Measuring the Magnitude of Internal Motion in a Complex Hexasaccharide

Soumya Ganguly,[1]* Junchao Xia,[2] Claudio Margulis,[2] Liz Stanwyck,[3] C. Allen Bush[1]

[1] *Department of Chemistry and Biochemistry, University of Maryland Baltimore County, Baltimore, MD 21250*

[2] *Department of Chemistry, University of Iowa, Iowa City, IA 52246*

[3] *Department of Mathematics and Statistics, University of Maryland Baltimore County, Baltimore, MD 21250*

Received 11 May 2010; revised 21 July 2010; accepted 22 July 2010
Published online 3 August 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bip.21532

ABSTRACT:

For the development of a scheme for quantitative experimental estimation of internal motion in the complex human milk hexasaccharide lacto-N-di-fuco hexose I (LNDFH I), we measured a large number of experimental residual dipolar couplings in liquid crystal orienting media. We present a total of 40 $^{13}C—^{1}H$ and $^{1}H—^{1}H$ dipolar coupling values, each representing distinct directions of internuclear vectors. The NMR data were interpreted with established methods for analysis of rigid subdomains of the oligosaccharide as well as a novel method in which dipolar couplings were calculated over an ensemble of conformers from a solvent Molecular Dynamics trajectory using multiple linear regression analysis. The Lewis^{b} epitope region of LNDFH I assumed a single unique conformation with internal motion described by fluctuations of 5–10° in glycosidic dihedral angles consistent with previous studies. Greater flexibility was observed for the remaining GlcNAc1→3-β-D-Gal and β-D-Gal1→4Glc linkages, with the former glycosidic linkage existing in a conformational exchange among three states. The results were also supported by similar results of calculations carried out with conformers obtained from a simple Monte Carlo simulation without explicit solvent. © *2010 Wiley Periodicals, Inc. Biopolymers 95: 39–50, 2011.*

Keywords: oligosaccharide; conformation; NMR; dipolar coupling; flexibility

This article was originally published online as an accepted preprint. The "Published Online" date corresponds to the preprint version. You can request a copy of the preprint by emailing the Biopolymers editorial office at biopolymers@wiley.com

INTRODUCTION

Defining the conformation and dynamics of macromolecules is one of the most interesting and challenging areas of structural studies. Among the different macromolecules, conformational details of oligosaccharides have been the most elusive because of the numerous linkage options and substitutions with sugars having varied functionality including amides, phosphates, carboxylates, and deoxy sugars.[1] For the last three decades, oligosaccharide conformation has been studied to gain insights into their structural characteristics.[2] Although many of these sugars exhibit conformational flexibility, it is not uncommon to find relatively rigid oligosaccharides such as the blood group determinants. The motion of these oligosaccharides when present in a physiological system has a major influence on various cellular recognitions mediated by both proteins and other carbohydrates. Experimental designs to determine these motions are difficult because of nanosecond to millisecond level motions of these oligosaccharides.[3,4]

Development of experimental methods to study motional characteristics in complex oligosaccharides has been particularly difficult. Some of these sugars can exhibit multiple

Correspondence to: C. Allen Bush; e-mail: bush@umbc.edu
**Present address:* Ganguly Vanderbilt University Centre for Structural Biology, 465 25th Avenue South., Nashville TN 37232.

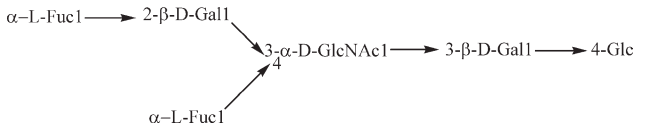

FIGURE 1 Schematic representation of the human milk oligosaccharide LNDFH I.

degrees of motion including various levels of conformational freedom at different linkages.[5] Molecular modeling simulation alone is insufficient to characterize these motions because accurate representation of the sugars requires large solvent boxes and time scales at the millisecond region, which are difficult to simulate. Parameterization of force fields also requires further refinement to accurately simulate the wide variety of complex carbohydrates. Experimental nuclear magnetic resonance (NMR) techniques have been extensively used to study carbohydrates, but the dependence of NMR observables like Nuclear Overhauser Effect on both molecular tumbling and internal motion complicates quantitative interpretation. Averaged values of scalar coupling can be more directly interpreted as averaged values of glycosidic dihedral angles but long-range $^{13}C—^{1}H$ coupling values necessary for conformation studies are difficult to measure accurately in the absence of isotope enrichment.

Recently, Prestegard and coworkers[6] used residual dipolar coupling (RDC) and generalized degree of order (GDO) to explain rigid and flexible regions present in a trisaccharide. As a scalar parameter, GDO can predict flexibility between two rigid fragments but does not explicitly reflect the amplitude of internal motion.[6] In the presence of internal motion, experimental parameters like RDC are averaged over all possible conformers and quantification of such motion is possible by identifying those conformers. Conformational averaging of RDC has recently been studied in sucrose and was found in a dynamic average of multiple conformers.[7] For branched and complex sugars consisting of four or more residues, similar studies have been very limited because of the lack of sufficient experimental data and a suitable methodology.

Here, we propose a scheme to study domain motion in the complex oligosaccharide. Lacto-*N*-difucohexose I (LNDFH I) (Figure 1) and its internal motion is discussed in detail. RDC was used as the experimental observable. For each sugar residue more than five independent RDC vectors were extracted to calculate the orientation tensors by fitting the experimental data. Molecular simulations were used to generate sets of co-ordinates for the hexasaccharide, which were later used to search the possible set of conformers responsible for averaging of the RDC. In this study, the Lewis^{b} tetrasaccharide epitope conformation was first separately studied using Powellten[8] and REsidual Dipolar Coupling Analysis Tool (REDCAT).[9] The linkages **d–e** and **e–f** that lie outside the Lewis^{b} tetrasaccharide epitope were studied by considering a shape induced alignment model of the entire hexasaccharide. TRAcking Moment of Inertia Tensor (TRAMITE)[10] was used to calculate dipolar couplings for selected models of the hexasaccharide, and multiple linear regression was used as the statistical method to determine which conformers best represent the average RDC data.

METHODS

Sample Preparation

The human milk oligosaccharide, LNDFH I was purchased from V-Labs, Inc (Covington, LA) and was used without further purification. A sample of 10 mg of the oligosaccharide was exchanged in 99.9% D_2O and lyophilized for three cycles. The final product was dissolved in 0.7 ml of 99.96% D_2O and used as an isotropic sample to determine scalar coupling.

A partially oriented sample of LNDFH I prepared using *n*-octyl pentaethylene glycol (C_8E_5) (Sigma-Aldrich, Milwaukee, WI) served as a neutral liquid crystalline medium.[11] The sample was prepared by dissolving 10 mg of lyophilized LNDFH I in 0.6 ml of 99.96% D_2O and 35.5 μl of C_8E_5, so that the final concentration of the liquid crystal in the solution is 5%. Finally, 18.07 μl of *n*-octanol was added to the solution to achieve a molar ratio of 0.87 for C_8E_5/*n*-octanol. Vigorous mixing resulted in an opalescent bluish, semiviscous, and homogeneous solution indicating formation of the liquid crystal.

Nuclear Magnetic Resonance

All NMR experiments for measuring scalar and RDCs were carried out on a 500 MHz Bruker Avance DRX spectrometer equipped with a cryoprobe at a temperature of 292 K. The experimental data were processed using NMRPipe/NMRDraw[12] and analyzed using NMRview[13] software on a Linux workstation. ^{2}H residual quadrupolar splitting of D_2O was used to check the homogeneity and concentration of the anisotropic sample. Homogeneous solution of 5% C_8E_5 resulted in a 40 Hz splitting at 292 K.

