Molecular recognition of complex-type biantennary *N*-glycans by protein receptors: a 3D view by using NMR

Ana Ardá,¹ Pilar Blasco,¹ Daniel Varón Silva,³ Volker Schubert,³ Sabine André,⁴ Marta Bruix,² Francisco J. Cañada,¹ Hans-Joachim Gabius,⁴ Carlo Unverzagt,^{3*} Jesús Jiménez-Barbero^{1*}

¹ Chemical and Physical Biology, Centro de Investigaciones Biológicas, CSIC, Ramiro de Maeztu 9, 28040 Madrid, Spain.

² Instituto de Química-Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain.

³ Bioorganische Chemie, Gebäude NWI, Universität Bayreuth, 95440 Bayreuth, Germany.

⁴ Institute of Physiological Chemistry, Faculty of Veterinary Medicine, Ludwig-Maximilians-University Munich, Veterinärstr. 13, 80539 Munich, Germany.

KEYWORDS: agglutinin – glycobiomarker – glycoprotein – lectin - sialylation

ABSTRACT: The current surge of interest to define glycobiomarkers by applying lectins rekindles the interest to define the docking sites of lectins at the highest possible resolution. Of note, natural complex-type *N*-glycans can present more than one potential binding site, posing the question on the actual interaction when interpreting e. g. lectin array data. By strategically combining chemoenzymatic *N*-glycan synthesis with STD NMR spectroscopy and modeling we illustrate that epitope recognition can depend on the structural context (for the sugar and the lectin, here wheat germ agglutinin and a single hev-*ein* domain). Equally important, we monitor branch-end substitutions by this strategy and describe a 3D structure which accounts for the accommodation of the α 2,6-sialylated terminus of a complex *N*-glycan by a plant lectin. The given approach is thus instrumental to pinpoint sites for lectin molecular recognition by glycans of natural complexity.

Introduction

Molecular recognition is at the heart of essential biological events. The recognition of saccharides presented as part of natural scaffolds (protein, sphingolipids) by protein receptors (lectins) mediates key processes underlying many aspects of life as reflected by the term ‘sugar code’.¹ Matching the wide spectrum of glycan structures (glycome complexity²), lectins are found in all branches of the phylogenetic tree and proved to be versatile tools for glycan detection and structural characterization.³ There is particular interest in the identification of glycoprotein glycan structures as a result of altered glycoprotein expression occurring in diseases, hereby defining functional glycobiomarkers of diagnostic value.⁴ Recent technological advances have lead to the employment of lectin and glycan microarrays, which offer a macroscopic view of the lectin-glycan interactions, and provide a high-throughput approach to screen a wide range of glycan profiles and glycan binding specificity of the lectin.⁵

With the growing realization that the way a glycan epitope is presented affects its reactivity for lectins^{3b} it is essential to study this interaction with glycans equalling their natural size and rigorously define the contact sites.

NMR and X-ray crystallography provide key information at atomic resolution on the studied molecules, including

complexes. In the protein-carbohydrate interaction field, these techniques have mainly focused on acquiring information on those complexes formed by lectins and small/medium size oligosaccharide fragments. Probably, the chemical complexity and flexibility of the whole glycan structures occurring in Nature have often impaired the detailed analysis of the corresponding complexes. Due to remarkable progress in chemical synthesis of *N*-glycans, which have facilitated to relate structural aspects of complex-type *N*-glycans such as core substitutions to lectin affinity,⁶ and the powerful NMR-based techniques to monitor the recognition process⁷ these studies are possible.

Herein, we demonstrate that NMR spectroscopy, applied from the perspective of the ligand, can be efficiently used to obtain high-resolution information on these molecular recognition processes. As model system, the interaction of complex-type biantennary *N*-glycans with four plant lectins of different specificities, topologies and folding features was investigated.

On the side of the glycans, we include the biantennary complex-type nonasaccharide (with *N*-acetylglucosamine (GlcNAc) termini) and its α 2,6-sialylated undecasaccharide, which were isolated from egg yolk, as well as the α 2,6-sialyl-GlcNAc trisaccharide (Figure 1). These structures represent common motifs of *N*-glycans on glycopro-

teins, α 2,6-sialylation acting as salient signal for masking/docking.⁸ From the lectin side, two agglutinins with reactivity to the *N,N'*-diacetylchitobiose ((GlcNAc)₂) of the core were chosen, i. e. wheat germ agglutinin (WGA)⁹ and a WGA-derived single domain similar to hevein.¹⁰ Hevein, one of the smallest plant lectins with 43 amino acids, was the first lectin whose structure in complex with its sugar ligand has been solved in solution.¹¹ Its reactivity is represented by the so-called hevein domain, present in WGA. Of note, WGA is also reactive with *N*-acetylneuraminic acid.⁹ Seeds of *Maackia amurensis* are the source of two agglutinins (the leukoagglutinin MAL and the haemagglutinin MAA), which differ in their requirement for α 2,3-sialylation of a LacNAc terminus.^{9, 12} In contrast, the *Viscum album* agglutinin (viscumin, VAA) is known to react with LacNAc-terminated and α 2,6-sialylated *N*-glycans.¹³

