



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

Terra Sztain^{1,8}, Surl-Hee Ahn^{1,8}, Anthony T. Bogetti², Lorenzo Casalino¹, Jory A. Goldsmith³, Evan Seitz⁴, Ryan S. McCool³, Fiona L. Kearns¹, Francisco Acosta-Reyes⁵, Suvrajit Maji⁵, Ghoncheh Mashayekhi⁶, J. Andrew McCammon^{1,7}, Abbas Ourmazd⁶, Joachim Frank^{4,5}, Jason S. McLellan³, Lillian T. Chong²✉ and Rommie E. Amaro¹✉

SARS-CoV-2 infection is controlled by the opening of the spike protein receptor binding domain (RBD), which transitions from a glycan-shielded ‘down’ to an exposed ‘up’ state to bind the human angiotensin-converting enzyme 2 receptor and infect cells. While snapshots of the ‘up’ and ‘down’ states have been obtained by cryo-electron microscopy and cryo-electron tomography, details of the RBD-opening transition evade experimental characterization. Here over 130 μ s of weighted ensemble simulations of the fully glycosylated spike ectodomain allow us to characterize more than 300 continuous, kinetically unbiased RBD-opening pathways. Together with ManifoldEM analysis of cryo-electron microscopy data and biolayer interferometry experiments, we reveal a gating role for the N-glycan at position N343, which facilitates RBD opening. Residues D405, R408 and D427 also participate. The atomic-level characterization of the glycosylated spike activation mechanism provided herein represents a landmark study for ensemble pathway simulations and offers a foundation for understanding the fundamental mechanisms of SARS-CoV-2 viral entry and infection.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped RNA virus and the causative agent of coronavirus disease 2019 (COVID-19), a disease that has caused substantial morbidity and mortality worldwide^{1,2}. The main infection machinery of the virus, the spike protein that sits on the outside of the virus, is the first point of contact that the virion makes with the host cell, and is a major viral antigen³. A substantial number of cryo-electron microscopy (cryo-EM) structures of the spike protein have been recently reported, collectively informing on structural states of the spike protein. The vast majority of resolved structures fall into either ‘down’ or ‘up’ states, as defined by the position of the receptor binding domain (RBD), which modulates interaction with the angiotensin-converting enzyme 2 (ACE2) receptor for cell entry^{4–6}.

The RBDs must transition from a ‘down’ to an ‘up’ state for the receptor binding motif (RBM) to be accessible for ACE2 binding (Fig. 1), and therefore the activation mechanism is essential for cell entry. Lu et al.⁷ used single-molecule fluorescence (Förster) resonance energy transfer (smFRET) imaging to characterize spike dynamics in real time. Their work showed that the spike dynamically visits four distinct conformational states, the populations of which are modulated by the presence of the human ACE2 receptor and antibodies. However, smFRET, as well as conventional structural biology techniques, are unable to inform on the atomic-level mechanisms underpinning such dynamical transitions. Recently, all-atom molecular dynamics (MD) simulations of the spike protein, with experimentally accurate glycosylation together with corroborating experiments, indicated the extensive shielding by spike glycans, as well as a mechanical role for glycans at positions N165

and N234 in supporting the RBD in the ‘open’ conformation⁸. Conventional MD simulations as performed in Casalino et al.⁸ also revealed microsecond-timescale dynamics to better characterize the spike dynamics but were limited to sampling configurations that were similar in energy to the cryo-EM structures. Several enhanced sampling MD simulations have been performed to study this pathway; however, these simulations lacked glycosylation for the spike protein⁹ or involved the addition of an external force¹⁰ or did not provide mechanistic detail¹¹.

