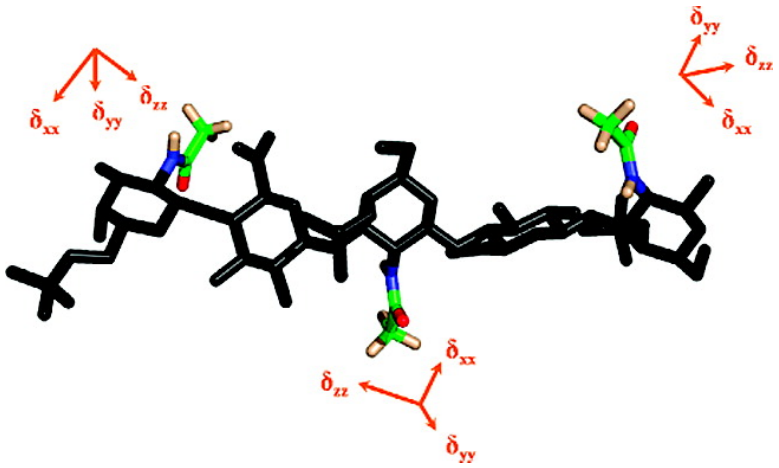
Article

Conformational Preferences of Chondroitin Sulfate Oligomers Using Partially Oriented NMR Spectroscopy of C-Labeled Acetyl Groups

Fei Yu, Jeremy J. Wolff, I. Jonathan Amster, and James H. Prestegard

J. Am. Chem. Soc., **2007**, 129 (43), 13288-13297• DOI: 10.1021/ja075272h • Publication Date (Web): 09 October 2007



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 10/09/2007

Conformational Preferences of Chondroitin Sulfate Oligomers Using Partially Oriented NMR Spectroscopy of ^{13}C-Labeled Acetyl Groups

Fei Yu,[†] Jeremy J. Wolff,[‡] I. Jonathan Amster,[‡] and James H. Prestegard*[,†]

Contribution from the Complex Carbohydrate Research Center and Department of Chemistry, University of Georgia, Athens, Georgia 30602-4712

Received July 15, 2007; E-mail: jpresteg@ccrc.uga.edu

Abstract: A new method is presented for the retrieval of information on the conformation of glycosaminoglycan oligomers in solution. The method relies on the replacement of acetyl groups in isolated native oligomers with ^{13}C labeled acetyl groups and the extraction of orientational constraints from residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA) offsets observed in NMR spectra of partially oriented samples. A novel method for assignment of resonances based on the correlation of resonance intensities with isotope ratios determined from mass spectrometric analysis is also presented. The combined methods are used in conjunction with more traditional NMR structural data to determine the solution structure of a pentasaccharide, GalNAc6*S*(β1−4)GlcA(β1−3)GalNAc4*S*(β1−4)GlcA(β1−3)GalNAc4S-ol, derived by enzymatic hydrolysis of chondroitin sulfate. The geometry derived is compared to that for similar molecules that have been reported in the literature, and prospects for use of the new types of data in the study of protein-bound oligosaccharides are discussed.

Introduction

Chondroitin sulfates (CSs) are widely distributed in extracellular matrices and at the surfaces of mammalian cells. Here, in addition to their structural role in connective tissues, they play crucial roles in neural development and regeneration, wound healing, infection, growth factor signaling, morphogenesis, and cell division.[1−7] They are also receptors for various pathogens.[8] Despite their ubiquitous nature and important roles, structural characterization of chondroitin sulfates and other glycosaminoglycan (GAG) polymers has been challenging.[9] Here we present a new approach to structural characterization based on a novel strategy for introducing ^{13}C labels into acetyl groups of isolated CS oligomers and observation of orientation dependent NMR parameters from these groups.

Part of the reason for the challenge in obtaining structural data is the heterogeneity of these polymers and their derived oligomers. Chondroitin sulfate is a linear, highly sulfated polysaccharide composed of repeating disaccharide units.[9] Like other members of the glycosaminoglycan (GAG) family, each unit contains an N-acetylated sugar and an acidic sugar; in CS these sugars are galactosamine and glucuronic acid. While such structures would seem to be simple, the galactosamine residues are differentially sulfated at C4 and/or C6 carbons and sometimes the glucuronic acid is sulfated at the 2-position, leading to significant structural diversity.

NMR, in particular proton NMR, has been used to characterize some CS oligomers in the past,[10,11] and this would be the natural choice for characterization in solution. However, proton NMR spectra are very complex, even when structurally well-defined oligomers can be isolated from digests of the normally heterogeneous polymers.[12] Recently, several structures of octasaccharides have been reported based on observations of nuclear Overhauser effects (NOEs), but interpretation is dependent to some extent on computer simulations.[13−15] The complexity of analysis only increases when the conformation of a CS-oligomer in complex with a CS-binding protein is of interest and lines are broadened by this interaction. Resolution of proton spectra then degrades, and use of resonance assignment

[†] Complex Carbohydrate Research Center.
[‡] Department of Chemistry.

(1) Viapiano, M. S.; Matthews, R. T. *Trends Mol. Med.* **2006**, *12*, 488−496.
(2) Mikami, T.; Sugahara, K. *Trends Glycosci. Glycotechnol.* **2006**, *18*, 165−183.
(3) Gowda, C.; Achur, R.; Muthusamy, A.; Takagaki, K. *Trends Glycosci. Glycotechnol.* **2004**, *16*, 407−420.
(4) Fthenou, E.; Zafiropoulos, A.; Tsatsakis, A.; Stathopoulos, A.; Karamanos, N. K.; Tzanakakis, G. N. *Int. J. Biochem. Cell Biol.* **2006**, *38*, 2141−2150.
(5) Shannon, J. M.; McCormick-Shannon, K.; Burhans, M. S.; Shangguan, X. F.; Srivastava, K.; Hyatt, B. A. *Am. J. Physiol.* **2003**, *285*, L1323−L1336.
(6) Wang, T. W.; Sun, J. S.; Wu, H. C.; Tsuang, Y. H.; Wang, W. H.; Lin, F. H. *Biomaterials* **2006**, *27*, 5689−5697.
(7) Matani, P.; Sharrow, M.; Tiemeyer, M. *Front. Biosci.* **2007**, *12*, 3852−3879.
(8) Morugova, T.; Morugova, I. *Diabetologia* **2005**, *48*, A382−A382.
(9) Imberty, A.; Lortat-Jacob, H.; Perez, S. *Carbohydr. Res.* **2007**, *342*, 430−439.
(10) Sugahara, K.; Tanaka, Y.; Yamada, S. *Glycoconjugate J.* **1996**, *13*, 609−619.
(11) Yamada, S.; Yoshida, K.; Sugiura, M.; Sugahara, K. *J. Biochem.* **1992**, *112*, 440−447.
(12) Kinoshita-Toyoda, A.; Yamada, S.; Haslam, S. M.; Khoo, K. H.; Sugiura, M.; Morris, H. R.; Dell, A.; Sugahara, K. *Biochemistry* **2004**, *43*, 11063−11074.
(13) Blanchard, V.; Chevalier, F.; Imberty, A.; Leeflang, B. R.; Basappa; Sugahara, K.; Kamerling, J. P. *Biochemistry* **2007**, *46*, 1167−1175.
(14) Mikhailov, D.; Linhardt, R. J.; Mayo, K. H. *Biochem. J.* **1997**, *328*, 51−61.
(15) Rodriguez-Carvajal, M. A.; Imberty, A.; Perez, S. *Biopolymers* **2003**, *69*, 15−28.

10.1021/ja075272h CCC: $37.00

strategies based on three-bond coupling of ^{1}H spin pairs becomes impractical. ^{13}C NMR spectra, and multidimensional versions of these spectra, offer higher resolution,[16] but without enrichment in ^{13}C, sensitivity is usually inadequate. There have been a few cases of isotope enrichment of GAGs, some based on fully synthetic methods[17] and some based on biosynthetic methods.[18−21] However, synthesis is time-consuming, and separation of the heterogeneous products produced in biosynthetic schemes can be difficult. Also, even when enriched products are produced, parameters directly connected with ^{13}C observation do not offer the wealth of structural information that normally comes from NOEs in ^{1}H spectra.

Here we present a partial solution to the sensitivity and structural information limitations of ^{13}C based approaches. The solution offered relies on introduction of observable ^{13}C labeled sites by replacing native acetyl groups in chondroitin sulfate oligomers, derived from natural products, with ^{13}C-labled acetyl groups. The resulting ^{13}C spectra return structural information through the orientational dependence of residual dipolar couplings (RDCs) and chemical shift anisotropy (CSA) offsets. We illustrate an assignment strategy for resonances in these spectra based on correlation of enrichment levels at various sites, as seen in NMR spectra, and isotope ratios seen in mass spectra of oligomers. Finally, we illustrate a solution structure determination for a simple CS pentasaccharide, GalNAc6*S*(β1−4)-GlcA(β1−3)GalNAc4*S*(β1−4)GlcA(β1−3)GalNAc4S-ol, using a combination of ^{1}H and ^{13}C data. We will refer to this molecule as CS5.

