

Targeted Molecular Dynamics Study of SARS-CoV-2 Spike Protein Head Protomer Opening and Closing Transitions

1 Introduction

The global COVID-19 pandemic caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) poses an incredible threat to human health, with a total of 147 million confirmed cases and 3.11 million deaths as of April 23rd, 2021. The spike protein (S-protein) of SARS-CoV-2 in the prefusion state is a trimer with three chains. Each has S1 and S2 subunits. S1 regulates receptor binding and S2 is responsible for viral fusion and entry. S1 subunit consists of four structural domains, namely, the N-terminal domain (NTD), the receptor binding domain (RBD), and two subdomains (SD1, SD2) [1]. Infection is initiated by the RBD binding to angiotensin converting enzyme (ACE2) receptors in human cells [2, 3]. The prerequisite for this binding event is a conformational change of the S-protein head trimer from the closed state with all three RBDs in the down position to the open state with at least one RBD erected [1, 4]. Understanding the transition mechanism enabling receptor binding would speed up the development of novel therapeutics and vaccines and hopefully terminate the pandemic. Therefore, investigating the transition path regarding the S-protein closed-to-open (or down-to-up) conformational change is currently a popular research topic.

Molecular dynamics (MD) simulations are widely used to study protein dynamics to gain atomic-level insights of the function machinery. However, to sample the opening or closing transition of the SARS-CoV-2 spike glycoprotein in a conventional MD simulation is extremely computationally expensive since the maximum achievable simulation time (up to microsecond for now [5]) for such a large system is still trivial compared with the length of the real biological process. There are some published works with efforts in accelerating the conformational sampling such that it is possible to observe the transition in a reasonable computational time. A path of the head trimer from the closed to the open state was first reported in April 2020 by Mert Gur et al. [6] via extracting the minimum energy path from a 2D principal component conformational space sampled by a series of steered MD (SMD) simulations and conventional MD simulations. Almost at the same time, another opening transition of the spike glycoprotein head trimer was captured by Maxwell I. Zimmerman et al. [7] by employing a goal-oriented adaptive sampling algorithm, FAST, in collaboration with the Folding@home distributed computing platform. The FAST search algorithm implemented a path optimization by iteratively building Markov state models (MSMs), ranking states based on the exploration (breadth of sampled conformational space) - exploitation (opening of RBD) tradeoffs, and restarting MD simulations from the top ranked state [8]. In addition, Hao Tian et al. [9] applied MSMs to two 10- μ s long MD trajectories of trimeric heads in open and closed states and then divided the conformational space into eight macrostates, based on which the most probable transition channel was derived via the transition path theory. Later on, in November 2020, Lorenzo Casalino et al. [10] used the weighted ensemble (WE) enhanced sampling method to simulate the opening transition of the spike glycoprotein head trimer by running many short MD simulations in parallel along two selected reaction coordinates, the distance between RBD and the spike helical core as well as the root mean square deviation (RMSD) of RBD from the target open structure. Despite the research listed above, there is not yet a universally accepted conformational transition path between the down and up states of S-protein. The underlying mechanism for the transition still remains a lot to be explored.

Targeted molecular dynamics (TMD) [11] is a perturbation method derived from MD. With selected atoms subject to a biasing potential, the structure is pulled from the initial configuration to the target in terms of predefined collective variables. Even though the large-scale-first bias was reported as a common problem residing in most perturbation sampling methods like SMD and TMD [12-14], TMD has been successfully applied to produce plausible conformational transition pathways for many large proteins [15-21]. Snapshots of conformations generated along the TMD path can be used in umbrella sampling and the

string method to create free energy landscape along reaction coordinates [12, 22, 23]. In this study, we focused on the opening and closing transitions of a single chain in the SARS-CoV-2 spike glycoprotein head. We used RMSD-based TMD simulations to guide the head protomers from down to up and up to down state. A series of analyses were performed to unravel both the global and local features displayed along the transition. Five replicates of opening and closing trajectories were produced with different initial and target structures sampled in the conventional MD simulations to show the repeatability and reliability of results.

2 Methods

Model preparation. The head protomers used in our simulations were extracted from CHARMM-GUI [24, 25] fully-glycosylated S-protein head-only trimeric models based on 6VXX [26] and 6VSB [1] in Protein Data Bank (PDB), with only one chain kept and the other two chains truncated. The down protomer was chain A of 6VXX structure (6VXX_PROA). The up protomer was the only chain with RBD in the up position in 6VSB structure (6VSB_UPprotomer). Both head protomers included residues 1-1146, with 20 glycosylation sites, namely, N17, N61, N74, N122, N149, N165, N234, N282, T323, N331, N343, N603, N616, N657, N709, N717, N801, N1074, N1098, and N1134. Amino acid sequence numbers in the five structural domains are: residues 27-304 for NTD, 305-328 for SD1, 335-521 for RBD, 522-672 for SD2 and 687-1146 for S2. The glycosylated head protomers in down and up states were solvated in a water box of $250 \times 140 \times 140 \text{ \AA}^3$ and $220 \times 180 \times 140 \text{ \AA}^3$, respectively, with at least 10 Å cushion of water in each direction to ensure enough space for the opening and closing transitions. TIP3P water molecules were used. Na^+ and Cl^- ions with a concentration of 0.15 M were added to neutralize the systems. With the glycans pruned, two non-glycosylated head protomers in the down and up states were prepared in the same way. In total, we have four systems. The sizes of systems in down and up states are ~535,000 and ~490,000 atoms, respectively.

MD simulations. All simulations were performed using NAMD 2.13 package [27] with CHARMM36 force field [28] in the NPT ensembles at 300 K and 1.01325 bar, controlled by Langevin dynamics with a damping coefficient of 5 ps^{-1} and the Nosé-Hoover Langevin piston with an oscillation period of 100 fs and a decay time of 50 fs. The cutoff for the non-bonded van der Waals interactions and short-range electrostatic interactions was set to 12 Å with the switching function active. Long-range electrostatic interactions were calculated by Particle Mesh Ewald (PME) method with a maximum grid spacing of 1 Å. To maintain the sugar chair conformations for the glycosylated systems, dihedral restraints with a force constant of 1 kcal/mol/rad² were applied. Periodic boundary conditions were used. Time step was set to 1 fs/step. The systems were first minimized for 50,000 steps followed by a short relaxation of 0.05 ps. The production run was performed for 50 ns for each system. Frames were output every 1 ps. The RMSD of the ordered regions (helices and sheets) in the MD simulations began to level off at around 20 ns. Starting from 20 ns, five configurations in up and down states at a spacing of 5 ns were saved and used as initial and target structures for the TMD runs.

TMD simulations. TMD simulations were performed by adding a harmonic biasing potential to the system using NAMD Collective Variable-based Calculations (Colvars) module. RMSD from reference positions was selected as the collective variable thus the potential energy function has the form,

$$U = \frac{1}{2} \times k \times (RMSD(t) - RMSD^*(t))^2$$

where k is the force constant. $RMSD(t)$ is the best-fit RMSD of current Cartesian coordinates at time t from the prescribed target coordinates of the TMD atoms. The desired RMSD at time t , $RMSD^*(t)$, decreases linearly from the prescribed initial value to the final value. In our TMD runs, biasing forces were applied to protein heavy atoms (8943 TMD atoms in total) to guide the protomer from the initial configuration

towards the target configuration. For each pair of opening and closing runs, the initial $RMSD^*$ was set to RMSD between the two starting configurations and the final $RMSD^*$ was set to zero. Since spike protein head protomers in up and down states were metastable and covered a relatively large conformational space without the stabilization from neighboring protomers, the initial $RMSD^*$ ranged from 9.68 Å to 11.77 Å for different replicates. Each TMD run was completed in 4 ns and followed by a 3-ns conventional MD run with biasing forces removed. Force constant k is set to 20,000 kcal/mol/Å² such that the RMSD from the target structure reached around 0.5 Å in the end of TMD simulation and the free structure in the following MD run would not have severe dihedral outliers compared with those detected in the previous 50-ns MD simulations using MolProbity geometry check [29].

Distance, angle and dihedral calculations. NTD-RBD distance is the distance between center of mass (COM) of NTD and COM of RBD. RBD-SD2 angle is defined by three points corresponding to the COM of RBD and C_α atoms of residue Q580 and N603 in the turns of two β-hairpins in SD2. RBD orientation dihedral is defined by C_α atoms of residues D405-G526-V622-V991, including one residue in RBD, two residues in SD2 and one in S2. Note Peng et al. [30] used angle D405-V622-V991 from normal mode analysis to describe the up and down state of SARS-CoV-2 spike protein. But in our MD runs of the head protomer, this angle could not distinguish up from down state very well due to the removal of the other two chains. Instead, we used pseudo dihedral D405-G526-V622-V991 in our study. The distance, angle, dihedral were calculated via Python scripts with MDAnalysis 1.0.0 [31, 32].

Hydrogen bond and salt bridge calculations. Hydrogen bond interactions between glycans and the protein residues were analyzed by MDAnalysis. Distance cutoff for donor heavy atom and acceptor heavy atom was set to 3.5 Å. 30° was used as the donor heavy atom, donor hydrogen atom and the acceptor angle cutoff. Detected hydrogen bonds were grouped by pairs of the N- or O-linked glycan and the unique protein residue forming the H-bond. The frequency for each interacting pair is defined as the percentage of simulation time when the hydrogen bonds within that pair existed. Salt bridge analyses in the opening and closing transitions were performed by Visual Molecular Dynamics (VMD) [33] with a distance cutoff of 6 Å between basic nitrogen and acidic oxygen.

Principal component analysis (PCA). PCA was conducted by scikit-learn 0.22.1 [34] in Python. First, all conformations sampled in the MD simulations of up and down glycosylated head protomers and TMD runs of the opening and closing transitions were aligned to the down state crystal structure based on C_α atoms with MDAnalysis. Stacking Cartesian coordinates of 1146 C_α atoms, each conformation ended up with 3438 features (1146 × 3). Principal components were calculated upon the MD data. Then all conformations in MD and TMD runs were projected onto the lower dimensional space using the first two principal components, PC1 and PC2. The explained variance ratio by the first two components reaches up to 83.5%. Porcupine plots of PC1 and PC2 showing the dominant motions were made by Tcl script and VMD.

