3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development

Oliver C. Grant, David Montgomery, Keigo Ito, Robert J. Woods*

Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd, Athens, GA 30602

Corresponding Author
*Mailing address: 315 Riverbend Road, Athens, GA 30602.
Tel.: 706-542-4454. Fax: 706-542-4412. E-mail: rwoods@ccrc.uga.edu

Abstract

Here we have generated 3D structures of glycoforms of the spike (S) protein SARS-CoV-2, based on reported 3D structures for the S protein and on reported glycomics data for the protein produced in HEK293 cells. We also report structures for glycoforms that represent those present in the nascent glycoproteins (prior to enzymatic modifications in the Golgi and ER), as well as those that are commonly observed on antigens present in other viruses.

These models were subjected to MD simulation to take into account protein and glycan plasticity, and to determine the extent to which glycan microheterogeneity impacts antigenicity. Lastly, we have identified peptides in the S protein that are likely to be presented in human leukocyte antigen (HLA) complexes, and discuss the role of S protein glycosylation in potentially modulating the adaptive immune response to the SARS-CoV-2 virus or to a related vaccine.

Introduction

The present COVID-19 pandemic has led to over a million confirmed infections with a fatality rate of approximately 5 percent [1] since the first reports of a severe acute respiratory syndrome (SARS) infection by a novel coronavirus (SARS-CoV-2) at the end of 2019. As of April 2020, there is still no vaccine or approved therapeutic to treat this disease. Here we examine the structure of the SARS-CoV-2 envelope spike (S) protein that mediates host cell infection, with a specific focus on the extent to which S protein glycosylation masks this virus antigen from the host immune response.

Viral envelope proteins are often modified by the attachment of complex glycans that can account for up to half of the molecular weight of these glycoproteins, as in HIV gp120 [2]. The glycosylation of these surface antigens helps the pathogen evade recognition by the host immune system by cloaking the protein surface from detection by antibodies, and can influence the ability of the host to raise an effective adaptive immune response [3,4] or even be exploited by the virus to enhance infectivity [5]. Fortunately, the innate immune system has evolved a range of strategies for responding to glycosylated pathogens [6], but antigen glycosylation nevertheless complicates the development of vaccines [7]. Over time, the protein sequences in viral antigens undergo mutations, which can alter the species specificity of the virus [8], modulate its infectivity [9], and alter the antigenicity of the surface proteins [10]. These mutations can also impact the degree to which the protein is glycosylated by creating new or removing existing locations of the glycans (glycosites) on the antigens [11,12]. Varying surface antigen glycosylation is thus a mechanism by which new virus strains can evade the host immune response [11], and attenuate the efficacy of existing vaccines [7].

Very recently, a cryo-EM structure of the SARS-CoV-2 S glycoprotein has been reported [13], which led to conclusion that, like the related protein from the 2002 - 2003 SARS pandemic (SARS-CoV-1) [14], the CoV-2 S protein is also extensively glycosylated [13]. Furthermore, an analysis of the glycan structures present at each glycosite in the S protein produced recombinantly in human embryonic kidney 293 cells has also been recently reported [15].

Here we have generated 3D structures of several glycoforms of the SARS-CoV-2 S glycoprotein, in which the glycans represent those present in the S protein produced in HEK293 cells [15], as well as those corresponding to the nascent glycoprotein (prior to enzymatic modifications in the Golgi), as well as those that are commonly observed on

antigens present in other viruses [16-18]. We have subjected these models to long MD simulations and compare the extent to which glycan microheterogeneity impacts epitope exposure. Additionally, we have identified peptides in the S protein that are likely to be presented in human leukocyte antigen (HLA) complexes, and discuss the role of S protein glycosylation in modulating the adaptive immune response to the SARS-CoV-2 virus or to a related vaccine.

Results

The impact of glycosylation on the ability of antibodies to bind to a pathogenic glycoprotein may be estimated by quantifying the fraction of the surface area of the protein antigen that is physically shielded by glycans from antibody recognition. Such an analytical approach to defining glycoprotein antigenicity led to the conclusion that glycosylation of the E2 envelope protein of hepatitis C virus reduces its surface area by up to 65% [19]. The structure of proteins may be readily determined experimentally using techniques such as x-ray crystallography or cryo-electron microscopy (cryo-EM), however, it is far more challenging to determine the structure of the glycans. This is because glycans are often significantly more flexible than the proteins to which they are attached, and also because multiple different glycan structures may be present at any given glycosite, giving rise to an ensemble of glycosylated protein variants (glycoforms). This glycan heterogeneity is a natural result of the mechanism of protein glycosylation. While all glycans that are added to proteins through attachment to asparagine side chains (N-linked glycans) are transferred *en bloc* as a single glycan precursor structure, subsequent passage of the nascent glycoprotein through the cellular machinery (Endoplasmic reticulum, Golgi apparatus) exposes the glycans to enzymes, glycosidases and glycosyl transferases, that modify their structure. The ability of enzymes to alter or process each glycan depends in part on the 3D shape of the protein to which the glycan is attached, particularly in the region of the glycosite [16,20,21], and on the time that any given glycan is exposed to these enzymes [20,22]. The presence of multiple glycosites and the variation in the composition of the glycans at a glycosite, can result in a staggering number of permutations of protein glycoforms. Although the particular structure of the glycans at the glycosites may not impact protein immunogenicity, as reported in the case of hepatitis C virus [19], the location and number or density of glycosites can have a profound impact on the innate immune response to virus [23] and on the antigenicity of viral envelope proteins [24,25], as observed in the case of influenza A. Lastly, to further complicate our understanding of the effect of glycosylation on protein antigenicity, glycans display large internal motions that prevents their accurate description by any single 3D shape, in contrast to proteins [26,27].

Fortunately, computational simulations can play a key role in characterizing glycoproteins by predicting the likely 3D shapes of the glycans [16,20,28-30]. In a typical simulation, the glycans are added to the experimentally-determined 3D structure of the protein, and the resulting static model of the glycoprotein is then subjected to molecular dynamics (MD) simulations that approximate the effects of water and temperature on the temporal and spatial properties of the glycoprotein over the simulation time frame. The 3D structures obtained from the simulation can then be used to determine the extent to which glycosylation has altered the protein's antigenicity [19] or other properties. When the structure of specific glycans at particular glycosites is known, these glycans can be employed in the simulation [19,20,31,32], however this data is often unavailable and plausible glycosylation states may have to be assumed, particularly in the case of proteins with multiple permutations of glycoforms [16].

