Delineating Binding Modes of Gal/GalNAc and Structural Elements of the Molecular Recognition of Tumor-Associated Mucin Glycopeptides by the Human Macrophage Galactose-Type Lectin

Dr. Filipa Marcelo, Dr. Fayna Garcia-Martin, Dr. Takahiko Matsushita, Dr. João Sardinha, Helena Coelho, Anneloes Oude-Vrielink, Dr. Christiane Koller, Dr. Sabine André, Prof. Dr. Eurico J. Cabrita, Prof. Dr. Hans-Joachim Gabius, Prof. Dr. Shin-Ichiro Nishimura, Prof. Dr. Jesús Jiménez-Barbero, Prof. Dr. F. Javier Cañada

Abstract

The human macrophage galactose-type lectin (MGL) is a key physiological receptor for the carcinoma-associated Tn antigen (GalNAc-α-1-*O*-Ser/Thr) in mucins. NMR and modeling-based data on the molecular recognition features of synthetic Tn-bearing glycopeptides by MGL are presented. Cognate epitopes on the sugar and matching key amino acids involved in the interaction were identified by saturation transfer difference (STD) NMR spectroscopy. Only the amino acids close to the glycosylation site in the peptides are involved in lectin contact. Moreover, control experiments with non-glycosylated MUC1 peptides unequivocally showed that the sugar residue is essential for MGL binding, as is Ca^{2+}. NMR data were complemented with molecular dynamics simulations and Corcema-ST to establish a 3D view on the molecular recognition process between Gal, GalNAc, and the Tn-presenting glycopeptides and MGL. Gal and GalNAc have a dual binding mode with opposite trend of the main interaction pattern and the differences in affinity can be explained by additional hydrogen bonds and CH–π contacts involving exclusively the NHAc moiety.

Introduction

Mucin MUC1 is a constituent of apical surfaces of glandular epithelial cells. In a normal tissue, MUC1 is an extensively *O*-glycosylated protein characterized by elaborated sugar chains such as the branched core 2 *O*-glycans extending from the GalNAc-unit that is directly α-O-linked to either serine (Ser) or threonine (Thr) of the mucin backbone (Figure 1).1 In adenocarcinoma cells, MUC1 can be overexpressed and aberrantly glycosylated, now mostly carrying shortened *O*-glycans, that is, GalNAc-α-1-O-Ser/Thr (Tn) or Galβ1–3GalNAc-α-1-O-Ser/Thr (T), which may also be sialylated.2 Evidently, the glycome signature is drastically altered, and examples are known that such remodeling is sensed by endogenous lectins translating the appearance of the new sugar signal into cellular activities.3 In this case, given the abundant presentation of T/Tn-antigens, members of the families of galectins and C-type lectins, two classes of tissue lectins,4 act as receptors. From the class of Ca^{2+}-dependent lectins, the human macrophage galactose-type lectin (MGL, also referred to as CD301 or CLEC10 A), expressed on the surface of monocyte-derived immature dendritic cells (DCs) and macrophages, has marked specificity to MUC1 (Figure 1).5 Along with contact formation with tumor cells in host defense, it can also attach to effector T cells via the mucin-type glycoproteins CD43 and CD15, whose O-glycan profile is reprogrammed depending on the activation status, leading to a decrease in CD45, tyrosine phosphatase activity and downstream attenuation of T cell

receptor-mediated signaling.6 As an endocytic receptor for Tn-presenting compounds MGL engagement increases DC activation, suggesting an enhanced ability for antigen presentation, along with effects on antigen-specific $CD8^+$ T cells.7 Therefore, acquiring a detailed knowledge on the fine structural details of the molecular recognition process between the tumor-associated glycopeptide antigens and MGL is of paramount importance to optimize structure-based design of novel types of anticancer vaccines. In fact, the study of how glycan targets are accommodated by their protein receptors is an important step toward this aim.8 Experimentally, NMR spectroscopy has proved to be notably useful to obtain structural insights into the solution structure and dynamics of ligand–protein complexes.9 However, owing to the intrinsic complexity and flexibility of glycoconjugates, a multidisciplinary strategy combining glycopeptide synthesis and lectin purification with obtaining NMR parameters and running molecular modeling will need to be taken.10 Recently, the combination of NMR data with results from applying molecular modeling was instrumental to disclose the effect of α-O-glycosylation on the conformation of mucin-type peptides and to detect an interaction with a biomedically relevant galectin.11 Underscoring the value of the concept, interaction of mucin-related glycopeptides with a model plant lectin, using saturation transfer difference (STD-NMR) measurements and molecular modeling, had been monitored.12 The STD NMR-spectroscopic approach has in this context further been introduced to determine the target epitope of short MUC1 glycopeptides for the breast cancer-selective monoclonal antibody SM3 and to map the contact site of the Le^x trisaccharide in MGL1, one of the two murine orthologues of human MGL, in solution.13

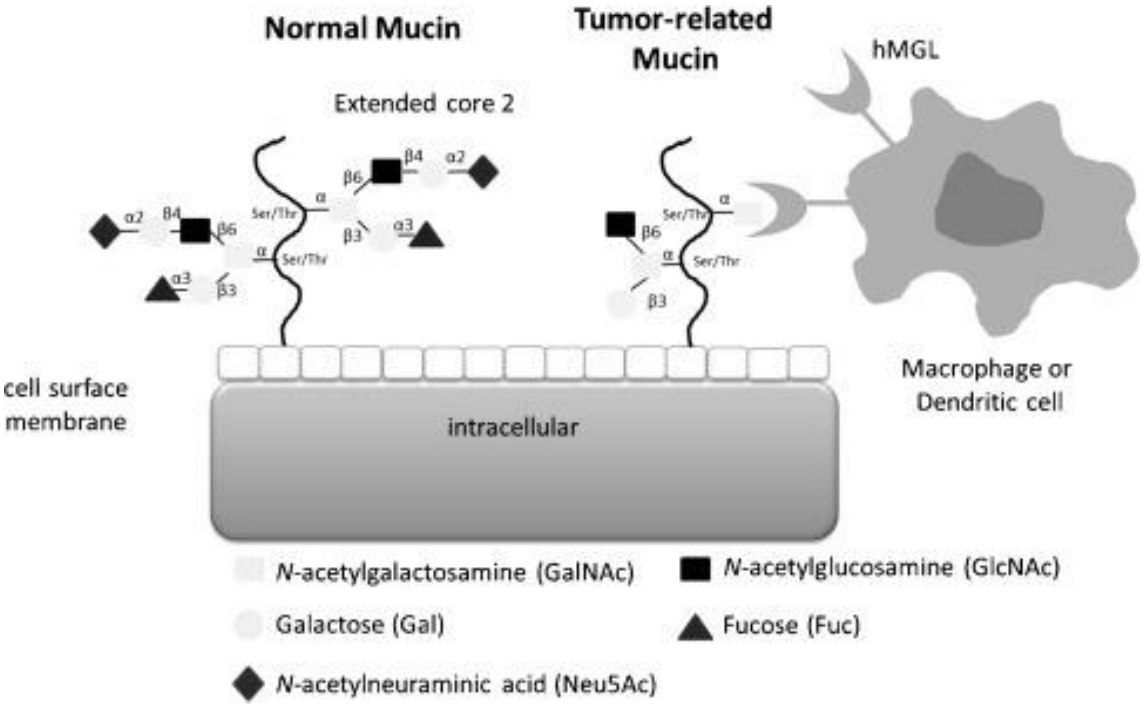

Figure 1. Representation of the interaction of MUC1 structures expressed on tumor cells by MGL, which specifically binds to MUC1-Tn on the cancer cell surface. In contrast, MUC1 in normal epithelial cells, which carries elongated and branched O-glycans, is much less reactive.

