

1

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

3

4 Terra Sztain^{1†}, Surl-Hee Ahn^{1†}, Anthony T. Bogetti², Lorenzo Casalino¹, Jory A. Goldsmith³,
5 Evan Seitz⁴, Ryan S. McCool³, Fiona L. Kearns¹, Francisco Acosta-Reyes⁵, Suvrajit Maji⁵,
6 Ghoncheh Mashayekhi⁶, J. Andrew McCammon^{1,7}, Abbas Ourmazd⁶, Joachim Frank^{4,5}, Jason S.
7 McLellan³, Lillian T. Chong^{2*}, Rommie E. Amaro^{1*}

8

9 1. Department of Chemistry and Biochemistry, UC San Diego, La Jolla, CA 92093
10 2. Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260
11 3. Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78712
12 4. Department of Biological Sciences, Columbia University, New York, NY, 10032, USA
13 5. Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center,
14 New York, NY 10032, USA

15 6. Department of Physics, University of Wisconsin-Milwaukee, 3135 N. Maryland Ave,
16 Milwaukee, WI 53211, USA

17 7. Department of Pharmacology, UC San Diego, La Jolla, CA 92093

18 † These authors contributed equally to this work.

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

20

21 **Abstract**

22 SARS-CoV-2 infection is controlled by the opening of the spike protein receptor binding domain
23 (RBD), which transitions from a glycan-shielded “down” to an exposed “up” state in order to

24 bind the human ACE2 receptor and infect cells. While snapshots of the “up” and “down” states
25 have been obtained by cryoEM and cryoET, details of the RBD opening transition evade
26 experimental characterization. Here, over 130 μs of weighted ensemble (WE) simulations of the
27 fully glycosylated spike ectodomain allow us to characterize more than 300 continuous,
28 kinetically unbiased RBD opening pathways. Together with ManifoldEM analysis of cryo-EM
29 data and biolayer interferometry experiments, we reveal a gating role for the N-glycan at position
30 N343, which facilitates RBD opening. Residues D405, R408, and D427 also participate. The
31 atomic-level characterization of the glycosylated spike activation mechanism provided herein
32 achieves a new high-water mark for ensemble pathway simulations and offers a foundation for
33 understanding the fundamental mechanisms of SARS-CoV-2 viral entry and infection.

34

35 **Main Text**

36 **Introduction**

37 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an enveloped RNA virus and
38 the causative agent of COVID-19, a disease that has caused significant morbidity and mortality
39 worldwide.^{1,2} The main infection machinery of the virus is the spike (S) protein, which sits on
40 the outside of the virus, is the first point of contact that the virion makes with the host cell, and is
41 a major viral antigen.³ A significant number of cryoEM structures of the spike protein have been
42 recently reported, collectively informing on structural states of the spike protein. The vast
43 majority of resolved structures fall into either “down” or “up” states, as defined by the position
44 of the receptor binding domain (RBD), which modulates interaction with the ACE2 receptor for
45 cell entry.^{4,5,6}

46

47 The RBDs must transition from a “down” to an “up” state for the receptor binding motif to be
48 accessible for ACE2 binding (**Fig. 1**), and therefore the activation mechanism is essential for cell
49 entry. Mothes et al.⁷ used single-molecule fluorescence (Förster) resonance energy transfer
50 (smFRET) imaging to characterize spike dynamics in real-time. Their work showed that the
51 spike dynamically visits four distinct conformational states, the population of which are
52 modulated by the presence of the human ACE2 receptor and antibodies. However, smFRET, as
53 well as conventional structural biology techniques, are unable to inform on the atomic-level
54 mechanisms underpinning such dynamical transitions. Recently, all-atom molecular dynamics
55 (MD) simulations of the spike protein with experimentally accurate glycosylation together with
56 corroborating experiments indicated the extensive shielding by spike glycans, as well as a
57 mechanical role for glycans at positions N165 and N234 in supporting the RBD in the “open”
58 conformation.⁸ Conventional MD simulations as performed in Casalino et al.⁸ also revealed
59 microsecond-timescale dynamics to better characterize the spike dynamics but were limited to
60 sampling configurations that were similar in energy to the cryoEM structures. Several enhanced
61 sampling MD simulations have been performed to study this pathway, however, these
62 simulations lacked glycosylation for the spike protein⁹ or involve the addition of an external
63 force¹⁰ or did not provide mechanistic detail.¹¹

64
65
66 In this study, we characterized the spike RBD opening pathway for the fully glycosylated SARS-
67 CoV-2 spike protein, in order to gain a detailed understanding of the activation mechanism. We
68 used the weighted ensemble (WE) path sampling strategy^{12,13} (**Supplementary Fig. 1**) to enable
69 the simulation of atomistic pathways for the spike opening process. As a path sampling strategy,
70 WE focuses computing power on the functional transitions between stable states rather than the

71 stable states themselves.¹⁴ This is achieved by running multiple trajectories in parallel and
72 periodically replicating trajectories that have transitioned from previously visited to newly
73 visited regions of configurational space,¹⁵ thus minimizing the time spent waiting in the initial
74 stable state for “lucky” transitions over the free energy barrier. Given that these transitions are
75 much faster than the waiting times,^{16,17} the WE strategy can be orders of magnitude more
76 efficient than conventional MD simulations in generating pathways for rare events such as
77 protein folding and protein binding.^{18,19} This efficiency is even higher for slower processes,
78 increasing *exponentially* with the effective free energy barrier.²⁰ Not only are dynamics carried
79 out without any biasing force or modifications to the free energy landscape, but suitable
80 assignment of statistical weights to trajectories provides an *unbiased* characterization of the
81 system’s time-dependent ensemble properties.¹³ The WE strategy therefore generates continuous
82 pathways with unbiased dynamics, yielding the most direct, atomistic views for analyzing the
83 mechanism of functional transitions, including elucidation of transient states that are too fleeting
84 to be captured by laboratory experiments. Furthermore, while the strategy requires a progress
85 coordinate toward the target state, the definition of this target state need not be fixed in advance
86 when applied under equilibrium conditions,²¹ enabling us to refine the definition of the target
87 “open” state of the spike protein based on the probability distribution of protein conformations
88 sampled by the simulation.

