Binding of the SARS-CoV-2 Spike Protein to Glycans

Wei Hao[1,‡], Bo Ma[2,‡], Ziheng Li[1], Xiaoyu Wang[2], Xiaopan Gao[1], Yaohao Li[2,3], Bo Qin[1], Shiying Shang[4], Sheng Cui[1,*], Zhongping Tan[2,*]

[1] NHC Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology and Center for AIDS Research, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100730, China.
[2] State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100050, China
[3] Department of Chemistry and Biochemistry and BioFrontiers Institute, University of Colorado, Boulder CO 80303, United States
[4] School of Pharmaceutical Sciences, Tsinghua University, Beijing 100084, China

*** Correspondence:**
Corresponding Author
cui.sheng@ipb.pumc.edu.cn
zhongping.tan@imm.pumc.edu.cn

Keywords: Heparan sulfate, Glycan array, Molecular docking, Sulfation, Sialic acid

Abstract

The 2019 novel coronavirus (SARS-CoV-2) is the seventh human coronavirus. The pandemic of this virus has caused a high number of deaths in the world. In order to more efficiently combat this pandemic, it is necessary to develop a better understanding of how the virus infects host cells. Infection normally starts with the initial attachment of the virus to cell-surface glycans like heparan sulfate (HS) proteoglycans and sialic acid-containing oligosaccharides. In this study, we used glycan microarray technology to study the binding of the SARS-CoV-2 spike protein (S protein) to HS and sialic acid. Our results indicated that the S protein can bind to HS in a sulfation-dependent manner and the length of HS appears not to be a critical factor for the binding. No binding with sialic acid residues was detected. In addition, we applied sequence alignment and molecular docking to analyze and explain the observed binding results. Our results suggested that HS may stabilize the open conformation of the S protein to promote the subsequent binding of the S protein to the virus entry receptor ACE2. Overall, this work supports the potential importance of HS in SARS-CoV-2 infection and in the development of antiviral agents.

Introduction

The 2019 novel coronavirus (CoV) is the seventh human coronavirus[1]. It is a deadly virus that is affecting the whole world in an unprecedented way. The global impact of the coronavirus disease 2019 (COVID-19) pandemic is far beyond that of

two other major coronavirus outbreaks in the past 20 years, the severe acute respiratory disease (SARS) in 2003[2,3] and the Middle East respiratory disease (MERS) in 2012[4]. Given that all three highly pathogenic CoVs were originated from bats and a large number of closely related CoVs are present in bats, future outbreak of this type of zoonotic virus remains possible. In order to avoid facing a similar pandemic in the future, it is necessary to develop a better understanding of these CoVs, especially the causative agent of the current COVID-19 pandemic.

Studies showed that the genome of the SARS-CoV-2 has about 80% nucleotide identity with that of SARS-CoV[5]. The major differences are found in the regions encoding the structural proteins (envelope E, membrane M, nucleocapsid N, and spike S) and accessory proteins (ORF3a/3b, 6, 7a/7b, 8 and 10), not the nonstructural proteins (nsp1 to nsp16). Based on this genetic similarity, the 2019 novel CoV was named by the International Committee on Taxonomy of Viruses (ICTV) as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[6]. However, despite their high homology at the nucleotide level, there are important differences between these two viruses. For example, SARS-CoV-2 has a relatively lower mortality rate than SARS-CoV, but is more transmissible among humans[7]. Such differences may indicate differences in viral attachment and subsequent viral entry into human cells.

Like other coronaviruses, in order to efficiently infect host cells, SARS-CoV-2 must bind with cell surface molecules in the lungs and other organs to mediate viral attachment and entry into host cells. Recent studies revealed that the receptor-binding domain (RBD) of SARS-CoV-2 S protein S1 subunit strongly interacts with receptor angiotensin-converting enzyme 2 (ACE2)[8-10], which may facilitate its infection of cells expressing this membrane protein. Previous studies of many other viruses suggested that, in addition to ACE2, SARS-CoV-2 S protein may use other molecules on host cell surface as attachment factors for binding to susceptible cells. Examples of such molecules include glycosaminoglycans (GAGs) and sialic acid-containing oligosaccharides.

GAGs are attachment factors for many different viruses[11,12], especially those in the β-coronavirus family, like SARS-CoV-2 and SARS-CoV. GAGs are primarily localized at the outer surface of cells. Such a location makes them particularly suitable for recruiting viruses to cell surfaces. HS is one of the most prevalent types of GAGs in mammals. It is a linear and sulfated polysaccharide that is abundantly expressed on the surface of almost all cell types and in the extracellular matrix. The HS chains are mostly covalently linked as side chains to core proteins to form HS proteoglycans (HSPGs) (Fig. 1). Recently, HS was suggested as an attachment factor for SARS-CoV-2[13].

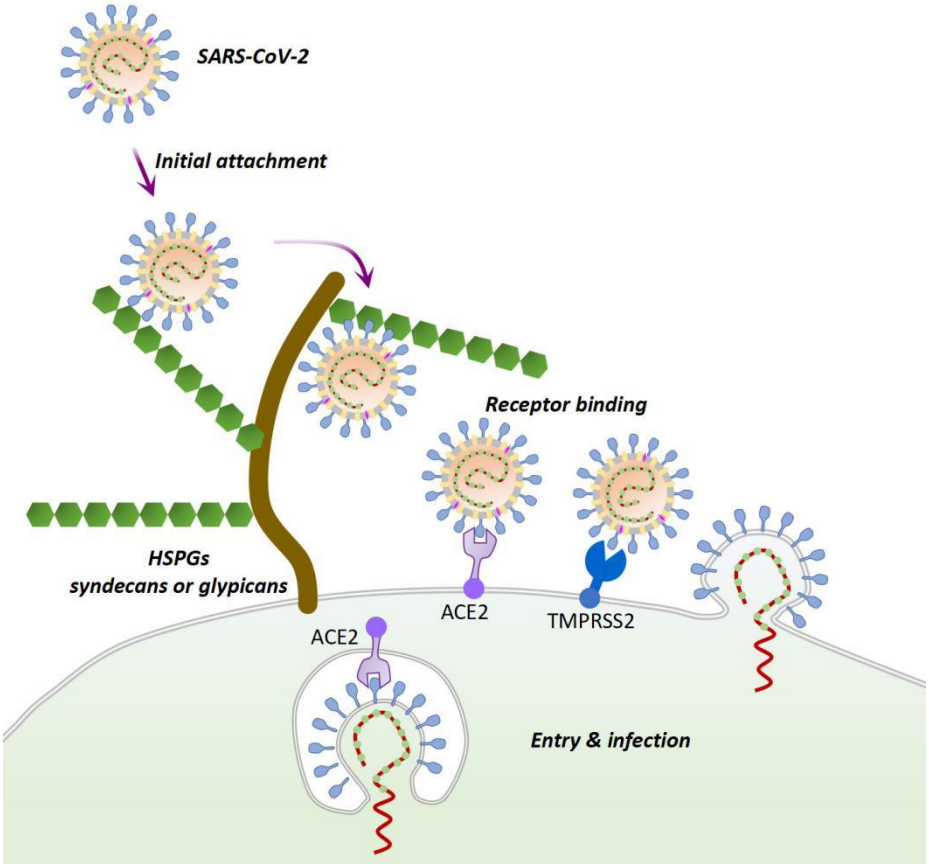

Figure 1. A possible mechanism for SARS-CoV-2 entry and infection. At the early stage of the process, SARS-CoV-2 may first interact with the HSPGs on the surface of susceptible cells using the spike protein protruding from the virus particle. This initial attachment may promote the subsequent binding of the virus to the entry receptor ACE2. The TMPRSS2 protease on host cell surface may assist in viral entry.

