Supplementary Information

A glycan gate controls opening of the SARS-CoV-2 spike protein

Terra Sztain[1†], Surl-Hee Ahn[1†], Anthony T. Bogetti[2], Lorenzo Casalino[1], Jory A. Goldsmith[3], Ryan S. McCool[3], Fiona L. Kearns[1], J. Andrew McCammon[1], Jason S. McLellan[3], Lillian T. Chong[2*], Rommie E. Amaro[1*]

1. Department of Chemistry and Biochemistry, UC San Diego, La Jolla, CA 92093
2. Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260
3. Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712

† These authors contributed equally to this work.

* contact authors: ramaro@ucsd.edu, ltchong@pitt.edu

Table of Contents

1. Supplementary Methods

1.1 Computational Methods

1.1.1 Model preparation of the initial closed state

A model of the closed conformation of the glycosylated spike structure and CHARMM36 force field parameters[29,30] were obtained from Casalino, et al.;[8] in this model hydrogen atoms were added using ionization states present in solution at pH 7.4. The stalk and membrane were excluded, and only residues 16-1140 of each trimer were used (**Figure 1A**). The system was solvated in a cubic box of TIP3P[31] explicit water molecules with at least 10 Å between the protein and box edges and 150 mM NaCl using VMD,[32] yielding a system size of 490,621 atoms. The GPU-accelerated Amber18[33,34,35,36] molecular dynamics (MD) engine was used, which gave a 16-fold speedup in dynamics propagation on a GPU vs. CPU. To enable the use of the Amber18 software package, the Chamber program[37] was used to convert the CHARMM36 force field parameters into an Amber readable format.

To relieve unfavorable interactions, the solvated system was subjected to a two-stage energy minimization followed by a two-stage equilibration. To minimize the energy of the system, the solvent was first minimized for 10,000 steps with harmonic position restraints (force constant of 100 kcal/mol /Å^2) applied to the sugars and proteins followed by an unrestrained minimization of the entire system for 100,000 steps. To equilibrate the energy-minimized system, the system was incrementally heated to 300 K over 300 ps in the NVT ensemble followed by a 1-ns equilibration in the NPT ensemble. A production simulation was then carried out in the NPT ensemble for 20 ns on the Triton Shared Computing Cluster at San Diego Supercomputer Center (SDSC). Equilibration and production simulations were carried out with a 2 fs timesteps and SHAKE[38] constraints on bonds to hydrogens. Pressure and temperature were controlled with the Monte Carlo barostat (with 100 fs between attempts to adjust the system volume) and the Langevin thermostat (1 ps^{-1} collision frequency), respectively. Long-range electrostatics were accounted for with the PME method[39] using a 10 Å cutoff for short-range, non-bonded interactions. To provide more extensive sampling of the closed state, we selected a set of 24 equally weighted conformations ("basis states") from the latter 5 ns of the production simulation for a weighted

ensemble (WE) simulation; this portion of the simulation exhibited reasonable convergence of the C-alpha root-mean-squared deviation (RMSD) from the initial, minimized conformation (**Figure S6**).

1.1.2 Weighted ensemble simulations

The weighted ensemble (WE) path sampling strategy orchestrates an ensemble of parallel trajectories with periodic communication to enhance the sampling of pathways for rare events without biasing the dynamics.[15] In particular, a resampling step is applied at fixed time intervals τ to enrich for promising trajectories that have advanced towards the target state – typically, along a progress coordinate that has been divided into bins. Trajectories are all initially assigned equal statistical weights and rigorously tracked to ensure that all weights sum to one at all times of the simulation, introducing no bias in the dynamics.[12] During the resampling step, trajectories that transition to empty bins are replicated and their corresponding weights split evenly between the resulting child trajectories; trajectories that do not make progress are occasionally terminated with their respective weights merged to other trajectories that will be continued. (**Figure S1**)

WE simulations can be run under non-equilibrium steady state or equilibrium conditions and can therefore provide equilibrium (e.g., state populations) and non-equilibrium observables (e.g., rate constants), respectively. To maintain non-equilibrium steady-state conditions, trajectories that reach the target state are "recycled" by initiating a new trajectory from the initial state with the same trajectory weight; steady-state WE simulations therefore require that the target state be defined in advance of the simulation, but are more efficient in generating successful events than equilibrium WE simulations. On the other hand, equilibrium WE simulations do not require a fixed definition of the target state and therefore enable refinement of the target-state definition at any time during the simulation. Here, we leveraged the advantages of both non-equilibrium steady state and equilibrium WE simulations: steady-state simulations were used to more efficiently generate successful pathways trajectories once the target state could be defined and equilibrium simulations were used to further explore and refine the definition of the target state.

All WE simulations were run using the open-source, highly scalable WESTPA software package[40] (**Figure S7**) with a fixed time interval τ of 100 ps for resampling and a target number of 8 trajectories/bin. Details of the progress coordinate and bin spacing for each WE simulation are provided below.

Extensive sampling of the initial, closed-state ensemble

To extensively sample the initial closed-state ensemble, we ran an equilibrium WE simulation starting from randomly selected conformations from the basis states discussed above. A two-dimensional progress coordinate was used consisting of 1) the distance between the center of mass (COM) of the alpha carbons and the structured region of the spike helical core, residues 747-784, 946-967, 986-1034 from each chain, to the alpha carbons and the four main beta sheets of the receptor binding domain (RBD) corresponding to residues 375-380, 394-404, 431-438, 508-517 (refers to RBD from chain A unless otherwise specified) and 2) the RMSD of the alpha carbons and the four main beta sheets of the RBD to the initial structure after 1 ns equilibration. Progress coordinates were calculated using CPPTRAJ.[41] This initial WE simulation was run for 8.77 days on 80 P100 GPUs on Comet at the San Diego Supercomputer Center (SDSC) collecting a comprehensive sampling of ~7.5 μs aggregate simulation time. Bin spacing was continuously monitored and adjusted to maximize efficient sampling.

Simulations of spike opening

After extensive sampling of the closed state, the second progress coordinate was adjusted to be the RMSD of the alpha carbons and the four main beta sheets of the RBD compared to the target up structure, modeled from 6VSB, after 1 ns of equilibration carried out with identical methods as the closed structure described above, which was initially calculated as 11.5 Å. This change allowed for more efficient sampling of the transition to the up state by focusing sampling on the states which are closer in rotational or translational space to the final state, rather than sampling all conformations which are distinctly different from the closed state. Bin spacing was continuously monitored and adjusted to maximize traversing the RMSD coordinate. The full transition was visually confirmed when the RMSD reached below 6 Å and the RBD-COM distance reached above 8.5 Å. The simulation was stopped for analysis after 1099 iterations, 26.7

days on 100 NVIDIA V100 GPUs on Longhorn at the Texas Advanced Computing Center (TACC), collecting an aggregate of ~70 μs of sampling.

