Communication

pubs.acs.org/Biomac

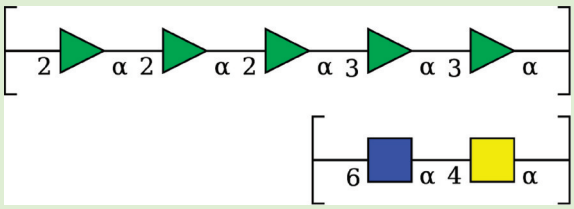
Automatic Structure Determination of Regular Polysaccharides Based Solely on NMR Spectroscopy

Magnus Lundborg, Carolina Fontana, and Göran Widmalm*

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

Supporting Information

ABSTRACT: The structural analysis of polysaccharides requires that the sugar components and their absolute configurations are determined. We here show that this can be performed based on NMR spectroscopy by utilizing butanolysis with (+)- and (−)-2-butanol that gives the corresponding 2-butyl glycosides with characteristic ^{1}H and ^{13}C NMR chemical shifts. The subsequent computer-assisted structural determination by CASPER can then be based solely on NMR data in a fully automatic way as shown and implemented herein. The method is additionally advantageous in that reference data only have to be prepared once and from a user's point of view only the unknown sample has to be derivatized for use in CASPER.



INTRODUCTION

Analysis of carbohydrate structure in general, and glycans of biological origin in particular, relies on a number of different techniques. The approach chosen depends on the question(s) to be answered but also on the amount of material available. Mass spectrometry (MS) techniques are available for analysis of N- and O-linked glycans, usually after a tryptic digest or liberation by chemical or glycan specific enzymes, using several fragmentations (MS^n) as part of the sequencing methodologies.[1] The high sensitivity of MS makes it extremely powerful, and it has been combined with other techniques such as HPLC, lectin affinity chromatography, microarrays, and specific chemical release in the study of glycoconjugate structure.[2−5] Although linkage positions and the anomeric configuration of sugar residues are possible to determine by MS,[6] its use in these aspects is still limited.

For a glycan of unknown structure NMR spectroscopy provides a way to determine its structure without any doubt; in practice, other techniques such as MS complement the de novo structural determination. A key part that needs to be determined is the absolute configuration of the sugar residues as well as (chiral) substituents when present. This is still usually carried out by gas chromatography using either derivatized samples so that diastereomeric pairs are analyzed[7] or sometimes using a chiral column phase. The absolute configuration of sugar residues may also be determined by ^{1}H NMR spectroscopy, once the oligo- or polysaccharide hydrolysate has been acylated by a chiral acyl ester.[8] The technique was recently used successfully for the determination of the absolute configurations of the sugar components of the exopolysaccharide from *Streptococcus thermophilus* ST1.[9] Recent technological developments in probe design and miniaturization promise significant increase in sensitivity of NMR spectroscopy such that NMR data may be acquired at low nanomolar concentrations, thereby alleviating the problem of low sensitivity of the NMR technique.[10]

In the present communication we present a protocol for rapid and automatic structure determination of oligo- and polysaccharides based solely on NMR spectroscopy. The conventional and manual analyses of gas chromatograms and 2D NMR spectra are time-consuming and prone to errors. We here present a method in which the structure of glycans can be determined solely by NMR spectroscopy in a highly automatic way using just a few NMR spectra. The CCPN Analysis software[11] is used to handle unassigned resonances in the spectra, prior to analysis with the CASPER (computer assisted spectrum evaluation of regular polysaccharides) program.[12] The sequence analysis is performed by CASPER, based on NMR chemical shifts and how they change when sugar residues are linked together, and in particular using information from HMBC spectra as described before.[12] Component analysis is the last step toward automatic structure determination by NMR spectroscopy.

EXPERIMENTAL SECTION

Preparation of (+)- and (−)-2-Butyl Glycosides of Non-Amino Sugars. The monosaccharide (5 mg) was dissolved in (+)-(*S*)- or (−)-(*R*)-2-butanol (0.5 mL), followed by addition of acetyl chloride (75 μL). Nitrogen was bubbled through the solution for 0.5 min, the tube was sealed, and the solution was heated at 80 °C overnight. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in water (1 mL), neutralized with 0.8 M NaOH, filtered through glass wool, and

Received: August 21, 2011
Revised: September 20, 2011
Published: September 28, 2011

freeze-dried. The derivatives thus obtained were analyzed by NMR spectroscopy without further purification.

