Breaking the Limits in Analyzing Carbohydrate Recognition by NMR Spectroscopy: Resolving Branch-Selective Interaction of a Tetra-Antennary *N*-Glycan with Lectins

Angeles Canales, Irene Boos, Lukas Perkams, Lukas Karst, Thomas Luber, Theodoros Karagiannis, Gemma Domínguez, F. Javier Cañada, Javier Pérez-Castells, Daniel Häussinger, Carlo Unverzagt,* and Jesus Jiménez-Barbero*

Dedicated to Professor Horst Kessler on the occasion of his 75th birthday

Abstract: *The biological recognition of complex-type N-glycans is part of many key physiological and pathological events. Despite their importance, the structural characterization of these events remains unsolved. The inherent flexibility of N-glycans hampers crystallization and the chemical equivalence of individual branches precludes their NMR characterization. By using a chemoenzymatically synthesized tetra-antennary N-glycan conjugated to a lanthanide binding tag, the NMR signals under paramagnetic conditions discriminated all four N-acetyl lactosamine antennae with unprecedented resolution. The NMR data revealed the conformation of the N-glycan and permitted for the first time the direct identification of individual branches involved in the recognition by two N-acetyllactosamine-binding lectins, Datura stramonium seed lectin (DSL) and Ricinus Communis agglutinin (RCA120).*

N-glycans are ubiquitous in nature and functionalize glycoproteins.[1] Protein glycosylation is required for proper biological and biophysical function and often, alterations in glycosylation are related to diseases.[2]

Complex glycosylation patterns containing multi-antennary *N*-glycans are typically found in mature glycoproteins. However, the structural characterization of these glycans is rather challenging. Usually, NMR spectroscopy and X-ray diffraction techniques fail to provide specific answers on the structure and molecular recognition features owing to the intrinsic attributes of the glycan. The properties of the glycosidic bond and especially the presence of 1–6 linkages endow a large flexibility to the molecule. This feature precludes crystallization or hampers the detection of enough electron density for most of the glycan part in the X-ray analysis of glycoproteins. Moreover, the standard use of the corresponding fitting programs to deduce three-dimensional structures frequently give rise to incorrect structures of the glycans.[3] Thus, any advance in this area is of high value.

As a promising approach, carbohydrates conjugated to lanthanide-binding tags have shown high potential toward this aim.[4,5] In its vicinity, a complexed paramagnetic ion induces significant chemical shift changes of the NMR signals of the glycan as a result of dipolar interactions involving the unpaired electron of the metal. These pseudocontact shifts (PCS) depend on the distance between each proton and the metal (proportional to $1/r^3$).[6] This methodology has first been applied to the study of small oligosaccharides (di-, tri-, and tetrasaccharides),[7–11] then to *N*-glycans. The conformational properties of complex-type bi-antennary and high-mannose-type *N*-glycans were elucidated and could be resolved in each case down to the level of individual branches.[4,5] Proceeding from this experimental basis, we herein extend this concept to the level of high-degree branching, and show that it is possible to experimentally characterize the conformational behavior and recognition properties of a galactosylated complex-type

[*] Prof. J. Jiménez-Barbero
CIC bioGUNE, Bizkaia Science and Technology Park
Building 801A, 48160 Derio (Spain)
and
IKERBASQUE, Basque Foundation for Science
48009 Bilbao (Spain)
E-mail: jjbarbero@cicbiogune.es

Dr. A. Canales
Dep. Química Orgánica I, Fac. C. C. Químicas, Univ. Complutense de Madrid, Avd. Complutense s/n, 28040 Madrid (Spain)

Dr. I. Boos, L. Perkams, L. Karst, T. Luber, T. Karagiannis, Prof. C. Unverzagt
Bioorganische Chemie, Gebäude NW1
Universität Bayreuth
95440 Bayreuth (Germany)
E-mail: carlo.unverzagt@uni-bayreuth.de

Dr. G. Domínguez, Prof. J. Pérez-Castells
Dep. Química, Fac. Farmacia, Univ. CEU-San Pablo
Madrid (Spain)

Prof. F. J. Cañada
Dep. Biología Físico Química, CIB-CSIC
C/ Ramiro de Maeztu 9, 28040 Madrid (Spain)

Dr. D. Häussinger
Department of Chemistry, University of Basel
Spitalstrasse 51, 4056 Basel (Switzerland)

tetra-antennary *N*-glycan, which is not possible by standard NMR methods owing to chemical shift degeneracy of the glycan. We have then tested the hypothesis of branch selectivity for two *N*-acetyllactosamine-binding lectins, *Datura stramonium seed lectin* (DSL) and *Ricinus Communis agglutinin* (RCA 120).

