Structure, Volume *30*

Supplemental Information

Uncovering cryptic pockets in the SARS-CoV-2 spike glycoprotein

Lorena Zuzic, Firdaus Samsudin, Aishwary T. Shivgan, Palur V. Raghuvamsi, Jan K. Marzinek, Alister Boags, Conrado Pedebos, Nikhil K. Tulsian, Jim Warwicker, Paul MacAry, Max Crispin, Syma Khalid, Ganesh S. Anand, and Peter J. Bond

Uncovering cryptic pockets in the SARS-CoV-2 spike glycoprotein – Supplemental Information

Lorena Zuzic,[1,2,†] Firdaus Samsudin,[1,†] Aishwary T. Shivgan,[1] Palur V. Raghuvamsi,[3] Jan K. Marzinek,[1] Alister Boags,[1,4] Conrado Pedebos,[4,5] Nikhil K. Tulsian[3,6], Jim Warwicker,[7] Paul MacAry[8], Max Crispin,[9] Syma Khalid,[4,5,*] Ganesh S. Anand[3,10,*] and Peter J. Bond[1,3,‡,*]

[1]Bioinformatics Institute (A*STAR), Singapore 138671, Singapore
[2]Department of Chemistry, Faculty of Science and Engineering, Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, United Kingdom
[3]Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore
[4]School of Chemistry, University of Southampton, Southampton SO17 1BJ, United Kingdom
[5]Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom
[6]Department of Biochemistry, National University of Singapore, Singapore 117546, Singapore
[7]School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, United Kingdom
[8]Life Sciences Institute, Centre for Life Sciences, National University of Singapore, 28 Medical Drive, Singapore 117546, Singapore
[9]School of Biological Sciences, University of Southampton, Southampton SO17 1BJ, United Kingdom
[10]Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, United States of America

[†]These authors contributed equally to the paper
[‡]Lead contact

*To whom correspondence should be addressed:
E-mail: peterjb@bii.a-star.edu.sg **(Lead Contact)**
E-mail: gsa5089@psu.edu
E-mail: syma.khalid@bioch.ox.ac.uk

Table S1: List of simulations (related to Star Methods)

S protein glycosylation state	Benzene concentration (M)	Simulation time (ns)	Number of replicates
Non-glycosylated	0	200	2
High mannose	0	200	2
Dominant glycans	0	200	2
Non-glycosylated	0.2	200	3
High mannose	0.2	200	3
Dominant glycans	0.2	200	3
Dominant glycans ECD only	0.2	200	3

