bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Manuscript submitted to **Biophysical** *Journal*
Article

Binding Mode of SARS-CoV2 Fusion Peptide to Human Cellular Membrane

D. Gorgun[1,2], M. Lihan[1,2], K. Kapoor[1,3], and E. Tajkhorshid[1,2,3,*]

ABSTRACT

Infection of human cells by the SARS-CoV2 relies on its binding to a specific receptor and subsequent fusion of the viral and host cell membranes. The fusion peptide (FP), a short peptide segment in the spike protein, plays a central role in the initial penetration of the virus into the host cell membrane, followed by the fusion of the two membranes. Here, we use an array of molecular dynamics (MD) simulations taking advantage of the Highly Mobile Membrane Mimetic (HMMM) model, to investigate the interaction of the SARS-CoV2 FP with a lipid bilayer representing mammalian cellular membranes at an atomic level, and to characterize the membrane-bound form of the peptide. Six independent systems were generated by changing the initial positioning and orientation of the FP with respect to the membrane, and each system was simulated in five independent replicas. In 60% of the simulations, the FP reaches a stable, membrane-bound configuration where the peptide deeply penetrated into the membrane. Clustering of the results reveals two major membrane binding modes, the *helix-binding* mode and the *loop-binding* mode. Taken into account the sequence conservation among the viral FPs and the results of mutagenesis studies establishing the role of specific residues in the helical portion of the FP in membrane association, we propose that the helix-binding mode represents more closely the biologically relevant form. In the helix-binding mode, the helix is stabilized in an oblique angle with respect to the membrane with its N-terminus tilted towards the membrane core. Analysis of the FP-lipid interactions shows the involvement of specific residues of the helix in membrane binding previously described as the fusion active core residues. Taken together, the results shed light on a key step involved in SARS-CoV2 infection with potential implications in designing novel inhibitors.

SIGNIFICANCE A key step in cellular infection by the SARS-CoV2 virus is its attachment to and penetration into the plasma membrane of human cells. These processes hinge upon the membrane interaction of the viral fusion peptide, a segment exposed by the spike protein upon its conformational changes after encountering the host cell. In this study, using molecular dynamics simulations, we describe how the fusion peptide from the SARS-CoV2 virus binds human cellular membranes and characterize, at an atomic level, lipid-protein interactions important for the stability of the bound state.

INTRODUCTION

Coronavirus disease 2019 (COVID-19) emerged in late 2019 as a significant threat to human health. It became a global pandemic by March 2020 (1, 2), and it continues to claim lives and to significantly impact all aspects of people's lives across the globe. COVID-19 is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), a positive-strand RNA virus that causes severe respiratory complications, among other symptoms, in humans (3). SARS-CoV2 recognizes and infects human cells that express a cell surface receptor termed angiotensin-converting enzyme 2 (ACE2) (4), which is specifically recognized by the viral spike glycoprotein (S-protein). Binding of the two proteins is a prerequisite for the fusion of the viral and cellular membranes (5), one of the first and required steps in viral infection facilitating the release of the viral genome into the infected cell (6–9).

The binding of the virus to the surface CAE2 receptor on the host cell is mediated by the S1 domain in the S-protein on the viral surface. The next key step, namly, virus-host membrane fusion is mediated by the S2 domain of the S-protein (10), the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

domain, which consists of multiple proteolytic cleavage sites, namely, one at the boundary of S1/S2 and one at the S2' site, which are cleaved as part of the fusion process. (11–13) Cleavages at the S1/S2 boundary and the S2' sites, downstream to the two heptad repeat regions (HR1 and HR2), induce the dissociation of the S1 subunits from the S-protein, followed by a series of conformational changes that trigger membrane fusion between the host cell membrane and the viral envelope (14, 15). The remaining S2 trimer, a post-fusion structural motif, is shared among all the class I viral fusion proteins (16, 17). A critical part of any viral fusion protein in the coronavirus family is the relatively apolar fusion peptide (FP), present in the S2 domain, which is responsible for directly inserting into and interacting with the host cell membrane, thereby initiating the fusion process (8, 18, 19).

There are several characteristics that viral FPs have in common and help locating the FP sequence. The sequences of the FPs are highly conserved within each family of viruses (but not between families), the frequency of glycines and alanines in the sequence is high, and bulky hydrophobic residues as well as hydrophilic residues usually flank the cleavage sites (20, 21). Some FPs have a central kink via a proline and a helix-turn-helix structure, as observed in influenza virus hemagglutinin (HA). In such cases, proteolytic cleavage occurs directly at the N-terminus to the FP, and the peptides are thus called external or N-terminal FPs. In other cases, the proteolytic cleavage site resides upstream from the FP which is relatively longer (25-30 amino acids) and contains a prolonged α-helix in the fusion-active state such as in the cases of Ebola virus or avian leukosis sarcoma virus (8, 20, 21). Such FPs are referred to as internal FPs as in the case of SARS-CoV2. To this date, three main FP regions have been suggested in SARS-CoV S-protein (22, 23), which are located in between HR1 and N-terminus of the S2 domain: (1) at the N-terminus of HR1, (2) near the S1/S2 cleavage site, and, (3) at the C-terminus of the cleavage site S2' (13). Based on the criteria stated above as well as experimental studies, most recent data suggest that immediately downstream of the S2' cleavage site of the SARS-CoV2 is the leading segment involved in the fusion process (11, 24–27). Mutagenesis experiments showed the significance of the FP in this region, specifically carrying a region termed the fusion active core (21).

The outer leaflet of human cell membranes is composed of a mixture of lipids including phosphatidylcholine (PC), phosphatidylethanolamine (PE), and cholesterol (CHL) (28). A fluorescence spectroscopy study has shown that CHL plays an important role in modulating the binding affinity and organization of the SARS-CoV FP in the membranes (29). Therefore, taking into account the natural lipid composition of a mammalian cell in simulation studies such as the present one is important. Characterizing how the FP binds to the membrane and how it interacts with specific lipids has been challenging experimentally. Computational methods, particularly atomistic molecular dynamics (MD) simulation, offer an alternative strategy to capture the membrane binding process of the FP and to probe the interface between the FP and the lipid membrane. One of the major challenges in simulating such processes lies in sufficient sampling of all possible FP membrane-binding poses. Due to the slow dynamics of membrane lipids, they are often insufficiently sampled on the timescales which atomistic MD simulations currently can access, causing the membrane binding and insertion of proteins to be heavily biased by the initial lipid distribution and protein placement. In this context, an alternative membrane model, termed the highly mobile membrane-mimetic (HMMM) model, has been developed to enhance lipid diffusion without compromising the atomistic description of lipid head groups (30–33). The HMMM model is based on the combination of a biphasic solvent system (30) with short-tailed lipids at the interface (31). Owing to its significantly enhanced lipid mobility the model has proven extremely efficient in describing mixed lipid bilayers, reproducibly capturing spontaneous (unbiased) membrane binding and insertion of a wide spectrum of peripheral proteins (31, 34–45), and collecting significantly improved sampling of lipid–protein interactions. Of particular interest to the present study, this method has also been successfully used to capture spontaneous membrane association of the influenza virus hemagglutinin FP (46).

In this study, we perform an extensive set of HMMM simulations, to investigate membrane binding of the SARS-CoV2 FP, also taking into account several initial, different positioning of the peptide with respect to the membrane to further improve sampling. The results provide a detailed mechanistic picture of the initial step in the fusion process, focusing on the biophysical aspects of the virus–lipid bilayer interactions taking place during this process. Characterizing the mechanism of the fusion-driving, FP-host membrane interactions is key to our understanding critical steps involved in viral infection, and might pave the way for development of novel therapeutic intervention strategies against the virus.

MATERIALS AND METHODS

Multiple sequence alignment and modeling of the SARS-CoV2 fusion peptide (FP)

As the first step for modeling the SARS-CoV2 FP, multiple sequence alignment was carried out for the S-proteins from different human coronaviruses (HCoVs). There are a total of seven known HCoVs, namely, 229E and NL63 belonging to the alpha subfamily of coronaviruses, and OC43, HKU1, Middle East respiratory syndrome (MERS), severe acute respiratory syndrome (SARS) and SARS-CoV2, belonging to the beta subfamily of coronaviruses. Out of these, S-protein structures are available for HKU1, MERS, SARS, as well as SARS-CoV2. Additionally, bat coronavirus RaTG13, a closely related homolog of

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SARS-CoV2, was also included in the sequence alignment. Multiple sequence alignment for the above eight sequences was carried out using the MAFFT program with the L-INS-i method (47) and visualized using Jalview (48) (Fig. S1).

