**Supplementary information for: Site-specific steric control of SARS-CoV-2 spike glycosylation**

Joel D. Allen1\*, Himanshi Chawla1\*, Firdaus Samsudin2\*, Lorena Zuzic2,18, Aishwary Tukaram Shivgan2,17, Yasunori Watanabe1, Wan-ting He3,4,5, Sean Callaghan3,4,5, Ge Song3,4,5, Peter Yong3,4,5, Philip J. M. Brouwer6, Yutong Song7,8, Yongfei Cai9, Helen M. E. Duyvesteyn11, Tomas Malinauskas11, Joeri Kint12, Paco Pino12, Maria J. Wurm12, Martin Frank13, Bing Chen9,10, David I. Stuart11,14, Rogier W. Sanders6,15, Raiees Andrabi3,4,5, Dennis R. Burton3,4,5,16, Sai Li7,8, Peter J. Bond2,17,†, Max Crispin1†

1 School of Biological Sciences, University of Southampton, Southampton, SO17 1BJ, UK

2 Bioinformatics Institute, Agency for Science, Technology and Research (A\*STAR), Singapore, 138671, Singapore

3 Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA 13 92037, USA

4 IAVI Neutralizing Antibody Center, The Scripps Research Institute, La Jolla, CA 92037, USA

5 Consortium for HIV/AIDS Vaccine Development (CHAVD), The Scripps Research Institute, La Jolla, CA 92037, USA

6 Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, Amsterdam Infection & Immunity Institute, Amsterdam, The Netherlands

7 Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China

8 Beijing Advanced Innovation Center for Structural Biology & Frontier Research Center for Biological Structure, Beijing 100084, China

9 Division of Molecular Medicine, Boston Children’s Hospital, 3 Blackfan Street, Boston, MA 02115, USA

10 Department of Pediatrics, Harvard Medical School, 3 Blackfan Street, Boston, MA 02115, USA

11 Division of Structural Biology, University of Oxford, The Wellcome Centre for Human Genetics, Headington, Oxford, OX3 7BN, UK

12 ExcellGene SA, CH1870 Monthey, Switzerland

13 Biognos AB, Generatorsgatan 1 41705, Göteborg, Sweden

14 Diamond Light Source Ltd, Harwell Science & Innovation Campus, Didcot, OX11 0DE, UK

15 Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY, USA

16 Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard University, Cambridge, MA 02139, USA

17 ﻿Department of Biological Sciences, National University of Singapore, Singapore 117543, Singapore

18 Department of Chemistry, Faculty of Science and Engineering, Manchester Institute of Biotechnology, The University of Manchester, Manchester M1 7DN, UK