Owing to the lack of structural knowledge on MGL binding to MUC1, we have studied these properties with MUC1 glycopeptides (Figure 2) of different length (10, 15, and 20 aa) with the GalNAc-unit linked to Thr4 (**1** and **2**) or Thr9 (**3** and **4**), and the carbohydrate recognition domain (CRD) of MGL by STD-NMR techniques. As controls, the binding modes of α-methyl Gal (**5**) and α-methyl GalNAc (**6**) have also been characterized. Also, we used the non-glycosylated peptides (**7** and **8**) to confirm if the presence of GalNAc moiety is essential for binding. Furthermore, to reveal affinity levels, the dissociation constants of **5**, **6**, and MUC1 glycopeptides **1**–**4** to MGL were estimated from STD titrations or STD competition experiments, and the crucial role of calcium ion for the recognition process was ascertained. Finally, a NMR-based 3D model of monosaccharide derivatives α-methyl Gal (**5**) and α-methyl GalNAc (**6**), as well the Tn antigen (**4**) bound to MGL, was derived by molecular dynamic simulations without constraints. For that purpose, a MGL homology model, built using the carbohydrate recognition domain of the hepatic asialoglycoprotein receptor as model reference (PDB code 1DV8),14 was used as the basis for the MD simulations.

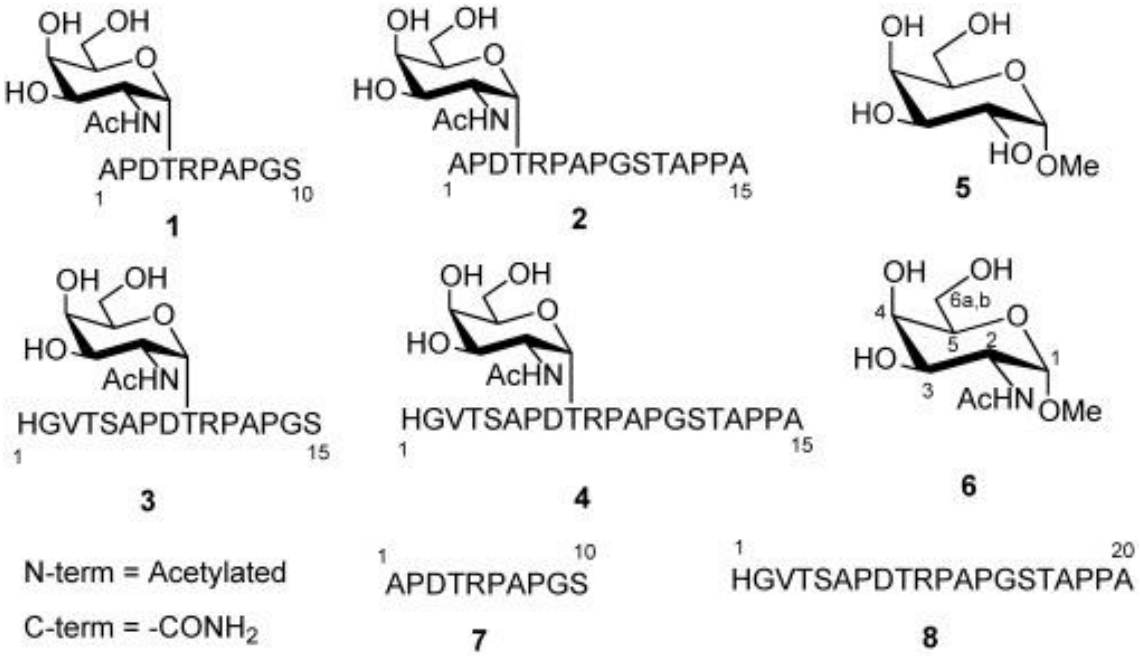

Figure 2. MUC1-derived (glycol)peptides and monosaccharide derivatives studied in this work along with the corresponding numbering.

Results and Discussion

STD application, suited to obtain reliable mapping of ligand epitopes,15 critically depends on the existence of fast exchange of the ligand, in the relaxation timescale, between the free and the protein-bound status. When such prerequisites are satisfied, STD will enable to deduce binding affinities.16 Considering the versatility of the method to characterize carbohydrate binding to distinct receptors,17 we tested its applicability to MGL interacting with the MUC1 peptides **1**–**4** and monosaccharide derivatives **5** and **6**.

Epitope mapping of MUC1 glycopeptides

To determine the structural elements of the MUC1 glycopeptides upon binding to MGL, STD experiments were performed on 1:12 molar ratio mixtures of MGL in the presence of the individual glycopeptides **1**–**4**. To establish the binding specificity of the system, different STD conditions had first to be verified. STD spectra of glycopeptides **2** and **3** in the presence and absence of MGL are compared in the Supporting Information, Figure S1. Proper control experiments were performed with the ligands in the presence and absence of MGL to optimize the frequency for protein saturation and at the same time to ensure that the ligand signals were minimally or not affected. MUC1 glycopeptides **3**, **4**, and **8** when irradiated at −1 ppm in absence of protein showed residual STD intensities on the aliphatic methyl groups in the STD spectra that were considered (subtracted) when analyzing the STD spectra in presence of protein. Furthermore, STD experiments recorded at different concentrations of MGL (Supporting Information, Figure S2) revealed that STD signal intensities clearly increased with the protein concentration proving the specificity of the binding process. Moreover, changing the temperature from 298 K to 310 K a significant increase of the STD signal occurred (Supporting Information, Figure S3), probably by increasing the off-rate constant (k_{off}) of the exchanging glycopeptide–lectin system.18

Figure 3 (Supporting Information, Figures S4–S6) compare the off resonance spectra (named as off res) with the corresponding STD spectra of **1**–**4**. STD signals were detected for all glycopeptides, demonstrating specific interaction with MGL. Interestingly, a clear epitope selection could be defined for all of them (Figure 4; Supporting Information, Figures S7–S10). Other studies using ELISA- and array-based glycan screening pointed out the preferential binding of MGL to Tn- and sialyl Tn-antigen, as well by STD-NMR it has been previously shown that the recognition occurred mainly through GalNAc moiety.19 Of note, the role of the peptide backbone and the amino acid side chains in binding remained undetermined. Looking at the glycopeptides, as depicted in Figure 4, the most intense signals originated in all cases are from the carbohydrate moiety, especially the GalNAc H2 proton. In fact, Mortezai and co-workers pointed out that Tn antigen binds mainly via the H-2, H-3 and H-4 regions of the GalNAc residue.19b Of note, protons of amino acids in the vicinity of GalNAc received much more saturation than those located further away. This evidence strongly suggests that, in addition to sugar-lectin interaction, MGL will also be in contact with amino acids.

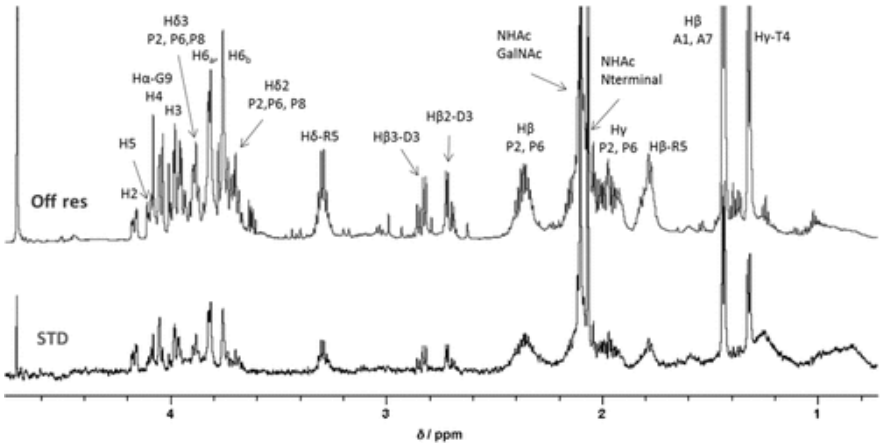

Figure 3. ^{1}H NMR 600 MHz spectra for APD[T]RPAPGS 1 (240 μM concentration) in the presence of MGL (20 μM) at 310 K. Top: reference spectrum (off res); bottom: the STD spectrum (STD). [T]=GalNAc is attached in T. The key proton resonances are marked.

