



OPEN

Analysis of the SARS-CoV-2 spike protein glycan shield reveals implications for immune recognition

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

Here we have generated 3D structures of glycoforms of the spike (S) glycoprotein from SARS-CoV-2, based on reported 3D structures and glycomics data for the protein produced in HEK293 cells. We also analyze structures for glycoforms representing those present in the nascent glycoproteins (prior to enzymatic modifications in the Golgi), as well as those that are commonly observed on antigens present in other viruses. These models were subjected to molecular dynamics (MD) simulation to determine the extent to which glycan microheterogeneity impacts the antigenicity of the S glycoprotein. Lastly, we have identified peptides in the S glycoprotein that are likely to be presented in human leukocyte antigen (HLA) complexes, and discuss the role of S protein glycosylation in potentially modulating the innate and adaptive immune response to the SARS-CoV-2 virus or to a related vaccine. The 3D structures show that the protein surface is extensively shielded from antibody recognition by glycans, with the notable exception of the ACE2 receptor binding domain, and also that the degree of shielding is largely insensitive to the specific glycoform. Despite the relatively modest contribution of the glycans to the total molecular weight of the S trimer (17% for the HEK293 glycoform) they shield approximately 40% of the protein surface.

The current COVID-19 pandemic has led to over 11 million confirmed infections globally with a fatality rate of approximately 4.5%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 July 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 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 gp120². The glycosylation of these surface antigens helps the pathogen evade recognition by the host immune system by cloaking the protein surface from detection by the humoral and cellular components of the innate immune system^{3–5}, and by altering the ability of the host to raise an effective adaptive immune response^{6,7} or even by enhancing infectivity⁸. Additionally, because the virus hijacks the host cellular machinery for replication and subsequent glycosylation, the viral glycan shield may be composed of familiar host glycans; thereby suppressing an anti-carbohydrate immune response⁹.

Fortunately, the innate immune system has evolved a range of strategies for responding to glycosylated pathogens¹⁰, but antigen glycosylation nevertheless complicates the development of vaccines¹¹. Over time, the protein sequences in viral antigens undergo mutations (antigenic drift), which can alter the species specificity of the virus¹², modulate its infectivity¹³, and alter the antigenicity of the surface proteins¹⁴. 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 surface antigens^{15,16}. Varying antigen glycosylation is thus a mechanism by which new virus strains can evade the host immune response¹⁵, and attenuate the efficacy of existing vaccines¹¹.

Recently, a cryo-EM structure of the SARS-CoV-2 S glycoprotein has been reported¹⁷, which led to conclusion that, like the related protein from the 2002–2003 SARS pandemic (SARS-CoV-1)¹⁸, the CoV-2 S protein is also extensively glycosylated¹⁷. Furthermore, an analysis of the glycan structures present at each glycosite in

Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Rd, Athens, GA 30602, USA. ✉email: rwoods@ccrc.uga.edu

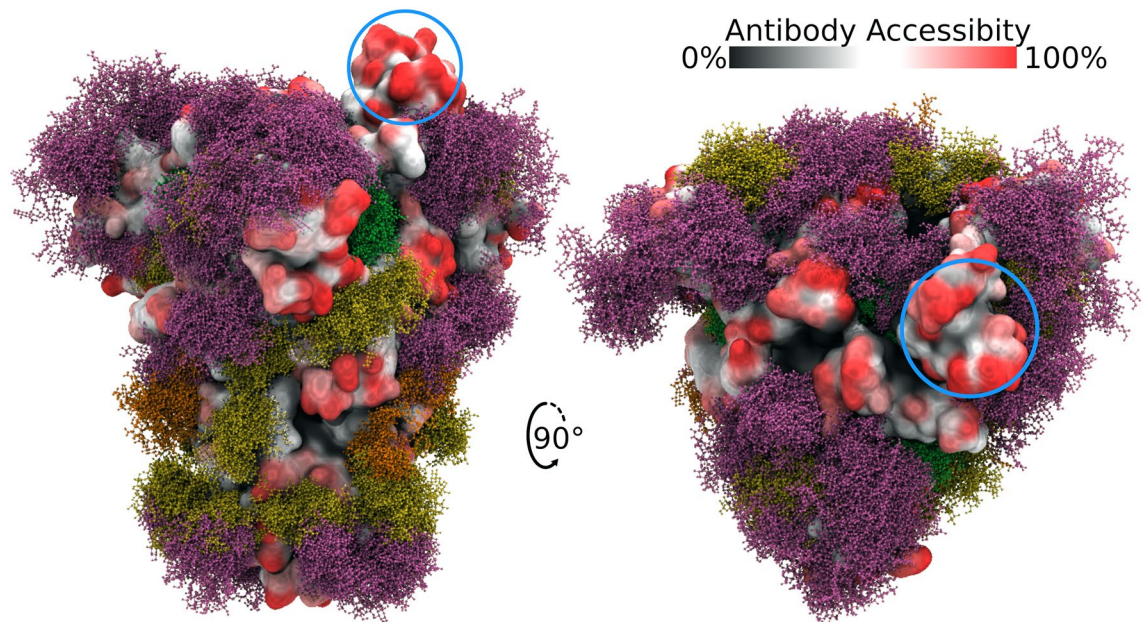


Figure 1. Side and top views of the S glycoprotein trimer with site-specific glycosylation shown as overlaid snapshots (moss surface) from MD simulations. The glycans are shown in ball-and-stick representation: M9 (green), M5 (dark yellow), hybrid (orange), complex (pink) (See Supplementary Table S1 for details). The protein surface is colored according to antibody accessibility from black to red (least to most accessible). The residues comprising the RBD in the “up” or “open” protomer are circled in blue. Images generated using Visual Molecular Dynamics (VMD)³⁵ version 1.9.3.

the S trimer produced recombinantly in human embryonic kidney 293 (HEK293) cells has also been recently reported¹⁹.

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. However, in contrast to proteins, glycans display large internal motions that prevents their accurate description by any single 3D shape^{20,21}. Fortunately, MD simulations allow accurate prediction of the 3D shapes and motions of glycans, as confirmed by comparison to solution NMR data^{22–24}, and such simulations have been widely applied to glycoproteins^{25–29}.

