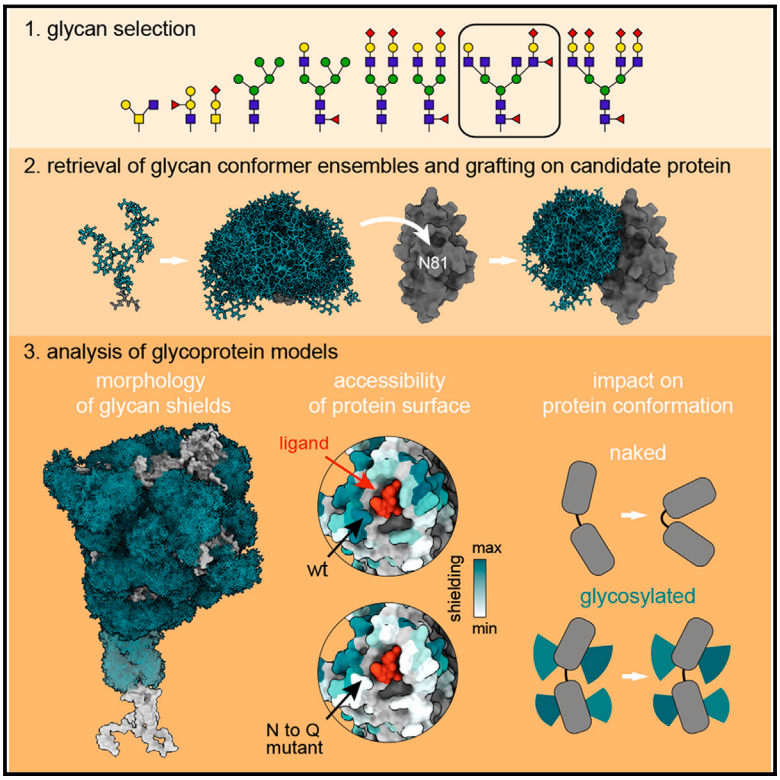
Cell

Resource

Rapid simulation of glycoprotein structures by grafting and steric exclusion of glycan conformer libraries

Graphical abstract



Highlights

- GlycoSHIELD predicts morphology and span of glycan shields on protein surface
- Predicts glycan impact on protein conformation and surface accessibility
- GlycoSHIELD models are consistent with cryo-EM datasets
- Enhances the interpretation of experimental density maps

Authors

Yu-Xi Tsai, Ning-En Chang, Klaus Reuter, ..., Shang-Te Danny Hsu, Cyril Hanus, Mateusz Sikora

Correspondence

sthsu@gate.sinica.edu.tw (S.-T.D.H.), cyril.hanus@inserm.fr (C.H.), mateusz.sikora@uj.edu.pl (M.S.)

In brief

GlycoSHIELD and its associated library of precomputed glycan conformers enable the generation of realistic glycosylated protein structure models at a fraction of the computing cost of conventional molecular dynamics simulations.

Tsai et al., 2024, Cell *187*, 1296–1311
February 29, 2024
https://doi.org/10.1016/j.cell.2024.01.034

CellPress

CellPress
OPEN ACCESS

Cell

Resource

Rapid simulation of glycoprotein structures by grafting and steric exclusion of glycan conformer libraries

Yu-Xi Tsai,[1,2,11] Ning-En Chang,[1,2,11] Klaus Reuter,[3,11] Hao-Ting Chang,[1,2] Tzu-Jing Yang,[1,2] Sören von Bülow,[4] Vidhi Sehrawat,[4,5] Noémie Zerrouki,[6] Matthieu Tuffery,[6] Michael Gecht,[4] Isabell Louise Grothaus,[7] Lucio Colombi Ciacchi,[7] Yong-Sheng Wang,[1,2] Min-Feng Hsu,[1] Kay-Hooi Khoo,[1,2] Gerhard Hummer,[4,8] Shang-Te Danny Hsu,[1,2,9,12,*] Cyril Hanus,[6,10,12,*] and Mateusz Sikora[4,5,12,13,*]

[1]Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan
[2]Institute of Biochemical Sciences, National Taiwan University, Taipei 10617, Taiwan
[3]Max Planck Computing and Data Facility, 85748 Garching, Germany
[4]Department of Theoretical Biophysics, Max Planck Institute for Biophysics, 60438 Frankfurt, Germany
[5]Malopolska Centre of Biotechnology, Jagiellonian University, 31-007 Kraków, Poland
[6]Institute of Psychiatry and Neurosciences of Paris, Inserm UMR1266, Université Paris-Cité, 75014 Paris, France
[7]Hybrid Materials Interfaces Group, Faculty of Production Engineering, Bremen Center for Computational Materials Science and MAPEX Center for Materials and Processes, University of Bremen, 28359 Bremen, Germany
[8]Institute of Biophysics, Goethe University, 60438 Frankfurt, Germany
[9]International Institute for Sustainability with Knotted Chiral Meta Matter (WPI-SKCM^2), Hiroshima University, Hiroshima 739-8526, Japan
[10]GHU Psychiatrie et Neurosciences de Paris, 75014 Paris, France
[11]These authors contributed equally
[12]These authors contributed equally
[13]Lead contact
*Correspondence: sthsu@gate.sinica.edu.tw (S.-T.D.H.), cyril.hanus@inserm.fr (C.H.), mateusz.sikora@uj.edu.pl (M.S.)
https://doi.org/10.1016/j.cell.2024.01.034

SUMMARY

Most membrane proteins are modified by covalent addition of complex sugars through N- and O-glycosylation. Unlike proteins, glycans do not typically adopt specific secondary structures and remain very mobile, shielding potentially large fractions of protein surface. High glycan conformational freedom hinders complete structural elucidation of glycoproteins. Computer simulations may be used to model glycosylated proteins but require hundreds of thousands of computing hours on supercomputers, thus limiting routine use. Here, we describe GlycoSHIELD, a reductionist method that can be implemented on personal computers to graft realistic ensembles of glycan conformers onto static protein structures in minutes. Using molecular dynamics simulation, small-angle X-ray scattering, cryoelectron microscopy, and mass spectrometry, we show that this open-access toolkit provides enhanced models of glycoprotein structures. Focusing on N-cadherin, human coronavirus spike proteins, and gamma-aminobutyric acid receptors, we show that GlycoSHIELD can shed light on the impact of glycans on the conformation and activity of complex glycoproteins.

INTRODUCTION

An estimated 60% of drugs currently available or under development target cell surface proteins.[1] It is therefore essential to characterize the three-dimensional structure of these proteins. In their cellular context, the overwhelming majority of these proteins are glycosylated.[2] Despite a general awareness that glycans modify protein stability and function, we still know surprisingly little on how glycan chemistry, structure, and dynamics drive these effects.

As shown by molecular dynamics simulation (MDS) of glycoproteins, glycans are much more dynamic than folded polypeptides.[3,4] Glycans sample extensive arrays of conformations over tens of nanoseconds, hence creating molecular shields that mask large swaths of protein surface. Such simulations provide key information on the impact of glycans on protein conformation and interactions with drugs and other biomolecules.[5,6] However, they require both expert knowledge and extended computing times on specialized supercomputers, which limits their use for routine evaluation of glycoprotein structures and comparison of distinct protein glycoforms.

Protein glycosylation encompasses an extremely diverse repertoire of chemical compositions and varies during tissue development and aging.[2,7] Protein glycosylation is strongly influenced by alteration of tissue homeostasis and cellular stress and

thus represents a promising but hitherto largely underexploited source of biomarkers. Depth and breadth of glycoproteomics analyses are rapidly expanding. A wealth of data is now available on the N- and O-glycomes of a variety of organisms and tissues in physiological or pathological conditions.[8–12] Progress in protein structure prediction enables structural prediction of complete proteomes and metagenomes.[13–18] These predictions, however, largely ignore glycans. There is thus a need for reliable and accessible tools to evaluate how glycan diversity impacts protein structure and shielding on a large scale.

Here, we use MDS to generate a library of representative conformers for a large cast of O- and N-glycan types and developed an open-source toolkit and a web application, GlycoSHIELD, to graft ensembles of glycan conformers onto any static protein structures. To provide proof-of-principle of the validity and predictive power of this approach, we focused on three classes of glycoproteins of particular interest whose structure, biochemistry, or pharmacology have been extensively studied: N-cadherin, an essential adhesion molecule,[19] coronavirus spike (S) proteins, which mediate coronavirus entry into target cells,[20] and type A γ-aminobutyric acid ($GABA_A$) receptors, the main inhibitory neurotransmitter receptor in mammals.[21]

Through comparison with extended simulations of glycosylated N-cadherin and SARS-CoV-2 S protein, we show that GlycoSHIELD captures key features of glycan impact on protein structure, in particular epitope masking and steric constraints on protein conformation. Because glycans are very dynamic, they are typically resolved only partially in protein structures elucidated by cryoelectron microscopy (cryo-EM)[22–24] and unaccounted for when using lower-resolution techniques such as small-angle X-ray scattering (SAXS). By comparing glycoprotein models with experimental density maps, we found that GlycoSHIELD can recover important information on glycan dynamics that is typically lost during cryo-EM data processing. We also provide evidence that GlycoSHIELD improves the fit of theoretical models to experimental data in SAXS experiments. By combining GlycoSHIELD with mass spectrometry (MS) and cryo-EM, we reconstructed the full glycan cover of the S proteins of six circulating human coronaviruses (hCoVs), namely, SARS-CoV-2, SARS-CoV, MERS-CoV, hCoV-HKU1, hCoV-NL63, and hCoV-229E, providing insight on S protein binding to cellular receptors and antibodies. Finally, we modeled fully glycosylated $GABA_A$ receptors. We found that glycans may regulate the binding of specific ligands and likely occupy a larger volume of the channel lumen than previously anticipated, raising hypotheses on the mechanics of this important neurotransmitter receptor.

GlycoSHIELD therefore provides quantitative insights on glycoprotein structure and function at a fraction of the computing cost of conventional MDS. GlycoSHIELD and its associated library of 68 glycan types to date are available online as a web application (www.glycoshield.eu) and stand-alone Python package (https://gitlab.mpcdf.mpg.de/dioscuri-biophysics/glycoshield-md/).

RESULTS

Rationale and glycan conformer sampling

As shown for SARS-CoV-2 S protein and other viral proteins by experimental data[25–28] and simulations,[3,4,6,29] glycan conformations are sterically constrained by local protein structure and strongly differ from one glycosylation site to another. Accessible glycan conformations are rapidly sampled through thermal agitation, depending mostly on extensive interactions with the solvent and non-specific self-interactions within individual glycans.[30] Excepting the case of specialized glycan-binding protein domains,[31–33] glycan interactions with the host protein surface are in most cases non-specific and transient.[34,35] We thus reasoned that these conformations could be reliably captured by simulations of individual glycans whose conformers may then be grafted on specific glycosylation sites and selected or rejected based on steric clashes with the protein structure. Similar grafting with steric exclusion has been used successfully to model disordered proteins[36] and reconstruct small N-glycans from individual disaccharides.[37] Here we generated a library of 68 N- and O-glycans of high prevalence and particular physiological relevance. For each of these glycans, we performed extensive MDS ($\geq$3 μs) in aqueous solution and obtained large conformation arrays for grafting on any static protein structure (Figure 1A; Table S1; see STAR Methods for details).

To verify that all regions of the glycan conformational space had been explored with the unbiased sampling used in our MDS, we performed simulations with enhanced sampling methods that were recently shown to exhaustively probe glycan conformations[38] (see STAR Methods). Simulations were performed with both methods for three glycans of increasing complexity (Man5, bi-antennary A2, and tetra-antennary Fuc1_Neu4; see Table S1 and Figure S1). Despite small differences in conformer distributions, unbiased simulations did sample the entire conformational space determined by enhanced sampling (Figure S1), confirming that unbiased MDS adequately sampled conformations even for the largest glycans in our library.

GlycoSHIELD provides realistic predictions of protein shielding

Next, we determined whether GlycoSHIELD faithfully reconstructed glycan shields modeled by full MDS. As noted, glycan dynamics seem to occur primarily without stable interactions with the host protein surface.[34,35] However, labile glycan-protein interactions may still influence the overall morphology of glycan shields and the conformation of the protein.[39] To quantify these effects, we turned to N-cadherin, an adhesion glycoprotein comprising five compact and stable immunoglobulin (Ig)-like domains (EC1-EC5) connected by linkers, which, in the absence of bound calcium, enable extensive interdomain movements.[40] To sample both glycan and domain dynamics, we performed MDS of a reduced system composed of EC4-EC5 domains, which was modeled over 3 μs in the absence of calcium in its non-glycosylated form or with four distinct glycans (Figure 1B). Over the course of these simulations, glycans remained mobile and probed extensive conformation arrays, covering a substantial fraction of the protein surface (Figure 1C). Next, we applied GlycoSHIELD to graft corresponding glycans onto a static non-glycosylated EC4-EC5 structure. The extent and shape of the glycan shields obtained by MDS were well reproduced by GlycoSHIELD (Figure 1D), with only small differences in the shielding of the protein surface (Figures 1C, 1D, and S2). Consistent with our previous observations on other systems,[6] analysis

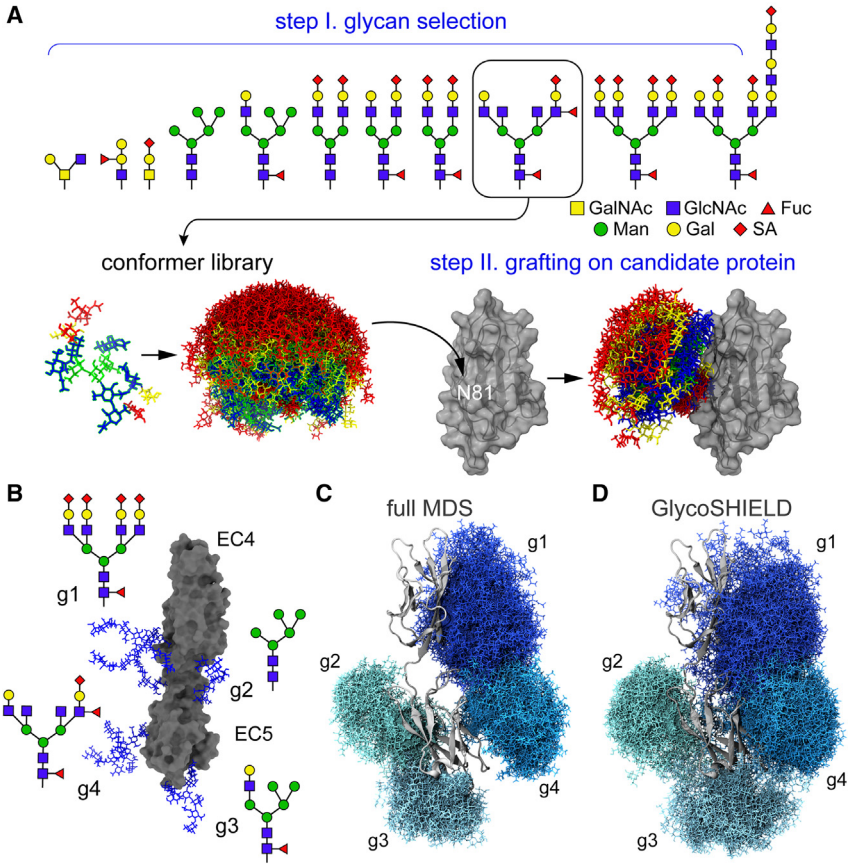


Figure 1. GlycoSHIELD generates realistic glycan shield models

(A) Overview of the pipeline: user provides a 3D protein structure with defined glycosites where glycans from the library of conformers not clashing with the protein are grafted and exported for visualization and analysis (GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Fuc, fucose; Man, mannose; Gal, galactose; SA, sialic acid; see color code in inset). See Table S1 for description of 68 glycans available to date.

(B) Structure of N-cadherin EC4-EC5 model system (surface representation, gray) with four distinct N-glycans as indicated at each glycosylation site (sticks, blue, g1–g4).

(C and D) Glycan conformers (sticks, shades of blue) generated by full MDS (C) or with GlycoSHIELD (D) after alignment on EC4-EC5 (cartoon, gray). Note the comparable morphology and span of the glycan shields obtained by the two approaches.

See also Figures S1, S2, S3, and S4.

of hydrogen bonds (H-bonds) showed that EC4-EC5 glycans interacted mostly with the solvent (11 ± 3, 24 ± 4, 299 ± 12; mean ± standard deviation; for glycan-protein, glycan-glycan, and glycan-water H-bonds, respectively). Small differences in protein shielding derived from GlycoSHIELD and full MDS (Figure S2) were thus not caused by direct glycan interactions with the protein. These differences rather resulted from the flexibility of the protein surface and steric hindrance between individual glycans, which, by design, are not accounted for in GlycoSHIELD. We thus conclude that GlycoSHIELD accurately captures the morphology and span of glycan shields modeled by MDS.

As a first step in assessing the validity of these models experimentally, we turned to SAXS, a technique that is widely used to characterize the structural dynamics of protein in solution.[41,42] We analyzed a fully glycosylated ectodomain of N-cadherin (EC1-EC5; Figures S3 and S4A–S4C) and a truncated variant encompassing EC4-EC5 domains to directly match the system that was used for MDS (Figures S3 and S4D–S4F). As shown in Figure S4, non-glycosylated protein models did not adequately match experimental SAXS profiles obtained for the two systems. By contrast, addition of N-glycans with GlycoSHIELD significantly improved agreement between theoretical and experimental SAXS profiles (Figures S4A–S4F), thus providing experimental evidence of the plausibility and applicability of GlycoSHIELD models.

Interestingly, we found for EC4-EC5 that the A2F complex N-glycans—a glycotype shown to be particularly abundant on N-cadherin by MS[43]—provided a better match with experimental data than less frequent glycotypes (Figure S4F), thus perhaps indicating glycan-type sensitivity.

GlycoSHIELD predicts glycan impact on N-cadherin conformation

As evident for N-cadherin (Figure 1B), N-glycans are often located in the vicinity of hinges linking mobile protein domains. It is thus likely that glycan-protein steric hindrance determines possible glycan conformational space, thus in return generating an entropic constraint on the conformation of the protein. To determine whether GlycoSHIELD could predict such effects, we first used MDS to compare the interdomain movements of our non-glycosylated and glycosylated N-cadherin EC4-EC5 systems (Figures 2A–2D). Consistent with other systems,[34,35,39] glycosylation had only a marginal effect on the internal motion of single EC domains (data not shown), indicating that glycan dynamics have little impact on the intrinsic stability of Ig-like domains.[44] By contrast, glycans affected the domain-scale motion of the whole protein. Non-glycosylated EC4-EC5 rapidly folded into a relatively compact and stable conformation, whereas the glycosylated protein remained extended and flexible (Figures 2A–2D). To determine whether the open conformation of glycosylated EC4-EC5 resulted from steric hindrance by glycans limiting close interactions between the EC4 and the EC5 domains, we grafted glycans onto individual snapshots from the MDS of non-glycosylated EC4-EC5 and quantified glycan conformer acceptance over the course of the simulation (Figure 2E). The compactness of EC4-EC5 measured by the gyration radius of the protein backbone negatively correlated with glycan acceptance (Figures 2B and 2E), providing a convenient proxy for assessing the entropic cost of steric hindrance on protein conformation (Figure 2F). This entropic effect depended on the composition of N-glycans (data not shown), indicating that the conformation of N-cadherin is modulated by the glycosylation status of the protein.

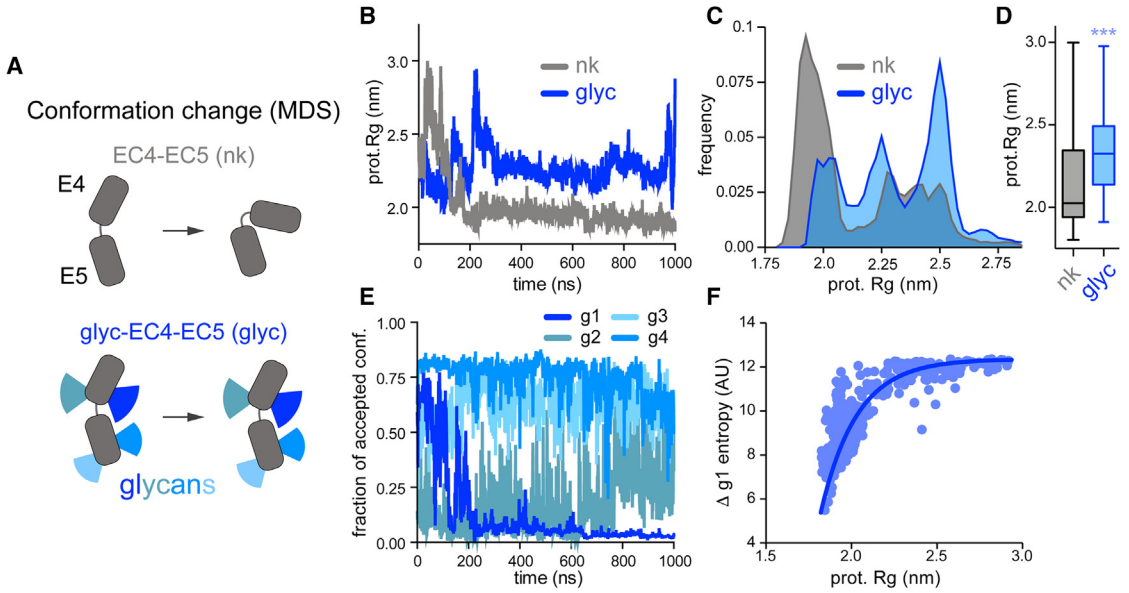


Figure 2. Glycans maintain N-cadherin in an open conformation

(A–D) Full MDS. (A) Schematics of protein movements during simulations of non-glycosylated EC4-EC5 (nk) or glycosylated EC4-EC5 (glyc-EC4-EC5). (B) Variation of protein gyration radius of nk (gray) or glyc-EC4-EC5 (blue) over time (1 μs simulation). (C and D) Distribution of protein gyration radius (Rg) over repeated simulations (3 × 1 μs, N = 3 × 10^4 conformers). Rg values in (D) are graphed as box and whiskers plots, In (C) and (D), note the more elongated conformation (higher Rg values) of glyc-EC4-EC5 compared with EC4-EC5. ***$p < 10^{-3}$, Kolmogorov-Smirnov test.

(E and F) GlycoSHIELD. (E) Fraction of accepted GlycoSHIELD conformers over time for g1–g4 glycans grafted on non-glycosylated EC4-EC5. Note the strong correlation between the compaction of the protein (reduced Rg values in the gray plot in C) and clashing with glycans (reduced fraction of accepted conformers for glycan shown in dark blue). (F) Variation of estimated partial entropy of glycan g1 (k_B ln[Ω], with k_B Boltzmann constant and Ω the number of glycan conformers) as a function of protein radius of gyration, showing increasing glycan steric constraints for Rg values below 2.25 nm.

These results show that GlycoSHIELD provides reliable predictions of glycan impacts on protein conformation. Previous structural studies have shown that calcium-dependent stiffness of the protein is required for strong cadherin-cadherin adhesion.[40] Glycans may thus be important to maintain the protein in an elongated conformation, which is more prone to oligomerization, and prevent its collapse in the absence of calcium.

GlycoSHIELD predicts SARS-CoV-2 S protein epitope accessibility

Extensive glycosylation of viral surface proteins contributes to the success of enveloped viruses in evading the immune system and hinders vaccine and drug development. In this context, the glycan shield of the SARS-CoV-2 S protein has been a focal point in the current effort to develop vaccines and antibodies targeting existing and emerging S protein variants.[45]

SARS-CoV-2 S protein is a homotrimer that comprises 66 N-glycosylation sites.[4,46] Its glycosylation patterns are organism- and cell type-dependent.[47] They impact how S protein interacts with its cellular receptor, angiotensin converting enzyme 2[5] (ACE2), and affect immune responses.[48] Large-scale MDS has provided compelling information on the structural impact and epitope-masking properties of specific N-glycans identified on S protein.[6,46,49–52] Due to prohibitive costs in computing resources, this approach is not amenable to screening of large arrays of tissue-specific N-glycans, calling for more versatile and less costly options. As a first step toward filling this gap, we used GlycoSHIELD to compare shields generated by various glycans on SARS-CoV-2 S protein. To this end, we compared glycan shields generated with GlycoSHIELD on the non-glycosylated protein with shields generated by extended simulation of the glycosylated protein. Although the glycan types used in GlycoSHIELD and MDS were not strictly identical (Figure S5A), the morphologies of the shields obtained by both methods were in very good agreement (Figure 3A). As expected, shield morphologies differed depending on glycan type and composition (Figures S5C and S5D), suggesting that glycan heterogeneity modulates S protein shielding.