A second, independent WE simulation was conducted to determine if the findings of the initial simulation were reproducible and to use the information about the energy landscape of the successful transition in the first WE simulation to inform bin spacing and target state definition to run an unsupervised simulation. The simulation was stopped for analysis after 1729 iterations, 19.64 days on 100 NVIDIA V100 GPUs on Longhorn at TACC, collecting an aggregate of ~51.5 μs of sampling and 106 pathways from the down to the open state. Finally, a third, independent WE simulation that was under steady-state conditions was conducted to maximize sampling of the down to the up-state transitions. The simulation was stopped for analysis after 3000 iterations, 39.84 days on 100 NVIDIA V100 GPUs on Longhorn at TACC, collecting an aggregate of ~106.6 μs of sampling and 204 pathways from the down to the open state. The WESTPA software was shown to scale almost linearly on these 100 NVIDIA V100 GPUs on Longhorn (**Figure S7**), which enabled fast and efficient simulation of the spike.

1.1.3 Weighted ensemble analysis

The successful pathways that reached the up (8.9 Å ≤ RBD-COM distance) or the open (9.9 Å ≤ RBD-COM distance) state were obtained by counting all arrivals to that particular state at every iteration, which yielded 204 and 106 pathways, respectively. We consider these pathways to be statistically independent pathways. The splitting trees for the 204 and 106 pathways, respectively, can be seen in **Figure S8** and **Figure S9**, respectively, which shows how much the pathways shared common segments with each other and when they split. The number of pathways is similar to the number of pathways obtained from calculating the autocorrelation function of arrivals to a particular state at a particular iteration. For instance, at the end of the WE simulation that sampled the open state, there were 1824 trajectories in total and 1193 trajectories that were part of the open state ensemble (defined in later sections as 9.0 Å ≤ RBD-COM distance). Out of the 1193 trajectories that reached the open state ensemble, 133 trajectories were calculated to be statistically independent from calculating the autocorrelation function of the number of arrivals to the open state ensemble[19] (**Figure S10**). The correlation time was calculated to be 16 iterations or 1.6 ns so the trajectories that did not share a common

segment for 16 iterations from the last point in the trajectory were considered to be statistically independent. By checking these multiple independent pathways that reached the up or the open state, we were able to confirm reproducibility of the identified glycan and residue interactions involved in the particular transition. For calculating the shortest and longest transition times, all successful pathways were taken into account. The first 25% of all successful pathways were disregarded to obtain the most probable transition times, however, since the initial transitions can skew the transition time to be shorter than it is normally (**Figure S11, Figure S12**).

We analyzed conformations from the entire WE, in addition to the successful pathways, by grouping state ensembles which satisfied selected criteria. The down ensemble consisted of structures with 11.0 Å ≤ RMSD and RBD-COM distance ≤ 7.5 Å, ~13 μs aggregate simulation time. Note that the entire progress coordinate array had to satisfy the criteria to be counted as part of the ensemble. The up ensemble was defined as 8.5 Å ≤ RBD-COM distance < 9.0 Å, ~ 6.5 μs aggregate simulation time; and the open ensemble 9.0 Å ≤ RBD-COM distance, ~ 4.9 μs aggregate simulation time. Two transition states between the down and up states were defined as: first transition RMSD < 11.0 Å and 7.4 Å ≤ RBD-COM distance ≤ 7.6 Å, ~ 3.1 μs aggregate simulation time; and second transition: 7.9 Å ≤ RBD-COM distance ≤ 8.1 Å, ~ 2.3 μs aggregate simulation time.

1.1.4 Trajectory analysis

Trajectories were visualized using VMD.[32] Glycans, salt bridge, and hydrogen bonding interactions involved in the down to up and open transition were first visually identified. Next, distances between the identified residues were calculated using cpptraj[41] for all 310 successful trajectories, and plotted with matplotlib.[42] To obtain the % of the most probable transition time that had a certain salt bridge, the distance between the atoms/residues that create the salt bridge was measured, and the total time in which the distance was less than 3.5 Å was calculated. The total time for each pathway was calculated and averaged to obtain the final %. To obtain the number of successful pathways that had a certain quantity, e.g., salt bridge, glycan-residue contact, the pathway was counted if the distance was less than 3.5Å in at least one of the frames, looking on at every 10th frame. Contact maps calculating the distance between the RBD A and all other residues and glycans were generated using MDAnalysis[43,44] (**Movie S3**). Structures for

figures and movies were generated using VMD, including NanoShaper[45] surface representation.

Analysis of residues mutated in emerging SARS-CoV-2 strains

As of the date of this submission, the following SARS-CoV-2 variants have been identified (with mutations to spike noted in parentheticals): B.1 (D614G), B.1.1.7 (H69-V70 deletion, Y144-Y145 deletions, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H), B.1.351: (L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V), P1 (L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y ,D614G, H655Y, T1027I) and CAL.20C (L452R, D614G).[27] To examine potential implications of these mutations on Spike opening mechanics, we have monitored the neighboring residues of key WT residues as a function of the opening mechanism. MDAnalysis[43,44] was used to identify residues whose center of mass was within 10 Å of the center of mass of the key residue of interest. For each contact, the fraction of frames in closed, up, and open ensembles containing the contact is provided. Contacts were only consider if they exist within > 5% of all frames and if the contacting pairs were separated by more than 3 peptide bonds in one-dimensional sequence.

1.2 Experimental Methods

Protein Expression and Purification

Substitutions N343A, D405A, R408A, and D427A were cloned into the HexaPro SARS-CoV-2 spike background.[25] A spike variant with all RBDs locked in the down position through the introduction of a disulfide bond was similarly produced through cysteine substitutions at residues S383C and D985C in the HexaPro protein.[23] All variants were expressed through polyethyleneimine-induced transient transfection of FreeStyle 293-F cells (Thermo Fisher). After 4 days, cell supernatant was clarified by centrifugation, passed through a 0.22 μm filter, and purified over StrepTactin resin (IBA). Variants were further purified by size-exclusion chromatography on a Superose 6 10/300 column (GE Healthcare) in a buffer consisting of 2 mM Tris pH 8.0, 200 mM NaCl and 0.02% NaN_3. Soluble ACE2 was produced and purified as previously described.[8]

Biolayer Interferometry

Anti-foldon IgG was immobilized to an anti-human Fc (AHC) Octet biosensor (FortéBio). Tips were then submerged into the specified HexaPro variants before being subsequently dipped into 200 nM ACE2 to observe variant association, followed by dissociation in buffer consisting of 20 mM Tris pH 7.5, 150 mM NaCl, 1 mg/mL bovine serum albumin, and 0.01% Tween-20. The relative proportion of RBD in an accessible state was quantified based on the binding level as previously described.[8] The S383C, D985C variant was used as a negative control. Data were collected in triplicate and replicate sensorgrams are shown in **Figure S13.**

