but in this case, the replicas that were initially at pH 1 to 4 started in the compact FPPR, while replicas at pH 5 to 8 started in the extended FPPR conformation. We analyzed the pH preference for each conformation by clustering the trajectory at each pH into two clusters using the k-means algorithm [14] (see Table 4, Figure S10). The two conformations are similarly populated within the uncertainty range except at the extreme of pH 1. No trend is seen to favor either conformation as a function of pH. These data are consistent with the similar titration curves obtained from the separate pH-REMD calculations for each FPPR conformation. Overall, the simulations suggest a very similar free energy for each FPPR conformation at the biologically relevant pH range.

Table 4. Data from pH-REMD with forced competition between different FPPR conformations. The table provides the fractions of simulation time spent in the compact and extended conformations at different pH values. Error bars were obtained from independent simulations.

pH	Fraction of Compact	Fraction of Extended
1	0.31 ± 0.09	0.69 ± 0.09
2	0.53 ± 0.04	0.47 ± 0.04
3	0.63 ± 0.04	0.37 ± 0.04
4	0.52 ± 0.09	0.48 ± 0.09
5	0.42 ± 0.13	0.58 ± 0.13
6	0.44 ± 0.06	0.56 ± 0.06
7	0.56 ± 0.15	0.44 ± 0.15
8	0.60 ± 0.19	0.40 ± 0.19

4. Conclusions

To test the hypothesis that the fusion peptide proximal region in the SARS-CoV-2 spike glycoprotein exhibits pH dependence and controls pH-dependent RBD positioning, a comprehensive analysis of existing experimental structures was conducted and was supplemented by new simulation data. Initial efforts focused on experimental structural analysis; the results did not support a difference in FPPR conformation as a function either of the experimental pH or RBD position, although the scarcity of spike structures determined at low pH or with the extended FPPR left uncertainty. The extensive use of artificially stabilized spike constructs for structure determination may also be a complicating factor in analyzing spike PDB structures; very few structures correspond to a biologically relevant spike sequence.

Calculating the pKa values of individual residues in selected experimental spike structures, with the aim of revealing distinct pKa values between conformations, yielded results that were poorly reproducible across different spike PDB structures with the same FPPR conformation type. The pKa differences between different examples of the same FPPR conformation were comparable to the pKa differences between different FPPR conformations. The poor reproducibility of pKa values from static spike structures is likely attributable to

the relatively low resolution of the cryo-EM structures coupled with the sensitivity of pKa to local structure details and dynamics.

In order to estimate pKa values in the context of a dynamic ensemble, molecular dynamics simulations were conducted across a range of pH levels utilizing pH-REMD. Even after extending the pH-REMD to 1 microsecond, transitions between FPPR conformations were not observed at any pH due to the slow timescale of FPPR reorganization. This prevented the direct estimation of the free energy difference between conformations. However, it enabled the calculation of distinct pKa values for titratable side chains in the alternate FPPR conformations.

The pH-REMD simulations revealed divergent pKa values for two residues, Asp 848 and Asp 839, in different FPPR conformations. However, the total protonation showed no significant dependence on conformation, opposing the hypothesis of a pH-dependent equilibrium between conformations. Similarly, when pH-REMD simulations were initiated using both conformations in a single simulation, no preference was observed for either conformation as a function of pH.

In this study, the highly flexible glycans were omitted from the model due to large uncertainties in their positions and their variable composition. The simulation results support conclusions that we drew from analyses of experimental data in which glycans were included and which suggested that the glycans do not play an important role in modulating the pH-dependent FPPR behavior. However, more work is certainly needed to explore the varied and complex effects of spike glycosylation.

