J Biomol NMR (2008) 42:241–256
DOI 10.1007/s10858-008-9279-6

ARTICLE

A tool for the prediction of structures of complex sugars

Junchao Xia · Claudio Margulis

Received: 30 April 2008 / Accepted: 11 September 2008 / Published online: 25 October 2008

Abstract In two recent back to back articles(Xia et al., J Chem Theory Comput 3:1620–1628 and 1629–1643, 2007a, b) we have started to address the problem of complex oligosaccharide conformation and folding. The scheme previously presented was based on exhaustive searches in configuration space in conjunction with Nuclear Overhauser Effect (NOE) calculations and the use of a complex rotameric library that takes branching into account. NOEs are extremely useful for structural determination but only provide information about short range interactions and ordering. Instead, the measurement of residual dipolar couplings (RDC), yields information about molecular ordering or folding that is long range in nature. In this article we show the results obtained by incorporation RDC calculations into our prediction scheme. Using this new approach we are able to accurately predict the structure of six human milk sugars: LNF-1, LND-1, LNF-2, LNF-3, LNnT and LNT. Our exhaustive search in dihedral configuration space combined with RDC and NOE calculations allows for highly accurate structural predictions that, because of the non-ergodic nature of these molecules on a time scale compatible with molecular dynamics simulations, are extremely hard to obtain otherwise (Almond et al., Biochemistry 43:5853–5863, 2004). Molecular dynamics simulations in explicit solvent using as initial configurations the structures predicted by our algorithm show that the histo-blood group epitopes in these sugars are relatively rigid and that the whole family of oligosaccharides derives its conformational variability almost exclusively from their common linkage (β-D-GlcNAc-(1→3)-β-D-Gal) which can exist in two distinct conformational states. A population analysis based on the conformational variability of this flexible glycosidic link indicates that the relative population of the two distinct states varies for different human milk oligosaccharides.

Keywords Oligosaccharide · Human milk sugar · Conformational analysis · Nuclear Overhauser Effect · Residual dipolar couplings · Systematic grid search · Molecular dynamics

Introduction

Carbohydrates are involved in a large number of biological recognition phenomena such as inflammatory processes, cancer metastasis, bacterial and viral infections, as well as other carbohydrate-mediated events (Rudd et al. 2001; Dwek 1996; Seeberger and Werz 2005; Galonic and Gin 2007). Structural determination in three dimensions is crucial for understanding these recognition processes and in order to build carbohydrate-based drugs. When compared to protein folding, this field is still in its infancy (French and Brady 1990; Vliegenthart and Woods 2006), but strong interest in carbohydrate-base drugs is arising (Klyosov et al. 2006; Wong 2003). This is because of the exquisite recognition ability that chirality provides to sugars. Due to the experimental difficulty involved in crystallization and the sensitivity that biologically relevant sugars show to environments, nuclear magnetic resonance (NMR) techniques are almost the only tool for the investigation of the structure of carbohydrates in aqueous

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9279-6) contains supplementary material, which is available to authorized users.

J. Xia · C. Margulis (✉)
Department of Chemistry, University of Iowa, Iowa City, IA 52242, USA
e-mail: claudio-margulis@uiowa.edu

solution (Wormald et al. 2002; Duus et al. 2000; Bush et al. 1999; Peters and Pinto 1996). Reliable solvation models and accurate force fields as well as fast prediction algorithms are in great need in the case of sugars. Some important advances have been made in the recent past (Xia et al. 2007a, b; Imberty and Perez 2000; Woods 1998, 1996; French and Brady 1990; Vliegenthart and Woods 2006) but much remains to be done. One of the main difficulties that are present in the case of sugars but not in the case of proteins is branching. If sampling of configuration space is hard in the case of a linear polymer like a protein, it is clear that sampling is much harder when branching is present, particularly at adjacent linkage points which are common in sugars. This is why often MD or MC simulations of oligosaccharides are trapped in local minima for much longer (computer wall time) than simulations of oligopeptides of comparable number of units.

