COMPUTATIONAL
ANDSTRUCTURAL
BIOTECHNOLOGY
J O U R N A L

Evidence of a putative glycosaminoglycan binding site on the glycosylated SARS-CoV-2 spike protein N-terminal domain

Zachariah P. Schuurs [a], Edward Hammond [b], Stefano Elli [c], Timothy R. Rudd [d], Courtney J. Mycroft-West [e], Marcelo A. Lima [e], Mark A. Skidmore [e], Richard Karlsson [f], Yen-Hsi Chen [f], Ieva Bagdonaite [f], Zhang Yang [f], Yassir A. Ahmed [g], Derek J. Richard [h], Jeremy Turnbull [f,g], Vito Ferro [i,j], Deirdre R. Coombe [j], Neha S. Gandhi [a,*]

[a] QUT, Centre for Genomics and Personalised Health, Cancer and Ageing Research Program, School of Chemistry and Physics, Faculty of Science and Engineering, Institute of Health and Biomedical Innovation, 2 George Street, Brisbane, QLD 4000, Australia
[b] Zucero Therapeutics Ltd, 1 Westlink Court, Brisbane, Queensland, Australia
[c] Istituto di Ricerche Chimiche e Biochimiche "G.Ronzoni", via Giuseppe Colombo 81, 20133 Milano, Italy
[d] National Institute for Biological Standards and Control, Analytical and Biological Sciences Division, Blanche Lane, South Mimms, Potters Bar, Hertfordshire EN6 3QG, UK
[e] Molecular & Structural Biosciences, School of Life Sciences, Keele University, Newcastle-Under-Lyme, Staffordshire ST5 5BG, UK
[f] Copenhagen Center for Glycomics, Department of Cellular & Molecular Medicine, University of Copenhagen, Copenhagen N 2200, Denmark
[g] Department of Biochemistry and Systems Biology, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, Liverpool L69 7ZB, UK
[h] QUT, Centre for Genomics and Personalised Health, Cancer & Ageing Research Program, Institute of Health and Biomedical Innovation at the Translational Research Institute (TRI), 37 Kent Street, Woolloongabba, Queensland 4102, Australia
[i] School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia
[j] Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD 4072, Australia

A R T I C L E   I N F O

A B S T R A C T

SARS-CoV-2 has rapidly spread throughout the world's population since its initial discovery in 2019. The virus infects cells via a glycosylated spike protein located on its surface. The protein primarily binds to the angiotensin-converting enzyme-2 (ACE2) receptor, using glycosaminoglycans (GAGs) as co-receptors. Here, we performed bioinformatics and molecular dynamics simulations of the spike protein to investigate the existence of additional GAG binding sites on the receptor-binding domain (RBD), separate from previously reported heparin-binding sites. A putative GAG binding site in the N-terminal domain (NTD) of the protein was identified, encompassing residues 245–246. We hypothesized that GAGs of a sufficient length might bridge the gap between this site and the PRRARS furin cleavage site, including the mutation S247R. Docking studies using GlycoTorch Vina and subsequent MD simulations of the spike trimer in the presence of dodecasaccharides of the GAGs heparin and heparan sulfate supported this possibility. The heparan sulfate chain bridged the gap, binding the furin cleavage site and S247R. In contrast, the heparin chain bound the furin cleavage site and surrounding glycosylation structures, but not S247R. These findings identify a site in the spike protein that favors heparan sulfate binding that may be particularly pertinent for a better understanding of the recent UK and South African strains. This will also assist in future targeted therapy programs that could include repurposing clinical heparan sulfate mimetics.

1. Introduction

The rapid spread of SARS coronavirus 2 (SARS-CoV-2) since its appearance in late 2019 has elicited a swift response from the scientific community to develop treatments and vaccines against this virus. Effective and efficient drug discovery requires under-standing the molecular events of the virus infection pathway and knowledge of the effects of the virus on host immunity. SARS-CoV-2 is an enveloped positive-sense RNA virus, one of several coronaviruses (*Coronaviridae*) that cause respiratory infections in humans. Before the emergence of SARS-CoV-2, there were two highly pathogenic coronaviruses, SARS-CoV and MERS-CoV, which caused severe respiratory disease in humans, and four other human coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKU1) which induced mild upper respiratory disease. SARS-CoV-2 is closely related to SARS-CoV and MERS-CoV, and, like

* Corresponding author.
  E-mail address: neha.gandhi@qut.edu.au (N.S. Gandhi).

these two viruses, it can cause very severe disease in some patients. Although the mutation rate of SARS-CoV-2 is considered moderate compared with other RNA viruses, numerous variants have been recorded; some of these have mutations in the spike (S) glycoprotein of the virus outer surface [1], the glycoprotein that is involved in virus infection of cells. The role of the SARS-CoV-2 S glycoprotein in virus infection makes it a key target for the development of antiviral drugs and vaccines. Accordingly, understanding how particular mutations in this glycoprotein may impact virus infection is a key part of the drug discovery process.

The SARS-CoV-2 S glycoprotein forms homotrimers on the virus surface where it is involved in the multistep receptor-mediated pathway of virus-host cell adhesion and virus-host cell membrane fusion, which culminates in cell infection. The S glycoproteins of other *Coronaviridae* family members also perform this role, but not all coronaviruses that are human pathogens recognize the same cell surface receptor. The cell surface receptor for SARS-CoV-2 is the human angiotensin-converting enzyme 2 (ACE2). Each S glycoprotein monomer consists of two main functional domains – S1 (residues 14–685) and S2 (residues 686–1273) (Fig. 1). It is well established that the S1 subunit mediates virus attachment to epithelial and other cell surfaces by binding to its receptor ACE2, while the S2 subunit mediates the fusion of the viral and human cell membranes [2,3]. Within S1, a region spanning residues 333–527 constitutes the receptor-binding domain (RBD) [4]. Cryo-EM studies revealed that the S glycoprotein trimer exists in several different conformational states. A significant fraction of the trimers are in a state with one of the three RBDs in an "up" or "open" conformation, whereas other trimers had the RBDs "down" or "closed" [2,5,6]. This has the effect of either masking (closed conformation) or exposing (open conformation) the ACE2

recognition interface on the RBD. The factors driving the conformational change are unclear, although a recent study suggested that interactions of the S glycoprotein with the glycosaminoglycans (GAGs), heparan sulfate (HS), or the structurally related heparin (HP), could be a contributing factor [7]. Specifically, it was found that the S glycoprotein could bind HP and ACE2 simultaneously. Moreover, HP oligomers enhance the binding of S glycoprotein to ACE2 [7]. Modelling studies led to the proposal that the site where HP and HS binds on the RBD is partially obscured in the closed conformation but completely exposed in the open state. These findings suggested that HP binding might increase the proportion of trimers in an open conformation, thereby assisting ACE2 binding [7]. Data from the same study indicated that HS side chains of cell surface proteoglycans (HSPGs) are necessary co-factors for infection by SARS-CoV-2, as removing cell surface HS with heparin lyases dramatically reduced S glycoprotein binding to cell surfaces, and SARS-CoV-2 infection [7]. Furthermore, another study found that whilst ACE2 is the primary receptor, the S glycoprotein can interact with cell surfaces in the absence of ACE2, suggesting that the initial interaction is independent of ACE2 [8].

Upon successful attachment to the cell surface, the S glycoprotein is cleaved by serine endoproteases at the S1/S2 site and at the S2′ site (Fig. 1). These events cause the dissociation of the S1 domain and a significant conformational change within S2, which brings the host and viral membranes together leading to fusion of the membrane bilayers and viral entry into the cell. This is a mechanism common to coronaviruses [9]. For SARS-CoV-2, the enzymes involved are believed to be the proprotein convertase, furin, and transmembrane serine protease 2 (TMPRSS2) [10]. Inhibitors of furin cleavage and TMPRSS2 cleavage have been shown to block virus entry and suppress virus production [11,12].
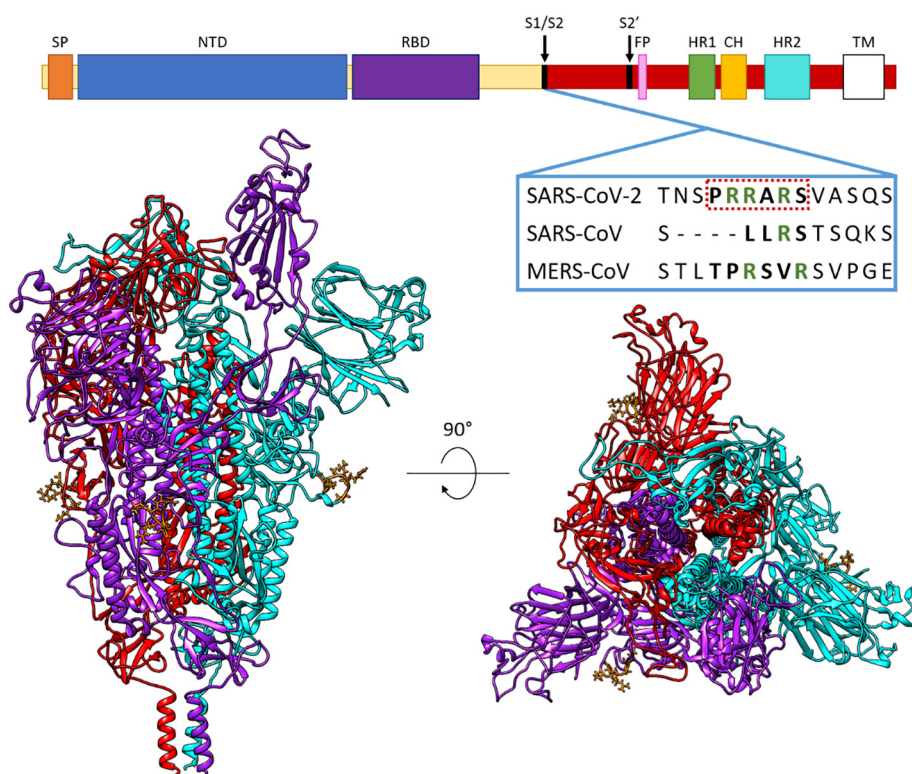

