analytical chemistry

LETTER

pubs.acs.org/ac

Structural Analysis of Glycans by NMR Chemical Shift Prediction

Magnus Lundborg and Göran Widmalm*

Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

Supporting Information

ABSTRACT: Structural determination of N- and O-linked glycans as well as polysaccharides is hampered by the limited spectral dispersion. The computerized approach CASPER, an acronym for computer assisted spectrum evaluation of regular polysaccharides, uses liquid state NMR data to elucidate carbohydrate structure based on agreement with predicted ^{1}H and ^{13}C chemical shifts. We here demonstrate developments based on multiple through-bond *J*-based correlations that significantly enhance the credence to the sequence connectivities proposed in the analysis exemplified by an oligosaccharide and a bacterial polysaccharide. The approach is also suitable for predicting ^{1}H and ^{13}C NMR chemical shifts of synthesized oligosaccharides and glycoconjugates, thereby corroborating a proposed structure.

Glycans, most often in the form of glycoconjugates such as N- or O-linked glycoproteins, proteoglycans, and polysaccharides, are an important class of biomolecules that are involved in molecular recognition, protein trafficking, regulation, and inflammation among other things.[1,2] The rapidly evolving glycomics area, in which the various relationships for an organism's complete repertoire of glycans are studied, relies on that structural information can be related to function.[3] As a prerequisite for detailed investigations, the glycan structures must be determined as accurately as possible. Whereas glycan profiling rapidly can be carried out by mass spectrometry, it relies on knowledge of conserved structures and to some extent on known biosynthetic pathways. NMR spectroscopy, on the other hand, is able to determine stereochemical relationships of completely novel sugars and in a reliable way also the anomeric configuration. The recent developments in sensitivity for NMR spectroscopy including cryoprobe technology,[4] "single-scan 2D techniques", and dynamic nuclear polarization[5] promise continued use of NMR for carbohydrates and other biomolecules.[6−8]

The structural determination of glycans is posed by the specific problem of limited spectral dispersion both for ^{1}H and ^{13}C nuclei. Using 1D and 2D as well as multidimensional NMR spectroscopy, the process of assignment of a certain resonance to a specific nucleus is still tedious and the most time-consuming part of the process. We developed a computerized method to determine carbohydrate structure from 1D ^{1}H and ^{13}C NMR spectra and ^{1}H,^{13}C one-bond correlations (HSQC/HETCOR spectra). It was termed CASPER, an acronym for computer assisted spectrum evaluation of regular polysaccharides, and can be used for oligosaccharides as well. Its advantage is that it uses unassigned NMR spectra; thus, the procedure is rapid. Its limitation is still that among the ranked structural suggestions the relative differences may be small and further experiments, but only a few, may be needed to differentiate the top ranked structures. Some time ago, CASPER was made available to the scientific community via the Internet (http://www.casper.organ.su.se/casper).[9]

We here present our recent developments and validations based on through-bond connectivities, i.e., ^{1}H,^{1}H-homonuclear and ^{1}H,^{13}C-heteronuclear spin−spin coupling constants. The advantageous approach of CASPER is still retained, i.e., unassigned NMR chemical shifts. Thus, in the present development, the analysis is, in addition to 1D ^{1}H and ^{13}C NMR spectra and 2D HSQC spectra (decoupled and coupled; the latter to facilitate determination of anomeric configuration), enhanced by ^{1}H,^{1}H-TOCSY experiments (typically having mixing times of 40 and 100 ms) from which sugar spin systems can be tied together, a ^{1}H,^{13}C-H2BC experiment[10] defining intraresidue correlations between adjacent proton and carbon atoms and a ^{1}H,^{13}C-HMBC experiment[11] (typical mixing time 50 ms) giving additional intramolecular three-bond correlations valuable in the assignments of ^{1}H and ^{13}C resonances within a specific sugar residue. The latter experiment also gives very important information on sequential relationships between sugar residues by trans-glycosidic heteronuclear three-bond correlations. Thus, the possibility to utilize the three above-described experiments in the structural analysis with CASPER enhances significantly the ability to assign each sugar spin system to a set of experimental resonances. CASPER does not require all chemical shifts but can assign multiple calculated signals to a single experimental signal if they are close enough. It is important to notice that all submitted experimental signals will be used in the assignment, which means it is better to leave out a resonance if it is suspected to originate from a contaminant. Another new function is employing knowledge about classes of glycans to generate only structures containing the structural elements of a specific kind of carbohydrate. Current implementations are biological rules for N-linked and

Received: December 14, 2010
Accepted: January 21, 2011
Published: January 31, 2011

