

Spike Protein and the Various Cell-Surface Carbohydrates: An Interaction Study

Nisha Grandhi Jayaprakash and Avadhesh Surolia*

Cite This: <https://doi.org/10.1021/acscchembio.1c00691>

Read Online

ACCESS |



Metrics & More

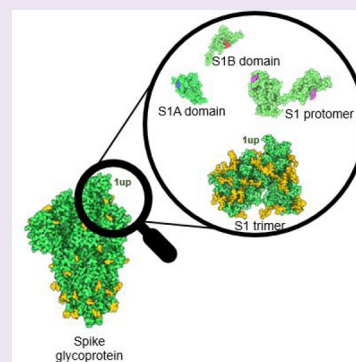


Article Recommendations



Supporting Information

ABSTRACT: The SARS-CoV-2 virus has been known to gain entry into the host cell through the spike protein that binds to the host ACE2 cell surface protein. However, the role of the putative sugar-binding sites in the spike protein has remained unclear. We provide a comprehensive *in silico* outlook into the infection initiation wherein the virus first recognizes the sialosides on the cell via its S1A domain of the spike protein as it surfs over the cell surface. This facilitates the subsequent interaction with the cellular glycosaminoglycans through the S1B domain of the spike protein as it binds to the ACE2 receptor. The unique coadaptation to recognize both the host protein and the cell-surface carbohydrate receptors provides an additional coupling mechanism for efficient viral attachment and infection.



INTRODUCTION

The covid-19 disease is a clinical syndrome associated with SARS-CoV-2 infection caused by a β -coronavirus.¹ These RNA viruses have a distinctive club-shaped spike protein on their envelope membrane and are highly glycosylated. The viral interaction with the host Angiotensin-Converting Enzyme 2 (ACE2) receptor is crucial in mediating attachment and viral entry initiating the infection. They exhibit strong species and tissue tropism demonstrated by their binding to the host cell receptors. Initially thought to infect the upper respiratory tract where the density of the ACE2 receptor is expectedly high, this virus has now been identified for its tropism for many other organs, including eyes, brain, GI tract, kidneys, heart, liver, etc., as seen from autopsy samples and numerous cell lines (~25 types).^{2,3}

Carbohydrates are particularly well suited to serve as receptors due to their localization on the surface of the cell and a rich repertoire of structural motifs that they can generate. Various cell-surface carbohydrates, including sialylated glycans, GAGs (glycosaminoglycans), and blood group antigens (BGAs), have thus been implicated to function as viral recruiting agents.⁴ Sialoglycans terminating with these acidic sugars have been known to have an essential role in transmission, organo-tropism, and viral virulence of several viruses, in particular, the polyomavirus, reovirus, paramyxovirus, orthomyxovirus, parvovirus, and coronavirus.^{5,6} The viruses have been known to engage with the sialic acid moiety primarily using a small number of contacts. The most common form of sialic acid is the 5-*N*-acetyl neuraminic acid (Neu5Ac), which is seen to undergo various modifications (acetylation, methylation, and sulfation), resulting in more than 50 different

sialic acid variants.^{7,8} Heparan sulfate, a linear sulfated glycosaminoglycan, is another highly anionic cell surface glycan suited for this role due to its length and flexibility. It offers multiple binding sites for viral recruitment in herpesvirus, adenovirus, and flavivirus.⁹ A considerable variation in the degree and sulfation pattern has been noted in this molecule across different species, organs, tissues, and even different disease stages. Reports have indicated its role in stabilizing the SARS CoV-2 spike protein's open conformation, promoting the subsequent binding of the ACE2 receptor.¹⁰ The absence of cellular heparan sulfate has been indicated to drastically reduce infection.^{11,12} Similarly, the uncharged blood group antigens (BGAs) are also known to be recognized by the enteric norovirus to mediate its infection.¹³ Thus, these molecules serve to engage the S1 subunit (residues 13–685) while the S2 (residues 686–1273) initiates the viral entry into the cells. Another important player in the recognition process is the aspartic acid residues that serve as a pH-dependent switch, which, when protonated, locks the S1B domain in the closed conformation and vice versa. They have been shown to provide a key structural transition across the physiological and endosomal pH by mediating the interprotomer movement of the two domains and the positioning of the S1B domain.¹⁴

Received: September 1, 2021

Accepted: December 6, 2021



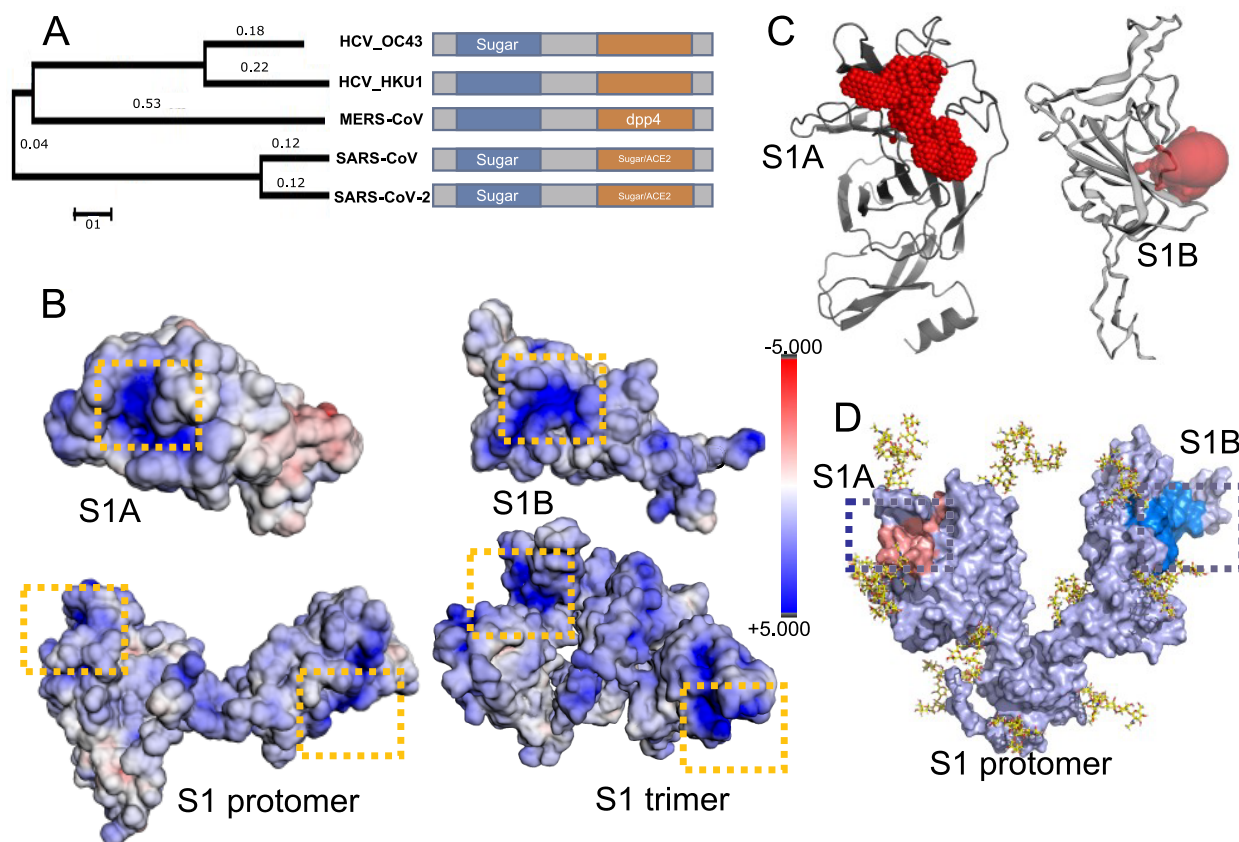


Figure 1. Sequence and structural characterization of the two domains of the S1 subunit of the spike protein. (A) Phylogenetic analysis of the five beta coronaviruses indicating their domain architecture and ligand preferences to infer the closest homologues of the SARS-CoV-2 spike protein. The neighbor-joining method was used to construct the tree using the MEGA software.²⁴ (B) Electrostatic surface rendering of the S1A and S1B domains, S1 protomer, and S1 trimer with the pockets lined with positively charged residues highlighted in the dotted box. (C) Results from the MetaPocket²² pocket prediction for the S1A domain and CASTp²³ pocket prediction for the S1B domain. Protein domains are shown in cartoon representation, with the identified pockets shown as spheres. (D) Modeling of the experimentally observed glycosylation (shown in stick representation) on the surface representation of the S1 subunit, SARS-CoV-2 spike protein (PDB ID: 6VSB), and the pockets, free from glycosylation highlighted in the dotted box; S1A binding site in pink and S1B binding site in blue.

The virus, therefore, appears to utilize elaborate strategies to orchestrate the binding to various receptors for its tissue tropism, in addition to its interaction with the ACE2 receptor, in order to gain access to the required host cell machinery. It is thus imperative to obtain a comprehensive outlook on the possible means of interaction between the virus and the various human cell receptors to fight this pandemic. At the same time, the evolution of SARS-CoV-2 has been a significant concern.¹⁵ A map of all the mutants seen to date has been seen to affect the spike proteins' receptor binding domain, which has served as the primary target of serum neutralizing activity of the vaccines designed. Reports have shown that a single mutation in the spike's receptor-binding domain (RBD; S1B domain, residues 319–541) or N-terminal domain (NTD; S1A domain, residues 14–304) was seen to reduce viral neutralization by polyclonal sera by as much as 10-fold.¹⁶ Studies have comprehensively mapped the mutations onto the RBD of the spike protein. Several of these mutations have been identified to increase the binding affinity to the ACE2 receptor. Drugs/antibodies designed to mimic or antagonize ACE2 binding or neutralization are thus susceptible to escape mutations led by the rapid appearance of resistance.¹⁶ Complementary experiments from the point of view of the interaction between the glycoprotein and the human lectins like DC/L-SIGN have also

been attempted to understand the lectin facilitated infection and immune exacerbation.^{17–19}

Here, we attempt to characterize the glycan-binding properties of the spike protein, which is less prone to mutation and has been shown to play a role in aiding its receptor binding ability. We have characterized its sugar-binding and affinity computationally. Our study envisions a model in which the SARS-CoV-2 virus first recognizes the sialosides on the cell surface via the S1A domain as it surfs over the cell surface and interacts with the GAGs through the S1B domain along with its binding to the ACE2 receptor. The S1 subunit thus facilitates the initial attachment to the receptors. This model also explains the relatively weaker affinity exhibited by the receptor to the sialosides compared to its binding to the GAGs, as evidenced in various experimental reports. Our molecular binding simulations of the complex corroborate this hypothesis. Through the study, we hope to understand the use of cell-surface sugars as a means of attachment factors indicating its role in extended tissue tropism, which act in conjunction with the ACE2 binding property of the receptor-binding domain to prompt the possibility of using carbohydrate mimetics as an adjunct therapeutic to treat the coronavirus infection.