One bond C—H heteronuclear couplings were measured from indirectly coupled gradient HSQC spectra (t_1-coupled gHSQC).[14] Two sets of t_1-folded experiments were acquired with carriers positioned around the center of the anomeric region (90 ppm) and the ring region (70 ppm) with the spectral width of the indirect dimension reduced to 5 KHz to increase the digital resolution of the spectra. The direct sweep width was maintained at 2.5 KHz. The INEPT delays corresponding to the nominal value of $^{1}J_{CH}$ were set to 160 Hz. Carbon decoupling during acquisition was carried out with the WALTZ-16 pulse sequence with field strength of 1612.9 Hz.

Long range H—H coupling values of LNDFH I were measured using ^{13}C resolved HH-TOCSY[15] experiment with MLEV-17 spinlock for the TOCSY transfer. HH-TOCSY experiments were carried out at different mixing times ranging from 15 ms to 105 ms with an increment of 5 ms mixing time for each experiment. A change in mixing times clearly resolves particular cross-peaks compared with others. No changes in coupling constants were observed between

different mixing periods. The down field anomeric cross peaks were clearly resolved at higher mixing times (70–105 ms) whereas lower mixing times resolved the cross peaks of the rest of the sugar rings. All the ^{13}C resolved HH-TOCSY experiments were carried out with a spectral width of 2.5 KHz in the direct dimension. The carbon carrier was positioned at 60 ppm with a spectral width of 12575 Hz. The inept delay corresponding to nominal value of $^{1}J_{CH}$ was set to 160 Hz. WALTZ−16 pulse sequence was used for carbon decoupling with field strength of 1612.9 Hz.

Alignment Tensors and RDC Calculation

Three different methods were evaluated to interpret the experimental RDC data from a collection of conformers generated using the two different simulation methods Molecular Dynamics (MD) and Monte Carlo (MC). RDC was calculated using the general expression[16,17] given by

$$D^{XY} = D^{XY}_{\max} \frac{\langle (3\cos^2\theta - 1) \rangle}{2} \tag{1}$$

where D_{max} is the maximum dipolar coupling value possible between the nucleus X and Y at a reference internuclear distance of 1 Å and θ defines the angle between the internuclear vector and the magnetic field. The brackets < > denote the averaging of the angular information with time.

Method I

In this method, the Powellten algorithm[18] was used to calculate the orientation tensors from the Cartesian coordinates of the molecular models. This method assumes a rigid conformation of the molecule under consideration and can be applied to any subdomain of a polymer that is rigid. We used this technique to calculate the orientation tensor of the Lewis[b] tetrasaccharide epitope of LNDFH I, which is considered to be fairly rigid. For each of the tetrasaccharide models obtained from MC and MD simulations, orientation tensors were calculated using the direction cosines of the bond vectors obtained from the Cartesian coordinates. A Powell optimization algorithm[19] was used to fit the orientation tensors to the observed data by minimizing a merit function χ, measuring the goodness of fit defined in Eq. (2):

$$\chi = \sqrt{\sum_{1}^{n} \frac{(D^{\exp} - D^{\text{calc}})^2}{(D^{\exp})^2}} \tag{2}$$

where n is the number of dipolar vectors. A variable width Gaussian error distribution was used to simulate the experimental error. Powellten was advantageous in iteratively calculating the orientation tensor of a large number of models and then ranking them on the basis of their fit to the observed data. The resulting data were used to plot a colored contour map where the rank (χ) for each model was plotted against glycosidic torsion angles.

Method II

To calculate the RDC from a collection of molecular conformers, the order matrix based approach REDCAT[9] was implemented. Similar to Powellten, REDCAT assumes a rigid molecular conformation of the fragment with small degrees of internal motions. Therefore, the order matrix calculation was used to derive the RDCs of the Lewis[b] tetrasaccharide epitope of LNDFH I similar to that of method I. With the knowledge of the molecular geometry and five or more distinct dipolar coupling data, the Saupe order matrix was diagonalized,[20] yielding the principal order parameter (S_{zz}), an asymmetry parameter (η) and the three Euler angles relating the molecular fragment from the arbitrary molecular co-ordinate to the principal co-ordinate frame.

One of the important features of REDCAT is that it can account for the experimental error of the calculated dipolar coupling data by adding uncertainties to generate several sets of dipolar coupling data by uniformly sampling using an MC simulation. Order tensors were calculated based on 10,000 cycles of MC simulation with RDC values randomly varied over a range of 1–1.5 Hz as described in Ref 9. From a collection of 3000–4000 order parameters, the best solution was selected based on the least mean square deviation from the experimental data. The order parameters were also used to back calculate the RDC for each of the conformers.

Internal motions of complex macromolecules of oligosaccharides have also been interpreted using the GDO ν.[21] As a scalar parameter, GDO values can characterize the degree of internal motion independent of the order frame and anisotropy of the molecule. Two rigid fragments with no internal motion should exhibit identical GDO values given by

$$\nu = \sqrt{\frac{2}{3}\sum s_{ij}^2} \tag{3}$$

where S_{ij} are the elements of Saupe order matrix.

Method III

The Methods I and II discussed above are not suitable to describe the conformation of flexible molecules. The shape of a single contributing conformer in an ensemble has also been used to estimate the alignment tensor. For neutral orienting media (C_8E_5), the Simulation of Sterically Induced Alignment method assumes that the molecular shape determines the alignment tensor.[22–24] TRAMITE[10] uses the moment of inertia tensor to approximate four of the five elements of the order tensor, leaving the magnitude to be adjusted using a scaling factor. TRAMITE proves advantageous in calculating order parameters for an ensemble of conformers based on the shape asymmetry, which makes this method independent of molecular rigidity. For each of the conformers, the RDCs can also be back calculated from the order parameters.

Molecular Modeling and Simulation

Two different methods of molecular modeling simulation were used to determine the three dimensional conformation of LNDFH I. The models from each simulation were treated independently to interpret the RDC data and identify combinations of conformers that can explain the measured experimental dipolar coupling values.

We have used the trajectories of MD simulation published earlier by Xia and Margulis.[25] This is a simulation of 10 ns with explicit SPC water solvent and OPLS-AA force field. A collection of 40,000 conformations each collected at an interval of 0.25 ps was obtained. A set of 3200 representative conformers was chosen for the present studies.

A simple and less elaborate simulation based on MC simulation was also performed on LNDFH I without explicit water. The simula-

tion was carried using Macromodel[26] with 10,000 MC steps[19] along with MM3[27] force field and GB/SA continuum solvent model.[28] The selected 2000 conformers had threshold energy values of 50 KJ/mol or less from the lowest energy structure. The random process of conformer generation using the MC algorithm can overcome energy barriers and sample different regions of the conformational space. Although it is quite possible that not all the conformers generated by such a simple simulation and force field are true local minima, this collection of conformers can serve as a repertoire for our conformational search.

The glycosidic torsion angle was used to analyze the conformational space of LNDFH I. The Φ and Ψ torsion angles are defined according to the IUPAC heavy atom convention in which Φ is defined by O_5-C_1-O_1-C_x and Ψ by C_{x-1}-C_x-O_1-C_1.

Statistical Analysis

One of the major goals of the present study was to determine internal motion as reflected by the experimental RDC data. In the presence of internal motion, the experimental RDC data are the result of average RDC of the conformers resulting from the motion. MD and MC simulation were used to search for possible sets of conformers that can explain the experimental RDC data. Multiple regression analysis technique was used to find the possible combination of conformers whose calculated RDC was the linear combination of the observed data.

Multiple linear regression attempts to fit a linear equation, which can relate two or more sets of explanatory (independent) variables to a response (dependent) variable. In this study, the independent variables were the calculated RDCs for each of the conformers from the molecular simulations whereas the dependent variables represented the observed RDC data.