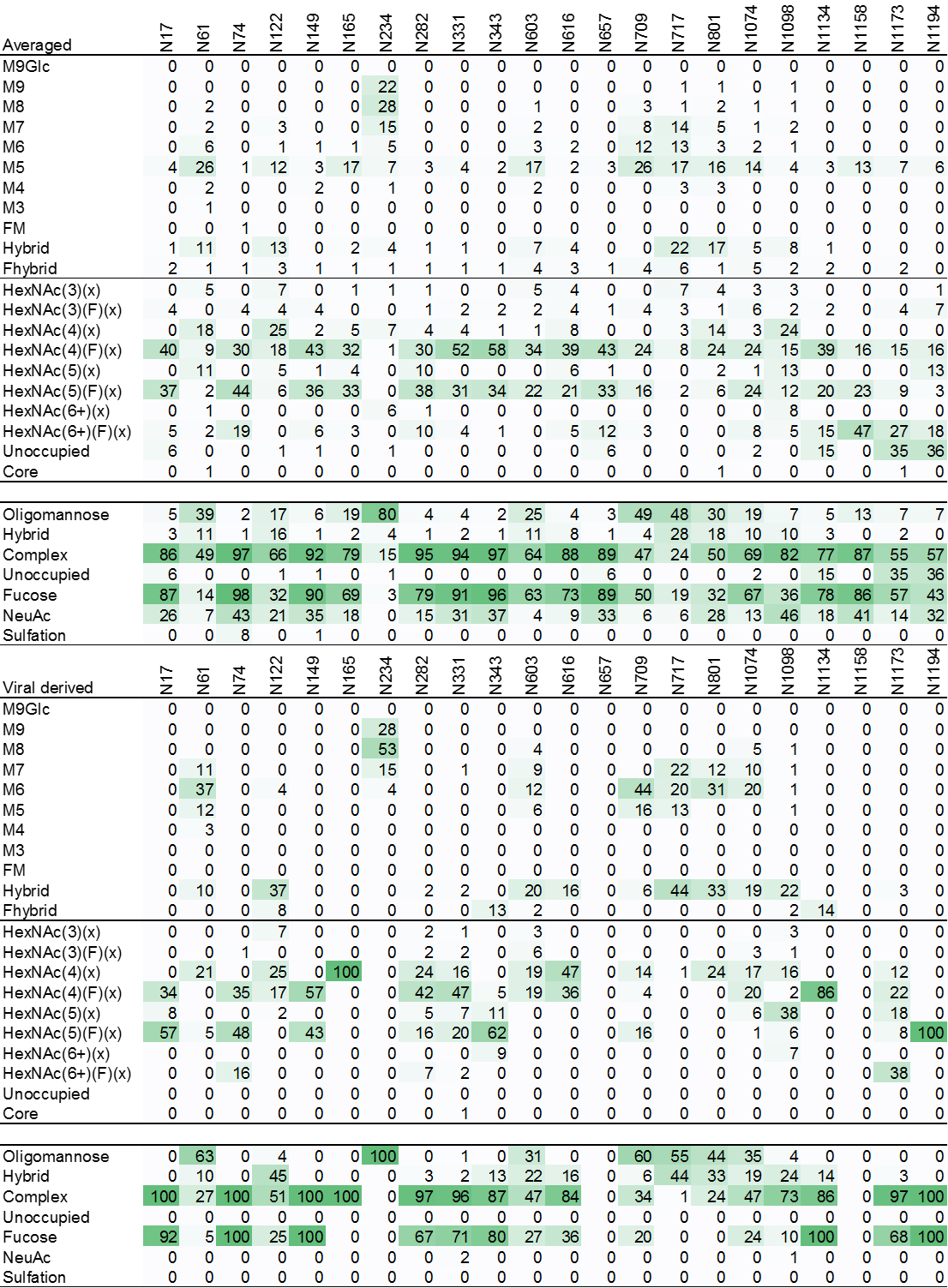
\* These authors contributed equally

†Correpsondence: Max Crispin (max.crispin@soton.ac.uk) and Peter Bond ([peterjb@bii.a-star.edu.sg](mailto:peterjb@bii.a-star.edu.sg)).

Supplemental Tables 1, 2 and 3

Supplemental Figures 1, 2, 3, 4, 5 and 6

Total number of pages: 10

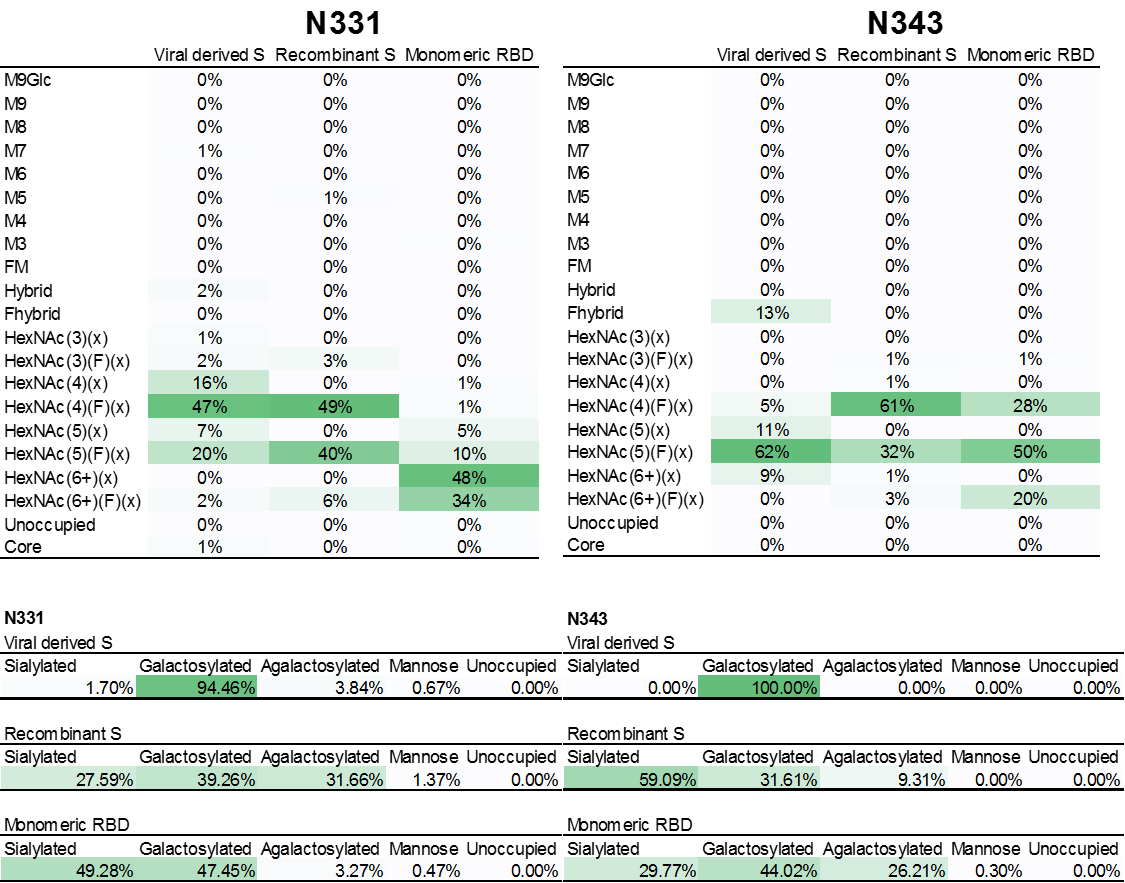
**Table S1: Average recombinant and viral-derived S protein site-specific glycan compositions averaged** 