Preparation of (+)- and (−)-2-Butyl Glycosides of N-Acetylated Amino Sugars. The N-acetylated amino sugar (5 mg) was dissolved in (+)-(*S*)- or (−)-(*R*)-2-butanol (0.5 mL), followed by addition of acetyl chloride (35−75 μL). Nitrogen was bubbled trough the solution for 0.5 min, the tube was sealed, and the solution was heated at 80 °C for 7 h. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in 1 mL of water, and Na_2CO_3 was added until the pH reached 11. Acetic anhydride (10 μL) was added at 0 °C, and the solution stirred for 2 h.[13] The resulting mixture was treated with Dowex 50W X8 (H^+ form) resin and filtered through glass wool. The solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in water (1 mL), neutralized with 0.8 M NaOH, treated with a mixed bed resin (TMD-8 H^+ and OH^- form, Sigma), filtered through glass wool, and freeze-dried. The derivatives thus obtained were analyzed by NMR spectroscopy without further purification.

Absolute Configuration Determination of the O-Antigen Polysaccharide of *Klebsiella pneumoniae* Strain 22535.[14] The polysaccharide (2.1 mg) was hydrolyzed with 2 M trifluoroacetic acid (1 mL) at 120 °C for 30 min. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (0.5 mL) was added and evaporated to dryness; the procedure was repeated twice. The crude material was dissolved in (+)-(*S*)-2-butanol (0.5 mL), followed by addition of acetyl chloride (75 μL). Nitrogen was bubbled trough the solution for 0.5 min, the tube was sealed, and the solution was heated at 80 °C for 16 h. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in water (1 mL), neutralized with 0.8 M NaOH, filtered through glass wool, and freeze-dried.

Absolute Configuration Determination of the O-Antigen Polysaccharide of *Escherichia coli* O101 (Containing N-Acetylated Amino Sugars).[15] The polysaccharide (2.4 mg) was hydrolyzed with 2 M trifluoroacetic acid (1 mL) at 120 °C for 30 min. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (0.5 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was re-N-acetylated with acetic anhydride (250 μL) for 4 h at room temperature. The solvent was evaporated to dryness with a stream of dry air. The crude material was dissolved in (+)-(*S*)-2-butanol (0.2 mL), followed by addition of acetyl chloride (30 μL). Nitrogen was bubbled through the solution for 0.5 min, the tube was sealed, and the solution was heated at 80 °C for 7 h. The reaction mixture was allowed to attain room temperature, and the solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in water (1 mL), and Na_2CO_3 was added until the pH reached 11. Acetic anhydride (10 μL) was added at 0 °C, and the solution was stirred for 2 h.[13] The resulting mixture was treated with Dowex 50W X8 (H^+ form) resin and filtered through glass wool. The solvent was evaporated to dryness with a stream of dry air. Methanol (2 mL) was added and evaporated to dryness; the procedure was repeated twice. The sample was dissolved in water (1 mL), neutralized with 0.8 M NaOH, treated with a mixed bed resin (TMD-8 H^+ and OH^- form, Sigma), filtered through glass wool, and freeze-dried.

NMR Spectroscopy. NMR experiments were carried out in D_2O (0.55 mL) on five different spectrometers: Bruker AVANCE 500 MHz and Bruker AVANCE III 700 MHz spectrometers equipped with 5 mm TCI Z-Gradient CryoProbes, Bruker AVANCE 400 MHz and Bruker AVANCE II 500 MHz spectrometers equipped with 5 mm BBI Z-Gradient probes, and a Bruker AVANCE III 600 MHz spectrometer equipped with a 5 mm inverse Z-gradient TXI ($^1H/^{13}C/^{15}N$) probe. Chemical shifts are reported in ppm using internal sodium 3-trimethylsilyl-(2,2,3,3-$^2H_4$)-propanoate (TSP, δ_H 0.00) or external 1,4-dioxane in D_2O (δ_C 67.40) as references.

1D and 2D NMR spectra of the O-antigen polysaccharide of *Klebsiella pneumoniae* strain 22535 (5.2 mg) and the (+)-2-butyl glycosides of the hydrolyzed polysaccharide were recorded at 700 MHz to reduce spectral overlap and 500 MHz, respectively. NMR spectra of the O-antigen polysaccharide of *Escherichia coli* O101 (9.6 mg) and the (+)-2-butyl glycosides of the hydrolyzed polysaccharide were carried out at 500 and 700 MHz, respectively. Measurements were made at 40 °C for the O-antigen polysaccharide of *Escherichia coli* O101 and at 70 °C for all the other samples.

