Fine-tuning the Spike: Role of the nature and topology of the glycan shield in the structure and dynamics of SARS-CoV-2 S

Aoife M. Harbison[1,‡], Carl A. Fogarty[1,‡], Toan K. Phung[2], Akash Satheesan[1], Benjamin L. Schulz[2] and Elisa Fadda[1,*]

[1]Department of Chemistry and Hamilton Institute, Maynooth University, Maynooth, Kildare, Ireland, [2]School of Chemistry and Molecular Biosciences, The University of Queensland, St Lucia, QLD, Australia

Abstract

The SARS-CoV-2 spike (S) is a type I fusion glycoprotein, responsible for initiating the infection leading to COVID19. As a feature unique of SARS-CoV-2, the thick glycan shield covering the S protein is not only essential for hiding the virus from immune detection, but it also plays multiple functional roles, stabilizing the S prefusion "open" conformation, which is competent for binding the ACE2 primary receptor, and gating the open-to-close transitions. This newly discovered functions of the glycan shield suggest the evolution of its sites of glycosylation is potentially intertwined with the evolution of the overall protein sequence to affect optimal activity. Furthermore, recent studies indicate that the occupancy and structures of SARS-CoV-2 S glycosylation depends not only on the host-cell, but also on the structural stability of the prefusion trimer; a point that raises important questions about the relative binding competence of different glycoforms. In this work we use multi-microsecond molecular dynamics simulations to characterize the structure and dynamics of different SARS-CoV-2 S models with different N-glycans at key functional sites, namely N234, N165 and N343. We also assessed the effect of a change in the SARS-CoV-2 S glycan shield's topology at N370, due to the recently acquired T372A mutation. Our results indicate that the structures of the N-glycans at N234, N165 and N343 affect the stability of the active (or open) S conformation, and thus its exposure and accessibility. Furthermore, while glycosylation at N370 stabilizes the open S conformation, we find that the N370 glycan binds the closed receptor binding domain (RBD) surface, essentially tying the closed protomers together. These results suggest that the loss of the N370 glycosylation site in SARS-CoV-2 may have increased the availability of the open S form, perhaps contributing to its higher

[*] Correspondence to: elisa.fadda@mu.ie
[‡] These authors equally contributed to the work.

infectivity relative to CoV1 and other variants carrying the sequon. Finally, we discuss these specific changes to the topology of the SARS-CoV-2 S glycan shield through ancestral sequence reconstruction of select SARS strains and discuss how they may have evolved to affect S activity.

Introduction

Spike (S) glycoproteins mediate the adhesion and fusion of enveloped viruses to the host cell, initiating viral infection[1-3]. The interaction with the cell-bound receptor leads to a complex conformational change, dependent on the S architecture[4, 5], that terminates with the fusion of the viral envelope with the host cell's membrane, giving the virus access to the cellular machinery for replication[3]. Viral envelope S are heavily coated with a dense layer of complex carbohydrates, also known as a glycan shield, that performs different intrinsic and extrinsic functions, including in protein folding, stability and trafficking, to masking the virus from the immune system and mediating contacts with lectins and antibodies[1]. The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has a trimeric S glycoprotein[6, 7] protruding from the viral envelope surface[8, 9]. The SARS-CoV-2 S has 22 N-glycosylation sequons per protomer, of which at least 18 appear to be consistently occupied in different constructs[8, 10-15], and O-glycosylation sites, with a lower level of occupancy[11-13, 16].

A unique and key feature of the SARS-CoV-2 S glycoprotein is the direct involvement of the glycan shield in its activation mechanism[17, 18]. Binding of SARS-CoV-2 S to its primary receptor, namely the angiotensin-converting enzyme 2 (ACE2)[19, 20], requires the opening of one (or more) receptor binding domains (RBDs), which emerge from the glycan shield[17, 18] and become accessible. The RBD opening creates a cavity within the SARS-CoV-2 S prefusion trimer's structure[6, 7], see **Figure 1**, which extends deeply into the trimer's core. In the absence of strategically positioned N-glycans as a support[17], upon RBD opening this large cleft would be filled by water molecules, likely weakening the S prefusion structure. This is especially important in view of the presence of the SARS-CoV-2 S pre-cleaved furin site at the S1/S2 boundary[21, 22]. Previous work based on multi-microsecond molecular dynamics (MD) simulations in conjunction with biolayer interferometry experiments[17], has shown that this structural weakness is effectively recovered by the N-glycan at position N234, which is able to fill the cavity, supporting the RBD open conformation[17]. Additionally, MD simulations have also shown that the N-glycans at positions N165 and N343, see **Figure 1,**

are directly involved in important interactions with residues of the open RBD, supporting its open conformation[17] and mediating (or gating) its transition from open to closed[18], respectively.

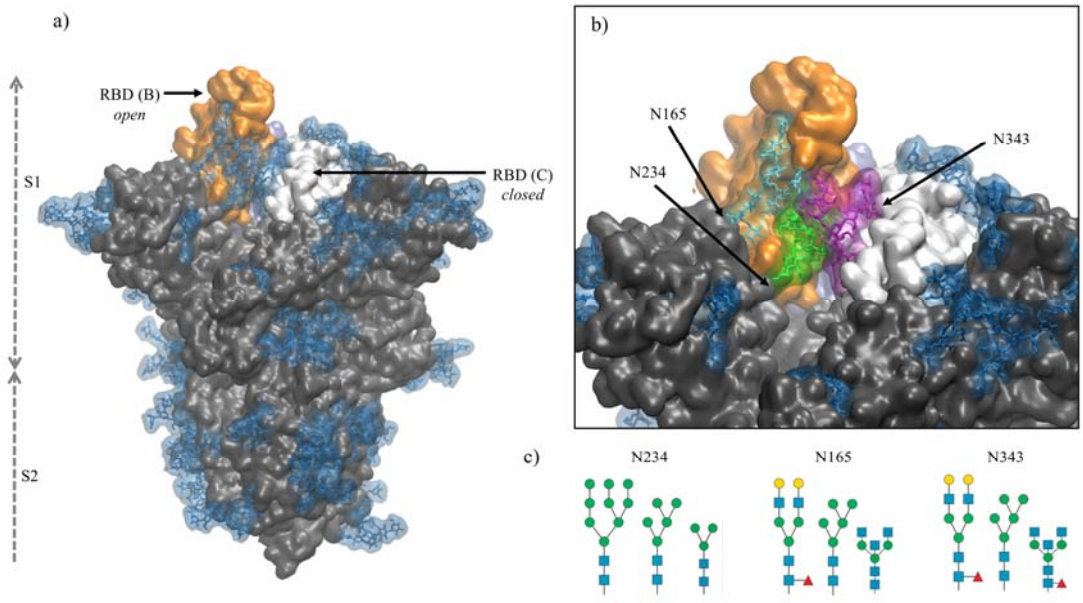

Figure 1. Panel a) Structure of the fully glycosylated SARS-CoV-2 S (PDBid 6VYB)[6] ectodomain. The protein is shown in grey with the RBDs of chain B and C highlighted in orange and white, respectively; the glycan shield is highlighted in blue. **Panel b)** Close up on the open pockets with the N-glycans at the strategic positions N234, N165, and N343 highlighted in green, cyan and purple, respectively. **Panel c)** N-glycans considered in the different models studied in this work, represented through the SNFG[23] and drawn with DrawGlycan[24] (http://www.virtualglycome.org/DrawGlycan/). Molecular rendering done with VMD[25].