89

90 Our work characterizes a series of transition pathways of the spike opening, in agreement with
91 conformations detected in cryoEM dataset by ManifoldEM,²² and identifies key residues,
92 including a glycan at position N343, that participate in the opening mechanism. Our simulation

93 findings are corroborated by biolayer interferometry experiments, which demonstrate a reduction
94 in the ability of the spike to interact with ACE2 after mutation of these key residues.

95

96 **Results and Discussion**

97

98 **Weighted ensemble simulations of spike opening**

99 As mentioned above, simulations of the spike opening process require an enhanced sampling
100 strategy as the process occurs beyond the microsecond timescale (*i.e.* seconds timescale⁷). We
101 therefore used the weighted ensemble (WE) path sampling strategy, which enabled the
102 generation of continuous, atomistic pathways for the spike opening process with unbiased
103 dynamics (**Fig. 2A-E, Supplementary Video 1**); these pathways were hundreds of ns long,
104 excluding the waiting times in the initial “down” state. The protein model was based on the head
105 region (residues 16 to 1140) of the glycosylated SARS-CoV-2 spike from Casalino et al.⁸ (**Fig.**
106 **1**), which in turn was built on the cryoEM structure of the three-RBD-down spike (PDB ID:
107 6VXX⁵). The entire simulation system, including explicit water and salt ions, reaches almost
108 half a million atoms. We focused sampling along a two-dimensional progress coordinate to track
109 RBD opening: the difference in the center of mass of the spike core to the RBD and the root
110 mean squared deviation (RMSD) of the RBD (**Fig. 2F-G**). On the SDSC Comet and TACC
111 Longhorn supercomputers, 100 GPUs ran the WE simulations in parallel for over a month,
112 generating over 130 μs of glycosylated spike trajectories and more than 200 TB of trajectory
113 data. We simulated a total of 310 independent pathways, including 204 pathways from the RBD-
114 down conformation (PDB ID: 6VXX⁵) to the RBD-up conformation (PDB ID: 6VSB⁴) and 106
115 pathways from the RBD-down to the RBD-open state, in which the RBD twists open beyond the

116 6VSB⁴ cryoEM structure. Remarkably, the RBD-open state that we sampled includes
117 conformations that align closely with the ACE2-bound spike cryoEM structure (PDB ID: 7A95⁶)
118 even though this structure was not a target state of our progress coordinate (**Fig. 2F-G**,
119 **Supplementary Video 1, Supplementary Figs. 2 and 3**). This result underscores the value of
120 using (i) equilibrium WE simulations that do not require a fixed definition of the target state and
121 (ii) a two-dimensional progress coordinate that allows the simulations to sample unexpected
122 conformational space along multiple degrees of freedom. The ACE2-bound spike conformation
123 has also been sampled by the Folding@home distributed computing project,¹¹ and RBD rotation
124 has been detected in cryoEM experiments.⁶

125

126 **Comparison to spike conformations detected by ManifoldEM**

127 To validate our simulated RBD-down to RBD-up pathway, the ManifoldEM framework²² was
128 applied using the cryo-EM dataset of PDB 6VSB from McLellan and colleagues.⁴ The
129 ManifoldEM method allows characterization of conformational variations as obtained from a
130 single-particle cryo-EM ensemble of a molecule in thermal equilibrium. Two conformational
131 coordinates (CCs) (i.e., collective motion coordinates) were discovered from this dataset, and
132 observed from several exemplary projection directions (PDs) showing a (1) RBD-down to RBD-
133 up pathway and (2) RBD outward opening pathway (**Supplementary Fig. 4 Supplementary**
134 **Videos 2 and 3**).

135

136 These projections were next aligned to corresponding 2D projections of Coulomb potential maps
137 generated with frames from the WE simulation (**Supplementary Fig. 5, Supplementary Videos**
138 **2 and 3**). Overall, there was very good agreement between the ManifoldEM conformational

139 coordinates and WE trajectory, aside from two discrepancies. First, the CC2 observed in the
140 ManifoldEM included concerted opening of all three RBDs, while the WE focused sampling on
141 the opening of a single RBD (**Supplementary Video 2**). Second, the WE trajectory ultimately
142 opens to an RBD — core distance 11 Å greater than the most open conformation in the
143 ManifoldEM. This is likely because the simulations sample the S1 subunit *en route* to the post-
144 fusion conformation, whereas the experimental dataset does not.

145

146 **The N343 glycan gates RBD opening**

147 In the “down” state, the RBD of the SARS-CoV-2 spike is shielded by glycans at positions
148 N165, N234, and N343.²⁴ While glycan shielding had been investigated for the RBD-down and
149 RBD-up states,⁸ our WE simulations allowed characterization of shielding *during* the opening
150 process, revealing an abrupt decrease in glycan shielding when the RBD transitions from the
151 “down” to the “up” state. The glycans at position N165 and N234 consistently shield the receptor
152 binding motif, while shielding by the N343 glycan decreases with RBD opening.

153 (**Supplementary Fig. 6**). Beyond shielding, a structural role for glycans at positions N165 and
154 N234 has been recently reported, stabilizing the RBD in the “up” conformation through a “load-
155 and-lock” mechanism.⁸

156

157 Our WE simulations reveal an even more specific, critical role of a glycan in the opening
158 mechanism of the spike: the N343 glycan acts as a “glycan gate” pushing the RBD from the
159 “down” to the “up” conformation by intercalating between residues F490, Y489, F456, and R457
160 of the ACE2 binding motif in a “hand-jive” motion (**Fig. 2A-E, Fig. 3, Supplementary Video**
161 **4**). Therefore, the N343 glycan plays an active role in initiating the transition, distinct from the

162 stabilizing roles of glycans N165 and N234. This gating mechanism was initially visualized in
163 several successful pathways of spike opening and then confirmed through analysis of all 310
164 successful pathways in which the N343 glycan was found to form contacts (within 3.5 Å) with
165 each of the aforementioned residues in every successful pathway (**Supplementary Fig. 7**). The
166 same mechanistic behavior of the N343 glycan was observed in two fully independent WE
167 simulations, suggesting the result is robust despite potentially incomplete sampling that can
168 challenge WE and other enhanced sampling simulation methods.¹⁵