Homogeneous oligomers suitable for structural characterization have been produced from native polymers a number of times before by digestion with hyaluronidase and separation by HPLC.[12,22] We shall follow these procedures in preparation of our starting material. The method we have chosen for the introduction of ^{13}C capitalizes on the presence of N-acetyl groups on every second sugar in chondroitin sulfate. These groups can be removed by either synthetic or biosynthetic methods,[23−25] providing a route to reintroduction of ^{13}C enriched N-acetyl groups. We have previously explored the possibility of introducing ^{13}C labeled acetyls in a dimer of *N*-acetylglucosamine using a carefully controlled base-catalyzed hydrolysis reaction to remove the original N-acetyl group and reaction with ^{13}C acetic anhydride to reacetylate the dimer.[26] However, for chondroitin sulfate, the O-sulfate groups are not stable in a strong base solution,[27] and improved methodology is required. The new methodology is based on hydrazinolysis. Hydrazinolysis of glycosaminoglycans has been used to aid in deaminative cleavage of GAGs, in particular heparan sulfate (HS).[23,28] It was also used to radio isotope label the N-acetyl groups in the degraded oligomers in order to monitor their separation.[23,29] The methodology proved to retain O-sulfation patterns and produce product in high yield.

The introduction of multiple acetyl labels does introduce an NMR resonance assignment problem, particularly if the complexity of proton spectra prevents unambiguous connection of acetyl resonances to resonances of a particular ring in an oligosaccharide. We will take advantage of the fact that hydrazinolysis, while efficient, does occur at different rates for different acetyl sites. If reacetylation with ^{13}C acetyl groups is done without complete hydrazinolysis, each acetyl contains a different $^{12}C/^{13}C$ isotope ratio and these ratios can be used to assign resonances. In NMR spectra the ratios are reflected in the different intensities of acetyl resonances; in MS/MS spectra the ratios are reflected in intensities of mass peaks for ^{12}C and ^{13}C isotopic peaks of the fragment ions. MS/MS of the oligomer of interest provides fragments in a manner that allows placement of various *N*-acetylgalactosamines in the oligomer sequence, and correlation of NMR derived isotope ratios with MS derived isotope ratios allows assignment of the NMR resonances.

^{13}C labeled acetyl groups carry a significant amount of structural information. When ^{13}C is in both the methyl and carbonyl groups, there is a $^{13}C-^{13}C$ residual dipolar coupling that can be measured. Dipolar couplings are orientation dependent through a $(1 - 3\cos^2\theta)$ function where θ is the angle between the magnetic field and the nucleus−nucleus through-space interaction vector. The function averages to zero for molecules that sample orientations uniformly as they tumble in solution, but when partially ordered, the function does not average to zero, and the resulting contributions to splittings of resonances return information on the angle, θ, between the $^{13}C-^{13}C$ bond vector and the magnetic field.[30]

CSA offsets provide additional orientational data. The offsets arise from the fact that chemical shifts are not isotropic parameters but depend on orientation of a molecule in the magnetic field. The carbonyl group has a particularly large shielding anisotropy.[31] Again the anisotropy averages to zero under uniform sampling of orientations, but in partially ordered systems they manifest themselves as offsets of resonances from their isotropic positions.[32] Interestingly, CSA offsets can be written in the form of two RDC-like equations.[33] The vectors used in describing the offsets are at angles of approximately 20° and 70° relative to the C−C bond vector. Thus, CSA offsets are quite complementary to the RDC data.

A great deal of structural information is required to fully determine the conformation of an oligosaccharide in solution or in complex with a protein. If determination was to be done solely on the basis of orientational information from RDCs and CSA offsets, at least five pieces of structural information per rigid entity would be needed. We will have six pieces of data from the RDCs and CSAs of acetyls in the pentasaccharide

(16) Scott, J. E.; Heatley, F. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4850−4855.
(17) Martin-Pastor, M.; Canales-Mayordomo, A.; Jimenez-Barbero, J. *J. Biomol. NMR* **2003**, *26*, 345−353.
(18) Milton, M. J.; Harris, R.; Probert, M. A.; Field, R. A.; Homans, S. W. *Glycobiology* **1998**, *8*, 147−153.
(19) Blundell, C. D.; DeAngelis, P. L.; Day, A. J.; Almond, A. *Glycobiology* **2004**, *14*, 999−1009.
(20) Tracy, B. S.; Avci, F. Y.; Linhardt, R. J.; DeAngelis, P. L. *J. Biol. Chem.* **2007**, *282*, 337−344.
(21) DeAngelis, P. L. *Polymer Biocatalysis and Biomaterials*; 2005; Vol. 900, pp 232−245.
(22) Kinoshita, A.; Yamada, S.; Haslam, S. M.; Morris, H. R.; Dell, A.; Sugahara, K. *J. Biol. Chem.* **1997**, *272*, 19656−19665.
(23) Rabenstein, D. L. *Nat. Prod. Rep.* **2002**, *19*, 312−331.
(24) Shaklee, P. N.; Conrad, H. E. *Biochem. J.* **1984**, *217*, 187−197.
(25) Grobe, K.; Ledin, J.; Ringvall, M.; Holmborn, K.; Forsberg, E.; Esko, J. D.; Kjellen, L. *Biochim. Biophys. Acta* **2002**, *1573*, 209−215.
(26) Yu, F.; Prestegard, J. H. *Biophys. J.* **2006**, *91*, 1952−1959.
(27) Fraidenraich, D.; Waisman; Cirelli, A. F. *Carbohydr. Polym.* **1992**, *17*, 111−114.
(28) Maccarana, M.; Casu, B.; Lindahl, U. *J. Biol. Chem.* **1993**, *268*, 23898−23905.
(29) Vives, R. R.; Pye, D. A.; Salmivirta, M.; Hopwood, J. J.; Lindahl, U.; Gallagher, J. T. *Biochem. J.* **1999**, *339*, 767−773.
(30) Prestegard, J. H.; Bougault, C. M.; Kishore, A. I. *Chem. Rev.* **2004**, *104*, 3519−3540.
(31) Burton, R. A.; Tjandra, N. *J. Am. Chem. Soc.* **2007**, *129*, 1321−1326.
(32) Lipsitz, R. S.; Tjandra, N. *J. Am. Chem. Soc.* **2001**, *123*, 11065−11066.
(33) Choy, W. Y.; Tollinger, M.; Mueller, G. A.; Kay, L. E. *J. Biomol. NMR* **2001**, *21*, 31−40.

studied here, (GalNAc6*S*(β1−4)GlcA(β1−3)GalNAc4*S*(β1−4)-GlcA(β1−3)GalNAc4S-ol), but these are distributed over three residues. Two of these are rigid to a reasonable approximation, but even these are connected by flexible glycosidic linkages. In applications to protein bound CS oligomers, the plan would be to supplement acetyl data with principal order parameters determined from the protein and with oligomer structural constraints derived from modeling. However, in the test case presented here (modeling of the preferred conformation in solution), we will supplement the acetyl data with a number of $^{13}C-^{1}H$ RDCs from ring carbons, $^{1}H-^{1}H$ RDCs within sugar rings, NOE distance constraints from $^{1}H-^{1}H$ trans-glycosidic pairs, and *J* coupling constraints from H2 to HN. The structure determined, using an iterative calculation of alignment parameters from an assumed structure and optimization of the structure using a simulated annealing protocol, gives an opportunity to understand the solution conformational preferences of CS5.

Materials and Methods

Materials. The sodium salt of chondroitin sulfate A (~70% of the residues have A type sulfation, and the balance have C type sulfation) from bovine trachea, hyaluronidase from sheep testes (type V), anhydrous hydrazine, hydrazine sulfate, iodic acid, hydroiodic acid, Sephadex G-15 resin, and pentaethylene glycol monododecyl ether (E_5C_{12}) were purchased from Sigma-Aldrich Co (St. Louis MO). Acetic anhydride-$^{13}C_4$ was purchased from Cambridge Isotope Laboratories (Andover, MA), and iatrobeads were purchased from Iatron Laboratories (Tokyo, Japan). A prepacked strong anion exchange (SAX) column was purchased from VWR Scientific (Rochester, NY).

