

Concerns Raised on Blood Group Determinants in Plasma Membrane Interaction of the SARS-CoV-2

Klaus Fiedler

Fribourg, Switzerland

contact@klausfiedler.ch

Keywords: Glycans, Coronaviridae, COVID-19, Organ Tropism

Words abstract: 199

Words text (without abstract): 8397

The SARS-CoV-2 pandemic has resulted in the generation of evolutionary-related variants. The S-protein of the B.1.1.7 variant (deletion N-terminal domain (NTD) His69Val70Tyr144) may contribute to altered infectivity. These mutations may have been presaged by animal mutations in minks housed in mink farms that according to the present analysis by modelling of protein ligand docking altered a high affinity binding site in the S-protein NTD. These mutants likely occurred only sporadically in humans. Tissue-adaptations and the size of the mink relative to the infected human population size back then may have comparatively increased the relative mutation rate. Simple, multi-threaded automated docking that is widely available, assigns increased binding of the blood type II A antigen to the SARS-CoV-2 S-protein NTD of B.1.1.7 with an overall increased docking interaction of blood group A harbouring glycolipids relative to group B or H (H, p=0.04). The top scoring glycan is identified as a DSGG (also classified as sialosyl-MSGG or disialosyl-Gb5) that may compete with heparin, which is similar to heparan sulfate linked to proteinaceous receptors on the tissue surface. Other glycolipids are found to interact with lower affinity, except long ligands that have suitable ligand binding poses to match the curved binding pocket.

Introduction

The cellular entry of viridae is shown to frequently include surface determinants of glycolipids and glycoproteins, whereas some viridae bind exclusively to proteinaceous receptors (1). Since genetic analyses have previously indicated, that surface loops of coronaviridae determine tissue-tropism in the animal (2), as imminent to simpler comparison in tissue-culture (3), the question of blood group glycolipid- or glycoprotein-determinant interaction has to be posed. Parvovirus, as one example of a DNA-virus (Erythrovirus) binds to the P-antigen (Globoside (Gb) 4) and can cause a transient aplastic anemia due to the abundance of Gb4 in red blood cells (1). Polyfucosylated N-linked glycopeptides and multiple glycolipids had previously been identified in the human intestine and have, moreover, suggested a high variability of individual O-glycomes, which may indicate individual differences in virus-receptor expression (4–6). Although the glycosphingolipid (GSL) and lipid variety in mammalian organisms and humans in particular is very high, succinct information on individual susceptibility to disease is still scarce (7). Transmission of SARS-CoV-2, a single-stranded RNA virus in mink farms has been recently studied (8), anthropozoonotic infection of humans has been proposed in spill-over from minks back to the original host in this infectious cycle. Moreover, it has been proposed, that mutations that arose in the mink propagation of SARS-CoV-2 had introduced novel mutants into the human population (9, 10). Since the multi-organ tropism of SARS-CoV-2 had been demonstrated, it is possible that prolonged anthropozoonotic amplification of host infections could alter the host and/or organ-range and tropisms that may increase disease lethality (11, 12). The association of blood groups with the SARS-CoV-2 disease (COVID-19) has recently been established in meta-analyses and suggests the likely increase in prevalence in blood group A individuals as well as linked elevated mortality (13, 14). A multitude of explanations for a role of determinants of individual blood groups has been put forward and it has been theorized that an indirect effect of blood group associated expression of clotting factors could contribute to the severity of COVID-19 (15, 16). Surface determinants alone, as shown in platelet clotting in vitro would provide the other line of thoughts to explain the ABO blood group-dependent aetiology, just as the above mentioned direct interaction of the virus with the cell surface of SARS-CoV and SARS-CoV-2 target cells could include a co-receptor next to the ACE2 protein (17, 18).

In the current work, a drug-docking-like approach is tested to analyse interaction of carbohydrates of a library of GSL headgroups with the SARS-CoV-2 N-terminal domains (NTDs) of the SARS-CoV-2 wildtype virus (MN908947, NC045512) and the British mutant B.1.1.7 (8, 19, 20).

Material and Methods

The computational screen of carbohydrates involved analysis with the preparation of glycans from Woods at <http://www.ccrc.uga.edu> (with multiple conformers) or preparation from pre-existing fragments from larger structures if not available as such. The PyRx modelling queue Version 0.8 was used with Intel processors on Windows 7, 8 or 10 operating systems. The MarvinView Dreiding force field utilized in some previous work was not utilized in the present experimental series, yet, files were processed by Chimera 1.14 (see (21)) and saved as mol2 file for import to PyRx docking. The Autodock VINA (22) implementation of PyRx from S. Dallakyan (<http://pyrx.scripps.edu>) was utilized with the grid size as indicated in single experiments. The algorithm installs OpenBabel (23) and a uff (united force field) for energy minimization, conjugate gradients with 200 steps and a cut-off for energy minimization of 0.1. Partial charges were added to receptors using PyBabel (MGL Tools; <http://mgltools.scripps.edu>). Authors mention the difference of this procedure to using OpenBabel for adding partial charges, and care should be taken especially for novel ligands that may not be recognized. No limits to torsions were allowed in the computational run. Single CPU time was up to 16 hours for longest/branched ligands in exhaustiveness 8. The analysed data were judged for surface binding in PyRx or in Chimera by the ViewDock import function. SQLite data were analysed using SQLite (Hipp, D. R.) and DB Browser for SQLite from <http://sqlitebrowser.org>. Autodock/Vina re-docking of ligands without torsional degrees of freedom was carried out to judge the top-scoring screen (exhaustiveness 3 or 6 with blood-group ligands). Re-dock of the top scoring ligand was also followed-up with the rotating side-chain function in Vina that allowed to validate the top scores independently and with slightly altered “poses”. For this step of the project, AutoDockTools Version 1.5.6 (<http://mgltools.scripps.edu>) was utilized to generate separate files of flexible and fixed amino-acid residues of the model (24). Further stepwise addition of “poses” was obtained with the flexdistance and autobox implemented in the SMINA program (<https://sourceforge.net/projects/smina/files>). Spreadsheet use and calculations were carried out in Microsoft Office 2013 Professional Plus. Further computational docking focused on the putative binding site was utilized to generate a high resolution of docking interaction, since the method is described to not only “home in” on the best interacting binding site but to stall on lowly evaluated interaction pockets if used in the “global” docking procedure. Therein the exhaustiveness was increased to 12. H-bonding was determined with ViewDock and with tolerances 0.4 Å, 20° (25) or 0.8 Å, 30° similar to calculations previously applied (26). Annotation of carbohydrates was from <http://www.lipidmaps.org> and from literature sources cited in the Results. Chimera 1.14 was used for further calculations and Coulombic surface charge presentations using default values. Structure files were scored as likely binding site ligands in *pdb-care* from <http://www.glycosciences.de> to test for structural intactness if not visually controlled.

Structures were downloaded from RCSB (<https://www.rcsb.org>) or PDBe (<https://www.ebi.ac.uk/pdbe>). The Swiss-Model Server on <http://www.sib.swiss> was applied to predict structures of the SARS-CoV-2 S-protein including several versions of the modelling: Either the automatic queue was utilized or direct selection of templates was applied in obtaining best fit of structure and template (27). BLAST (28) and HHBlits (29) were used for the homology modelling. Templates that matched the primary sequence model query (amino acids 1-291) excluding the 13 residues of signal-sequence were used for modelling. These were represented by 7a25 A/B/C and 328 other templates for a general approach of ligand binding. The top templates corresponded to these 7a25 chains, chain A of 7cab and three chains of 7cai. Nine amino-acids were subjected to loop modelling although the structures of the S-protein was nearly complete (30). Previous models were not utilized, since the 6vxx and 6vsb structures were

not completing the NTD and contained some gaps (31, 32). The SwissModel7C_26J matched preferentially the C chain of 7a25 with RMSD of 0.129 Å and a QMean -2.07. Specific models matching 7a25 A, B or C were generated to compare the ligand binding characteristics of each conformer (SwissModel7A, 7B and 7C of QMean -1.72, -1.64 and -2.22. Evaluation of similarity included 1705 templates. RMSDs and further characteristic found for the NTD and RBD are listed in the graphical description of models. Energy minimization of structures was carried through with a minimum of 100 steps of conjugate gradients applying the amber ff14SB force field (33) and further AM1-BCC charges. Molecular dynamics to generate random conformers in the first step was utilized with equilibration of 5000 steps and a production phase of further 5000 steps, and was visually controlled by the movie output. A Nosé thermostat with 298 K was applied (relaxation time 0.2). For the mutants generated in Modeller Version 9.12 (34, 35) with a single structural template (and for the wildtype protein) the last third of the output was clustered and judged in frequent occurrence, the top scoring clusters with a maximal member number were selected. Automodel was applied in the Modeller suite for this procedure and full length NTD sequence 14-291 or 69Del70Del144Del of 14-291 (20) was used as input to the structural match of above described self-generated template (SwissModel7C_26J). The potential energy for the wildtype protein 7C_Mod-wt reached -15544.9 and for the mutant 7C_Mod-B-1-1-7 -14974.6 following the heating in the molecular dynamics, and -16429.9 and -15663.3 after the production procedure, respectively. Automodel (Modeller) and Swiss-Model (WWW) results were judged differently in energy and could not be comparatively analysed. They are indicated with RMSD values: SwissModel7C_26J - 7C_Mod-wt 0.190 Å, 7C_Mod-wt - 7C_Mod-B-1-1-7 0.341 Å and molecular dynamics clusters (high population number) 7C_Mod-wt-MD - 7C_Mod-B-1-1-7-MD 2.403 Å. The SwissModel7C_26J models themselves differed by 0.084 Å RMSD from energy-minimised and 1.741 Å RMSD from molecular dynamics simulated form used for some experiments. Following the described model generation, ClusPro was used for further docking of heparin with rotating side-chains and generated best scoring ligand-bound “poses” with the SwissModel7A, 7B and 7C input files (36). Lowest energies are indicated in the Figure. Some genetic and epidemiological data were gleaned from www.datamonkey.org and www.nextstrain.org to confirm the spread of the wildtype and mutant SARS-CoV-2 sub-strains or clades.

Results

In a first approach, the SARS-CoV-2 S-protein was subjected to molecular docking of a tetrasaccharide heparin using the ClusPro queue (36) to confirm the results on the S-protein RBD (37–39)(see SARS and protective role of lactoferrin (40)). The trimer of the S-protein is shown in Fig. 1 to demonstrate the different binding sites within S-protein RBDs and NTDs that can be described by docking each of chain A, B and C conformers of the SwissModel 7a25 (SwissModel7A, B, C) generated by the queue on 11 February 2021 (27, 30). ClusPro delivers several high scoring docking solutions some of which largely correspond to the previously described ligand binding simulations (Fig. 2, B RBD and C RBD). The Autodock re-dock energies corresponded to the -14.4 kcal/mol (B RBD) and -13.5 kcal/mol (C RBD) which could not be directly compared to the entropic energy evaluations used in the original ClusPro docking poses. Novel to this docking analysis is the pose of the heparin bound to the A conformer of the SwissModel here found interacting with the “up” conformation of the S-protein, which is slightly displaced towards α -helix 304-308 of the RBD A, with an increased Autodock affinity of -15.8 kcal/mol. Although elongated heparin molecules or antennae of proteoglycans could span and connect the RBD with the NTD, the data do not provide an indication for the proximity of the tetrasaccharide to both, each RBD and neighbouring NTDs. The described bridging of RBD and ACE2 wherein the hexasaccharide heparan sulfate (GlcA(2S)-GlcNS(6S))₃ suggested to interact with the RBD, would connect to ACE2, could not be demonstrated, since other binding sites showed highly increased affinity relative to the proposed interaction. A summary of potentially interacting residues (proximity 5 Å) is shown in Table 1 (SwissModel of residues 334-524 of S-protein). With vastly increased ClusPro affinity, a further binding site in the NTD of each SARS-CoV-2 S-protein protomer could be demonstrated and is shown with lowest energies in Fig. 2. The lowest energy of -944.4 corresponded to the Autodock re-dock energy of -14.3 kcal/mol for the B NTD, the A NTD had a re-dock affinity of -14.3 kcal/mol and the C chain of -15.2 kcal/mol. As compared by ClusPro energies, the binding to the N-terminal domain would be highly likely, more prevalent or of higher affinity than the interaction previously described, i.e. the binding to the RBD. The conformer of SwissModel NTD C docked to heparin was studied in the later analysis with docked CARB115 library residues to demonstrate the influence of side-chain rotamers (Suppl. Fig. 1) and/or sufficiency of the procedure. Residues within 5 Å distance of docked heparin for the SwissModel NTDs A, B, C (residues 14-291) of the S-protein are shown in Table 1. Evident from analysing the preliminary data with regard to natural heparan sulfate interaction, is the slightly different pose of the B NTD ligand, which is fully covered by the S-protein loop 245-251. This terminal interaction does not correspond to the interaction of the nitrous acid depolymerized isolate of heparin and may constitute the reducing end of heparin produced in an enzymatic digest (see (41)). As a note of caution, it should be stated, that only the interaction of heparin with the RBDs is currently validated by the full structure of the 7a25 trimer, whereas several of the NTD residues indicated in Fig. 1 that were introduced by the protein modelling show heparin interaction (5 of 9 for NTD A, 7 of 9 for NTD B, 6 of 9 for NTD C).