As shown by us and others, glycan shields may impact the recognition of the S protein by antibodies,[6,46,53] thereby affecting host immune responses and the outcome of clinical interventions with neutralizing antibodies. We thus determined whether GlycoSHIELD captured the masking of S protein epitopes seen in extended MDS.[6] To this end, we performed a ray accessibility analysis, i.e., quantified the shading of protein surfaces by glycans upon illumination with randomly oriented light rays (see STAR Methods), a method previously shown to be a strong predictor of epitope masking.[6] As shown in Figure 3B, epitope shielding predicted by full MDS was reproduced almost completely by GlycoSHIELD (Figure S5B).

A recent study of densely glycosylated segments of ACE2[54] has shown that the volumes sampled by individual glycans are in most cases not significantly affected by that of neighbor glycans, supporting the notion that individual glycans can be treated in isolation

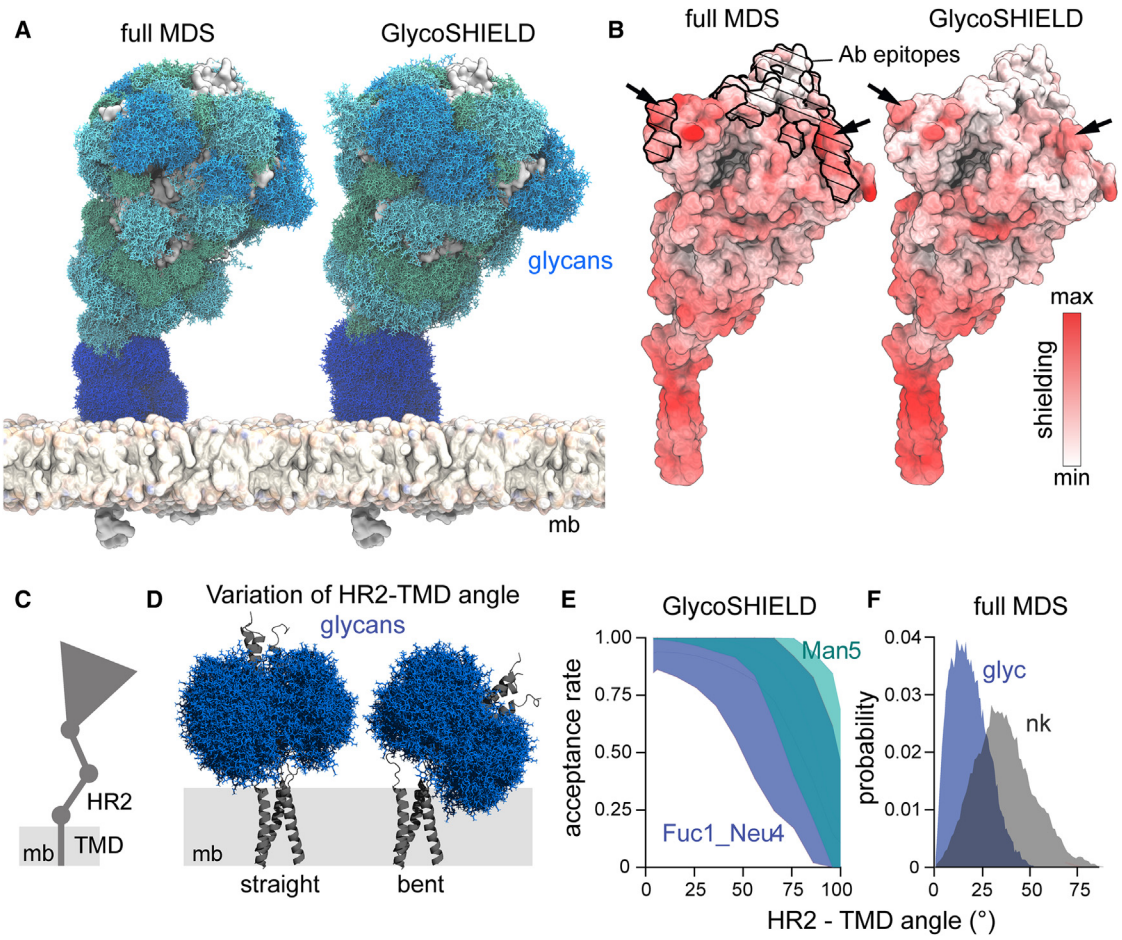


Figure 3. N-glycans may impact the conformation of SARS-CoV-2 Spike protein and its recognition by antibodies

(A) S protein glycan conformers (blue or green sticks) sampled from 10 μs of MDS or reconstructed with GlycoSHIELD. In both structures, 160 glycan conformers are displayed for each glycosite. Membranes were added for visualization purposes (surface representation, mb).

(B) Glycan shielding of S protein extracellular domain (surface representation) calculated from MDS vs. GlycoSHIELD. Shown are 3D heatmaps of accessibility scores calculated by ray analysis. Higher color intensities indicate higher shielding. Arrows indicate predicted shielded areas within antibody epitopes (black lines and hatched areas).

(C) Scheme of S protein main domains.

(D) Model of S protein HR2 and TMD domains in upright and inclined positions with glycan shields shown for N1194 (Fuc1_Neu4 glycans, blue). Note the higher proportion of glycan conformers clashing with the membrane (mb, gray band) for the bent protein.

(E) Rate of glycan conformer acceptance as a function of HR2 bending with all Fuc1_Neu4 (blue) or all Man5 N-glycans (green). Shown are minimum and maximal acceptance rates for the 3 HR2-TMD protomers. Note the increase of steric constraints (reduced conformer acceptance) upon bending and the distinct impact of the two N-glycan types.

(F) Angle distribution of non-glycosylated (nk) and glycosylated (glyc) S protein stalk in 3.5 μs MDS showing decreased bending of the glycosylated protein.

See also Figures S5 and S6.

for prediction of protein shielding. However, as glycan-glycan interactions are not taken into account by GlycoSHIELD, we sought to assess how it approximated glycan ensembles in crowded environments. To this end, we performed full MDSs of S protein heptad repeat 2 (HR2) domain, which hosts three large complex N-glycans in close proximity at position N1194 (Figure S6A). N-glycans were simulated individually on each of the 3 protein protomers or together as a bundle and resulting shields compared with GlycoSHIELD models. We found that glycan-glycan interactions did not intrinsically limit and, somewhat counterintuitively, may in some instances enhance glycan conformation space (Figures S6B and S6C). Thus, although one cannot exclude that steric hindrance between glycans may occasionally reduce available conformation space in highly glycosylated protein domains,[55,56] these simulations indicate that weak and generic glycan—glycan binding[57] may also expand sampled volumes. Importantly, although GlycoSHIELD tended to overestimate glycan overlap (Figure S6B), conformer sampling and predicted impact on protein accessibility were very similar to results obtained using full MDS (Figures S6D).

These results indicate that GlycoSHIELD generates realistic glycan shields, reliably recapitulating MDS predictions of the protein surfaces available for interactions with other molecules and in particular antibodies.

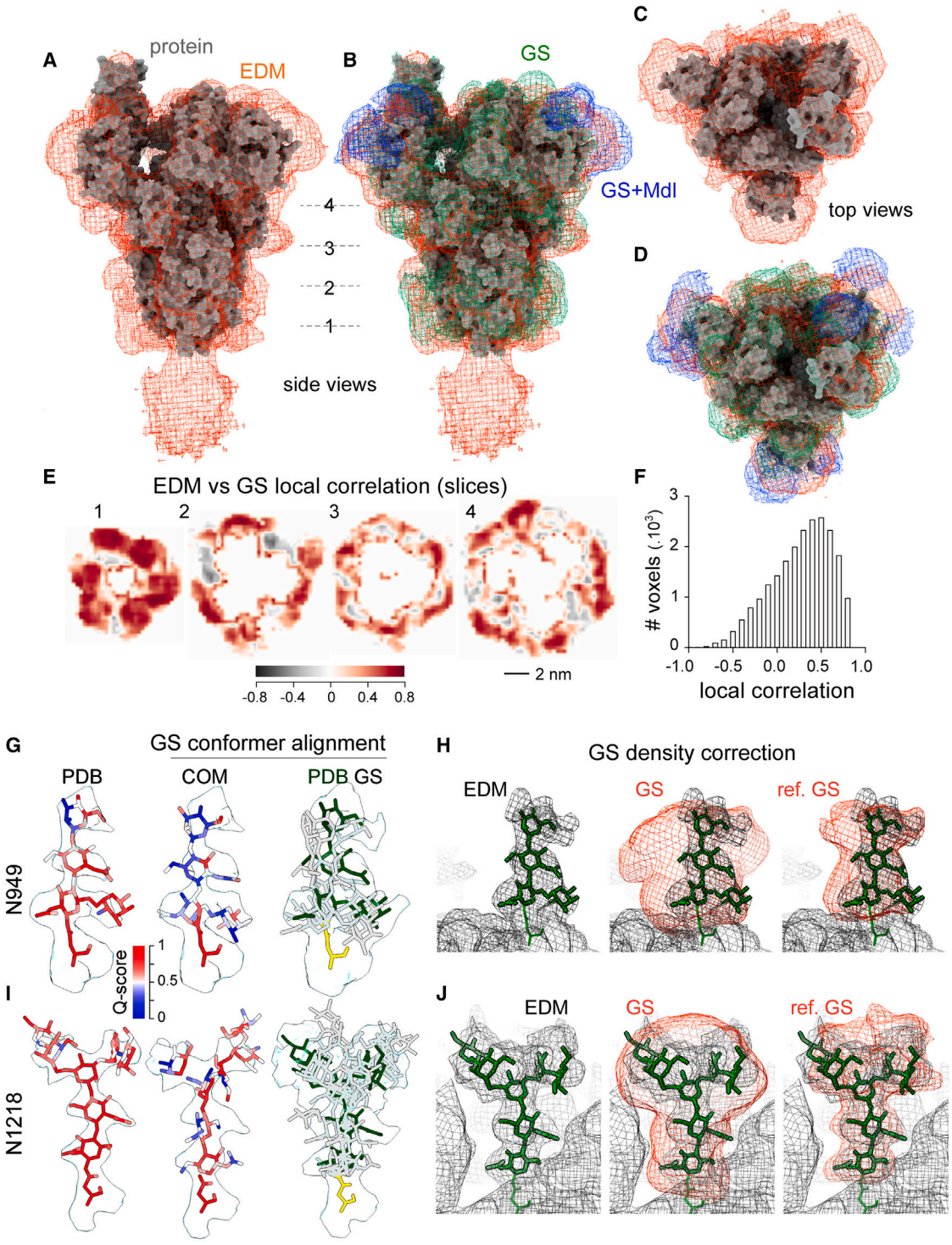


Figure 4. Agreement between GlycoSHIELD models and cryo-EM density maps

(A–F) Glycan shields fill unresolved volumes of cryo-EM density maps.

(A–D) Partial structure of the extracellular domains of SARS-CoV-2 (D614G mutant) S protein (PDB: 7EAZ, gray surface)[62] shown with unsharpened experimental EM density map (EDM, orange mesh) and glycan density maps modeled with GlycoSHIELD (green and blue mesh). Density maps colored in green were directly obtained by grafting glycans on the protein structure (GS). Density maps colored in blue were obtained by grafting glycans on unresolved parts of the proteins that were modeled with Modeler (GS + Mdl). (A and C) Side views. (B and D) Top views (see STAR Methods for details).

(E) Coronal sections ("slices") of correlation heatmap showing local agreement between EDM and GS synthetic map at positions (1–4) indicated in (A).

(legend continued on next page)

GlycoSHIELD predicts N-glycan impact on SARS-CoV-2 S protein orientation

Cryoelectron tomography *in situ* and MDS have shown that S protein stalk remains very flexible[4,58,59] likely facilitating S protein interaction with ACE2 and thus increasing SARS-CoV-2 avidity for target cells (Figure 3C). As shown in Figure 3A, N-glycans located close to S protein transmembrane segments may clash with the virus envelope upon excessive tilting, hence constraining the orientation of the protein. We used GlycoSHIELD to test this hypothesis. To this end, we generated a truncated model comprising S protein HR2 region and transmembrane domain (TMD, Figures 3C and 3D). To mimic the flexibility of the HR2-TMD hinge, we generated a range of protein conformers, varying the angle between HR2 and TMD, and used GlycoSHIELD on each protein conformer to count glycan conformers rejected from grafting due to clash with the virus envelope (Figure 3E). We found a steep decrease of conformer acceptance for HR2 tilted toward the envelope (Figure 3E), which indicates that stalk glycans favor upright positions of S protein.[4] To validate this prediction, we used full MDS to simulate the glycosylated and non-glycosylated models of truncated S protein embedded in a realistic membrane (Figure 3F). We found that the glycosylated protein remained more upright than its non-glycosylated form (Figure 3F), consistent with the typical orientation of the native protein observed *in situ*.[4]

As seen for EC4-EC5, this shows that GlycoSHIELD provides reliable estimates of glycan impact on large-scale protein dynamics.

GlycoSHIELD recovers structural information lost during cryo-EM data processing

Recent advances in cryo-EM have enabled detailed structural characterizations of fully glycosylated proteins at an unprecedented rate.[22–24,60,61] However, most cryo-EM structures provide limited structural information on N-glycans as reported glycan structures are generally trimmed to the first two N-acetylglucosamine (GlcNAc) moieties stemming from N-glycosylated asparagines (Figure S7A). Rapid cooling of protein samples during vitrification for cryo-EM analysis captures glycans and other mobile protein elements in multiple conformations. This results in poor alignment of the cryo-EM images of glycan structures present at individual glycosylation sites, thereby hindering their resolution. Corresponding volumes of experimental EM density maps (EDMs) are thus typically considered "noise" and almost entirely removed during processing. In other words, it is likely that a significant fraction of experimental densities that is lost during the processing of glycoprotein structures may correspond to glycan shields.

To assess this, the glycan shield of a recently resolved SARS-CoV-2 S protein D614G mutant[24] was reconstituted with GlycoSHIELD. Missing sections of the protein models were reconstructed and experimentally determined glycan types grafted on the protein using GlycoSHIELD (see STAR Methods). Synthetic EDMs were then generated from GlycoSHIELD models and compared with unsharpened experimental EDMs. As previously described,[24] N-glycans were either completely missing (as in the case of glycans decorating the flexible loops of the N-terminal domain [NTD]) or heavily truncated (with only the first two GlcNAc fitted to processed EDMs, Figure S7A), resulting in the elucidated protein structure seemingly floating in the much larger volume of the unsharpened EDM (Figures 4A and 4C). By contrast, synthetic density maps generated with GlycoSHIELD faithfully accounted for the unsharpened EDM, including that of the hyperantigenic NTD that is missing in the majority of reported SARS-CoV-2 S protein structures (Figures 4B and 4D). Accordingly, local correlation between the experimental and the synthetic map voxels was overall high throughout the hitherto unassigned volume surrounding the S protein body EDM (Figures 4E and 4F).

As mentioned above, glycan densities are typically resolvable only partially and only in the immediate proximity of protein cores, reflecting the high intrinsic dynamics of glycan chains (Figure S7A). Nevertheless, there are examples of larger and well-defined N-glycan structures, probably reflecting particularly strong constraints of the protein on glycan conformation. To assess to what extent GlycoSHIELD recapitulated such structures, we turned to a feline coronavirus (FIPV) S protein where N-glycans were particularly well resolved.[28] Focusing on glycans at N949 (four resolved monosaccharides) and N1218 (seven resolved monosaccharides), we aligned GlycoSHIELD conformers to manually built structures elucidated in this previous study (see STAR Methods) and quantified Q scores[63] to assess agreement with the EDM (see STAR Methods). Despite clear differences in specific sections of the N-glycan structures that were considered (Figures 4G and 4I), agreement between experimentally defined and GlycoSHIELD conformers was overall good considering the high number of permitted degrees-of-freedom when building glycan structures into ECM contours.

(F) Correlation with EDM across all GS map voxels from position 1 to position 4 indicated in (A).

In (A)–(E), note the significant volume difference between protein structure and experimental EDM and the good agreement between the experimental and modeled EDMs.

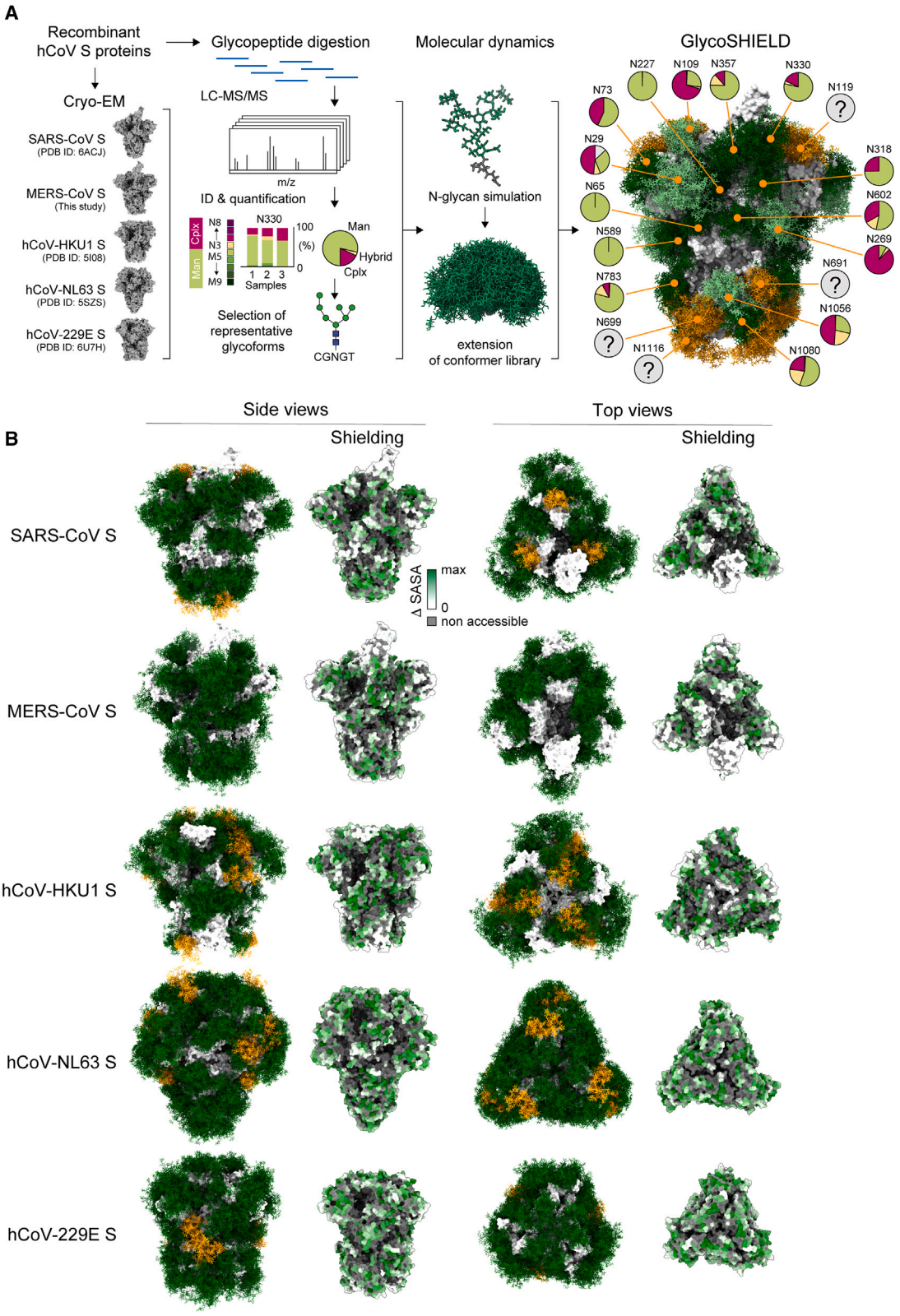
(G–J) Agreement of GS conformer ensembles with experimentally resolved N-glycan structures and refinement of GS models obtained for two N-glycans of feline infectious peritonitis virus (FIPV) S protein (PDB: 6JX7[28]: N949) (4 resolved monosaccharides, G and H) and N1218 (7 resolved monosaccharides, I and J).

(G and I) Left: manually built "reference" N-glycan structures (PDB) at these two positions.[28] Middle: examples of GS conformers aligned to reference structures (see STAR Methods). Monosaccharides are colored according to local correlation with the EDM (Q score heat scale, see inset) and shown with high-threshold EDM contours. Right: superposition of GS conformers (black outlines) and manually built N-glycan structures (Asn side chain is shown in yellow, monosaccharides in green) with low-threshold EDM contours.

(H and J) Refinement of GS density maps by maximization of fit with EDMs using GlycoDENSITY (see STAR Methods). Shown are overlays of EDMs (black mesh) and GS density maps (orange meshes) before (GS) and after refinement (ref. GS).

In (G)–(J), note the overall good agreement of GS conformers with the EDM and the improvement of GS density maps after reweighting of GS conformer ensembles based on local agreement with cryo-EM density.

See also Figure S7.



(legend on next page)

Excepting cases where glycan dynamics is restricted by the protein surface, glycan densities are typically observed as extended "mushrooms" (see Figures 4A and 4B). Although the contours of these densities are too loose to allow for the placement of single glycan conformers, they contain information about glycan conformational dynamics. Capitalizing on this, we developed an algorithm to weight grafted glycan conformer ensembles based on local agreement with EDMs and thereby refine synthetic density maps (see description of GlycoDENSITY in STAR Methods). We found that refined maps closely matched glycan EDMs, not only for the well-resolved glycan structures described above (see Figures 4H and 4J), but also for a more typical glycan at position N491 (Figure S7), where only two GlcNAcs at the glycan stem were originally matched to the EDM (Figure S7B). Interestingly, narrowing of the contours of the refined synthetic density revealed the probable position of three additional mannose moieties of this N-glycan in the EDM (Figures S7C and S7D).

Taken together, these data provide direct experimental evidence that the span and morphology of shields modeled by GlycoSHIELD are realistic and that GlycoSHIELD may be used to enrich existing cryo-EM datasets and aid the elucidation of glycoprotein structures.

GlycoSHIELD adds depth to cryo-EM and glycoproteomics analyses of coronavirus glycan camouflage

Combined analyses of protein structures and glycan compositions by cryo-EM and MS provide important information on glycan impact on protein accessibility.[28,46,64–67] Yet, specific glycan trees identified by MS are typically modeled on protein structure as single glycan conformers,[28,46,64] providing only partial information on protein shielding. Due to the prohibitive cost of MDS, the span and dynamics of glycan shields have been so far characterized by extensive simulations only for a few proteins, notably for SARS-CoV-2 S protein due to its immediate relevance to the COVID-19 pandemic.[3,6,20] To characterize the structural impact of glycans on hCoV S proteins more systematically, we combined GlycoSHIELD with MS and cryo-EM for comparison of the S protein of five coronaviruses: SARS-CoV,[68] MERS-CoV (this study), hCoV-HKU1,[69] hCoV-NL63,[70] and hCoV-229E[71] (Figure 5A). The receptor-binding domains (RBDs) of SARS-CoV-2 switch between a downward, closed conformation, and an upward, open conformation. Only the open conformation is accessible to receptor ACE2 binding.[58,62] In the case of hCoV-HKU1, hCoV-NL63, and hCoV-229E S proteins, the RBDs in the respective cryo-EM structures were captured in the downward conformations (Figure 5B). By contrast, one of the three RBDs of SARS-CoV and MERS-CoV S proteins adopted the upward conformation (Figure 5B). Quantitative MS glycopeptide analyses (see STAR Methods) indicated that these S proteins displayed a high level of glycan heterogeneity, both across N-glycosylation sites and at individual sites (Figures 5A and S8). The 17 most frequent N-glycans identified among the five S proteins were used to reconstruct representative glycoforms of these proteins. In all instances, the span of glycan covers was large, leaving only small portions of protein surfaces fully accessible (Figure 5B). Interestingly, glycan shields covered the two alphacoronavirus (hCoV-NL63 and hCoV-229E) almost entirely (Figure 5B). The shields of the three betacoronavirus (SARS-CoV, MERS-CoV, and hCoV-HKU1) were relatively less extended, leaving larger surfaces of the protein apex accessible (Figure 5B). Consistent with predictions made for SARS-CoV-2,[6] the elevated RBD of SARS-CoV and MERS-CoV S protein remained relatively unshielded and available for interactions with cellular receptors and antibodies (Figure 6A).

The 5-*N*-acetyl neuraminic acid (NeuAc) binding site of MERS-CoV S protein was clearly accessible. By contrast, the equivalent 9-*O*-acetylated sialic acid (9-O-Ac-Sia) binding site of hCoV-HKU1 seemed less accessible (Figures 6B and 6C). Interestingly, a previous study has shown that removal of hCoV-HKU1 S protein glycans at N29 and N251 increases sialic acid binding,[72] thus indicating that corresponding N-glycans hinder sialic acid access to the protein. Indeed, this difference in accessibility between wild-type and mutant hCoV-HKU1 was well predicted by GlycoSHIELD (Figure 6C), thus providing a plausible structural explanation for the known effect of these mutations.