The tetra-antennary conjugate **4** (Scheme 1) required for the NMR studies was obtained by chemical and enzymatic synthesis based on recent advances in modular *N*-glycan assembly.[12] The protected *N*-glycan **1** was synthesized from modular building blocks followed by selective deprotection to **2** bearing an azido group at the reducing end (Scheme 1). This azide was subjected to enzymatic galactosylation.[13] reduced and coupled to the ligand **3**[4] followed by deprotection of the ethylesters. The tagged conjugate **4** was purified by HPLC and complexed in D$_2$O with diamagnetic LaCl$_3$ (**4a**) or paramagnetic DyCl$_3$ (**4b**) followed by buffer addition prior to NMR measurements.

For the conformational study, as initial step, the ^{1}H–^{13}C HSQC NMR spectrum of **4a** (diamagnetic conditions) was acquired. The NMR spectrum displays all carbohydrate signals in a narrow region in the ^{1}H dimension as expected for a pseudosymmetric complex-type *N*-glycan. The galactose moieties (Gal) at the termini of the four arms A–D cannot be distinguished, since the resonances of the four Gal moieties are isochronous in the ^{1}H as well as the ^{13}C dimension, resulting in complete signal equivalence (see Figure 1). The signals of the GlcNAc residues in the branches also severely overlap.

For the paramagnetic compound **4b**, the HSQC spectrum displayed a strong increase in overall ^{1}H signal dispersion to 5.3 ppm (from 4.9 ppm (H1αMan4) to −0.36 ppm (H5GlcNAc1)). In contrast, for diamagnetic **4a** the corre-

sponding dispersion for the ring protons of the carbohydrates is only 1.8 ppm. In the case of **4b**, the signals of the four terminal Gal and penultimate GlcNAc residues are clearly separated. Figure 1 presents an expansion of the anomeric region of the HSQC spectra (Figure 1a, lower panels). The H1 signals of the four Gal units of **4a** are merged to one signal (blue spectrum, Figure 1) whereas for paramagnetic **4b** they are shifted to a different extent. This results in four individual signals separated by 0.1–0.3 ppm in the ^{1}H dimension (red spectrum, Figure 1). The anomeric signals of GlcNAc moieties in the antennae (GlcNAc A–D) appear in two groups under diamagnetic conditions; one for the two 1–2 linked GlcNAc units (A and C arm) and the other for the 1–4 and 1–6 linked moieties (B and D arm). Under paramagnetic conditions, four well-separated anomeric signals for GlcNAc A–D (separated ca. 0.1–0.3 ppm in the ^{1}H dimension) are found. The unprecedented overall signal dispersion for the tetra-antennary *N*-glycan conjugate **4b** (Figure 1) allowed the assignment of 58 pseudocontact shifts (PCS) for ^{1}H signals. PCSs were measured as the difference in chemical shift of the NMR signals under diamagnetic (**4a**) and paramagnetic conditions (**4b**). The conformational behavior of the *N*-glycan was subsequently deduced by comparison of the experimental PCS with those estimated from back-calculating the expected PCS values for the individual conformational energy minima or their ensembles, following the method described in the experimental section (Supporting Information).

For the Manα(1–6)Man linkage of the *N*-glycan core, an excellent correlation was obtained for the *gauche–gauche* (*gg*) conformers 1 and 2 (Figure 2c, left panel) with a quality factor of 0.08. In contrast, the *gauche–trans* (*gt*) conformers 3 and 4 (Figure 2c, right panel) produced a poor fit (QF 0.2).