Here we have generated 3D structures of several glycoforms of the SARS-CoV-2 S trimer glycoprotein, in which the glycans represent those present in the S protein produced in HEK293 cells¹⁹, as well as those corresponding to the nascent glycoprotein (prior to processing in the Golgi apparatus), as well as those that are commonly observed on antigens present in other viruses^{28,30,31}. We have subjected these models to long replicate explicitly solvated molecular dynamics (MD) simulations and compared the extent to which glycan microheterogeneity impacts protein epitope exposure. The simulations were performed with the GLYCAM06/AMBER force field, which was developed for modeling carbohydrates, carbohydrate-protein complexes and glycoproteins^{32–34}, and we use the data to assess the impact of glycosylation on the immunogenic and antigenic properties of the S glycoprotein. 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

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 for 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, led to the observation of profound differences in the compositions of the glycans at a given glycosite; with structures varying from paucimannose (Sf9 cells) to core-fucosylated hybrid with bisecting N-acetylglucosamine (Egg) to sialylated biantennary glycans (HEK293)³¹. For these reasons, we have generated five model glycoforms for the S glycoprotein: with reported site-specific glycosylation (HEK293)¹⁹, and 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 between the glycoforms permits an assessment of the impact of differential levels of glycan processing on S protein antigenicity.

Assessment of the impact of glycosylation on antigenicity. We subjected the five glycoforms of the CoV-2 S glycoprotein to MD simulation and interpret the results in terms of the impact of glycan structure on the theoretical antigenic surface area of the S glycoprotein trimer (Fig. 1, Supplementary Figure S1, Table 1).

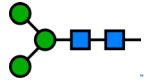
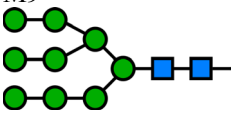
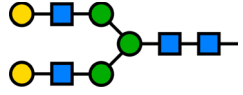
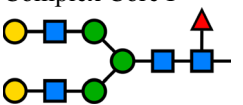
Glycoform	Average antibody accessible surface area (AbASA) ^a	Percentage of trimer surface shielded from antibody recognition by glycans ^b	Glycan percentage of total molecular weight
M3 	58,492 ± 2.2%	31 ± 0.7	11.7
M9 	47,578 ± 3.7%	44 ± 1.6	21.6
Complex 	47,824 ± 4.8%	43 ± 2.1	19.4
Complex Core F 	46,407 ± 2.5%	45 ± 1.1	20.7
HEK293 site-specific glycosylation	50,262 ± 2.6%	40 ± 1.0	17.0
Nude (non-glycosylated)	84,255 ± 2.2%	0	0

Table 1. SARS-CoV-2 S glycoprotein antigenic surface areas (\AA^2) as a function of glycoform. ^a \AA^2 , computed with the naccess software³⁸, version 2.1.1. ^b $(1 - \text{AbASA}/84,255) \times 100\%$.

A series of 3D structure snapshots of the simulation were taken at 1 ns intervals and analysed in terms of their ability to interact with a spherical probe based on the average size of hypervariable loops present in an antibody complementarity determining region (CDR) (see “Methods”). The surface accessibility of each amino acid in the trimer to the antibody probe was computed over the course of the simulations as the average antibody accessible surface area (AbASA). The per-residue AbASA values were plotted onto both the trimer 3D structure (Fig. 1) and sequence (Fig. 4), and the aggregate AbASA values reported for each glycoform in Table 1.

The data indicate that uniform glycosylation with the smallest of the glycans (paucimannose, M3), which is a sub-structure common to all *N*-linked glycans, provided the least shielding of the trimer (31% coverage), leaving 69% of the surface exposed to an antibody probe relative to the same protein with no glycosylation. In contrast, the largest high mannose *N*-linked glycans (M9), which corresponds to the nascent glycoform that would exist prior to processing through the Golgi apparatus, led to a higher level of surface shielding (44%). The degree of cloaking offered by the two types of complex glycans are not significantly different from that of M9 at 43–45%. Glycosylation based on the HEK293 glycoform model resulted in 40% of the surface being shielded from antibody recognition. Unlike the extremely high level of glycan shielding in gp120 that hampers HIV vaccine development^{36,37}, the level of shielding by glycans in the S protein is more moderate, with approximately 60% of the surface potentially accessible to antibodies.

The results from the AbASA analysis suggest that the overall antigenicity of the S protein is largely insensitive to glycan microheterogeneity, with the exception of the glycoform composed solely of M3 glycans. Nevertheless, differences in glycan microheterogeneity would be expected to modulate the binding to lectins associated with innate immune response and may impact the exposure of specific protein epitopes. Additionally, variations in glycan structure may affect local structural fluctuations in either the protein or glycan conformations, a feature that is poorly captured by the surface accessibility analysis.

A visual examination of the glycoform 3D structures (Fig. 1 and Supplementary Figure S1) indicates that the most exposed protein epitopes comprise the ACE2 receptor site, specifically the apex region of the S1 domain when that domain is in the “up” or “open” conformation. It can also be observed that a ring of antigenic sites appears to encircle the S1 domain, independent of glycoform. To corroborate these theoretical predictions we then analysed the experimentally reported epitopes for related S trimer–monoclonal antibody (mAb) interactions.