The crucial role of the N-glycans at N234, N165 and N343 is exerted through the contacts they are able to make with protein residues, both in the open RBD (of chain B in the models studied here) and in the adjacent closed RBD (chain C) at either side of the empty cleft[17], see **Figure 1**. Their ability to engage in effective interactions is intrinsically linked to the nature, size, sequence and branching of the N-glycans at these sites, opening the floor to a broader discussion on the relative structural stability and dynamics of different S glycoforms. This is a very important point to explore, especially in view of the design of specific antiviral therapeutic strategies targeting glycosylation[26], yet a very difficult one to systematically address experimentally.

In this work, we present the results of a set of multi-microsecond MD simulations of different SARS-CoV-2 S glycoforms aimed at studying the effect of changes in the type of glycosylation at positions N234, N165 and N343, see **Figure 1**, while the rest of the glycan shield is represented consistently with a stable recombinant S prefusion trimer (S_{trimer})[10, 17].

More specifically, we investigated if shorter oligomannoses structures at N234, such as Man5 and paucimannose (Man3), see **Figure 1**, could occupy the empty cleft in the open SARS-CoV-2 S as effectively as larger oligomannoses such as Man9, which are characteristic of a highly stable prefusion S_{trimer}[10]. As an important note, Man5, rather than Man7/8/9 is found to be present, or even to be the dominant glycosylation type at N234 in the virus, vaccine epitopes and in some recombinant S constructs[8, 11, 12]. We also explored how the role of oligomannose structures at N234 is supplemented by the complex N-glycans at N165 and N343, which are able to form a tight network of glycan-glycan and glycan-protein interactions that stabilize the orientation and dynamics of the open RBD across different possible orientations.

Finally, we also investigated the effect of a recently acquired mutation of the SARS-CoV-2 Wuhan strain that changes the RBD glycan shield's topology. More specifically, in SARS-CoV and MERS[27], as well as in the bat RaTG13 and pangolin CoV SARS-CoV-2 variants[22, 28, 29], position N370 on the RBD is part of an occupied NST sequon[27], which is lost in SARS-CoV-2 due to the recently acquired T372A mutation. We performed ancestral sequence reconstruction of selected SARS S sequences to investigate the gain and loss of glycosylation sequons during S evolution. To address the effect of this change in the glycan shield's topology, we re-introduced the N-glycan at N370 in the SARS-CoV-2 S native sequence and ran MD simulations to study its effect on the stability of both, the open and closed RBDs. Our simulations indicate that within the SARS-CoV-2 S architecture, glycosylation at N370 does not interfere with the glycan network at N234, N165 and N343, but actively contributes to it by very effectively stabilizing the RBD open conformation.

Interestingly, analysis of the closed protomers shows that the N370 glycan from RBD (A) is firmly bound to the RBD (C) surface, where it occupies a specific cleft. This interaction results in tying the closed RBD (C) to the adjacent closed RBD (A), very much like the laces in a shoe, thus potentially hindering the opening of the RBDs. Based on these findings, we propose that the recent loss of N-glycosylation at N370 allows for a higher availability of open S conformations by lowering the energetic cost of the opening reaction, which is likely to be beneficial by providing higher infectivity of SARS-CoV-2 relative to closely related variants carrying this sequon. We also discuss how the cleft on the closed RBD surface in SARS-CoV-2 S, which is occupied by the N370 glycan in S glycoforms with the sequon, may be used by other glycans, such as glycosaminoglycans[30-33] and/or sialic acid-containing

glycans and glycolipids[34, 35], where these interactions may contribute to increasing the S cell-surface localization.

Results

In this section we will present the results of extensive conformational sampling based on multi-microsecond MD simulations of SARS-CoV-2 S models with different glycosylation at N234, N165 and N343, see **Table 1**. Specifically, we determined the effects of site-specific glycan structure on the SARS-CoV-2 S ectodomain's structure and dynamics by systematically reducing the size of the oligommanose at N234; from Man9 (N234-Man9), found in highly stable prefusion SARS-CoV-2 S constructs[10, 12, 13], and also studied in previews work[17, 18], we progressed to a shorter Man5 (N234-Man5), found on the virus and on other SARS-CoV-2 S constructs[11, 12], and to paucimannose (N234-Man3) as a hypothetical size limit. In these models the glycosylation at N165 and N343 is biantennary complex (FA2G2), and in one case we have also considered bisecting GlcNAc F/A2B forms at N343 and N165, respectively, see **Figure 1 and Table S.1**. The results obtained for the N234-Man5 models were compared to a SARS-CoV-2 S model with a uniform immature glycosylation[36], i.e. in which all N-glycans are Man5 (all-Man5). Finally, to assess the effect of loss of glycosylation at N370 on SARS-CoV-2 S, we added a complex N-glycan (FA2G2) at N370 in the N234-Man9 model with bisecting GlcNAc (F/A2B) N-glycans at N165 and N343, respectively. In all of these models the glycosylation at all sites other than the ones listed above is consistent and the same as the profile experimentally determined for the stable prefusion S_{trimer}[10], see **Table S.1**. Results are based on the analysis of multiple uncorrelated MD trajectories (replicas) run in parallel for each model (see details in the Computational Methods section in Supplementary Material).