In this study, we characterized the spike RBD-opening pathway for the fully glycosylated SARS-CoV-2 spike protein to gain a detailed understanding of the activation mechanism. We used the weighted ensemble (WE) path-sampling strategy^{12,13} (Supplementary Fig. 1) to enable the simulation of atomistic pathways for the spike-opening process. As a path-sampling strategy, WE focuses computing power on the functional transitions between stable states rather than the stable states themselves¹⁴. This is achieved by running multiple trajectories in parallel and periodically replicating trajectories that have transitioned from previously visited to newly visited regions of configurational space¹⁵, thus minimizing the time spent waiting in the initial stable state for ‘lucky’ transitions over the free energy barrier. Given that these transitions are much faster than the waiting times^{16,17}, the WE strategy can be orders of magnitude more efficient than conventional MD simulations in generating pathways for rare events such as protein folding and protein binding^{18,19}. This efficiency is even higher for slower processes, increasing exponentially with the effective free energy barrier²⁰. Not only are dynamics carried out without any biasing force or modifications to the free energy landscape, but suitable assignment of statistical weights to

¹Department of Chemistry and Biochemistry, University of California–San Diego, La Jolla, CA, USA. ²Department of Chemistry, University of Pittsburgh, Pittsburgh, PA, USA. ³Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX, USA. ⁴Department of Biological Sciences, Columbia University, New York, NY, USA. ⁵Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY, USA. ⁶Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI, USA. ⁷Department of Pharmacology, University of California–San Diego, La Jolla, CA, USA. ⁸These authors contributed equally: Terra Sztain, Surl-Hee Ahn. ✉e-mail: ltchong@pitt.edu; ramaro@ucsd.edu

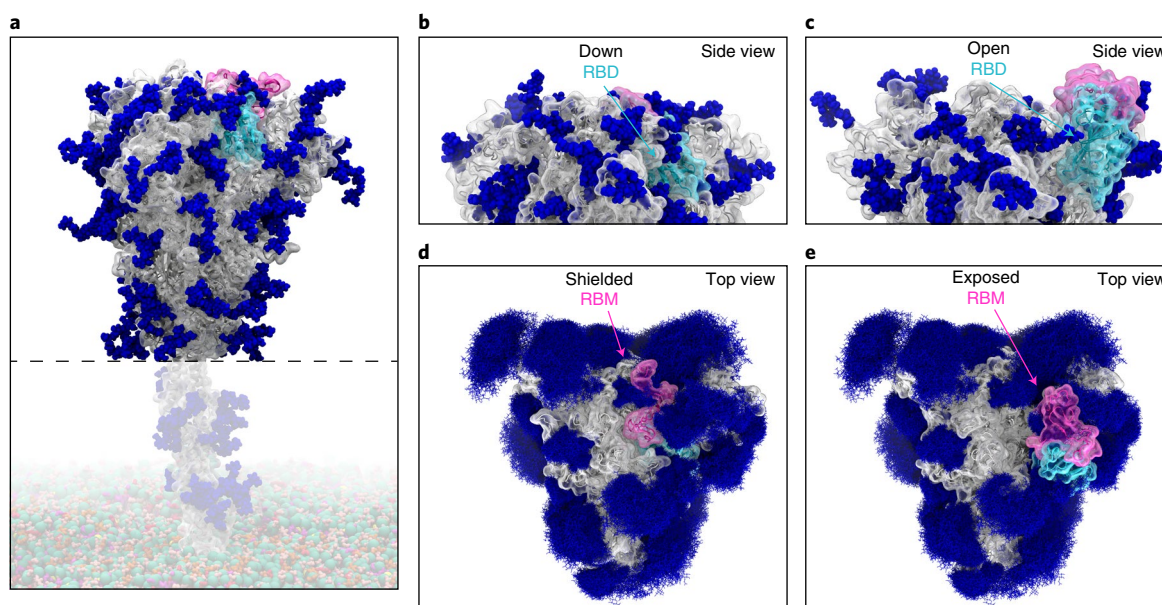


Fig. 1 | Glycosylated spike RBD ‘down’ and ‘open’ conformations. **a**, The SARS-CoV-2 spike head (grey) with glycans (dark blue) as simulated, with the stalk domain and membrane (not simulated here, but shown as transparent for completeness). RBD shown in cyan, RBM in pink. **b,c**, Side view of the RBD_{down} (shielded, **b**) and RBD_{open} (exposed, **c**). **d,e**, Top view of the RBM_{closed} (shielded, **d**) and RBM_{open} (exposed, **e**). Composite image of glycans (dark blue lines) shows many overlapping snapshots of the glycans over the microsecond simulations.

trajectories provides an unbiased characterization of the system’s time-dependent ensemble properties¹³. The WE strategy therefore generates continuous pathways with unbiased dynamics, yielding the most direct, atomistic views for analysing the mechanism of functional transitions, including elucidation of transient states that are too fleeting to be captured by laboratory experiments. Furthermore, while the strategy requires a progress coordinate towards the target state, the definition of this target state need not be fixed in advance when applied under equilibrium conditions²¹, enabling us to refine the definition of the target ‘open’ state of the spike protein on the basis of the probability distribution of protein conformations sampled by the simulation.