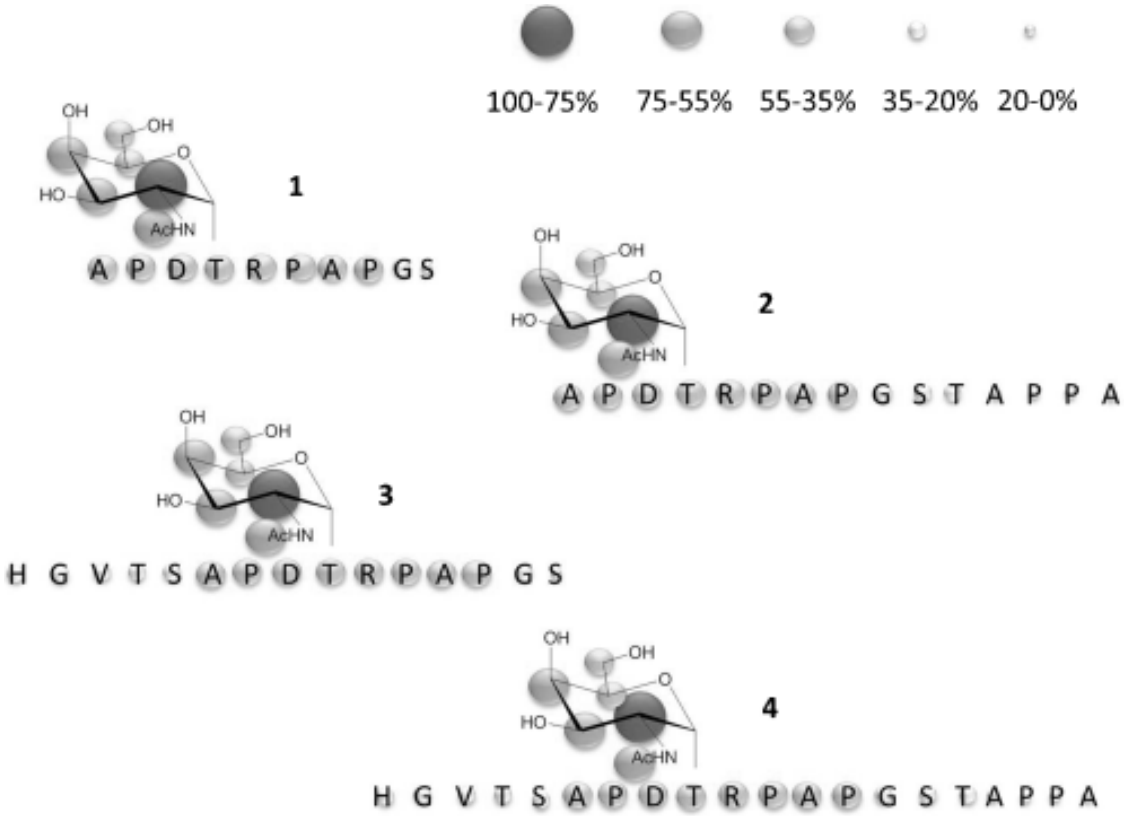

Figure 4. Epitope mapping obtained for glycopeptides **1**–**4** with MGL. For clarity, the STD response of each amino acid corresponds to the average of STD percentages of all amino acid proton resonances that were measured with sufficient accuracy. The percentage saturation of each proton of the amino acids in peptide backbone is given in the Supporting Information, Figures S7–S10. H2 of GalNAc is the proton that receives the maximum percentage saturation from the protein (normalized to 100 %). The signal of the anomeric proton of GalNAc could not be analyzed in the STD spectra owing to their proximity to the HDO signal.

The difference on the STD response for distinct amino acids depending to the distance to the Tn antigen are documented in the Supporting Information, Figures S11 and S12. Fittingly, the STD response of the terminal methyl acetamide at the N-terminus significantly decreased when its distance to the Tn determinant was increased. Among the different Ala amino acid residues, those that received a higher percentage of saturation were always those closer to the Tn fragment (Supporting Information, Figure S12). As negative control, STD spectra of the MUC1-peptides **7** and **8** in the presence of MGL were recorded. As no signal was detected under any experimental conditions, it can be concluded that the carbohydrate is crucial for recognition by MGL (Supporting Information, Figures S13, S14), in contrast to the interaction with certain MUC1 antibodies, namely the SM3 breast-cancer-related antibody that has been shown capable to bind non-glycosylated peptides.20 Therefore the binding properties of the free sugar as the α-methyl derivative were investigated.

α-Methyl Gal 5 and α-methyl GalNAc 6 recognition by MGL

The binding modes of Gal (**5**) and GalNAc (**6**) to MGL were investigated by STD (Supporting Information, Figures S15, S16). Interestingly, epitope mapping for the two monosaccharide derivatives tracked down a remarkable difference (Figure 5). For **6**, the H2 received the maximum of saturation from the protein, closely followed by the H4 (77 %), H3 (79 %), and NHAc (78 %) protons. Overall, a similar trend was observed for the GalNAc residue in the glycopeptides **1**–**4**. In contrast, for **5**, H4 received much more saturation transfer when compared to the other sugar protons (H2 and H3 only receive 53 %). These differences in the contact pattern strongly suggest the existence of a distinct presentation of **5** and **6** towards MGL. Using STD competition experiments,21 differences in binding mode of **5** and **6** were indeed shown to be associated with rather different binding affinity of these monosacharides to MGL (see below). Furthermore, STD competition experiments revealed that glycopeptide **1** competes with **5** and **6** for the same MGL binding site (data not shown).

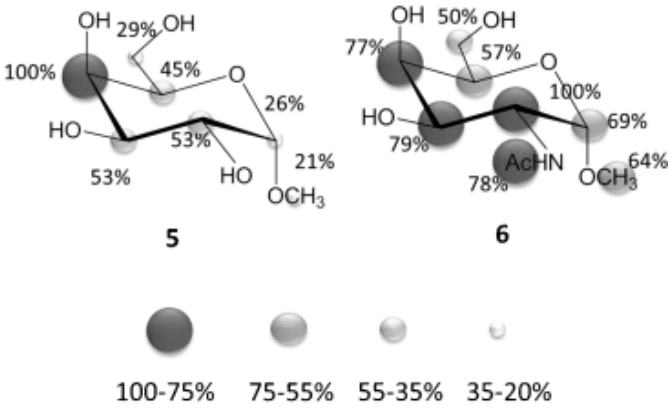

Figure 5. Epitope mapping obtained for Gal **5** and GalNAc **6** with MGL.

Dissociation constants for the binding of 1–6 to MGL

The dissociation constants (K_d) of **1–4** and **6** versus MGL were estimated using STD displacement experiments, using **5** as reference ligand. As the estimation

of K_d through this procedure involves the employment of a reference ligand of known affinity, the K_d value of **5** with MGL was first estimated employing STD titrations experiments based on time dependence of saturation build-up curves.15c, 16 The detailed procedure of the STD titration is described in the Supporting Information (Supporting Information S11–S13 and Table S1). Beyond the intrinsic experimental error, STD titration facilitated us to deduce a K_d value in the mM range (ca. 0.9±0.3 mM) for **5**. As a next step, the intensity variations in the STD spectrum of **5** in the presence of MGL were monitored by adding either **6** or any of the glycopeptides **1**–**4**. In particular, the decrease of the STD signal intensities of H4, H2, and H3 of **5** was measured. The detailed procedure is summarized in the Supporting Information (Supporting Information S14–S16).