Model glycoforms. It is well established that there is a strong dependence of both the composition and relative glycan abundance (glycan microheterogeneity) on the cell type used in glycoprotein production. And there is a large body of data relating to the influence of host cell line on viral envelop protein glycosylation. For example, a glycomics analysis of influenza A virus produced in five different cell lines, all of relevance to vaccine production, let to the observation of profound differences in the compositions of the glycans at a given site; with structures varying from paucimannose (Sf9 cells) to core-fucosylated hybrid with bisecting N-acetylglucosamine (Egg) to sialylated biantennary glycans (HEK293) [18]. For these reasons, we have modeled the S glycoprotein with reported site-specific glycosylation [15], as well as hypothetical homogeneously glycosylated glycoforms of the high mannose (M9), paucimannose (M3), biantennary complex (Complex) and core-fucosylated biantennary complex (Complex Core F) types. Comparisons among the glycoforms permits an assessment of the impact of cell-based differential glycan processing on S protein antigenicity.

Notably, differences in glycan microheterogeneity have been shown not to significantly affect the potency of influenza antigens, as determined in single radial immunodiffusion (SRID) assays; an assay that is typically used to measure the potency of commercial influenza vaccines [17]. That the precise composition of the glycans may not alter antigenicity, suggests that glycan microheterogeneity may not heavily impact the exposed surface area of the protein. Nevertheless, such glycan variations have been shown to impact the ability of innate immune lectins, such as surfactant protein-D (SPD), to respond to the antigen [16,23]. Because SPD requires a cluster of high mannose glycans on the pathogenic protein in order to bind [16], variations in the glycan composition can impair such a response. Again, in the case of influenza, increases in the number of glycosites have been shown to impact receptor binding, protein folding, virulence, antigenicity, and shielding of immunogenic sites [24,33-36]. To assess the impact glycosylation on S protein antigenicity, an initial comparison with other more-fully characterized coronaviruses was undertaken.

Comparison with epitopes in related coronavirus S glycoproteins. Viral adhesion to host cells is enabled by proteins on the virus surface (or envelope) that evolve to recognize specific receptors on the host cell. The receptors may be glycans, as is well known in the case of influenza, endemic human CoVs, and MERS-CoV [37-41], or proteins, as in the case of the SARS-CoV-1 [42] and current SARS-CoV-2 viruses [43,44], or both, as in human immunodeficiency virus (HIV) [45]. The receptor binding domain (RBD) within domain 1 of the S protein (S1) in SARS-CoV-1 and -2 has been shown to interact with the host protein angiotensin-converting enzyme 2 (ACE2). The affinity of S1 for ACE2 enables the virus to adhere to and infect the host cell [43,44]. Being exposed on the viral surface, S proteins are a major target for host antibodies and are referred to as viral antigens; these antigens are therefore targets for vaccine development.

The S glycoproteins of SARS-CoV-1 and CoV-2 share a high degree of structural similarity, with an average root-mean-squared difference (RMSD) in the Cα positions of only 3.8 Å [13]. They also share a relatively high sequence homology of 75.5% [46]. The MERS S glycoprotein also shares a similar trimeric structure with CoV-1 and Cov-2. Cryo-EM studies show that the receptor binding domains of the S1 subunits of the CoV-2 S glycoprotein can exist in at least two conformations by undergoing a hinge-like movement, which has also been reported in SARS-CoV-1 and MERS, as well as other more distantly related coronaviruses [14,47,48]. Despite the structural and sequence similarities between the SARS-CoV-1 and CoV-2 S

glycoproteins, sequence based epitope prediction indicates that approximately 85% of the theoretical sequential epitopes are unique between the two proteins [46].

To better understand the antigenic properties of the SARS-CoV-2 S protein from a structural perspective, we aligned the 3D structure of the S glycoprotein from SARS-CoV-2 [13] with those from co-crystal structures of SARS-CoV-1 and MERS that contained bound antibody fragments. From this alignment, the extent to which epitopes in the CoV-2 S glycoprotein might be inaccessible to known antibodies on the basis of structural differences in the glycoproteins or due to shielding by glycans on the CoV-2 S glycoprotein surface (Figure 1) was inferred. The previously reported antibodies bind almost exclusively to the S1 RBD. When aligned on the glycosylated CoV-2 S protein, it is evident that approximately 50% of the corresponding CoV-1 or MERS epitopes are shielded by glycans from antibody binding in the CoV-2 S glycoprotein, and only areas of the protein surface at the apex of the S1 domain appear to be accessible to known antibodies. The epitope surface area in the known antibody co-complexes ranges between approximately 800 to 1000 $Å^2$, which is comparable to the contact surface between the S glycoprotein and its ACE2 receptor [49].