As computational power has increased, various methods have been proposed to perform conformational analysis of carbohydrates. These include but are not limited to the construction of adiabatic maps(Imberty et al. 1990a, b, 1991), the CICADA method (Koca 1998), Monte Carlo schemes (Peters et al. 1993), simulated annealing (Kiddle and Homans 1998), and genetic algorithms (Nahmany et al. 2005; Strino et al. 2005). In general, these methods search for the global energy minimum in the $\phi - \psi$ inter-residue glycosidic space in vacuo or in implicit solvents. Our results (Veluraja and Margulis 2005; Xia et al. 2007a, b) and many other computational studies (Almond 2005; Almond et al. 2004, 2001; Kirschner and Woods 2001; Naidoo and Brady 1999; Liu et al. 1997; Brady and Schmidt 1993) in explicit solvent show the importance of water mediated hydrogen-bonds in determining 3D structures in solution. Explicit solvation does matter. Unfortunately, conventional MD simulations in explicit solvent often fail to produce correct structural predictions in current simulation time scale unless initial configurations are carefully chosen. This is particularly true when dihedral angles are strongly coupled such as in the case when branching is present on adjacent linkages. For example excellent experimental work coupled with explicit solvent MD simulations (Almond et al. 2004) found that relatively small oligosaccharides can stay trapped in metastable states for times longer than 50 ns! Systems may remain essentially non-ergodic on the time scale of MD simulations and such studies are unlikely to produce correct NOE values consistent with experiment. This problem is not unique to sugars, de novo folding of proteins is not easy, but most simulations of proteins start with an initial 3D structure obtained from the Protein Data Bank. No such luxury is available for carbohydrates, although some useful tools (Engelsen et al. 1996; Imberty et al. 1990a, 1991; Bohne et al. 1998) are available.

In order to deal with this sampling problem, we have recently developed a fully automatic structural prediction tool (Xia et al. 2007a, b) for oligosaccharides and polysaccharides in solution (FSPS). This tool consists of a ring perception algorithm that automatically detects rotable dihedrals, a systematic and exhaustive coupled dihedral space search, the use of a substructure matching algorithm that recognizes a branch within a complex sugar when that branch has already been studied and stored in a database, the optimization of sterically allowed structures in the gas phase or using implicit solvation models, the calculation of NMR observables to produce a structural rank in comparison with experimental data, and explicit solvent molecular dynamics simulations for structural refinement and if desired to produce thermal averages in the vicinity of the global free energy minimum. Unlike other molecular builders (Engelsen et al. 1996; Imberty et al. 1990a, 1991; Bohne et al. 1998), our method produces a rank of unique structures according to the RMSD with respect to NMR observables such as NOEs and has been shown to be fast and accurate in predicting structures in solution without the need of expensive MD simulations. The reason why we have chosen to use a combined computational/experimental approach for our predictions is that force fields available as well as high level MP2 calculations in implicit solvent including quadratic corrections to obtain free energies are unable to correctly determine the proper conformational structure of even simple disaccharides in solution (Xia et al. 2007a, b). De novo computational predictions are extremely difficult since implicit solvent models fail even in the case of the simplest disaccharides while explicit solvent simulations are commonly unable to sample the full configuration space of sugars that are moderately complex which are commonly the ones that are biologically relevant.

It is important to emphasize that our approach is not based on NOE restrictions. We actually produce structural candidates (without the use of NOEs) that get ranked in comparison to the experimental NOE data (Xia et al. 2007a, b). While NOE restrictions may not produce a structure if only a few NOEs are detectable, it is frequently the case that our algorithm is able to distinguish between likely candidates for free energy minimum in solution using these few NOE values.

Other NMR observables complement the information that can be extracted from NOEs. Since the first demonstration of their use as a source of structural information on proteins (Tjandra and Bax 1997), RDCs (Tolman and Ruan 2006; Prestegard et al. 2004) have been used to study nucleic acids, carbohydrates, and protein–ligand interactions. NOE interactions scale as $1/r^6$ which means that they can only report on distances between atoms that are a few angstroms apart (Neuhaus and Williamson 1989). In

contrast, RDC values depend on the angle subtended between internuclear vectors and the alignment tensor in an aqueous dilute liquid crystalline medium. Since the alignment tensor is determined by the shape of the molecular object in question, RDCs are long-range in nature (Prestegard et al. 2004). RDCs and NOEs are ideally complementary in a 3D structure prediction tool for carbohydrates in solution and in this paper we present results of using both NOEs and RDCs obtained from different experimental sources to computationally predict complex oligosaccharide structures in solution using our FSPS algorithm and database.