Preparation of Chondroitin Sulfate Oligomers. Preparation of chondroitin sulfate oligomers was adapted from a literature procedure.[22] Chondroitin sulfate A (1 g) was digested with 100 mg of hyaluronidase (451units/mg) in 10 mL 50 mM phosphate buffer, pH 6, containing 150 mM NaCl at 37 °C for 48 h. The digest was crudely separated into fractions I−III on an iotrabeads column with H_2O/ACN as an elution solvent. Fraction II, which was confirmed by MS to mostly consist of tetramer and pentamer, was treated with 0.1 mL of $NaBH_4$ in H_2O to reduce the terminal sugars to the corresponding galactitol form, and then fraction II was subfractionated on a strong anion exchange (SAX) HPLC column (4.6 mm × 250 mm) using a linear NaCl gradient from 0 to 700 mM over a 50 min period at a flow rate of 3.0 mL/min. The separation was monitored by absorption at 215 nm. Fractions V and VI (elution times of 85 min and 87 min) were collected, desalted on a Sephadex G-15 column (1.5 cm × 100 cm), and lyophilized.

Fraction VI was analyzed by mass spectrometry and NMR. It proved to be a pentamer of mass 1221.20. This is consistent with the following molecule: GalNAc6*S*(β1−4)GlcA(β1−3)GalNAc4*S*(β1−4)GlcA(β1−3)GalNAc4S-ol, denoted CS5. Terminal residues and positions of sulfates were determined by NMR analysis as described below.

Isotopic Labeling of CS Oligomers. Isotope labeling at the galactosamine acetyls relied on the selective removal of N-acetyl groups using a procedure adapted from Shaklee et al.,[24] followed by re-acetlyation with ^{13}C-labeled acetic anhydride. The reduced CS5 pentamer (5 mg) was dissolved in 1 mL of anhydrous hydrazine containing 1% (w/v) hydrazine sulfate. The solution was sealed in a 1 mL vial and heated at 90 °C for 5 h. After that, the hydrazine was evaporated under vacuum overnight, and the dry sample was dissolved in 3 mL of H_2O and moved to a 15 mL separatory funnel where 500 μL of 0.2 M HIO_3 solution in H_2O were added slowly at room temperature to convert the uronic acid hydrazide back to a uronic acid residue. Diethyl ether (5 mL) was added to extract I_2 formed. The excess HIO_3 was then reduced by adding 0.2 M HI/H_2O in 10 μL increments until no more I_2 formed in the aqueous layer. The ether layer was removed, and 2 mL more of diethyl ether were added to extract any remaining I_2. The aqueous solution was collected, and the diethyl ether layers were washed with two 1 mL portions of water to retrieve any additional deacetylated product. All the aqueous solutions were combined and lyophilized.

The deacetylated sample was dissolved in 1 mL of water, and Na_2CO_3 was added until the pH reached 11. Acetic anhydride-$^{13}C_4$ (10 μL) was added at 0 °C, and the solution was stirred for 2 h. The product was purified on a Sephadex G-15 column (1.5 cm × 100 cm) monitored by absorption at 215 nm.

Preparation of NMR Samples. All the NMR samples were buffered to the same pH value of 7 with 20 mM phosphate. For extraction of residual dipolar couplings and chemical shift anisotropy offsets from ^{13}C spectra of the N-acetyl group, a sample in an anisotropic medium was prepared by dissolving isotopically labeled CS5 in 500 μL of 10% (w/w) alkylethyleneglycol detergent, E_5C_{12}, and 20 μL of hexanol in aqueous phosphate buffer. Tetramethylammonium bromide (5 mg) was added as a chemical shift reference. The isotropic reference sample was prepared by simply adding extra hexanol (2 μL) to the anisotropic sample to transform the aligned medium into an isotropic one. For the measurement of $^{13}C-^{1}H$ RDCs and $^{1}H-^{1}H$ RDCs from other parts of the sugar residues, an anisotropic sample was prepared by dissolving nonlabeled CS5 in a lower concentration of E_5C_{12} medium (4%). Another 5 mM unlabeled sample in D_2O was made for the isotropic measurement of the same $^{13}C-^{1}H$ and $^{1}H-^{1}H$ couplings. For 1D HN−H2 *J* coupling measurements an additional 5 mM unlabeled sample was prepared in 10% D_2O/H_2O.

NMR Spectroscopy. All NMR experiments were recorded at 25 °C on a Varian Inova spectrometer operating at 800 MHz for ^{1}H and 200 MHz for ^{13}C. For ^{13}C spectra, proton homonuclear decoupling was achieved using WALTZ-16 with a power level of 30 dB. The recycling delay, d1, was set to as long as 30 s (about twice the T1s of carbonyls measured in an analogous hexamer) to allow the magnetization to fully relax before each scan. 1D spectra were recorded with a spectral width of 50 kHz and an acquisition time of 1 s. The FIDs accumulated over 3 h were apodized with a Gaussian weighting function (0.5) before Fourier transformation.

Heteronuclear single quantum coherence (HSQC) spectra were acquired without proton decoupling in the indirect dimension at spectral widths of 5000 Hz for the direct proton dimension and 8000 Hz for the indirect carbon dimension using 80 scans per t1 increment to achieve a time domain matrix of 4096 × 512 complex points. This was linear predicted to 4096 × 800, apodized in both dimensions with a 90° shifted sinebell, and zero filled to 4096 × 2048. The couplings were extracted directly from the indirect frequency domain.

The $^{1}H-^{1}H$ homonuclear couplings were measured by using the magnitude form of the constant time COSY (CT-COSY) experiment.[31] The data were acquired using the same sequence as the phase sensitive CT-COSY except processed in a magnitude mode ($spectra = \sqrt{Re^2+Im^2}$). Spectra were acquired with a spectral width of 1800 Hz for both direct and indirect proton dimensions, 8 scans per t1 increment, 1024 × 64 complex points. Spectra were then linear predicted to 1024 × 128, apodized in both dimensions with a 90° shifted sinebell, and zero filled to 1024 × 256. The constant time delays were 40−340 ms with a 20 ms increment. The curve fitting of data as a function of delay was performed using the curve fitting tool in MATLAB.

The NOESY spectra used to supplement structural data were acquired with a spectral width of 1800 Hz for both direct and indirect proton dimensions, 64 scans per t1 increment, 512 × 400 complex points. These were linear predicted to 1024 × 512, apodized in both dimensions with a 90° shifted squared sinebell, and zero filled to 1024 × 1024.

The 1D proton spectra for HN−H2 *J* coupling measurements were acquired using a Watergate sequence with a spectral width of 12 kHz and an acquisition time of 1 s. Assignments of proton resonances of ring protons were based on standard COSY, TOCSY, and NOE spectra.

Mass Spectrometry (MS). Experiments were performed with a 7 T Bruker Apex IV QeFTMS (Billerica, MA) fitted with an Apollo II

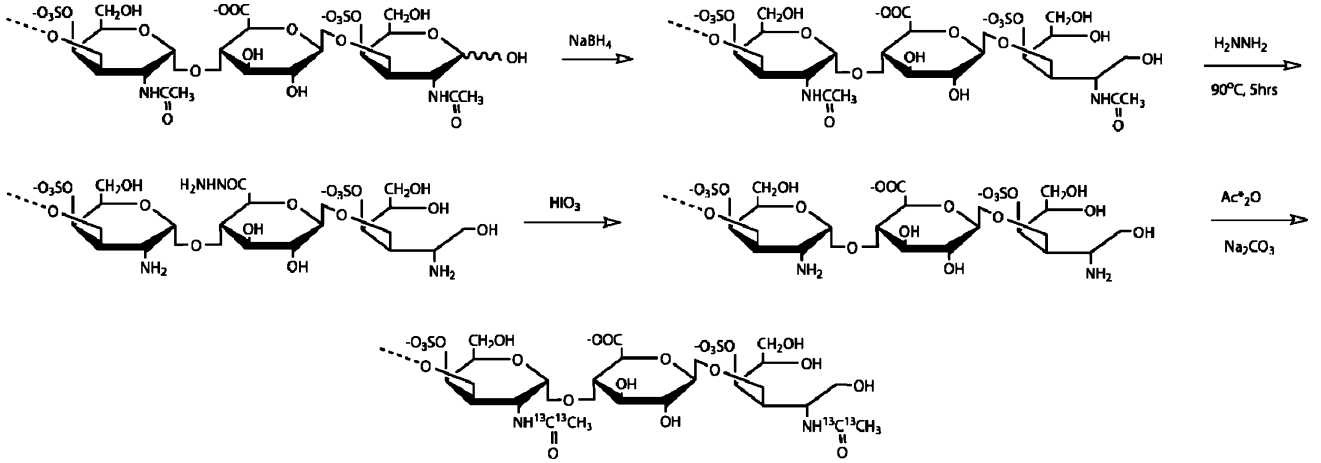


Figure 1. ^{13}C enrichment strategy for isolated CS oligomers.

