FULL PAPER

This is the peer-reviewed version of the following article: Structural characterization of the N-linked glycans in the receptor binding domain of the SARS-CoV-2 spike protein and their interactions with human lectins using NMR spectroscopy, which has been published in final form at doi: 10.1002/anie.202011015

FULL PAPER

Structural characterization of the N-linked glycans in the receptor binding domain of the SARS-CoV-2 spike protein and their interactions with human lectins using NMR spectroscopy

Maria Pia Lenza, [a] Iker Oyenarte,[a] Tammo Diercks,[a] Jon Imanol Quintana,[a], Ana Gimeno,[a] Helena Coelho,[d] Ana Diniz,[d] Francesca Peccati,[a] Sandra Delgado,[a] Alexandre Bosch,[a] Mikel Valle,[a] Oscar Millet,[a] Nicola G. A. Abrescia,[a],[b],[c] Asís Palazón,[a],[b] Filipa Marcelo,[d] Gonzalo Jiménez Osés,[a] Jesús Jiménez-Barbero,*[a][b][e] Ana Ardá,*[a] and June Ereño-Orbea*[a][b]

[a]   M. P. Lenza, I. Oyenarte, Dr. T. Diercks, J. I. Quintana, Dr. A. Gimeno, Dr. F. Peccati, S. Delgado, A. Bosch, Dr. M. Valle, Dr. O. Millet, Prof. Dr. N. G. A. Abrescia, Dr. A. Palazón, Dr. G. Jimenez-Osés, Prof. Dr. J. Jiménez-Barbero, Dr. A. Ardá, Dr. J. Ereño-Orbea
        CIC bioGUNE, Basque Research and Technology Alliance, BRTA, Bizkaia Technology Park, 48162 Derio, Spain
        jjbarbero@cicbiogune.es, aarda@cicbiogune.es, jereno@cicbiogune.es
[b]   Prof. Dr. N. G. A. Abrescia, Dr. Asís Palazón, Prof. Dr. J. Jiménez-Barbero, Dr. June Ereño-Orbea
        Ikerbasque, Basque Foundation for Science, 48013 Bilbao, Spain
[c]   Prof. Dr. N. G. A. Abrescia Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBERehd), Instituto de Salud Carlos III, Madrid, (Spain)
[d]   Dr. F. Marcelo, Dr. H. Coelho, A. Diniz
        UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia,
        Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal
[e]   Prof. Dr. J. Jiménez-Barbero
        Department of Organic Chemistry II, University of the Basque Country
        UPV/EHU, 48940 Leioa, Spain

Dedicated to Professor Christian Griesinger on the occasion of his 60th birthday

Supporting information for this article is given via a link at the end of the document.

Abstract: The glycan structures of the receptor binding domain of the SARS-CoV2 spike glycoprotein expressed in human HEK293F cells have been studied by using NMR. The different possible interacting epitopes have been deeply analysed and characterized, providing evidence of the presence of glycan structures not found in previous MS-based analyses. The interaction of the RBD ^{13}C-labelled glycans with different human lectins, which are expressed in different organs and tissues that may be affected during the infection process, has also been evaluated by NMR. In particular, ^{15}N-labelled galectins (galectins-3, -7 and -8 N-terminal), Siglecs (siglec-8, siglec-10), and C-type lectins (DC-SIGN, MGL) have been employed. Complementary experiments from the glycoprotein perspective or from the lectin's point of view have permitted to disentangle the specific interacting epitopes in each case. Based on these findings, 3D models of the interacting complexes have been proposed.

Introduction

The current COVID-19 pandemic caused by SARS-CoV-2 coronavirus represents an enormous health and social problem.[1,2] The virus employs a glycosylated spike protein (S) to bind the angiotensin-converting enzyme 2 (ACE2) of the host.[3,4] In many viral infections (influenza, Ebola, SARS-CoV, among others), glycan-mediated interactions are essential for the initial contact between the virus and the host.[5–7] In fact, glycans modulate molecular recognition events not only in host–pathogen recognition or infections, but also in tissue differentiation, cell signalling, immune response, and cancer, besides contributing to proper protein folding.[8] In SARS-CoV-2, a receptor binding domain (RBD) has been identified that efficiently binds ACE2. Both ACE2 and the RBD are glycosylated, although the RBD glycans do not seem to be directly involved in the interaction, according to the structural data available so far.[3,4,9] Additionally, our immune system contains a variety of glycan-binding proteins (lectins) that are able to specifically detect and bind diverse glycan-epitopes, triggering innate responses in a glycan-dependent manner.[10,11] In the SARS-Cov-2 context, a recent study[12] has suggested the existence of lectin-mediated molecular pathways that may contribute to viral infection and immune exacerbation, identifying some lectins that bind to the RBD. From the molecular recognition perspective, unravelling these viral glycan-host lectin interactions at high resolution represents a tremendous scientific challenge. Since N-glycosylation is not template-driven, the hallmark is chemical heterogeneity. Glycoprofile analysis remains technically difficult given the huge range of possible monosaccharide combinations and the different ways they can link to each other. Advances in mass spectrometry (MS) allows achieving a global perspective of the glycoprofile of the target protein. In fact, the glycosylation profile of the spike glycoprotein S has been recently described.[13–16] However, given the need of digestion protocols for MS-based methods, molecular recognition studies should be carried out with procedures that only minimally alter the test samples. In this context, we have herein applied an NMR protocol[17] to characterize the precise glycan structures of the two N-linked uniformly ^{13}C-labeled glycans (at N331 and N343) in the domain B of subunit S1 (S[B]) from the RBD (hereinafter referred to as RBD)

of SARS-CoV-2,[4,18] produced in human HEK293F cells.[19] Additionally, we have dissected the glycan-mediated interactions of RBD with a variety of human lectins, which are expressed in different organs and tissues that may be affected during the infection (Figure 1). For this task, two complementary protocols have been employed. On the one hand, signal changes in the 2D $^1H,^{13}C$-HSQC NMR spectra of the ^{13}C-labeled glycans on the RBD have been monitored upon addition of the lectins. Alternatively, signal perturbations in the $^1H,^{15}N$-TROSY/HSQC spectra of the ^{15}N-labelled lectins in the presence of the RBD have been assessed to provide a complementary view on their specific interactions. Our study provides key structural details on the N-glycan content on the RBD of SARS-CoV-2 spike protein, especially respect to the exposed glyco-epitopes at the terminal chains, prone to participate in lectin recognition. We have identified the specific glycans in the RBD that are recognized by the corresponding lectin.
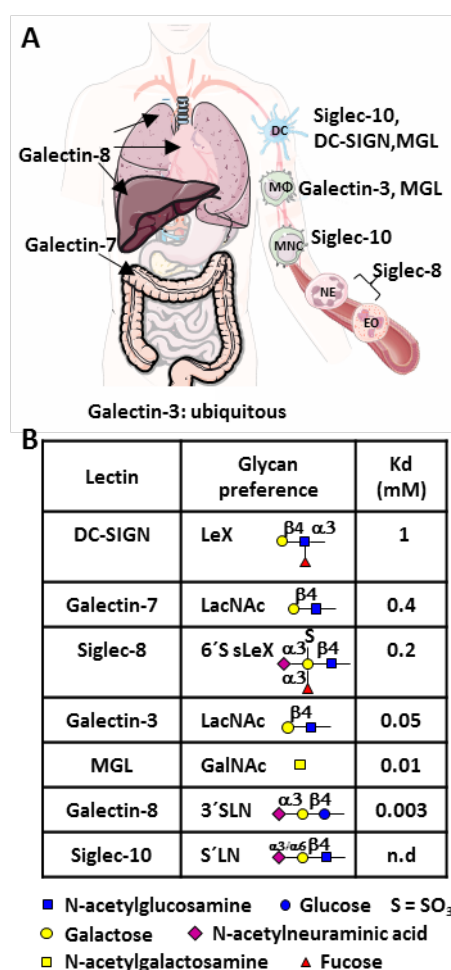
Figure 1. A. Panel of human lectins employed herein along with their locations in human organs and tissues. B. Major glycan specificities and binding affinities for DC-SIGN,[20] galectin-7,[21] Siglec-8,[22] galectin-3,[23] MGL,[24] galectin-8[21] and siglec-10[25] are given. Glycans are represented in SNFG symbols.[26]