Hundreds of oligosaccarides exist in human milk. Although the underlying mechanisms are still not clear, it is well known that breastfeeding improves the immune system of infants in comparison with formula-feeding (Newburg 2005) and recent studies have shown that lactose-derived oligosaccarides in human milk are likely to provide mechanisms of protection for infants against enteric pathogens (Newburg et al. 2005). Although conformational studies based on NOEs, *J* coupling, RDCs and molecular dynamics simulations, (Landersjo et al. 2005, 2000; Almond et al. 2004; Martin-Pastor et al. 2005; Martin-Pastor and Bush 2000a, b) have been carried out in the case of a few of these important human milk oligosaccharides much more remains to be done to shed light on their 3D conformational structure in solution and its relation with biological functions. In this article we focus on the sugars shown in Fig. 1.

Computational methods

Coarse-graining systematic search

The FSPS algorithm as well as the fully coupled rotameric topology database algorithm have been thoroughly described in two back to back articles (Xia et al. 2007a, b). In these two articles we also argued that a smart brute force exhaustive search algorithm combined with experimental input is a much more practical, fast and reliable predictor of complex oligosaccharide structures in solution than MD. This is particularly true for complex branched oligosaccharides which are relevant in biological recognition phenomena. All local minima are captured and can then be sorted according to any desired criteria for the estimation of a best or a set of best solution structures. For completion Fig. 2, shows a flow chart of the FSPS algorithm. Throughout this paper, glycosidic angles $\phi - \psi$ are defined as $\phi = H_1 - C_1 - O_x - C_x$ and $\psi = C_1 - O_x - C_x - H_x$. In our exhaustive systematic search the scanning increment for each glycosidic linkage was 10°. As in previous articles (Xia et al. 2007a, b), in order to reduce the number of conformations studied, structures have been pooled so that four adjacent points on ϕ and four adjacent points on ψ are converted into a single geometry-averaged point. This was done for each glycosidic linkage. The first dihedral angle of the longest side chain (NAc group) was also rotated with increments of 60°.

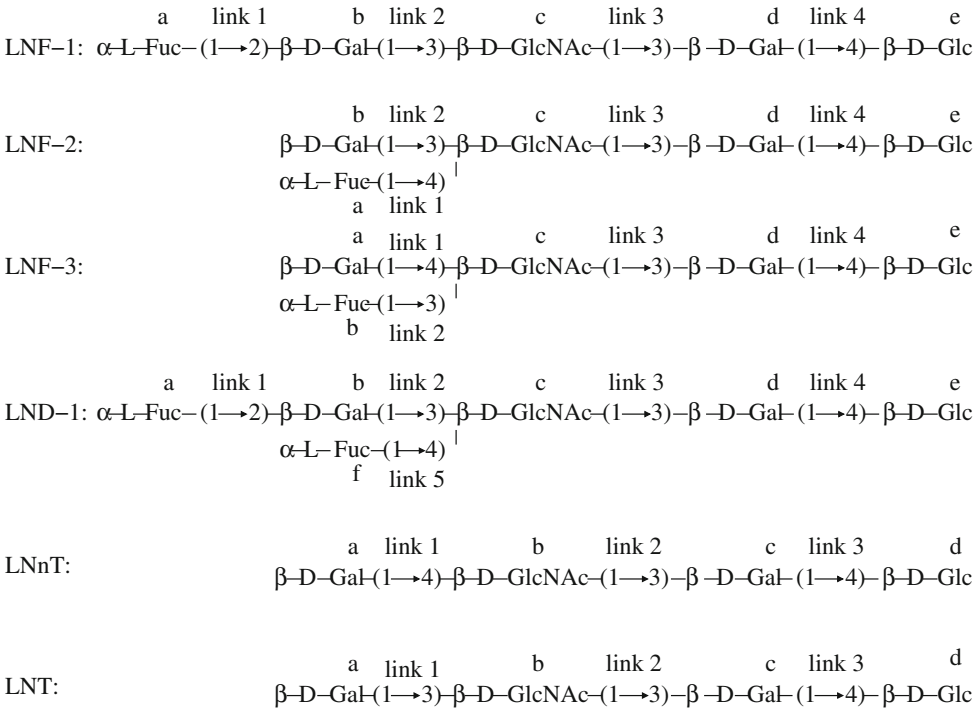

Fig. 1 Chemical sequences of human milk sugars LNF-1, LNF-2, LNF-3, LND-1, LNnT and LNT and corresponding definitions for residues and linkages used throughout this article