HS is synthesized in the Golgi apparatus by many different enzymes. During and after its assembly, HS undergoes extensive series of modifications including sulfation, acetylation and epimerization, which leads to glycan structures with high heterogeneity in length, sulfation, and glucuronate/iduronate ratio[14]. Considerable variation in the sulfation pattern and degree of HS was noted in different species, organs, tissues, and even at different ages and disease stages. The sequence and sulfation pattern of HS has been shown to be able to regulate the binding of many viruses to host cells during infection[15,16]. A similar trend was also observed for glycan sialylation. These findings implicate that a possible relationship may exist between the distribution of HS/sialylated glycans and the viral tropism[17-19].

Better understanding of viral tropism can potentially contribute to the design of new antiviral strategies. Although currently the data on the viral tropism of SARS-CoV-2 are limited, results from recent studies suggest that its tropism may not be correlated with the ACE2 expression and some other receptors may be the determinant of cellular susceptibility to the infection with this virus[20,21]. The intriguing possibility that variation in HS and sialic acid characteristics could impact the tropism of viruses prompts us to investigate the binding of SARS-CoV-2 toward a series of HS and sialic acid containing oligosaccharides [22]. Since the S1 subunit and its RBD are mainly responsible for

mediating the viral attachment to the host cell surface, to a reasonable extent, they can be used to assess the tropism of SARS-CoV-2.

In this study, we compared the binding of SARS-CoV-2 S1 subunit and its RBD to different HS and sialic acid-containing oligosaccharides using microarray experiments. Our results suggested that the S1 subunit and RBD had similar binding preferences for HS oligomers, with higher affinity toward molecules with higher degrees of sulfation. The binding was also shown to be related to the level of 6-O-sulfation and the chain length. Moreover, our study suggested that the S1 subunit might not be able to bind sialic acid residues. Molecular dynamics simulation of the interaction of the RBD with HS provided a possible explanation for the observed binding and highlighted its importance in viral infection. Overall, our study laid a foundation for future studies to explore whether the binding specificity to HS can serve as an important contributor to the viral tropism of SARS-CoV-2 and to explore the possibility of exploiting HS for therapeutic or diagnostic strategies.

Results

Binding of RBD and S1 subunit to a HS microarray. Earlier studies have demonstrated that the S1 subunit is the dominant part of the S protein with respect to binding to cell surface attachment factors [9]. In order to determine if there is any preference of the SARS-CoV-2 S protein for particular HS structures, we investigated the binding of the RBD (the C-terminal region of the S1 subunit. Here termed as SARS-CoV-2-RBD-His) and S1 subunit (Here termed as SARS-CoV-2-S1-His) to a HS microarray containing 24 synthetic heparan sulfate oligosaccharides. These oligosaccharides have systematic differences in their length, monosaccharide composition, and sulfation pattern (Fig. 2). The microarray experiment was performed using a previously established standard protocol [23-25]. Briefly, the proteins were labeled with Cy3 fluorescent dye and incubated with the microarray at different concentrations (SARS-CoV-2-RBD-His at concentrations of 10 μg/ml, 5 μg/ml, 2.5μg/ml, 1.25 μg/ml; SARS-CoV-2-S1-His at concentrations of 4 μg/ml, 2 μg/ml, 1 μg/ml, 0.5 μg/ml). After washing away the unbound HS molecules, a highly sensitive fluorescence method was used to detect the binding of the SARS-CoV-2-S1-His to HS. Under the experimental conditions, the binding can be detected at concentrations higher than 0.5 μg/ml. Increasing the concentration of the protein was not found to produce noticeable changes in binding.

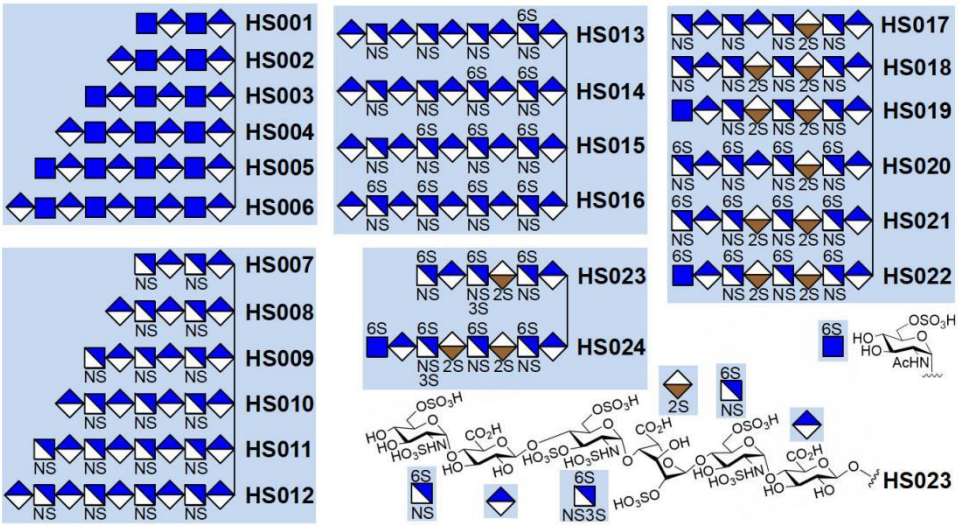

Figure 2: The numbering and structures of the HS oligosaccharides on the microarray. Each HS chain is covalently attached to the microarray slide via the reducing end.

Quantification of fluorescence revealed that the SARS-CoV-2-RBD-His is able to bind to almost half of the molecules on the microarray, and not surprisingly, the binding is strongly affected by the sulfation level, which is a trend that has been previously noted for many HS-binding proteins [26-28]. As shown in Figure 3, the HS oligosaccharides with higher sulfation degree (**HS020-HS024**, the number of sulfate groups per monosaccharide unit >1.00) exhibit higher fluorescence intensity. The highest fluorescence intensity is observed for **HS023** (1.35 sulfate groups per monosaccharide), which is followed by those of **HS021** and **HS024** (1.25 sulfate groups per monosaccharide) and those of **HS020** and **HS022** (1.15 sulfate groups per monosaccharide). Binding of SARS-CoV-2-RBD-His to HS seems to not be affected by the monosaccharide composition (compare **HS020** with **HS022**, and **HS021** with **HS024**).