Comparison with epitopes in related coronavirus S glycoproteins. To illustrate the location of known epitopes and to assess the impact of glycosylation on epitope exposure, we aligned the 3D structures of the spike proteins from SARS-CoV-2 (CoV-2) with those from co-crystal structures of SARS-CoV-1 (SARS) and Middle East Respiratory Syndrome CoV (MERS) that contained bound mAb fragments. The S trimers of SARS and CoV-2 share a high degree of structural similarity, with an average root-mean-squared difference (RMSD) in the C α positions of only 3.8 \AA ¹⁷. The MERS S glycoprotein also shares a similar trimeric structure with SARS and CoV-2. From this alignment, the preference of neutralizing antibodies to bind to the RBD in these coronaviruses is apparent. The extent to which epitopes in the CoV-2 S trimer might be inaccessible to known antibodies on

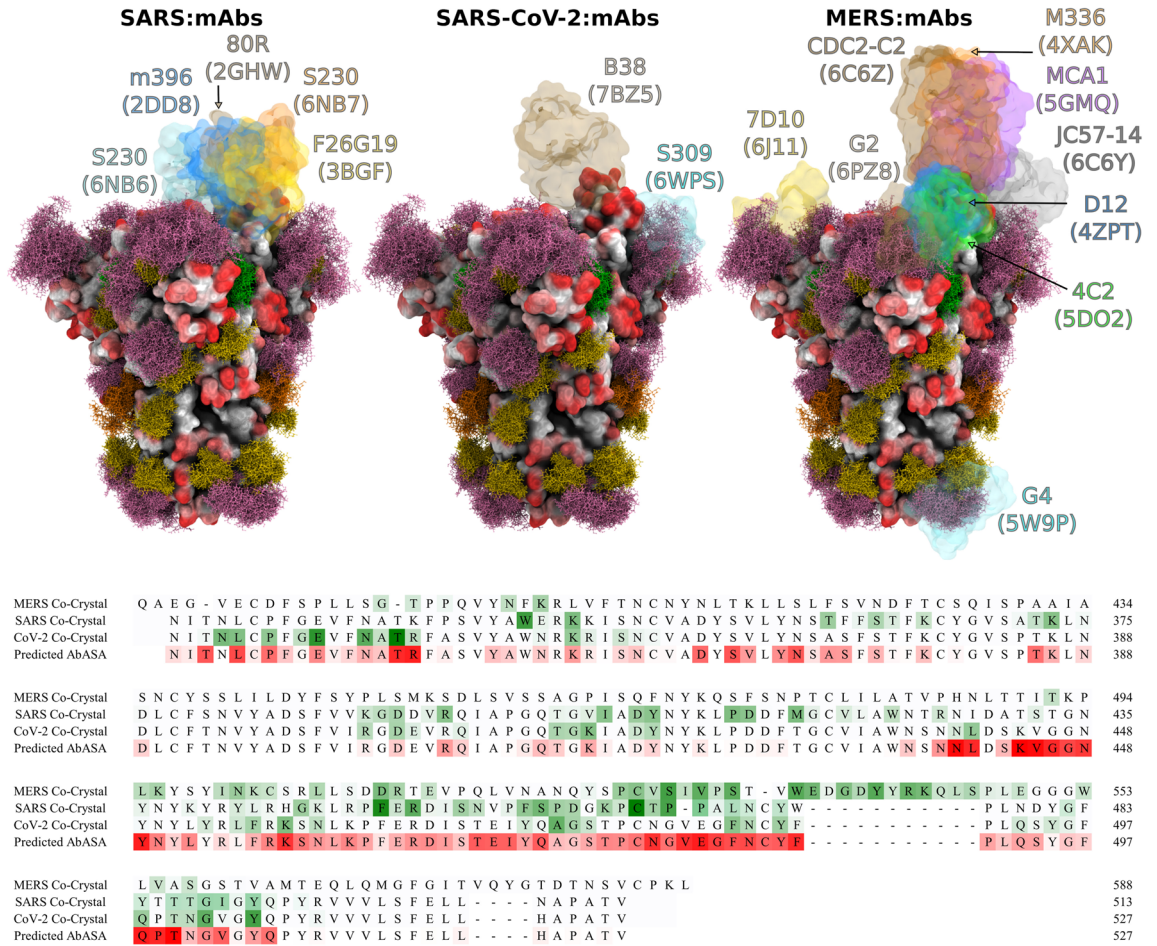


Figure 2. Superimpositions of neutralizing antibodies from co-complexes for the SARS, MERS and SARS-CoV-2 S proteins onto the HEK293 S trimer model for SARS-CoV-2. Upper panels, the antibody fragments are shown as pastel transparent surfaces with the mAb name and PDB ID for each co-complex shown in the same color^{40–52}. Lower panel, the alignments of the RBD sequences of MERS, SARS, and CoV-2 spike proteins⁵³ with the experimentally derived antibody contact areas shaded from white to green (least to most contact) compared to the predicted AbASA values for the HEK293 glycoform, shaded from white to red (least to most exposed). Glycosites in the SARS-CoV-2 sequence are indicated with an asterisk above the aligned sequences. Images generated using VMD³⁵ version 1.9.3; antibody contact areas computed with the naccess software³⁸ using a water-sized probe.

the basis of structural differences in the RBD orientations or due to shielding by glycans on the CoV-2 S trimer surface may also be inferred (Fig. 2).

Of the S glycoprotein residues observed to interact with neutralizing CoV-2 antibodies, 94% were predicted to be accessible to antibody binding from the AbASA analysis of the HEK293 glycoform (Fig. 2 and Supplementary Figure S3). With the exception of mAb G4 (PDB ID 5W9P), these epitopes are all in or proximal to the RBD, providing a structural rationale for the neutralizing ability of these mAbs. However, the AbASA analysis also predicted a significant number of epitopes distal to the RBD for which crystallographic antibody co-complexes have not yet been reported. This may reflect the fact that neutralizing antibodies are most often the focus of such structural studies. It is worth noting that a recent screening of serum from confirmed COVID-19 patients³⁹ confirmed the presence of additional immunodominant linear epitopes that are not in the RBD, as predicted by the AbASA analysis to be antigenic sequences (Supplementary Figure S3).