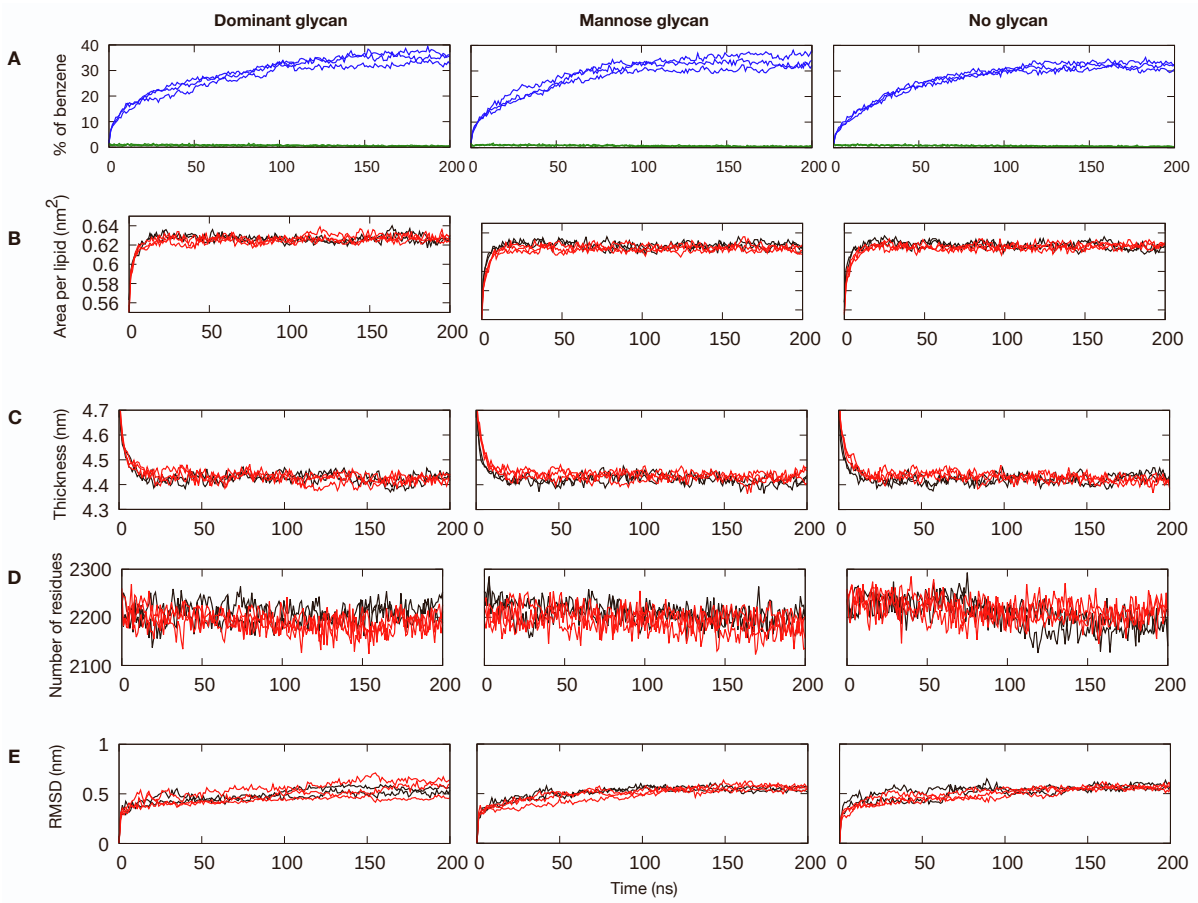

Figure S1: Membrane and protein structural properties during simulations of S protein with and without benzene (related to Figure 1). (A) Percentage of benzene in contact with the membrane lipids (green) or S protein (blue) throughout the simulations with benzene. Cut-off for contact is 0.4 nm. **(B), (C)** Area per lipid and membrane thickness in simulations with (red) and without benzene (black). **(D)** Total number of residues with secondary structural elements (α-helix, β-sheet, β-bridge and turn). **(E)** Backbone RMSDs for the ECD (residue 27-1146).

