

Supplementary Information

Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2

Kapil Gupta^{1,2,†*}, Christine Toelzer^{1,2,†}, Maia Kavanagh Williamson^{3,†}, Deborah K. Shoemark^{1,2,†}, A. Sofia F. Oliveira^{1,4}, David A. Matthews³, Abdulaziz Almuqrin³, Oskar Staufer^{5,6,7,8}, Sathish K.N. Yadav^{1,2}, Ufuk Borucu^{1,2}, Frederic Garzoni⁹, Daniel Fitzgerald¹⁰, Joachim Spatz^{5,6,7,8}, Adrian J. Mulholland⁴, Andrew D. Davidson³, Christiane Schaffitzel^{1,2,10,*}, Imre Berger^{1,2,4,8,10,*}

¹ School of Biochemistry, University of Bristol, 1 Tankard's Close, Bristol BS8 1TD, UK

² Bristol Synthetic Biology Centre BrisSynBio, 24 Tyndall Ave, Bristol BS8 1TQ, UK.

³ School of Cellular and Molecular Medicine, University of Bristol, University Walk, Bristol, BS8 1TD, UK.

⁴ School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

⁵ Department for Cellular Biophysics, Max Planck Institute for Medical Research, Jahnstraße 29, 69120 Heidelberg, Germany.

⁶ Institute for Physical Chemistry, Department for Biophysical Chemistry, University of Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany.

⁷ Max Planck School Matter to Life, Jahnstraße 29, D-69120 Heidelberg, Germany.

⁸ Max Planck Bristol Centre for Minimal Biology, Cantock's Close, Bristol BS8 1TS, UK.

⁹ Imophoron Ltd, St. Philips Central, Albert Rd, St. Philips, Bristol, BS2 0XJ, UK.

¹⁰ Halo Therapeutics Ltd, St. Philips Central, Albert Rd, St. Philips, Bristol, BS2 0XJ, UK.

† These authors contributed equally to this study.

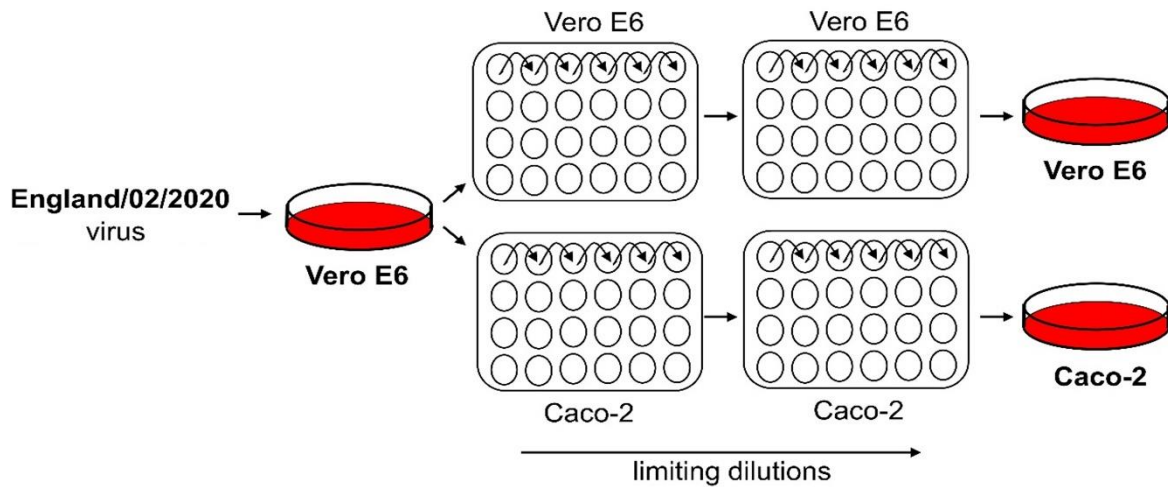
* Corresponding authors: Kapil Gupta, 0044 (0)117 394 1251 kapil.gupta@bristol.ac.uk

Christiane Schaffitzel, 0044 (0)117 394 1869 cb14941@bristol.ac.uk

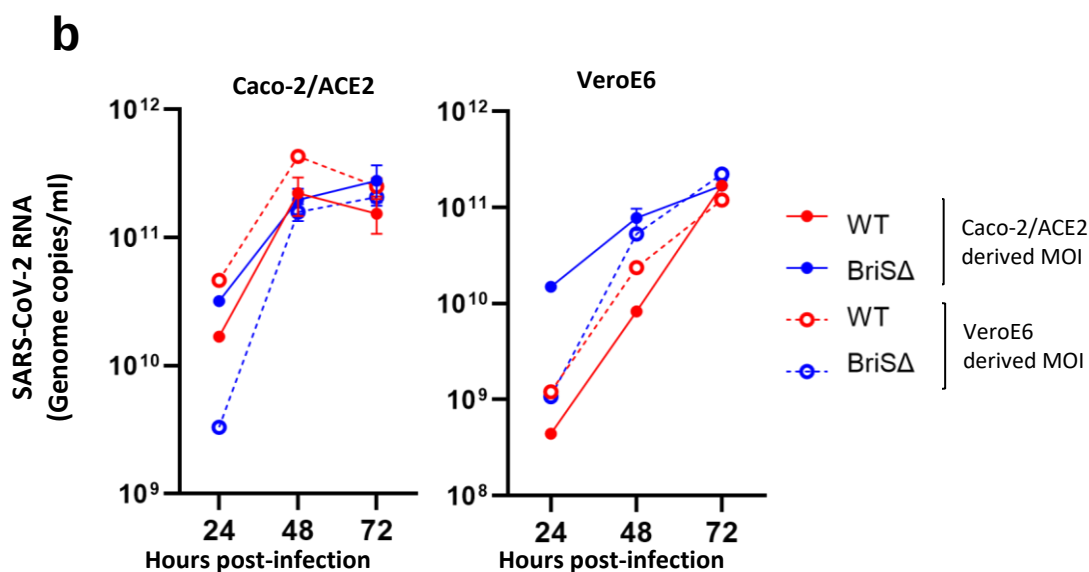
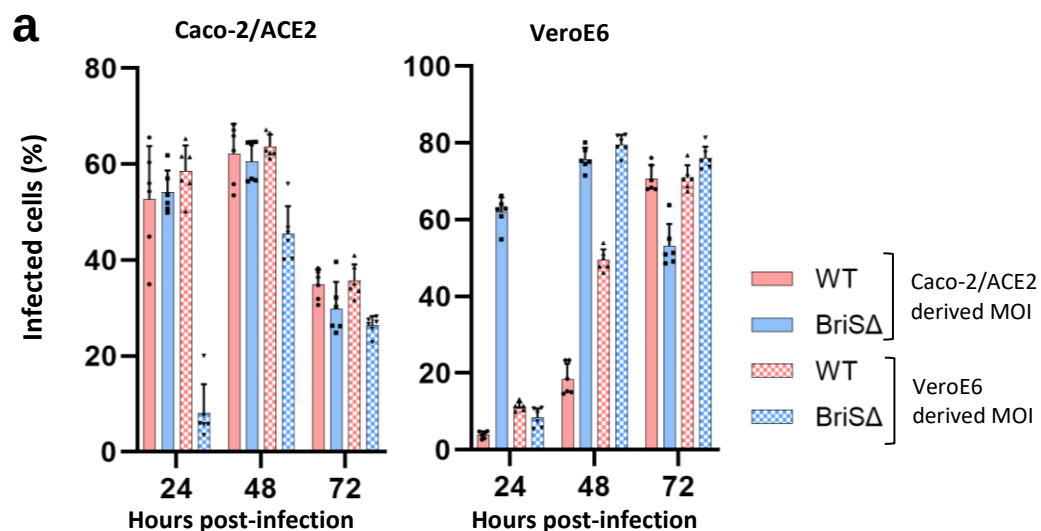
Imre Berger, 0044 (0)117 394 1857 imre.berger@bristol.ac.uk

26 **Brief description of what this file includes:**

- 27 Supplementary Fig. 1: Isolation of the wildtype and BriSA viruses.
- 28 Supplementary Fig. 2: SARS-CoV-2 WT and BriSA growth assays using MOIs determined on different
29 cell types result in differences in infectivity.
- 30 Supplementary Fig. 3: Human convalescent serum efficiently neutralizes WT and BriSA viruses.
- 31 Supplementary Fig. 4: Purification and quality control of the BriSA glycoprotein.
- 32 Supplementary Fig. 5: Cryo-EM image processing workflow.
- 33 Supplementary Fig. 6: Cryo-EM structure validation.
- 34 Supplementary Fig. 7: BriSA RBD structural organization.
- 35 Supplementary Fig. 8: Masked 3D classification.
- 36 Supplementary Fig. 9: Spike proteins binding to ACE2 by Surface Plasmon Resonance (SPR).
- 37 Supplementary Fig. 10: Nonequilibrium simulations of wildtype S and BriSA proteins.
- 38 Supplementary Fig. 11: Average C α -positional deviation in the five nanoseconds after removing LA
39 from wildtype S protein.
- 40 Supplementary Fig. 12: Average C α -positional deviation in the five nanoseconds after removing LA
41 from BriSA protein.
- 42 Supplementary Fig. 13: Distributions of the distances between R408-V395, F377-Y369 and I434-L368 in WT.
- 43 Supplementary Fig. 14: Distributions of the distances between R408-V395, F377-Y369 and I434-L368 in
44 BriSA.
- 45 Supplementary Fig. 15: Example of the FFA pocket in the beginning (t= 0 ns) and end (t=5 ns) of a
46 nonequilibrium simulation of BriSA.
- 47 Supplementary Fig. 16: Distribution of the distance between R634 and Y837 in BriSA.
- 48 Supplementary Fig. 17: Number of hydrogen bonds between the R1039 and the protein in BriSA.
- 49 Supplementary Fig. 18: Distribution of the R1039-F1042 and R1039-E1031 distances in BriSA.
- 50 Supplementary Table 1: Cryo-EM data collection and refinement statistics.
- 51 Supplementary Table 1: Cryo-EM data collection and refinement statistics.
- 52 Supplementary Table 2: N-linked glycosylation sites in recombinant SARS-CoV-2 S proteins.
- 53 Supplementary Table 3: Alignment of SARS-CoV-2 S protein sequences used in this study
- 54
- 55
- 56
- 57