The antibody contact data in Fig. 2 show a remarkable degree of epitope conservation among each of the coronaviruses, and a strong correlation between the predicted S glycoprotein antibody accessibility and the observed antibody epitopes. However, a closer examination also indicates a contraction between the 3D glycoform model and the observed binding of the neutralizing antibody S309 (PDB ID 6WPS). The moss plot (Fig. 2) of the glycan at N343 (sequence VFNATR) indicates that there would be a high degree of overlap expected between the glycan and mAb S309. Despite this visual inference, the quantitative AbASA accessibility analysis is in good agreement with the experimental epitope contact areas. While the moss plot representation provides a clear indication of the extent of glycan motion and illustrates that no single static model could fully capture the degree of glycan shielding, it can also overemphasize the extent to which a glycan might cloak the protein surface from mAb recognition. While many glycan poses observed during the MD simulation were incompatible with mAb S309

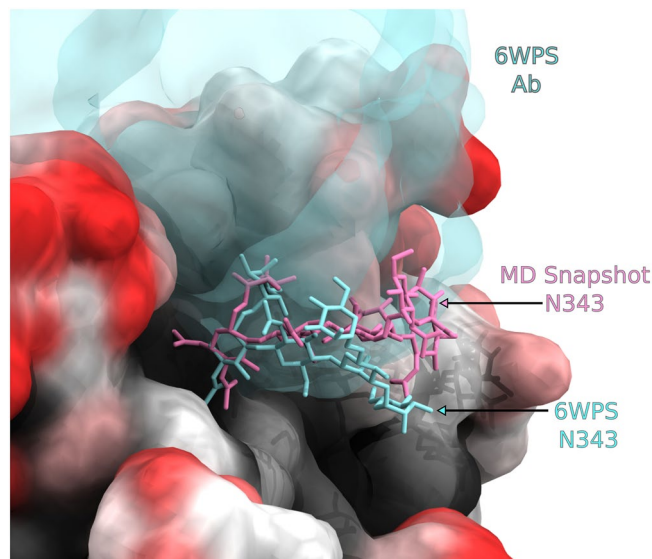


Figure 3. Image of the S309 antibody (cyan) observed in the crystallographic co-complex (PDB ID 6WPS⁵⁴) compared to a single pose from an MD simulation of the S-glycoprotein trimer. While numerous poses of the glycan at N343 were incompatible with antibody binding, there are poses within the MD trajectory that are similar to that found in the crystal structure co-complex with S309 that would permit binding.

binding, a subset of poses could be identified in the MD trajectory that would permit the binding of this mAb (Fig. 3). Thus, the degree of shielding is better captured by the more detailed AbASA analysis rather than by a visual overlay of structures.

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 pathogen are presented on antigen presenting cells by major histocompatibility complex II molecules, also known as human leukocyte antigen (HLA) complexes. HLA complexes have preferred peptide antigen motifs, and based on a knowledge of these preferences it is possible to predict which peptides in a pathogenic protein are likely to be HLA antigens^{55,56}. 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 if the glycan is small enough⁵⁷ or if it is found on the end of the peptide antigen where it doesn't interfere with HLA binding⁵⁸. The glycan-mediated shielding of predicted HLA antigens (Supplementary Table S2) derived from the S glycoprotein are shown in Fig. 4 and Supplementary Figures S2 and S4 for all HLA peptide sequences that also contain a glycosite.

As expected, glycosylation consistently decreased the surface exposure of the residues proximal to the glycosites (Fig. 4c), but also led to non-sequential changes in exposure, as a result of the 3D topology of the protein surface in the vicinity of each glycosite. Of the 18 glycosites in the 3D structure, 16 are present in putative HLA peptides. Although the glycans may occur throughout the HLA sequences (Supplementary 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, and may lead to antibodies that target these carbohydrates on the S glycoprotein⁵⁷. As a case in point, glycosite N343 is predicted to exist in an immunogenic HLA sequence (Supplementary Table S2), and recently a co-crystal structure of the SARS-CoV-2 S trimer has been reported in which a neutralizing antibody (S309) interacts with this epitope and with the glycan at N343⁵⁴. Anti-carbohydrate antibodies have been shown to be neutralizing in other viruses, such as HIV⁵⁹, and therefore glycosylated peptides can offer an alternative to more traditional peptide epitopes. Glycan immunogenicity is enhanced when the glycans or their clusters are significantly different from self, and thus are less immunologically tolerated⁹. Although viruses exploit the host glycosylation machinery in their biosynthesis, differences from typical host glycan distributions can occur when for example the virus cloaks itself so densely in glycosites that the glycans are not accessible to glycan processing enzymes, due to steric crowding, and remain in their high mannose form²⁸. Examples of this are seen in the high-mannose clusters in some strains of influenza²⁸ and in HIV⁶⁰. From the perspective of vaccine development⁶¹, 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 appropriate consideration of the choice of cell type for vaccine production³¹.