Table 1. Ligands Used in the Study

sialosides	neu	Neu	
	aSia	aDNeu5Ac	
	bSia	bDNeu5Ac	
	NEU5GC	bDNeu5Gc	
	ACE9NEU	bDNeu9Ac	
	ACE7NEU	bDNeu7Ac	
	ACE4NEU	bDNeu4Ac	
	KDN	deaminated neuraminic acid	
	23SIA	aDNeu5Ac(2→3)bDGal	
	26SIA	aDNeu5Ac(2→6)bDGal	
	26SIL	aDNeu5Ac(2→6)bDGal(1→4)bDGlc	
	23SIL	aDNeu5Ac(2→3)bDGal(1→4)aDGlc	
	S26GalNAc	aDNeu5Ac(2→6)aDGalNAc	
	diNANA	aDNeu5Ac(2→8)aDNeu5Ac	
	sLeX	aDNeu5Ac(2-3)bDGal(1-4)[aFuc(1-3)]GlcNAc	
	gangliosides	GM1	bDGalp(1-3)bDGalNAc[aNeu5Ac(2-3)]bDGalp(1-4)bDGlc
		GM1B	aNeu5Ac(2-3)bDGalp(1-3)bDGalNAc(1-4)bDGalp(1-4)bDGlc
GD1A		aNeu5Ac(2-3)bDGalp(1-3)bDGalNAc(1-4)[aNeu5Ac(2-3)]bDGalp(1-4)bDGlc	
GD1B		bDGalp(1-3)bDGalNAc(1-4)[aNeu5Ac(2-8)aNeu5Ac(2-3)]bDGalp(1-4)bDGlc	
GT1B		aNeu5Ac(2-3)bDGalp(1-3)bDGalNAc(1-4)[aNeu5Ac(2-8)aNeu5Ac(2-3)]bDGalp(1-4)bDGlc	
GT1C		bDGalp(1-3)bDGalNAc(1-4)[aNeu5Ac(2-8)aNeu5Ac(2-8)aNeu5Ac(2-3)]bDGalp(1-4)bDGlc	
GQ1C		aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-3)bDGalNAc(1-4)[aNeu5Ac(2-8)aNeu5Ac(2-3)]bDGalp(1-4)bDGlc	
glycosaminoglycans	Hept	aLIdoA(1→4)aDGlc(1→4)aLIdoA(1→4)aDGlc	
	Hepsul	aDGlcNAc(1→4)bDGlcA	
	Hep	aLIdoA(1→4)aDGlc	
blood group antigens	Aant-I	aLFuc(1→2)[aDGalNAc(1→3)]bDGal(1→3)bDGlcNAc(1→3)aDGal	
	Aant-II	aLFuc(1→2)[aDGalNAc(1→3)]bDGal(1→4)bDGlcNAc(1→3)aDGal	
	Aant-III	aLFuc(1→2)[aDGalNAc(1→3)]bDGal(1→3)aDGlcNAc(1→3)aDGal	
	Aant-IV	aLFuc(1→2)[aDGalNAc(1→3)]bDGal(1→3)bDGalNAc(1→3)aDGal	
	Bant-I	aLFuc(1→2)[aDGal(1→3)]bDGal(1→3)bDGlcNAc(1→3)aDGal	
	Bant-II	aLFuc(1→2)[aDGal(1→3)]bDGal(1→4)bDGlcNAc(1→3)aDGal	
	Hant-I	aLFuc(1→2)bDGal(1→3)bDGlcNAc(1→3)aDGal	
	Hant-II	aLFuc(1→2)aDGal(1→4)bDGlcNAc(1→3)aDGal	

RESULTS AND DISCUSSION

Preliminary Characterization of the S1 Subunit of the Spike Protein. Herein, we attempt to characterize the domains in the S1 subunit of the spike protein to delineate the various carbohydrate-binding sites through sequence analysis. To date, seven coronavirus species have been known to be human pathogens. Two of them are alphaviruses: the HCoV-229E and the HCoV-NL63l ACoV genus. They are endemic and cause mild respiratory symptoms, similar to the HCoV-HKU1 and HCoV-OC43, which are beta-coronaviruses. However, the other three beta-coronaviruses have caused severe outbreaks: SARS-CoV-1, MERS-CoV, and the current SARS-CoV-2.²⁰ The proteins from the beta-coronaviruses have been compared owing to their high sequence similarity, and indeed their S1 domain is found to be similar to the SARS-CoV-1 and MERS-CoV domains. Despite lacking significant structural similarities, the regions around residues 58–70 and 255–265 display high sequence similarity scores. The hallmarks of sialic-acid-binding pockets are the presence of a shallow pocket lined by positively charged amino acids that can accommodate the negatively charged sugar molecules. This pocket prediction is in line with the previous reports where surface iso-electron density mapping predicted a sialoside binding pocket²¹ (Figure 1, Figure S1, Table S1)

In the S1A domain, we find such a shallow pocket at the tip of the structure. Metapocket²² server predicted a pocket with residues 64–72, 95–100, 145–152, 180–186, and 245–266

CASTp²³ failed to identify this shallow pocket. The S1B domain of SARS-CoV-2 is believed to have characteristics of the heparin-binding region. It shares ~52% sequence identity with the spike protein from SARS-CoV-1, and both the proteins have been known to bind heparan sulfate. HCoV-NL63 and SARS-CoV-1 use GAGs for adhesion on the host cell and the ACE2 receptor. Studies have revealed that GAGs binding usually occurs through electrostatic interactions between a cluster of positively charged basic residues. CASTp server, which maps the surface topography of the proteins, revealed a positively charged pocket with a volume of 377 Å³. The residues in the predicted pocket include 353–355, 394–398, 426–431, 462–464, and 512–518. The Metapocket server, in the case of S1B, did not identify this pocket. Sequence analysis of this domain revealed an insert in the amino acid sequence (XRKRXXNX) that follows the Cardin Weintraub GAGs binding motif (XBBBXXBX) where B represents positively charged amino acid; this difference between the homologues suggests that these viruses have different binding preferences/affinities toward the glycosaminoglycans. We next calculated the electrostatic potential map of the S1 subunit, which, as expected, revealed an extended electropositive surface constituting extended loops at both the predicted pockets of S1A and the S1B domain. It is important to note that these putative carbohydrate regions are adjacent to the ACE2 binding site on the S1B domain, suggesting a possibility of simultaneous binding of both the ACE2 receptor

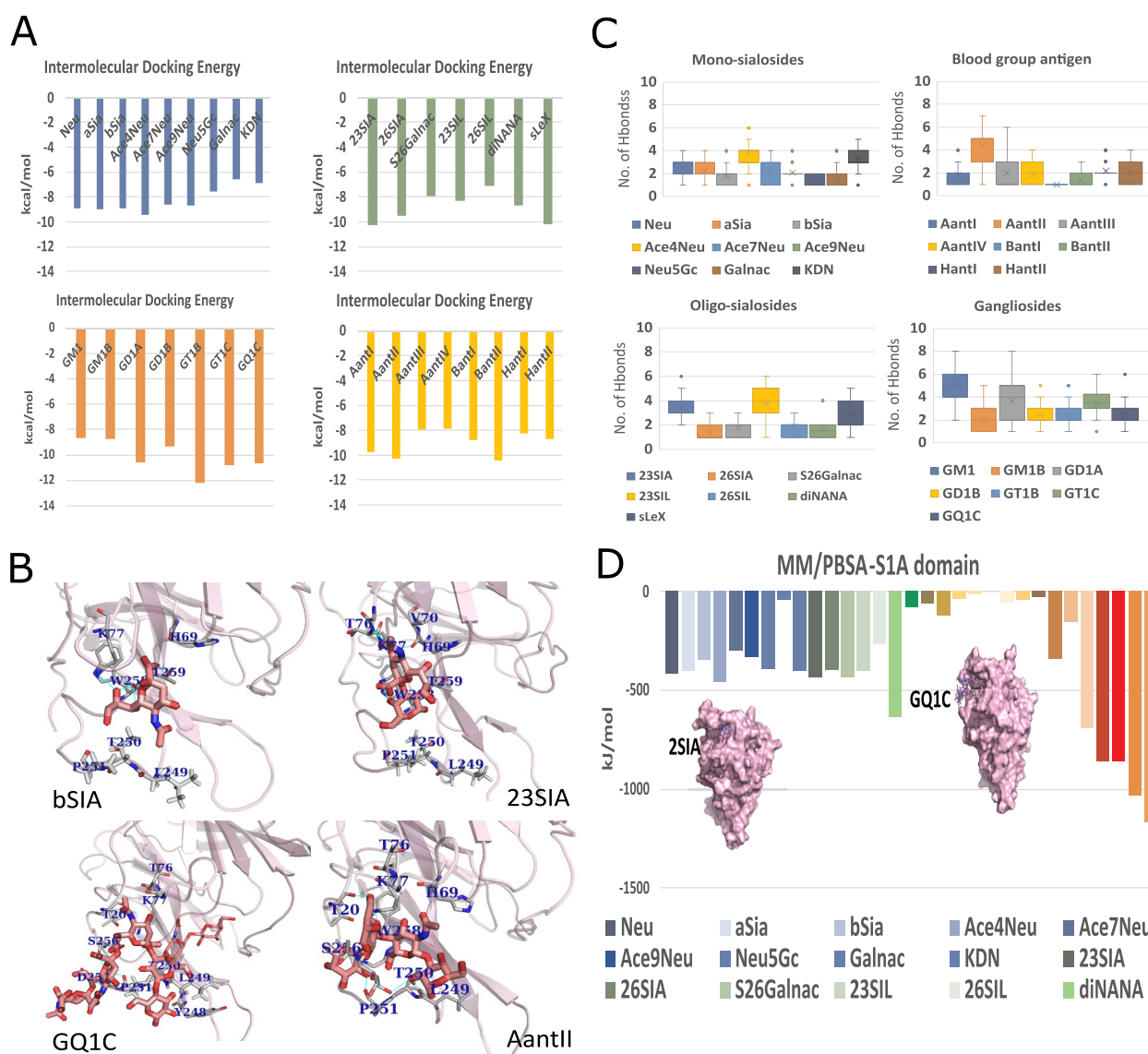


Figure 2. Docking and simulation studies of the S1A domain with the various saccharides. (A) Intermolecular docking energy from the molecular docking studies for all the ligands of the various complexes with the S1A domain. The set is subdivided into various ligand subtypes—sialosides (mono and oligo), gangliosides, and blood group antigens. (B) Cartoon representation of the four S1A domain complexes highlighting the receptor residues involved in hydrogen bonds and hydrophobic interactions (shown as sticks) with ligands in orange stick representation and hydrogen bonds shown in cyan (Figure S3 includes detailed Ligplot representations for all S1A complexes). (C) Average number of intermolecular hydrogen bonds as indicated by box plots for the S1A-ligand complexes subdivided according to the ligand subtypes. The best binders were seen to have a higher number of hydrogen bonds. (D) Binding energy calculation using MMPBSA analysis of the various complexes with the S1A domain. The energy values are in kJ/mol. Inset: Surface representation of the protein with the best binding ligand in stick representation.

protein and the cell-surface sugars. We further accessed the possible glycan cover of the respective domain, which can shield the predicted pockets, preventing their accessibility. Spike proteins are known to be heavily glycosylated, and even a light coverage of the glycans on the spike protein can hinder the accessibility of the pocket. It was observed that the residues in the S1 subunit that are predicted to be glycosylated are 17, 61, 74, 122, 149, 165, 234, 282, 331, and 343, and the pockets were seen to be glycosylation-free, allowing an unhindered binding of the carbohydrate ligands to the S1 subunit (Figure 1).

To understand the interaction capacity of both the domains of the spike protein with the cell-surface receptors, we employed a two-step approach of molecular docking and simulation to selectively dock the representative sialosides,

gangliosides, glycosaminoglycans, and the blood group antigens to the respective domains of the spike protein.