The oligosaccharides with relatively lower sulfation levels (**HS001-HS019**, the number of sulfate groups per monosaccharide unit <1.00) have lower or almost no fluorescence signals. An analysis of the effect of the variation in sulfation revealed that the position of sulfation is another factor that strongly influence the binding. As shown in the Figure 3, removal of the 6-O-sulfate group from the glucosamine units significantly reduced the binding (compare **HS017** with **HS020**, **HS018** with **HS021,** and **HS019** with **HS022**), suggesting that the 6-O-sulfate in HS plays a crucial role in determining its interaction with SARS-CoV-2-RBD-His. The importance of the 6-O-sulfate for binding is further supported by comparing the binding of **HS012**, **HS013**, **HS014**, **HS015,** and **HS016**, which shows that the one-by-one addition of sulfate to the 6-O-position of the glucosamine residues gradually increased the binding of HS with SARS-CoV-2-RBD-His. The microarray study also indicated that the binding is related to the length of the HS chain, with a preference for shorter ones (compare **HS001**-**HS006**, and **HS007**-**HS0012**).

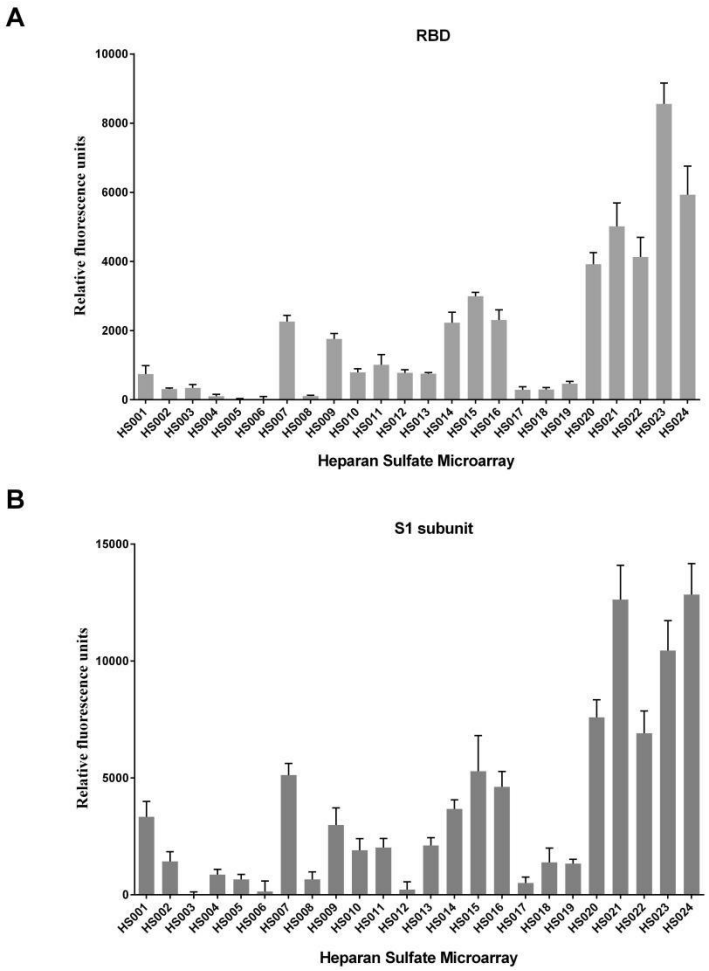

Figure 3. The contributions of different HS characteristics to the binding to (A) SARS-CoV-2-RBD-His at the concentration of 2.5μg/ml and (B) SARS-CoV-2-S1-His at the concentration of 2 μg/ml. The fluorescence intensity was measured at an excitation wavelength of 532 nm. All error bars are standard deviation of more than three replicates.

Using a surface plasmon resonance binding assay, we determined the binding affinity (K_D) of the SARS-CoV-2-RBD-His with a commercially available heparin, a highly sulfated HS (Supporting Information, Fig. S1). The value is 626 μM, which agrees well with previous observations that HS is a weak binder to viral RBD [29]. The binding results of the SARS-CoV-2-S1-His follow a similar trend as those of the SARS-CoV-2-RBD-His. This is consistent with the assumption that RBD is the major determinant for viral S protein binding to HS.

Binding of S1 subunit and RBD to a sialylated N-glycan microarray. Many human viruses, including MERS-CoV, can interact with sialic acid-containing glycans present on the cell surface [30,31]. Such an interaction is normally mediated by the N-terminal domain of the S1 subunit [32]. In order to find out if SARS-CoV-2 can bind to sialic acid residues, we carried out a microarray analysis of the SARS-CoV-2-RBD-His and SARS-CoV-2-S1-His. The microarray used here contains 100 different N-glycans that may be found on the surface of cells. 49 of them are terminated with α2,3- and α2,6-linked sialic acid, also known as N-acetylneuraminic acid (Neu5Ac), 8 with α2,3- and

α2,6-linked N-Glycolylneuraminic acid (Neu5Gc), and the rest with other glycan residues (Supporting Information, Table S1). The experiment was carried out in the similar way as described above using different concentrations of SARS-CoV-2-RBD-His and SARS-CoV-2-S1-His (SARS-CoV-2-RBD-His at concentrations of 10 μg/ml, 5 μg/ml, 2.5μg/ml, 1.25 μg/ml; SARS-CoV-2-S1-His at concentrations of 4 μg/ml, 2 μg/ml, 1 μg/ml, 0.5 μg/ml). The microarray results showed that both the SARS-CoV-2-RBD-His and SARS-CoV-2-S1-His gave no binding signal, suggesting that SARS-CoV-2 may not be able to interact with sialylated N-glycans on cell surface.

Amino acid sequence alignment of the RBD regions. In order to better understand the observed binding between RBD and HS, it is necessary to get more information on the binding sites for HS and the binding poses for RBD-HS complexes. One powerful method for predicting the binding sites is the alignment of the target sequence to homologous proteins [33]. Following this line of thinking, we compared the amino acid sequence of the SARS-CoV-2 RBD with those of SARS-CoV, MERS-CoV, bovine coronaviruses (BCoV) and mouse hepatitis virus (MHV). The reason for selecting the RBDs of these viruses to align with the sequence of the SARS-CoV-2 RBD was that all of these viruses belong to the β-coronavirus family and the RBDs of their S proteins have high sequence and structural similarity, enabling them to be compared with each other. Two of them, SARS-CoV-2 and SARS-CoV, have been found to be able to bind to HS, while the rest of them were demonstrated to not have such binding capacity [29,34]. Past studies have revealed that HS binding usually occurs through electrostatic interactions between a cluster of positively charged basic residues (either K or R) located in one region of a protein and the negatively charged sulfate/carboxylate groups of the HS chains. All of these findings taken together suggested that it may be possible to locate the binding site of the SARS-CoV-2 RBD to HS by analyzing the positional similarities and differences of the Lys and Arg residues in the amino acid sequences of these RBDs.

The sequence alignment was performed using software MUSCLE [35,36]. From the alignment shown in Figure 4, it is not difficult to predict the possible HS binding site in the RBD of SARS-CoV-2. Our results suggests that the binding site may contain three of more of the following basic residues, R355, K356, R357, K462, and K466. The alignment also showed that the arrangement of these basic amino acids is quite different between SARS-CoV-2 and SARS-CoV, which may suggest that these two viruses have different binding preferences toward the HS glycans.