ESI source and a CO_2 laser for infrared multiphoton dissociation (IRMPD). Solutions of each oligossaccharide were made at a concentration of 0.1 mg/mL in 50:50 methanol/H_2O and ionized by nanospray using a pulled fused silica tip (model no. FS360-75-15-D-5, New Objective, Woburn, MA). The sample solutions were infused at a rate of 10 μL/h. All oligosaccharides were examined in negative ion mode. For MS/MS experiments, precursor ions were isolated in the external quadrupole and then transferred to the ICR analyzer cell. The isolation/cell fill was repeated up to 6 times per acquisition. The selection of the precursor ion was further refined by using in-cell isolation with a coherent harmonic excitation frequency (CHEF) event.[34] Precursor ions were then fragmented by IRMPD inside the analyzer cell. A total of 24 acquisitions were signal averaged per mass spectrum. For each mass spectrum, 512K points were acquired, padded with one zero fill, and apodized using a sinebell window. All fragments are reported using the Domon and Costello nomenclature.[35]

Structure Calculation. Structure determination was accomplished using a combination of the program REDCAT (residual dipolar coupling analysis tool) to estimate alignment parameters[36] and XPLOR-NIH to optimize the structure using a simulated annealing protocol.[37,38] REDCAT was originally designed for the analysis of RDCs. Coordinates for pairs of atoms (taken from a trial structure) are entered along with a maximum coupling calculated from nuclear moments separated by 1A (−60 400 for a C−H pair, −15 200 for a C−C pair, and −240 200 for a H−H pair) and an experimental RDC. Error limits are also entered and are usually set to 10% of the range of measured couplings, a number that adequately accounts for structural variations in geometry that are not well modeled by common force fields. The program returns a large set of allowed order tensor solutions, a best solution that can be interpreted in terms of principal elements describing the level and asymmetry of order, and Euler angles describing the orientation of the principal frame relative to the original molecular frame. It can also back-calculate RDCs for a given structure and order parameter set.

The program was not originally intended to handle CSA offsets. However, based on the similarity of equations relating observables to orientational constraints for RDCs and CSA offsets, it is possible to use REDCAT for the analysis of CSA offsets as well. The dependence of dipolar coupling on orientation can also be described by eq 1.

$$D = D_{\max}(r)^{-3} \sum_{ij} S_{ij} \cos(\theta_i) \cos(\theta_j) \tag{1}$$

Here $D_{\max}$ is a constant that depends on the properties of nuclei k and l. The S_{ij} are elements of an order matrix written in an arbitrary molecular frame, and the $\cos(\theta_i)$ are angles of the internuclear vector relative to molecular axes i,j.

The dependence of the CSA offset on orientation is described by eq 2.

$$\delta_{\mathrm{CSA}} = (^2/_3) \sum_{ijk} S_{ij} \cos(\theta_{ik}) \cos(\theta_{jk}) \delta_{kk} = (^1/_3) \begin{bmatrix} 2\delta_{xx} \sum_{ij} S_{ij} \cos(\theta_{ix}) \cos(\theta_{jx}) \\ + 2\delta_{yy} \sum_{ij} S_{ij} \cos(\theta_{iy}) \cos(\theta_{jy}) \\ + 2\delta_{zz} \sum_{ij} S_{ij} \cos(\theta_{iz}) \cos(\theta_{jz}) \end{bmatrix} \tag{2}$$

Here the δ_{kk} are principal elements of the chemical shift tensor, and the $\cos(\theta_{ik})$ are angles of the CSA principal frame axes k relative to molecular frame axes i,j. Comparing the expression for RDCs and CSAs, a similar $\cos(\theta)$ dependent form is observed, and the three terms in eq 2 can be expressed as three pseudo RDC entries that result in a measured average offset. The $D_{\max}$ entered for the RDCs is simply replaced by twice the chemical shift anisotropy component, and the CSA principal tensor vectors replace coordinates for RDC internuclear vectors. The CSA tensor values used, which originated with data on peptides systems, were converted to Hz at 200 Hz/ppm (−29 600 Hz, −400 Hz, 30 000 Hz)[39] and assumed to be the same for all acetyl groups (in proteins, variations in tensor values can reach 20% of the range[31]). Using the averaging tool available in the REDCAT program, an entry of "AVG" is placed after the first two pseudo RDC values and the observed chemical shift offset in Hz is input for the last value. The back-calculated CSA offset (using the back-calculation tool in REDCAT) is equal to the average value of the three pseudo RDCs.

One additional subtly in application is that scaling of both the entry and $D_{\max}$ can be used to vary the weight of various types of data in the REDCAT calculation. This is particularly important when experimental data have very different magnitudes or have differences in experimental precision. In our case $^{13}C-^{13}C$ RDCs were scaled up by a factor of 10

(34) Heck, A. J. R.; Dekoning, L. J.; Pinkse, F. A.; Nibbering, N. M. M. *Rapid Commun. Mass Spectrom.* **1991**, *5*, 406−414.
(35) Domon, B.; Costello, C. E. *Glycoconjugate J.* **1988**, *5*, 397−409.
(36) Valafar, H.; Prestegard, J. H. *J. Magn. Reson.* **2004**, *167*, 228−241.
(37) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *J. Magn. Reson.* **2003**, *160*, 65−73.
(38) Schwieters, C. D.; Kuszewski, J. J.; Clore, G. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 47−62.
(39) Oas, T. G.; Hartzell, C. J.; Dahlquist, F. W.; Drobny, G. P. *J. Am. Chem. Soc.* **1987**, *109*, 5962−5966.

to achieve similar weighting of different types of data. The observed CSAs and chemical shift tensor elements δ_{ii} were not scaled as the conversion to Hz put them in an appropriate range.

REDCAT calculations were carried out for a number of structures selected from low energy conformers identified in the literature. In each case an rmsd of back calculated RDCs and CSAs relative to experimental data was returned. The structure with minimum rmsd was chosen as the starting model in the following XPLOR-NIH refinement, and the order parameters determined for that structure were converted to the anisotropy (D_a) and Rhombicity (R_h) parameters typically used in XPLOR-NIH. They are related to the REDCAT order parameters as follows: $D_a = D_{max} * {}^1/_2 S_{zz}$ and $R_h = {}^2/_3\eta$.

XPLOR-NIH can accomplish simulated annealing under restraints from experimental data including RDCs, CSA offsets, NOE derived distances, and NH−H2 *J* couplings. A modification was made to the topology file for carbohydrates[40] to allow application to a chondroitin sulfate oligomer. Additional force field parameters needed were obtained from the GLYCAM website (http://glycam.ccrc.uga.edu/AMBER/index.html). The starting model from the above REDCAT calculation was heated to temperatures of 3500 K and slowly cooled in steps of 25 K to 300 K. The process was repeated 2000 times, and 10 structures with the lowest energies were selected and analyzed in VMD.[41] Again, RDCs and CSAs were back-calculated using REDCAT on XPLOR refined structures. A new set of alignment parameters was calculated, and simulated annealing in XPLOR-NIH was repeated.

Results

Selection of the Oligomer. The particular chondroitin sulfate oligomer to be studied here was selected based on its modest size and modest degree of structural heterogeneity. Long oligomers may well provide too much of a challenge in resolution and assignment of labeled acetyl groups, and too short an oligomer would limit utility as a basis for future studies of protein−condroitin sulfate interactions. The primary structure of the oligomer, GalNAc6*S*(β1−4)GlcA(β1−3)GalNAc4*S*(β1−4)GlcA(β1−3)GalNAc4S-ol was deduced from the NMR resonances found in $^{13}C-^1H$ HSQC spectra of the native material. The residues were classified as glucuronic or *N*-acetylgalactosamine by patterns in TOCSY/COSY spectra, and the sequence of residues was established by trans-glycosidic NOE patterns. The sulfation sites are easily identified by a characteristic downfield shift (~0.5 ppm) of the H4 resonance for C4 sulfation and the H6 resonances for C6 sulfation. Considering the inhomogeniety of the polymer precursor (~30% chondroitin C type sulfation), the occurrence of some 6-sulfation is not unexpected. A pentamer is, however, unanticipated as hyaluronidase is expected to have specificity for hydrolysis after an *N*-acetylgalactosamine residue leading to a series of oligomers with an even number of residues. We note that under long periods of hydrolysis the fraction of odd oligomers rises, suggesting that production of odd numbered oligomers may be due to the presence of minor amounts of exosidic enzymes. We used the pentamer simply because we could more easily isolate a pure preparation.