Thus, combining GlycoSHIELD with MS and cryo-EM provides missing information on glycan impact on protein accessibility.

GlycoSHIELD provides insights into the function of an essential ionotropic neurotransmitter receptor

N-glycosylation is particularly marked in the brain.[11,73] Yet, little information is available on the structural impact of glycans on neuronal proteins and in particular neurotransmitter receptors. As a first step toward filling this gap, we focused on $GABA_A$ receptors, the target of drugs widely used in the clinic, including benzodiazepines and general anesthetics, whose structure has been extensively characterized in its non-glycosylated form.[21] We first focused on a homopentameric structure obtained with human β3 subunits.[23] The receptor was modeled with Man5, one of the two most prevalent glycotypes observed for $GABA_A$ receptors in the mammalian brain.[10] Resulting models showed that glycan shields masked a significant fraction of the protein surface, partially occluding access to the agonist (benzamidine [BZ]) binding sites of the receptor (Figures 7A and 7B).

Next, we focused on an $\alpha_1\beta_3\gamma_2$ $GABA_A$ heteropentamer, the generic form of the receptor in the mammalian brain.[22] The receptor was modeled with Man9 to mimic the glycosylation status

Figure 5. GlycoSHIELD adds missing dimension to cryo-EM and MS analyses of coronavirus glycan camouflage

(A) S proteins from 5 coronaviruses were analyzed by cryo-EM (previous and this study) and MS for identification and quantification of N-glycan types and positions. A set of 17 most representative N-glycans identified across all 5 S proteins was selected for MDS and added to GlycoSHIELD library. Shades of green in the glycosylated protein model shown in the right-hand panel indicate distinct N-glycan types (i.e., high-mannose vs. complex N-glycan types). Unresolved glycan types were modeled with Man5 and colored in orange.

(B) Side and top views and proteins with glycan shields and 3D maps of surface shielding (ΔSASA). Models of N-glycans identified by MS are colored in green. Unresolved glycan types are modeled with Man5 and colored in orange. ΔSASA scales are shown as white-to-green gradients. Note the almost complete coverage of NL63 and 229E S protein surface and, in increasing order, the larger accessible areas at the apex of HKU1, MERS, and SARS S proteins.

See also Figure S8.

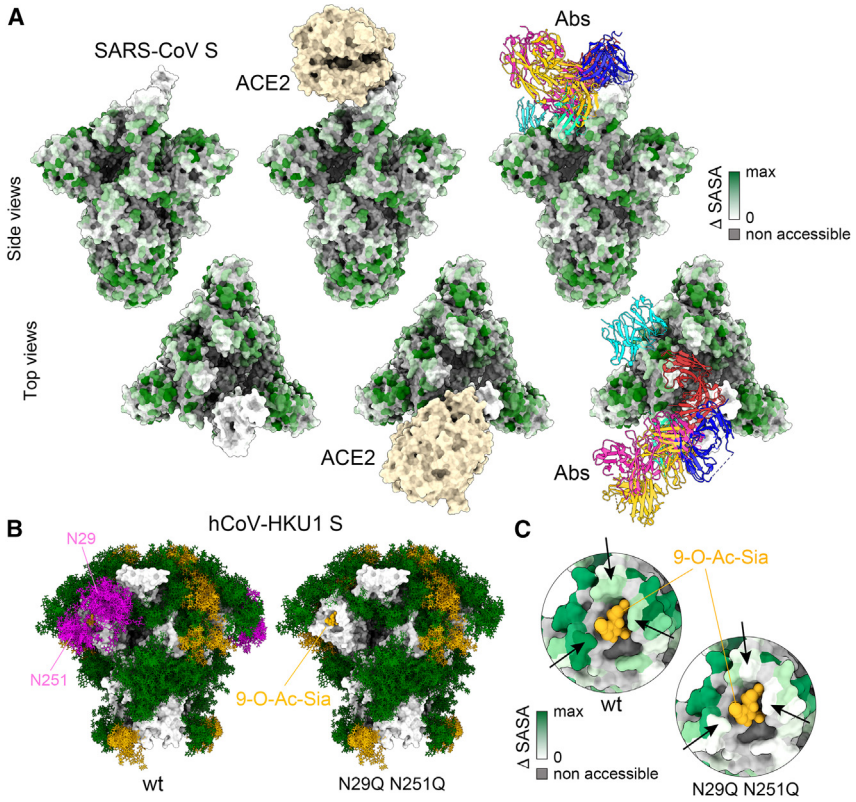


Figure 6. GlycoSHIELD provides structural insights into S protein binding to interacting molecules

(A) SARS-CoV S protein shielding 3D maps (ΔSASA, side and top views, white-to-green gradients) shown for S protein alone and for the protein bound to ACE2 or antibodies. Note the binding of ACE2 and antibodies to the unshielded area of S protein.

(B and C) Glycan shields (B) and shielding 3D maps (ΔSASA, C) of wild-type and N29Q-N251Q (C) hCoV-HKU1 S protein. (B) Glycans unresolved by MS are shown in orange. Glycans resolved by MS are shown in green or magenta. N29 and N251 glycans are shown in magenta. 9-O-acetylated sialic acid (9-O-Ac-Sia) bound to the protein is shown in orange. (C) ΔSASA 3D maps shown at higher magnification. Note the higher accessibility of the sialic acid binding pocket in HKU1 N29Q-N251Q double mutant (arrows).

of the receptor used for structural elucidation as this receptor was produced while blocking N-glycan maturation with Kifunensine.[22]

As observed for the β3 homopentamer, N-glycan shields occupied large volumes around the receptor (Figure 7C). The two GABA-binding sites of the receptor were largely accessible (Figure 7D). By contrast, and as seen for the BZ binding site of the homopentameric receptor (Figure 7B), access to the alprazolam (Xanax) binding site was partially masked by glycans, suggesting that glycans may regulate the binding of benzodiazepines (Figure 7D).

A common feature of functional $GABA_A$ receptors in the brain is the presence of two α subunits per pentamer. These subunits all harbor one N-glycan facing the inner side of the channel vestibule, imposing a strong constraint on receptor assembly and composition (i.e., two α subunits per receptor and α subunits at opposite sides of the main axis of the molecule).[22] The stems of these N-glycans resolved in the original structure seem to interact with the interior of the vestibule,[22] thus probably limiting glycan mobility. To account for this, we grafted and aligned N-glycans to the glycan stems that were experimentally resolved in the original structure (see STAR Methods). Resulting models indicated that, despite their reduced mobility, N-glycans may occupy up to 45% of the volume of the channel lumen (Figures 7E–7G), thus potentially limiting ion diffusion through the receptor. As previously reported,[22] a lateral tunnel was visible at one of the two α_1-β_3 subunit interfaces and was located above the plasma membrane and below glycan shields protruding in the central channel (Figure 7F).

These results thus indicate that N-glycans may impact the physiology of $GABA_A$ receptors by favoring ion conduction through a lateral conduction tunnel, raising exciting possibilities regarding the inner workings of the receptor.

DISCUSSION

This study shows that GlycoSHIELD provides a straightforward and accessible approach for obtaining quantitative information on glycoprotein morphology and structural dynamics.

This reductionist approach greatly reduces required computing resources, time, and technical know-how. Yet, by design, it ignores non-steric interactions of glycans with the protein and reciprocal interactions between glycans. Glycans modeled with GlycoSHIELD are, therefore, not expected to recapitulate specific glycan-protein binding, as occurs, for example, in glycan binding to lectins[32] or S proteins.[5,72] However, reconstituted glycan shields provide a probabilistic view of glycan positions integrated over the nearly complete ensemble of possible conformations and, as shown here, are very similar to shields predicted by full MDS. Thus, although GlycoSHIELD cannot substitute for extended MDS of complete systems, it fills an important gap by enabling expert and non-expert users to generate reliable models of glycan shields for the overwhelming majority of glycoproteins for which such shields have not been modeled yet.

Importantly, glycan conformational dynamics and glycan binding to cognate binding sites are co-dependent and should not be perceived as opposite processes. Indeed, as recently shown with GlycoSHIELD and full MDS, high-conformational freedom is a prerequisite for efficient glycan capture by processing enzymes.[74,75] Tools exist to predict protein glycan-binding sites.[76] It will be interesting to combine such tools with GlycoSHIELD to better understand how conformational exploration and glycan transient immobilization and modification determine one another.

By combining GlycoSHIELD models with unsharpened cryo-EM density maps, we provide direct experimental evidence that the span and morphology of shields modeled by

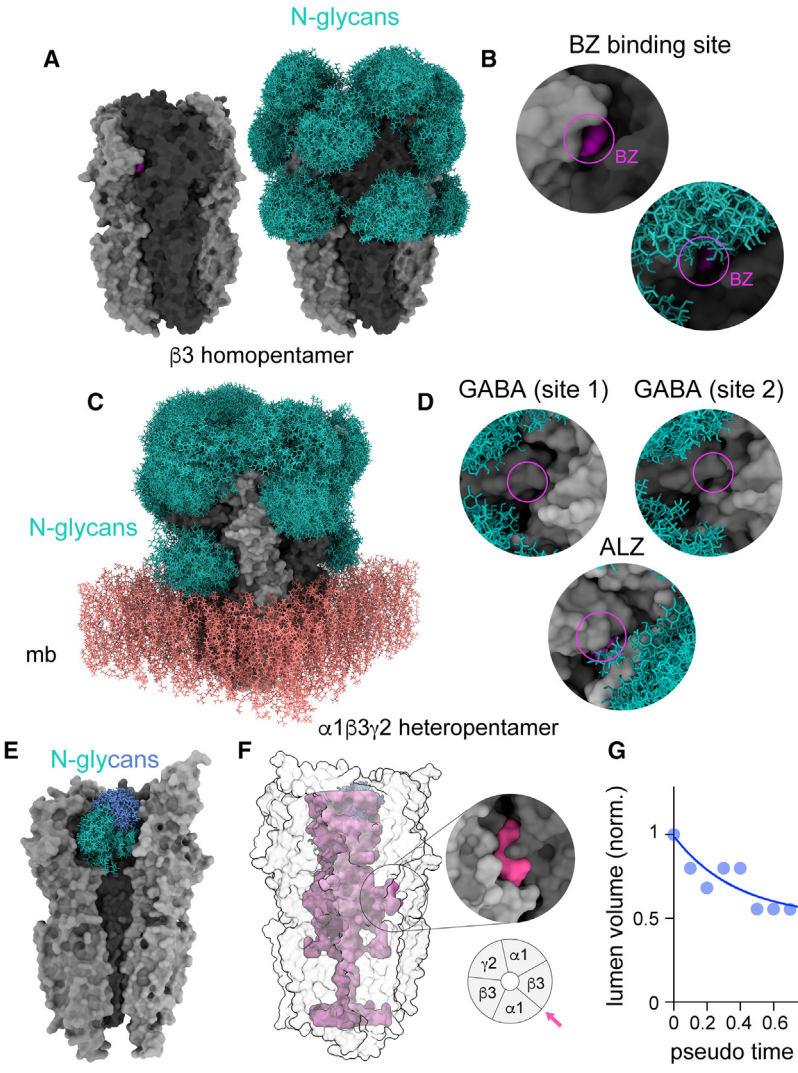


Figure 7. Glycans sculpt the outer surface and the inner vestibule of $GABA_A$ receptors

(A) Structure of a homopentameric $GABA_A$ receptor (β_3 subunits, gray)[23] shown with and without reconstituted glycan shields (Man5, green).

(B) Higher magnification of one agonist binding site of the receptor (benzamidine or BZ, magenta) shown without or with glycans (green). Note the partial occlusion of the BZ binding site.

(C) Structure of a heteropentameric $GABA_A$ receptor ($\alpha_1\beta_3\gamma_2$ subunits, gray)[22] shown with glycans (Man9, green) in a patch of membrane (mb, pink).

(D) Higher magnification of the three agonist binding sites of the glycosylated receptor (magenta circles; the 2 GABA sites and single alprazolam or ALZ binding sites are in close configuration).

(E) Inner vestibule of the $\alpha_1\beta_3\gamma_2$ receptor (exposed after removal of 2 subunits) with the N139 glycan shields (Man9, blue or green) of the two α_1 subunits.

(F) Inner cavity of the receptor (magenta volume). Upper inset: higher magnification of the lateral opening of the receptor cavity (magenta) on the protein surface (gray). Lower inset: position of the lateral opening in the receptor (top view).

(G) Variation of accessible inner cavity volume as a function of glycan conformer number (1–10 ns in pseudotime, i.e., 1–10 conformers per glycosylated site).

GlycoSHIELD are realistic. From a broader perspective, this demonstrates that structural information is typically lost during processing of glycosylated protein in cryo-EM analysis due to the high-conformational dynamics of glycans. This information may be, at least partially, recovered with GlycoSHIELD.

Impressive progress has been made in predicting protein structure,[15,17,18] enabling structure prediction for vast proportions of reported proteomes.[16] However, these approaches largely ignore glycan contribution to protein conformation.[77] Quantitative insights on glycan dynamics will thus be instrumental to improve structural predictions made for the large fraction of the proteome that is glycosylated. To facilitate this, we developed GlycoALPHAFOLD (see STAR Methods) to enable automated glycosylation of protein structures available in the AlphaFold2 database.

The important research focus on SARS-CoV-2 over the past 3 years has provided compelling evidence that N-glycans are key determinants of virus recognition by the immune system and entry into host cells.[20] How glycans impact the structure and function of other CoV S proteins is less understood. Here we provide realistic structural models for five related S proteins, contributing to a better understanding of how enveloped viruses maximize their glycan camouflage while leaving important binding sites accessible for interaction with cognate cellular receptors. Yet, the shielding of specific protein segments does not imply that binding partners with strong binding affinities cannot overcome the entropic cost and displace glycans to access these segments. The argument can be made, however, that glycan explore their conformation space at rates that are orders of magnitude faster than, say, the rotational and translational diffusion of an IgG.[78] In other words, although only a single glycan conformer may exist at a given time point, "observers" perceiving their environment at the timescale of protein motion will "see" glycans as blurred and delocalized shields. Glycan mass content at any given point in space will consequently be relatively low, but this mass will be distributed and limit physical interactions over a relatively large surface.

Through extended MS analysis, we show that the glycosylation profiles of S proteins may be more diverse than previously anticipated. Previous studies required tens of millions of computing hours to characterize the glycan shields of single S protein glycoforms.[3,4] Complete analysis of the hundreds of possible glycoforms of such proteins by standard MDS would be prohibitively costly. This study provides proof of principle that GlycoSHIELD constitutes a reliable proxy to do so at a simplified level but at a fraction of the computing cost.

It has long been recognized that neurotransmitter receptors are extensively N-glycosylated.[73] The impact of N-glycans on

the structure and function of these proteins is, however, largely unknown. Here we provide models of fully glycosylated $GABA_A$ receptors, the main fast inhibitory receptor of the mammalian brain. These models predict that N-glycans may occupy a much larger fraction of the channel vestibule than previously anticipated,[22] possibly favoring ion conduction through lateral tunnels located below glycans. Intriguingly, the top opening of the channel is largely obstructed in EDMs obtained at synapses, while lateral openings are clearly visible.[79] This thus suggests that these tunnels may be present in native receptors *in vivo*. Consistent with this, recent MDSs and electrophysiological recordings indicate that such lateral openings also exist in the closely related glycine receptor and constitute the main conduit for chloride permeation in the channel vestibule.[80]

Tools exist to add single glycan conformers onto existing protein structures.[81–83] Yet, these do not fully account for possible glycan conformational space and, when used for system preparation for MDS, may be biased and excessively constrained by initial glycan conformation. Selecting glycan conformers generated by GlycoSHIELD may alleviate this problem, allowing multiple locally equilibrated structures to be used as replicates of a given glycosylated site.

As shown in particular for SARS-CoV-2 and other CoVs, naturally occurring mutations may influence virus cellular tropism.[71] This may, in turn, affect virus glycosylation status and recognition by the immune system. The issue is also critical for cancer immunotherapy as cancers are typically associated with aberrant protein glycosylation.[8,84] Tools enabling rapid assessment of the impact of diverse glycans on protein shielding will thus likely be of value in this context as well. As illustrated by RosettaAntibodyDesign,[85] dynamic structural modeling greatly accelerates the design and generation of custom antibodies. Because of its modular and open-source structure, GlycoSHIELD may easily be integrated into such pipelines, thus adding an important but yet still missing dimension to *in silico* screening of antigens and antibodies.

Limitations of the study

Here, we provide experimental evidence that GlycoSHIELD generates realistic structural models at a fraction of the computing cost of conventional MDSs. As described above, some predictions derived from these models can be compared and are consistent with experimental data. Others, for which such data are not available yet, thus remain theoretical and will need to be ascertained experimentally.

Although a similar simulation setup has been shown to provide complete sampling of glycan conformation space,[38] we cannot rule out that residual bias may still impact conformer ensembles as a result of imperfect force fields, choice of water model, or peptide protonation states. Available and emerging enhanced sampling methods and clustering algorithms[38] may allow improving GlycoSHIELD further.

Most of the shielding maps that are described here were generated with static protein frames. Although we show for a few examples that this may not always be required, shielding predictions may be improved by taking protein conformational ensembles into consideration. Significant progress has been made in deciphering protein conformational dynamics from X-ray or cryo-EM experiments[86,87] and *ab initio* predictions.[88] A more comprehensive description of glycan shielding will require taking these datasets into account (as is already possible in GlycoSHIELD), in particular for glycans located close to flexible protein domains.

Despite good agreement with glycoprotein structures elucidated at low and medium resolution, GlycoSHIELD is limited by the fact that it only takes into account glycan-protein interactions at a very coarse level and ignores glycan-glycan interactions. Overcoming these limitations will probably require the development of more exhaustive approaches including machine learning to combine heuristics derived from the mining of cryo-EM datasets and the simulation of larger "elementary" protein-glycan blocks and a broader repertoire of complete glycoprotein structures.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS
  - Bacteria
  - Cell lines
- METHOD DETAILS
  - Molecular dynamics simulations
  - GlycoSHIELD software suite
  - Ray analysis of protein surface accessibility
  - Solvent accessible surface area (SASA)
  - Generation and refinement of synthetic density maps
  - 3D structure rendering
  - Expression and purification of S proteins
  - Glycan analysis by mass spectrometry
  - Cryo-EM structural analyses
  - Expression and purification of recombinant N-cadherin variants
  - Size-exclusion chromatography-coupled small-angle X-ray scattering (SEC-SAXS)
  - Molecular volume calculations
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2024.01.034.

ACKNOWLEDGMENTS

We thank IDRIS for the allocation of HPC resources on the Jean Zay supercomputer (allocations #2020-AP010711998 and #2021-A0100712343 to C.H.). We thank the Max Planck Computing and Data Facility for providing computational resources and support for the web application, and the Max Planck Society for support. We thank the North German Supercomputing Alliance (HLRN) for the allocation of HPC resources (project hbb00001). We are indebted to Radu Aricescu for his encouragement and his insightful comments

on the manuscript. We thank Erin Schuman, Ludger Johannes, Michael Ehlers, and Tomek Wlodarski for their comments on the manuscript. We thank the Academia Sinica Biophysics Core Facility (AS-CFII-111-201), Academia Sinica Common Mass Spectrometry Facilities (AS-CFII-111-209), and Academia Sinica Cryo-EM Center (AS-CFII-111-210) for data collection, all of which are funded by the Academia Sinica Core Facility and Innovative Instrument Project. Taiwan Protein Project (AS-KPQ-109-TPP2) is acknowledged for supporting the Academia Sinica Cryo-EM Center. We thank the mammalian cell culture facility of the Institute of Biological Chemistry, Academia Sinica, for supporting recombinant protein production. We thank Camille Penet (IPNP, Paris) for assistance with system preparation for MDSs. Work in the laboratories of S.-T.D.H. and K.-H.K. is supported by Academia Sinica intramural funds to S.-T.D.H. and K.-H.K., an Academia Sinica Career Development Award, Academia Sinica to S.-T.D.H. (AS-CDA-109-L08), the Infectious Disease Research Supporting Grants to S.-T.D.H. (AS-IDR-110-08, AS-IDR-111-03, and AS-IDR-112-04), and the Ministry of Science and Technology (MOST), Taiwan (MOST 109-3114-Y-001-001, MOST 110-2113-M-001-050-MY3, and MOST 110-2311-B-001-013-MY3) to S.-T.D.H. Work in the laboratory of C.H. is supported by the Agence Nationale de la Recherche (grants #ANR-16-CE16-0009-01 and #ANR-21-CE16-0021-01) and the Foundation NRJ. M.S.: project financed under Dioscuri, a program initiated by the Max Planck Society, jointly managed with the National Science Centre in Poland, and mutually funded by the Polish Ministry of Education and Science and German Federal Ministry of Education and Research (UMO-2021/03/H/NZ1/00003), with additional support by the Fonds zur Förderung der wissenschaftlichen Forschung Schrödinger fellowship (J4332-B28) and Clusterproject ENABLE funded by the Hessian Ministry for Science and the Arts. The open-access publication of this article was funded by the Priority Research Area BioS under the program "Excellence Initiative – Research University" at the Jagiellonian University in Krakow, Poland.

AUTHOR CONTRIBUTIONS

Y.-X.T., N.-E.C., H.-T.C., T.-J.Y., S.v.B., V.S., N.Z., M.T., M.G., I.L.G., L.C.C., Y.-S.W., M.-F.H., K.-H.K., C.H., and M.S. prepared samples, collected data, performed analyses, and edited the manuscript. K.-H.K., L.C.C., G.H., S.-T.D.H., and C.H. provided resources and data. S.-T.D.H., C.H., and M.S. designed the project and wrote the manuscript.

DECLARATION OF INTERESTS

S.v.B., G.H., and M.S. have filed a patent application related to this work: European Patent Application No. 22203671.7 (2022). S.-T.D.H., C.H., and M.S. have filed a patent application related to this work: European Patent Application No. 23307151.3 (2023).

Received: October 25, 2022
Revised: October 18, 2023
Accepted: January 22, 2024
Published: February 29, 2024

REFERENCES

1. Yin, H., and Flynn, A.D. (2016). Drugging Membrane Protein Interactions. Annu. Rev. Biomed. Eng. *18*, 51–76.
2. Moremen, K.W., Tiemeyer, M., and Nairn, A.V. (2012). Vertebrate protein glycosylation: diversity, synthesis and function. Nat. Rev. Mol. Cell Biol. *13*, 448–462.
3. Casalino, L., Gaieb, Z., Goldsmith, J.A., Hjorth, C.K., Dommer, A.C., Harbison, A.M., Fogarty, C.A., Barros, E.P., Taylor, B.C., McLellan, J.S., et al. (2020). Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. ACS Cent. Sci. *6*, 1722–1734.
4. Turoňová, B., Sikora, M., Schürmann, C., Hagen, W.J.H., Welsch, S., Blanc, F.E.C., von Bülow, S., Gecht, M., Bagola, K., Hörner, C., et al. (2020). In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. Science *370*, 203–208.
5. Mehdipour, A.R., and Hummer, G. (2021). Dual nature of human ACE2 glycosylation in binding to SARS-CoV-2 spike. Proc. Natl. Acad. Sci. USA *118*, e2100425118.
6. Sikora, M., von Bülow, S., Blanc, F.E.C., Gecht, M., Covino, R., and Hummer, G. (2021). Computational epitope map of SARS-CoV-2 spike protein. PLOS Comput. Biol. *17*, e1008790.
7. Rutledge, J., Oh, H., and Wyss-Coray, T. (2022). Measuring biological age using omics data. Nat. Rev. Genet. *23*, 715–727.
8. Costa, A.F., Campos, D., Reis, C.A., and Gomes, C. (2020). Targeting Glycosylation: A New Road for Cancer Drug Discovery. Trends Cancer *6*, 757–766.
9. Khoo, K.H. (2019). Advances toward mapping the full extent of protein site-specific O-GalNAc glycosylation that better reflects underlying glycomic complexity. Curr. Opin. Struct. Biol. *56*, 146–154.
10. Riley, N.M., Hebert, A.S., Westphall, M.S., and Coon, J.J. (2019). Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. Nat. Commun. *10*, 1311.
11. Williams, S.E., Noel, M., Lehoux, S., Cetinbas, M., Xavier, R.J., Sadreyev, R.I., Scolnick, E.M., Smoller, J.W., Cummings, R.D., and Mealer, R.G. (2022). Mammalian brain glycoproteins exhibit diminished glycan complexity compared to other tissues. Nat. Commun. *13*, 275.
12. Zhang, Q., Ma, C., Chin, L.S., and Li, L. (2020). Integrative glycoproteomics reveals protein N-glycosylation aberrations and glycoproteomic network alterations in Alzheimer's disease. Sci. Adv. *6*, eabc5802.
13. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303.
14. Pieper, U., Eswar, N., Webb, B.M., Eramian, D., Kelly, L., Barkan, D.T., Carter, H., Mankoo, P., Karchin, R., Marti-Renom, M.A., et al. (2009). MODBASE, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. *37*, D347–D354.
15. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.
16. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., et al. (2021). Highly accurate protein structure prediction for the human proteome. Nature *596*, 590–596.
17. Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science *373*, 871–876.
18. Lin, Z., Akin, H., Rao, R., Hie, B., Zhu, Z., Lu, W., Smetanin, N., Verkuil, R., Kabeli, O., Shmueli, Y., et al. (2023). Evolutionary-scale prediction of atomic-level protein structure with a language model. Science *379*, 1123–1130.
19. Troyanovsky, S.M. (2023). Adherens junction: the ensemble of specialized cadherin clusters. Trends Cell Biol. *33*, 374–387.
20. Chawla, H., Fadda, E., and Crispin, M. (2022). Principles of SARS-CoV-2 glycosylation. Curr. Opin. Struct. Biol. *75*, 102402.
21. Scott, S., and Aricescu, A.R. (2019). A structural perspective on GABAA receptor pharmacology. Curr. Opin. Struct. Biol. *54*, 189–197.
22. Masiulis, S., Desai, R., Uchański, T., Serna Martin, I.S., Laverty, D., Karia, D., Malinauskas, T., Zivanov, J., Pardon, E., Kotecha, A., et al. (2019). GABAA receptor signalling mechanisms revealed by structural pharmacology. Nature *565*, 454–459.
23. Miller, P.S., and Aricescu, A.R. (2014). Crystal structure of a human GABA A receptor. Nature *512*, 270–275.
24. Yang, T.-J., Yu, P.-Y., Chang, Y.-C., and Hsu, S.-T.D. (2021). D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. J. Biol. Chem. *297*, 101238.