Native contact analysis (NCA). We define a native contact is formed if the distance between C_α atoms of two protein residues is less than 8 Å [35, 36]. The fraction of native contacts at time t , Q(t), is the percentage of native contacts in current structure with respect to those in the reference structure. Evolutions of native contacts along all transition paths were plotted in a 2D space made up of fraction Q1 and Q2, of which the references structures are the initial and final conformations, respectively. NCA was performed using MDAnalysis in Python.

3 Results

3.1 Glycans role in restraining the motion of RBD.

It has been reported that interactions with neighboring protomers play an important role in stabilizing the head trimer in both the down and up states [37, 38]. Thus, in our 50-ns MD simulations of the S-protein head protomers, without those interactions, the structures were metastable and expected to occupy a large conformational space. For the down head protomers, the non-glycosylated one was observed to be even more flexible with a continuous large motion of RBD moving towards and away from NTD and S2. Figure 1 (a) shows that the NTD-RBD distance of the glycosylated head protomer fluctuated around a steady value after 13 ns, while the distance for the non-glycosylated structure kept undergoing large changes in the entire simulation. The snapshots of the two systems at 13 ns are shown in Fig. 1 (b) and (c), respectively. Besides stabilizing neighboring protomers in the timer, we expect the N- and O-linked glycans also contribute to restricting the motion of the protomer itself.

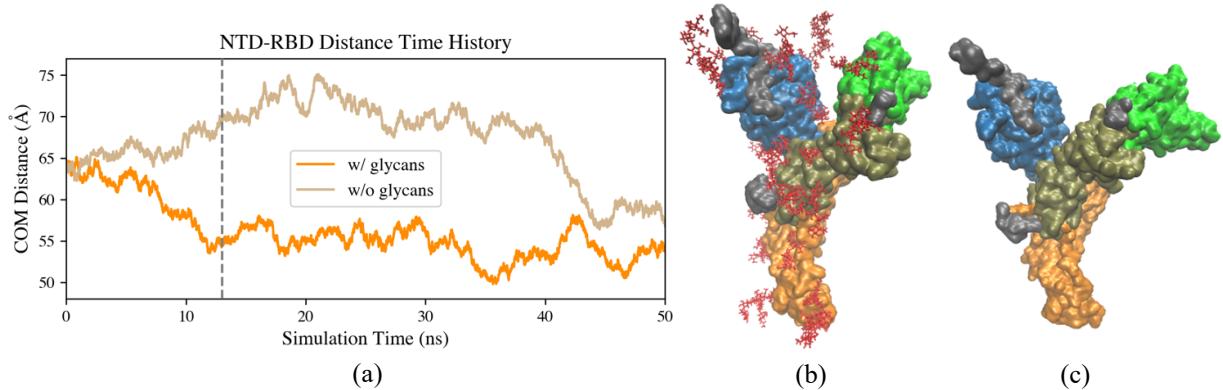


Fig. 1 (a) Time histories of NTD-RBD distances of the glycosylated and non-glycosylated S-protein head protomers in the down state MD simulations. The dashed line at 13 ns marks when the glycoprotein reached a nearly stable NTD-RBD distance. **(b)** The glycosylated down head protomer at 13 ns. **(c)** The non-glycosylated down head protomer at 13 ns. Proteins and glycans are shown with surface and licorice representation, respectively. Color scheme: glycans in red, NTD in blue, SD1 and SD2 in tan, RBD in green, S2 in orange and residues of other undefined regions in gray.

We calculated the hydrogen bond interactions between glycans and protein residues in the MD simulation of the down head protomer. In Fig. 2 (a), results are shown for interacting pairs with frequencies larger than 20%. Most of the glycans primarily formed H-bonds with protein residues in the same domain as the glycosylation site, including glycans linked to N17, N61, N122, N149, N234, N282 in NTD, N343 in RBD, T323 in SD1, N657 in SD2 and N709, N717, N801, N1074, N1098, N1134 in S2. There were three glycans frequently interacting with protein residues in other domains, forming inter-domain H-bonds between protein and glycan. Critical inter-domain H-bonds existed between O-glycan linked to T323 in NTD and residue H625 (29%) in SD2, N-glycan linked to N603 in SD2 and residues E281 (22%), T307 (22%) in NTD, N-glycan linked to N331 near RBD and residue Q580 (61%) in SD2. Among them, H-bonds formed by G0F at N331 and residues in SD2 illustrated in Fig. 2 (c) are more interesting since they may explain the less flexible RBD-SD2 hinge angle observed in the MD simulation of the glycosylated down head protomer compared with the non-glycosylated one. As shown in Fig. 2 (b), the RBD-SD2 angle converged to around 90° after 13 ns for the glycosylated structure, while the one without glycan displayed a relatively large range of angle values and continuous changes. In our 50-ns MD simulation of the down head protomer, the glycan G0F worked like a “hinge stiffener” to prevent RBD from rotating around the hinge towards a further open RBD-SD2 angle. The restriction of G0F linked to N331 on the motion of RBD is also correlated with the observation regarding NTD-RBD distance mentioned previously. To validate this hypothesis, more replicates of MD simulations and the investigation of the restraining effect with varying glycan sizes are required.

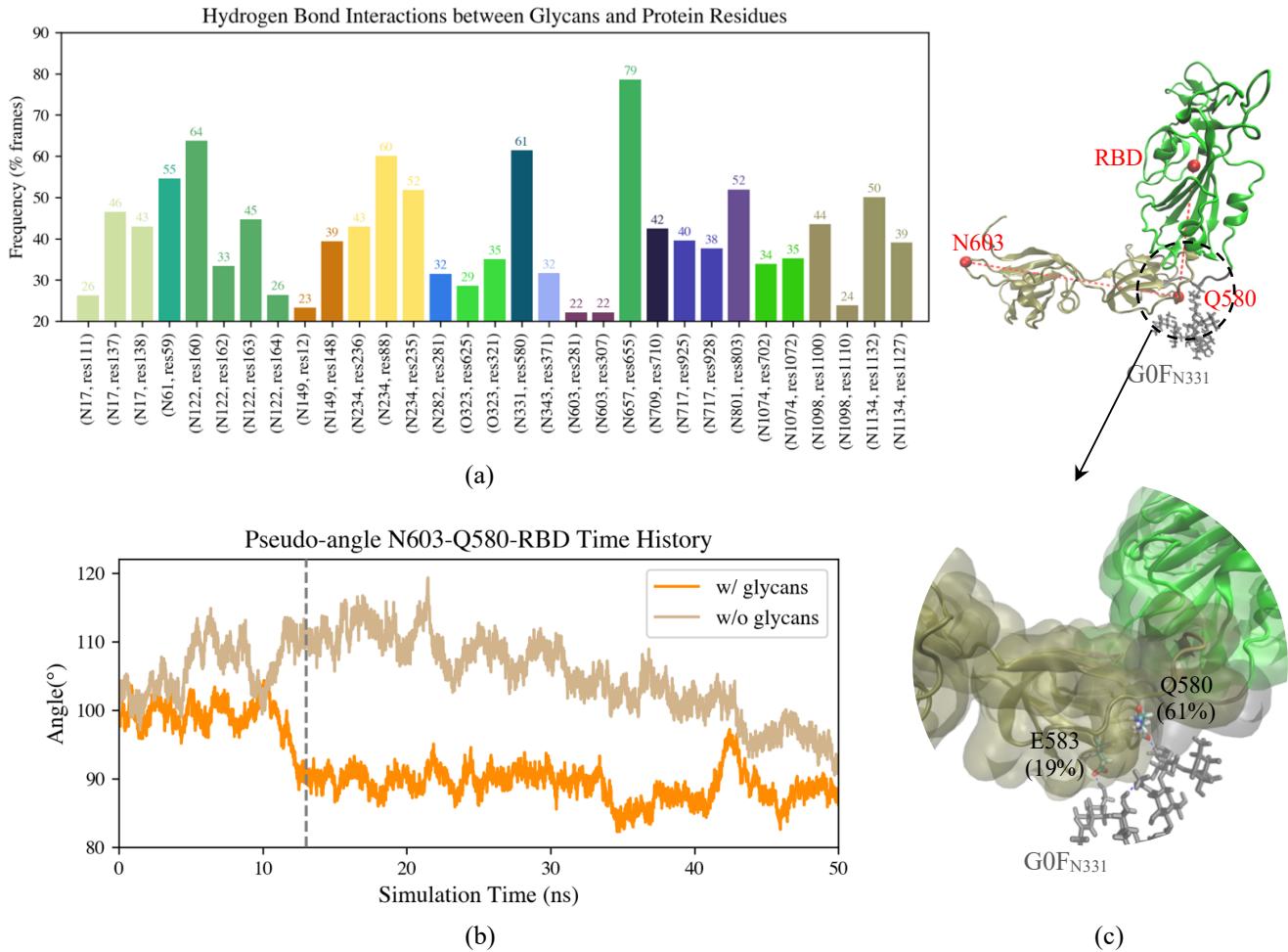


Fig. 2 (a) Critical hydrogen bond interactions between glycans and protein residues (frequency larger than 20%) in MD simulations of the down head protomer. H-bonds formed by the same glycan are shown in the same color. The first entry in each tuple under the bar plot stands for the N- or O-linked glycan and the second is the protein residue forming H-bonds with that glycan. (b) Time histories of RBD-SD2 pseudo angle of down-state glycosylated and non-glycosylated S-protein head protomers in the MD simulations. The dashed line at 13 ns marks when the glycoprotein reached a nearly stable RBD-SD2 angle. (c) Top: RBD-SD2 angle and glycan GOF linked to N331 interacting with residues in SD2. Bottom: GOF forming H-bonds with E583 and Q580. Color scheme: RBD in green, SD2 in tan, undefined residues F329 to N334 in gray.

3.2 PCA and TMD path completion.

We performed PCA for the S glycoprotein head protomers in the up and down states using the Cartesian coordinates of C_α atoms after aligned to the down crystal structure (6VXX_PROA). The first and second principal components explain 70.5% and 13.0% of the data variance, respectively. Figure 3 (c) shows the projection of up and down head protomer C_α atom positions onto PC1 and PC2 as time evolved. PC1 successfully divides the data into two clusters with a large spacing. Along PC2, these two clusters occupy a shared region. Therefore, PC1 mainly accounts for the difference between up and down clusters while PC2 is more responsible for the variance within the same state displayed over time. PC1 and PC2 are visualized on the 30-ns down structure using porcupine plots in (a) and (b) panels of Fig. 3, respectively. PC2 is shown with generally smaller arrows for the less explained variance ratio compared with PC1.