In conclusion, the comprehensive analysis presented here, including a meta-analysis of experimental conditions and structure correlations as well as all-atom constant pH simulations, suggests that the FPPR is unlikely to be a pH-sensitive switch controlling the RBD position in the SARS-CoV-2 spike. However, it is important to note that other parts of the spike protein might be pH-sensitive, as with the influenza hemagglutinin protein.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v16071066/s1>, List S1 : List of 700 SARS-CoV-2 full spike protein (Uniprot ID P0DTC2) cryo-EM structures with resolution 4 Å or better retrieved from the PDB on 08/3/2023; Table S1: pKa values of 13 titratable residues of a Spike fragment calculated using PROPKA3 using PDB files in both FPPR conformations; Figure S1: Time series of the cumulative protonated fraction of 13 titratable residues in initially compact conformation in pH-REMD. Simulations were repeated in triplicate; Figure S2: Time series of the cumulative protonated fraction of 13 titratable residues in initially extended conformation in pH-REMD. Simulations were repeated in triplicate; Figure S3: Time series of the pH for individual replica trajectories in pH-REMD in extended conformation. Simulations were repeated in triplicate; Figure S4: Time series of the pH for individual replica trajectories in pH-REMD in compact conformation. Simulations were repeated in triplicate; Figure S5: FPPR in compact conformation in PDB 7DZW. FPPR (chain C) is shown with blue ribbon and other regions use orange. The FPPR has significant clash with the CTD1 domain above it; Figure S6: FPPR in compact conformation in PDB 6XLU (left) and 7WGX (right). FPPR (chain B) is shown with red ribbon and other regions use gray. In 6XLU Asp843 is surrounded by Asp586, Thr553 and Thr588 and has a pKa of 6.0. In 7WGX Asp843 is surrounded by Asn556 and has a pKa of 3.8; Figure S7: Time series of the RMSD for FPPR backbone atoms with a histogram on the right. Data are shown at different pH values during two pH-REMD simulations, each initiated with a different FPPR conformation. Red symbols—initially extended simulation, with backbone RMSD calculated using the experimental extended conformation; blue symbols—initially compact simulation, with backbone RMSD calculated using to the experimental compact conformation. Low values indicate that the initial FPPR conformation is stable at all pH values. 3 repeats; Figure S8: Trajectory frames from both simulations across all pH ranges were combined and clustered with hierarchical clustering using symmetry-corrected RMSD for all non-hydrogen FPPR atoms. The fractional population for each cluster is color-coded. Each row sums to 1.0. Frames from simulations starting in the extended conformation fall into a single cluster, while frames from the initially-compact conformation form two clusters; Figure S9: Titration curves of 13 titratable residues in two different FPPR conformations from pH-REMD simulations. Red—extended conformation, blue—compact conformation. Error bars are obtained from the independent simulations; Table S2: Mean pKa values

and Hill coefficients with standard deviation for 13 titratable residues in both conformations from 3 repeats of pH-REMD simulations. Red highlights 2 residues that have different pKas between the two conformations; Table S3: Average protonation of 13 titratable residues in 3 repeats of pH-REMD simulations of compact and extended conformations; Figure S10: Time series of the fraction of each conformation at different pH values along the pH-REMD simulation started with replicas of both conformations.

Author Contributions: Conceptualization, D.S., Y.W. and C.S.; performed the research, D.S., Y.W. and C.S.; wrote the article, D.S. and C.S.; reviewed and edited the manuscript, D.S., Y.W. and C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Research Corporation for Science Advancement (COVID Initiative grant #27350). We gratefully acknowledge support from Henry and Marsha Laufer.