25. Berndsen, Z.T., Chakraborty, S., Wang, X., Cottrell, C.A., Torres, J.L., Diedrich, J.K., López, C.A., Yates, J.R., van Gils, M.J., Paulson, J.C., et al. (2020). Visualization of the HIV-1 Env glycan shield across scales. Proc. Natl. Acad. Sci. USA *117*, 28014–28025.

26. Huang, C.-Y., Draczkowski, P., Wang, Y.-S., Chang, C.-Y., Chien, Y.-C., Cheng, Y.-H., Wu, Y.-M., Wang, C.-H., Chang, Y.-C., Chang, Y.-C., et al. (2022). In situ structure and dynamics of an alphacoronavirus spike protein by cryo-ET and cryo-EM. Nat. Commun. *13*, 4877.

27. Watanabe, Y., Berndsen, Z.T., Raghwani, J., Seabright, G.E., Allen, J.D., Pybus, O.G., McLellan, J.S., Wilson, I.A., Bowden, T.A., Ward, A.B., et al. (2020). Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. Nat. Commun.. 11, 2688.

28. Yang, T.-J., Chang, Y.-C., Ko, T.-P., Draczkowski, P., Chien, Y.-C., Chang, Y.-C., Wu, K.-P., Khoo, K.-H., Chang, H.-W., and Hsu, S.-T.D. (2020). Cryo-EM analysis of a feline coronavirus spike protein reveals a unique structure and camouflaging glycans. Proc. Natl. Acad. Sci. USA *117*, 1438–1446.

29. Sztain, T., Ahn, S.-H., Bogetti, A.T., Casalino, L., Goldsmith, J.A., Seitz, E., McCool, R.S., Kearns, F.L., Acosta-Reyes, F., Maji, S., et al. (2021). A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat. Chem. *13*, 963–968.

30. Tams, J.W., Vind, J., and Welinder, K.G. (1999). Adapting protein solubility by glycosylation: N-Glycosylation mutants of Coprinus cinereus peroxidase in salt and organic solutions. Biochim. Biophys. Acta *1432*, 214–221.

31. Bänfer, S., and Jacob, R. (2022). Galectins. Curr. Biol. *32*, R406–R408.

32. Dingjan, T., Imberty, A., Pérez, S., Yuriev, E., and Ramsland, P.A. (2017). Molecular Simulations of Carbohydrates with a Fucose-Binding Burkholderia ambifaria Lectin Suggest Modulation by Surface Residues Outside the Fucose-Binding Pocket. Front. Pharmacol. *8*, 393.

33. Kearns, F.L., Sandoval, D.R., Casalino, L., Clausen, T.M., Rosenfeld, M.A., Spliid, C.B., Amaro, R.E., and Esko, J.D. (2022). Spike-heparan sulfate interactions in SARS-CoV-2 infection. Curr. Opin. Struct. Biol. *76*, 102439.

34. Erbel, P.J.A., Karimi-Nejad, Y., van Kuik, J.A., Boelens, R., Kamerling, J.P., and Vliegenthart, J.F.G. (2000). Effects of the N-Linked Glycans on the 3D Structure of the Free α-Subunit of Human Chorionic Gonadotropin. Biochemistry *39*, 6012–6021.

35. Wormald, M.R., and Dwek, R.A. (1999). Glycoproteins: glycan presentation and protein-fold stability. Structure *7*, R155–R160.

36. Pietrek, L.M., Stelzl, L.S., and Hummer, G. (2023). Structural ensembles of disordered proteins from hierarchical chain growth and simulation. Curr. Opin. Struct. Biol. *78*, 102501.

37. Turupcu, A., and Oostenbrink, C. (2017). Modeling of Oligosaccharides within Glycoproteins from Free-Energy Landscapes. J. Chem. Inf. Model. *57*, 2222–2236.

38. Grothaus, I.L., Bussi, G., and Colombi Ciacchi, L. (2022). Exploration, Representation, and Rationalization of the Conformational Phase Space of N-Glycans. J. Chem. Inf. Model. *62*, 4992–5008.

39. Rosenau, J., Grothaus, I.L., Yang, Y., Kumar, N.D., Ciacchi, L.C., Kelm, S., and Waespy, M. (2022). N-glycosylation modulates enzymatic activity of Trypanosoma congolense trans-sialidase. J. Biol. Chem. *298*, 102403.

40. Sotomayor, M., and Schulten, K. (2008). The Allosteric Role of the Ca2+ Switch in Adhesion and Elasticity of C-Cadherin. Biophys. J. *94*, 4621–4633.

41. Blanchet, C.E., and Svergun, D.I. (2013). Small-Angle X-ray Scattering on Biological Macromolecules and Nanocomposites in Solution. Annu. Rev. Phys. Chem. *64*, 37–54.

42. Brosey, C.A., and Tainer, J.A. (2019). Evolving SAXS versatility: solution X-ray scattering for macromolecular architecture, functional landscapes, and integrative structural biology. Curr. Opin. Struct. Biol. *58*, 197–213.

43. Carvalho, S., Reis, C.A., and Pinho, S.S. (2016). Cadherins Glycans in Cancer: Sweet Players in a Bitter Process. Trends Cancer *2*, 519–531.

44. Rief, M., Gautel, M., Schemmel, A., and Gaub, H.E. (1998). The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. Biophys. J. *75*, 3008–3014.

45. Lawton, G. (2021). Sights set on universal vaccine. New Sci. *249*, 8–9.

46. Watanabe, Y., Allen, J.D., Wrapp, D., McLellan, J.S., and Crispin, M. (2020). Site-specific glycan analysis of the SARS-CoV-2 spike. Science *369*, 330–333.

47. A. Varki, R.D. Cummings, J.D. Esko, P. Stanley, G.W. Hart, M. Aebi, D. Mohnen, T. Kinoshita, N.H. Packer, and J.H. Prestegard, et al., eds. (2022). Essentials of Glycobiology, Fourth Edition (Cold Spring Harbor Laboratory Press).

48. Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T.H.O., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G.A., Jurczyszak, D., Polanco, J., et al. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat. Med. *26*, 1033–1036.

49. Tian, W., Li, D., Zhang, N., Bai, G., Yuan, K., Xiao, H., Gao, F., Chen, Y., Wong, C.C.L., and Gao, G.F. (2021). O-glycosylation pattern of the SARS-CoV-2 spike protein reveals an "O-Follow-N" rule. Cell Res. *31*, 1123–1125.

50. Sanda, M., Morrison, L., and Goldman, R. (2021). N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. Anal. Chem. *93*, 2003–2009.

51. Zhang, Y., Zhao, W., Mao, Y., Chen, Y., Wang, S., Zhong, Y., Su, T., Gong, M., Du, D., Lu, X., et al. (2021). Site-specific N-glycosylation Characterization of Recombinant SARS-CoV-2 Spike Proteins. Mol. Cell. Proteomics *20*, 100058.

52. Shajahan, A., Supekar, N.T., Gleinich, A.S., and Azadi, P. (2020). Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology *30*, 981–988.

53. Grant, O.C., Montgomery, D., Ito, K., and Woods, R.J. (2020). Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition. Sci. Rep. *10*, 14991.

54. Mugnai, M.L., Shin, S., and Thirumalai, D. (2023). Entropic contribution of ACE2 glycans to RBD binding. Biophys. J. *122*, 2506–2517.

55. Chakraborty, S., Berndsen, Z.T., Hengartner, N.W., Korber, B.T., Ward, A.B., and Gnanakaran, S. (2020). Quantification of the Resilience and Vulnerability of HIV-1 Native Glycan Shield at Atomistic Detail. iScience *23*, 101836.

56. Ghorbani, M., Brooks, B.R., and Klauda, J.B. (2021). Exploring dynamics and network analysis of spike glycoprotein of SARS-COV-2. Biophys. J. *120*, 2902–2913.

57. Kav, B., Grafmüller, A., Schneck, E., and Weikl, T.R. (2020). Weak carbohydrate–carbohydrate interactions in membrane adhesion are fuzzy and generic. Nanoscale *12*, 17342–17353.

58. Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., Nakane, T., Zivanov, J., Neufeldt, C.J., Cerikan, B., et al. (2020). Structures and distributions of SARS-CoV-2 spike proteins on intact virions. Nature *588*, 498–502.

59. Klein, S., Cortese, M., Winter, S.L., Wachsmuth-Melm, M., Neufeldt, C.J., Cerikan, B., Stanifer, M.L., Boulant, S., Bartenschlager, R., and Chlanda, P. (2020). SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography. Nat. Commun. *11*, 5885.

60. Mannar, D., Saville, J.W., Zhu, X., Srivastava, S.S., Berezuk, A.M., Tuttle, K.S., Marquez, A.C., Sekirov, I., and Subramaniam, S. (2022). SARS-CoV-2 Omicron variant: Antibody evasion and cryo-EM structure of spike protein–ACE2 complex. Science *375*, 760–764.

61. Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.-L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science *367*, 1260–1263.

62. Yang, T.-J., Yu, P.-Y., Chang, Y.-C., Liang, K.-H., Tso, H.-C., Ho, M.-R., Chen, W.-Y., Lin, H.-T., Wu, H.-C., and Hsu, S.-T.D. (2021). Effect of SARS-CoV-2 B.1.1.7 mutations on spike protein structure and function. Nat. Struct. Mol. Biol. *28*, 731–739.

63. Pintilie, G., Zhang, K., Su, Z., Li, S., Schmid, M.F., and Chiu, W. (2020). Measurement of atom resolvability in cryo-EM maps with Q-scores. Nat. Methods *17*, 328–334.

64. Kuo, C.-W., Yang, T.-J., Chien, Y.-C., Yu, P.-Y., Hsu, S.-T.D., and Khoo, K.-H. (2022). Distinct shifts in site-specific glycosylation pattern of SARS-CoV-2 spike proteins associated with arising mutations in the D614G and Alpha variants. Glycobiology *32*, 60–72.

65. Newby, M.L., Fogarty, C.A., Allen, J.D., Butler, J., Fadda, E., and Crispin, M. (2023). Variations within the Glycan Shield of SARS-CoV-2 Impact Viral Spike Dynamics. J. Mol. Biol. *435*, 167928.

66. Perrett, H.R., Brouwer, P.J.M., Hurtado, J., Newby, M.L., Liu, L., Müller-Kräuter, H., Müller Aguirre, S., Burger, J.A., Bouhuijs, J.H., Gibson, G., et al. (2023). Structural conservation of Lassa virus glycoproteins and recognition by neutralizing antibodies. Cell Rep. *42*, 112524.

67. Allen, J.D., Ivory, D., Ge Song, S.G., He, W.T., Capozzola, T., Yong, P., Burton, D.R., Andrabi, R., and Crispin, M. (2022). The diversity of the glycan shield of sarbecoviruses closely related to SARS-CoV-2. Cell Rep. *42*, 112307.

68. Song, W., Gui, M., Wang, X., and Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. PLoS Pathog. *14*, e1007236.

69. Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M., Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., and Ward, A.B. (2016). Pre-fusion structure of a human coronavirus spike protein. Nature *531*, 118–121.

70. Walls, A.C., Tortorici, M.A., Frenz, B., Snijder, J., Li, W., Rey, F.A., DiMaio, F., Bosch, B.J., and Veesler, D. (2016). Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. Nat. Struct. Mol. Biol. *23*, 899–905.

71. Li, Q., Wu, J., Nie, J., Zhang, L., Hao, H., Liu, S., Zhao, C., Zhang, Q., Liu, H., Nie, L., et al. (2020). The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell *182*, 1284–1294.e9.

72. Hulswit, R.J.G., Lang, Y., Bakkers, M.J.G., Li, W., Li, Z., Schouten, A., Ophorst, B., van Kuppeveld, F.J.M., Boons, G.-J., Bosch, B.-J., et al. (2019). Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proc. Natl. Acad. Sci. USA *116*, 2681–2690.

73. Mathew, C., Weiß, R.G., Giese, C., Lin, C.W., Losfeld, M.-E., Glockshuber, R., Riniker, S., and Aebi, M. (2021). Glycan–protein interactions determine kinetics of N-glycan remodeling. RSC Chem. Biol. *2*, 917–931.

74. Mikolajczyk, K., Sikora, M., Hanus, C., Kaczmarek, R., and Czerwinski, M. (2022). One of the two N-glycans on the human Gb3/CD77 synthase is essential for its activity and allosterically regulates its function. Biochem. Biophys. Res. Commun. *617*, 36–41.

75. Bonnardel, F., Mariethoz, J., Pérez, S., Imberty, A., and Lisacek, F. (2021). LectomeXplore, an update of UniLectin for the discovery of carbohydrate-binding proteins based on a new lectin classification. Nucleic Acids Res. *49*, D1548–D1554.

76. Bagdonas, H., Fogarty, C.A., Fadda, E., and Agirre, J. (2021). The case for post-predictional modifications in the AlphaFold Protein Structure Database. Nat. Struct. Mol. Biol. *28*, 869–870.

77. García-Sánchez, M., Castro, M., and Faro, J. (2023). B cell receptors and free antibodies have different antigen-binding kinetics. Proc. Natl. Acad. Sci. USA *120*, e2220669120.

78. Scott, H., and Panin, V.M. (2014). The role of protein N-glycosylation in neural transmission. Glycobiology *24*, 407–417.

79. Liu, Y.-T., Tao, C.-L., Zhang, X., Xia, W., Shi, D.-Q., Qi, L., Xu, C., Sun, R., Li, X.-W., Lau, P.-M., et al. (2020). Mesophasic organization of GABAA receptors in hippocampal inhibitory synapses. Nat. Neurosci. *23*, 1589–1596.

80. Cerdan, A.H., Peverini, L., Changeux, J.-P., Corringer, P.-J., and Cecchini, M. (2022). Lateral fenestrations in the extracellular domain of the glycine receptor contribute to the main chloride permeation pathway. Sci. Adv. *8*, eadc9340.

81. Bohne-Lang, A., and von der Lieth, C.-W. (2005). GlyProt: in silico glycosylation of proteins. Nucleic Acids Res. *33*, W214–W219.

82. Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J. Comput. Chem. *29*, 622–655.

83. Park, S.J., Lee, J., Qi, Y., Kern, N.R., Lee, H.S., Jo, S., Joung, I., Joo, K., Lee, J., and Im, W. (2019). CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. Glycobiology *29*, 320–331.

84. Moffett, S., Shiao, T.C., Mousavifar, L., Mignani, S., and Roy, R. (2021). Aberrant glycosylation patterns on cancer cells: Therapeutic opportunities for glycodendrimers/metallodendrimers oncology. Wiley Interdiscip. Rev. Nanomed. Nanobiotechnol. *13*, e1659.

85. Adolf-Bryfogle, J., Kalyuzhniy, O., Kubitz, M., Weitzner, B.D., Hu, X., Adachi, Y., Schief, W.R., and Dunbrack, R.L. (2018). RosettaAntibodyDesign (RAbD): A general framework for computational antibody design. PLoS Comput. Biol. *14*, e1006112.

86. Nass, K., Gorel, A., Abdullah, M.M., V Martin, A., Kloos, M., Marinelli, A., Aquila, A., Barends, T.R.M., Decker, F.J., Bruce Doak, R., et al. (2020). Structural dynamics in proteins induced by and probed with X-ray free-electron laser pulses. Nat. Commun. *11*, 1814.

87. Punjani, A., and Fleet, D.J. (2023). 3DFlex: determining structure and motion of flexible proteins from cryo-EM. Nat. Methods *20*, 860–870.

88. Ellaway, J.I.J., Anyango, S., Nair, S., Zaki, H.A., Nadzirin, N., Powell, H.R., Gutmanas, A., Varadi, M., and Velankar, S. (2023). Automated pipeline for comparing protein conformational states in the PDB to AlphaFold2 predictions. https://doi.org/10.1101/2023.07.13.545008.

89. Harrison, O.J., Jin, X., Hong, S., Bahna, F., Ahlsen, G., Brasch, J., Wu, Y., Vendome, J., Felsovalyi, K., Hampton, C.M., et al. (2011). The Extracellular Architecture of Adherens Junctions Revealed by Crystal Structures of Type I Cadherins. Structure *19*, 244–256.

90. Park, Y.J., Walls, A.C., Wang, Z., Sauer, M.M., Li, W., Tortorici, M.A., Bosch, B.-J., DiMaio, F., and Veesler, D. (2019). Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. Nat. Struct. Mol. Biol. *26*, 1151–1157.

91. Zivanov, J., Nakane, T., and Scheres, S.H.W. (2020). Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. IUCrJ *7*, 253–267.

92. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

93. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221.

94. Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

95. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82.

96. Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., et al. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J. Appl. Crystallogr. *54*, 343–355.

97. Shih, O., Liao, K.-F., Yeh, Y.-Q., Su, C.-J., Wang, C.-A., Chang, J.-W., Wu, W.-R., Liang, C.-C., Lin, C.-Y., Lee, T.-H., et al. (2022). Performance

of the new biological small- and wide-angle X-ray scattering beamline 13A at the Taiwan Photon Source. J. Appl. Crystallogr. *55*, 340–352.

98. Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX *1–2*, 19–25.
99. Briones, R., Blau, C., Kutzner, C., de Groot, B.L., and Aponte-Santamaría, C. (2019). GROmaps: A GROMACS-Based Toolset to Analyze Density Maps Derived from Molecular Dynamics Simulations. Biophys. J. *116*, 4–11.
100. Gowers, R.J., Linke, M., Barnoud, J., Reddy, T.J.E., Melo, M.N., Seyler, S.L., Domański, J., Dotson, D.L., Buchoux, S., Kenney, I.M., et al. (2016). MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In Proceedings of the Python in Science Conference, pp. 98–105.
101. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.
102. Hunter, J.D. (2007). Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. *9*, 90–95.
103. Burnley, T., Palmer, C.M., and Winn, M. (2017). Recent developments in the CCP-EM software suite. Acta Crystallogr. Sect. J. Struct. Biol. *73*, 469–477.
104. Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinformatics *54*, 5.6.1–5.6.37.
105. Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. *29*, 1859–1865.
106. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926–935.
107. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B.L., Grubmüller, H., and MacKerell, A.D. (2017). CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nat. Methods *14*, 71–73.
108. Guvench, O., Hatcher, E.R., Venable, R.M., Pastor, R.W., and MacKerell, A.D., Jr. (2009). CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. J. Chem. Theor. Comput. *5*, 2353–2370.
109. Guvench, O., Mallajosyula, S.S., Raman, E.P., Hatcher, E., Vanommeslaeghe, K., Foster, T.J., Jamison, F.W.I., and MacKerell, A.D., Jr. (2011). CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. J. Chem. Theor. Comput. *7*, 3162–3180.
110. Hakansson-McReynolds, S., Jiang, S., Rong, L., and Caffrey, M. (2006). Solution Structure of the Severe Acute Respiratory Syndrome-Coronavirus Heptad Repeat 2 Domain in the Prefusion State. J. Biol. Chem. *281*, 11965–11971.
111. Jacquemyn, J., Cascalho, A., and Goodchild, R.E. (2017). The ins and outs of endoplasmic reticulum-controlled lipid biosynthesis. EMBO Rep. *18*, 1905–1921.
112. Evans, D.J., and Holian, B.L. (1985). The Nose–Hoover thermostat. J. Chem. Phys. *83*, 4069–4074.
113. Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. *52*, 7182–7190.
114. Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J. Chem. Phys. *126*, 014101.
115. Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R.A., et al. (2009). PLUMED: A portable plugin for free-energy calculations with molecular dynamics. Comput. Phys. Commun. *180*, 1961–1972.
116. Bussi, G. (2014). Hamiltonian replica exchange in GROMACS: a flexible implementation. Mol. Phys. *112*, 379–384.
117. Gil-Ley, A., and Bussi, G. (2015). Correction to Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. J. Chem. Theor. Comput. *11*, 5554.
118. Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. *18*, 1463–1472.
119. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. J. Chem. Phys. *81*, 3684–3690.
120. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.
121. Ross, C., Nizami, B., Glenister, M., Sheik Amamuddy, O., Atilgan, A.R., Atilgan, C., and Tastan Bishop, Ö. (2018). MODE-TASK: large-scale protein motion tools. Bioinformatics *34*, 3759–3763.
122. Kluyver, T., Ragan-Kelley, B., Pérez Rez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., et al. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In Positioning and Power in Academic Publishing: Players, Agents and Agendas (IOS Press), pp. 87–90.
123. Li, Z., Tomlinson, A.C., Wong, A.H., Zhou, D., Desforges, M., Talbot, P.J., Benlekbir, S., Rubinstein, J.L., and Rini, J.M. (2019). The human coronavirus HCoV-229E S-protein structure and receptor binding. eLife *8*, e51230.
124. Brown, A., Long, F., Nicholls, R.A., Toots, J., Emsley, P., and Murshudov, G. (2015). Tools for macromolecular model building and refinement into electron cryo-microscopy reconstructions. Acta Crystallogr. D Biol. Crystallogr. *71*, 136–153.
125. von Bülow, S., Sikora, M., Blanc, F.E.C., Covino, R., and Hummer, G. (2023). Antibody accessibility determines location of spike surface mutations in SARS-CoV-2 variants. PLOS Comput. Biol. *19*, e1010822.
126. Schrödinger, LLC (2020). The PyMOL Molecular Graphics System, Version 2.0. https://pymol.org/2/.
127. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38.
128. Yu, P.-Y., Yang, T.-J., and Hsu, S.-T.D. (2022). Chapter Twelve – Biophysical and structural characterizations of the effects of mutations on the structure–activity relationships of SARS-CoV-2 spike protein. In Methods in Enzymology Integrated Methods in Protein Biochemistry, A. Part and A.K. Shukla, eds. (Academic Press), pp. 299–321.
129. Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.
130. Liu, D.-G., Chang, C.-H., Lee, M.-H., Liu, C.-Y., Chang, C.-F., Chiang, L.-C., Hwang, C.-S., Huang, J.-C., Sheng, A., Kuan, C.-K., et al. (2019). Advanced small- and wide-angle X-ray scattering beamline for frontier research in biological structures at the Taiwan photon source. AIP Conf. Proc. *2054*, 060021.
131. Svergun, D.I., Petoukhov, M.V., and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. Biophys. J. *80*, 2946–2953.
132. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys. J. *105*, 962–974.
133. Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2016). FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. Nucleic Acids Res. *44*, W424–W429.
134. Wagner, J.R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., and Amaro, R.E. (2017). POVME 3.0: Software for Mapping Binding Pocket Flexibility. J. Chem. Theor. Comput. *13*, 4584–4592.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
E. coli DH5 alpha	YB Biotech	Cat# ECOS101
Chemicals, peptides, and recombinant proteins		
Trypsin	Promega	Cat# V5111
Chymotrypsin	Promega	Cat# V1061
α-Lytic Protease	New England BioLabs	Cat# P8113
Glu-C	Promega	Cat# V1651
FreeStyle 293 Expression Medium	ThermoFisher	Cat# 12338018
Expi293 Expression Medium	ThermoFisher	Cat# A1435101
ExpiFectamine 293 Transfection Kit	ThermoFisher	Cat# A14524
Polyethylenimine	Polysciences	Cat# 23966-1
LB Broth	Merck	Cat# 1.10285.0500
Critical commercial assays		
EasyPrep EndoFree Maxi Plasmid Extraction kit	BIOTOOLS	Cat# DPT-BA17
Deposited data		
SAXS data of EC1-EC5	This paper	SASBDB: SASDT35
SAXS data of EC4-EC5	This paper	SASBDB: SASDT45
Glycan conformer database	This paper	Zenodo: https://doi.org/10.5281/zenodo.8422323
Mass Spectrometry dataset	This paper	MassIVE IDL: MSV000092889
Ancillary MS figure dataset	This paper	Zenodo: https://doi.org/10.5281/zenodo.10000411
EC4-5 modeled with Man5 fitted to SAXS data	This paper	PDBDEV_00000224
EC4-5 modeled with Man9 fitted to SAXS data	This paper	PDBDEV_00000228
EC4-5 modeled with F1N4 fitted to SAXS data	This paper	PDBDEV_00000222
EC4-5 modeled with F1N6 fitted to SAXS data	This paper	PDBDEV_00000223
EC1-5 modeled with Man5 fitted to SAXS data	This paper	PDBDEV_00000225
EC1-5 modeled with Man9 fitted to SAXS data	This paper	PDBDEV_00000229
EC1-5 modeled with F1N4 fitted to SAXS data	This paper	PDBDEV_00000226
EC1-5 modeled with F1N6 fitted to SAXS data	This paper	PDBDEV_00000227
Prefusion 2019-nCoV spike glycoprotein with a single receptor-binding domain up	Wrapp et al.[61]	PDB: 6VSB
Cryo-EM structure of SARS-CoV-2 Spike D614G variant, one RBD-up conformation 1	Yang et al.[24]	PDB: 7EAZ