2. Supplementary Figures

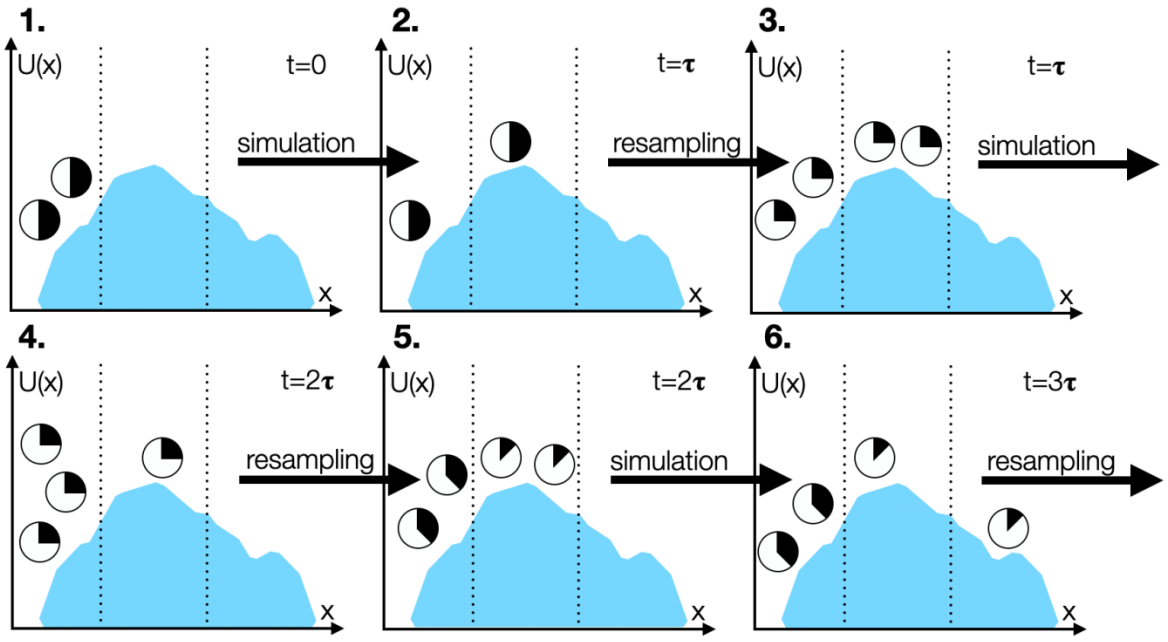

Figure S1 Schematic of the weighted ensemble (WE) strategy. The WE strategy is illustrated for a three-state system with a one-dimensional progress coordinate x that is divided into bins. U(x) represents the potential of the system dependent on x, which can be seen from the curve of the shaded region. 1. WE initiates two equally weighted trajectories (represented as circles) from the first bin, each with a statistical weight of 0.5 (represented as filled parts of the circles), for a fixed time interval **τ**. 2. Resampling is then performed, replicating or terminating trajectories to maintain a target number of two trajectories in each bin (*e.g.,* in the first and second bins, splitting the weight among the two child trajectories with a weight of 0.25 for each trajectory). 3. Trajectories are run for another fixed time interval **τ**. 4. After running, resampling is performed (*e.g.,* in the first bin, terminating two of the three trajectories and in the second bin, replicating the one trajectory to yield two trajectories). 5. The system ends up with two trajectories in each of the visited bins. 6. One of the trajectories ends up in the third bin. Rounds of simulation and resampling are performed until a desired number of continuous pathways into the target state are generated.

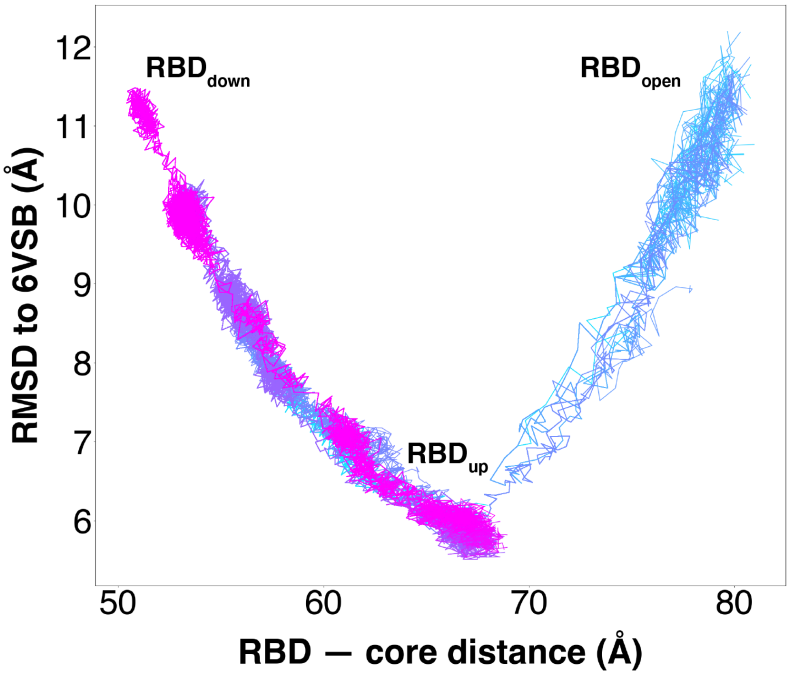

Figure S2 Successful pathways of spike opening. Overlay of 310 successful pathways including 204 pathways of the RBD transitioning from the down to up state and 106 pathways from the down to open state. Continuous trajectories plotted with the RMSD of the RBD to the 6VSB "up" state versus the RBD — core distance.

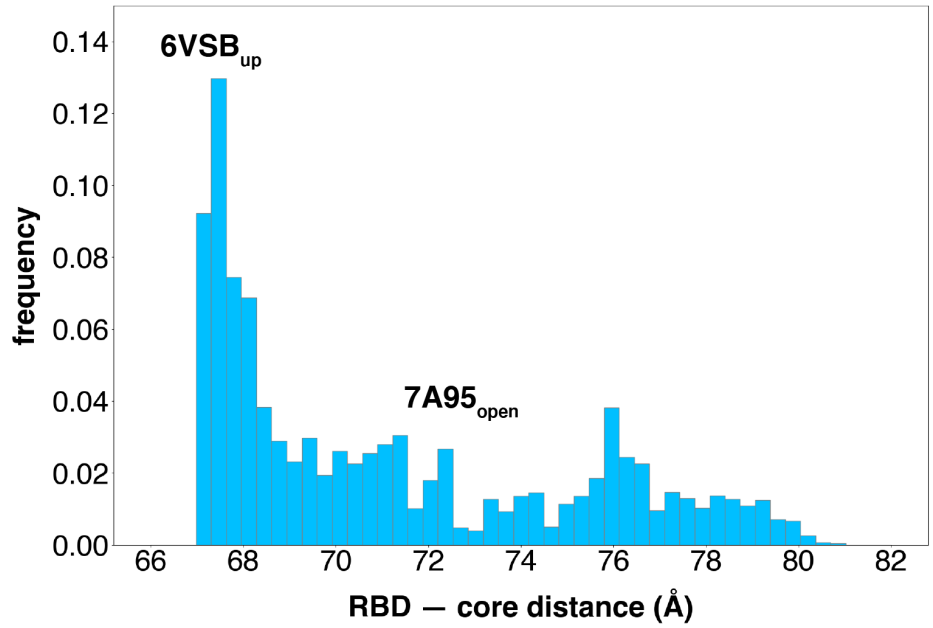

Figure S3 Diversity of RBD "open" state sampled. Probability distribution of RBD — core distances greater than the RBD "up" conformation defined by PDB 6VSB (67.2 Å). The ACE2-bound structure from PDB 7A95 distance is 72.1 Å.