Our work characterizes a series of transition pathways of the spike opening, in agreement with conformations detected in the cryo-EM dataset by ManifoldEM²², and identifies key residues, including a glycan at position N343, that participate in the opening mechanism. Our simulation findings are corroborated by biolayer interferometry (BLI) experiments, which show a reduction in the ability of the spike to interact with ACE2 after mutation of these key residues.

Results and discussion

WE simulations of spike opening. As mentioned above, simulations of the spike-opening process require an enhanced sampling strategy as the process occurs beyond the microsecond timescale (that is, the seconds timescale⁷). We therefore used the WE path-sampling strategy, which enabled the generation of continuous, atomistic pathways for the spike-opening process with unbiased dynamics (Fig. 2a–e and Supplementary Video 1); these pathways were hundreds of nanoseconds long, excluding the waiting times in the initial ‘down’ state. The protein model was based on the head region (residues 16 to 1,140) of the glycosylated SARS-CoV-2 spike from Casalino et al.⁸ (Fig. 1), which in turn was built on the cryo-EM structure of the three-RBD-down spike (Protein Data Bank (PDB) ID, 6VXX (ref. 5)). The entire simulation system, including explicit water and salt ions, reaches almost half a million atoms. We focused sampling along a two-dimensional progress coordinate to track RBD

opening: the difference in the centre of mass of the spike core to the RBD and the root-mean-square deviation of the RBD from the RBD_{up} state (Fig. 2f,g). On the San Diego Supercomputer Center (SDSC) Comet and Texas Advanced Computing Center (TACC) Longhorn supercomputers, 100 graphics processing units (GPUs) ran the WE simulations in parallel for over a month, generating over 130 μ s of glycosylated spike trajectories and more than 200 TB of trajectory data. We simulated a total of 310 independent pathways, including 204 pathways from the RBD_{down} conformation (PDB ID, 6VXX (ref. 5)) to the RBD_{up} conformation (PDB ID, 6VSB (ref. 4)) and 106 pathways from the RBD_{down} to the RBD_{open} state, in which the RBD twists open beyond the 6VSB (ref. 4) cryo-EM structure. Remarkably, the RBD_{open} state that we sampled includes conformations that align closely with the ACE2-bound spike cryo-EM structure (PDB ID, 7A95 (ref. 6)) even though this structure was not a target state of our progress coordinate (Fig. 2f,g, Supplementary Video 1 and Supplementary Figs. 2 and 3). This result underscores the value of using (1) equilibrium WE simulations that do not require a fixed definition of the target state and (2) a two-dimensional progress coordinate that allows the simulations to sample unexpected conformational space along multiple degrees of freedom. The ACE2-bound spike conformation has also been sampled by the Folding@home-distributed computing project¹¹, and RBD rotation has been detected in cryo-EM experiments⁶.

Comparison with spike conformations detected by ManifoldEM.

To validate our simulated RBD_{down} to RBD_{up} pathway, the ManifoldEM framework²² was applied using the cryo-EM dataset of PDB 6VSB from McLellan and colleagues⁴. The ManifoldEM method allows characterization of conformational variations as obtained from a single-particle cryo-EM ensemble of a molecule in thermal equilibrium. Two conformational coordinates (that is, collective motion coordinates) CC1 and CC2 were discovered from this dataset, and observed from several exemplary projection directions (PDs) showing a (1) RBD_{down} to RBD_{up} pathway and (2) RBD outward opening pathway (Supplementary Fig. 4 and Supplementary Videos 2 and 3).