169
170 To test the role of the N343 glycan as a key-gating residue, we performed biolayer
171 interferometry (BLI) experiments. BLI experiments assess the binding level of the spike receptor
172 binding motif (RBM, residues 438 to 508) to ACE2, acting as a proxy for the relative proportion
173 of RBDs in the “up” position for each spike variant. No residues directly involved in the binding
174 were mutated (i.e., at the RBM-ACE2 interface), to ensure controlled detection of the impact of
175 RBD opening in response to mutations. Although previous results have shown reduced binding
176 levels for N165A and N234A variants in the SARS-CoV-2 S-2P protein,⁸ the N343A variant
177 displayed an even greater decrease in ACE2 binding, reducing spike binding level by ~56 %
178 (**Fig. 4, Supplementary Table 1**). As a negative control, the S383C/D985C variant,²⁵ which is
179 expected to be locked by disulfides into the three-RBD-down conformation, showed no
180 association with the ACE2 receptor. These results support the hypothesis that the RBD-up
181 conformation is significantly affected by glycosylation at position N343.

182
183
184 **Atomic details of the opening mechanism**
185 The RBD-down state features a hydrogen bond between T415 of the RBD (chain A) and K986 of
186 chain C, a salt bridge between R457 of RBD_A and D364 of RBD_B, and a salt bridge between

187 K462 of RBD_A and D198 of NTD_C (**Fig. 5A-C,E Supplementary Fig. 8**). The hydrogen bond
188 T415_A - K986_C spends an average of 12% of the successful pathways to the “up” state before
189 K986_C makes a short lived (2% average duration to “up” state) salt bridge with RBD_A D427.
190 (**Fig. 5B,E Supplementary Fig. 8**) Next, K986_C forms salt bridges with E990_C and E748_C as the
191 RBD_A continues to open. These contacts are formed in all 310 successful pathways
192 (**Supplementary Fig. 8**). Mutation of K986 to proline has been used to stabilize the pre-fusion
193 spike,^{23,26} including in vaccine development,²⁷ and these simulations provide molecular context
194 to an additional role of this residue in RBD opening.

195
196 Subsequently, at an average of 16% of the way through the successful pathways to the “up” state,
197 the R457_A - D364_B salt bridge is broken, prompting the RBD_A to twist upward, away from RBD_B
198 towards RBD_C and forming a salt bridge between R408 of RBD_A and D405 of RBD_C (**Figs. 5C**
199 **and 5E, Supplementary Fig. 8**). This salt bridge persists for 20% of the successful trajectories
200 to the “up” state and is present in all 310 successful pathways.

201
202 A salt bridge between R466 of RBD_A and E132 from NTD_B is present in 189 out of 204
203 successful pathways to the “up” state, and all 106 pathways to the “open” state. This contact is
204 most prevalent during the transition between the “up” and “open” state. Finally, the salt bridge
205 between D428 of RBD_A and R454 of RBD_C is present only in all 106 pathways from the “up” to
206 the “open” state and is the last salt bridge between the RBD and the spike in the “open” state
207 before the S1 subunit begins to peel off (**Fig. 5D and 5E, Supplementary Fig. 8**), at which point
208 the last remaining contact to the RBD_A is the glycan at position N165 of NTB_B.

209

210 Additional BLI experiments of the key identified spike residues R408A, D405A, and D427A
211 corroborate the pathways observed in our simulations. Each of these reduces the binding
212 interactions of the spike with ACE2 by ~13%, ~27%, and ~52%, respectively (**Fig. 4**,
213 **Supplementary Table 1**). We also note that identified residues D198, N343, D364, D405, R408,
214 T415, D427, D428, R454, R457, R466, E748, K986 and E990 are conserved between SARS-
215 CoV and SARS-CoV-2 spikes, supporting their significance in coordinating the primary spike
216 function of RBD opening. The emerging mutant SARS-CoV-2 strains B.1: D614G; B.1.1.7:
217 H69-V70 deletion, Y144-Y145 deletions, N501Y, A570D, D614G, P681H, T716I, S982A,
218 D1118H; B.1.351: L18F, D80A, D215G, R246I, K417N, E484K, N501Y, D614G, A701V; P1:
219 L18F, T20N, P26S, D138Y, R190S, K417T, E484K, N501Y, D614G, H655Y, T1027I and
220 CAL.20C: L452R, D614G²⁸ do not contain mutants in the residues we identified here to facilitate
221 RBD opening. Analysis of neighboring residues and glycans to those mutated in the emerging
222 strains along the opening pathway is detailed in **Supplementary Table 2**, and distances between
223 each residue and glycan to RBD_A is summarized in **Supplementary Video 5**.

224
225
226

Conclusions

227 We report extensive weighted ensemble molecular dynamics simulations of the glycosylated
228 SARS-CoV-2 spike head characterizing the transition from the “down” to “up” conformation of
229 the RBD. Over 130 microseconds of simulation provide more than 300 independent RBD
230 opening transition pathways. The simulated opening pathways align very well to conformations
231 detected from cryoEM with the ManifoldEM method. Analysis of these pathways from
232 independent WE simulations indicates a clear gating role for the glycan at N343, which lifts and
233 stabilizes the RBD throughout the opening transition. We also characterize an “open” state of the

234 spike RBD, in which the N165 glycan of chain B is the last remaining contact with the RBD *en*
235 *route* to further opening of S1. Biolayer interferometry experiments of residues identified as key
236 in the opening transitions, including N343, D405, R408, and D427, broadly supported our
237 computational findings. Notably, a 56% decrease in ACE2 binding of the N343A mutant,
238 compared to a 40% decrease in N234A mutant, and 10% decrease in the N165A mutant reported
239 previously⁸ evidenced the key role of N343 in gating and assisting the RBD opening process,
240 highlighting the importance of sampling functional transitions to fully understand mechanistic
241 detail. None of the individual mutations fully abolished ACE2 binding, indicating the virus has
242 evolved a mechanism involving multiple residues to coordinate spike opening. Our work
243 indicates a critical gating role of the N343 glycan in spike opening and provides new insights to
244 mechanisms of viral infection for this important pathogen.