The blood group antigens or elongated glycolipids (with Glc at the reducing end) were tested for interaction in the next step. The glycolipids displaying antigenic determinants (Table 2) can be grouped into lacto (type I), neolacto (type II and type III) and globo (type IV) series of glycosphingolipids (GSL). A variety of different linkages generates at least 15 different GSL-headgroups that could be recognized by anti-blood group antibodies. For this approach, Autodock Vina was used with the localized binding pocket scrutinized in the Figs. 1-5 with the S-protein NTDs. The model used for heparin docking was further modified by the Modeller routine (34, 42) to mutate the wildtype to the His69Val70Tyr144

deletion mutant B.1.1.7. High-energy conformers were produced by molecular dynamics in Chimera (298 K) that could likely mimic one major binding mode of the S-protein NTD to be used for the interaction analyses. Localized docking shows, that the elongated blood type determinants have interaction energies (Autodock re-dock) of -15.0 to -21.6 kcal/mol (Fig. 3 A). Overall, a significantly stronger interaction of A versus H (0) blood group determinants could be determined with these procedures for the B.1.1.7 mutant S-protein NTDs which is shown in the comparison of blood type averages in Fig. 3 B. Although the result could be considered preliminary, one of the blood type II A presenting glycolipids (No. 5) shows clearly increased affinities to the B.1.1.7 binding pocket. Regardless of whether the minimized energy model only (not shown) or the molecular dynamics (cluster) model was subjected to docking, a highly increased interaction was simulated.

Previous analyses have suggested that the S-protein NTD may interact with ganglioside GM1 although the structure of the SARS-CoV-2 S-protein available was then including large gaps in loops and in particular at the N-terminal region (43). In determining the different binding sites of the entire N-terminal domain, which is subject to algorithmic hindrance due to a multitude of possible interaction sites, the half molecule (NTD) exposed to the viral exterior was here used with Autodock Vina (Fig. 4). Both, the elongated binding site demonstrated in Fig. 1, 2 and 5 and an N-terminal site could be shown. In Fig. 4 the top score of the carbohydrate screen Di-Sialosyl Galactosyl Globoside (DSGG) or di-sialosyl-Gb5 (44) which interacted with the affinity of -7.8 kcal/mol is displayed in violet and residues within 3 Å proximity are indicated. The top-score GalNAc-GM1b that was found to interact at the N-terminus with relatively high affinity of -6.6 kcal/mol was discarded as low affinity ligand. In previous screens with the similar procedure interactions of identical affinity were considered to be false-positives or nearly unreliable (45, 46). This was proposed in cognate or non-cognate docking poses but would be exceeded in tetrasaccharides that serially interacted with larger binding pockets. Previously identified residues (43) are shown, yet, did only partially overlap with the here identified novel binding site which apparently includes the N-terminal Gln14 itself (H-bonded). Residues overlapping with the GM1 binding site are signified in grey (Fig. 4). Also here three amino acids are within 5 Å distance that were included from the modelling queue, and the result should thus not be considered as final.

In the final analysis of refinement of interactions, SwissModel7C_26J was used to generate docking in local binding mode. This included the area surrounding His69 which has a deepened, curved shape surface morphology. Table 4 lists the top-scoring glycans of the CARB115 library that could be visualized and placed ligands at appropriate distance within the binding pocket. Top-scoring is Di-Sialosyl Galactosyl Globoside (DSGG) or di-sialosyl-Gb5, a globoside, which showed a high affinity of -25.4 kcal/mol (refined). Although the blood group I H (0) antigen scored with -15.5 kcal/mol (refined), the ganglioside GalNAc-GM1b interacted in this place with the refined affinity of -21.3 kcal/mol (-7.6 kcal/mol original score) exceeding the interaction energy defined in the approach above (Fig. 4). Ganglioside GM1b was found to interact with the affinity of -18.2 kcal/mol, several neolacto and lacto series GSLs scored with the affinity of -14.2 kcal/mol to -25.6 kcal/mol, and globo series GSL Gb4 (named P antigen / belonging to another “blood group system”), which is a precursor of the top-scoring DSGG, was defined in Autodock Vina with the re-dock affinity of -14.3 kcal/mol. Overall, when analysed with the hexameric heparin (gathered from 3ina), the increased energy of -29.7 kcal/mol could imply competitive interactions in the binding site of gangliosides, globosides etc. and heparins that may aid to deter the virus from cell binding.

The docking queue results are presented for the top-score DSGG in Fig. 5 with the Coulombic surface presentation of the S-protein NTD. The side chain locations of charged residues are named and indicated (left) and demonstrate the likely large binding area that is formed in-between. Very

demanding in computational task of docking is the large number of rotational degrees of freedom in particular with these positively charged residues and binding poses can only be approximated in the panel to the right (Fig. 5). For this task serial docking was applied where rigid receptor – flexible ligand and flexible receptor – rigid ligand docking was alternated to obtain the final pose. It was seen that the ligand was moving within the pocket from the left to right (Fig. 5, right panel) with side chains adapting to the new pose of similar energy (underlined). Moreover, terminal two saccharides were rotating with respect to the five residues at the reducing, ceramide end. If interaction with the globoside would prevail for a longer time-period, it could be envisioned, that conformational changes within the backbone of the SARS-CoV-2 NTD would be generated. These could be transmitted to another binding site or to the rest of the molecule. The interaction with ligands in this binding site is expected to tolerate few changes, the His69 is found in tyrosine His69Tyr sub-strains or as the discussed deletion B.1.1.7 mutant (in combination with the Val70 deletion since 2/20) that was studied with blood groups in detail above (Table 2). More work is necessary to elucidate the full panel of carbohydrates and glycolipid-headgroups that vastly exceeds computational capacities of even cluster-computations or supercomputing, since even several thousand ligands that harbour the very high torsional degrees of freedom would have to be docked to the entire surface. The first glimpse provided here and the data from datamonkey.org as well as the nextstrain.org list of mutants suggests that the loop with the Tyr145 and Trp152 indicated in the binding site – ligand interactions, is polymorph; it includes deletions of Val143 and Val143Phe replacements as well as the insertion of 2-15 amino acids, which makes it highly unlikely that a quick computational solution to the binding task will be installed.

In the next analysis, the top-scoring ligands of the SwissModel7C_26J (Table 4) were tested for interaction with the surface pocket of the SARS-CoV-1 S (47). The structure nearly corresponded to the energy minimized conformer with little change (RMSD 0.099 Å) and only Lys142, Glu174, and Asp204 in the putative binding site subject to minimal side-chain rotation when energy minimized. Although the 5X4S structure contained gaps and some amino acids had not been resolved, the ligands docked to the structurally resolved surface area within the neighbourhood of these four residues. In the Autodock approach the distinctly lower binding affinities of both, heparin (3ina) and DSGG, are shown (Fig. 6). In comparison, the Gb4 (P-Antigen) and GalNAc-GM1b interacted also stronger with the SwissModel7C_26J than with the SARS-CoV-1 S-protein. Other ligands showed mostly comparable affinities.

Finally, the recently published convalescent sera study was used to comparatively analyze the glycan binding site (48) (Suppl. Fig. 2). It appears, that the major antigenic site in the NTD (S-protein) would extend from Tyr144, His146 to Val143 and Leu141 that has now been defined. Only the first two residues are exposed, the residual amino-acids that grossly alter antigenicity are located to the interior of the domain and none of the amino acids in the binding site within direct proximity in rotamers of side-chains or side-chains themselves alter the antigenicity.

Discussion

Based on two recent analyses, I would like to suggest, that the putative glycan binding site established with this work on Autodock and carbohydrate ligands is not directly involved in “immune-escape”. This theory holds, that surface residues of viral proteins, constantly evade immune recognition by mutation and structural change and surface patches may also be indirectly affected by altering internal residues. Two most recent studies have mapped the immune epitopes recognized by the antibodies in humans. These are consistent with the assumption that monoclonal antibodies and convalescent sera against the SARS-CoV-2 Wuhan isolate bind to a surface area distinctly different from the surface patch surrounding His69 of the SARS-CoV-2 S-protein (48, 49), the putative glycan binding pocket.

Previous analyses in genetics have supported the role of glycans in the susceptibility of the human population to SARS-CoV-1 and -2 infection and/or severity of disease (COVID-19). Although different models have been suggested that could explain the relative or absolute protection of individuals with blood group H or O, the interaction of glycans with the S-protein itself had not been demonstrated. In this approach, the SwissModel generated conformer SwissModel7C_26J with a highest similarity to SARS-CoV-2 S-protein structure 7a25 C was automatically generated to maximize the fit to any structural entry available in the end of January 2021 (30). The model differed by only 9 amino acids to the reported structure 7a25 with residues introduced by the modelling (amino acids 71-75 and 248-251). Since it is to be expected, that SARS-CoV-2 just as many other viridae that incorporated a lectin domain during evolution, may bind to carbohydrates of distinct structure the Autodock Vina approach was further tested for the carbohydrate interaction. The approach is criticized by some due to the lack of modelling of pi-interactions and force field changes have been introduced in the novel modelling methods (50) wherein each carbohydrate-pi interaction may, however, contribute 0.8–1.0 kcal/mol. In the described binding site (Figs. 4 and 5) glycans in the vicinity could (with the static structure) contribute only little. These can possibly contact the rings of Trp64, Tyr145, Phe186 and Trp258, but the glycans are, in the docking poses, positioned at or largely exceeding the d_{Cx} distance exclusion limit of 4.5 Å (51). In contrast, with blood type antigens several poses have been found that would allow some pi interactions in particular with Gal and Fuc to Tyr145 in the wildtype S-protein, or of Fuc with the Trp152 or Phe186 (according to wildtype numbering). Whereas the expected energies in scoring would thus not differ in the screening run with the general CARB115 library, it may be worthwhile and affordable to use high-precision force fields and molecular dynamics to generate a sufficient ranking of blood type antigen interactions. Visually inspecting the binding site environment, it could be inferred from Coulombic surface colouring (Fig. 5), that non-blood group ligands would be attracted by low-affinity, transient binding events that may include charged groups of heparin, proteoglycans or sialylated molecules. Low-affinity interaction would then be followed by high-affinity induced fit.