The best sets of independent variables explaining the experimental data were calculated using the R^2 (R-square) selection method. The R^2 is a measure of the proportion of variation explained by the model and ranges between $0 < R^2 < 1$, where values closer to 1 indicate a better fit for the model. This method calculates R^2 for all possible subsets of conformers. The subset with the largest R^2 is declared as the best linear fit for the model. From a combinatorial approach, the R^2 selection technique is ideal because it searches through all possible combinations of conformers to best explain the observed data. However, this process is also computationally challenging because an increase in the number of conformers in the subset increases the computational time exponentially. Inclusion of fewer than five models gave a smaller R^2 value whereas 6 models gave similar values and computations with 7 or more models were not practical. We concluded that 5 models is an optimal choice for the experimental data available.

Statistical averaging of the five candidate conformers selected during multiple linear regression was calculated by assigning weights randomly to each one of them for 10,000 cycles using the following equation:

$$\sum D^{\text{pred}} = a_1 \sum D_1 + a_2 \sum D_2 + \ldots + a_5 \sum D_5 \qquad (4)$$

$$a_1 + a_2 + \ldots a_5 = 1 \qquad (5)$$

where D^{Pred} is the predicted dipolar coupling, D_1, D_2, . . ., D_5 are the sum of the dipolar couplings of the selected conformers and a_1, a_2, . . ., a_5 are the corresponding statistical weights. Squared deviation from the experimental data was calculated using the sum of the predicted dipolar coupling values from Eq. (6):

$$\sigma = \sqrt{\left(\sum D^{\text{pred}} - \sum D^{\text{exp}}\right)^2} \qquad (6)$$

where D^{Exp} is the experimental RDC data. The weights of the least squared deviation were used for statistical averaging of the five conformers.

Table I One bond C—H Experimental Residual Dipolar Couplings in Hz of LNDFH I (A: Fucose; B: Galactose; C: Fucose; D: *N*-acetylglucosamine, E: Galactose; F: Glucose)

Atom Pairs	C—H RDC ($^1D_{CH}$) in Hz
H1—C1A	−22.5
H2—C2A	2.70
H3—C3A	0.31
H5—C5A	2.48
H1—C1B	−9.16
H2—C2B	−1.17
H3—C3B	−1.43
H4—C4B	12.28
H5—C5B	−3.23
H1—C1C	8.56
H2—C2C	2.13
H5—C5C	2.77
H1—C1D	14.12
H2—C2D	14.4
H5—C5D	15.5
H1—C1E	11.6
H2—C2E	3.86
H3—C3E	6.23
H4—C4E	−3.78
H5—C5E	4.92
H1—C1Fα	−10.2
H1—C1Fβ[a]	8.47
H3—C3F	4.9
H5—C5F	7.33

[a] Denotes the beta anomer of glucose.

RESULTS

The scalar coupling values were measured for LNDFH I in natural abundance using D_2O as the isotropic media. An anisotropic sample consisting of 5% C_8E_5/*n*-octanol liquid crystal media was used to orient the hexasaccharide to measure the scalar plus dipolar coupling. RDCs were finally obtained by subtracting individual splitting obtained in isotropic media from that of anisotropic ones.

One bond $^{13}C—^1H$ heteronuclear dipolar couplings (Table I) were measured using t_1 coupled gradient HSQC (Figure 2) spectra. We report 24 $^{13}C—^1H$ RDC data meas-

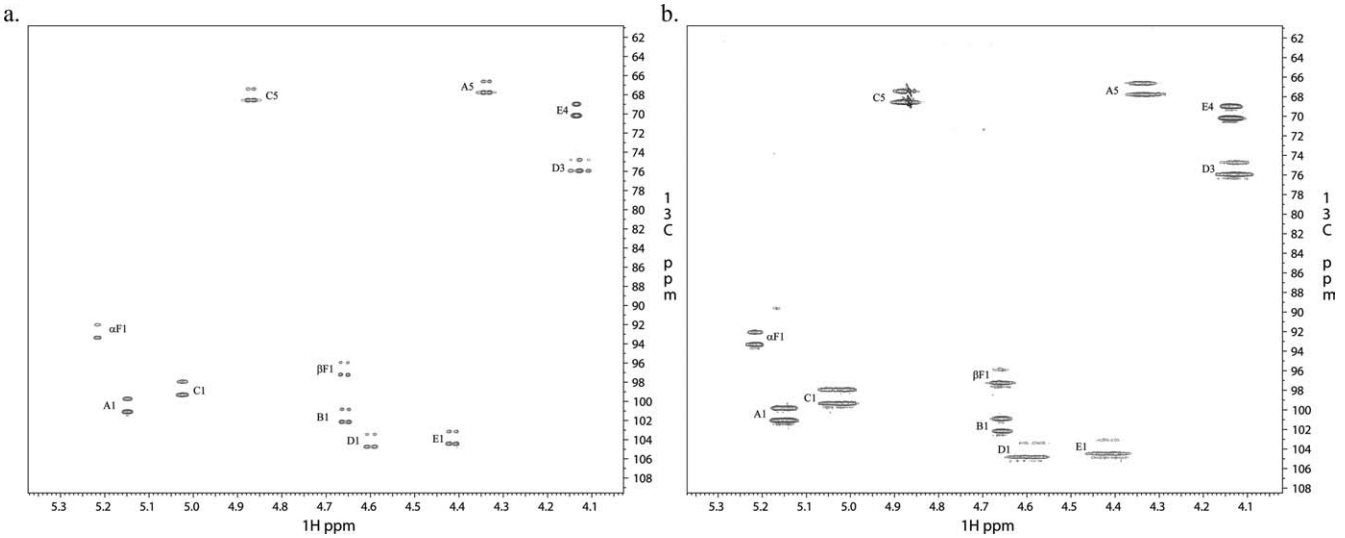

FIGURE 2 Expanded region of the ^{13}C coupled HSQC spectrum of LNDFH I in (a) isotropic and (b) anisotropic media. Cross-peaks corresponding to the C—H correlation in the down field anomeric region are annotated.

ured for LNDFH I. It is to be noted that the glucose residue at the reducing terminal can have two different anomeric configurations with similar ring geometry except for the anomeric region. We have used the dipolar couplings of anomeric carbon and proton from both α and β anomers in this study.

Long range $^1H—^1H$ RDCs ($^ND_{HH}$) were measured using ^{13}C resolved HH-TOCSY[15] (Figure 3). The offsets from the E.COSY type cross peaks in the proton detected dimension were used to determine the sign and magnitude of the $^ND_{HH}$. The homonuclear couplings were measured for only those peaks that could be unambiguously assigned in both isotropic and anisotropic spectra. Although the f_1 splittings of the HH-TOCSY spectra could be used to measure one bond $^{13}C—^1H$ dipolar couplings, the resolution of such splittings are lower than that of HSQC experiments and thus not con-