Results and Discussion

Disentangling the glycoprofile of the RBD produced in HEK293F cells.

A NMR-based approach[17,27–31] was employed to perform the glycoprofile analysis of the N-linked glycans at N331 and N343 at the S^B domain of RBD (residues 328-533). Key regions of the spectra are shown in Figure 2 and Figure S1. In particular, the combined analysis of 3D H',CH NOESY-HSQC, H',CH TOCSY-HSQC, and H'[C'],CH and [H']C',CH edited HSQC-[$^{13}C,^{13}C$]TOCSY-HSQC (Figure 2) allowed to determine the precise structure of the glycans and their glycosidic linkages (Table S1). The [H']C',CH edited HSQC-TOCSY-HSQC was instrumental to assign all carbon resonances for every spin system (Figure 2), identifying glycosylated or otherwise chemically modified positions The presence of terminal N-acetyllactosamine (LacNAc) units was assessed by the analysis, and appeared also decorated with α2,3- (3'SLacNAc or 3´SLN) and α2,6-linked sialyl (6'SLacNAc or 6´SLN) moieties. The presence of GalNAc-containing epitopes, β1-4 linked to GlcNAc, was also evident: terminal GalNAcβ1-4GlcNAc (LacdiNAc or LDN) was found, along with their α2,6-sialylated and 4-O-sulfated derivatives (6´SLDN and 4SulLDN) that had not been identified in previous MS analyses of the S protein.[13,14,16] 4SulLDN was identified due to the exclusive $^1H/^{13}C$ chemical shifts (Table S1) of position 4 of GalNAc.[32] Another relevant observation was the presence of a high degree of fucosylation, both at the core and at terminal positions, corresponding to LewisX (LeX) and fucosylated LDN (LDNF). The presence of this last epitope was somehow unexpected as it has usually been related to parasites, and is thought to cause immunogenic response in humans.[33–35] With respect to the N-glycan architecture, although a quantitative analysis is out of the scope of this study, biantennary complex N-glycans are the prevalent scaffolds. Signals corresponding to high-mannose-type N-glycans display almost undetectable intensity, while the presence of core bisecting GlcNAc was discarded due to the absence of its characteristic signals (downfield shift of the H4-C4 correlation for βMan).[36] On the other hand, a minor degree of additional branching to give tri- and tetra-antennae was also verified. The branching at the α3-antenna occurs through further β1-4-linked GlcNAc glycosylation of αMan3, whose anomeric carbon becomes upfield shifted (Figure 2).[36] The branching at the α6 antenna is produced by a GlcNAc β1-6 linkage to αMan6, whose anomeric correlation can be now distinguished.[36] Additionally, the low intensity of the signals corresponding to the anomeric positions of βGal moieties in type I poly-LacNAc structures, strongly suggested the absence of these elongations.

Protein N-glycosylation is a highly complex and tightly regulated event, hitherto not fully understood.[37,38] The use of glycoproteins as therapeutics has fostered the development of novel methods to control glycosylation, with special focus on producer-cell lines, which strongly influence the N-glycosylation outcome.[39] Different MS-based studies on the glycosylation pattern of the spike protein in HEK293/F cells have been recently presented.[13–16] However, despite employing the same expression system, these studies yielded slightly different results, which could be partially explained by the use of diverse S proteins (either the trimeric form or the separate S1 and S2 subunits). Indeed, the specific protein structure has been proposed as one of the multiple factors influencing protein glycosylation.[40,41] Our results for the RBD of the S protein, which contains two glycosylation sites at N331 and N343, show very important levels of fucosylation (Fuc) and N-acetyl galactosylation (GalNAc), also reported in some of the MS-based studies.[13–16] However, our NMR methodology allows

defining the precise chemical nature and structural details (glycosidic linkages, sulfation) of the epitopes in which these residues (Fuc, GalNAc) are found. Indeed, glycan motifs not described earlier, as 4SulLDN, 6´SLDN, LeX, and LDNF were evident by NMR.
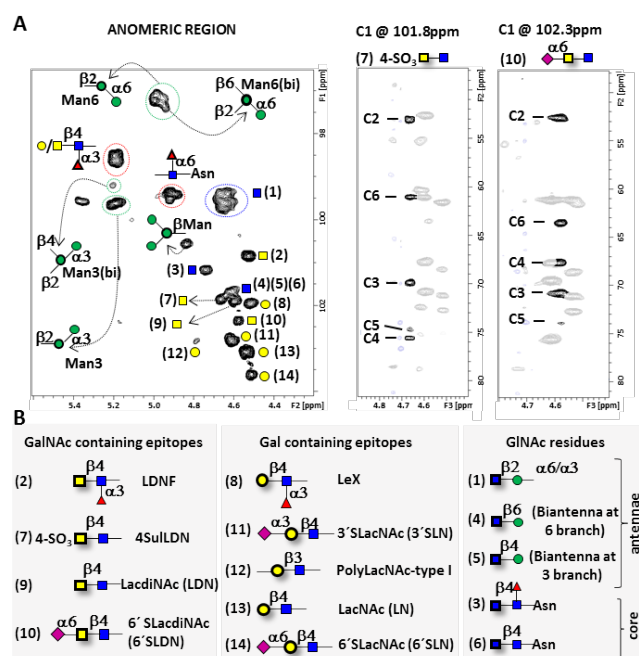
Figure 2. NMR identification of glycan structures on SARS-CoV-2 RBD glycoprotein. A. Left: detail of the $^1H,^{13}C$-HSQC of RBD showing the assignment for most anomeric correlations, represented as SNFG symbols.[26] Anomeric correlations for Gal, GalNAc and GlcNAc are identified with a number in brackets. Right: 3D [H']C',CH edited HSQC-TOCSY-HSQC of RBD, selected planes for C1 GalNAc on the 4SulLDN fragment and for C1 GalNAc on 6´SLDN, showing the correlations to all ^{13}C atoms within the pyranose spin system. Nearby cross peaks belonging to other spin systems have been veiled for clarity. B. GalNAc, Gal and GlcNAc containing epitopes in N-linked glycans on RBD, represented as SNFG symbols.