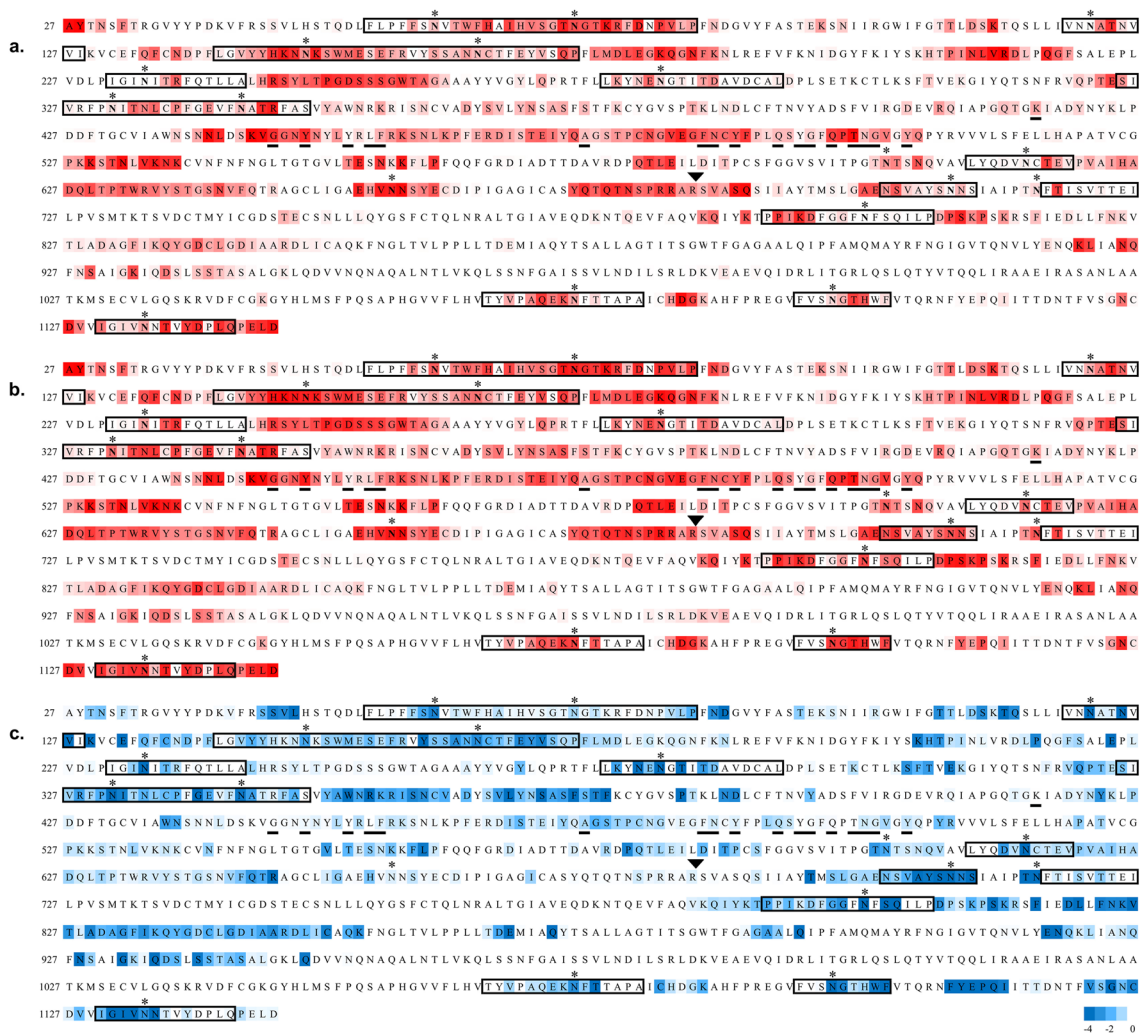


Figure 4. Sequence of the S protein (NCBI: YP_009724390.1) used to generate the 3D model of the glycoprotein. Residues 1–26 and 1,147–1,273 were not included in the 3D structure due to a lack of relevant template structures. Sequences within a rectangle were predicted to consist of one or more HLA antigens using the RankPep server (imed.med.ucm.es/Tools/rankpep^{55,56}). Glycosites are indicated with asterisks, residues reported to interact with the ACE2 receptor⁶² 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 (nude) protein. (c) The difference in accessibilities between the site-specific and non-glycosylated glycoforms is plotted as the fold change in epitope accessibility during the simulation, from – 4 (blue) to 0 (white), where blue indicates glycosylation-dependent surface shielding.

Discussion

The present study indicates that glycans shield approximately 40% of the underlying protein surface of the S glycoprotein trimer from antibody recognition, and that this value is relatively insensitive to glycan type. This suggests that although viral glycan microheterogeneity varies according to host cell type and potentially therefore between infected individuals, the efficacy of antisera should not be profoundly impacted by such differences. This conclusion is consistent with emerging data that indicate the efficacy of convalescent sera^{63–65} in treating COVID-19 patients.

The observation that homogeneously glycosylated glycoforms are predicted to display approximately the same level of antibody shielding as those computed for the more relevant site-specific glycoform suggests that hypothetical models of glycosylation can be usefully applied in advance of the reporting of experimental glycomics data. As in all MD simulations, accurate and adequate sampling of the conformational poses of the glycans and the protein is essential to avoid anecdotal results. Here, multiple independent simulations were performed into the μ s regime. The ability to predict the impact of protein glycosylation on epitope recognition is significant as it enables the effects of glycosite alterations to be estimated in anticipation of seasonal antigenic drift. By analogy with influenza^{66,67}, variations in glycosite location arising from antigenic drift can be expected to have a profound effect on SARS-CoV-2 S protein antigenicity and potentially vaccine efficacy. Fortunately, the most accessible and largest epitope in the S trimer consists of the ACE2 binding domain, where the virus cannot

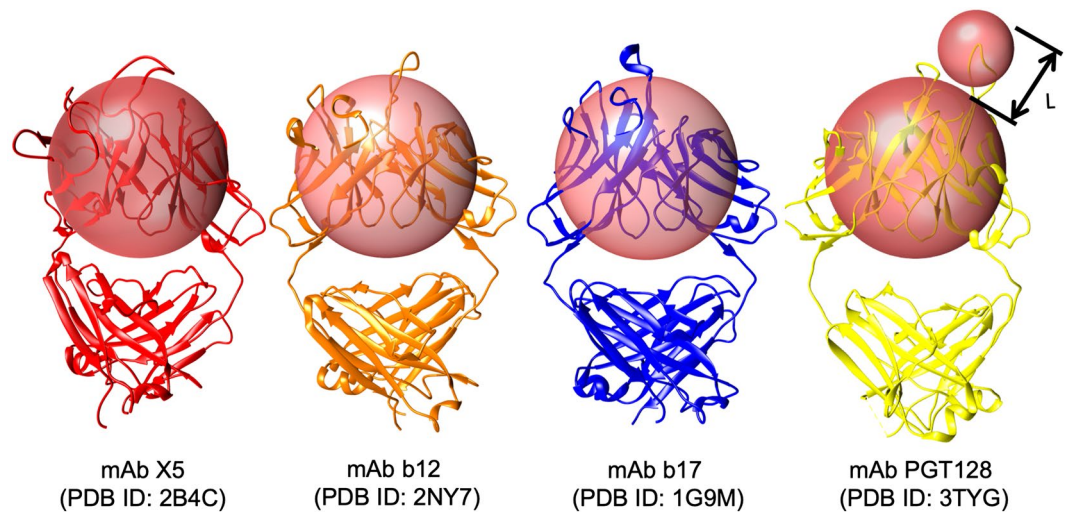


Figure 5. Antibody accessible surface area estimation using a pair of spherical probes. To estimate the AbASA, a CDR spherical probe was derived (radius 7.2 Å, smaller sphere) that approximates the average size of the hypervariable loops in the CDR from four anti-gp120 antibodies, in which the epitopes were either protein surface residues (PDB IDs: 2B4C⁸⁷, 2NY7⁸⁸, 1G9M⁸⁹) or both carbohydrate and protein residues: (3TYG⁹⁰). 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. Images generated with UCSF Chimera⁹¹.

exploit glycan shielding or mutational changes to evade host immune response without potentially attenuating its fitness. The requirement that the virus maintain the integrity of the ACE2 RBD suggests that a vaccine that includes this epitope may maintain efficacy despite antigenic drift, as long as the virus continues to target the same host receptor.