The site-specific compositions categorized as outlined in the materials and methods for the average of all recombinant samples and of the viral S protein. These tables are the figures used for the bottom panel in Figure 1A and also Figure 1B

**Table S2: Percentage point change in glycosylation on viral-derived S protein compared to recombinant S protein**



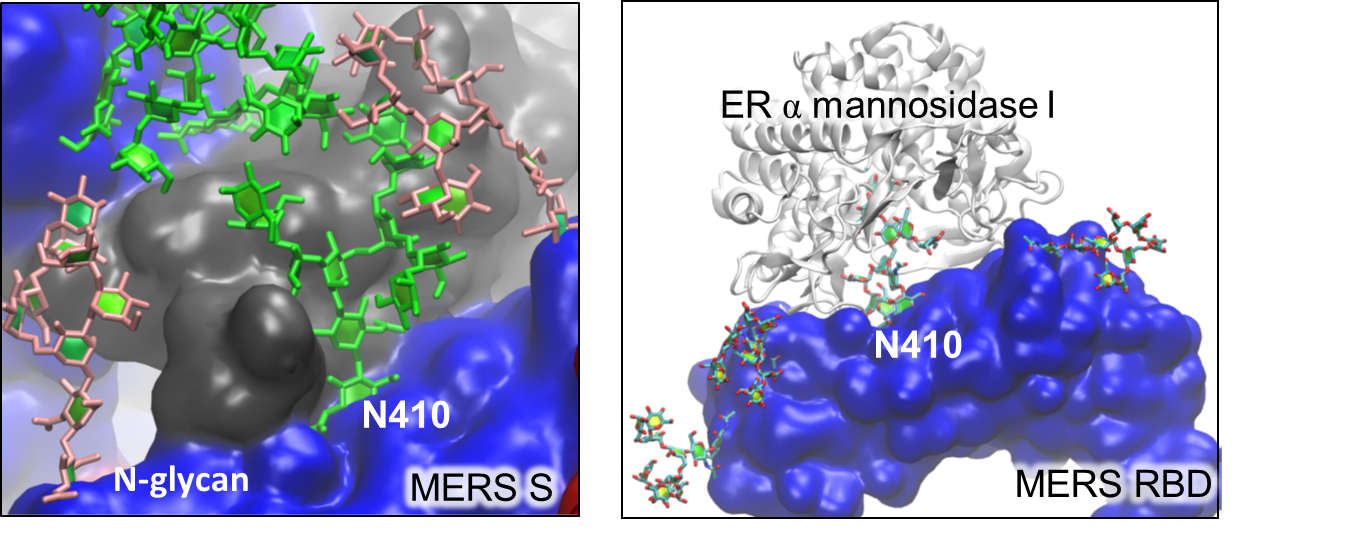
The tabulated values represent the percentage point change in site-specific composition between the two tables in **Table S1**. A positive value in the table represents a category of glycan that was enriched in the viral-derived S protein and a negative value represents a category of glycan that is less abundant on viral-derived S protein.