Screening of the S1A Domain. The S1A domain (residues 14–301) constitutes the predicted sialoside binding pocket seen to be formed by amino acid residues Leu 18–Gln 23, His 66–Thr 78 of the $\beta 4$ – $\beta 5$ loop, and Gly 252–Ser 254 of the $\beta 14$ – $\beta 15$ loop (Figure S2). We selectively docked the representative human sialosides listed in Table 1 onto the modeled structure of the S1A domain (Figure S2). We employed a blind docking strategy to ensure no bias in choosing the sialoside binding pocket. One-hundred conformations were evaluated for each ligand. The best conformation was determined based on cluster analysis, cluster rank, lowest binding energy, and the number of conformations in each cluster (Figure 2, Figure S2). Though a blind docking

Table 2. MMPBSA Analysis for the S1A Domain Complexes

(A) monosialosides									
kJ/mol	Neu	aSia	bSia	Ace4Neu	Ace7Neu	Ace9Neu	Neu5Gc	Galnac	KDN
Van der Waals energy	-49.695	-56.582	-39.306	-74.145	-29.078	-63.582	-42.115	-18.43	-61.191
electrostatic energy	-580.69	-527.16	-453.632	-635.628	-352.292	-339.75	-467.63	-26.81	-475.81
polar solvation energy	224.756	192.861	151.388	264.236	89.182	80.422	130.168	5.32	144.579
SASA energy	-9.165	-9.045	-6.289	-11.094	-5.39	-8.743	-7.194	-2.83	-8.82
binding energy	-414.97	-399.45	-345.544	-455.009	-297.307	-331.771	-388.33	-42.73	-402.69
(B) oligosialosides									
kJ/mol	23SIA	26SIA	S26Galnac	23SIL	26SIL	diNANA	sLeX		
Van der Waals energy	-92.773	-75.269	-98.909	-90.884	-62.199	-46.663	-92.12		
electrostatic energy	-549.23	-464.25	-556.288	-575.028	-334.009	-716.961	-117.91		
polar solvation energy	223.535	159.633	235.175	279.462	139.69	137.515	144.734		
SASA energy	-12.662	-12.463	-13.795	-13.202	-8.887	-7.389	-12.845		
binding energy	-433.06	-393.17	-434.407	-400.503	-265.902	-637.133	-77.631		
(C) blood group antigens									
kJ/mol	Aant-I	Aant-II	Aant-III	Aant-IV	Bant-I	Bant-II	Hant-I	Hant-II	
Van der Waals energy	-75.23	-142.8	-42.028	-49.51	-33.498	-86.273	-98.877	-76.745	
electrostatic energy	-104.25	-216.19	-55.853	-33.736	-38.656	-106.81	-79.8	-85.172	
polar solvation energy	122.647	257.29	70.916	79.318	82.333	146.19	154.494	141.49	
SASA energy	-10.673	-18.407	-8.011	-8.31	-6.103	-11.949	-15.3	-10.512	
binding energy	-63.924	-120.84	-33.326	-14.153	3.937	-58.606	-39.739	-30.419	
(D) gangliosides									
kJ/mol	GM1	GM1B	GD1A	GD1B	GT1B	GT1C	GQ1C		
Van der Waals energy	-119.03	-38.715	-145.707	-87.833	-88.008	-131.961	-157.55		
electrostatic energy	-485.36	-211.99	-829.505	-1053.24	-1052.07	-1208.32	-1289.3		
polar solvation energy	279.429	97.991	309.601	300.089	301.187	335.268	299.496		
SASA energy	-18.557	-4.582	-22.037	-15.695	-15.642	-20.814	-23.101		
binding energy	-341.83	-155.48	-690.003	-860.067	-856.84	-1029.81	-1167.8		

strategy was used, ligands docked into the putative sialoside binding pocket were identified in the previous section. Our studies revealed 2,3-linked sialoside to being a better binder based on the lowest binding energy. The next best binder was sialyl Lewis X (sLeX), demonstrating the S1A domain's ability to accommodate a larger sialoside. The list of interacting sites of the docked sialosides is tabulated (Table S2). The presence of crucial hydrogen bonds, hydrophobic interactions, and salt bridges with these sialosides implies the possibility of a physiological interaction with this domain of the spike protein. Neu5Gc, on the other hand, occupied a slightly different region of the pocket aided by residues Pro 251, Ser 256, and Trp 258, in addition to the Lys 77, which suggests a possible broad range of selectivity enabling interaction with various forms of the sialosides. Additionally, similar to the sialic-acid binding lectins, siglecs, the interaction site at the S1A domain contains a lysine (K77) residue that is seen to be essential for interaction with sialic acid and sialic-acid-containing ligands.²⁵ This was validated in our docking analysis with the best binders where the mutant K77A had a considerable loss in interaction energy (Figure S2)

We next performed molecular dynamics simulations to verify further and understand the sialosides and S1A domain interaction. The system's stability was determined by the RMSD values of protein backbone atoms. The average RMSD values of sialoside complexes were ~ 0.25 nm. Residue-wise, RMSF analysis highlights the flexibility of the binding site, wherein the highest mobility is caused by the residues forming structural loops of the binding pockets. In particular, the residues in the $\beta 14$ – $\beta 15$ loop impart an extended binding site allowing the accommodation of larger sugars. On interaction,

the predicted binding site witnessed a considerable loss in atomic fluctuations of the residues lining the binding site. Sialosides offered the most rigidity to the binding pocket in comparison to the blood group antigens (Figure 2). Detailed data about the lifetime of binding events are reported in the Supporting Information (Figure S2, Table S3).

Subsequently, we performed the binding free energy calculations through the MM-PBSA method for all of the complexes. The different types of energies which contribute to binding, such as van der Waals, electrostatic, polar solvation, and SASA, were calculated; the average values of these energies are depicted in Table 2. In the case of sialo-monosaccharides, N-acetyl neuraminic acid was found to be a better binder in comparison to the nonacetylated version. Differential acetylation was also seen to play a role in the interaction with the protein. The interacting residues were Thr 20, His 69, Gly 75, Lys 77, Leu 249, Thr 250, Pro 251, Ser 256, Gly 257, Trp 258, and Thr 259. In sialo-oligosaccharides, the simulation analyses also revealed 2,3-linked sialic acids to a better binder based on the time the ligand spends interacting with the domain. Additionally, disialic acid with a 2,8-linkage, diNANA, was the best binder with a binding energy of -637.13 kJ/mol. The residues Thr 19, Arg 21, Thr 22, Val 70, Tyr 145, Trp 152, Ser 247, and Gly 252 were seen to offer interaction in addition to the ones previously mentioned. Average free binding energy was found to be $-341.97 + 40$ kJ/mol and $-377.40 + 71$ kJ/mol for mono and oligo-sialosides, respectively. In the case of diNANA, the binding energy roughly scales to the additive number of the individual sialic-acid binding energies.

In addition to its interaction with the sialosides, higher glycans like blood group antigens, GAGs, and gangliosides

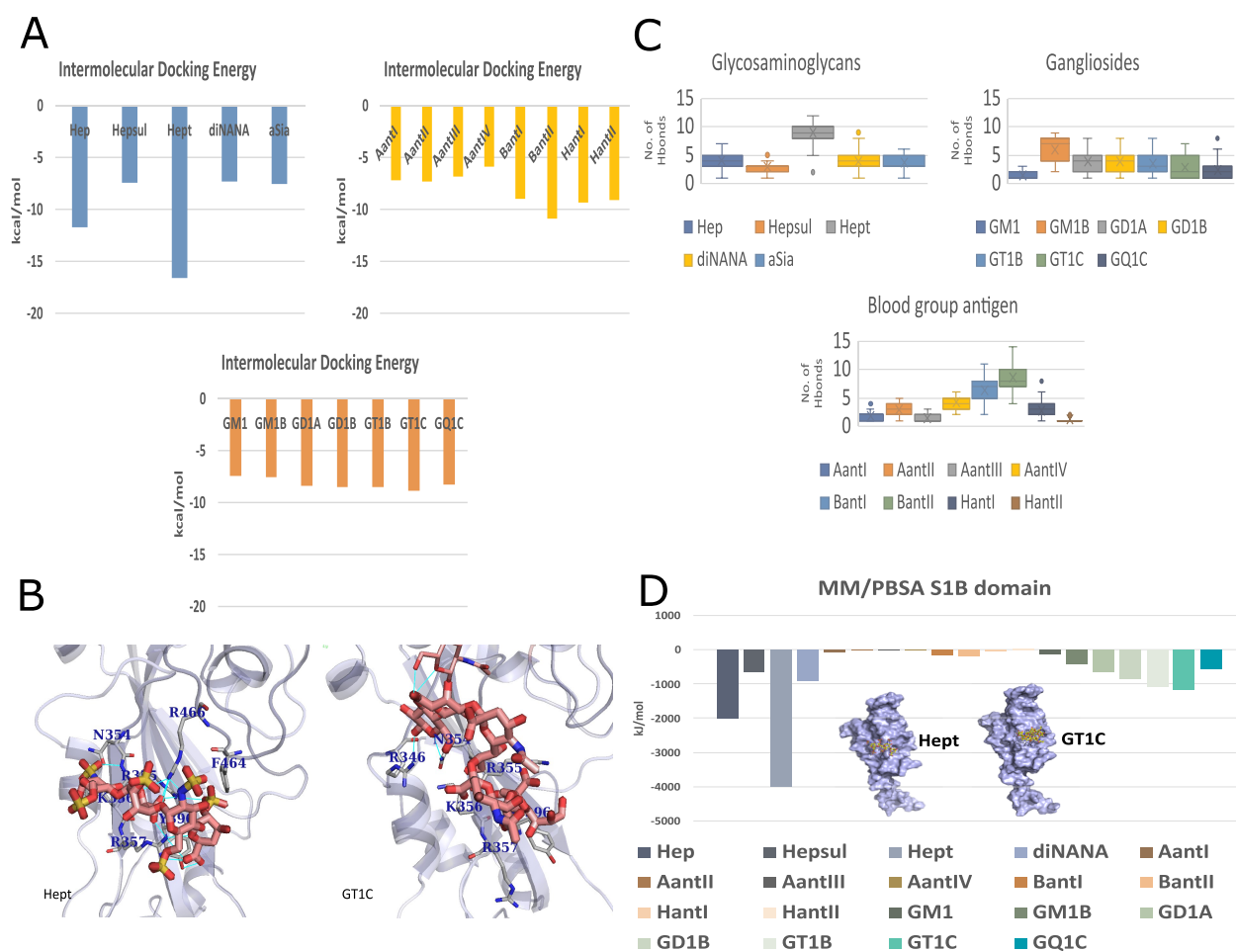


Figure 3. Docking and simulation studies of the S1B domain with the various saccharides. (A) Intermolecular docking energy from the molecular docking studies for all the ligands of the their complexes with the S1B domain. The set is subdivided into various ligand subtypes—glycosaminoglycans, gangliosides, and blood group antigens. (B) Cartoon representation of the two S1B domain complexes highlighting the receptor residues involved in hydrogen bonds and hydrophobic interactions (shown as sticks) with ligands in orange stick representation and hydrogen bonds shown in cyan (Figure S5 includes detailed Ligplot representations for all S1B complexes). (C) Average number of intermolecular hydrogen bonds as indicated by box plots for the S1B-ligand complexes subdivided according to the ligand subtypes. The best binders were seen to have a higher number of hydrogen bonds. (D) Binding energy calculation using MMPBSA analysis of the various complexes with the S1B domain. The energy values are in kJ/mol. Inset: Surface representation of the protein with the best binding ligand in stick representation.

were also investigated. In the case of gangliosides, the interaction energy was seen to proportionately increase with the increase in the number of glycone units, with GQ1C having the least binding energy of -1167.8 kJ/mol. Blood group A antigen, type II, in particular among the blood group antigen, was found to be the best binder with an energy of -120.84 kJ/mol. The residues Gly 72, Thr 73, Asn 74, Glu 180, Gly 181, and Gln 182 were also involved in interacting with BGAs. The average free binding energy was found to be -44.63 kJ/mol, which was considerably lower than the sialosides. Taking the poor binding energies into account, the BGAs can be ignored for the purposes of determining their role in SARS-CoV-2 infection (Figure 2, Figure S3).