These moieties, along with terminal LacNAc, LDN, 3´SLN and 6´SLN fragments are predominant epitopes on the outer chains of the RBD N-glycans. These unexpected findings prompted us to produce a different ^{13}C-labelled glycoprotein, the α subunit of the human high-affinity Fc receptor for IgE (FcεRIα),[17,42] using exactly the same conditions used for the RBD. Interestingly, the superimposition of the $^1H,^{13}C$-HSQC spectra of the glycans of RBD and FcεRIα evidenced the lack of all GalNAc-containing cross peaks as well as fucosylated LDNF and LeX signals in FcεRIα, as previously observed by employing other conditions (Figure S2). Among other factors, these results suggest that the precise protein structure could influence the glycosylation pattern.[40,41,43–45] Nevertheless, the NMR methodology described herein allows detecting key features of the epitopes, as sulfation, rather difficult to detect by the potent MS approach.[46–48]

Molecular recognition studies: The interaction with lectins.
It has been recently proposed that the spike glycoprotein is specifically recognized by C-type lectins (Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin (DC-SIGN) and Macrophage Galactose-type lectin (MGL)) and by Siglecs expressed in the lung microbiota.[12] Thus, we focused on deducing the specific epitopes responsible for these interactions. Moreover, we also studied the interaction with several human galectins (galectins-3, 7 and 8), which are involved in inflammation.[49,50] It has been proposed that galectin inhibitors may modulate the cytokine storm associated to COVID-19 as well as interfering with viral attack.[51] Thus, NMR experiments both from the glycan and from the lectin perspectives were carried out to monitor the lectin/RBD interactions. From the glycan point of view, the cross peak intensities of the ^{13}C-labelled glycans described above were compared to those recorded in the presence of the lectins (Figures 3, 4, 6, and 7). Alternatively, $^1H,^{15}N$-TROSY/HSQC experiments allowed analysing the line width perturbations of the amide NMR signals of the ^{15}N-labelled lectins upon RBD addition (Figure 5 and Figures S4-7). Although signal broadening depends on many factors and cannot be directly related with binding affinity, it allowed us to discern the glycan specificities between the lectins studied.

The interaction with galectins. Upon addition of galectin-3, the $^1H,^{13}C$-HSQC spectrum of the RBD glycans showed significant, but selective, reductions in the intensity of diverse cross peaks (Figures 3 and 4). Intensity attenuation was more pronounced for peaks from LDN, LacNAc and 3'SLacNAc. The observed dramatic signal broadening evidence the presence of dynamic processes in the intermediate exchange regime in the NMR chemical shift time scale and thus, the presence of a significant interaction.

Additional information was obtained by observing the changes in the lectin $^1H,^{15}N$ cross peaks (Figure 5). The results strongly suggest that the interaction with the RBD affects the canonical LacNAc binding site of galectin-3.[23] In fact, the $^1H,^{15}N$-TROSY spectrum of the galectin-3/RBD sample (Figure 5) exhibited the complete disappearance of specific cross peaks, such as that for H158, conserved among galectins and participating in hydrogen bonding with 4OH-Gal, or R186, involved in hydrogen bonding with the GlcNAc moiety in the complexes with LacNAc. W181, conserved and key for the CH-π stacking interaction with Gal, was also disturbed. Additionally, the whole S5 or S6 β-strand (V172-L177) along the loop 177-180 were highly affected. In particular, the T175-N179 region is hardly modified by LacNAc, but greatly perturbed when LacNAc is α2-3 sialylated (Figure S3). Thus, these data confirm that galectin-3 binds the RBD through the canonical Gal binding site by specifically recognising terminal LN, LDN and 3'SLN epitopes on the RBD.[23]

We next studied the interaction of the RBD with the N-terminal domain of galectin-8 (galectin-8N), a tandem repeat lectin whose N-terminal domain has partially overlapping glycan binding preferences with galectin-3,[52,53] although with diverse affinities for the same epitopes.[54,55] Strikingly, when the RBD and galectin-8N were mixed in a 1:1 ratio, the same conditions for galectin-3, the sample became cloudy and unmanageable for NMR experiments. Thus, a 1:0.2 ratio (^{13}C-RBD:galectin) was used, resulting in a clear sample that allowed recording the $^1H,^{13}C$-HSQC (Figure 4). The observed cross peak signal reduction was now more selective than for galectin-3, and showed that galectin-8N binds mainly to the 3'SLN RBD glycan epitope. The interaction

from the lectin perspective (Figure S4) showed that the most affected residues on galectin-8N were around the canonical glycan binding site. These results permit not only underline the different glycan binding preferences between both galectins towards the RBD in terms of epitopes, but also the different recognition phenomena that take place when the binding epitopes are differently exposed or hidden, especially in multivalent presentations, as also highlighted by others.[52]
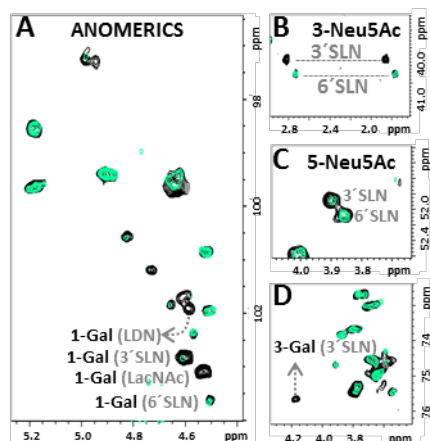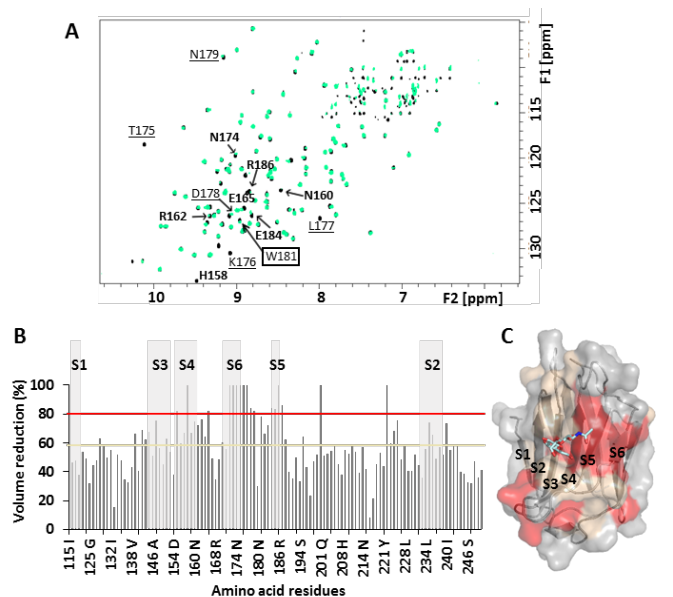

Figure 3.