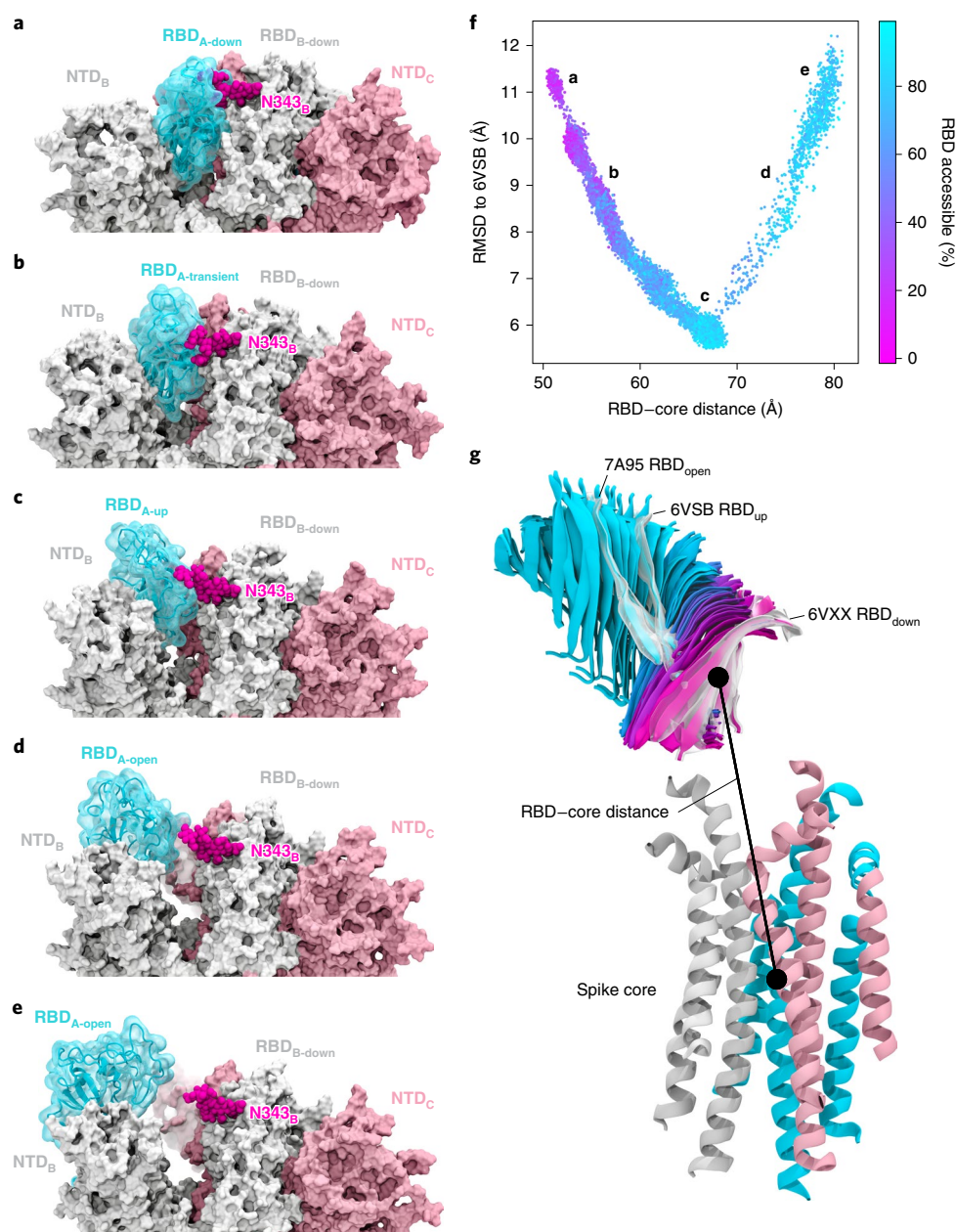


Fig. 2 | Atomically detailed pathways of spike opening. **a–e**, Snapshot configurations along the opening pathway with chain A shown in cyan, chain B in grey, chain C in pink and the glycan at position N343 in magenta. Each RBD and N-terminal domain (NTD) are subscripted with their chain ID (A, B or C). RBDs are also subscripted with their conformation from initial conformation with all three RBDs in the 'down' state (6VXX) (**a**), RBD_A in a 'transient' state in between the 'down' and 'up' state (6VSB) (**b**) RBD_A in the 'up' state (**c**), RBD_A in the 'open' state (beyond 6VSB) (**d**) and RBD_A in the furthest open state sampled (**e**). **f**, Scatter plot of data from the 310 continuous pathways with the Cα-root-mean-square deviation (RMSD) of the RBD from the RBD_{up} state plotted against the RBD-core distance. Data points are coloured on the basis of the percentage RBD solvent-accessible surface area compared with the RBD_{down} state. The locations of the snapshots shown in **a–e** are labelled. **g**, Primary regions of spike defined for tracking progress of the opening transition. The spike core is composed of three central helices per trimer, coloured according to chains as in **a–e**. The RBD contains a structured pair of antiparallel beta-sheets, and an overlay of snapshots from a continuous WE simulation are shown coloured along a spectrum resembling the palette in **f**. Overlaid cryo-EM structures are highlighted and labelled including the initial RBD_{down} state (6VXX), the target RBD_{up} state and the ACE2-bound RBD_{open} state (7A95).