The inhibition constants values are summarized in Table 1. GalNAc (**6**) exhibits an inhibition constant value two orders of magnitude lower (in the low μM range) than the K_d value of Gal (**5**), reflecting the high degree of MGL selectivity towards GalNAc-containing species. This property is shared by the mammalian hepatic asialoglycoprotein receptor.14 Testing neoglycoconjugates with MGL, the Tn epitope was fittingly strong, the T antigen considerably less reactive, and the weak ligand N-acetyllactosamine turned into an avid binding partner by turning the terminal Gal to GalNAc to form LacdiNAc.22 Similar competition experiments with Tn-glycopeptides **1** and **4** showed inhibition constants in the same order of magnitude as **6** (Table 1). Fittingly, the STD-derived epitope mapping had already inferred that only the neighboring amino acids from Tn fragment appeared to contribute to the interaction with MGL.

Table 1. Dissociation (K_d) and inhibition constants (K_i) determined by STD-NMR at 298 K.

Ligand	**K_d or K_i [μM]**
α-methyl Gal (**5**)	K_d=900±300
α-methyl GalNAc (**6**)	K_i=12±1
APD[T]RPAPGS (**1**)	K_i=35±5
HGVTSAPD[T]RPAPGSTAPPA (**4**)	K_i=23±2

Following the epitope mapping, STD-NMR can also assess the Ca^{2+} dependence. STD experiments in presence and absence of calcium unequivocally ascertaining the key role of calcium in the recognition process (Supporting Information, Figure S20).

3D view of the interaction of α-methyl Gal 5 and α-methyl GalNAc 6 by MGL

First, a homology model of the carbohydrate recognition domain (CRD) of MGL (Cys181-Leu308) was built using as template the structure of the CRD of the mammalian asialoglycoprotein receptor (PDB code 1DV8)14 with which MGL shares 73 % identity (Supporting Information, Figure S21). Next, crystal structures of other C-type lectins in complex with galactose-related ligands were analyzed to dock the ligands **5** and **6** in the binding site of the MGL model. At least two binding modes for galactoside derivatives, herein defined as Mode A

and Mode B, can be observed from the crystal structures of other C-type lectins. The sugar-face orientation towards the protein can be rather different, close to a 180° rotation, defining interaction modes, A or B. Mode A is present in a mutant form of the collectin mannose-binding lectin, which contains a dodecapeptide with QPDWG in the place of residues 185–191, switching specificity from Man to Gal/GalNAc (PDB code 1BCH),23 in the C-type lectin CEL-I isolated from sea cucumber and complexed with GalNAc (PDB code 1WMZ),24 in the murine scavenger receptor, a C-type lectin, complexed with the Le^x epitope (PDB code 2OX9)25 and in the galactoside-specific C-type lectin isolated from rattlesnake venom and complexed with lactose (PDB code 1 JZN),26 all of them harboring the signature sequences QPD and WND,27 along with the tryptophan residue located one amino acid apart from QPD signature (QPDxW) that stacks with the galactopyranoside. As a consequence, when **5** or **6** are docked in mode A to the MGL model a stacking interaction of protons H3, H4, H5, and H6a,b with the aromatic system of W271 (numbering in full-length MGL), a part of the QPDxW signature located in the long loop of the binding site, becomes apparent. In contrast, for mode B the stacking interaction with the π-electronic system of W271 involves H2, H1, and OCH3. Mode B is observed in another galactose-binding lectin, CEL-IV, from sea cucumber and complexed with raffinose (PDB code 3ALU),28 in the human langerin, a C-type lectin expressed by Langerhans cells, complexed with a sulfated galactoside (PDB code 3P5I)29 and in a tunicate C-lectin complexed with galactose (PDB code 1TLG).30 In each of these cases, the sequences lack the QPD signature and the equivalent tryptophan (W271 in MGL) in the loop. However, CEL-IV and the tunicate lectin contain a tryptophan residue at the opposite site of the binding site (W79 and W100, respectively), enabling CH–π interactions with the ligand, which is rotated by about 180° when compared to binding mode A. Interestingly, a His residue (H286) in MGL at the equivalent position to W100 in the tunicate lectin and a Tyr moiety (Y236) is equivalent to W79 in CEL-IV. With respect to the Ca^{2+} ion at the binding site, contacts with the 3′ and 4′ hydroxy groups, the discriminatory set, are mandatory for Gal-/GalNAc specificity. In both orientations of the galactoside ring, the hydroxy groups at positions 3 and 4 are indeed close to Ca^{2+}. Hence, these two starting model geometries of the complexes of MGL with Gal **5** and GalNAc **6** were subjected by molecular dynamics (MD) simulations. These simulations without experimental restraints of the selected complexes were recorded during 20 ns at 298 K in explicit water using the Amber 12 package.31 Selected structures from MD runs, for both binding modes A and B and in presence of Gal **5** (Figure 6, complex 1 and 2) and GalNAc **6** (Figure 6, complex 3 and 4) were analyzed by Corcema-ST (complete relaxation and conformational exchange matrix-saturation transfer).32 This program enables the saturation of each proton of the ligand to be predicted depending on the geometry of the complex, dissociation constant, and irradiation conditions. The sugar protons that receive a higher percentage of saturation are rather different for modes A and B (Supporting Information, Figures S22 and S23). Mode A (Figure 6, complex 1 and 3) depicts H4 as the proton that receives the maximum of transfer of saturation from the protein, while in mode B (Figure 6, complex 2 and 4) the H2 of the sugar collects the highest percentage of saturation. With these models at hand together with the Corcema-ST analysis (Figure 7 A), the experimental data of Gal (**5**) can be explained based on binding mode A, given as complex 1 in Figure 6, with a minor contribution of mode B (Supporting Information, Figure S22). However, the STD data of GalNAc (**6)** can

only be explained if both binding geometries, mode A and B corresponding to complexes 3 and 4 in Figure 6, are present. Only considering the participation of both modes, in around 4:1 proportions of modes B and A, achieve the profile saturation of the protons of compound **6** (Supporting Information, Figure S23). In fact, the combination of STD-NMR data with CORCEMA-ST calculations has already been employed to demonstrate the dual binding character of a mannose disaccharide to DC-SIGN.33 The complex with **5** (complex 1) and complexes with **6** (complexes 3 and 4) were stable during of the entire simulation period (Supporting Information, Figures S24–26). Intermolecular interactions, namely hydrogen bond (H-bonds) and CH–π interactions, were monitored along of the MD trajectory, as well as the role of Ca^{2+}. For the complexes 1, 3, and 4, highly stable H-bonds were deduced, namely between OH3, OH4/OH6 and the carboxylate groups of E280 (E100 in CRD template) and D269 (D89 in CRD template) of MGL (96-99 % of occurrence; Supporting Information, Figures S27–S29). Other less populated H-bonds (10–30 % of occurrence), including O4, O3, O6, and Q267 (Q87 in CRD template) and N292 (N112 in CRD template) of MGL, were also detectable. Additional H-bonds between the acetamido group NHAc and MGL were observed in the complex MGL/**6** (Supporting Information, Figures S28 and S29). One involves the carbonyl of NHAc and the amino group of K264 (K84 in CRD template) with 16 % of occurrence for mode A, another the NH of NHAc and the carboxylate function of D269 (D89 in CRD template), with 10 % of occurrence for mode B. These interactions were absent in the models containing **5**.