NMR identification of glycan epitopes on the RBD recognized by galectin-3. Different regions of the ^{1}H,^{13}C-HSQC spectrum of RBD alone (in black) and in the presence of 1 equivalent of galectin-3 (in green). A. Anomeric region: signals for terminal epitopes mostly affected by the presence of galectin-3 are annotated. B and C. Regions of the H3-C3 and H5-C5 (respectively) correlations of terminal Neu5Ac residues: signals for α2-3 linked residues completely disappear, while those for the α2-6 linked are barely affected. D. Region showing the signal of H3-C3 Gal in 3´SLN epitope. The graphical bar representation for the % of volume reduction of selected cross peaks on ^{1}H,^{13}C-HSQC of RBD upon addition of galectin-3 is given in Figure 4 upper panel.

Finally, the prototype galectin-7 was tested. Galectin-7 contains two identical glycan binding sites, forms non-covalent homodimers, and displays the lowest glycan affinities reported among galectins.[52,56] The NMR analysis of the RBD glycans upon addition of 1 equivalent of galectin-7 revealed that the most perturbed signals correspond to the LDN epitope, with those for LN and 3´SLN also affected to a lesser extent (Figure 4). These results agree with reported data that showed that acetylation of terminal Gal moieties increased the affinity for Gal-7,[21] although contrast respect to 3´SLN which was reported to bind weaker than LN.[21,57] From the lectin perspective, the cross peak intensity loss upon RBD addition affected not only amino acids at the lactose binding pocket (W70, H50, V60, V61, N63, N52, E74), but also far from this site (T57, S58, Q67, loops), at the F face (G76, R113, Y106), and even at the dimer interface (V89, L90), reflecting that most of the protein is actually affected by the interaction (Figure S5). This evidence agrees with previous studies that claimed an inter-domain communication upon lactose binding.[58]
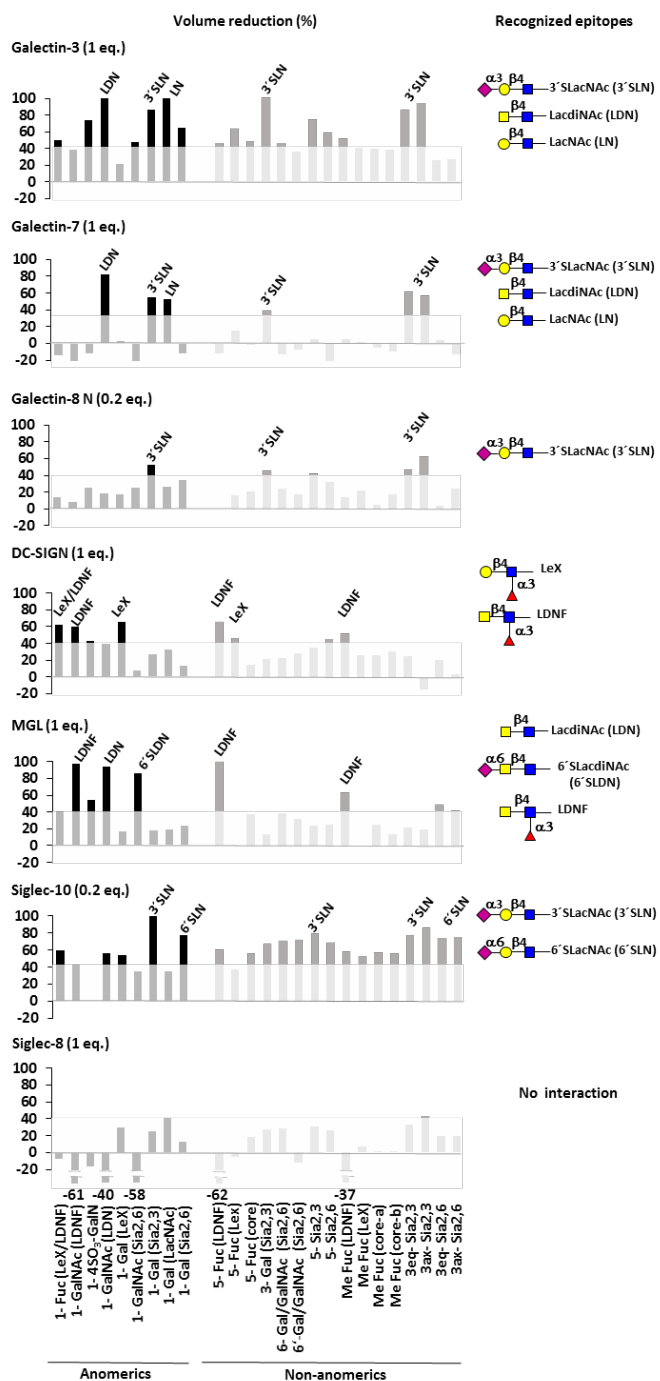
Figure 4. Graphical bar representation for the % of cross peak volume reduction of cross peaks on ^{1}H,^{13}C-HSQC of RBD upon adding the different lectins. From top to bottom: galectin-3 (1 eq), galectin-7 (1 eq), galectin-8N (0.2 eq), DC-SIGN (1 eq), MGL (1eq), Siglec-10 (0.2 eq), Siglec-8 (1 eq). Each cross peak is identified with a number corresponding to the position on the pyranose, the residue, and the epitope. An arbitrary threshold (transparent square) is used to highlight the most affected signals. SNFG representation of the major deduced epitopes on the RBD are on the right.
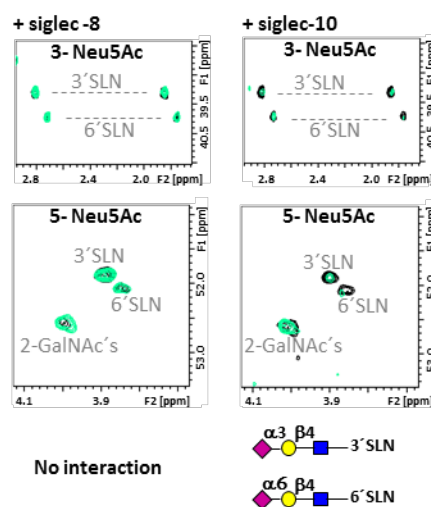
Figure 5.

Galectin-3/RBD interaction deduced by NMR from the lectin perspective. A. Superimposition of the ^{1}H,^{15}N-TROSY for free galectin-3 (black) and galectin-3/RBD (green). Some affected cross peaks are annotated: amino acids involved in LacNAc interactions are bolded; those interacting with α2-3 sialic acid are underlined. The key signal of W181 is squared. B. % cross peak volume reduction on the ^{1}H,^{15}N-TROSY upon addition of galectin-3. S1-S6 β-strands are depicted with grey boxes. Red and wheat horizontal lines are baselines for cross peaks suffering 80-100% reduction and 60-80% reduction, respectively. C. Cartoon and surface representation of galectin-3 bound to LacNAc (PDB 1A3K) according to the X-Ray structure. Amino acids are coloured based on their perturbation (%volume reduction) due to RBD binding (thresholds in B).