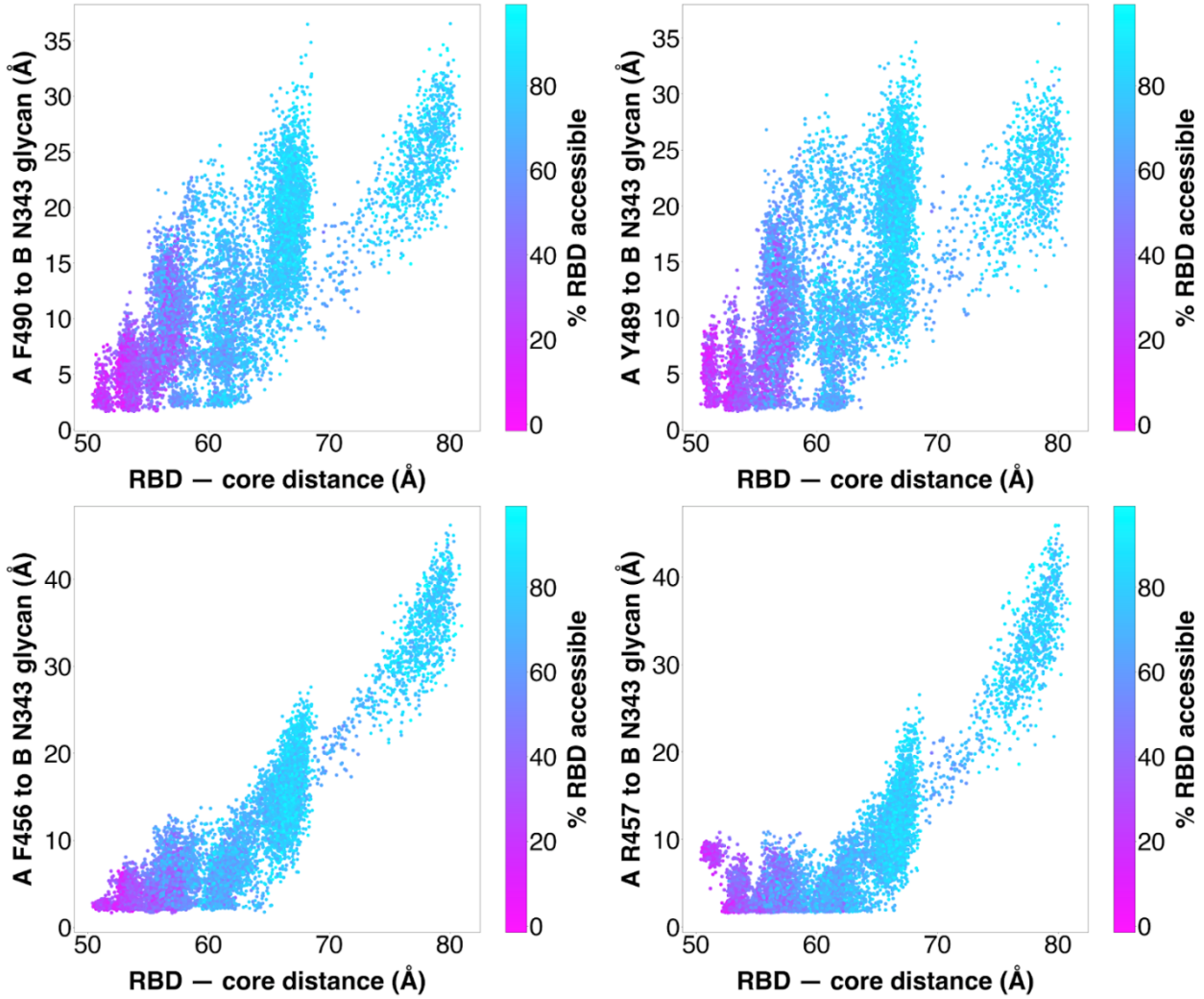

Figure S4 Distance between N343 glycan and RBD residues. Scatter plot of data from the 310 continuous pathways with the minimum distance between the N343 glycan and RBD A residues F490, Y489, F456, or R457 plotted against RBD — core distance. Data points are colored based on % RBD solvent accessible surface area compared to the RBD "down" state 6VXX.