These projections were next aligned to corresponding two-dimensional projections of coulomb potential maps generated with frames from the WE simulation (Supplementary Fig. 5 and Supplementary Videos 2 and 3). Overall, there was very good agreement between the ManifoldEM conformational coordinates and the WE trajectory, aside from two discrepancies. First, the CC2 observed in the ManifoldEM included concerted opening of all three RBDs, while the WE focused sampling on the opening

of a single RBD (Supplementary Video 2). Second, the WE trajectory ultimately opens to an RBD-core distance 11 Å greater than the most open conformation in the ManifoldEM. This is probable because the simulations sample the S1 subunit en route to the post-fusion conformation, whereas the experimental dataset does not.

The N343 glycan gates RBD opening. In the 'down' state, the RBD of the SARS-CoV-2 spike is shielded by glycans at positions

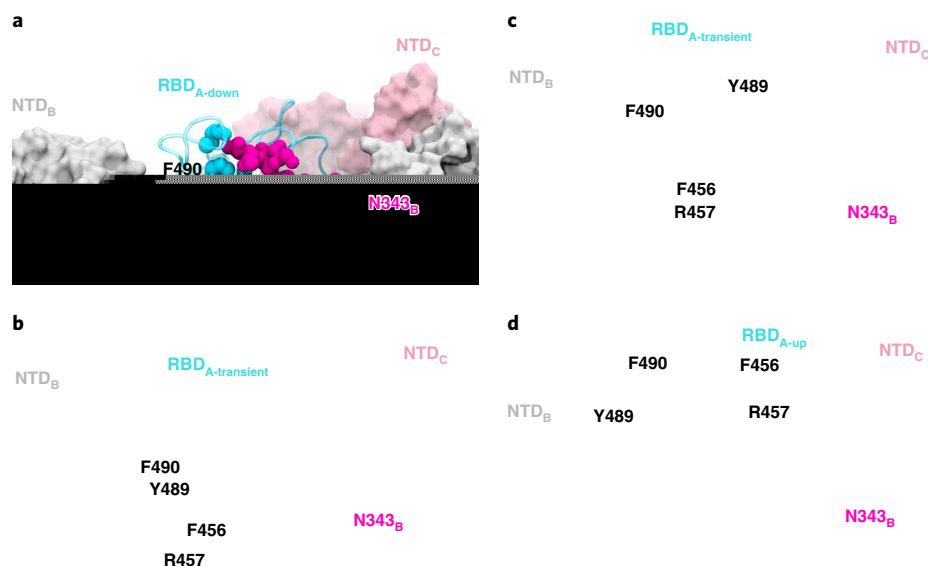


Fig. 3 | Glycan gating by N343. a–d, Snapshot configurations along the opening pathway with chain A shown in cyan, chain B in grey, chain C in pink and the glycan at position N343 in magenta. RBD_A in the ‘down’ conformation is shielded by the glycan at position N343 of the adjacent RBD_B (**a**). The N343 glycan intercalates between (**b**) and underneath (**c**) residues F490, Y489, F456 and F457 to push the RBD up and open (**d**).