Data Availability Statement: Files frcmod.constph.ff14SB, AS4.ff14SB.off and GL4.ff14SB.off are publicly available to facilitate further research via the following link <https://github.com/csimmerling/ff14SB-cphmd>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Lu, M.; Uchil, P.D.; Li, W.; Zheng, D.; Terry, D.S.; Gorman, J.; Shi, W.; Zhang, B.; Zhou, T.; Ding, S.; et al. Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles. *Cell Host Microbe* **2020**, *28*, 880–891.e8. [[CrossRef](#)] [[PubMed](#)]
2. Sztain, T.; Ahn, S.H.; Bogetti, A.T.; Casalino, L.; Goldsmith, J.A.; Seitz, E.; McCool, R.S.; Kearns, F.L.; Acosta-Reyes, F.; Maji, S.; et al. A glycan gate controls opening of the SARS-CoV-2 spike protein. *Nat. Chem.* **2021**, *13*, 963–968. [[CrossRef](#)] [[PubMed](#)]
3. Shah, M.; Ahmad, B.; Choi, S.; Woo, H.G. Mutations in the SARS-CoV-2 spike RBD are responsible for stronger ACE2 binding and poor anti-SARS-CoV mAbs cross-neutralization. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3402–3414. [[CrossRef](#)] [[PubMed](#)]
4. Wrapp, D.; Wang, N.; Corbett, K.S.; Goldsmith, J.A.; Hsieh, C.L.; Abiona, O.; Graham, B.S.; McLellan, J.S. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020**, *367*, 1260–1263. [[CrossRef](#)] [[PubMed](#)]
5. Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G.Y.; Katsamba, P.S.; Sampson, J.M.; Schön, A.; Bimela, J.; 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*, 867–879.e5. [[CrossRef](#)] [[PubMed](#)]
6. Di Lella, S.; Herrmann, A.; Mair, C.M. Modulation of the pH Stability of Influenza Virus Hemagglutinin: A Host Cell Adaptation Strategy. *Biophys. J.* **2016**, *110*, 2293–2301. [[CrossRef](#)]
7. Russell, C.J. Acid-Induced Membrane Fusion by the Hemagglutinin Protein and Its Role in Influenza Virus Biology. In *Current Topics in Microbiology and Immunology*; Springer International Publishing: Berlin/Heidelberg, Germany, 2014; pp. 93–116. [[CrossRef](#)]
8. Qu, K.; Chen, Q.; Ciazynska, K.A.; Liu, B.; Zhang, X.; Wang, J.; He, Y.; Guan, J.; He, J.; Liu, T.; et al. Engineered disulfide reveals structural dynamics of locked SARS-COV-2 spike. *PLoS Pathog.* **2022**, *18*, e1010583. [[CrossRef](#)] [[PubMed](#)]
9. Zhang, J.; Cai, Y.; Xiao, T.; Peng, H.; Sterling, S.; Walsh, R., Jr.; Rawson, S.; Volloch, S.; Chen, B. Distinct Conformational States of SARS-CoV-2 Spike Protein. 2020. Available online: <https://doi.org/10.2210/pdb6xr8/pdb> (accessed on 14 May 2024).
10. Zhou, T.; Tsybovsky, Y.; Olia, A.; Kwong, P. Consensus Structure of SARS-CoV-2 Spike at pH 5.5. 2020. Available online: <https://doi.org/10.2210/pdb6xm0/pdb> (accessed on 14 May 2024).
11. Cai, Y.; Zhang, J.; Xiao, T.; Peng, H.; Sterling, S.M.; Walsh, R.M.; Rawson, S.; Rits-Volloch, S.; Chen, B. Distinct conformational states of SARS-CoV-2 spike protein. *Science* **2020**, *369*, 1586–1592. [[CrossRef](#)] [[PubMed](#)]
12. Olia, A.S.; Tsybovsky, Y.; Chen, S.J.; Liu, C.; Nazzari, A.F.; Ou, L.; Wang, L.; Kong, W.P.; Leung, K.; Liu, T.; et al. SARS-CoV-2 S2P spike ages through distinct states with altered immunogenicity. *J. Biol. Chem.* **2021**, *297*, 101127. [[CrossRef](#)]
13. Stepanenko, D.; Budhan, S.; Simmerling, C. SpikeScape: A tool for analyzing structural diversity in experimental structures of the SARS-COV-2 spike glycoprotein. *J. Chem. Inf. Model.* **2023**, *63*, 1087–1092. [[CrossRef](#)]
14. Roe, D.R.; Cheatham, T.E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095. [[CrossRef](#)]
15. Olsson, M.H.M.; Søndergaard, C.R.; Rostkowski, M.; Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537. [[CrossRef](#)]
16. Søndergaard, C.R.; Olsson, M.H.M.; Rostkowski, M.; Jensen, J.H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295. [[CrossRef](#)]
17. Schrödinger, LLC. *The PyMOL Molecular Graphics System*; Version 2.5.0; Schrödinger, LLC: New York, NY, USA, 2015.
18. Casalino, L.; Gaieb, Z.; Goldsmith, J.A.; Hjorth, C.K.; Dommer, A.C.; Harbison, A.M.; Fogarty, C.A.; Barros, E.P.; Taylor, B.C.; McLellan, J.S.; et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722–1734. [[CrossRef](#)]
19. Grant, O.C.; Montgomery, D.; Ito, K.; Woods, R.J. Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. *Sci. Rep.* **2020**, *10*, 14991. [[CrossRef](#)]