Supplementary Fig. 1: Isolation of the wildtype and BriSΔ viruses. Schematic illustrating how the mixed population of SARS-CoV-2 viruses with wildtype (WT) and the “Bristol” variant BriSΔ were separated from each other by two rounds of limiting dilution in different cell types. After initial growth of the mixed virus population in Vero E6 cells, the virus titer was determined by TCID50 assay and serially diluted in 96 well plates until statistically just one infectious virus particle was present in the wells with the highest dilution. In the wells with evidence of viral growth at the highest dilutions, the virus in the supernatant was assayed by site-specific PCR to determine if the virus was WT or the BriSΔ variant. This process was repeated to isolate a pure preparation of either virus which were then grown up to provide sufficient virus to sequence by dRNA-seq.

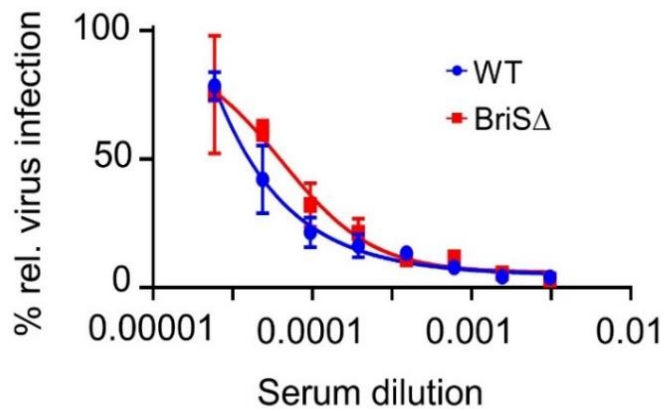


69

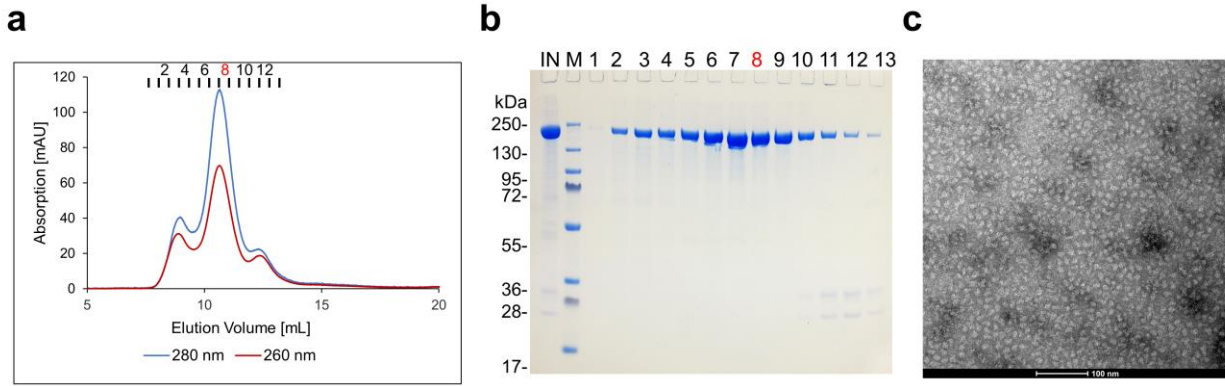
70

71 **Supplementary Fig. 2: SARS-CoV-2 WT and BriSΔ growth assays using MOIs determined on**
 72 **different cell types result in differences in infectivity.** Caco-2/ACE2 or Vero E6 cells were seeded in 96
 73 well plates and infected with two different volumes of the WT and the BriSΔ virus stocks, each representing
 74 a MOI of 0.5 calculated from infectious titres previously determined on either Caco-2/ACE2 (red) or Vero
 75 E6 (blue) cells. At the indicated time points post-infection, supernatants were collected from the infected
 76 cells, which were then fixed in 4% PFA. **a.** The % of virus infected cells was determined by staining the
 77 fixed cells using an antibody against the SARS-CoV-2 N protein and DAPI to detect nuclear DNA, before
 78 analysis using the ImageXpress Pico automated imaging system. The data points for n=6 biological

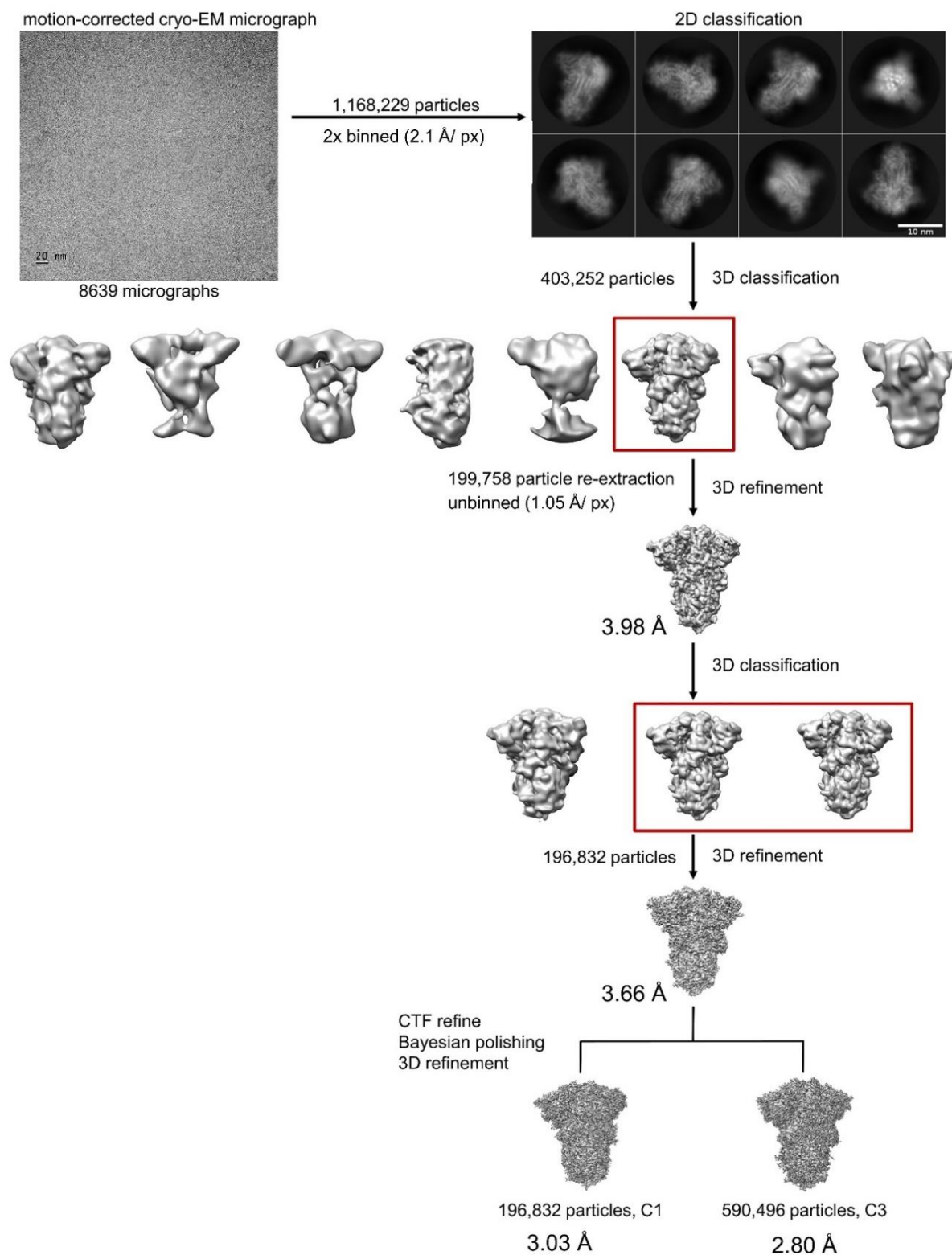
79 replicates are overlaid on box-plots with the top of the box plot showing the mean values and error bars
80 shown as + SD. **b.** Viral RNA was extracted from pooled culture supernatants from each replicate time point
81 and the amount of viral RNA quantitated by qRT-PCR and expressed as genome copies/ml using a calibrated
82 reference standard. qRT-PCR assays were done in technical triplicate. Data are presented as mean values
83 with error bars showing +/- SD.