N165, N234 and N343 (ref. ²³). While glycan shielding had been investigated for the RBD_{down} and RBD_{up} states⁸, our WE simulations allowed characterization of shielding during the opening process, revealing an abrupt decrease in glycan shielding when the RBD transitions from the ‘down’ to the ‘up’ state. The glycans at position N165 and N234 consistently shield the RBM, while shielding by the N343 glycan decreases with RBD opening (Supplementary Fig. 6). Beyond shielding, a structural role for glycans at positions N165 and N234 has been recently reported, stabilizing the RBD in the ‘up’ conformation through a ‘load and lock’ mechanism⁸.

Our WE simulations reveal an even more specific, critical role of a glycan in the opening mechanism of the spike: the N343 glycan acts as a ‘glycan gate’ pushing the RBD from the ‘down’ to the ‘up’ conformation by intercalating between residues F490, Y489, F456 and R457 of the ACE2 binding motif in a ‘hand-jive’ motion (Fig. 2a–e, 3 and Supplementary Video 4). Therefore, the N343 glycan plays an active role in initiating the transition, distinct from the stabilizing roles of glycans N165 and N234. This gating mechanism was initially visualized in several successful pathways of spike opening and then confirmed through analysis of all 310 successful pathways in which the N343 glycan was found to form contacts (within 3.5 Å) with each of the aforementioned residues in every successful pathway (Supplementary Fig. 7). The same mechanistic behaviour of the N343 glycan was observed in two fully independent WE simulations, suggesting the result is robust despite potentially incomplete sampling that can challenge WE and other enhanced sampling simulation methods¹⁵.

To test the role of the N343 glycan as a key-gating residue, we performed BLI experiments. BLI experiments assess the binding level of the spike RBM (residues 438 to 508) to ACE2, acting as a proxy for the relative proportion of RBDs in the ‘up’ position for each spike variant. No residues directly involved in the binding were mutated (that is, at the RBM–ACE2 interface) to ensure controlled detection of the impact of RBD opening in response to mutations. Although previous results have shown reduced binding levels for N165A and N234A variants in the SARS-CoV-2 S-2P protein⁸, the N343A variant displayed an even greater decrease in ACE2 binding, reducing the spike binding level by ~56% (Fig. 4 and Supplementary Table 1). As a negative control, the S383C/D985C variant²⁴, which is

expected to be locked by disulfides into the three-RBD-down conformation, showed no association with the ACE2 receptor. These results support the hypothesis that the RBD_{up} conformation is substantially affected by glycosylation at position N343.

Atomic details of the opening mechanism. The RBD_{down} state features a hydrogen bond between T415 of the RBD_A and K986 of chain C, a salt bridge between R457 of RBD_A and D364 of RBD_B, and a salt bridge between K462 of RBD_A and D198 of N-terminal domain C (NTD_C) (Fig. 5a–c,e and Supplementary Fig. 8). The hydrogen bond T415_A–K986_C spends an average of 12% of the successful pathways to the ‘up’ state before K986_C makes a short lived (2% average duration to the ‘up’ state) salt bridge with RBD_A–D427. (Fig. 5b,e and Supplementary Fig. 8). Next, K986_C forms salt bridges with E990_C and E748_C as the RBD_A continues to open. These contacts are formed in all 310 successful pathways (Supplementary Fig. 8). Mutation of K986 to proline has been used to stabilize the prefusion spike^{25,26}, including in vaccine development²⁷, and these simulations provide molecular context to an additional role of this residue in RBD opening.

Subsequently, at an average of 16% of the way through the successful pathways to the ‘up’ state, the R457_A–D364_B salt bridge is broken, prompting the RBD_A to twist upward, away from RBD_B towards RBD_C and forming a salt bridge between R408 of RBD_A and D405 of RBD_C (Fig. 5c,e and Supplementary Fig. 8). This salt bridge persists for 20% of the successful trajectories to the ‘up’ state and is present in all 310 successful pathways.