Percentage contributions from different residues to PC1 and PC2 are shown in the top and bottom panels of Fig. 3 (d), respectively. It reveals that PC1 is characterized by the predominant motion of RBD and its extension into part of the SD2 linked to RBD (residues 522-595). The porcupine plot of PC1 further illustrates this motion as a combination of rotating and rising. As for PC2, there are two major movements observed regardless of some large fluctuations of the floppy loop regions in the N-terminal patch. One is the RBD moving towards NTD thus slightly closing the pocket formed by NTD, SD1, SD2 and RBD. The fluctuations of NTD-RBD distance depicted in Fig. 2 (a) also reflects this behavior to some extent. The other is the pocket approaching central helices of S2, dragged by SD1 and SD2. Since the head protomer was neither sterically protected by RBD from a neighboring protomer inserted in its pocket nor by central helices from other chains, this pocket closing and drifting motion was inevitable in our MD simulations. In the 2D PC1-PC2 space, the head protomers in the up cluster take up a broader area along PC2, which indicates the up protomers are more flexible regarding this motion. Notably, these additional motions of domains other than RBD due to the loss of stabilization from neighboring protomers found in the our systems also participated in shaping the initial and target structures of TMD thus affecting the transition paths generated accordingly.

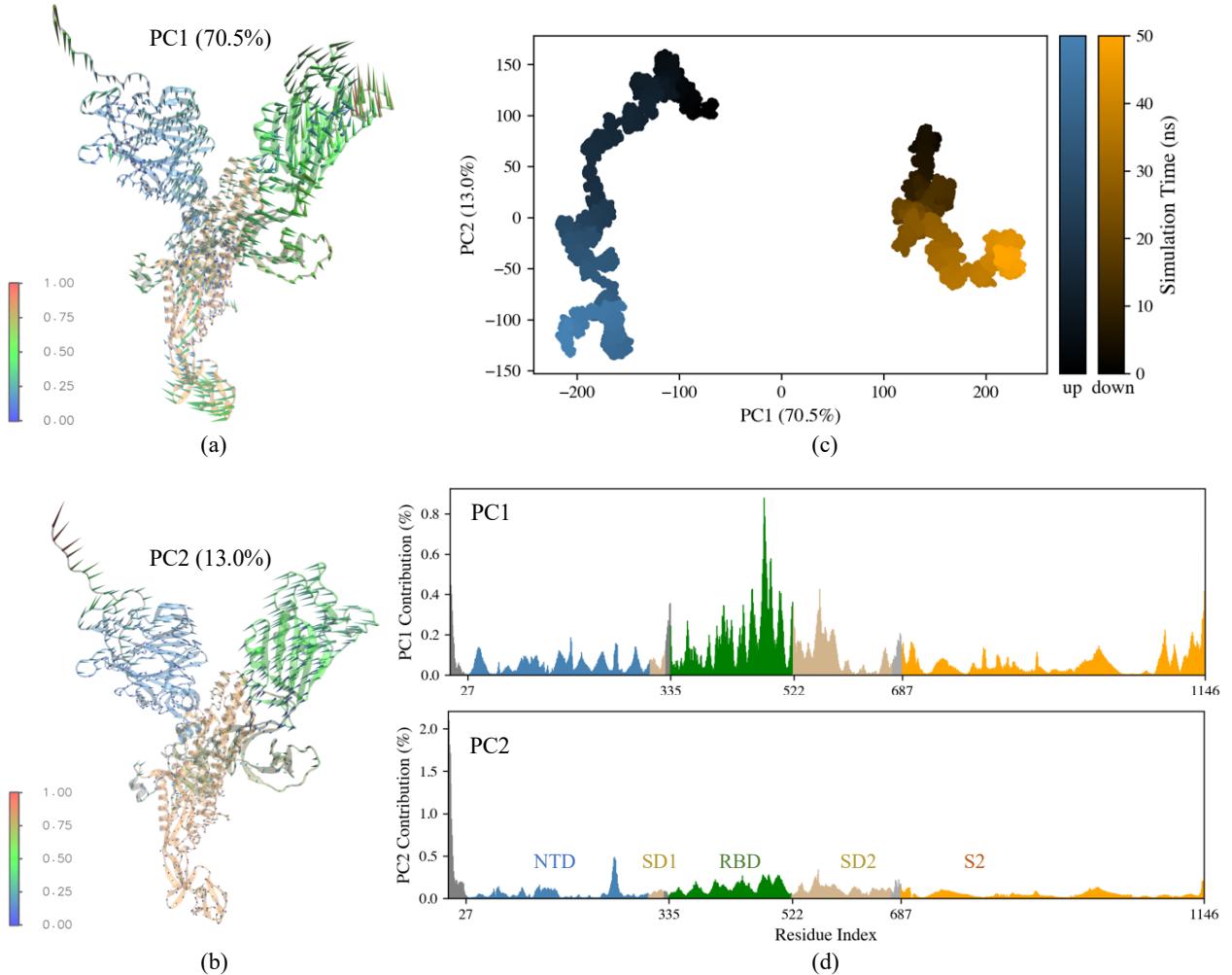


Fig 3 (a) PC1 and (b) PC2 shown on the 30-ns down structure using porcupine plots. Arrow direction and length show the direction and magnitude of motion for each C_α atom. Color bar is scaled such that the magnitude of the

maximum movement is 1. (c) Sampled glycosylated head protomers in up (blue) and down (orange) states in PC1-PC2 space. Colors show the time evolution. (d) PC1 (top) and PC2 (bottom) contributions from protein residues. Protein color scheme: NTD in blue, SD1 and SD2 in tan, RBD in green, S2 in orange and residues of other undefined regions in gray.

We obtained five replicates of spike glycoprotein head protomer opening and closing trajectories using TMD with RMSD colvar. Initial and target structures were snapshots extracted from MD runs of the up and down systems at 20 ns, 25 ns, 30 ns, 35 ns, and 40 ns. For example, for replicate 1, the opening trajectory recorded the pulling process of 20-ns down head protomer towards the position of 20-ns up head protomer and the closing one was from 20-ns up to 20-ns down conformer. In total, we have 10 trajectories. All TMD simulations decreased RMSDs (to targets) from slightly different initial values to a value between 0.51 Å and 0.54 Å as shown in the RMSD histories in Fig. 4 (a), which was close enough (less than 1 Å [39, 40]) to the target in terms of the Cartesian coordinates of the protein heavy atoms (TMD atoms). In Fig. 4 (b), we projected the sampled structures along transitions onto the 2D PC1-PC2 space after alignment. Starting from different conformations in the up or down cluster, all structures finally reached the target down or up state. Figure 4 (c) shows the structures at 0, 1, 2, 3 and 4 ns in the opening transition of replicate 3 with the target structure shown in translucent gray.

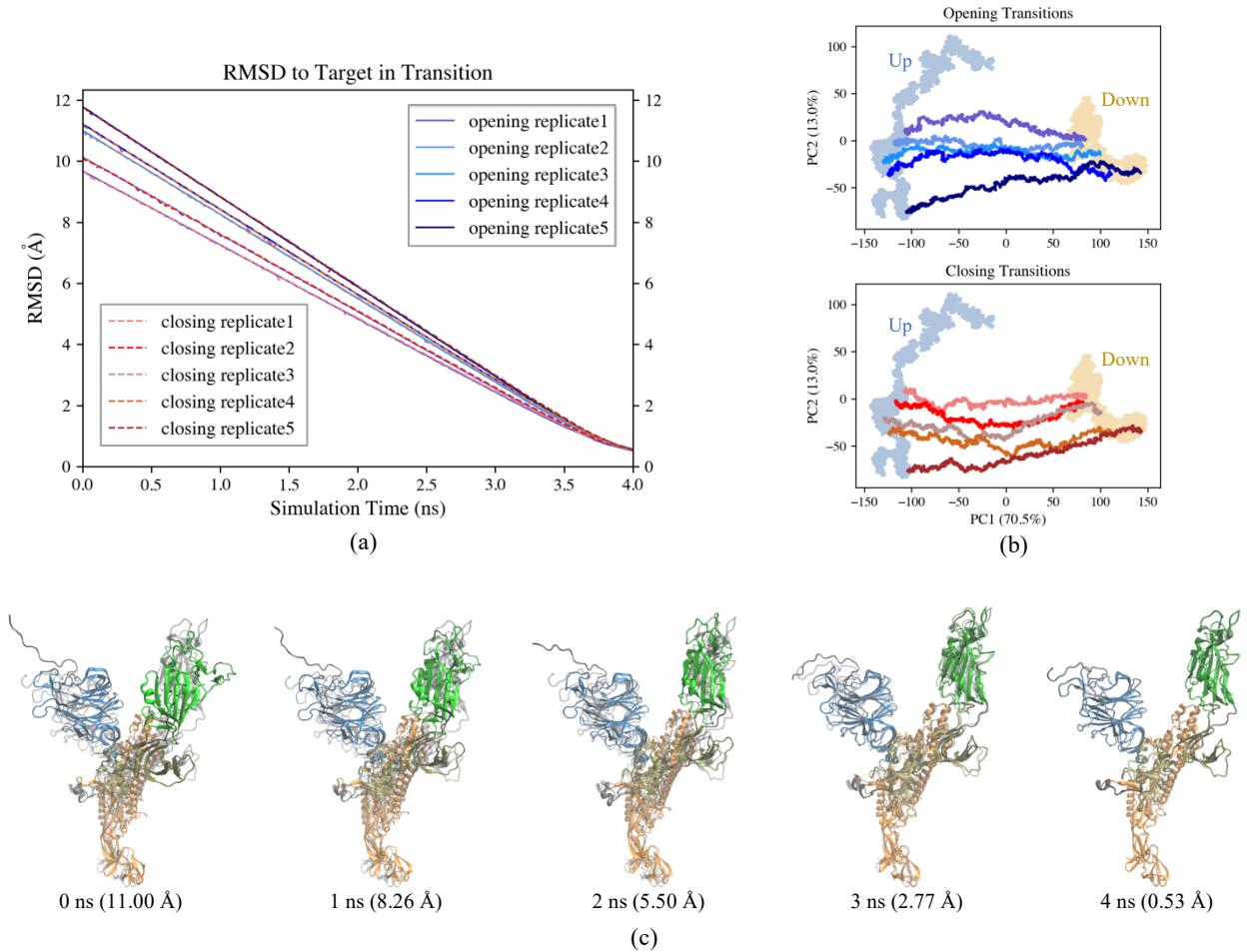


Fig. 4 (a) RMSD from the target structure in five replicates of opening and closing transitions in 4-ns TMD simulations. (b) The opening (top) and closing (bottom) trajectories in PC1-PC2 space along with the up and down clusters shown in light blue and orange, respectively. Opening trajectories are in blue-based cool colors and closing trajectories are in red-based warm colors. (c) Snapshots of replicate 3 opening transition with RMSDs to the target marked in the parentheses. Color scheme for the protomer in transition is the same as Fig. 3 (a) and (b). The target structure in the up state is shown in translucent gray with heavy atoms aligned to current structure.