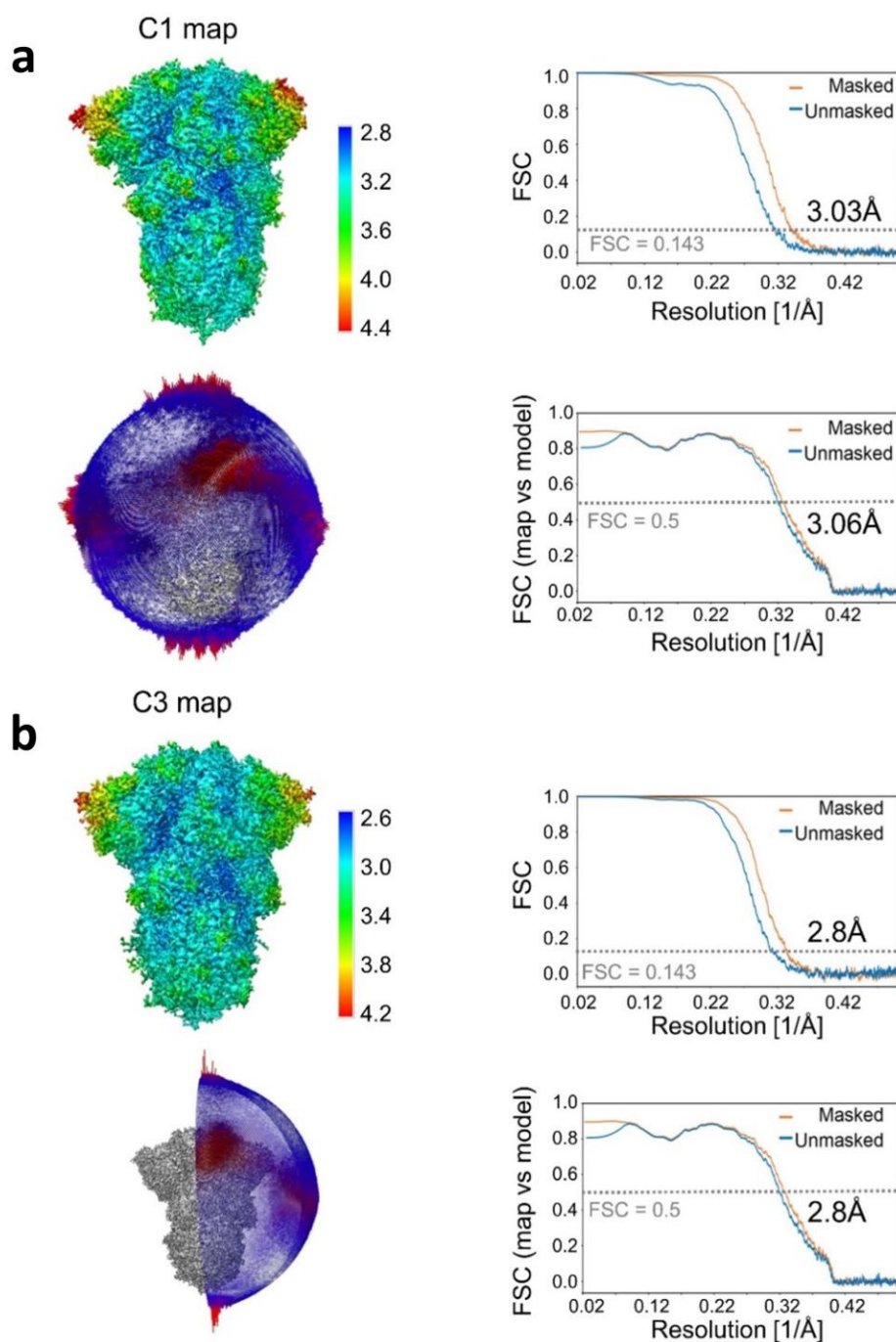
89
90
91 **Supplementary Fig. 3: Human convalescent serum efficiently neutralizes WT and BriSΔ viruses.**
92 Serum dilutions as indicated were utilized for virus neutralization using Vero E6 cells. WT: SARS-CoV-2
93 wildtype; BriSΔ : SARS-CoV-2 S deletion variant. Data are presented as mean values with error bars
94 showing +/- SD. N=3 biological replicates.



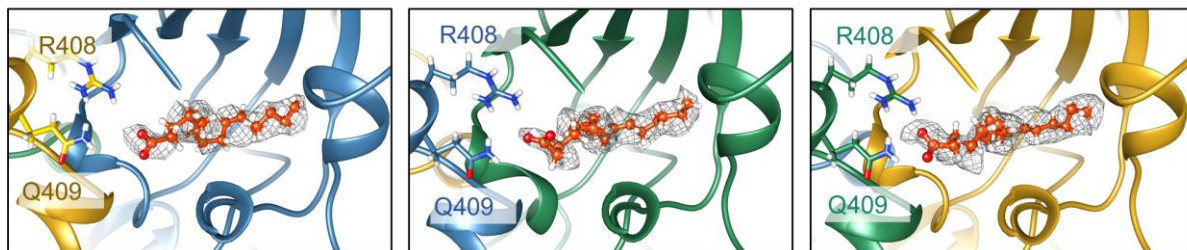
95
 96 **Supplementary Fig. 4: Purification and quality control of the BriSA glycoprotein.** **a** Size-exclusion
 97 chromatography (SEC) of affinity-purified BriSA spike protein using a Superdex 200 column. Absorption
 98 was detected at 280 nm (blue line) and 260 nm (red line). Peak fractions are indicated. **b** SDS PAGE analysis
 99 of the SEC fractions from panel A. lane 1: input fraction, lane 2: molecular weight marker, lanes 3-15:
 100 fractions 1 to 13 from SEC. **c** Negative stain EM micrograph of SEC fraction 8 (scale bar: 100 nm).
 101 Preparations were performed 4 times (n=4).



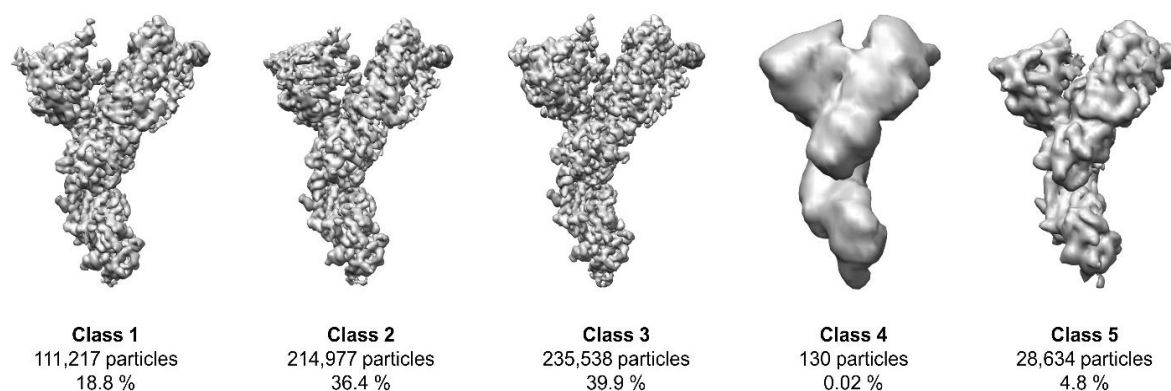
Supplementary Fig. 5: Cryo-EM image processing workflow. A motion-corrected cryo-EM micrograph is shown (scale bar 20 nm, particles circled in red), reference-free 2D class averages (scale bar 10 nm), 3D classifications and refinements resulting in a not symmetrized C1 and a C3-symmetrized cryo-EM map.



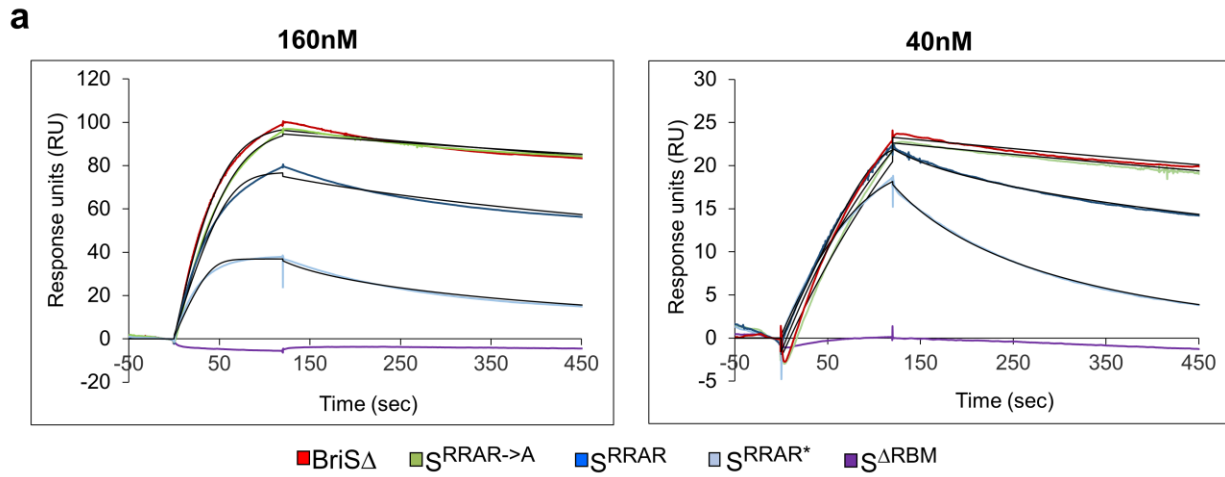
Supplementary Fig. 6: Cryo-EM structure validation. Top left: Cryo-EM reconstruction colored according to the local resolution from a side view. Top right: Fourier Shell Correlation (FSC) curve after gold standard refinement. Below left: Orientation distribution of views that contributed to this map. Longer red rods represent orientations that comprise more particles. Below right: Cross-validation FSC curves for the refined model versus the final masked and unmasked maps, shown for **a.** the unsymmetrized C1 map and **b.** the C3-symmetrized map.