The available cryoEM structure of the SARS-CoV2 S-protein at the time (PDB: 6VSB) (49) contained only a partial structure (12 residues) of the FP, making it unsuitable for developing the initial SARS-CoV2 FP model (Fig. 1A). Since the S2 domain of the S-protein containing the FP is well-conserved among the SARS coronaviruses, we used the S-protein from SARS-CoV as a guide for modeling the FP of the SARS-CoV2. Accordingly, the cryoEM structure of the S-protein from SARS-CoV (PDB: 5XLR) (50) containing the FP structure was used as a template for constructing the initial SARS-CoV2 FP model. Our SARS-CoV2 FP model also contained the loop connecting the FP to the neighboring proximal region, suggested to be important in the fusion process (24–26), but not modeled fully in the current study due to lack of a suitable template structure at the time. The only two different residues between the sequences of SARS-CoV and SARS-CoV2 FPs (I834/M816 and D839/E821) were mutated to the SARS-CoV2 residues (Fig. 1B). The initial FP model was then solvated and ionized with 0.15 M NaCl in VMD (51). Energy minimization was carried out using the steepest descent method for 100 timesteps followed by solution equilibrium MD for 20 ns, in order to obtain a fully equilibrated SARS-CoV2 FP in an aqueous environment. The resulting equilibrated FP was used in the subsequent simulations investigating the membrane binding characteristics of the peptide.

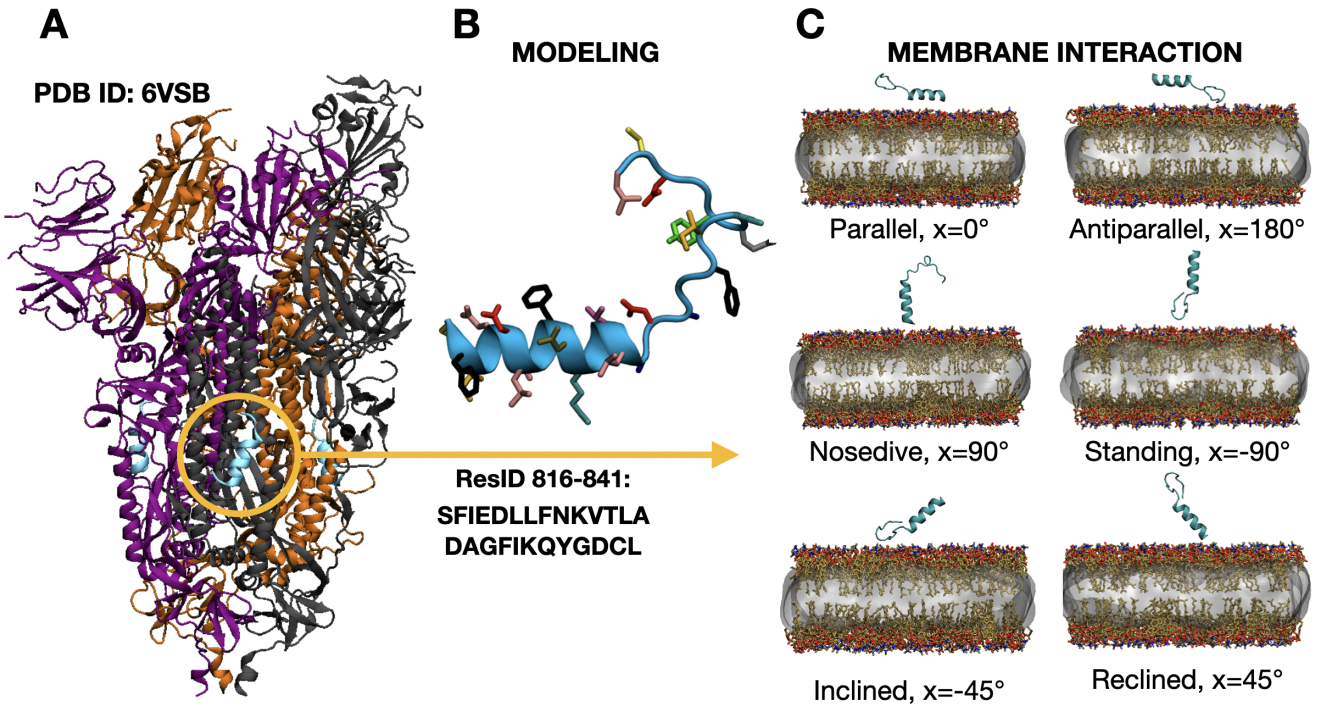

Figure 1: **Modeling and design of membrane binding simulations of the SARS-CoV2 FP.** A) Cryo-EM structure of pre-fusion trimeric SARS-CoV2 S-protein with the highlighted FP (blue). Each monomer in the trimer is drawn in a different color (grey, purple, and orange). B) The SARS-CoV2 FP. The missing residues of the SARS-CoV2 FP structure were modeled using SARS-CoV FP (PDB 5XLR), with the mutations of I834/M816 and D839/E821. Residues 817-828 are in an alpha helical conformation and residues 829-841 form a loop structure. C) HMMM membrane binding simulation setup. After simulating the modeled FP in solution for 20 ns, we placed the equilibrated peptide above the HMMM lipid bilayers in several different orientations. The membrane lipid composition is PC/PE/CHL (60/6/34 mol% which is representative of the outer leaflet of the human plasma membrane (carbon as tan, oxygen as red, phosphorus as blue, and nitrogen as orange). We use six different initial orientations rotating the peptide around its x axis with respect the parallel orientation: parallel (P, x=0°), antiparallel (A, x=180°), nosedive (N, x=90°), standing (S, x=−90°), inclined (I, x=−45°), and reclined (R, x=45°)), for the FP placement above the membrane to randomize the initial interactions. Lipid bilayers were randomly and individually built. Each orientation was simulated in five independent replicas, with the total replica number of 30 for membrane-binding simulations.

Membrane preparation

We developed membrane-bound models of the SARS-CoV2 FP, utilizing multiple simulation replicas using the HMMM model (Fig. 1C) (30, 31, 33). Symmetric conventional (full) membranes were first constructed using CHARMM-GUI (52) with a lipid composition of PC/PE/CHL (60/6/34 mol%), resembling the outer leaflet of human plasma membrane (28). The membranes were then converted to HMMM membranes by removing the atoms after the fifth carbon in the phospholipid acyl chains while

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

keeping the cholesterol molecules intact. To mimic the membrane core, a previously developed *in silico* solvent, termed SCSE (including two carbon-like interaction centers) (33), was used to match the number of heavy atoms removed from the lipid tails in the previous step. The resulting HMMM membranes contained 2,178 SCSE molecules (33) and 150 lipids in each leaflet.

In order to further expand the sampling of the phase space, we varied the initial placement and orientation of the FP, i.e., six rotated orientations (P, A, N, S, I, R) (Fig. 1C), to further reduce the initial bias. We solvated the systems using the SOLVATE plugin in VMD (51), with 0.15 M NaCl resulting in total system sizes of 70,000 to 90,000 atoms and box sizes of 103×103×90 to 103×103×110 $Å^3$, respectively. Multiple replicas were simulated for each FP orientation, as the diffusion and mixing of lipids and the process of membrane binding and insertion of the FP can be slow even when using HMMM membranes, and more sampling would ensure the meaningfulness of the obtained membrane-bound configurations. Five independent HMMM membranes with each specified FP orientation were generated using a Monte Carlo based lipid mixing protocol developed in our group to further enhance variation of initial membrane configurations for each FP orientation.

Membrane binding simulations

The systems were energy minimized for 10,000 steps and simulated for 10 ns with the Cα atoms of the peptide harmonically restrained (k = 0.5 kcal mol^{-1} $Å^{-2}$), followed by a production run of 100 ns where Cα restraints were removed. A harmonic restraint along the z-axis, with a force constant k = 0.05 kcal mol^{-1} $Å^{-2}$, was applied to the C26, C36, and C2 atoms of the phospholipids and to the O3 atoms of cholesterol to mimic the atomic distributions of a full lipid bilayer more closely, and to prevent the occasional diffusion of short-tailed lipids into the aqueous phase, which is expected for these surfactant-like molecules. To prevent SCSE molecules from diffusing out of the core of the membrane, we subjected them to a grid-based restraining potential, applied using the gridForce (53) feature of NAMD (54, 55). Five replicas, each with an independently generated HMMM membrane and a starting orientation of the FP, were simulated, resulting in a total of 30 independent membrane binding simulations.