3.3 RBD orientation dihedral showing the global motion in transitions.

As mentioned in PCA in 3.2, PC1 is dominated by the rotation and rising of the RBD. To further investigate this motion, we aligned SD2 of the initial down and target up head protomers used in replicate 3 opening transition (i.e., the 30-ns structures in MD simulations of down and up states), then plotted PC1 of RBD (in opaque black) on the down conformer in Fig. 5 (a). The RBD and SD2 of the up structure are shown in translucent gray. Arrows indicate the RBD in the down state would rotate up to reach the target state. To quantify this global motion in PC1 separating up and down clusters, we defined an orientation dihedral involving C_α atoms of four residues of the head protomer (see Fig. 5 (c) inset). Figure 5 (c) shows the distribution of this pseudo dihedral in our MD simulations. The head protomer in down state displayed a smaller dihedral with a mean of 17.86° , while that of the up protomer was much higher, centered at 80.54° , without overlap with the down cluster. Thus this RBD orientation dihedral becomes a good global reaction coordinate in the opening and closing transitions. In Fig. 5 (d), we show 10 TMD trajectories along this reaction coordinate. In the opening process, the pseudo dihedral D405-G526-V622-V991 monotonically increased during 0 to 3 ns. The closing transitions displayed the opposite trend. After 3 ns, the RBD orientation dihedral in all trajectories nearly reached a plateau without further increase or decrease. This can be explained by the reported large-scales-first bias of TMD simulations. The structure tends to first undergo global motions (low-frequency modes) followed by local motions (high-frequency modes). Here, the global rotation and rising of the RBD from the down to up state was almost completed in first 3 ns of the TMD run as shown by snapshots of the opening process of replicate 3 in Fig. 4 (c), thus the orientation dihedral characterizing this motion stopped growing after 3 ns. The last 1-ns simulation in the TMD trajectory mainly consisted of local adjustments of atom positions to further approach the target.

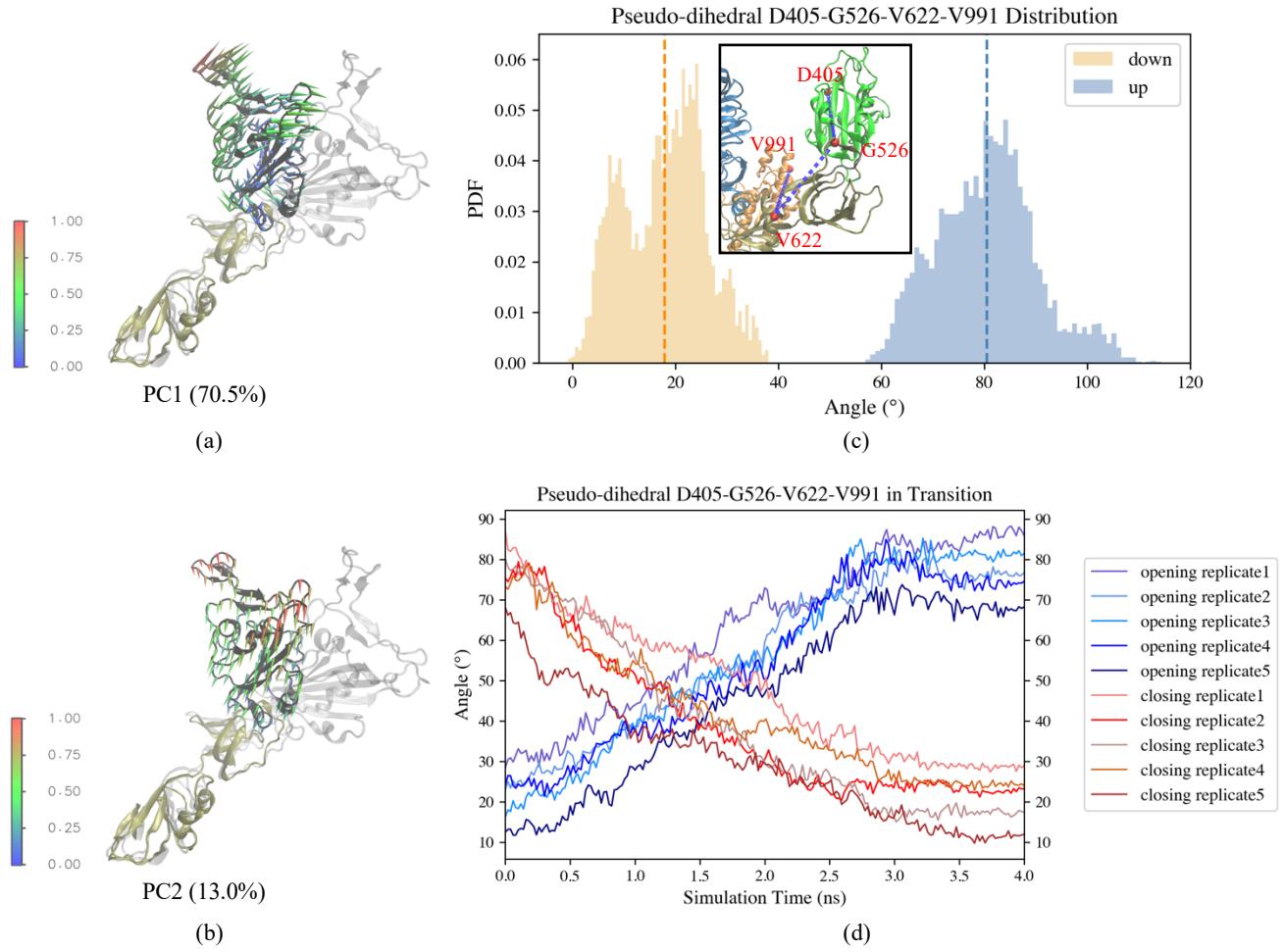


Fig. 5 (a) PC1 and (b) PC2 of RBD (black) drawn on the starting conformation of replicate 3 in the opening transition with SD2 (tan) aligned to the target up structure shown in translucent gray. (c) The RBD orientation dihedral distribution in the MD production runs. Dashed orange and blue lines mark the mean orientation dihedral in the down (17.86°) and up (80.54°) states, respectively. The inset shows this pseudo dihedral and its constituent C_α atoms. Protein color scheme is the same as Fig. 3 (a) and (b). (d) The evolution of the RBD orientation dihedral along simulation time in 10 TMD runs.

3.4 Critical backbone dihedrals in transition.

Besides the global RBD orientation dihedral, we also calculated distributions of 1145 Φ and 1145 Ψ backbone dihedrals in the MD runs of the 1146-residue up and down head protomers from 2 to 50 ns. We found there are 24 critical backbone dihedrals, among which each could successfully separate the up and down states, namely, Φ of residues 142, 256, 381, 431, 570, 601, 614, 679, 688, 813, 814, 828, 849, and Ψ of residues 122, 137, 380, 430, 569, 600, 613, 813, 827, 830, 888. For those backbone dihedrals, we investigated their evolutions in the TMD opening and closing transitions. During the transition, some of these critical dihedrals left the cluster of initial state and reached the target distribution either in a gradual fashion from 3 to 4 ns or suddenly fell into the target cluster at the end of the simulation, as shown in Fig. 6 (a) and (b), respectively. This correlates with our observations in the previous section that the last 1-ns TMD trajectories are mainly made up of local motions compared with the global RBD dihedral evolution

from 0 to 3 ns. However, some backbone dihedrals were very stable in their initial states and failed to move towards the target in most replicates, like Φ of residue A570 in Fig. 6 (c). It indicates that even though the final structures of TMD simulations reached the target states in terms of the local measurement, RMSD, and the global measurement, RBD orientation dihedral, some critical backbone dihedrals still remained in the initial state, which shows the limitation of RMSD-based TMD simulation in sampling the transition path of such a large protein molecule. It seems that this fast sampling method tends to complete the conformational change at the cost of losing some local features.