A salt bridge between R466 of RBD_A and E132 from NTD_B is present in 189 out of 204 successful pathways to the ‘up’ state, and all 106 pathways to the ‘open’ state. This contact is most prevalent during the transition between the ‘up’ and ‘open’ state. Finally, the salt bridge between D428 of RBD_A and R454 of RBD_C is present only in all 106 pathways from the ‘up’ to the ‘open’ state and is the last salt bridge between the RBD and the spike in the ‘open’ state before the S1 subunit begins to peel off (Fig. 5d,e and Supplementary Fig. 8), at which point the last remaining contact to the RBD_A is the glycan at position N165 of NTD_B.

Additional BLI experiments of the key identified spike residues R408A, D405A and D427A corroborate the pathways observed in

Data availability

Data supporting the findings of this study are included in the article and its Supplementary Information files. We endorse the community principles around open sharing of COVID-19 simulation data²⁹. All simulation input files and data are available at the NSF MolSSI COVID-19 Molecular Structure and Therapeutics Hub at <https://covid.molssi.org> and the Amaro Lab website <http://amarolab.ucsd.edu>. Source data are provided with this paper.

Code availability

This study utilized the standard builds of the simulation software WESTPA 2020.02 (<https://github.com/westpa/westpa>) and AMBER 18 (<https://ambermd.org>) according to best practices for running WE simulations³⁰ with no special modifications.

Acknowledgements

We are grateful for the efforts of the Texas Advanced Computing Center (TACC) Longhorn team and for the computing time made available through a Director's Discretionary Allocation (made possible by National Science Foundation (NSF) award OAC-1818253). We thank Z. Gaieb for helpful discussions around system construction. We thank M. Tatineni for help with computing on SDSC Comet, as well as a COVID-19 HPC Consortium Award for computing time. We also thank C. Simmerling and his research group (SUNY Stony Brook), and A. Mulholland and his research group (University of Bristol), for helpful discussions related to the spike protein, as well as D. Zuckerman, J. Copperman, M. Zwier and A. Saglam for helpful methodological discussions. T.S. is funded by an NSF GRFP grant (DGE-1650112). This work was supported by: a National Institutes of Health (NIH) grant (GM132826); an NSF RAPID grant (MCB-2032054); an award from the RCSA Research Corp. and a UC San Diego Moores Cancer Center 2020 SARS-CoV-2 seed grant to R.E.A.; an NIH grant (R01-GM31749) to J.A.M.; an NIH grant (R01-AI127521) to J.S.M.; an NIH grant (R01 GM115805) and an NSF grant (CHE-1807301) to L.T.C.; and NIGMS grants (R01 GM29169 and R35 GM139453) to J.F. A.O. and G.M. acknowledge support by

the US Department of Energy, Office of Science, Basic Energy Sciences under award DE-SC0002164 (underlying dynamical algorithms), and by the US National Science Foundation under awards STC 1231306 (underlying data-analytical techniques) and DBI-2029533 (underlying data-analytical models).

Author contributions

T.S. and S.-H.A. contributed equally to this work. R.E.A. and L.T.C. oversaw the project. T.S. and L.C. prepared the simulation model. T.S. and S.-H.A. performed WE simulations and A.T.B. provided WESTPA scripts. S.-H.A., A.T.B., T.S. and L.T.C. carried out WE analysis. T.S. and F.L.K. performed simulation analyses. L.C., T.S. and F.L.K. created figures and videos. J.S.M. designed and oversaw BLI experiments. R.S.M. and J.A.G. performed BLI experiments and wrote the corresponding parts in Results and Methods. J.F. and A.O. directed the ManifoldEM study, and E.S., F.A.-R., S.M. and G.M. performed the ManifoldEM study. E.S. and F.A.-R. analysed the results and compared them with WE simulations. E.S. and J.F. described the ManifoldEM methods and results. T.S., S.H.-A., L.T.C. and R.E.A. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41557-021-00758-3>.

Correspondence and requests for materials should be addressed to L.T.C. or R.E.A.

Peer review information Nature Chemistry thanks Syma Khalid and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.