While overall shielding of the underlying protein surface does not appear to be highly sensitive to glycan microheterogeneity, it could certainly impact the innate immune response by altering the ability of collectins and other lectins of the immune system to effectively bind to the S glycoprotein and neutralize the virus²⁸, and may impact the adaptive immune response by altering the number of viable HLA antigens. Given that in humans, glycan microheterogeneity varies between individuals, and depends on many factors, including age^{68,69}, underlying disease^{70,71} and ethnicity⁷², access to 3D models of the S glycoprotein may aid in defining the molecular basis for the observed differential susceptibilities among individuals to COVID-19^{73,74}.

The level of agreement between the predicted epitopes based on MD simulation of the S trimer and those observed by epitope mapping or in co-crystal structures, suggests the AbASA approach to epitope identification could play a useful role in guiding the development of vaccine strategies as well as in interpreting differences in the responses of individuals to either the disease severity or vaccine efficacy.

Methods

SARS-CoV-2 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)¹⁷, was obtained from the SWISS-MODEL server (swissmodel.expasy.org). The model has 95% coverage (residues 27 to 1,146) of the S protein.

S trimer 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 ranges⁷⁵ to relieve any atomic overlaps with the core protein, as described previously^{66,76}. The site specific glycans used to model a glycoform representative of the data obtained from the S glycoprotein expressed in HEK293 cells⁷⁷, are presented in Supplementary Table S1.

Energy minimization and molecular dynamics simulations. Each glycosylated structure was placed in a periodic box of approximately 130,000 TIP3P water molecules⁷⁸ 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 descent, followed by 10,000 conjugant gradient) under constant pressure (1 atm) and temperature (300 K) conditions. All MD simulations were performed under nPT conditions with the CUDA implementation of the PMEMD^{79,80} simulation code, as present in the Amber14 software suite⁸¹. The GLYCAM06j force field⁸² and Amber14SB force field⁸³ 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) method⁸⁴. Covalent bonds involving hydrogen were constrained with the SHAKE algorithm, allowing an integration time step of 2 fs⁸⁵ to be employed. The energy minimized coordinates were equilibrated at 300 K over 400 ps with restraints on the solute heavy atoms. Each system was then equilibrated with restraints on the Ca atoms of the protein for 1 ns, prior to initiating 6 independent production MD simulations with no restraints applied for a total time of 3 μ s for the HEK293 glycoform, or 3 runs for a total of 1 μ s for the homogeneously glycosylated theoretical glycoforms.

Antigenic surface analysis. The antigenic surface area was calculated as the sum of the surface areas of any protein residues that make contact with the CDR probe, provided that the CDR probe is proximal to the Fv probe (Fig. 5). This latter requirement is governed by “L”, 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. These probe sizes may be compared to values of 5 and 10 Å employed previously to estimate antigenic surface area⁸⁶. Changes in the solvent accessible surface area (SASA) showed no significant shielding by glycans and thus a simple SASA model (radius 1.4 Å) was not useful for this analysis⁸⁶. Failure to include the Fv domain probe, that is, probing the surface only with the smaller 7.2 Å sphere, led to the detection of contacts in narrow deep crevasses on the protein surface (Supplementary Figure S5), that while theoretically possible, would require an exceptionally long hypervariable loop.

Received: 29 April 2020; Accepted: 17 August 2020

Published online: 14 September 2020

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. <https://doi.org/10.1074/jbc.M112.371898> (2012).
3. Tate, M. D. *et al.* Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection. *Viruses* **6**, 1294–1316. <https://doi.org/10.3390/v6031294> (2014).
4. Helle, F., Duverlie, G. & Dubuisson, J. The hepatitis C virus glycan shield and evasion of the humoral immune response. *Viruses* **3**, 1909–1932. <https://doi.org/10.3390/v3101909> (2011).
5. Marth, J. D. & Grewal, P. K. Mammalian glycosylation in immunity. *Nat. Rev. Immunol.* **8**, 874–887 (2008).
6. Pereira, M. S. *et al.* Glycans as key checkpoints of T cell activity and function. *Front. Immunol.* **9**, 2754. <https://doi.org/10.3389/fimmu.2018.02754> (2018).
7. Baum, L. G. & Cobb, B. A. The direct and indirect effects of glycans on immune function. *Glycobiology* **27**, 619–624. <https://doi.org/10.1093/glycob/cwx036> (2017).
8. Vigerust, D. J. & Shepherd, V. L. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol.* **15**, 211–218. <https://doi.org/10.1016/j.tim.2007.03.003> (2007).
9. Crispin, M. & Doores, K. J. Targeting host-derived glycans on enveloped viruses for antibody-based vaccine design. *Curr. Opin. Virol.* **11**, 63–69. <https://doi.org/10.1016/j.coviro.2015.02.002> (2015).
10. 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. <https://doi.org/10.3389/fimmu.2018.01998> (2018).
11. Hutter, J. *et al.* Toward animal cell culture-based influenza vaccine design: viral hemagglutinin N-glycosylation markedly impacts immunogenicity. *J. Immunol.* **190**, 220–230. <https://doi.org/10.4049/jimmunol.1201060> (2013).
12. Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410. <https://doi.org/10.1126/science.1124513> (2006).
13. 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. <https://doi.org/10.1371/journal.ppat.1003831> (2014).
14. 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. <https://doi.org/10.1128/JVI.01023-13> (2013).
15. Altman, M. O. *et al.* Human influenza A virus hemagglutinin glycan evolution follows a temporal pattern to a glycan limit. *mBio* <https://doi.org/10.1128/mBio.00204-19> (2019).
16. 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. USA* **114**, 12578–12583. <https://doi.org/10.1073/pnas.1712377114> (2017).
17. Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* <https://doi.org/10.1126/science.abb2507> (2020).
18. 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. <https://doi.org/10.1038/ncomms15092> (2017).
19. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* <https://doi.org/10.1126/science.abb9983> (2020).
20. Homans, S. W., Dwek, R. A. & Rademacher, T. W. Solution conformations of N-linked oligosaccharides. *Biochemistry* **26**, 6571–6578. <https://doi.org/10.1021/bi00395a001> (1987).
21. Homans, S. W. *et al.* Conformational transitions in N-linked oligosaccharides. *Biochemistry* **25**, 6342–6350. <https://doi.org/10.1021/bi00368a076> (1986).
22. Woods, R. J. Predicting the structures of glycans, glycoproteins, and their complexes. *Chem. Rev.* **118**, 8005–8024. <https://doi.org/10.1021/acs.chemrev.8b00032> (2018).
23. Yamaguchi, T. *et al.* Exploration of conformational spaces of high-mannose-type oligosaccharides by an NMR-validated simulation. *Angew. Chem. Int. Ed.* **53**, 10941–10944. <https://doi.org/10.1002/anie.201406145> (2014).
24. Sawen, E., Svensson, B., Ostervall, J., Maliniak, A. & Widmalm, G. Molecular conformations in the pentasaccharide LNF-1 derived from NMR spectroscopy and molecular dynamics simulations. *J. Phys. Chem. B* **115**, 7109–7121. <https://doi.org/10.1021/jp2017105> (2011).
25. Jo, S., Qi, Y. & Im, W. Preferred conformations of N-glycan core pentasaccharide in solution and in glycoproteins. *Glycobiology* **26**, 19–29. <https://doi.org/10.1093/glycob/cwv083> (2016).