Fig. 1. The main domains of the SARS-CoV-2 spike protein, highlighting the unique polybasic furin cleavage site that doubly functions as a GAG-binding motif, absent in the aligned SARS-CoV and MERS-CoV. Annotated domains include SP, signal peptide; NTD, N-terminal domain; RBD, receptor binding domain; S1/S2, cleavage site between S1 and S2 domains; FP, fusion peptide; HR1, heptad repeat 1; CH, central helix; HR2, heptad repeat 2; TM, transmembrane domain. The spike trimer structure was built from PDB: 6VSB and 6LZG. The up conformation of RBD is in purple; the down conformations in cyan and red; and the PRRARS domain is in orange. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Heparan sulfate is a ubiquitous component of both the glycocalyx on cell surfaces and the extracellular matrix (ECM) that surrounds and supports cells in tissues. The ECM and a cell's glycocalyx must be traversed before viruses, and infectious organisms can engage their cell-surface receptors to mediate their entry into cells. In addition to SARS-CoV-2, the related coronaviruses SARS-CoV and HCoV-NL63 also use HS as attachment factors to facilitate binding to their receptor, ACE2, and infection of cells [13–15]. Glycosaminoglycans are a family of anionic carbohydrates of which HS and HP are members. Both these GAGs are linear polysaccharides of repeating disaccharides consisting of an *N*-acetylated or *N*-sulfated glucosamine alternating with a uronic acid that is either a D-glucuronic acid or an L-iduronic acid [16,17]. Each monosaccharide in the repeating region of HP and HS may also possess varying degrees of sulfation due to further modifications during biosynthesis. Generally, epimerization of glucuronic acid to iduronic acid is relatively low in HS, whereas for HP the reverse is true. In HP, most disaccharides carry near complete modifications (except of the rare 3-*O*-sulfate) with iduronic acid residues sulfated at carbon-2 (2S), alternating with *N*-sulfated glucosamine that is *O*-sulfated at carbon-6 (6S). These tri-sulfated disaccharides account for ~70–80% of HP. In contrast, approximately ~ 50% of HS chains are composed of unsulfated D-glucuronic acid β $(1 \rightarrow 4)$ *N*-acetylated D-glucosamine. In HS, there are also highly sulfated regions containing tri-sulfated disaccharides, such as L-iduronic acid (2S) $\alpha(1 \rightarrow 4)$ D-*N*-sulfoglucosamine (6S), amongst other sulfated variants. The rare structure of 3-*O*-sulfated, *N*-sulfated, and 6-*O*-sulfated glucosamine, can occur in the highly sulfated regions of both HS and HP [17]. The unsulfated and highly sulfated regions of HS are typically flanked by transition regions which consist of partially modified disaccharides separated by completely unmodified *N*-acetylated domains. This domain structure is a dominant feature of HS chains that is largely lacking in heparin chains. It is an important factor of HS-protein interactions due to the flexibility of the non-sulfated regions. Together with variations in their underlying disaccharide structure, HS and HP chains also vary in length and collectively these variations in HS and HP chains give rise to structures that are extremely heterogeneous. The SARS-CoV-2 spike protein displays a higher binding affinity to certain HS or HP sulfation patterns [18]. Chain-length and 6-*O*-sulfation have the greatest effect on affinity [18,19].

The S protein is heavily glycosylated, with the *N*-glycans attached to it accounting for up to half of the protein's molecular weight. It is likely that glycosylation shields the amino acid residues and epitopes of the S protein from recognition by cells and antibodies [2]. This probably impairs the host's ability to raise an adaptive immune response targeting the S protein, as generally glycans are poor immunogens. Glycosylation may also increase the infectivity of the virus. To date, most computational simulations of the spike trimer in the presence of ligands/co-receptors have ignored its glycosylation despite readily available information on the glycosylation pattern [19–26]. Understanding the glycosylation pattern is crucial when developing inhibitors of a protein, as the glycosylation chains can sterically hinder the binding of candidate ligands [24]. Ideally, therapies should target accessible surface regions on the coronavirus that are highly conserved and unlikely to mutate [27]. Hence, in this study we carried out MD simulations that considered the glycosylation.

Most studies looking to develop treatments like antibodies [28–30], repurposed drugs [31], GAGs [18,32,33], or fatty acid-like molecules [34], have been targeting RBD sites within the S glycoprotein. However, it has been determined that mutations distal from the RBD influence the transmissibility of SARS-CoV-2. One mutation that has attracted interest is D614G [35]. This mutation appears to increase viral infectivity by favoring the open RBD conformational state, and it is now a major virus variant globally [36–38]. Therefore, it is astute to study new mutations that are appearing, such as mutations to basic amino acids. One such mutation – S247R – was documented in the first patient diagnosed with COVID-19 in Australia (GenBank: QHR84449.1). Basic residues have a greater affinity for GAG molecules, as the negatively charged sulfates and carboxylates of the GAG chains can interact with the positive charges of appropriately positioned basic amino acids [39]. The S1/S2 boundary of SARS-CoV-2 glycoprotein contains a unique polybasic furin cleavage site PRRARS (681–686). This site does not appear in SARS-CoV-1 or MERS [2] (Fig. 1) and is one of several suggested GAG-binding motifs in the S glycoprotein from analyses of the amino acid sequence. Another non-RBD HP binding sequence is located at 810–816 (SKPSKRS) [19], but this is not exposed for binding to GAGs on monomers in either the up or down conformations. The sequence YRLFRKS is in the RBD domain and binds short oligosaccharides. To date, most predicted HP binding sites are located in the RBD. These include amino acid sequences 345–348; 354–360; 400–411; 416–426; 443–447; 453–459; 461–468; 507–513; 517–522 and 681–686. These sites have been published by Kim et al. [19], Mycroft-West et al. [33] and Paiardi et al. [40].The PRRARS site is exposed, and it appears to contribute to the S glycoprotein-ACE2 interaction in an auxiliary role through electrostatic interactions and its effects on hydration [21]. Importantly, cleavage at this site by furin or related proprotein convertases is essential for activating the S glycoprotein for its role in the fusion of virus and cell membranes, thereby contributing to virus production and syncytium formation [41].

Many studies have focused on the structure and function of SARS-CoV-2-neutralizing antibodies targeting the association of ACE2 and the S glycoprotein RBD domain [5,30,42–46]. The S glycoprotein can rapidly accumulate escape mutations in an experimental system using a replicating VSV-SARS-CoV-2-S virus under selection with a single antibody targeting the RBD. To avoid loss of antibody binding affinity and drug efficacy because of mutations in the RBD, alternatives such as using antibody cocktails as therapies are being investigated [47,48]. Of particular interest are some antibodies that have been experimentally shown to bind to non-RBD sites of the S glycoprotein. The antibody 4A8 demonstrates potent neutralizing activity. It binds to the NTD residues (R246, Y145-K147, N149, K150, W152) that restrain conformational changes (PDB: 7C2L) [49]. [49]The antibody 2G12 binds to a region that includes the furin cleavage site but does not neutralize the spike protein (PDB: 7L06) [50].

Understanding the dynamic structure of the S protein provides insight into the molecular mechanisms of the protein and its recognition by the host immune system, and it may reveal potential therapeutic intervention points. Recent studies have shown heparin binding to the RBD domain [33,51] and furin cleavage sites [19]; however, binding sites appropriate for GAG oligosaccharides longer than an octasaccharide remain elusive. Here, using structural bioinformatics methods, docking, and molecular dynamics simulations (MD), we identified HS binding sites on the NTD of the S protein. We then studied, in the presence of glycosylation, the interaction of HS and HP molecules with the S247R mutant protein, a variant of the S protein that first emerged in Melbourne, Australia (GenBank: QHR84449.1) [52]. We propose longer GAG molecules are able to bridge the gap between the PRRARS furin cleavage site and the 245H-246R site with the S247R mutation improving this binding. The emergence of specific mutations in the S protein of UK B.1.1.7 [53] and South African B.1.351 [54] variants has caused these strains to be more virulent. How these mutant residues contribute to virulence is not yet clear. It is possible that GAG binding to the NTD site we identified may be influenced by some of these mutations and so contribute to the virulence of these new strains. Targeting the new binding site with antibodies or structurally tailored GAG dodecasaccharides (or

longer GAGs) capable of bridging the gap between the furin site and residues 245H-246R may prevent proteolytic cleavage at the S1/S2 domain boundary of the S protein, thereby preventing infection of host cells.

2. Material and methods

2.1. Trimer model building

A complete trimer model was built to investigate GAG oligosaccharide binding sites. To do this, the cryo-EM structure of the trimeric S protein in its prefusion conformation (PDB: 6VSB [6]) was used as the base model. A homology model was built using the SWISS-MODEL webserver [55] to account for missing loops in the PRRARS furin cleavage site and RBD domains in the down conformation. The RBD domain from the RBD-ACE2 crystal structure (PDB: 6LZG [56]) was isolated and used to replace the RBD subunit in the up conformation from the homology model. The combined model was merged using Modeller in UCSF Chimera v. 1.13.1 [57]. While there are now multiple trimer structures available, this was the only trimer model structure available at the time of this work [57].

To run simulations as close to biological conditions as possible, glycosylations and disulfide bonds were added to the model using the CHARMM-GUI web-server Glycan Reader & Modeler input generator [58–60]. N- and C-termini were treated as being neutral, and disulfide bonds were added between cysteine residues 131–166, 291–301, 336–361, 379–432, 391–525, 480–488, 538–590, 617–649, 662–671, 738–760, 743–749, 840–851, 1032–1043 and 1082–1126. Glycosylations were added according to Table S1, adapted from the site-specific model by Grant et al. [25].

2.2. Modelling of spike trimer-heparin tetrasaccharides in cosolvent MD simulations

Cosolvent molecular dynamics (MD) simulations were carried out on the trimeric model described above to search for new oligosaccharide binding locations. During a cosolvent simulation, the protein local environment and bulk are sampled to characterize potential binding sites on the protein surface. In our experiment, ten HP tetrasaccharides with the sequence [IdoA2S-α-(1 $\rightarrow$ 4)-GlcNS6S-α-(1 $\rightarrow$ 4)]$_2$ were used as the cosolvent. These were randomly placed around the trimer model. The CHARMM-GUI input generator [59,60] was used to place the protein at the center of a triclinic simulation box, with 15 Å edge space. The box was solvated with TIP3P water [61] and 0.15 M NaCl placed with the Monte-Carlo method to neutralize the system charge. Long-range electrostatics were treated using the particle-mesh Ewald (PME) method. Periodic boundary conditions (PBC) were applied throughout the simulations. A non-bonded cut-off of 12 Å was used and the non-bonded neighbor list was updated at every time step. The SHAKE algorithm [62] was used to constrain all bonds involving hydrogen atoms. Minimization, equilibration and production simulations were performed with GROMACS v. 2020.1 [63,64] using the CHARMM36 forcefield [65] on V100 GPU of the NCI Gadi supercomputer. The system was first minimized for 5,000 steps followed by equilibration for 10 ns with 2 fs steps. The final production simulation was run for 300 ns, saving a snapshot each 10 ps. Temperature coupling was at 303.15 K and used the Nose-Hoover extended ensemble. Pressure coupling used the Parrinello-Rahman extended ensemble at 1 bar. The final snapshot was analyzed using VMD v. 1.9.3 [66] to determine the locations on the trimer where HP molecules bound. To confirm the results, a second simulation on the same system was conducted for 100 ns starting with different velocities, and a third saturated system with

20 HP tetrasaccharides was built and run for 100 ns with the same settings as described above.

2.3. HS and HP ligand docking and MD simulation to S glycoprotein monomer

It is understood that HP favors binding to basic residues. This rationalized a search for mutations involving basic residues that have appeared within global SARS-CoV-2 spike amino acid sequences. The sequences published on the COVID-19 Viral Genome Analysis Pipeline were searched for the site mutations. This database has been assembled using data from the Global Initiative on Sharing Avian Influenza Data (GISAID) database [67].