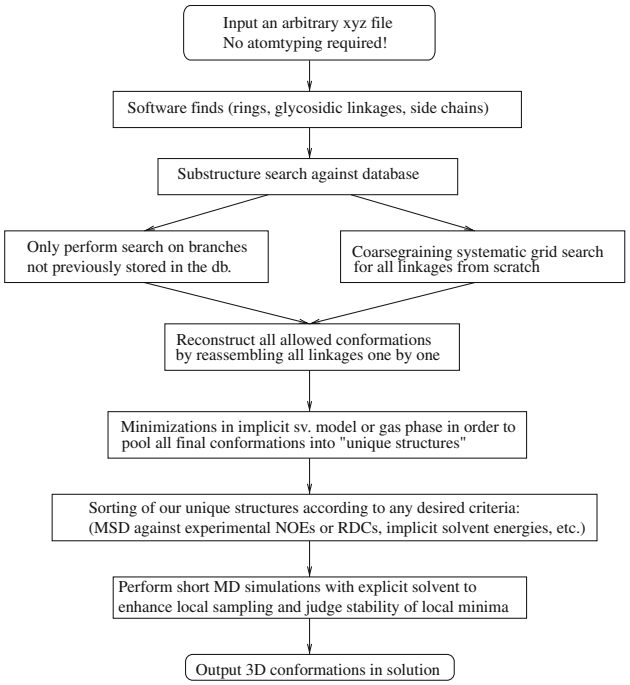

Fig. 2 Flow chart description of the FSPS algorithm

As we have shown earlier, (Xia et al. 2007a, b) global energy minima obtained in the gas phase or in implicit solvents even when using a high level MP2 ab-initio calculation do not usually coincide with the real free energy minimum in solution. Hence, an energy rank is less useful than ranks based on the calculation of experimental observables. Minimizations in the FSPS are only performed to group structures into families (not to predict the free energy minimum in solution) and therefore gas phase minimizations are in many cases sufficient for this purpose. Currently, the FSPS carries out energy minimizations by interfacing to the following software packages: TINKER (Ponder 2004; Ren and Ponder 2003; Ponder and Richards 1987), GROMACS (Van der Spoel et al. 2006, 2005; Lindahl et al. 2001) and AMBER (Case et al. 2006, 2005; Pearlman et al. 1995).

Force fields for carbohydrates lag behind those for proteins. Some of the challenges in this area are the subtle balance between inter- and intramolecular hydrogen bonding, the wide range of possible monosaccharides, functional groups and glycosidic linkages as well as having to account for the gauche, anomeric, and exoanomeric effects. (Mackerell 2004; Ponder and Case 2003) Several force fields such as MM3 (Allinger et al. 1991, 1989), AMBER (Ponder and Case 2003; Cheatham and Young 2000), CHARMM (MacKerell et al. 1998) and OPLS (Kaminski et al. 2001; Damm et al. 1997; Jorgensen et al. 1996) are available for simulating sugars. An important development based on the AMBER force field, is that of Woods and coworkers who have developed GLYCAM (Woods et al. 1995; Woods and Chappelle 2000; Basma et al. 2001). Still, for complex oligosaccharides, validation tests against NMR observables remain challenging (see for example the following interesting review articles (Mackerell 2004; Vliegenthart and Woods 2006)). The use of different force fields may result in different number of minimized structures and conformational families. This affects the predictive ability of the FSPS. A comparison of the predictive power of each of these different force fields within the context of our algorithm is discussed in the supporting materials.

Solvent effects can be included at the minimization stage by using an implicit model such as GBSA (Hawkins et al. 1996). This can be easily done since GBSA is already implemented in AMBER and TINKER. Since no implicit solvent models are implemented in GROMACS, we only performed energy minimizations in the gas phase when using the OPLS-AA force field. As shown in the supporting materials, our detailed comparison of the performance of the FSPS with different force fields and solvent models appears to indicate that OPLS-AA in the gas phase does best.