The blood groups associated with the SARS-CoV-2 infection and severity of disease could not be identified in this study and interpreted in an easy way. However, when comparing the protein conformers of the predicted wildtype S-protein NTD with the mutant B.1.1.7 which harbours the His69Val70Tyr145 deletion, a consistent observation is the highly increased affinity of a glycolipid of the A type II antigen (No. 5). Apparently, a H (0) type III antigen interacted less in the mutant B.1.1.7 strain. The type III B antigen that was included in this study, was measured to complete the series of lipidic antigens that may be produced in the human body, but is described so far linked to O-glycans: The enzymatic reaction of the A- or B-transferase (ABO) may link terminal Gal- just as GalNAc-residues to the type III precursor. Since the type III A GSL has been found (LipidMaps) it is a matter of further research, to elucidate the full sphingolipid glycome. This particular GSL, however, interacted less with

the B.1.1.7 mutant clade S-protein NTD and it may allow to speculate, that a large variety of change to tropism may set in once a glycan binding site has altered in specificity, even if single linkages only were recognized differently. I would like to suggest that the terminal GalNAc of blood group A would be bound, yet, the affinity of interaction does currently not allow to pinpoint towards the exact binding site geometry. Only the large screen with the CARB115 library has allowed to collect ligands of highest binding affinity that may allow to conclude, that the His69<->Lys182 central binding area is most often filled with Neu5Ac or N-acetylated glycan residues. However, results of the previous docking study on the S-protein, demonstrating Neu5Ac bound to the NTD (52) were found to be largely discordant with my present result (Suppl. Fig. 3). The S-protein structure that was used at that time included larger gaps and depended on simulation for a large fraction of residues including the N-terminal domain. Yet, since the structures of ABH determinants are found on N-, O-glycans as well as glycolipids and the type I, II and III form is, for example, expressed in gastrointestinal tissues (4–6) this study could alert to a change in tissue tropism that may adapt the SARS-CoV-2 to conform to the clinical view on other coronaviridae including SARS-CoV-1 (53). Gastrointestinal symptoms had been more often reported with the ancient SARS-CoV or MERS-CoV.

The ligand with the current top scoring affinity of -26.1 kcal/mol (Fig. 5) DSGG fully fills the binding pocket and likely would contact residues in similar locations to the Asp72Asn and Ala219Ser that have been defined previously in the Transmissible GastroEnteritis corona Virus (TGEV) of piglets (2). These mutations have been found to alter tissue tropism from the respiratory and gastrointestinal system towards the respiratory tract. Growth of the TGEV was measured in different tissues and established a correlation to define the tropism measured. Binding of viral S-protein to the cell surface aminopeptidase N, the proteinaceous viral receptor, may be enhanced by bivalent interaction of the S-protein to the protein receptor and to glycans on the host. Expression of MSGG (Mono Sialosyl Galactosyl Globoside), the desialylated DSGG, and of DSGG is found in human erythrocytes and in kidney within the distal tubule and Henle's loop (44). GSL expression can vary in different tissues and MSGG has, for example, been characterized in embryonic stem cells, dorsal root ganglia and tumour tissues. Parvovirus B19 (54), in contrast to SARS-CoV-2, causes anemia due to erythrocyte infection. This is likely due to binding of Gb4s (P antigen), Gb5 and MSGG among others. Although the similar binding profile could be ascribed to the SARS-CoV-2 virus with a differential binding mode, the aplastic anemia has only been observed in a single case (55, 56) and clearly co-receptors are the major determinant of the observed respiratory tract interaction and viral uptake, the ACE2 receptor. Complexity increases, when relegating part or all of the initial SARS-CoV-2 interactions to the glycan shield and glycan-glycan interactions of coronaviridae which is essentially unexplored, in simulations as well as in biochemical studies (57–59). Finally, when considering zoonosis and anthropozoonotic cycles of infection, it remains to be shown whether influenza viridae are teaching a lesson suggesting, that although lectin domains are displayed on the viral surface, glycan interactions seem sometimes non-essential (60–62). The differences of lectin-activities of SARS-CoV-1, if any, and SARS-CoV-2 S-protein (Fig. 6) remain to be analysed in high resolution and structurally in the future.

References

1. L. Cooling, Blood Groups in Infection and Host Susceptibility. *Clin. Microbiol. Rev.* **28**, 801–870 (2015).
2. M. L. Ballesteros, C. M. Sánchez, L. Enjuanes, Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism. *Virology*. **227**, 378–388 (1997).
3. H. Chu, J. F.-W. Chan, T. T.-T. Yuen, H. Shuai, S. Yuan, Y. Wang, B. Hu, C. C.-Y. Yip, J. O.-L. Tsang, X. Huang, Y. Chai, D. Yang, Y. Hou, K. K.-H. Chik, X. Zhang, A. Y.-F. Fung, H.-W. Tsoi, J.-P. Cai, W.-M. Chan, J. D. Ip, A. W.-H. Chu, J. Zhou, D. C. Lung, K.-H. Kok, K. K.-W. To, O. T.-Y. Tsang, K.-H. Chan, K.-Y. Yuen, Comparative tropism, replication kinetics, and cell damage profiling of SARS-CoV-2 and SARS-CoV with implications for clinical manifestations, transmissibility, and laboratory studies of COVID-19: an observational study. *The Lancet. Microbe*. **1**, e14–e23 (2020).
4. J. Finne, M. E. Breimer, G. C. Hansson, K. A. Karlsson, H. Leffler, J. F. Vliegenthart, H. van Halbeek, Novel polyfucosylated N-linked glycopeptides with blood group A, H, X, and Y determinants from human small intestinal epithelial cells. *J. Biol. Chem.* **264**, 5720–5735 (1989).
5. M. E. Breimer, G. C. Hansson, K.-A. Karlsson, G. Larson, H. Leffler, Glycosphingolipid composition of epithelial cells isolated along the villus axis of small intestine of a single human individual. *Glycobiology*. **22**, 1721–1730 (2012).
6. C. Jin, D. T. Kenny, E. C. Skoog, M. Padra, B. Adamczyk, V. Vitizeva, A. Thorell, V. Venkatakrishnan, S. K. Lindén, N. G. Karlsson, Structural diversity of human gastric mucin glycans. *Mol. Cell. Proteomics*. **16**, 743–758 (2017).
7. D. Lingwood, K. Simons, Lipid rafts as a membrane-organizing principle. *Science*. **327**, 46–50 (2010).
8. P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. **579**, 270–273 (2020).
9. ECDC, “European Centre for Disease Prevention and Control. Risk related to spread of new SARS-CoV-2 variants of concern in the EU/EEA – 29 December 2020” (Stockholm, 2020).
10. S. A. Kemp, B. Meng, I. A. T. M. Ferriera, R. Datir, W. T. Harvey, D. A. Collier, S. Lytras, G. Papa, The COVID-19 Genomics UK (COG-UK) Consortium, A. Carabelli, J. Kenyon, A. Lever, L. C. James, D. Robertson, R. Gupta, Recurrent Emergence and Transmission of a SARS-CoV-2 Spike Deletion H69/V70 (2021), doi:10.2139/ssrn.3780277.
11. V. G. Puelles, M. Lütgehetmann, M. T. Lindenmeyer, J. P. Sperhake, M. N. Wong, L. Allweiss, S. Chilla, A. Heinemann, N. Wanner, S. Liu, F. Braun, S. Lu, S. Pfefferle, A. S. Schröder, C. Edler, O. Gross, M. Glatzel, D. Wichmann, T. Wiech, S. Kluge, K. Pueschel, M. Aepfelbacher, T. B. Huber, Multiorgan and Renal Tropism of SARS-CoV-2. *N. Engl. J. Med.* (2020), doi:10.1056/NEJMc2011400.
12. A. T. Irving, M. Ahn, G. Goh, D. E. Anderson, L.-F. Wang, Lessons from the host defences of bats, a unique viral reservoir. *Nature*. **589**, 363–370 (2021).
13. N. Liu, T. Zhang, L. Ma, H. Zhang, H. Wang, W. Wei, H. Pei, H. Li, The impact of ABO blood

group on COVID-19 infection risk and mortality: A systematic review and meta-analysis. *Blood Rev.*, **100**785 (2020).

- 14. D. Golinelli, E. Boetto, E. Maietti, M. P. Fantini, The association between ABO blood group and SARS-CoV-2 infection: A meta-analysis. *PLoS One.* **15** (2020), p. e0239508.
- 15. M. Franchini, F. Capra, G. Targher, M. Montagnana, G. Lippi, Relationship between ABO blood group and von Willebrand factor levels: from biology to clinical implications. *Thromb. J.* **5**, 14 (2007).
- 16. V. Ramlall, P. M. Thangaraj, C. Meydan, J. Foox, D. Butler, J. Kim, B. May, J. K. De Freitas, B. S. Glicksberg, C. E. Mason, N. P. Tatonetti, S. D. Shapira, Immune complement and coagulation dysfunction in adverse outcomes of SARS-CoV-2 infection. *Nat. Med.* **26**, 1609–1615 (2020).
- 17. J. G. Kelton, C. Hamid, S. Aker, M. A. Blajchman, The amount of blood group A substance on platelets is proportional to the amount in the plasma. *Blood.* **59**, 980–985 (1982).
- 18. J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* **581**, 215–220 (2020).
- 19. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature.* **579**, 265–269 (2020).
- 20. A. Rambaut, N. Loman, O. Pybus, W. Barclay, J. Barrett, A. Carabelli, T. Connor, T. Peacock, D. L. Robertson, E. Volz, C.-19 G. C. UK, Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *arXiv* (2020), (available at <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>).
- 21. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- 22. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
- 23. N. M. O’Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, Open Babel: An open chemical toolbox. *J. Cheminf.* **3**, 33 (2011).
- 24. M. F. Sanner, Python: A programming language for software integration and development. *J. Mol. Graph.* **17**, 57–61 (1999).
- 25. J. Mills, P. Dean, Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput. Aided Mol. Des.* **10**, 607–622 (1996).
- 26. E. Krissinel, Crystal contacts as nature’s docking solutions. *J. Comput. Chem.* **31**, 133–143 (2010).
- 27. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis.* **18**, 2714–2723 (1997).
- 28. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **216**, 403–410 (1990).
- 29. M. Steinegger, M. Meier, M. Mirdita, H. Vöhringer, S. J. Haunsberger, J. Söding, HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics.* **20**, 473 (2019).

30. T. F. Custódio, H. Das, D. J. Sheward, L. Hanke, S. Pazicky, J. Pieprzyk, M. Sorgenfrei, M. A. Schroer, A. Y. Gruzinov, C. M. Jeffries, M. A. Graewert, D. I. Svergun, N. Dobrev, K. Remans, M. A. Seeger, G. M. McInerney, B. Murrell, B. M. Hällberg, C. Löw, Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. *Nat. Commun.* **11**, 5588 (2020).

31. A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell.* **180**, 1–12 (2020).

32. D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science.* **367**, 1260–1263 (2020).

33. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **25**, 247–260 (2006).

34. A. Fiser, A. Sali, Modeller: generation and refinement of homology-based protein structure models. *Methods Enzymol.* **374**, 461–91 (2003).

35. M.-Y. Shen, A. Sali, Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **15**, 2507–24 (2006).

36. S. R. Comeau, D. W. Gatchell, S. Vajda, C. J. Camacho, ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics.* **20**, 45–50 (2004).

37. S. Y. Kim, W. Jin, A. Sood, D. W. Montgomery, O. C. Grant, M. M. Fuster, L. Fu, J. S. Dordick, R. J. Woods, F. Zhang, R. J. Linhardt, Characterization of heparin and severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) spike glycoprotein binding interactions. *Antiviral Res.* **181**, 104873 (2020).

38. P. S. Kwon, H. Oh, S.-J. Kwon, W. Jin, F. Zhang, K. Fraser, J. J. Hong, R. J. Linhardt, J. S. Dordick, Sulfated polysaccharides effectively inhibit SARS-CoV-2 in vitro. *Cell Discov.* **6**, 50 (2020).