Isotopic Labeling. Our strategy for the introduction of ^{13}C-labeled acetyl groups to the chondroitin sulfate pentamer (CS5) involves two steps: deacetylation and reacetylation. The deacetylation strategy, which is shown in Figure 1, was based on the hydrazinolysis mechanism. Hydrazinolysis has proven to be a high yield reaction that leaves the O-sulfate and carboxyl groups that characterize glycosaminoglycans intact.[24] However, two side reactions possible in the hydrazinolysis of GAGs are of some concern. One is the reaction between hydrazine and the aldehyde exposed at the reducing end of an oligosaccharide on opening of the pyranose ring.[42] To prevent this side reaction, the sugar was reduced to the corresponding galactol before the hydrazinolysis reaction. Leaving this as a galactol in the final product also helps reduce the complexity of the NMR spectra due to elimination of the α/β anomeric equilibrium at the reducing end of the oligomer. The other side reaction results in the formation of a hydrazide derivative of the carboxyl moieties in the glucuronic acid residues.[24] To remove products of this second reaction the initial hydrazinolysis product was treated with HIO_3 to convert the uronic acid hydrazides back into uronic acid residues.

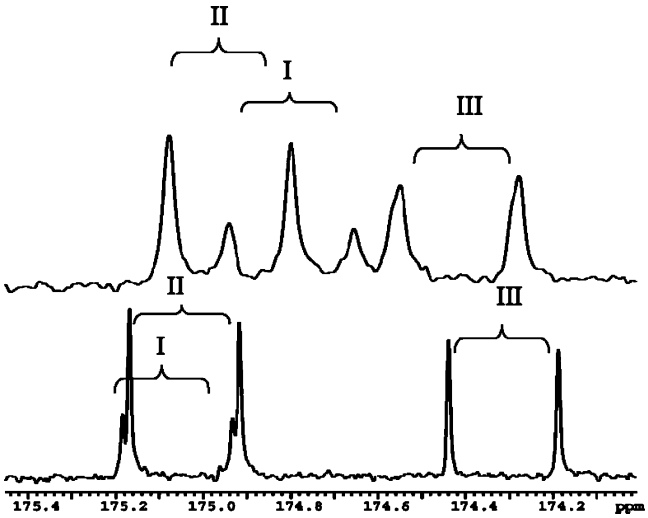


Figure 2. Carbonyl region of ^{13}C spectra for an isotopically labeled CS pentamer in 10% $C_{12}E_5$/hexanol/water. The bottom one corresponds to the spectra acquired for sample in isotropic phase, and the top one corresponds to the spectra acquired for the sample in anisotropic phase (with extra hexanol). Three doublets with different intensities represent three acetyl groups with different isotopic ratios.

Conversion of the amino sugar products back to ^{13}C-labeled N-acetyl sugars using ^{13}C-acteic anhydride was nearly quantitative. To facilitate assignment, however, it was desirable to have different $^{13}C/^{12}C$ isotopic ratios for each N-acetyl group. To accomplish this, we choose to partly remove the native N-acetyl groups at the hydrazinolysis stage, leaving different amounts of ^{12}C-actetyls. Due to the chemical environment difference for each N-acetyl group, the deacetylation ratio is different. Here we optimized the hydrazinolysis reaction conditions (90 °C for 5 h) to achieve deacetylation ratios ranging from 90% to 30%. Deacetylation was followed by fully reacetylating the amino sugars with 99% ^{13}C-acetic anhydride. The overall recovery of CS was ~50%.

$^{13}C-^{13}C$ RDCs and Carbonyl Carbon CSA Offsets. The carbonyl regions of the ^{13}C NMR spectra for the CS5 sample in an isotropic phase and anisotropic phase of an aqueous solution of E_5C_{12} are presented in Figure 2. Both spectra were acquired with the resonance from tetramethyl ammonium bromide as the chemical shift reference. Three doublets with 50−60 Hz splittings are observed in both spectra. These

(40) Weis, W. I.; Brunger, A. T.; Skehel, J. J.; Wiley, D. C. *J. Mol. Biol.* **1990**, *212*, 737−61.
(41) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.
(42) Tipson, R. S. *J. Org. Chem.* **1962**, *27*, 2272-&.

Table 1. Observed and Unified $^{13}C-^{13}C$ RDCs and Carbonyl Carbon CSA Offsets

	observed (10% C12E5)		unified (4% C12E5)	
doublet/residue	CSA (Hz)	RDC (Hz)	CSA (Hz)	RDC (Hz)
II/A	−20.5	4.8	−1.6	0.39
I/C	−54.6	7.3	−4.4	0.58
III/E	18.2	4.9	1.5	0.39

correspond to the three N-acetyl groups with splittings arising from the coupling to ^{13}C's in the directly bonded acetyl methyl groups. In comparing spectra it is clear that the doublets in the aligned spectrum (2b) have both shifted and changed splittings. The doublet labeled I is shifted upfield by 54.6 Hz, the doublet labeled II is shifted upfield by 20.5 Hz, and the doublet labeled III is shifted downfield by 18.2 Hz. The splittings of the doublets have changed by 7.3 Hz for doublet I, 4.8 Hz for doublet II, and 4.9 Hz for doublet III. The change in splitting arises from a residual dipolar coupling, and the change in chemical shift arises from a chemical shift anisotropy offset. The changes are typical of the effects of incomplete averaging of anisotropic parameters in aligned media. Changes can also be seen in the acetyl methyl region of the spectrum (not shown). However, the changes in splittings are identical to those measured from the carbonyl resonances, and the changes in chemical shift are negligible because of the small chemical shift anisotropy of the methyl carbons. The data, as summarized in Table 1, clearly contain structural information on the average orientation of each acetyl group, but this information is not useful without assignment of doublets to specific acetyls.

Assignment of ^{13}C Spectra. Assignments of carbonyl resonances would normally be undertaken by transfer of carbonyl ^{13}C magnetization to the amide proton two bonds away using an INEPT sequence and then transfer to other sugar ring protons using TOCSY techniques. Since all sugars containing acetyls are *N*-acetylglactosamines, they would need to be distinguished by sequentially linking them with NOE or HMBC transfers across the glycosidic linkages. While this would likely work for simple solution samples, the small coupling between the carbonyl carbon and the amide proton makes this a difficult procedure and one that is not likely to work for the broader lines encountered in protein complexes, or the broader lines encountered when CS5 is aligned in a E_5C_{12} medium.

Here we use a new assignment strategy based on the different level of isotopic labeling for each N-acetyl. This strategy was introduced previously in our work on an *N*-acetylglucosamine disaccharide.[26] In the ^{13}C spectra of Figure 2, the difference in the intensity of the doublets reflects the different extent of ^{13}C label introduction. Integration of the peaks shows the relative extent of ^{13}C label introduction for the three N-acetyl groups I/II/III to be 3:10:8. To decide which doublet belongs to which N-acetyl group, we correlate these intensity ratios with mass ratios found for each sugar residue in MS/MS spectra of the CS5 oligomer.

Expansions of the mass regions of MS spectra corresponding to B_2, Y_1, and Y_3 ions of CS5 are shown in Figure 3. These were collected in a negative ion mode on an FT/MS spectrometer after fragmentation using IRMPD. The B_2 and Y_1 ions correspond to fragments containing just a single *N*-acetylglactosamine residue. In these regions two sets of mass peaks were observed. The monoisotopic peaks at lower mass, *M*, represents ^{12}C containing products. The peaks two mass units higher, $M + 2$, primarily represent those molecules containing a single acetyl group with two ^{13}C atoms. Their ratio is, in principle, correlated with the NMR intensity for a particular doublet seen in Figure 2. It must be noted, however, that the natural abundance ^{13}C isotopes from all nonlabeled carbons also contribute to the $M + 2$ peaks. Therefore, to extract an accurate isotopic labeling ratio, the contribution from natural abundance ^{13}C was simulated and subtracted from the peaks labeled $M + 2$. The isotopic labeling ratio was then calculated as $R = I'_{M+2}/(I_M + I'_{M+2})$, where I'_{M+2} represents the intensity of the corrected $M + 2$ peaks and I_M represents the intensity of *M* peaks. The calculated results suggest the isotopic ratio for the N-acetyl groups in residue A (B_2 ion) and residue E (Y_1 ion) are 61% and 72%, respectively.