Analysis

We define stable membrane binding with the criteria described below (see Fig. 3). First, a contact between the FP and the membrane is defined for any heavy atom of an FP residue that is within 3.5 Å of any lipid heavy atoms. Any contiguous segment of the simulation trajectory with a length of at least 10 ns during which at least one contact between the lipids and the FP existed was considered stable binding. The mid-points of these 10-ns segments are marked as a red point in Fig.3.

To characterize the binding orientation of the FP with respect to the membrane, the first of the three principal axes (PA) of the FP helical segment (residues 816-825) was computed (Fig. S2). The angle between the first helical PA and the membrane normal, θ_{PA}, was used to describe the tilting of the helical segment. Two auxiliary angles, θ_{F817} and θ_{F823}, were further defined by the angle between the membrane normal and the vector component of the respective selected phenylalanine C_α-C_β vector perpendicular to the first PA, which together describe the rotational degrees of freedom of the FP about the helical first PA.

K-medoids clustering algorithm (56) was performed to categorize the membrane-bound poses of the FP, resulting in two major membrane binding modes. For all the frames identified as membrane-bound (see above for definition), a vector composed of the z-distances between the COM of individual FP residues and the lipid phosphate plane was used as the dissimilarity metric for clustering. The cluster approximate centers, i.e., medoids, were viewed as the representative structures for each FP membrane binding mode identified. To better visualize the distribution of membrane-bound poses as well as the clustering results, principal component analysis (PCA) was performed to reduce the dimensionality of the original data set. In PCA, the covariance matrix of the z-distances between the COM of individual residues and the phosphate plane was computed and diagonalized. The resulting eigenvectors, i.e., the principal components (PCs), represents the coordinates that maximize the variance of projected data. The first two PCs were selected to project the original z-distance data of membrane-bound frames along with the clustering result onto the reduced dimension. The PCA was performed using the scikit-learn package (57).

RESULTS AND DISCUSSION

The SARS-CoV2 attachment to and its penetration into the human plasma membrane are key steps in viral infection that rely on the interaction of the viral FP with the host cellular membrane. This aspect is the central phenomenon that we study here using atomistic simulations, in order to characterize the binding pose and conformation of the SARS-CoV2 FP when bound to the membrane. In the following sections we first describe the results of our spontaneous membrane binding simulations in terms of the overall binding of the FP using some coarse parameters. Then, we use the interaction of individual residues from the FP in clustering and examine them in a reduced dimension to classify better the different microstates that arise during the binding process and discuss their biological significance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Spontaneous Membrane Binding and Insertion of SARS-CoV2 FPs

In order to characterize the membrane binding mode of the SARS-CoV2 FP to a lipid bilayer representing human cellular membranes, we performed 30 independent copies of HMMM, membrane-binding simulations. Spontaneous diffusion and membrane binding of the FP in each simulation replica can be monitored by tracking the center of mass (COM) of the peptide. The z component of the COM was tracked with respect to the phosphate layer (the average z position of all the phosphate groups) of each leaflet (blue and red lines, respectively, in Fig. 2). Due to the applied periodic boundary conditions, and the free diffusion of the FP in the solution, the FP was able to diffuse towards either the upper or the lower leaflet of the membrane (Fig. 2), which both include a lipid composition resembling the outer leaflet of human plasma membrane (28). Among the 30 performed membrane-binding simulations, instances of both stable (majority) and transient binding events, as well as cases with no membrane binding, were observed. Because the lipid molecules in the membrane are neutral, there are no dominating electrostatic driving forces between the FP and the lipid bilayer. Therefore, the major driving forces between the two are hydrophobic effects, and the FP diffusion in the solution can make the peptide take a longer time to make the initial encounter with the membrane.

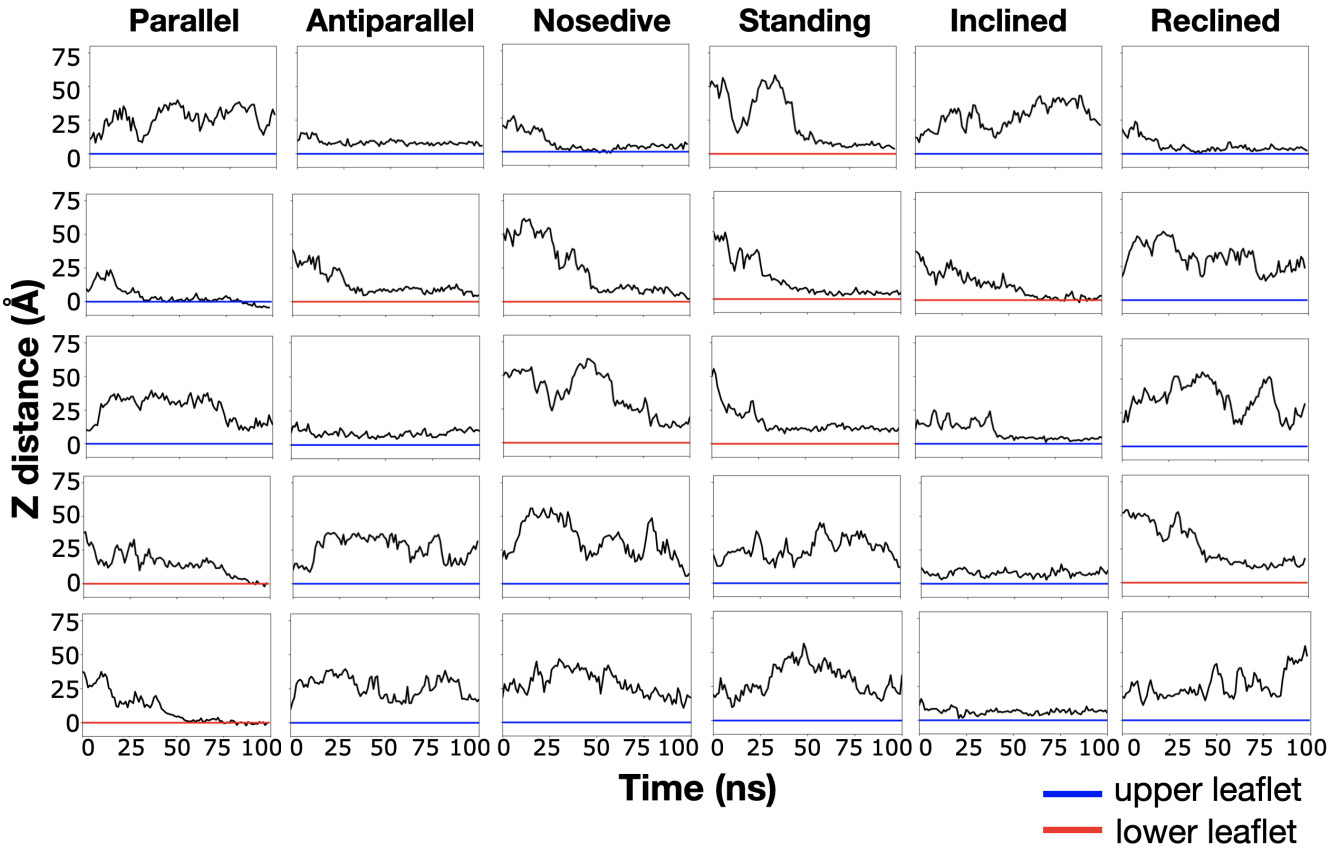

Figure 2: **Solution diffusion and membrane binding of the SARS-CoV2 FP during the 30 independent membrane-binding simulations.** The z-distance of center of mass (COM) of the FP with respect to the average z position of the lipid phosphates in each leaflet, referred to as phosphate layer. The peptide can diffuse within the simulation box and reach the distal leaflet, due to the periodicity of the simulation box. Replicas where the FP interacts with the lower leaflet are shown with a phosphate layer colored in red, and those binding to the upper leaflet in blue.