N234-Man5

The structure and dynamics of the RBD (residues 330 to 530) from a model with a Man5 at N234 and FA2G2 at N165 and N343 was analysed based on three replicas with sampling times indicated in **Table 1**. Note that the first 300 ns of each replica were omitted from the analysis to allow conformational equilibration. The dynamics of the RBD is rather complex and for consistency with previous work[17] we defined it here in terms of lateral and axial angle values populations. The lateral angle indicates an in-plane motion of the open RBD along a hypothetical circle centred on the central helices of the spike. This angle is defined by three

points, corresponding to 1) the centre of mass of the open RBD core β-sheets at frame 0, i.e. the orientation of the RBD in PDB 6VYB, 2) the centre of mass of the top section of the central helices, and 3) the centre of mass of the open RBD core β-sheets at each frame along the trajectory[17]. The axial-angle defines a tilting motion of the open RBD either away from (negative values) or toward (positive values) the central helices of the spike. The axial angle is defined by 1) the centre of mass of the open RBD core β-sheets, 2) the centre of mass of the central helices, and 3) the centre of mass of the top section of the central helices[17].

Results shown in **Figure 2** and **Table 1** indicate that in the N234-Man5 S model the open RBD can access a larger conformational space relative to the N243-Man9 model along the lateral displacement coordinate, while it is still able to adopt a stable open conformation, indicated by the axial tilt values. These open conformations are slightly diverse, encompassing different degrees of opening, as shown in **Figure S.1**. More specifically, the RBD can be 'wide-open', see in **Figure 2 panel b**, i.e. in a similar orientation seen in the cryo-EM structures[6], stabilized by interactions involving different RBD residues and the N165 and N234 N-glycans, while the N343 N-glycan is relatively free on the opposite side of the pocket, or alternatively the RBD can be in an intermediate conformation between the wide-open and the closed. In the latter the N-glycans at all three positions N165, N234 and N343 are involved in complex interactions with protein residues in the RBD, namely with the loop from L460 and F489, and in and around the pocket, and with each other, see **Figure S.2**. Within this framework, the higher lateral degree of flexibility of the open RBD in the N234-Man5 S model is due to the smaller size of the Man5 glycan and its specific conformational propensity relative to Man9, which tends to adopt a more tree-like conformation supported by extensive inter-arm interactions[37] that fills the pocket much more effectively, see **Figure 2 panels e** and **f**. More specifically, the Man5 at N234 appears to be less competent than Man9 at forming interactions that bridge both sides of the pocket as it accesses the cavity left open by the RBD, while it mainly interacts with residues at the base of the open RBD, leaving it 'unhinged' when in the "wide-open" conformation, relative to the N234-Man9 model. As an interesting point to note, in all of these trajectories the core fucose of the N343 FA2G2 glycan is exposed to the solvent, potentially allowing for its recognition, see **Figure 2**. This is in agreement with cryo-EM studies reporting interactions involving the core fucose of the N-glycan at N343 and human neutralizing antibodies[38].

To further assess the role of the type of glycosylation around the N234-Man5 in the RBD dynamics, we ran three MD simulations of a model with uniform Man5 glycosylation (all Man5), see **Table 1** and **Figure S.3**. The results indicate that the replacement of the complex glycans with immature Man5 structures is problematic for the stability of the open RBD. More specifically, the shorter size and 3D architecture of the Man5 at N343 does not allow it to engage effectively with residues in the RBD, while the Man5 at N234, as seen for the N234-Man5 model, is not able to effectively engage with the closed RBD (chain C) that flank the opposite side of the empty pocket. These results agree with recent work indicating that immature glycosylation, achieved in in GnTI-/- mutant cells, leads to a less competent S glycoprotein relative to the fully glycosylated variant[36]. For comparison we tested a N234-Man9 model where only the N-glycans at N165 and N343 were Man5. The results shown in **Table 1** and **Figure S.4** indicate that the presence of a large oligomannose structure at N243 helps recover the instability by more effectively occupying the empty pocket. In the equilibrium conformation sampled through 2.1 μs, the RBD conformation is more closed, shown also by the positive value of the axial angle, see **Table 1**, where both Man5 at N165 and N343 can interact with each other and with the RBD residues, see **Figure S.4.**

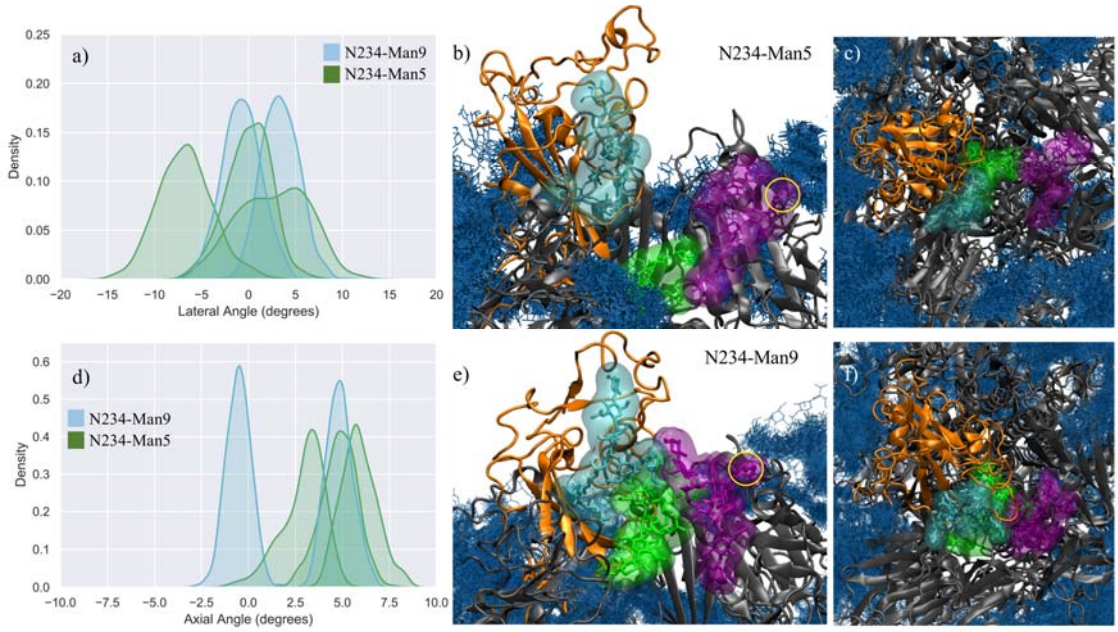

Figure 2. Panels a) and **d)** Kernel density estimation (KDE) analysis of the lateral and axial angles distributions calculated through the uncorrelated MD trajectories obtained for N234-Man5 (green), replicas R1-3, and for N234-Man9 (cyan), replicas R1,2 for comparison. **Panels b)** and **c)** Close-ups on a representative snapshot of the N234-Man5 (R3) simulation from the side and top, respectively. Man5 at N234 is shown in green, while the FA2G2 N-glycans at N165 and N343 are shown in cyan and purple, respectively. The open RBD is shown in orange cartoons, while the rest of the protein is shown in grey. All other glycans are shown in blue, as an overlay of snapshots collected every 10 frames. The position of the core fucose in the FA2G2 N-glycans at N343 is highlighted within a yellow circle. **Panels e)** and **f)** Close-ups on a representative

snapshot of the N234-Man9 (R2) simulation from the side and top, respectively. Man9 at N234 is shown in green, while the FA2G2 N-glycans at N165 and N343 are shown in cyan and purple, respectively. The position of the core fucose in the FA2G2 N-glycans at N343 is highlighted within a yellow circle. Data analysis and graphs were done with *seaborn* (www.seaborn.pydata.org) and molecular rendering with VMD[25].