The combination of conformers did not improve the results. These results point out that the Manα(1–6)Man linkage in the tetra-antennary *N*-glycan (ω1 in Figure 2a) is less flexible than in the corresponding bi-antennary compound (only A and C arms), where both *gg* and *gt* conformers are simultaneously populated in solution (65% and 35%, respectively).[4] In contrast, for the D arm, which is attached via a β(1–6) linkage (ω2 in Figure 2a), the best correlation between experimental and calculated PCSs was obtained by combining *gg* (40%) and *gt* (60%) conformers (PCSs are given in Supporting Information). Indeed, in this particular case, *J* values for H5-H6a (3 Hz) and H5–H6b (6.9 Hz) could be measured, in agreement with the existence of a conformational equilibrium. Thus, the β(1–6) GlcNAc linkage (ω2) of branch D is more flexible than the α(1–6) linkage (ω1) orienting the α1,6 arm. The predominant *gt* rotamer of the Manα(1-6)Man linkage has also been found for tri-antennary and penta-antennary *N*-glycans with a branched α(1–6) arm.[14,15]

The measurement of residual dipolar couplings (RDCs) induced by the lanthanide were also attempted to complement the PCS values.[16]
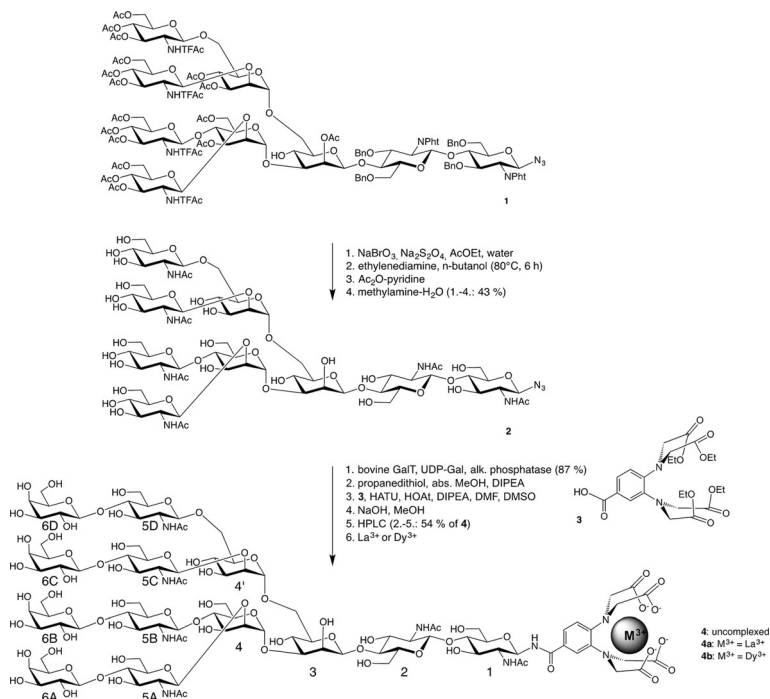
Scheme 1. Synthesis of the tagged *N*-glycan **4** and the complexes **4a,b**.
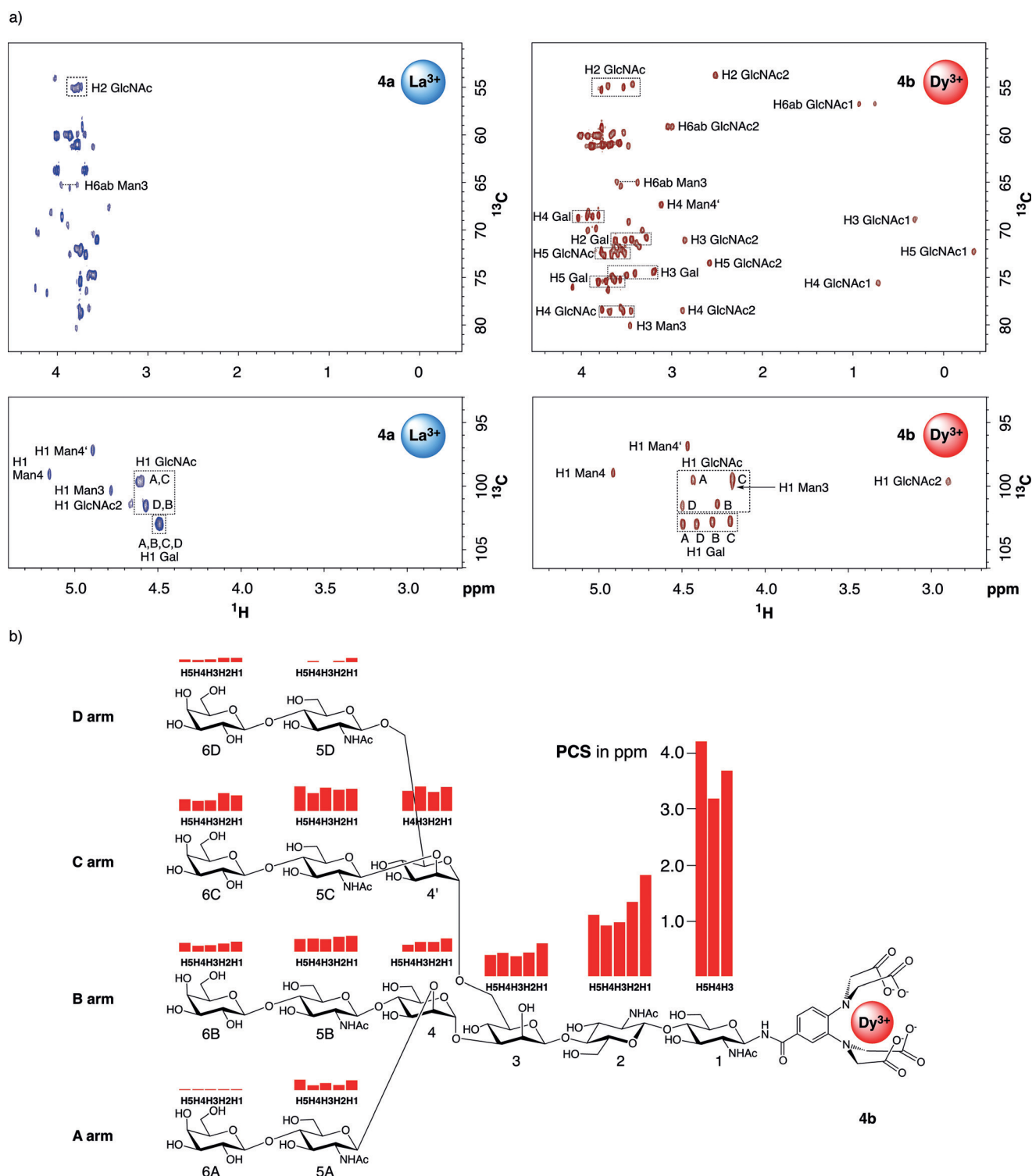
Figure 1. a) Expansions of the aliphatic (upper panels) and anomeric regions (lower panels) of the ^{1}H–^{13}C HSQC spectra of the tetra-antennary conjugate **4** loaded with lanthanum (compound **4a**, diamagnetic reference in blue) and loaded with dysprosium (compound **4b**, paramagnetic sample in red). In the spectra of **4b** the signals for Gal and GlcNAc marked with boxes indicate complete resolution of all four branches. Under diamagnetic conditions (**4a** spectrum) the chemical shifts are isochronous. b) Graphical visualization of the pseudocontact shifts, PCS (to scale), throughout the paramagnetic conjugate **4b**.