**Table S3: Site-specific glycan compositions of the two RBD sites, N331 and N343**

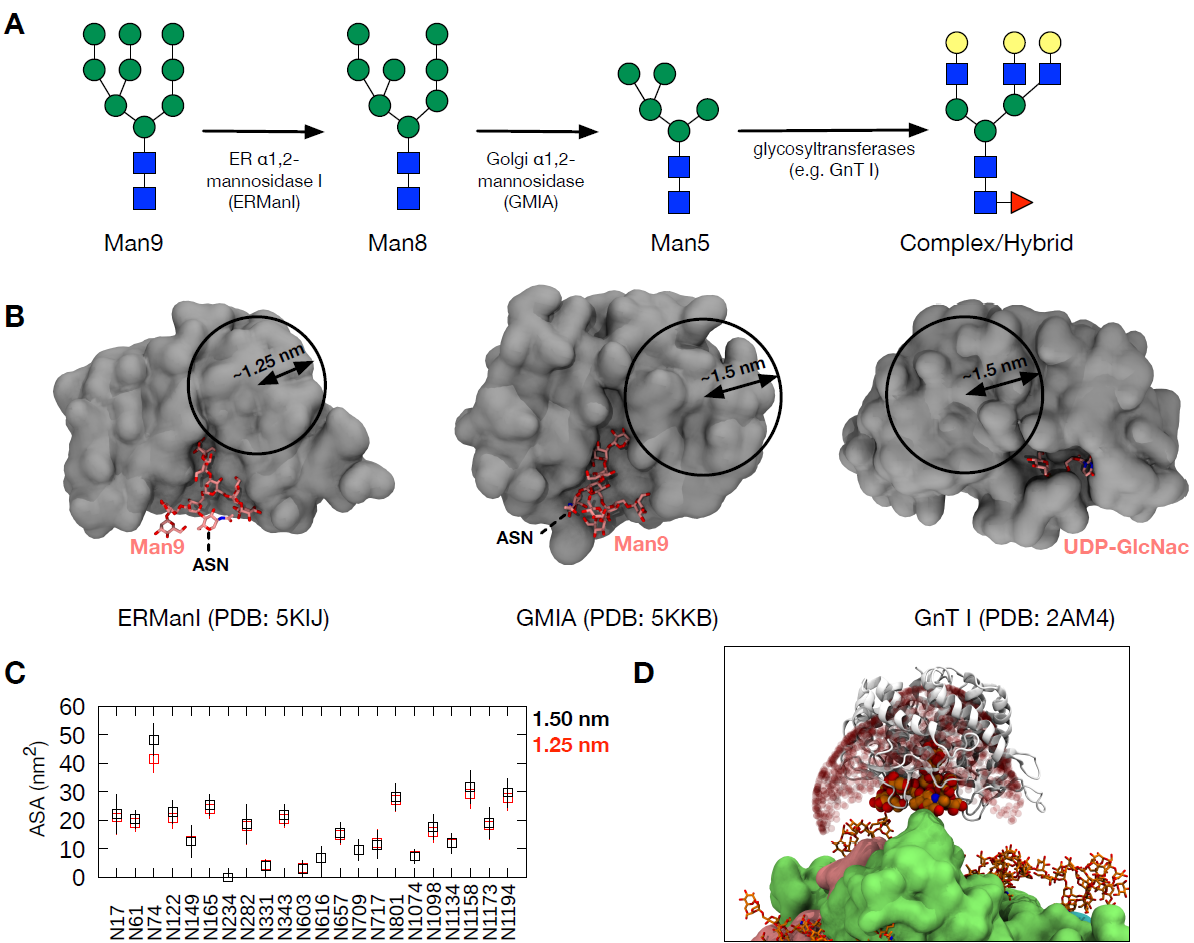
The site-specific composition of the two N-linked glycan sites, N331 and N343, are displayed, with the categories identical to **Table S1**. The values for N331 and N343 for recombinant and virion are reproduced from **Table S1**, the monomeric RBD values are those represented as bars in **Figure 3.**