Furthermore, we observed various binding and unbinding events of the various sialosides tested. In particular, the unbinding event of the 2,6-linked sialosides was higher in comparison to their 2,3-coupled counterparts. As observed in the docking studies, the binding and recognition of sialosides in the S1A domain is not unique, probably due to the low affinity and the avidity of the interaction. The ligands that did not occupy the pocket for at least 5 ns were treated as false

positives and were not considered for further analysis. Despite the ability to bind to different sialosides, the residues participating in the interaction were fairly similar, indicating a similar binding site for the various ligands. The residues include Thr 19, Thr 20, Arg 21, Thr 22, His 69, Val 70, Gly 72, Thr 73, Asn 74, Gly 75, Lys 77, Tyr 145, Trp 152, Glu 180, Gly 181, Gln 183, Ser 247, Leu 249, Thr 250, Pro 251, Gly 252, Ser 256, Gly 257, Trp 258, and Thr 259. The residue Lys 77 was seen to be very crucial in its interaction with the sialosides. Some of the studies that have not used this region could have missed the signal in their experiment.

Screening of the S1B Domain. The binding of the GAGs is known to occur to the S1B domain (residues 319–541) of the spike protein. Microarray binding experiments have indicated an extensive interaction with the heparan sulfate oligosaccharide and the absence of binding to sialic-acid-like molecules in this region.²⁶ The heparin-binding site is seen to be a groove formed by loops $\beta 1$ – $\beta 2$, lined by the residues in the antiparallel β sheet, $\beta 7$ and $\beta 12$, and the loops $\beta 5$ – $\beta 6$ and $\beta 11$ – $\beta 12$ and a loop between $\alpha 3$ and $\beta 5$ (Figure S4). The binding groove is comprised of four basic residues, Arg 355,

Table 3. MMPBSA Analysis for the S1B Domain Complexes

(A) glycosaminoglycans								
kJ/mol	Hept	Hepsul	Hep	diNANA	aSia			
Van der Waals energy	-69.207	-58.186	-55.397	-51.85	-36.537			
electrostatic energy	-5097.65	-953.06	-2501.3	-1184.2	-840.17			
polar solvation energy	1164.89	362.28	533.443	340.528	307.95			
SASA energy	-14.162	-9.636	-9.413	-8.219	-8.007			
binding energy	-4018.12	-657.31	-2027.1	-905.68	-573.05			
(B) gangliosides								
kJ/mol	GM1	GM1B	GD1A	GD1B	GT1B	GT1C	GQ1C	
Van der Waals energy	-12.052	-114.45	-43.337	-64.539	-95.843	-76.68	-112.13	
electrostatic energy	-204.685	-571.02	-766.27	-1086.6	-1354.6	-1362.1	-1236.4	
polar solvation energy	83.618	263.17	144.904	321.289	362.707	182.27	186.53	
SASA energy	-3.093	-19.257	-6.436	-11.053	-14.848	-12.027	-16.83	
binding energy	-139.97	-441.28	-672.13	-841.18	-1092.9	-1257.6	-1178.6	
(C) blood group antigens								
kJ/mol	AantI	AantII	AantIII	AantIV	BantI	BantII	HantI	HantII
Van der Waal energy	-92.374	-87.85	-15.081	0	-151.46	-137.16	-57.889	-16.078
electrostatic energy	-120.016	-143.19	-17.291	-0.265	-307.13	-387.6	-115.24	-8.539
polar solvation energy	161.502	226.531	38.988	22.776	309.57	357.55	136.716	54.202
SASA energy	-11.717	-13.653	-2.088	0.18	-20.032	-21.1	-9.77	-1.79
binding energy	-63.76	-20.908	4.432	22.801	-170.09	-189.86	-49.52	29.2

Lys 356, Arg 357, and Arg 466, which have been predicted to be essential for binding the negatively charged heparin. These residues form a highly positively charged patch which suggests that the binding is most likely to be mediated by the nonspecific electrostatic interactions. To obtain insights into how GAGs and GSLs bind to the spike protein, we conducted docking and simulation studies using representative GAGs and GSLs. Pocket prediction studies led to the prediction of a site where heparin binds on the S1B domain. In the case of the trimer, this pocket is partially obscured in the closed conformation but wholly exposed in the open state. We selectively docked the ligands listed in Table 1 onto the modeled structure of the S1B domain (Figures S1, S4). We conducted docking simulations on similar lines as mentioned for the S1A domain. The best conformation from the 100 docked conformers was chosen based on the lowest binding energy and statistical mechanical analysis (Figure 3, Figure S4, and Table S4).

Heparin and di- and tetrasaccharide were found to be the best binders based on the intermolecular interaction energies of -16.61 and -11.73 kcal/mol, followed by the gangliosides GT1C, -8.82 kcal/mol, which were seen to be dependent on the sialic acid at the termini. On the basis of these results, we also conducted docking studies of the diNANA, 2,8-linked disialic acid, and sialic acid monomer to study its interaction with the S1B domain, which had weaker interaction energies. In an attempt to decipher the blood group specificity, we also docked the S1B domain with the blood group antigens, where the type II saccharides had better energy scores in comparison to their type I counterparts (Figure S4).

To further verify and understand the ligand spike protein interaction, we next performed molecular dynamics simulations. The system stability was again evaluated by the RMSD. GAGs and the GSLs were seen to have reduced deviation in comparison to the BGAs (average RMSD nm, respectively). In the case of GAGs, the hydrogen bond and interaction fingerprint analysis of the heparin complex reveals its continued binding through hydrogen bonding interactions

(with over 90% occupancy throughout the MD trajectories) to the basic residues of the S1B domain. The differences in the binding of di- and tetra-heparin molecules highlighted in this present study are in accordance with the results of previous experimental studies that indicate more favorable binding to specific lengths and sulfation levels of GAGs.⁹ We also observed heparin to employ induced-fit binding during the course of the trajectory. The complementarity of the interactions and the change in width of the basic groove in both the closed and open spike bound to heparin suggest an induced fitting upon binding that results in the well-defined partially grooved basic path where heparin lies (Figure 3, Figure S4). MM/PBSA analysis of the heparin, di- and tetrasaccharide, and the heparan sulfate S1B domain revealed the average free binding energy to be -3022.64 ± 55 kJ/mol for the heparin complex and -657 ± 15.26 kJ/mol for the heparan sulfate complex. The residues seen to be crucial for interacting with the GAGs were found to be Arg 355, Lys 356, Arg 357, and Arg 466. The gangliosides, however, were seen to occupy a site adjacent to the heparin-binding region, which is seen to be neutral on the electrostatic map of the protein as predicted by the CASTp server. We also noted the role of the 2,8-linked terminal of the GT1C ganglioside playing a role in interaction, imparting a better binding affinity to the complex. The most preferred residues for GSL binding are as follows: Arg 346, Asn 354, Arg 355, Lys 356, Arg 357, Asn 394, Tyr 396, Asp 428, Arg 457, Lys 458, Lys 462, Pro 463, Phe 464, Glu 465, Arg 466, Ile 468, Ser 469, Thr 470, Ser 514, Phe 515, Glu 516, Leu 518, and His 519. The BGAs were seen to share similar binding pockets but had weaker affinities than the GSLs. We observed a clear distinction wherein blood group type II was a better binder than type I antigen; however, they demonstrated a better interaction with the S1A domain than the S1B domain (Figure 3, Figure S4, Figure S5, Table 3, Table S4).

Reports have previously shown the role of Lys 417, Asn 487, Tyr 489, Gln 493, Tyr 449, Gly 446, Thr 500, Asn 501, and Gly 502 residues of the S1B domain in ACE2 recognition and

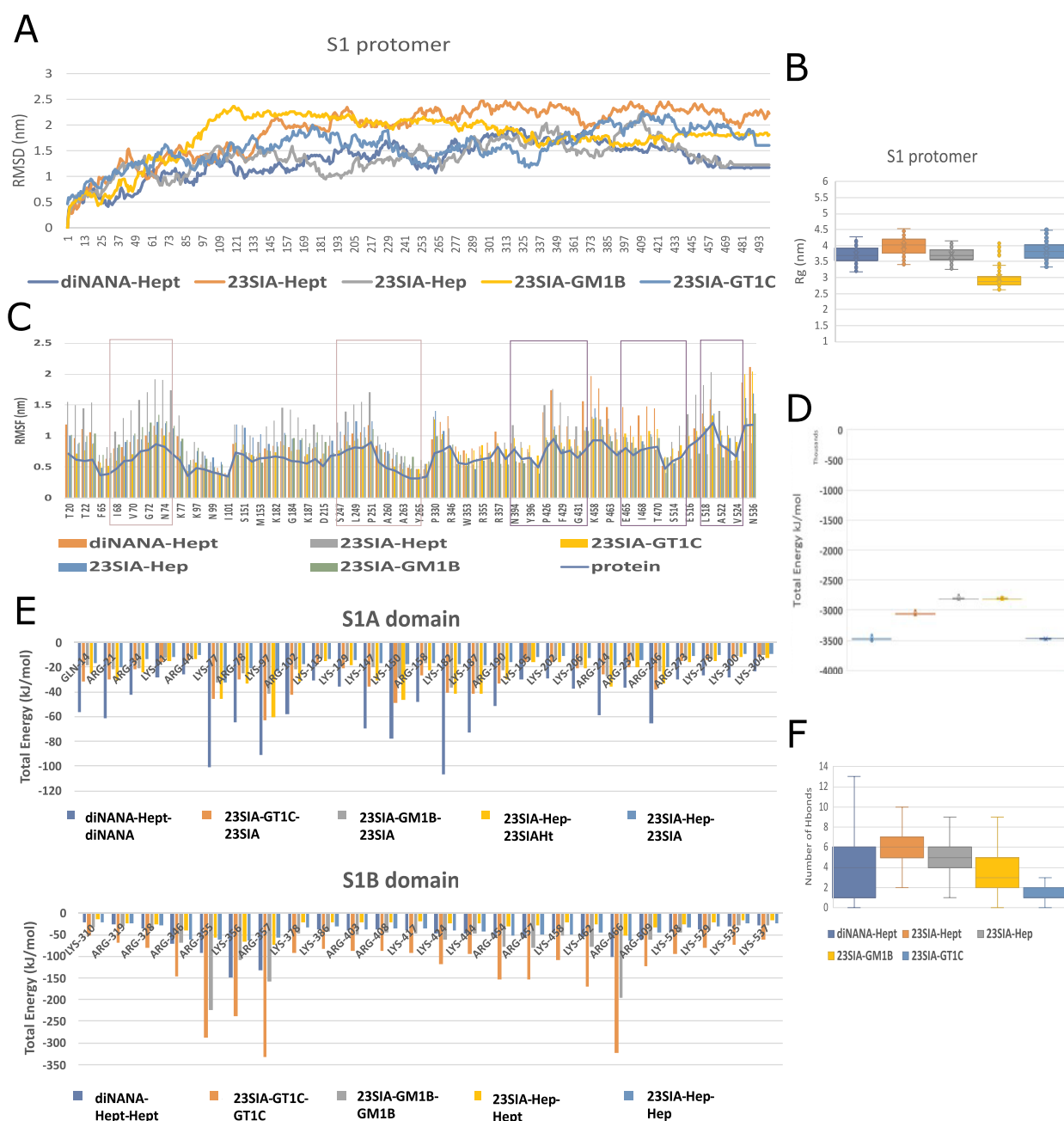


Figure 4. Simulation studies of the S1 subunit with the various saccharides. (A) Average RMSD analysis of the various complexes of the S1 subunit and the protomer with the distribution plots indicating the range sampled. The five complexes are color-coded. (B) The radius of gyration plot for the five complexes of the S1 subunit and protomers for 50 ns of time with the distribution plots indicating the range sampled. (C) RMSF analysis for the $C\alpha$ atoms of the S1 subunit for the five complexes in comparison to the free protein highlighting the reduction in fluctuation offered by the ligand interaction. The binding site residues are indicated in the boxes. (D) The total energy of the five complexes of the S1 protomer color-coded across the 50 ns simulation. (E) Average number of intermolecular hydrogen bonds as indicated by box plots for the S1 protomer for the five complexes subdivided according to the ligand subtypes. The best binders were seen to have a higher number of hydrogen bonds. (F) Decomposition of the energetic contribution of the binding site residues to ligand binding. The complexes are color-coded. The energy values are given in kJ/mol. The residue-wise energy is calculated using a free energy decomposition scheme using `g_mmpbsa` tool.