In this paper, we only performed energy minimizations using GROMACS and the OPLS-AA force field in the gas phase for all sugars. Typical CPU times for a full conformational search and energy minimizations were less than a day on a single processor (Intel(R) Pentium(R) 4 CPU 2.80 GHz) computer. All minimized structures are pooled into unique conformational families. We consider that two conformations belong to a unique family if their energy difference is $\Delta E < 5.0$ Kcal/mol and the difference in each of the glycosidic dihedral angles is $<10°$. The structure with lowest energy in each family is stored and defined as a "unique" conformer representing that family of structures. Based on these unique structures, we have produced NOE and RDC rankings.

The calculation of the Nuclear Overhauser Effects

Following Cumming and Carver (1987a, b), steady state NOE values can be computed for each of the unique structures produced by our algorithm from the set of equations below

$$R_i f(i,k) + \sum_{j \neq i} \sigma_{ij} f(j,k) = \sigma_{ik}, \quad i \neq k. \tag{1}$$

Here $f(i,k)$ is the NOE value for proton H_i on saturation of proton H_k, R_i is the total dipolar relaxation rate of H_i given by the equation

$$R_i = \frac{\mu_0^2\gamma^4\hbar^2}{(4\pi)^2 10}\left[\frac{3\tau_c}{1+(\omega_0\tau_c)^2}+\tau_c+\frac{6\tau_c}{1+(2\omega_0\tau_c)^2}\right]\sum_{j\neq i} r_{ij}^{-6} + R^s, \tag{2}$$

and σ_{ij} is the cross relaxation rate between H_i and H_j, described by

$$\sigma_{ij} = \frac{\mu_0^2\gamma^4\hbar^2}{(4\pi)^2 10}\left[\frac{6\tau_c}{1+(2\omega_0\tau_c)^2}-\tau_c\right] r_{ij}^{-6}. \tag{3}$$

In the equations above, τ_c is the overall molecular tumbling correlation time which can be found from experimental data or estimated roughly by the size of sugars (Almond et al. 2004), r_{ij} is the interproton distance, ω_0 is the Larmor frequency of a proton, μ_0 is the magnetic permeability of free space, γ is the proton gyromagnetic ratio, and R_s represents the relaxation rate due to other mechanisms which usually close to zero. The NOE value between proton i and j, $f(i,j)$, is approximately proportional to r_{ij}^{-6} or the ensemble averaged value $<r_{ij}^{-6}>$. Unfortunately, it is sometimes the case that only a few NOE signals can be detected. Furthermore, some NMR studies do not provide the actual NOE values but instead provide qualitative symbols indicating the pair of atoms that show NOE signals. It is therefore very important, in order to make an accurate prediction, to be able to compute as many observables as possible, not simply the NOEs.

The calculation of residual dipolar couplings

As opposed to the case of NOEs, RDCs provide information about long range correlations. Therefore these two techniques are highly complementary. For a weakly coupled pair of spins, i and j, the dipolar contribution to the observed resonance splitting can be described by a Hamiltonian as follows (Prestegard et al. 2000),

$$H_{ij}^D(t) = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^3} I_{iz} I_{jz} P_2(\cos\theta(t)), \tag{4}$$

where r_{ij} is the internuclear distance between spins, γ_i and γ_j are the gyromagnetic ratios of spins i and j, and I_{iz} and I_{jz} are spin angular momentum operators. The angular portion of the dipolar Hamiltonian is described using the second rank Legendre function, $P_2(\cos\theta(t))$, which is a function of the angle θ subtended by the magnetic field and the ijth internuclear vector.

In solution, RDCs stem from the time-averaged resultant of overall molecular reorientations and internal motions, denoted by

$$D_{ij}^{\text{res}} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2} < \frac{P_2(\cos\theta(t))}{r_{ij}^3} > . \tag{5}$$

This equation can be further simplified if one assumes that internal motions are negligible. In this case, RDCs are only angular dependent as below

$$D_{ij}^{\text{res}} = -\left(\frac{\mu_0}{4\pi}\right)\frac{\gamma_i\gamma_j h}{2\pi^2 r_{ij}^3} < P_2(\cos\theta(t)) > . \tag{6}$$

The average of the angular part reflects the orientational distribution of the effective ij internuclear vector relative to the magnetic field. This average will disappear under an isotropic angular distribution. However, if the molecules have magnetic susceptibility anisotropies or the molecules are surrounded by weakly ordered media such as a diluted liquid crystal, the average will not vanish and produce a measurable quantity.