Deducing the ^{13}C content of residue C is more complicated. The Y_3 product ion reveals the presence of two partially labeled N-acetyl groups. Here we observe three sets of peaks, *M*, $M + 2$, and $M + 4$. The peak *M* represents the molecules with two ^{12}C-acetyl groups. Peak $M + 2$ represents the Y_3 product with only one ^{13}C-acetyl group (in either residue C or residue D). Peak $M + 4$ represents the product with two ^{13}C-acetyl groups. The natural abundance refinement was also performed to improve the accuracy of the isotopic ratio calculation. Since we already knew the isotopic ratio for residue E to be 72%, the isotopic ratio for residue C can be calculated from any pair among the *M*, $M + 2$, and $M + 4$ peaks. Here we performed the calculation through these three combinations, *M* and $M + 2$; $M + 2$ and $M + 4$; and *M* and $M + 4$. The isotopic ratios from a fit of the data are 29%, 32%, and 33%, or an average of 31% for residue C. In summary, the isotopic labeling ratios for the three N-acetyl groups were calculated as 72%, 31%, and 61%. Correlating this information with the intensity ratios seen in the ^{13}C spectra, we can easily make the assignments shown in Table 1. For the anisotropic spectra, the assignment was performed in a completely analogous way.

Complementary Structural Information. While the ^{13}C labeled N-acetyl groups provide useful structural information with high sensitivity and resolution, to independently determine the molecular conformation of CS5 in solution, at least five pieces of RDC and CSA data would be required for each rigid unit.[43] This is less of a problem in the case of structural characterization of a protein-associated carbohydrate because protein bound carbohydrates share the order parameters S_{zz} and η with the protein, and these can usually be obtained from protein spectra. This reduces the number of pieces of structural information needed for each unit to three, and any restriction on glycosidic torsion angles from modeling the interaction will reduce the number further. Here, since we are going to probe the structure of CS5 in free solution, some other structural information, including $^{13}C-^{1}H$ RDCs, $^{1}H-^{1}H$ RDCs, and NOEs are employed as complementary data.

The one-bond $^{13}C-^{1}H$ RDCs were obtained from a series of sensitivity enhanced constant time HSQC spectra (SE-CT-HSQC) taken under partially oriented and isotropic conditions. These RDCs are listed in Table 2. For a sample in the 4% E_5C_{12} medium, values ranging from +9.3 to −10.0 Hz are observed.

(43) Prestegard, J. H.; Al-Hashimi, H. M.; Tolman, J. R. *Q. Rev. Biophys.* **2000**, *33*, 371−424.

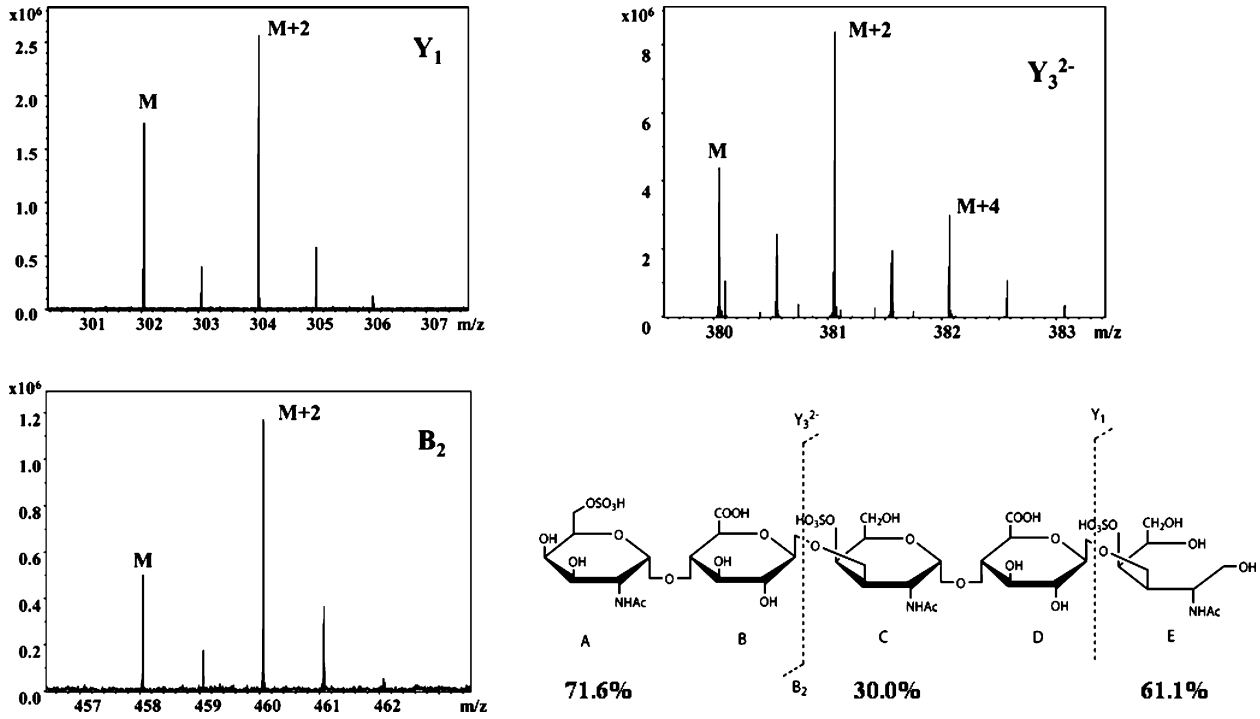


Figure 3. MS/MS spectra of Y_1, Y_3^{2-}, and B_2 ions. The different isotopic ratio for each N-acetylated residue is reflected in intensities of peaks $M/M + 2/M + 4$. After fitting the data, the isotopic labeling ratios calculated for the three N-acetyl groups A, C, and E were 72%, 31%, and 61%.

Table 2. Complementary Structural Information

Vector	RDC (Hz)	Vector	RDC (Hz)
Residue A		Residue D	
C1H1	6.1	C1H1	7.9
C2H2	5.8	C2H2	7.7
C4H4	-10.0	C3H3	9.3
C5H5	7.3	C4H4	8.1
H1H2	0.6	H1H2	1.5
Residue B			
C1H1	4.4	CH3(A)	-1.3
C2H2	5.0	CH3(C)	-2.1
C3H3	4.8	CH3(E)	-1.3
C4H4	4.2		
H1H2	-0.2		NOE(Å)
Residue C		H1(A)-H4(B)	2.20
C1H1	7.9	H1(B)-H3(C)	2.47
C2H2	5.7	H1(C)-H4(D)	2.21
C3H3	7.1		
C4H4	-6.5	Residue	JH2-HN(Hz)
H1H2	1.8	A	9.7
H3H4	0.8	C	9.5
		E	8.5

The spread of the data reflect the different orientations of each sugar residue in the oligomer. For the glucuronic acid, the RDCs for C1−H1, C2−H2, C3−H3, and C4−H4 are of similar size and sign because of the nearly parallel orientation of these vectors. The same thing is true for the galactosamine residue except for the RDC of the equatorial C4−H4 vector. While the data are of high quality, the near degeneracy in vectors still makes it difficult to get five independent RDC measurements.

The vectors connecting proximate protons of the sugar rings are not typically parallel to any of the one-bond C−H vectors and represent additional independent pieces of data. The homonuclear $^1H-^1H$ coupling can be accurately measured by a simple magnitude form of the constant time COSY (CT−COSY) experiment. Here several spectra with various constant time delays were recorded, and data were fit to an equation relating the intensity ratios of cross-peaks and auto-peaks to a function of the constant time delay (t).

$$\frac{I_{\text{cross}}}{I_{\text{auto}}} = |A \tan(\pi(J + D)t)| \qquad (3)$$

The extracted $^1H-^1H$ RDCs are also listed in Table 2. Due to the interference from the E_5C_{12} signal, only the coupling for H1−H2 of each residue and H3−H4 of residue C are reported. The signs of couplings cannot be determined from the CT-COSY experiments unless large scalar couplings of known sign exist for the pair of protons. Here, the magnitude of the scalar couplings for H1−H2 and H3−H4 are more than 8 Hz, which is large enough to resolve issues of sign ambiguity.

Besides the RDC data, other complementary structural data included NOE data on the distance between proton pairs across the glycosidic linkages and the H2−HN scalar coupling between the H2 ring proton of galactosamine residues and the amide proton of the acetyl amide bond. The glycosidic NOE distances function as additional constraints on the glycosidic torsion angles, and the H2−HN couplings function as constraints on the relative orientation of N-acetyl groups to the galactosamine rings. The NOE distance constraints are shown in Table 2.

The configurations of the C2−N torsion angles in the *N*-acetylgalactosamine residues are particularly critical to our use of data from acetyl groups in our structure determination. These torsions can be constrained using scalar couplings from HN to H2 of the N-acetylgalactosamine residues. The three H2-HN J couplings measured for residue A, C and E are 9.7 Hz, 9.5 Hz and 8.5 Hz respectively (shown in Table 2). The relative smaller value for residue E compared with the other two might be due to the loss of ring structure for this residue. Using the Karplus equation with coefficients from Wang and Bax,[44] the

(44) Wang, A. C.; Bax, A. *J. Am. Chem. Soc.* **1996**, *118*, 2483−2494.

calculated torsion angle for the H−N−C2−H2 angle for residue A and C would be 169° and 164°, values close to the trans configuration found in previous MD calculations.[13] The values were thus used to justify selection of a trans orientation for acetyls in all starting structures, and an average value of 9.6 Hz was used to restrain the A and C N-acetyl groups in refinement steps.