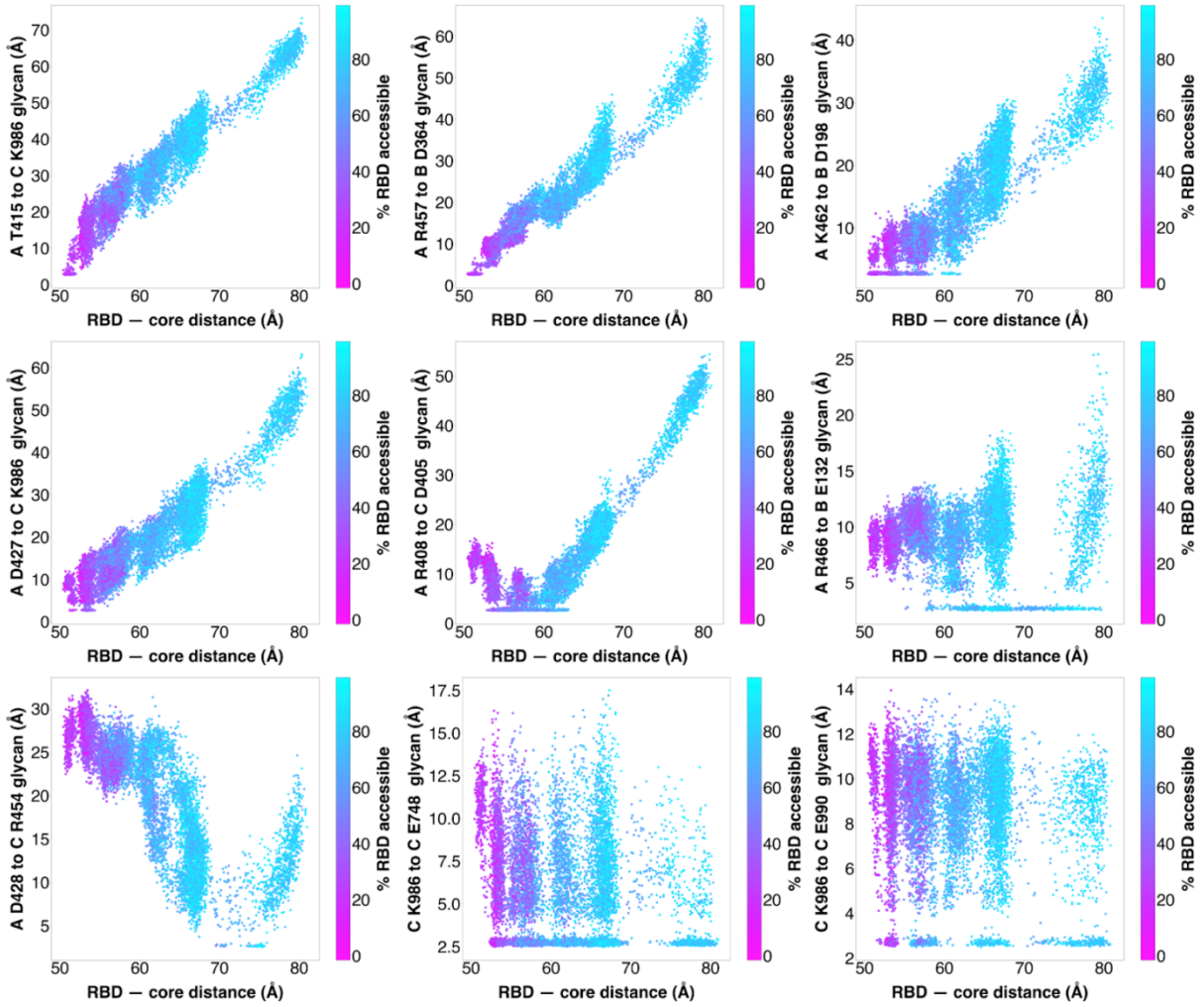

Figure S5 Distance between salt bridge and hydrogen bond residues along the opening pathway. Scatter plot of data from the 310 continuous pathways with the minimum distance between the residues shown in Figure 4 plotted against RBD — core distance. Data points are colored based on % RBD solvent accessible surface area compared to the RBD "down" state 6VXX.

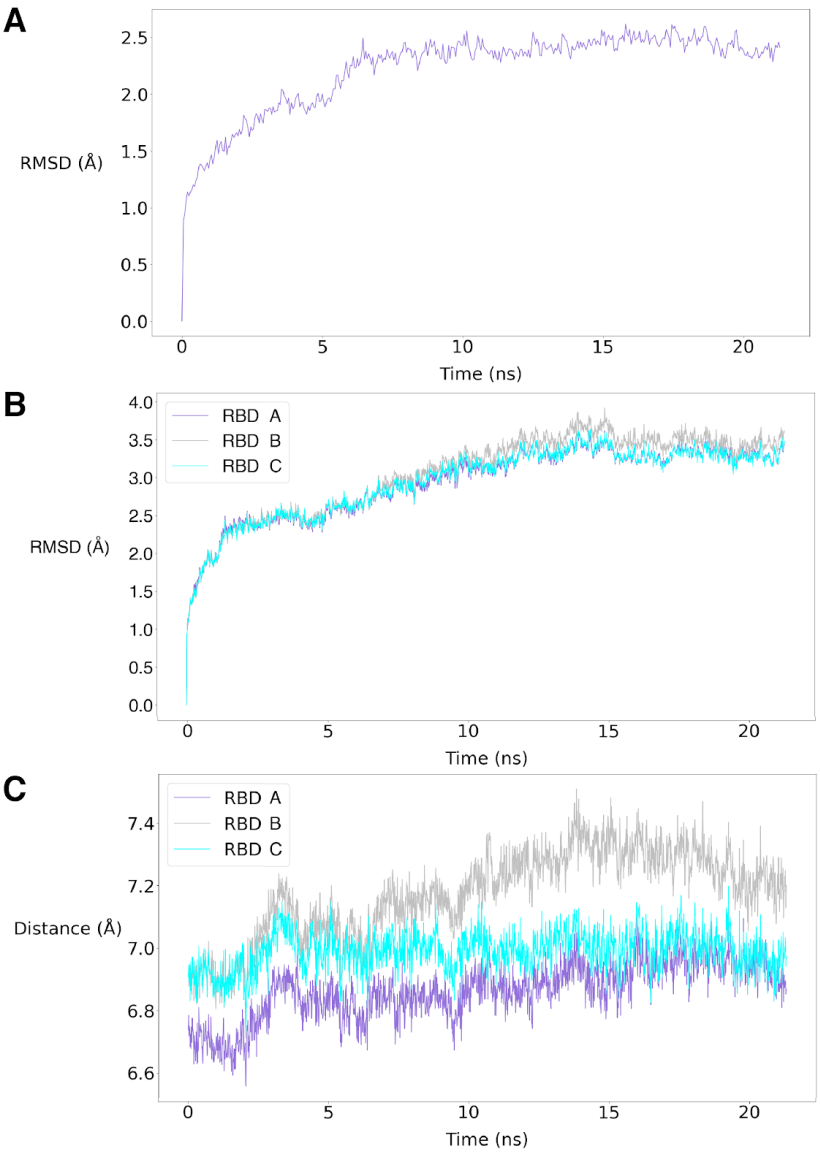

Figure S6 Initial equilibration of a closed-state structure using a standard MD simulation. (A) RMSD of Cα atoms of protein residues vs time. (B) RMSD of Cα atoms of structured region of RBD after alignment of core to first frame. (C) Distance of RBD center of mass to center of mass of core.

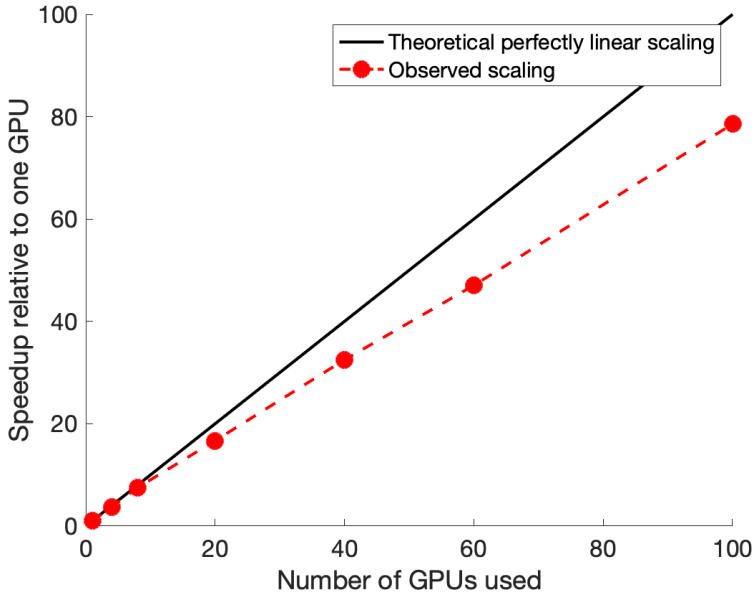

Figure S7 Scaling of WESTPA on 100 NVIDIA V100 GPUs on the Longhorn supercomputer vs. theoretical perfectly linear scaling.