39. T. M. Clausen, D. R. Sandoval, C. B. Spliid, J. Pihl, H. R. Perrett, C. D. Painter, A. Narayanan, S. A. Majowicz, E. M. Kwong, R. N. McVicar, B. E. Thacker, C. A. Glass, Z. Yang, J. L. Torres, G. J. Golden, P. L. Bartels, R. N. Porell, A. F. Garretson, L. Laubach, J. Feldman, X. Yin, Y. Pu, B. M. Hauser, T. M. Caradonna, B. P. Kellman, C. Martino, P. L. S. M. Gordts, S. K. Chanda, A. G. Schmidt, K. Godula, S. L. Leibel, J. Jose, K. D. Corbett, A. B. Ward, A. F. Carlin, J. D. Esko, SARS-CoV-2 Infection Depends on Cellular Heparan Sulfate and ACE2. *Cell.* **183**, 1043–1057.e15 (2020).

40. J. Lang, N. Yang, J. Deng, K. Liu, P. Yang, G. Zhang, C. Jiang, Inhibition of SARS pseudovirus cell entry by lactoferrin binding to heparan sulfate proteoglycans. *PLoS One.* **6**, e23710–e23710 (2011).

41. Y.-H. Han, M.-L. Garron, H.-Y. Kim, W.-S. Kim, Z. Zhang, K.-S. Ryu, D. Shaya, Z. Xiao, C. Cheong, Y. S. Kim, R. J. Linhardt, Y. H. Jeon, M. Cygler, Structural Snapshots of Heparin Depolymerization by Heparin Lyase I. *J. Biol. Chem.* **284**, 34019–34027 (2009).

42. N. Eswar, D. Eramian, B. Webb, M.-Y. Shen, A. Sali, Protein structure modeling with MODELLER. *Methods Mol. Biol.* **426**, 145–59 (2008).

43. J. Fantini, H. Chahinian, N. Yahi, Leveraging coronavirus binding to gangliosides for innovative vaccine and therapeutic strategies against COVID-19. *Biochem. Biophys. Res. Commun.* (2020), doi:10.1016/j.bbrc.2020.10.015.

44. S. Saito, S. B. Levery, M. E. Salyan, R. I. Goldberg, S. Hakomori, Common tetrasaccharide epitope NeuAcalpha(2-3)Galbeta(1-3)[NeuAcalpha(2-6)]GalNAc, presented by different carrier

glycosylceramides or O-linked peptides, is recognized by different antibodies and ligands having distinct specificities. *J. Biol. Chem.* **269**, 5644–5652 (1994).

- 45. K. Fiedler, The Wnt segment polarity pathway and TMED2 protein may interact via a lectin- and decoy-type mechanism. *bioRxiv* (2016), doi:10.1101/056531.
- 46. K. Fiedler, VIP36 preferentially binds to core-fucosylated N-glycans: a molecular docking study. *bioRxiv* (2016), doi:10.1101/092460.
- 47. Y. Yuan, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, X. Zhang, G. F. Gao, Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat. Commun.* **8**, 15092 (2017).
- 48. K. R. McCarthy, L. J. Rennick, S. Nambulli, L. R. Robinson-McCarthy, W. G. Bain, G. Haidar, W. P. Duprex, Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **371**, 1139–1142 (2021).
- 49. P. Wang, M. S. Nair, L. Liu, S. Iketani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, P. D. Kwong, B. S. Graham, J. R. Mascola, J. Y. Chang, M. T. Yin, M. Sobieszczyk, C. A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, D. D. Ho, Antibody Resistance of SARS-CoV-2 Variants B.1.351 and B.1.1.7. *Nature* (2021), doi:10.1038/s41586-021-03398-2.
- 50. D. Neumann, O. Kohlbacher, Structural Glycomics – Molecular Details of Protein-Carbohydrate Interactions and their Prediction. *Glyco-Bioinformatics, Beilstein-Institut*, 101–122 (2010).
- 51. A. Kerzmann, D. Neumann, O. Kohlbacher, SLICK - Scoring and Energy Functions for Protein-Carbohydrate Interactions. *J. Chem. Inf. Model.* **46**, 1635–1642 (2006).
- 52. M. Awasthi, S. Gulati, D. P. Sarkar, S. Tiwari, S. Kateriya, P. Ranjan, S. K. Verma, The Sialoside-Binding Pocket of SARS-CoV-2 Spike Glycoprotein Structurally Resembles MERS-CoV. *Viruses* **12**, 909 (2020).
- 53. B. Chen, E.-K. Tian, B. He, L. Tian, R. Han, S. Wang, Q. Xiang, S. Zhang, T. El Arnaout, W. Cheng, Overview of lethal human coronaviruses. *Signal Transduct. Target. Ther.* **5**, 89 (2020).
- 54. J. Qiu, M. Söderlund-Venermo, N. S. Young, Human Parvoviruses. *Clin. Microbiol. Rev.* **30**, 43–100 (2017).
- 55. P. E. Taneri, S. A. Gómez-Ochoa, E. Llanaj, P. F. Raguindin, L. Z. Rojas, Z. M. Roa-Díaz, D. Salvador, D. Groothof, B. Minder, D. Kopp-Heim, W. E. Hautz, M. F. Eisenga, O. H. Franco, M. Glisic, T. Muka, Anemia and iron metabolism in COVID-19: a systematic review and meta-analysis. *Eur. J. Epidemiol.* **35**, 763–773 (2020).
- 56. M. Figlerowicz, A. Mania, K. Lubarski, Z. Lewandowska, W. Służewski, K. Derwich, J. Wachowiak, K. Mazur-Melewska, First case of convalescent plasma transfusion in a child with COVID-19-associated severe aplastic anemia. *Transfus. Apher. Sci.* **59**, 102866 (2020).
- 57. Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* **369**, 330–333 (2020).
- 58. H. Yao, Y. Song, Y. Chen, N. Wu, J. Xu, C. Sun, J. Zhang, T. Weng, Z. Zhang, Z. Wu, L. Cheng, D. Shi, X. Lu, J. Lei, M. Crispin, Y. Shi, L. Li, S. Li, Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **183**, 730-738.e13 (2020).
- 59. A. C. Walls, M. A. Tortorici, B. Frenz, J. Snijder, W. Li, F. A. Rey, F. DiMaio, B.-J. Bosch, D. Veesler, Glycan shield and epitope masking of a coronavirus spike protein observed by cryo-electron microscopy. *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
- 60. X. Sun, Y. Shi, X. Lu, J. He, F. Gao, J. Yan, J. Qi, G. F. Gao, Bat-derived influenza hemagglutinin H17 does not bind canonical avian or human receptors and most likely uses a unique entry

mechanism. *Cell Rep.* **3**, 769–778 (2013).

61. S. Tong, X. Zhu, Y. Li, M. Shi, J. Zhang, M. Bourgeois, H. Yang, X. Chen, S. Recuenco, J. Gomez, L.-M. Chen, A. Johnson, Y. Tao, C. Dreyfus, W. Yu, R. McBride, P. J. Carney, A. T. Gilbert, J. Chang, Z. Guo, C. T. Davis, J. C. Paulson, J. Stevens, C. E. Rupprecht, E. C. Holmes, I. A. Wilson, R. O. Donis, New World Bats Harbor Diverse Influenza A Viruses. *PLOS Pathog.* **9**, e1003657 (2013).

62. X. Zhu, W. Yu, R. McBride, Y. Li, L.-M. Chen, R. O. Donis, S. Tong, J. C. Paulson, I. A. Wilson, Hemagglutinin homologue from H17N10 bat influenza virus exhibits divergent receptor-binding and pH-dependent fusion activities. *Proc. Natl. Acad. Sci.* **110**, 1458–1463 (2013).

Figure Legends

Figure 1:

SARS-CoV-2 S-protein interaction with heparin. S-protein domains NTD (amino acids 14-291) and RBD (amino acids 334-524) were submitted for molecular ligand docking and results overlaid on the complete S-protein structure. The side view lacking the membrane proximal, transmembrane and cytoplasmic domains is presented on the left, the top “crown-view” is shown on the right with heparin presented with the pose that was obtained from ClusPro docking with lowest energy indicated. The current number of amino acids in Swiss-Model queue prediction is indicated (green) and more SARS-CoV-2 high-resolution structures are expected to validate heparin interaction in the future. Monomers are indicated with the chain A, B or C, separate coloring is shown in RBD and NTD backbone with the “crown-view”.

Figure 2:

Lowest energy interactions analysed in Autodock-re-dock. Only partial overlap of low energy poses obtained were confirmed in local re-docking and some amino-acids did not coincide with the lowest energy ligand conformer and/or energy of side-chain rotamers. Autodock energies in refinement are indicated and comparable to energies shown in the rest of the work. A, B, C NTDs and A, B, C RBDs are displayed.

Figure 3:

The molecular dynamics conformer of SARS-CoV-2 S-protein and blood group type interactions. **(A)** Hypothetical interactions are demonstrated by drug docking using a multithreaded procedure that is only partially available for glycan docking: Small glycan residues have previously been tested, the procedure is here used for glycans, that may be exceeding the computational capacity/force-field adjustments of Autodock (22) with difficult binding sites. The NTD was subjected to Autodock docking, re-docking in refinement with the model generated by Modeller of the SARS-CoV-2 (wildtype, B.1.1.7 mutant) S-protein NTD. The molecular dynamics conformer was obtained by a standard run in Chimera with a thermostat of 298 K and clustering with conformers in the equilibrated phase. The graph shows the binding energy of re-docking of each individual glycolipid “blood type” with underlaid green (type I), in blue (type II), red (type III) and ochre (type IV). The British S-protein NTD (lineage B.1.1.7 in orange) mutant and wildtype S-protein NTD (blue) are indicated. Numbering and structural (IUPAC) formulae are shown in the accompanying table. **(B)** A significant difference is found with the British S-protein NTD (lineage B.1.1.7 in orange) mutant for interaction of type A and H (0) ($p=0.04$ Mann-Whitney test). The wildtype S-protein NTD results are shown in blue. “Attached” molecular dynamics with fixed residues did not allow to model a suitable ligand binding pose, and model molecular dynamics of the full length trimer of SARS-CoV-2 S were not yet available from covid.molssi.org. Error bars are indicated with the confidence interval (CI) presented with an $\alpha=0.05$. The significant difference of type A and H (0) was also obtained when glycolipid 11 was left out in (B), one of the duplicates 6 incorporated for testing exhaustiveness (A) was deleted from results for the graph (B) and only top-scores were retained.

Figure 4:

Docking to SARS-CoV-2 N-terminal domain (NTD) residues. Autodock Vina was utilized for interaction screen (box size in Å x = 45.5, y = 31.1, z = 53.4) of carbohydrates shown in the accompanying tables. The DSGG (sialosyl-MSGG, also called di-sialosyl-Gb5, di-sialosyl-Gal β 1-3-Gb4) and GalNAc β 1-4-GM1b are shown for comparison. Sites and amino acids within the proximity of 3 Å are listed. Previously identified residues are shown for comparison and printed in grey if found in proximity with GalNAc-GM1b.

Figure 5:

Surface presentation of the SARS-CoV-2 NTD with half-side view onto putative binding sites of glycans. The surface is colored by Coulombic electrostatic surface charges, the ligand is colored by the indicated IUPAC code and major side-chain rotations in refinement are: Asn74, Trp152, Lys182, Gln183, Asn185, Arg214 and Arg246 (underlined). Energies gathered in the refined poses were increased from -7.7 kcal/mol to -10.2 kcal/mol and corresponded to the -26.1 kcal/mol and -26.0 kcal/mol obtained in the local or freely-rotating side-chain poses, respectively. Computational resources for the overall approach of no restriction to backbone movements and/or freely rotating side-chains in ligands docked without restricted torsional degrees of freedom were not available. Charged six Lys and one Arg amino acids in the binding site are denoted. The likely location of ceramide is indicated. Glycans are colored in IUPAC style yellow Gal and GalNAc, blue Glc and purple NeuAc.