245

246 **References**

- 247 (1) Chan, J. F.-W.; Yuan, S.; Kok, K.-H.; To, K. K.-W.; Chu, H.; Yang, J.; Xing, F.; Liu, J.;
248 Yip, C. C.-Y.; Poon, R. W.-S.; Tsui, H.-W.; Lo, S. K.-F.; Chan, K.-H.; Poon, V. K.-M.;
249 Chan, W.-M.; Ip, J. D.; Cai, J.-P.; Cheng, V. C.-C.; Chen, H.; Hui, C. K.-M.; Yuen, K.-Y.
250 A Familial Cluster of Pneumonia Associated with the 2019 Novel Coronavirus Indicating
251 Person-to-Person Transmission: A Study of a Family Cluster. *The Lancet* **2020**, 395
252 (10223), 514–523. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9).
- 253 (2) Lu, R.; Zhao, X.; Li, J.; Niu, P.; Yang, B.; Wu, H.; Wang, W.; Song, H.; Huang, B.; Zhu,
254 N.; Bi, Y.; Ma, X.; Zhan, F.; Wang, L.; Hu, T.; Zhou, H.; Hu, Z.; Zhou, W.; Zhao, L.;
255 Chen, J.; Meng, Y.; Wang, J.; Lin, Y.; Yuan, J.; Xie, Z.; Ma, J.; Liu, W. J.; Wang, D.; Xu,
256 W.; Holmes, E. C.; Gao, G. F.; Wu, G.; Chen, W.; Shi, W.; Tan, W. Genomic

- 257 Characterisation and Epidemiology of 2019 Novel Coronavirus: Implications for Virus
258 Origins and Receptor Binding. *Lancet* **2020**, *395* (10224), 565–574.
259 [https://doi.org/10.1016/s0140-6736\(20\)30251-8](https://doi.org/10.1016/s0140-6736(20)30251-8).
- 260 (3) Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*
261 **2016**, *3* (1), 237–261. <https://doi.org/10.1146/annurev-virology-110615-042301>.
- 262 (4) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham,
263 B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion
264 Conformation. *Science* **2020**, *367* (6483), 1260–1263.
265 <https://doi.org/10.1126/science.abb2507>.
- 266 (5) Walls, A. C.; Park, Y.-J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure,
267 Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2),
268 281-292.e6. <https://doi.org/10.1016/j.cell.2020.02.058>.
- 269 (6) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J.
270 J.; Gamblin, S. J. Receptor Binding and Priming of the Spike Protein of SARS-CoV-2 for
271 Membrane Fusion. *Nature* **2020**, *588* (7837), 327–330. [https://doi.org/10.1038/s41586-020-2772-0](https://doi.org/10.1038/s41586-020-
272 2772-0).
- 273 (7) Lu, M.; Uchil, P. D.; Li, W.; Zheng, D.; Terry, D. S.; Gorman, J.; Shi, W.; Zhang, B.;
274 Zhou, T.; Ding, S.; Gasser, R.; Prévost, J.; Beaudoin-Bussières, G.; Anand, S. P.; Laumaea,
275 A.; Grover, J. R.; Liu, L.; Ho, D. D.; Mascola, J. R.; Finzi, A.; Kwong, P. D.; Blanchard, S.
276 C.; Mothes, W. Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus
277 Particles. *Cell Host Microbe* **2020**, *28* (6), 880-891.e8.
278 <https://doi.org/10.1016/j.chom.2020.11.001>.

- 279 (8) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.;
280 Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E.
281 Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.*
282 **2020.** <https://doi.org/10.1021/acscentsci.0c01056>.
- 283 (9) Gur, M.; Taka, E.; Yilmaz, S. Z.; Kilinc, C.; Aktas, U.; Golcuk, M. Conformational
284 Transition of SARS-CoV-2 Spike Glycoprotein between Its Closed and Open States. *J.*
285 *Chem. Phys.* **2020**, *153* (7), 075101. <https://doi.org/10.1063/5.0011141>.
- 286 (10) Fallon, L.; Belfon, K.; Raguette, L.; Wang, Y.; Corbo, C.; Stepanenko, D.; Cuomo, A.;
287 Guerra, J.; Budhan, S.; Varghese, S.; Rizzo, R.; Simmerling, C. Free Energy Landscapes
288 for RBD Opening in SARS-CoV-2 Spike Glycoprotein Simulations Suggest Key
289 Interactions and a Potentially Druggable Allosteric Pocket. **2020.**
290 <https://doi.org/10.26434/chemrxiv.13502646.v1>.
- 291 (11) SARS-CoV-2 Simulations Go Exascale to Capture Spike Opening and Reveal Cryptic
292 Pockets Across the Proteome | bioRxiv
293 <https://www.biorxiv.org/content/10.1101/2020.06.27.175430v3> (accessed Jan 13, 2021).
- 294 (12) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein
295 Association Reactions. *Biophys. J.* **1996**, *70* (1), 97–110. [https://doi.org/10.1016/S0006-3495\(96\)79552-8](https://doi.org/10.1016/S0006-3495(96)79552-8).
- 297 (13) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The “Weighted Ensemble” Path Sampling
298 Method Is Statistically Exact for a Broad Class of Stochastic Processes and Binning
299 Procedures. *J. Chem. Phys.* **2010**, *132* (5), 054107. <https://doi.org/10.1063/1.3306345>.