The interaction with Siglecs. The combination of the RBD with Siglec-8 produced no changes neither on the ^{1}H,^{13}C-HSQC of the ^{13}C-glycans at the RBD (Figure 6 and Figure S6), nor in the ^{1}H,^{15}N-TROSY of ^{15}N-Siglec-8 (Figure S6). Thus, Siglec-8 does not recognize any glycan on the RBD, in agreement with the tight glycan binding selectivity of this lectin that binds terminal 3´SLN and SLeX, only when they are sulfated at Gal 6.[22,59] This chemical modification is not present in our glycosylated RBD since it would be readily identified due to the characteristic chemical shifts of a sulfated C6-Gal. A completely different situation was found for Siglec-10, for which the addition of 0.2 equivalents of lectin to the RBD caused a general reduction of the cross-peak intensities of the ^{13}C-labelled glycans. The effect was more pronounced for the signals of terminal 3´SLN and 6´SLN epitopes (Figure 6), in agreement with the reported selectivity for this lectin.[60] Although a preference for 6'SLN over 3'SLN has been described, this is not appreciable from the NMR data. Information from the lectin perspective was not possible in this case due to the lack of access to a suitable ^{15}N-labelled lectin for NMR.

The interaction with C-type lectins DC-SIGN and MGL. The presence of 1 equivalent of DC-SIGN caused a selective intensity decrease on specific glycan ^{1}H,^{13}C-HSQC cross peaks of the RBD (Figure 7). In terms of terminal epitopes, the LeX and LDNF signals were the most affected (Figures 4 and 7).


Figure 6. The interaction of RBD with Siglecs -8 and -10 from the glycan perspective. Different regions of the ^{1}H,^{13}C-HSQC of RBD alone (in black) and with 1 eq. of Siglec-8 (left, superimposed in green), and with 0.2 equivalents of Siglec-10 (right, superimposed in green). Top and middle: regions for the C3-H3 and C5-H5 correlations of Neu5Ac. In the presence of Siglec-8, no signal is affected, indicating that there is no interaction, while in the presence of Siglec-10, the signals of Neu5Ac, both α2-3 and α2-6 linked are affected, indicating that Siglec-10 interacts with the RBD through these epitopes. The graphical bar representation for % of volume reduction of cross peaks on the ^{1}H,^{13}C-HSQC of RBD upon adding Siglec-8 and -10 are in Figure 4.

This fact is in agreement with the reported preference of DC-SIGN for these moieties, in which key interactions are provided by the Fuc residue that binds at the calcium binding site of the lectin.[20,61] The comparison with the effects produced by 1 equivalent of MGL were markedly different (Figures 4 and 7), permitting to identify the diverse binding preferences of both lectins. Indeed, the presence of MGL produced an exquisite selective reduction of the signals corresponding to GalNAc-containing epitopes, with the exception of 4SulLDN (Figures 4 and 7). Thus, terminal LDN, its α2-6 sialylated version (6´SLDN), and the fucosylated LDN (LDNF) are the glycans specifically recognized by MGL.

With respect to the lectin binding site, the cross peaks on the ^{1}H,^{15}N-TROSY of DC-SIGN (Figure S7) exhibited differential intensity loss upon addition of RBD. The most affected residues belong to the calcium binding site, directly involved in interactions with the bound Fuc (N365, D366, N367, K368). Additionally, the signals for F313 and F374 were completely absent in the presence of the RBD, confirming the placement of Gal/GalNAc close to F313.[20,61,62] Interestingly, a number of residues at a secondary calcium site (D320, L321, Q323, G325, T326 and W327) were also affected. The results for MGL were completely different, reflecting the different dynamic properties of both lectins. The presence of 0.5 equivalents of RBD produced the homogeneous intensity reduction for most of the lectin with the exception of the C-terminal fragment, while 1 equivalent produced the complete disappearance of all the NMR signals in the ^{1}H,^{15}N-HSQC (Figure S8). In order to confirm that the MGL glycan binding site was indeed involved, competition experiments with a simple GalNAc sugar were performed (Figure S9). Suitably, the

FULL PAPER

addition of 1 equivalent of GalNAc produced the recovery of the NMR signals of the lectin, confirming that the RBD and GalNAc compete for the same binding site.
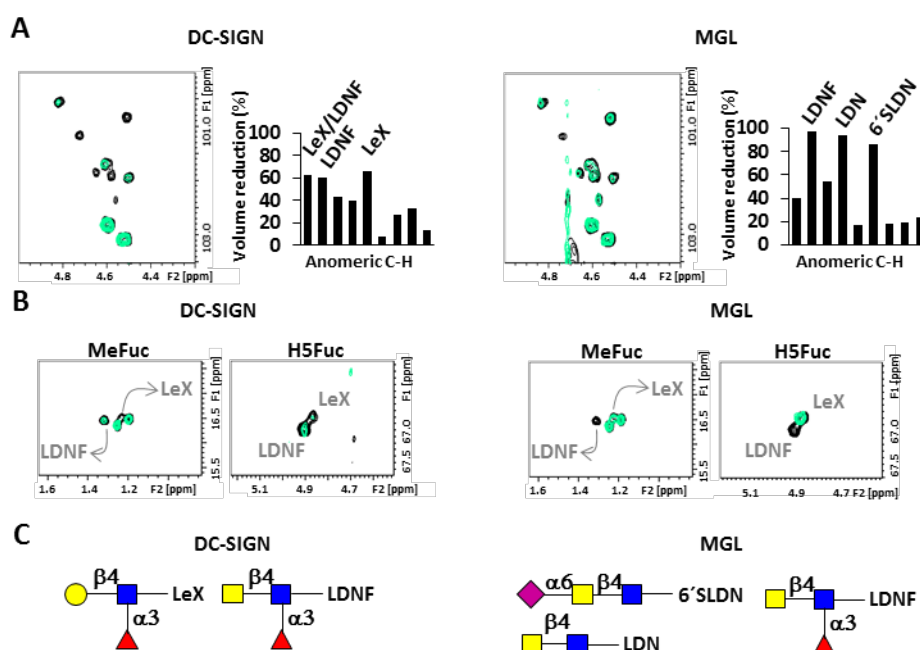
Figure 7. Different interactions of the RBD with C-type lectins DC-SIGN and MGL from the glycan perspective. Selected regions of the ^{1}H,^{13}C-HSQC of RBD alone (in black), with 1 eq. of DC-SIGN (left, superimposed in green), and with 0,2 equivalents of MGL (right, superimposed in green). A. Anomeric region and graphical bar representation for the % of volume reduction (also in Figure 4 with additional cross peaks) B. Specific regions for C6-H6 (Me) and C5-H5 correlations of Fuc in LDNF and LeX. C. Epitopes on RBD recognised by both lectins, as SNFG symbols.