20. Watanabe, Y.; Allen, J.D.; Wrapp, D.; McLellan, J.S.; Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**, *369*, 330–333. [[CrossRef](#)]
21. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [[CrossRef](#)]
22. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [[CrossRef](#)]
23. Case, D.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham, T., III; Cruzeiro, V.; Darden, T.; Duke, R.; Giambasu, G.; et al. *Amber 2020*; University of California: San Francisco, CA, USA, 2020.
24. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [[CrossRef](#)]
25. Onufriev, A.; Bashford, D.; Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins Struct. Funct. Bioinform.* **2004**, *55*, 383–394. [[CrossRef](#)]
26. Mongan, J.; Case, D.A.; McCammon, J.A. Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048. [[CrossRef](#)]
27. Swails, J.M.; York, D.M.; Roitberg, A.E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J. Chem. Theory Comput.* **2014**, *10*, 1341–1352. [[CrossRef](#)]
28. Yeager, A.V.; Swails, J.M.; Miller, B.R.I. Improved Accuracy for Constant pH-REMD Simulations through Modification of Carboxylate Effective Radii. *J. Chem. Theory Comput.* **2017**, *13*, 4624–4635. [[CrossRef](#)] [[PubMed](#)]
29. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [[CrossRef](#)]
30. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [[CrossRef](#)]
31. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [[CrossRef](#)]
32. Ma, J.; Su, D.; Sun, Y.; Huang, X.; Liang, Y.; Fang, L.; Ma, Y.; Li, W.; Liang, P.; Zheng, S. Cryo-electron Microscopy Structure of S-Trimer, a Subunit Vaccine Candidate for COVID-19. *J. Virol.* **2021**, *95*, e00194-21. [[CrossRef](#)]
33. Cerutti, G.; Rapp, M.; Guo, Y.; Bahna, F.; Bimela, J.; Reddem, E.R.; Yu, J.; Wang, P.; Liu, L.; Huang, Y.; et al. Structural basis for accommodation of emerging B.1.351 and B.1.1.7 variants by two potent SARS-CoV-2 neutralizing antibodies. *Structure* **2021**, *29*, 655–663.e4. [[CrossRef](#)]
34. Sun, D.; Sang, Z.; Kim, Y.J.; Xiang, Y.; Cohen, T.; Belford, A.K.; Huet, A.; Conway, J.F.; Sun, J.; Taylor, D.J.; et al. Potent neutralizing nanobodies resist convergent circulating variants of SARS-COV-2 by targeting diverse and conserved epitopes. *Nat. Commun.* **2021**, *12*, 4676. [[CrossRef](#)]
35. Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. *Nat. Struct. Mol. Biol.* **2003**, *10*, 980. [[CrossRef](#)]
36. Huang, W.; Taylor, D. CryoEM Structure of SARS-CoV-2 Spike in Complex with Nb17. 2021. Available online: <https://doi.org/10.2210/pdb7n9t/pdb> (accessed on 14 May 2024).
37. Gobeil, S.M.C.; Henderson, R.; Stalls, V.; Janowska, K.; Huang, X.; May, A.; Speakman, M.; Beaudoin, E.; Manne, K.; Li, D.; et al. Structural diversity of the SARS-CoV-2 Omicron spike. *Mol. Cell.* **2022**, *83*, 2050–2068.e6. [[CrossRef](#)]
38. Gobeil, S.; Acharya, P. Delta (B.1.617.2) SARS-CoV-2 Variant Spike Protein (S-GSAS-Delta) in the M1 Conformation, D4. 2022. Available online: <https://doi.org/10.2210/pdb7tpl/pdb> (accessed on 14 May 2024).
39. Liu, Y.; Soh, W.; Li, S.; Kishikawa, J.; Hirose, M.; Kato, T.; Standley, D.; Okada, M.; Arase, H. Apo Spike Protein from SARS-CoV2. 2021. Available online: <https://doi.org/10.2210/pdb7dzw/pdb> (accessed on 14 May 2024).
40. Liu, Y.; Soh, W.; Li, S.; Kishikawa, J.; Hirose, M.; Kato, T.; Standley, D.; Okada, M.; Arase, H. Spike Protein from SARS-CoV2 with Fab Fragment of Enhancing Antibody 8D2. 2021. Available online: <https://doi.org/10.2210/pdb7dzy/pdb> (accessed on 14 May 2024).
41. Liu, Y.; Soh, W.; Li, S.; Kishikawa, J.; Hirose, M.; Kato, T.; Standley, D.; Okada, M.; Arase, H. Spike Protein from SARS-CoV2 with Fab Fragment of Enhancing Antibody 2490. 2021. Available online: <https://doi.org/10.2210/pdb7dzy/pdb> (accessed on 14 May 2024).
42. Liu, Y.; Soh, W.T.; Tada, A.; Arakawa, A.; Matsuoka, S.; Nakayama, E.E.; Li, S.; Ono, C.; Torii, S.; Kishida, K.; et al. An infectivity-enhancing site on the SARS-CoV-2 spike protein is targeted by COVID-19 patient antibodies. *BioRxiv* **2020**. [[CrossRef](#)]
43. Tanford, C. Ionization-linked changes in protein conformation. I. theory. *J. Am. Chem. Soc.* **1961**, *83*, 1628–1634. [[CrossRef](#)]
44. Chaudhury, S.; Ripoll, D.R.; Wallqvist, A. Structure-based pKa prediction provides a thermodynamic basis for the role of histidines in pH-induced conformational transitions in dengue virus. *Biochem. Biophys. Rep.* **2015**, *4*, 375–385. [[CrossRef](#)]
45. Ingemar Olsen, J.; Sauer, S.P.A.; Marcus Pedersen, C.; Bols, M. Exploring the relationship between the conformation and pKa: Can a pKavalue be used to determine the conformational equilibrium? *Org. Biomol. Chem.* **2015**, *13*, 3116–3121. [[CrossRef](#)]
46. Di Russo, N.V.; Estrin, D.A.; Martí, M.A.; Roitberg, A.E. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental pKas: The Case of Nitrophorin 4. *PLoS Comput. Biol.* **2012**, *8*, e1002761. [[CrossRef](#)]
47. Di Russo, N.V.; Martí, M.A.; Roitberg, A.E. Underlying Thermodynamics of pH-Dependent Allostery. *J. Phys. Chem. B* **2014**, *118*, 12818–12826. [[CrossRef](#)]