The initial cosolvent simulations formed the rationale behind the docking and subsequent experiments to test the binding of longer HS to the PRRARS site and S247R. To achieve this, the monomer in the up conformation was isolated from the model. We used USCF Chimera v. 1.13.1 [57] to mutate Ser-247 to Arg-247 and built two systems with the HP and HS dodecasaccharides. GlycoTorch Vina [68] was used to perform docking of oligosaccharides. The dodecasaccharide HP ([IdoA2S-α-(1 $\rightarrow$ 4)-GlcNS6S-α-(1 $\rightarrow$ 4)]$_6$) and HS ([IdoA2S-α-(1 $\rightarrow$ 4)-GlcNS6S-α-(1 $\rightarrow$ 4)]$_2$-[IdoA-α-(1 $\rightarrow$ 4)-GlcNAc-α-(1 $\rightarrow$ 4)]$_2$-[IdoA2S-α-(1 $\rightarrow$ 4)-GlcNS6S-α-(1 $\rightarrow$ 4)]$_2$ were built using the Glycam GAG builder [69], with the IdoA2S residues considered in the $^1C_4$ conformation. The online GlycoTorch tool was used to convert the PDB input files to PDBQT. This tool accounts for ϕ and ψ glycosidic torsions, and contains parameters to model $^2S_0$ and $^1C_4$ for iduronic/glucuronic acids [68]. All sulfate and hydroxyl groups, and glycosidic torsion angles were treated as flexible. The box size was 60, 53, 80 centered between the furin cleavage domain and S247R. For docking we used an energy range of 12, an exhaustiveness of 12, chi_cutoff = 1, chi_coeff = 2 and set the number of modes to 100.

The two systems of the S247R S protein monomer with HP and HS dodecasaccharides were prepared for MD simulations using the CHARMM glycan input generator according to the same settings outlined in 2.2. The MD simulations were performed with GROMACS for 100 ns using the methodology as described in section 2.2 above. Due to the computational resources available, only a single MD run was conducted.

2.4. ClusPro Docking

To confirm the MD simulation results, ClusPro [70] was used to dock heparin to an unglycosylated monomer subunit of the S protein in the "up" conformation and with the S247R mutation. This was to confirm the binding sites observed in the MD simulation described in section 2.3. The resulting models were superimposed, visualized and atom contacts determined using ChimeraX version 1.1 [71] (Fig. S2).

2.5. Statistical analysis of heparin-binding motifs

The putative GAG binding sequence within the SARS-CoV-2 NTD were analyzed by methods described in Rudd et al. [72]. A brief description of the method follows. In summary, 776 heparin binding proteins, were fragmented into sequences of a minimium of 3 residues with at least one basic amino acid in the following combinations; BXA, BXS, BXP, BXAS, BXAP and BXPS (where, B = basic, X = hydrophobic, A = acidic, P = polar and S = special). A comparison of these sequences was made with a metric using the Levenshtein distance. This is a measure of similarity between character strings based on the minimum number of insertions, deletions or substitutions that a string needs to undergo to alter one amino acid sequence to match the other. Sequences with a similarity score of greater than 0.7 (70% similarity) were considered highly conserved. Using this same Levenshtein similarity cut-off

of 0.7, sets of highly conserved basic amino acid containing sequences extracted from heparin binding proteins were used to support the validity of the proposed binding site in the SARS-CoV-2 S247R mutant S protein NTD.

2.6. Trajectory, MM/PBSA analysis and electrostatic potential surface

Tools built into GROMACS were used to process the trajectories, including re-centering, fitting, periodicity treatments and concatenation before analyzing them with an array of tools. MDAnalysis [73], VMD 1.9.3 [66] and USCF Chimera [57] were used to visualize and analyze the output trajectories. The final 50 ns of the HS and HP trajectories was used to calculate per-residue energy decompositions with the g_mmpbsa [74] module compatible with GRO-MACS distribution. A frame was extracted every 250 ps from each system and a total of 200 frames were used for the MM/PBSA calculations. The system enthalpy is calculated using the molecular mechanics method (MM). The polar part of the solvent effect is obtained by solving the finite difference Poisson-Boltzmann (PB) equation, and the non-polar part is fitted by estimating the surface area that is solvent-accessible (SA). This is demonstrated in Equation (1):

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (1)$$

The binding free energy is ΔG_{bind}, and the intramolecular energy under vacuum is represented by ΔE_{MM}. The solvation free energy difference is ΔG_{sol}, which is the sum of polar and non-polar solvation free energies. T is the absolute temperature and ΔS is the change in entropy of the system. We did not carry an out entropy calculations owing to the size of the systems. CHARMM radii were used and an ionic strength of 0.15 M was set for the salt. The outer dielectric constant was set to 80, and solute dielectric constant set to 2. Further details of the MM/PBSA methods have been published by Genheden and Ryde [75].

The electrostatic potential surface was visualized by preparing files with PDB2PQR webserver [76]. The CHARMM forcefield, with an internal dielectric constant of 2 and a constant of 80 for water was used to map the potential on the DelPhi web server [77,78], using an internal dielectric constant of 2. ChimeraX version 1.1 [79] was used to visualize the electric potential surface.

2.7. Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was conducted on 1 μg S1 in PBS (pH 7.6) with 1.25 X SYPROTM Orange (Invitrogen) in the presence of 100 μg of either unfractionated porcine mucosal heparin (UFH; Celsus) or size defined heparin oligosaccharides
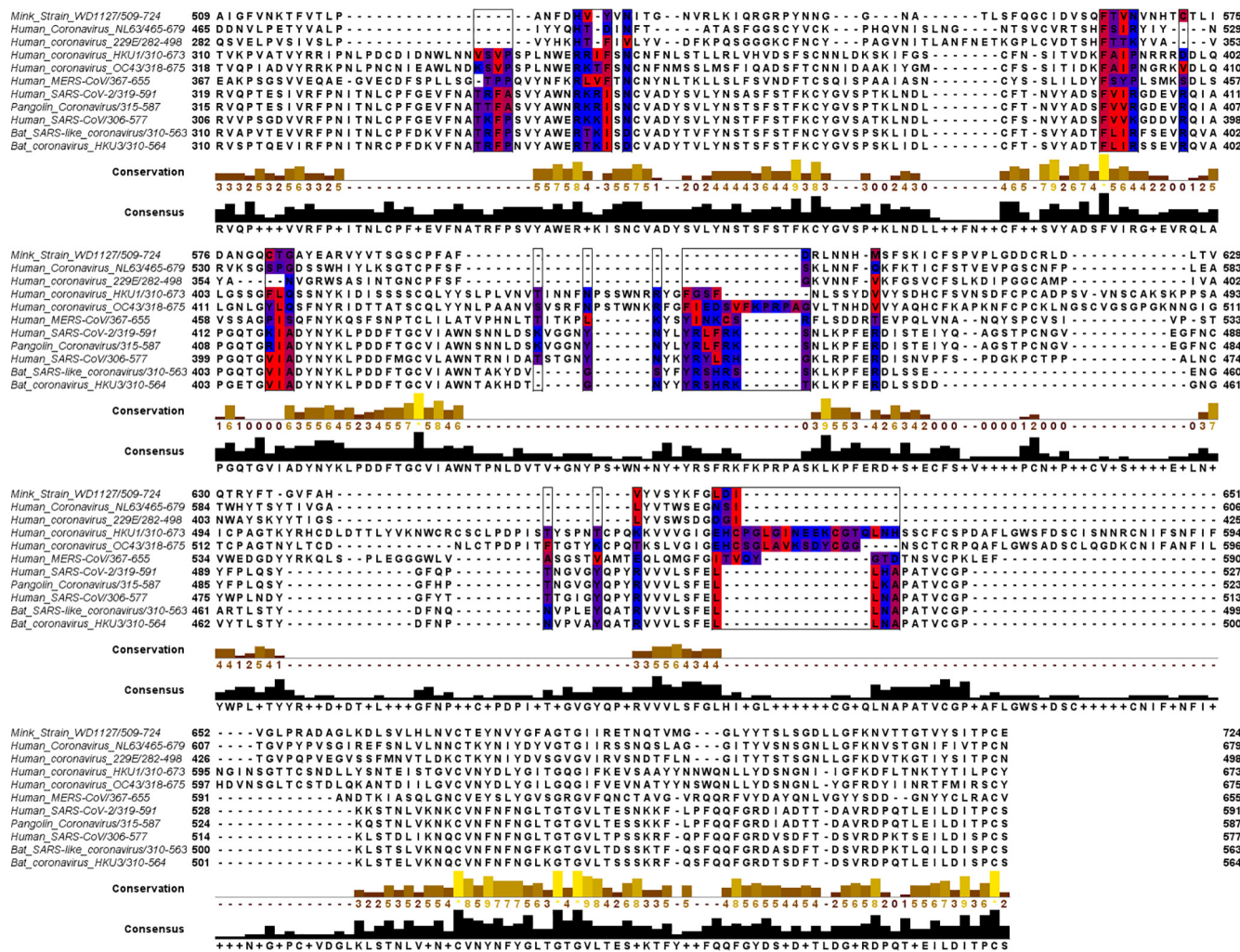

Fig. 2. Multiple sequence alignment using Clustal Omega of the RBD of related virus strains. Heparin binding sites identified in previous papers [19,33] are highlighted and colored according to hydrophobicity in Jalview [101]. Bat, mink, and pangolin strains were compared to the human coronaviruses, as these species are theorized to be the zoonotic hosts of the viruses [80,81].

(dp 6, 12 or 16), with H_2O used as a control. Reaction mixtures were made up to a volume of 40 μL, in 96-well qPCR plates (AB Biosystems) before being subjected to DSF using an AB Biosystems, StepOnePlus, qPCR machine with the TAMRA filter employed. Melt curve experiments were performed by increasing the temperature from 25° C to 90° C by 0.5° C increments every 30 s. Following, smoothing (Savitxky-Golay, 2nd-order polynomial, 9 neighbors) the first differential of the melt curves was calculated (Prism 8; GraphPad) and the peaks used to determine the melt temperature (T_m) of S1 alone or in the presence of UFH or size defined oligosaccharides (MatLab software; R20018a, MathWorks).