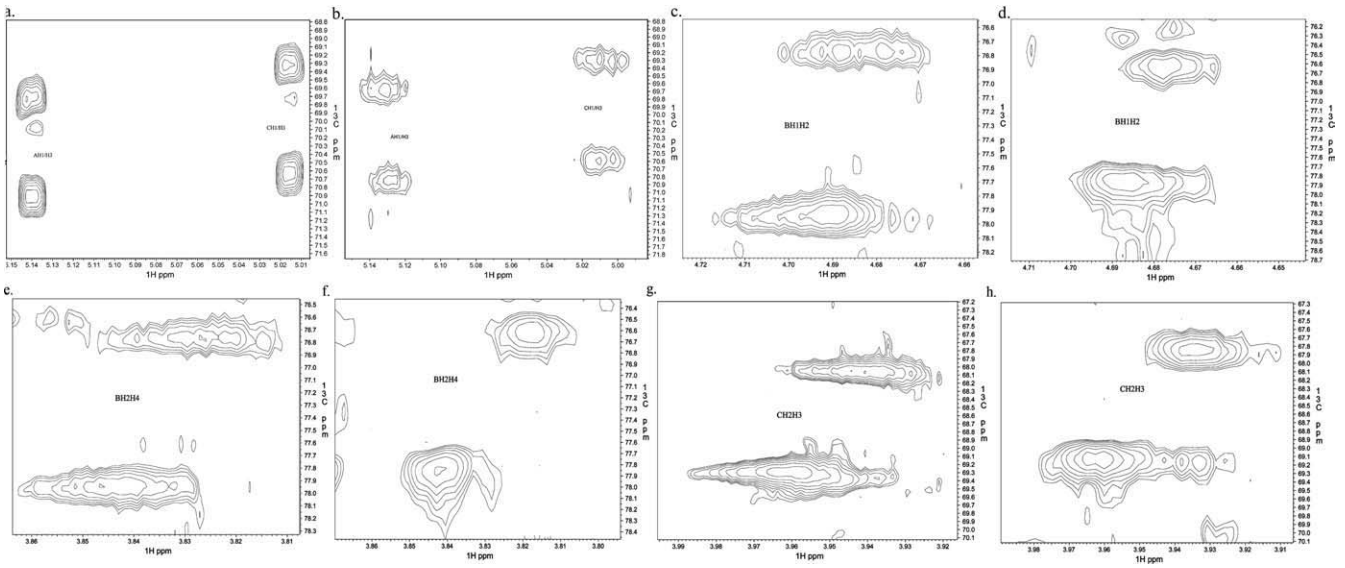

FIGURE 3 Expanded regions from the ^{13}C resolved HH-TOCSY spectra in (a, c, e, and g) D_2O and (b, d, f, and h) orienting media 5% C_8E_5 used to measure long range homonuclear ^1H, ^1H RDCs in LNDFH I. For each of the peaks the first letter and number denotes the proton connected to the carbon and the second set of letter and number is the proton to which the former is coupled. The overlapping peaks were resolved using spectra at different mixing times.

Table II Long Range H—H Experimental Residual Dipolar Couplings of LNDFH I (A: Fucose; B: Galactose; C: Fucose; D: *N*-acetylglucosamine, E: Galactose; F: Glucose)

Atom Pairs	H—H RDC ($^{n}D_{HH}$) in Hz
H1—H2A	−1.56
H1—H3A	−1.93
H1—H4A	0.40
H3—H4A	−0.57
H1—H2B	−3.8
H2—H3B	2.25
H3—H4B	−2.40
H1—H2C	−2.15
H1—H3C	1.30
H2—H3C	2.03
H1—H2D	−2.56
H1—H3D	−2.42
H1—H4D	0.10
H2—H3D	3.26
H2—H4D	1.00
H4—H5D	−2.41
H2—H5D	−1.20
H1—H2E	1.26
H1—H3E	1.75
H2—H3E	−0.86
H3—H4E	0.61
H1—H2Fα	−0.50
H1—H2Fβ[a]	−9.3
H3—H4F	0.60
H2—H3F	−4.67

[a] Denotes the beta anomer of glucose.

sidered for the present calculations. The 26 ^{1}H—^{1}H dipolar couplings (Table II) measured for LNDFH I included both H1(α)—H2 and H1(β)—H2 splittings from the glucose residue at the reducing terminal.

Both the gHSQC and HH-TOCSY experiment produced a two-dimensional (2D) matrix of 4096 × 512 complex FID data points. The digital resolution of the resulting spectra was 1.2 Hz/pt in the direct dimension and 2.4 Hz/pt in the indirect dimension. Each coupled cross peak column was then extracted, phased, and stored. An inverse Hilbert transformation and zero filling of the resulting 1D vector further increased the digital resolution. The RDC was calculated by superimposing two 1D vector components collected from an isotropic and weakly aligned sample of the oligosaccharide (Figure 4). One of the vector components was adjusted so that one wing of the vector completely coincided with the wing of the corresponding vector. The offsets arising in the other wing of the vectors as a result of the adjustment was the measure of C—H RDC ($^{1}D_{CH}$) and long range H—H RDC ($^{N}D_{HH}$) from gHSQC and HH-TOCSY experiments, respectively. Error in the measurement attributed mostly to the difference in lineshapes of the two spin doublets of a single resonance[29] was estimated in both the homonuclear and heteronuclear coupling to be approximately ±1 Hz to ±1.2 Hz. The experimental data include RDC values for at least six distinct vector directions for each of the sugar residues. The number of experimental observables recorded for LNDFH I was much higher than in previous studies[19,30] on similar oligosaccharides providing a more accurate prediction of the molecular geometry of LNDFH I.

The hexasaccharide conformation was studied using a 10 ns molecular dynamic simulation in explicit water as has been discussed earlier.[25] The results indicated a single unique conformation for each of the three linkages in the Lewis[b] tetrasaccharide epitope in the LNDFH I hexasaccharide. For the **d–e** linkage, two different conformers were observed whereas the lactose linkage (**e–f**) exhibited some flexibilities in the Ψ angle (77–174°).

The hexasaccharide conformation was also studied using MC simulations. The 10,000 step MC simulation was a less rigorous approach than the MD simulation and resulted in a little more than 2000 conformers. Conformers with energy above 50 KJ/mol (12 Kcal/mol) threshold energy level were rejected during this calculation. Unlike the MD calculation, each of the linkages of the hexasaccharide indicated more than one family of conformers (Figures 5a–5e), but the conformational spaces included conformer families for each of the linkages described by the MD simulation.

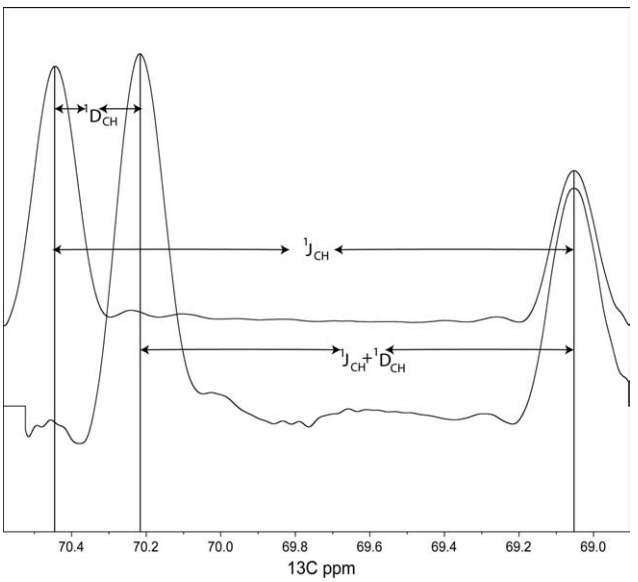

FIGURE 4 A 1D slice extracted from 2D gHSQC spectra. The two spectra recorded from isotropic and anisotropic samples of β-D-Gal1→4 C4-H4 couplings are superimposed to calculate the RDC.