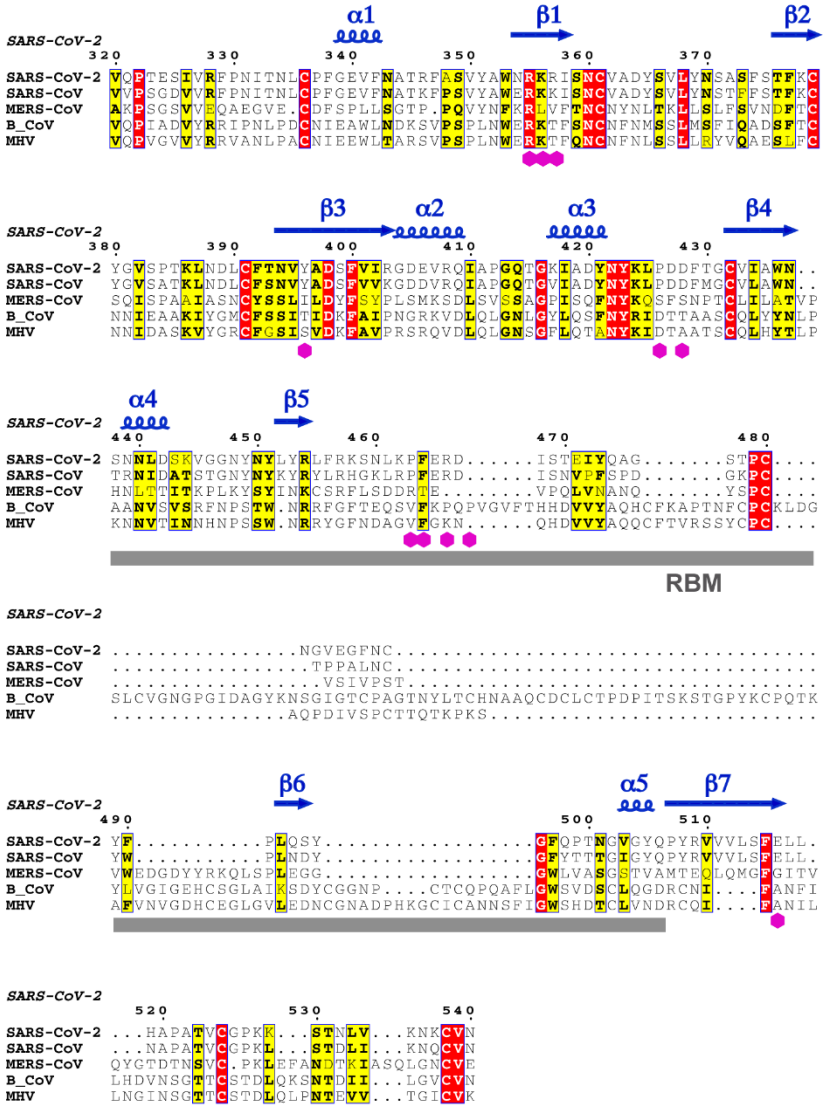

Figure 4: Structure based multiple sequence alignment of CoV RBDs. The UniProt accession number for the RBD of SARS-CoV-2 is P0DTC2, SARS-CoV Q202E9, MERS-CoV K9N5Q8, BCoV P25194, MHV P11224. The secondary structure elements shown above the alignment are assigned according to the crystal structure of the SASR-CoV-2 RBD (PDB: 6M0J). The invariant amino acid residues are highlighted by a red background and the conserved residues by yellow. The receptor-binding motif (RBM) of the RBD is underline by a gray line and the residues predicted to contact HS are indicated by the magenta hexagons.

***In silico* molecular docking of the RBD of SARS-CoV-2 with HS023.** In order to gain structural insights into the binding of the RBD of SARS-CoV-2 to HS, we performed a molecular docking simulation. The simulation was carried out by docking one of the representative HS molecules, **HS023**, to a crystal structure of the RBD using the UCSF DOCK software package [37]. All docking conformations were ranked according to the score assigned by the scoring function and the conformation with the highest score (Grid_Score: 92.153641) was selected to analyze the binding

characteristics.

