JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

^{13}C-Labeled Heparan Sulfate Analogue as a Tool To Study Protein/Heparan Sulfate Interactions by NMR Spectroscopy: Application to the CXCL12α Chemokine

Cédric Laguri,[*,†,‡,§] Nicolas Sapay,[∥] Jean-Pierre Simorre,[†,‡,§] Bernhard Brutscher,[†,‡,§] Anne Imberty,[∥] Pierre Gans,[†,‡,§] and Hugues Lortat-Jacob[†,‡,§]

[†]CEA, Institut de Biologie Structurale Jean-Pierre Ebel, F-38027 Grenoble Cedex 1, France

[‡]CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, F-38027 Grenoble Cedex 1, France

[§]Université Joseph Fourier − Grenoble 1, Institut de Biologie Structurale Jean-Pierre Ebel, F-38027 Grenoble Cedex 1, France

[∥]Centre de Recherche sur les Macromolécules Végétales − CNRS, 601 rue de la chimie, BP 53, 38041 Grenoble Cedex 9, France

Supporting Information

ABSTRACT: Heparan sulfate (HS), a polysaccharide of the glycosaminoglycan family characterized by a unique level of complexity, has emerged as a key regulator of many fundamental biological processes. Although it has become clear that this class of molecules exert their functions by interacting with proteins, the exact modes of interaction still remain largely unknown. Here we report the engineering of a ^{13}C-labeled HS-like oligosaccharide with a defined oligosaccharidic sequence that was used to investigate the structural determinants involved in protein/HS recognition by multidimensional NMR spectroscopy. Using the chemokine CXCL12α as a model system, we obtained experimental NMR data on both the oligosaccharide and the chemokine that was used to obtain a structural model of a protein/HS complex. This new approach provides a foundation for further investigations of protein/HS interactions and should find wide application.

Heparan sulfate (HS), one of the most abundant molecules on the cell surface, is a key regulator of many essential processes, including cell growth and development, cell adhesion, migration, immune response, and inflammation as well as pathological conditions such as cancer and infectious diseases. This polysaccharide exerts its biological functions by interacting with a large number of proteins (e.g., cytokines, growth factors, viral envelopes) to modify their conformation, stability, and reactivity and provide scaffold structures to promote molecular encounters.[1,2] Consistent with its functional diversity, HS is structurally complex. It consists of a disaccharide repeat unit composed of a glucuronic acid (GlcA) and an *N*-acetylglucosamine (GlcNAc), which can be enzymatically modified during biosynthesis. These modifications include N-deacetylation/N-sulfation of GlcNAc to GlcNS, C5 epimerization of GlcA to iduronic acid (IdoA), and a variable number of sulfations at position 2 of GlcA/IdoA or positions 6 and 3 of GlcN residues.[3,4] These modifications, which are essential for protein recognition,[5] occur in restricted domains along the chain and give rise to specific chemical structures and spatial arrangements that show a high degree of sequence variability. It has thus been thought that information for protein recognition resides within specific HS epitopes that are characterized by precise *N*- and *O*-sulfate group distributions. However, efforts to correlate the structure with binding activity have been hampered by the extreme complexity of this class of molecules; therefore, high-resolution structural studies of HS/protein complexes remain very limited.[6]

In this report, we describe the first production of ^{13}C-labeled HS with a defined sulfation sequence, which, in combination with an ^{15}N-labeled protein, yields a powerful experimental tool for investigating the structural determinants involved in protein/HS binding by multidimensional NMR.