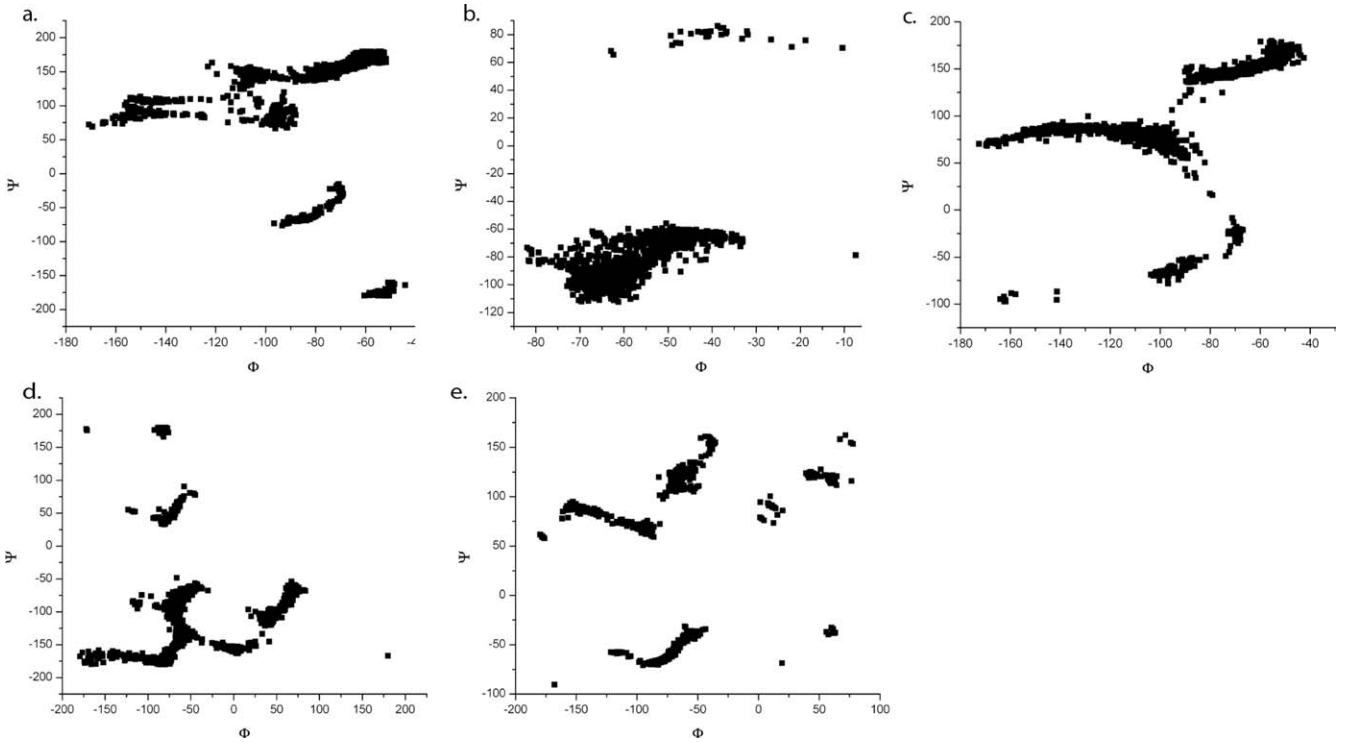

FIGURE 5 Scatter plot of five Φ/Ψ dihedral angles for 2000 conformers of LNDFH I with a threshold energy of 50KJ/mol calculated from MC simulation. (a) α-L Fuc1→2-β-D-Gal, (b) β-D-Gal1→3-β-D-GlcNAc, (c) α-L-Fuc1→4-β-D-GlcNAc, (d) β-D-GlcNAc1→3-β-D-Gal, and (e) β-D-Gal1→4D-Glc.

Measuring Internal Motion of the Tetrasaccharide Epitope in the Hexasaccharide

Powellten was used to assign ranks to each model based on the discrepancies between the experimental and calculated RDC (χ) value of the optimally oriented tetrasaccharide model. These model conformers were ranked for both MD and MC simulation results by plotting χ values on a contour map against the Φ/Ψ dihedral angles for each of the glycosidic linkages. An experimental error of 1.0 Hz was used for all the model structures in the calculation.

Figure 6 shows two dimensional plots where the minimum discrepancy of the calculated dipolar coupling from the observed values (χ) from the Powellten calculation are plotted as a function of the dihedral angle for all the conformers

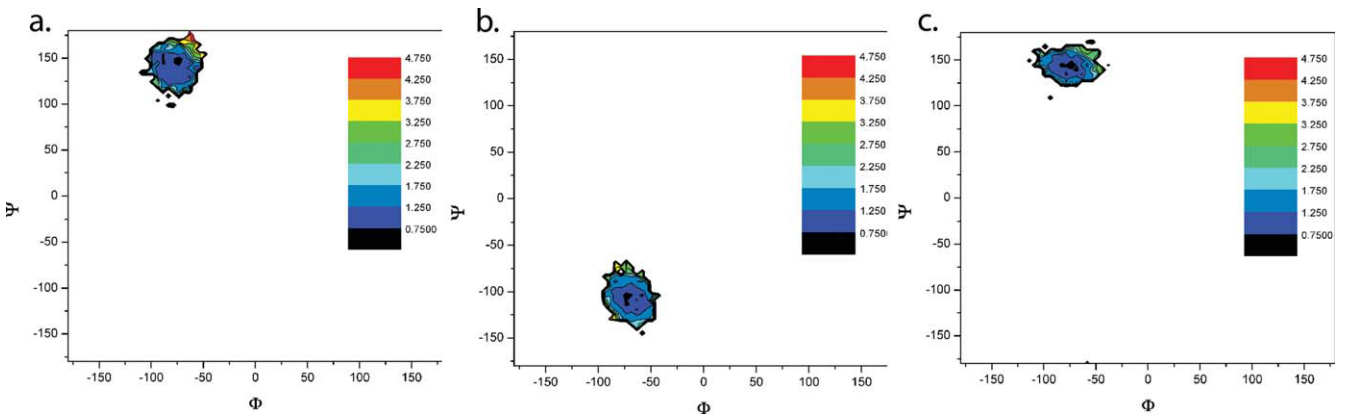

FIGURE 6 Contour plots from the MD models of the tetrasaccharide epitope with the ranking of conformers based on the χ_{min} value calculated using Powellten for individual dihedral linkages. (a) α-L-Fuc1–>2-β-D-Gal, (b) β-D-Gal1–>3-β-D-GlcNAc, and (c) α-L-Fuc1–>4-β-D-GlcNAc.

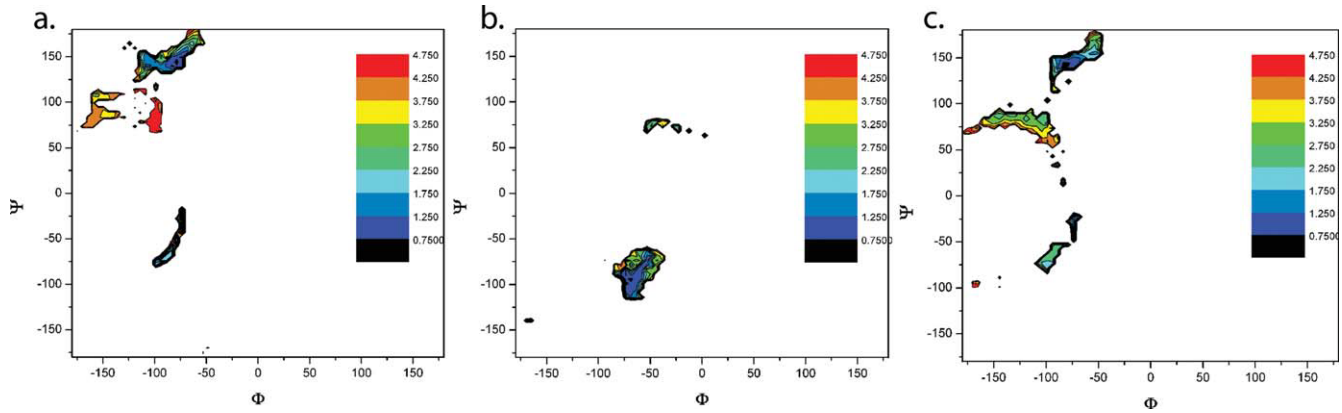

FIGURE 7 Contour plots from the MC models of the tetrasaccharide epitope with the ranking of conformers based on the χ_{min} value calculated using Powellten for individual dihedral linkages. (a) α-L-Fuc1→2-β-D-Gal, (b) β-D-Gal1→3-β-D-GlcNAc, and (c) α-L-Fuc1→4-β-D-GlcNAc.