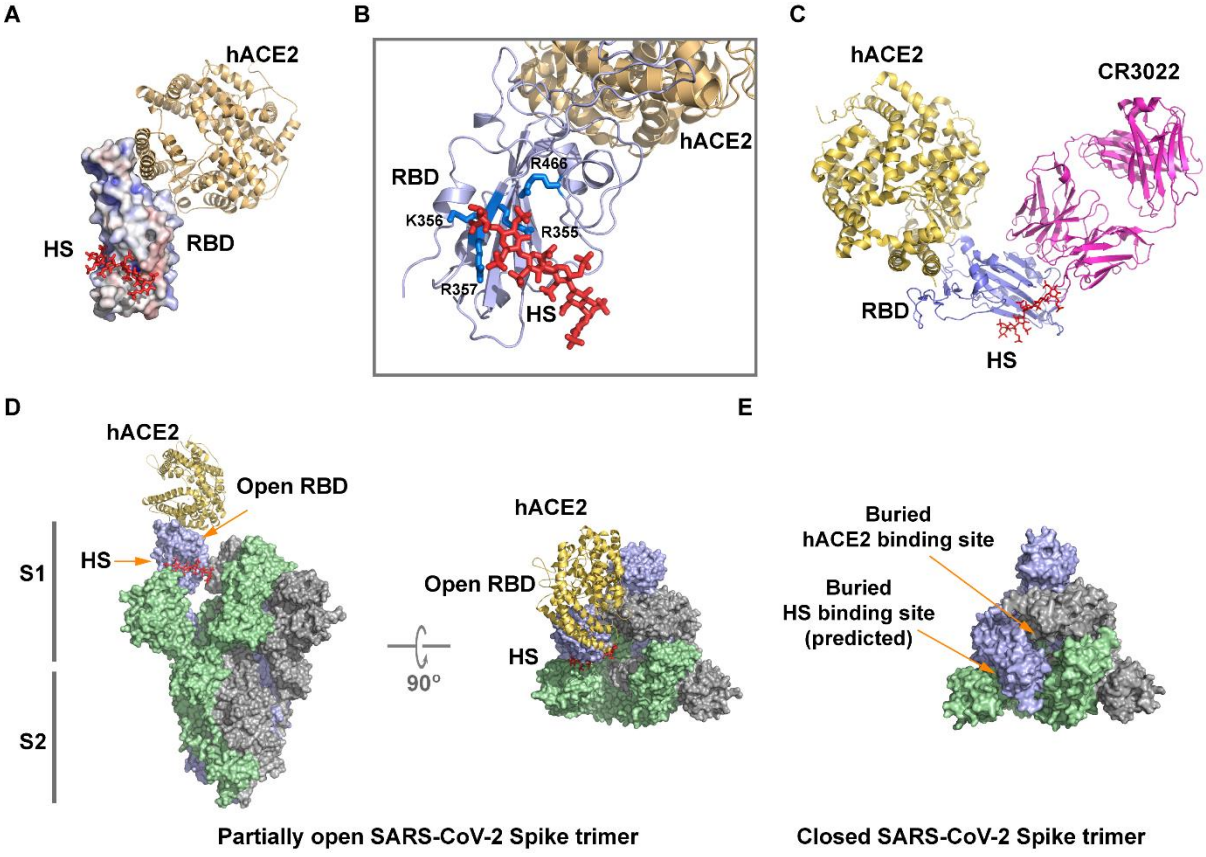

Figure 5. The modeled structures of the SARS-CoV-2 RBD and S protein in complex with HS. The HS023 was docked to the crystal structure of the SARS-CoV-2 RBD (PDB ID: 6M0J) and the cryo-EM structure of the S protein trimer in close and open conformations (PDB: 6VXX and 6VYB). (A) The surface electron potential of the SARS-CoV-2 RBD bound with the human ACE2 (yellow ribbon) and HS (red stick). (B) The magnified view of the interactions shown in (A). The 4 basic residues forming the positively charged patch are highlighted in blue stick representation. (C) The ribbon view of the SARS-CoV-2 RBD bound with the human ACE2 (yellow), HS (red) and a SARS-CoV neutralizing antibody CR3022 (magenta). (D) Right: The side view of the SARS-CoV-2 S protein trimer in open conformation. Left: The top view of the same trimer. The HS molecule (red) and the human ACE2 (yellow) are modeled to the S1-CTD that is rotated up. (E) The top view of the SARS-CoV-2 S protein trimer in close conformation. The receptor binding site for the human ACE2 and the predicted binding site for the HS molecule are both buried inside the trimer (indicated by arrows).

In the highest-scoring conformation, the HS binds a groove formed by three antiparallel β sheets, β1, β3, β7, and two loops, a loop between 458 and 471 and a loop between α3 and β4 (Fig. 4 and 5). The groove contains 4 basic residues that have been predicted to be important for HS binding by the sequence alignment method, R355, K356, R357 and R466. The basic residues form a highly positively charged patch. The close contact between the HS and this basic patch suggests the binding is most likely to be mediated by the nonspecific electrostatic interactions. Analysis also revealed that the putative HS binding site does not overlap with the site occupied by ACE2 or the site targeted by a SARS-CoV neutralizing antibody, CR3022, which can

cross-react with the RBD of SARS-CoV-2 with a much lower binding affinity (Fig. 5C).

We also modeled the HS molecule to the cryo-EM structure of the SARS-CoV-2 S protein trimer in the open conformation using the software Coot [38]. Similar to what have been observed in the structure of the SARS-CoV S protein, the S1 subunits of SARS-CoV-2 also exhibit pronounced structure heterogeneity, in which different S1-CTDs (the S1 subunit C-terminal domain that contains the RBD) adopt two different conformations, closed and open [39]. In the open conformation, the S1-CTD rotates up and the RBD becomes more accessible to its binding partners. In the close conformation, both the HS binding site and the ACE2 binding site are buried inside the S protein trimer (Fig. 5E). Because HS may interact with the SARS-CoV-2 RBD at a very early stage, it is not unreasonable to assume that such interaction could stabilize the open conformation of the S1-CTD, thus promoting the subsequent binding of S protein to its receptor ACE2.

Discussion

For a virus like SARS-CoV-2 to establish infection, it must first attach itself to the surface of target cells in different organs and tissues. The S protein plays an essential role in this attachment process. Recently, the structure of SARS-CoV-2 S protein in the prefusion conformation was determined by the cryo-EM technique[10]. It shows that the overall structure of the SARS-CoV-2 S protein is very similar to that of the closely related SARS-CoV S protein, which is organized as a homotrimer. Each monomer can be divided into an N-terminal receptor-binding S1 subunit and a C-terminal fusion-mediating S2 subunit. The S1 subunits are located at the apex of the spike, making them more accessible for binding to target molecules (most likely through their RBD).

Although similar, there are some notable differences between the SARS-CoV-2 and SARS-CoV S proteins[8,40]. For example, the SARS-CoV-2 S protein contains a polybasic furin protease recognition motif (RRAR) at the junction of S1 and S2, while SARS-CoV and other closely related CoVs do not possess such cleavage sites[9]. Although SARS-CoV-2 also uses its RBD to bind ACE2, the key amino acid residues involved in the binding are largely different from those of SARS-CoV[9,41,42].These differences may be related to the clinical, epidemiological and treatment characteristics of COVID-19. Accumulating evidence appears to support this hypothesis. For example, the SARS-CoV-2 RBD was found to bind to the viral entry receptor ACE2 with a ~20-folds lower Kd than that of the SARS-CoV RBD, which, together with the furin-like cleavage site in the S protein sequence and the transmembrane serine protease 2 (TMPRSS2), may partially account for the increased infectivity of SARS-CoV-2 relative to SARS-CoV[10,43]. Studies also demonstrated that the binding of the SARS-CoV-2 S protein to neutralizing antibodies is different from that of SARS-CoV. Almost all SARS-CoV neutralizing antibodies fail to cross-react with SARS-CoV-2. Although there are exceptions, such as CR3022, its binding affinity for the SARS-CoV-2 RBD is significantly lower (by a factor of >100) and it has no neutralizing ability against SARS-CoV-2[44]. This finding agree well with the fact that the S1 subunit of the SARS-CoV-2 S protein is highly variable. It also in part explains why COVID-19 may require different

treatment modalities.

In addition to binding protein-based receptors, many coronaviruses can interact with cell surface glycans, including GAGs and sialic acid-containing oligosaccharides. Depending on the virus, the glycan molecules can act as attachment factors, co-receptors or primary receptors [45]. Coronaviruses typically bind GAGs through non-specific charge-based interactions on the surface of RBDs. As one of the most abundant GAGs, HS appears to be the preferred binding partner for many viruses [19,46-49]. Binding of coronavirus to sialic acid often occurs in a groove located on the surface of the N-terminal domain of the S1 subunit (S1-NTD) [30]. Sialic acids are normally terminal monosaccharide residues linked to glycans decorating cell surface glycoproteins, glycolipids, or other glycoconjugates [50]. In general, the interactions of coronaviruses with HS or sialic acids are responsible for the first contact with host cells. Such contact may serve to concentrate viruses on the surface of target cells, facilitate their binding to more specific high-affinity protein receptors and promote their entry into host cells[51,52]. It has been demonstrated that virus binding and infection can be reduced by enzymatic removal HS or sialic acid from cell surface, or by treating virus with soluble HS or multivalent sialic acid conjugates[53-56]. Therefore, in order to better understand and treat COVID-19, it is necessary to carry out research to investigate the possible interactions between SARS-CoV-2 and HS and sialic acid-containing glycans, and to assess if such interactions could represent a target for therapeutic intervention.

Similar to studies that have been successfully conducted for many other viruses, we used the microarray technology to study the binding of the RBD and the S1 subunit of SARS-CoV-2 S protein to HS and sialic acid[53]. Detection of binding was achieved using a sensitive fluorescence technique. Changes in fluorescence intensity reflect changes in binding of the RBD to HS and sialic acid. The results showed that about half of the HS microarray spots have fluorescence clearly above background. In contrast, only background levels of fluorescence were detected on the sialylated N-glycan microarray. This observation supports the previous finding that the RBD of SARS-CoV-2 is able to bind to HS [29]. It was suggested that SARS-CoV could also bind to HS [34]. This well agrees with the fact that these two viruses are genetically closely related. Since the SARS-CoV outbreak in 2002, there has been no report that it can bind sialic acid-containing glycans. Because of its similarity to SARS-CoV-2, it is not surprise to find that no binding was observed between the S1 subunit of SARS-CoV-2 S protein and α2,3- and α2,6-sialylated N-glycans.