Escherichia coli K5 bacteria,[7,8] which produce a capsular polysaccharide composed of GlcA−GlcNAc repeats identical to unsulfated HS (heparosan), were grown in culture medium containing ^{13}C-labeled glucose and ^{15}N-labeled NH_4Cl as the sole sources of carbon and nitrogen, respectively.[9,10] Uniformly ^{13}C/^{15}N-labeled polysaccharide was isolated, and the GlcNAc residues were chemically N-deacetylated, N-resulfated, and then 6-O-sulfated essentially as previously described.[11] The polysaccharide was depolymerized with heparinase II, and the resulting mixture was fractionated by gel-filtration chromatography to produce size-defined oligosaccharides from di- to dodecasaccharides [see the Supporting Information (SI)]. The octasaccharide fraction, which is of the minimum size necessary to encompass the HS binding site of the protein used in this study (see below), was selected and further separated according to charge using strong anion exchange HPLC (Figure 1a). NMR analysis of these materials showed that among the four major well-resolved species, one was homogeneously N- and 6-O-sulfated on the GlcN residues while the GlcA units remained unmodified (Figure 1b). This species, termed $dp8_{NS,6S}$ ("dp" stands for degree of polymerization), was retained for interaction studies. $dp8_{NS,6S}$ has the chemical structure: $GlcN_{NS6S}$−GlcA−$GlcN_{NS6S}$−GlcA−$GlcN_{NS6S}$−GlcA−$GlcN_{NS6S}$−ΔHexA (Figure 1c), where ΔHexA possesses a double bond between C4 and C5 as a result of enzymatic cleavage. ^{13}C−^{1}H resonances of $dp8_{NS,6S}$ were assigned by conventional heteronuclear NMR experiments (Figure SI-1 in the SI). Only the sugar resonances of GlcN at position 1 (GlcN-1) and ΔHexA at position 8 (ΔHexA-8)

Received: February 25, 2011
Published: June 02, 2011

ACS Publications © 2011 American Chemical Society

dx.doi.org/10.1021/ja201753e | *J. Am. Chem. Soc.* 2011, 133, 9642–9645

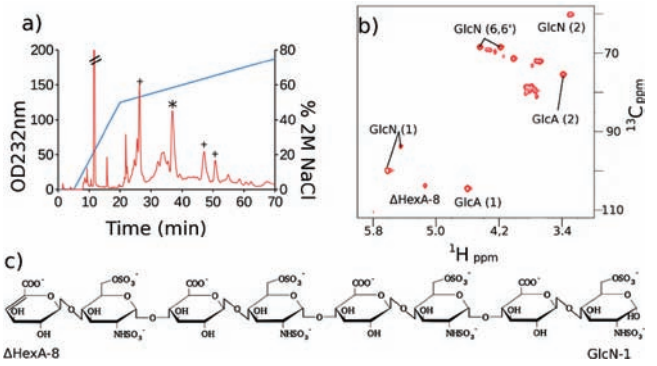

Figure 1. Purification and NMR characterization of $^{13}C/^{15}N$-labeled N-sulfated, 6-O-sulfated octasaccharide. (a) Strong anion exchange chromatogram of modified heparosan-derived octasaccharides. $dp8_{NS,6S}$ is the only homogeneously N- and 6-O-sulfated octasaccharide (*), and three other species are inhomogeneously sulfated (+). (b) $^{13}C-^{1}H$ correlation spectrum of $dp8_{NS,6S}$ octasaccharide showing complete N- and 6-O-sulfation of GlcN sugars with characteristic ^{13}C signals at 60 and 69 ppm.

were well-resolved in the NMR spectra. GlcN-3, -5, and -7 on one side and, conversely, GlcA-2, -4 and, -6 have highly overlapping NMR signals, suggesting that the central residues have identical chemical environments.

The ^{1}H and ^{13}C chemical shift assignments for $dp8_{NS,6S}$ provided an opportunity to study the mode of interaction of an HS analogue with proteins using NMR chemical shift perturbation methods. Here we chose CXCL12α as an example of an HS-binding protein. CXCL12α belongs to the chemokine family, a group of 50 proteins that display important chemotactic activity and for which binding to HS is key to ensure their correct positioning within tissues and maintain concentration gradients along which cells can migrate directionally.[12,13] ^{15}N-CXCL12α, prepared as described previously,[14] was used to titrate a solution of ^{13}C-labeled $dp8_{NS,6S}$, during which the binding was followed by $^{1}H-^{15}N$ SOFAST-HMQC NMR spectroscopy.[15]

When CXCL12α binds to $dp8_{NS,6S}$, its NMR spectrum becomes more complex than that of the free form. The resonances of R^{12}, K^{24}, A^{40}, and N^{46}, which correspond to the previously identified CXCL12α HS binding site,[16] showed significant chemical shift changes upon interaction with $dp8_{NS,6S}$. These data indicate an equilibrium between the free and oligosaccharide-bound proteins in fast exchange on the NMR time scale. Furthermore, for 20 other CXCL12α residues that are mostly located outside of the HS binding site, additional peaks appeared in the course of the titration (e.g., L^{55}) (Figure SI-2b). These additional peaks and the free-form resonances were in slow exchange on the NMR time scale, as revealed by longitudinal-exchange NMR spectroscopy (see Experimental Procedures in the SI). This observation can be ascribed to a heparin-induced dimerization process of the CXCL12 chemokine, as described previously.[17] This interpretation was further supported by the apparent increase in the overall rotational correlation time of the protein in the complex with respect to the free form, as estimated from ^{15}N spin relaxation measurements.