For further analysis of the membrane interactions, we selected only the portions of the trajectories where "stable membrane binding" or "membrane-bound state" was defined (see Methods for details). Residue contacts between any heavy atom of the FP and any heavy atom of the lipid bilayer are shown in Fig. 3 where the red segments of the graphs represent stable binding. Consisting of six independent systems and five replicas of each, 60% of the 30 replicas contain stably bound configurations (Fig. 3). We observe stable binding in, e.g., parallel replicas P2, P4, P5, and in antiparallel replicas A1, A2, A3 (Fig. 3). In some simulations, nearly the entire length of the FP was observed to be engaged with lipids (e.g., replicas P2, P5, N1, S2, and R1), while in other cases, only a specific part of the peptide makes contact with the membrane, specifically via either the loop or the helical part. Biasing the initial placement of the peptides allowed the first interaction of the FP to diversify. Some peptides

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

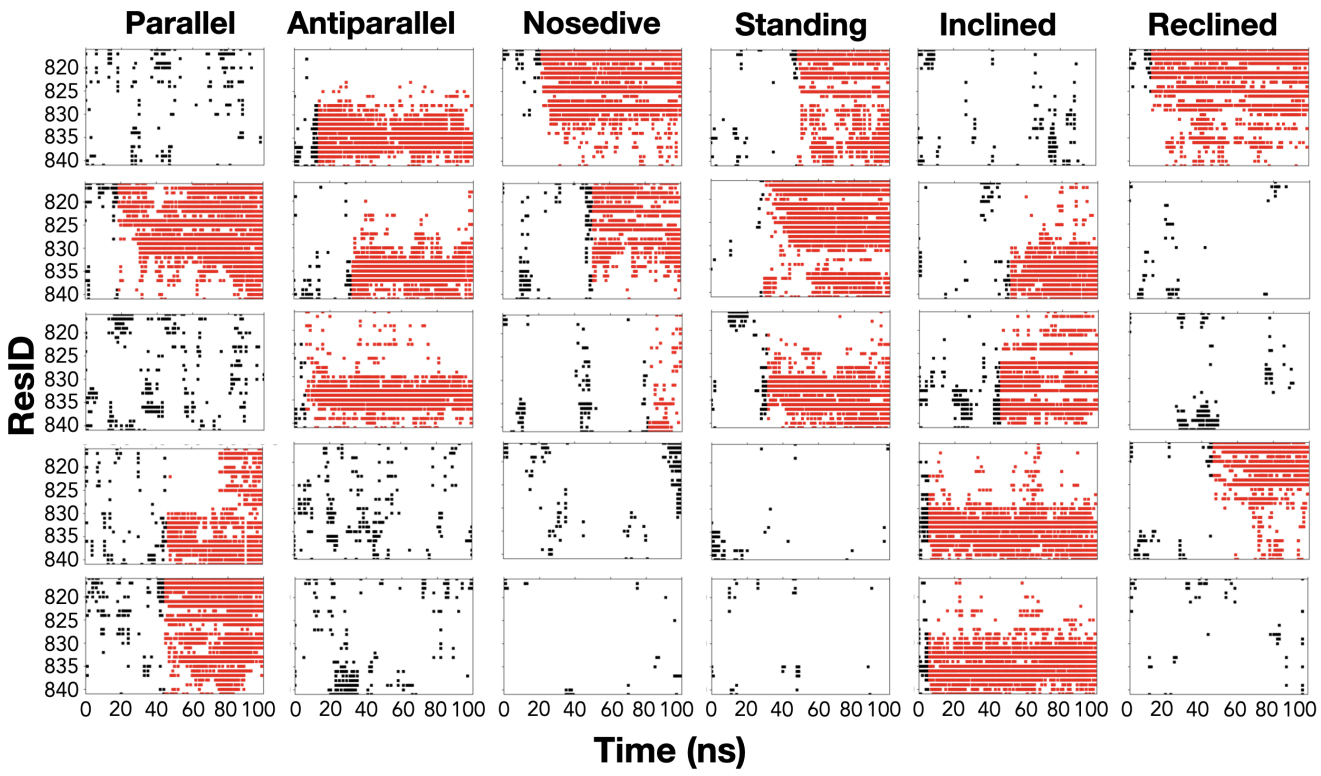

Figure 3: **Defining stable membrane binding based on the contacts between the FP and the lipid bilayer.** The plots show the contact of individual residues (y-axis) with the lipid bilayer over time (x-axis) in sampling step size of 1 ns. If the FP is in contact with the membrane (heavy atom distance of less than 3.5 Å) for at least 10 ns, the midpoint of that segment of the trajectory is marked with a black dot. Short-term, transient contacts are labeled with black dots. We observe stable binding in 60% of the replicas.

bound to the lipid bilayer as soon as in 10 ns, while others diffused and tumbled longer prior to interacting with the membrane, naturally allowing the peptide to unbias itself more from the effect of initial placement. Examples include nosedive-placed peptides resulting in membrane-interacting FP via the loop part, or standing-placed peptides buried in the bilayer via their helical segment (Fig. 3).

Using the stable membrane-bound states defined above, we further examined the average position and lipid interaction of individual FP residues. We calculated the z-distance of the side chain COM to the phosphate layer of the membrane (Fig 4). Each ensemble average is calculated from the membrane-bound portions of the individual trajectory. The average depth of insertion of individual residues reaches as deep as 10 Å below the phosphate layer, indicating interaction with the hydrophobic core of the membrane.

We observe two major classes of membrane-bound states where either the alpha helical part (approximately residues 817-827) or the loop part (the rest of the residues at C-terminus) of the FP interact with the lipid bilayer. For example, replicas P2, P5, and N1, the FP are bound to the membrane primarily through its alpha helical segment whereas in systems P4, P5, A1, A2, and A3, the C-terminal end (the loop) is mostly engaged with the lipids. In order to better classify these binding modes in a reduced space, we will next do clustering of the obtained membrane-bound configurations.

Clustering and characterization of FP binding modes

To analyze the ensemble of FP membrane binding configurations captured in the 18 HMMM simulations where stable membrane binding was observed, clustering was performed using the z-distances between the the side chain COM and the membrane lipid phosphate plane. Despite the large variance in membrane-bound poses among MD snapshots from different simulation replicas, two major clusters can be identified, which represent the two distinct binding patterns mentioned briefly above. We call the first cluster where the N-terminal, helical part of the FP interacts with the membrane, the “helix-binding” mode. In the other major cluster, the FP interacts with the bilayer primarily from the C-terminal loop, hereby referred to as the “loop-binding” mode. For better visualization, we performed principal component analysis (PCA) and reduced the dimension to the first two principal components (PCs), which together explain 84% of the variance. The nature of these two PCs, PC1 and PC2, was examined by