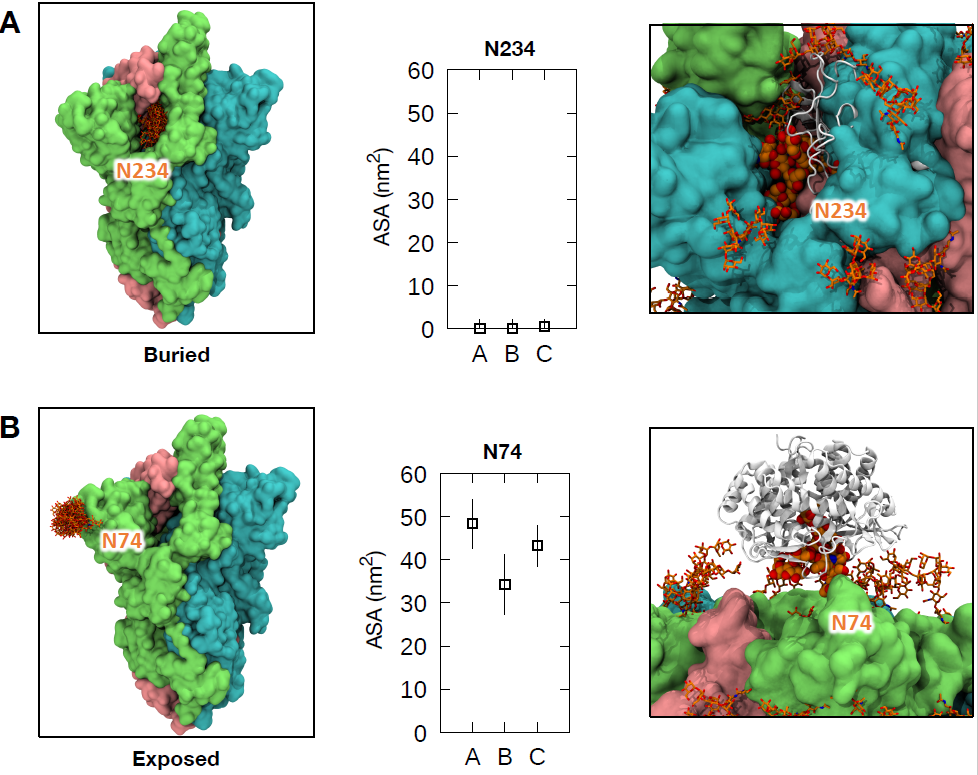
**Figure S1: BLAST alignment of recombinant S proteins.** Sequences are labelled according to the principal investigator responsible for expression and purification. Protein analyzed from the binding site used the McLellan construct