Our results also suggested that RBD/S1-HS binding is likely to be related to HS size, sulfation position and degree. It seems that more 6-O-sulfate groups and higher sulfation degree lead to better binding. Because HSPGs exhibit different sulfation patterns in different tissues, such a binding specificity may contribute to the tropism of SARS-CoV-2 for human cells [57]. The length of HS appears not to be a critical factor for the binding. Short HS chains could have comparable binding signals. This finding implies that it may be possible to reduce the attachment of SARS-CoV-2 to the surface of host cells by low-molecular-weight-heparin (LMWH). This is in agreement with a recent study showing that LMWH treatment may be associated with better prognosis

in some severe COVID-19 patients[58]. While these initial findings are encouraging, further research is required to determine if HS could be used for the inhibition of SARS-CoV-2 infection and to determine if the binding to HS could affect the tropism and pathogenesis of this virus.

As a first step to understand the possible mechanism and role of the binding of the SARS-CoV-2 S protein to HS, we performed multiple sequence alignment and molecular docking studies using available softwares [35-37]. Our results indicate that the SARS-CoV-2 RBD binds to HS through electrostatic interactions and the binding site is distinct from the binding site for ACE2. Similar to what have been observed in the structures of the SARS-CoV and MERS-CoV S proteins, the structure of the S1-CTD of the SARS-CoV-2 S protein, which is located at the apex of the S protein trimer, is very flexible and can adopt at least two different conformations (close and open) in solution [8,41]. Our docking study revealed that both HS and ACE2 can bind to the S1-CTD only when it is in the open conformation (Fig. 5). Since the attachment of the S protein to HS may occur earlier than the binding to ACE2, this finding raises an interesting possibility that the HS attachment may serve to stabilize the open conformation, thus promote the binding of the S protein to entry receptor ACE2.

Comparative alignment of the sequence of SARS-CoV-2 S1-NTD with those of coronaviruses with known sialic acid binding sites suggests that this N-terminal domain may not be able to bind sialic acid residues, which, to some extent, explains the results we obtained in the sialylated N-glycan microarray experiment (Supporting Information, Fig. S2 and S3). Previous structural studies of S proteins found that the S1-NTDs from different coronaviruses use highly conserved amino acids for the recognition of the same sialic acid molecules. For example, in a recent work, it was shown that the NTD of the human coronavirus OC43 (HCoV-OC43) S protein utilized amino acid residues Leu80, Lys81, Ser83, Leu85, Leu86, Trp90, and Phe95 to interact with 9-O-acetylated sialic acid (9-O-Ac-Sia). Analysis of the NTDs of coronaviruses that can bind the same monosaccharide, such as bovine coronavirus (BCoV) and human coronavirus HKU1 (HCoV-HKU1), revealed that these domains relied on very similar sets of amino acids to form the binding pockets (Supporting Information, Fig. S2) [30]. This finding supports the reasonability of applying multiple-sequence alignment to predict and explain the possible binding of the SARS-CoV-2 S1-NTD to different sialic acid forms. Based on this assumption, we compared the sequences of the S1-NTDs of SARS-CoV-2 and SARS-CoV with each other, as well as with that of MERS-CoV, which can bind sialic acid [59]. Neither SARS-CoV-2 nor SARS-CoV contains the amino acids that form the binding pocket in the MERS-CoV S1-NTD for sialic acid (Supporting Information, Fig. S3). This agrees well with our results that no significant fluorescence signals were detected after the Cy3-labeled SARS-CoV-2 S1-subunit were incubated with the sialylated N-glycan microarray, and the fact that no data is available on the binding of SARS-CoV to sialic acid since the outbreak of this virus in 2002. Comparative sequence alignment also revealed SARS-CoV-2 and SARS-CoV do not have the amino acid residues conserved for the binding of 9-O-Ac-Sia, indicating that these two viruses may not be able to bind this modified sialic acid (Supporting Information, Fig.

S2).

In conclusion, through our study, we provided experimental evidence for the binding of the S protein of SARS-CoV-2 to two types of cell-surface glycans, HS and sialic acid-containing glycans, which are commonly utilized by human viruses for attachment to target cells. Our data revealed that the S protein can weakly bind to HS in a sulfation-dependent manner. No binding with sialic acid residues was detected. This result suggests that HS may act as an attachment factor that serve to concentrate the virus at the cell surface and affect its tropism. Through computation, we predicted the binding site of HS and modeled the possible mechanism by which HS facilitates the virus infection. This study suggests that HS is bound to a positively charged surface patch on the SARS-CoV-2 RBD. This binding may stabilize the S protein trimer in an open conformation, and thus can promote the subsequent binding of the S protein to ACE2. Our study also indicated that sialic acid residues may not contribute to the attachment of SARS-CoV-2 to cells. Overall, our findings support the potential importance of HS in SARS-CoV-2 infection and in the development of antiviral agents.

Experiments

Reagents and cell lines. High Five™ Cells for baculovirus expression were purchased from ThermoFisher Scientific and maintained in Express Five™ Medium. The SARS-CoV2-S1 subunit (16-685) with a C-terminus polyhistidine tag (termed SARS-CoV-2-S1-His) was purchased from ACROBiosystems. It was expressed in HEK293 cells and was confirmed by ELISA and SDS-PAGE. The glycan microarray experiments were performed by Creative Biochip Ltd (Nanjing, China).

Protein expression and purification. DNA containing the coding sequence for an N-terminal hemo signal peptide, the receptor binding domain (RBD, residues 319-541) of SARS-CoV-2 S protein and a C-terminal polyhistidine tag was amplified and inserted into a pFasebac1 vector for expression in High-5 insect cells using the Bac-to-Bac expression system (Invitrogen). The resulting recombinant protein, termed SARS-CoV-2-RBD-His, was secreted into cell culture medium, and subsequently purified on a nickel-nitrilotriacetic acid (Ni-NTA) affinity column, followed by a Superdex 200 gel filtration column (GE Healthcare). The final buffer for the protein contains 10 mM Hepes (pH=7.0) and 100 mM NaCl. The purified SARS-CoV-2-RBD-His was concentrated to 3.5 mg/ml and flash frozen in liquid N_2 and stored at -80 degrees Celsius.

Binding of S1 subunit and RBD to glycan microarrays. SARS-CoV-2-S1-His and SARS-CoV-2-RBD-His were first labeled with Cy3 fluorescent dye (10 mg/ml in DMSO). After dialysis, they were incubated at different concentrations (0.5 μg/ml, 1 μg/ml, 2 μg/ml, 4 μg/ml) with microarrays for 1 hour in the dark at room temperature. After incubation, the microarray slide was gently washed using washing buffer (20 mM Tris-Cl containing 0.1% tween 20, pH 7.4) to remove unbound proteins. Finally, the slide was scanned with a microarray scanner LuxScan-10K/Aat an excitation wavelength of 532 nm and evaluated by the Microarray Image Analyzer software.