However, the obtained RDCs were too small to warrant their proper use in the conformational studies.

After achieving the characterization of the geometry and shape of *N*-glycan **4b**, the capacity to monitor the existence of

arm selectivity in recognition events was explored using two model lectins. *Datura stramonium* seed lectin (DSL) was selected as the first candidate for binding studies since this lectin recognizes *N*-glycans with multiple *N*-acetyllactos-

amine units with high affinity. In addition, tri- and tetra-antennary *N*-glycans containing the C–D arm pentasaccharide motif showed highest binding to immobilized DSL.[17,18] ¹H and HSQC spectra of **4b** were measured in the presence of DSL and the changes in signal linewidth were monitored. Clear line broadening of several NMR signals was detected upon lectin addition. Moreover, the magnitude of the observed NMR effect was different for the four individual antennae (Figure 3). The detailed analysis of the line broadening effects allowed to assign the major lectin interactions to the Gal and GlcNAc residues at the A, C, and D arms. In fact, the signals showing the highest decrease in intensity belong to the sugar moieties located at the D arm. In contrast, the B arm is less affected, strongly suggesting that it is less involved in protein interactions. This result is in agreement with a previous binding study of DLS that showed that *N*-glycan fragments containing only the A and B arms displayed reduced binding compared to those containing only C and D arms.[19]

Ricinus communis agglutinin (RCA120) was chosen as second model lectin. According to the available X-ray structure, RCA120 has an exposed and accessible sugar binding site, specific for terminal Gal monosaccharides.[20]

Indeed, with respect to DSL, smaller line broadening of the NMR signals of **4b** after addition of RCA120 was detected in the ¹H and HSQC NMR spectra. The more affected signals belong to the Gal residues at the termini of the arms indicating that these units are the main interacting points with the protein. In contrast, the NMR signals of the neighboring GlcNAc moieties are affected to a lesser extent.