- 300 (14) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-Sampling Strategies for Simulating
301 Rare Events in Biomolecular Systems. *Curr Opin Struct Biol* **2017**, *43*, 88–94.
302 <https://doi.org/10.1016/j.sbi.2016.11.019>.
- 303 (15) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology,
304 Applications, and Software. *Annu Rev Biophys* **2017**, *46*, 43–57.
305 <https://doi.org/10.1146/annurev-biophys-070816-033834>.
- 306 (16) Pratt, L. R. A Statistical Method for Identifying Transition States in High Dimensional
307 Problems. *J. Chem. Phys.* **1986**, *85* (9), 5045–5048. <https://doi.org/10.1063/1.451695>.
- 308 (17) Zuckerman, D. M.; Woolf, T. B. Transition Events in Butane Simulations: Similarities
309 across Models. *J. Chem. Phys.* **2002**, *116* (6), 2586–2591.
310 <https://doi.org/10.1063/1.1433501>.
- 311 (18) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.;
312 Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding
313 Times. *J Am Chem Soc* **2019**, *141* (16), 6519–6526. <https://doi.org/10.1021/jacs.8b10735>.
- 314 (19) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate
315 Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chem. Sci.* **2019**, *10* (8),
316 2360–2372. <https://doi.org/10.1039/C8SC04811H>.
- 317 (20) DeGrave, A. J.; Ha, J.-H.; Loh, S. N.; Chong, L. T. Large Enhancement of Response Times
318 of a Protein Conformational Switch by Computational Design. *Nature Communications*
319 **2018**, *9* (1), 1013. <https://doi.org/10.1038/s41467-018-03228-6>.
- 320 (21) Suárez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.;
321 Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information

- 322 Using a Weighted Ensemble of Trajectories. *J. Chem. Theory Comput.* **2014**, *10* (7), 2658–
323 2667. <https://doi.org/10.1021/ct401065r>.
- 324 (22) Dashti, A.; Schwander, P.; Langlois, R.; Fung, R.; Li, W.; Hosseiniyadeh, A.; Liao, H. Y.;
325 Pallesen, J.; Sharma, G.; Stupina, V. A.; Simon, A. E.; Dinman, J. D.; Frank, J.; Ourmazd,
326 A. Trajectories of the ribosome as a Brownian nanomachine. *Proceedings of the National
327 Academy of Sciences* **2014**, *111*, 17492–17497.
- 328 (23) Hsieh, C.-L.; Goldsmith, J. A.; Schaub, J. M.; DiVenere, A. M.; Kuo, H.-C.; Javanmardi,
329 K.; Le, K. C.; Wrapp, D.; Lee, A. G.; Liu, Y.; Chou, C.-W.; Byrne, P. O.; Hjorth, C. K.;
330 Johnson, N. V.; Ludes-Meyers, J.; Nguyen, A. W.; Park, J.; Wang, N.; Amengor, D.;
331 Lavinder, J. J.; Ippolito, G. C.; Maynard, J. A.; Finkelstein, I. J.; McLellan, J. S. Structure-
332 Based Design of Prefusion-Stabilized SARS-CoV-2 Spikes. *Science* **2020**, *369* (6510),
333 1501–1505. <https://doi.org/10.1126/science.abd0826>.
- 334 (24) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan
335 Analysis of the SARS-CoV-2 Spike. *Science* **2020**.
336 <https://doi.org/10.1126/science.abb9983>.
- 337 (25) Henderson, R.; Edwards, R. J.; Mansouri, K.; Janowska, K.; Stalls, V.; Gobeil, S. M. C.;
338 Kopp, M.; Li, D.; Parks, R.; Hsu, A. L.; Borgnia, M. J.; Haynes, B. F.; Acharya, P.
339 Controlling the SARS-CoV-2 Spike Glycoprotein Conformation. *Nature Structural &
340 Molecular Biology* **2020**, *27* (10), 925–933. <https://doi.org/10.1038/s41594-020-0479-4>.
- 341 (26) Pallesen, J.; Wang, N.; Corbett, K. S.; Wrapp, D.; Kirchdoerfer, R. N.; Turner, H. L.;
342 Cottrell, C. A.; Becker, M. M.; Wang, L.; Shi, W.; Kong, W.-P.; Andres, E. L.; Kettenbach,
343 A. N.; Denison, M. R.; Chappell, J. D.; Graham, B. S.; Ward, A. B.; McLellan, J. S.

344 Immunogenicity and Structures of a Rationally Designed Prefusion MERS-CoV Spike
345 Antigen. *PNAS* **2017**, *114* (35), E7348–E7357. <https://doi.org/10.1073/pnas.1707304114>.
346 (27) Ryan Cross. The Tiny Tweak behind COVID-19 Vaccines. *Chemical & Engineering News*.
347 September 29, 2020.
348 (28) Corum, J.; Zimmer, C. Coronavirus Variants and Mutations. *The New York Times*. February
349 10, 2021.
350 (29) Amaro, R. E.; Mulholland, A. J. A Community Letter Regarding Sharing Biomolecular
351 Simulation Data for COVID-19. *J. Chem. Inf. Model.* **2020**, *60* (6), 2653–2656.
352 <https://doi.org/10.1021/acs.jcim.0c00319>.
353 (30) Bogetti, A. T.; Barmak, M.; Dickson, A.; Pratt, A. J.; Saglam, A. S; Harrison, P. O.;
354 Adelman, J. L.; Dudek, M.; Torrillo, P. A.; DeGrave, A. J.; Adhikari, U.; Zwier, M.C.;
355 Zuckerman, D. M.; Chong, L.T. " *Living J. Comput. Mol. Sci.* **2019**, *1* (2), 10607.
356 <https://doi.org/10.33011/livecoms.1.2.10607>
357
358 **Acknowledgements**
359 We are grateful for the efforts of the Texas Advanced Computing Center (TACC)
360 Longhorn team and for the compute time made available through a Director's Discretionary
361 Allocation (made possible by the National Science Foundation award OAC-1818253). We thank
362 Dr. Zied Gaieb for helpful discussions around system construction. We thank Mahidhar Tatineni
363 for help with computing on SDSC Comet, as well as a COVID19 HPC Consortium Award for
364 compute time. We also thank Prof. Carlos Simmerling and his research group (SUNY Stony
365 Brook), and Prof. Adrian Mulholland and his research group (University of Bristol) for helpful

366 discussions related to the spike protein, as well as Prof. Daniel Zuckerman, Dr. Jeremy
367 Copperman, Dr. Matthew Zwier, and Dr. Sinam Saglam for helpful methodological discussions.

368

369 **Funding**

370 T.S. is funded by NSF GRFP DGE-1650112. This work was supported by NIH GM132826, NSF
371 RAPID MCB-2032054, an award from the RCSA Research Corp., and a UC San Diego Moores
372 Cancer Center 2020 SARS-COV-2 seed grant to R.E.A.; NIH grant R01-GM31749 to J.A.M.;
373 NIH grant R01-AI127521 to J.S.M; NIH R01 GM115805 and NSF CHE-1807301 to L.T.C;
374 and NIGMS R01 GM29169 and R35 GM139453 to J.F.