Once the interacting glycan epitopes were experimentally assessed, putative 3D structures were generated for the complexes formed between the glycosylated RBD and several lectins (galectin-3, galectin-7, galectin-8N and DC-SIGN) using the coordinates of the X-ray crystal structures of the lectins (PDB 4R9A, 4GAL, 5GZF, 1SL5, respectively) and that of the RBD within the full S glycoprotein structure (PDB 6VSB) as described in the supporting information. For MGL, an homology model was built since no crystallographic structures are available. Given the existence of two glycosylation sites at the RBD, two 1:1 complexes were generated for each RBD/lectin system, one for the glycan at N331 and a second for that at N343. Molecular dynamics simulations (1 µs) were run for each complex to produce fully equilibrated structures in water solution. The simulations revealed a multitude of highly dynamical glycan-receptor contacts in addition to those established at the canonical sugar binding sites, in agreement with the NMR observations. The formation of fleeting unspecific interactions between RBD glycans and the receptor can be appreciated, often creating an interface between the two proteins. As an example, a 3D perspective of the possible interaction of the glycosylated RBD to galectin-3 is shown in Figure 8, while those for the other lectins are gathered in the SI. Interestingly, interaction with galectin-3 takes place on the opposite side of RBD with respect to the ACE2 recognition region. As a general trend, complexes involving glycans at N343 are more compact and display larger intermolecular contacts than those involving the solvent-exposed glycans at N331 (see SI).

Overall, our study has allowed identifying N-linked glycan epitopes located in the RBD of the spike protein from SAR-CoV2 that serve for the recognition of host lectins, which may contribute to viral infection and subsequent immune exacerbation. An additional analysis of binding and competition using the fully glycosylated spike trimer, will help us test the possible interfering ability of these human lectins either by ACE2 binding competition or by hampering the fusion of the virus with human cells.
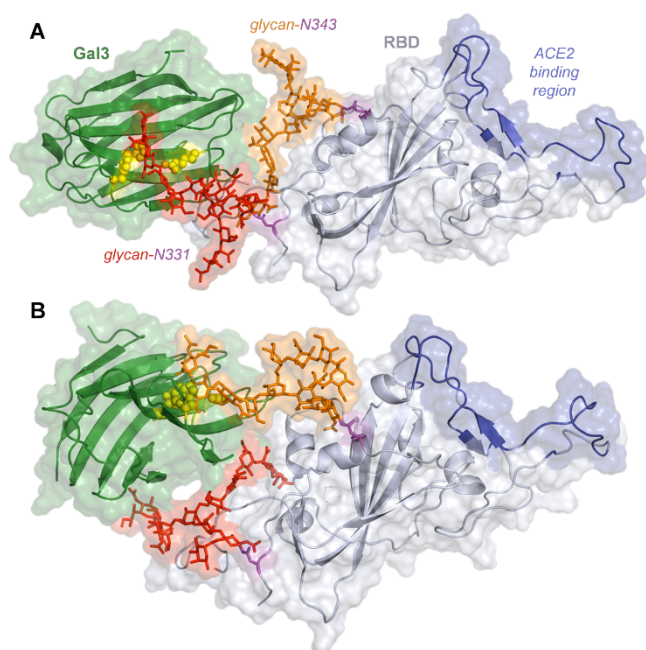
Figure 8. 3D models for 1:1 complexes of galectin-3 (PDB 4R9A) with the 3'SLacNAc epitope attached at both RBD glycosylation sites (PDB 6VSB), according to selected snapshots of 1 μs MD simulations. Binding through glycans at N331 (A) and N343 (B). Galectin-3 and RBD are shown as green and grey cartoons, respectively. The ACE2 binding region of RBD is in blue. Glycans at N331 and N343 are in red and orange sticks, respectively. Glycosylated Asn are in magenta. Key binding residues H162, R62, W181 are shown as yellow spheres.

Conclusion

The RBD fragment of the SARS-CoV-2 S glycoprotein with ^{13}C glycan labelling has been generated. The great sensitivity provided by ^{13}C opens the door to significant opportunities for exhaustive NMR analysis of its glycoprofile and its molecular recognition features. Thus, by employing an NMR-based methodology, which avoids sample digestions and derivatizations, most of the ^{1}H and ^{13}C NMR glycan resonances of the intact (folded) glycoprotein in solution have been assigned, allowing to characterize the specific terminal glycan epitopes exposed on the antennae of the RBD N-glycans. Although the current analysis do not allow for fully quantitative occupancy determination and site specific identification at N331 and N343, it has provided unprecedented structural details. Thus, besides the expected LN, 3'SLN, and 6'SLN terminal moieties, the presence of LDN and its fucosylated LDNF derivative have been assessed. Whereas the former has been detected in a trimer-stabilized version of the SARS-CoV-2 S protein, the presence of the LDNF epitope was unexpected. Indeed,[15] LDN motifs have been found on several mammalian glycoproteins and observed in HEK293-produced glycoproteins.[63] In contrast, the LDNF epitope has been mainly related to pathogens. Additionally, 4-O-sulfated and α2-6 sialylated LDN derivatives, not previously reported either, have also been identified as terminal epitopes, together with the LeX epitope. Overall, our analysis highlights the presence of important levels of N-acetyl-galactosylation and hyper-fucosylation at the terminal chains of the RBD N-glycans, revealing glyco-epitopes not observed in previous MS-based analysis.[13–16] Interestingly, the comparison with a different glycoprotein produced exactly under the same conditions suggests a relationship between the observed high levels of GalNAc and Fuc contents with the protein structure. The exhaustive NMR analysis has also allowed disclosing the main N-glycan scaffold, being complex biantennary, core fucosylated, while lacking bisecting GlcNAc and elongated antennas involving type I polyLacNAc sequences.

The interaction of the glycosylated RBD with a panel of human lectins has also been scrutinized. The ^{13}C-glycan labeling of the RBD has permitted to exploit the ^{1}H,^{13}C-HSQC spectrum of the RBD to report on the specific glycan epitopes recognized by each lectin, affording the corresponding glycan binding selectivity. Thus, while galectins-3 and -7 recognize the LN, LDN and 3'SLN motifs on the RBD, galectin-8-N seems to prefer exclusively the 3'SLN epitope. Siglecs-8 and -10 demonstrated markedly differences, with Siglec-8 unable to recognize any of the glycan epitopes on the RBD, while Siglec-10 interacting with both 3'SLN and 6'SLN. For the C-type lectins, DC-SIGN exhibited selectivity for the two fucosylated terminal epitopes LeX and LDNF, while MGL showed exquisite selectivity for all GalNAc containing epitopes, except for the 4-O-sulfated derivative. The complementary information obtained from the ^{15}N-lectin based experiments permitted to assess that binding occurs through the canonical glycan binding site for each of the lectins (except for Siglec-10). Importantly, all the binding studies have been carried out by using the intact (folded) form of the RBD glycoprotein in solution, allowing to propose atomistic 3D models for the corresponding complexes.

This study paves the way to unveiling the interlaces roles of glycosylation patterns and cell receptors in SARS-CoV-2 infection mechanisms in the cell, particularly the recognition of tissue-dependent ACE2 by full-length glycosylated spike protein (S). Such studies are currently ongoing in our labs.

Additional details are given in the Supporting Information.

[1] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, *Lancet* **2020**, *395*, 497–506.

[2] X. Zhang, Y. Tan, Y. Ling, G. Lu, F. Liu, Z. Yi, X. Jia, M. Wu, B. Shi, S. Xu, J. Chen, W. Wang, B. Chen, L. Jiang, S. Yu, J. Lu, J. Wang, M. Xu, Z. Yuan, Q. Zhang, X. Zhang, G. Zhao, S. Wang, S. Chen, H. Lu, *Nature* **2020**, *583*, 437–440.