Figure 6:

The surface binding of glycans to the S-protein of SARS-CoV-1 and -2 was comparatively analyzed. The S-protein NTD of SARS-CoV-1 5x4s was docked to the top-scores of the SwissModel7C_26J docking run glycolipid headgroup glycans. Number 1 to 16 are labelled and graphed to the right in IUPAC style colors yellow Gal and GalNAc, blue Glc and GlcNAc, red Fuc, purple NeuAc, white/blue GlcN and brown/white IdoA.

Tables

Table 1:

Original poses of ClusPro high affinity interactions and residues in the proximity (5 Å). S-protein domains NTD (amino acids 14-291) and RBD (amino acids 334-524) were analysed for proximity to residues in 5 Å, chains are denoted with A, B, C and colored as shown in the molecular overview (Fig. 2).

Table 2:

Blood group type antigens presented on glycolipids. Various blood type antigens terminally linked in glycosidic bonding on sphingolipids are shown and grouped as defined. Type I [B-D-Galp-(1-3)-B-D-GlcNAc-r], type II [B-D-Galp-(1-4)-B-D-GlcNAc-r], type III [B-D-Galp-(1-3)-A-D-GalpNAc-r] and type IV [B-D-Galp-(1-3)-B-D-GalpNAc-r] are indicated and denoted with the respective categories. Listed blood type antigens and the numbering is used throughout the work. The glycan 11 determinant is presented on protein-linked O-glycans, in biosynthesis the same transferase likely uses the ligand A-L-Fucp-(1-2)-B-D-Galp-r for transfer of A-D-GalpNAc (blood type III A, number 10) or A-D-Galp (blood type III B) in structural isoform (transferases A and B), the enzymatic interaction with ligands hinges upon the B-D-Galp interaction, water may be displaced if type III ligands are converted instead, for example, in a reaction with structurally characterised transfer to B-D-Galp-(1-4)-B-D-GlcNAc-r (type II ligand).

Table 3:

Carbohydrate-interaction screen of the SARS-CoV-2 S-protein NTD. Carbohydrate ligands utilized in Vina are indicated and listed with their common names. Ligands not expressed or metabolically produced in humans, or only found in very rare cell types and as human polymorphisms are indicated (*). Formulae (IUPAC style) are provided in the Supplementary Table.

Table 4:

Carbohydrate screen for the local docking to the identified binding site, list of top-scores. The box size x = 39.2, y = 26.5, z = 28.9 was used for the Autodock Vina screen, screen energies are listed (black) and refined local autodocking energies are indicated in green. These correspond to local energies obtained in SMINA. The shared terminal epitope of DSSG (Sialosyl-MSGG or Disialosyl-Gb5) found in GD1 α was bound in grossly similar configuration to the S-protein NTD with N-acetylated residue GalNAc within the central binding pocket and with Autodock Vina affinity of -6.8/-18.4 kcal/mol. In this binding, the reducing end was likely not available and only partial low affinity binding to GD1 α would be expected. Categories of glycolipids are denoted with series name and IUPAC formulae are indicated.

Supplementary Figures:**Supplementary Fig. 1:**

The conformer obtained from the ClusPro queue was analysed and compared. SwissModel7C_26J was subjected to ClusPro docking with heparin. The differential of interaction of SwissModel7C_26J with “SwissModel7C_26J heparin” is graphed and side-chains that do not overlap and exclude or facilitate ligand binding are indicated for some CARB115 examples.

Supplementary Fig. 2:

Amino-acids identified in McCarthy et al. (48) are labelled on the SARS-CoV-2 S-protein NTD and viewed from different angles. The “imprint” of top-scoring 15 glycans neglecting heparin is shown at the surface (see Tables) to demonstrate the size of the binding-site. Residues identified in (48) are labelled with an asterisk if without effect in the tissue-culture assay (antisera binding or antisera neutralization, cf. Table S1). Residues labelled are visible from the respective side. Small font is applied if residues cannot be detected at the model or protein (7a25) surface. Glycans are colored in IUPAC style yellow Gal and GalNAc, blue Glc and purple NeuAc.

Supplementary Fig. 3:

A previous study suggested interactions of sialic-acid residues with S-proteins and in particular with the SARS-CoV-2 and was compared with present docking results (52). The ligands found here did largely not overlap and did not contact N-terminal residues shown (Leu18, Thr20). Structures varied, results were not comparable since not generated from a modelling queue, and the present SwissModel includes only 9 residues that were subject to modelling and the previous attempt modelled a quarter of the entire N-terminal domain.

Figure 1

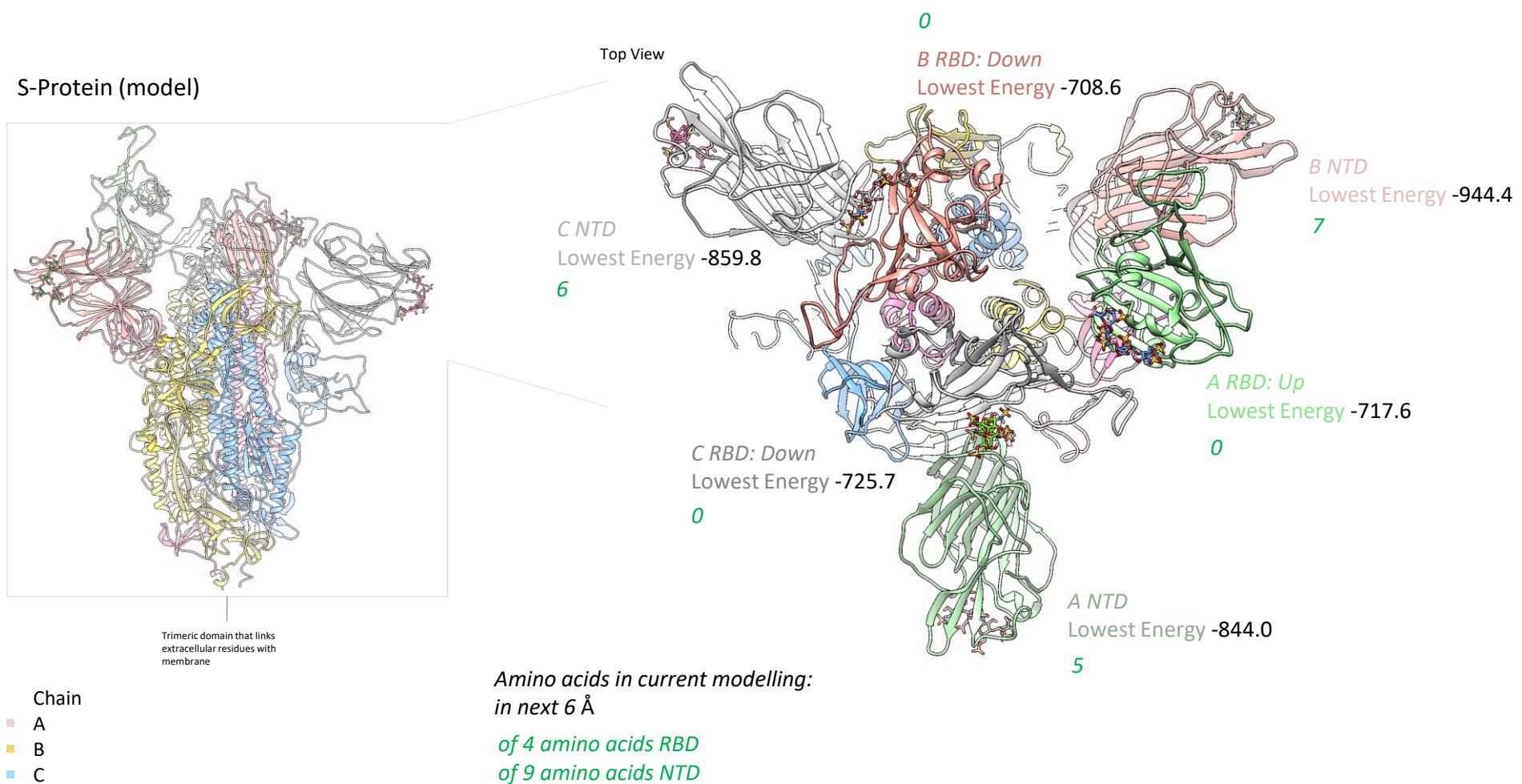
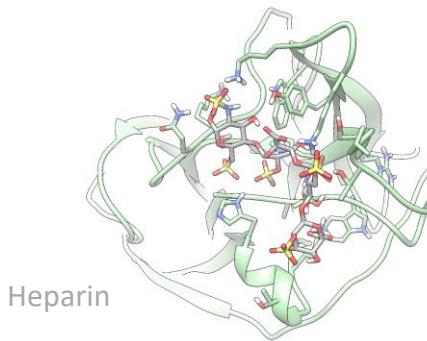
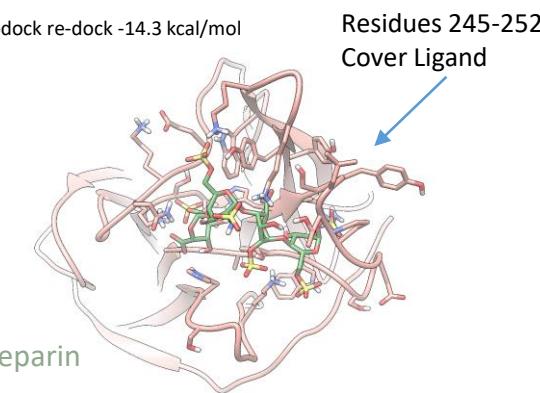


Figure 2

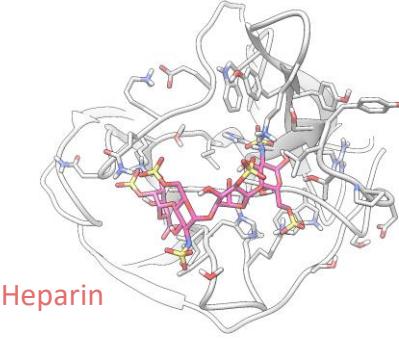
A NTD
Lowest Energy **-844.0**
Autodock re-dock -15.5 kcal/mol



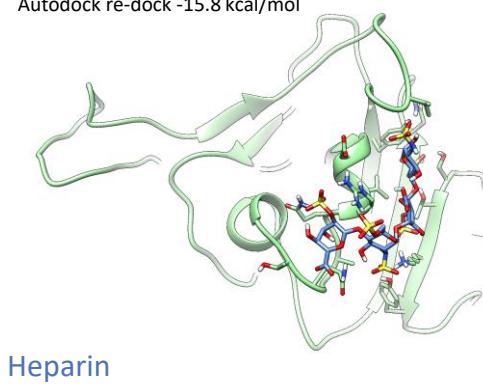
B NTD
Lowest Energy **-944.4**
Autodock re-dock -14.3 kcal/mol



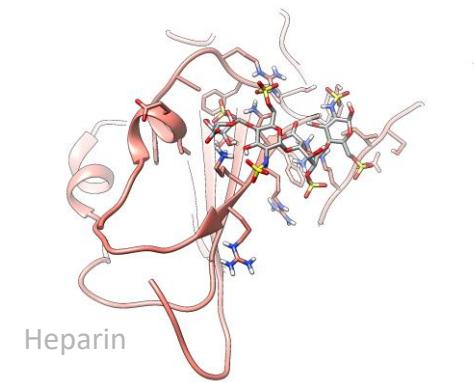
C NTD
Lowest Energy **-859.8**
Autodock re-dock -15.2 kcal/mol