3. Results and discussion

It has been reported that the interaction of HS with SARS-CoV-2 is a requirement for the virus to infect cells [7]. The key binding sites on the S protein reported in this earlier study were all located in the RBD domain near the ACE2 binding site. Studies investigating where HP binds to the spike protein have predicted multiple binding sites on the RBD [19,33,40]. The RBD protein sequence of several coronavirus strains theorized to be zoonotically related to SARS-CoV-2 [80,81] were aligned to allow the amino acid sequences of these predicted HP binding sites to be compared. This alignment (Fig. 2) revealed relative conservation of the HP binding sites between viral strains examined. To further understand HS interactions with the SARS-CoV-2 S trimer, we first conducted an unbiased molecular dynamics (MD) simulation of a glycosylated SARS-CoV-2 S trimer with ten HP tetrasaccharides randomly placed around the protein. Such cosolvent MD simulations have literature precedence to detect novel binding sites [82]. Other published computational studies (using CHARMM [83], or Glycam [25] force fields) looked at glycosylation patterns that included N- and O-linked glycoforms [83] and glycoforms in the Golgi prior to enzymatic modification [25], but for the purposes of this study we considered the model used in MD simulations by Grant et al. [25]. In our study, seven of the ten HP tetrasaccharides bound to the trimer during the initial MD simulation (Fig. 3). Regions where the HP tetrasaccharides bound to the protein contained the positively charged residues arginine (R246, 249, 577, 634, 682, 683, 685), lysine (K147, 150, 444), and a protonated histidine (H245), as well as the polar uncharged serine (S71), and the small non-polar residues glycine (G72) and alanine (A684). Two further short cosolvent simulations (100 ns) produced the same results, providing good evidence for the binding of HP tetrasaccharides to R246 and the PRRARS domain (Fig. S1). These regions in the S1 NTD give access to the tetrasaccharides to the protein surface between the glycosylation sites.

A common method to provide support to data from MD simulations of GAGs binding to a protein is to dock a generic HP tetramer using the online server ClusPro [70]. We docked heparin to the unglycosylated S247R monomer in the "up" conformation. HP molecules bound to R246-S247R and the furin cleavage site (Fig. S2). This aligns with the results we observed in the MD simulations of the trimer and monomer.

Interestingly, a few of the amino acid residues found to interact with HS in the MD simulations are insertions in the regions 72–82, 144–147 and 244–246, which are shared by SARS-CoV-2 and the bat coronavirus RatG13, but not SARS-CoV. Moreover, these regions are known to play a role in host receptor binding [85–86]. Hence, we predict they play a significant role in controlling the binding of SARS-CoV-2 to cells, including the process affecting the conformation of the RBD binding site.

With mutations to the spike appearing as the pandemic spreads across the globe, we felt it was pertinent to investigate some of these changes because newly arising basic mutations may promote GAG binding. Of particular interest was the S247R mutation, which is physically located close to the PRRARS furin cleavage site and close to some of the deletions in the UK mutant strain (B.1.1.7) (Table S5). The S247R mutation is adjacent to two charged amino acids (H245-R246) that have been targeted by neutralizing antibodies [49]. We hypothesized that the S247R mutation would augment the basic nature of this region containing H245-R246 and favor bridging of GAG molecules between this region and the furin cleavage site. The GAG docking program GlycoTorch Vina [68] was used to test this hypothesis, with HS and HP of different lengths being examined. This determined that GAGs 12 saccharides long, i.e. dodecasaccharides, would bridge the gap in a biased docking (Fig. 4B). An analysis of the electrostatic surface potential (Fig. 4A) revealed that the region HS binds is positively charged. As predicted, the positive region near the H245-R246 is made more basic by the S247R mutation, and this facilitated HS binding.

Using GROMACS [63,64], we ran MD simulations with the HS and HP dodecasaccharides to conduct an in depth analysis of the binding to the mutated site, S247R, and the PRRARS furin cleavage site. Fig. 4B (HS) and C (HP) show a clear differentiation between the binding modes of the two GAG molecules. Only HS bridges the gap between the furin site and S247R, while HP sits in the pocket between the PRRARS loop and the loop containing R634. The less sulfated HS spans several domains, while HP has limited contact with the S protein as indicated by the distance between residue centers of mass (Fig. S3) according to calculations using MDAnalysis [73,87]. This analysis also highlighted the residues that appeared to play key roles in binding the oligosaccharides. The distances between the centers of mass are indicative of electrostatic interactions, such as hydrogen bonding, between the atoms involved. The data in Fig. S3 supports previously published data [18] that variation in the sulfation levels of molecules has a
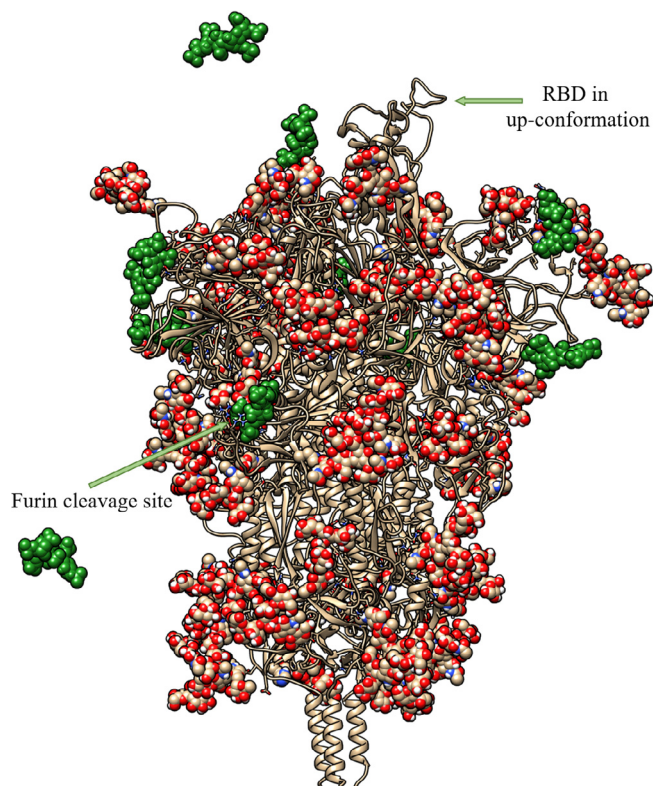

Fig. 3. A representative snapshot showing the binding of heparin tetrasaccharides to the glycosylated SARS-CoV-2 S protein, obtained from the unbiased MD simulations. The tan ribbons are the representation of the protein; the glycosylations are the spheres coloured.

RBD in up-conformation

Furin cleavage site

significant impact on binding to the S protein. Pulmonary HS is quite heterogeneous [88], with functional studies suggesting it regulates signaling proteins involved in lung development, homeostasis and injury [89]. Given our data, it would be interesting to model the interactions of specific HS structures known to be present in the lung with the S protein to examine whether these structures bind in the way suggested in our model, and so link the furin cleavage site with the region around S247R or H245-R246.

Analysis of the hydrogen bonds formed between the S protein and the two ligands (HS and HP) revealed that glycans on the S protein are involved in binding the HP molecule (Fig. 5). Both HS and HP bind to the PRRARS furin site and interact with R634 (Tables S2

and S3). They otherwise differently bind to the S protein: HS bridges the gap between the PRRARS domain and S247R and does not have any strong interactions with the glycan shield, whereas the N-glycans near the furin cleavage site exert a shielding effect on the short HP dodecasaccharide examined here. Longer HP chains ignore this effect, forming a bridge between RBD heparin binding residues (T345, R346, N354, R355, N360) and the furin cleavage site [40]. This supports the use of UFH and longer heparin chains over LMWH for preventing the binding of SAR-CoV-2 to the cell surface. The higher sulfation density of the HP molecule shifts the preferred binding of this molecule to the PRRARS furin cleavage site and surrounding N-glycans. Compare this to the sulfates of the
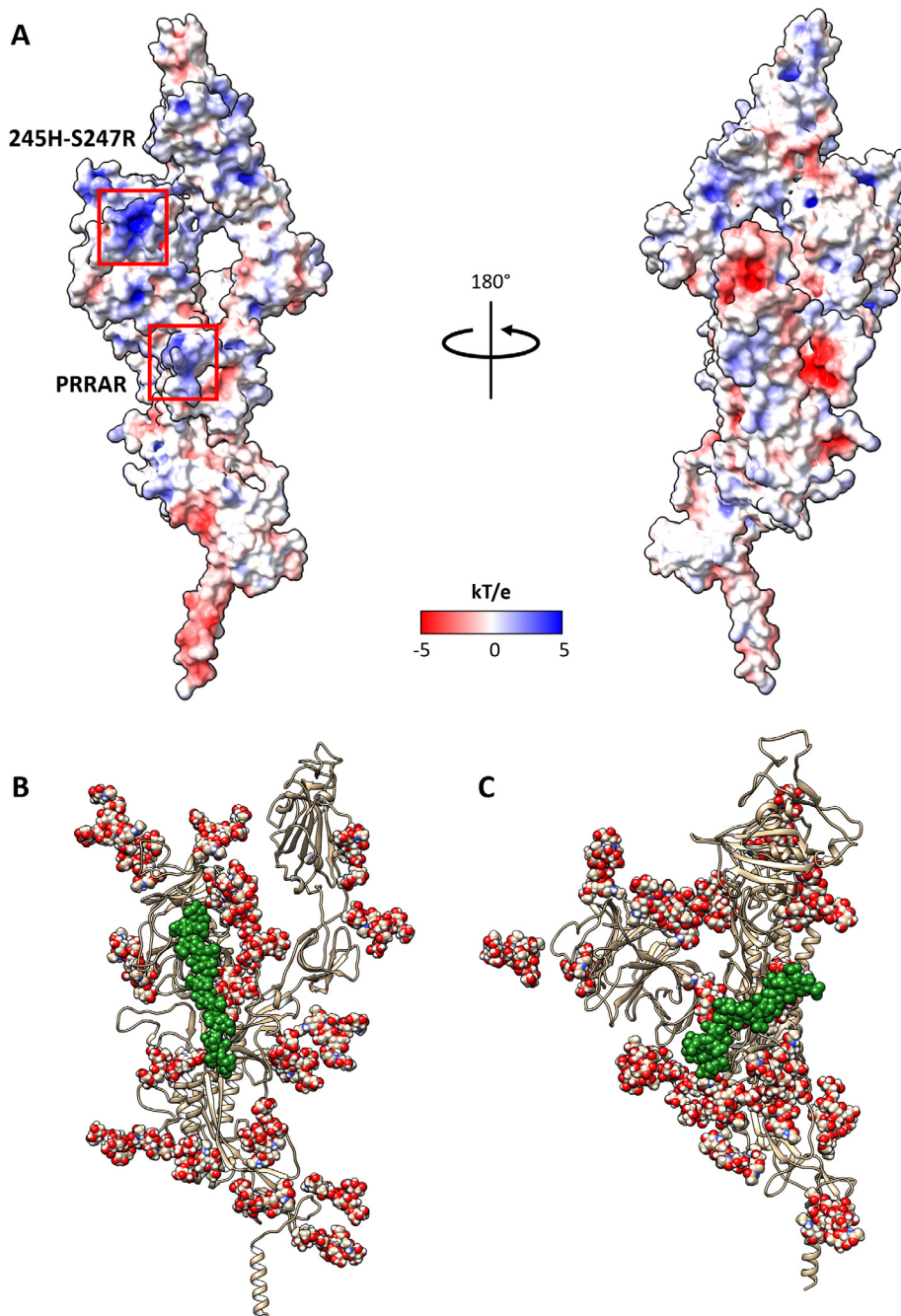
Fig. 4. (A) Electrostatic potential surface representation of the spike monomer, calculated using the DelPhi web-server [77,78]. The left hand side presents the face containing the PRRARS furin cleavage site and 245H-S247R site, while the right hand side is the opposite face. Representative snapshots obtained from MD frames at 100 ns of the (B) HS and the (C) HP dodecasaccharides bound to the S protein in the up conformation. The spheres represent glycosylations and the GAG oligosaccharides as green spheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

HS molecule, which were primarily located at either end of the chain, giving rise to a binding mode involving both the furin site and binding to the distal 245H-S247R site. Post-MM/PBSA analysis of the complexes support this (Fig. 6A). In Fig. 6A, HS shows a strong affinity for residues H66 and H69; interestingly H69 is deleted in the recent UK strain (B.1.1.7). Those two histidines are unique in the decomposition as only HS had an effect on them. The residues H66, H69, K77, R78, R246, R247, R634, R646, R682, R683 and R685 were calculated to demonstrate favorable free energy contributions to HS binding, whilst the eight arginines in this list and K77 were found to have favorable free energy for binding to HP (Fig. 6). We believe these differences are a result of a higher flexibility in the non-sulfated region of the HS molecule, compared to the more fully sulfated HP. It is this flexibility that

allows the bridging of the two regions (furin cleavage site and the 245H-S247R region), in a manner similar to what was observed in the studies by Perkins [90] and Rashid [91].