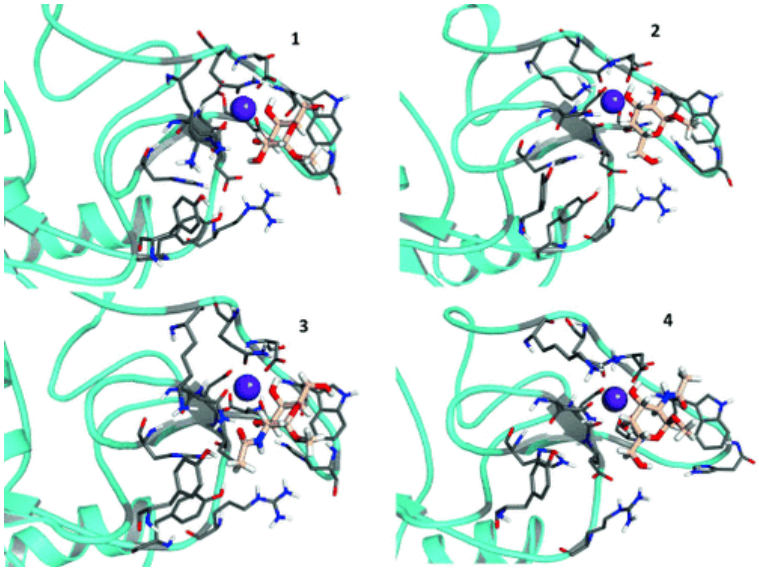

Figure 6. Complexes of MGL selected from MD run considering both binding modes A (left panel) and B (right panel) in presence of α-methyl Gal **5** (complex 1 and 2) and α-methyl GalNAc **6** (complex 3 and 4).

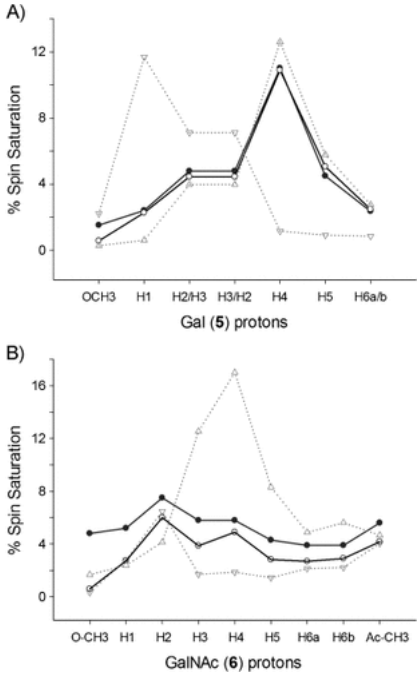

Figure 7. Representation of the percentage of spin saturation for each proton of A9 Gal **5** and B) GalNAc **6**. Experimental value obtained after 2 s irradiation of MGL (•,—, irradiation frequency, −0.5 ppm), calculated value using Corcema-ST for each calculated binding mode geometry A (▵ ,••••) and B (▿ ,••••), calculated value for a linear combination (0.85 A+0.15B for **5** and 0.2 A+0.8B for **6**) of both calculated binding modes (○,—).

Owing to their importance for sugar binding,34 presence of CH–π interactions was also scrutinized. The distances between sugar protons and the π-electron systems of W271, Y236, H284, and H286 (W91, Y56, H104, and H106 in CRD template) were monitored. A CH–π distance of below 4.5 Å and an angle larger than 120° were considered as prerequisites.35 In binding mode A, CH–π interactions between the α-face of **5** and **6** (H-3, H-4, H-5, and H-6a and H6b) and W271 showed up (Supporting Information, Figures S30–S33), in full agreement with X-ray structures mentioned above (PDB codes 1BCH, 1WMZ, 2OX9, and 1JZN). Additionally, for the complex of **6**, a CH–π interaction between the CH_3 of the NHAc moiety and the aromatic system of Y236 (Supporting Information, Figure S34) was observed. Alternatively, from the analysis of mode B MD trajectories (MGL complexed with **6**), a different CH–π interaction was deduced for the CH_3 of the NHAc moiety, that can now be proposed to occur with W271 (Supporting Information, Figure S35). Fittingly, since MGL has aromatic amino acids at both sides of the sugar-binding site, CH–π interactions may take place between the CH_3 of the NHAc moiety and these residues. The difference in affinity constants for **5** and **6** is caused by engaging the N-acetyl group in contacts, both H-bond and stacking. The spatial features of Ca^{2+} in binding ligands **5** and **6** by MGL were evaluated (Supporting Information, Figures S36–S38).

3D view of the interaction of MUC1 glycopeptides by MGL

The 3D model of the complex formed between the Tn antigen (GalNAc-α-1-*O*-Thr) and MGL was generated considering the binding modes A and B. MD simulations were carried out using glycopeptide **4** as ligand. As in the case of **6**, both A- and B-type structures of the MGL/**4** complex appeared to be very stable during of the entire simulation period (Supporting Information, Figures S39, S40). Apparently, MGL can accommodate both GalNAc binding modes (A and B), even if it is bound to a peptide chain as in these Tn-antigen structures. Figure 8 presents the spatial distribution function (SDF) of the carbohydrate unit and the edges of the peptide chain for both binding modes. Obviously, the GalNAc residue is kept within the binding site during the complete MD simulation, while the peptide edges sample a wide range of the conformational space.

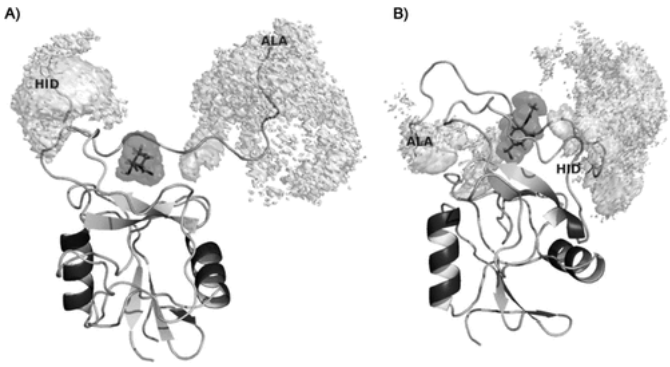

Figure 8. Spatial distribution function (SDF) of the carbohydrate unit and edges of the peptide chain along the trajectory of 20 ns of the MD simulation. Panel A: binding mode **A**, Panel B: binding mode **B**.

When counting intermolecular interaction, for the A or B binding modes, an increased number of H-bonds was identified for the Tn-glycopeptide **4** compared to **6** (Supporting Information, Figures S41 and S42). However, this increase is not reflected in the relative K_d values (Table 1), which is probably due to the delicate enthalpy-entropy balance that takes place in any ligand–protein recognition process. Noteworthy, multivalent presentation of the multiple Tn epitopes and binding of each CRD of the MGL trimer can lower the off rates, as known from other human and also plant lectins.36 From a structural perspective, the bound glycopeptide **4** adopts an extended conformation, which is in complete agreement with the uniform transfer of magnetization to the Hα protons along the peptide backbone. This result is also in accordance with previous X-ray structures of complexes involving MUC1 glycopeptides (PDB code 1SM3 and 2FO4).20, 37, 38

Conclusion

The strategic combination of glycopeptide synthesis and recombinant protein production with NMR spectroscopy and molecular modeling provides fine-structural insights into the interaction between a GalNAc derivative and GalNAc-bearing MUC1-derived glycopeptides with MGL. Contacts by hydrogen and coordination bonds as well as CH–π interactions of the carbohydrate underlie the

measured affinity, the peptide portion apparently playing no major role. Design of additional contacts in this area would improve recognition to this endocytic receptor in vitro/in vivo, our structural models providing a platform for respective efforts, which then could also help blocking virus (filovirus such as Ebola or Marburg or influenza) entry39 or attenuate effector T cell activity in autoimmune diseases and chronic inflammation.

Experimental Section

General methods for synthesis of MUC1 (glyco)peptides (1–4, 7, 8)

MUC1 (glyco)peptides were synthesized by solid-phase peptide synthesis. The detailed description of the synthesis is summarized in the Supporting Information. The crude compounds were purified by preparative RP-HPLC to give **1**–**4**, **7**, and **8** (Supporting Information, Figures S43–S48). They were characterized by analytical RP-HPLC, ESI-HRMS, and amino acid analysis (Supporting Information, Figures S49–S54).