from the MD simulation. For all three plots, the χ minimum was restricted to a small region within the dihedral plot. For the α-L-Fuc1→2-β-D-Gal ($\Phi_{a\text{-}b}/\Psi_{a\text{-}b}$) linkage (Figure 6a) this region lies around −70°/135°, a result similar to that reported by Azurmendi and Bush for the same linkage in blood group A trisaccharide.[18] The regions indicating χ minima for the other two linkages, α-L-Fuc1→4-β-D-GlcNAc and β-D-Gal1→3 β-D-GlcNAc were approximately ($\Phi_{b\text{-}d}/\Psi_{b\text{-}d}$) −75°/145° and ($\Phi_{c\text{-}d}/\Psi_{c\text{-}d}$) −65°/110°, respectively.

The χ minima ranking was also carried out independently for all the conformers from the MC simulation and were plotted similarly in a contour plot as described above (Figure 7). The Powellten calculations were repeated with MC data to investigate whether the results were dependent on a high quality set of simulations (e.g with explicit solvent). A small single region from each of the contour plots indicated conformers with least deviation from the observed RDC data. The conformers with lowest RDC discrepancies occupied the same dihedral angle regions as that of the MD simulation.

To gain further insight into the dynamics of the tetrasaccharide epitope of LNDFH I, we implemented GDO techniques described in previous studies[6] using the REDCAT algorithm. Unlike Powellten, where the entire tetrasaccharide epitope was oriented, the GDO was calculated by independently orienting individual pyranose rings of the tetrasaccharide. The GDO values of each pyranoside ring of the tetrasaccharide were calculated from a random frame of the MD trajectory. For the sugar rings **a**, **b**, **c,** and **d** (Fuc1→2, Gal, Fuc1→4, and GlcNAc), the GDOs were 0.00047, 0.00043, 0.00032, and 0.00081, respectively. Similarities in the values for the first three rings indicate that these fragments tumble together because of a single rigid conformation. However, different results for the GlcNAc residue indicate significant internal motion between the rings **b**–**d** and **c**–**d**. Comparable GDO values were also obtained from the lowest energy MC simulation model. The results from the GDO calculation were not consistent with earlier studies in which the tetrasaccharide was found to exist in a single rigid conformation. The GDO results thus proved to be inconsistent with previous molecular modeling and NMR studies of this epitope, perhaps as a result of the complexity of the branched tetrasaccharide used in the present study when compared with an earlier GDO analysis of a trimannoside.[6]

As an alternative approach, REDCAT was used to measure the order parameters from the RDC data to calculate the amplitude of dihedral angle motion within the Lewis[b] tetrasaccharide epitope. REDCAT assumes that the fragment for which order tensors are to be calculated is rigid.[31] Based on our Powellten calculation and also from earlier studies on the epitope, we assumed a relatively rigid conformation of the tetrasaccharide with motions of the first kind.[32] It is also assumed that any sub fragments within the tetrasaccharide should also share the same order parameters and the principal order frame. This should also be true for the individual sugar pyranoside residues which are part of the tetrasacchar-

Table III The Intercept and Linear Combination Coefficients of the Models Resulting in an Average Conformer of the Tetrasaccharide Epitope of LNDFH1

	MD	MC
Intercept	−0.41	−0.34
Conformer 1	0.74	−1.49
Conformer 2	0.71	1.41
Conformer 3	−0.54	−0.95
Conformer 4	−0.81	−0.86
Conformer 5	0.53	1.23
R^2	0.98	0.98

Table IV Dihedral Angles of the Average Conformer of the Tetrasaccharide Epitope with the Degrees of Fluctuation

	Fuc1→2Gal		Gal1→3GlcNAc		Fuc1→4Gal	
	Phi	Psi	Phi	Psi	Phi	Psi
MD	−69.4°±7°	139.8°±10°	−63.2°±8°	103.8°±10°	−77.8°±8°	144.4°±8°
MC	−78.0°±8°	145.4°±7°	−65.9°±2°	93.4°±6°	−76.7°±5°	143.4°±2°

ide and are considered to be fairly rigid. Because the β-D-GlcNAc resides at the center of the tetrasaccharide fragment, we have used the order tensor parameters of this residue to orient the entire tetrasaccharide epitope along the principal order frame. The order tensor parameters were calculated for the β-D-GlcNAc residue from 100 random MD conformers using REDCAT. At least eight independent C—H and H—H RDC vectors of the β-D-GlcNAc residue were used for this calculation to avoid an underdetermined system. The best order parameter is the least RMSD solution of all possible outcomes for the β-D-GlcNAc residue. The order parameter and three Euler angles were then used to orient the entire tetrasaccharide epitope from which the RDCs were back calculated for each of the 100 models.

Determination of the average RDC required identification of the conformers and their linear combination coefficients. Multiple linear regression analysis was used to calculate conformational averaging and estimate the degree of internal motion. RDC data back-calculated for the 100 tetrasaccharide models were used to explain the experimental data. Recursive combinations of all possible conformers were performed within a regression model until the best R^2 value of 0.98 was achieved for a set of five conformers. The linear combination coefficients of the five selected conformers are shown in Table III. The average dihedral angles of the five best fit conformers and their standard deviations were calculated (Table IV). Dihedral angles of the final conformers were found in a very small conformational range with average deviation of the Φ angle to be ∼ ±7.5° and that of Ψ ∼±9° around the average conformer.

The results of RDC averaging performed with MD simulation data sets were validated by repeating the calculation using conformers from MC simulation. The five conformers of the tetrasaccharide representing the average RDC value of the experimental data were distributed in a small conformational space. Three average dihedral angles (Table IV) of the tetrasaccharide region of LNDFH I was similar to those obtained from the comparable calculation done with the MD models. The deviations in dihedral angles between the glycosidic linkages among the five different conformers were also minimal. Thus, this analysis of the RDC data for the tetrasaccharide epitope, while it allowed for multiple conformers, showed a good fit with a group of very similar conformers regardless of the quality of the molecular models used.

Measuring Internal Motion in LNDFH I Hexasaccharide

The concept of finding combinations of conformers to explain the experimental data was extended for the hexasaccharide to determine the conformations of β-D-GlcNAc1→3-β-D-Gal (**d–e**) and β-D-Gal1→4Glc (**e–f**) linkages. As this region unlike the tetrasaccharide is considered to exhibit domain motion, Methods I and II are not expected to be useful for calculation of order parameters. Order parameters for the hexasaccharide were predicted based on a shape induced steric obstruction method, TRAMITE.

To calculate the order parameters, a collection of 100 conformers from the MD simulation was generated by using the tetrasaccharide dihedral angles from the previous section (Table IV) to select the possible hexasaccharide conformations. Thus different possibilities for the glycosidic dihedral angles of **d–e** and **e–f** were considered. The selected conformers had the same tetrasaccharide conformation reported in Table IV, with different dihedral angles for the **d–e** and **e–f** linkages. The RDCs were then back calculated for each of the 100 conformers.