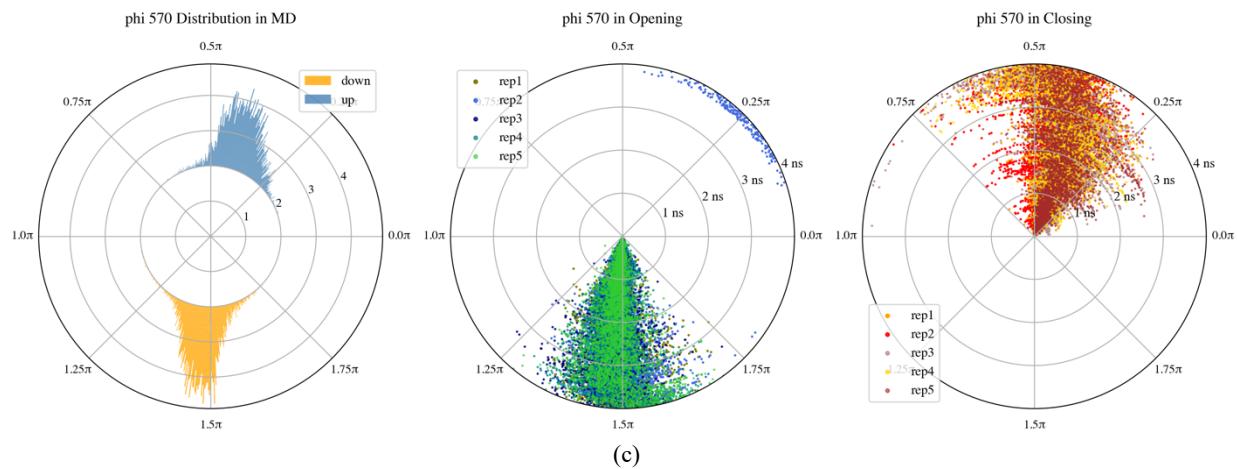
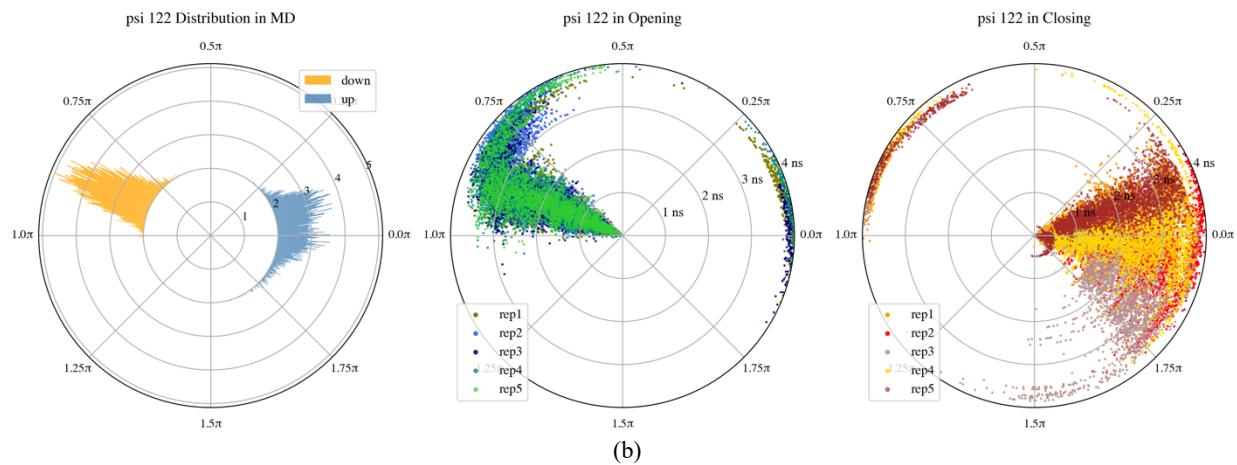
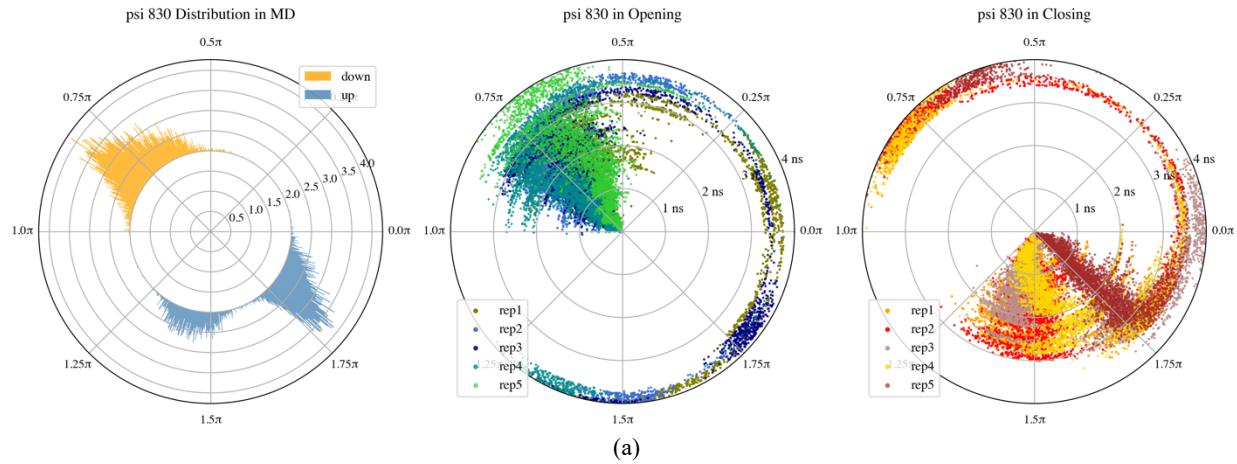


Fig. 6 Typical critical backbone dihedrals distinguishing head protomers in down state from up state. The dihedral distribution in MD simulations are shown on the left. Dihedral time history in the TMD opening and closing transitions are shown in the middle and right panel, respectively. (a) Ψ of D830 gradually arrived at the target cluster from 3 to 4 ns. (b) Ψ of N122 suddenly reached the target state at the end of TMD simulations. (c) Φ of A570 failed to leave the cluster of initial state and move towards the target distribution for most replicates.

3.5 Salt bridges forming and breaking in transitions.

Again, with only one head protomer in our simulations, we did not observe salt bridges forming across protomers as reported by Mert Gur et al. [6]. Instead, we tracked the evolution of intra-domain and inter-domain salt bridges in the head protomer opening and closing transitions obtained from TMD. Distances between basic nitrogen and acidic oxygen of detected intra-domain salt bridges are shown in Fig. 7 (a)-(f). In the down-to-up transitions, salt bridges D111-K113 in NTD, E780-R1019, E819-R815, and E1072-R1107 in S2 were formed with a fast decrease of N-O distance from 3 to 4 ns. At the same time, salt bridges D253-K77 in NTD and D820-R815 in S2 were broken. These salt bridges in the closing transitions show opposite breaking and forming events. Figure 7 (h) and (g) illustrate the locations of intra-domain salt bridges in NTD and S2, respectively. The breaking and forming of salt bridges due to the increase and decrease of N-O distances in the last 1-ns TMD runs is likely related to the aforementioned local motions observed at the end of transition. Figure 8 (a)-(f) show the N-O distance time histories of inter-domain salt bridges forming and breaking along the paths. Two salt bridges, namely, D950-K310 (S2-SD1) and R847-E281 (S2-NTD), were broken, along with salt bridges D830-K304, R847-D40, D848-R44, D848-K304 between S2 and NTD formed in opening transitions. In contrast to the intra-domain salt bridges, the breaking and forming events of inter-domain salt bridges happened either along the entire process or the first 3 ns of TMD runs, together with the global motion of the head protomer. Recall the global motion PC2 in Fig. 3 (b), displayed by structures in the same state over time. As mentioned, without steric barriers from central helices in the other two protomers of the head, NTD-SD1/SD2-RBD pocket moved towards S2 central helices in our 50-ns MD simulations. Depending on the relative positions of the pocket and central helices, different inter-domain salt bridges were formed in the up and down states. The NTD in the down protomer occupied a higher position with respect to S2 central helices as shown in Fig. 8 (h), thus NTD slightly slid downwards in the opening transition, which explains the S2-NTD salt bridge switch mechanism from the down to up state.

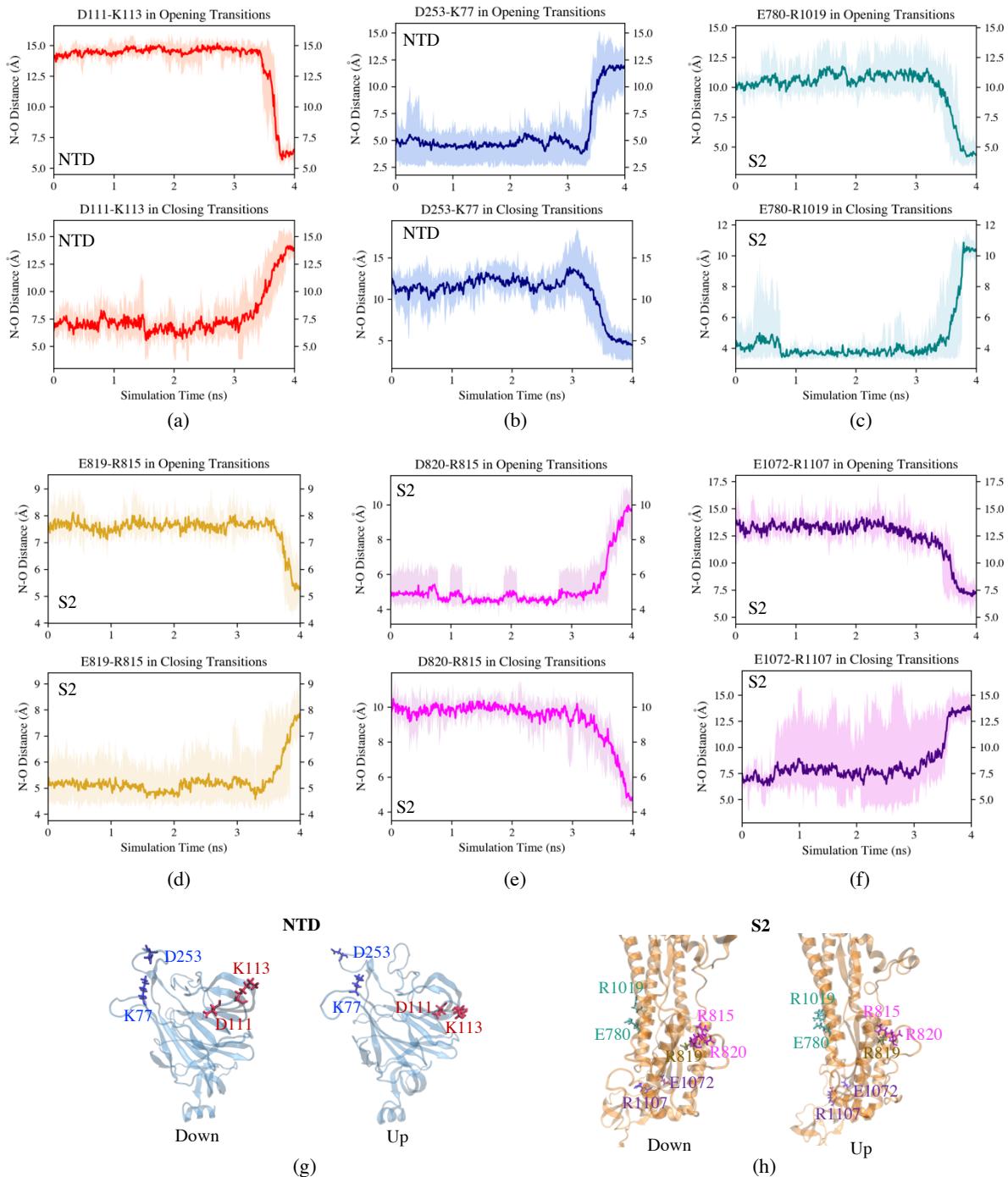


Fig. 7 Time evolution of intra-domain salt bridges in the opening (top) and closing (bottom) transitions. (a) and (b): Salt bridges in NTD. (c)-(f): Salt bridges in S2. The solid line in each panel is averaged over five replicates. The regions between minimum and maximum values of five replicates along the path are filled with the same color. Residues forming salt bridges in NTD and S2 are shown in licorice in (g) and (h) on the 30-ns down and up structures, respectively. NTD (in blue) and S2 (in orange) are drawn with cartoons. Color schemes for salt bridges in (g) and (h) are the same as (a)-(f).