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cryo-EM structure of spike protein of feline infectious peritonitis virus strain UU4	Yang et al.[28]	PDB: 6JX7
Crystal structure of mouse N-cadherin ectodomain	Harrison et al.[89]	PDB: 3Q2W
Crystal structure of a human gamma-aminobutyric acid receptor, the GABA(A)R-beta3 homopentamer	Miller and Aricescu[23]	PDB: 4COF
CryoEM structure of human full-length heteromeric alpha1beta3gamma2L GABA(A)R in complex with alprazolam (Xanax), GABA and megabody Mb38.	Masiulis et al.[22]	PDB: 6HUO
MERS-CoV S structure in complex with 5-N-acetyl neuraminic acid	Park et al.[90]	PDB: 6Q04
Cryo-EM structure of MERS-CoV spike protein, all RBD-down conformation	This paper	PDB: 7YN0
Cryo-EM structure of MERS-CoV spike protein, intermediate conformation	This paper	PDB: 7YMZ
Cryo-EM structure of MERS-CoV spike protein, One RBD-up conformation 1	This paper	PDB: 7YMY
Cryo-EM structure of MERS-CoV spike protein, One RBD-up conformation 2	This paper	PDB: 7YMX
Cryo-EM structure of MERS-CoV spike protein, all RBD-down conformation	This paper	EMD-33949
Cryo-EM structure of MERS-CoV spike protein, intermediate conformation	This paper	EMD-33948
Cryo-EM structure of MERS-CoV spike protein, One RBD-up conformation 1	This paper	EMD-33947
Cryo-EM structure of MERS-CoV spike protein, One RBD-up conformation 2	This paper	EMD-33946
Fourier-cropped cryo-EM map of SARS-CoV-2 S-D614G (corresponding to PDB:7EAZ and EMDB:EMD-31047)	This paper	EMD-38650
Experimental models: Cell lines		
Expi293F™ Cells	ThermoFisher	Cat# A14527
HEK293 Freestyle	ThermoFisher	Cat# R79007
Recombinant DNA		
pcDNA3.1 CD33-N-cadherin (EC1-EC5; residues 160-712)	This paper	Custom synthesis based on Uniprot ID: NP_031690.3
pcDNA3.1 CD33-N-cadherin (EC4-EC5; residues 331-542)	This paper	Custom synthesis based on Uniprot ID: NP_031690.3
SARS-CoV-2 S protein with the D614G mutation in pcDNA3.4-TOPO	This paper	Custom synthesis based on Uniprot ID: P0DTC2
SARS-CoV S protein with the 2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40150-G-N (GenBank ID: AAP13567.1)
MERS-CoV S protein with the 2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40069-NF (GenBank ID: APF29071.1)

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MERS-CoV S protein with the fm2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40069-NF (GenBank ID: AFS88936.1)
hCoV-NL63 S protein with the 2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40604-CF (GenBank ID: APF29071.1)
hCoV-229E S protein with the 2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40605-CF (GenBank ID: APT69883.1)
hCoV-HKU1 S protein with the fm2P mutation in pcDNA3.4-TOPO	This paper	Subcloned from Sino Biological Cat# VG40606-UT (GenBank ID: Q0ZME7.1)
Software and algorithms		
GlyCONFORMER	Grothaus et al.[38]	https://github.com/IsabellGrothaus/GlyCONFORMER
Relion 3.1	Zivanov et al.[91]	https://www3.mrc-lmb.cam.ac.uk/relion//index.php?title=Main_Page
CryoSPARC v3.2	Punjani et al.[92]	https://cryosparc.com/
Phenix 1.19.2-4158	Adams et al.[93]	http://www.phenix-online.org/
Win Coot 0.9.4.1	Emsley et al.[94]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
UCSF-ChimeraX 1.2.5	Pettersen et al.[95]	https://www.rbvi.ucsf.edu/chimerax/
Byonic v3.9.6	Protein Metrics, USA	https://proteinmetrics.com/byos/
Byos v3.11	Protein Metrics, USA	https://proteinmetrics.com/byos/
ATSAS-3.0.3-1	Manalastas-Cantos et al.[96]	https://www.embl-hamburg.de/biosaxs/software.html
SAXS deduction program	Shih et al.[97]	N/A
Gromacs 2018.2 or newer	Abraham et al.[98]	www.gromacs.org
Gromaps	Briones et al.[99]	https://mptg-cbp.github.io/gromaps.html
MDAnalysis-1.1.1	Gowers et al.[100]	https://www.mdanalysis.org/
numpy-1.21.1	Harris et al.[101]	Numpy.org
Matplotlib-2.2.3	Hunter[102]	Matplotlib.org
mrcfile-1.4.3	Burnley et al.[103]	https://mrcfile.readthedocs.io/en/stable/index.html
GlycoSHIELD program suite: GlycoSHIELD, GlycoTRAJ, GlycoSASA, GlycoALPHAFOLD	This paper	Zenodo: https://zenodo.org/records/10668375
GlycoDENSITY	This paper	www.glycoshield.eu
Modeller-10.4	Webb and Sali[104]	https://salilab.org/modeller/
Other		
Superdex 200 Increase 10/300 GL	Cytiva	Cat# 28990944
HiLoad 16/600 Superdex 75 pg	Cytiva	Cat# 28989333
HisPur™ Cobalt Resin	ThermoFisher	Cat# 89966
Superose 6 Increase 10/300 GL	Cytiva	Cat# 29091596
Carbon film 300 mesh Cu Grids	Electron Microscopy Sciences	Cat# CF300-Cu-50
Quantifoil R1.2/1.3 300-mesh Cu Grids.	Ted Pella	Cat# 658-300-CU-100

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Mateusz Sikora (mateusz.sikora@uj.edu.pl).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The cryo-EM maps and atomic coordinates for MERS-CoV and SARS-CoV-2 S proteins have been deposited in the Electron Microscopy Data Bank (EMDB) and in the Protein Data Bank (PDB) and are publicly available as of the date of publication. SAXS data reported in this paper have been deposited in the Small Angle Scattering Biological Data Bank (SASBDB) and are publicly available as of the date of publication. The atomic coordinates of glycosylated models of EC1-EC5 and EC4-EC5 used for SAXS analyses have been deposited in the Prototype Archiving System for Integrative Structures databank (PDB-Dev) and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. MS data as well as additional figures describing MS results obtained for S proteins have been deposited at Zenodo and MassIVE (see key resources table) and are publicly available as of the date of publication.
- All source code related to GlycoSHIELD, GlycoTRAJ, GlycoSASA, GlycoDENSITY and GlycoALPHAFOLD has been deposited at Zenodo or www.glycoshield.eu, and is publicly available as of the date of publication. Reference structures and molecular dynamics trajectories of single glycans have been deposited at Zenodo and are publicly available as of date of publication. DOIs are listed in the key resources table. Files needed to regenerate simulations of N-cadherin, SARS-CoV-2 S full length protein as well as fragments, together with complete trajectories are available upon request from the lead contact.
- Any additional information required to reanalyze data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODELS AND STUDY PARTICIPANT DETAILS

Bacteria

DH5 Alpha *E. coli* (YB Biotech) used for propagation of expression plasmids were cultured at 37 ºC in LB broth (Merck) under agitation at 180 rpm.

Cell lines

Recombinant proteins were produced in Expi293F^{TM} or HEK293 Freestyle. Expi293 cells (ThermoFisher) were cultured in Expi293 expression medium at 37°C in an 8% CO_2 atmosphere under agitation at 125 rpm. Cells were transfected with plasmid DNA using the ExpiFectamine293 transfection kit (ThermoFisher). 20 hr post-transfection, enhancers were added to the cell culture as per manufacturer's instruction and recombinant protein expression continued for another 3–5 days before harvesting. HEK293 FreeStyle cells (ThermoFisher) were cultured in FreeStyle 293 expression medium at 37°C in an 8% CO_2 atmosphere under agitation at 125 rpm. Cells were transfected with plasmid DNA using polyethylenimine (Polysciences) and recombinant proteins harvested 3-5 days post-transfection.

METHOD DETAILS

Molecular dynamics simulations

System preparation

Glycans were modeled using averaged structures from the glycan fragment database as implemented in Glycan Modeler.[83] CHARMM-GUI[105] was used to add glycans to protein structures and insert proteins in membrane models, unless mentioned otherwise. All systems were solvated using the TIP3P water model[106] in the presence of 150mM NaCl and were prepared for simulations with the CHARMM36m force field.[107–109]

Single glycan systems for unbiased simulations. Individual glycans provided in the GlycoSHIELD glycan library (Table S1) were prepared as follows: N-glycans were attached to the central residue of GLY-ASN-GLY tripeptides with NH^--COOH extremities; O-glycans were attached to ALA-THR-ALA tripeptides with neutralized N- and C-termini (O-glycans) and placed in rectangular simulation boxes enlarged by 15-20 Å to prevent self-interactions.

The Man5, A2 and Fuc1_Neu4 glycans used for the analyses presented here were prepared in an earlier version of our glycan library and were simulated on GNG tripeptides with NH_3^+-COOH extremities. For consistency with other entries, these systems were not included in the glycan library, but are available on request. The influence of peptide capping on N-glycan conformer ensemble was tested - and found to be negligible in the present case - by comparing glycan simulations on peptides with NH_3^+-COO^-; NH_3^+-COOH, NH^--COOH or NH^-- COO^- capping, using the analysis pipeline described in "comparison of unbiased and biased sampling" (data not shown).

Single glycan systems for biased simulations. Individual N-glycans were prepared as described above, but without the tripeptide, and solvated including neutralising sodium ions whenever required.

EC4-EC5. EC4-EC5 domains (residues 331-542) were derived from N-cadherin structure PDB: 3Q2W[89] and were prepared with neutral N- and C-termini, in a non-glycosylated and a glycosylated form, with the N-glycans shown in Figure 1B.

Full-length glycosylated SARS-CoV-2 S protein. The preparation of this system has been described elsewhere[6] and experimentally validated by cryo-electron tomography.[4] In brief, the system contained four membrane-embedded S protein trimers assembled from available structures and *de novo* models for the missing parts. S protein head was modeled based on PDB: 6VSB[61] with one RBD domain in an open conformation. The stalk connecting S protein head to the membrane was modeled *de novo* as two trimeric coiled coils.[110] The TMD as well as the cytosolic domain were modeled *de novo*. Glycans were modeled according to glycoproteomics data[46] (See Figure S5A for glycan composition).

S protein stalk (HR2-TMD). Truncated non-glycosylated and glycosylated systems comprising the heptad repeat region 2 (HR2) and the fully palmitoylated transmembrane domains (TMD) of S protein (residues 1161-1273) were generated as described above. N-termini were neutralized and the protein was inserted in a hexagonal patch of ER-like, symmetric plasma membrane with lipid ratios: 25:25:20:15:5:5:5 DOPC:POPC:POPE:POPI:POPS:CHOL:CER160,[111] with the membrane oriented in the XY-plane and the TMD oriented along the Z-axis of the simulation box.

S proteins HR2 domain. HR2 domain was extracted from the full-length S protein model (residues 1176-1206) with all termini set to neutral. Fuc1_Neu4 glycan was added to N1194 for single protomers individually or all 3 protomers (*i.e.,* 4 independent systems).

Minimization, equilibration and production

MDS were performed with GROMACS engine 2018.2 or newer versions.[98] The potential energy of each system was first minimized (steepest descent algorithm, 5000 steps), before equilibration procedures (see below) were started. Cut-off radii for interactions were applied at 1.2 nm, with a force switch at 1 nm for van der Waals interactions. The Particle Mesh Ewald algorithm was used to treat long-range electrostatic interactions. All simulations were performed at ambient pressure and 310 K temperature. Atom positions were stored at 100 ps intervals.

Unbiased simulations of single glycan systems. After minimization, 12.5 ps NVT equilibration runs were performed using a Nosé-Hoover thermostat,[112] restraining all bonded interactions involving hydrogen atoms. Both atom positions and dihedral angles were restrained during equilibration, with initial force constants of 400, 40 and 4 kJ/mol/nm^2 for restraints on backbone positions, side chain positions and dihedral angles, respectively. Force constants were gradually reduced to 0. NPT production runs were performed using a Parrinello-Rahman[113] barostat and a velocity-rescaling thermostat.[114] Total simulation time per glycan system was between 3 to 4 μs.

Biased simulations of single glycan systems. GROMACS 2018.4 patched with PLUMED 2.6[115] was used with 2 fs integration time steps and the LINCS algorithm was employed to constrain bonds connected to hydrogen atoms. Temperature coupling was performed with velocity rescaling and a 0.1 ps time constant, setting the reference temperature at 310 K. The Parrinello-Rahman barostat was used to set the pressure to 1 bar with a compressibility of 4.5 x 10^{-5} bar^{-1}, using a time constant of 5 ps. Electrostatic interactions were calculated by the Particle Mesh Ewald (PME) method using a cut-off distance of 1.2 nm. Atom positions were stored every 8 ps. Keeping N-glycan atoms restrained, energy minimization using the steepest descent algorithm with a tolerance of 1000 kJ mol^{-1}nm^{-1} was performed, followed by equilibration in an NVT and an NPT ensemble for 1 ns, respectively. After disabling all restraints, the equilibration procedure was repeated.

Enhanced sampling simulation runs were performed with the REST-RECT method, combining replica exchange with solute scaling (REST2)[116] and collective variable tempering (RECT).[38,117] All N-glycan atoms were defined as the solute part in REST2, scaling interaction energies (and, consequently, effective temperature) in higher replicas by $\lambda \in (0\text{-}1)$, acting only on long-range electrostatics, Lennard-Jones potentials as well as dihedral angles. 12 replicas were used in total, with a temperature ladder following a geometric progression at 310.150, 338.050, 368.460, 401.610, 437.739, 477.118, 520.039, 566.822, 617.814, 673.393, 733.972, 800 K. The solvent was kept at a ground temperature of $T_0 = 310.15$ K in all replicas. In the RECT part, all n_t glycosidic bond torsion angles (φ, ψ, ω) of each glycan were defined as collective variables (CVs) and biased simultaneously by n_t -times one dimensional energy potentials in each replica i. Torsion angles were defined as φ = O5'−C1'−Ox−Cx, ψ = C1'−Ox– Cx−C(x−1) and ω = O6−C6−C5−O5, with x being the carbon number of the linkage at the non-reducing end. 2−6 linkages between Gal and Neu5Ac represented an exception with different definitions as φ = O6'−C2'−O6−C6, ψ = C2'−O6−C6−C5 and ω = O6−C6−C5−O5 (see Figure S1B). Man5, A2 and Fuc1_Neu4 had n_t equal to 14, 23 and 37, respectively. The bias factor (γ) acting on each CV was scaled over the different replicas i. The ground replica was kept unbiased having $\gamma_0 = 1$ and the highest replica $\gamma_{12} = 14$, following a geometric progression between these replicas. Gaussian hills were deposited every t = 1 ps with a width of 0.35 rad and τ, the characteristic time for the bias evolution in the RECT part, equal to 4 ps. Replica exchanges were attempted every 400 steps, following a Metropolis-Hastings acceptance criterion. Total simulation time per replica was 500 ns, resulting in 6 μs of cumulative time (12 replicas).

EC4-EC5. The glycosylated and non-glycosylated systems were equilibrated in NVT runs for 18.75 ps with 1 fs time-steps using the Nosé-Hoover thermostat. Atom positions and dihedral angles were restrained during the equilibration, with initial force constants of 400, 40 and 4 kJ/mol/nm^2 for restraints on backbone positions, side chain positions and dihedral angles, respectively. During equilibration, force constants were gradually reduced to 0. Systems were additionally equilibrated in NPT runs (Parinello-Rahman pressure coupling with a time constant of 5 ps and compressibility of 4.5x10^{-5} bar) over the course of 10 ns, with a time step of 2 fs. Bonds containing H atoms were restrained using the LINCS[118] algorithm. During the production runs, a velocity-rescaling thermostat was used and the temperature was kept at 310 K. Three independent production MDS of non-glycosylated and glycosylated EC4-EC5 were performed for a total duration of 3 μs for each system.

S protein stalk (HR2-TMD). Preparation and production runs were performed similarly to EC4-EC5. The total NVT equilibration time was extended to 125 ps, and the NPT equilibration comprised an additional 1 ns step using the Berendsen barostat[119] (with semi-isotropic coupling, time constant of 5 ps and compressibility equal to 4.5x10^{-5} bar^{-1}) to resolve the packing of lipids around the TMD.

Restraint force constants were gradually reduced from 4000, 2000, 1000 to 0 kJ/mol/nm^2 for backbone positions, side chain positions and dihedral angles, respectively. Non-glycosylated and glycosylated HR2-TM systems were each simulated for 3.5 μs.

S protein HR2 domain. Preparation and production runs were performed similarly to EC4-EC5. A total of 3 μs simulated time was acquired for each system.

Comparison of unbiased and biased sampling

Glycan conformers obtained by REST-RECT simulations were analyzed using GlyCONFORMER (https://github.com/IsabellGrothaus/GlyCONFORMER).[38] For comparison, tripeptides were removed from unbiased sampling simulations to match conditions used for REST-RECT sampling. In short, each recorded conformer was assigned a conformer string, based on the occupancy of the three torsion angles φ, ψ and, if applicable, ω (see Figure S1). The conformer string is characterized by n_t digits, equal to the number of torsion angles in the glycan. The assignment starts at the free reducing end of the glycan and continues along all branches, proceeding first along the branch with the highest C atom number (e.g., 1-6 linkage) at a junction. Once the branch is assigned until its terminus, labeling continues from the last junction following the second highest C atom number (e.g., 1-3 linkage). Each branch is separated within the string by a line, harboring the C atom number of that linkage for reference. Primary branches of N-glycan core: 1-3 and 1-6 are labeled by numbers in bold. Within one linkage, the string reports ordered digits assigned to φ, ψ and ω, respectively (and if applicable). The attachment of e.g. fucose to the chitobiose core of N-glycans is mentioned last. For the string assignment, first, the free energy profile along each torsion angle is determined from the REST-RECT simulation and the positions of individual minima are recorded. Torsion angles belonging to the same free energy minimum (between two maxima) are labeled equally. The digits corresponding to φ and ψ angles are assigned the letters C, G_+, A_+, T, A- and G-, depending on the occupied minima in each frame, in line with the IUPAC nomenclature for dihedrals (https://goldbook.iupac.org). In contrast, digits corresponding to ω torsion angles are labelled by gg, gt or tg.

Probability distributions for each conformer were calculated and block averaging was performed (using 10 blocks) to calculate error bars defined as the standard deviation of the sampling distribution. The convergence of simulations was assessed by calculating the moving average of conformer probabilities, with a window size of 1.2 μs for unbiased MDS and 100 ns for REST-RECT to encompass identical effective sampling periods.

Trajectory analyses

Root mean square deviations (RMSD) and root mean square fluctuations (RMSF) of atom positions, radius of gyration of the protein backbone, and numbers of hydrogen bonds were calculated with GROMACS.[98] Angle and distance measurements were performed with MDAnalysis.[100] HR2 angle to the membrane normal direction was calculated by measuring the deflection of a vector defined by the centres of mass of the residues flanking the coiled-coil region (residues 1160 and 1181) from the z-axis of the simulation box.

Principal component analysis (PCA): the PCA shown in Figure S1D was performed using the scikit-learn package.[120] To take torsional periodicity into account, sin and cos of torsion angles were used as input features. When comparing unbiased MDS and REST-RECT simulations, both datasets were combined prior to performing PCA. Free energy (ΔG) surfaces along the two first principal components were computed in two-dimensional histograms using 35 x 35 bins, after conversion of the probabilities (P) to ΔG using the relationship:

$$\Delta G = -k_B T \, ln(P) \qquad \text{(Equation 1)}$$

where k_B is the Boltzmann constant.

The PCA shown in Figure S6D was done on all N-glycan atoms using MODE-TASK.[121]

GlycoSHIELD software suite

Rationale and code description

GlycoSHIELD. A library of conformers of 68 distinct glycans (30,000 conformers per glycan, corresponding to 100 ps sampling intervals and a total of up to 3 μs simulated time) was generated and is publicly available on Zenodo (See Key Resources List and https://doi.org/10.5281/zenodo.8422323). Each conformer contains structural information about the glycan and a tripeptide to which the glycan was attached. All trajectories were aligned to a single tripeptide structure to facilitate comparison between ensembles. The library is an open access resource and additional glycan types and more extensive sets of conformers will be made available in the future.

For every glycosylation site, users select a glycan to be grafted from the available library. GlycoSHIELD aligns every glycan conformer using the backbone atoms of its tripeptide anchor onto the corresponding residues on the protein, i.e., the selected ASN or THR/SER residue ± 1 flanking residue. The selected array (excluding the flanking residues) is then tested for clashes with protein atoms within a threshold radius, and clashing conformers are excluded from the glycan array. The rejection criterion can be based on all protein and glycan atoms or on protein Cα-carbons and glycan ring-oxygens. The latter, dubbed coarse-grained (CG) threshold, is especially recommended for larger protein systems, where clash search may require extended calculations. Based on comparison with atomistic simulations (Figure S2B), "soft" threshold values of 0.7 Å and 3.5 Å are recommended for atomistic and CG clash tests, respectively, which implicitly takes into account the dynamic nature of amino acids lateral chains at the protein surface. In addition, the presence of a biological membrane can be mimicked after orienting the protein with the z-axis perpendicular to the membrane by setting up glycan rejection above or below defined values along the z-axis using z_{max} or below z_{min} thresholds, respectively.

GlycoSHIELD output consists of glycan frames successfully grafted onto the static protein for each probed glycosylation site which are exported in XTC and PDB formats. In case of multiple glycosylation sites, the script "GlycoTRAJ.py" may be used to merge this

information into a single master trajectory with all grafted glycans. Typically, the smallest number of glycan frames grafted across all glycosylation sites of the protein can be selected, thus generating a "trajectory" with static protein and mobile glycans.

GlycoSASA is a convenience script to calculate ΔSASA with a selected probe radius using grafted glycan frames as an input (see below).