Merging CSA and RDC Data. Because of resolution problems in observing proton spectra, the CSA offset and $^{13}C-^{13}C$ RDC data were acquired under different conditions than those for $^{13}C-^{1}H$ and $^{1}H-^{1}H$ RDC data, namely 10% and 4% E_5C_{12}. To perform the structural analysis using all the information simultaneously, the two sets were unified, assuming that the only difference would be a scaling of the measured coupling and CSA offsets. In an acetyl group, the dipolar coupling between the methyl protons and the methyl carbon (D_{CH_3}) actually is a redundant measure of the orientation of the C−C bond obtained from the carbonyl carbon−methyl carbon RDC (D_{CC}). This is a result of the fast rotation of the methyl group about the C−C axis. Therefore, the D_{CH_3} can be expressed in terms of D_{CC} according to the equation $^{1}D_{CC} = {}^{1}D_{CH_3}(-3\gamma_C/\gamma_H)(r_{CH}^3/r_{CC}^3)$.[30] Based on this relationship, the three D_{CH_3} values observed in 4% E_5C_{12} were transformed to three D_{CC} values as 0.38, 0.61, and 0.38 Hz. When compared with the corresponding D_{CC} observed in 10% E_5C_{12}, the scale factors obtained were 12.6, 12.0, and 12.9, respectively. An average of 12.5 was then used to scale all 10% E_5C_{12} orientational data to 4% E_5C_{12} conditions. The unified RDCs and CSAs are shown in Table 1.

Structure Determination. One of the most widely practiced approaches to the utilization of a variety of experimental data in structure determination is to represent those data in terms of pseudo-energy functions added to the molecular energy representations of molecular simulation programs. A search can then be done for a minimum energy structure using simulated annealing or Monte Carlo methods. The XPLOR-NIH program has been developed to accommodate a variety of NMR data, including RDCs and CSA offsets.[32,45] However, in order to utilize the RDC and CSA data in XPLOR-NIH, an input of anisotropy and rhombicity parameters for the molecular order is required. In applications to proteins, values for these parameters are normally obtained from analysis of the distributions of large numbers of data. However, with the limited number of data available for oligosaccharides, parameter estimation from distributions is not reliable. Therefore we adopted a hybrid approach in which a second program, REDCAT (residual dipolar coupling analysis tool), is used to estimate alignment parameters using an initial structure, and then XPLOR-NIH and REDCAT are used iteratively to arrive at a refined structure. REDCAT calculates the best least-squares fit of an order matrix to experimental data, given an initial structure. It too can handle RDCs and CSAs in evaluations of structural models. The order matrix REDCAT returns is transformed to a principal frame, and the principal order and asymmetry parameters, S_{zz} and η (simply related to anisotropy and rhombicity), are extracted. REDCAT also provides a very convenient way to evaluate a structure by back-calculating RDCs and CSAs for comparison to experimental data.

(45) Tjandra, N.; Marquardt, J.; Clore, G. M. *J. Magn. Reson.* **2000**, *142*, 393−396.

Table 3. The Glycosidic Torsion Angles φ/ψ Sampled for the Structural Approximation by REDCAT

	α	β	γ	δ	ϵ
A−B β(1−4)	260/190	280/250	280/70	60/240	80/270
B−C β(1−3)	280/90	280/130	60/110	270/280	
C−D β(1−4)	260/210	280/250	270/60	60/240	80/270

To begin the process, reasonable initial structures must be selected. Recently J. P. Kamerling and co-workers carried out a combined NMR/computational study on a set of eight CS oligomers.[13] This work does not include the molecule studied here, but the set does include similar linkages. Based on energy diagrams for the common linkages, we can select reasonable starting conformers for our analysis. These diagrams typically have several local minima. Table 3 shows the torsion angles for the local minima (α−ϵ) of the first three glycosidic bonds of our pentamer (the fourth, to the reduced sugar was not considered because of the expected flexibility of the open ring structure). Structures were initially generated using the GLYCAM utilites,[46] and conformations were adjusted to the required torsions using PYMOL.[47] The error limits for data used in REDCAT were initially adjusted to values large enough to allow order matrix solutions for each conformation (typically the error limit for $^{13}C-^{13}C$ RDCs and C═O CSAs were set to 5 Hz), and the best set of order parameters was used to calculate an rmsd relative to experimental data. The same procedure was repeated for the random structural sampling around the minimum rmsd region. This time, the errors limits were reduced (typically 2.5 Hz for $^{13}C-^{13}C$ RDCs and CSAs). All combinations for the first two glycosidic bonds were evaluated. The best conformation for the first two glycosidic bonds was then chosen, and all possibilities for the third torsion were explored in combination with this conformation. The results showed the structure with torsion angles in the region β-α-β to achieve a minimum rmsd. The final structure (structure 1) has glycosidic torsion angles of 270°/250°, 305°/105°, and 270°/250° and an rmsd of 0.96 Hz. The final numbers we obtained for the order parameters S_{zz} and η are 3.87×10^{-4} and 7.17×10^{-2}, respectively.

The order parameters from the initial structure were transformed to the anisotropy (D_a) and rhombicity (R_h) parameters used in XPLOR-NIH, and a simulated annealing routine was implemented. Structures (10) with minimum final energies were collected from 2000 generated structures. The low-energy structure (structure 2) had torsion angles 291°/220°, 298°/97°, and 301°/219°. REDCAT analysis of this structure gave an S_{zz} of 3.1×10^{-4} and an η of 2.1×10^{-1} with a back-calculated rmsd relative to experiment of 0.82 Hz. The new order parameters were used in a second round of XPLOR-NIH refinement. The new best structure (structure 3) torsion angles were 290°/239°, 299°/109°, and 296°/234° with an S_{zz} of 3.3×10^{-4} and $\eta = 2.0 \times 10^{-4}$. This last refinement resulted in changes of angles by less than 20° and changes in order parameters by less than 5%. Convergence was judged adequate at this point. Figure 4 shows the superposition of the 10

(46) Woods, R. J.; Dwek, R. A.; Edge, C. J.; Fraser-Reid, B. *J. Phys. Chem.* **1995**, *99*, 3832−3846.
(47) DeLano, W. L. *http://www.pymol.org*; 2002.

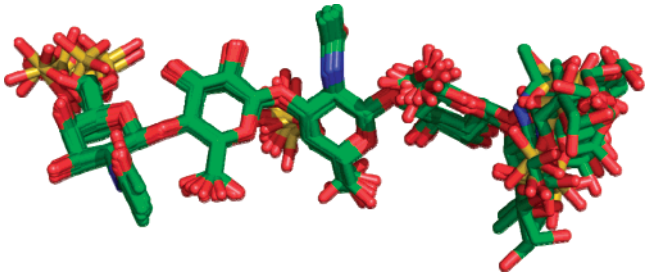


Figure 4. Superposition of 10 structures with minimum energy from the XPLOR-NIH simulated annealing calculation.

Table 4. Back-Calculated Values of CSA Offsets and RDCs Produced by REDCAT at Each Structure Refining Stage

residue	vector	exptl (Hz)	structure 1 (Hz)	structure 2 (Hz)	structure 3 (Hz)
A	C1H1	6.1	4.6	5.3	6.0
	C2H2	5.8	5.8	5.5	6.6
	C4H4	−10.0	−11.7	−11.5	−9.2
	C5H5	7.3	6.7	5.6	6.4
	H1H2	0.6	0.5	0.5	0.5
B	C1H1	4.4	4.1	4.9	3.7
	C2H2	5.0	5.7	5.4	4.6
	C3H3	4.8	6.0	5.2	5.6
	C4H4	4.2	3.4	4.8	4.0
	H1H2	−0.2	0.1	0.1	−0.1
C	C1H1	7.9	7.2	6.8	7.3
	C2H2	5.7	7.3	6.8	7.1
	C3H3	7.1	6.9	6.5	6.8
	C4H4	−6.5	−5.7	−6.7	−7.2
	H1H2	1.8	0.7	0.7	0.8
	H3H4	0.8	0.4	1.8	1.7
D	C1H1	7.9	6.6	7.8	8.2
	C2H2	7.7	8.5	8.1	8.4
	C3H3	9.3	8.2	8.1	8.4
	C4H4	8.1	7.6	7.9	8.2
	H1H2	1.5	0.9	1.0	1.1
A	CC X 10	3.9	2.0	3.3	3.4
C	CC X 10	5.8	6.3	5.3	6.1
A	C(O)	−1.6	−2.8	−3.1	−2.8
C	C(O)	−4.4	−5.0	−4.6	−4.9
	rmsd		0.96	0.82	0.68

minimum energy structures from the last refinement. Table 4 summarizes the course of the structure search. Back-calculated values of CSA offsets and RDCs have been produced by REDCAT at each stage.