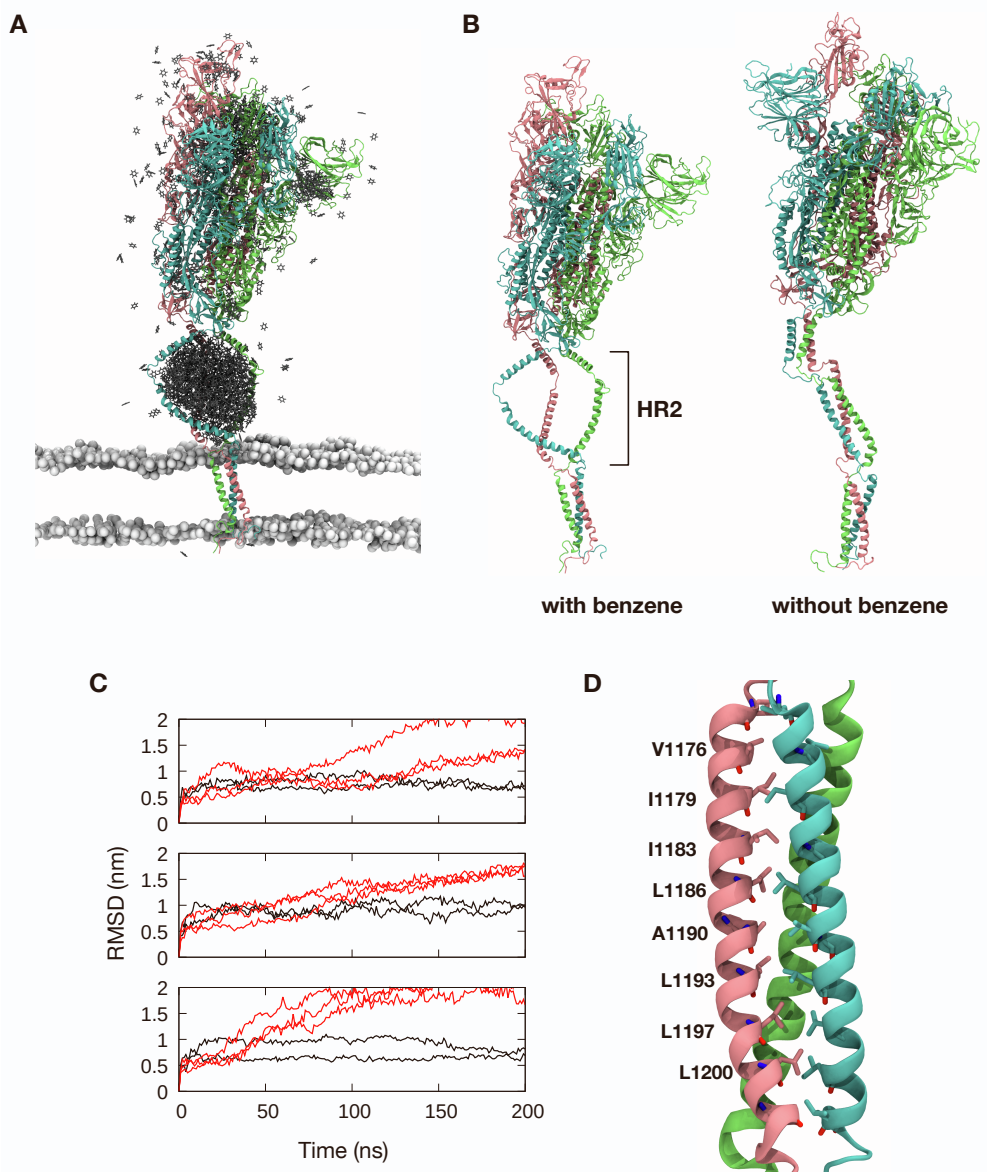

Figure S2: Aggregation of benzene around the HR2 domain (related to Figure 1). (A) The final snapshot from one of the simulations of glycosylated S protein open state with 0.2 M benzene. Benzene is shown in stick representation and coloured grey. For clarity, only benzene found within 2.0 nm of the protein is shown. **(B)** Comparison of protein structures at the end of simulations with and without benzene. HR2 domain is labelled. **(C)** Backbone RMSDs for the HR2 domain from simulations with (red) and without benzene (black). Dominant glycans (top); mannose glycans (middle); no glycans (bottom). **(D)** Hydrophobic residues at the interface of the HR2 domain.