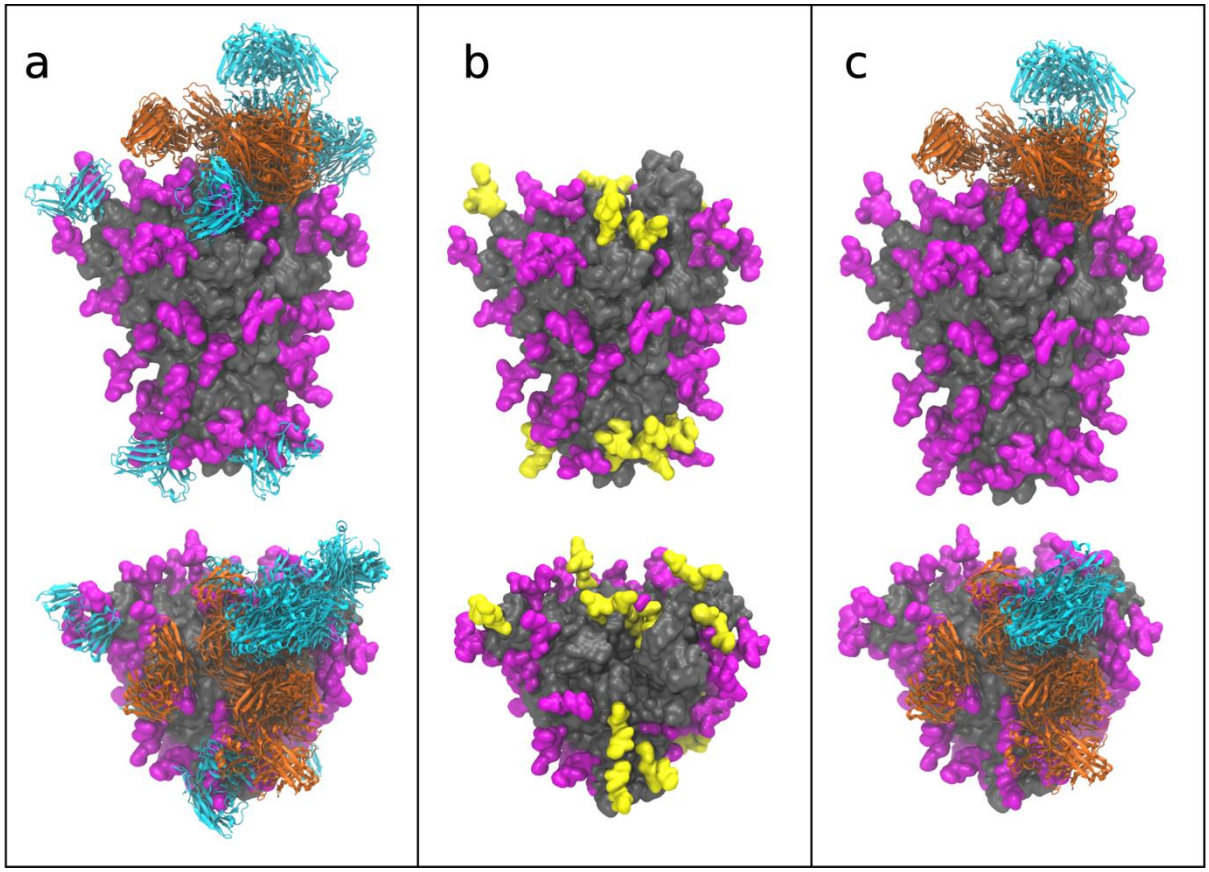


Figure 1. a. Side view (upper panels) and Top view (lower panels) of the SARS-CoV-2 S protein (grey surface) with homogeneous Complex glycosylation (magenta) showing aligned antibody fragments (ribbons) from co-complexes with the S glycoproteins from SARS-CoV-1 (orange) and MERS (cyan) [48,50-61]. **b.** Glycans present on SARS-CoV-2 S glycoprotein that are incompatible with known antibody positions due to steric overlap are shown in yellow. **c.** Potential antibody poses after elimination of epitopes blocked by S protein glycosylation. Images generated using Visual Molecular Dynamics (VMD) [62] version 1.9.3.

This preliminary epitope analysis does not take into account variations in site-specific glycan composition or the plasticity of the glycans or the protein. Furthermore, the number of antibodies that have been co-crystallized with the related S proteins represents only a minute fraction of the possible repertoire. To address these limitations, we subjected multiple glycoforms of the CoV-2 S glycoprotein to MD simulation and interpreted the results in terms of the impact of glycan structure on the theoretical S glycoprotein antigenic surface area (Figure 2).

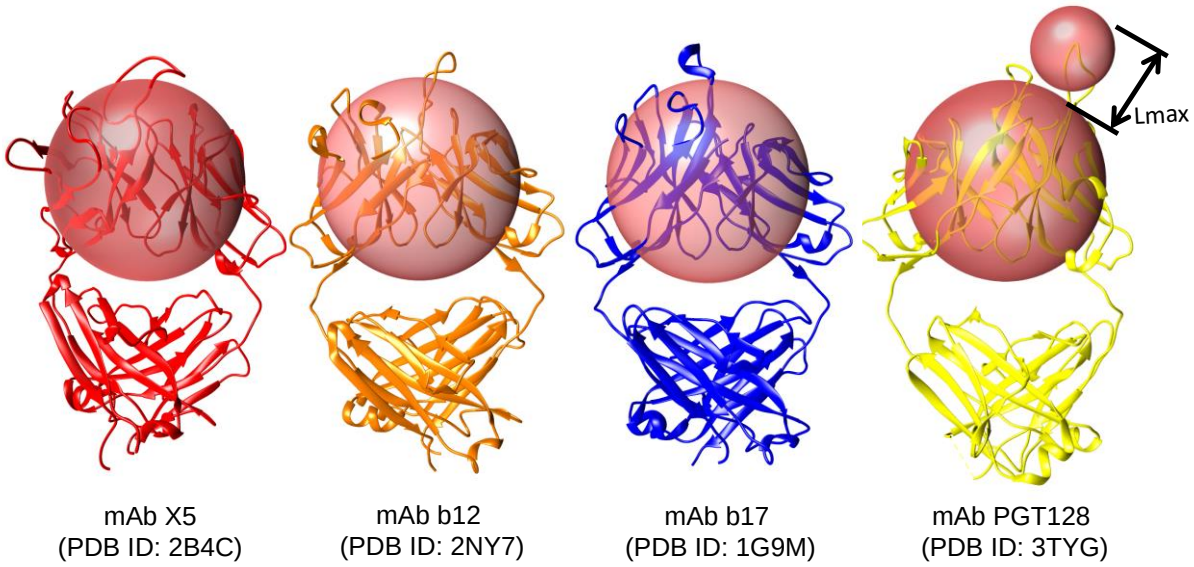
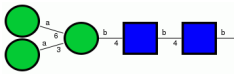
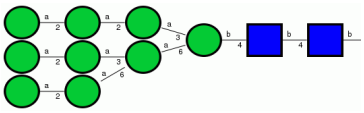
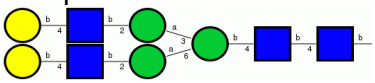
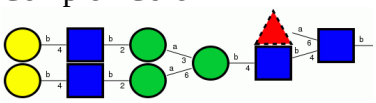