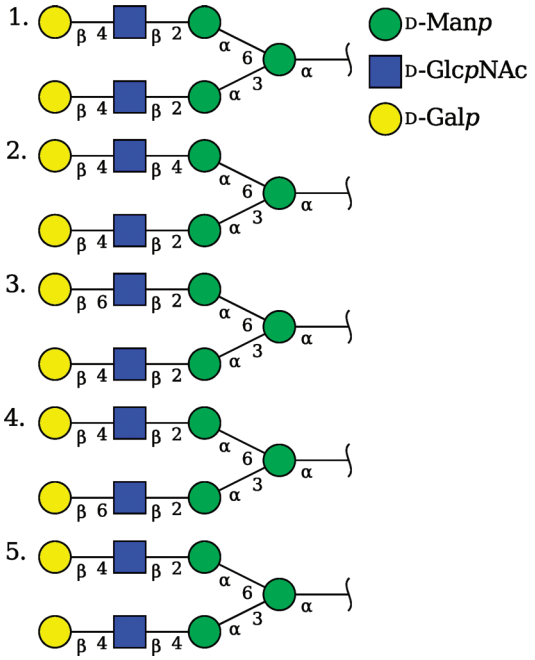

Figure 1. CASPER output of the five top ranked structures in CFG format utilizing GlycanBuilder[15] for their generation (for standard carbohydrate listing and NMR data from a $^{1}H,^{13}C$-HSQC NMR spectrum see Supplementary Table S1.1 and Scheme S2.1 in the Supporting Information) for a heptasaccharide corresponding to a partial structure present in N-linked glycoproteins. That the sugar at the reducing end is a mannopyranose residue is denoted by a vertical tilde. The relative deviations for structures 1−5 are 1.00, 1.46, 1.49, 1.50, and 1.50, respectively.

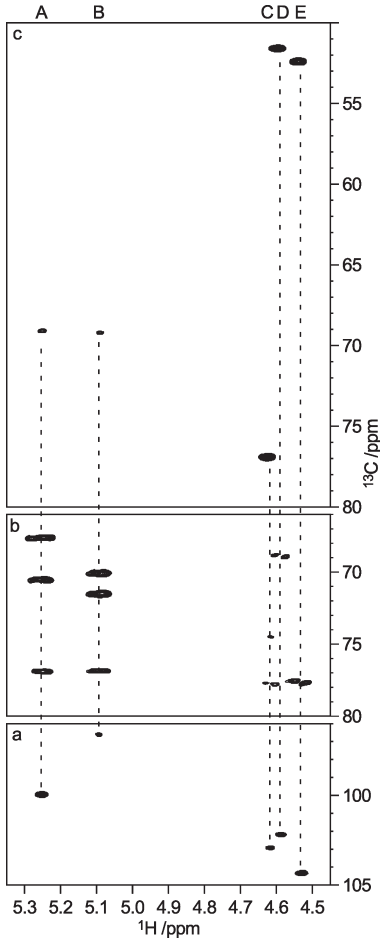

Figure 2. Selected NMR spectral regions (^{1}H anomeric chemical shift region) used by CASPER in the structural analysis of the *E. coli* O128 O-antigen polysaccharide having pentasaccharide repeating units: (a) $^{1}H,^{13}C$-HSQC, (b) $^{1}H,^{13}C$-HMBC (mixing time 50 ms), and (c) $^{1}H,^{13}C$-H2BC. Capital letters A−E at the top denote constituent sugar residues labeled by decreasing ^{1}H NMR chemical shifts of their anomeric protons.

O-linked glycans, Wzx/Wzy dependent WecA assembled *Escherichia coli* O-antigen repeating units,[12] *Shigella flexneri* O-antigen repeating units, and *Haemophilus influenzae* lipopolysaccharides. More biological rules can easily be added on request. In previous versions of CASPER, information from component and methylation analyses were needed as input. The present implementation can approach structural problems without the need of this information (only the absolute configurations of the sugar residues, i.e., D- or L-sugar, need to be determined at some point as described[13,14]) since the NMR chemical shifts of two enantiomers are identical, but the calculation time increases drastically if many residues, or linkage positions, are unknown. We herein validate the present developments with two key types of carbohydrate structures, viz., a branched oligosaccharide being part of an N-linked glycan and a bacterial polysaccharide having pentasaccharide repeating units.