48. Beychok, S.; Steinhardt, J. Increase in Acid-binding Sites on Denaturation of Horse Ferrihemoglobin at 0°1. *J. Am. Chem. Soc.* **1959**, *81*, 5679–5687. [[CrossRef](#)]
49. Cerutti, G.; Guo, Y.; Zhou, T.; Gorman, J.; Lee, M.; Rapp, M.; Reddem, E.R.; Yu, J.; Bahna, F.; Bimela, J.; et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **2021**, *29*, 819–833.e7. [[CrossRef](#)]
50. Zhao, M.M.; Zhu, Y.; Zhang, L.; Zhong, G.; Tai, L.; Liu, S.; Yin, G.; Lu, J.; He, Q.; Li, M.J.; et al. Novel cleavage sites identified in SARS-CoV-2 spike protein reveal mechanism for cathepsin L-facilitated viral infection and treatment strategies. *Cell Discov.* **2022**, *8*, 53. [[CrossRef](#)]
51. Brini, E.; Simmerling, C.; Dill, K. Protein storytelling through physics. *Science* **2020**, *370*, eaaz3041. [[CrossRef](#)]
52. Wallace, J.A.; Shen, J.K. Predicting pKa Values with Continuous Constant pH Molecular Dynamics. *Methods Enzymol.* **2009**, *466*, 455–475. [[CrossRef](#)] [[PubMed](#)]
53. Mongan, J.; Case, D.A. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.* **2005**, *15*, 157–163. [[CrossRef](#)] [[PubMed](#)]
54. Meng, Y.; Roitberg, A.E. Constant pH Replica Exchange Molecular Dynamics in Biomolecules Using a Discrete Protonation Model. *J. Chem. Theory Comput.* **2010**, *6*, 1401–1412. [[CrossRef](#)] [[PubMed](#)]
55. Swails, J.M.; Roitberg, A.E. Enhancing Conformation and Protonation State Sampling of Hen Egg White Lysozyme Using pH Replica Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 4393–4404. [[CrossRef](#)] [[PubMed](#)]
56. Itoh, S.G.; Damjanovic, A.; Brooks, B.R. pH Replica-Exchange Method Based on Discrete Protonation States. *Biophys. J.* **2012**, *102*, 40a. [[CrossRef](#)]
57. Fallon, L.; Belfon, K.A.A.; Raguette, L.; Wang, Y.; Stepanenko, D.; Cuomo, A.; Guerra, J.; Budhan, S.; Varghese, S.; Corbo, C.P.; et al. Free Energy Landscapes from SARS-CoV-2 Spike Glycoprotein Simulations Suggest that RBD Opening Can Be Modulated via Interactions in an Allosteric Pocket. *J. Am. Chem. Soc.* **2021**, *143*, 11349–11360. [[CrossRef](#)]
58. Doig, A.J.; Baldwin, R.L. N- and C-capping preferences for all 20 amino acids in α -helical peptides. *Protein Sci.* **1995**, *4*, 1325–1336. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.