Supplementary Fig. 7: BriSA RBD structural organization. LA binding in each bipartite free fatty acid binding pocket in the unsymmetrized C1 structure is shown. EM density is shown as grey mesh.



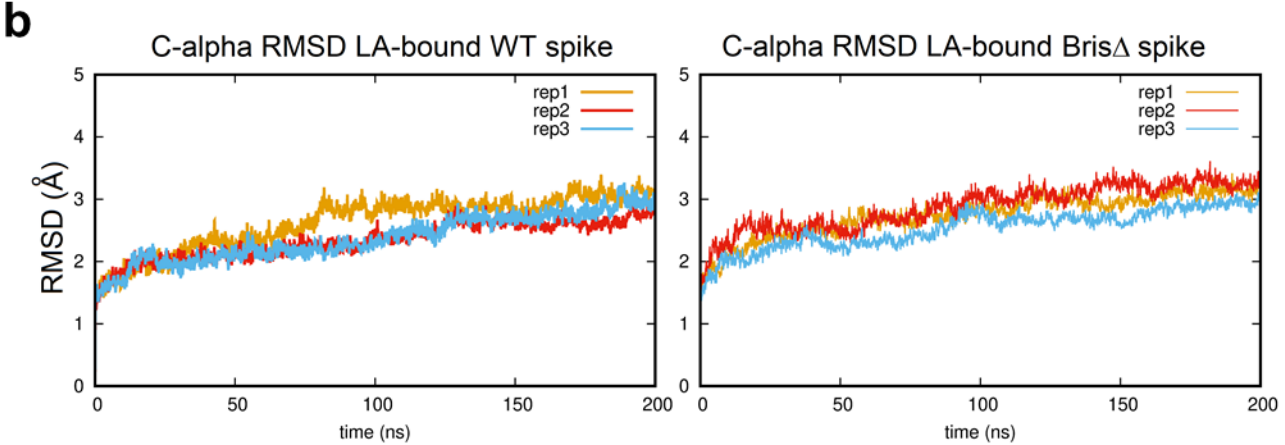
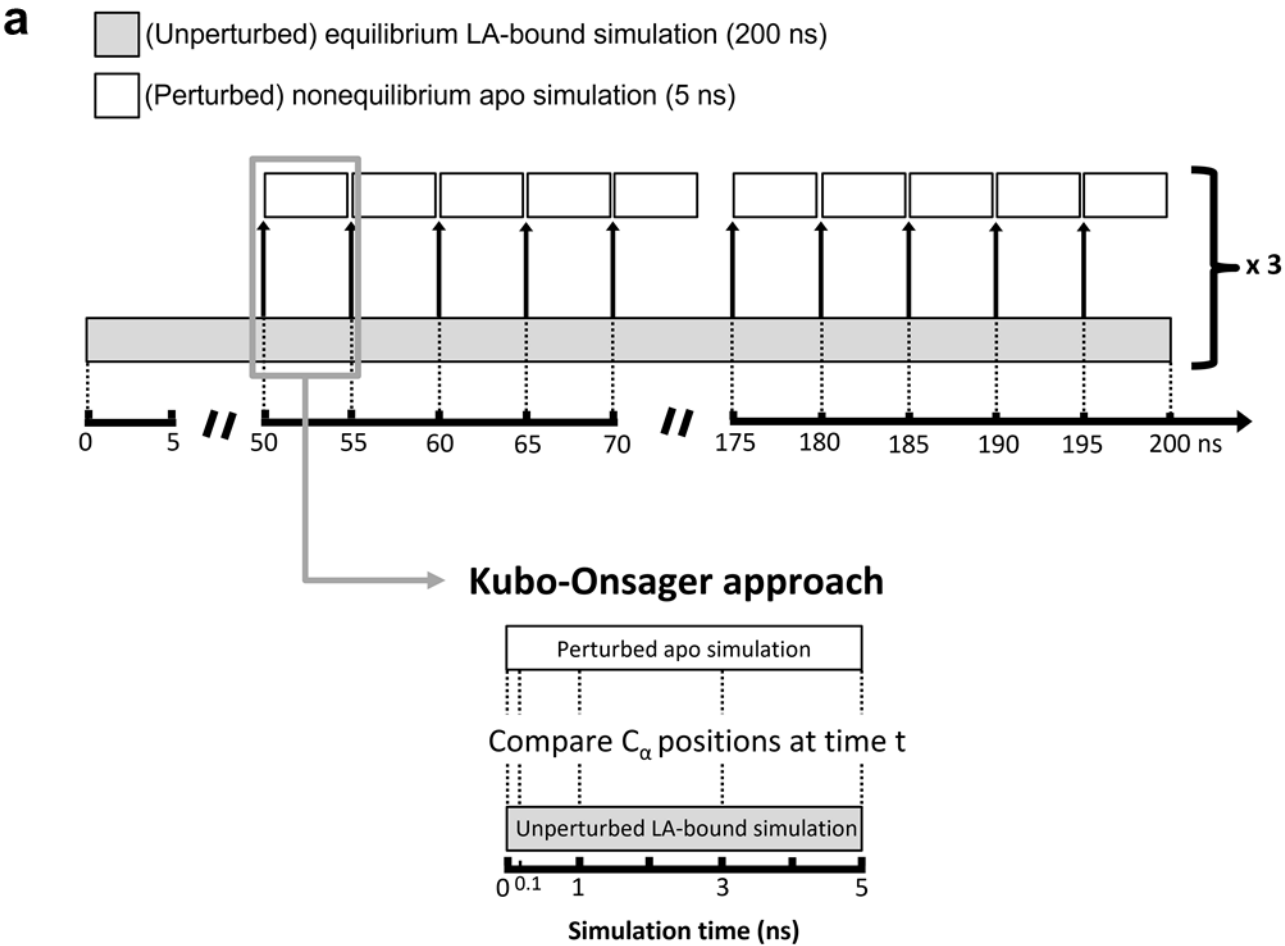
Supplementary Fig. 8: Masked 3D classification. Masked 3D classification focusing of individual chains within the S trimer into 5 classes. Class 1-3 comprised 95% of all particles. These classes all present LA-bound RBDs. The other 3D classes did not reach sufficient resolution to determine the presence or absence of LA in the RBD reconstructions.



b

	K_D (nM)	k_{on} M ⁻¹ s ⁻¹	k_{off} = s ⁻¹
BriSΔ	1.8 ± 0.34	2.22 ± 0.49 × 10 ⁵	3.9 ± 0.4 × 10 ⁻⁴
S ^{RRAR->A}	1.4 ± 0.3	2.5 ± 1.2 × 10 ⁵	3.6 ± 2.2 × 10 ⁻⁴
S ^{RRAR}	2.51 ± 0.06	4.55 ± 0.04 × 10 ⁵	11.4 ± 0.28 × 10 ⁻⁴
S ^{RRAR*}	12.2 ± 4.36	6.33 ± 2.18 × 10 ⁵	70.69 ± 5.94 × 10 ⁻⁴

Supplementary Fig. 9: Spike proteins binding to ACE2 by Surface Plasmon Resonance (SPR). **a.** Spike proteins BriSΔ, S^{RRAR->A}, S^{RRAR} and furin-cleaved S^{RRAR*} were analyzed for binding to ACE2 receptor immobilized on a streptavidin-coated sensor chip. Sensorgrams for representative concentrations (160 nM and 40 nM) are shown including S^{ΔRBM} lacking the receptor-binding motif as control. K_D values, k_{on} and k_{off} are listed in **b**. Values for S^{RRAR->A} are from ¹.