26. Harbison, A. & Fadda, E. An atomistic perspective on ADCC quenching by core-fucosylation of IgG1 Fc N-glycans from enhanced sampling molecular dynamics. *Glycobiology* <https://doi.org/10.1093/glycob/cwz101> (2019).
27. Hang, I. *et al.* Analysis of site-specific N-glycan remodeling in the endoplasmic reticulum and the Golgi. *Glycobiology* **25**, 1335–1349. <https://doi.org/10.1093/glycob/cwv058> (2015).
28. 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. <https://doi.org/10.1074/mcp.M116.058016> (2016).
29. Amaro, R. E. & Li, W. W. Molecular-level simulation of pandemic influenza glycoproteins. *Methods Mol. Biol.* **819**, 575–594. https://doi.org/10.1007/978-1-61779-465-0_34 (2012).
30. An, Y. *et al.* N-glycosylation of seasonal influenza vaccine hemagglutinins: implication for potency testing and immune processing. *J. Virol.* <https://doi.org/10.1128/JVI.01693-18> (2019).
31. An, Y. *et al.* Comparative glycomics analysis of influenza hemagglutinin (H5N1) produced in vaccine relevant cell platforms. *J. Proteome Res.* **12**, 3707–3720. <https://doi.org/10.1021/pr400329k> (2013).
32. Singh, A. *et al.* Extension of the GLYCAM force field parameters for glycosaminoglycans and their validation. *Can. J. Chem.* **94**, 1–9 (2016).
33. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655. <https://doi.org/10.1002/jcc.20820> (2008).
34. Tessier, M. B., DeMarco, M. L., Yongye, A. B. & Woods, R. J. Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Mol. Simul.* **34**, 349–364. <https://doi.org/10.1080/08927020701710890> (2008).
35. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
36. Horiya, S., MacPherson, I. S. & Krauss, I. J. Recent strategies targeting HIV glycans in vaccine design. *Nat. Chem. Biol.* **10**, 990–999. <https://doi.org/10.1038/nchembio.1685> (2014).
37. Doores, K. J. The HIV glycan shield as a target for broadly neutralizing antibodies. *FEBS J.* **282**, 4679–4691. <https://doi.org/10.1111/febs.13530> (2015).
38. Hubbard, S. J. & Thornton, J. M. NACCESS v. 2.1.1. Computer Program. Department of Biochemistry and Molecular Biology, University College London (1993).
39. Farrera, L. *et al.* Identification of immunodominant linear epitopes from SARS-CoV-2 patient plasma. *MedRxiv.* <https://doi.org/10.1101/2020.06.15.20131391> (2020).
40. Prabhakaran, P. *et al.* Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J. Biol. Chem.* **281**, 15829–15836. <https://doi.org/10.1074/jbc.M600697200> (2006).
41. 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. <https://doi.org/10.1074/jbc.M603275200> (2006).
42. 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. <https://doi.org/10.1016/j.jmb.2009.03.042> (2009).
43. 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. <https://doi.org/10.1038/ncomms9223> (2015).
44. Wang, L. *et al.* Evaluation of candidate vaccine approaches for MERS-CoV. *Nat. Commun.* **6**, 7712. <https://doi.org/10.1038/ncomms8712> (2015).
45. 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. <https://doi.org/10.1038/cr.2015.113> (2015).
46. 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. <https://doi.org/10.1093/infdis/jix209> (2017).
47. Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci. USA* **114**, E7348–E7357. <https://doi.org/10.1073/pnas.1707304114> (2017).
48. 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. <https://doi.org/10.1016/j.celrep.2018.06.041> (2018).
49. 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.* <https://doi.org/10.1128/JVI.02002-17> (2018).
50. Zhou, H. *et al.* Structural definition of a neutralization epitope on the N-terminal domain of MERS-CoV spike glycoprotein. *Nat. Commun.* **10**, 3068. <https://doi.org/10.1038/s41467-019-10897-4> (2019).
51. Walls, A. C. *et al.* Unexpected receptor functional mimicry elucidates activation of coronavirus fusion. *Cell* **176**, 1026–1039 e1015. <https://doi.org/10.1016/j.cell.2018.12.028> (2019).
52. Wang, N. *et al.* Structural definition of a neutralization-sensitive epitope on the MERS-CoV S1-NTD. *Cell Rep.* **28**, 3395–3405 e3396. <https://doi.org/10.1016/j.celrep.2019.08.052> (2019).
53. Tai, W. *et al.* Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol. Immunol.* **17**, 613–620. <https://doi.org/10.1038/s41423-020-0400-4> (2020).
54. Pinto, D. *et al.* Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *BioRxiv.* <https://doi.org/10.1101/2020.04.07.023903>.
55. 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. <https://doi.org/10.1007/s00251-004-0709-7> (2004).
56. Reche, P. A., Glutting, J. P. & Reinherz, E. L. Prediction of MHC class I binding peptides using profile motifs. *Hum. Immunol.* **63**, 701–709. [https://doi.org/10.1016/s0198-8859\(02\)00432-9](https://doi.org/10.1016/s0198-8859(02)00432-9) (2002).
57. 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. *Nat. Med.* **17**, 1602–1609. <https://doi.org/10.1038/nm.2535> (2011).
58. 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. <https://doi.org/10.1021/acs.jproteome.6b00496> (2017).
59. Haji-Ghassemi, O., Blackler, R. J., Martin Young, N. & Evans, S. V. Antibody recognition of carbohydrate epitopes. *Glycobiology* **25**, 920–952. <https://doi.org/10.1093/glycob/cwv037> (2015).
60. Bonomelli, C. *et al.* The glycan shield of HIV is predominantly oligomannose independently of production system or viral clade. *PLoS ONE* **6**, e23521. <https://doi.org/10.1371/journal.pone.0023521> (2011).
61. Chang, D. & Zaia, J. Why glycosylation matters in building a better flu vaccine. *Mol. Cell Proteomics* **18**, 2348–2358. <https://doi.org/10.1074/mcp.R119.001491> (2019).
62. Lan, J. *et al.* Crystal structure of the 2019-nCoV spike receptor-binding domain bound with the ACE2 receptor. *BioRxiv.* <https://doi.org/10.1101/2020.02.19.956235>.
63. Li, L. *et al.* Effect of convalescent plasma therapy on time to clinical improvement in patients with severe and life-threatening COVID-19: a randomized clinical trial. *JAMA* <https://doi.org/10.1001/jama.2020.10044> (2020).
64. Casadevall, A., Joyner, M. J. & Pirofski, L. A. A randomized trial of convalescent plasma for COVID-19-potentially hopeful signals. *JAMA* <https://doi.org/10.1001/jama.2020.10218> (2020).
65. Casadevall, A. & Pirofski, L. A. The convalescent sera option for containing COVID-19. *J. Clin. Invest.* **130**, 1545–1548. <https://doi.org/10.1172/JCI138003> (2020).