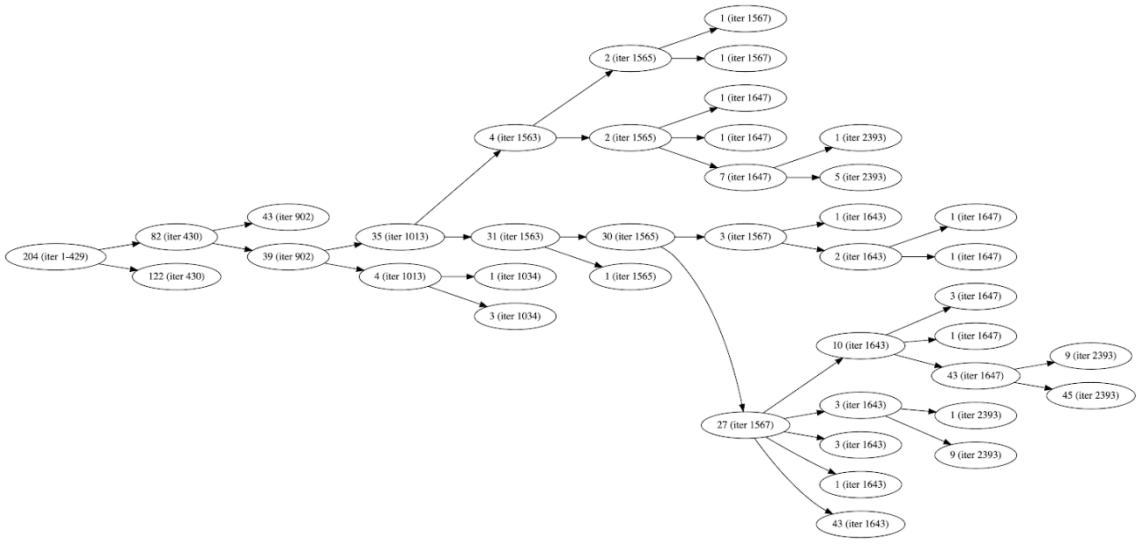

Figure S8 Trajectory splitting tree of the 204 independent pathways that reached the up state. The numbers inside each node indicate the number of pathways at the given iteration, which is specified in parentheses. All trajectories shared the same parents up until iteration 429, with the first main

split occurring at iteration 430. Subsequent splitting occurred at later iterations. Note that the sum of the child pathways does not necessarily match up with the parent's number of pathways due to splitting and merging with other trajectories not shown in the graph.

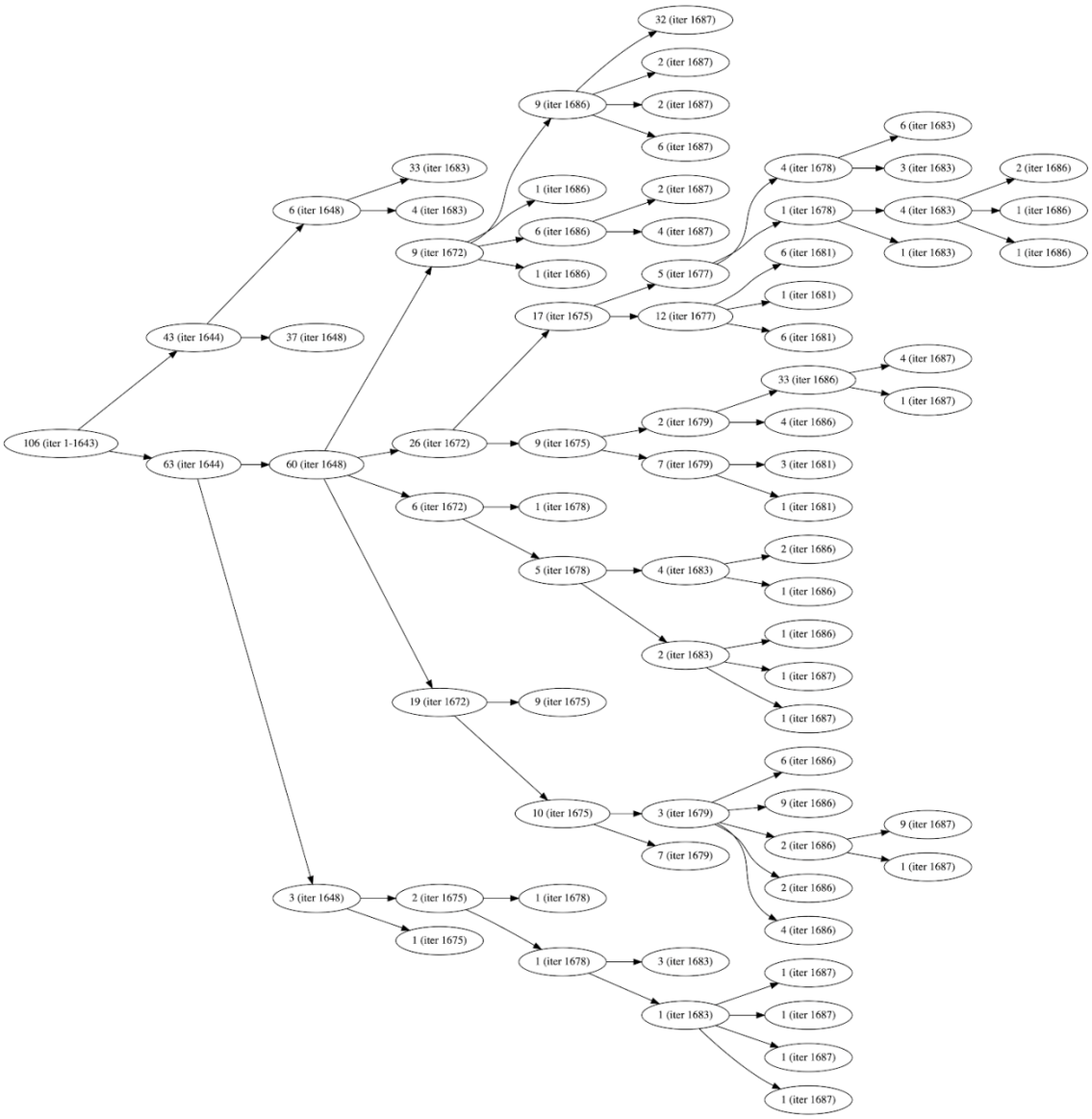

Figure S9 Trajectory splitting tree of the 106 pathways that reached the open state. The numbers inside each node indicate the number of pathways at the given iteration, which is specified in parentheses. All trajectories shared the same parents until iteration 1643, the first main split occurring at iteration 1644. Note that the sum of the child pathways does not necessarily match up

with the parent's number of pathways at subsequent iterations due to splitting and merging with other trajectories not shown in the graph.