We show that the mode of glycan presentation is crucial for recognition, and that the selectivity observed in assays with free sugars might change with subtle alterations in the *N*-glycan structure. Additionally, the importance of multivalency could also be deduced from our experiments.

Methods.

NMR.- The ¹H-NMR resonances of the ligands were completely assigned through standard TOCSY (60 ms mixing time), NOESY (300 and 500 ms mixing time) and HSQC experiments at 800 MHz (cryo) and 600 MHz spectrometers. Typical concentrations were 1 mM for the homonuclear and 2 mM for the heteronuclear experiments.

The samples for STD experiments as adapted to lectins¹⁴ were prepared in PBS buffer (MAA and WGA pH=5.7, and pH=7.3 for viscumin) using ligand/lectin ratios varying from 1:20 to 1:100. The applied temperatures varied between 283 K and 303 K. Molar ratio and temperature were optimized in each case. Representative experiments with significant STD responses are presented in the figures.

Data from 1D STD experiments were acquired at 600 and 500MHz using a Gaussian pulse (49 ms) cascade separated by 1 ms delays. STD-TOCSY experiments were acquired at 500MHz, using 256 increments and a isotropic mixing time of 60 ms. In all cases, the on-resonance frequency was set at the aromatic or aliphatic regions (6.9 ppm or 0.5 ppm, respectively), and a series of experiments were recorded, where the saturation time was varied between 200 ms and 2s. The off-resonance frequency was always set at 100 ppm. Under these conditions, the free ligand in solution showed residual STD intensities in the 1D STD spectra that were taking into account when analysing the STD spectra.

Molecular modelling.- The starting structures for the docking procedures were built by taking as starting points X-ray structures of the lectin complexed with small oligosaccharides pdb codes: 2UVO and 2CWG for WGA, 1DBN for *maackia amurensis* and 1PUU for viscumin. The structure of the *N*-glycan was manually docked in the binding site working with the most populated conformation found for the free state (according to a standard NOE/molecular modelling approach). The complex structure was then submitted to a short MD run, followed by energy minimization with a low gradient convergence threshold (0.02) in

5000 steps. In all cases, the OPL2005 force field¹⁵ was employed, as integrated in the Schroedinger MAESTRO suite of programmes.¹⁶

Lectins.- WGA and MAA were purchased from Sigma-Aldrich, VAA was obtained and controlled for purity as described,¹⁷ and WGA-B domain was also obtained as previously described.¹⁸

Ligands.- Compound **5** was purchased from Sigma Aldrich. Compounds **1**,¹⁹ **2** and **4** were obtained as previously described.²⁰ Compound **3** was prepared from its corresponding sialylated undecasaccharide-glycopeptide, that was obtained from egg yolk in analogy to the published procedures.²¹ This undecasaccharide-glycopeptide 5.7 mg (2 μ mol) was dissolved in 400 μ L of phosphate buffer (75 mM; pH 6) and a solution of 0.2 mg (0.6 U) of neuraminidase from *Clostridium perfringens* (Sigma Aldrich, EC: 3.2.1.18) in 50 μ L of phosphate buffer (75 mM, pH 6) was added. The mixture was incubated at ambient temperature for 3d (TLC: 2-propanol/ 1 M ammonium acetate 1.5:1), cleared by centrifugation and the supernatant was purified by gel filtration (Superdex 30 (1.6 x 60cm); flow rate 1mL/min; eluent 0.1 M NH₄HCO₃; detection 214 nm). The fractions eluting at 77 min were collected, lyophilized and desalted by gel filtration (Sephadex G25 (2.5 x 70 cm); flow rate 1 mL/min; eluent 5% ethanol in water; detection 214 nm). The fractions eluting at 201 min were lyophilized and yielded 4.05 mg of nonasaccharide-glycopeptide **3** (98.2 %)

Results.

Monovalent versus multivalent lectins: hevein domain and WGA.