66. 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. <https://doi.org/10.1016/j.chom.2016.11.004> (2017).
67. 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. <https://doi.org/10.1128/JVI.00593-13> (2013).
68. Dall'Olio, F. *et al.* N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res. Rev.* **12**, 685–698. <https://doi.org/10.1016/j.arr.2012.02.002> (2013).
69. Kristic, J. *et al.* Glycans are a novel biomarker of chronological and biological ages. *J. Gerontol. A Biol. Sci. Med. Sci.* **69**, 779–789. <https://doi.org/10.1093/gerona/glt190> (2014).
70. Pavic, T. *et al.* N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *J. Transl. Med.* **16**, 323. <https://doi.org/10.1186/s12967-018-1695-0> (2018).
71. Rudman, N., Gornik, O. & Lauc, G. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. *FEBS Lett.* **593**, 1598–1615. <https://doi.org/10.1002/1873-3468.13495> (2019).
72. Gebrehiwot, A. G. *et al.* Healthy human serum N-glycan profiling reveals the influence of ethnic variation on the identified cancer-relevant glycan biomarkers. *PLoS ONE* **13**, e0209515. <https://doi.org/10.1371/journal.pone.0209515> (2018).
73. Davies, N. G. *et al.* Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0962-9> (2020).
74. Millett, G. A. *et al.* Assessing differential impacts of COVID-19 on black communities. *Ann. Epidemiol.* <https://doi.org/10.1016/j.annepidem.2020.05.003> (2020).
75. 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. <https://doi.org/10.1002/jcc.23517> (2014).
76. 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. <https://doi.org/10.1093/glycob/cww094> (2016).
77. Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific analysis of the SARS-CoV-2 glycan shield. *BioRxiv.* <https://doi.org/10.1101/2020.03.26.010322> (2020).
78. 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. <https://doi.org/10.1021/ja00392a016> (1981).
79. 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. <https://doi.org/10.1021/ct400314y> (2013).
80. Gotz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized born. *J. Chem. Theory Comput.* **8**, 1542–1555. <https://doi.org/10.1021/ct200909j> (2012).
81. Case, D. A. *et al.* AMBER 14. Computer Program. University of California, San Francisco (2014).
82. Kirschner, K. N. *et al.* GLYCAM06: a generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.* **29**, 622–655. <https://doi.org/10.1002/jcc.20820> (2008).
83. Maier, J. A. *et al.* ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713. <https://doi.org/10.1021/acs.jctc.5b00255> (2015).
84. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: an N⁻¹ log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993).
85. 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).
86. 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.* <https://doi.org/10.1128/JVI.01403-18> (2019).
87. Huang, C. C. *et al.* Structure of a V3-containing HIV-1 gp120 core. *Science* **310**, 1025–1028. <https://doi.org/10.1126/science.1118398> (2005).
88. Zhou, T. *et al.* Structural definition of a conserved neutralization epitope on HIV-1 gp120. *Nature* **445**, 732–737. <https://doi.org/10.1038/nature05580> (2007).
89. Kwong, P. D. *et al.* Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **8**, 1329–1339. [https://doi.org/10.1016/s0969-2126\(00\)00547-5](https://doi.org/10.1016/s0969-2126(00)00547-5) (2000).
90. Pejchal, R. *et al.* A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. *Science* **334**, 1097–1103. <https://doi.org/10.1126/science.1213256> (2011).
91. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

Acknowledgements

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.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71748-7>.

Correspondence and requests for materials should be addressed to R.J.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020