GlycoDENSITY enables the reweighting of glycan conformers generated by GlycoSHIELD based on correlation with cryo-EM maps. To this end, the experimental map is normalised to zero mean and unit variance. Then, for each glycan, a sub-volume is extracted from the experimental map based on the maximum bounding box defined by the ensemble of grafted glycans, and expanded by a given margin. In the next step, a fitness score F(conformer) is defined for each conformer of the ensemble, reflecting its local fit to the map. For typical cryo-EM maps where the grid spacing is larger than the radius of gyration of a monosaccharide (Rs, usually ~3 Å), F is computed according to (2):

$$F(conformer) = \frac{1}{N}\sum_{i=1}^{N}\rho(x_{0,i}, y_{0,i}, z_{0,i}) \quad \text{(Equation 2)}$$

where i=1,...,N iterates over the number of monosaccharides in a glycan; ρ(x,y,z) is the density value at a given x,y,z coordinate; ($x_{0,i}$, $y_{0,i}$, $z_{0,i}$) is the center of mass of monosaccharide i. If grid spacing is significantly smaller than R_s, using a single value leads to a high susceptibility to noise. In this case, a spherical binary mask with radius R_s is placed at ($x_{0,i}$, $y_{0,i}$, $z_{0,i}$) for each monosaccharide unit of the glycan. Mask edges are then smoothed with a gaussian filter and values in the voxels included or overlapping with the mask are weighed by the percentage of the voxel covered by the mask, averaged, and then taken as a monosaccharide's contribution to the glycan fitness according to (3):

$$F(conformer) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sum_{x,y,z} w_i(x,y,z)\rho(x,y,z)}{\sum_{x,y,z} w_i(x,y,z)}\right) \quad \text{(Equation 3)}$$

where

$$w_i(x,y,z) = \begin{cases} e^{-\frac{(x-x_{0,i})^2+(y-y_{0,i})^2+(z-z_{0,i})^2}{2\sigma^2}} & \text{if } (x-x_{0,i})^2+(y-y_{0,i})^2+(z-z_{0,i})^2 \le R_s^2;\ 0 \quad \text{otherwise} \end{cases}$$

and σ defines the standard deviation of a gaussian (by default set to 0.5 of a voxel length).

Finally, members of the ensemble are ranked by fitness value and a defined number of "best-fit" conformers is selected as a re-weighted ensemble.

GlycoALPHAFOLD is a convenience software that automatically N-glycosylates AlphaFold2[15] models, requiring only their Uniprot ID. As described in the Jupyter notebook[122] tutorial GlycoAlphaFold.ipynb, it fetches information on its glycosylation status from Uniprot, automatically generates GlycoSHIELD input and grafts selected glycan (Man9 by default) onto selected sites. The grafting then enables a ΔSASA analysis and a simplified 3D visualization of the glycan shields.

The software is available as a python library and standalone user scripts (see key resources table).

User script description

GlycoSHIELD.py. As input, the user provides a static structure file (PDB format) or series of structures (PDB or GROMACS XTC format). Glycosylation sites and glycans to be grafted are provided by users as batch input files as follows: the first column defines protein chain symbol as defined in PDB, followed by comma-separated glycosylation site residue number together with its two flanking residues, the third column defines the tripeptide residue numbers as defined in the glycan conformer library (typically 1, 2, 3). The fourth and the fifth columns provide a path to the glycan conformer reference structure (PDB) and trajectory files (XTC), respectively. Finally, the last two columns define names of the PDB and XTC output files generated by the script, containing static protein structure and a number of the accepted glycan conformers.

The syntax of the main GlycoSHIELD is as follows, with optional parameters in square brackets:

GlycoSHIELD.py [-h] –protpdb PROTPDB –protxtc PROTXTC –inputfile INPUTFILE [–threshold THRESHOLD] [–mode MODE] [–zmin ZMIN] [–zmax ZMAX] [–dryrun] [–no-dryrun] [–shuffle-sugar] [–no-shuffle-sugar] [–ignorewarn] [–no-ignorewarn] [–skip SKIP]

options:

--Inputfile	–Path to the batch input file as defined above
--Mode	–Mode of trimming of the glycan conformer library. CG (coarse-grained): distances between alpha carbons of amino acids and ring oxygens of glycans are calculated to remove clashes, less precise but faster for larger systems. All: distances between all protein and sugar atoms are calculated to remove clashes.
--Threshold	–a minimum distance between protein and glycan atoms below which glycan conformer is rejected

(*Continued on next page*)

Continued	
--Protpdb	–Structure file containing protein coordinates in PDB format
--Protxtc	–Optional, sequence of conformations of a protein (i.e., a trajectory). Each of the frames will be grafted separately.
--Dryrun	–Does not produce grafted structures, but informs about the number of accepted glycan conformers per site and per protein conformer. Useful, e.g., when entropic penalty for a number of distinct protein conformations is calculated.
--Ignorewarn	–Does not stop the grafting if no glycan conformers not clashing with the protein were found.
--Zmax and --Zmin	–Force rejection of glycans that extend beyond these values (in angstrom) along the Z axis. This option is useful for mimicking steric hindrance by cellular membranes.
--Shuffle-sugar	–By definition glycan conformers are taken from the library in a sequential fashion. This option introduces additional randomization by shuffling the order of the library of conformers for every glycosite.
--Skip n	–Only read every n-th frame. This option can be used when very large systems are glycosylated.
--Help	–Displays a help message

GlycoTRAJ.py is a convenience program that allows for merging multiple glycan conformer clouds into one trajectory. In order to obtain a consistent trajectory, i. e., containing the same number of atoms in each frame, the maximum trajectory length is limited to the smallest number of conformers grafted among all glycosylation sites.

GlycoTRAJ.py [-h] –maxframe MAXFRAME –outname OUTNAME –pdblist PDBLIST –xtclist XTCLIST –chainlist CHAINLIST –reslist RESLIST

options:

-H, Help	–Show this help message and exit
--Maxframe MAXFRAME	–Number of frames to include for each glycan
--Outname OUTNAME	–Root name of the output file
--Pdblist PDBLIST	–List of coma-separated reference pdb files
--Xtclist XTCLIST	–List of coma-separated trajectory files
--Chainlist CHAINLIST	–List of coma-separated chain descriptors where each glycan was attached
--Reslist RESLIST	–List of coma-separated residue numbers where each glycan was attached

GlycoSASA.py. Shielding calculation requires a set of glycan trajectories generated by the GlycoSHIELD script (see above), each containing the same, static protein conformer, and a mobile glycan. Script syntax is as follows:

GlycoSASA.py [-h] –pdblist PDBLIST –xtclist XTCLIST [–probelist PROBELIST] [–plottrace] [–no-plottrace] [–keepoutput] [–no-keepoutput] [–ndots NDOTS] [–mode MODE] [–endframe ENDFRAME]

Options:

--Pdblist	–Comma-separated list of paths to the PDB reference files
--Xtclist	–Comma-separated list of paths to the XTC trajectories
--Probelist	–Comma-separated list of the probe radii, for which SASA should be calculated (in nm)
--Plottrace	–Boolean, should a plot of per-residue SASA be also produced
--Keepoutput	–Whether the temporary output files of the SASA calculation should be kept. For debugging purposes
--Ndots	–How many dots should be used per atom in the SASA algorithm (see GROMACS manual for the details)
--Mode	–By default shielding is calculated as a maximum (max) of the shieldings at each glycosylation site. User can choose to calculate also the average shielding (avg)
--Endframe	–The last frame to read for glycan trajectories, assumes 1ps time step

The output consists of a text file that assigns a shielding value to each residue. These values multiplied by 100 are also written into the b-factor column of the PDB file. Residues that are not visible to the probe at all (i.e., inaccessible) are further marked by 0 in the occupancy column of the PDB file. The output files are named: maxResidueSASA_probe_{RPROBE}.txt maxResidueSASA_probe_{RPROBE}.pdb, where RPROBE is a probe radius. If a per-residue shielding plot was requested (–plottrace), a plot will be created under maxResidueSASA_probe_{RPROBE}.pdf.

GlycoDENSITY.py can be used on any protein structures with an available corresponding cryo-EM / electron density map. As a prerequisite, the protein model has to be aligned within the map and standard GlycoSHIELD grafting has to be performed on chosen sites using selected glycans as described above. The full GlycoSHIELD output should be available, including glycan conformers

without the protein structure ({chain}_{residue}_glycan.pdb and {chain}_{residue}_glycan.xtc files in the examples provided. As input, the user provides a density map as well as the list of glycan sites and respective chains where reweighing should be performed. GlycoDENSITY syntax is as follows:

GlycoDENSITY.py [-h] –path PATH –expmap EXP_MAP –chainlist CHAINLIST –reslist RESLIST [–mxfrmlst MXFRMLST] –fitframelist FITFRAMELIST [–usegmx] [–normosecount] [–normr2] [–weighedgridsum] [–verbose] [–sigma SIGMA]

options:

-h, –help	- show this help message and exit
–path PATH	- path to where GlycoSHIELD full output is stored. Expecting that GlycoTRAJ was also run, but not requiring GlycoSASA. The map is expected to reside in the same folder. This is also where the results will be stored.
–expmap EXP_MAP	- experimental density map (has to have ccp4 extension!)
–chainlist CHAINLIST	- coma-separated list of chain names for each glycan
–reslist RESLIST	- coma-separated list of residue numbers of ASN for each glycan
–mxfrmlst MXFRMLST	- how many frames to use. -1 uses all. (define for each glycan)
–fitframelist FITFRAMELIST	- for each glycan define how many most fit frames enter the synthetic density map calculation. Selected manually by user for each glycan.
–usegmx	- If Gromaps is installed, do a full set of analysis. If not, the analysis is done but the diagnostic maps are not produced.
–normosecount	- normalize by the number of sugar monomers in glycan
–normr2	- normalize fitness score by the distance from the ASN alpha carbon
–weighedgridsum	- do fine sampling of a grid (relevant for grid size << glycan size, slow)
–verbose	- be loud and include all possible outputs
–sigma SIGMA	- maptide sigma to use [nm]. To select an adequate value, the easiest is to make a synthetic map with several sigma values and compare to the experimental one in the region of interest. The level of detail should look similar.

Dependences

GlycoSHIELD was tested with the MDAnalysis-1.1.1,[100] numpy-1.21.1[101] and Matplotlib-2.2.3[102] Python packages. Calculation of the glycan-dependent Solvent Accessible Surface Area (SASA) requires GROMACS-2019.2 (*gmx sasa* tool, only tested by us in version 2019.2). GlycoDENSITY requires that the python mrcfile library[103] is installed. The generation of synthetic maps requires Gromaps.[99]

Web application

Users are encouraged to use the GlycoSHIELD web application (www.glycoshield.eu). The application uses High Performance Computing resources in the Cloud and provides access to the majority of the GlycoSHIELD, GlycoTRAJ and GlycoSASA functionalities *via* an intuitive web graphical user interface. Technically, the web application is implemented in pure Python using the Streamlit (https://streamlit.io) package and runs on a scalable Jupyterhub/Binderhub infrastructure. The web application streamlines the handling of the GlycoSHIELD pipeline by unifying input definition *via* PDB/CIF structure file upload, glycan selection, computation runs, result visualization in 2D plots and interactive 3D model visualization windows, and output files download. More advanced users may directly access tutorials implemented as Jupyter notebooks from the web application portal to evaluate and use GlycoSHIELD via a typical JupyterLab interface. The glycan selection window of the web application is tethered to the latest version of the glycan database. The ‘‘input generator’’ interface can also be controlled through a text box to facilitate the handling of more complex cases requiring direct editing (e.g., copy and paste to propagate to multiple chains) and full control of grafting parameters. A detailed report on the design, implementation and benefits of scalable Cloud-based web applications for computational sciences is in preparation and will be published elsewhere.

Glycan shield modeling with GlycoSHIELD

EC4-EC5. Non-glycosylated and glycosylated EC4-EC4 trajectories were aligned to select similar protein conformers from the two simulations. For comparison of shields derived from actual MDS and GlycoSHIELD, all the glycan conformers from the simulation of the glycosylated protein were locally aligned to the selected protein conformer. The non-glycosylated protein was glycosylated with GlycoSHIELD with the same glycans as the glycosylated protein, with a CG distance threshold of 3.5 Å.

$GABA_A$ receptors. The β3 homopentameric models shown in Figures 7A and 7B were generated from structure PDB: 4COF.[23] The 11 accessible N-glycosylation sites of the protein structure were glycosylated with Man5 and glycan shields modeled with an ‘‘All’’ threshold of 0.8 Å and taking into account steric hindrance by the upper leaflet of the membrane. The $\alpha1\beta3\gamma2$ heteropentameric models shown in Figures 7C–7G were generated from structure PDB: 6HUO.[22] Missing segments of the protein were reconstructed with AlphaFold2.[15] The 12 accessible N-glycosylation sites of the protein structure were glycosylated with Man9[22] and glycan shields

modeled with a "CG" threshold of 3.5 Å and taking into account steric hindrance by the upper leaflet of the membrane using GlycoSHIELD—zmax option. The glycan conformers of the two N-glycosylation sites of the channel vestibule were further refined by alignment with the monosaccharides resolved in the PDB file (2x GlcNAc, β1-4 mannose and the first α1-3 and α1-6 mannoses after the branching point). Only conformers for which the centre of mass of individual monosaccharides was closer than 3 Å to the centre of mass of corresponding monosaccharides in the original PDB file were selected. To take into account a possible rotation of the glycan trees along the GlcNAc-GlcNAc axis, alignment of α1-3 and α1-6 mannoses was done with the closest α1-3 or α1-6 mannose of the original PDB file (*i.e.,* regardless of whether these mannoses belonged to the A or B/C branches of the tree).

SARS-CoV-2 S protein. Grafting was performed on a full-length S protein trimer, either using the heterogeneous glycans as shown in Figure S5 or a homogeneous coverage with Man5 or Fuc1_Neu4 (see glycan definitions in Table S1).

For the ray analysis, heterogeneous grafting was further performed on the first 10 ns of the full S protein simulation at 1 ns intervals, resulting in 160 glycan conformers grafted on each glycosylation site and each protein conformation (CG mode with 3.5 Å threshold, glycans grafted at 100 ps intervals).

S protein stalk (HR2-TMD). Models of the S protein stalk with different bending angles were generated from non-glycosylated HR2-TMD as follows. The linker region from the HR2-TMD system (residues 1205-1211) was removed. TMD was oriented along the Z-axis and HR2 placed above it at the correct distance. Next, HR2 was rotated by varying the angle between the Z-axis and a vector defined by the centres of mass of HR2 N-terminal residues, with the point of rotation fixed above TMD N-terminal residues. Structures corresponding to a range of angles were generated and the linker region reconstructed using Modeller.[104] Each of these structures were then processed with GlycoSHIELD to reconstruct glycan shields with Man5 or Fuc1_Neu4 for all glycosylation sites (CG mode, 3.5 Å threshold, glycans grafted at 1 ns intervals).

S protein HR2 domain. The protein frames of the simulations were stripped of glycans and solvent and used as input for GlycoSHIELD and re-glycosylated with the same glycan type (CG mode, 3.5 Å threshold). The reference frame of the protein that was used for comparison of MD simulations and GlycoSHIELD was selected by maximizing the number of accepted glycan conformers grafted by GlycoSHIELD and by minimizing variability between protomers. The number of accepted conformers was then adjusted to have an equal number of conformers on each protomer (*i.e.,* 303). In parallel, glycans conformers were extracted from MD simulations of the mono or tri-glycan systems, re-aligned on this reference frame and down-sampled to the same number of conformers grafted with GlycoSHIELD.

Other hCoV S proteins. Grafting was performed with the default settings using the following S protein models: SARS-CoV,[68] MERS (present study), hCoV-HKU1,[69] hCoV-NL63,[70] hCoV-229E.[123] Most abundant glycan species according to the MS data were grafted at each site. Missing loops, N- and C-termini in reported PDB structures were manually built and refined using Coot.[124] Man5 was used for N-glycosylation sites for which glycans were not detected/resolved by MS (see below). Shielding of each S protein was calculated using GlycoSASA with default parameters and 20 glycan conformers per sequon.

FIPV S protein glycans. As a reference structure, PBD ID: 6JX7[28] was used. Man9 was grafted on N491 and N1218 and NA2F was grafted on N949, using default GlycoSHIELD settings and resulting in 7417, 393 and 2756 conformers, respectively.

Ray analysis of protein surface accessibility

The accessibility of SARS-CoV-2 S protein surfaces (Figure 3B) was probed by illuminating the protein with diffuse light as described elsewhere.[6] In brief, 10^6 light rays of random orientation were generated from the inner surface of a 25 nm radius spherical dome centred on the centre of mass of the protein. Rays were considered absorbed when located $\leq$ 2 Å from any heavy atom. The shielding effect of glycans was quantified by calculating the difference in protein surface irradiation with and without including glycans. Representative antibody epitopes were taken from Sikora et al.[6]

To determine the relative contributions of protein and glycan dynamics to glycan shielding, snapshots of S protein conformations were extracted over short and extended segments of the full MDS. Specifically, 1, 10 and 100 conformers were extracted from the first simulation frame ("0 ns"), from the first 9 ns of the simulation in 1 ns intervals ("9 ns"), or from equally spaced snapshots from the entire 10 μs-long simulation ("10 μs"). Ray analysis was then performed after grafting 160 glycan conformers (i.e., the smallest number of accepted conformations across all the glycosylation sites of the protein) onto these 1-100 protein conformers (Figure S5B).

As seen by comparing squared residuals of GlycoSHIELD versus full MDS shielding plots (Figure S5B), the grafting of single glycan conformers onto a static S protein structure was already sufficient to predict glycan shielding at a coarse level (0 ns/1glyc). As expected however, grafting arrays of glycan conformers with GlycoSHIELD substantially improved shielding prediction (0 ns/160glyc). Sampling protein conformers over short segments of the protein trajectory did not improve the prediction made on a single protein conformer (compare 9 ns/160glyc to 0 ns/160glyc). In contrast, averaging over protein conformers sampled over extended durations further improved shielding predictions (10 μs/160glyc). This extended analysis was performed on a dataset used in a previous publication.[125]

Differences in shielding scores along the protein between single protein structures and extended protein conformer sampling were apparent in the first 30 N-terminal residues of the protein and in the loops of the receptor-binding domain (RBD, residues 450-500) (see also the 3D shielding heat map in Figure 3B). These differences may be attributed to the substantial conformational changes that occur in these protein regions over μs time scales and which are thus only captured by extended simulations.

Solvent accessible surface area (SASA)

To quantify the impact of glycans on protein surface accessibility, per-residue SASA was calculated in parallel for non-glycosylated and glycosylated proteins ($SASA_{nogly}$ and $SASA_{gly}$, respectively). Calculations were performed with GROMACS *gmx sasa* and custom Python scripts (see code description and availability below). When considering multiple glycan conformers, $SASA_{gly}$ was calculated and averaged over all conformations. Shielding was calculated as a relative SASA:

$$\Delta SASA = \left(SASA_{nogly} - SASA_{gly}\right) / SASA_{nogly} \quad \text{(Equation 4)}$$

A $\Delta SASA$ value of 1 denotes the full shielding of a residue, and 0 denotes no shielding. For glycans grafted on EC4-EC5, SASA was calculated separately for each glycosylation site: trajectories containing the static protein structure and mobile grafted glycans were used to calculate the average shielding stemming from a specific glycan. The final shielding score of a residue was then defined as its maximum shielding score obtained with all glycan conformers. Grafting was performed with 100 ps time steps, CG mode and a 3.5 Å cutoff, yielding at least 7750 conformers per site (see GlycoSHIELD details below).

To ensure a fair comparison with the grafted system, glycans from EC4-EC5 MDS were locally aligned onto the initial protein conformer, yielding a trajectory containing a static protein and mobile glycans. Shielding values were calculated for a random subset of 7750 frames per site.

Generation and refinement of synthetic density maps

Synthetic maps were generated using Gromaps[99] *maptide* function. Whenever a synthetic map was generated for comparison with an experimental map, the later was used as reference to ensure the same map dimensions and voxel size. Gaussian spread parameter (-sigma) was in each case determined by testing several values and comparing the level of detail with the experimental map. To calculate the local correlation between synthetic and experimental maps, protein core and regions beyond reach of glycans had to be masked.

SARS-CoV-2 S protein. Synthetic EDMs shown in Figure 4 were generated with SARS-CoV-2 S protein model PDB: 7EAZ.[24] The protein was glycosylated with GlycoSHIELD with the glycans described in Table S2 using CG mode and a 3.5 Å cutoff. The resulting glycosylated protein model with 1603 glycan conformers per glycosylation site was used to generate synthetic EDMs with Gromaps *maptide* function.[99] Map position and voxel dimensions were defined according to the experimental (*i.e.,* processed EDM) associated with PDB: 7EAZ (EMD-31047).

Missing sections of the N-terminal loops (residues 1-27, 69-80, 143-159, 173-186, 245-264) harbouring the N17, N74 and N149 sites, absent in the original protein structure, were reconstructed and processed separately. In brief, protein structure was modeled with Modeller, keeping disulphide bonds identified in the original structure (C15-C136, C131-C166, C291-C301). To mimic the high mobility of these loops, the reconstruction was repeated 25 times and each of the models used as input for GlycoSHIELD. The numbers of glycan conformers grafted on these 25 models were downsampled to match sampling pseudotime used for the sites present in the original structure. Synthetic EDMs were generated as described above for other parts of the protein.

The unsharpened experimental EDM of SARS-CoV-2 S protein was generated without imposing B-factor sharpening after the refinement steps.[24]

Local correlation map between synthetic and experimental maps were performed using *gmx mapcompare* package, part of Gromaps, using default local correlation range of 0.3 nm, mask cutoff of $1x10^{-5}$, and an external mask (-mask) set as a product of protein only mask (binarized synthetic protein only map, taken at map level > 2.56.) and outer region (binarized synthetic protein+glycan map taken at level $<1x10^{-8}$).

HR2 of S protein. Synthetic cryo-EM maps generated to visualise glycan overlap on the HR2 domain were generated based on 303 conformers grafted with GlycoSHIELD, both for 1- and 3-glycan system). For MD trajectory, glycans together with underlying tripeptides were extracted from frames taken at 1 ns intervals from MDS trajectory, randomized and 303 conformers were grafted back on the original sites with GlycoSHIELD (with no threshold applied). *gmx maptide* was used in each case to generate synthetic maps for glycans, with sigma value 0.1 nm.

FIPV S protein glycans. Synthetic maps corresponding to glycan ensembles used in GlycoDENSITY for the selected glycans were calculated using *gmx maptide* wrapped into a python script, with sigma value 0.1 nm and cutouts of the experimental map used as reference map (-refmap) to ensure the same grid size and spatial location of the map. As the reference map, EMD-9891 dataset was used. Correlation between calculated maps and reference map was performed using *gmx mapcompare* as described above.

Refinement of glycan ensemble and Q-score calculation. GlycoSHIELD conformers were aligned to experimentally elucidated glycan structures by step-wise minimization of the distance between centers-of-mass (COM) of corresponding monosaccharides in the PDB reference structure and in grafted conformers, starting from the GlcNAc residue attached to the Asn anchor and allowing for ambiguity in the glycan branching order (if resolved in the PDB structure). Conformers were rejected from the final ensembles if the distance between COMs exceeded 3.5 Å at each iteration.

Agreement between selected glycan conformer and corresponding segments of the cryo-EM density map, *i.e.*, their "resolvability" was assessed by quantification of Q-scores as described.[63]

3D structure rendering

Rendering of molecular structures and EMDs were performed with PyMOL,[126] VMD[127] or UCSF-ChimeraX.[95] Glycan conformers were subsampled for visualisation of representative conformers on displayed renders.

Multiple glycan conformers (assuming multi-frame PDB file containing glycan ensemble has been generated) can be visualized as follows:

- *Pymol*: open PDB file and issue command set all_states, on in the console to see all glycan conformers.
- *VMD*: Open PDB file, navigate to Graphics > Representations and look for a trajectory tab. There, in the "Draw Multiple Frames" field put 1:1:X where X is the total number of conformers you would like to see.
- ChimeraX: by default ChimeraX fails to connect glycan bonds if CONNECT term is missing in PDB file. To fix this issue, we recommend to first open the PDB file in pymol and then export it as a new multi-frame PDB file (remember to tick all states as well as write CONECT records from all bonds). The new file should be properly recognised by ChimeraX.