A RBD: Up
Lowest Energy **-717.6**
Autodock re-dock -15.8 kcal/mol



B RBD: Down
Lowest Energy **-708.6**
Autodock re-dock -14.4 kcal/mol



C RBD: Down
Lowest Energy **-725.7**
Autodock re-dock -13.5 kcal/mol

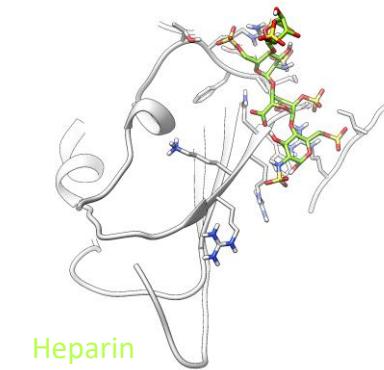


Figure 5

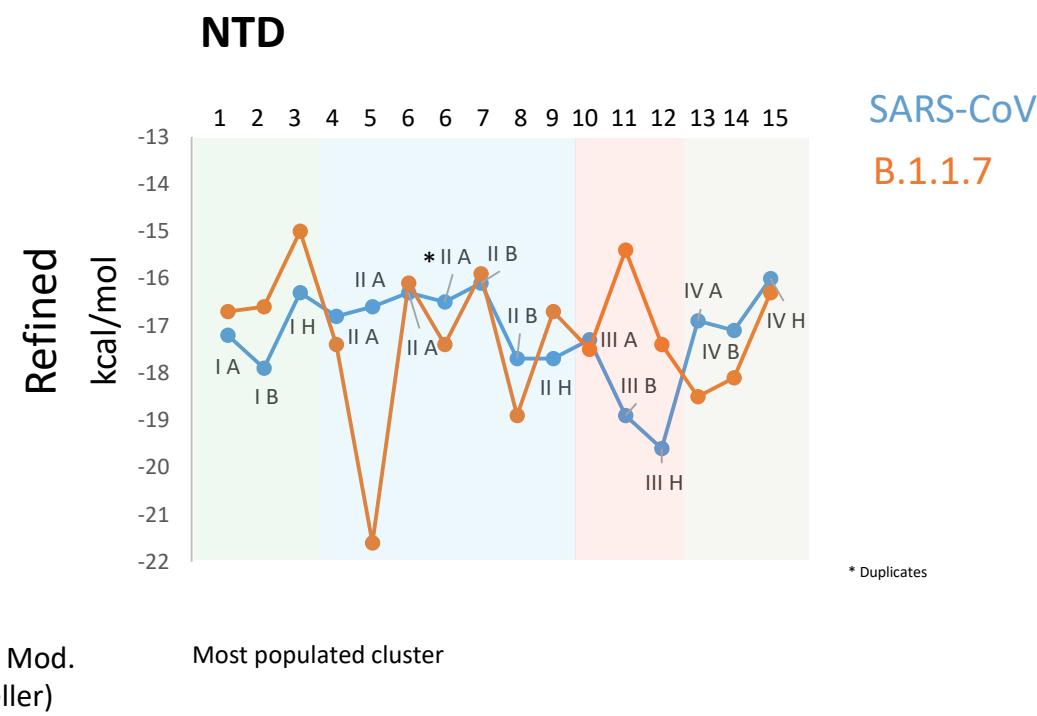
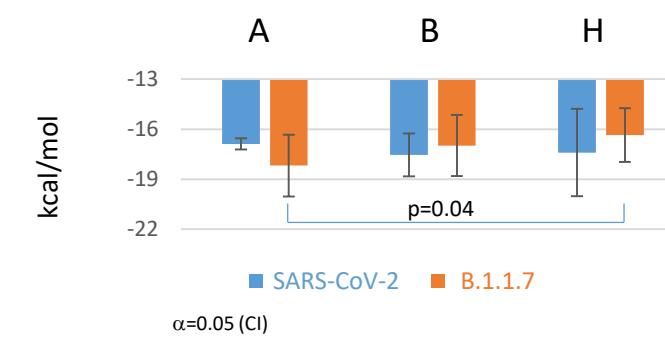
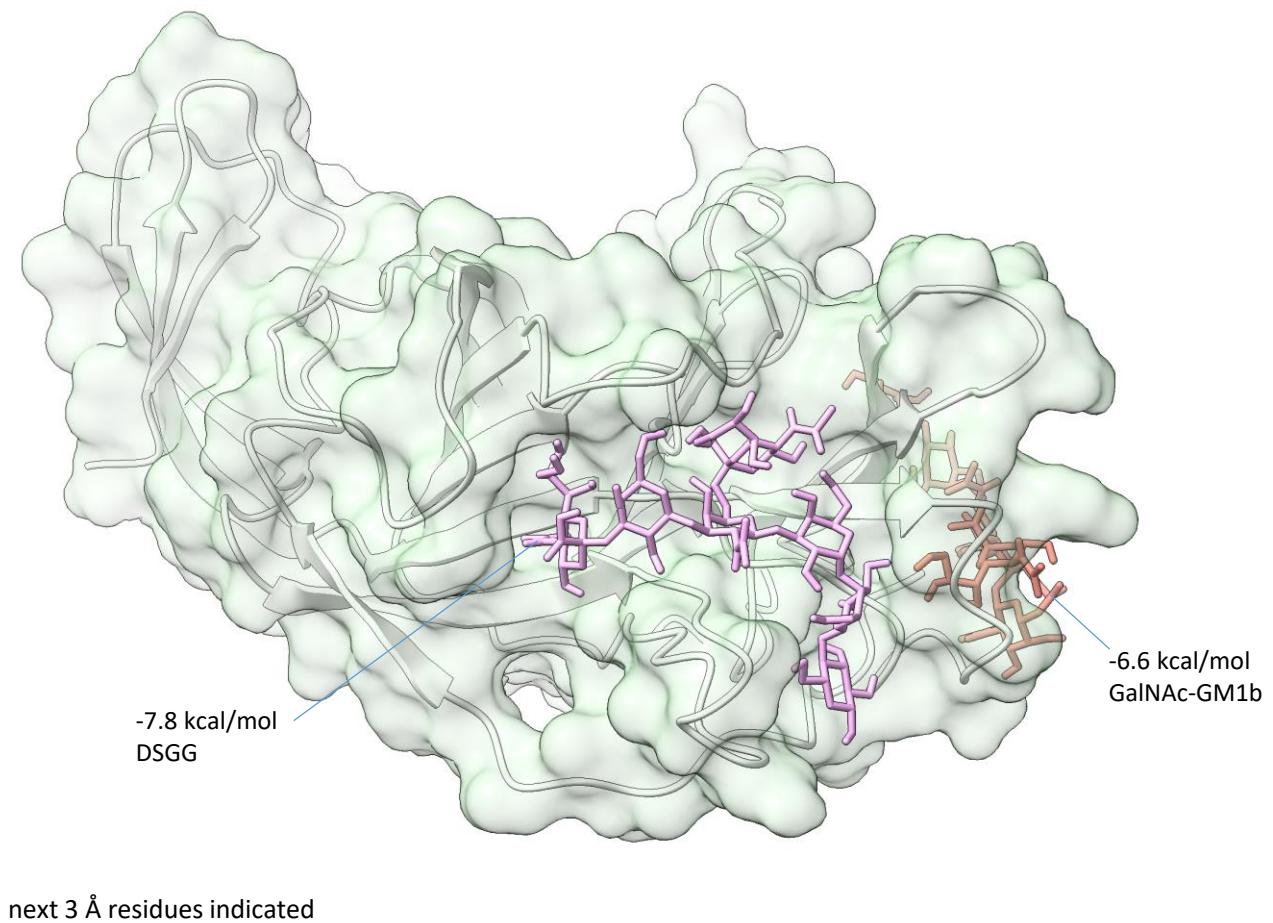
A**B**

Figure 4



DSGG Thr20, Ile68, His69, Lys77, Lys97, Ser98, Gly184, Phe186, Arg214, His245, Leu249, Thr250, Ser256, Trp258, Thr259, Ala262

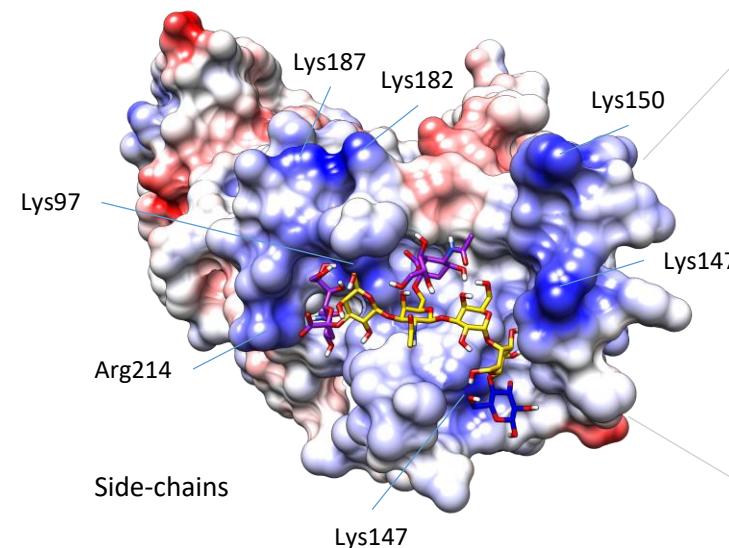
GalNAc-GM1b Gln14 (N-ter), Asn122, Glu154, Ser155, Glu156, Phe157, Arg158, Arg246, Ser247, Gly252, Ser254, Gly257

Previously identified residues in the absence of the modelled N-terminus may include:
Asp111, Ser112, Lys113, Gln134, Phe135, Cys136, Asn137, Phe140, Gly142, Glu156, Phe157, Arg158, Tyr160, Ser161, Ser162

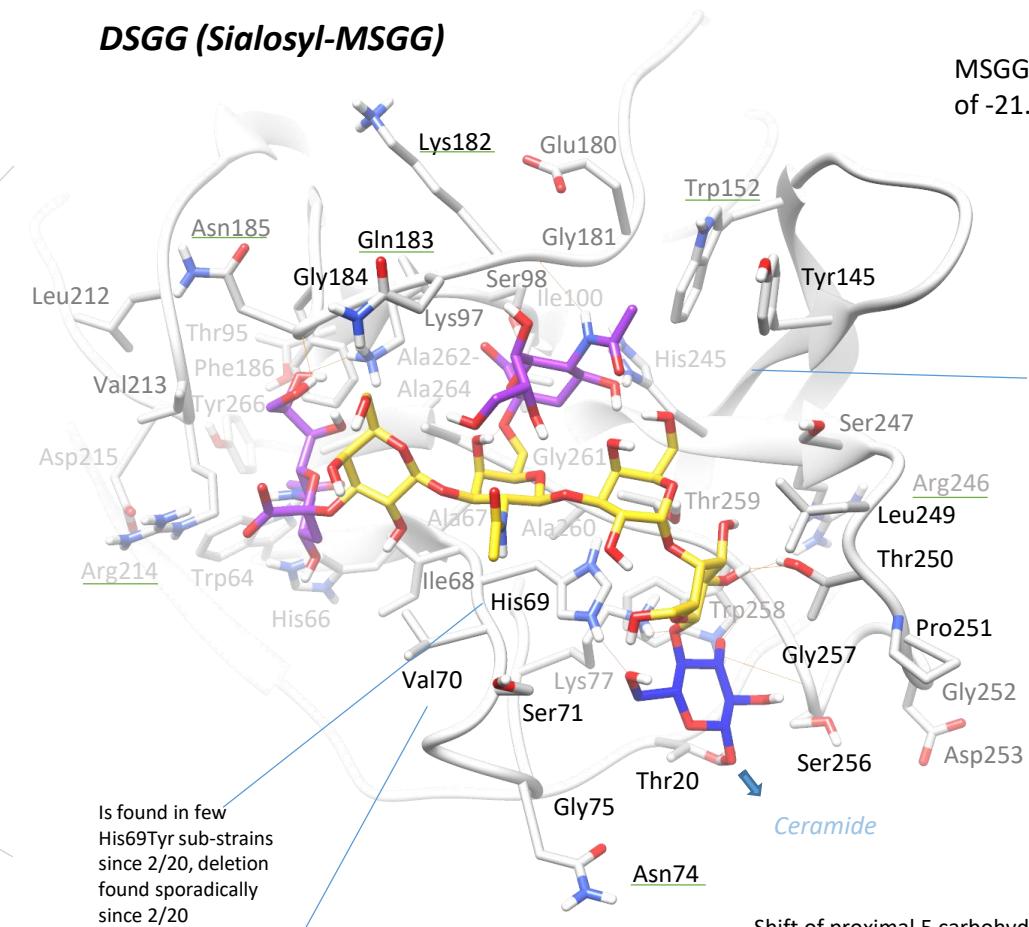
Figure 5

S Protein NTD

7a25 chain C



● Gal ■ GalNAc ● Glc ◆ NeuAc

DSGG (Sialosyl-MSGG)