binding. Hence, it corroborated that heparin and ACE2 can bind at adjacent loci of the S1B domain simultaneously and in a cooperative manner. These results suggest a mechanism in which the S1B domain can independently interact with the surface glycosaminoglycans in conjunction with its interaction with the ACE2 domain. This binding is, however, dependent on the open conformation of the S1B domain.^{27–29}

Analysis of the S1 Subunit. The interaction data from individual simulations presented the necessary screen to evaluate the best binders in the context of simultaneous occupancy of the respective ligands to the two domains in their monomer and trimeric forms. To further investigate the binding of these cell-surface glycans, we modeled five systems with the S1 subunit of the spike protein, a single protomer with sialosides occupying the S1A domain, and the GAGs

occupying the S1B domain. From the docking and simulation studies of the individual domains, we screened for the most suitable saccharide with good binding affinity and low RMSD for the critical amino acids, indicative of relatively favorable binding energies. These molecules were thus docked further to the protomer and trimer complexes. The aim was to obtain convergence of structural clustering and the free energy of binding and analyze the maintenance of the native contacts between the glycans and the target domains. MD simulations for the S1 protomer involved the following combinations: 23SIA-Hep, 23SIA-Hept, 23SIA-GM1B, 23SIA-GT1C, and diNANA-Hept were conducted for 50 ns. The complexes were chosen and designed based on the binding energy preferences where diNANA-Hept was the best binder from our analysis and the previous experimental reports where 2,3-linked sialic acid and Hep have the maximum cellular occurrence.

Here, the RMSD of the complexes relative to each starting structure cannot be taken as a stability measure owing to the inherent flexibility offered by the hinge region, which introduces considerable fluctuations in the entire complex. The evidence of the structural stabilization of the sugar-binding pockets was provided by the RMSF, which tended to lower values for the residues in the S1B domain vis-à-vis S1A domain and the hinge connecting the two domains (Figure 4, Figure S6). Representative structures from the clustering analysis reveal that the heparin ligand maintains its alignment to the extended positively charged area of the S1B domain, as witnessed in our independent simulations. The MM-PBSA analysis revealed a similar pattern observed in the individual domain simulations. Sialosides exhibited a weaker binding than the glycosaminoglycans and the gangliosides (Table 4). These

Table 4. MMPBSA Analysis for the S1 Subunit Complexes

binding energy	S1A domain	S1B domain
diNANA-Hept	-996.391 ± 43.584 kJ/mol	-1246.155 ± 52.947 kJ/mol
23SIA-Hept	-522.267 ± 18.801 kJ/mol	-1258.782 ± 24.144 kJ/mol
23SIA-Hep	-285.431 ± 35.673 kJ/mol	-1913.800 ± 27.995 kJ/mol
23SIA-GT1C	-563.118 ± 5.963 kJ/mol	-570.308 ± 64.290 kJ/mol
23SIA-GM1B	-377.654 ± 24.122 kJ/mol	-529.669 ± 17.235 kJ/mol

results and the interaction pattern demonstrated the ability of the different cell surface glycan molecules to cooperatively and simultaneously interact with both the domains of the protomer (Table S6). We next ventured to understand the dynamics in the trimer to check if the results from the protomer simulation hold good for the trimer.

Analysis of the S1 Trimer. Experiments using transmission electron microscopy (TEM) have shown the presence of a significant fraction of the S1B domains in the one-up conformation followed by three-up and two-up conformations in the sample incubated with the ACE2 receptor and the spike protein.³⁰ We next conducted MD simulations for the S1 subunit trimer in the following combinations: 1-up S1B domain and 2-up S1B domain conformation with 2,3-sialic acid ligand in all three protomers and the heparin tetrasaccharide ligand in the protomers of the opened/up conformation for 50 ns each. We also conducted simulations with both the complexes in the absence of the ligands. In the 1-up and 2-

up states, the apo trimer was seen to be less stable, indicating its transient form as an “up or open” conformation in the absence of the cell-surface sugars or the ACE2 binding protein in comparison to its holo forms. The trimer exhibited lower energy values in comparison to its protomer counterpart. The complex also had lower RMSF fluctuations in the regions interacting with the ligand, and the overall stability of the complex was higher in the trimer conformation. We also observed a higher RMSF at the ends of the S1B domains in the open/up conformations. These residues were seen to correlate with the residues in the RBM, which are known to interact with the ACE2 receptor. It is also noted that the S1B domain in the down/closed conformation lacked this behavior, implicating the role of open/up conformation in the spike protein–host cell ACE2 receptor interaction (Figure 5, Figure S6).

It is important to note that physiologically the spike protein is extensively glycosylated (~22 glycosylation sites), which often enables its escape from experimental characterization. Beyond being structurally necessary to bind ACE2, this glycan shield allows the virus to elude the host immune response.³¹ Notably, none of the binding residues are close to these glycosylation sites. Sugar-binding pockets are unencumbered and free of glycosylation and therefore capable of interacting with the host-cell sugars. We further analyzed the conformational space of the glycans from the glycosylated asparagine residues. They were indeed observed to fill the void created by the opening of the protomers, supporting the S1B domain in its up conformation (Figure S6).

The ability of the S1 domain to interact with the surface carbohydrates suggests that the spike protein from the SARS-CoV-2 virus might have evolved independently to recognize both sialosides and the glycosaminoglycans, aiding in its interaction with the ACE2 binding and the latching onto the host cell surface. Differential distribution of the sialosides on the various organs and varied ACE2 expression would explain differential infectivity and transmission. The acquired ability of the spike protein to engage diverse glycans on the host cell might be the reason for its broad tissue tropism and high rate of infectivity.

Effect of the Spike Variants on Its Sugar-Binding Ability. Several variants have emerged in recent times. These have accumulated one or more nonsynonymous mutations in the spike protein wherein the nucleotide mutation alters the amino acid, enabling their transmission and resistance to several neutralizing antibodies. Amino acid substitutions alter the epitopes; increase receptor binding avidity; change glycosylation, deletions, and insertions; and induce allosteric structural effects. We, in this study, have explored the specific mutations for the variants of concern (WHO), for example, alpha variant (Del 69–70, Del 144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H), beta variant ((L18F), D80A, D215G, (Del 242–244), K417N, E484K, N501Y, D614G, A701 V), gamma variant (L18F, T20N, P26S, D138Y, R190S, K417N/T, E484 K, N501Y, D614G, H655Y, T1027I, V1176F), delta variant (L452R, T478 K, D614G, P681R), delta plus variant AY.1 (K417N, L452R, T478 K, D614G, P681R), and delta plus variant AY.2 (A222 V, K417N, L452R, T478 K, D614G, P681R), which have been implicated to have altered binding affinity and new interprotein contacts that perhaps change the internal structural dynamics to increase their binding to the host cell and eventually the infectivity^{15,32,33} (Table 5, Figure 6).

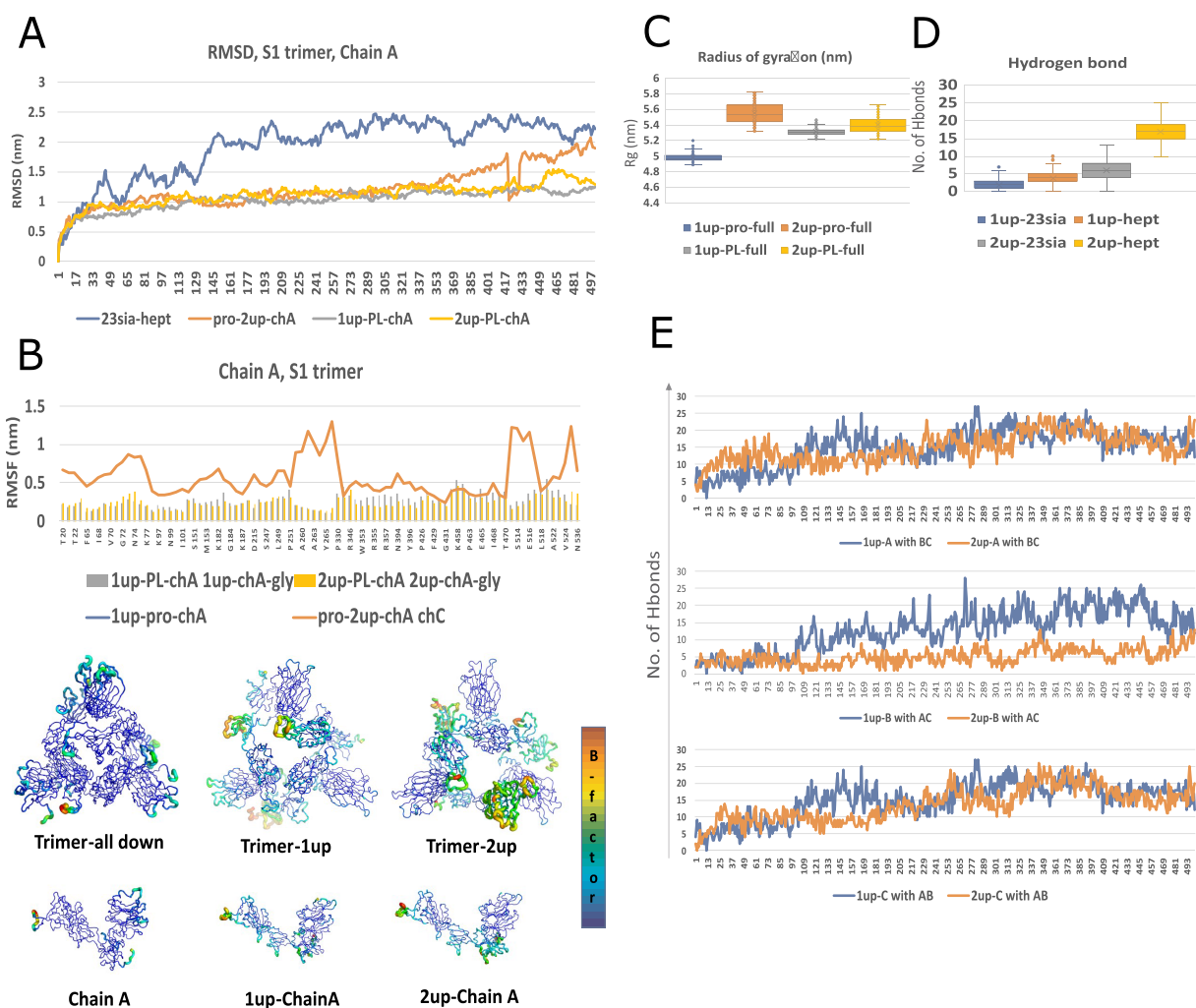


Figure 5. Simulation studies of the S1 trimer with sialosides in the S1A domain and a glycosaminoglycan in the S1B domain. (A) Average RMSD analysis of chain A of the protomer and holo and apo trimers in the 1-up and 2-up conformations are color-coded. The presence of the carbohydrate and the trimer presentation is seen to stabilize the systems. (B) RMSF analysis for the $C\alpha$ atoms of the S1 subunit trimer for the apo and holo forms are color-coded. The values are represented as bfactor plots and mapped onto the cartoon representation. (C) The radius of gyration plot for the full apo and holo trimers of the S1 subunit in the 1-up and 2-up conformations color-coded accordingly. (D) Average number of intermolecular hydrogen bonds as indicated by box plots for the S1 trimer with its respective carbohydrate ligands (color-coded). (E) Average number of interchain hydrogen bonds seen as a function of time for the 1-up and 2-up trimer conformations subdivided chain-wise. The chains A with BC and C with AB are seen to have similar interaction patterns.