375

376 **Author contributions**

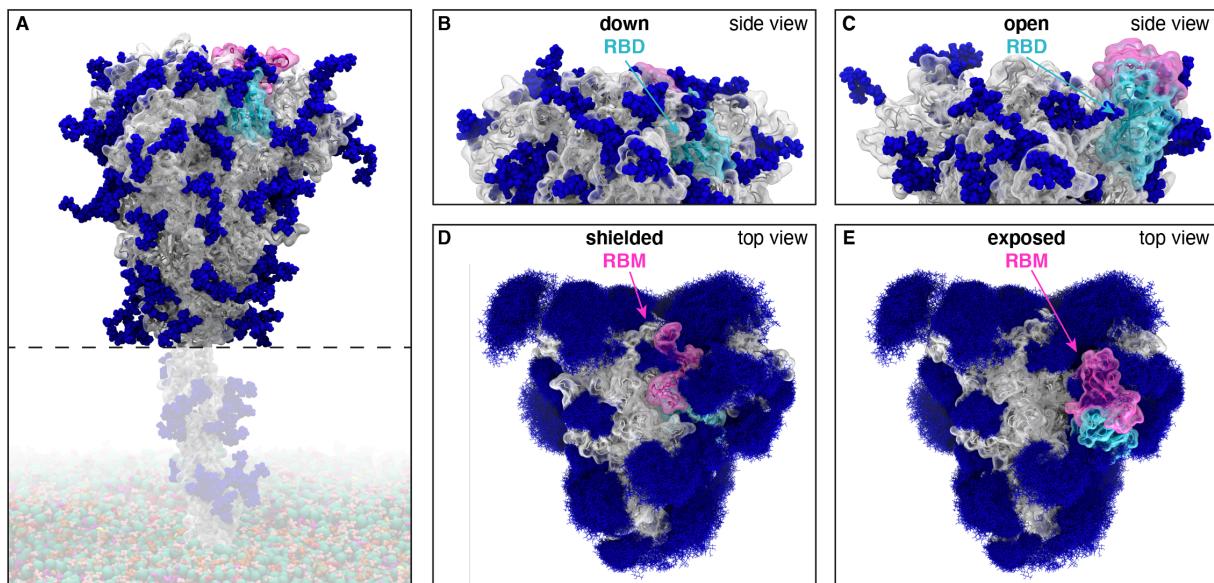
377 T.S. and S.A. contributed equally to this work. R.E.A. and L.T.C. oversaw the project. T.S. and
378 L.C. prepared the simulation model. T.S. and S.A. performed WE simulations and A.B. provided
379 WESTPA scripts. S.A., A.B., T.S., and L.T.C. carried out WE analysis. T.S., and F.L.K.
380 performed simulation analyses. L.C., T.S., and F.L.K. created figures and movies. J.S.M.
381 designed and oversaw biolayer interferometry experiments. R.S.M. and J.A.G. performed
382 biolayer interferometry experiments and wrote the corresponding parts in the Results and
383 Methods sections. J.F. and A.O. directed the ManifoldEM study, E.S., F.A., S.M. and G.M.
384 performed the ManifoldEM study. E.S. and F.A. analyzed the results and compared them with
385 WE simulations. E.S. and J.F. described the ManifoldEM methods and results. T.S., S.A.,
386 L.T.C., and R.E.A. wrote the manuscript with contributions from all authors.

387

388 **Competing interests statement**

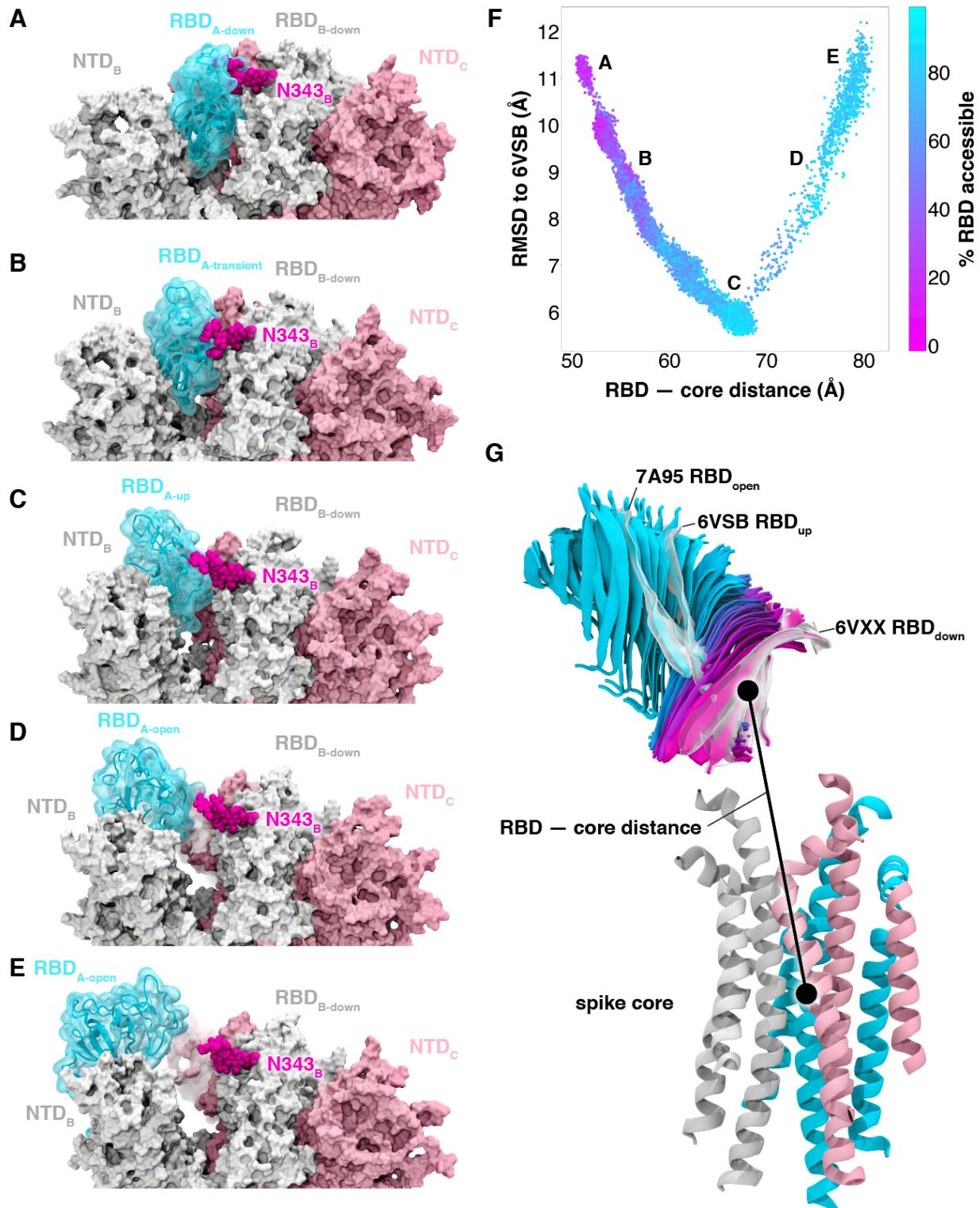
389 The authors declare no competing financial interests.