Amino acid sequence alignment. The multiple sequence alignment was performed

using software MUSCLE and ESPript. The amino acid sequences of the proteins were retrieved from the UniProt database. The secondary structure elements were determined using the structures deposited in the PDB database.

***In silico* molecular docking.** The HS structure corresponding to the HS023 spot on the HS microarray was docked to the RBD of of SARS-CoV-2 S protein (PDB ID: 6M0J) to predict the binding poses of RBD-bound HS using the UCSF DOCK 6.8 program following the provided instructions. The HS structure was extracted from the PDB file (PDB ID: 4R9W). Its partial atomic charges were assigned using the semi-empirical AM1-BCC method with the Antechamber program. Because the docking center was unknown, the whole receptor was enclosed in the docking box. A flexible docking strategy based on an algorithm called Anchor-and-Grow was used to model the structure of HS bound to RBD, with the max_orientations, anchor_max _iterations and grow_max_iterations being set to 500. The docking conformations were ranked using the scoring function. The highest-scoring conformation in the first cluster was considered as the most probable prediction of the binding pose. To model the HS molecule to the cryo-EM structure of SARS-CoV-2 S trimer (PDB id: 6VYB), Coot software was used to superimpose the HS-RBD model (results from UCSF DOCK) to the opened CTD domain of the trimer.

Acknowledgement

We would like to thank the National Natural Science Foundation of China (Grant number: 91853120), the National Major Scientific and Technological Special Project of China (Grant number: 2018ZX09711001-013), the National Key R&D Program of China (Grant number: 2018YFE0111400), the State Key Laboratory of Bioactive Substance and Function of Natural Medicines, Institute of Materia Medica, and the Chinese Academy of Medical Sciences and Peking Union Medical College CRP-ICGEB Research Grant 2019 (Grant number: CRP/CHN19-02) for funding. We also thank Mr. Kai Zhang and Mrs. Wenjie Zhou for their support to the research.

References

1 Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *The New England journal of medicine* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).

2 Drosten, C. *et al.* Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *The New England journal of medicine* **348**, 1967-1976, doi:10.1056/NEJMoa030747 (2003).

3 Ksiazek, T. G. *et al.* A novel coronavirus associated with severe acute respiratory syndrome. *The New England journal of medicine* **348**, 1953-1966, doi:10.1056/NEJMoa030781 (2003).

4 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *The New England journal of medicine* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).

5 Chan, J. F. *et al.* A familial cluster of pneumonia associated with the 2019 novel

coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet (London, England)* **395**, 514-523, doi:10.1016/s0140-6736(20)30154-9 (2020).

6 Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature microbiology* **5**, 536-544, doi:10.1038/s41564-020-0695-z (2020).

7 Ceccarelli, M., Berretta, M., Venanzi Rullo, E., Nunnari, G. & Cacopardo, B. Differences and similarities between Severe Acute Respiratory Syndrome (SARS)-CoronaVirus (CoV) and SARS-CoV-2. Would a rose by another name smell as sweet? *European review for medical and pharmacological sciences* **24**, 2781-2783, doi:10.26355/eurrev_202003_20551 (2020).

8 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220, doi:10.1038/s41586-020-2180-5 (2020).

9 Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e286, doi:10.1016/j.cell.2020.02.058 (2020).

10 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.)* **367**, 1260-1263, doi:10.1126/science.abb2507 (2020).

11 Kalia, M., Chandra, V., Rahman, S. A., Sehgal, D. & Jameel, S. Heparan sulfate proteoglycans are required for cellular binding of the hepatitis E virus ORF2 capsid protein and for viral infection. *Journal of virology* **83**, 12714-12724, doi:10.1128/jvi.00717-09 (2009).

12 Trybala, E., Liljeqvist, J. A., Svennerholm, B. & Bergström, T. Herpes simplex virus types 1 and 2 differ in their interaction with heparan sulfate. *Journal of virology* **74**, 9106-9114, doi:10.1128/jvi.74.19.9106-9114.2000 (2000).

13 Mycroft-West, C. *et al.* The 2019 coronavirus (SARS-CoV-2) surface protein (Spike) S1 Receptor Binding Domain undergoes conformational change upon heparin binding. 2020.2002.2029.971093, doi:10.1101/2020.02.29.971093 %J bioRxiv (2020).

14 Lyon, M. & Gallagher, J. T. Bio-specific sequences and domains in heparan sulphate and the regulation of cell growth and adhesion. *Matrix biology : journal of the International Society for Matrix Biology* **17**, 485-493, doi:10.1016/s0945-053x(98)90096-8 (1998).

15 Zautner, A. E., Jahn, B., Hammerschmidt, E., Wutzler, P. & Schmidtke, M. N- and 6-O-sulfated heparan sulfates mediate internalization of coxsackievirus B3 variant PD into CHO-K1 cells. *Journal of virology* **80**, 6629-6636, doi:10.1128/jvi.01988-05 (2006).

16 Sasaki, M. *et al.* The Role of Heparan Sulfate Proteoglycans as an Attachment Factor for Rabies Virus Entry and Infection. *The Journal of infectious diseases* **217**, 1740-1749, doi:10.1093/infdis/jiy081 (2018).

17 Shukla, D. & Spear, P. G. Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry. *The Journal of clinical investigation* **108**, 503-510, doi:10.1172/jci13799 (2001).

18 Hilgard, P. & Stockert, R. Heparan sulfate proteoglycans initiate dengue virus infection of hepatocytes. *Hepatology (Baltimore, Md.)* **32**, 1069-1077, doi:10.1053/jhep.2000.18713 (2000).

19 Tamhankar, M. *et al.* Heparan sulfate is an important mediator of Ebola virus infection in

polarized epithelial cells. *Virology journal* **15**, 135, doi:10.1186/s12985-018-1045-0 (2018).

20 Guo, Y. R. *et al.* The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. *Military Medical Research* **7**, 11, doi:10.1186/s40779-020-00240-0 (2020).

21 Bojkova, D. *et al.* SARS-CoV-2 and SARS-CoV differ in their cell tropism and drug sensitivity profiles. 2020.2004.2003.024257, doi:10.1101/2020.04.03.024257 %J bioRxiv (2020).

22 Wickramasinghe, I. N., de Vries, R. P., Gröne, A., de Haan, C. A. & Verheije, M. H. Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *Journal of virology* **85**, 8903-8912, doi:10.1128/jvi.05112-11 (2011).

23 Ahuja, S. Lectin microarray profiling and relative quantification of glycome associated with proteins of neonatal wt and rd1 mice retinae. *Investigative ophthalmology & visual science* **54**, 3272-3280, doi:10.1167/iovs.12-11363 (2013).

24 Amonsen, M., Smith, D. F., Cummings, R. D. & Air, G. M. Human parainfluenza viruses hPIV1 and hPIV3 bind oligosaccharides with alpha2-3-linked sialic acids that are distinct from those bound by H5 avian influenza virus hemagglutinin. *Journal of virology* **81**, 8341-8345, doi:10.1128/jvi.00718-07 (2007).

25 Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T. & Dermody, T. S. The sweet spot: defining virus-sialic acid interactions. *Nature reviews. Microbiology* **12**, 739-749, doi:10.1038/nrmicro3346 (2014).