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

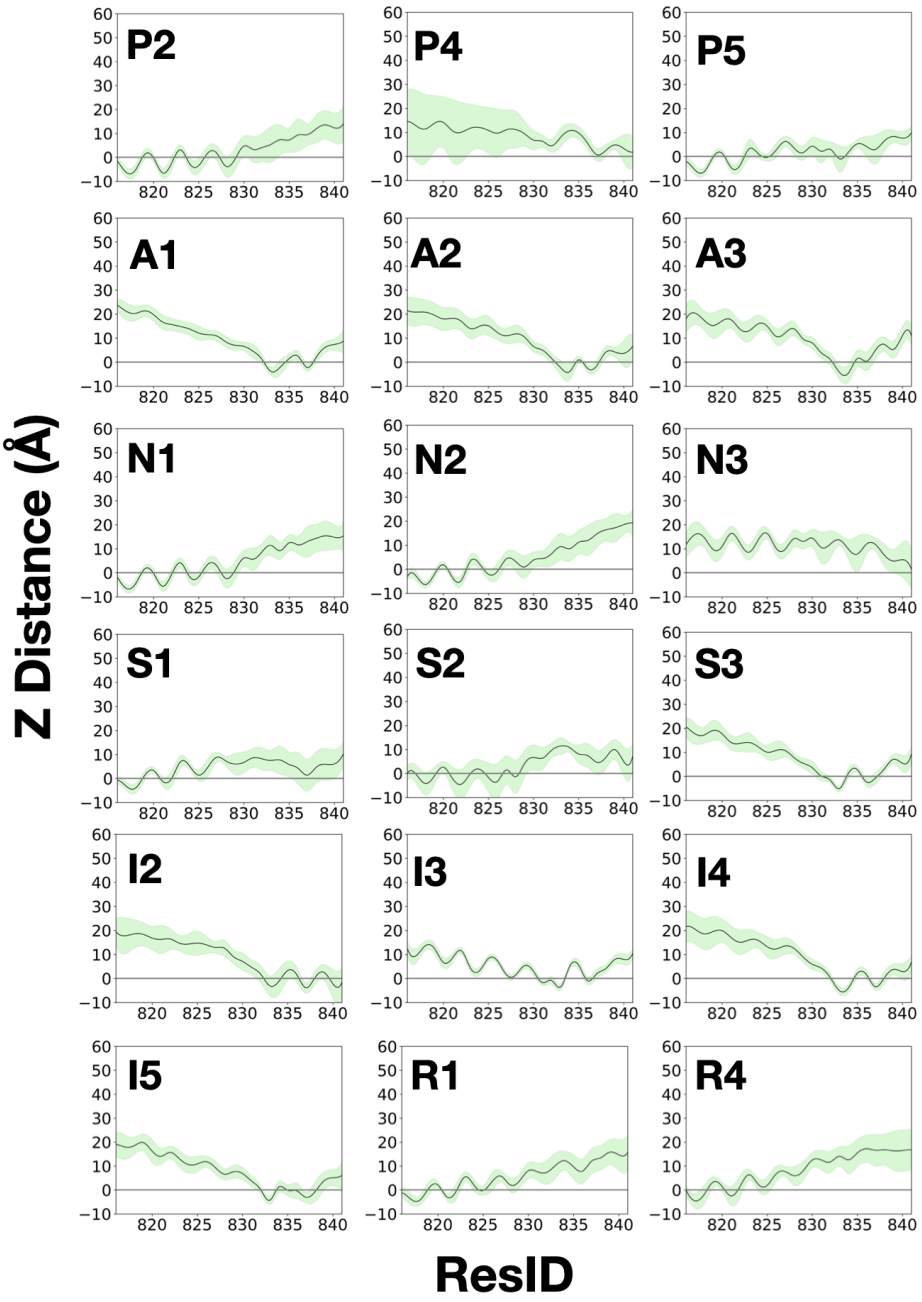

Figure 4: **Insertion of individual FP residues into the membrane.** Plots show the ensemble average of the side chain COM z with respect to the phosphate plane of the bilayer in replicas with stable FP-membrane binding. Each data point is the average from the stably bound segments of each individual trajectory (red areas in Fig. 3). Transparent green areas represent the standard deviation of the distance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the eigenvectors, which quantitatively evaluate the contribution of membrane insertion by each residue (Fig. S3). From the eigenvectors, PC1 measures the contrast of membrane insertion by the helix and the C-terminal loop, whereas PC2 measures mostly how much the C-terminal loop is inserted. All membrane binding configurations were then projected and visualized in the reduced dimension defined by PC1 and PC2 (Fig. 5).

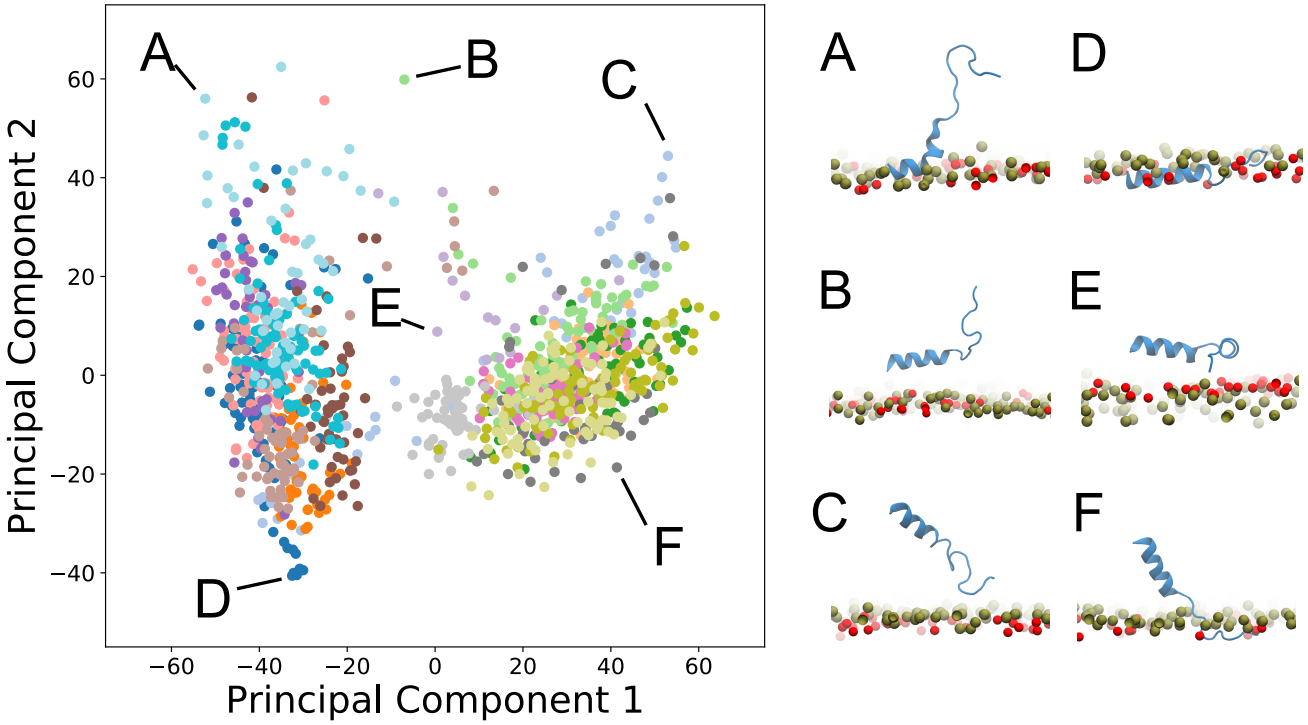

Figure 5: **Membrane-bound poses of the FP projected onto the reduced space formed by PC1 and PC2.** Different replicas of membrane-binding simulations are denoted by different colors. Representative membrane-binding snapshots are shown in the right panel to demonstrate different membrane-bound poses of the FP observed, with their locations in the 2D space labeled in the left panel. Membrane-bound poses exhibit a range of membrane binding preferences between the helix (A, D) and the loop (C, F), and varied degrees of membrane interaction at the loop, e.g., A vs. D, or B vs. E, or C vs. F.

Examination of the projections and selected MD snapshots are consistent with the presence of two major membrane-binding modes as introduced earlier, i.e., either the helix (Fig. 5A,D) or the loop (Fig. 5C,F), with a few involving both segments only at the membrane surface (Fig. 5B,E). The flexibility of the C-terminal loop resulted in a variety of scenarios in terms of its membrane interaction, including its deep membrane insertion (e.g., D and F in Fig. 5), tangential interaction (e.g., C and E in Fig. 5), or no interaction (e.g., A and B in Fig. 5). The two major clusters of membrane binding poses are also visualized in the reduced dimension (Fig. 6A), along with the two cluster centers showing the representative membrane binding modes (Fig. 6B,C). As expected, the cluster corresponding to the helix-binding mode demonstrated a larger variance of the loop involvement in the membrane binding compared to the cluster representing the loop-binding mode.

Orientation of the FP in its Membrane Bound Configuration

To further analyze the membrane binding configurations of the FP in our simulations, we analyzed the orientation of the helical segment of the FP (residues 816-825), which remains alpha helical throughout all the simulation replicas (Fig. S4). We traced the internal angles θ_{PA}, θ_{F817}, and θ_{F823} defined to describe the FP helix orientation (see Methods for the definition of the angles) over the simulation trajectories in Fig. 7A, for replicas with stably bound configurations. In the simulations where the FP binds the membrane in a helix-binding mode (P2, P5, N1, N2, S1, S2, R1, R4), θ_{PA} stabilizes at ~102° which corresponds to an oblique position relative to the membrane where the N-terminus is tilted towards the membrane core. Since the loop is more flexible and the reference internal angle is defined at the α-helix moving in the solution, it is more difficult to provide an equally accurate description for the loop-binding modes (P4, A1, A2, A3, N3, S3, I2, I3, I4, I5). In most of the replicas towards the end of the trajectory, θ_{PA} is less than 90°. This means that the N-terminal of the helix is facing up, and as the angle decreases the