Recently, RDCs in neutral diluted liquid crystal media have been measured for some carbohydrates (Landersjo et al. 2005; Martin-Pastor and Bush 2000a, b; Almond et al. 2001; Almond and Duus 2001) to determine 3D conformations in solutions. The analysis of RDCs is based on the additional assumption that the liquid crystal molecules have only steric alignment effects and do not change the internal structure of the sugar. Due to the alignment effect, sugar molecules have certain order and the residual dipolar coupling can be simplified as below (Prestegard et al. 2004, 2000)

$$D_{ij}^{\text{res}} = -\left(\frac{\mu_0}{16\pi^2}\right)\frac{\gamma_i\gamma_j\hbar}{r_{ij}^3}\kappa\left[S_{zz}(3\cos^2\theta_{ij}^z-1) + (S_{xx}-S_{yy})(\cos^2\theta_{ij}^x-\cos^2\theta_{ij}^y)\right]. \tag{7}$$

The order parameters, $S_{\alpha\alpha}$, describe the average transformation from the molecular frame of reference to the phase director of the liquid crystal medium, κ is a scaling factor related to the order parameters which can be fitted to produce best RDCs, and θ_{ij}^{α} (α = x, y, and z) are the angles between the spin-spin vectors and the molecular coordinate frame. For a much more in depth explanation the reader is referred to reference (Prestegard et al. 2004) and citations therein.

There are several methods to estimate the order parameters from the 3D conformations computationally obtained from our software. (Landersjo et al. 2005; Almond and Axelsen 2002; Azurmendi and Bush 2002). We have used the one proposed by Almond and Axelsen (Almond and Axelsen 2002) as follows,

$$(S_{xx}, S_{yy}, S_{zz}) = \left(-\frac{1}{2}-\frac{1}{2}\delta, \delta-\frac{1}{2}, 1-\frac{1}{2}\delta\right), \tag{8}$$

where

$$\delta = \frac{\rho_{yy} - \rho_{xx}}{\rho_{zz} - \rho_{xx}}. \quad (9)$$

$\rho_{\alpha\alpha}$ ($\rho_{zz} > \rho_{yy} > \rho_{xx}$) represent characteristic lengths of the ellipsoid shape of a sugar molecule and can be obtained from the square roots of the eigenvalues of the radius of gyration of the whole molecule defined as

$$R_{ij}^2 = \frac{1}{N}\sum_{r=1}^{N} x_i^r x_j^r, \quad (10)$$

where x_i^r are the positions of the N atoms in the molecule.

NOE and RDC rankings

In order to rank the different conformers generated by our FSPS we have calculated the root mean square deviation between computed RDCs (NOEs) and those reported experimentally using Eq. 11.

$$R_{\mathrm{msd}} = \sqrt{\frac{\sum_i^N \left(\frac{Q_i - Q_{0i}}{Q_{0i}}\right)^2}{N}}, \quad (11)$$

Here Q_i is our calculated RDC value for the ith spin nuclear pair and Q_{0i} is the corresponding experimental value. The summation is over all available N experimental RDCs (NOEs).

Since both RDCs and NOEs are experimentally available for some of the systems studied, comparison against experiment of our RDCs (NOEs) predicted only based on NOEs (RDCs) provides a stringent test of our algorithm and sheds light on the type of structural information that we can expect to obtain from each of these two approaches.

Molecular dynamics simulation

After ranking our unique structures obtained from the FSPS via the RMSD criterion, short explicit solvent MD simulations on the order of 10 ns were run using as initial conformations the structures with best RDCs RMSD. We know from previous experience that these MD runs will usually not visit all reasonable configuration space points, in fact the FSPS search will find regions that MD will never visit unless a simulation is started in the vicinity of that region. Our MD simulations are performed in order to test for the kinetic stability of the FSPS predicted structures, in order to sample the predicted global free energy minimum and as a tool for refinement. In general, MD refined structures have lower RMSDs with respect to experimental data than the raw FSPS predictions. MD in explicit solvent is a good refinement tool once structures matching the experimental RDCs or NOEs are found by the FSPS. This is because at this stage effects such as water mediated