^1H and ^{13}C NMR chemical shifts assignments of the (+)- and (−)-2-butyl glycosides were performed on different spectrometers (400, 500, and 600 MHz) at 70 °C. ^1H chemical shifts assignments were performed using $^1H,^1H$-TOCSY experiments with two different mixing times (10 and 70 ms). ^{13}C NMR chemical shifts were assigned using $^1H,^{13}C$-HMBC and proton decoupled $^1H,^{13}C$-HSQC experiments.

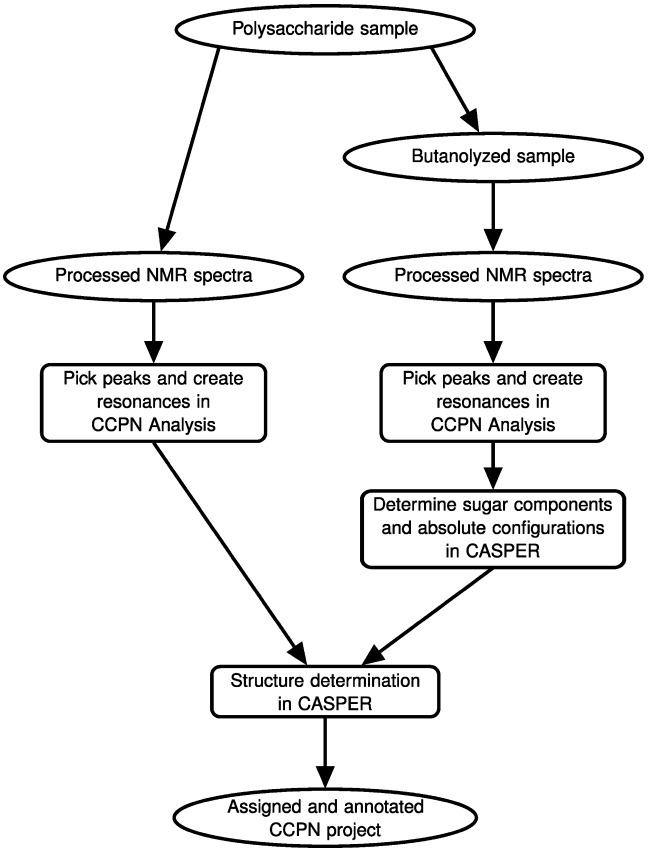
RESULTS AND DISCUSSION

A very common way in the analysis of absolute configuration of sugar residues is to derivatize the unknown components of a sample with an optically active alcohol, e.g., (+)- or (−)-2-butanol. The glycosides that are obtained are subsequently compared to standards made from the (+)- or (−)-forms of the alcohol and analyzed by gas chromatography.[7] Furthermore, monosaccharides have been used to determine the absolute configuration of 2-butanol by ^1H NMR spectroscopy.[16] We here extend the above approaches to a general method, in which the sugar residues of an oligo- or polysaccharide may be determined solely by NMR spectroscopy including the absolute configurations of the constituent sugar residues. Once the NMR data have been acquired for the reference compounds and stored in the database utilized by CASPER, reference material do not have to be prepared each time; only the unknown sample has to be analyzed by 1D or 2D NMR experiments.

The CASPER program uses the ^1H and ^{13}C NMR chemical shifts of the monosaccharides and methyl glycosides as part of the process to predict NMR chemical shifts of oligo- or polysaccharides.[17] Thus, from a hydrolysate or methanolysis of a carbohydrate-containing material, one should be able to identify the sugar components present, albeit without the determination of their absolute configurations(s). In the present developments we have prepared reference substances based on butanolysis of monosaccharides with (+)- and (−)-2-butanol that give the corresponding 2-butyl glycosides with characteristic ^1H and ^{13}C NMR chemical shifts (Tables S2.1 and S2.2). Depending on the monosaccharide the α- and β-anomeric forms of the pyranosides as well as the furanosides will be present to a different extent. Since the methodology is based on a computerized analysis, the mixture of anomeric and ring forms is not regarded as a limitation but will instead be used in advantageous way, acting as a fingerprint, to identify the sugar residues and their absolute configurations. Note that the (+)-D-sugar will give an identical NMR spectrum to that of the (−)-L-sugar, since they are enantiomers. Consequently, from a practical point of view, either of the 2-butanol enantiomers may be used in the analysis of an unknown sample (provided that reference data of the sugar are available for CASPER to analyze). The methodology is outlined in Scheme 1.