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

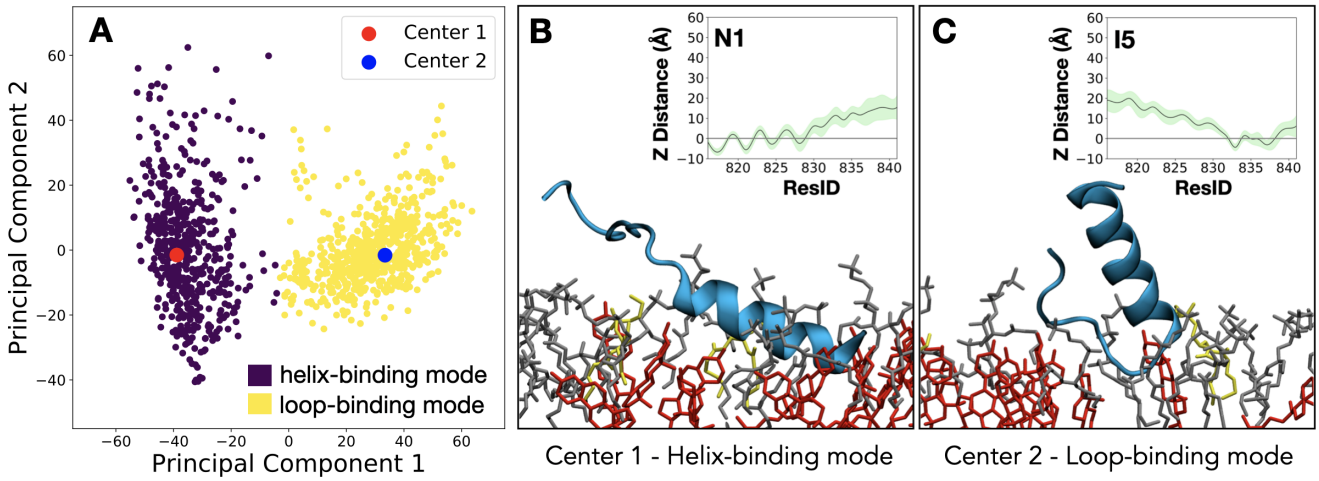

Figure 6: **Clustered FP membrane-bound poses projected on PC1 and PC2 and representative binding modes of the FP.** A) The helix-binding and the loop-binding modes are colored in purple and yellow, respectively, with their clustering centers highlighted. B) The configuration where the helix interacts with the bilayer in an oblique manner (helix-binding mode). C) The binding mode where the loop is inserted in the membrane (loop-binding mode). The insets of panels B and C represent the average COM distance of each residue with respect to the membrane phosphate layer.

helix becomes more orthogonal to the membrane. In the rest of the loop-binding modes, θ_{PA} is larger than 90°, meaning while the loop is inserted into the membrane, the helix is also interacting with the membrane but not fully inserted. The θ_{F817} angles in the helix-binding mode averaged around ~131° where F817 is facing towards the lipid bilayer. This preference can be clearly seen in Fig. 7B. Since F823 is nearly located on the opposite side of the helix, this residue is mostly facing up in the membrane with θ_{F823} averaging around 45°. However, there is no specific preference for θ_{F817} and θ_{F823} angles in the loop-binding mode (Fig. 7B).

Physiologically Relevant Membrane Bound Configuration of the FP

Within each family of viruses, the sequence of the FPs is highly conserved. In addition, the alpha helical part of the SARS-CoV2 FP (SFIEDLLFNKV) is highly conserved among the coronaviridae family (20, 21), as indicated by the multiple sequence alignment (Fig. S1), highlighting its importance in the viral life cycle.

In our simulations we observed two membrane binding modes for the SARS-CoV2 FP, the helix-binding mode in which the helical segment of the FP engages with the membrane, and the loop-binding mode where the C-terminal loop of the peptide is the part primarily interacting with the lipids (Fig. 6). Mutagenesis experiments have clearly shown that residues L821, L822, and F823 play a major role in viral fusion. Given their importance, these residues are termed the “fusion active core” of the FP (21, 58). Consistent with these results, in the helix-binding mode found in our simulations, residues L821, L822, and F823 are closely interacting with the membrane lipids with the two leucine side chains deeply inserted into the hydrophobic core of the membrane (Fig. 8). Notably, the same residues are found to be highly conserved among all human corononaviruses (Fig. S1). Given the high sequence conservation and established role of several residues in the helical segment of the FP that we simulated, we strongly believe and propose that the helix-binding mode represents the physiologically relevant membrane-bound form of the SARS-CoV2 FP. Notably, from an energetic perspective, insertion of helical segments into the membrane might be less costly, since backbone amide groups in a helix satisfy their hydrogen bonds internally (59).

Based on structural data currently available, the loop-binding mode might become relevant to potential interaction of other segments of the protein with the membrane upon structural transition and cleavage of the S protein. In addition to the simulated segment in this study (which we refer to as the FP), there are other pieces proximal to the FP (named FP2 in SARS-CoV and FPPR in SARS-CoV2) (11, 24–27), which are implicated in membrane fusion. The fusion peptide proximal region (FPPR) of SARS-CoV2 downstream to the FP was later resolved and claimed to be involved in the structural rearrangement of the S protein prior to membrane fusion (27). An internal disulfide bond within the FPPR, between C840 and C851, was observed and is suggested to increase membrane-ordering activity (25, 27). The membrane-ordering activity of the FP, due to the fusion active core, is significantly higher than the FPPR and the activity of FP/FPPR together is only slightly increased compared to the activity of FP and FPPR separately (25). The loop-binding mode might support the formation of a "fusion platform" where both FP and FPPR interact with the membrane simultaneously as two subdomains (25). To characterize such platform

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

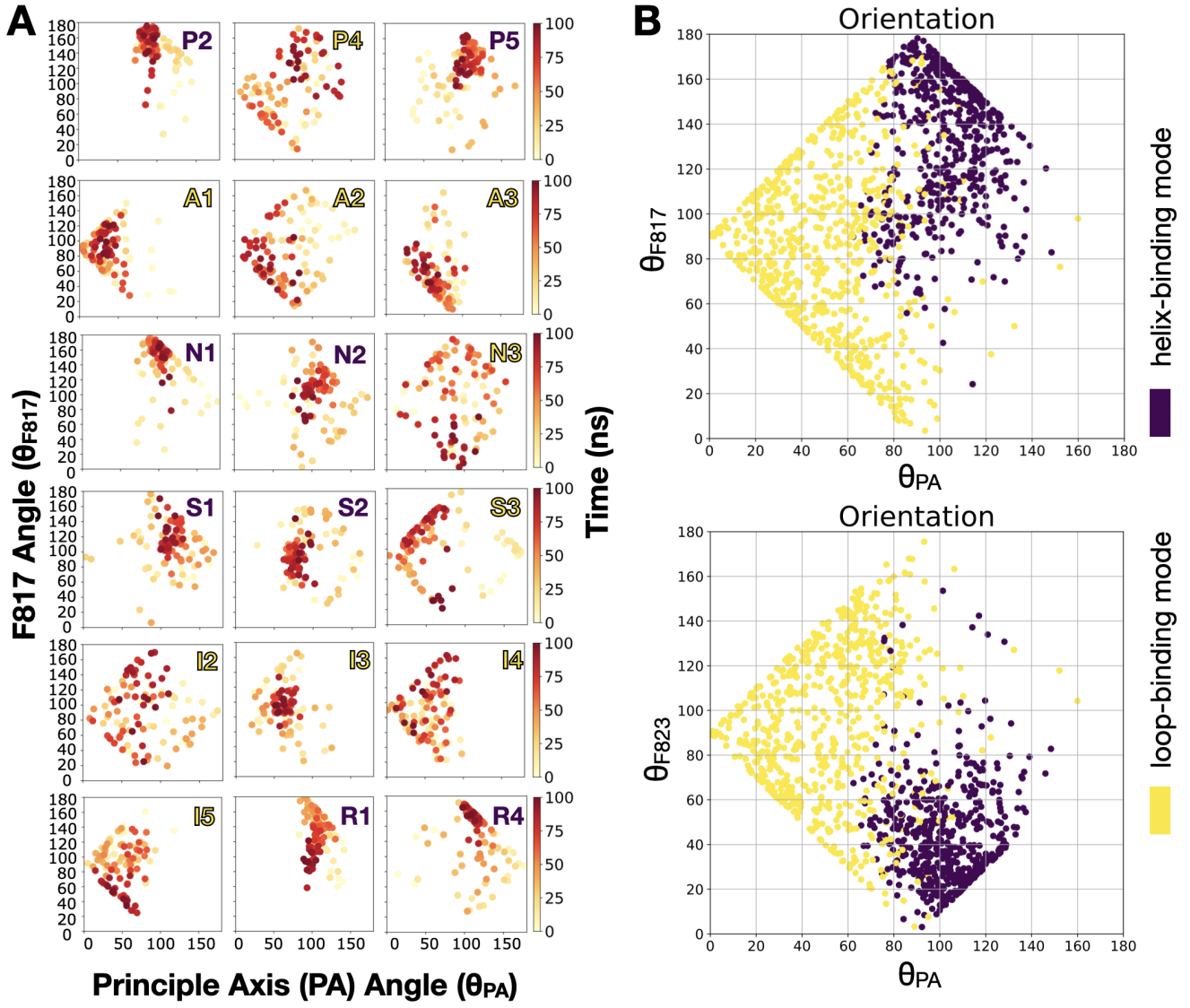

Figure 7: **Angular analysis of the membrane-bound FPs.** A) Two defined internal angles (Fig. S2) relative to the membrane normal are represented in x-y axis changing with color in time. Replica names are colored according to whether the FP is in a helix- (purple) or in a loop-binding mode (yellow). B) Different helix orientations displayed by the two FP membrane-binding modes. The helix-binding mode prefers a θ_{PA} angle between 70° and 140°, accumulated more around 100° whereas the loop-binding mode prefers a wider range of θ_{PA} angles, spanning from 0° to 100°. Unlike the loop-binding mode, the helix-binding mode demonstrates a strong preference for F817 rather than F823 to face towards the membrane.

interacting with the membrane, a longer peptide including the FPPR should be studied in future.