[3] J. Shang, G. Ye, K. Shi, Y. Wan, C. Luo, H. Aihara, Q. Geng, A. Auerbach, F. Li, *Nature* **2020**, *581*, 221–224.

[4] A. C. Walls, Y. J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, D. Veesler, *Cell* **2020**, *181*, 281-292.e6.

[5] I. Bagdonaite, H. H. Wandall, *Glycobiology* **2018**, *28*, 443–467.

[6] Y. Watanabe, T. A. Bowden, I. A. Wilson, M. Crispin, *Biochim. Biophys. Acta - Gen. Subj.* **2019**, *1863*, 1480–1497.

[7] W. Van Breedam, S. Pöhlmann, H. W. Favoreel, R. J. de Groot, H. J. Nauwynck, *FEMS Microbiol. Rev.* **2014**, *38*, 598–632.

[8] A. Varki, *Glycobiology* **2017**, *27*, 3–49.

[9] J. Lan, J. Ge, J. Yu, S. Shan, H. Zhou, S. Fan, Q. Zhang, X. Shi, Q. Wang, L. Zhang, X. Wang, *Nature* **2020**, *581*, 215–220.

[10] J. L. Johnson, M. B. Jones, S. O. Ryan, B. A. Cobb, *Trends Immunol.* **2013**, *34*, 290–298.

[11] R. L. Schnaar, *J. Allergy Clin. Immunol.* **2015**, *135*, 609–615.

[12] F. Chiodo, S. C. M. Bruijns, E. Rodriguez, R. J. E. Li, A. Molinaro, A. Silipo, F. Di Lorenzo, D. Garcia-Rivera, Y. Valdes-Balbin, V. Verez-Bencomo, Y. van Kooyk, *bioRxiv* **2020**, 2020.05.13.092478.

[13] A. Shajahan, N. T. Supekar, A. S. Gleinich, P. Azadi, *Glycobiology* **2020**, 1–20.

[14] Y. Watanabe, J. D. Allen, D. Wrapp, J. S. McLellan, M. Crispin, *Science (80-.).* **2020**, *369*, 330–333.

[15] P. Zhao, J. L. Praissman, O. C. Grant, Y. Cai, T. Xiao, K. E. Rosenbalm, K. Aoki, B. P. Kellman, R. Bridger, D. H. Barouch, M. A. Brindley, N. E. Lewis, M. Tiemeyer, B. Chen, R. J. Woods, L. Wells, *bioRxiv* **2020**, 2020.06.25.172403.

[16] Y. Zhang, W. Zhao, Y. Mao, S. Wang, Y. Zhong, T. Su, M. Gong, X. Lu, J. Cheng, H. Yang, *bioRxiv* **2020**, 2020.03.28.013276.

[17] L. Unione, M. P. Lenza, A. Ardá, P. Urquiza, A. Laín, J. M. Falcón-Pérez, J. Jiménez-Barbero, O. Millet, *ACS Cent. Sci.* **2019**, *5*, 1554–1561.

[18] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C.-L. Hsieh, O. Abiona, B. S. Graham, J. S. McLellan, *Science (80-.).* **2020**, *367*, 1260–1263.

[19] A. W. Barb, D. J. Falconer, G. P. Subedi, in *Methods Enzymol.*, Academic Press Inc., **2019**, pp. 239–261.

[20] K. Pederson, D. A. Mitchell, J. H. Prestegard, *Biochemistry* **2014**, *53*, 5700–5709.

[21] J. Iwaki, J. Hirabayashi, *Trends Glycosci. Glycotechnol.* **2018**, *30*, SE137–SE153.

[22] J. M. Pröpster, F. Yang, S. Rabbani, B. Ernst, F. H. T. Allain, M. Schubert, *Proc. Natl. Acad. Sci.* **2016**, *113*, E4170–E4179.

[23] A. Gimeno, S. Delgado, P. Valverde, S. Bertuzzi, M. A. Berbís, J. Echavarren, A. Lacetera, S. Martín-Santamaría, A. Surolia, F. J. Cañada, J. Jiménez-Barbero, A. Ardá, *Angew. Chemie* **2019**, *131*, 7346–7350.

[24] F. Marcelo, F. Garcia-Martin, T. Matsushita, J. Sardinha, H. Coelho, A. Oude-Vrielink, C. Koller, S. André, E. J. Cabrita, H. J. Gabius, S. I. Nishimura, J. Jiménez-Barbero, F. J. Cañada, *Chem. - A Eur. J.* **2014**, *20*, 16147–16155.

[25] L. D. Movsisyan, M. S. Macauley, *Org. Biomol. Chem.* **2020**, *18*, 5784–5797.

[26] S. Neelamegham, K. Aoki-Kinoshita, E. Bolton, M. Frank, F. Lisacek, T. Lütteke, N. O'Boyle, N. H. Packer, P. Stanley, P. Toukach, A. Varki, R. J. Woods, *Glycobiology* **2019**, *29*, 620–624.

[27] Y. Yamaguchi, K. Kato, M. Shindo, S. Aoki, K. Furusho, K. Koga, N. Takahashi, Y. Arata, I. Shimada, *J. Biomol. NMR* **1998**, *12*, 385–394.

[28] K. Kato, Y. Yamaguchi, Y. Arata, *Prog. Nucl. Magn. Reson. Spectrosc.* **2010**, *56*, 346–359.

[29] Y. Yamaguchi, M. Wälchli, M. Nagano, K. Kato, *Carbohydr. Res.* **2009**, *344*, 535–538.

[30] S. Shahzad-ul-Hussan, M. Sastry, T. Lemmin, C. Soto, S. Loesgen, D. A. Scott, J. R. Davison, K. Lohith, R. O'Connor, P. D. Kwong, C. A. Bewley, *ChemBioChem* **2017**, *18*, 764–771.

[31]   G. Nestor, T. Anderson, S. Oscarson, A. M. Gronenborn, *J. Am. Chem. Soc.* **2017**, *139*, 6210–6216.

[32]   D. Z. Vinnitskiy, N. E. Ustyuzhanina, A. S. Dmitrenok, A. S. Shashkov, N. E. Nifantiev, *Carbohydr. Res.* **2017**, *438*, 9–17.

[33]   A. v. Remoortere, C. H.Hokke, G. J. v. Dam, I. v. Die, A. M.Deelder, D. H. v. d. Eijnden, *Glycobiology* **2000**, *10*, 601–609.

[34]   A. E. Luyai, J. Heimburg-Molinaro, N. S. Prasanphanich, M. L. Mickum, Y. Lasanajak, X. Song, A. K. Nyame, P. Wilkins, C. A. Rivera-Marrero, D. F. Smith, I. Van Die, W. E. Secor, R. D. Cummings, *Glycobiology* **2014**, *24*, 602–618.

[35]   M. L. Mickum, N. S. Prasanphanich, J. Heimburg-Molinaro, K. E. Leon, R. D. Cummings, *Front. Genet.* **2014**, *5*, 262.