390



391

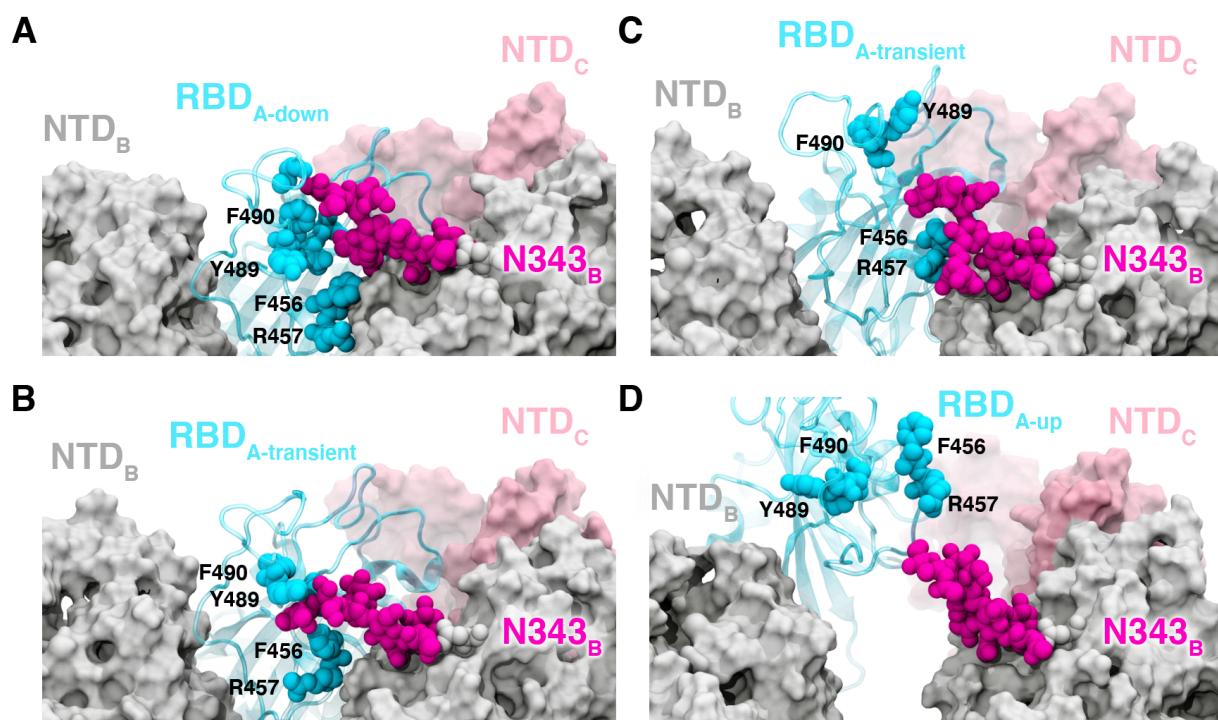
392 **Figure 1** Glycosylated spike RBD “down” and “open” conformations. (A) The SARS-CoV-2
393 spike head (gray) with glycans (dark blue) as simulated, with the stalk domain and membrane
394 (not simulated here, but shown in transparent for completeness). RBD shown in cyan, receptor
395 binding motif (RBM) in pink. Side view of the “down” (shielded, B) and “open” (exposed, C)
396 RBD. Top view of the closed (shielded, D) and “open” (exposed, E) RBM. Composite image of
397 glycans (dark blue lines) shows many overlapping snapshots of the glycans over the microsecond
398 simulations.



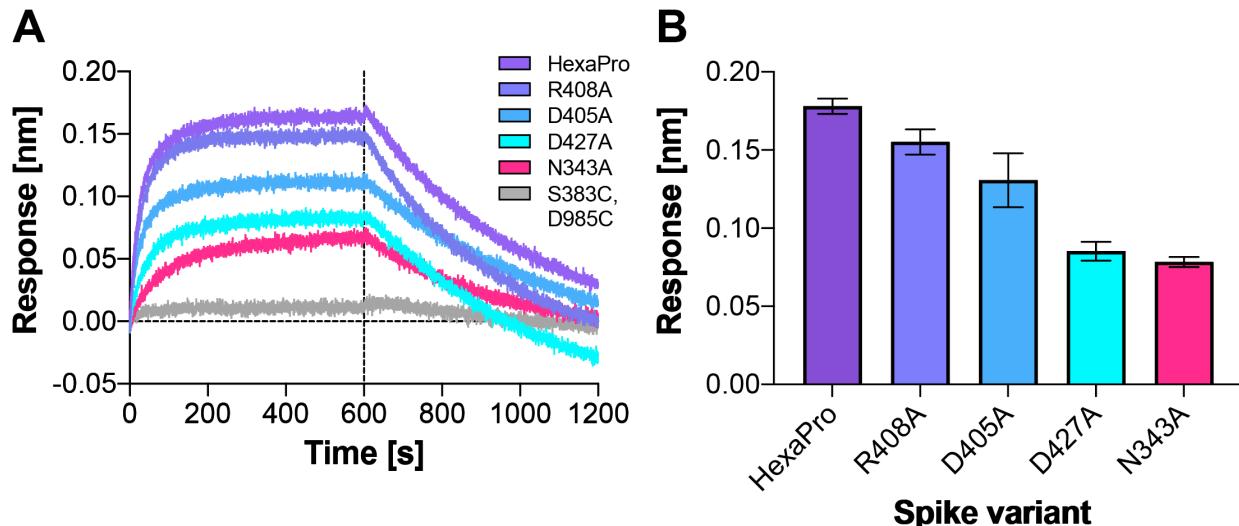
399

400 **Figure 2** Atomically detailed pathways of spike opening. (A-E) Snapshot configurations along
401 the opening pathway with chain A shown in cyan, chain B in gray, and chain C in pink and the