CONCLUSION

COVID-19, which emerged as a severe pandemic worldwide, calls for a need to accelerate the development of novel therapeutic intervention strategies. The S-protein of the SARS-CoV2 contains the key machinery necessary for the infection of human cell, including the FP, a highly conserved segment that inserts into the human cellular membrane initiating the fusion of the virus. In this study, using a large set of simulations, we describe how the SARS-CoV2 FP binds mammalian cellular membranes and characterize, at atomic details, lipid-protein interactions important for the stability of the bound state. Characterizing the mechanism of the fusion driving FP-host membrane interactions is key to our understanding critical steps involved in the process of viral infection, paving way for potential development of novel therapeutics against SARS-CoV2. These include modulation of FP-membrane binding interface through small molecules showing high specificity for this region of the S-protein, or inhibiting the key lipid-protein interactions observed.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

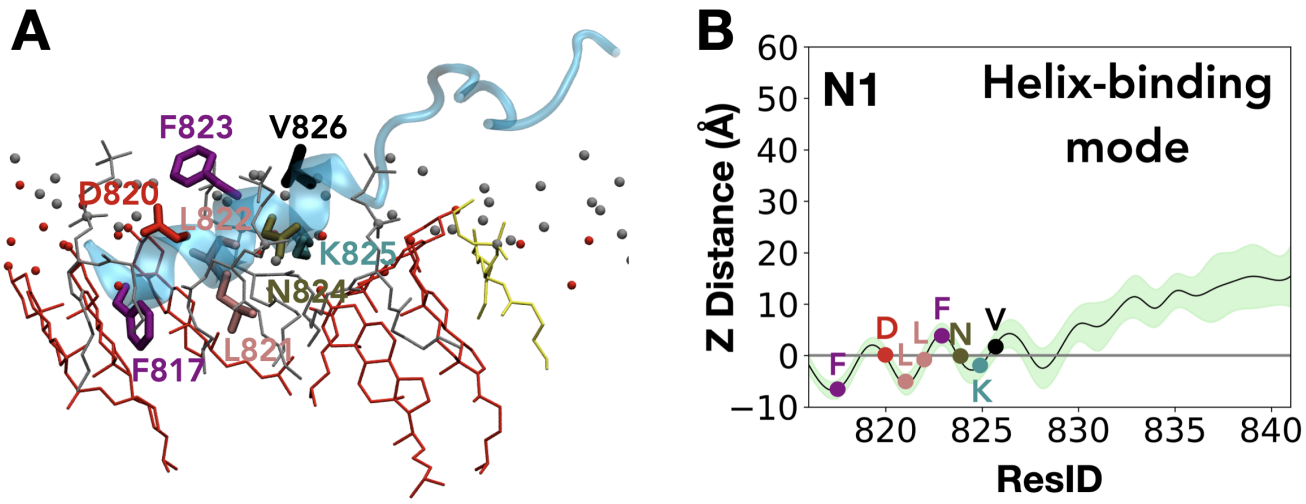

Figure 8: **Detailed representation of the proposed helix-binding mode.** A) Helix-binding mode is shown in a detailed manner, both lipids such as cholesterol in red, PE in yellow and PC in silver and side chains are shown that are interacting closer than 3 Å. Phosphorus atoms of PC and PE lipids (grey) and oxygen atoms of cholesterol (red) are shown as spheres. B) Average z-distance of each residue COM with respect to the phosphate bilayer is shown with marking of residues that are involved in membrane interaction. The residues with heavy atoms closer than 3 Å are marked in the plot.

Based on the suggested binding mode elucidated in our study, mutagenesis experiments can be designed to further confirm the role of the important residues implicated in membrane binding. Given the close similarity of the fusion peptides in coronaviruses in general, these results can also be applicable to infections caused by other members of this life-threatening family of pathogens.

AUTHOR CONTRIBUTIONS

DG, KK and ET designed the research. DG carried out all simulations, DG and ML analyzed the data. All authors wrote the article.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (Grants P41-GM104601 and R01-GM123455 to ET). MD simulations were performed using Blue Waters and computational resources provided by Microsoft Azure.

REFERENCES

1. Huang, C., Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, 2020. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet* 395:497–506.

2. Chan, J. F.-W., S. Yuan, K.-H. Kok, K. K.-W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C.-Y. Yip, R. W.-S. Poon, H.-W. Tsoi, S. K.-F. Lo, K.-H. Chan, V. K.-M. Poon, W.-M. Chan, J. D. Ip, J.-P. Cai, V. C.-C. Cheng, H. Chen, C. K.-M. Hui, and K.-Y. Yuen, 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *The Lancet* 395:514–523.

3. Singhal, T., 2020. A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics* 1–6.

4. Wan, Y., J. Shang, R. Graham, R. S. Baric, and F. Li, 2020. Receptor recognition by the novel coronavirus from Wuhan: an analysis based on decade-long structural studies of SARS coronavirus. *Journal of virology* 94.

5. Tang, T., M. Bidon, J. A. Jaimes, G. R. Whittaker, and S. Daniel, 2020. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res.* 178:104792.

6. Skehel, J. J., and D. C. Wiley, 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* 69:531–569.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

7. Eckert, D. M., and P. S. Kim, 2001. Mechanisms of viral membrane fusion and its inhibition. *Annual review of biochemistry* 70:777–810.

8. White, J. M., S. E. Delos, M. Brecher, and K. Schornberg, 2008. Structures and mechanisms of viral membrane fusion proteins: multiple variations on a common theme. *Critical reviews in biochemistry and molecular biology* 43:189–219.

9. Kielian, M., and F. A. Rey, 2006. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat. Rev. Microbiol.* 4:67–76.

10. Wentworth, D. E., and K. V. Holmes, 2007. Coronavirus binding and entry. *Coronaviruses: molecular and cellular biology. Caister Academic Press, Norfolk, United Kingdom* 3–32.

11. Belouzard, S., J. K. Millet, B. N. Licitra, and G. R. Whittaker, 2012. Mechanisms of coronavirus cell entry mediated by the viral spike protein. *Viruses* 4:1011–1033.

12. Heald-Sargent, T., and T. Gallagher, 2012. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* 4:557–580.

13. Basso, L. G., E. F. Vicente, E. Crusca Jr, E. M. Cilli, and A. J. Costa-Filho, 2016. SARS-CoV fusion peptides induce membrane surface ordering and curvature. *Sci. Rep.* 6:37131.

14. Belouzard, S., V. C. Chu, and G. R. Whittaker, 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences* 106:5871–5876.

15. Beniac, D. R., S. L. Devarennes, A. Andonov, R. He, and T. F. Booth, 2007. Conformational reorganization of the SARS coronavirus spike following receptor binding: implications for membrane fusion. *PLoS One* 2:e1082.

16. Supekar, V. M., C. Bruckmann, P. Ingallinella, E. Bianchi, A. Pessi, and A. Carfí, 2004. Structure of a proteolytically resistant core from the severe acute respiratory syndrome coronavirus S2 fusion protein. *Proc. Natl. Acad. Sci. USA* 101:17958–17963.

17. Bosch, B. J., R. Van der Zee, C. A. De Haan, and P. J. Rottier, 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. *J. Virol.* 77:8801–8811.

18. Earp, L., S. Delos, H. Park, and J. White, 2004. The many mechanisms of viral membrane fusion proteins. *In* Membrane trafficking in viral replication, Springer, 25–66.

19. Tamm, L. K., and X. Han, 2000. Viral fusion peptides: a tool set to disrupt and connect biological membranes. *Bioscience reports* 20:501–518.

20. Lai, A. L., Y. Li, and L. K. Tamm, 2005. Interplay of proteins and lipids in virus entry by membrane fusion. *Protein-lipid interactions, in press. Wiley-VCH, Weinheim, Germany* 279–305.

21. Madu, I. G., S. L. Roth, S. Belouzard, and G. R. Whittaker, 2009. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J. Virol.* 83:7411–7421.