Figure 2. To estimate the antibody accessible surface area (AbASA), a spherical probe was derived (radius 7.2 Å, smaller sphere) that approximates the average size of the hypervariable loops from four anti-gp120 antibodies, in which the epitopes were either protein surface residues (PDB IDs: 2B4C [63], 2NY7 [64], 1G9M [65]) or both carbohydrate and protein residues: (3TYG [66]). This probe size may be compared to values of 5 and 10 Å employed previously to estimate antigenic surface area [19]. Changes in the solvent accessible surface area (SASA) showed no shielding by glycans and thus a simple SASA model was not useful for this analysis[19]. Additionally, to account for the presence of the beta-sheet framework in the antibody variable fragment (Fv), we introduced a second larger probe (18.6 Å) sufficient to approximately enclose that domain. The antigenic surface area is then defined as sum of the surface areas of any protein residues that make contact with the CDR probe, provided that the CDR probe is proximal to residues accessible to the Fv probe. This latter requirement is governed by L_{max}, which requires that the distance between the CDR-antigen contact site and the Fv probe surface be less than the length (10.4 Å) of the longest CDR loop in mAb PGT128. PGT128 was chosen for this reference as it contains a particularly long CDR loop that penetrates the glycan shield of gp120. Images generated with UCSF Chimera [67].

Assessment of the impact of glycosylation on antigenicity. Three independent MD simulations of five glycoforms of the S protein were performed for a combined total of 0.25 μs per glycoform and used to compute the extent to the antigenic surface area was sensitive to glycan microheterogeneity (Table 1). The data indicate that uniform glycosylation with the smallest of the glycans (paucimannose, M3), which is a sub-structure within all N-linked glycans, provided the least shielding of the S protein surface, leaving 71% of the surface exposed to an antibody probe. In contrast, the largest high mannose N-linked glycans (M9), which corresponds to the nascent glycoform that would exist prior to processing through the Endoplasmic reticulum and Golgi apparatus, led to the highest level of surface shielding.

The extent of shielding offered by the two complex types of glycans are not statistically different from that of M9 at 53-55% antigenic surface exposure. A site-specific model generated from the most commonly observed glycans at each glycosite for the S glycoprotein produced recombinantly in HEK293 cells was also included for comparison. This glycoform resulted in shielding of just over 40% of the S protein antigenic surface; a value similar to the models based on uniform M9 and Complex glycans. These results are broadly consistent with the conclusion that antigenicity of the S protein is insensitive to glycan microheterogeneity, with the exception of the glycoform composed solely of M3 glycans. Nevertheless, differences in glycosylation may impact other features, such as local interactions between the glycan and the protein surface, or local structural fluctuations in either the protein or glycan conformations that are only partially captured by the CRD probe analysis.

Table 1. SARS-CoV-2 S glycoprotein antigenic surface areas (Å^2) as a function of glycoform.

Glycoform	Average antibody accessible surface area (AbASA)[a]	Exposed fraction of AbASA
M3 	58,579 ± 2.8%	0.71
M9 	44,184 ± 1.1%	0.53
Complex 	45,571 ± 1.6%	0.55
Complex Core F 	43,943 ± 2.0%	0.53
HEK293 site-specific glycosylation	48,322 ± 0.7%	0.58
Non-glycosylated	83,041 ± 2.8%	1.00

[a]Surface areas were computed with the Naccess software [68], version 2.1.1.

A visual examination of the structures from MD simulation (Figures 3 and S1) broadly confirms the observation (Figure 1) that the most exposed epitopes comprise the ACE2 receptor site, specifically the apex region of the S1 domain when that domain is in the open conformation. Moreover, the extensive motion displayed by each glycan illustrates that no single static model can fully capture the extent of glycan shielding. It can also be observed that a ring of antigenic sites appears to encircle the S1 domain, independent of glycoform. Unlike the extremely high level of glycan shielding in gp120 that challenges HIV vaccine development [69,70], the level of shielding by glycans in the S protein is more moderate, with approximately 60% of the surface potentially accessible to antibodies.

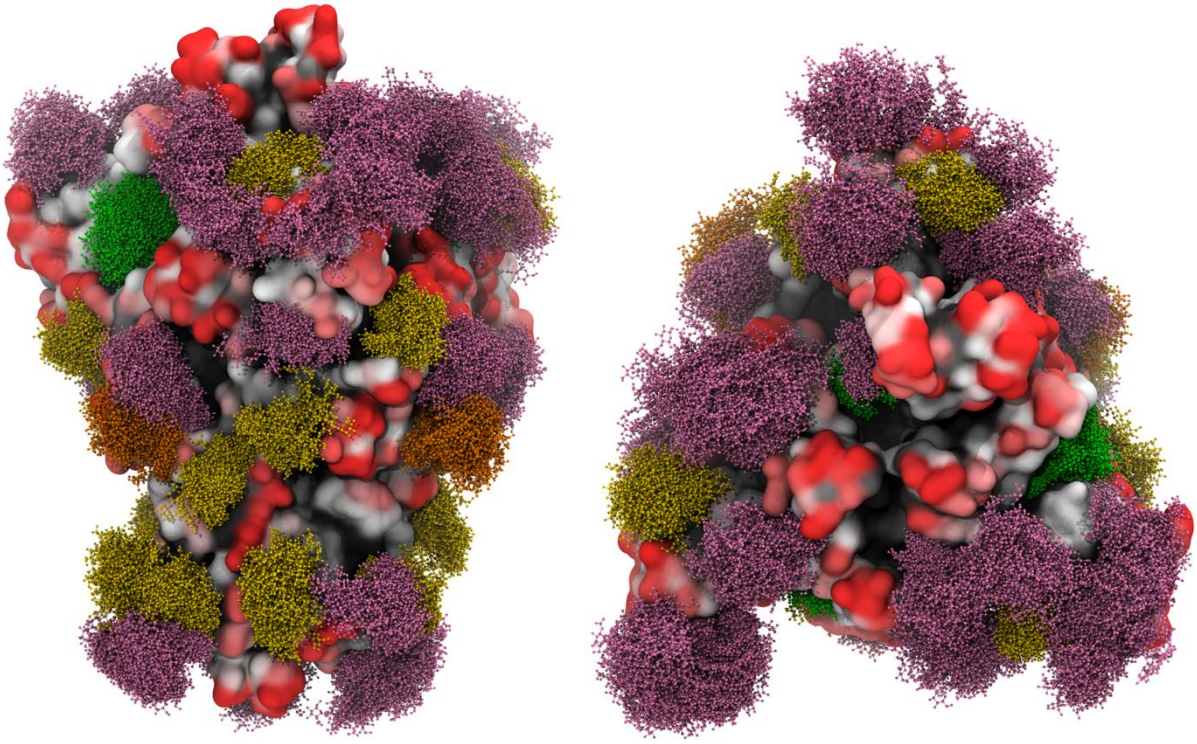


Figure 3. Overlay of snapshots from MD simulation of the S glycoprotein with site-specific glycosylation. The glycans are shown in ball-and-stick representation: M9 (green), M5 (dark yellow), hybrid (orange), complex (pink) (See Table S1 for details). The protein surface is colored according to antibody accessibility from black to red (least to most accessible). Images generated using Visual Molecular Dynamics (VMD) [62] version 1.9.3.