Expression and purification of S proteins

Codon-optimized DNA sequences corresponding to residues 1-1209 of SARS-CoV-2 S protein harboring the D614G mutation (defined as S-D614G), residues 1-1190 of SARS-CoV S protein (UNIPROT: P59594), residues 17-1291 of MERS-CoV S protein (UNIPROT: K0BRG7), residues 1-1276 of HKU1-CoV S protein (UNIPROT: Q0ZME7), residues 1-1291 of NL63-CoV S protein (UNIPROT: A0A1L2YVI8), residues 1-1111 of 229E-CoV S protein (UNIPROT: A0A1L7B942) were individually cloned into the mammalian expression vector pcDNA3.4-TOPO (Invitrogen, U. S. A.), which contains a foldon trimerization domain based on phage T4 fibritin followed by a c-Myc epitope and a hexa-repeat histidine tag as previously described.[24,62,128] All constructs contained the fm2P modification, which is defined as the tandem proline replacement (2P; ^{968}KV^{969} → ^{968}PP^{969} for SARS-CoV, ^{1060}VL^{1061} → ^{1060}PP^{1061} for MERS-CoV, ^{1067}NL^{1068} → ^{1067}PP^{1068} for hCoV-HKU1, ^{1052}SI^{1053} → ^{1052}PP^{1053} for hCoV-NL63, and ^{871}TI^{872} → ^{871}PP^{872} for hCoV-229E) and the furin cleavage site mutation where applicable (fm, ^{748}RSVR^{751} → ^{748}ASVG^{751} for MERS-CoV, ^{752}RRKRR^{756} → ^{752}GGSGS^{756} for hCoV-HKU1), for stabilizing the S protein in a prefusion state. Plasmids of all S variants were transiently expressed in HEK293 Freestyle cells or Expi293 cells and purified as previously described.[24,62] In brief, the culture medium was incubated with HisPur Cobalt Resin (Thermo Fisher Scientific, U. S. A.) in binding buffer (50 mM Tris-HCl (pH 7.6), 300 mM NaCl, 10 mM imidazole, and 0.02% NaN_3) at 4ºC overnight. The resin was thoroughly washed with wash buffer (50 mM Tris-HCl (pH 7.6), 300 mM NaCl, and 10 mM imidazole). The recombinant S proteins were eluted by elution buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl, and 150 mM imidazole). The eluted S proteins were further purified by size-exclusion chromatography (SEC) using a Superose 6 increase 10/300 GL column (GE Healthcare, U.S.A.) in TBS buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl, and 0.02 % NaN_3). The protein concentrations were determined using the UV absorbance at 280 nm using a UV-Vis spectrometer (Nano-photometer N60, IMPLEN, Germany).

Glycan analysis by mass spectrometry

In-solution proteolytic digestion of S proteins

Five micrograms of recombinant S proteins were reduced with 10 mM dithioerythritol (DTT) and 6 M urea in 25 mM ammonium bicarbonate at 37°C for 1 hr, then alkylated with 50 mM iodoacetamide (IAA) in 25 mM ammonium bicarbonate for 1 hr in the dark at room temperature (RT). DTT was added to a final concentration of 50 mM and buffer-exchanged to 25 mM ammonium bicarbonate using a centrifugal concentrator (Amicon Ultra-0.5, 10 kDa MWCO, Merck Millipore, Ireland) and treated overnight with a sequencing grade trypsin and chymotrypsin cocktail (Promega, Madison, WI), alpha-lytic protease, or endoproteinase Glu-C at an enzyme-to-substrate ratio of 1:30 at 37 °C. The digests were diluted with formic acid to a final concentration of 0.1% for MS analyses.

In-gel proteolytic digestion of S proteins

Three micrograms of recombinant S proteins were separated by gel electrophoresis and stained by One-Step Blue (Biotium, U.S.A.) for visualization. The target protein bands were excised and diced into small pieces (1 mm^3). The diced gel pieces were reduced by 10 mM DTT in 25 mM ammonium bicarbonate at 37 °C for 1 hr, followed by alkylation with 50 mM iodoacetamide in 25 mM ammonium bicarbonate in the dark at RT for 1 hr. The gel pieces were destained by incubation in 25% acetonitrile in 25 mM ammonium bicarbonate at 37 °C for 15 minutes. The destaining steps were repeated until the gel pieces were fully destained. The gel pieces were then dehydrated in 100% acetonitrile (can) and dried in a Speed Vac. The dried gel pieces were then digested overnight by sequencing grade trypsin at an enzyme-to-substrate ratio of 1:50 at 37 °C, followed by a second digestion step using sequencing grade chymotrypsin at an enzyme-to-substrate ratio of 1:30 at 37 °C. The digests were placed in a clean tube for peptide extraction. Deionized water was added to the samples followed by sonication for 15 minutes, and supernatants containing the peptide digests were collected in a clean tube. The same extraction step was repeated twice by replacing deionised water with 1% formic acid (FA) and 50% ACN/1% FA. The extracted peptide digests were pooled and dried in a Speed Vac. The final samples were dissolved in 0.1% FA for further experiments.

Liquid chromatography-mass spectrometry

MS data were acquired using an Orbitrap Fusion Lumos Tribrid mass spectrometer (Thermo Fisher Scientific, San Jose, CA) equipped with an Easy-nLC 1200 system (Thermo Fisher Scientific). For each LC–MS/MS analysis, 0.5 μg of glycoprotein digests was loaded onto an Acclaim PepMap RSLC C18 column (Thermo Fisher Scientific) and separated at a flow rate of 300 nL/min using a

gradient of 5% to 40% of 80% acetonitrile plus 0.1% formic acid over 200 min. The mass spectrometer was operated in the Top speed (3 s) data-dependent acquisition mode. Glycopeptide fragmentation was achieved by collision-induced dissociation (CID) and higher-energy collision dissociation (HCD) as described previously[64] but without electron transfer dissociation (ETD). Briefly, survey full scan MS spectra were acquired in the Orbitrap from 400 to 1800 *m/z* at a mass resolution of 120,000. The highest charge state ions within charge state 2–6 were sequentially isolated for MS2 analysis using the following settings: HCD MS2 with AGC target at 5×10^4, isolation window 2, orbitrap resolution 30,000, step collision energy (%): 25, 28 and 32.

Glycopeptide identification and quantification

Raw files derived from the LC-MS/MS analyses were processed in Byonic v 3.9.6 for N-glycopeptide identification using the following parameters: the sequences of individual S proteins were defined in the search library together with the definitions of individual cleavage sites specific to the enzyme(s) used for sample preparation (F, K, L, R, W, Y residues for trypsin plus chymotrypsin digestion samples; E, K, R residues for trypsin plus endoproteinase Glu-C digestion samples; and A, S, T, V residues for alpha-lytic protease digestion samples). Up to two miscleavages were allowed in the data search. The "Both: CID & HCD" option of the fragmentation type was selected, with the precursor mass tolerance at 5 ppm and the fragment mass tolerance at 10 ppm. Fixed modification included cysteine carbamidomethylation (+57.0215 Da, at C), while variable common modification contained methionine oxidation (+15.9949 Da, at M), and variable rare modifications consisted of asparagine and glutamine deamidation (+0.9840 Da, at N, Q). For N-glycopeptide identification, the built-in N-glycan library of "182 human" was used. Next, the unfiltered Byonic search results and the MS raw files were analyzed by the Byologic module of the Byos suite (v3.11, Protein Metrics Inc., USA) to quantify the N-glycopeptide populations of individual glycosylation sites, based on the integration of peak areas of extracted ion chromatograms (XIC-AUC). The quantification results derived from Byonic and Byologic were analyzed by an in-house python script. Positive matches required a score > 200 and PEP2D < 0.001. Key attributes, including "N-glycosylation site", "sequence", "glycan composition" and "calculated m/z", were summed by using the number of PSMs, within which the highest score match was considered as "unique glycopeptide", and its peak area under the extracted ion chromatograms was treated as a quantified entry.

For the final outputs, all unique glycopeptides were classified into glycan categories as described previously.[46] The peak areas of individual glycopeptides were summed and normalized to generate bar charts (see ancillary figures in key resources table). The results obtained from different sample treatments were compiled side-by-side for individual N-glycosylation sites to document the reproducibility of our analyses.[9] For each N-glycosylation site, the relative proportions of high-mannose, hybrid, and complex type glycans were presented in pie charts (Figures 5 and S8). The complete dataset has been deposited in the MassIVE database (see key resources table).

Selection of representative S protein glycoforms

The representative glycoform selection for further modeling with GlycoSHIELD was based on the following principles. Because not all of the S proteins contained the same digestion protocols with the same enzyme combinations, we first focused on the samples treated with the same enzyme combinations, within which we selected the dataset that covered the most N-glycosylation sites as the representative glycoforms. For sites that were not covered by the representative dataset, we chose the glycoforms derived from the second most complete dataset as representatives, and the third most representative dataset to fill in the gaps that remained, etc. After exhausting all the available datasets, putative N-glycosylation sites that lacked experimental data were modeled with Man5 in the final GlycoSHIELD modeling steps.

Cryo-EM structural analyses

Data collection and processing

SARS-CoV-2 S-D614G. The collection and processing of cryo-EM data for this protein have been described previously.[24] The unsharpened EM density map of the protein was created in CryoSPARC[92] v3.2 by three-dimensional variability analysis followed by heterogeneous refinement and downsampling using Fourier-cropping from 384 to 128 pixels. The map and corresponding particles were then used and re-extracted to generate the final high-resolution map (PDB: 7EAZ / EMD-31047).

MERS-CoV S protein. Cryo-EM imaging of the protein was performed on a 200 kV Talos Arctica microscope (ThermoFisher Scientific) equipped with a Falcon III electron detector in a linear detection mode. The micrograph stacks were collected with a defocus range between -1.5 and -1.7 μm, and a magnification of 92000x, corresponding to a pixel size of 1.1 Å. A total dose of 40 $e^-/Å^2$ was distributed over 50 frames with an exposure time of 2.5 s at each imaged location on the sample grid. A total of 2886 super-resolution micrograph stacks were imported to Relion-3.1, and analysed with dose-weighting and 5x5 patch-based alignment using motion-correction features in Relion.[129] All motion-corrected micrographs were transferred to CryoSPARC v3.2 for contrast transfer function (CTF) estimation.[91] The micrographs were filtered by the ''CTF_fit_to_Res'' with a cut-off distance between 2.6 and 8 Å. The 2808 micrograph stacks that filled this criterion were used for particle picking with a minimum and maximum particle diameter of 120 and 160 Å, respectively. A total of 1,690,870 particles were picked with a box size of 256x256 pixels, and Fourier-cropped, binned to a box size of 128x128 pixels and then subjected to iterations of 2D classification. The resulting 720,333 particles were used for *ab-initio* reconstruction with C1 symmetry, followed by heterogeneous refinement to generate five distinct classes. The particles from the best 3D-classes were subjected to non-uniform (NU) refinement to generate an initial map and its corresponding mask. The initial 3D-map with a refined mask from the NU refinement was used for three-dimensional variability analysis. Five clusters generated from the 3DVA as the templates were used for further heterogeneous refinement and NU refinement. Finally, the particles from different 3D classes were re-extracted, un-binned, and applied to NU refinement with C1 symmetry (except for the particles from ''All-RBD

down'' conformations with C3 symmetry) to generate maps. The 3D maps were further applied to a second round NU refinement with the corresponding local CTF refined particles. Then followed by local refinement to generate the final eight cryo-EM maps based on which seven protein structural models were built (see figure Methods S1 and Table S3).

Building and refinement of MERS-CoV S models

The initial model of the ''All-RBD down'' conformation of MERS-CoV S protein was generated based on the previously reported structure (PDB: 6Q04)[90] aided by Swiss-Model.[13] The atomic coordinates were divided into individual domains, manually fitted into the cryo-EM map in UCSF-ChimeraX and optimized using Coot.[94] Next, the model was processed by real space refinement in Phenix[93] with the default parameters, additional non-crystallographic constraints, and rigid body refinements. To improve the correlation between the model and cryo-EM map, iterative rounds of optimization in Coot and real space refinement in Phenix were carried out without rigid body refinement. Finally, real-space refinements were performed in Phenix, with an additional non-bonded weight parameter of 500 to decrease clash scores. N-linked glycans were added onto asparagine side-chains by using the extension module ''Glyco'' within Coot. The final model was validated using Phenix. The same procedures were applied to generate the atomic models of ''Intermediate'', ''One-RBD up'', and ''Two-RBD up'' conformations. Structural visualization and rendering of structural representations were accomplished by using UCSF-ChimeraX and Pymol 2.3.4. All MERS structures and corresponding cryo-EM maps are available in PDB and EMDB with the IDs described in Table S3.

Expression and purification of recombinant N-cadherin variants

The open reading frame of mouse N-cadherin cDNA (NCBI accession number: NP_031690.3) encompassing its entire ectodomain (EC1-EC5, residues 160-712) and domains EC4-EC5 (residues 331 to 542) were synthesized by GenScript. These sequences were modified by addition of the N-terminal signal peptide of CD33 and a C-terminal hexahistidine (His_6) tag, edited by codon optimization for expression in human cells and subsequently cloned into the mammalian expression vector pcDNA3.1 (+) (Invitrogen) as *NheI* and *XhoI* inserts.

Expi293 cells were transfected with resulting plasmids with ExpiFectamine™ 293 (Invitrogen) according to the manufacturer's instructions. The recombinant proteins secreted into the culture medium were purified by binding to Ni-NTA affinity resin (Cytiva). Washes were performed with a buffer containing 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 20 mM Imidazole, and eluted in a washing buffer supplemented with 250 mM Imidazole. Eluted EC1-EC5 and EC4-EC5 proteins were further purified by size-exclusion chromatography (SEC) on Superdex 200 increase 10/300 GL columns (Cytiva) and Superdex 75 increase 10/300 GL columns (Cytiva), respectively, on an AKTA FPLC system (Cytiva) in a final buffer containing 10 mM HEPES (pH 8.0), 150 mM NaCl, 3 mM $CaCl_2$. Protein concentrations were determined by measuring UV absorbance at 280 nm according to their extinction coefficients derived from their amino acid sequences.

Size-exclusion chromatography-coupled small-angle X-ray scattering (SEC-SAXS)

SAXS profiles of the N-cadherin constructs (EC1-EC5 and EC4-EC5) were collected at the BioSAXS beamline 13A of Taiwan Photon Source (TPS) at the National Synchrotron Radiation Research Center (NSRRC) in Hsinchu, Taiwan.[130] The apparatus is coupled to an on-line high-performance liquid chromatography (HPLC) system (Agilent) allowing separation of distinct oligomeric states of the protein by size-exclusion chromatography prior to SAXS data collection. 100 μL of sample solution at a concentration of 10 mg/ml were injected in the HPLC system and separated on an BioSEC-3 300Å 4.6×300 mm analytical size-exclusion chromatography column (Agilent, USA) at a flow rate of 0.35 mL/min. The flow rate was reduced to 0.1 mL/min immediately before SAXS data collection to obtain sufficient data frames for post-processing. The exposure time of each frame was set to two seconds. SAXS data were collected with a momentum transfer (q) ranged between 0.006 and 1.770 $Å^{-1}$ at a wavelength of 0.8265 Å, 15 keV, using an integrated detection system comprising an Eiger X 9M SAXS detector (Dectris, Switzerland). An in-house *LabVIEW*-based software was used for data reduction, solvent subtraction and merging of the SAXS data.[97] SAXS data were further processed and analyzed using PRIMUS (ATSAS 3.2.1 software package) to estimate the R_g and D_{max} value based on Guinier approximation and the pairwise distance distribution function P(r).[96] The P(r) profiles were used as inputs for GASBOR to generate the Molecular envelops of the protein were generated from P(r) profiles using GASBOR.[131] ''Bead models'' were superimposed to the crystal structures of full-length N-cadherin (EC1-EC5) and truncated EC4-EC5 with or without GlycoSHIELD ensembles obtained with distrinct glycan types. Theoretical SAXS profiles of the atomic glycan-conjugated models were back-calculated using FoXS* for comparison with the experimental data.[132,133]

Molecular volume calculations

Volumes of the $\alpha1\beta3\gamma2$ $GABA_A$ receptor inner pore shown in Figures 7F and 7G were computed with POVME[134] using a 1.7Å grid spacing and a distance cut-off of 1.09 Å from protein Van der Waals atom surface. Volumes were calculated with 1 to 42 glycan conformers per glycosylation site and plotted as fractions of the pore volume of the non-glycosylated receptor as a function of pseudotime.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using Prism9 (Graphpad, USA). Differences in the distributions of EC4-EC5 gyration radius shown in Figures 3C and 3D were assessed using an impaired Kolmogorov-Smirnov test. Average residual χ^2 values in Figures S4C and S4F were compared using Kruskal-Wallis' test for multiple comparisons with Dunn's correction. The significance of correlation plots shown in Figures S6E and S6F was determined using Spearman correlation tests. Details of the different quantitative analyses reported appear in the relevant section of the method details.

ADDITIONAL RESOURCES

A standalone web application available under www.glycoshield.eu provides access to GlycoSHIELD capabilities to non-expert users. Gitlab web page of the project (https://gitlab.mpcdf.mpg.de/dioscuri-biophysics/glycoshield-md) contains additional tutorials and convenience scripts. Listed resources will also contain any future updates and/or extensions to the GlycoSHIELD method and glycan conformer database.

CellPress
OPEN ACCESS

Supplemental figures

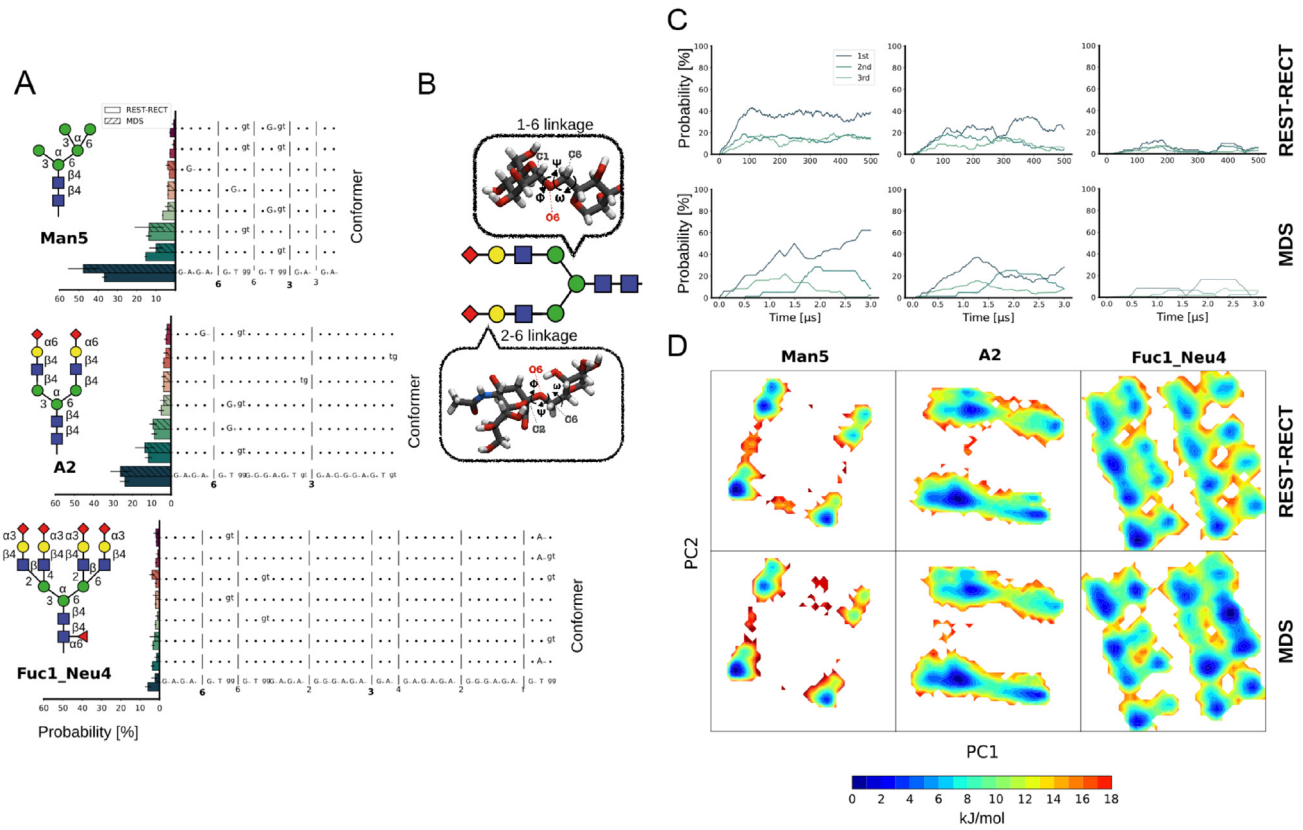


Figure S1. Comparison of REST-RECT with unbiased MDS, related to Figure 1 and STAR Methods

(A) Conformer distributions for both simulation types. The conformer string is given on the y axis where each digit stands for a torsion angle, the letter representing the occupied minima (T, *trans*; C, *cis*; G, gauche; A, anticlinal). Dots are used instead of letters when no change in occupancy could be observed in comparison with the most occupied conformer (see details in STAR Methods). Shown in the probability histograms are means ± SD.

(B) Definitions of torsional angles used in (A).

(C) Moving average of conformer probabilities for the three most populated conformers as defined in (A) over a temporal window of 100 ns (REST-RECT) or 1.2 µs (MDS), corresponding to the same effective sampling time.

(D) Principal component analysis of the conformer distribution converted to free energy surface.

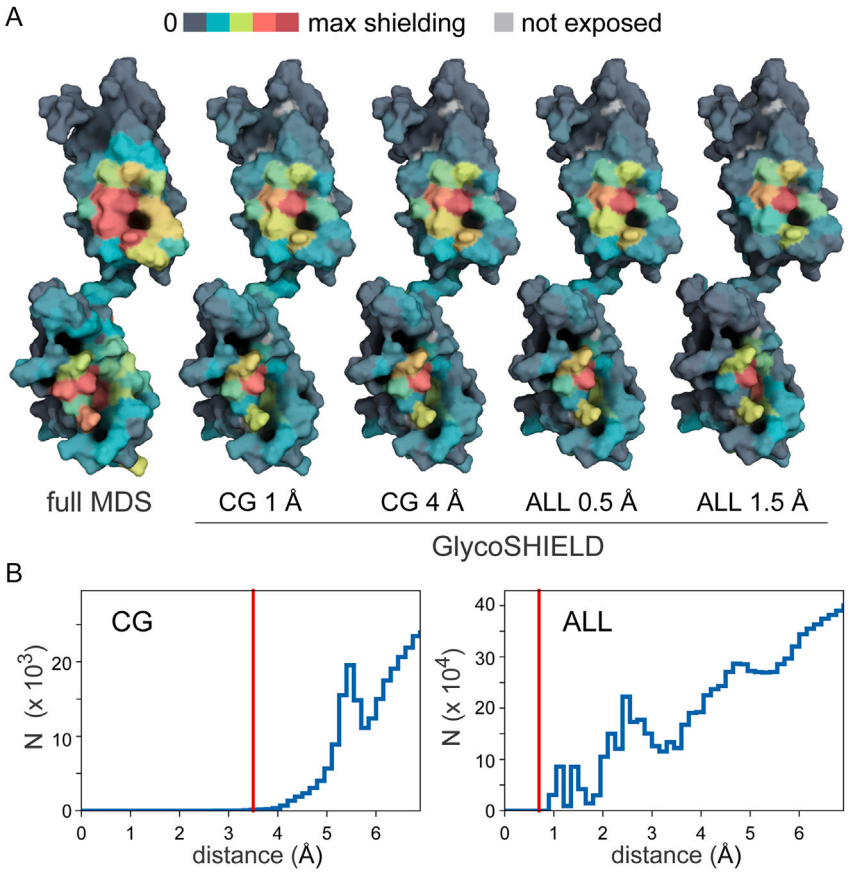


Figure S2. EC4-EC5 shielding estimated by MDS and GlycoSHIELD, related to Figure 1

(A) 3D-heatmaps of protein shielding (i.e., glycan-dependent reduction of solvent accessible surface area; see STAR Methods) by glycan conformers computed by MDS or GlycoSHIELD for multiple exclusion distances between protein α-carbons and glycan ring oxygens (coarse grain or CG) or all protein and glycan atoms (ALL).

(B) Left: distribution of glycan oxygen atoms at increasing distance from protein α-carbons (CG) calculated from 1 μs trajectory at 1 ns intervals; N: occurrence counts; right: distribution of glycan atoms at increasing distance from protein atoms. Threshold values selected for shield reconstructions (e.g., Figure 1C) are indicated in red. In (A), note the good agreement of shielding heatmaps derived from MDS and GlycoSHIELD.

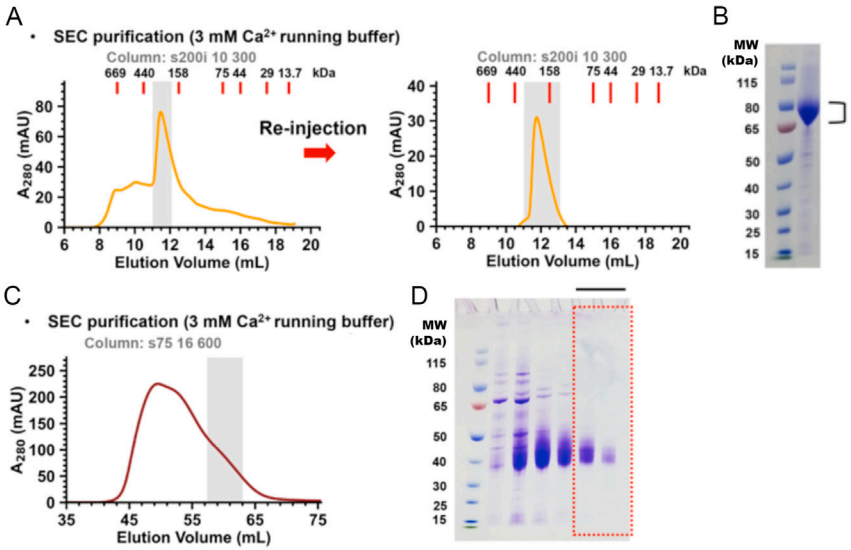


Figure S3. Purification of N-cadherin ectodomain and EC4-EC5 after expression in HEK cells, related to Figure 1

(A and B) Chromatograms (A) and SDS-polyacrylamide gel picture (B) obtained after purification of mouse cadherin-2 ectodomain on Nickel resin (see STAR Methods).