The recognition features of chitoooligosaccharides by hevein domains have been well established through NMR.^{10, 11} More recently, we have demonstrated that the trisaccharide *N*-glycan core (**1**) is also recognised by hevein, with GlcNAc1 (linked to Asn) on subsite +2 (on top of W21), GlcNAc2 on subsite +1 (on top of W23) and β Man at subsite -1, leaving the Asn residue outside the binding site, without establishing interactions with the lectin.¹⁹ This structural model was used as template for analyzing recognition of complex-type *N*-glycans. However, when we titrated a 0.3mM solution of a single hevein domain (the B-domain of WGA¹⁸ with increasing amounts of the sialylated *N*-glycan **4** (figure SupplInfo), no changes at all were observed on the ¹H NMR resonances of the lectin. Similarly, attempts to dock the undecasaccharide **4** using the binding mode of the core trisaccharide-Asn glycopeptide (**1**) (Figure 2) were unsuccessful. In fact, branching at position 6 of the central β Man residue causes a steric clash of the sugar extension with different protein residues, independent of the presence of either *gg* or *gt* rotamers at the C5-C6 linkage. By contrast, WGA is known to react with *N*-glycans.⁹

WGA is a dimeric lectin where each subunit consists of an assembly of 4 hevein domains.²² Therefore, WGA is composed of 8 hevein domains. It has been described to be specific for terminal Neu5Ac and GlcNAc moieties, by inhibition studies,⁹ crystallography²² and NMR spectroscopy.²³ The structure of complexes, also with a sialoglycopeptide,^{22c} reveals that the primary binding site is constituted

by three conserved aromatic residues from one hevein domain, but further polar residues (S114 and E115) from the neighbouring hevein domain contribute to the stabilization of the complex, providing structural evidence for an intra-protein cooperation of lectin domains.

Strikingly, in contrast with the observations for the single hevein domain, the STD experiments performed for the sialylated *N*-glycan **4** in the presence of the multidomain lectin showed STD signals, evidencing an interaction. Analysis of the STD spectra (Figures 3 and 4) clearly showed that recognition does not involve the terminal Neu5Ac residue. Only a weak effect is observed in the methyl group of the acetamide, while the ring hydrogens do not show up in the STD spectrum. In fact, the residues in closer contact with the protein are at the opposite end of the molecule, namely GlcNAc1 and Asn. Thus, the binding mode is different to that occurring between a single hevein domain (Figure 2A) and the trisaccharide core (**1**), in which the Asn moiety was far from the lectin (positive NOEs in the NOESY spectrum of the complex).¹⁹ For hevein itself, GlcNAc1 and GlcNAc2 are located at subsites +2 and +1 respectively. However, for WGA, the docking procedure based on the experimental observations showed that the binding mode is shifted by one subsite: the Asn residues now occupies subsite +2 (on top of W21), GlcNAc1 is on subsite +1 (on top of W23) and GlcNAc2 sits on subsite -1. In this situation, the β Man residue is shifted one subsite away from the center of binding to the protein and therefore the typical branching at O6 now does not cause steric clashes with the protein (Figure 5). Obviously, one question arises: why a single hevein domain does not use this binding mode for recognizing the large *N*-glycan **4**. Detailed comparison points to the presence of the additional polar residues in the neighbouring hevein domain in WGA, i. e. S114 and E115, to appear as key factor to stabilize the complex. They form hydrogen bonds with the *N*-glycan (Figure 5), those which are naturally absent in the single hevein domain. Probably, this binding mode is energetically not favorable, if these interactions do not contribute to the binding enthalpy, as is the case of just a single hevein domain.

Analogous STD experiments were carried out with the non-sialylated *N*-glycan **2** (Figure 6). The results were very similar to those obtained with **4**: Asn and GlcNAc1 display the strongest STD intensities, with lower intensities for the branches. This result supports the data discussed above, which showed that Neu5Ac is not involved in stable interactions with the protein.

As mentioned above, WGA is often included among the sialic acid-binding lectins. However, our experimental STD data have shown that although **4** is indeed recognised by WGA, the terminal sialic acid of the undecasaccharide does not establish interactions with the lectin. Instead, the interactions occur via the GlcNAc1 and Asn moieties at the stem region of the glycan.

To clarify this issue an additional STD experiment with 6'-sialylLacNAc (**5**) was acquired. The STD spectrum clearly shows that (Figure 7) only protons from Neu5Ac show STD intensity, especially those from the acetyl group, in agreement with the X-ray crystallographic data.²² The other STD signals correspond to H5, H6, H7 and H4 of Neu5Ac. When presented in the context of an *N*-glycan, the

sialic acid is therefore not a primary contact site, what it is for the branch-end trisaccharide. Explicitly, in the **4**/WGA complex, only the acetyl group of Neu5Ac showed weak STD, and protons H5, H6 and H7 did not show STD (Figure 8).