Multiple linear regression analysis, as described above, was then used to determine the collection of models that best explains the experimental data. For the MD simulation data,

Table V The Intercept and Linear Combination Coefficients of the Models Resulting in an Average Conformer of the Hexasaccharide LNDFH I

	MD	MC
Intercept	1.94	1.18
Conformer 1	−3.15	−4.06
Conformer 2	−2.45	−3.74
Conformer 3	−3.28	2.42
Conformer 4	2.09	4.52
Conformer 5	4.35	−4.57
R^2	0.70	0.69

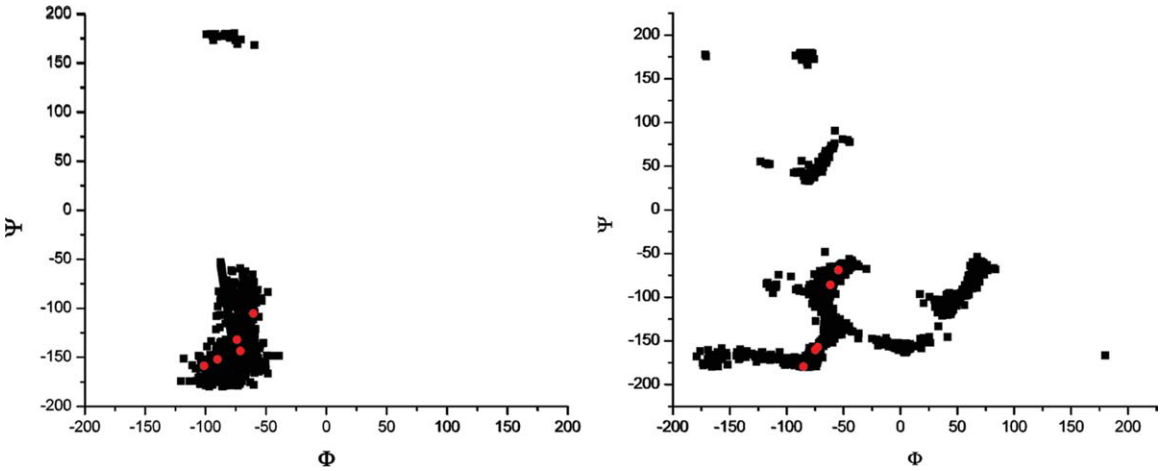

FIGURE 8 Scatter plot of the glycosidic torsion angles for β-D-GlcNAc1→3-β-D-Gal linkage. The conformers (in black) are generated using (a) MD simulation and (b) MC simulation. The red points represent the best combination of conformers explaining the experimental RDC data.

the regression model resulted in the identification of five distinct conformers with R^2 of 0.7 (Table V). When the same calculation was carried out on the models from MC data, a set of five models was obtained whose linear combination equaled the observed RDC values. The R^2 for the models from the MC simulation was 0.69 (Table V). The decrease in R^2 value when compared with the tetrasaccharide calculation can be attributed to the increase in the number of RDCs incorporated in the hexasaccharide model thereby increasing the degrees of freedom for the resulting fit.

The five conformers resulting from our calculation can be broadly classified within three distinct regions in the dihedral angle space of the **d–e** linkage (Figure 8a). The Φ and Ψ angles for the conformers in these three regions are −60°/−105° (region I), −74°/−131° (region II), and −90°/−152° (region III), respectively (Table VI). The differences among the conformers in these three regions were much larger in the Ψ dihedral space than in the Φ dihedral angle. A similar calculation using MC simulation models also identified three different conformational regions whose Φ and Ψ dihedral angles are −61.32°/−85.86° (region I), −72.27°/−157.24° (region II), and −85.00°/−179.4° (region III), respectively (Table VI). The conformers in these three regions were much more widely dispersed in the Ψ dihedral region compared to that of MD outcomes (Figure 8b). The study of **d–e** linkage from both MD and MC data thus indicated conformational averaging among three different conformers. When statistical averaging was carried out on the five conformers from MD and MC simulation, we found that 80% of the conformers were represented by the regions II and III whereas only 20% of the population were observed in region I.

The results from both MD and MC simulation data showed a broad single conformational region occupied by the β-D-Gal1→4-Glc (**e–f**) dihedral angle. The five conformers resulting from the multiple linear regression calculation were found to be spread throughout this broad conformational region. For the results obtained from MD data, the Φ dihedral angles ranges between −95° and −65° and the Ψ dihedral angle ranges between 74° and 144° (Figure 9a). Similar results were obtained from the MC calculation where the Φ and Ψ dihedral ranged between −70° and −40° and 108° and 154°, respectively (Figure 9b). Although these conformers were found in discrete regions within the conformational space, it is reasonable to assume that these conformers give a sense of average degree of motion within this conformational region. The amplitude of this motion was found within a broad range of ±30° (Φ) and ±20° (Ψ) indicating large continuous motion. Large fluctuations of the lactose linkage can be attributed to the absence of steric interaction for the glucose with the rest of the sugars because it is positioned further away from the tetrasaccharide.

For the Lewis^b tetrasaccharide region of LNDFH I, our results indicated a single conformation with small local fluctuation between the glycosidic linkages. The fact that glyco-

Table VI Major Conformers of GlcNAc1→3Gal Linkage in LNDFH1 Calculated From MD and MC Data

	Conf I	Conf II	Conf III
MD	−60.55°/−105.09°	−74.04°/−131.84°	−90.17°/−151.97°
MC	−61.32°/−85.86°	−72.27°/−157.24°	−85.00°/−179.4°

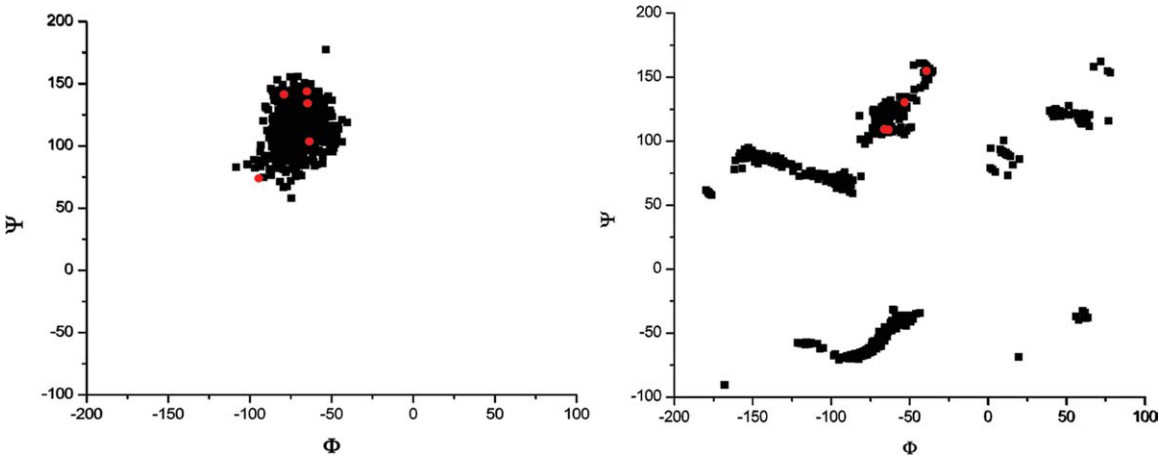

FIGURE 9 Scatter plot of the glycosidic torsion angles for β-D-Gal1→4-D-Glc linkage. The conformers (in black) are generated using (a) MD simulation and (b) MC simulation. The red points represent the best combination of conformers explaining the experimental RDC data.