Given the further signal dispersion of the ¹H NMR spectrum of **4b** in the presence of the paramagnetic metal, STD experiments were also carried out to complement the epitope mapping provided by the HSQC approach (Supporting Information). It was hypothesized that the growth of STD should not be precluded since the paramagnetic metal is relatively far from the ends of the antennae. Fittingly, STD signals were detected for **4b** in the presence of RCA120. Higher STD values were found for Gal H4, whereas small STD effects were measured for the neighboring GlcNAc signals. These data highlight that only the terminal Gal units are involved in recognition, as expected for the sugar specificity of this lectin.[20] In analogy to the HSQC data, the STD percentages were similar for all the branches of **4b** pointing out that there is only a marginal branch selectivity for RCA120. The geometry of the binding pocket of RCA120 permits that any the four arms of **4b** can independently interact with no selectivity.

In contrast, in the presence of DSL, **4b** showed major STD signals not only for the Gal moieties, but also for the GlcNAc units, and very significantly for the *N*-acetyl group of the GlcNAc located at the D arm. Fittingly, these STD results agree with the HSQC data, where the NMR signals of the residues at the D arm displayed the largest broadening effects.

Thus, the employed PCS methodology has allowed the different branch selectivity of these two lectins to be easily deduced, revealing their interaction profiles within their individual branches, even at the residue level. The addition of STD results, obtained from lectin complexes of paramagnetic **4b**, further complement the data analysis and provide a fairly detailed description of the recognition process. The chemical synthesis of suitably labelled *N*-glycan-conjugates is the prerequisite for the proposed NMR spectroscopy approach, which opens new avenues for advancing in glycosciences. The method provides breakthroughs in two areas: the characterization of the conformation of highly complex multi-antennary *N*-glycans and the elucidation of their recognition features with unprecedented resolution. By combining the flexibility of synthetic chemistry with the robustness of this NMR-based method, further advances in how the recognition of glycans take place will be facilitated.
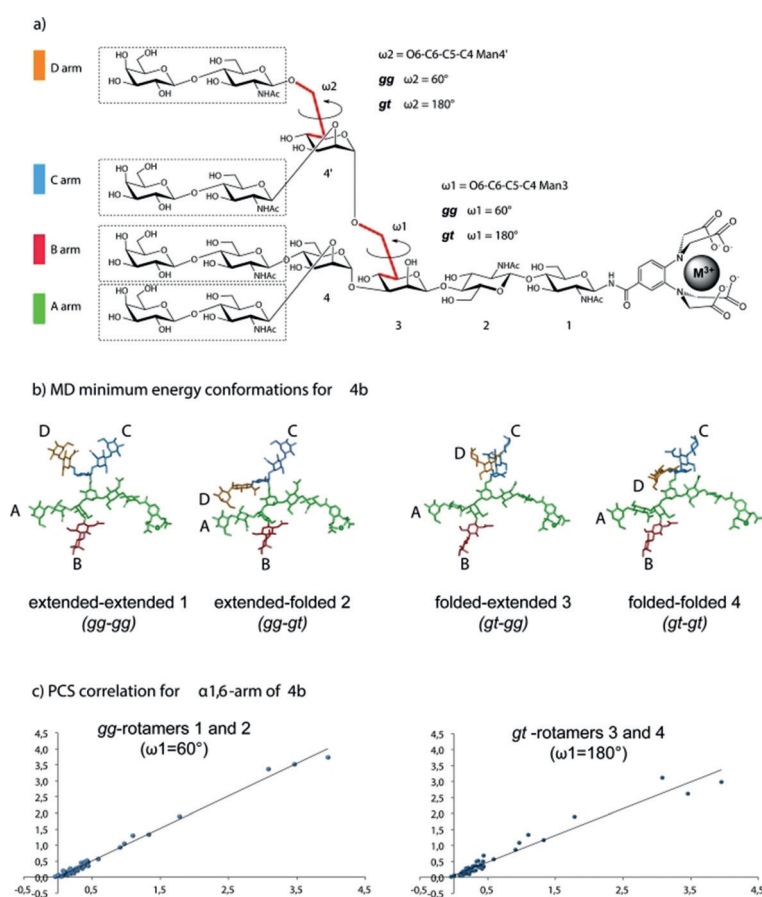