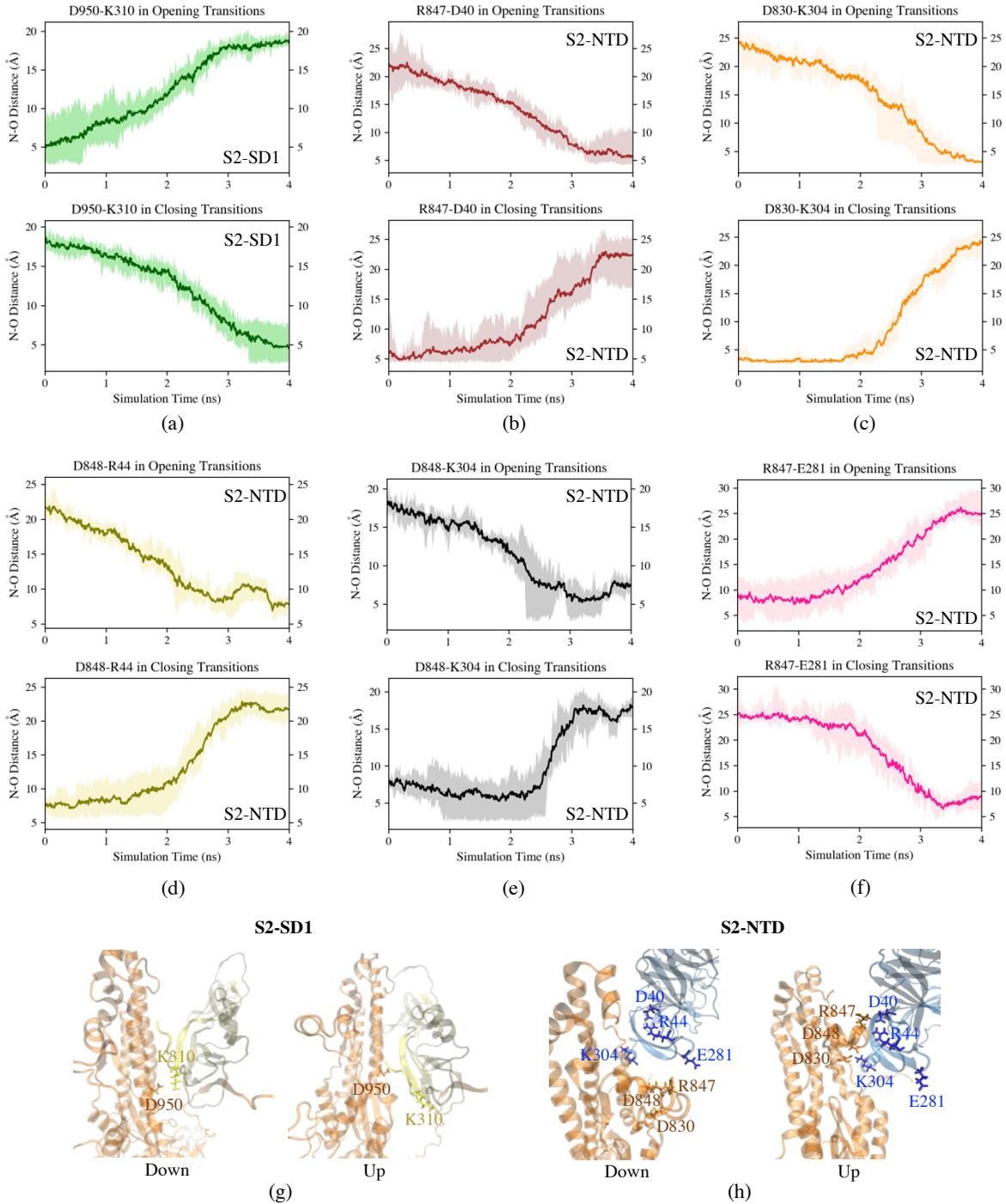


Fig. 8 Time evolution of inter-domain salt bridges in the opening (top) and closing (bottom) transitions. (a) Salt bridge between S2 and SD1. (c)-(f): Salt bridges between S2 and NTD. The solid line in each panel is averaged over five replicates. The regions between minimum and maximum values of five replicates along the path are filled with the same color. Residues forming S2-SD1 salt bridge and S2-NTD salt bridges are shown in licorice in (g) and (h) on the 30-ns down and up structures, respectively. Domains are drawn using cartoons with color scheme: NTD in blue, SD1 in tan, SD2 in yellow and S2 in orange. Salt bridge residues are colored by domains.

3.6 Native contacts evolving along the path.

Figure 9 shows the evolution of native contacts of the up (in blue) and down (in orange) protomers in our MD runs as fractions of those in the down (6VXX_PROA) and up (6VSB_UPprotomer) head protomers in crystal structures. After minimization and a short relaxation, the starting points for the up and down protomers are $(Q1, Q2) = (93.5\%, 94.7\%)$ and $(95.3\%, 94.0\%)$, respectively, both less than 100% of the corresponding crystal structures initiating the simulations. From 0 to 50 ns, head protomers in up and down states moved towards the direction of losing native contacts appearing in both crystal structures quickly and started to converge at around 20 ns. It is reasonable considering the fact that protomers were first extracted from the trimeric crystal structures then equilibrated alone in the solvent. In this way, the crystal structures may not be good reference structures for NCA of the transition due to the overlap of clusters where initial and target structures were obtained from. Therefore, instead of projecting the transition paths onto $Q1, Q2$ based on the native contacts in crystal structures, we used the initial and final frames of each trajectory as the reference structures. More specifically, $Q1$ always measures the fraction of native contacts in current conformation with respect to the initial conformation, and $Q2$ is calculated upon the final conformation. With $Q1$ and $Q2$ defined in this way, evolutions of native contacts in five replicates of down-to-up and up-to-down transitions are shown in Fig. 10 (a) and (b), respectively. Starting from $Q1 = 100\%$, the head protomers in both states underwent a decrease in $Q1$, with $Q2$ nearly kept constant, which means the structure was moving away from the initial conformation but not approaching the target in terms of native contacts. This lasted for around 3 ns. After that, the structure was guided almost in a straight line towards $Q2 = 100\%$ on the 2D $Q1$ - $Q2$ plot by forming native contacts required to achieve the final conformation, along with a further decrease in $Q1$. A similar trend for all replicates is observed, which indicates the transition paths are not reversible as a result of the TMD large-scale-first bias.

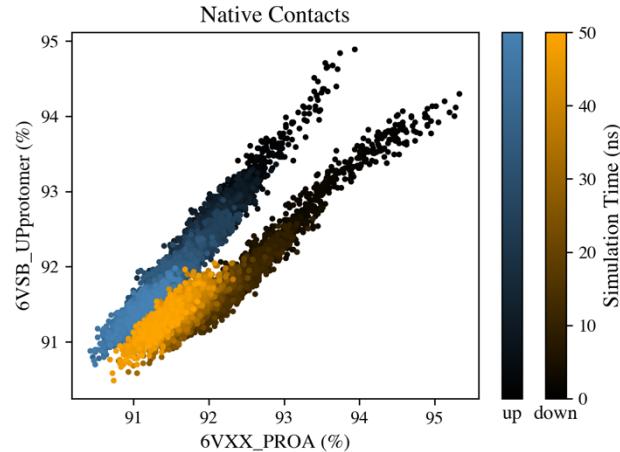
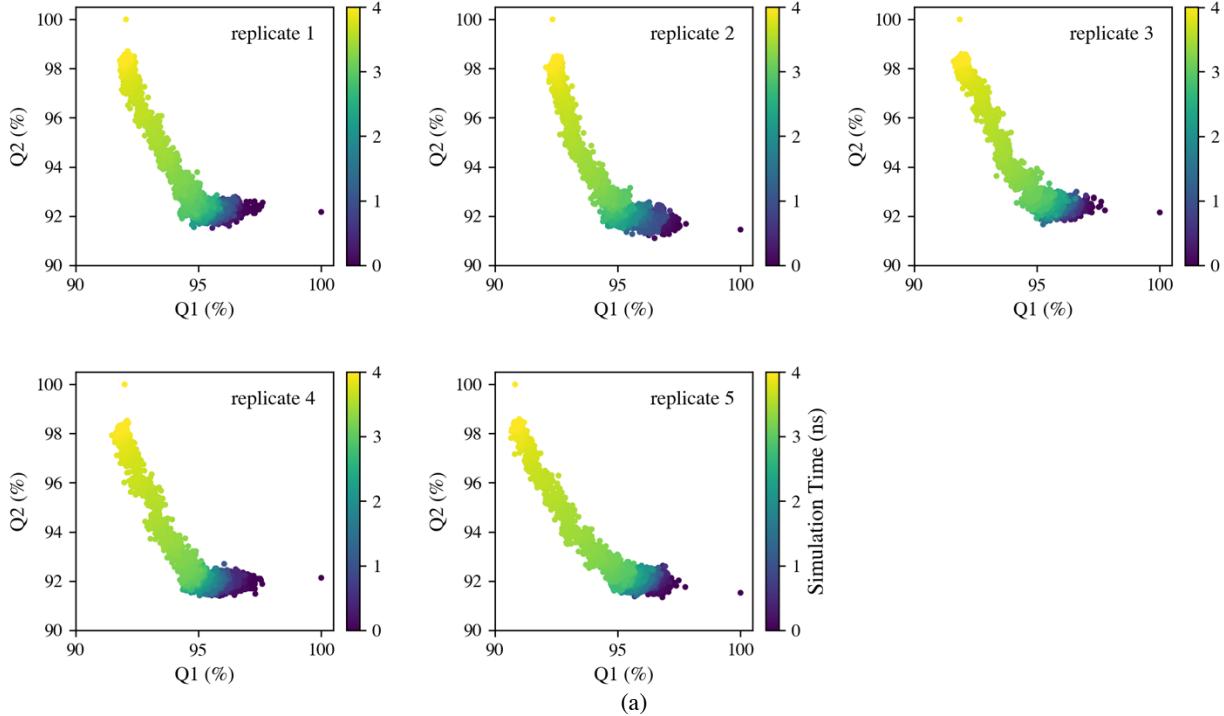


Fig. 9 Fraction of native contacts with respect to the down (6VXX_PROA) and up (6VSB_UPprotomer) head protomers in crystal structures. Data points are colored by time with up state in blue and down state in orange.