Table 1. Orientation of the RBD in different SARS-CoV-2 S glycoforms in terms of average lateral and axial angle values relative to the SARS-CoV-2 S ectodomain, as described in ref.[17]. Standard deviation values are indicated in parentheses. Total simulation time for each replica is indicated in parentheses. Note: the first 300 ns of each trajectory were omitted from the analysis and considered as part of conformational equilibration.

Model System	Replicas	Lateral Angle (°)	Axial Angle (°)
N234-Man5	R1 (2 µs)	0.0 (2.4)	2.9 (1.1)
	R2 (2 µs)	-6.7 (2.9)	4.8 (0.9)
	R3 (1.5 µs)	2.7 (3.7)	5.9 (0.9)
N234-Man3	R1 (2 µs)	4.4 (3.1)	1.8 (1.1)
	R2 (1.9 µs)	1.3 (2.5)	-0.1 (1.0)
	R3 (1.4 µs)	2.5 (2.5)	6.7 (1.1)
N234-Man9	R1 (2 µs)	3.4 (2.0)	-0.5 (0.7)
	R2 (1.4 µs)	-0.7 (2.0)	4.8 (0.7)
N234-Man9 (+ N370)	R1 (1.4 µs)	-3.4 (2.4)	6.0 (0.7)
	R2 (1.3 µs)	-0.6 (2.1)	-1.0 (0.7)
All-Man5	R1 (1.4 µs)	-14.1 (2.6)	3.2 (0.9)
	R2 (1.1 µs)	-3.5 (3.6)	6.9 (1.0)
	R3 (1.9 µs)	-1.9 (6.4)	-0.2 (3.8)
N234-Man9* (Man5 N165/N343)	R1 (2.1 µs)	0.2 (2.5)	7.3 (1.0)

N234-Man3

In order to gauge the implications for the orientation and dynamics of the open RBD of the size of the oligomannose at N243, we studied a model were N243 is modified with a paucimannose (Man3). Note, to our knowledge, the presence of Man3 at this or at any sequon in the SARS-CoV-2 S has not been detected to date[8, 10-15]; nevertheless, as Man3 is one of the smallest oligomannose structures that could potentially occupy this position, this S glycoform represents a case study to account for the effect of the reduction in the size of the N-glycan to an extreme at this strategic position. The results shown in **Table 1** and **Figure 3** in addition to the visual analysis of all trajectories, show that Man3 at N234 is less competent than Man 5 and Man9 in supporting a fully wide-open RBD. In only one of the three replicas, namely R3 shown in **Figure 3 panels b** and **d**, the Man3 accesses the interior of the open pocket, although its reduced size does not allow it to form any functional interactions that would contribute to stability. Meanwhile, in both R1 and R2, Man3 interacts with residues outside or at the edge of the pocket, see **Figure S.5**.

In the N234-Man3 model, the open RBD does not show any significant lateral excursions, see **Figure 3 panel a)**, as its position is firmly held in place by interactions with the complex N-glycans at N165 and N343, which contribute to pulling it towards a conformation intermediate between the fully open (seen in the simulations of the N234-Man5 and N234-Man9 models) and the closed (as shown in **Figure S.1** and **S.5**, were the N165 and N343 N-glycans interact extensively with residues in the receptor binding motif (RBM)). As an interesting point, when the FA2G2 N-glycan at N343 interacts with the RBD, its core fucose becomes less accessible to a potential recognition, see **Figure 3 panel b**.