Adaptive immune response to SARS-CoV-2. Beyond a role in shielding the underlying protein from recognition by antibodies, the glycans on pathogenic proteins may also attenuate the ability of the host immune system to raise antibodies against any epitopes that include the glycan. In a T-cell dependent adaptive immune response, peptides from the endocytosed pathogen are presented on antigen presenting cells by major histocompatibility complex II molecules, known as human leukocyte antigen (HLA) complexes. HLA complexes have preferred peptide motifs, and based on a knowledge of these preferences it is possible to predict which peptides in a protein are likely to be HLA antigens [71,72]. However, when that peptide contains a glycosylation site, the ability of the peptide to be presented in an HLA complex may be compromised, if for example the peptide cannot bind to the HLA molecule due to the steric presence of the glycan. However, glycopeptides may be presented in HLA complexes [73] if the glycan is small enough or if it is found on the end of the peptide antigen where it doesn't interfere with HLA binding [74]. The glycan-mediated shielding of predicted HLA antigens (Table S2) derived from the S protein are shown in Figures 4, S2 and S3 for all HLA peptide sequences that also contain a glycosite.

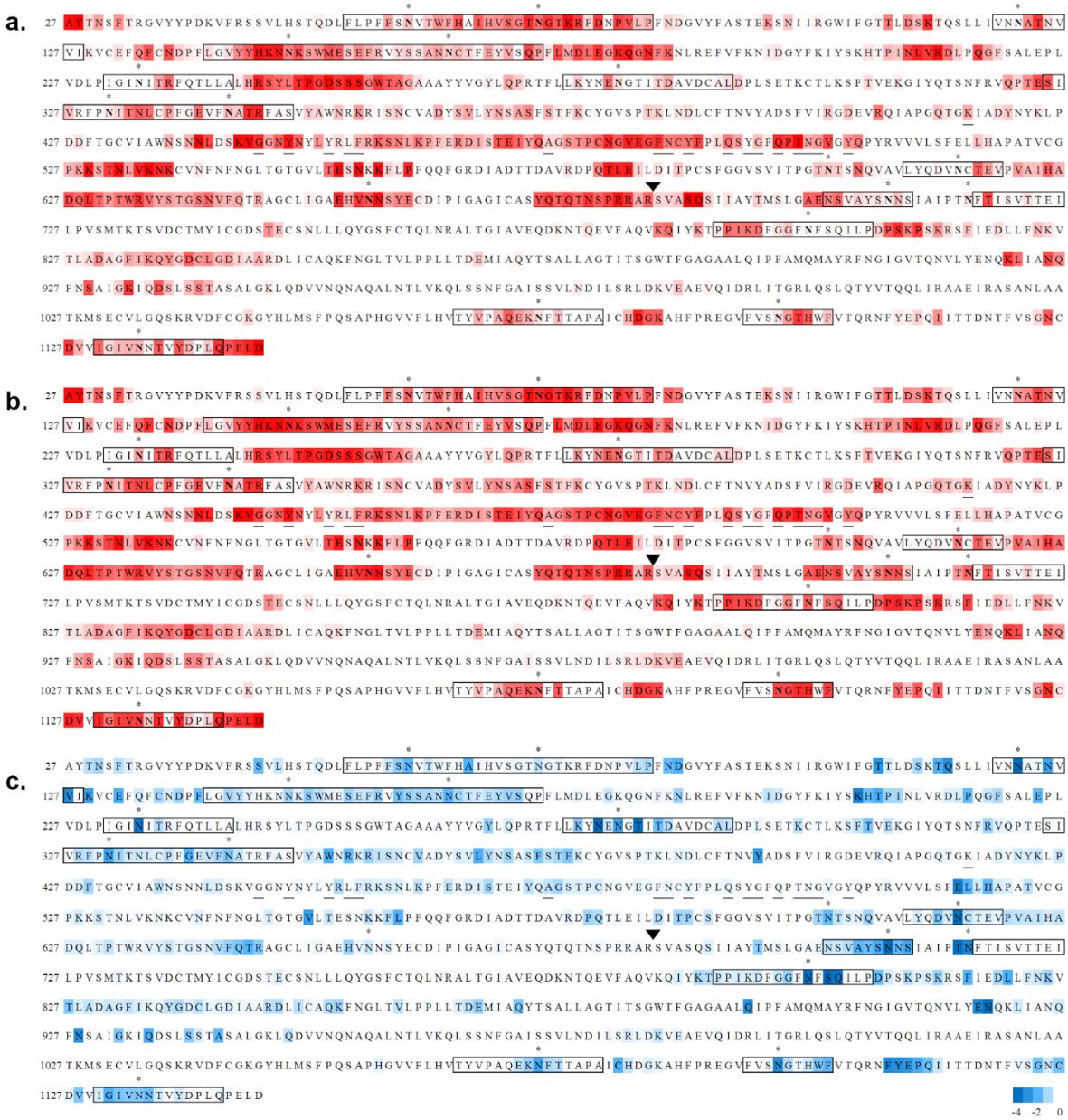


Figure 4. Sequence of the S protein (NCBI: YP_009724390.1) used to generate the 3D model of the glycoprotein. Residues 1-26 and 1147-1273 were not included in the 3D structure due to a lack of relevant template structures. Sequences within a rectangle were predicted to consist of one or more HLA antigens using the RankPep server (imed.med.ucm.es/Tools/rankpep [71,72]). Glycosites are indicated with asterisks, residues reported to interact with the ACE2 receptor [49] are underlined, and the protease cleavage site is indicated with a triangle above the RS junction. **a.** The sequence is colored according to antibody accessibility computed for the site-specific glycoform from white to red (least to most accessible). **b**. Antibody accessibility computed for the non-glycosylated protein. **c**. The difference in accessibilities between the site-specific and non-glycosylated glycoforms is plotted as the fold change in antigen accessibility during the simulation from -4 to 0 (blue to white), where blue indicates glycosylation-dependent surface shielding.