22. Guillén, J., P. K. Kinnunen, and J. Villalaín, 2008. Membrane insertion of the three main membranotropic sequences from SARS-CoV S2 glycoprotein. *Biochim. Biophys. Acta, Biomembr.* 1778:2765–2774.

23. Guillén, J., A. J. Pérez-Berná, M. R. Moreno, and J. Villalaín, 2005. Identification of the membrane-active regions of the severe acute respiratory syndrome coronavirus spike membrane glycoprotein using a 16/18-mer peptide scan: implications for the viral fusion mechanism. *J. Virol.* 79:1743–1752.

24. Millet, J. K., and G. R. Whittaker, 2018. Physiological and molecular triggers for SARS-CoV membrane fusion and entry into host cells. *Virology* 517:3–8.

25. Lai, A. L., J. K. Millet, S. Daniel, J. H. Freed, and G. R. Whittaker, 2017. The SARS-CoV fusion peptide forms an extended bipartite fusion platform that perturbs membrane order in a calcium-dependent manner. *J. Mol. Biol.* 429:3875–3892.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26. Kirchdoerfer, R. N., N. Wang, J. Pallesen, D. Wrapp, H. L. Turner, C. A. Cottrell, K. S. Corbett, B. S. Graham, J. S. McLellan, and A. B. Ward, 2018. Stabilized coronavirus spikes are resistant to conformational changes induced by receptor recognition or proteolysis. *Scientific reports* 8:1–11.

27. Cai, Y., J. Zhang, T. Xiao, H. Peng, S. M. Sterling, R. M. Walsh, S. Rawson, S. Rits-Volloch, and B. Chen, 2020. Distinct conformational states of SARS-CoV-2 spike protein. *Science* .

28. Ingólfsson, H. I., M. N. Melo, F. J. van Eerden, C. Arnarez, C. A. López, T. A. Wassenaar, X. Periole, A. H. de Vries, D. P. Tieleman, and S. J. Marrink, 2014. Lipid Organization of the Plasma Membrane. *J. Am. Chem. Soc.* 136:14554–14559.

29. Meher, G., S. Bhattacharjya, and H. Chakraborty, 2019. Membrane Cholesterol Modulates Oligomeric Status and Peptide-Membrane Interaction of Severe Acute Respiratory Syndrome Coronavirus Fusion Peptide. *The Journal of Physical Chemistry B* 123:10654–10662.

30. Arcario, M. J., Y. Z. Ohkubo, and E. Tajkhorshid, 2011. Capturing Spontaneous Partitioning of Peripheral Proteins using a Biphasic Membrane-Mimetic Model. *J. Phys. Chem. B* 115:7029–7037.

31. Ohkubo, Y. Z., T. V. Pogorelov, M. J. Arcario, G. A. Christensen, and E. Tajkhorshid, 2012. Accelerating Membrane Insertion of Peripheral Proteins with a Novel Membrane Mimetic Model. *Biophys. J.* 102:2130–2139.

32. Pogorelov, T. V., J. V. Vermaas, M. J. Arcario, and E. Tajkhorshid, 2014. Partitioning of Amino Acids into a Model Membrane: Capturing the Interface. *J. Phys. Chem. B* 118:1481–1492.

33. Vermaas, J. V., T. V. Pogorelov, and E. Tajkhorshid, 2017. Extension of the Highly Mobile Membrane Mimetic to Transmembrane Systems through Customized *in Silico* Solvents. *J. Phys. Chem. B* 121:3764–3776.

34. Baylon, J. L., I. L. Lenov, S. G. Sligar, and E. Tajkhorshid, 2013. Characterizing the Membrane-Bound State of Cytochrome P450 3A4: Structure, Depth of Insertion, and Orientation. *J. Am. Chem. Soc.* 135:8542–8551.

35. Vermaas, J. V., and E. Tajkhorshid, 2014. A Microscopic View of Phospholipid Insertion into Biological Membranes. *J. Phys. Chem. B* 118:1754–1764.

36. Arcario, M. J., and E. Tajkhorshid, 2014. Membrane-Induced Structural Rearrangement and Identification of a Novel Membrane Anchor in Talin F2F3. *Biophys. J.* 107:2059–2069. http://linkinghub.elsevier.com/retrieve/pii/S0006349514009965.

37. Blanchard, A. E., M. J. Arcario, K. Schulten, and E. Tajkhorshid, 2014. A Highly Tilted Membrane Configuration for the Pre-Fusion State of Synaptobrevin. *Biophys. J.* 107:2112–2121.

38. Vermaas, J. V., and E. Tajkhorshid, 2014. Conformational heterogeneity of α-synuclein in membrane. *Biochim. Biophys. Acta, Biomembr.* 1838:3107–3117.

39. Wu, Z., and K. Schulten, 2014. Synaptotagmin's Role in Neurotransmitter Release Likely Involves Ca^{2+}-induced Conformational Transition. *Biophys. J.* 107:1156–1166.

40. Rhéault, J.-F., E. Gagné, M. Guertin, G. Lamoureux, M. Auger, and P. Lagüe, 2015. Molecular Model of Hemoglobin N from *Mycobacterium tuberculosis* Bound to Lipid Bilayers: A Combined Spectroscopic and Computational Study. *Biochemistry* 54:2073–2084.

41. Madsen, J. J., Y. Z. Ohkubo, G. H. Peters, J. H. Faber, E. Tajkhorshid, and O. H. Olsen, 2015. Membrane Interaction of the Factor VIIIa Discoidin Domains in Atomistic Detail. *Biochemistry* 54:6123–6131.

42. Vermaas, J. V., J. L. Baylon, M. J. Arcario, M. P. Muller, Z. Wu, T. V. Pogorelov, and E. Tajkhorshid, 2015. Efficient Exploration of Membrane-Associated Phenomena at Atomic Resolution. *J. Membr. Biol.* 248:563–582.

43. Baylon, J. L., J. V. Vermaas, M. P. Muller, M. J. Arcario, T. V. Pogorelov, and E. Tajkhorshid, 2016. Atomic-level description of protein-lipid interactions using an accelerated membrane model. *Biochim. Biophys. Acta, Biomembr.* 1858:1573–1583.

44. Muller, M. P., Y. Wang, J. H. Morrissey, and E. Tajkhorshid, 2017. Lipid Specificity of the Membrane Binding Domain of Coagulation Factor X. *J. Thromb. Haem.* 15:2005–2016.

bioRxiv preprint doi: https://doi.org/10.1101/2020.10.27.357350; this version posted October 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

45. Vermaas, J. V., and E. Tajkhorshid, 2017. Differential Membrane Binding Mechanics of Synaptotagmin Isoforms Observed at Atomic Detail. *Biochemistry* 56:281–293.

46. Baylon, J. L., and E. Tajkhorshid, 2015. Capturing Spontaneous Membrane Insertion of the Influenza Virus Hemagglutinin Fusion Peptide. *J. Phys. Chem. B* 119:7882–7893.

47. Katoh, K., K. Misawa, K. Kuma, and T. Miyata, 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30:3059–3066. http://nar.oxfordjournals.org/content/30/14/3059.abstract.

48. Clamp, M., J. Cuff, S. M. Searle, and G. J. Barton, 2004. The Jalview Java alignment editor. *Bioinformatics* 20:426–427.

49. Wrapp, D., N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, 2020. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367:1260–1263.

50. Gui, M., W. Song, H. Zhou, J. Xu, S. Chen, Y. Xiang, and X. Wang, 2017. Cryo-electron microscopy structures of the SARS-CoV spike glycoprotein reveal a prerequisite conformational state for receptor binding. *Cell research* 27:119–129.

51. Humphrey, W., A. Dalke, and K. Schulten, 1996. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* 14:33–38.

52. Jo, S., T. Kim, V. G. Iyer, and W. Im, 2008. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 29:1859–1865.

53. Wells, D. B., V. Abramkina, and A. Aksimentiev, 2007. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *J. Chem. Phys.* 127:125101.

54. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802.

55. Phillips, J., D. Hardy, J. Maia, J. Stone, J. Ribeiro, R. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. dos Reis Melo, B. Radak, R. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. Kale, K. Schulten, C. Chipot, and E. Tajkhorshid, 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* 153.

56. Bauckhage, C., 2015. Numpy/scipy recipes for data science: k-medoids clustering. *Researchgate. Net, February* .

57. Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12:2825–2830.

58. Madu, I. G., S. Belouzard, and G. R. Whittaker, 2009. SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion. *Virology* 393:265–271.

59. Wimley, W. C., and S. H. White, 1996. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* 3:842–848.