[36]   M. Weiss, D. Ott, T. Karagiannis, M. Weishaupt, M. Niemietz, S. Eller, M. Lott, M. Martínez-Orts, A. Canales, N. Razi, J. Paulson, C. Unverzagt, *ChemBioChem* **2020**, cbic.202000268.

[37]   E. Bieberich, in *Glycobiol. Nerv. Syst. Adv. Neurobiol.* (Eds.: R. Yu, C.L. Schengrund), Springer, New York, NY, **2014**, pp. 47–70.

[38]   P. Wang, H. Wang, J. Gai, X. Tian, X. Zhang, Y. Lv, Y. Jian, *Sci. Rep.* **2017**, *7*, 40301.

[39]   J. B. Goh, S. K. Ng, J. B. Goh, S. K. Ng, *Crit. Rev. Biotechnol.* **2018**, *38*, 851–867.

[40]   A. Suga, M. Nagae, Y. Yamaguchi, *Glycobiology* **2018**, *28*, 774–785.

[41]   I. Hang, C. W. Lin, O. C. Grant, S. Fleurkens, T. K. Villiger, M. Soos, M. Morbidelli, R. J. Woods, R. Gauss, M. Aebi, *Glycobiology* **2015**, *25*, 1335–1349.

[42]   S. C. Garman, S. Sechi, J. P. Kinet, T. S. Jardetzky, *J. Mol. Biol.* **2001**, *311*, 1049–1062.

[43]   S. M. Dharmesh, T. P. Skelton, J. U. Baenziger, *J. Biol. Chem.* **1993**, *268*, 17096–17102.

[44]   E. Miller, D. Fiete, N. M. J. Blake, M. Beranek, E. L. Oates, Y. Mi, D. S. Roseman, J. U. Baenziger, *J. Biol. Chem.* **2008**, *283*, 1985–1991.

[45]   D. Bonar, F. G. Hanisch, *J. Biol. Chem.* **2014**, *289*, 29677–29690.

[46]   E. P. Go, G. Hewawasam, H.-X. Liao, H. Chen, L.-H. Ping, J. A. Anderson, D. C. Hua, B. F. Haynes, H. Desaire, *J. Virol.* **2011**, *85*, 8270–8284.

[47]   W. R. J. Alley, B. F. Mann, M. V Novotny, *Chem. Rev.* **2013**, *113*, 2668–2732.

[48]   J.-R. Wang, W.-N. Gao, R. Grimm, S. Jiang, Y. Liang, H. Ye, Z.-G. Li, L.-F. Yau, H. Huang, J. Liu, M. Jiang, Q. Meng, T.-T. Tong, H.-H. Huang, S. Lee, X. Zeng, L. Liu, Z.-H. Jiang, *Nat. Commun.* **2017**, *8*, 631.

[49]   M. F. Brinchmann, D. M. Patel, M. H. Iversen, *Mediators Inflamm.* **2018**, *2018*, 1–11.

[50]   L. Johannes, R. Jacob, H. Leffler, *J. Cell Sci.* **2018**, *131*, jcs208884.

[51]   J. L. Caniglia, M. R. Guda, S. Asuthkar, A. J. Tsung, K. K. Velpula, *PeerJ* **2020**, *8*, e9392.

[52]   M. I. Nielsen, J. Stegmayr, O. C. Grant, Z. Yang, U. J. Nilsson, I. Boos, M. C. Carlsson, R. J. Woods, C. Unverzagt, H. Leffler, H. H. Wandall, *J. Biol. Chem.* **2018**, *293*, 20249–20262.

[53]   J. Hirabayashi, T. Hashidate, Y. Arata, N. Nishi, T. Nakamura, M. Hirashima, T. Urashima, T. Oka, M. Futai, W. E. G. Muller, F. Yagi, K. I. Kasai, *Biochim. Biophys. Acta - Gen. Subj.* **2002**, *1572*, 232–254.

[54]   H. Ideo, A. Seko, I. Ishizuka, K. Yamashita, *Glycobiology* **2003**, *13*, 713–723.

[55]   S. Carlsson, C. T. Öberg, M. C. Carlsson, A. Sundin, U. J. Nilsson, D. Smith, R. D. Cummings, J. Almkvist, A. Karlsson, H. Leffler, *Glycobiology* **2007**, *17*, 663–676.

[56]   D. D. Leonidas, E. H. Vatzaki, H. Vorum, J. E. Celis, P. Madsen, K. R. Acharya, *Biochemistry* **1998**, *37*, 13930–13940.

[57]   C. F. Brewer, *Glycoconj. J.* **2004**, *19*, 459–465.

[58]   E. Ermakova, M. C. Miller, I. V. Nesmelova, L. López-Merino, M. A. Berbís, Y. Nesmelov, Y. V. Tkachev, L. Lagartera, V. A. Daragan, S. André, F. J. Cañada, J. Jiménez-Barbero, D. Solís, H. J. Gabius, K. H. Mayo, *Glycobiology* **2013**, *23*, 508–523.

[59]   H. Yu, A. Gonzalez-Gil, Y. Wei, S. M. Fernandes, R. N. Porell, K. Vajn, J. C. Paulson, C. M. Nycholat, R. L. Schnaar, *Glycobiology* **2017**, *27*, 657–668.

[60]   O. Blixt, S. Han, L. Liao, Y. Zeng, J. Hoffmann, S. Futakawa, J. C. Paulson, *J. Am. Chem. Soc.* **2008**, *130*, 6680–6681.

[61]   Y. Guo, H. Feinberg, E. Conroy, D. A. Mitchell, R. Alvarez, O. Blixt, M. E. Taylor, W. I. Weis, K. Drickamer, *Nat. Struct. Mol. Biol.* **2004**, *11*, 591–598.

[62]   P. Valverde, S. Delgado, J. D. Martínez, J.-B. Vendeville, J. Malassis, B. Linclau, N.-C. Reichardt, F. J. Cañada, J. Jiménez-Barbero, A. Ardá, *ACS Chem. Biol.* **2019**, *14*, 1660–1671.

[63]   D. Sugahara, A. Tomioka, T. Sato, H. Narimatsu, H. Kaji, *Proteomics* **2015**, *15*, 2921–2933.

Entry for the Table of Contents
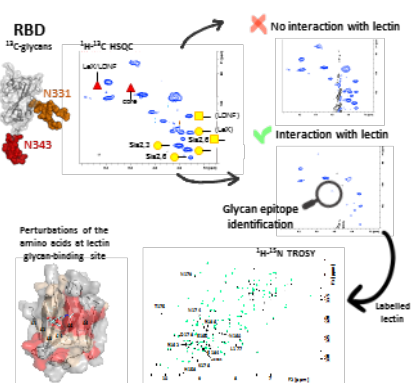
Unprecedented structural details of the glycans of the RBD of SARS-CoV-2 spike glycoprotein have been revealed by NMR. Unexpected and non-previously reported glycoepitopes have been detected. The interaction of the RBD glycoprotein with diverse human lectins has been scrutinised by exploiting the NMR signature of the ^{13}C-glycans. Our analysis permitted to identify the corresponding glycan epitopes responsible for the interaction with each lectin.

Institute and/or researcher Twitter usernames: GlycoNMR@Bilbao, @CICbioGUNE, @jjbarber0