A graph showing the decomposition of glycosylations interacting with the ligands is presented in Fig. S4. The figure indicates that the glycosylations on the S protein interact with the HP molecule more than the HS. Specifically, *D*-mannose and *N*-acetylgalactosamine (Table S3) glycan subunits attached to N61 and N616. In our simulation, the glycosylations attached to these residues formed hydrogen bonds with the HP dodecasaccharide. Previous experimental molecular dynamics simulations have evaluated the effects of N-linked glycosylation on serpins and the consequences of heparin binding to both proteases function and dynamics [92–94]. Similarly, our work alludes to a role that these
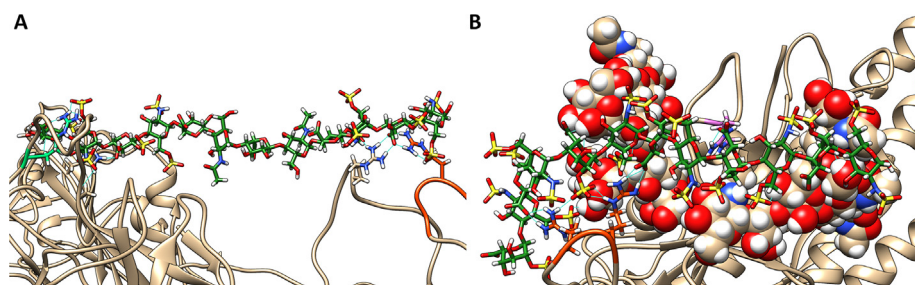

Fig. 5. Frames from 100 ns of MD simulations. Hydrogen bonds (solid cyan lines) between residues of the S protein or glycan shield (spheres) and the GAG molecules. The PRRARS site is in orange, and the 247H-S247R is in green. (A) shows the surface binding of the HS molecule to the S247R mutation and residue R683 of the PRRARS domain. Later frames show further binding of PRRARS domain residues. Other residues binding the HS molecule include H69, S71, A262 and R634. (B) shows the hydrogen bonds between the HP molecule and the S protein or glycan shield. Here the residues R683 and R685 of the PRRARS domain form hydrogen bonds. Other residues include R634 and saccharides from the glycan shield. Full information on the hydrogen bonds including residue, atom and distance information are in Table S2 (HS) and S3 (HP). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
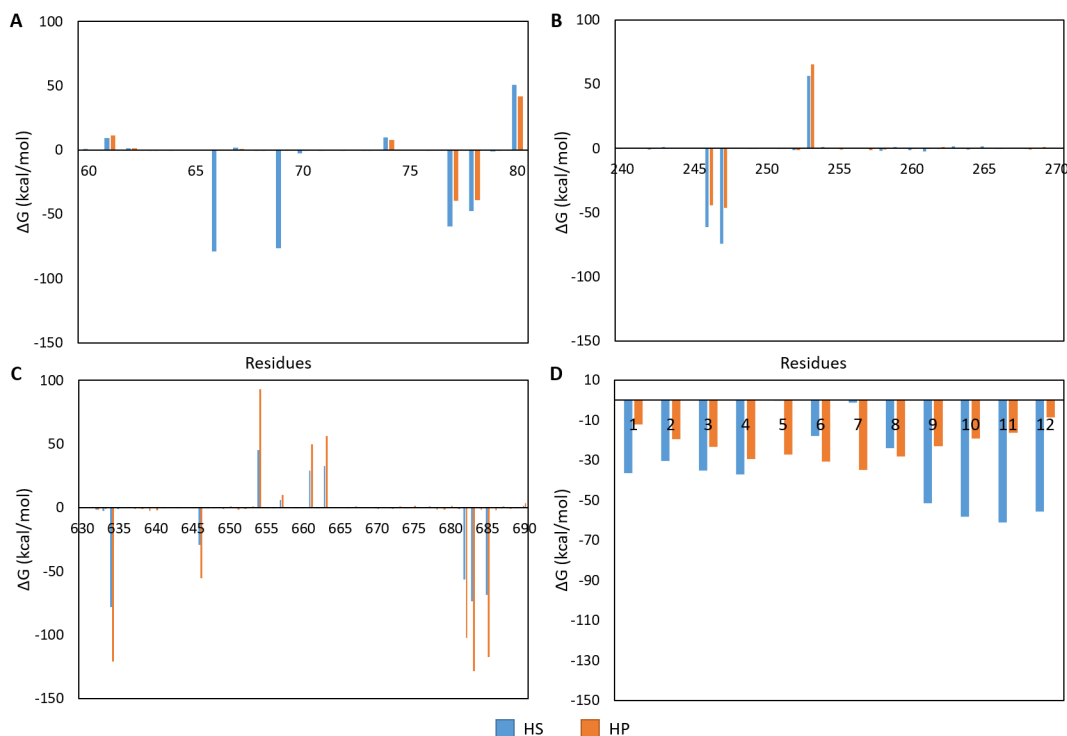
Fig. 6. MM/PBSA-binding free energy per residue decomposition [ΔG kcal/mol] for the protein–ligand interactions. The x-axis of each is numbered according to the residue in the protein or GAG. The residues displayed were selected as they were the ones with the largest affinity for the HP/HS, and were therefore of interest. A) the decomposition of the S glycoprotein residues 60–80B) decomposition of the S glycoprotein residues 240–270; C) decomposition of the S glycoprotein residues 630–690; D) decomposition of the HP and HS ligands.

glycosylations might play in contributing to GAG specificity. This alludes to a role that these glycosylations might play in contributing to GAG specificity.

The differences in the binding of HS and HP highlighted in this present study is in accordance with the results of other studies that indicate more favorable binding to certain lengths and sulfation levels of GAGs [18]. Post-MM/PBSA analysis of the HP- and the HS-S glycoprotein complexes revealed the relative free binding energy and standard deviation of that average to be −461.91 ± 3.97 kcal/mol for the HP complex and −788.39 ± 3.06 kcal/mol for the HS complex.

A key aspect in our approach to identifying novel binding sites on the S glycoprotein is the consideration of the site-specific glycan shield in the MD simulations. This revealed significant interactions between the docked GAG molecules and the glycan side chains, which would not be apparent if these studies had been performed in the absence of S protein glycosylation. In light of this, we were able to identify a gap in the glycan shield at 245H-S247R that may imply an increased viral infectivity of the S247R mutant. At the same time, this new virus strain is likely to bind HS more strongly than the wild type strain. Therefore, its sequestration to the cell surface by proteoglycans like syndecans could be vulnerable to inhibition of by free HS chains or HS fragments. A search of the COVID-19 Viral Genome Analysis Pipeline database revealed other sites of mutation to basic residues. Table S4 shows the number of times such mutations have been recorded. It may be pertinent to investigate these mutations as they could promote the ability of SARS-CoV-2 to bind to cell surfaces, thus increasing virulence. These basic mutations could also act as alternative targets for the development of HP-based therapies. The UK strain B.1.1.7 that is spreading rapidly throughout the UK has a deletion in the NTD at residues 69–70. This is an escape mutation that is believed to have arisen as a result of a treatment regime – in this case, convalescent plasma [95,96]. Our data identified H69 and S71 to be HS binding residues (Table S1). We have included a commentary of the amino acid residue mutations and their likely roles in HS/HP binding that have occurred in the recent British (B.1.17) and South African (B.1.351) mutant strains of SARS-CoV-2 (Table S5). HS octasaccharides optimally bind to the RBD [97], but it is not yet known where longer HS oligosaccharides bind. As a result of our analyses, we believe they bind to the NTD, whereas the RBD has a higher affinity for shorter oligosaccharides closer to dp8.

The thermal stability of full-length trimeric S1 was investigated using differential scanning fluorimetry (DSF), a technique which utilizes a conventional real time PCR machine to determine the mid-point of unfolding (or melt temperature, T_m) of a given protein, through the use of a hydrophobic fluorescent probe [98]. When heated, protein unfolding exposes buried hydrophobic regions where the fluorescent dye can bind. This results in an increase in fluorescence that can be used to produce a melt curve. The T_m can then be obtained from the peak of the first differential. In this way, protein unfolding can be monitored. The T_m of full-length trimeric S1 using this method was determined to be 44° C, approximately 2° C lower than previously reported for the glycosylated RBD of S1 [99].

Changes in a protein's T_m value in the presence of a ligand is also indicative of binding affinity, therefore, the ΔT_m of S1 with size defined heparin oligosaccharides was determined to investigate the size dependency of binding. Oligosaccharides consisting of 16 saccharide units when mixed with S1 caused a statistically significant 1° C reduction in the T_m of S1. Shorter oligosaccharides consisting of 12 or 6 units did not exhibit a statistically significant change in the T_m of S1 (Fig. 7), suggesting that heparin chains longer than dodecamers may be required for optimal binding to S1. It is noteworthy that UFH did not show a statistically significant difference in T_m, thus potentially indicating that there is a limit to the optimal length of HP that S1 can bind or that the sulfation patterns present in the UFH were not optimal for binding.

A comparison of full-length S protein amino acid sequences was made with an extensive library of amino acid sequences from known HP binding proteins. The degree of similarity observed suggests regions on the S protein that are likely to bind to GAGs. The putative binding site at 245H-S247R did not appear with a Levenshtein cut-off of 0.7, but did at 0.65 (Table 1). This is because the sequence 241–246 (LLALHR) extended by the S247R mutation (LLALHRR) is quite rare and not conserved in the library of heparin-binding proteins used for the analysis. The analysis also identified that the 69–71 region is likely to be involved in HP binding.