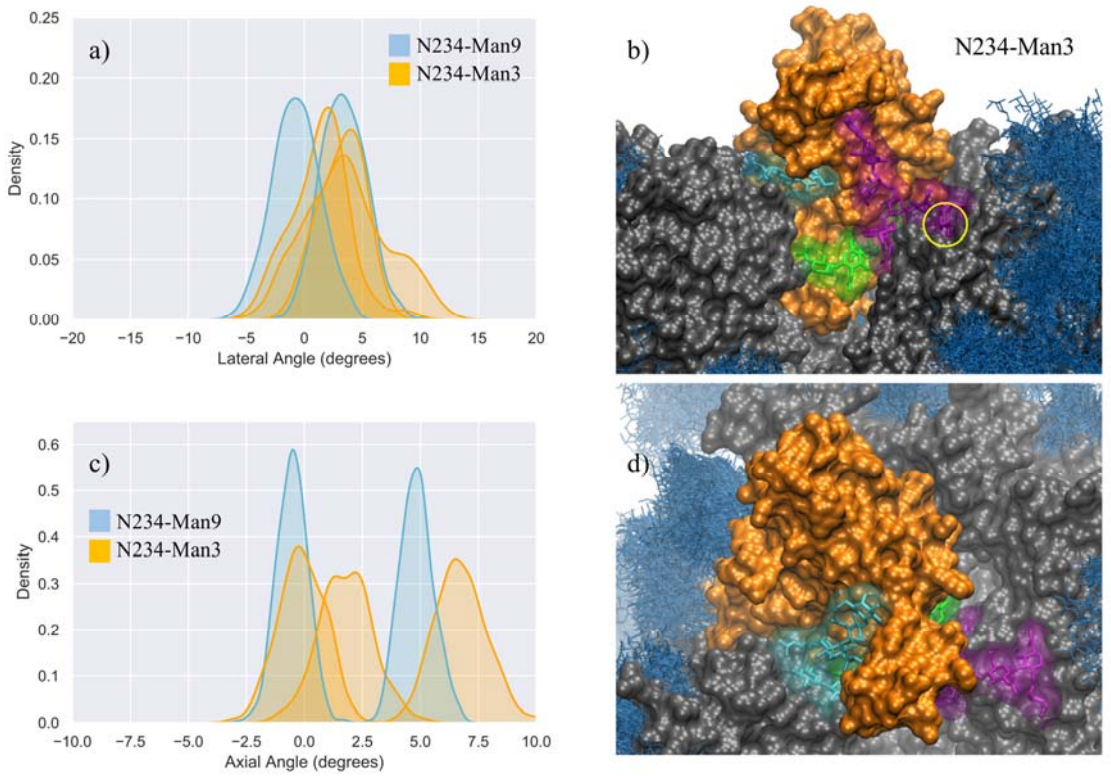

Figure 3. Panels a) and **c)** Kernel density estimation (KDE) analysis of the lateral and axial angle distributions calculated through the uncorrelated MD trajectories obtained for N234-Man3 (orange), replicas R1-3, and for N234-Man9 (cyan), replicas R1,2 for comparison. **Panels b)** and **d)** Close-ups on a representative snapshot of the N234-Man3 (R3) simulation from the side and top, respectively. Man3 at N234 is shown in green, while the N-glycans at N165 and N343 are shown in cyan and purple, respectively. The solvent accessible surface of the open RBD is shown in orange, while the rest of the protein is shown in grey. All other glycans are shown in blue, as an overlay of snapshots collected every 10 frames. The position of the core fucose in the FA2G2 N-glycans at N343 is highlighted within a yellow circle. Data analysis and graphs were done with *seaborn* (www.seaborn.pydata.org) and molecular rendering with VMD[25].

N234-Man9 with FA2G2 at N370

Analysis of the glycan shield topology through the ancestral sequence reconstruction of select SARS S sequences, see **Figures S.6** and **S.7**, indicates a recent loss of a sequon at N370 in

SARS-CoV-2, which is due to a mutation to NSA of the conserved NST present in closely related strains of current interest, namely SARS-CoV, the bat RaTG13G S and the pangolin CoV S. To study the effect on SARS-CoV-2 S structure and dynamics of this additional ancestral glycan at N370, we restored the sequon at N370 and glycosylated this position with a complex FA2G2 N-glycan, as reported in the literature for SARS-CoV[39]. In this model, position N234 is modified with Man9, while positions N343 and N165 are glycosylated with complex N-glycans with a bisecting GlcNAc and core fucosylation at N343 (FA2B), and without core-fucosylation at N165 (A2B). Our results, shown in **Figure 4** and **Table 1**, indicate that the wide-open RBD in a N370-glycosylated SARS-CoV-2 S is as stable as the one observed for the corresponding N234-Man9, where the N-glycan at N370 fills the interior of the empty pocket together with the Man9. In this model, the Man9 at N234 is only able to access the entrance of the pocket due to steric hindrance with the N370 glycan that occupies the core. This suggest that once the RBD opens, an S glycoform with glycosylated N370 would be equally competent in exposing the RBD as the SARS-CoV-2 S.

The analysis of the dynamics of the closed RBDs of protomers A and C (PDB 6VYB numbering) shows that the N370 N-glycan of RBD (A) is tightly bound to the surface of the adjacent closed RBD (C), threading the two closed RBDs together, see **Figure 4 panel b** and **Figure S.8**. The 3D architecture of the complex FA2G2 N-glycan at N370, characterized by independent dynamics of the arms[40], allows for stable interactions of the (1-6) arm within a cleft in the RBD, flanked by residues between N448 and Y453 on one side and F490 and Y495 on the other, see **Figure S.8**, that support binding through hydrogen bonding and hydrophobic interactions with different monosaccharide units along the (1-6) arm. Because of the stability of the N370 glycan-protein interaction in the closed RBD, N370 glycosylation may hinder the opening mechanism. Thus, the loss of glycosylation at N370 likely contributes to enhancing the binding activity of S and infectivity of SARS-CoV-2 relative to other variants with an N-glycosylation sequon at this position.

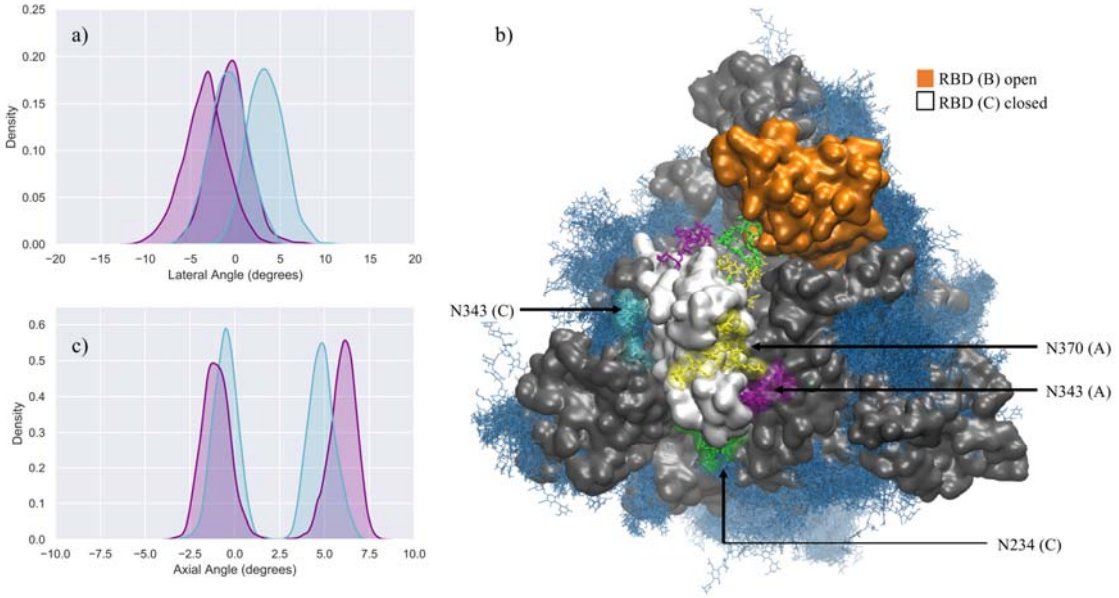

Figure 4. Panels a) and **c)** Kernel density estimation (KDE) analysis of the lateral and axial angle distributions calculated through the uncorrelated MD trajectories obtained for N370-glycosylated N234-Man9 (purple), replicas R1,2, and for N234-Man9 (cyan), replicas R1,2 for comparison. **Panel b)** Close-ups on a representative snapshot of the N370-glycosylated N234-Man3 (R2) simulation from top. Man9 at N234 (B,C) is shown in green, while the complex N-glycans at N165 (B,C), N343 (C,A) and N370 (C,A) are shown in cyan, purple and yellow, respectively. The surface of the open RBD (B) is shown in orange and of the closed RBD (C) is shown in white; the rest of the protein is shown in grey. All other glycans are shown in blue, as an overlay of snapshots collected every 10 frames. Data analysis and graphs were done with *seaborn* (www.seaborn.pydata.org) and molecular rendering with VMD[25]