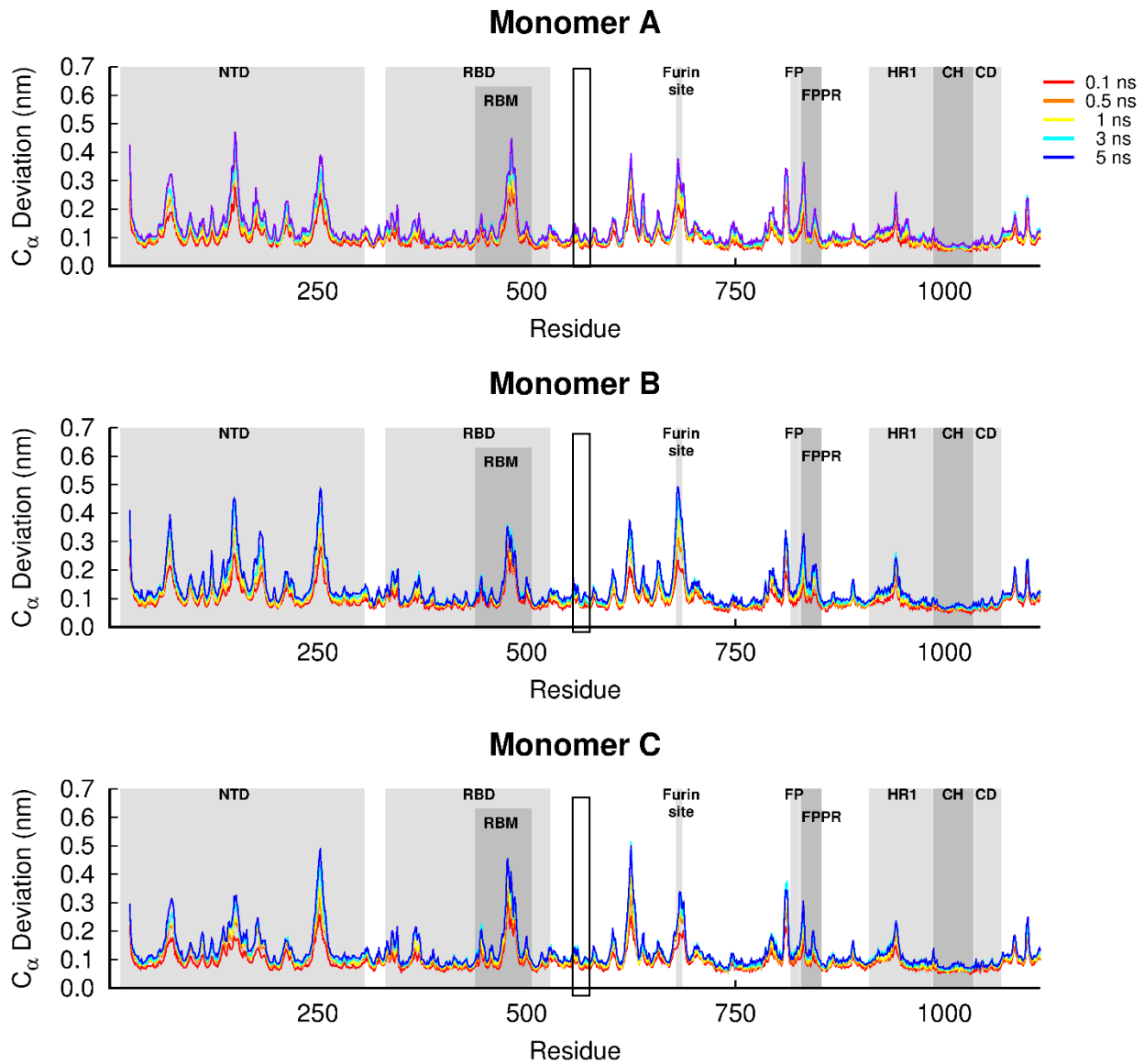


136

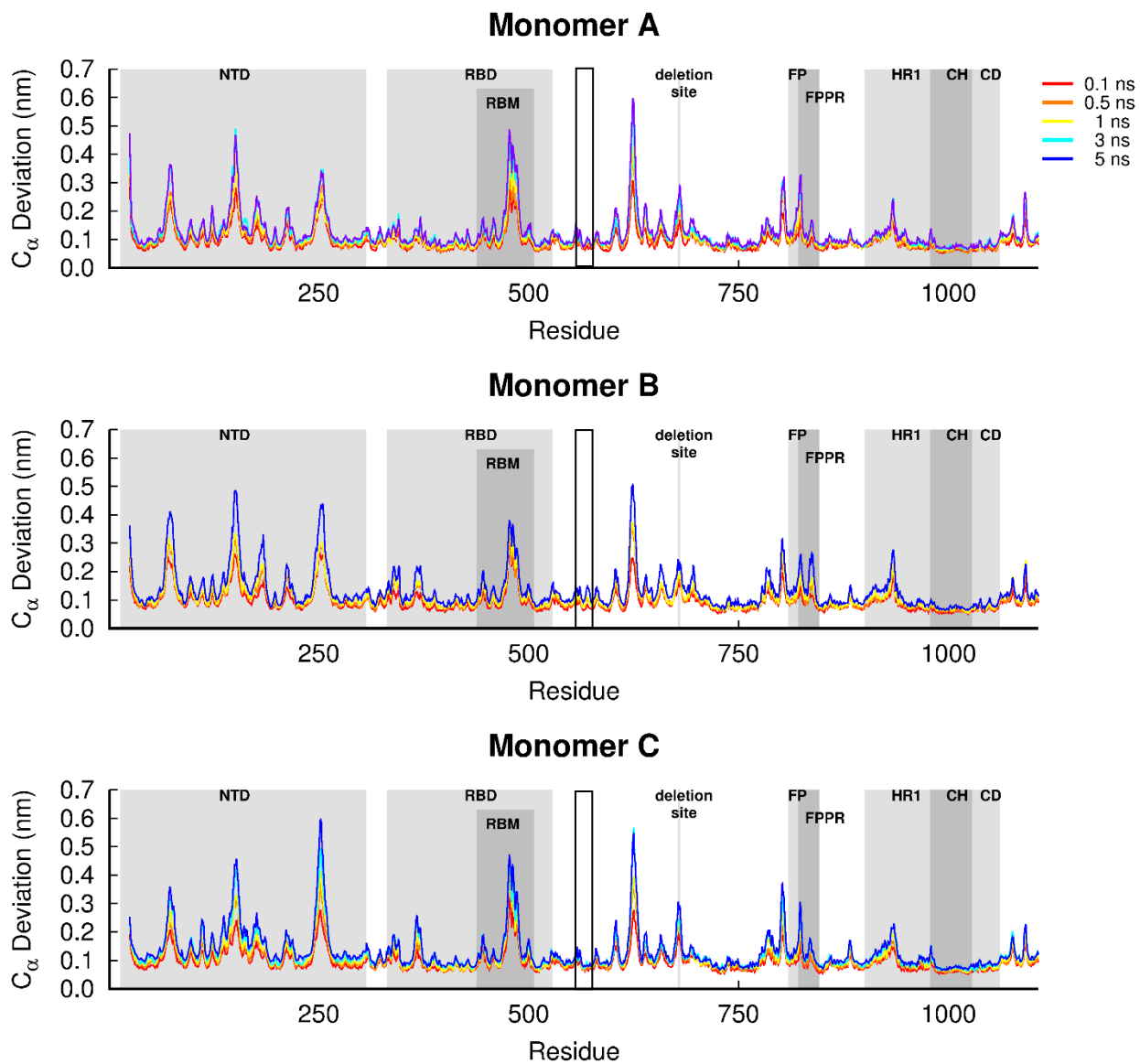
137 **Supplementary Fig. 10: Nonequilibrium simulations of wildtype S and BriSΔ proteins. a.** A schematic
138 description of the procedure used to set up and analyze the nonequilibrium MD simulations. From the cryo-
139 EM structure of the S proteins with LA bound, three equilibrium MD simulations, 200 ns each, were
140 performed. These equilibrium simulations (indicated by rectangles filled in grey) were used to generate
141 starting conformations for the nonequilibrium simulations (rectangles filled in white). **b.** shows the RMSDs

142 of the equilibration runs for WT and BriSΔ used as the start point for the nonequilibrium MD runs. The
143 systems were considered sufficiently equilibrated after 50 ns and conformations were extracted every five
144 nanoseconds beyond 50 ns from each LA-bound simulation (from 50-200 ns), and LA was removed. Each
145 nonequilibrium simulation was run for five nanoseconds. The Kubo-Onsager approach ²⁻⁵ was used to
146 extract the response of the system to LA removal (bottom panel). For each pair of unperturbed LA-bound
147 and perturbed apo simulations, the positional deviations of each Cα at equivalent times (namely 0, 0.1, 1, 3
148 and 5 ns) were determined and averaged over all 90 simulations.