Table 5. List of the SARS-CoV-2 Variants of Concern Used in the Study

mutant name	country	mutations
wild type	China	0
Variants of Concern		
B.1.1.7 (ALPHA)	UK	Del 69–70, Del 144, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H
B.1.351 (BETA)	South Africa	(L18F), D80A, D215G, (Del 242–244), K417N, E484K, N501Y, D614G, A701V
P.1 (GAMMA)	Brazil	L18F, T20N, P26S, D138Y, R190S, K417N/T, E484K, N501Y, D614G, H655Y, T1027I, V1176F
B.1.617.2 (DELTA)	India	L452R, T478K, D614G, P681R
B.1.617.2 (DELTA plus AY.1)		L452R, K417N, T478K, D614G, P681R
B.1.617.2 (DELTA plus AY.2)		A222V, L452R, K417N, T478K, D614G, P681R

Previous studies have shown that deletion of residues His 69–Val 70 appears to compensate for infectivity deficits associated with affinity-boosting RBM mutations despite its inability to reduce neutralization by a panel of convalescent sera. This deletion is seen to alter the loop length and thereby the conformation of the N3 NTD loop (140–156). It is imperative to study new mutations appearing, such as mutations to basic amino acids, to understand their influence on sugar-binding. Here, we have used the experimental structure of the spike protein cocrystallized with the ACE2 receptor and have mapped the numerous data reported on the various amino acid substitutions in this protein. The spike variants noted so far have no significant structural changes reported. We calculated the Shannon entropy for the mutant variants that have been identified thus far. The mutants have either improved binding by destabilizing the S1B's down conformation and favoring the up conformation or increasing the stability of the prefusion complex. To visualize the mutants, we mapped the structural variability onto the protomer. The

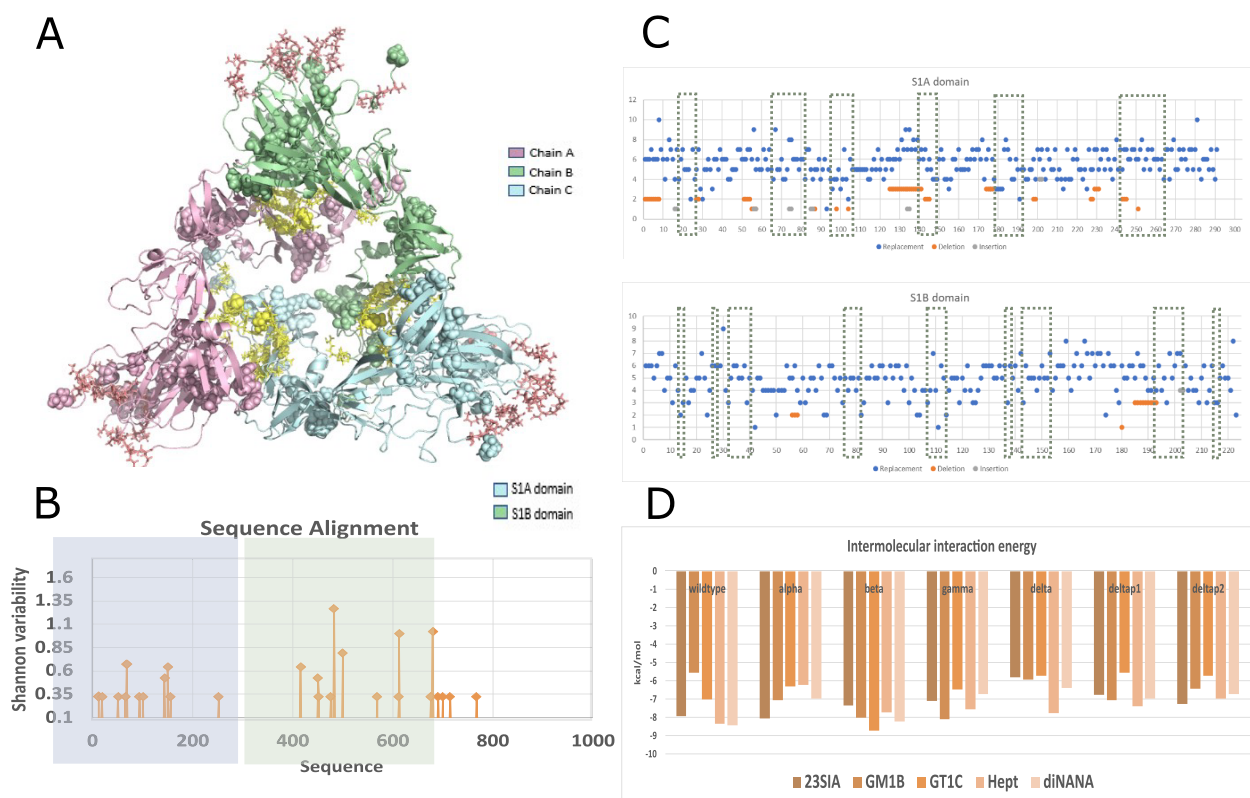


Figure 6. Sequence analysis of the spike protein variants of SARS-CoV-2. (A) Modeling of the experimentally observed glycosylation (shown in stick representation) in a cartoon representation of the S1 subunit trimer; SARS-CoV-2 spike protein (PDB ID: 6VSB) trimer with chains color-coded as chain A in green, chain B in pink, and chain C in cyan. The mutations across the chains are indicated as spheres. (B) Shannon entropy plot obtained using the protein variability server for the aligned variants of SARS-CoV-2. The bars indicate the frequency of variation at different residue positions. Lower entropic frequency implies higher residue conservation. (C) Residue-wise scatter plot representing the frequency of mutations observed at various residues of the variants of the spike protein obtained using the web application CoV-GLUE. The variations are indicated as replacements, deletions, or insertions. The binding site residues are indicated in the green dotted box (<http://cov-glue.cvr.gla.ac.uk>). (D) Intermolecular docking scores for the variants of concern with the respective carbohydrate ligands docked against the S1A and S1B domains of the S1 spike protein subunit clustered according to the variants.

Shannon entropy calculation reveals that most of the mutations have been found in the S1B binding region. The data thus show the S1B's vulnerability to a mutation in comparison to the S1A domain, which is seen to hinder the vaccine design targeting this region. Another important piece of information that we obtain here is that, despite higher Shannon entropy at the S1B domain, the residues in the putative GAG binding site have not been affected.

To better understand the role of the deletion of the residues in the S1A domain, we conducted a short docking analysis. The deletion of the residues 69–70 renders the loop adjacent to the sialic acid binding site short. The alpha variant exhibited a slight increase in the binding energy in its interaction with the 2,3-linked sialic acid compared to the wildtype protomer but was not found to be significant (Figure S7). The mutations in the S1B domain were seen to be distant to the sugar-binding sites and are reflected as similar binding energies with the four variants of concern. Our analysis reveals that despite a variety of mutations, the cell-surface binding ability of the spike protein does not appear to be affected, thus highlighting the desirability of incorporating strategies of inhibiting carbohydrate-binding site (CBS) by glycan analogs or their neutralization by immune mechanisms in our armamentarium for treating SARS-CoV-2 infection. Notably, a recent work that appeared during the review of the study also highlighted the

role of the attachment cell surface receptors in ranking of the SARS-CoV-2-neutralizing antibodies supporting our finding.¹⁹

Through the study, we aim to understand and explain the SARS-CoV-2 infection. The viral entry can be understood as a complex multistep process. The initial interactions with the glycans terminating with sialic acid can be seen as of moderate specificity and low affinity. Nonetheless, they may help to concentrate the virus on the cell surface, which can then recruit specific receptors that initiate the conformational change required to drive the reactions leading to the entry. Here, more extended and widely available charged sugar molecules like heparin or heparan sulfate can drive the stronger interaction together with the maintenance of an open conformation for the adjacent ACE2 receptor site, after which the virus is internalized.

In general, the carbohydrate moieties have been known to lack drug-like properties like adequate bioavailability in terms of therapeutically relevant plasma half-lives, low metabolic stability, and high hydrophilicity. Understanding the structure–activity relationship between the virus and the carbohydrate is thus crucial in the identification of the carbohydrate lead; specifically, the role of each functional group in binding is essential to determining the presentation of the pharmacophores. This biological information is utilized to design a glycomimetic with the ability to overcome these challenges and has the required pharmacodynamic and pharmacokinetic

properties. In this regard, some of the recently approved drugs are mainly glycosidases and sulfated glycosaminoglycans, synthetic heparin-like drugs, and heparin.^{34,35} Thus, the carbohydrate epitopes on the cell surface could now be explored for designing glycomimetic entry inhibitors for use against SARS-CoV-2 infection.

CONCLUSIONS

The molecular models investigating the sugar-binding ability of the SARS-CoV-2 spike protein have been partial to either the S1A domain or the S1B domain independently, thus limiting the comprehension. However, to understand the effect exerted by the cell-surface sugars comprehensively, it is necessary to study the protomer *en-bloc*. Using structural bioinformatics methods, docking, and molecular dynamics simulations (MD), we identified simultaneous glycan-binding to both S1A and S1B domains of the spike protein and their specificities for a given carbohydrate ligand. We determined the overall stability of each complex through the MD simulation, where the molecules were further filtered by the MM-PBSA analysis, which is a reliable method for binding free energy calculations. We studied the dynamic and structural properties of the interacting residues involved in binding. On the basis of our findings, we propose a crucial role to the S1A domain, which first interacts with the sialoside as it scoots over the cell surface. Subsequently, the S1B domain undergoes a conformational switch wherein the heparan sulfate may interact with the S1B domain, followed by the ACE2 binding site. The glycosaminoglycan binding could stabilize the open conformation of the S1B domain, thus promoting the subsequent binding of S protein to its receptor ACE2. Therefore, the sialoside binding S1A domain initially serves to concentrate viruses on the cell surface, facilitating their interaction to more high-affinity protein receptors that promote the viral entry into host cells. Despite the shorter-length simulations, we have captured the essential elements of the interactions between the domains and the glycans. The *in silico* structural analysis from this study thus provides a basis for the functional role of the two parts in conjunction with the interaction with the ACE2 receptor. Furthermore, our analysis into the possible impacts of the mutants on the interaction of the spike protein with the host cell-surface glycans reveals that the putative sugar-binding sites are, for now, unaffected by the various substitutions, making it all the more necessary to study these sites to use them as an adjunct in therapeutics or designing glycomimetics to counter this infection.