Discussion

The combined data clearly lead to a reasonably well-defined description of a conformation of our CS pentamer in solution. The minimum energy structures resulting from simulated annealing under experimental restraints cluster well except for residue E. The rmsd for all heavy atoms in residues A−D is 0.53 Å; with residue E included the rmsd rises to 2.58 Å. The increased deviation in E results from the limited number of RDCs and the absence of an NOE distance constraint between the anomeric proton of residue D and any proton of residue E. However, the loss of ring structure for this residue on reduction introduces considerable motional freedom, and attempts to define a single structure would have been inappropriate in any event.

The validity of the structures depicted in Figure 4 is best discussed in comparison to data in the literature. A comparison of glycosidic torsion angles found in our structure and others is presented in Table 5. Direct comparisons are not actually

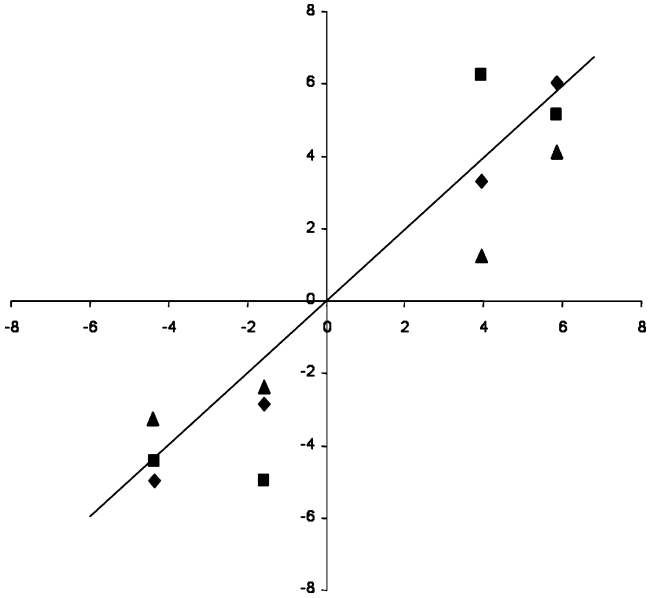


Figure 5. Correlation of experimental and calculated $^{13}C-^{13}C$ RDCs and C(O) CSA offsets for simulated docked CS structures. The NMR structure is assumed to be the correct docked structure (◆), and order parameters for this structure were used to back-calculate data for all structures. The other two docked structures are the X-ray structures from Cael et al. (■)[48] and Michel et al. (▲).[52] Residues B (glucouronic acid residues) in these two structures were superimposed on residue B in the NMR structure before back-calculating data.

Table 5. Comparisons of the NMR Structure of CS5 to Structures in the Literature

structure	torsions	A–B $\beta(1–4)$	B–C $\beta(1–3)$	C–D $\beta(1–4)$
NMR CS5	φ	−70	−61	−64
	ψ	−121	109	−126
Xray (fiber)[48]	φ	−98	−80	−98
	ψ	−174	107	−174
Xray (chondroitinase)[52]	φ	−69*	−89	−69
	ψ	−108*	108	−108
computation (MM3)[15]	φ	−79	−79	−78
	ψ	−110	90	−111
computation (CHARMM25)[49]	φ	−70	−70	−70
	ψ	−120	90	−120
computation/NMR[13]	φ	−80	−80	−80
	ψ	−110	90	−110

possible because of the differences in sulfation, differences in environment such as the presence or absence of counterions, or differences due to the presence or absence of protein binding (chondroitinase B). However, in each case, molecules share a $\beta(1-3)$ bond between a glucuronic acid residue and an *N*-acetylgalactosamine residue and a $\beta(1-4)$ bond between an *N*-acetylglactosamine residue and a glucuronic acid residue. In the table we have entered data corresponding to the analogous disaccharide linkages.

The earliest experimental data in the literature come from structures modeled to X-ray fiber diffraction data.[48] Coordinates for the CS4 tetramer have been deposited (PDB code 2C4S).[48] There are significant departures from the ϕ/φ angles we observe for both the $\beta(1-4)$ and $\beta(1-3)$ linkages (28°/53° and 34°/48°). However, the angles in this model also depart significantly in several cases from other observations in the literature. This may well be the result of intermolecular packing of polymeric

(48) Cael, J. J.; Winter, W. T.; Arnott, S. *J. Mol. Biol.* **1978**, *125*, 21−42.

chains or the presence of Ca^{2+} in these structures. Our data for the $\beta(1-4)$ linkage agree well with other structures; for example, ϕ and ψ depart by just 0°/1° and 6°/6° from values reported by Almond and Sheehan.[49] Our values for the ϕ/φ angles of the $\beta(1-3)$ linkage depart more significantly from other structures (9°/19° from the value given by Almond and Sheehan and 21°/19° from the value given by Blanchard et al.[13]). However, examination of the energy maps suggests that the energy predicted for our conformer is less than 2 kcal higher that the minimum energy structure; plus the energy maps provided are for a non-sulfated analogue *N*-acetylgalactosamine in one case and an unsaturated analogue of the glucuronic acid in the other case.

It is also important to point out that our experimental data are averages over conformations sampled by a potentially flexible molecule. We have already pointed to the fact that the effects would be severe for the E residue, which we do not attempt to model. For the rest of the residues, however, the fact that a large amount of data fit to a single representation of the conformation argues against flexibility being a large contributor.

More important than the identification of a solution conformation for CS5 is the establishment of a protocol for the evaluation of conformations of CS oligomers bound to proteins. This is an important issue as chondroitin sulfates (and other glycosaminoglycans) are known to modulate signaling by chemokines[50] as well as play roles in modulation of neural development through interactions with growth factors.[51] The proposed protocol would rely primarily on $^{13}C-^{13}C$ (or $^{13}C-^{1}H$) RDCs and CSA offsets from ^{13}C-labeled acetyl groups introduced into natural products. These are parameters that can be extracted from spectra even when resonances are broadened by strong interactions with macromolecules. The NMR/MS assignment strategy that we have introduced also allows specific assignment of resonances from multiple acetyl groups without the aid of triple or double resonance experiments that require passing magnetization though multiple scalar coupling pathways.

A key concern for potential application to protein-bound oligomers is whether the limited amount of data available from $^{13}C-^{13}C$ labeled acetyl groups can provide sufficient information to guide the identification of protein bound conformers. We do not as yet have data on protein bound oligomers, but we can assess our ability to distinguish sets of conformers that might be generated by docking programs. We can simulate this situation using some of the representative conformers in Table 5. We will mimic docking by superimposing the glucuronic acid residue between the two *N*-acetyl galactosamines of tetramers constructed using torsions listed in the Table 5. We will also assume that we have obtained an alignment tensor from data on the protein that coincided with the tensor determined from data on CS5 presented above. The acetyl RDCs and CSA offsets were then back-calculated for each of the structures using this alignment tensor in the program REDCAT. The correlation plots are presented in Figure 5. As expected, our structure shows a very good fit. Both of the other structures show sizable deviations. While this is purely an exercise in simulation, it does illustrate an ability to select among a set of acceptable conformations based on limited data available from acetyl $^{13}C-^{13}C$ RDCs and CSA offsets introduced into isolated glycosaminoglycan oligomers. Given an assignment strategy that does not require observation and resolution of resonances from other NMR active sites on the residues, we believe this strategy will be applicable to rather large protein complexes and oligosaccharides in the hexamer to octamer range. The limitation on oligosaccharide size is very likely to be a similarity of deacetylation rates near the center of oligomers and a resulting inability to make specific assignments.

Acknowledgment. This work was supported by a grant from the National Institute of General Medical Sciences, GM33225, and used computational facilities supported by a grant from the National Centers for Research Resources, RR05351.

JA075272H

(49) Almond, A.; Sheehan, J. K. *Glycobiology* **2000**, *10*, 329−338.
(50) Proudfoot, A. E. I. *Biochem. Soc. Trans.* **2006**, *34*, 422−426.
(51) Snow, D. M.; Mullins, N.; Hynds, D. L. *Microsc. Res. Tech.* **2001**, *54*, 273−286.
(52) Michel, G.; Pojasek, K.; Li, Y. G.; Sulea, T.; Linhardt, R. J.; Raman, R.; Prabhakar, V.; Sasisekharan, R.; Cygler, M. *J. Biol. Chem.* **2004**, *279*, 32882−32896.