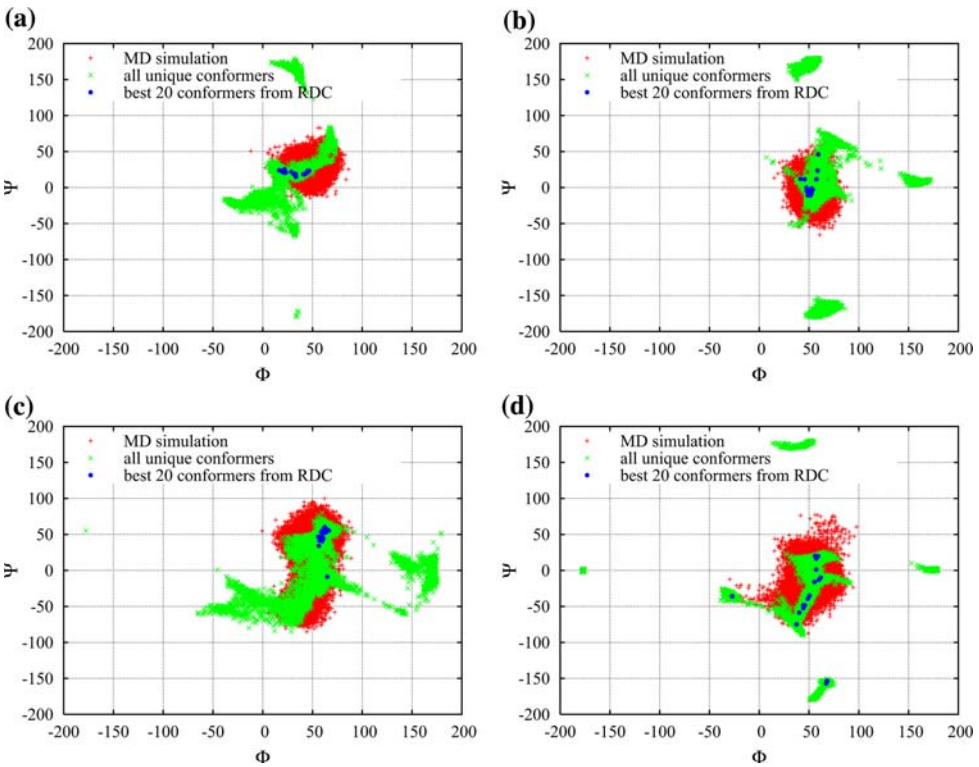

Fig. 3 The distribution of conformations in $\phi - \psi$ glycosidic space for LNF-1. The green points show our 6607 sterically allowed and energy minimized "unique structures" generated by the FSPS search. The blue points display the best 20 unique conformers based on an RDC rank with respect to experimental values. The red points correspond to conformations found by the explicit MD simulation with initial configurations corresponding to structures with best RDCs. It is important to understand that these maps are fully coupled, not every point in one of the subfigures can be compatible with points in the others. A full representation of this coupling would require a multidimensional graph. **a** Link 1, α-L-Fuc-(1→2)-β-D-Gal, **b** Link 2, β-D-Gal-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc