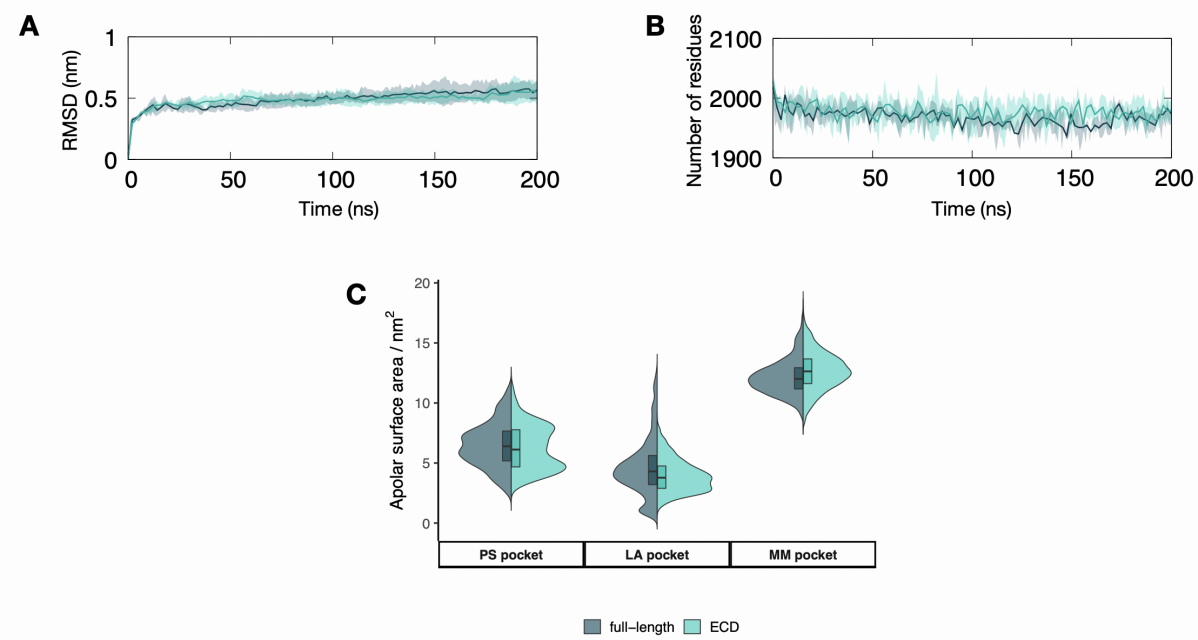

Figure S3: Comparison of simulations with benzene of full-length S protein versus ECD only (related to Figure 1). (A) Backbone RMSDs for the ECD from simulations of isolated ECD dominant glycans (green) and full-length S protein dominant glycans (black). (**B**) Total number of residues with secondary structural elements (α-helix, β-sheet, β-bridge and turn). **(C)** Apolar surface area of the pockets described in Figure 3 calculated from both sets of simulations.

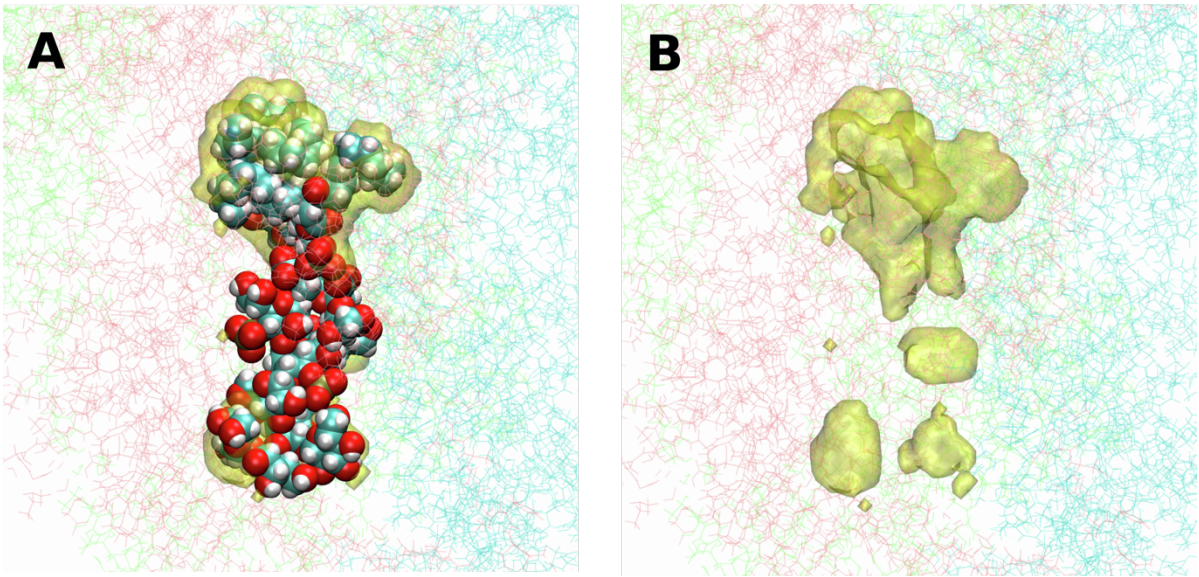

Figure S4: Location and properties of proposed LPS pocket (related to Figure 3). LPS pocket is defined as any pocket area within 0.2 nm of the LPS molecule. **(A)** LPS molecule shown in spheres docked on a previously proposed LPS pocket (Petruk et al., 2021). Mapped pocket area is shown as yellow surface. LPS pocket predominantly outlines the fatty acid tail portion of the LPS, while the hydrophilic sugar components are largely outside any detected pocket regions. **(B)** The outline of the LPS pocket without a bound ligand.