This result underscores the inherent value of the given approach to team up binding-site identification by STD NMR spectroscopy with *N*-glycan synthesis. When more than one possible contact sites is presented the given strategy will decide on the preferences in the context of the *N*-glycan.

Since the recognition of **2** and **4** by WGA mainly involves GlcNAc1 and the Asn residue, it remained to be clarified whether there is an influence of a peptidic moiety, as occurring in glycoproteins. Thus, an STD experiment was carried out under the same experimental conditions for a mixture of glycopeptide **3** and WGA. The spectrum showed STD intensities similar to those obtained for **2** (Figure 9). In particular, the protons of the additional peptide residues did not show up in the spectrum, indicating exclusive recognition of **3** by WGA through the GlcNAc-Asn region.

Maackia amurensis agglutinin was used as the mixture of the two lectin species, MAL and MAH.¹² Both isolectins are known to be able to bind to α 2,3-sialylgalactose, although with different preferences: the leucoagglutinin (MAL) is known to preferentially bind Neu5Ac α 2,3Gal β 1,4 β GlcNAc moieties present in *N*-glycans,²⁴ while the hemagglutinin (MAH) has higher affinity for the disialylated tetrasaccharide Neu5Ac α 2,3Gal β 1,3[Neu5Ac(2,6)] α GalNAc present in *O*-glycans.^{12, 25} The binding sites of both isolectins are very similar, the only difference being the amino acids Y221 and E222 at MAL, that are both Ala in MAH. The X-Ray crystal structure of MAL complexed with 3'-sialyllactose has been published.²⁶

1D STD and 2D STD-TOCSY experiments were acquired for **2** and **4**, using samples with a protein/ligand 1:20 (50 μ M/1mM) ratio. The corresponding spectra are shown in Figure 10. The comparison of the spectra on the left (sialylated *N*-glycan **4**) and on the right (non-sialylated **2**) enables to assess that MAA indeed does not recognise the Neu5Ac α (2,6)Gal-containing molecule (**4**), a negative control for specificity, while it does efficiently bind to non-sialylated **2**. From the analysis of this STD spectrum, it was deduced that the residues interacting with the lectin are the terminal galactose and GlcNAc residues (LacNAc unit), while in this case the GlcNAc1 and Asn stem is not in contact with the protein. This observation is in agreement with the previously reported specificity data and with the published crystallographic structure (pdb code 1DBN). Indeed, the binding site of MAL is unusually narrow for a lectin. Both the terminal galactose and GlcNAc residues are sandwiched between Y136 and Y221, leaving position Gal O6 deep inside the binding site, and establishing hydrogen bonding with D137, clearly precluding the extension through Gal O-6, as would occur in a 2,6 sialylated *N*-glycan. The absence of sialic acid allows the terminal LacNAc unit to perfectly fit inside the binding pocket, establishing the combination of CH- π stacking interactions and hydrogen bonds with Y136, Y221, and Y131, D87,

D137 and E224 respectively,²⁶ a typical feature for lectin binding.^{3b} Further interactions are established with GlcNAc2 and ManB through Y136 and Y221 (Figure 11).

Viscumin. This lectin is a toxic AB-type protein, which in addition to the A chain with rRNA *N*-glycosidase activity, contains a B chain with lectin activity to galactosides. The B-chain harbors two carbohydrate-binding sites, characterised by W38 and Y249. In solution, the lectin activity of VAA can be attributed primarily to the Tyr site.²⁷ The recognition mode of lactose by VAA has been characterized by X-ray crystallography, chemical mapping and modeling.²⁸ It has also been reported that VAA strongly binds Neu5Ac α 2-6Gal β 1-4GlcNAc on gangliosides and glycoproteins, while gangliosides with terminal galactoses are poorly recognised,²⁹ but a structural rationalization has not yet been given.

Figure 12 shows the STD spectra, along with their corresponding reference spectra, for VAA in the presence of the sialylated *N*-glycan **4** (on top) and of the non-sialylated analogue **2** (below). Both compounds are bound by VAA. Since strictly the same experimental conditions were kept, and the STD intensities are significantly stronger for **4** than for **2**, the results suggest respective grading of interaction for VAA. Furthermore, the distinct protons of the Neu5Ac residue (H3ax, H3eq and Ac) clearly revealed that this residue is in close contact with VAA. This fact was confirmed by the STD-TOCSY spectrum (Figure 13). Thus, in this case, the Neu5Ac moiety is clearly involved in the recognition process and its absence weakens the glycan-ligand interaction.