Figure 2. a) The flexibility of two 1,6 linkages leads to rotamers with preferred ω1 and ω2 values (angle definition is shown). b) Minimum-energy conformations of the tetra-antennary *N*-glycan **4b** obtained by MD calculations. Conformer nomenclature: (*gg-gt*) implies ω1: *gauche–gauche*, and ω2: *gauche–trans*. c) Correlation between experimental and computed PCSs for the protons of the A, B, and C arms calculated for the *gg* (left panel) and *gt* (right panel) conformations of the α(1–6) linkage (ω1).

Conflict of interest

The authors declare no conflict of interest.

[1] _The Sugar Code. Fundamentals of glycosciences_ (Ed.: H.-J. Gabius), Wiley-VCH, Weinheim, **2009**.

[2] a) Y. Miura, T. Endo, _Biochim. Biophys. Acta Gen. Subj._ **2016**, _1860_, 1608–1614; b) N. Taniguchi, Y. Kizuka, _Adv. Cancer Res._ **2015**, _126_, 11–51.

[3] J. Agirre, J. Iglesias-Fernandez, C. Rovira, G. J. Davies, K. S. Wilson, K. D. Cowtan, _Nat. Strut. Mol. Biol._ **2015**, _22_, 833–834.

[4] A. Canales, A. Mallagaray, J. Perez-Castells, I. Boos, C. Unverzagt, S. André, H. J. Gabius, F. J. Cañada, J. Jiménez-Barbero, _Angew. Chem. Int. Ed._ **2013**, _52_, 13789–13793; _Angew. Chem._ **2013**, _125_, 14034–14038.

[5] T. Yamaguchi, Y. Sakae, Y. Zhang, S. Yamamoto, Y. Okamoto, K. Kato, _Angew. Chem. Int. Ed._ **2014**, _53_, 10941–10944; _Angew. Chem._ **2014**, _126_, 11121–11124.

[6] G. Otting, _J. Biomol. NMR_ **2008**, _42_, 1–9.

[7] A. Canales, A. Mallagaray, M. A. Berbis, A. Navarro-Vazquez, G. Domínguez, F. J. Cañada, S. Andre, H. J. Gabius, J. Pérez-Castells, J. Jiménez-Barbero, _J. Am. Chem. Soc._ **2014**, _136_, 8011–8017.

[8] M. Erdélyi, E. d'Auvergne, A. Navarro-Vázquez, A. Leonov, C. Griesinger, _Chem. Eur. J._ **2011**, _17_, 9368–9376.

[9] S. Yamamoto, T. Yamaguchi, M. Erdelyi, C. Griesinger, K. Kato, _Chem. Eur. J._ **2011**, _17_, 9280–9282.

[10] S. Yamamoto, Y. Zhang, T. Yamaguchi, T. Kameda, K. Kato, _Chem. Commun._ **2012**, _48_, 4752–4754.

[11] Y. Zhang, S. Yamamoto, T. Yamaguchi, K. Kato, _Molecules_ **2012**, _17_, 6658–6671.

[12] a) M. Mönnich, S. Eller, T. Karagiannis, L. Perkams, T. Luber, D. Ott, M. Niemietz, J. Hoffman, J. Walcher, L. Berger, M. Pischl, M. Weishaupt, C. Wirkner, R. G. Lichtenstein, C. Unverzagt, _Angew. Chem. Int. Ed._ **2016**, _55_, 10487–10492; _Angew. Chem._ **2016**, _128_, 10643–10648; b) M. Niemietz, L. Perkams, J. Hoffman, S. Eller, C. Unverzagt, _Chem. Commun._ **2011**, _47_, 10485–10487.

[13] C. Unverzagt, H. Kunz, J. C. Paulson, _J. Am. Chem. Soc._ **1990**, _112_, 9308–9309.

[14] T. Taguchi, K. Kitajima, Y. Muto, S. Yokoyama, S. Inoue, Y. Inoue, _Eur. J. Biochem._ **1995**, _228_, 822–829.