Discussion

The SARS-CoV-2 S glycosylation profile with expression in (or infection of) mammalian cells has been reported by several studies[8, 10-15], almost all of which find a large oligomannose N-glycan, such as Man7-Man9, as the most common structure at N234. This is especially true in highly stable prefusion SARS-CoV-2 S trimer glycoforms[10, 12], which bear the 2P mutation[6, 41]. Meanwhile, a shorter Man5 appears to be present or even the dominant structure at N234 in the virus[8, 12] and in the secreted ChAdOx1 nCoV-19 (AZD1222) vaccine epitope[12], suggesting a higher degree of accessibility to that site by alpha-mannosidases in the ER. The results of multi-microsecond simulations presented in this work indicate that a reduced degree of filling of the cleft left empty by the opening of the RBD by smaller N-glycan at N243 leads to reduced filling of the cleft left empty by the opening of the RBD, and progressively higher degree of instability of the wide-open conformation of the RBD. More specifically, while the N234-Man5 model appears to be as competent as the N234-Man9 model in exposing an open RBD, despite its higher dynamics, the N234-Man3 model leads to a dominant “more closed” conformation, see **Figure S.1**, where the paucimannose cannot

form stable interactions within the pocket, interacting only with residues at the pocket's gate or below it, see **Figures 3** and **S.5**. The progressive destabilization upon reduction of the N-glycan size at N234 is in agreement with the results obtained for the mutant N234A[17], designed to account for the complete removal of glycosylation at N234, which leads to 60% less binding to the ACE2 receptor.

In agreement with recent work that discusses the roles of the glycan at N343 in supporting the RBD intermediate dynamics between open and closed conformations[18], and of the glycan at N165 in supporting the open RBD[17], we observed that the stability of the N234-Man5 is significantly reduced when all the N-glycans in the shield are Man5 (All-Man5). This destabilization is due primarily to the lack of interactions that the shorter Man5 at N343 can make with the open RBD, and in particular with residues in the disordered loop within the RBM (400 to 508), in addition to the Man5 at N234 allowing the RBD to be relatively unhinged. We have seen that this destabilization is partially recovered when the N234 is occupied by a larger oligomannose, such as Man9. However, within this framework, the RBD adopts a more closed conformation, where the Man5 at both N165 and N343 can interact with the open RBD and with each other, see **Figure S.4**.

The introduction of an additional glycosylation site at N370, which SARS-CoV-2 S has lost due to the recent T372A mutation, illustrates the importance of an effective filling of the cavity left empty by the opening of the RBD. Indeed, the complex FA2G2 glycan at N370 is able to enter the pocket, in addition to Man9 at N234, which in this case can only partially fill the pocket due to steric hindrance. We have shown that the presence of an N370 glycan contributes effectively to the stability of the wide-open RBD state, which nevertheless needs to be achieved starting from a closed S conformation[18]. To this end, we find that the N370 N-glycan on the SARS-CoV-2 S throughout all of our simulations occupies a specific cleft on the surface of the closed RBDs. This binding mode is stabilized by a network of hydrogen bonds and hydrophobic interactions between the protein and the monosaccharides of the FA2G2 N370 glycan (1-6) arm, see **Figure 4** and **S.7**. Because the N370 glycan involved in this interaction originates from the adjacent closed RBD, its binding results in tying the closed RBDs together, likely hindering opening in the first place. This is in agreement with recent work[42] suggesting that the introduction of a N370 sequon in SARS-CoV-2 S negatively affects binding to human ACE2, contributing to increased replication of SARS-CoV-2 S in human cells relative to its putative ancestral variant. Furthermore, it is also

interesting to note that cryo-EM S structures from the bat RatG13 and pangolin CoV variants, both carrying the N370 sequon, have been only solved in their closed states[22, 28, 29], possibly also suggesting opening is less favoured in these proteins relative to SARS-CoV-2 S.

Ultimately, the tight binding we observed between the N370 N-glycan and the surface of the SARS-CoV-2 S closed RBD surface is not only interesting in terms of the implications for higher ACE2-binding activity through loss of the corresponding sequon suggests, but also indicates the presence of a glycan binding site in that cleft, which is occupied in CoV variants that retain that sequon. In this context, recent work[30] provided evidence that heparan sulfate (HS) binds the SARS-CoV-2 S RBD in a ternary complex with ACE2, as an essential interaction for cell infection, meanwhile unfractionated heparin, non-anti-coagulant heparin, heparin lyases, and lung HS are found to potently block SARS-CoV-2 S binding to ACE2 and infection. Notably, the same study provides evidence that SARS-CoV S binding to heparin-BSA is significantly reduced, yet not completely negated[30]. Our work suggests that, in addition to the overall lower degree of positive charge of the SARS-CoV RBD relative to SARS-CoV-2, the same glycan-binding cleft on the RBD surface is available to be occupied by HS (or heparin) in SARS-CoV-2, while in SARS-CoV this cleft is occupied by the N370 glycan. Further investigation on this topic is underway. Recent work has also shown evidence that SARS-CoV-2 S can specifically bind oligosaccharides containing sialic acid[35], which could also occupy the same cleft on the RBD.

Analysis of reconstructed SARS S ancestral sequences indicates that while the N370 sequon was recently lost in SARS-CoV-2 S, this sequon was only quite recently acquired within the phylogeny. However, two proximal sequons at positions 381 (G381-VS) and 364 (D364-YS) have been gained and lost alternatively, **see Figures S.7** and **S.8**. It is reasonable to think that glycosylation within this S topological region may have been evolutionarily conserved, because of its functional role in stabilizing the open RBD, despite the higher energetic cost involved in the transition from the RBD down-to-up state. In SARS-CoV-2, glycosylation at N234, which according to our analysis also appeared relatively recently, would then functionally take the role of a glycan at N370. This evolutionary flexibility in the precise positions of glycosylation sites would make the presence of any specific glycan dispensable, with the consequent advantage of ensuring easier RBD opening reaction, and thus a more active S.