Thus far, by looking at the chemokine NMR spectrum, we have established that $dp8_{NS,6S}$ interacts in the HS binding region of CXCL12α[16] and at the same time promotes dimerization of the protein, as previously described for natural heparin-derived oligosaccharides.

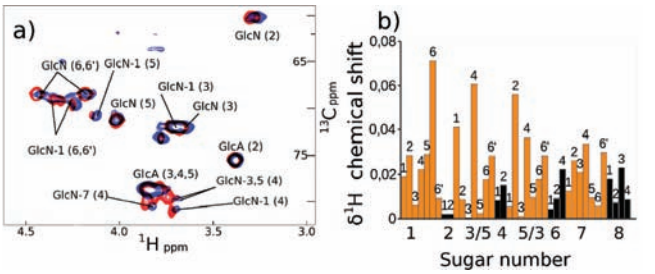

Figure 2. (a) $^{13}C-^{1}H$ correlation spectrum of free $dp8_{NS,6S}$ (red) and $dp8_{NS,6S}$ in the presence of a 1.45 molar excess of CXCL12α (blue). (b) ^{1}H chemical shift variations of $dp8_{NS,6S}$ upon CXCL12α addition, with data for GlcN and GlcA colored orange and black, respectively; the assignments of sugars 3 and 5 are ambiguous.

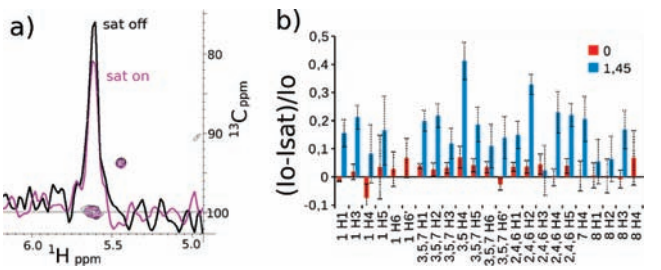

Figure 3. STD experiments. (a) One-dimensional ^{1}H trace for the GlcN-3, -5, and -7 H1−C1 correlation peak at a protein/HS molar ratio of 1.45 with (purple) or without (black) protein saturation. (b) Saturation of $dp8_{NS,6S}$ ^{1}H without protein (red) and at a CXCL12α/$dp8_{NS,6S}$ molar ratio of 1.45:1 (blue).

The availability of a ^{13}C-labeled glycan also enabled the study of the same interaction from the point of view of the polysaccharide. We were able to define the sugars participating in the interaction by titrating protein to the labeled HS analogue. Increasing chemical shift variations up to a CXCL12α/$dp8_{NS,6S}$ ratio of 1.45 were observed in the $^{1}H-^{13}C$ correlation spectra (Figure 2), accompanied by significant line broadening. An apparent binding affinity constant of 50 μM for the complex formation was derived from the change in the peak volume of the GlcN H1 signal during the ligand titration, considering the protein to be a dimer in the complex. The $^{13}C-^{1}H$ groups in the vicinity of sulfation positions H6 and H2 of the GlcN residues and H4 in the vicinity of the glycosidic linkages exhibited the most significant chemical shift perturbations. This suggests that all of the N-sulfated, 6-O-sulfated GlcN residues collectively contribute to the interaction (Figure 2b).