METHODS AND MATERIALS

Protein Structure Preparation. The CryoEM structure of the spike glycoprotein (PDB ID: 6VSB) was used as the starting conformation of this study. We have combined experimentally available structure data with computational predictions to build the model of the spike protein. The trimeric, prefusion protein is composed of two domains S1 and S2. We considered residues 14–304 as the N-terminal domain (NTD/S1A) and residues 319–541 as the receptor-binding domain (RBD/S1B), residues 14–541 as the S1 subunit protomer, and the entire S1 subunit (1–685) for the trimer complexes. The full-length model of the spike protein is primarily based on the cryoEM structure, while the missing residues and the extended loops were modeled using MODELER version 9v24.³⁶ The models were ranked using the DOPE statistical potentials. Amino acid mutant models in the spike protein were generated using the wildtype protein as the template and generated using MODELER. Pocket prediction servers like Metapocket²² and CASTp²³ were used to

analyze the surface topography of the two domains and identify cavities which had a higher propensity to function as ligand binding sites.

Ligands Preparation. The saccharides (as indicated in the Supporting Information) were obtained using MolView server (MolView) to generate the “mol” files. These were then converted into the PDB forms using Online SMILES Translator and Structure File Generator (Online SMILES Translator (nih.gov)).

Docking Studies. The molecular docking was performed using Autodock v4.2.³⁷ The Lamarckian genetic algorithm was utilized for docking and screening. The initial ligand conformations were generated using the GLYCAM server,³⁸ and the ligands were prepared using Autodock Tools by adding hydrogens and fixing the bonds; a grid box of 80 × 80 × 80 Å was defined as centered on the predicted binding pocket to allow the ligands to rotate freely using the AutoGrid program. Autodock4 using the Lamarckian genetic algorithm was used to generate various docking conformations with a maximum of 5 000 000 evaluations with default docking parameters. One-hundred conformations were evaluated for each ligand. The best conformation was determined based on cluster analysis, cluster rank, lowest binding energy, and the number of conformations in each cluster.

MD Simulations. MD simulations were performed with Gromacs software using the CHARMM36 force field.³⁹ CHARMM-GUI, a glycan reader module, was used to generate the topology and the starting conformation obtained from the docking results.^{40–42} An octahedral water box was used with 10Å as the edge distance. The complexes were solvated in TIP3P (Transferable Intermolecular Potential with 3 Points), and ions were added to neutralize the system charge using the Monte Carlo algorithm. The steepest descent algorithm was used for energy minimization until the force converged at Fmax <1000 kJ/mol/nm. The system was then allowed to equilibrate initially at a constant number of particles, volume, and temperature (300 K) using a Nose–Hoover thermostat followed by a constant number of particles, pressure (1 bar), and temperature using a Parrinello–Rahman barostat wherein positional restraints of $k = 1000$ kJ/mol/nm² were applied to heavy atoms in the protein. The equilibration was followed by production simulations at a constant temperature of 300 K for 10 ns with 2 fs time intervals for each of the complexes for screening and studying the interaction data. Analysis was carried using the inbuilt packages in GROMACS⁴³ to process the trajectories, eliminating periodicity issues and fitting, recentering, and concatenation before analyzing them with various tools.

Sequence Analysis, Phylogenetic Tree, and Sequence Alignment. Sequence conservation among these spike proteins was estimated by comparing the domains using the BlastP⁴⁴ server and subjecting them to alignment tools. The sequences of the seven human coronaviruses were retrieved from UniProt.⁴⁵ Multiple sequence alignments of these sequences was performed using the T-COFFEE⁴⁶ server using the M-COFFEE algorithm, which uses a structure-based alignment protocol. Phylogenetic analyses of these spike proteins were conducted using MEGA-6²⁴ via the neighbor-joining (NJ) method with 100 bootstrap replicates.

Sequence Variability. The sequence variability was calculated using the protein variability server.⁴⁷ This server calculates the sequence variability within a multiple sequence alignment using several variability metrics. Here, the first sequence was taken as the reference sequence and was set as the standard by which the mutants were compared. The mutants were mapped onto the structure of the S1 subunit. Shannon entropy (H ; 0–4.322) is calculated for every position with equal representation for all of the residues. A position with $H > 2.0$ is considered variable, whereas those with $H < 2$ are considered conserved. Highly conserved positions are seen to have $H < 1.0$

Electrostatic Potential. The electrostatic potential map of the S1 subunit was constructed using the APBS server.

Interactive structural analyses were conducted using PyMol,⁴⁸ UCSF Chimera,⁴⁹ and 3D protein imager⁵⁰

■ ASSOCIATED CONTENT

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscchembio.1c00691>.

Additional information regarding the complex preparation and interaction energy is provided in the Supplementary Methods, followed by Supplementary Figures and Tables (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Avadhesh Surolia – Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India; orcid.org/0000-0002-2466-2514; Phone: (+91)-080-22932714; Email: surolia@iisc.ac.in

Author

Nisha Grandhi Jayaprakash – Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Complete contact information is available at: <https://pubs.acs.org/10.1021/acscchembio.1c00691>

Author Contributions

N.G.J. and A.S. designed the experiments. N.G.J. performed the experiments and carried out the analysis. N.G.J. and A.S. analyzed the data, compiled the results, and wrote the manuscript.

Funding

N.G.J. is a Research Associate in the Department of Biotechnology (DBT), Government of India, Grant No. BT/PR24443/MED/29/1220/2017 (to A.S.). A.S. is a Science and Engineering Research Board (SERB) Distinguished Fellow, Government of India. N.G.J. received a Department of Science Technology, Government of India, INSPIRE Fellowship for initial training. This study was supported in part by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India, Grant SB/DF-003/2018 (to A.S.) and by the Department of Biotechnology (DBT), Government of India, Grant No. BT/PR27659/BID/7/829/2018 (to A.S.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We would like to thank MBU for a high-performance computing facility in the MBU department.

■ ABBREVIATIONS

ACE2, Angiotensin Converting Enzyme 2; APBS, Advanced Poisson–Boltzmann Solver; BGAs, Blood Group Antigens; BLAST, Basic Local Alignment Search Tool; CBS, Carbohydrate Binding Site; CHARMM, Chemistry at Harvard Macromolecular Mechanics; GAGs, Glycosaminoglycans; GROMACS, Groningen Machine for Chemical Simulations; GSLs, Gangliosides; MD, Molecular Dynamics; MM/PBSA, Molecular Mechanics/Poisson–Boltzmann Surface Area; MSA, Multiple Sequence Alignment; NTD, N-Terminal Domain; PDB, Protein Data Bank; RBD, Receptor Binding Domain; RBM, Receptor Binding Motif; RMSD, Root Mean Square Deviation; RMSF, Root Mean Square Fluctuation; S1A, Spike Protein, S1 Subunit, A Domain; S1B, Spike Protein,

S1 Subunit, B Domain; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; SASA, Solvent Accessibility Surface Area; sLeX, Sialyl Lewis X; TIP3P, Transferable Intermolecular Potential with 3 Points; VOC, Variant of Concern

■ REFERENCES

- (1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382* (8), 727–733.
- (2) Trypsteen, W.; Van Cleemput, J.; van Snippenberg, W.; Gerlo, S.; Vandekerckhove, L. On the Whereabouts of SARS-CoV-2 in the Human Body: A Systematic Review. *PLoS Pathog.* **2020**, *16* (10), No. e1009037.
- (3) Liu, J.; Li, Y.; Liu, Q.; Yao, Q.; Wang, X.; Zhang, H.; Chen, R.; Ren, L.; Min, J.; Deng, F.; Yan, B.; Liu, L.; Hu, Z.; Wang, M.; Zhou, Y. SARS-CoV-2 Cell Tropism and Multiorgan Infection. *Cell Discovery* **2021**, *7* (1), 1–4.
- (4) Baker, A. N.; Richards, S. J.; Guy, C. S.; Congdon, T. R.; Hasan, M.; Zwetsloot, A. J.; Gallo, A.; Lewandowski, J. R.; Stansfeld, P. J.; Straube, A.; Walker, M.; Chessa, S.; Pergolizzi, G.; Dedola, S.; Field, R. A.; Gibson, M. I. The SARS-CoV-2 Spike Protein Binds Sialic Acids and Enables Rapid Detection in a Lateral Flow Point of Care Diagnostic Device. *ACS Cent. Sci.* **2020**, *6* (11), 2046–2052.
- (5) Stehle, T.; Khan, Z. M. Rules and Exceptions: Sialic Acid Variants and Their Role in Determining Viral Tropism. *J. Virol.* **2014**, *88* (14), 7696–7699.
- (6) 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*; Nature Publishing Group, 2014; pp 739–749. DOI: 10.1038/nrmicro3346.
- (7) Sun, X.-L. The Role of Cell Surface Sialic Acids for SARS-CoV-2 Infection. *Glycobiology* **2021**, *31*, 1245.
- (8) Dhar, C.; Sasmal, A.; Diaz, S.; Verhagen, A.; Yu, H.; Li, W.; Chen, X.; Varki, A. Are Sialic Acids Involved in COVID-19 Pathogenesis? *Glycobiology* **2021**, *31* (9), 1068–1071.
- (9) Yu, M.; Zhang, T.; Zhang, W.; Sun, Q.; Li, H.; Li, J. P. Elucidating the Interactions Between Heparin/Heparan Sulfate and SARS-CoV-2-Related Proteins—An Important Strategy for Developing Novel Therapeutics for the COVID-19 Pandemic. *Frontiers in Molecular Biosciences*; Frontiers Media S.A., 2021; p 628551. DOI: 10.3389/fmolb.2020.628551.
- (10) Linhardt, R. J. 2003 Claude S. Hudson Award Address in Carbohydrate Chemistry. Heparin: Structure and Activity. *J. Med. Chem.* **2003**, *46* (13), 2551–2564.
- (11) Clausen, T. M.; Sandoval, D.; Spleid, C.; Pihl, J.; Painter, C.; Thacker, B.; Glass, C.; Narayanan, A.; Majowicz, S.; Zhang, Y.; Torres, J.; Golden, G.; Porell, R.; Garretson, A.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B.; Caradonna, T.; Kellman, B.; Martino, C.; Gordts, P.; Leibel, S.; Chanda, S.; Schmidt, A.; Godula, K.; Jose, J.; Corbett, K.; Ward, A.; Carlin, A.; Esko, J. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *bioRxiv*2020. DOI: 10.1101/2020.07.14.201616.
- (12) Nguyen, L.; McCord, K. A.; Bui, D. T.; Bouwman, K. M.; Kitova, E. N.; Kumawat, D.; Daskhan, G. C.; Tomris, I.; Han, L.; Chopra, P.; Yang, T.-J.; Willows, S. D.; Mason, A. L.; Lowary, T. L.; West, L. J.; Hsu, S.-T. D.; Tompkins, S. M.; Boons, G.-J.; de Vries, R. P.; Macauley, M. S.; Klassen, J. S. Sialic Acid-Dependent Binding and Viral Entry of SARS-CoV-2. *bioRxiv* 2021, 2021.03.08.434228. DOI: 10.1101/2021.03.08.434228.
- (13) Kato, K.; Ishiwa, A. The Role of Carbohydrates in Infection Strategies of Enteric Pathogens. *Trop. Med. Health* **2015**, *43* (1), 41.
- (14) Zhou, T.; Tsybovsky, Y.; Gorman, J.; Rapp, M.; Cerutti, G.; Chuang, G. Y.; Katsamba, P. S.; Sampson, J. M.; Schön, A.; Bimela, J.; Boyington, J. C.; Nazzari, A.; Olia, A. S.; Shi, W.; Sastry, M.; Stephens, T.; Stuckey, J.; Teng, I. T.; Wang, P.; Wang, S.; Zhang, B.; Friesner, R.