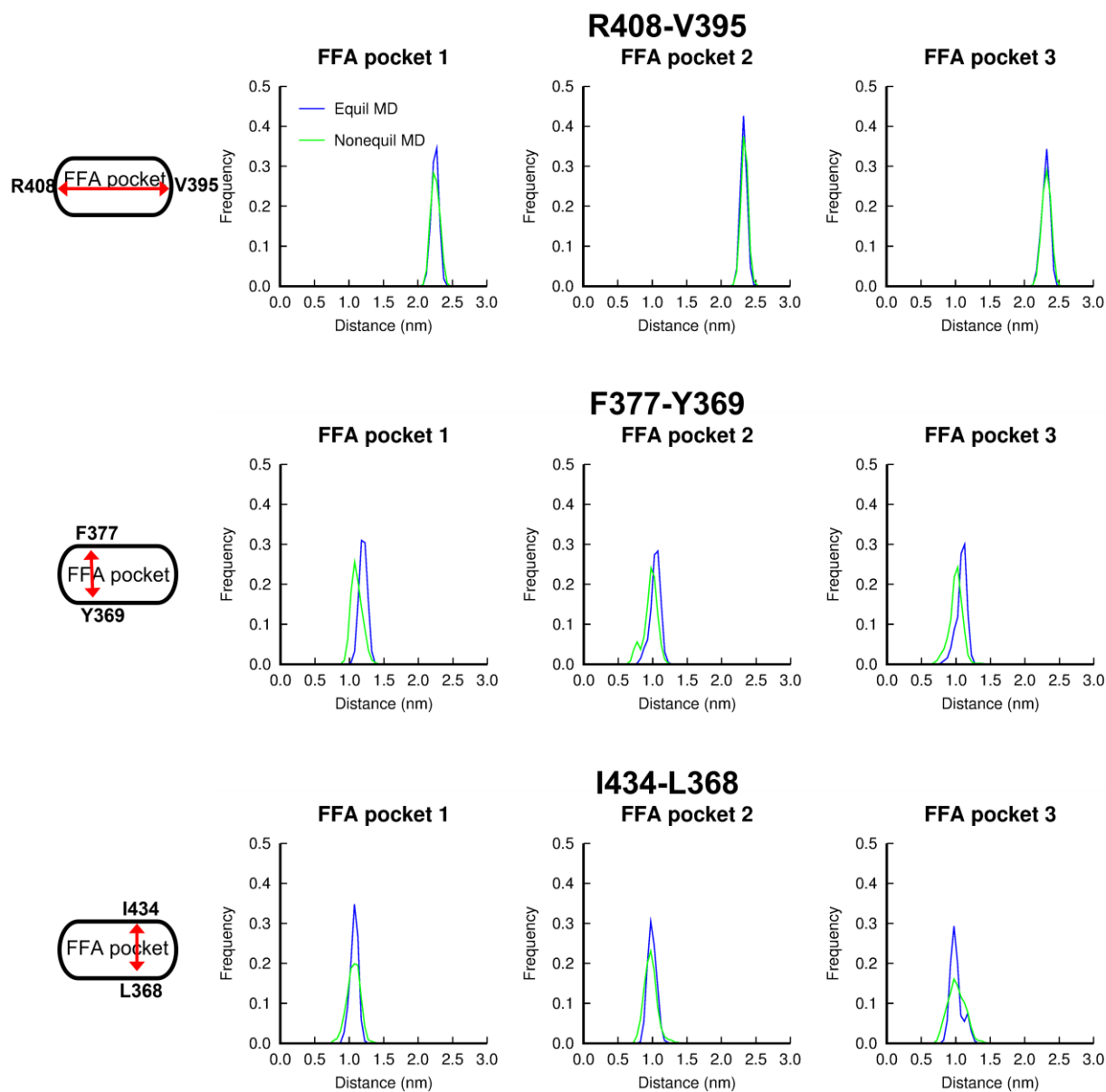
149



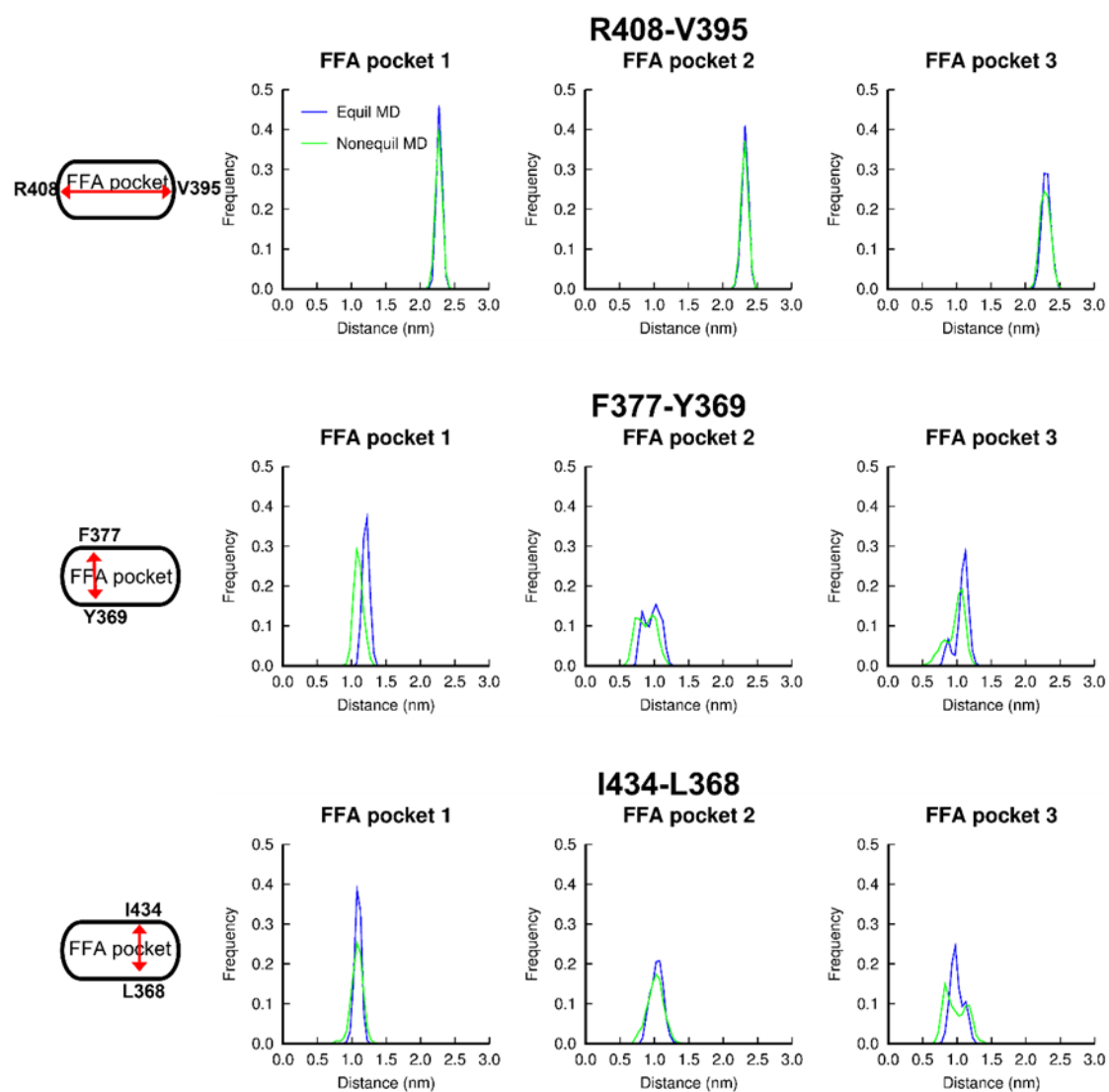
Supplementary Fig. 11: Average C_{α} -positional deviation in the five nanoseconds after removing LA from wildtype S protein. The average deviations were calculated using the Kubo-Onsager approach for the pairwise comparison between the nonequilibrium apo and equilibrium LA-bound simulations. The positions of important structural motifs are highlighted in grey, namely N-terminal domain (NTD), receptor-binding domain (RBD), receptor-binding motif (RBM), fusion peptide (FP), fusion peptide proximal region (FPPR), heptad repeat 1 (HR1), central helix (CH) and connector domain (CD). The V622-L629 region adjacent to R634 is boxed in black.



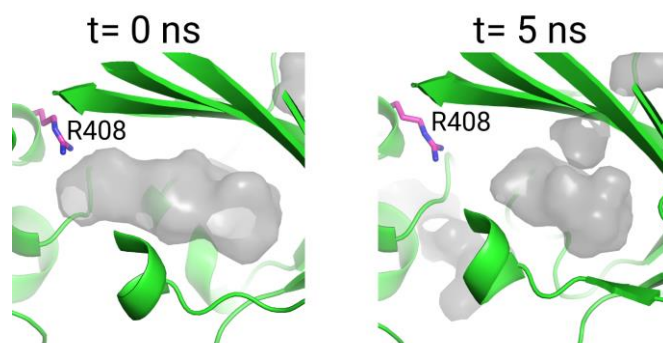
Supplementary Fig. 12: Average C_{α} -positional deviation in the five nanoseconds after removing LA from BriSA protein. The average deviations were calculated using the Kubo-Onsager approach for the pairwise comparison between the nonequilibrium apo and equilibrium LA-bound simulations. The positions of relevant structural motifs are highlighted in grey, namely the N-terminal domain (NTD), receptor-binding domain (RBD), receptor-binding motif (RBM), fusion peptide (FP), fusion peptide proximal region (FPPR), heptad repeat 1 (HR1), central helix (CH), connector domain (CD). The V622-L629 region adjacent to R634 is boxed in black.



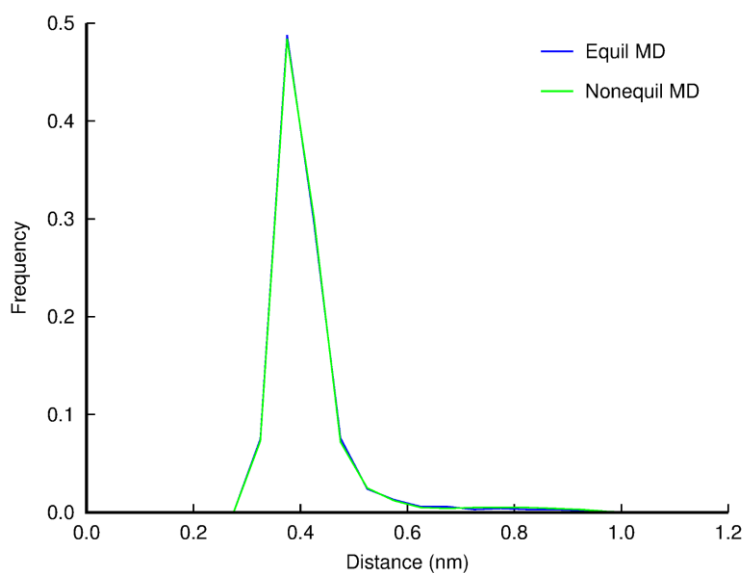
Supplementary Fig. 13: Distributions of the distances between R408-V395, F377-Y369 and I434-L368 in WT. Overall distribution of the distance between the center of mass of R408 and V395, F377 and Y369 and I434 and L368 in the equilibrium and nonequilibrium WT simulations. The histograms contain the distances for all three FFA pockets.