4. Conclusions

Overall, our work shows the importance of taking into account the glycan shield when conducting MD simulations of proteins, as it can act to prevent binding to certain regions of the protein and directly interact with some docked ligands. Such interactions may be missed when conducting studies without glycosylation being included. By considering the glycan shield, the prediction of important sites involved in molecular interactions and possibly immune recognition is improved. Indeed, our study identified a putative GAG binding site at residues 241–246 of the SARS-CoV-2 S glycoprotein. This is particularly pertinent to the South African strain (B.1.351) mutations of this virus, with its deletion at L242-L244. Finally, this work has confirmed the preferential affinity for particular GAG sulfation patterns and lengths based on bridging between different oligosaccharide binding sites that may act cooperatively as anchoring points to prevent or support conformational changes in the protein. The repurposing of drugs is becoming a common approach to developing novel therapies. This work, among others [7], suggests the potential use against SARS-CoV-2 of HS mimetics such as Sanofi's SR123781, which reached phase II clinical trials (NCT00123565, NCT00338897), or PI-88 which reached phase III trials as an anticancer agent (NCT00268593) and has an antiviral effect against Dengue virus and flavivirus encephalitis [100]. Future *in vitro* studies on the affect that HS moieties have on cell infection by SARS-CoV-2 would be interesting. To conclude, this work has revealed a putative multi-contact binding mechanism of HS to the SARS-CoV-2 spike protein. This highlights alternative ways that HP and HS mimetics could contribute to treatment of COVID-19, other than preventing coagulation and micro-thrombi formation.
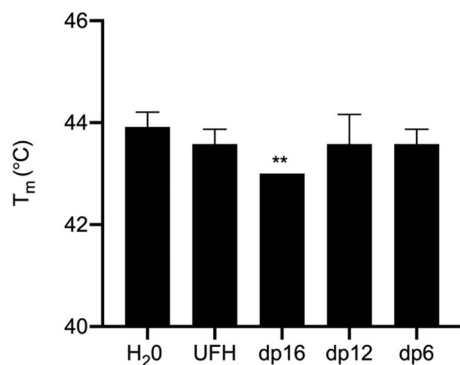

Fig. 7. T_m of S1 with size defined oligosaccharides or heparin; H_2O was used as a control for no addition. ** Significant difference between the T_m value of S1 H_2O control (44 °C ± 0.3; n = 3) compared to RBD plus dp16 (43 °C ± 0; n = 3), t(4) = 5.5, p ≤ 0.01 0. ΔT_m = 1 °C. Where no error bars are shown SD = 0.

Table 1

Sequence of SARS-CoV-2 S glycoprotein residues with a Levenshtein Distance above 0.7. The normalized count for similar sequences as found in the library of 776 heparin binding proteins. The higher the normalized count, the more often a sequence was identified among the protein set. Basic amino acids are identified; arginine (blue), lysine (red) and histidine (green). AA No; amino acid number. AA; amino acid identity. HBP; relative frequency of sequence among heparin binding proteins. The region highlighted in blue shows residues 241–247 count with a cut-off of 0.65.

AA No.	34	35	36	37	38	39	40	41	42	43	44	…	47	48	49	…	64	65	66	67	68	69	70	…	76	77	78
AA	R	G	V	Y	Y	P	D	K	V	F	R		V	L	H		W	F	H	A	I	H	V		T	K	L
HBP	0.005	0.005	0.005	0.005	0.005	0.005	0.032	0.632	0.632	0.632	0.632		0.353	0.353	0.353		0.005	0.005	0.005	0.005	0.005	0.005	0.005		0.184	0.758	0.758

AA No.	79	80	…	100	101	102	…	126	127	128	129	130	131	…	143	144	145	146	147	…	156	157	158	159	160	…	189
AA	F	D		I	I	R		V	V	I	K	V	C		V	Y	Y	H	K		E	F	R	V	Y		L
HBP	0.758	0.332		0.274	0.274	0.274		0.068	0.068	0.068	0.068	0.068	0.005		0.011	0.011	0.011	0.011	0.011		0.005	0.332	0.332	0.332	0.332		0.005

AA No.	190	191	192	193	194	195	…	200	201	202	203	204	…	210	211	212	213	214	215	216	…	241	242	243	244	245	246
AA	R	E	F	V	F	K		Y	F	K	I	Y		I	N	L	V	R	D	L		L	L	A	L	H	R
HBP	0.005	0.005	0.347	0.347	0.347	0.347		0.005	0.005	0.005	0.005	0.005		0.037	0.037	0.600	0.600	0.600	0.074	0.074		0	0	0	0	0	0

AA No.	247	248	249	…	275	276	277	278	279	…	296	297	298	299	300	…	302	303	304	305	306	307	308	309	310	311	312
AA	R	Y	L		F	L	L	K	Y		L	S	E	T	K		T	L	K	S	F	T	V	E	K	G	I
HBP	0	0	0		0.042	0.042	0.042	0.042	0.042		0.005	0.005	0.005	0.005	0.005		0.005	0.005	0.005	0.005	0.005	0.005	0.321	0.316	0.553	0.237	0.237

AA No.	313	…	318	319	320	…	323	324	325	326	327	328	329	330	…	346	347	348	…	355	356	357	358	385	386	387	388
AA	Y		F	R	V		T	E	S	I	V	R	F	P		R	F	A		R	K	R	I	T	K	L	N
HBP	0.237		0.226	0.226	0.226		0.005	0.005	0.032	0.458	0.458	0.458	0.458	0.047		0.263	0.263	0.263		0.658	0.658	0.658	0.658	0.600	0.200	0.200	0.200

AA No.	…	399	400	401	402	403	404	405	406	407	408	409	410	411	…	416	417	418	419	420	421	422	423	424	425	426	…
AA		S	F	V	I	R	G	D	E	V	R	Q	I	A		G	K	I	A	D	Y	N	Y	K	L	P	
HBP		0.021	0.532	0.532	0.532	0.532	0.026	0.342	0.342	0.358	0.358	0.016	0.016	0.016		0.442	0.784	0.784	0.784	0.021	0.037	0.016	0.642	0.642	0.642	0.332	

AA No.	443	444	445	446	447	…	461	462	463	464	465	466	467	468	…	507	508	509	510	511	512	513	…	515	516	517	518
AA	S	K	V	G	G		L	K	P	F	E	R	D	I		P	Y	R	V	V	V	L		F	E	L	L
HBP	0.174	0.405	0.405	0.232	0.232		0.468	0.468	0.468	0.479	0.011	0.011	0.011	0.011		0.026	0.068	0.068	0.068	0.068	0.068	0.068		0.005	0.005	0.468	0.468

AA No.	519	520	521	522	…	533	534	535	…	537	538	539	…	557	558	559	560	…	565	566	567	568	569	570	571	…	574
AA	H	A	P	A		L	V	K		K	C	V		K	K	F	L		F	G	R	D	I	A	D		D
HBP	0.468	0.468	0.016	0.016		0.758	0.758	0.758		0.111	0.111	0.111		0.732	0.732	0.732	0.732		0.168	0.168	0.084	0.021	0.021	0.021	0.021		0.021

AA No.	575	576	577	578	…	622	623	624	625	626	627	…	633	634	635	636	…	653	654	655	656	…	681	682	683	684	685
AA	A	V	R	D		V	A	I	H	A	D		W	R	V	Y		A	E	H	V		P	R	R	A	R
HBP	0.537	0.537	0.537	0.021		0.021	0.021	0.021	0.021	0.021	0.005		0.200	0.200	0.200	0.200		0.189	0.189	0.189	0.189		0.042	0.853	0.853	0.853	0.853

AA No.	…	761	762	763	764	765	766	767	768	…	776	777	778	779	…	788	789	790	…	792	793	794	795	796	797	…	831
AA		T	Q	L	N	R	A	L	T		K	N	T	Q		I	Y	K		P	P	I	K	D	F		A
HBP		0.005	0.005	0.005	0.005	0.747	0.747	0.747	0.005		0.179	0.179	0.179	0.179		0.258	0.258	0.258		0.189	0.189	0.174	0.474	0.284	0.284		0.021

AA No.	832	833	834	835	836	837	…	844	845	846	847	…	852	853	854	855	856	…	902	903	904	905	906	…	921	922	923
AA	G	F	I	K	Q	Y		I	A	A	R		A	Q	K	F	N		M	A	Y	R	F		K	L	I
HBP	0.021	0.321	0.321	0.321	0.005	0.005		0.563	0.563	0.563	0.563		0.005	0.005	0.005	0.005	0.005		0.005	0.005	0.005	0.005	0.005		0.689	0.689	0.689

AA No.	924	…	933	934	935	…	944	945	946	947	948	949	…	962	963	964	…	980	981	982	983	984	985	…	993	994	995
AA	A		K	I	Q		A	L	G	K	L	Q		L	V	K		I	L	S	R	L	D		I	D	R
HBP	0.689		0.068	0.068	0.068		0.116	0.116	0.116	0.363	0.363	0.247		0.758	0.758	0.758		0.042	0.042	0.042	0.358	0.358	0.316		0.021	0.021	0.700

AA No.	996	997	998	999	1000	1001	…	1012	1013	1014	1015	1016	…	1018	1019	1020	…	1036	1037	1038	1039	1040	1041	1042	…	1047	1048
AA	L	I	T	G	R	L		L	I	R	A	A		I	R	A		Q	S	K	R	V	D	F		Y	H
HBP	0.700	0.700	0.311	0.363	0.363	0.363		0.047	0.047	0.047	0.047	0.047		0.274	0.274	0.274		0.011	0.011	0.374	0.374	0.374	0.011	0.011		0.158	0.158

AA No.	1049	1050	1051	1052	…	1060	1061	1062	1063	1064	1065	…	1078	1079	1080	1081	1082	1083	…	1085	1086	1087	1088	1089	1090	1091	…
AA	L	M	S	F		V	V	F	L	H	V		A	P	A	I	C	H		G	K	A	H	F	P	L	
HBP	0.158	0.158	0.005	0.005		0.011	0.011	0.011	0.011	0.011	0.011		0.005	0.005	0.005	0.005	0.005	0.005		0.021	0.479	0.479	0.479	0.479	0.021	0.021	

AA No.	1101	1102	1103	1104	…	1147	1148	1149	1150	1151	1152	1153	1154	1155	1156	1157	…	1181	1182	1183	1184	1185	1186	1187	1188	1189	1190
AA	H	W	F	V		S	F	K	E	E	L	D	K	Y	F	K		K	E	I	D	R	L	N	E	V	A
HBP	0.079	0.079	0.079	0.079		0.189	0.195	0.195	0.005	0.005	0.005	0.405	0.405	0.405	0.405			0.026	0.026	0.026	0.026	0.126	0.126	0.100	0.489	0.411	0.411

AA No.	1191	1192	1193	1194	…	1203	1204	1205	1206	1207	1208	1209	1210	1211	1212	…	1253	1254	1255	1256	…	1264	1265	1266	…	1268	1269
AA	K	N	L	N		L	G	K	Y	E	Q	Y	I	K	W		C	C	K	F		V	L	K		V	K
HBP	1	0.016	0.016	0.016		0.400	0.400	0.558	0.558	0.158	0.005	0.126	0.126	0.126	0.126		0.058	0.058	0.058			0.742	0.742	0.742		0.021	0.021

AA No.	1270	1271	1272	Cut-off 0.65	241	242	243	244	245	246	247	248	249
AA	L	H	Y		L	L	A	L	H	R	R	Y	L
HBP	0.021	0.021	0.021		0.00431	0.00431	0.00431	0.00431	0.00431	0.00431	0.00431	0.00431	0.00431

CRediT authorship contribution statement

Zachariah P. Schuurs: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Edward Hammond:** Conceptualization, Writing - review & editing. **Stefano Elli:** Formal analysis, Investigation, Writing - review & editing, Investigation. **Timothy R. Rudd:** Formal analysis, Investigation, Writing - review & editing, Investigation. **Courtney J. Mycroft-West:** Methodology, Investigation, Formal analysis, Visualization. **Marcelo A. Lima:** Methodology, Investigation, Formal analysis, Visualization. **Mark A. Skidmore:** Methodology, Investigation, Formal analysis, Visualization. **Richard Karlsson:** Methodology, Investigation, Formal analysis, Visualization. **Yen-Hsi Chen:** Methodology, Investigation, Formal analysis, Visualization. **Ieva Bagdonaite:** Methodology, Investigation, Formal analysis, Visualization. **Zhang Yang:** Methodology, Investigation, Formal analysis, Visualization. **Yassir A. Ahmed:** Methodology, Investigation, Formal analysis, Visualization. **Derek J. Richard:** Writing - review & editing. **Jeremy Turnbul:** Methodology, Validation, Funding acquisition. **Vito Ferro:** Conceptualization, Validation, Writing - review & editing. **Deirdre R. Coombe:** Conceptualization, Validation, Writing - review & editing. **Neha S. Gandhi:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.05.002.