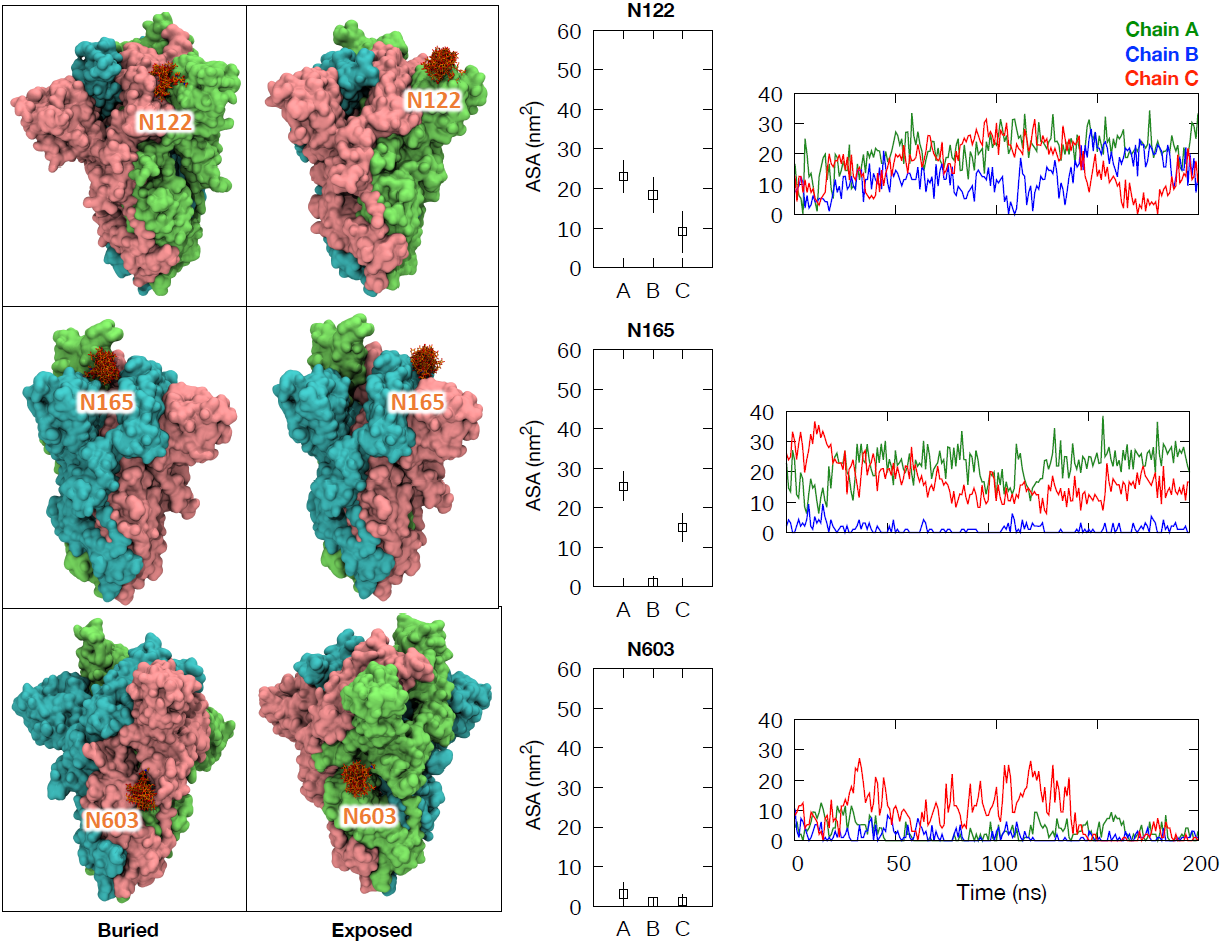
**Figure S2: Model demonstrating the differential accessibility of the N410 glycan for glycan processing enzymes between trimeric MERS S protein and monomeric MERS RBD.** In trimeric MERS-CoV (left) the accessibility of the oligomannose-type N-glycan at N410 is significantly restricted due to interactions with the RDB of a neighboring S protein chain (shown in dark grey). Additionally, there are surrounding N-glycans that provide further shielding. As a consequence glycan N410 is protected from processing by mannosidase enzymes. This is in contrast to monomeric MERS RBD, were the glycan can be accessed e.g. by ER α mannosidase I (right), which results in highly processed N-glycans. These observations highlight how the quaternary structure of a glycoprotein is a key determinant of the glycan processing state.