MSGG interacts with an affinity of -21.1 kcal/mol

Polymorphic site Val143 currently without contact, deletion found since 3/20, Val143Phe since 10/20 and infrequent insertion of 2-15 amino acids

Is found in few His69Tyr sub-strains since 2/20, deletion found sporadically since 2/20

Combination of His69del with the Val70del since 2/20

Shift of proximal 5 carbohydrates by 1-1.5 Å and rotation of glycosidic linkage of terminal 2 sugar residues in refinement

Figure 6

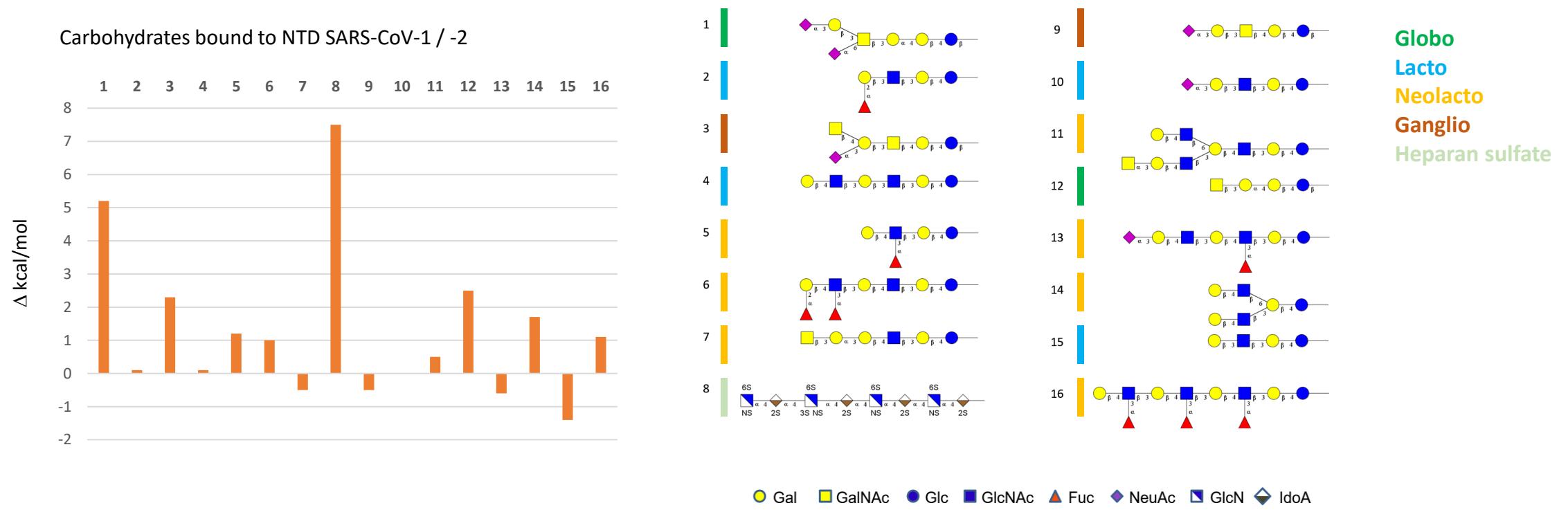


Table 1

S-Protein

NTD amino acid 14-291 RBD amino acid 334-524

NTD A	NTD B	NTD C	RBD A	RBD B	RBD C
HIS 69	HIS 69	ILE 68	SER 375	GLU 340	THR 345
GLY 72	SER 71	HIS 69	THR 376	VAL 341	ARG 346
THR 73	GLY 72	VAL 70	LYS 378	PHE 342	PHE 347
LYS 77	LYS 77	SER 71	TYR 380	ALA 344	ALA 348
SER 98	LYS 97	GLY 72	GLY 404	ARG 346	SER 349
TYR 145	SER 98	LYS 77	ASP 405	PHE 347	ALA 352
LYS 147	ILE 100	LYS 97	VAL 407	ALA 348	TRP 353
LYS 150	TYR 144	TYR 145	ARG 408	SER 349	ASN 354
TRP 152	TYR 145	LYS 147	GLN 409	TYR 351	ARG 355
GLY 181	LYS 147	TRP 152	ALA 411	ALA 352	LYS 356
LYS 182	LYS 150	GLY 181	GLY 413	TRP 353	ARG 357
GLN 183	TRP 152	LYS 182	GLN 414	ASN 354	TYR 451
GLY 184	GLU 180	GLN 183	THR 415	ARG 355	ARG 466
HIS 245	GLY 181	GLY 184	GLY 416	LYS 356	ILE 468
ARG 246	LYS 182	ASN 185	ALA 435	ARG 357	
SER 247	GLN 183	HIS 245	TRP 436	SER 399	
LEU 249	HIS 245	ARG 246	ASN 437	ARG 466	
THR 250	ARG 246	SER 247	VAL 503	ASP 467	
TRP 258	SER 247	LEU 249	GLY 504	ILE 468	
THR 259	TYR 248	THR 250	TYR 508	SER 469	
ALA 260	LEU 249	SER 256			
GLY 261	THR 250	GLY 257			
ALA 262	PRO 251	TRP 258			
	GLY 252	THR 259			
	ASP 253	ALA 262			
	SER 256				
	GLY 257				
	TRP 258				
	THR 259				
	ALA 260				
	GLY 261				
	ALA 262				

Table 2

Blood Group

			Sphingolipid Category
1	I A	A-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
2	I B	A-D-Galp-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
3	I H	A-L-Fucp-(1-2)-B-D-Galp-(1-3)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
4	II A	B-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
5	II A	A-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
6	II A	A-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
7	II B	A-D-Galp-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
8	II B	A-D-Galp-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
9	II H	A-L-Fucp-(1-2)-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
10	III A	A-D-GalpNAc-(1-3)[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-A-D-GalpNAc-(1-3)[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-OH	
11	III B	A-D-Galp-(1-3)[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-A-D-GalpNAc-(1-3)[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-OH	
12	III H	A-L-Fucp-(1-2)-B-D-Galp-(1-3)-A-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-4)-B-D-GlcpNAc-(1-3)-B-D-Galp-(1-4)-B-D-Glcp-OH	
13	IV A	A-D-GalpNAc-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-B-D-GalpNAc-(1-3)-A-D-Galp-(1-4)-B-D-Galp-(1-4)-B-D-Glcp-OH	
14	IV B	A-D-Galp-(1-3)-[A-L-Fucp-(1-2)]-B-D-Galp-(1-3)-B-D-GalpNAc-(1-3)-A-D-Galp-(1-4)-B-D-Galp-(1-4)-B-D-Glcp-OH	
15	IV H	A-L-Fucp-(1-2)-B-D-Galp-(1-3)-B-D-GalpNAc-(1-3)-A-D-Galp-(1-4)-B-D-Galp-(1-4)-B-D-Glcp-OH	

Other blood groups or tissular determinants have not been tested if not otherwise indicated, units of glycans were limited to 8 (10, 11 do not correspond to complete glycolipids). O-glycans may present the Type III B determinant exclusively / the entry is currently not listed in LipidMaps.

Table 5

Carbohydrate Screen ⁽¹¹⁵⁾

2'-Fucosyllactose	GG1ba	Type II B antigen
3-Fucosyllactose	GG1c	Type II H antigen
3KDNLN	GT1a	Type III A antigen
3'-Sialyl-3-fucosyllactose	GT1aa	Type III H antigen
3'-Sialylactose	GT1b	Type IV A antigen
6'-Sialylactose	GT1b Ac	Type IV B antigen
Asialo GM1	GT1ba	Type IV H
Asialo GM2	GT1c	VI 3GalNAca-IV 6klndoLcOse8
BdGalNac-neolacto	GT3	VI 3(Galb 1-4GlcNAcb)-Lc4
cis GM1, GM1b	Heparin	VIM-II
Difucosyllactose	Isoglobotriglycosyl	X3 ganglioside
Disialylacto-N-tetraose	IV 3-nLcOse4	X3 ganglioside
DSSG (Sialosyl-MSGG)	KDN	X4 ganglioside
*Forssman antigen	KDN-GD1a	X-haptens, SSEA-1, Lex-5
*Forssman Branched	KDN-GD1a	
Forssman-like iGb4	KDN-GM1	
Fuc-Gal-GD1b	KDN-GM2	
Fuc-GM1	KDN-lactotetraosylceramide	
GalNAc-GD1a	KDN LewisC	
*GalNAc-GD1a(Neu5Ac/Neu5Gc)	KDN-neolacto (short)	
*GalNAc-GD1a(Neu5Gc/Neu5Ac)	Lacto-N-fucopentaose I	
GalNAc-GM1b	Lacto-N-fucopentaose V	
Gb3	Lacto-N-neohexaose	
Gb4	Lacto-N-neotetraose	
Gb5	Lacto-N-tetraose	
GD1a	Lactosialyltetraose	
*GD1a (NeuAc/NeuGc)	LeC	
*GD1a (NeuGc/NeuGc)	Lex-7	
GD1a, GD1e	Lex-9	
GD1aa	Ley-6	
GD1b	Ley-8	
GD1c	Ley-A-9	
GD2	LM1, iso-LM1	
GD2	Man	
GD3	*Neu5Gc aD OH1	
GD3	*Neu5Gc aD OMe	
GD3	Para-Forssman x3b	
GD3 9OAc	Paragloboside, nLc4Cer	
Glc	Para-Lacto-N-neohexaose	
Globo H	Polymeric Lex	
Globo-Lex-9	Polymeric Lex	
Gb4 (P antigen)	Polymeric Lex	
GM1	Polymeric Lex	
GM2	Type I A antigen	
GM2	Type I B antigen	
GM3	Type I H antigen	
GM4	Type II A antigen	
GP1c	Type II A antigen	
GG1aa	Type II A antigen	
GG1b	Type II A antigen	
GG1b 6	Type II B antigen	