References

[1] Ahamad S, Kanipakam H, Gupta D. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. J Biomol Struct Dyn 2020. https://doi.org/10.1080/07391102.2020.1811774.

[2] Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020;181 (2):281–292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

[3] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579 (7798):270–3. https://doi.org/10.1038/s41586-020-2012-7.

[4] Lan J, Ge J, Yu J, Shan S, Zhou H, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 2020;581 (7807):215–20. https://doi.org/10.1038/s41586-020-2180-5.

[5] Wrapp D, De Vlieger D, Corbett KS, Torres GM, Wang N, et al. Structural basis for potent neutralization of betacoronaviruses by single-domain camelid antibodies. Cell 2020;181(5):1004–1015.e15. https://doi.org/10.1016/j.cell.2020.04.031.

[6] Wrapp D, Wang N, Corbett KS, Goldsmith JA, Hsieh C-L, et al. Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. Science 2020;367(6483):1260–3. https://doi.org/10.1126/science.abb2507.

[7] Clausen TM, Sandoval DR, Spliid CB, Pihl J, Painter CD, et al. SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. Cell 2020;183 (4):1043–1057.e15. https://doi.org/10.1016/j.cell.2020.09.033.

[8] Partridge L J, Urwin L, Nicklin M J H, James D C, Green L R, et al. ACE2-independent interaction of SARS-CoV-2 spike protein to human epithelial cells can be inhibited by unfractionated heparin. bioRxiv 2020. https://doi.org/10.1101/2020.05.21.107870. Preprint:2020.05.21.107870.

[9] Belouzard S, Millet JK, Licitra BN, Whittaker GR. Mechanisms of coronavirus cell entry mediated by the viral spike protein. Viruses 2012;4(6):1011–33. https://doi.org/10.3390/v4061011.

[10] Bestle D, Heindl M R, Limburg H, Van Lam van T, Pilgram O, et al. TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. Life Sci Alliance 2020;39:. https://doi.org/10.26508/lsa.202000786e202000786.

[11] Cheng Y-W, Chao T-L, Li C-L, Chiu M-F, Kao H-C, et al. Furin inhibitors block SARS-CoV-2 spike protein cleavage to suppress virus production and cytopathic effects. Cell Rep 2020;33(2):. https://doi.org/10.1016/j.celrep.2020.108254108254.

[12] Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 2020;181(2):271–280.e8. https://doi.org/10.1016/j.cell.2020.02.052.

[13] Milewska A, Nowak P, Owczarek K, Szczepanski A, Zarebski M, et al. Entry of human coronavirus NL63 into the cell. J Virol 2018;92(3):e01933–e2017. https://doi.org/10.1128/JVI.01933-17.

[14] Milewska A, Zarebski M, Nowak P, Stozek K, Potempa J, et al. Human coronavirus NL63 utilizes heparan sulfate proteoglycans for attachment to target cells. J Virol 2014;88(22):13221–30. https://doi.org/10.1128/jvi.02078-14.

[15] Lang J, Yang N, Deng J, Liu K, Yang P, et al. Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. PLoS ONE 2011;6(8):. https://doi.org/10.1371/journal.pone.0023710e23710.

[16] Lindahl U, Couchman J, Kimata K, Esko J. Proteoglycans and sulfated glycosaminoglycans. In: Varki A, Cummings R, Esko J, Stanely P, Hart G, Aebi M, Darvill A, Kinoshita T, Packer N, Prestegard J, Schnaar R, Seeberger P, editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.

[17] Coombe DR, Kett WC. Heparan sulfate-protein interactions: therapeutic potential through structure-function insights. Cell Mol Life Sci 2005;62 (4):410–24. https://doi.org/10.1007/s00018-004-4293-7.

[18] Hao W, Ma B, Li Z, Wang X, Gao X, et al. (2020) Binding of the SARS-CoV-2 Spike Protein to Glycans. bioRxiv; Preprint:2020.05.17.100537. 10.1101/2020.05.17.100537.

[19] Kim SY, Jin W, Sood A, Montgomery DW, Grant OC, et al. Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. Antiviral Res 2020;181:. https://doi.org/10.1016/j.antiviral.2020.104873104873.

[20] Chowdhury P. In silico investigation of phytoconstituents from Indian medicinal herb 'Tinospora cordifolia (giloy)' against SARS-CoV-2 (COVID-19) by molecular dynamics approach. J Biomol Struct Dyn 2020. https://doi.org/10.1080/07391102.2020.1803968.

[21] Qiao B, Olvera de la Cruz M. Enhanced binding of SARS-CoV-2 spike protein to receptor by distal polybasic cleavage sites. ACS Nano 2020;14(8):10616–23. https://doi.org/10.1021/acsnano.0c04798.

[22] Shamsi A, Mohammad T, Anwar S, Alajmi Mohamed F, Hussain A, et al. Glecaprevir and maraviroc are high-affinity inhibitors of SARS-CoV-2 main protease: possible implication in COVID-19 therapy. Biosci Rep 2020;40(6). https://doi.org/10.1042/bsr20201256.

[23] Basit A, Ali T, Rehman SU. Truncated human angiotensin converting enzyme 2; a potential inhibitor of SARS-CoV-2 spike glycoprotein and potent COVID-19 therapeutic agent. J Biomol Struct Dyn 2020. https://doi.org/10.1080/07391102.2020.1768150. In press.

[24] Shajahan A, Supekar NT, Gleinich AS, Azadi P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. Glycobiology 2020;30(12):981–8. https://doi.org/10.1093/glycob/cwaa042.

[25] Grant OC, Montgomery D, Ito K, Woods RJ. Analysis of the SARS-CoV-2 spike protein glycan shield: implications for immune recognition. Sci Rep 2020;10:. https://doi.org/10.1038/s41598-020-71748-714991.

[26] Casalino L, Gaieb Z, Goldsmith JA, Hjorth CK, Dommer AC, et al. Beyond shielding: the roles of glycans in the SARS-CoV-2 spike protein. ACS Cent Sci 2020;6(10):1722–34. https://doi.org/10.1021/acscentsci.0c01056.

[27] Robson B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. Comput Biol Med 2020;121(103749). https://doi.org/10.1016/j.compbiomed.2020.103749.

[28] Florindo HF, Kleiner R, Vaskovich-Koubi D, Acúrcio RC, Carreira B, et al. Immune-mediated approaches against COVID-19. Nat Nanotechnol 2020;15 (8):630–45. https://doi.org/10.1038/s41565-020-0732-3.

[29] Brouwer PJM, Caniels TG, van der Straten K, Snitselaar JL, Aldon Y, et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. Science 2020;369(6504):643–50. https://doi.org/10.1126/science.abc5902.

[30] Cao Y, Su B, Guo X, Sun W, Deng Y, et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. Cell 2020;182(1):73–84.e16. https://doi.org/10.1016/j.cell.2020.05.025.

[31] Trezza A, Iovinelli D, Santucci A, Prischi F, Spiga O. An integrated drug repurposing strategy for the rapid identification of potential SARS-CoV-2 viral inhibitors. Sci Rep 2020;10(1):13866. https://doi.org/10.1038/s41598-020-70863-9.

[32] Cagno V, Tseligka ED, Jones ST, Tapparel C. Heparan sulfate proteoglycans and viral attachment: true receptors or adaptation bias? Viruses 2019;11(7):596. https://doi.org/10.3390/v11070596.

[33] Mycroft-West CJ, Su D, Pagani I, Rudd TR, Elli S, et al. 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):1700–15. https://doi.org/10.1055/s-0040-1721319.

[34] Toelzer C, Gupta K, Yadav SKN, Borucu U, Davidson AD, et al. Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. Science 2020;370(6517):725–30. https://doi.org/10.1126/science.abd3255.

[35] Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, et al. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 Virus. Cell 2020. https://doi.org/10.1016/j.cell.2020.06.043.

[36] Weissman D, Alameh MG, de Silva T, Collini P, Hornsby H, et al. D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. Cell Host Microbe 2020;29(20):1–9. https://doi.org/10.1016/j.chom.2020.11.012.

[37] Mansbach R A, Chakraborty S, Nguyen K, Montefiori D C, Korber B, et al. (2020) The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational State. bioRxiv; Preprint:2020.07.26.219741. 10.1101/2020.07.26.219741

[38] Yurkovetskiy L, Wang X, Pascal KE, Tomkins-Tinch C, Nyalile TP, et al. Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. Cell 2020;183(3):739–751.e8. https://doi.org/10.1016/j.cell.2020.09.032.

[39] Esko J. Glycosaminoglycan-binding proteins. In: Varki A, Cummings R, Esko J, Freeze H, Hart G, Marth J, editors. Essentials of Glycobiology. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1999.

[40] Giulia Paiardi S R, Marco Rusnati, Rebecca C Wade (2021) Mechanism of inhibition of SARS-CoV-2 infection by the interaction of the spike glycoprotein with heparin. arXiv; Preprint:2103.07722.

[41] Hoffmann M, Kleine-Weber H, Pöhlmann S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. Mol Cell 2020;78(4):779–784.e5. https://doi.org/10.1016/j.molcel.2020.04.022.

[42] Wang C, Li W, Drabek D, Okba NMA, van Haperen R, et al. A human monoclonal antibody blocking SARS-CoV-2 infection. Nat Comm 2020;11 (1):2251. https://doi.org/10.1038/s41467-020-16256-y.

[43] Chen X, Li R, Pan Z, Qian C, Yang Y, et al. Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. Cell Mol Immunol 2020;17(6):647–9. https://doi.org/10.1038/s41423-020-0426-7.

[44] Yuan M, Wu NC, Zhu X, Lee C-CD, So RTY, et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science 2020;368(6491):630–3. https://doi.org/10.1126/science.abb7269.