In the first case, a $^{1}H,^{13}C$-HSQC spectrum of a heptasaccharide[16] was acquired to provide information about one-bond proton−carbon correlations. A $^{1}H,^{1}H$-TOCSY spectrum was used to identify the spin systems, starting from the anomeric signals. $^{1}H,^{13}C$-HMBC, and $^{1}H,^{13}C$-H2BC spectra gave anomeric inter- and intraresidue correlations. Chemical shifts from 2D spectra are entered as chemical shift pairs in the web interface, ^{13}C followed by ^{1}H chemical shifts for heteronuclear experiments. Apart from that, there are no specific requirements for how to separate the chemical shifts or chemical shift pairs. The constituent sugar residues were submitted, viz., 3 D-Man*p*, 2 D-Glc*p*NAc, and 2 D-Gal*p*. On the basis of conserved structures of N-linked oligosaccharides, one of the mannose residues was specified as 3,6-substituted; the remaining two mannose residues were chosen to specifically link to the mannose branch point, and the galactose residues were chosen as the nonreducing ends of the oligosaccharide by specifying that they should be linked only from their anomeric position. The remaining possible substitution positions were then selected as unknowns. In addition to this, one anomeric $^{3}J_{H,H}$ coupling >7 Hz was observed, in a ^{1}H NMR spectrum, and submitted to CASPER. The calculations took 100 s on a standard workstation, and the correct structure was found at the top of the list of results (Figure 1) with a large score difference to the structure ranked second, which differs in one linkage position but has the same anomeric configurations. Note that the submitted linkage possibilities were deliberately set relatively specific in this example to reduce the calculation times. It is possible to submit all linkages, as well as residues, as completely unknown, but it would take too long to use the present web interface setup for that structure determination.

In the second example, correlations from the same NMR experiment types (Figure 2) as above were used on the O-antigen polysaccharide from *E. coli* O128 having pentasaccharide repeating units.[17] The structure information submitted as input to CASPER was 2 D-Gal*p*NAc, 2 D-Gal*p*, and 1 L-Fuc*p*. All linkage

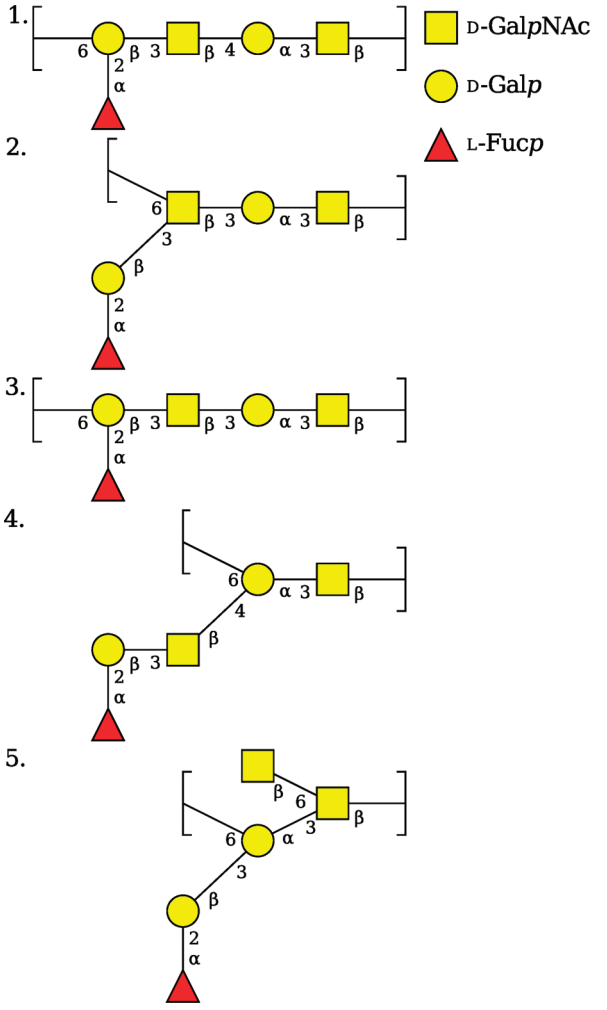

Figure 3. CASPER output of the five top ranked structures in CFG format (for standard carbohydrate listing and NMR data see Supplementary Table S1.2 and Scheme S2.2 in the Supporting Information) for the *E. coli* O128 O-antigen polysaccharide. The square brackets denote that the enclosed structures represent repeating units. The relative deviations for structures 1−5 are 1.00, 1.09, 1.09, 1.11, and 1.12, respectively.