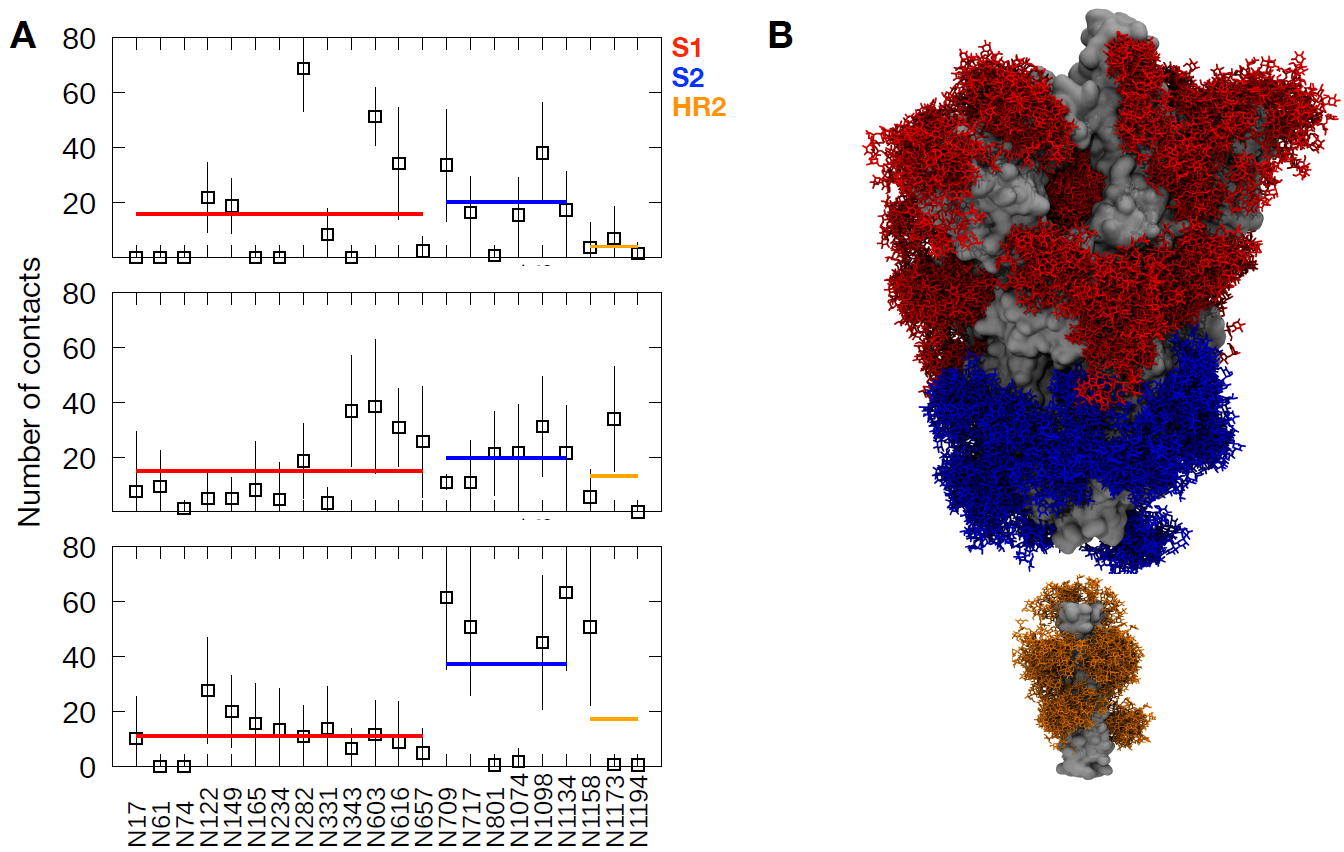
**Figure S3: Validation of probe size for accessible surface area (ASA) measurement.** (A) A schematic of glycan processing from Man9 to complex/hybrid glycans and examples of enzymes involved. (B) Cross-section view of the crystal structures of these enzymes (grey in surface representation) bound to their substrate or substrate analogue (pink in stick representation). The circles show the approximate probe radii that could be used for glycan ASA calculation. (C) Comparison of average ASA values measured using 1.25 and 1.5 nm probe radii for Man9 in chain A of the S protein. Error bars indicate standard deviation along the simulation trajectory. (D) Accessible points (shown as red spheres) using a probe of 1.5 nm around glycan N74 (orange; van der Waals representation). S protein show in surface representation in green (chain A), cyan (chain B) and pink (chain C) and other glycans shown in stick representation in orange. ERManI enzyme shown in cartoon representation in white.



**Figure S4: Examples of glycans either buried or exposed in all three chains.** (Left) Protein shown in surface representation with the individual chains coloured differently: chain A, green; chain B, cyan; chain C, pink. The overlay of glycan snapshots taken every 5 ns for the last 50 ns of the simulation is shown in stick representation. (Middle) Average ASA values for all three chains with error bars indicating the standard deviation along the trajectory. (Right) The structure of ERManI (PDB: 5KIJ) docked onto the glycans to show overlap with the S protein in A and full accessibility of the glycan in B. ERManI is shown in white and cartoon representation.



**Figure S5: Examples of glycans with bimodal accessibility properties.** (Left) Protein is shown as in Supplementary Figure 3, with glycan snapshots taken from the portion of the simulation during which they were buried and exposed. (Middle) Average ASA values for all three chains with error bars indicating the standard deviation along the trajectory. (Right) ASA values for these glycans throughout the 200 ns simulation.



**Figure S6: Glycan-glycan contacts from MD simulation.** (A) The average number of contacts made by glycans on each glycosylation site with any other glycans on the S protein throughout the last 50 ns of the simulation. The error bars show standard deviation along the trajectory. The thick red, blue and orange horizontal lines show the average contacts made by all glycans in the S1, S2 and HR2 regions, respectively. (B) An overlay of glycan snapshots taken every 10 ns along the 200 ns trajectory. Glycans are shown in stick representation and colored red, blue and orange for S1, S2 and HR2 regions, respectively. S protein is shown in surface representation and colored grey.