3D structures of the corresponding complexes were generated by using molecular modeling techniques, taking the available experimental X-ray coordinates for the VAA/lactose complex as starting geometry. In the resulting 3D structure, the galactose-binding site is formed by amino acids D335, N256 and Q238 that establish hydrogen bonding with the Gal residue and by Y249, which establishes stacking interaction with the α -face of the Gal moiety, explicitly with H3, H4, H5 (Figure 14A). Inspection of this experimental geometry detects space for sugar extension from Gal O6. This location is surrounded by polar amino acids that may efficiently interact with sialic acid residues. They are S200, R245 and Q238. Docking of **4** at this locus, using the lactose-binding model as a template, provided a stable pose, which was further optimized through an energy minimization and MD protocol. Of note, further interactions can then take place between Neu5Ac and the above mentioned polar residues (Figure 14B): R245 establishes a bidentate hydrogen bond with the glycerol chain of the sialic acid residue, S200 is hydrogen bonded to the carboxylic acid, and Q238 establishes a cooperative hydrogen bond with OH7 of Neu5Ac and OH3 of the branching GlcNAc. Evidently, these interactions appear to underlie the preferential reactivity reported above.

Discussion

The availability of synthetic *N*-glycans facilitates to define the docking sites of lectins. Shown herein, the nature of the target epitope strictly depends on the nature of the tested lectin, which can select distinct fragments of the

complex-type saccharide chain with exquisite specificity, depending on the relative presentation of the different residues and the architecture of the binding site.

We have underscored the importance of multivalency for achieving glycan binding in the case of a comparison between a single hevein domain and WGA. Where the monomeric unit does not recognise the large glycan, the branching at the β Man residue being the key structural element for precluding the interaction the multi-domain lectin binds the terminal reducing end of either **2** or **4**, namely the GlcNAc-Asn epitope, owing to the possibility of achieving additional inter-site stabilizing interactions from one of the neighboring hevein domains. In this context, it has to be emphasized that WGA does not react with **2** or **4** through the (GlcNAc)₂ moiety, at it would be expected. The chemical nature of this branched glycan excludes the possibility of efficient interactions of the diacetylchitobiose stem with the two consecutive aromatic residues, typical of hevein domains, the recognized epitope is shifted one position to the GlcNAc-Asn fragment. Clearly, the sialyl residue of **4** is not involved in the binding process, although it has been demonstrated that WGA also can accommodate sialic acid residues. Indeed, using α 2,6-sialyllactose, we clearly demonstrate that this small trisaccharide is recognized by WGA. Thus, although presented at the spatially readily accessible non-reducing end, binding occurs exclusively via the glycan-Asn stem region, a result of conspicuous importance for interpretation of data using WGA in array platforms. The same holds true for MAA, for which reactivity to the LacNAc terminus is detectable.

As conclusion, the wide variety with the glycome and of lectin architectures is a challenge for analyzing protein-lectin interactions. Advances in macroscopic analysis of binding using array-based and other technologies³⁰ can strategically be complemented by the given approach, to rigorously define the contact site in naturally complex glycans. Work with VAA yielded insights into how α 2,6-sialylation is accommodated, here identifying resonances from protons of the sialic acid and a network of interactions. Thus, the combination of *N*-glycan synthesis, STD NMR spectroscopy and modeling proves its potency to detect and characterize epitope interaction.

ASSOCIATED CONTENT

¹H NMR spectra for the titration of WGA-B domain with increasing amounts of the sialylated *N*-glycan **4**. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Author Contributions

The manuscript was written through contributions of all authors. / All authors have given approval to the final version of the manuscript.

Funding Sources

Financial support from the Ministry of Economy and Competitiveness of Spain (Grants CTQ2011-22514, CTQ2009-08536,

and CTQ2012-30820) and from the EU (Projects GlycoHit and GlycoPharm, and COST action CM1102 is acknowledged.

ABBREVIATIONS

WGA, wheat germ agglutinin; VAA, *Viscum album* agglutinin.