Expression and purification of extracellular domain of MGL

The extracellular domain of MGL (UNIPROT entry number Q8IUN9) expressed in this work corresponds to the sequence from D68 to H316 with an additional starting methionine and includes a potential coiled coil segment and the C-type lectin domain (from 85 to 176 and from 188 to 305 amino acid residues, respectively; Supporting Information, Figure S18). The cloning of cDNA and recombinant production was performed as described,5a except for use of pET3d vector, induction with IPTG at 0.1 mM, and further incubation at 30 °C. Following processing of the inclusion bodies, solubilization and dialysis5a affinity chromatography was performed using lactose-Sepharose 4B obtained after divinyl sulfone activation, product analysis by gel electrophoretic separation, and activity by a solid-phase assay using GalNAc-presenting neoglycoprotein as matrix.40 Note that analytical ultracentrifugation (AUC) preliminary data (data not shown) indicates that the expressed extracellular domain of MGL exists in oligomeric forms higher than dimers in solution. Before running the NMR spectra, the buffer of the protein samples were exchange by means of several cycles of dilution–ultrafiltration (Vivaspin6, 5.000 Dalton molecular weight cutoff, Sartorius Stedim Biotech, Germany) to 10 mM perdeuterated Tris(D11)–DCl (Sigma–Aldrich) in deuterated water with 1 mM $CaCl_2$ and 75 mM NaCl, and uncorrected pH meter reading pH 7.5.

NMR spectroscopy

NMR experiments were recorded on a Bruker Avance 500 MHz equipped with a triple channel TXI probe or a 600 MHz spectrometer equipped with a triple-channel cryoprobe head. The detailed procedure for ^{1}H NMR characterization are described in the Supporting Information (Figures S55–S60). In the case of α-methyl Gal (**5**), a series of STD experiments were performed at 298 K with ca. 20 μM of MGL on a 500 MHz spectrometer at various saturation times (0.5, 1, 1.5, 2, 3.5, and 5.5 s) and ligand concentrations (0.5, 1, 2, 3.5, and 6 mM). The on-resonance frequency was set at −0.5 ppm and off-resonance frequency at 100 ppm and the spectra were recorded in an interleaved fashion using “stddiff” pulse

sequence included in TOPSPIN 2.1 software without water suppression. In the case of α-methyl GalNAc (**6**), the molar ratio was adjusted between 20:1 to 100:1 ligand/MGL and STD-NMR experiments recorded at 298 K and 310 K and with 2 s irradiation time using pulse sequences with or without water suppression in a 600 MHz spectrometer. In the case of the glycopeptides the STD-NMR experiments were performed for 12:1 glycopeptide/MGL molar ratios at 310 K with 30 μM of MGL (otherwise indicated). The negative controls with non-glycosylated MUC1-peptides were recorded using the same experimental conditions. A standard STD-NMR spectrum for the glycopeptides was acquired with 2048 transients in a matrix with 16 K data points in the acquisition channel T2 in a spectral window of 12019.23 Hz centered at 2825.27 Hz. An excitation sculpting module with gradients was employed to suppress the water proton signals. Selective saturation of the protein resonances (on resonance spectrum) was performed by irradiating at −1 ppm using a series of Eburp2.1000-shaped 90° pulses (50 ms, 1 ms delay between pulses) for a total saturation time of 2.0 s. For the reference spectrum (off resonance), the samples were irradiated at 100 ppm. Proper control experiments were performed with the ligands in the presence and absence of MGL in order to optimize the frequency for protein saturation and to ensure that the ligand signals were not affected (−1 ppm for glycopeptides and −0.5 ppm for monosaccharides). MUC1 glycopeptides **3**, **4**, and **8** in solution showed residual STD intensities in the STD spectra that were taken into account (subtracted) when analyzing the STD spectra. In all cases, to accomplish the epitope mapping of each ligand, the STD intensities were normalized with respect to that with the highest response. The signal of the anomeric proton and also the Hα protons of the amino acids of glycopeptides could not be analyzed in the STD spectra that were run applying water suppression pulse sequences, which was due to their close distance to the HDO signal.

Molecular dynamics simulations

The homology model was constructed with the SwissProteinModel Portal.41 The selected structural template protein, the carbohydrate recognition domain of the H1 subunit of the hepatic asialoglycoprotein receptor (PDB code 1DV8 chain A, aa C180–L308),14 has a 73 % sequence identity with the carbohydrate recognition domain of the MGL (CRD, aa H187–H316 of MGL). The overlay of the amino acid sequences of both proteins is shown in the Supporting Information, Figure S19. As this procedure does not include Ca^{2+}, the divalent cation was then introduced in the model using the equivalent position found in the mutated mannose binding protein complexed with GalNAc (PDB code 1BCH).23 3D model of MGL was minimized using the AMBER 12 package31 with the ff12SB force field.42 Calcium parameters (frcmod and prep files)43a were freely downloaded from Amber parameter database of Bryce group.43b

The eicosapeptide **4**, HGVTSAPDT*RPAPGSTAPPA, O-glycosylated with α-GalNAc at Thr-9, was constructed using the Glycam0644 and ff12SB force fields as implemented in Amber 12. The built geometry of compound **4** was then minimized to relieve the possible clashes, starting from an extended conformation of the peptide backbone. Available X-ray structures of distinct C-type lectin complexes clearly points out two distinct binding modes for galactoside derivatives. Thus, to build the molecular complexes of MGL, its structure was first

aligned with two different C-type lectin complexes: a) loaded with GalNAc (PDB code 1BCH);23 and b) loaded with Gal (PDB code 3ALU). Complexes of MGL with α-methyl Gal **5** and α-methyl GalNAc **6** were generated considering both binding modes found on the literature and then submitted to MD simulations to evaluate their stability. Thereafter, the starting 3D geometries were placed into a 10Å octaedric box of explicit TIP3P waters,45 and counterions were added to maintain electroneutrality. Two consecutive minimizations were performed: 1) involving only the water molecules and ions, and 2) involving the entire system. Then, molecular dynamics simulations without constraints were recorded, using a NPT ensemble with periodic boundary conditions and the particle mesh Ewald method to treat long-range electrostatic interactions. The systems were then heated and equilibrated in two steps: 1) 50 ps of MD heating the whole system from 100 to 298 K, followed by 2) equilibration of the entire system during 100 ps at 298 K. The equilibrated structures were the starting points for MD simulations (20 ns) at constant temperature (298 K) and pressure (1 atm). The constraint algorithm SHAKE was used to keep bonds involving H atoms at their equilibrium length, allowing a 2 fs time step for the integration of Newton's equations of motion. At the end, the structures from the MD simulations were selected according to the NMR experimental data, namely the STD data. A detailed analysis of each MD trajectory (for example r.m.s.d. evaluation, hydrogen-bond and CH–π analysis) was accomplished using the cpptraj module included in AmberTools 12 package. A cluster analysis (up to a maximum of 5 clusters) was also carried out to select the MD frames along the 20 ns of production to be submitted to Corcema-ST calculations. At end, a glycopeptide **4**/MGL molecular complex was built taking in account in MD calculations without constraints and Corcema-ST results from α-methyl GalNAc **6**.