sidic dihedral angles of the Lewis[b] fragment occupy a narrow potential energy well correlates with previous studies.[33,34] The dihedral angles of the tetrasaccharide linkages predicted by our study matches with the data obtained from lectin (GS4) bound Lewis[b][35] except for the α-L-Fuc1→4-β-D-GlcNAc linkage, where the Φ angle differed by ∼15°. Similarities were also observed with the dihedral angles of the tetrasaccharide calculated by Martin-Pastor and Bush,[30] except for the α-L-Fuc1→2-β-D-Gal linkage. Conformational exchange among three different conformers observed in this study for the β-GlcNAc1→3-β-D-Gal linkage contrasts with earlier NMR studies indicating a single unique conformation for this linkage.[30,36] However, these results for the GlcNAc1→3-β-D-Gal (**e–f**) linkage are consistent with similar studies on another milk oligosaccharide, Lacto-*N*-neotetraose (LNnT).[37] Substantial flexibility of the lactose glycosidic linkage was confirmed by the present study. In LNDFH I, the lactose linkage is in continuous motion within a single broad conformational well. Minimal steric interaction of the glucose residue with other pyranose rings in the hexasaccharide is a plausible explanation for such motion. These motions can be represented as conformers with a broad potential energy well, in contrast to those with more than one narrow potential well separated by energy barriers. Similar motion of the lactose linkage has also been proposed by Widmalm and coworkers.[3]

CONCLUSION

We investigated the internal motion of the hexasaccharide LNDFH I using $^{1}D_{CH}$ and $^{N}D_{HH}$ RDCs. The approach used is related to that of Blackledge for describing domain motion in partially folded proteins as reviewed in Ref. 38. For that system, which has been much more extensively studied, one can rely on a library of known polypeptide substructures. For oligosaccharides, we used simple molecular modeling calculations to construct candidate structures.

We present methods to extract a large number of RDC data from an oligosaccharide molecule, which is enough to test either rigid or flexible models for best fit to experimental data. The use of more RDC data than any of the previous studies helps in accurate determination of the orientation tensors. Instead of using the RDCs as constraints,[39] we have exploited the fact that these NMR observables exist as an average value of all the possible conformers. Our model for flexible conformation is represented by an ensemble of conformers exchanging at a rate faster than few 100 hertz. Such an assumption is certainly valid for oligosaccharides.

Three distinct internal motions were observed in LNDFH I. Small local libration (∼±10°) around the average conformer resulted in a relatively rigid conformation of the Lewis[b] tetrasaccharide region. Three different conformers resulted in conformational exchange of the β-D-GlcNAc1→3-β-D-Gal linkage. Large internal motion of the β-D-Gal1→4-Glc linkage within a single conformational well indicates highly flexible linkage. Our methods were also able to quantify the degree of internal motions of this complex sugar. The results indicate sensitivity of RDC to small changes in the dihedral angles. The conclusions about the three dimensional structure do not depend on availability of high quality MD simulation and were supported by results from a simple MC data. The fact that our results were able to narrow down the rigid conformational spaces for the three dihedral angles of the tetrasaccharide from a multi-confor-

mational space generated by MC simulation indicates the robustness of our method. This method proved efficient for prediction of internal motion using a simple set of molecular simulations to sample the conformational space. An important result of this method is a quantitative estimate of the degree of motion in this oligosaccharide.

REFERENCES

1. Adeyeye, J.; Azurmendi, H. F.; Stroop, C. J. M.; Sozhamannan, S.; Williams, A. L.; Adetumbi, A. M.; Johnson, J. A.; Bush, C. A. Biochemistry 2003, 42, 3979–3988.
2. Lemieux, R. U.; Bock, K.; Delbaere, L. T. J.; Koto, S.; Rao, V. S. Can J Chem 1980, 58, 631–653.
3. Rundloef, T.; Venable, R. M.; Pastor, R. W.; Kowalewski, J.; Widmalm, G. J Am Chem Soc 1999, 121, 11847–11854.
4. Martin-Pastor, M.; Bush, C. A. Biopolymers 2000, 54, 235–248.
5. Xia, J.; Daly, R. P.; Chuang, F.-C.; Parker, L.; Jensen, J. H.; Margulis, C. J. J Chem Theory Comput 2007, 3, 1620–1628.
6. Tian, F.; Al-Hashimi, H. M.; Craighead, J. L.; Prestegard, J. H. J Am Chem Soc 2001, 123, 485–492.
7. Venable, R. M.; Delaglio, F.; Norris, S. E.; Freedberg, D. I. Carbohydr Res 2005, 340, 863–874.
8. Martin-Pastor, M.; Bush, C. A. Carbohydr Res 2000, 323, 147–155.
9. Valafar, H.; Prestegard, J. H. J Magn Reson 2004, 167, 228–241.
10. Azurmendi, H. F.; Bush, C. A. J Am Chem Soc 2002, 124, 2426–2427.
11. Rueckert, M.; Otting, G. J Am Chem Soc 2000, 122, 7793–7797.
12. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. J Biomol NMR 1995, 6, 277–293.
13. Johnson, B. A.; Blevins, R. A. J Biomol NMR 1994, 4, 603–614.
14. John, B. K.; Plant, D.; Hurd, R. E. J Magn Reson A 1993, 101, 113–117.
15. Willker, W.; Leibfritz, D. J Magn Reson 1992, 99, 421–425.
16. Tjandra, N.; Bax, A. Science 1997, 278, 1111–1114.
17. Bax, A. Protein Sci 2003, 12, 1–16.
18. Azurmendi, H. F.; Bush, C. A. Carbohydr Res 2002, 337, 905–915.
19. Azurmendi, H. F.; Martin-Pastor, M.; Bush, C. A. Biopolymers 2002, 63, 89–98.
20. Saupe, A. Angew Chem Int Ed Engl 1968, 7, 97–112.
21. Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. J Am Chem Soc 2001, 123, 1416–1424.
22. Zweckstetter, M.; Bax, A. J Am Chem Soc 2000, 122, 3791–3792.
23. Fernandes, M. X.; Bernado, P.; Pons, M.; Garcia de la Torre, J. J Am Chem Soc 2001, 123, 12037–12047.
24. Almond, A.; Axelsen, J. B. J Am Chem Soc 2002, 124, 9986–9987.
25. Xia, J.; Margulis, C. J Biomol NMR 2008, 42, 241–256.
26. Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. J Comput Chem 1990, 11, 440–467.
27. Allinger, N. L.; Yuh, Y. H.; Lii, J. H. J Am Chem Soc 1989, 111, 8551–8566.
28. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127–6129.
29. Pham, T. N.; Liptaj, T.; Bromek, K.; Uhrin, D. J Magn Reson 2002, 157, 200–209.
30. Martin-Pastor, M.; Bush, C. A. Biochemistry 2000, 39, 4674–4683.
31. Losonczi, J. A.; Andrec, M.; Fischer, M. W. F.; Prestegard, J. H. J Magn Reson 1999, 138, 334–342.
32. Gunawardena, S.; Fiore, C. R.; Johnson, J. A.; Bush, C. A. Biochemistry 1999, 38, 12062–12071.
33. Cagas, P.; Bush, C. A. Biopolymers 1990, 30, 1123–1138.
34. Mukhopadhyay, C.; Bush, C. A. Biopolymers 1991, 31, 1737–1746.
35. Delbaere, L. T. J.; Vandonselaar, M.; Prasad, L.; Quail, J. W.; Wilson, K. S.; Dauter, Z. J Mol Biol 1993, 230, 950–965.
36. Almond, A.; Petersen, B. O.; Duus, J. O. Biochemistry 2004, 43, 5853–5863.
37. Landersjoe, C.; Jansson, J. L. M.; Maliniak, A.; Widmalm, G. J Phys Chem B 2005, 109, 17320–17326.
38. Blackledge, M. Prog Nucl Magn Reson Spectrosc 2005, 46, 23–61.
39. Bax, A.; Kontaxis, G.; Tjandra, N. Methods Enzymol 2001, 339, 127–174.

Reviewing Editor: Eric J. Toone