Native Contacts in Down-to-up Transitions



Native Contacts in Up-to-down Transitions

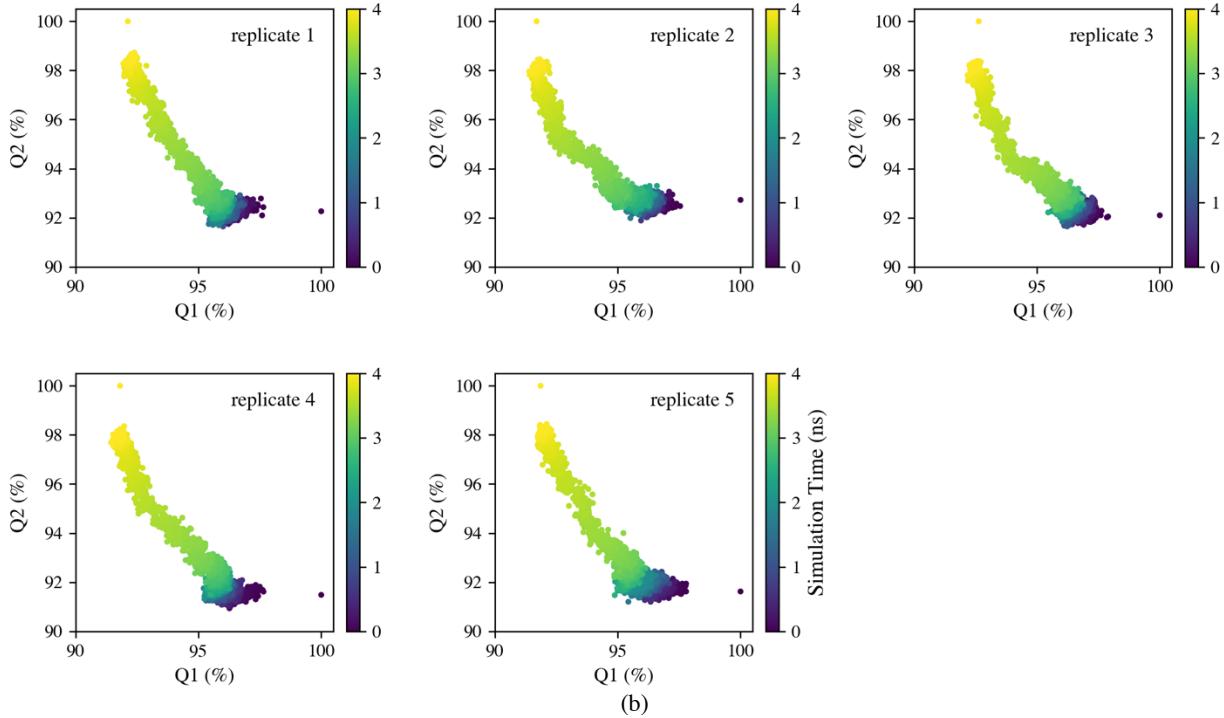


Fig. 10 Evolution of native contacts in five replicates of (a) opening and (b) closing trajectories. Q1 and Q2 measure the fraction of native contacts in the current structure with respect to the initial and final structure, respectively. Data points are colored by time.

4 Discussion and Conclusions

Main findings. In this study, we conducted RMSD-based TMD simulations on glycosylated S-protein head protomers to complete the opening and closing transitions in a short simulation time. PCA was performed on the conformations sampled in 50-ns MD simulations, from which the initial and target structures used in TMD runs were generated. PC1 is characterized by the RBD rotating up and down. PC2 captures the NTD-SD1/SD2-RBD pocket closing and drifting. In TMD simulations, each initial structure finally reached the target conformation with an RMSD around 0.5 Å. The path completion was also validated in the 2D PC1-PC2 space. We defined an RBD orientation dihedral to quantify the global movement of RBD in the protomer, which increased or decreased nearly monotonically in the opening and closing transitions, respectively. The intra- and inter-domain salt bridges related to the protomer conformational change were investigated. Also detected were some critical backbone dihedrals reluctant to undergo the transition and remaining in the initial state.

TMD large-scale-first bias. The transition was divided into two stages artificially by the bias in TMD, mainly involving the global motion or the local motion. In a real biological process, the protein should undergo these two motions in parallel. As a result, any arguments regarding the event happening sequence along the path could be problematic. Nevertheless, this artifact would barely affect observations of features displayed either by the global motion like the RBD orientation dihedral change and the inter-domain salt bridge formation, or the local motion such as the critical backbone dihedrals in transitions.

Head protomer vs. trimer. In a TMD trial run of the S-protein head trimer with the biasing potential applied to all 13,755 protein backbone atoms, one of the three protomers was guided from down to up position by decreasing the RMSD from 7.32 Å to around 0.5 Å (backbone). Tracking the trajectory of the protomer being pulled up during the simulation, the RBD orientation dihedral increased monotonically as expected and the native contacts evolved in the similar trend as the protomer system. However, we did not observe the aforementioned salt bridges forming and breaking within the protomer undergoing down-to-up transition. The different conformations one chain adopted in the protomer and trimer systems and the resulting different initial and target structures account for this inconsistency. The structural domains of neighboring protomers and their decorated glycans stabilized the relative positions of NTD, SD1 and SD2 with respect to S2 central helices and sterically prevented them from approaching S2. With small positional changes of NTD, SD1, SD2 and S2 in the up state compared with the down state, the trimer would complete the transition in a more constrained fashion, in which mainly the motion of “free-to-move” RBD and its interactions with neighboring chains could be observed. To the contrary, without these protomer-protomer interactions, the equilibrated head protomer exhibited a more compact conformation with some strong interactions formed between S2 and the NTD-SD1/SD2-RBD pocket. The locations for these strong interactions could be different for the up and down states, thus requiring the switch of interactions, like salt bridges, to complete the conformational change.

References

- [1] Wrapp, Daniel, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. McLellan. 2020. "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation." *Science (American Association for the Advancement of Science)* 367 (6483): eabb2507-1263. <https://doi.org/10.1126/science.abb2507>.
- [2] Wan, Yushun, Jian Shang, Rachel Graham, Ralph S. Baric, and Fang 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 (7): e00127-20. <https://doi.org/10.1128/jvi.00127-20>. <https://jvi.asm.org/content/jvi/94/7/e00127-20.full.pdf>.
- [3] Li, Wenhui, Michael J. Moore, Natalya Vasilieva, Jianhua Sui, Swee Kee Wong, Michael A. Berne, Mohan Somasundaran, John L. Sullivan, Katherine Luzuriaga, Thomas C. Greenough, Hyeryun Choe, and Michael Farzan. 2003. "Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus." *Nature* 426 (6965): 450-454. <https://doi.org/10.1038/nature02145>. <https://doi.org/10.1038/nature02145>.
- [4] Benton, Donald J., Antoni G. Wrobel, Pengqi Xu, Chloë Roustan, Stephen R. Martin, Peter B. Rosenthal, John J. Skehel, and Steven J. Gamblin. 2020. "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion." *Nature (London)*. <https://doi.org/10.1038/s41586-020-2772-0>.
- [5] D. E. Shaw Research, 2020. "Molecular Dynamics Simulations Related to SARS-CoV-2." *D. E. Shaw Research Technical Data*. http://www.deshawresearch.com/resources_sarscov2.html
- [6] Gur, Mert, Elhan Taka, Sema Zeynep Yilmaz, Ceren Kilinc, Umut Aktas, and Mert Golcuk. 2020. "Conformational transition of SARS-CoV-2 spike glycoprotein between its closed and open states." *The Journal of chemical physics* 153 (7): 75101. <https://doi.org/10.1063/5.0011141>.
- [7] Zimmerman, Maxwell I., Justin R. Porter, Michael D. Ward, Sukrit Singh, Neha Vithani, Artur Meller, Upasana L. Mallimadugula, Catherine E. Kuhn, Jonathan H. Borowsky, Rafal P. Wiewiora, Matthew F. D. Hurley, Aoife M Harbison, Carl A Fogarty, Joseph E. Coffland, Elisa Fadda, Vincent A. Voelz, John D. Chodera, and Gregory R. Bowman. 2020. "Citizen Scientists Create an Exascale Computer to Combat COVID-19." *bioRxiv*: 2020.06.27.175430. <https://doi.org/10.1101/2020.06.27.175430>.
- [8] Zimmerman, Maxwell I., and Gregory R. Bowman. 2015. "FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs." *Journal of chemical theory and computation* 11 (12): 5747-5757. <https://doi.org/10.1021/acs.jctc.5b00737>.
- [9] Tian, Hao, and Peng Tao. 2020. "Deciphering the protein motion of S1 subunit in SARS-CoV-2 spike glycoprotein through integrated computational methods." *Journal of Biomolecular Structure and Dynamics*: 1-8. <https://doi.org/10.1080/07391102.2020.1802338>. <https://doi.org/10.1080/07391102.2020.1802338>.
- [10] Casalino, L., A. Dommer, Z. Gaieb, E. P. Barros, T. Sztain, S. H. Ahn, A. Trifan, A. Brace, A. Bogetti, H. Ma, H. Lee, M. Turilli, S. Khalid, L. Chong, C. Simmerling, D. J. Hardy, J. D. C. Maia, J. C. Phillips, T. Kurth, A. Stern, L. Huang, J. McCalpin, M. Tatineni, T. Gibbs, J. E. Stone, S. Jha, A. Ramanathan, and R. E. Amaro. 2020. "AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics." *bioRxiv*. <https://doi.org/10.1101/2020.11.19.390187>.
- [11] Schlitter, J., M. Engels, and P. Krüger. 1994. "Targeted molecular dynamics: A new approach for searching pathways of conformational transitions." *Journal of molecular graphics* 12 (2): 84-89. [https://doi.org/10.1016/0263-7855\(94\)80072-3](https://doi.org/10.1016/0263-7855(94)80072-3).
- [12] Ovchinnikov, Victor, and Martin Karplus. 2012. "Analysis and Elimination of a Bias in Targeted Molecular Dynamics Simulations of Conformational Transitions: Application to Calmodulin." *The journal of physical chemistry. B* 116 (29): 8584-8603. <https://doi.org/10.1021/jp212634z>.
- [13] Noé, Frank, and Stefan Fischer. 2008. "Transition networks for modeling the kinetics of conformational change in macromolecules." *Current opinion in structural biology* 18 (2): 154-162. <https://doi.org/10.1016/j.sbi.2008.01.008>.