*not in humans or rarely expressed
in genetic variants

Table 4

GSL-Headgroup	SWModel 7a25	5x4S	IUPAC Name	Category	No.		
DSGG #	-7.7	-25.4	-6.2	-20.2	A-D-Neu5Ac-(2-3)-B-D-Galp-(1-3)-[A-D-Neu5Ac-(2-6)]-B-D-GalpNAc-(1-3)-A-D-Galp-(1-4)-B-D-Galp-(1-4)-B-D-GlcP-OH / 113	Globo	1
I H antigen	-7.6	-15.5	-7	-15.4	A-L-Fucp-(1-2)-B-D-Galp-(1-3)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Lacto	2
GalNAc-GM1b	-7.6	-21.3	-6.8	-19	B-D-GalpNAc-(1-4)-[A-D-Neu5Ac-(2-3)]-B-D-Galp-(1-3)-B-D-GalpNAc-(1-4)-B-D-Galp-(1-4)-B-D-GlcP-OH	Ganglio	3
VI3(Galb 1-4GlcNAcb)-Lc4	-7.4	-18.9	-7.4	-18.8	B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-3)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Lacto	4
X-hapten, SSEA-1, Le ^x -5 *	-7.3	-16.1	-6.7	-14.9	B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	5
Le ^y -8	-7.3	-21.4	-7	-20.4	A-L-Fucp-(1-2)-B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	6
IV 3-nLcOse4	-7.3	-18.9	-7.5	-19.4	B-D-GalpNAc-(1-3)-A-D-Galp-(1-3)-B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	7
Heparin (3ina †)	-7.3	-29.7	-5.5	-22.2	a-D-GlcPNSO36SO3-(1-4)-a-L-IdoP2SO3-(1-4)-a-D-GlcPNSO33SO36SO3-(1-4)-a-L-IdoP2SO3-(1-4)-a-D-GlcPNSO36SO3-(1-4)-a-L-IdoP2SO3-(1-4)-a-D-GlcPNSO36SO3-(1-4)-a-L-IdoP2SO3	Heparan sulfate	8
GM1b	-7.2	-18.2	-7.4	-18.7	A-D-Neu5Ac-(2-3)-B-D-Galp-(1-3)-B-D-GalpNAc-(1-4)-B-D-Galp-(1-4)-B-D-GlcP-OH	Ganglio	9
Lactosialyltetraose	-7.0	-17.2	-6.8	-17.2	A-D-Neu5Ac-(2-3)-B-D-Galp-(1-3)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Lacto	10
VI 3GalNAca-IV 6kladoLcOse8	-7.0	-24.1	-6.9	-23.6	A-D-GalpNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-[B-D-Galp-(1-4)-B-D-GlcPNAc-(1-6)]-B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	11
Gb4 (P antigen)	-7.0	-14.3	-5.8	-11.8	B-D-GalpNAc-(1-3)-A-D-Galp-(1-4)-B-D-Galp-(1-4)-B-D-GlcP-OH	Globo	12
VIM-II	-7.0	-21.4	-6.8	-22	A-D-Neu5Ac-(2-3)-B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	13
Lacto-N-neohexaose	-7.0	-18.1	-6.4	-16.4	B-D-Galp-(1-4)-B-D-GlcPNAc-(1-6)-[B-D-Galp-(1-4)-B-D-GlcPNAc-(1-3)]-B-D-Galp-(1-4)-B-D-Glc-OH	Neolacto	14
LM1, iso-LM1	-7.0	-14.2	-7.7	-15.6	B-D-Galp-(1-3)-B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Lacto	15
Polymeric Le ^x ‖	-7.0	-25.6	-6.6	-24.5	B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-[A-L-Fucp-(1-3)]B-D-GlcPNAc-(1-3)-B-D-Galp-(1-4)-B-D-GlcP-OH	Neolacto	16

Affinity in kcal/mol

Found as Di-Sialosyl Galactosyl Globoside or Sialosyl-Monosialosyl Galactosyl Globoside (Sialosyl-MSGG) and classifiable as Di-Sialosyl-Gb5 or Di-sialosyl-Galb1-3-Gb4

* Type II isomer of Le^a

† charge correction pending

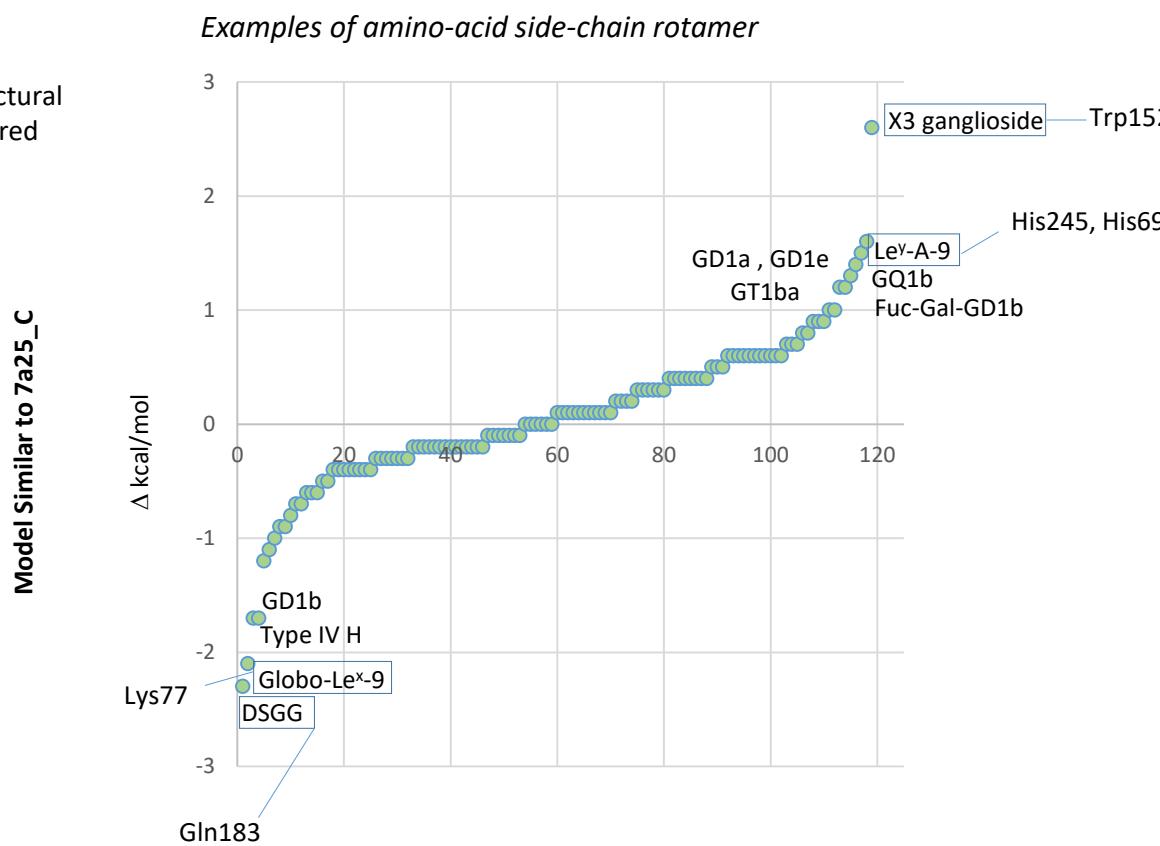
|| Also named „Trimeric Le^x“

refined/redocked (exhaustiveness 3)

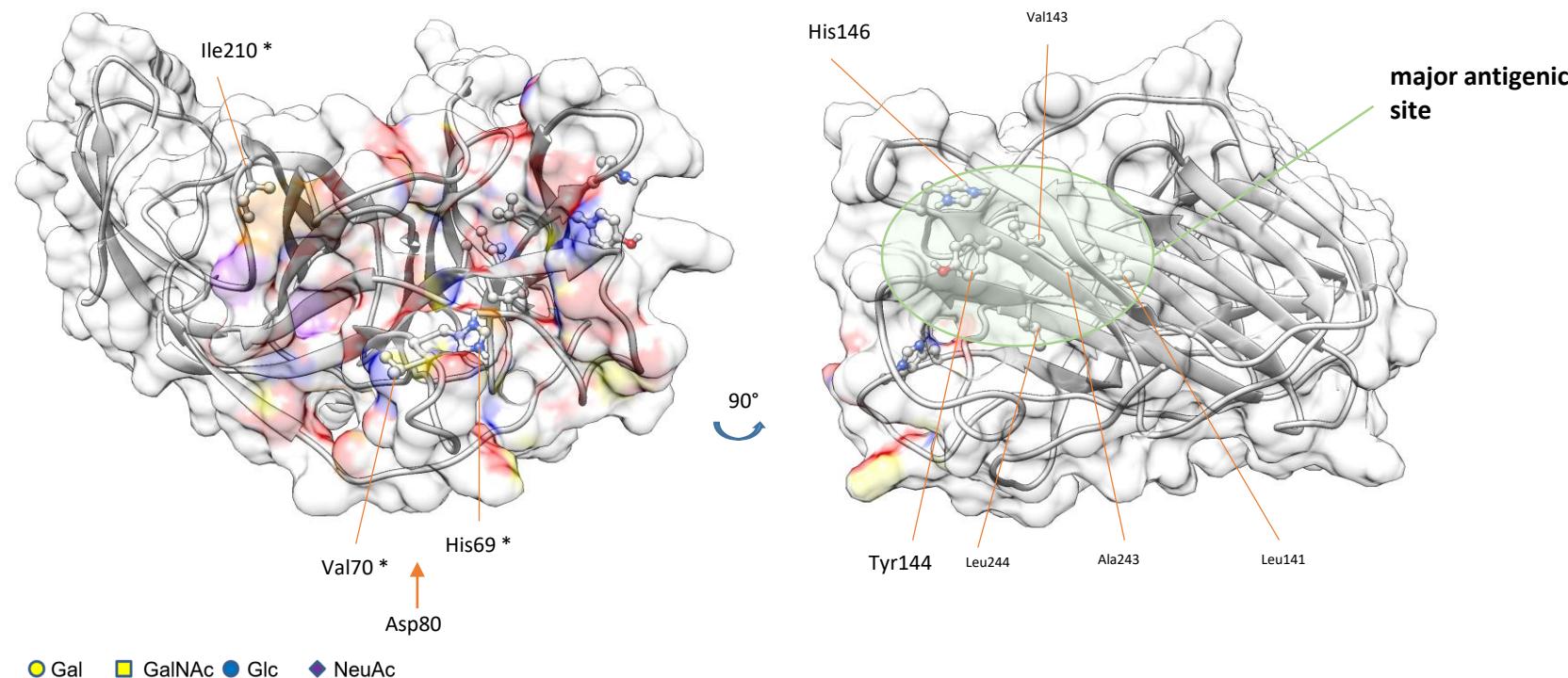
refined/redocked (exhaustiveness 4)

Supplementary Figure 1

The Δ values of two structural conformers were compared

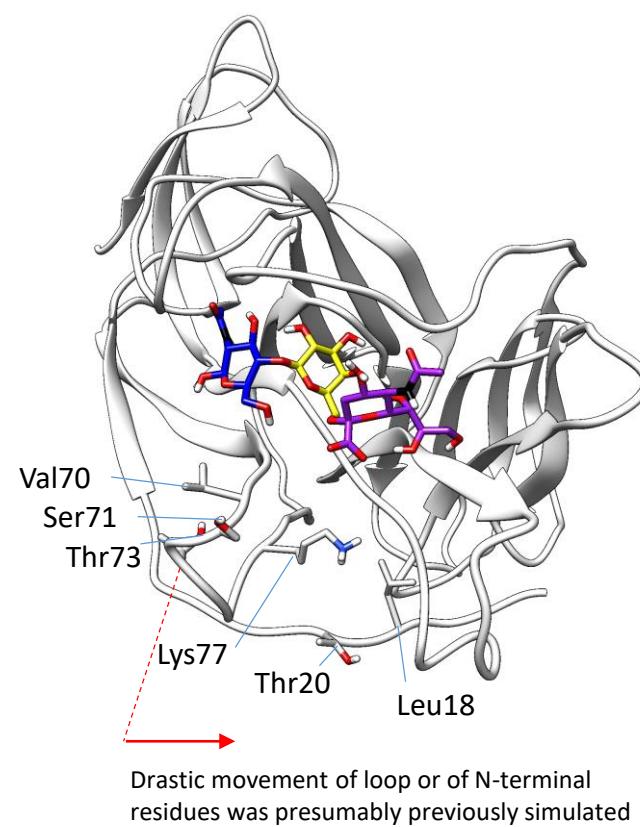


Supplementary Figure 2



* No alteration in antigenicity in McCarthy et al. (2021)
Small font: Non-exposed residues with likely indirect structural effects

Supplementary Figure 5



● Gal ● GlcNAc ● NeuAc

Example of docked glycan:

A-D-Neu5Ac-(2-6)-B-D-Galp-(1-4)-B-D-GlcNAc-OH