As expected, glycosylation consistently decreased the surface exposure of the residues proximal to the glycosites (Figure 4.c), but also led to non-sequential changes in exposure, as a result of the 3D topology in the vicinity of each glycosite. Of the 18 glycosites in the 3D structure, 16 are predicted to be present in HLA peptides. Although the glycans may occur throughout the HLA sequences (Table S2), in 12 of these sequences the glycans are predicted to be present at the terminus of at least one putative HLA antigen. This observation suggests that these 12 glycosites may not interfere with antigen presentation in an HLA complex. This property is essential for the potential generation of antibodies against the underlying epitopes, but moreover, may lead to antibodies that target these carbohydrates on the S glycoprotein [73]. Anti-carbohydrate antibodies have been shown to be neutralizing in other viruses, such as HIV [75], and therefore glycosylated peptides can offer an alternative to more traditional peptide epitopes. From the perspective of vaccine development [76], targeting glycans as epitopes would be expected to benefit from matching the glycan microheterogeneity in the vaccine to that in the circulating virus, which requires additional consideration of the choice of cell type for vaccine production.

Discussion

The present study indicates that glycans shield approximately 40% of the underlying protein surface of the S glycoprotein from antibody recognition, and that this value is relatively insensitive to glycan type. This suggests that although glycan microheterogeneity varies according to host cell type, the efficacy of antisera should not be impacted by such differences. In contrast, by analogy with influenza hemagglutinin [33,77], variations in glycosite location arising from antigenic drift can be expected to have a profound effect on S protein antigenicity and potentially vaccine efficacy. Fortunately, the most accessible and largest antigenic surface in the S protein consists of the ACE2 binding domain, where the virus can't exploit glycan shielding or mutational changes to evade host immune response without potentially attenuating viral fitness. The requirement that the virus maintain the integrity of the ACE2 RBD suggests that a vaccine that includes this epitope may be effective, as long as the virus continues to target the same host receptor.

Glycan microheterogeneity may impact the innate immune response by altering the ability of collectins and other lectins of the immune system to neutralize the virus, and may impact the adaptive immune response by altering the number of viable HLA antigens. But such heterogeneity has little impact on the antigenic surface area of the S protein. From a vaccine perspective, efficacy may benefit from ensuring that the production method results in glycosylation profiles that match those of the circulating virus, or by engineering a vaccine so as to avoid glycosylated sequences.

Lastly, the observation that homogeneously glycosylated glycoforms are predicted to display approximately the same antigenic properties as those computed for the more relevant site-specific glycoform suggests that such models can be usefully applied in advance of the report of experimental glycomics data. This final conclusion is significant as it enables the effects of glycosite alterations to be estimated in anticipation of antigenic shift or drift.

Methods

SARS-CoV2 spike (S) protein structure – A 3D structure of the prefusion form of the S protein (RefSeq: YP_009724390.1, UniProt: P0DTC2 SPIKE_SARS2), based on a Cryo-EM structure (PDB code 6VSB) [13], was obtained from the SWISS-MODEL server

(swissmodel.expasy.org). The model has 95% coverage (residues 27 to 1146) of the S protein.

S protein glycoform generation – Five unique 3D models for the glycosylated glycoprotein were generated using the glycoprotein builder available at GLYCAM-Web (www.glycam.org) together with an in-house program that adjusts the asparagine side chain torsion angles and glycosidic linkages within known low-energy ranges [78] to relieve any atomic overlaps with the core protein, as described previously [33,79]. The site specific glycans used to model a glycoform representative of the data obtained from the S glycoprotein expressed in HEK293 cells [15], are presented in Table S1.

Energy minimization and Molecular dynamics (MD) simulations – Each glycosylated structure was placed in a periodic box of approximately 130,000 TIP3P water molecules [80] with a 10 Å buffer between the glycoprotein and the box edge. Energy minimization of all atoms was performed for 20,000 steps (10,000 steepest decent, followed by 10,000 conjugant gradient) under constant pressure (1 atm) and temperature (300 K) (nPT) conditions. All MD simulations were performed under nPT conditions with the CUDA implementation of the PMEMD [81,82] simulation code, as present in the Amber14 software suite [83]. The GLYCAM06j force field [84] and Amber14SB force field [85] were employed for the carbohydrate and protein moieties, respectively. A Berendsen barostat with a time constant of 1 ps was employed for pressure regulation, while a Langevin thermostat with a collision frequency of 2 ps^{-1} was employed for temperature regulation. A nonbonded interaction cut-off of 8 Å was employed. Long-range electrostatics were treated with the particle-mesh Ewald (PME) method [86]. Covalent bonds involving hydrogen were constrained with the SHAKE algorithm, allowing an integration time step of 2 fs [87] to be employed. The energy minimized coordinates were equilibrated at 300K over 400 ps with restraints on the solute heavy atoms. Each system was then equilibrated with restraints on the Cα atoms of the protein for 1ns, prior to initiating 3 independent production MD simulations with random starting seeds for a total time of 0.25 μs, with no restraints applied.

Supporting Information

Coordinates in pdb format for each glycoform will be made available for download from GLYCAM-Web (www.glycam.org).

Acknowledgments

R.J.W. thanks the National Institutes of Health (U01 CA207824 and P41 GM103390) for financial support.

Author Contributions

O.C.G. and R.J.W. designed the research; O.C.G., D. M., K. I., and R.J.W. performed the research; R.J.W. wrote the paper.

Competing Interests

The authors declare no competing interests.