Conclusions

In this work we have used multi-microsecond MD simulations to determine the effect of changes in the nature and topology of the SARS-CoV-2 S N-glycosylation at sites known to be involved in its function. Our results indicate that reducing the size of the N-glycans at N234 led to the instability of the "wide-open" RBD conformation, with a consequent increase in RBD dynamics and a progressive stabilization of conformations favouring the closed protomer. Additionally, the structure of the N-glycans at N165 and N343 also affects the stability of the open RBD with shorter structures unable to effectively interact with the RBD disordered loop within the RBM. This effect is especially dramatic when a shorter N-glycan, such as Man5, is also present at N234. To account for changes in the glycan shield topology, we explored the effect of re-introducing N-glycosylation to a recently lost sequon at N370. Our results indicate that while the N-glycan at N370 is highly effective in stabilizing the open RBD in conjunction with the N-glycan at N234, it tightly binds a specific cleft on the surface of the closed RBD, tying the closed protomers together and likely increasing the energetic cost of the RBD opening. Because the architecture of this RBD cleft is particularly able to bind multiple monosaccharides through a network of hydrogen bonds and dispersion interactions, we suggest that in SARS-CoV-2 it can be occupied by other diverse glycan structures, such as glycosaminoglycans, proteoglycans or sialylated species, which have also been shown to bind S. This flexibility in glycan-binding preference would provide an additional advantage in terms of increasing localization at the host cell surface. Finally, comparative analysis of reconstructed SARS ancestral sequences suggests that specific changes in the glycan shield topology at and around N370, in conjunction with the gain of N-glycosylation at N234, may have contributed to an increase in S activity, and thus of the infectivity of the SARS-CoV-2 Wuhan strain.

Acknowledgements

The authors would like to thank Dr Lorenzo Casalino in Prof Rommie Amaro's lab (UCSD) for sharing the *tcl* script for the calculation of the axial and lateral angle displacement of the RBD and Prof Amaro for insightful discussions. The Partnership for Advanced Computing in Europe (PRACE) is gratefully acknowledged for generous allocation of computational resources on Marconi100 hosted by CINECA, under the PRACE COVID-19 Fast Track scheme. CAF acknowledges the Irish Research council for funding under the Government of Ireland Postgraduate Scholarship scheme. AMH acknowledges Maynooth University (MU) for funding under the John and Pat Hume Postgraduate Scholarship scheme.

References

1. Watanabe, Y.; Bowden, T. A.; Wilson, I. A.; Crispin, M., Exploitation of glycosylation in enveloped virus pathobiology. *Biochim Biophys Acta Gen Subj* **2019,** *1863* (10), 1480-1497.
2. Li, F., Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* **2016,** *3* (1), 237-261.
3. Plemper, R. K., Cell entry of enveloped viruses. *Curr Opin Virol* **2011,** *1* (2), 92-100.
4. Tortorici, M. A.; Veesler, D., Structural insights into coronavirus entry. *Adv Virus Res* **2019,** *105*, 93-116.
5. Walls, A. C.; Tortorici, M. A.; Snijder, J.; Xiong, X.; Bosch, B. J.; Rey, F. A.; Veesler, D., Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proc Natl Acad Sci U S A* **2017,** *114* (42), 11157-11162.
6. Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C. L.; Abiona, O.; Graham, B. S.; McLellan, J. S., Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **2020,** *367* (6483), 1260-1263.
7. Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D., Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**.
8. Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.; Wu, Z.; Cheng, L.; Shi, D.; Lu, X.; Lei, J.; Crispin, M.; Shi, Y.; Li, L.; Li, S., Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020,** *183* (3), 730-738.e13.
9. Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W.; Welsch, S.; Blanc, F.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; Trevisan Doimo de Azevedo, N.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M.; Hummer, G.; Krijnse Locker, J.; Beck, M., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, eabd5223.
10. Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M., Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **2020**.
11. Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P., Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **2020**.
12. Brun, J.; Vasiljevic, S.; Gangadharan, B.; Hensen, M.; Chandran, A.; Hill, M.; Kiappes, J.; Dwek, R.; Alonzi, D.; Struwe, W.; Zitzmann, N., Assessing Antigen Structural Integrity of an Approved COVID Vaccine through Glycosylation Analysis of the SARS-CoV-2 Viral Spike *bioRxiv* **2020**.
13. Sanda, M.; Morrison, L.; Goldman, R., N- and O-Glycosylation of the SARS-CoV-2 Spike Protein. *Anal Chem* **2021,** *93* (4), 2003-2009.
14. Watanabe, Y.; Mendonça, L.; Allen, E. R.; Howe, A.; Lee, M.; Allen, J. D.; Chawla, H.; Pulido, D.; Donnellan, F.; Davies, H.; Ulaszewska, M.; Belij-Rammerstorfer, S.; Morris, S.; Krebs, A. S.; Dejnirattisai, W.; Mongkolsapaya, J.; Supasa, P.; Screaton, G. R.; Green, C. M.; Lambe, T.; Zhang, P.; Gilbert, S. C.; Crispin, M., Native-like SARS-CoV-2 spike glycoprotein expressed by ChAdOx1 nCoV-19/AZD1222 vaccine. *bioRxiv* **2021**.

15. Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L., Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host Microbe* **2020,** *28* (4), 586-601.e6.
16. Bagdonaite, I.; Thompson, A. J.; Wang, X.; Søgaard, M.; Fougeroux, C.; Frank, M.; Diedrich, J. K.; Yates, J. R.; Salanti, A.; Vakhrushev, S. Y.; Paulson, J. C.; Wandall, H. H., Site-specific O-glycosylation analysis of SARS-CoV-2 spike protein produced in insect and human cells. *bioRxiv* **2021**.
17. Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E., Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci* **2020,** *6* (10), 1722-1734.
18. Sztain, T.; Ahn, S. H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; McCool, R. S.; Kearns, F. L.; McCammon, J. A.; McLellan, J. S.; Chong, L. T.; Amaro, R. E., A glycan gate controls opening of the SARS-CoV-2 spike protein. *bioRxiv* **2021**.
19. Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F., Cell entry mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **2020,** *117* (21), 11727-11734.
20. Letko, M.; Marzi, A.; Munster, V., Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol* **2020,** *5* (4), 562-569.
21. Coutard, B.; Valle, C.; de Lamballerie, X.; Canard, B.; Seidah, N. G.; Decroly, E., The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res* **2020,** *176*, 104742.
22. Wrobel, A. G.; Benton, D. J.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J., SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* **2020,** *27* (8), 763-767.
23. Neelamegham, S.; Aoki-Kinoshita, K.; Bolton, E.; Frank, M.; Lisacek, F.; Lütteke, T.; O'Boyle, N.; Packer, N. H.; Stanley, P.; Toukach, P.; Varki, A.; Woods, R. J.; Group, S. D., Updates to the Symbol Nomenclature for Glycans guidelines. *Glycobiology* **2019,** *29* (9), 620-624.
24. Cheng, K.; Zhou, Y.; Neelamegham, S., DrawGlycan-SNFG: a robust tool to render glycans and glycopeptides with fragmentation information. *Glycobiology* **2017,** *27* (3), 200-205.
25. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996,** *14* (1), 33-8, 27-8.
26. Williams, S. J.; Goddard-Borger, E. D., α-glucosidase inhibitors as host-directed antiviral agents with potential for the treatment of COVID-19. *Biochem Soc Trans* **2020,** *48* (3), 1287-1295.
27. Watanabe, Y.; Berndsen, Z.; Raghwani, J.; Seabright, G.; Allen, J.; McLellan, J.; Wilson, I.; Bowden, T.; Ward, A.; Crispin, M., Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *bioRxiv* **2020**.
28. Wrobel, A. G.; Benton, D. J.; Xu, P.; Calder, L. J.; Borg, A.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J., Structure and binding properties of Pangolin-CoV spike glycoprotein inform the evolution of SARS-CoV-2. *Nat Commun* **2021,** *12* (1), 837.
29. Zhang, S.; Qiao, S.; Yu, J.; Zeng, J.; Shan, S.; Tian, L.; Lan, J.; Zhang, L.; Wang, X., Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat Commun* **2021,** *12* (1), 1607.

30. Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Painter, C. D.; Thacker, B. E.; Glass, C. A.; Narayanan, A.; Majowicz, S. A.; Zhang, Y.; Torres, J. L.; Golden, G. J.; Porell, R.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Leibel, S. L.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D., SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *bioRxiv* **2020**.
31. Liu, L.; Chopra, P.; Li, X.; Bouwman, K.; Tompkins, S.; Wolfert, M.; de Vries, R.; Boons, G., Heparan sulfate proteoglycans as attachment factor for SARS-CoV-2. *bioRxiv* **2020**.
32. Tandon, R.; Sharp, J. S.; Zhang, F.; Pomin, V. H.; Ashpole, N. M.; Mitra, D.; McCandless, M. G.; Jin, W.; Liu, H.; Sharma, P.; Linhardt, R. J., Effective Inhibition of SARS-CoV-2 Entry by Heparin and Enoxaparin Derivatives. *J Virol* **2021,** *95* (3).
33. Mycroft-West, C. J.; Su, D.; Pagani, I.; Rudd, T. R.; Elli, S.; Gandhi, N. S.; Guimond, S. E.; Miller, G. J.; Meneghetti, M. C. Z.; Nader, H. B.; Li, Y.; Nunes, Q. M.; Procter, P.; Mancini, N.; Clementi, M.; Bisio, A.; Forsyth, N. R.; Ferro, V.; Turnbull, J. E.; Guerrini, M.; Fernig, D. G.; Vicenzi, E.; Yates, E. A.; Lima, M. A.; Skidmore, M. A., Heparin Inhibits Cellular Invasion by SARS-CoV-2: Structural Dependence of the Interaction of the Spike S1 Receptor-Binding Domain with Heparin. *Thromb Haemost* **2020,** *120* (12), 1700-1715.
34. Baker, A. N.; Richards, S. J.; Guy, C. S.; Congdon, T. R.; Hasan, M.; Zwetsloot, A. J.; Gallo, A.; Lewandowski, J. R.; Stansfeld, P. J.; Straube, A.; Walker, M.; Chessa, S.; Pergolizzi, G.; Dedola, S.; Field, R. A.; Gibson, M. I., The SARS-COV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent Sci* **2020,** *6* (11), 2046-2052.
35. Nguyen, L.; McCord, K.; Bui, D.; Bouwman, K.; Kitova, E.; Kumawat, D.; Daskhan, G.; Tomris, I.; Han, L.; Chopra, P.; Yang, T.; Willows, S.; Mason, A.; Lowary, T.; West, L.; Hsu, S.; Tompkins, S.; Boons, G.; de Vries, R.; Macauley, M.; Klassen, J., Sialic acid-Dependent Binding and Viral Entry of SARS-CoV-2. *bioRxiv* **2021**.
36. Bouwman, K. M.; Tomris, I.; Turner, H. L.; van der Woude, R.; Shamorkina, T. M.; Bosman, G. P.; Rockx, B.; Herfst, S.; Snijder, J.; Haagmans, B. L.; Ward, A. B.; Boons, G. J.; de Vries, R. P., Multimerization- and glycosylation-dependent receptor binding of SARS-CoV-2 spike proteins. *PLoS Pathog* **2021,** *17* (2), e1009282.
37. Fogarty, C. A.; Fadda, E., Oligomannose N-Glycans 3D Architecture and Its Response to the FcγRIIIa Structural Landscape. *J Phys Chem B* **2021**.
38. Barnes, C. O.; Jette, C. A.; Abernathy, M. E.; Dam, K. A.; Esswein, S. R.; Gristick, H. B.; Malyutin, A. G.; Sharaf, N. G.; Huey-Tubman, K. E.; Lee, Y. E.; Robbiani, D. F.; Nussenzweig, M. C.; West, A. P.; Bjorkman, P. J., SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **2020,** *588* (7839), 682-687.
39. Watanabe, Y.; Berndsen, Z. T.; Raghwani, J.; Seabright, G. E.; Allen, J. D.; Pybus, O. G.; McLellan, J. S.; Wilson, I. A.; Bowden, T. A.; Ward, A. B.; Crispin, M., Vulnerabilities in coronavirus glycan shields despite extensive glycosylation. *Nat Commun* **2020,** *11* (1), 2688.
40. Harbison, A. M.; Brosnan, L. P.; Fenlon, K.; Fadda, E., Sequence-to-structure dependence of isolated IgG Fc complex biantennary N-glycans: a molecular dynamics study. *Glycobiology* **2019,** *29* (1), 94-103.
41. Pallesen, J.; Wang, N.; Corbett, K. S.; Wrapp, D.; Kirchdoerfer, R. N.; Turner, H. L.; Cottrell, C. A.; Becker, M. M.; Wang, L.; Shi, W.; Kong, W. P.; Andres, E. L.; Kettenbach, A. N.; Denison, M. R.; Chappell, J. D.; Graham, B. S.; Ward, A. B.;

McLellan, J. S., Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A* **2017,** *114* (35), E7348-E7357.
42. Kang, L.; He, G.; Sharp, A.; Wang, X.; Brown, A.; Michalak, P.; Weger-Lucarelli, J., A selective sweep in the Spike gene has driven SARS-CoV-2 human adaptation. *bioRxiv* **2021**.