[15] T. Taguchi, Y. Muto, K. Kitajima, S. Yokoyama, S. Inoue, Y. Inoue, _Glycobiology_ **1997**, _7_, 31–36.

[16] M. Erdélyi, E. d'Auvergne, A. Navarro-Vázquez, A. Leonov, C. Griesinger, _Chem. Eur. J._ **2011**, _17_, 9368–9376.

[17] J. F. Crowley, I. J. Goldstein, J. Arnarp, J. Lönngren, _Arch. Biochem. Biophys._ **1984**, _231_, 524–533.

[18] K. Yamashita, K. Totani, T. Ohkura, S. Takasaki, I. J. Goldstein, A. Kobata, _J. Biol. Chem._ **1987**, _262_, 1602–1607.

[19] J. U. Baenziger, D. Fiete, _J. Biol. Chem._ **1979**, _254_, 9705–9799.

[20] A. G. Gabdoulkhakov, Y. Savochkina, N. Konareva, R. Krauspenhaar, S. Stoeva, S. V. Nikonov, W. Voelter, C. Betzel, A. M. Mikhailov, **2004**, PDB: 1RZO.
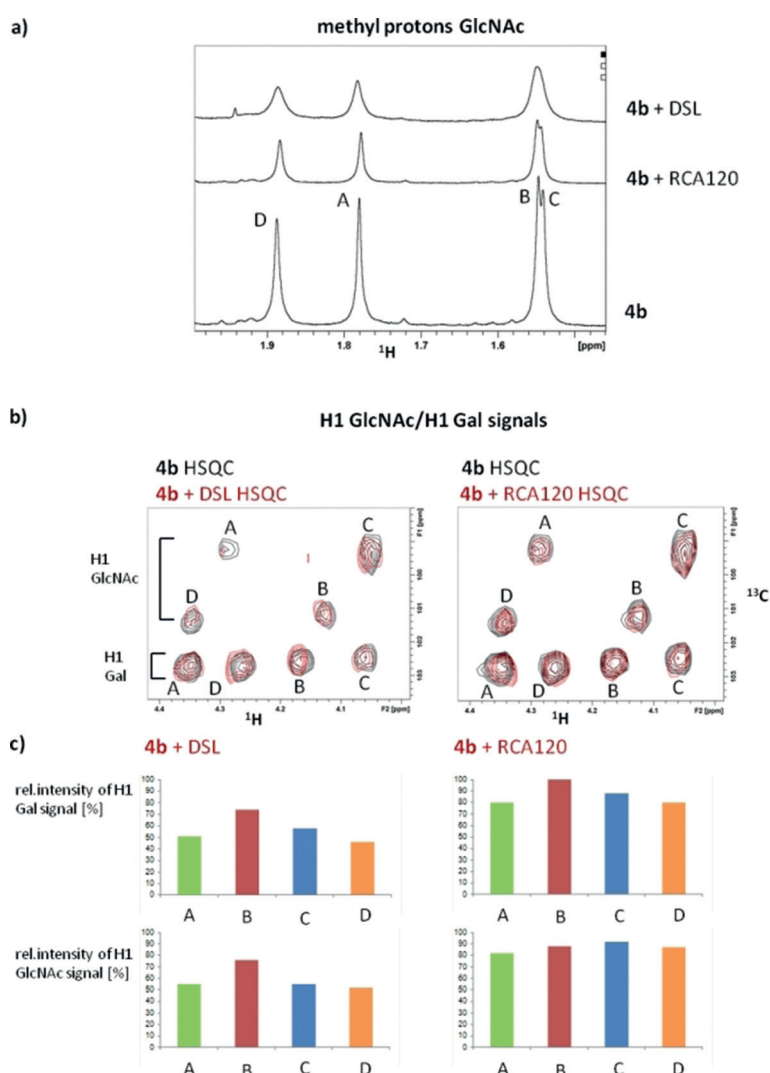
Figure 3. Lectin binding experiments. a) The intensities of the methyl signals of the GlcNAc units of **4b** in the absence and in the presence of lectin. b) H1 Gal and H1 GlcNAc region from an ^1H–^{13}C HSQC of the tetra-antennary conjugate **4b** in the absence (black) and in the presence (red) of DSL and RCA120. c) Intensities of the HSQC signals of H1 Gal and H1 GlcNAc after lectin addition relative to the spectrum without protein.