In absence of intermolecular nuclear Overhauser effect (NOE) data and to obtain additional information on the structure of the complex in fast exchange with the free molecules, we performed saturation transfer difference (STD) experiments.[18] In an STD experiment, the selective saturation of some protein protons is transferred to the bound ligand via dipolar proton interactions. As a result, a decrease in signal intensity in the NMR spectrum of the ligand is observed for all protons that are in close vicinity to the binding interface. The ^{13}C labeling of the octasaccharide allows the detection of such intensity changes in a 2D $^{1}H-^{13}C$ correlation spectrum, yielding site-specific information. Figure 3 shows the STD results obtained for the free $dp8_{NS,6S}$ octasaccharide and for a protein/ligand mixture at a molar ratio of 1.45. We observed a strong intensity decrease for most of

dx.doi.org/10.1021/ja201753e |J. Am. Chem. Soc. 2011, 133, 9642–9645

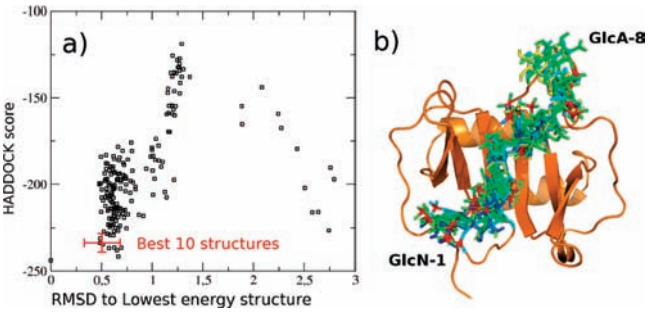

Figure 4. Results of NMR-restrained CXCL12α/dp8$_{NS,6S}$ docking study. (a) Plot of the Haddock score vs the rmsd with respect to the lowest-energy structure for the 200 water-refined complexes. (b) Superimposition of the 10 structures with the best Haddock scores (only the CXCL12α from the best structure is shown).

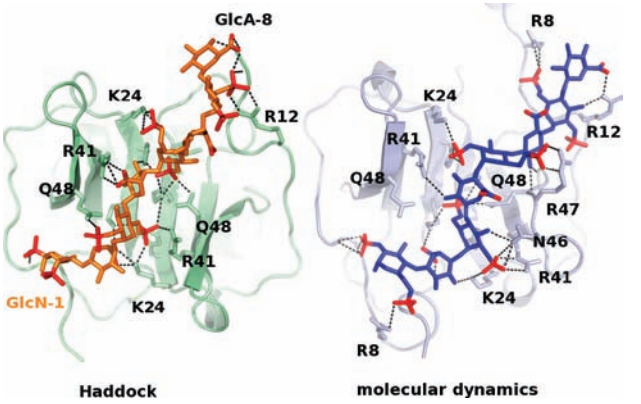

Figure 5. Comparison of the best Haddock structure and an MD structure extracted from simulation 1. The rmsd between the two octasaccharides was 2.4 Å, taking into account heavy atoms of the glycosidic ring.

the oligosaccharide signals. Both GlcN and GlcA sugars were efficiently saturated, particularly H4 of the central GlcN-3 and -5 and H2 of GlcA-2, -4, and -6. These data suggest close contacts with the protein side chains, particularly at the center of the oligosaccharide.

We then used the Haddock program[19] to derive a model of the CXCL12α/dp8$_{NS,6S}$ complex based on our NMR data. The Haddock protocol makes use of ambiguous distance restraints derived from NMR chemical shift perturbation data or any other available NMR and biochemical information on the binding site of the two partners to drive the docking process. To date, the Haddock protocol has been successfully applied to modeling of protein−protein and protein−DNA complexes.[20,21] In the Haddock program, two sets of residues from the CXCL12α/dp8$_{NS,6S}$ titrations had to be defined to set up the ambiguous interaction restraints (AIRs). AIRs are established from a set of active residues of one molecule to both active and passive residues of the second molecule. For the protein, active residues are those showing large chemical shift changes in fast exchange with the ligand (CXCL12α[12,20,24,25,45,48]), whereas passive residues are neighbors to the first set and are solvent-accessible (CXCL12α[8,13,,27,29,39,41,43,46,47]). In the case of dp8$_{NS,6S}$, GlcN residues showing significant chemical shift changes constituted the active set, and GlcA residues (which are also efficiently saturated in STD experiments) represented the passive set. Initial input coordinates for the docking protocol were the dimeric structure of CXCL12α (PDB entry 2NWG)[22] and a dp8$_{NS,6S}$ structure extracted from a molecular dynamics (MD) trajectory of the octasaccharide alone in explicit solvent (see Experimental Procedures). Docking and refinement in water with the Haddock protocol produced mainly one large ensemble of structures very close in both Haddock score and root-mean-square deviation (rmsd) to the structure with the lowest Haddock score (Figure 4a).