[45] Wu Y, Wang F, Shen C, Peng W, Li D, et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. Science 2020;368(6496):1274–8. https://doi.org/10.1126/science.abc2241.

[46] Ju B, Zhang Q, Ge J, Wang R, Sun J, et al. Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature 2020;584(7819):115–9. https://doi.org/10.1038/s41586-020-2380-z.

[47] Chan KK, Dorosky D, Sharma P, Abbasi SA, Dye JM, et al. Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. Science 2020;369(6508):1261–5. https://doi.org/10.1126/science.abc0870.

[48] Baum A, Fulton B O, Wloga E, Copin R, Pascal K E, et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. Science 2020; In press. doi:10.1126/science.abd0831.

[49] Chi X, Yan R, Zhang J, Zhang G, Zhang Y, et al. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science 2020;369(6504):650–5. https://doi.org/10.1126/science.abc6952.

[50] Acharya P, Williams W, Henderson R, Janowska K, Manne K, et al. (2020) A glycan cluster on the SARS-CoV-2 spike ectodomain is recognized by Fab-dimerized glycan-reactive antibodies. bioRxiv; Preprint:2020.06.30.178897. doi:10.1101/2020.06.30.178897.

[51] Kwon PS, Oh H, Kwon S-J, Jin W, Zhang F, et al. Sulfated polysaccharides effectively inhibit SARS-CoV-2 in vitro. Cell Disc 2020;6(1):50. https://doi.org/10.1038/s41421-020-00192-8.

[52] Caly L, Druce J, Roberts J, Bond K, Tran T, et al. Isolation and rapid sharing of the 2019 novel coronavirus (SARS-CoV-2) from the first patient diagnosed with COVID-19 in Australia. Med J Aust 2020;212(10):459–62. https://doi.org/10.5694/mja2.50569.

[53] Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations 2020; COVID-19 Genomics Consortium UK: Virological.

[54] Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, et al. (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. medRxiv; Preprint:2020.12.21.20248640. doi:10.1101/2020.12.21.20248640.

[55] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46(W1):W296–303. https://doi.org/10.1093/nar/gky427.

[56] Wang Q, Zhang Y, Wu L, Niu S, Song C, et al. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 2020;181(4):894–904.e9. https://doi.org/10.1016/j.cell.2020.03.045.

[57] Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, et al. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 2004;25(13):1605–12. https://doi.org/10.1002/jcc.20084.

[58] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem 2008;29(11):1859–65. https://doi.org/10.1002/jcc.20945.

[59] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, et al. CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field. J Chem Theory Comput 2016;12(1):405–13. https://doi.org/10.1021/acs.jctc.5b00935.

[60] Park S-J, Lee J, Qi Y, Kern NR, Lee HS, et al. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. Glycobiology 2019;29(4):320–31. https://doi.org/10.1093/glycob/cwz003.

[61] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79(2):926–35. https://doi.org/10.1063/1.445869.

[62] Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23(3):327–41. https://doi.org/10.1016/0021-9991(77)90098-5.

[63] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001.

[64] Bekker H, Berendson H, Dijkstra E, Achterop S, van Drunen R, et al. (1993) Gromacs: A parallel computer for molecular dynamics simulations. In Physics Computing '92, World Scientific Publishing: Singapore.

[65] Huang J, MacKerell Jr AD. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. J Comput Chem 2013;34 (25):2135–45. https://doi.org/10.1002/jcc.23354.

[66] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14(1):33–8. https://doi.org/10.1016/0263-7855(96)00018-5.

[67] Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global Challenges 2017;1(1):33–46. https://doi.org/10.1002/gch2.1018.

[68] Boittier ED, Burns JM, Gandhi NS, Ferro V. GlycoTorch Vina: docking designed and tested for glycosaminoglycans. J Chem Inf Model 2020;60(12):6328–43. https://doi.org/10.1021/acs.jcim.0c00373.

[69] Woods Group (2005-2021) (2005-2021) GLYCAM Web. http://glycam.org.

[70] Mottarella SE, Beglov D, Beglova N, Nugent MA, Kozakov D, et al. Docking server for the identification of heparin binding sites on proteins. J Chem Inf Model 2014;54(7):2068–78. https://doi.org/10.1021/ci500115j.

[71] Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 2021;30(1):70–82. https://doi.org/10.1002/pro.3943.

[72] Rudd TR, Preston MD, Yates EA. The nature of the conserved basic amino acid sequences found among 437 heparin binding proteins determined by network analysis. Mol Biosyst 2017;13(5):852–65. https://doi.org/10.1039/C6MB00857G.

[73] Gowers R, Linke M, Barnoud J, Reddy T, Melo M, et al. (Year) In MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations, SciPy2016, Austin, Texas, Austin, Texas, pp 98–105.

[74] Kumari R, Kumar R, Lynn A. g_mmpbsa—A GROMACS tool for high-throughput MM-PBSA calculations. J Chem Inf Model 2014;54(7):1951–62. https://doi.org/10.1021/ci500020m.

[75] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov 2015;10(5):449–61. https://doi.org/10.1517/17460441.2015.1032936.

[76] Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 2004;32:W665–7. https://doi.org/10.1093/nar/gkh381.

[77] Sarkar S, Witham S, Zhang J, Zhenirovskyy M, Rocchia W, et al. DelPhi web server: a comprehensive online suite for electrostatic calculations of biological macromolecules and their complexes. Commun Comput Phys 2013;13(1):269–84. https://doi.org/10.4208/cicp.300611.201011s.

[78] Smith N, Witham S, Sarkar S, Zhang J, Li L, et al. DelPhi web server v2: incorporating atomic-style geometrical features into the computational protocol. Bioinformatics 2012;28(12):1655–7. https://doi.org/10.1093/bioinformatics/bts200.

[79] Goddard TD, Huang CC, Meng EC, Pettersen EF, Couch GS, et al. UCSF ChimeraX: meeting modern challenges in visualization and analysis. Protein Sci 2018;27(1):14–25. https://doi.org/10.1002/pro.3235.

[80] Lam TT-Y, Jia N, Zhang Y-W, Shum MH-H, Jiang J-F, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 2020;583 (7815):282–5. https://doi.org/10.1038/s41586-020-2169-0.

[81] van Dorp L, Tan C C, Lam S D, Richard D, Owen C, et al. Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. bioRxiv 2020. https://doi.org/10.1101/2020.11.16.384743.

[82] Ghanakota P, Carlson HA. Driving structure-based drug discovery through cosolvent molecular dynamics. J Med Chem 2016;59(23):10383–99. https://doi.org/10.1021/acs.jmedchem.6b00399.

[83] Woo H, Park S-J, Choi YK, Park T, Tanveer M, et al. Developing a fully glycosylated full-length SARS-CoV-2 spike protein model in a viral membrane. J Phys Chem B 2020;124(33):7128–37. https://doi.org/10.1021/acs.jpcb.0c04553.

[84] Saputri DS, Li S, van Eerden FJ, Rozewicki J, Xu Z, et al. Flexible, functional, and familiar: characteristics of SARS-CoV-2 spike protein evolution. Front Microbiol 2020;11(2112). https://doi.org/10.3389/fmicb.2020.02112.

[85] Seyran M, Takayama K, Uversky V N, Lundstrom K, Palù G, et al. The structural basis of accelerated host cell entry by SARS-CoV-2. FEBS 2020:1–11. https://doi.org/10.1111/febs.15651.

[86] Behloul N, Baha S, Shi R, Meng J. Role of the GTNGTKR motif in the N-terminal receptor-binding domain of the SARS-CoV-2 spike protein. Virus Res 2020;286:. https://doi.org/10.1016/j.virusres.2020.198058 198058.

[87] Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J Comput Chem 2011;32 (10):2319–27. https://doi.org/10.1002/jcc.21787.

[88] Smits NC, Robbesom AA, Versteeg EM, van de Westerlo EM, Dekhuijzen PN, et al. Heterogeneity of heparan sulfates in human lung. Am J Respir Cell Mol Biol 2004;30(2):166–73. https://doi.org/10.1165/rcmb.2003-0198OC.

[89] Haeger SM, Yang Y, Schmidt EP. Heparan sulfate in the developing, healthy, and injured lung. Am J Respir Cell Mol Biol 2016;55(1):5–11. https://doi.org/10.1165/rcmb.2016-0043TR.

[90] Perkins SJ, Fung KW, Khan S. Molecular Interactions between Complement Factor H and Its Heparin and Heparan Sulfate Ligands. Front Immunol 2014;5:126. https://doi.org/10.3389/fimmu.2014.00126.

[91] Rashid Q, Abid M, Jairajpuri M. Elucidating the specificity of non-heparin-based conformational activators of antithrombin for factor Xa inhibition. J Nat Sc Biol Med 2014;5(1):36–42. https://doi.org/10.4103/0976-9668.127282.

[92] Pol-Fachin L, Verli H. Structural glycobiology of heparin dynamics on the exosite 2 of coagulation cascade proteases: implications for glycosaminoglycans antithrombotic activity. Glycobiology 2013;24 (1):97–105. https://doi.org/10.1093/glycob/cwt095.

[93] Pol-Fachin L, Franco Becker C, Almeida Guimarães J, Verli H. Effects of glycosylation on heparin binding and antithrombin activation by heparin 2011;9:2735–45. https://doi.org/10.1002/prot.23102.

[94] Olson ST, Frances-Chmura AM, Swanson R, Björk I, Zettlmeissl G. Effect of individual carbohydrate chains of recombinant antithrombin on heparin affinity and on the generation of glycoforms differing in heparin affinity. Arch Biochem Biophys 1997;341(2):212–21. https://doi.org/10.1006/abbi.1997.9973.

[95] Choi B, Choudhary MC, Regan J, Sparks JA, Padera RF, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. New Engl J Med 2020;383(23):2291–3. https://doi.org/10.1056/NEJMc2031364.

[96] EDCDC (2020) Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the United Kingdom; ECDC: Stockholm: 20 December 2020.

[97] Liu L, Chopra P, Li X, Bouwman K M, Tompkins S M, et al. (2021) Heparan sulfate proteoglycans as attachment factor for SARS-CoV-2. bioRxiv; Preprint:2020.05.10.087288. doi:10.1101/2020.05.10.087288.

[98] Niesen FH, Berglund H, Vedadi M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2007;2 (9):2212–21. https://doi.org/10.1038/nprot.2007.321.

[99] Tree J, Turnbull J, Buttigieg K, Elmore M, Coombes N, et al. Unfractionated heparin inhibits live wild-type SARS-CoV-2 cell infectivity at therapeutically relevant concentrations. Br J Pharmacol 2020. https://doi.org/10.1111/bph.15304.

[100] Lee E, Pavy M, Young N, Freeman C, Lobigs M. Antiviral effect of the heparan sulfate mimetic, PI-88, against dengue and encephalitic flaviviruses. Antiviral Res 2006;69(1):31–8. https://doi.org/10.1016/j.antiviral.2005.08.006.

[101] Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 2009;25(9):1189–91. https://doi.org/10.1093/bioinformatics/btp033.