The sugar component analysis method uses NMR chemical shifts from $^1H,^{13}C$ and/or $^1H,^{13}C$-HSQC spectra as input. The

Scheme 1. Methodology Used by CASPER To Determine the Structure of an Oligo- or Polysaccharide Based Solely on NMR Spectroscopy



chemical shifts of the α- and β-anomeric forms as well as different ring forms (if available) for the monosaccharide residues in the CASPER database that match the lysis method used, i.e., hydrolysis, methanolysis, or butanolysis, are compared to the experimental data. The tests are performed for one monosaccharide at a time, and the chemical shifts and/or correlations from 2D HSQC spectra are matched to the closest experimental signal, allowing the same signal to be used more than once. This means that it is not necessary to integrate the signals or completely resolve overlapping peaks, but on the other hand, some submitted signals might not be used. 1D ^{1}H and ^{13}C chemical shift deviations are the absolute difference between experimental data and the database entries, whereas the HSQC difference is the distance between the 2D coordinates multiplying the ^{13}C chemical shift separation by 0.2. If the difference of the best match of a peak/cross-peak from an atom/atom pair in a residue is larger than 0.05, 0.25, or 0.25 ppm for ^{1}H,^{13}C and ^{1}H,^{13}C-HSQC data, respectively, the deviation is set to that upper limit and no match is made. The different experiment types have distinctive thresholds: the cutoffs are 0.15, 0.75, and 0.75 ppm for ^{1}H,^{13}C and ^{1}H,^{13}C-HSQC data, respectively. The maximum deviation per signal is based upon the signal assignment tolerances used by CASPER, whereas the thresholds for each experiment type were chosen after testing the sugar component analysis module. If at least half of the signals of each experiment type can be assigned for a component form (α- and β-anomeric forms as well as pyranose and furanose ring forms), the deviations of that form will be counted. This means that a sugar component will not be discarded just because, e.g., the signals from the furanose form are not observed. If the resonance deviations of a component (divided by the number of forms, of which deviations were counted) are below the total thresholds of all experimental data types submitted, it will be proposed by CASPER as a component and listed in the results along with its deviations and the experimental chemical shifts used for matching this residue. If there are experimental signals that are not used for matching any of the suggested residues, they will also be listed. It is common that there are unused signals, especially if they are close to each other, since only the best match between experimental and database chemical shifts are used. The presence of several unused chemical shifts may indicate that an additional sugar component, which was not identified by CASPER, is present in the investigated material or that a substituent has not been accounted for, which highlights that in such cases further detailed analysis of the NMR spectra is required (Scheme S1.1).

The fully computerized structure determination approach based on unassigned NMR spectra is presented for two O-antigen polysaccharides, from *Klebsiella pneumoniae* strain 22535 and *Escherichia coli* O101,[14,15] shown in Figure 1. In order to obtain an efficient way of handling NMR data we have developed the possibility of uploading CCPN projects directly into the CASPER interface. In the current implementation[12] a set of NMR data, acquired on the oligo- or polysaccharide, e.g., the 1D ^{13}C NMR spectrum, 2D ^{1}H,^{13}C-HSQC (Figure 2a–c), ^{1}H,^{13}C-H2BC, and ^{1}H,^{13}C-HMBC spectra, from which ^{1}H and ^{13}C NMR chemical shifts have been obtained by peak-picking using the CCPN Analysis software,[11] is used for the structure determination procedure.

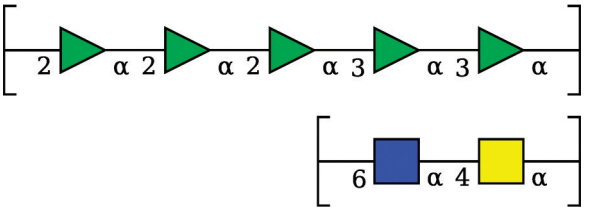


Figure 1. Structures of the repeating units of the O-antigen polysaccharides from *Klebsiella pneumoniae* strain 22535 (top) and *Escherichia coli* O101 (bottom) represented in CFG format.