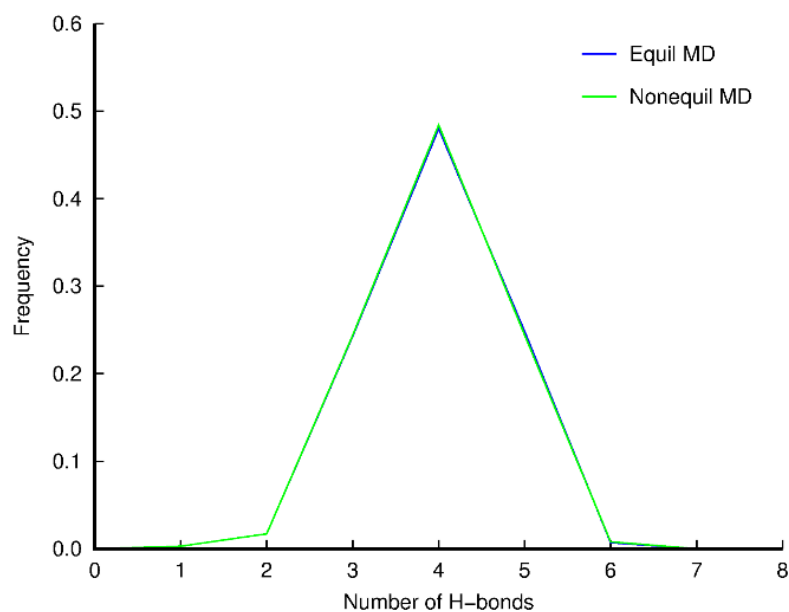
Supplementary Fig. 14: Distributions of the distances between R408-V395, F377-Y369 and I434-L368 in BriSA. Overall distribution of the distance between the center of mass of R408 and V395, F377 and Y369 and I434 and L368 in the equilibrium and nonequilibrium BriSA simulations. The histograms contain the distances for all three FFA pockets.



Supplementary Fig. 15: Example of the FFA pocket in the beginning ($t=0$ ns) and end ($t=5$ ns) of a nonequilibrium simulation of BriSA. Note the clear volume reduction of the pocket after 5 ns.

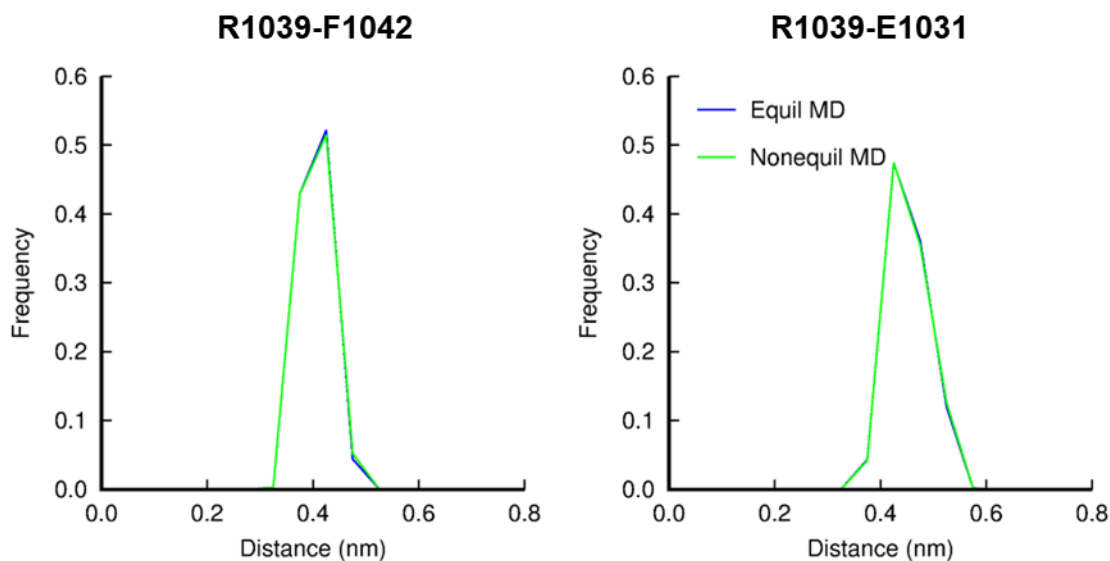


Supplementary Fig. 16: Distribution of the distance between R634 and Y837 in BriSA. Overall distribution of the distance between the positively charged sidechain of R634 and the aromatic sidechain of Y837 in the equilibrium and nonequilibrium MD simulations of BriSA. The histograms reflect the distances over the three chains.



Supplementary Fig. 17: Number of hydrogen bonds between the R1039 and the protein in BriSA.

Overall distribution of the number of hydrogen bonds formed by R1039 in the equilibrium and nonequilibrium MD simulations of BriSA. The histograms reflect the distances over the three chains.



Supplementary Fig. 18: Distribution of the R1039-F1042 and R1039-E1031 distances in BriSA.

Overall distribution of the distance between the sidechains of R1039, F1042 and E1031 in the equilibrium and nonequilibrium MD simulations of BriSA. The histograms reflect the distances over the three chains.

Supplementary Table 1: Cryo-EM data collection and refinement statistics.

	C1	C3
Data collection and processing		
Voltage (kV)	200	200
Magnification (nominal)	130,000	130,000
Pixel size (Å / px)	1.05 (0.525)	1.05 (0.525)
Flux (e ⁻ /pix/sec)	6.4	6.4
Frames per exposure	55	55
Exposure (e ⁻ / Å ²)	63.8	63.8
Defocus range (µm)	-0.8 to -2.0	-0.8 to -2.0
Micrographs collected	9,519	9,519
Particles final	196,832	590,496
Map sharpening B-factor (Å ²)	-97.6	-106.7
Masked resolution at 0.143 FSC (Å)	3.03	2.8
Refinement		
Composition		
Amino acids	3066	3036
Glycans	30	33
Ligands	3	3
RMSD bonds (Å)	0.003	0.004
RMSD angles (°)	0.564	0.555
Mean B-factors (Å ²)		
Amino acids	18.75	29.26
Ligands	33.61	41.97
Ramachandran		
Favored (%)	94.06	95.22
Allowed (%)	5.84	4.78
Outliers (%)	0.1	0.00
Rotamer outliers (%)	0.64	0.23
Clash score	2.77	2.49
C-beta outliers (%)	0.00	0.00
CaBLAM outliers (%)	3.14	2.51
CC (mask)	0.82	0.80
MolProbity score	1.46	1.36
EMRinger score	3.77	3.95
Model resolution (Å), 0.5 FSC threshold	3.0	2.8

201 **Supplementary Table 2: N-linked glycosylation sites in recombinant SARS-CoV-2 S proteins.**

WT SARS-CoV-2 S*	Recombinant SARS-CoV-2 S expressed in		
	Freestyle 293F** ⁶	Hi5** ¹	Hi5** (BriSA, this study)
N ₁₇ LT		N ₁₇ LT	
N ₆₁ VT	N ₆₁ VT	N ₆₁ VT	N ₆₁ VT
N ₇₄ GT			
N ₁₂₁ NA [#]			
N ₁₂₂ AT	N ₁₂₂ AT	N ₁₂₂ AT	N ₁₂₂ AT
N ₁₄₉ KS			
N ₁₆₅ CT	N ₁₆₅ CT	N ₁₆₅ CT	N ₁₆₅ CT
N ₂₃₄ IT	N ₂₃₄ IT	N ₂₃₄ IT	N ₂₃₄ IT
N ₂₈₂ GT	N ₂₈₂ GT	N ₂₈₂ GT	N ₂₈₂ GT
N ₃₃₁ IT	N ₃₃₁ IT	N ₃₃₁ IT	N ₃₃₁ IT
N ₃₄₃ AT	N ₃₄₃ AT	N ₃₄₃ AT	N ₃₄₃ AT
N ₃₇₀ SA [#]			
N ₆₀₃ TS	N ₆₀₃ TS		
N ₆₁₆ CT	N ₆₁₆ CT	N ₆₁₆ CT	N ₆₁₆ CT
N ₆₅₇ NS	N ₆₅₇ NS		N ₆₅₇ NS
N ₇₀₉ NS	N ₇₀₉ NS	N ₇₀₆ NS	N ₇₀₉ NS
N ₇₁₇ FT	N ₇₁₇ FT	N ₇₁₄ FT	N ₇₁₇ FT
N ₈₀₁ FS	N ₈₀₁ FS	N ₇₉₈ FS	N ₈₀₁ FS
N ₁₀₇₄ FT	N ₁₀₇₄ FT	N ₁₀₇₁ FT	N ₁₀₇₄ FT
N ₁₀₉₈ GT	N ₁₀₉₈ GT	N ₁₀₉₅ GT	N ₁₀₉₈ GT
N ₁₁₃₄ NT	N ₁₁₃₄ NT	N ₁₁₃₁ NT	N ₁₁₃₄ NT
N ₁₁₅₈ HT			
N ₁₁₇₃ AS			
N ₁₁₉₄ ES			