REFERENCES

- (1) (a) Haltiwanger, R.S.; Lowe, R. B. *Annu. Rev. Biochem.* **2004**, 73, 491-537; (b) Gabius, H.-J. (ed) *The Sugar Code. Fundamentals of glycosciences*. Wiley-VCH, Weinheim, 2009.
- (2) Cummings, R. D. *Mol. BioSyst.* **2009**, 5, 1087-1104.
- (3) (a) Kiessling, L. L.; Splain, R. A. *Annu. Rev. Biochem.* **2010**, 79, 619-653; (b) Gabius H.-J.; André, S.; Jiménez-Barbero, J.; Romero, A.; Solís, D. *Trends Biochem. Sci.* **2011**, 36, 298-313.
- (4) (a) Arnold, J. N.; Saldova, R.; Hamid, U. M.; Rudd, P. M. *Proteomics* **2008**, 8, 3284-3293; (b) Drake, P. M.; Cho, W.; Li, B.; Prakobphol, A.; Johansen, E.; Anderson, N. L.; Regnier, F. E.; Gibson, B. W.; Fisher, S. J. *Clin. Chem.* **2010**, 56, 223-236; (c) Reis, C. A.; Osorio, H.; Silva, L.; Gomes, C.; David, L. J. *Clin. Pathol.* **2010**, 63, 322-329; (d) Gabius, H.-J. *Biochem. Soc. Trans.* **2011**, 39, 557-565.
- (5) (a) Etzebarria, J.; Calvo, J.; Martin-Lomas, M.; Reichardt, N. C. *ACS Chem. Biol.* **2012**, 10.1021/cb300262x; (b) Song X.; Yu H.; Chen X.; Lasanajak Y.; Tappert M. M.; Air G. M.; Tiwari V. K.; Cao H.; Chokhawala H. A.; Zheng H.; Cummings R. D.; Smith D. F. *J Biol Chem.* **2011**, 286, 31610-31622 (c) Song, X. Z.; Xia, B. Y.; Stowell, S. R.; Lasanajak, Y.; Smith, D. F.; Cummings, R. D. *Chem. Biol.* **2009**, 16, 36-47; (d) Blixt, O.; Head, S.; Mondala, T.; Scanlan, C.; Huflejt, M. E.; Alvarez, R.; Bryan, M. C.; Fazio, F.; Calarese, D.; Stevens, J.; Razi, N.; Stevens, D. J.; Skehel, J. J.; van Die, I.; Burton, D. R.; Wilson, I. A.; Cummings, R.; Bovin, N.; Wong, C.-H.; Paulson, J. C. *Proc. Natl. Acad. Sci. USA* **2004**, 101, 17033-17038.
- (6) (a) André, S.; Kožár, T.; Schuberth, R.; Unverzagt, C.; Kojima, S.; Gabius, H.-J. *Biochemistry* **2007**, 46, 6984-6995; (b) André, S.; Kožár, T.; Kojima, S.; Unverzagt, C.; Gabius, H.-J. *Biol. Chem.* **2009**, 390, 557-565.
- (7) Roldós, V.; Cañada, F. J.; Jiménez-Barbero, J. *ChemBioChem* **2011**, 12, 990-1005.
- (8) (a) Zuber, C.; Roth, J. In: *The Sugar Code. Fundamentals of glycosciences* (Gabius, H.-J., ed), Wiley-VCH, Weinheim, pp. 87-110. (b) Schauer, R. *Curr. Opin. Struct. Biol.* **2009**, 19, 507-514.
- (9) Goldstein, I. J.; Poretz, R. D. In: *The Lectins. Properties, Functions and Applications in Biology and Medicine* (Liener, I. E.; Sharon, N.; Goldstein, I. J., eds), Academic Press, San Diego, 1986, pp. 33-247.
- (10) (a) Jiménez-Barbero, J.; Cañada, F. J.; Asensio, J. L.; Aboitiz, N.; Vidal, P.; Canales, A.; Groves, P.; Gabius, H.-J.; Siebert, H.-C.; Derek, H. *Adv. Carbohydr. Chem. Biochem.* **2006**, 60, 303-354.
- (11) Asensio, J. L.; Cañada, F. J.; Bruix, M.; Rodríguez-Romero, A.; Jiménez-Barbero, J. *Eur. J. Biochem.* **1995**, 230, 621-633.
- (12) (a) Kawaguchi, T.; Matsumoto, I.; Osawa, T. *J. Biol. Chem.* **1974**, 249, 2786-2792; (b) Knibbs, R. N.; Goldstein, I. J.; Ratcliffe, R. M.; Shibuya, N. *J. Biol. Chem.* **1991**, 266, 83-88; (c) Geisler, C.; Jarvis, D.L. *Glycobiology* **2011**, 21, 988-993.
- (13) (a) André, S.; Unverzagt, C.; Kojima, S.; Dong, X.; Fink, C.; Kayser, K.; Gabius, H.-J. *Bioconj. Chem.* **1997**, 8, 845-855; (b) Galanina, O. E.; Kaltner, H.; Khraltsova, L. S.; Bovin, N. V.; Gabius, H.-J. *J. Mol. Recognit.* **1997**, 10, 139-147; (c) Unverzagt, C.; André, S.; Seifert, J.; Kojima, S.; Fink, C.; Srikrishna, G.; Freeze, H.; Kayser, K.; Gabius, H.-J. *J. Med. Chem.* **2002**, 45, 478-491.