References

1 W.H.O. Coronavirus disease 2019 (Covid-19) Situation Report. Report No. 77, (2020).
2 Depetris, R. S. *et al.* Partial enzymatic deglycosylation preserves the structure of cleaved recombinant HIV-1 envelope glycoprotein trimers. *J Biol Chem* **287**, 24239-24254 (2012).
3 Pereira, M. S. *et al.* Glycans as Key Checkpoints of T Cell Activity and Function. *Front Immunol* **9**, 2754 (2018).
4 Baum, L. G. & Cobb, B. A. The direct and indirect effects of glycans on immune function. *Glycobiology* **27**, 619-624 (2017).
5 Vigerust, D. J. & Shepherd, V. L. Virus glycosylation: role in virulence and immune interactions. *Trends in microbiology* **15**, 211-218 (2007).
6 Casals, C., Campanero-Rhodes, M. A., Garcia-Fojeda, B. & Solis, D. The Role of Collectins and Galectins in Lung Innate Immune Defense. *Front Immunol* **9**, 1998 (2018).
7 Hutter, J. *et al.* Toward animal cell culture-based influenza vaccine design: viral hemagglutinin N-glycosylation markedly impacts immunogenicity. *J Immunol* **190**, 220-230 (2013).
8 Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404-410 (2006).
9 Cotter, C. R., Jin, H. & Chen, Z. A single amino acid in the stalk region of the H1N1pdm influenza virus HA protein affects viral fusion, stability and infectivity. *PLoS Pathog* **10**, e1003831 (2014).
10 Li, Y. *et al.* Single hemagglutinin mutations that alter both antigenicity and receptor binding avidity influence influenza virus antigenic clustering. *J Virol* **87**, 9904-9910 (2013).
11 Altman, M. O. *et al.* Human Influenza A Virus Hemagglutinin Glycan Evolution Follows a Temporal Pattern to a Glycan Limit. *mBio* **10** (2019).
12 Zost, S. J. *et al.* Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. *Proc Natl Acad Sci U S A* **114**, 12578-12583 (2017).
13 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* (2020).
14 Yuan, Y. *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun* **8**, 15092 (2017).
15 Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific analysis of the SARS-CoV-2 glycan shield. Preprint at https://www.biorxiv.org/content/10.1101/2020.03.26.010322v1 (2020).
16 Khatri, K. *et al.* Integrated Omics and Computational Glycobiology Reveal Structural Basis for Influenza A Virus Glycan Microheterogeneity and Host Interactions. *Mol. Cell. Proteomics* **15**, 1895-1912 (2016).

17 An, Y. *et al.* N-Glycosylation of Seasonal Influenza Vaccine Hemagglutinins: Implication for Potency Testing and Immune Processing. *J Virol* **93** (2019).
18 An, Y. *et al.* Comparative Glycomics Analysis of Influenza Hemagglutinin (H5N1) Produced in Vaccine Relevant Cell Platforms. *J. Proteome Res.* **12**, 3707-3720 (2013).
19 Urbanowicz, R. A. *et al.* Antigenicity and Immunogenicity of Differentially Glycosylated Hepatitis C Virus E2 Envelope Proteins Expressed in Mammalian and Insect Cells. *J Virol* **93** (2019).
20 Hang, I. *et al.* Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology* **25** (2015).
21 Losfeld, M. E. *et al.* Influence of protein/glycan interaction on site-specific glycan heterogeneity. *FASEB J* **31**, 4623-4635 (2017).
22 Arigoni-Affolter, I. *et al.* Mechanistic reconstruction of glycoprotein secretion through monitoring of intracellular N-glycan processing. *Sci Adv* **5**, eaax8930 (2019).
23 York, I. A., Stevens, J. & Alymova, I. V. Influenza virus N-linked glycosylation and innate immunity. *Biosci Rep* **39** (2019).
24 Skehel, J. J. *et al.* A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Proc Natl Acad Sci U S A* **81**, 1779-1783 (1984).
25 An, Y., McCullers, J. A., Alymova, I., Parsons, L. M. & Cipollo, J. F. Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites. *J Proteome Res* **14**, 3957-3969 (2015).
26 Homans, S. W., Dwek, R. A. & Rademacher, T. W. Solution Conformations of N-Linked Oligosaccharides. *Biochemistry* **26**, 6571-6578 (1987).
27 Homans, S. W. *et al.* Conformational Transitions in N-Linked Oligosaccharides. *Biochemistry* **25**, 6342-6350 (1986).
28 Jo, S., Qi, Y. & Im, W. Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins. *Glycobiology* **26**, 19-29 (2016).
29 Harbison, A. & Fadda, E. An atomistic perspective on ADCC quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics. *Glycobiology* (2019).
30 Amaro, R. E. & Li, W. W. Molecular-level simulation of pandemic influenza glycoproteins. *Methods Mol Biol* **819**, 575-594 (2012).
31 Bernardi, A., Kirschner, K. N. & Faller, R. Structural analysis of human glycoprotein butyrylcholinesterase using atomistic molecular dynamics: The importance of glycosylation site ASN241. *PLoS One* **12**, e0187994 (2017).

32 Yanaka, S. *et al.* Dynamic Views of the Fc Region of Immunoglobulin G Provided by Experimental and Computational Observations. *Antibodies (Basel)* **8** (2019).
33 Peng, W. *et al.* Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. *Cell Host Microbe* **21**, 23-34 (2017).
34 Vigerust, D. J. *et al.* N-linked glycosylation attenuates H3N2 influenza viruses. *J Virol* **81**, 8593-8600 (2007).
35 Lin, Y. P. *et al.* Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc Natl Acad Sci U S A* **109**, 21474-21479 (2012).
36 Gulati, S. *et al.* Human H3N2 Influenza Viruses Isolated from 1968 To 2012 Show Varying Preference for Receptor Substructures with No Apparent Consequences for Disease or Spread. *PLoS One* **8**, e66325 (2013).
37 Hulswit, R. J. G. *et al.* 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 U S A* **116**, 2681-2690 (2019).
38 Tortorici, M. A. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nature Structural & Molecular Biology* **26**, 481-489 (2019).
39 Vlasak, R., Luytjes, W., Spaan, W. & Palese, P. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci U S A* **85**, 4526-4529 (1988).
40 Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci U S A* **114**, E8508-E8517 (2017).
41 Park, Y. J. *et al.* Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nature Structural & Molecular Biology* **26**, 1151-1157 (2019).
42 Li, W. *et al.* Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**, 450-454 (2003).
43 Yan, R. *et al.* Structural basis for the recognition of the SARS-CoV-2 by full-length human ACE2. *Science* (2020).
44 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020).
45 Connell, B. J. & Lortat-Jacob, H. Human immunodeficiency virus and heparan sulfate: from attachment to entry inhibition. *Front Immunol* **4** (2013).
46 Zheng, M. & Song, L. Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. *Cell Mol Immunol* (2020).