(C and D) Chromatogram (C) and SDS-polyacrylamide gel picture (D) obtained after purification of mouse cadherin-2 EC4-EC5 on Nickel resin (see STAR Methods). The orange rectangle in the bottom right picture shows the eluted fractions that were used for structural evaluation.

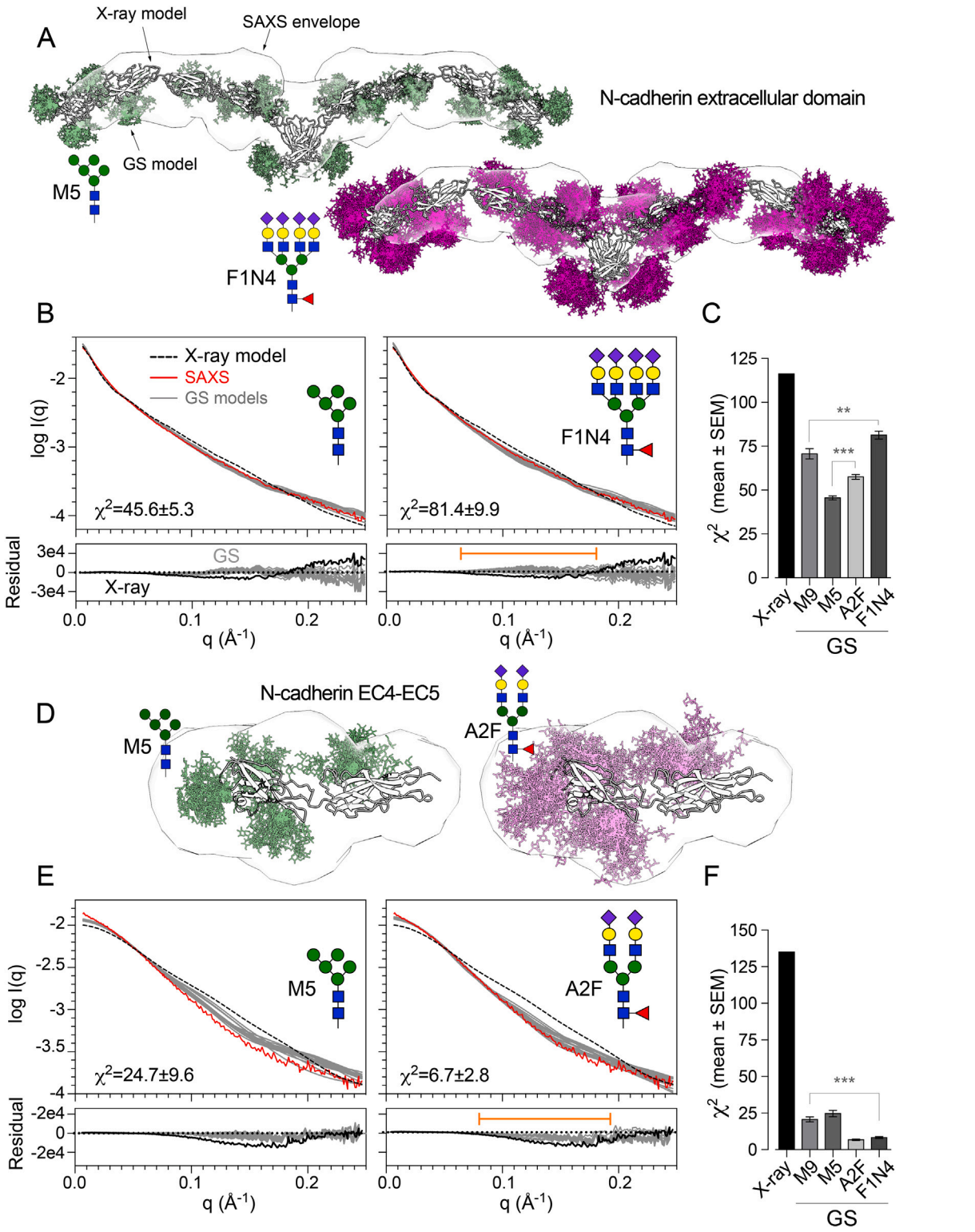


Figure S4. GlycoSHIELD improves fitting of SAXS data, related to Figure 1

(A–C) SAXS analysis of glycosylated N-cadherin extracellular domain (EC1-EC5). (A) SAXS molecular envelop and structure of N-cadherin (PDB: 3Q2W) modeled with GlycoSHIELD with Man5 (M5) or Fuc1_Neu4 (F1N4) N-glycans. (B) Upper: SAXS intensity in log scale as a function of the scattering vector q (red trace) and curve fit with non-glycosylated (naked X-ray model, "X-ray," black trace) or glycosylated protein models (gray) generated with Man5 (M5) or Fuc1_Neu4 (F1N4)

(legend continued on next page)

N-glycans. χ^2 values are indicated (mean ± SD). Lower: corresponding residual values (error-weighted) as a function of q after curve fitting with non-glycosylated (X-ray, black) or glycosylated (GS, gray) protein models. (C) Average (mean ± SEM, n = 23 conformers) of χ^2 values after SAXS plot fit with naked (X-ray) or glycosylated models (GS) obtained with distinct N-glycans. **$p < 0.001$; ***$p < 10^{-4}$, Kruskal-Wallis' and Dunn's tests for multiple comparisons.
(D–F) SAXS analysis of glycosylated EC4-EC5. (D) SAXS molecular envelop and structure of EC4-EC5 (taken from PDB: 3Q2W) modeled with GS with M5 or A2F. (E) Upper: SAXS intensity in log scale as a function of the scattering vector q (red trace) and curve fit with naked (X-ray, black trace) or glycosylated (gray) protein models with Man5 (M5) or A2F. χ^2 values are indicated (mean ± SD). Lower: corresponding variations of residual values (error-weighted) as a function of q after curve fitting with naked (X-ray, black) or glycosylated (GS, gray) protein models. (F) Average (mean ± SEM, n = 23 conformers) of χ^2 values after SAXS plot fit with naked (X-ray) or glycosylated models (GS) obtained with distinct N-glycans. **$p < 0.001$; ***$p < 10^{-4}$; ANOVA and Kruskal-Wallis' and Dunn's tests for multiple comparisons.
Note the improved fitting of experimental data with glycosylated models generated with GS (see in particular plot sections highlighted by orange bars in B and E and the sensitivity of the curve fitting to N-glycan type).

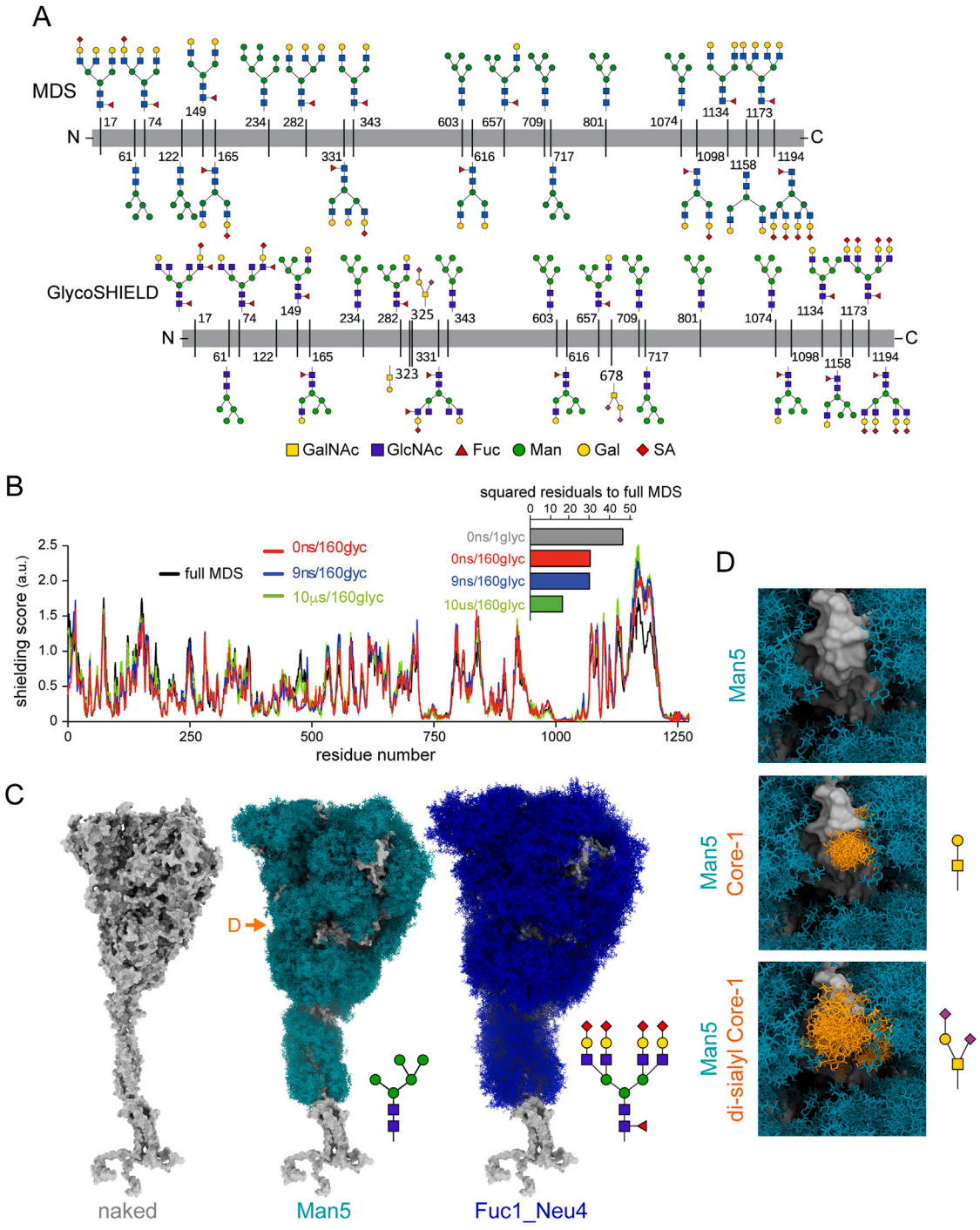


Figure S5. Glycosylation profile and shielding of SARS-CoV-2 S-protein, related to Figure 3

(A) Glycosylation profile of S-protein used for MDS and GlycoSHIELD in Figures 3A and 3B.

(B) Shielding of S-protein residues (ray method) calculated from 10 μs MDS (black) and GlycoSHIELD models obtained with either a single protein conformer and single glycan conformer (gray), a single protein conformer and 160 glycan conformers (red), 9 protein conformers and 160 glycan conformers (blue), and 100 protein conformers sampled at 100 ns intervals with 160 glycan each. Inset shows corresponding sums of squared residuals of shielding in respect to full MDS.

(C and D) S-protein modeled without glycans or with GlycoSHIELD with all Man5 or all Fuc1_Neu4 N-glycans, without (C) and with O-glycans. The orange arrow indicates the position of the protein segment shown at higher magnification in (D) with Man5 alone and with two distinct types of O-glycans.

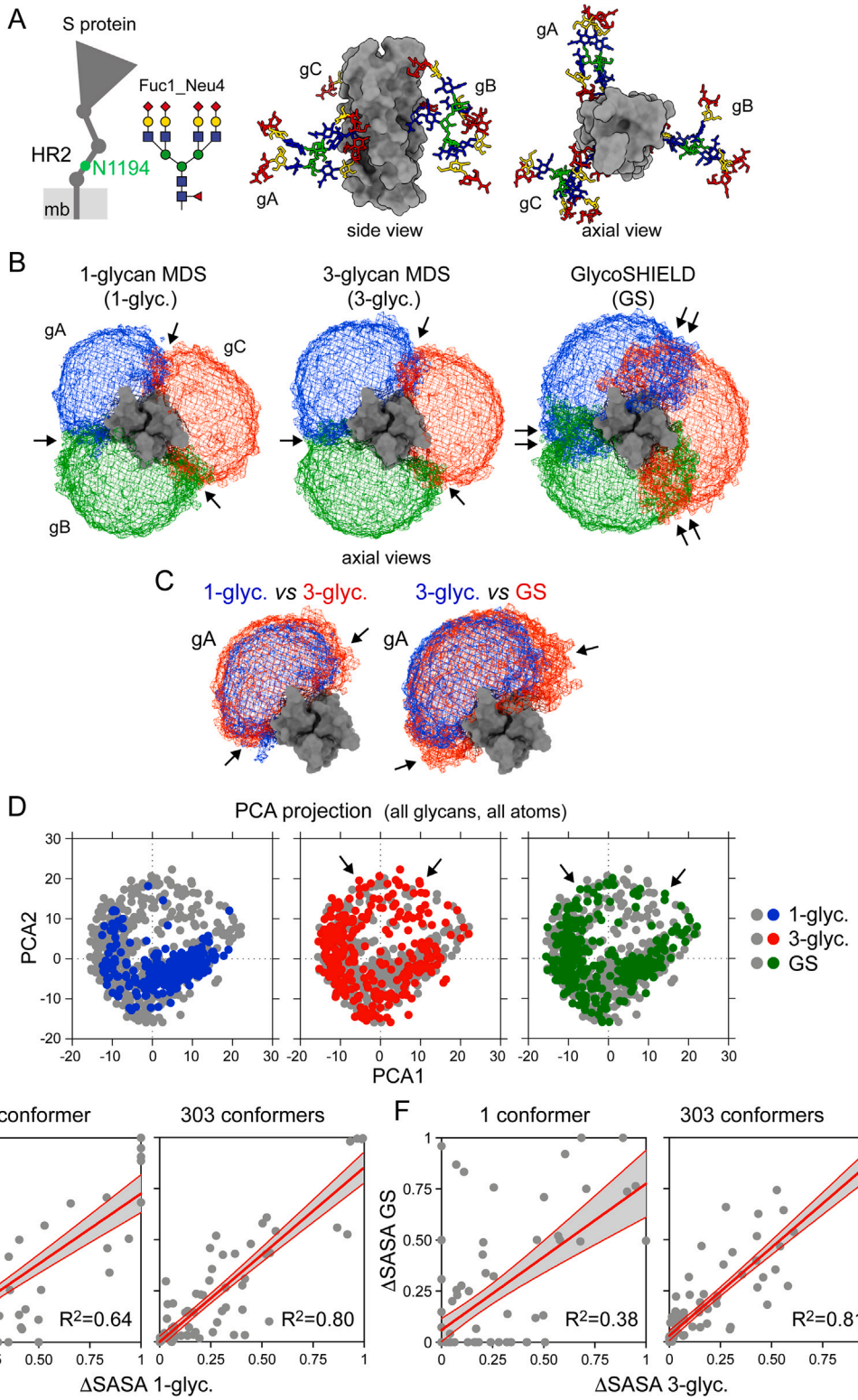


Figure S6. Glycan overlap in shields generated by GlycoSHIELD or MDS, related to Figure 3

(A) Scheme of SARS-CoV-2 S protein and 3D models of the HR2 domain trimer with Fuc1_Neu4 N-glycans at N1194 on each monomer (gA, gB, and gC).

(B) Contours of the glycan synthetic density maps generated from MDS of the HR2 domain trimer with a single N-glycosylated monomer (1-glyc.) or 3 glycosylated monomers (3-glyc.) and GlycoSHIELD model (GS). Note the larger overlap in shields generated with GS (arrows).

(legend continued on next page)

(C) Overlaid contours of glycan density maps generated by 1-glyc. (blue) and 3-glyc. (red) MDS; or 3-glyc. MDS (blue) and GS (red). Note the extended sides of contours obtained with 3- vs. 1-glyc. and GS vs. 3-glyc. (arrows).
(D) Principal component analysis (PCA2-PCA1 plots, all atoms) of corresponding N-glycan conformer arrays. Shown are values obtained for 1-glyc., 3-glyc., and GS conformers after projection on the same PCA space. Note the broader conformer sampling after MDS of the 3-glycans together or modeling with GS (arrows).
(E and F) Correlation of surface shielding (ΔSASA, see STAR Methods) determined after MDS of the HR2 trimer with individual glycans or the 3 glycans together (1-glyc. vs. 3-glyc.) (E), and 3-glyc. MDS vs. GS using single or all available conformers (F). Linear fits and 95% confidence intervals are shown by red lines and gray areas. R^2 values are indicated. In (E) and (F), note the improvement of shielding prediction and agreement between GS and MDS when considering multiple conformers (see STAR Methods).

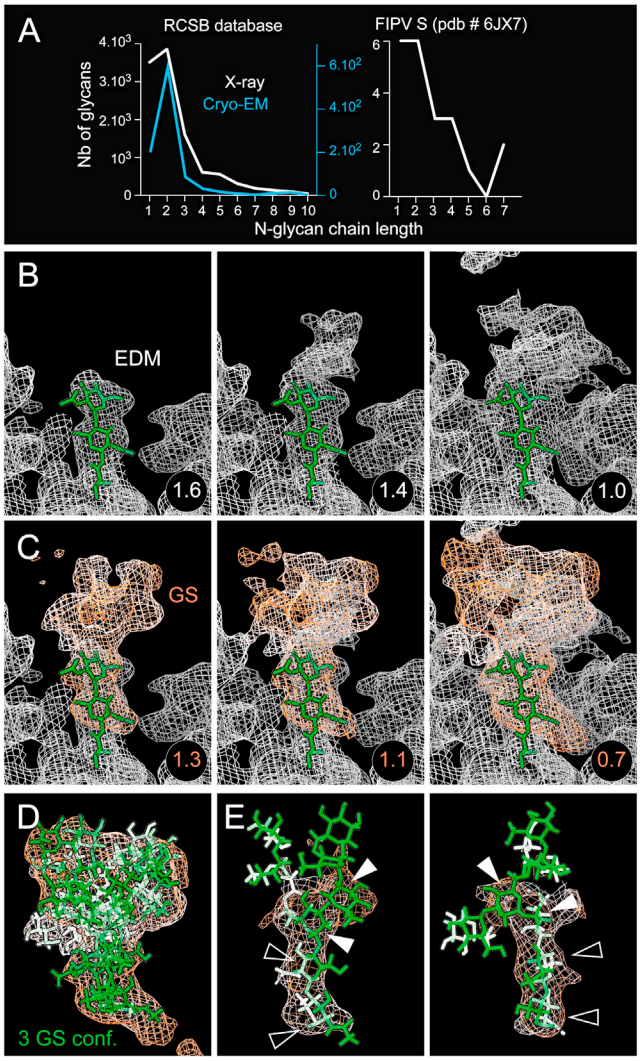


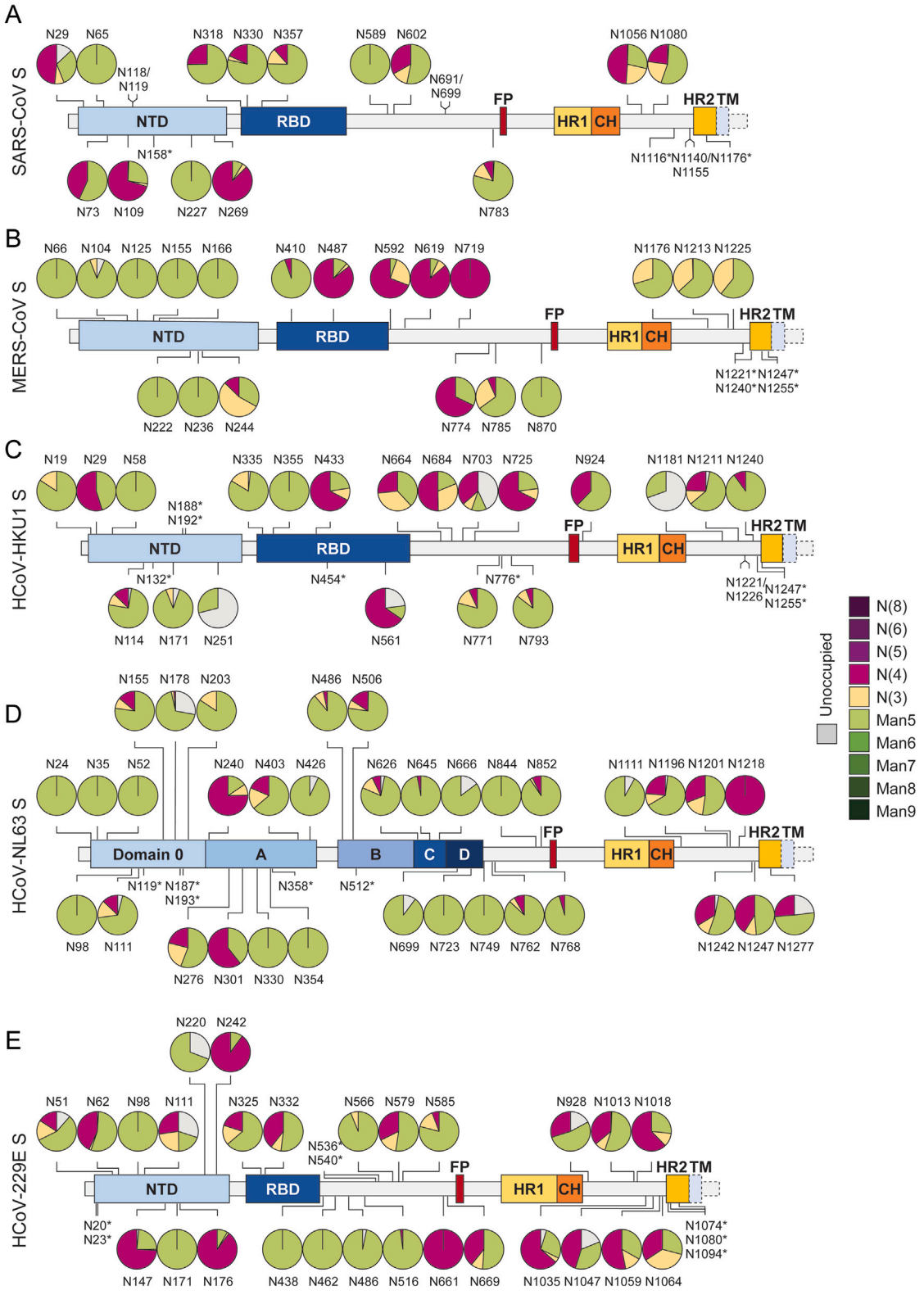
Figure S7. Correlation between GlycoSHIELD and cryo-EM density maps reveals unresolved features of glycan structures, related to Figure 4

(A) Distribution of N-glycan length (number of monosaccharides) in glycoprotein structures resolved by X-ray crystallography or cryo-EM available at the RCSB Protein Data Bank (left) compared with resolved glycans of the feline infectious peritonitis virus (FIPV) S protein (PDB: 6JX7,[28] right).

(B and C) Manually built structure of FIPV S protein N491 N-glycan (2 resolved GlcNAc monosaccharides) shown with experimental EM density map (EDM, gray meshes in B and C) and GlycoSHIELD density map obtained with Man9 and refined with GlycoDENSITY (ref. GS, red in B, see STAR Methods) at different contour levels. Contour thresholds are indicated.

(D) High-level contours of refined GS map shown with 3 superimposed Man9 GS conformers with high local correlation with cryo-EM density map.

(E) Low-level contours of refined GS map shown with a single Man9 conformer at 2 different angles, allowing capture of 3 monosaccharides (filled white triangles) in addition to the initial 2 resolved in the original structure (empty triangles).



(legend on next page)

Figure S8. Quantitative glycosylation analyses of SARS-CoV, MERS-CoV, hCoV-HKU1, hCoV-NL63, and hCoV-229E S proteins, related to Figure 5

The pie charts summarize the relative quantities of N-glycans as high-mannose type (green), hybrid type (yellow), and complex type (purple), determined by MS analyses (see STAR Methods). The positions of N-glycosylation sites are indicated in the schematic representations of the protein primary structures (the regions that were not included in the protein expression constructs are highlighted with dash lines). NTD, N-terminal domain; RBD, receptor-binding domain; FP, fusion peptide; HR1/HR2, heptad repeat 1/2; CH, central helix; TMD, transmembrane domain.