From butanolysis of the *Klebsiella pneumoniae* polysaccharide one set of resonances is observed in the ^{13}C and/or ^{1}H,^{13}C-HSQC spectra (Figure 2d–f), whereas the butanolyzed monosaccharides give distinctly different NMR spectra for the (+)- and (−)-forms (Figure 2g–i), and this chemical shift information is used for the analysis procedure. The data from the butanolyzed polysaccharide are utilized in a second CCPN project which is submitted to CASPER and execution of the absolute configuration module determines the sugars present in the polysaccharide and their absolute configuration(s) in ∼5 s. The lists of the signals expected to come from the sugars in the CCPN projects are presented in Tables S3.1 and S3.2. The resulting sugar components (for a detailed list of results see Tables S4.1 and S4.2) are subsequently transferred to the structure determination module/interface where now the sugar residues automatically are entered with their appropriate absolute configuration(s), but where the linkage positions are unknown. The number of the constituent residues is not determined from the component analysis, requiring the user to

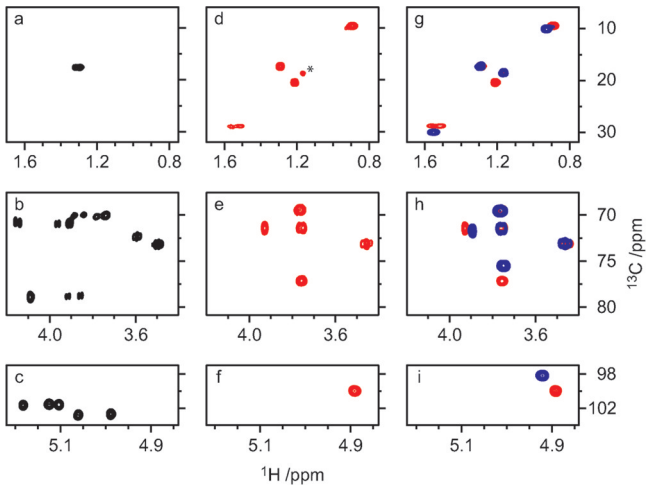


Figure 2. Selected regions of ^{1}H,^{13}C-HSQC NMR spectra of (a–c) the polysaccharide from *Klebsiella pneumoniae* strain 22535; (d–f) (+)-2-butyl glycosides from the polysaccharide from *Klebsiella pneumoniae* strain 22535 showing that the homopolymer contains L-rhamnose; and (g–i) (+)- and (−)-2-butyl glycosides from L-rhamnose in red and blue color, respectively. An impurity is marked with an asterisk in (d).

specify that in the structure determination input. In this case, that meant that five L-rhamnose residues were selected instead of one, which was identified from the component analysis of the *Klebsiella pneumoniae* polysaccharide. The subsequent structure determination then returns a list of structures ranked according to their agreement to experimental data. The correct structures of both polysaccharides are top ranked in the results (see Tables S5.1 and S5.2) with large margins to the second candidate. For the *E. coli* O101 polysaccharide the relative difference between the two highest ranked structures is 1.25, and the second *Klebsiella pneumoniae* polysaccharide structure suggestion has a relative deviation, compared to the best structure, of 1.30, both of which are significantly large to rule out the second option as a probable structure candidate.[12] The structure determinations took approximately 5 and 35 s for the *E. coli* O101 and *Klebsiella pneumoniae* polysaccharides, respectively, and the results, which are in full agreement with the previous structural studies,[14,15] are shown in Tables S5.1 and S5.2. Subsequently, ^{1}H and ^{13}C NMR chemical shift assignments (listed in Tables S6.1 and S6.2) are then transferred back into the CCPN project, and further analysis may be carried out in CCPN analysis in which the ^{1}H and ^{13}C NMR resonances now are assigned to atoms of the sugar residues. The limitation of the methodology is the absence of NMR chemical shifts in the database for some sugar components or oligosaccharides, but this is continuously addressed by adding assigned NMR chemical shift data; e.g., the ^{1}H and ^{13}C NMR chemical shifts of butanolyzed uronic acids are presently being analyzed and will subsequently be included as part of the database. At the present time the method requires ∼1 mg of polysaccharide which may be complex, but it should have a regular structure with repeating units. Impure and/or heterogeneous samples may lead to larger differences between experimentally acquired chemical shifts and those predicted by CASPER; consequently, the credence to the proposed structure will then be lower. Thus, robust purification procedures and techniques are essential to apply prior to utilizing the computerized CASPER approach based on unassigned NMR data. The slightly larger deviations between calculated and experimental ^{13}C NMR chemical shifts for the *E. coli* O101 polysaccharide compared to the *Klebsiella pneumoniae* polysaccharide are due to model approximations and available oligosaccharides in the current database. To a small extent also the different temperature (40 °C) of acquisition of NMR experiments for the *E. coli* O101 polysaccharide compared to the temperature (70 °C) at which the database is compiled may lead to small deviations. However, this example shows that it is possible to acquire the experimental NMR data at a temperature different from that of the database and still determine the polysaccharide structure in an efficient way. In the present study two polysaccharides having linear repeating units were presented, but the absolute configuration methodology is likewise applicable to branched polysaccharides. The recent developments have been made available to the scientific community via the Internet (http://www.casper.organ.su.se/casper).[12,17]