47 Gui, M. *et al.* Electron microscopy studies of the coronavirus ribonucleoprotein complex. *Protein Cell* **8**, 219-224 (2017).
48 Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A* **114**, E7348-E7357 (2017).
49 Lan, J. *et al.* Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. Preprint at https://www.biorxiv.org/content/10.1101/2020.02.19.956235v1 (2020).
50 Prabakaran, P. *et al.* Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem* **281**, 15829-15836 (2006).
51 Hwang, W. C. *et al.* Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *J Biol Chem* **281**, 34610-34616 (2006).
52 Pak, J. E. *et al.* Structural insights into immune recognition of the severe acute respiratory syndrome coronavirus S protein receptor binding domain. *J Mol Biol* **388**, 815-823 (2009).
53 Ying, T. *et al.* Junctional and allele-specific residues are critical for MERS-CoV neutralization by an exceptionally potent germline-like antibody. *Nat Commun* **6**, 8223 (2015).
54 Wang, L. *et al.* Evaluation of candidate vaccine approaches for MERS-CoV. *Nat Commun* **6**, 7712 (2015).
55 Li, Y. *et al.* A humanized neutralizing antibody against MERS-CoV targeting the receptor-binding domain of the spike protein. *Cell Res* **25**, 1237-1249 (2015).
56 Chen, Z. *et al.* Human Neutralizing Monoclonal Antibody Inhibition of Middle East Respiratory Syndrome Coronavirus Replication in the Common Marmoset. *J Infect Dis* **215**, 1807-1815 (2017).
57 Zhang, S. *et al.* Structural Definition of a Unique Neutralization Epitope on the Receptor-Binding Domain of MERS-CoV Spike Glycoprotein. *Cell Rep* **24**, 441-452 (2018).
58 Wang, L. *et al.* Importance of Neutralizing Monoclonal Antibodies Targeting Multiple Antigenic Sites on the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein To Avoid Neutralization Escape. *J Virol* **92** (2018).
59 Zhou, H. *et al.* Structural definition of a neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein. *Nat Commun* **10**, 3068 (2019).
60 Walls, A. C. *et al.* Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion. *Cell* **176**, 1026-1039 e1015 (2019).
61 Wang, N. *et al.* Structural Definition of a Neutralization-Sensitive Epitope on the MERS-CoV S1-NTD. *Cell Rep* **28**, 3395-3405 e3396 (2019).

62 Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33-38 (1996).
63 Huang, C. C. *et al.* Structure of a V3-containing HIV-1 gp120 core. *Science* **310**, 1025-1028 (2005).
64 Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732-737 (2007).
65 Kwong, P. D. *et al.* Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **8**, 1329-1339 (2000).
66 Pejchal, R. *et al.* A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the HIV Glycan Shield. *Science* **334**, 1097-1103 (2011).
67 Pettersen, E. F. *et al.* UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comp. Chem.* **25**, 1605-1612 (2004).
68 NACCESS v. 2.1.1 (University College London, London, 1993).
69 Horiya, S., MacPherson, I. S. & Krauss, I. J. Recent strategies targeting HIV glycans in vaccine design. *Nat. Chem. Biol.* **10**, 990-999 (2014).
70 Doores, K. J. The HIV glycan shield as a target for broadly neutralizing antibodies. *FEBS J* **282**, 4679-4691 (2015).
71 Reche, P. A., Glutting, J. P., Zhang, H. & Reinherz, E. L. Enhancement to the RANKPEP resource for the prediction of peptide binding to MHC molecules using profiles. *Immunogenetics* **56**, 405-419 (2004).
72 Reche, P. A., Glutting, J. P. & Reinherz, E. L. Prediction of MHC class I binding peptides using profile motifs. *Hum Immunol* **63**, 701-709 (2002).
73 Avci, F. Y., Li, X., Tsuji, M. & Kasper, D. L. A mechanism for glycoconjugate vaccine activation of the adaptive immune system and its implications for vaccine design. *Nature: Medicine* **17**, 1602-1609 (2011).
74 Malaker, S. A. *et al.* Identification and Characterization of Complex Glycosylated Peptides Presented by the MHC Class II Processing Pathway in Melanoma. *J Proteome Res* **16**, 228-237 (2017).
75 Haji-Ghassemi, O., Blackler, R. J., Martin Young, N. & Evans, S. V. Antibody recognition of carbohydrate epitopes. *Glycobiology* **25**, 920-952 (2015).
76 Chang, D. & Zaia, J. Why Glycosylation Matters in Building a Better Flu Vaccine. *Mol Cell Proteomics* **18**, 2348-2358 (2019).
77 Sun, X. *et al.* N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses. *J Virol* **87**, 8756-8766 (2013).
78 Nivedha, A. K., Makeneni, S., Foley, B. L., Tessier, M. B. & Woods, R. J. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *J Comput Chem* **35**, 526-539 (2014).

79 Grant, O. C. *et al.* Gly-Spec: a webtool for predicting glycan specificity by integrating glycan array screening data and 3D structure. *Glycobiology* **26**, 1027-1028 (2016).
80 Jorgensen, W. L. Quantum and Statistical Mechanical Studies of Liquids. 10. Transferable Intermolecular Potential Functions for Water, Alcohols, and Ethers. Application to Liquid Water. *J. Am. Chem. Soc.* **103**, 335-340 (1981).
81 Salomon-Ferrer, R., Götz, A. W., Poole, D., Le Grand, S. & Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **9**, 3878-3888 (2013).
82 Gotz, A. W. *et al.* Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542-1555 (2012).
83 AMBER 14 (University of California, San Francisco, 2014).
84 Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622-655 (2008).
85 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11** (2015).
86 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *The Journal of chemical physics* **98**, 10089 (1993).
87 Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **23**, 327-341 (1977).