* YP_009724390.1
 **Sites lacking glycosylation in cryo-EM maps are omitted (boxes colored in grey).
 # Predicted based on SARS-CoV-2 S protein ⁶

	1	10	20	30	40	50	60
S WT	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
BriSΔ	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
S RRAR->A	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
S RRAR	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
S RRAR*	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
S ΔRMB	MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFS						
	70	80	90	100	110	120	
S WT	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
BriSΔ	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
S RRAR->A	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
S RRAR	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
S RRAR*	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
S ΔRMB	NVTWFHAIHVSGTNGTKRFDNPVLPFNDGVYFASTSEKSNIIRGWIFGTTLDSKTQSLIIV						
	130	140	150	160	170	180	
S WT	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
BriSΔ	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
S RRAR->A	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
S RRAR	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
S RRAR*	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
S ΔRMB	NNATNVVIKVCEFQFCNDPFLGVYHKNKSWMESEFRVYSSANNCTFEYVSQPFLMDLE						
	190	200	210	220	230	240	
S WT	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
BriSΔ	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
S RRAR->A	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
S RRAR	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
S RRAR*	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
S ΔRMB	GKQGNFKNLREFVFNIDGYFKIYSKHTPINLVRDLPQGFSALEPLVDLPIGINITRFQT						
	250	260	270	280	290	300	
S WT	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
BriSΔ	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
S RRAR->A	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
S RRAR	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
S RRAR*	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
S ΔRMB	LLALHRSYLTTPGDSSSGWTAGAAAYVGYLQPRTFLLKYNENGTITDAVDCALDPLSETK						
	310	320	330	340	350	360	
S WT	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
BriSΔ	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
S RRAR->A	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
S RRAR	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
S RRAR*	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
S ΔRMB	CTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITNLCPPFGEVFNATRFASVYAWNRRKRISN						
	370	380	390	400	410	420	
S WT	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
BriSΔ	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
S RRAR->A	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
S RRAR	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
S RRAR*	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
S ΔRMB	CVADYSVLYNSASFSTFKCYGVSPTKLNDLCFTNVYADSFVIRGDEVQRQIAPGQTGKIAD						
	430	440	450	460	470	480	
S WT	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						
BriSΔ	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						
S RRAR->A	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						
S RRAR	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						
S RRAR*	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						
S ΔRMB	YNYKLPDDFTGCVIAWNSNNLDSKVGGN						

204
205

	490	500	510	520	530	540
S WT	NGVE	GFNCYF	PLQSYGF	QPTN	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
BriSΔ	NGVE	GFNCYF	PLQSYGF	QPTN	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
S RRAR->A	NGVE	GFNCYF	PLQSYGF	QPTN	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
S RRAR	NGVE	GFNCYF	PLQSYGF	QPTN	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
S RRAR*	NGVE	GFNCYF	PLQSYGF	QPTN	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN
S ΔRMB	..GS	GGSGGS	PLQSYGF	GGG	GVGYQ	PYRVVVLSFELLHAPATVCGPKKSTNLVKNKCVN

	550	560	570	580	590	600
S WT	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		
BriSΔ	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		
S RRAR->A	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		
S RRAR	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		
S RRAR*	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		
S ΔRMB	FNFNGLTGTGVL	TESNKKFLPFQ	QFGRDIADTTDAVRDPQTLEILDITPCSF	GGVSVITP		

	610	620	630	640	650	660
S WT	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			
BriSΔ	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			
S RRAR->A	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			
S RRAR	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			
S RRAR*	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			
S ΔRMB	GTNTSNQVAVLYQ	DVNCTEVPVAIHADQLTPTWRVYSTGSNVFQTRAGCLIGAEHVNN	SY			

	670	680	690	700	710	720
S WT	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI
BriSΔ	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI
S RRAR->A	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI
S RRAR	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI
S RRAR*	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI
S ΔRMB	ECDIPIGAGICASYQT	TNTSP	RRAR	SVASQSI	IAYTMSLGAENSVAYSNN	SIAIPTNFTI

	730	740	750	760	770	780
S WT	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			
BriSΔ	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			
S RRAR->A	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			
S RRAR	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			
S RRAR*	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			
S ΔRMB	SVTTEILPVSMKT	TSVDCTMYICGDSTEC	SNLLQYGSFCTQLNRALTGIAVEQDKNTQE			

	790	800	810	820	830	840
S WT	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		
BriSΔ	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		
S RRAR->A	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		
S RRAR	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		
S RRAR*	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		
S ΔRMB	VFAQVKQIYKTPPIK	DFFGGFNFSQILPDPSKPSKRSFIEDLLFNKVT	LADAGFIKQY	GDC		

	850	860	870	880	890	900
S WT	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			
BriSΔ	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			
S RRAR->A	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			
S RRAR	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			
S RRAR*	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			
S ΔRMB	LGDI AARDLICAQK	FNGLTVLPPLL	TDEMIAQYTSALLAGTITSGWTFGAGAALQIPFAM			

	910	920	930	940	950	960
S WT	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	
BriSΔ	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	
S RRAR->A	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	
S RRAR	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	
S RRAR*	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	
S ΔRMB	QMAYRFNGIGV	TQNVLYENQKLI	ANQFN	SAIGKIQDSL	SSSTASALGKLQDVVNQNAQALN	

* Furin cleavage in S^{RRAR*} is marked with a triangle filled in black

	970	980	990	1000	1010	1020
S WT	TLVKQLSSNF	GAISSVLNDILSR	LDKV	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		
BriSΔ	TLVKQLSSNF	GAISSVLNDILSR	LDKV	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		
S RRAR->A	TLVKQLSSNF	GAISSVLNDILSR	LDKV	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		
S RRAR	TLVKQLSSNF	GAISSVLNDILSR	LDPP	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		
S RRAR*	TLVKQLSSNF	GAISSVLNDILSR	LDPP	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		
S ΔRMB	TLVKQLSSNF	GAISSVLNDILSR	LDPP	EAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRA		

	1030	1040	1050	1060	1070	1080
S WT	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					
BriSΔ	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					
S RRAR->A	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					
S RRAR	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					
S RRAR*	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					
S ΔRMB	SANLAATKMSECVLGQSKRVDFCGKGYHLMSPQSAPHGVVFLHVTYVPAQEKNFETTAPA					

	1090	1100	1110	1120	1130	1140
S WT	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					
BriSΔ	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					
S RRAR->A	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					
S RRAR	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					
S RRAR*	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					
S ΔRMB	ICHDGKAHFPREGVFVSNNGTHWFVTQRNFYEPQIIITDNTFVSGNCDVVIGIVNNTVYDP					

	1150	1160	1170	1180	1190	1200
S WT	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					
BriSΔ	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					
S RRAR->A	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					
S RRAR	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					
S RRAR*	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					
S ΔRMB	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDL					

	1210	1220	1230	1240
S WT	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			
BriSΔ	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			
S RRAR->A	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			
S RRAR	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			
S RRAR*	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			
S ΔRMB	QELGKYEQYIKWPWYIWLGFIA...GLIAIVMV.....TIMLCCMTSCCSCCLKG			

	1250	1260	1270
S WT	CCSCGSCCKFDEDDSEPVLKGVKLHHT...		
BriSΔ	GGGSGSEQKLISEEDLGSGSGSHHHHHHHH		
S RRAR->A	GGGSGSEQKLISEEDLGSGSGSHHHHHHHH		
S RRAR	GGGSGSEQKLISEEDLGSGSGSHHHHHHHH		
S RRAR*	GGGSGSEQKLISEEDLGSGSGSHHHHHHHH		
S ΔRMB	GGGSGSEQKLISEEDLGSGSGSHHHHHHHH		

Supplementary Movie 1: An example of a targeted dynamics simulation. In this movie one of the 50 (10 ns) replicates shows the opening of the RBD of the WT apo SARS-CoV-2 S protein trimer when 0.2 kJ/mol/nm force constant was applied to atoms of the RBD to pull it from a starting, closed, equilibrated (following 200 ns simulation) conformation to the raised RBD position in EMD-1114.

217 **References**

218 1. Toelzer, C. et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein.
219 *Science* **370**, 725-730 (2020).

220 2. Ciccotti, G., Jacucci, G. & McDonald, I.R. Thought-experiments by molecular dynamics. *J Stat Phys*
221 **21**, 1-12 (1979).

222 3. Ciccotti, G. Computer simulation in material science. *P. V. Meyer M, Ed. (Kluwer Academic*
223 *Publishers)*, 119–137 (1991).

224 4. Ciccotti, G. & Ferrario, M. Non-equilibrium by molecular dynamics: a dynamical approach. *Mol Simul*
225 **42**, 1385-1400 (2016).

226 5. Oliveira, A.S.F., Ciccotti, G., Haider, S. & Mulholland, A.J. Dynamical nonequilibrium molecular
227 dynamics reveals the structural basis for allostery and signal propagation in biomolecular systems. *Eur.*
228 *Phys. J. B* **94**, 144 (2021).

229 6. Walls, A.C. et al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell*
230 **181**, 281-292 e6 (2020).