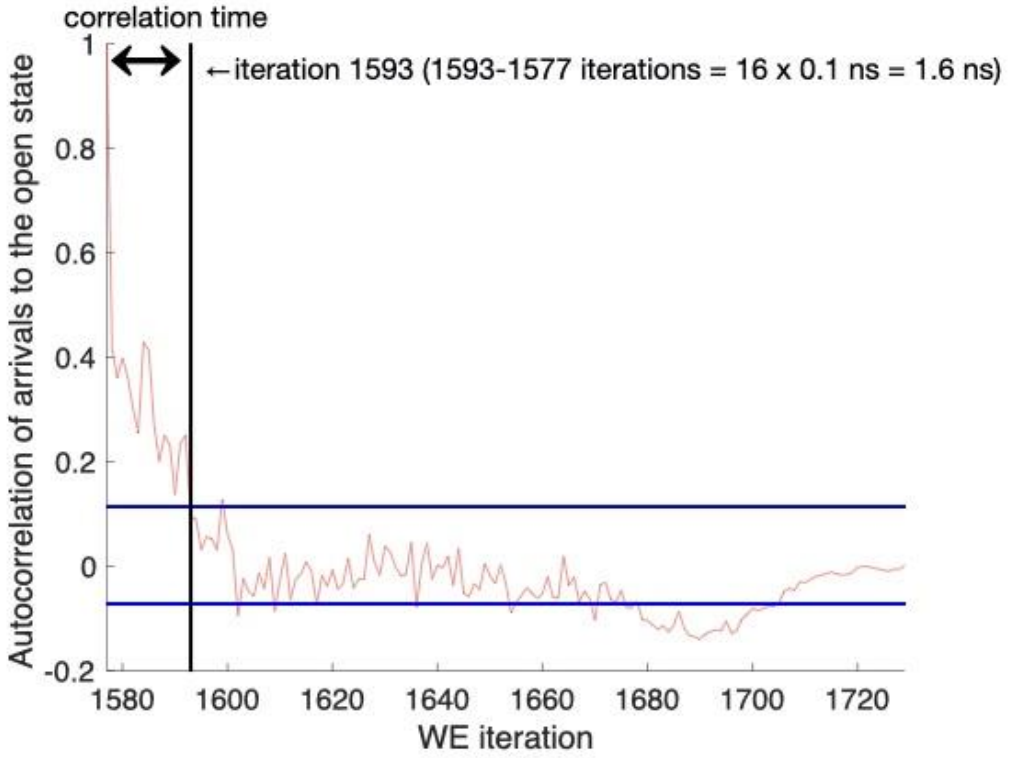

Figure S10 Autocorrelation of arrivals to the open state ensemble (red) with a 95% confidence interval (blue). The confidence interval was generated by calculating the autocorrelation function of 1000 datasets (drawn with replacement) with bootstrapping technique. The first point at which the autocorrelation falls within the confidence is marked, which is used to calculate the correlation time.

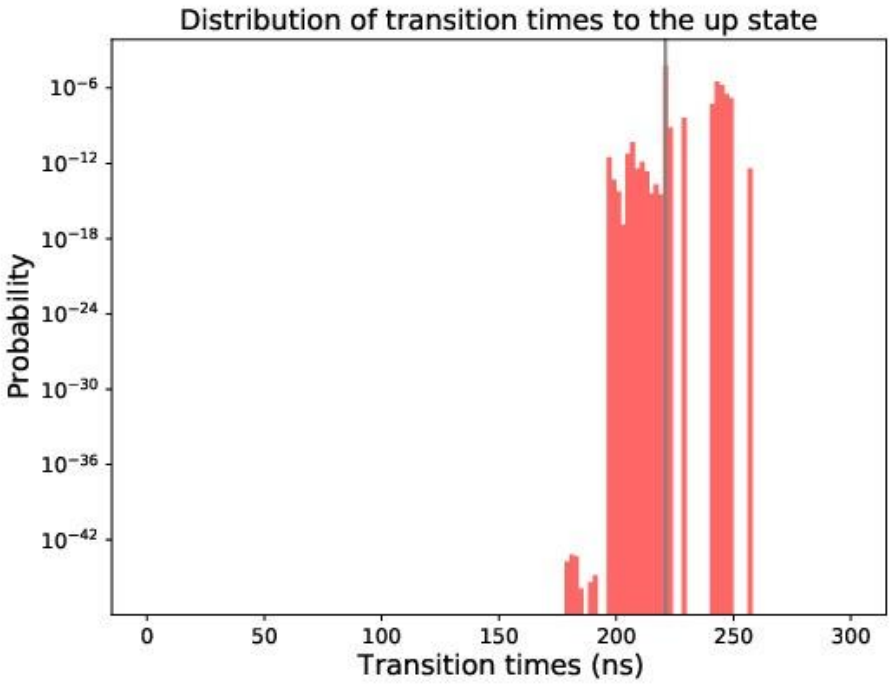

Figure S11 Distribution of transition times to the up state. The most probable transition time is marked in grey. Note that the first 25% of the "fast" transitions are discarded here to calculate the most probable transition time.