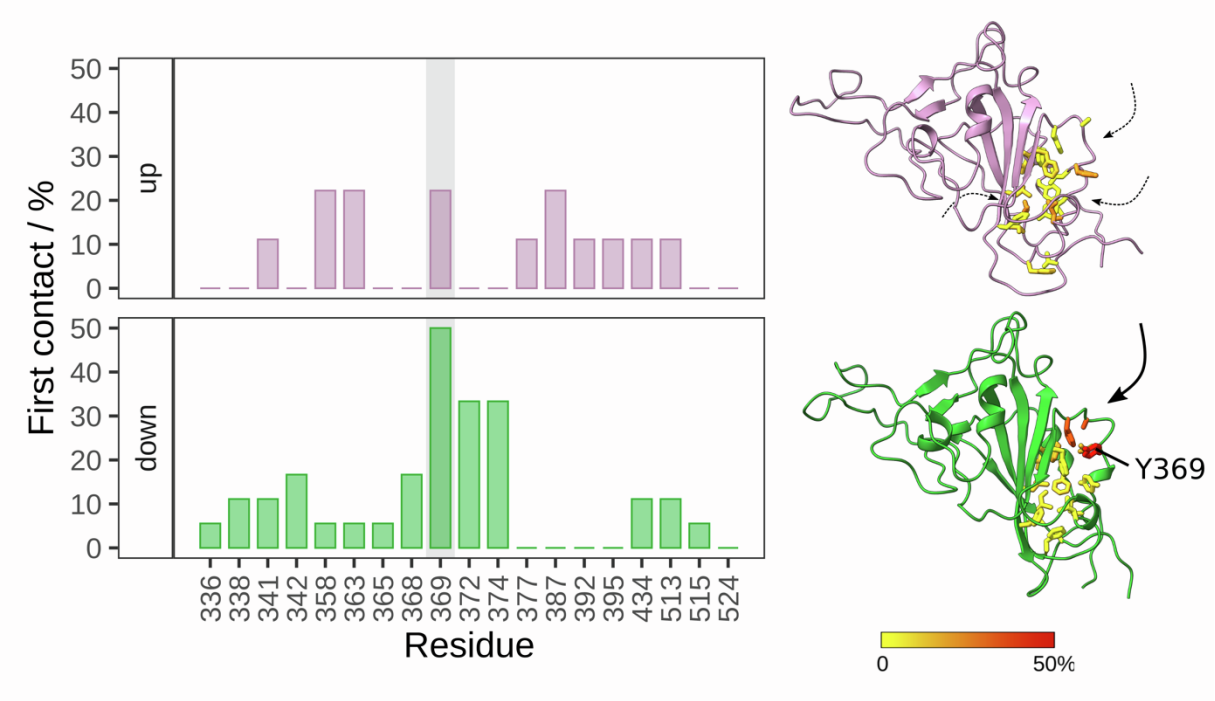

Figure S5. The first point of contact for benzene probes entering the LA pocket and likely ligand entry routes, as shown with arrows (related to Figure 5). The first contact is expressed as a percentage of simulations in which the residue in question establishes initial contact with the ligand, thus marking the entry route for benzene. The highest contact residue, Tyr369, is highlighted in grey on the plot and labelled on the structure. The RBD structures are shown in ribbon representation – 'up' in purple and 'down' in green. LA pocket residues are displayed as sticks and coloured according to the first contact frequency. Down conformation presents a single entry route for the ligands, whereas the up conformation cannot offer a single determinate mode of binding.