In conclusion, we have presented a fully automatic procedure to determine the primary structure of a regular polysaccharide based on NMR spectroscopy only in which the absolute configuration determination of the sugar components is obtained from unassigned NMR spectra. Furthermore, once the monosaccharide NMR reference data have been added to the database, references do not have to be prepared for each new and unknown sample. Only one butanolyzed sample of the material whose structure should be determined needs to be prepared.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The sugar component analysis method flowchart, ^{1}H and ^{13}C NMR chemical shifts of 2-butyl glycopyranosides and the corresponding ones from the polysaccharides, results from the sugar component analyses and structure determinations as well as NMR chemical shift predictions by CASPER for the polysaccharides. 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.

■ ACKNOWLEDGMENTS

This work was funded by the sixth Research Framework Program of the European Union (Contract: RIDS contract no. 011952) as part of the EUROCarbDB project. It was also supported by grants from the Swedish Research Council and The Knut and Alice Wallenberg Foundation. The research leading to these results has received funding from the European Commission's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 215536.

■ REFERENCES

(1) North, S. J.; Hitchen, P. G.; Haslam, S. M.; Dell, A. *Curr. Opin. Struct. Biol.* **2009**, *19*, 498–506.

(2) Mariño, K.; Bones, J.; Kattla, J. J.; Rudd, P. M. *Nat. Chem. Biol.* **2010**, *6*, 713–723.

(3) Kubota, K.; Sato, Y.; Suzuki, Y.; Goto-Inoue, N.; Toda, T.; Suzuki, M.; Hisanaga, S.-i.; Suzuki, A.; Endo, T. *Anal. Chem.* **2008**, *80*, 3693–3698.

(4) Patwa, T.; Li, C.; Simeone, D. M.; Lubman, D. M. *Mass Spectrom. Rev.* **2010**, *29*, 830–844.

(5) Nilsson, J.; Rüetschi, U.; Halim, A.; Hesse, C.; Carlsohn, E.; Brinkmalm, G.; Larsson, G. *Nature Methods* **2009**, *6*, 809−811.
(6) Daikoku, S.; Widmalm, G.; Kanie, O. *Rapid Commun. Mass Spectrom.* **2009**, *23*, 3713−3719.
(7) Leontein, K.; Lönngren, J. *Methods Carbohydr. Chem.* **1993**, *9*, 87−89, and references therein.
(8) York, W. S.; Hantus, S.; Albersheim, P.; Darvill, A. G. *Carbohydr. Res.* **1997**, *300*, 199−206.
(9) Säwén, E.; Huttunen, E.; Zhang, X.; Yang, Z.; Widmalm, G. *J. Biomol. NMR* **2010**, *47*, 125−134.
(10) Maguire, Y.; Chuang, I. L.; Zhang, S.; Gershenfeld, N. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9198−9203.
(11) 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.
(12) Lundborg, M.; Widmalm, G. *Anal. Chem.* **2011**, *83*, 1514−1517.
(13) Yu, F.; Wolff, J. J.; Amster, I. J.; Prestegard, J. H. *J. Am. Chem. Soc.* **2007**, *129*, 13288−13297.
(14) Ansaruzzaman, M.; Albert, M. J.; Holme, T.; Jansson, P.-E.; Rahman, M. M.; Widmalm, G. *Eur. J. Biochem.* **1996**, *237*, 786−791.
(15) Staaf, M.; Urbina, F.; Weintraub, A.; Widmalm, G. *Carbohydr. Res.* **1997**, *297*, 297−299.
(16) Seroka, P.; Płosiński, M.; Czub, J.; Sowiński, P.; Pawlak, J. *Magn. Reson. Chem.* **2006**, *44*, 132−138.
(17) Jansson, P.-E.; Stenutz, R.; Widmalm, G. *Carbohydr. Res.* **2006**, *341*, 1003−1010.