Corcema-ST calculation

Corcema-ST matlab scripts for predicting proton ligand saturation transfer provided by Dr. Rama Krishna32 (with some in house modifications to include experimental and calculated saturation values in the input and output PDB formatted files) were applied to the modeled structures of the complexes, obtained after molecular dynamics calculations, between α-methyl-Gal **5**, or α-methyl-GalNAc, **6**, with MGL. Representative structures from all clusters were selected and at the end an overall of 10 structures were analyzed by Corcema-ST. The input parameters used in the calculations were: 2 s saturation time; amino acid in a radius of 10 Å around the ligand; direct irradiation on methyl groups of Ile, Leu, and Val (as an approximation to −0.5 ppm experimental irradiation frequency); calculated K_d dissociation constants, 0.88 mM for **5** and 0.013 mM for **6**; experimental concentration conditions, 0.025 mM and 0.5 mM in the case of **5** and 0.01 mM MGL and 0.2 mM in the case of **6**. A k_{on} of between 10^8–10^{10} L mol^{-1} s^{-1} was used assuming a diffusion controlled kinetic model; correlation times of 0.5 and 48 ns for the ligand in the free and bound form, respectively, were estimated following an empirical approximation46 and considering a 84 kDa trimeric oligomeric form for the extracellular domain of MGL.19a

Acknowledgements

The authors thank FCT Portugal for a post-doc research grant (SFRH/BPD/65462/2009) and for the financial support of EXPL/QEQ/MED/0799/2012 project, CQFB Strategic Project PEst-C/EQB/LA0006/2013 and MINECO, Spain, for the financial support of project CTQ2012–32025. The NMR spectrometer at FCT-UNL is part of the National NMR Network (RNRMN) and is funded by FCT-Portugal RECI/BBB-BQB/0230/2012. The generous financial support by the EC-funded Marie Curie Initial Training Network GLYCOPHARM (PITN-GA-2012–317297) and the GLYCOHIT program (grant agreement 260600) are also gratefully acknowledged. F.G.M. acknowledges the Japanese Special Coordination Fund for Promotion of Science and Technology for accelerated innovation of the fostering system for female scientists and JSPS Wakatte B KAKENHI Grant Number 24710242. We also thank S. Oka and A. Tokumitsu for ESI-HRMS measurements and T. Hirose for amino acid analysis at Instrumental Analysis Division, Equipment Management Center, Creative Research Institution, Hokkaido University. MUC1 (glyco)peptides syntheses of this work was supported by Project for Developing Innovation Systems of the Japanese Ministry of Education, Culture, Science and Technology and JSPS KAKENHI Grant Number 25220206.

References

1. a) M. A. Hollingsworth, B. J. Swanson, Nat. Rev. Cancer 2004, 4, 45-60; b) D. W. Kufe, Nat. Rev. Cancer 2009, 9, 874-885.
2. a) J. M. Burchell, A. Mungul, J. Taylor-Papadimitriou, J. Mammary Gland Biol. Neoplasia 2001, 6, 355-364; b) G. Patsos, A. Corfield, "O-Glycosylation: structural diversity and functions" In The Sugar Code. Fundamentals of glycosciences (Ed.: H.-J. Gabius), Wiley-VCH, Weinheim, 2009, pp. 111-137.
3. a) H.-J. Gabius, Biochem. Soc. Trans. 2011, 39, 399-405; b) M. Amano, H. Eriksson, J. C. Manning, K. M. Detjen, S. André, S.-I. Nishimura, J. Lehtiö, H.-J. Gabius, FEBS J. 2012, 279, 4062-4080.
4. a) H.-J. Gabius, S. André, J. Jiménez-Barbero, A. Romero, D. Solís, Trends Biochem. Sci. 2011, 36, 298-313; b) K. Smetana, Jr., S. André, H. Kaltner, J. Kopitz, H.-J. Gabius, Exp. Opin. Ther. Targets 2013, 17, 379-392.
5. a) N. Suzuki, K. Yamamoto, S. Toyoshima, T. Osawa, T. Irimura, J. Immunol. 1996, 156, 128-135; b) S. J. van Vliet, E. Saeland, Y. van Kooyk, Trends Immunol. 2008, 29, 83-90.
6. S. J. van Vliet, S. I. Gringhuis, T. B. H. Geijtenbeek, Y van Kooyk, Nat. Immunol. 2006, 7, 1200-1208.

7. a) N. Higashi, K. Fujioka, K. Denda-Nagai, S. Hashimoto, S. Nagai, T. Sato, Y Fujita, A. Morikawa, M. Tsuiji, M. Miyata-Takeuchi, Y Sano, N. Suzuki, K. Yamamoto, K. Matsushima, T. Irimura, J. Biol. Chem. 2002, 277, 20686-20693; b) E. Saeland, S. J. van Vliet, M. Bäckström, V. C. M. van den Berg, T. B. H. G. Geijtenbeek, A. Meijer, Y. van Kooyk, Cancer Immunol. Immunother. 2007, 56, 1225-1236; c) C. Napoletano, A. Zizzari, H. Rughetti, T. Rahini, H. Irimura, H. Clausen, F. Wandall, F. Belleudi, L. I. Bellati, G. Pierelli, L. Frati, M. Nuti, Eur. J. Immunol. 2012, 42, 936-945.
8. D. Serruto, R. Rappuoli, FEBS Lett. 2006, 580, 2985-2992.
9. a) F. Marcelo, F. J. Cañada, J. Jiménez-Barbero, "The interaction of saccharides with antibodies. A 3D view by using NMR" in Anticarbohydrate antibodies: from molecular basis to clinical application (Eds.: P. Kosma, S. Müller-Leonnies), Springer, Wien, 2012, pp. 385-402; b) A. Arda, P. Blasco, D. Varón Silva, V. Schubert, S. André, M. Bruix, F. J. Cañada, H.-J. Gabius, C. Unverzagt, J. Jiménez-Barbero, J. Am. Chem. Soc. 2013, 135, 2667-2675; c) D. Solís, N.V. Bovin, A. P. Davis, J. Jiménez-Barbero, A. Romero, R. Roy, K. Smetana Jr., H.-J. Gabius, Biochim. Biophys. Acta 2014, dx.doi.org/10.1016/j.bbagen.2014.03.016.
10. a) M. C. del Carmen Fernandez-Alonso, D. Diaz, M. A. Berbis, F. Marcelo, F.J. Cañada, J. Jiménez-Barbero, Curr Protein Pept. Sci. 2012, 13, 816-830; b) S. Martín-Santamaría, H.-J. Gabius, J. Jiménez-Barbero, Pure Appl. Chem. 2012, 84, 49-64.
11. a) A. Borgert, J. Heimburg-Molinaro, X. Song, Y Lasanajak, J. Tongzhong, M. Liu, P. Thompson, G. Ragupathi, G. Barany, D. F Smith, R. D. Cummings, D. Live, ACS Chem. Biol. 2012, 7, 1031-1039; b) A. B. Yongye, L. Calle, A. Arda, J. Jiménez-Barbero, S. André, H.-J. Gabius, K. Martínez-Mayorga, M. Cudic, Biochemistry 2012, 51, 7278-7289.
12. a) F. Corzana, J. H. Busto, F. Marcelo, J. L. Asensio, S. Martín-Santamaría, M. Garcia de Luis, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, Chem. Eur. J. 2011, 17, 3105-3110; b) F Corzana, J. H. Busto, F. Marcelo, M. Garcia de Luis, J. L. Asensio, S. Martín-Santamaría, J. Jiménez-Barbero, A. Avenoza, J. M. Peregrina, Chem. Commun. 2011, 47, 5319-5321.
13. a) H. Möller, N. Serttas, H. Paulsen, J. M. Burchell, J. Taylor-Papadimitriou, B. Meyer, Eur J Biochem 2002, 269, 1444-1455; b) M. Sakakura, S. Oo-Puthinan, C. Moriyama, T. Kimura, J. Moriya, T. Irimura, I. Shimada, J. Biol. Chem. 2008, 283, 33665-33673.

14. M. Meier, M. D. Bider, V. N. Malashkevich, M. Spiess, P. Burkhard, J. Mol. Biol. 2000, 300, 857-865.

15. a) M. Mayer, B. Meyer, Angew. Chem. Int. Ed. 1999, 38, 1784-1788. Angew. Chem. 1999, 111, 1902-1906; b) M. Mayer, B. Meyer, J. Am. Chem. Soc. 2001, 123, 6108-6117; c) B. Meyer, T. Peters, Angew. Chem. Int. Ed. 2003, 42, 864-890. Angew. Chem. 2003, 115, 890-918.