- (14) (a) Mayer M.; Meyer, B. *Angew. Chem. Int. Ed.* **1999**, 38, 1784-1788; (b) Jiménez-Barbero, J.; Dragoni, E.; Venturi, C.; Nannucci, F.; Ardá, A.; Fontanella, M.; André, S.; Cañada, F. J.; Gabius, H.-J.; Nativi, C. *Chem. Eur. J.* **2009**, 15, 10423-10431; (c) Martín-Santamaría, S.; André, S.; Buzamet, E.; Caraballo, R.; Fernández-Cureses, G.; Morando, M.; Ribeiro, J. P.; Ramírez-Gualito, K., de Pascual-Teresa, B., Cañada, F. J., Menéndez, M.; Ramström, O.; Jiménez-Barbero, J.; Solís, D.; Gabius, H.-J. *Org. Biomol. Chem.* **2011**, 9, 5445-5455.
- (15) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. J. *J. Phys. Chem. B* **2001**, 105, 6474.
- (16) *Maestro, A powerful, all-purpose molecular modeling environment*, version 8.5, Schroedinger, LLC, New York, NY, 2008.
- (17) Gabius, H.-J.; Darro, F.; Rammelink, M.; André, S.; Kopitz, J.; Danguy, A.; Gabius, S.; Salmon, I.; Kiss, R. *Cancer Invest.* **2001**, 19, 114-126.
- (18) Espinosa, J. F.; Asensio, J. L.; García, J. L.; Laynez, J.; Bruix, M.; Wright, C.; Siebert, H.-C.; Gabius, H.-J.; Cañada, F. J.; Jiménez-Barbero, J. *Eur. J. Biochem.* **2000**, 267, 3965-3978.
- (19) Hernández-Gay, J. J.; Ardá, A.; Eller, S.; Mezzato, S.; Leeftang, B. R.; Unverzagt, C.; Cañada, F. J.; Jiménez-Barbero, J. *Chem. Eur. J.* **2010**, 16, 10715-10726.
- (20) Kajihara, Y.; Suzuki, Y.; Yamamoto, N.; Sasaki, K.; Sakakibara, T.; Juneja, L. R. *Chem. Eur. J.* **2004**, 10, 971-985.
- (21) Seko, A.; Koketsu, M.; Nishizono, M.; Enoki, Y.; Ibrahim, H. R.; Juneja, L. R.; Kim, M.; Yamamoto, T. *Biochim. Biophys. Acta Gen. Subj.* **1997**, 1335, 23-32.
- (22) (a) Wright, C. S. *J. Mol. Biol.* **1984**, 178, 91-104; (b) Wright, C. S. *J. Mol. Biol.* **1990**, 215, 635-651; (c) Wright, C. S. *J. Biol. Chem.* **1992**, 267, 14345-14352.
- (23) (a) Kronis, K. A.; Carver, J. P. *Biochemistry* **1982**, 21, 3050-3057; (b) Kronis, K. A.; Carver, J. P. *Biochemistry* **1985**, 24, 826-833.
- (24) Wang, W. C.; Cummings, R. D. *J. Biol. Chem.* **1988**, 263, 4576-4585.
- (25) Konami, Y.; Yamamoto, K.; Osawa, T.; Irimura, T. *FEBS Lett.* **1994**, 342, 334-338.
- (26) Imbert, A.; Gautier, C.; Lescar, J.; Perez, S.; Wyns, L.; Loris, R. *J. Biol. Chem.* **2000**, 275, 17541-17548.
- (27) Jiménez, M.; André, S.; Siebert, H.-C.; Gabius, H.-J.; Solís, D. *Glycobiology* **2006**, 16, 926-937.
- (28) (a) Mikeska, R.; Wacker, R.; Arni, R.; Singh, T. P.; Mikhailov, A.; Gabdoulkhakov, A.; Voelter, W.; Betzel, C. *Acta Crystallogr. Sect. F: Struct. Biol. Cryst. Commun.* **2005**, 61, 17-25; (b) Jiménez M.; André, S.; Barillari, C.; Romero, A.; Rognan, D.; Gabius, H.-J.; Solís, D. *FEBS Lett.* **2008**, 582, 2309-2312.
- (29) Müthing, J.; Meisen, I.; Bulau, P.; Langer, M.; Witthohn, K.; Lentzen, H.; Neumann, U.; Peter-Katalinic, J. *Biochemistry* **2004**, 43, 2996-3007.
- (30) (a) Solís, D.; Romero, A.; Menéndez, M.; Jiménez-Barbero, J. In: *The Sugar Code. Fundamentals of glycosciences* (Gabius, H.-J., ed), Wiley-VCH, Weinheim, pp. 233-245; (b) Tateno, H.; Kuno, A.; Hirabayashi, J. In: *The Sugar Code. Fundamentals of glycosciences* (Gabius, H.-J., ed), Wiley-VCH, Weinheim, pp. 247-259.