- A.; Ho, D. D.; Mascola, J. R.; Shapiro, L.; Kwong, P. D. Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **2020**, *28* (6), 867.
- (15) Tracking SARS-CoV-2 variants. <https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/> (accessed Jul 11, 2021).
- (16) Greaney, A. J.; Loes, A. N.; Crawford, K. H. D.; Starr, T. N.; Malone, K. D.; Chu, H. Y.; Bloom, J. D. Comprehensive Mapping of Mutations in the SARS-CoV-2 Receptor-Binding Domain That Affect Recognition by Polyclonal Human Plasma Antibodies. *Cell Host Microbe* **2021**, *29* (3), 463–476.
- (17) Lenza, M. P.; Oyenarte, I.; Diercks, T.; Quintana, J. I.; Gimeno, A.; Coelho, H.; Diniz, A.; Peccati, F.; Delgado, S.; Bosch, A.; Valle, M.; Millet, O.; Abrescia, N. G. A.; Palazon, A.; Marcelo, F.; Jimenez-Oses, G.; Jimenez-Barbero, J.; Arda, A.; Ereno-Orbea, J. Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and Their Interactions with Human Lectins. *Angew. Chem., Int. Ed. Engl.* **2020**, *59* (52), 23763–23771.
- (18) Thépaut, M.; Luczkowiak, J.; Vivès, C.; Labiod, N.; Bally, I.; Lasala, F.; Grimoire, Y.; Fenel, D.; Sattin, S.; Thielens, N.; Schoehn, G.; Bernardi, A.; Delgado, R.; Fieschi, F. DC/L-SIGN Recognition of Spike Glycoprotein Promotes SARS-CoV-2 Trans-Infection and Can Be Inhibited by a Glycomimetic Antagonist. *PLoS Pathog.* **2021**, *17* (5), e1009576.
- (19) Lempp, F. A.; Soriaga, L. B.; Montiel-Ruiz, M.; Benigni, F.; Noack, J.; Park, Y.-J.; Bianchi, S.; Walls, A. C.; Bowen, J. E.; Zhou, J.; Kaiser, H.; Joshi, A.; Agostini, M.; Meury, M.; Dellota, E.; Jaconi, S.; Cameroni, E.; Martinez-Picado, J.; Vergara-Alert, J.; Izquierdo-Useros, N.; Virgin, H. W.; Lanzavecchia, A.; Veesler, D.; Purcell, L. A.; Telenti, A.; Corti, D. Lectins Enhance SARS-CoV-2 Infection and Influence Neutralizing Antibodies. *Nature* **2021**, *598* (7880), 342–347.
- (20) Li, F. Receptor Recognition Mechanisms of Coronaviruses: A Decade of Structural Studies. *J. Virol.* **2015**, *89* (4), 1954–1964.
- (21) Milanetti, E.; Miotto, M.; Di Rienzo, L.; Nagaraj, M.; Monti, M.; Golbek, T. W.; Gosti, G.; Roeters, S. J.; Weidner, T.; Otzen, D. E.; Ruocco, G. In-Silico Evidence for a Two Receptor Based Strategy of SARS-CoV-2. *Front. Mol. Biosci.* **2021**, *8*, 1–10.
- (22) Zhang, Z.; Li, Y.; Lin, B.; Schroeder, M.; Huang, B. Identification of Cavities on Protein Surface Using Multiple Computational Approaches for Drug Binding Site Prediction. *Bioinformatics* **2011**, *27* (15), 2083–2088.
- (23) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: Computed Atlas of Surface Topography of Proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W363–W367.
- (24) Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **2013**, *30* (12), 2725–2729.
- (25) Varki, A.; Angata, T. Siglecs—the Major Subfamily of I-Type Lectins. *Glycobiology* **2006**, *16* (1), 1R–27R.
- (26) Liu, Y. M.; Shahed-Al-Mahmud, M.; Chen, X.; Chen, T. H.; Liao, K. S.; Lo, J. M.; Wu, Y. M.; Ho, M. C.; Wu, C. Y.; Wong, C. H.; Jan, J. T.; Ma, C. A Carbohydrate-Binding Protein from the Edible Lablab Beans Effectively Blocks the Infections of Influenza Viruses and SARS-CoV-2. *Cell Rep.* **2020**, *32* (6), 108016.
- (27) Brandt, M. W.; Santa-Clara, P. Simulated Likelihood Estimation of Diffusions with an Application to Exchange Rate Dynamics in Incomplete Markets. *J. financ. econ.* **2002**, *63* (2), 161–210.
- (28) Kim, S. Y.; Jin, W.; Sood, A.; Montgomery, D. W.; Grant, O. C.; Fuster, M. M.; Fu, L.; Dordick, J. S.; Woods, R. J.; Zhang, F.; Linhardt, R. J. Characterization of Heparin and Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Spike Glycoprotein Binding Interactions. *Antiviral Res.* **2020**, *181*, 104873.
- (29) Kalra, R. S.; Kandimalla, R. Engaging the Spikes: Heparan Sulfate Facilitates SARS-CoV-2 Spike Protein Binding to ACE2 and Potentiates Viral Infection. *Signal Transduction and Targeted Therapy*; Springer Nature, 2021; pp 1–2. DOI: 10.1038/s41392-021-00470-1.
- (30) Clausen, T. M.; Sandoval, D. R.; Spliid, C. B.; Pihl, J.; Perrett, H. R.; Painter, C. D.; Narayanan, A.; Majowicz, S. A.; Kwong, E. M.; McVicar, R. N.; Thacker, B. E.; Glass, C. A.; Yang, Z.; Torres, J. L.; Golden, G. J.; Bartels, P. L.; Porell, R. N.; Garretson, A. F.; Laubach, L.; Feldman, J.; Yin, X.; Pu, Y.; Hauser, B. M.; Caradonna, T. M.; Kellman, B. P.; Martino, C.; Gordts, P. L. S. M.; Chanda, S. K.; Schmidt, A. G.; Godula, K.; Leibel, S. L.; Jose, J.; Corbett, K. D.; Ward, A. B.; Carlin, A. F.; Esko, J. D. SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell* **2020**, *183* (4), 1043–1057.
- (31) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6* (10), 1722–1734.
- (32) Singer, J.; Gifford, R.; Cotten, M.; Robertson, D. CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. **2020**. DOI: 10.20944/PREPRINTS202006.0225.V1.
- (33) Khan, A.; Zia, T.; Suleman, M.; Khan, T.; Ali, S. S.; Abbasi, A. A.; Mohammad, A.; Wei, D.-Q. Higher Infectivity of the SARS-CoV-2 New Variants Is Associated with K417N/T, E484K, and N501Y Mutants: An Insight from Structural Data. *J. Cell. Physiol.* **2021**, *236*, 7045.
- (34) Ernst, B.; Magnani, J. L. From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Discovery* **2009**, *8* (8), 661–677.
- (35) Schuurs, Z. P.; Hammond, E.; Elli, S.; Rudd, T. R.; Mycroft-West, C. J.; Lima, M. A.; Skidmore, M. A.; Karlsson, R.; Chen, Y.-H.; Bagdonaite, I.; Yang, Z.; Ahmed, Y. A.; Richard, D. J.; Turnbull, J.; Ferro, V.; Coombe, D. R.; Gandhi, N. S. Evidence of a Putative Glycosaminoglycan Binding Site on the Glycosylated SARS-CoV-2 Spike Protein N-Terminal Domain. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2806–2818.
- (36) Webb, B.; Sali, A. Protein Structure Modeling with MODELLER. In *Methods Mol. Biol.*; Humana Press Inc., 2017; Vol. 1654, pp 39–54. DOI: 10.1007/978-1-4939-7231-9_4.
- (37) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Software News and Updates AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791.
- (38) Glycam Web. <http://glycam.org>.
- (39) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.
- (40) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**, *29* (4), 320–331.
- (41) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; Mackerell, A. D.; Klauda, J. B.; Im, W. 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), 405–413.
- (42) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865.
- (43) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56.
- (44) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
- (45) Bateman, A.; Martin, M. J.; O'Donovan, C.; Magrane, M.; Alpi, E.; Antunes, R.; Bely, B.; Bingley, M.; Bonilla, C.; Britto, R.; Bursteinas, B.; Bye-Ajee, H.; Cowley, A.; Da Silva, A.; De Giorgi, M.; Dogan, T.; Fazzini, F.; Castro, L. G.; Figueira, L.; Garmiri, P.; Georghiou, G.; Gonzalez, D.; Hatton-Ellis, E.; Li, W.; Liu, W.; Lopez, R.; Luo, J.; Lussi, Y.; MacDougall, A.; Nightingale, A.; Palka, B.;

Pichler, K.; Poggioli, D.; Pundir, S.; Pureza, L.; Qi, G.; Rosanoff, S.; Saidi, R.; Sawford, T.; Shypitsyna, A.; Speretta, E.; Turner, E.; Tyagi, N.; Volynkin, V.; Wardell, T.; Warner, K.; Watkins, X.; Zaru, R.; Zellner, H.; Xenarios, I.; Bougueleret, L.; Bridge, A.; Poux, S.; Redaschi, N.; Aimo, L.; ArgoudPuy, G.; Auchincloss, A.; Axelsen, K.; Bansal, P.; Baratin, D.; Blatter, M. C.; Boeckmann, B.; Bolleman, J.; Boutet, E.; Breuza, L.; Casal-Casas, C.; De Castro, E.; Coudert, E.; Cucho, B.; Doche, M.; Dornevil, D.; Duvaud, S.; Estreicher, A.; Famiglietti, L.; Feuermann, M.; Gasteiger, E.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz-Gumowski, N.; Hinz, U.; Hulo, C.; Jungo, F.; Keller, G.; Lara, V.; Lemercier, P.; Lieberherr, D.; Lombardot, T.; Martin, X.; Masson, P.; Morgat, A.; Neto, T.; Noupikel, N.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pozzato, M.; Pruess, M.; Rivoire, C.; Roechert, B.; Schneider, M.; Sigrist, C.; Sonesson, K.; Staehli, S.; Stutz, A.; Sundaram, S.; Tognolli, M.; Verbregue, L.; Veuthey, A. L.; Wu, C. H.; Arighi, C. N.; Arminski, L.; Chen, C.; Chen, Y.; Garavelli, J. S.; Huang, H.; Laiho, K.; McGarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Q.; Wang, Y.; Yeh, L. S.; Zhang, J. UniProt: The Universal Protein Knowledgebase. *Nucleic Acids Res.* **2017**, *45* (D1), D158–D169.

(46) Di Tommaso, P.; Moretti, S.; Xenarios, I.; Orobittg, M.; Montanyola, A.; Chang, J.-M.; Taly, J.-F.; Notredame, C. T-Coffee: A Web Server for the Multiple Sequence Alignment of Protein and RNA Sequences Using Structural Information and Homology Extension. *Nucleic Acids Res.* **2011**, *39*, W13–W17.

(47) Garcia-Boronat, M.; Diez-Rivero, C. M.; Reinherz, E. L.; Reche, P. A. PVS: A Web Server for Protein Sequence Variability Analysis Tuned to Facilitate Conserved Epitope Discovery. *Nucleic Acids Res.* **2008**, *36*, W35.

(48) PyMOL *The PyMOL Molecular Graphics System*, version 1.2r3pre; Schrödinger, LLC.

(49) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.

(50) Tomasello, G.; Armenia, I.; Molla, G. The Protein Imager: A Full-Featured Online Molecular Viewer Interface with Server-Side HQ-Rendering Capabilities. *Bioinformatics* **2020**, *36* (9), 2909–2911.