16. J. Angulo, P.M. Enriquez-Navas, P.M. Nieto, Chem. Eur. J. 2010, 16, 7803-7812.

17. a) F. Marcelo, F. J. Cañada, S. André, C. Colombo, F. Doro, H.-J. Gabius, A. Bernardi, J. Jiménez-Barbero, Org. Biomol. Chem. 2012, 10, 5916-5923; b) L. Calle, V. Roldós, F. J. Cañada, M. L. Uhrig, A. J. Cagnoni, V. E. Manzano, O. Varela, J. Jiménez-Barbero, Chem. Eur. J. 2013, 19, 4262-4270; c) B. Fiege, C. Rademacher, J. Cartmell, P. I. Kitov, F. Parra, T. Peters, Angew. Chem. Int. Ed. 2012, 51, 928-932. Angew. Chem. 2012, 124, 952-956; d) P. Enríquez-Navas, M. Marradi, D. Padro, J. Angulo, S. Penadés, Chem. Eur. J. 2011, 17, 1547-1560.

18. P. Groves, K. E. Kövér, S. André, J. Bandorowicz-Pikula, G. Batta, M. Bruix, R. Buchet, A. Canales, F. J. Cañada, H.-J. Gabius, D. V. Laurents, J. R. Naranjo, M. Palczewska, S. Pikula, E. Rial, A. Strzelecka-Kiliszek, J. Jiménez-Barbero, Magn. Reson. Chem. 2007, 45, 745-748.

19. a) S. A. F. Jégouzo, A. Quintero-Martínez, X. Ouyang, A. dos Santos, M. E. Taylor, K. Drickamer, Glycobiology 2013, 23, 853-864; b) N. Mortezai, H. N. Behnken, A.-K. Kurze, P. Ludewig, F. Buck, B. Meyer, C. Wagener, Glycobiology 2013, 23, 844-852.

20. P. Dokurno, P. A. Bates, H. A. Band, L. M. D. Stewart, J. M. Lally, J. M. Burchell, J. Taylor-Papadimitriou, D. Snary, M. J. E. Sternberg, P. S. Freemont, J. Mol. Biol. 1998, 284, 713-728.

21. a) Y. S. Wang, D. Liu, D. F. Wyss, Magn. Reson. Chem. 2004, 42, 485-489; b) J.P. Ribeiro, S. André, F. J. Cañada, H.-J. Gabius, A. P. Butera, R. J. Alves, J. Jiménez-Barbero, ChemMedChem 2010, 5, 415-419.

22. S. J. van Vliet, E. van Liempt, E. Saeland, CA. Aarnoudse, B. Appelmelk, T. Irimura, T. B. Geijtenbeek, O. Blixt, R. Alvarez, I. van Die, Y van Kooyk, Int. Immunol. 2005, 17, 661-669.

23. A. R. Kolatkar, A. K. Leung, R. Isecke, R. Brossmer, K. Drickamer, W. I. Weis, J. Biol. Chem. 1998, 273, 19502-19508.

24. H. Sugawara, M. Kusunoki, G. Kurisu, T. Fujimoto, H. Aoyagi, T. Hatakeyama, J. Biol. Chem. 2004, 279, 45219-45225.

25. H. Feinberg, M. E. Taylor, W.I. Weis, J. Biol. Chem. 2007, 282, 17250-17258.

26. J. R. Walker, B. Nagar, N. M. Young, T. Hirama, J. M. Rini, Biochemistry 2004, 43, 3783-3792.

27. a) K. Drickamer, Nature 1992, 360, 183-186; b) J. E. Gready, A. N. Zelensky, "Routes in lectin evolution: case study on the C-type lectin-like domain" in The Sugar Code. Fundamentals of glycosciences (Ed.: H.-J. Gabius), Wiley-VCH, Weinheim, 2009, pp. 329-346.

28. T. Hatakeyama, T. Kamiya, M. Kusunoki, S. Nakamura-Tsuruta, J. Hirabayashi, S. Goda, H. Unno, J. Biol. Chem. 2011, 286, 10305-10315.

29. H. Feinberg, M. E. Taylor, N. Razi, R. McBride, Y. A. Knirel, S. A. Graham, K. Drickamer, W. I. Weis, J. Mol. Biol. 2011, 405, 1027-1039.

30. S. F Poget, G. B. Legge, M. R. Proctor, P.J. Butler, M. Bycroft, R. L. Williams, J. Mol. Biol. 1999, 290, 867-879.

31. D. Case, TA. Darden, TE. Cheatham, CL. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 12, University of California, San Francisco, 2012.

32. a) V. Jayalakshmi, N. R. Krishna, J. Magn. Reson. 2004, 168, 36-45; b) N. R. Krishna, V. Jayalakshmi, Top. Curr. Chem. 2007, 273, 15-54.

33. J. Angulo, I. Diaz, J.J. Reina, G. Tabarani, F. Fieschi, J. Rojo, P. M. Nieto, ChemBioChem 2008, 9, 2225-2227.

34. J. L. Asensio, A. Arda, F. J. Cañada, J. Jiménez-Barbero, Acc. Chem. Res. 2013, 46, 946-954.

35. I. A. Tayubi, R. Sethumadhavan, Int. J. PharmTech Res. Int. J. Pharm. Pharm. Sci. 2011, 3, 212-218.

36. a) T. K. Dam, TA. Gerken, CF. Brewer, Biochemistry 2009, 48, 3822-3827; b) P. V. Murphy, S. André, H.-J. Gabius, Molecules 2013, 18, 4026-4053.

37. C. L. Brooks, A. Schietinger, S. N. Borisova, P. Kufer, M. Okon, T. Hirama, C. R. Mackenzie, L.-X. Wang, H. Schreiber, S. V. Evans, Proc. Natl. Acad. Sci. USA 2010, 107, 10056-10061.

38. E. Lazoura, J. Lodding, W. Farrugia, P. A. Ramsland, J. Stevens, I. A. Wilson, G. A. Pietersz, V. Apostolopoulos, Immunology 2006, 119, 306-316.

39. W. C. Ng, S. Liong, M. D. Tate, T. Irimura, K. Denda-Nagai, A. G. Brooks, S. L. Londrigan, P. C. Reading, J. Virol. 2014, 88, 1659-1672.

40. a) G.-N. Wang, S. André, H.-J. Gabius, P. V. Murphy, Org. Biomol. Chem. 2012, 10, 6893-6907; b) J. Kopitz, Z. Fik, S. André, K. Smetana Jr, H.-J. Gabius, Mol. Pharm. 2013, 10, 2054-2061.

41. http://swissmodel.expasy.org/.

42. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould Jr., K. M. Merz, D. M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, J. Am. Chem. Soc. 1995, 117, 5179-5197.

43. a) G. M. Bradbrook, T. Gleichmann, S. J. Harrop, J. Habash, J. Raftery, J. Kalb, J. Yariv, I. H. Hillier, J. R. Helliwell, J. Chem. Soc. Faraday Trans. 1998, 94, 1603;
b) http://www.pharmacy.manchester.ac.uk/bryce/amber/.

44. K. N. Kirschner, A. B. Yongye, S. M. Tschampel, J. González-Outeiriño, C. R. Daniels, B. L. Foley, R. J. Woods, J. Comput. Chem. 2008, 29, 622-655.

45. D. J. Price, C. L. Brooks, J. Chem. Phys. 2004, 121, 10096-10103.

46. C. R. Cantor, P. R. Schimmel, Biophysical Chemistry. Part II. Techniques for the study of biological structure and function, W. H. Freeman, Oxford, 1980, pp. 503.