positions were marked as unknowns. Information from anomeric coupling constants was submitted as follows: two $^{3}J_{H,H}$ coupling constants in the range of 2−7 Hz and three larger than 7 Hz. Since this was known to be an O-antigenic polysaccharide from *E. coli*, the biological rule for Wzx/Wzy WecA repeating units was used, i.e., a 3- or 4-substituted D-Glc*p*NAc or D-Gal*p*NAc residue is present at the reducing end of the biological repeating unit,[12] thereby reducing the number of possible structures. The structure calculations (taking 110 s) returned the correct structure as the top candidate (Figure 3). The score difference between the top five structures is smaller than in the first case, but comparing them to known *E. coli* O-antigens supports the structures ranked first and third, since four sugars in the backbone of the polymer is common, i.e., structures with four sugar residues in the backbone and one residue as a side chain in the repeating unit, whereas the other combinations of branches are rare.[12] A closer inspection and analysis of the already obtained NMR spectra as well as acquiring an additional NMR spectrum if needed would settle this uncertainty.

In conclusion, the developments presented in this work have led to rapid automatic structural determination of complex carbohydrates and polysaccharides based on NMR chemical shift predictions. The additional input data needed besides unassigned and peak-picked 2D NMR spectra are few which facilitate rapid analysis and ease of use. Furthermore, the procedure may also be employed efficiently for NMR chemical shift assignments of known carbohydrate structures but where the assignments are needed for another study, e.g., on carbohydrate−protein interactions. Future developments will be targeted toward covering carbohydrate structures found on glycoproteins in general, integration of NMR softwares such as CCPN,[18] utilization of recent advances in which pure chemical shift correlations are produced in ^{1}H,^{1}H-TOCSY spectra devoid of homonuclear couplings,[19] and high throughput automated structural analysis using efficient data transfer and handling.

■ ASSOCIATED CONTENT

Supporting Information. Full results lists from the CASPER structure determinations as well as CASPER assignments of the top-ranked correct structures. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: gw@organ.su.se.

■ ACKNOWLEDGMENT

This work was funded by the sixth Research Framework Program of the European Union (Contract: RIDS Contract Number 011952) as part of the EUROCarbDB project. It was also supported by grants from the Swedish Research Council, The Knut and Alice Wallenberg Foundation, The Lars Hierta Memorial Foundation, and Magn. Bergvall Foundation.

■ REFERENCES

(1) Ohtsubo, K.; Marth, J. D. *Cell* **2006**, *126*, 855–867.
(2) Dam, T. K.; Brewer, C. F. *Glycobiology* **2010**, *20*, 270–279.
(3) Prescher, J. A.; Bertozzi, C. R. *Cell* **2006**, *126*, 851–854.
(4) Kovacs, H.; Moskau, D.; Spraul, M. *Prog. NMR Spectr.* **2005**, *46*, 131–155 .
(5) Mishkovsky, M.; Frydman, L. *ChemPhysChem* **2008**, *9*, 2340–2348.
(6) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.
(7) Aramini, J. M.; Rossi, P.; Anklin, C.; Xiao, R.; Montelione, G. T. *Nat. Methods* **2007**, *4*, 419–493.
(8) Zhang, Q.; Al-Hashimi, H. M. *Nat. Methods* **2008**, *5*, 243–245.
(9) Jansson, P.-E.; Stenutz, R.; Widmalm, G. *Carbohydr. Res.* **2006**, *341*, 1003–1010.
(10) Nyberg, N. T.; Duus, J. Ø.; Sørensen, O. W. *J. Am. Chem. Soc.* **2005**, *127*, 6154–6155.
(11) Bax, A.; Summers, M. F. *J. Am. Chem. Soc.* **1986**, *108*, 2093–2094.
(12) Stenutz, R.; Weintraub, A.; Widmalm, G. *FEMS Microbiol. Rev.* **2006**, *30*, 382–403.
(13) Leontein, K.; Lönngren, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87–89.
(14) Säwén, E.; Huttunen, E.; Zhang, X.; Yang, Z.; Widmalm, G. *J. Biomol. NMR* **2010**, *47*, 125–134.

(15) Ceroni, A.; Dell, A.; Haslam, S. M. *Source Code Biol. Med.* **2007**, *2*, 3.
(16) Bock, K.; Arnarp, J.; Lönngren, J. *Eur. J. Biochem.* **1982**, *129*, 171–178.
(17) Sengupta, P.; Bhattacharyya, T.; Shashkov, A.; Kochanowski, H.; Basu, S. *Carbohydr. Res.* **1995**, *277*, 283–290.
(18) Vranken, W. V.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct. Funct. Bioinf.* **2005**, *59*, 687–696.
(19) Morris, G. A.; Aguilar, J. A.; Evans, R.; Haiber, S.; Nilsson, M. *J. Am. Chem. Soc.* **2010**, *132*, 12770–12772.