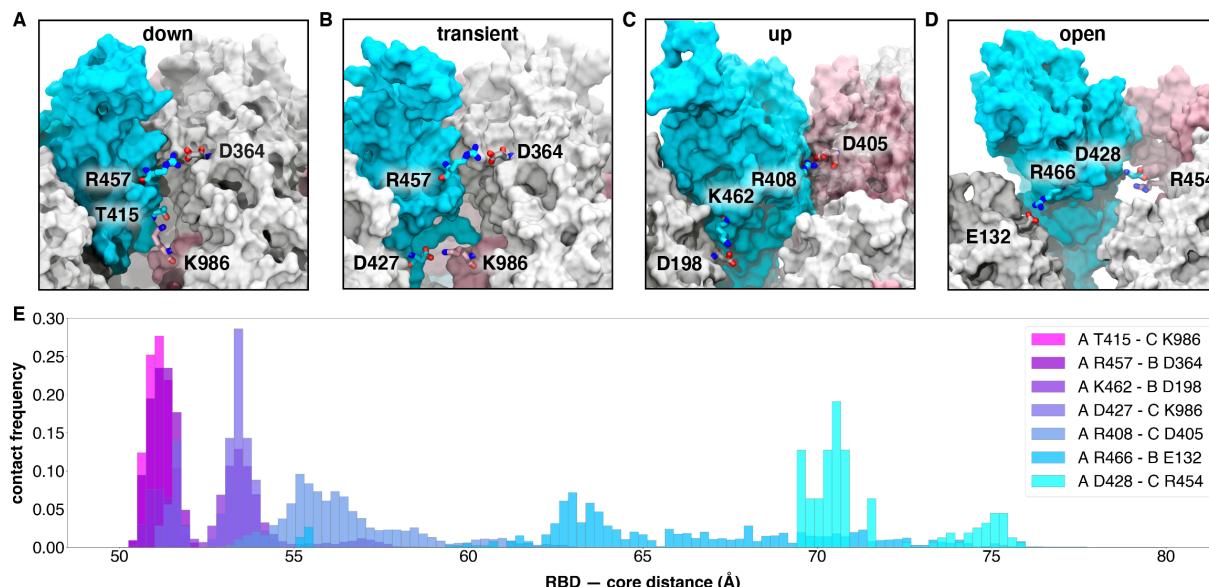
402 glycan at position N343 is shown in magenta. (F) Scatter plot of data from the 310 continuous
403 pathways with the C α RMSD of the RBD from the 6VSB “up” state plotted against the RBD —
404 core distance. Data points are colored based on % RBD solvent accessible surface area compared
405 to the RBD “down” state 6VXX. Location of snapshots shown in A-E are labeled. (G) Primary
406 regions of spike defined for tracking progress of the opening transition. The spike core is
407 composed of three central helices per trimer, colored according to chain as in (A-E). The RBD
408 contains a structured pair of antiparallel beta sheets and an overlay of snapshots from a
409 continuous WE simulation are shown colored along a spectrum resembling the palette in (F).
410 Overlayed cryoEM structures are highlighted and labeled including the initial RBD “down”
411 state, 6VXX, the target RBD “up” state and the ACE2 bound “open” state, 7A95.



412
413 **Figure 3** Glycan gating by N343. (A-D) Snapshot configurations along the opening pathway
414 with chain A shown in cyan, chain B in gray, chain C in pink, and the glycan at position N343 is
415 shown in magenta. (A) RBD A in the “down” conformation is shielded by the glycan at position
416 N343 of the adjacent RBD B. (B-D) The N343 glycan intercalates between and underneath the
417 residues F490, Y489, F456, F457 to push the RBD up and open (D).
418



419
420 **Figure 4** ACE2 binding is reduced by mutation of N343 glycosylation site and key salt bridge
421 residues. (A) Biolayer interferometry sensorgrams of HexaPro spike variants binding to ACE2.
422 For clarity, only the traces from the first replicate are shown. (B) Graph of binding response for
423 BLI data collected in triplicate with error bars representing the standard deviation from the mean.



424
425 **Figure 5** Salt bridges and hydrogen bonds along the opening pathway. (A-D) salt bridge or
426 hydrogen bond contacts made between RBD A, shown in blue, with RBD B, shown in gray, or
427 RBD C shown in pink within the “down”, transient, “up”, and “open” conformations,
428 respectively. (E) Histogram showing the frequency at which residues from (A-D) are within 3.5
429 Å of each other relative to RBD — core distance. Frequencies are normalized to 1.
430

431 **Data availability**

432 Data supporting the findings of this study are included in the Article and its Supplementary
433 Information files. We endorse the community principles around open sharing of COVID19
434 simulation data.²⁹ All simulation input files and data are available at the NSF MolSSI COVID19
435 Molecular Structure and Therapeutics Hub at <https://covid.molssi.org> and the Amaro Lab
436 website <http://amarolab.ucsd.edu>.

437

438 **Code availability**

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

442

443 **Supplementary information**

444 Sztain_Ahn_Supplementary_Information.pdf

445 Supplementary_Table2.xlsx

446 Supplementary_Video1.mp4

447 Supplementary_Video2.mp4

448 Supplementary_Video3.mp4

449 Supplementary_Video4.mp4

450 Supplementary_Video5.mp4

451 **Source data**

452 Source_data_fig2.zip

453 Source_data_fig4.zip

454 Source_data_fig5.zip

455