The 10 best structures (Figure 4b) presented low AIR restraint violations (residues outside the HS binding site R^{20} and N^{45}) (Figure SI-4). During the Haddock calculations, the structure of CXCL12α was maintained, as only side chains of residues close to the ligand (<5 Å) were allowed to rotate. The φ/ψ torsion angles, which define the dp8$_{NS,6S}$ glycosidic linkages, remained close to their initial values (Figure SI-5). These data are also consistent with the angles observed in the structures of N-sulfated heparosan that were calculated from NOE-derived restraints.[23] Therefore, the Haddock protocol produced an ensemble of complexes consistent with the NMR data recorded on the ^{15}N-CXCL12α/^{13}C-dp8$_{NS,6S}$ complex. These modeling results were also in agreement with mutagenesis data[16] and converged to a configuration of the octasaccharide that is centered on a basic groove formed at the interface of the CXCL12α dimer.

It was of interest to compare the Haddock structures to the results of a naive unrestrained modeling method. We performed a series of molecular docking calculations followed by MD simulations in water, according to a recently described method.[24] The combination of molecular docking and MD ensured a large exploration of the docking modes of the ligand on the protein surface. The dp8$_{NS,6S}$ compound was first docked onto the basic groove of the CXCL12α dimer (PDB entry 2NWG) with three different alignments of the octasaccharide: one with aligned mass centers for both the protein and the oligosaccharide (1) and two others in which the mass center of the protein was aligned with GlcA-4 (2) or GlcN-5 (3). Each docking model was eventually used as the starting point for a 40 ns MD simulation in water. No substantial difference between the three MD simulations was observed, and the contacts observed were highly similar (Figures SI-6 and SI-7). For the three trajectories, the core of the protein/HS interaction involved the sugar units GlcN-3, GlcA-4, and GlcN-5 and the CXCL12α residues K^{24}, H^{25}, K^{27}, and R^{41}, with additional contacts with residues N^{46} and Q^{48} (Figure 5 and Figure SI-7); these results were similar to the Haddock results. The other protein/HS contacts involved the two first and last glycan units and the N-terminal residues of the protein, especially with residues R^{8} and R^{12} (Figure SI-7). Consistent with the fact that CXCL12α is a symmetrical homodimer, the best Haddock structures fit best with simulation 1, for which the centers of mass of both partners were aligned (Figure 5 and Figure SI-6). The main difference in the orientation observed in the last three glycan residues accounted for the variation of the protein structure during the MD trajectory, in which the 10 N-terminal residues of CXCL12α presented a particularly large rmsd (∼2.5 Å) (Figure SI-8).

Understanding the mechanism by which HS interacts with proteins remains a major challenge in glycosaminoglycan research and has far-reaching implications, particularly for the design of specific HS-derived therapeutic compounds.[4] NMR spectroscopy is an invaluable technique for extracting data on protein/HS complexes.[25] Using CXCL12α as a model system, we developed a method based on NMR analysis of a ^{13}C-labeled

HS in combination with a ^{15}N-labeled protein that enabled the identification of the main contributors of the interaction and from which a model of the complex could be derived. The fact that two distinct approaches (HADDOCK calculations based on experimental NMR data and MD simulations in explicit water) converged toward a very similar result validates our method.

The availability of HS-modifying enzymes (including C5-epimerase and various sulfotransferases) will allow the generation of differentially modified, isotopically labeled octasaccharides to increase the NMR signal dispersion and record more site-specific information. Our strategy based on the ^{13}C labeling of an HS-like oligosaccharide has been successfully applied in our case to a weak-affinity complex and could be easily adapted to stronger-affinity complexes by measuring NOE distances between the two labeled molecules.

■ ASSOCIATED CONTENT

Supporting Information. Experimental procedures, titration data, and statistics for Haddock and MD simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
cedric.laguri@ibs.fr

■ ACKNOWLEDGMENT

We acknowledge PSB platforms for NMR time and amino acid quantification and the "Centre Informatique National de l'Enseignement Supérieur" (www.cines.fr) for providing computation time. The MD calculations were performed as part of the DOSCA Project from "Pole de Compétitivité MEDICEN". We also thank R. Vivès for helpful discussions, A. Favier for useful scripts and assistance with NMR experiments, and Noé Dumas for great technical help.