26 Corti, F. *et al.* N-terminal syndecan-2 domain selectively enhances 6-O heparan sulfate chains sulfation and promotes VEGFA(165)-dependent neovascularization. *Nature communications* **10**, 1562, doi:10.1038/s41467-019-09605-z (2019).

27 Meneghetti, M. C. *et al.* Heparan sulfate and heparin interactions with proteins. *Journal of the Royal Society, Interface* **12**, 0589, doi:10.1098/rsif.2015.0589 (2015).

28 Zong, C. *et al.* Heparan Sulfate Microarray Reveals That Heparan Sulfate-Protein Binding Exhibits Different Ligand Requirements. *Journal of the American Chemical Society* **139**, 9534-9543, doi:10.1021/jacs.7b01399 (2017).

29 Liu, L. *et al.* SARS-CoV-2 spike protein binds heparan sulfate in a length- and sequence-dependent manner. 2020.2005.2010.087288, doi:10.1101/2020.05.10.087288 %J bioRxiv (2020).

30 Tortorici, M. A. *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nature structural & molecular biology* **26**, 481-489, doi:10.1038/s41594-019-0233-y (2019).

31 Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E8508-e8517, doi:10.1073/pnas.1712592114 (2017).

32 Li, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *Journal of virology* **89**, 1954-1964, doi:10.1128/jvi.02615-14 (2015).

33 Pang, B. *et al.* An Integrated Approach to Sequence-Independent Local Alignment of Protein Binding Sites. *IEEE/ACM transactions on computational biology and bioinformatics* **12**, 298-308, doi:10.1109/tcbb.2014.2355208 (2015).

34 Lang, J. *et al.* Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PloS one* **6**, e23710, doi:10.1371/journal.pone.0023710 (2011).

35 Robert, X. & Gouet, P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic acids research* **42**, W320-324, doi:10.1093/nar/gku316 (2014).

36 Madeira, F. *et al.* The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic acids research* **47**, W636-w641, doi:10.1093/nar/gkz268 (2019).

37 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).

38 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132, doi:10.1107/s0907444904019158 (2004).

39 Song, W., Gui, M., Wang, X. & Xiang, Y. Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS pathogens* **14**, e1007236, doi:10.1371/journal.ppat.1007236 (2018).

40 Wrapp, D. *et al.* Cryo-EM Structure of the 2019-nCoV Spike in the Prefusion Conformation. 2020.2002.2011.944462, doi:10.1101/2020.02.11.944462 %J bioRxiv (2020).

41 Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science (New York, N.Y.)* **309**, 1864-1868, doi:10.1126/science.1116480 (2005).

42 Ge, X. Y. *et al.* Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535-538, doi:10.1038/nature12711 (2013).

43 Dockendorff, C. *et al.* Synthetic Analogues of the Snail Toxin 6-Bromo-2-mercaptotryptamine Dimer (BrMT) Reveal That Lipid Bilayer Perturbation Does Not Underlie Its Modulation of Voltage-Gated Potassium Channels. *Biochemistry* **57**, 2733-2743, doi:10.1021/acs.biochem.8b00292 (2018).

44 Yuan, M. *et al.* A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science (New York, N.Y.)* **368**, 630-633, doi:10.1126/science.abb7269 (2020).

45 Raman, R., Tharakaraman, K., Sasisekharan, V. & Sasisekharan, R. Glycan-protein interactions in viral pathogenesis. *Current opinion in structural biology* **40**, 153-162, doi:10.1016/j.sbi.2016.10.003 (2016).

46 Chang, A., Masante, C., Buchholz, U. J. & Dutch, R. E. Human metapneumovirus (HMPV) binding and infection are mediated by interactions between the HMPV fusion protein and heparan sulfate. *Journal of virology* **86**, 3230-3243, doi:10.1128/jvi.06706-11 (2012).

47 Mandl, C. W. *et al.* Adaptation of tick-borne encephalitis virus to BHK-21 cells results in the formation of multiple heparan sulfate binding sites in the envelope protein and attenuation in vivo. *Journal of virology* **75**, 5627-5637, doi:10.1128/jvi.75.12.5627-5637.2001 (2001).

48 Kesari, A. S., Sharkey, C. M. & Sanders, D. A. Role of heparan sulfate in entry and exit of Ross River virus glycoprotein-pseudotyped retroviral vectors. *Virology* **529**, 177-185, doi:10.1016/j.virol.2019.01.022 (2019).

49 Gao, H. *et al.* Role of heparan sulfate in the Zika virus entry, replication, and cell death.

Virology **529**, 91-100, doi:10.1016/j.virol.2019.01.019 (2019).

50 A. Varki *et al.* Essentials of Glycobiology . (Cold Spring Harbor Laboratory Press Copyright 2015-2017 by The Consortium of Glycobiology Editors, La Jolla, California. All rights reserved., 2015).

51 Aquino, R. S. & Park, P. W. Glycosaminoglycans and infection. *Frontiers in bioscience (Landmark edition)* **21**, 1260-1277, doi:10.2741/4455 (2016).

52 Kamhi, E., Joo, E. J., Dordick, J. S. & Linhardt, R. J. Glycosaminoglycans in infectious disease. *Biological reviews of the Cambridge Philosophical Society* **88**, 928-943, doi:10.1111/brv.12034 (2013).

53 Chandra, N. *et al.* Sulfated Glycosaminoglycans as Viral Decoy Receptors for Human Adenovirus Type 37. *Viruses* **11**, doi:10.3390/v11030247 (2019).

54 Techaarpornkul, S., Collins, P. L. & Peeples, M. E. Respiratory syncytial virus with the fusion protein as its only viral glycoprotein is less dependent on cellular glycosaminoglycans for attachment than complete virus. *Virology* **294**, 296-304, doi:10.1006/viro.2001.1340 (2002).

55 Nilsson, E. C., Jamshidi, F., Johansson, S. M., Oberste, M. S. & Arnberg, N. Sialic acid is a cellular receptor for coxsackievirus A24 variant, an emerging virus with pandemic potential. *Journal of virology* **82**, 3061-3068, doi:10.1128/jvi.02470-07 (2008).

56 Reuter, J. D. *et al.* Inhibition of viral adhesion and infection by sialic-acid-conjugated dendritic polymers. *Bioconjugate chemistry* **10**, 271-278, doi:10.1021/bc980099n (1999).

57 Nagamine, S. *et al.* Organ-specific sulfation patterns of heparan sulfate generated by extracellular sulfatases Sulf1 and Sulf2 in mice. *The Journal of biological chemistry* **287**, 9579-9590, doi:10.1074/jbc.M111.290262 (2012).

58 Tang, N. *et al.* Anticoagulant treatment is associated with decreased mortality in severe coronavirus disease 2019 patients with coagulopathy. *Journal of thrombosis and haemostasis : JTH* **18**, 1094-1099, doi:10.1111/jth.14817 (2020).

59 Park, Y. J. *et al.* Structures of MERS-CoV spike glycoprotein in complex with sialoside attachment receptors. *Nature structural & molecular biology* **26**, 1151-1157, doi:10.1038/s41594-019-0334-7 (2019).