- [14] Huang, He, Elif Ozkirimli, and Carol Beth Post. 2009. "Comparison of Three Perturbation Molecular Dynamics Methods for Modeling Conformational Transitions." *Journal of chemical theory and computation* 5 (5): 1304-1314. <https://doi.org/10.1021/ct9000153>.
- [15] Yoon, Hyun Jung, Sungmin Lee, Sun Joo Park, and Sangwook Wu. 2018. "Network approach of the conformational change of c-Src, a tyrosine kinase, by molecular dynamics simulation." *Scientific reports* 8 (1): 5673-10. <https://doi.org/10.1038/s41598-018-23964-5>.
- [16] Schulz, Robert, Attilio V. Vargiu, Paolo Ruggerone, and Ulrich Kleinekathöfer. 2015. "Computational Study of Correlated Domain Motions in the AcrB Efflux Transporter." *BioMed research international* 2015: 487298-12. <https://doi.org/10.1155/2015/487298>.
- [17] Xiao, Xiuchan, Xiaojun Zeng, Yuan Yuan, Nan Gao, Yanzhi Guo, Xuemei Pu, and Menglong Li. 2014. "Understanding the conformation transition in the activation pathway of β 2 adrenergic receptor via a targeted molecular dynamics simulation." *Physical chemistry chemical physics : PCCP* 17 (4): 2512-2522. <https://doi.org/10.1039/c4cp04528a>.
- [18] Weng, Jingwei, Kangnian Fan, and Wenning Wang. 2012. "The conformational transition pathways of ATP-binding cassette transporter BtuCD revealed by targeted molecular dynamics simulation." *PloS one* 7 (1): e30465-e30465. <https://doi.org/10.1371/journal.pone.0030465>.
- [19] Provasi, Davide, Marta Murcia, Barry S. Coller, and Marta Filizola. 2009. "Targeted molecular dynamics reveals overall common conformational changes upon hybrid domain swing-out in β 3 integrins." *Proteins, structure, function, and bioinformatics* 77 (2): 477-489. <https://doi.org/10.1002/prot.22463>.
- [20] Cheng, Xiaolin, Hailong Wang, Barry Grant, Steven M. Sine, and J. Andrew McCammon. 2006. "Targeted molecular dynamics study of C-loop closure and channel gating in nicotinic receptors." *PLoS computational biology* 2 (9): e134-e134. <https://doi.org/10.1371/journal.pcbi.0020134>.
- [21] van der Vaart, Arjan, Jianpeng Ma, and Martin Karplus. 2004. "The Unfolding Action of GroEL on a Protein Substrate." *Biophysical journal* 87 (1): 562-573. <https://doi.org/10.1529/biophysj.103.037333>.
- [22] Ficici, Emel, Daun Jeong, and Ioan Andricioaei. 2017. "Electric-Field-Induced Protein Translocation via a Conformational Transition in SecDF: An MD Study." *Biophysical journal* 112 (12): 2520-2528. <https://doi.org/10.1016/j.bpj.2017.04.034>.
- [23] Banavali, Nilesh K., and Benoît Roux. 2005. "Free Energy Landscape of A-DNA to B-DNA Conversion in Aqueous Solution." *Journal of the American Chemical Society* 127 (18): 6866-6876. <https://doi.org/10.1021/ja050482k>.
- [24] Jo, Sunhwan, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. 2008. "CHARMM-GUI: A web-based graphical user interface for CHARMM." *Journal of computational chemistry* 29 (11): 1859-1865. <https://doi.org/10.1002/jcc.20945>.
- [25] Lee, Jumin, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, Charles L. Brooks, Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil Im. 2016. "CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field." *Journal of chemical theory and computation* 12 (1): 405-413. <https://doi.org/10.1021/acs.jctc.5b00935>.
- [26] Walls, Alexandra C., Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. 2020. "Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein." *Cell* 181 (2): 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>.
- [27] Phillips, James C., Rosemary Braun, Wei Wang, James Gumbart, Emad Tajkhorshid, Elizabeth Villa, Christophe Chipot, Robert D. Skeel, Laxmikant Kalé, and Klaus Schulten. 2005. "Scalable molecular dynamics with NAMD." *Journal of computational chemistry* 26 (16): 1781-1802. <https://doi.org/10.1002/jcc.20289>.
- [28] Best, Robert B., Xiao Zhu, Jihyun Shim, Pedro E. M. Lopes, Jeetain Mittal, Michael Feig, and Alexander D. MacKerell. 2012. "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the

- Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles." *Journal of chemical theory and computation* 8 (9): 3257-3273. <https://doi.org/10.1021/ct300400x>.
- [29] Williams, B.J. Hintze, J.J. Headd, N.W. Moriarty, V.B. Chen, S. Jain, S.M. Prisant MG Lewis, L.L. Videau, D.A. Keedy, L.N. Deis, I.I.I. Arendall WB, V. Verma, J.S. Snoeyink, P.D. Adams, S.C. Lovell, J.S. Richardson, and D.C. Richardson. 2018. "MolProbity: More and better reference data for improved all-atom structure validation." *Protein Science* 27: 293-315.
- [30] Peng, Cheng, Zhengdan Zhu, Yulong Shi, Xiaoyu Wang, Kaijie Mu, Yanqing Yang, Xinben Zhang, Zhijian Xu, and weiliang zhu. 2020. "Exploring the Binding Mechanism and Accessible Angle of Sars-cov-2 Spike and ACE2 by Molecular Dynamics Simulation and Free Energy Calculation". *ChemRxiv*. <https://doi.org/10.26434/chemrxiv.11877492.v1>.
- [31] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, D. L. Dotson, J. Domanski, S. Buchoux, I. M. Kenney, and O. Beckstein. 2016. "MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations." In S. Benthall and S. Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98-105. SciPy, doi:10.25080/majora-629e541a-00e.
- [32] Michaud-Agrawal, Naveen, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. 2011. "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations." *Journal of computational chemistry* 32 (10): 2319-2327. <https://doi.org/10.1002/jcc.21787>.
- [33] Humphrey, W.; Dalke, A.; Schulten, K. 1996. "VMD: Visual Molecular Dynamics." *J. Mol. Graph.* [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [34] Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. 2011. "Scikit-learn: Machine Learning in Python." *Journal of machine learning research*. <https://doi.org/10.5555/1953048.2078195>.
- [35] Best, Robert B., Gerhard Hummer, and William A. Eaton. 2013. "Native contacts determine protein folding mechanisms in atomistic simulations." *Proceedings of the National Academy of Sciences* 110 (44): 17874-17879. <https://doi.org/10.1073/pnas.1311599110>. <https://www.pnas.org/content/pnas/110/44/17874.full.pdf>.
- [36] Seyler, Sean L., Avishek Kumar, M. F. Thorpe, and Oliver Beckstein. 2015. "Path Similarity Analysis: A Method for Quantifying Macromolecular Pathways." *PLoS computational biology* 11 (10): e1004568-e1004568. <https://doi.org/10.1371/journal.pcbi.1004568>.
- [37] Casalino, Lorenzo, Zied Gaieb, Jory A. Goldsmith, Christy K. Hjorth, Abigail C. Dommer, Aoife M. Harbison, Carl A. Fogarty, Emilia P. Barros, Bryn C. Taylor, Jason S. McLellan, Elisa Fadda, and Rommie E. Amaro. 2020. "Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein." *ACS central science* 6 (10): 1722-1734. <https://doi.org/10.1021/acscentsci.0c01056>.
- [38] Choi, Yeol Kyo, Yiwei Cao, Martin Frank, Hyeonuk Woo, Sang-Jun Park, Min Sun Yeom, Tristan I. Croll, Chaok Seok, and Wonpil Im. 2021. "Structure, Dynamics, Receptor Binding, and Antibody Binding of the Fully Glycosylated Full-Length SARS-CoV-2 Spike Protein in a Viral Membrane." *Journal of chemical theory and computation* 17 (4): 2479-2487. <https://doi.org/10.1021/acs.jctc.0c01144>.
- [39] Zhou, Hongyu, Feng Wang, and Peng Tao. 2018. "t-Distributed Stochastic Neighbor Embedding Method with the Least Information Loss for Macromolecular Simulations." *Journal of chemical theory and computation* 14 (11): 5499-5510. <https://doi.org/10.1021/acs.jctc.8b00652>.
- [40] Pande, Vijay S., Kyle Beauchamp, and Gregory R. Bowman. 2010. "Everything you wanted to know about Markov State Models but were afraid to ask." *Methods (San Diego, Calif)* 52 (1): 99-105. <https://doi.org/10.1016/j.ymeth.2010.06.002>.