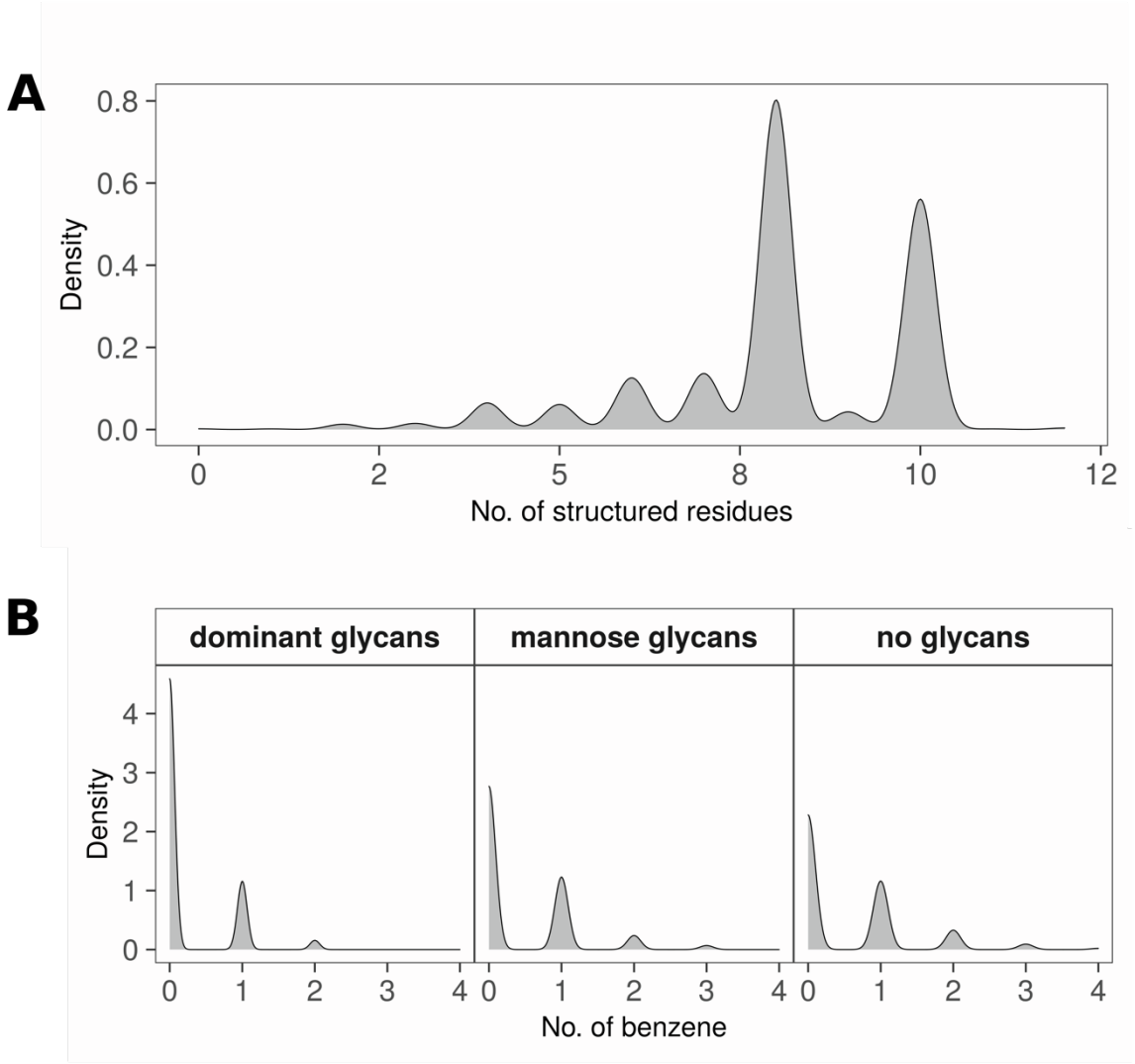

Figure S6: Properties of MM pocket segment (related to Figure 6). (A) Cumulative density distribution of number of residues in the 617-628 loop that are structured (either as α-helix or turn). **(B)** Number of benzene molecules that occupy a segment of the MM pocket underneath the 617-628 loop (see Figure 6A). The maximum number of benzenes occupying the pocket at any given point was three.