FIGURES

FIGURE 1. Glycan structures used in this study.

FIGURE 2. A: The complex of hevein with the trisaccharide core (**3**) as deduced from NMR.¹⁹ B: Attempted docking of the extended *N*-glycan structure **4** obtained by superimposition of the sialylated *N*-glycan in the trisaccharide core template (see panel A). The extension at position O6 of the central Man moiety in the core causes steric clashes with the protein. The gg orientation is shown at Man C5-C6. The gt rotamer is sterically also forbidden. Subsites are indicated on top, and the key amino acids, i. e. Trp21 and Trp23, are annotated.

FIGURE 3. STD (below) and off-resonance spectrum (on top) of the sample containing WGA and the sialylated *N*-glycan **4**. The region of the methyl protons of the Ac groups is enlarged on the right.

FIGURE 4. Left: TOCSY of the ligand **4**. Right: STD-TOCSY of the mixture of WGA + sialylated *N*-glycan **4**. The sialyl residue does not contribute to the binding event. In contrast, the Asn residue provided clear STD signals indicating its direct involvement in the recognition process.

FIGURE 5. The deduced complex of the sialylated *N*-glycan (**4**, in orange) in the primary binding site of WGA, as deduced from the docking, minimization and MD process. Amino acids Y66, Y 64 and S62 (equivalents to W21, W23 and S19 in hevein) from one hevein domain are represented in grey CPK, while S114 and E115 from the neighbouring hevein domain are represented in balls and sticks, in purple. The sialyl residues are exposed to the solvent and not directly involved in WGA binding.

FIGURE 6. Left: TOCSY spectrum of non-sialylated *N*-glycan (**2**). Right: STD-TOCSY of WGA/non-sialylated *N*-glycan, **2**. As for **4**, the Asn and GlcNAc1 residues provided clear STD signals indicating their direct involvement in the recognition process.

FIGURE 7. STD spectrum (below) of 6'-sialyllactose (**5**) with WGA. On top, off-resonance reference spectrum. The hydrogens of the sialyl residue clearly appear in the STD, indicating that they are directly involved in WGA binding.

FIGURE 8. HQC(edited) spectrum of compound 4, and superimposed STD (in red) and off resonance spectrum (in black) of 4 / WGA. The hydrogens of the sialyl residue do not appear in the STD, indicating that they are not directly involved in WGA binding.

FIGURE 9. STD (below) and off-resonance spectrum (on top) of glycopeptide **3** and WGA. As for **2** and **4**, the Asn and GlcNAc1 residues provided clear STD signals indicating their direct involvement in the recognition process.

FIGURE 10. Left: STD and STD-TOCSY spectra corresponding to MAA/sialylated *N*-glycan (**4**). No STD is detected. Right: spectra corresponding to MAA/non-sialylated *N*-glycan (**2**). The terminal *N*-acetylglucosamine residues provided clear STD signals indicating their direct involvement in the recognition process.

FIGURE 11. Perspective of the modeled structure (using a docking + MD protocol) for the complex formed between MAL and **2**, using the deposited X-ray structure (pdb code 1DBN) as template structure. The *N*-acetylglucosamine at the 1→6 branch has been docked at the binding site.

FIGURE 12. STD spectra of the mixtures of VAA with **4** (panel A) and **2** (panel B). The methyl groups regions are enlarged at the right hand side and the resonances are annotated for each case. The STD spectra are shown below, while the corresponding off-resonance spectra are on top. In all cases, VAA binds both molecules (see discussion).

FIGURE 13. A. TOCSY spectrum of sialylated *N*-glycan **4**. B. STD-TOCSY of VAA with **4**. The sialyl residue is in contact to the lectin surface.

FIGURE 14. A. X-ray structure (pdb code 1PUU) of VAA complexed with lactose. The amino acids implicated in interactions with the ligand are highlighted. B. Docking pose of the sialylated *N*-glycan **4** on the Tyr-site in subdomain 2 γ . The new amino acids providing further interactions (hydrogen bonds depicted in green) with the Neu5Ac residues are highlighted.