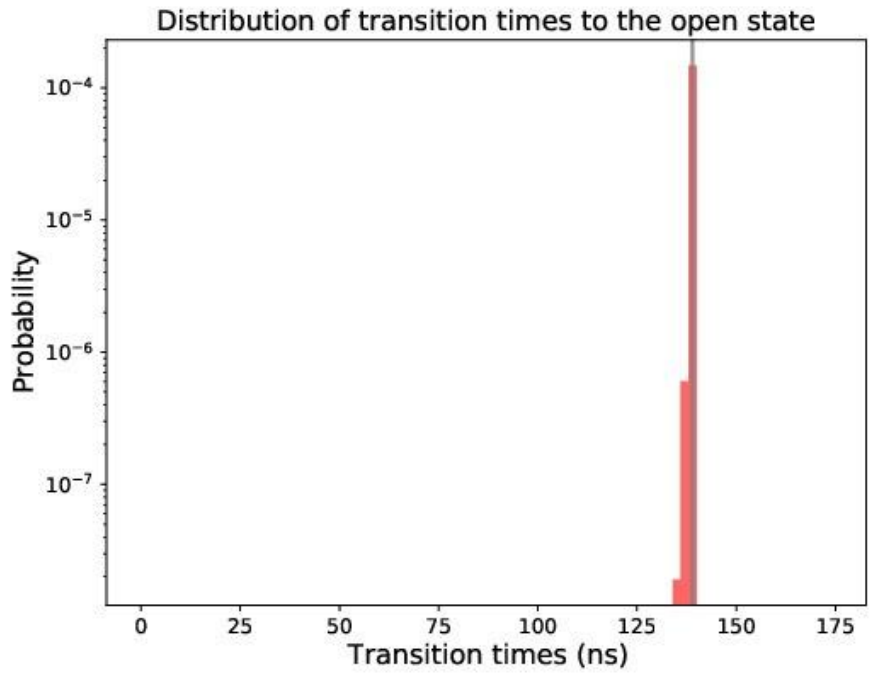

Figure S12 Distribution of transition times to the open state. The most probable transition time is marked in grey. Note that the first 25% of the "fast" transitions are discarded here to calculate the most probable transition time.

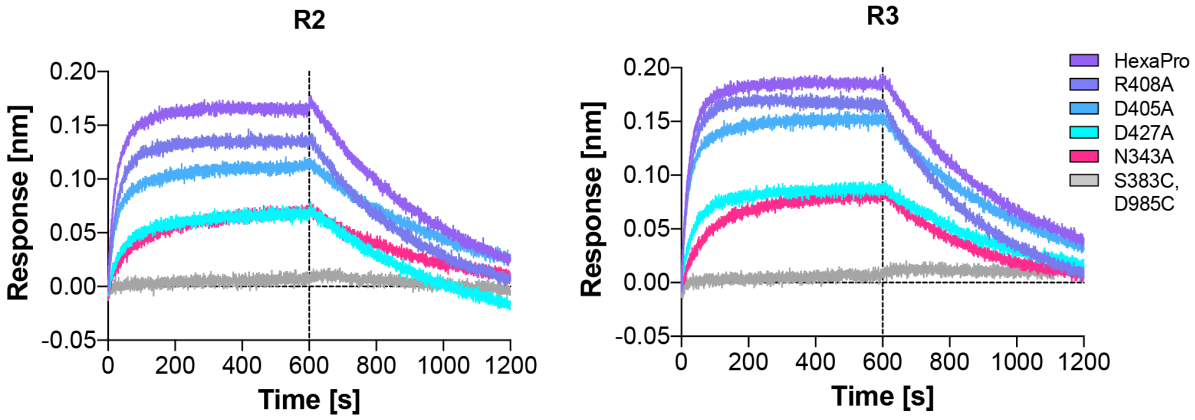

Figure S13 BLI sensorgrams of spike variants binding to ACE2 from duplicate (R2) and triplicate (R3) experiments.

3. Supplementary Tables

Table S1 Biolayer interferometry data of spike variants binding to ACE2.

VARIANT	HEXAPRO	R408A	D405A	D427A	N343A
R1 - Binding level (nm)	0.1733	0.1560	0.1206	0.0913	0.0783
R2 - Binding level (nm)	0.1776	0.1467	0.1208	0.0793	0.0751
R3 - Binding level (nm)	0.1831	0.1629	0.1506	0.0849	0.0816
Minimum (nm)	0.1733	0.1467	0.1206	0.07932	0.07512
Maximum (nm)	0.1831	0.1629	0.1506	0.09131	0.0816
Range (nm)	0.0098	0.0162	0.03	0.01199	0.00648
Mean (nm)	0.1780	0.1552	0.1307	0.0852	0.0783
Std. Deviation (± nm)	0.0049	0.0081	0.0173	0.0060	0.0032
Response (% to HexaPro)	**100.00**	**87.19**	**73.43**	**47.85**	**44.01**
Response decrease (%)	**0.00**	**12.81**	**26.57**	**52.15**	**55.99**

Table S2 Evaluation of neighboring residues of key altered residues in circulating SARS-CoV-2 Spike variants. <TableS2.xlsx>

4. Supplementary Movies

Movie S1 Continuous pathway of RBD opening. This movie shows one of the continuous, unbiased pathways obtained from the WE simulations. All glycans are shown in blue except the N343 glycan which is colored magenta. Starting from all three RBDs in the down conformation, the chain A RBD lifts and twists counterclockwise into the up conformation, facilitated through interactions with the two adjacent RBDs, especially the N343 glycan gate on the chain B RBD. Upon reaching the up conformation, the RBD continues to twist into an open conformation en route to S1 dissociation.

Movie S2 Glycan gate at position N343 intercalates with residues to facilitate RBD opening. This movie zooms in closer to the glycan at position N343 to show how RBD opening is facilitated through intercalation between and underneath the residues F490, Y489, F456, F457 of RBD A. The glycan also transiently interacts with other residues of the RBD which are shown when they are within Å from the glycan.

Movie S3 Mapping of residue contacts to RBD throughout opening pathway. Distances between residues throughout a continuous opening pathway calculated for the trajectory shown in **Movies S1-2**. Distances to each residue from RBD_A are shown for each chain in panels A-C and each of the glycans in panel D. Select regions are labeled, and N165, N234, and N343 are labeled with +, ++, +++, respectively.

5. Supplementary References

(29) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J Comput Chem* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(30) Guvench, O.; Hatcher, E. R.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J Chem Theory Comput* **2009**, *5* (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(32) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J Mol Graph* **1996**, *14* (1), 33–38, 27–28.

(33) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8* (5), 1542–1555. https://doi.org/10.1021/ct200909j.

(34) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(35) D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden,; R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang,; S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J.; Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R.; Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. AMBER 2018. *University of California, San Francisco.* **2018**.

(36) Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J Chem Inf Model* **2018**, *58* (10), 2043–2050. https://doi.org/10.1021/acs.jcim.8b00462.

(37) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive Support for CHARMM Force Fields within the AMBER Software. *International Journal of Quantum Chemistry* **2009**, *109* (15), 3767–3772. https://doi.org/10.1002/qua.22372.

(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(39) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(40) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J. Chem. Theory Comput.* **2015**, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(41) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(42) Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering* **2007**, *9* (3), 90–95. https://doi.org/10.1109/MCSE.2007.55.

(43) 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. 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*. **2016**, 98-105. Austin, TX, SciPy, doi:10.25080/majora-629e541a-00e.

(44) N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, 32 (10), 2319-2327. doi:10.1002/jcc.21787.

(45) Decherchi, S.; Spitaleri, A.; Stone, J.; Rocchia, W. NanoShaper-VMD Interface: Computing and Visualizing Surfaces, Pockets and Channels in Molecular Systems. *Bioinformatics* **2019**, *35* (7), 1241–1243. https://doi.org/10.1093/bioinformatics/bty761.