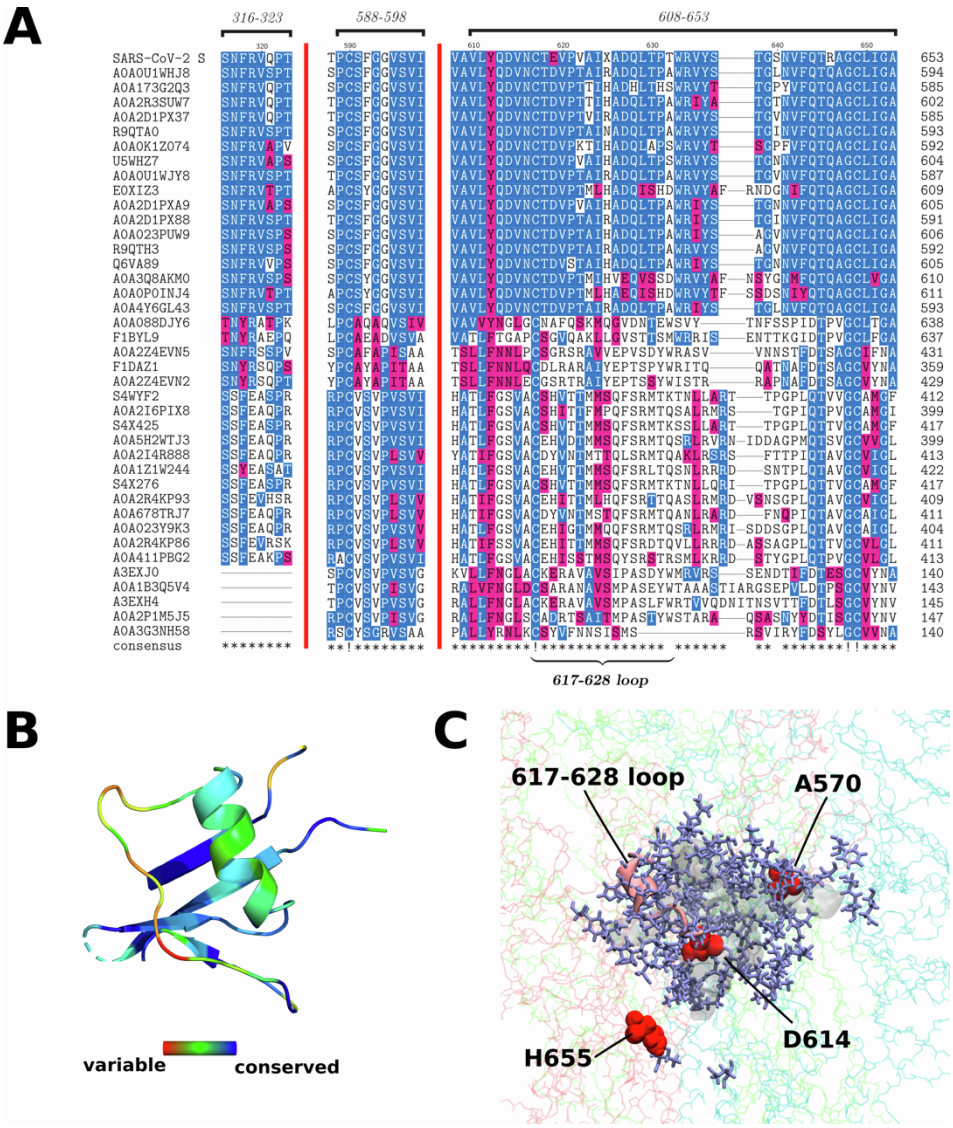

Figure S7: Conservation of the MM pocket segment underneath the 617-628 loop (related to Figure 6). Sequences were selected and scored using CONSURF webserver (details in Methods). **(A)** Multiple sequence alignment of the MM pocket segment. Residues are shaded based on similarity, using a threshold of 30%. **(B)** MM pocket segment is shown in cartoon representation and is coloured according to CONSURF score. **(C)** The MM pocket is shown in transparent surface representation, residues surrounding the pocket are shown in licorice, and the 617-628 loop in cartoon. The remainder of the S protein is shown in pink, cyan, and lime lines. Residues close to the MM pocket and mutated in novel SARS-CoV-2 strains are shown in red spheres.