■ REFERENCES

(1) Lander, A. D. *Matrix Biol.* **1998**, *17*, 465.
(2) Bishop, J. R.; Schuksz, M.; Esko, J. D. *Nature* **2007**, *446*, 1030.
(3) Gandhi, N. S.; Mancera, R. L. *Chem. Biol. Drug Design* **2008**, *72*, 455.
(4) Lindahl, U.; Li, J. P. *Int. Rev. Cell. Mol. Biol.* **2009**, *276*, 105.
(5) Esko, J. D.; Lindahl, U. *J. Clin. Invest.* **2001**, *108*, 169.
(6) Imberty, A.; Lortat-Jacob, H.; Pérez, S. *Carbohydr. Res.* **2007**, *342*, 430.
(7) Vann, W. F.; Schmidt, M.; Jann, B.; Jann, K. *Eur. J. Biochem.* **1981**, *116*, 359.
(8) Chen, J.; Jones, C. L.; Liu, J. *Chem. Biol.* **2007**, *14*, 986.
(9) Linhardt, R. J.; Kim, J. *Chem. Biol.* **2007**, *14*, 972.
(10) Nguyen, T. K. N.; Tran, V. M.; Victor, X. V.; Skalicky, J. J.; Kuberan, B. *Carbohydr. Res.* **2010**, *345*, 2228.
(11) Casu, B.; Grazioli, G.; Razi, N.; Guerrini, M.; Naggi, A.; Torri, G.; Oreste, P.; Tursi, F.; Zoppetti, G.; Lindahl, U. *Carbohydr. Res.* **1994**, *263*, 271.
(12) Handel, T. M.; Johnson, Z.; Crown, S. E.; Lau, E. K.; Proudfoot, A. E. *Annu. Rev. Biochem.* **2005**, *74*, 385.
(13) Lortat-Jacob, H. *Curr. Opin. Struct. Biol.* **2009**, *19*, 543.
(14) Laguri, C.; Sadir, R.; Rueda, P.; Baleux, F.; Gans, P.; Arenzana-Seisdedos, F.; Lortat-Jacob, H. *PLoS One* **2007**, *2*, No. e1110.
(15) Schanda, P.; Kupce, E.; Brutscher, B. *J. Biomol. NMR* **2005**, *33*, 199.
(16) Sadir, R.; Baleux, F.; Grosdidier, A.; Imberty, A.; Lortat-Jacob, H. *J. Biol. Chem.* **2001**, *276*, 8288.
(17) Veldkamp, C. T.; Peterson, F. C.; Pelzek, A. J.; Volkman, B. F. *Protein Sci.* **2005**, *14*, 1071.
(18) Mayer, M.; Meyer, B. *J. Am. Chem. Soc.* **2001**, *123*, 6108.
(19) Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. *J. Am. Chem. Soc.* **2003**, *125*, 1731.
(20) Titushin, M. S.; Feng, Y.; Stepanyuk, G. A.; Li, Y.; Markova, S. V.; Golz, S.; Wang, B.; Lee, J.; Wang, J.; Vysotski, E. S.; Liu, Z. *J. Biol. Chem.* **2010**, *285*, 40891.
(21) Lou, Y.; Wei, S.; Rajasekaran, M.; Chou, C.; Hsu, H.; Tai, J.; Chen, C. *Nucleic Acids Res.* **2009**, *37*, 2381.
(22) Murphy, J. W.; Cho, Y.; Sachpatzidis, A.; Fan, C.; Hodsdon, M. E.; Lolis, E. *J. Biol. Chem.* **2007**, *282*, 10018.
(23) Zhang, Z.; McCallum, S. A.; Xie, J.; Nieto, L.; Corzana, F.; Jiménez-Barbero, J.; Chen, M.; Liu, J.; Linhardt, R. J. *J. Am. Chem. Soc.* **2008**, *130*, 12998.
(24) Sapay, N.; Cabannes, E.; Petitou, M.; Imberty, A.; *Glycobiology* [Online early access]. DOI:10.1093/glycob/cwr052. Published Online: May 13, 2011.
(25) Canales, A.; Angulo, J.; Ojeda, R.; Bruix, M.; Fayos, R.; Lozano, R.; Giménez-Gallego, G.; Martín-Lomas, M.; Nieto, P. M.; Jiménez-Barbero, J. *J. Am. Chem. Soc.* **2005**, *127*, 5778.