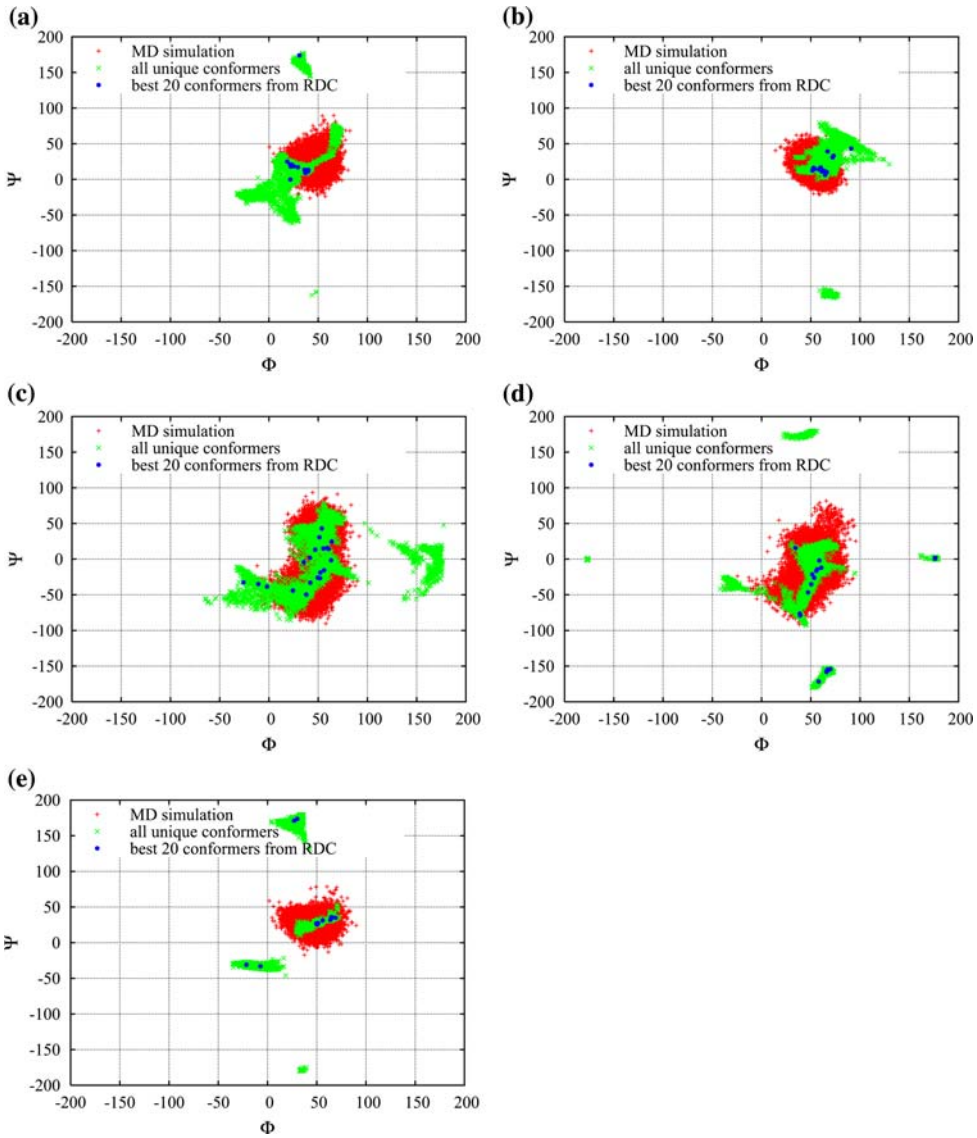

Fig. 4 Same as Fig. 3 but for LND-1. Link 1–4 are as in LNF-1 but Link 5 is an additional branch. **a** Link 1, α-L-Fuc-(1→2)-β-D-Gal, **b** Link 2, β-D-Gal-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc, **e** Link 5, α-L-Fuc-(1→4)-β-D-GlcNAc

hydrogen bonds can be incorporated. Explicit solvent MD simulations were carried out using the software GROMACS (Van der Spoel et al. 2006, 2005; Lindahl et al. 2001) with the OPLS-AA force field (Kaminski et al. 2001; Damm et al. 1997; Jorgensen et al. 1996) and the SPC (Berendsen et al. 1981) water model. In each case the simulation box was about 5 nm × 5 nm × 5 nm in size under periodic boundary conditions in the constant pressure, temperature and number of particles (NPT) ensemble. Temperature and pressure were kept at values $T =$ 300 K and $P =$ 1 atm using the Nose-Hoover thermostat (Nose 1984; Hoover 1985) and the Berendsen pressure coupling scheme (Berendsen et al. 1984). Previous to the 10 ns production runs, equilibration runs of 1 ns were performed with position constraints of all non-hydrogen sugar atoms. This was done in order to equilibrate explicit water around each of the selected unique conformers without distorting their initial structure. Our integration time step was 0.001 ps.

Conformations from our MD simulations were stored every 0.25 ps (40,000 conformational structures were saved). RDCs for all of these conformers were computed and compared with experimental values.

Results and discussions

Scoring unique structures via RDCs

Figures 3–8 show the distribution of all sterically allowed and energy minimized conformations in $\phi - \psi$ glycosidic space obtained from the FSPS search in the case of LNF-1,

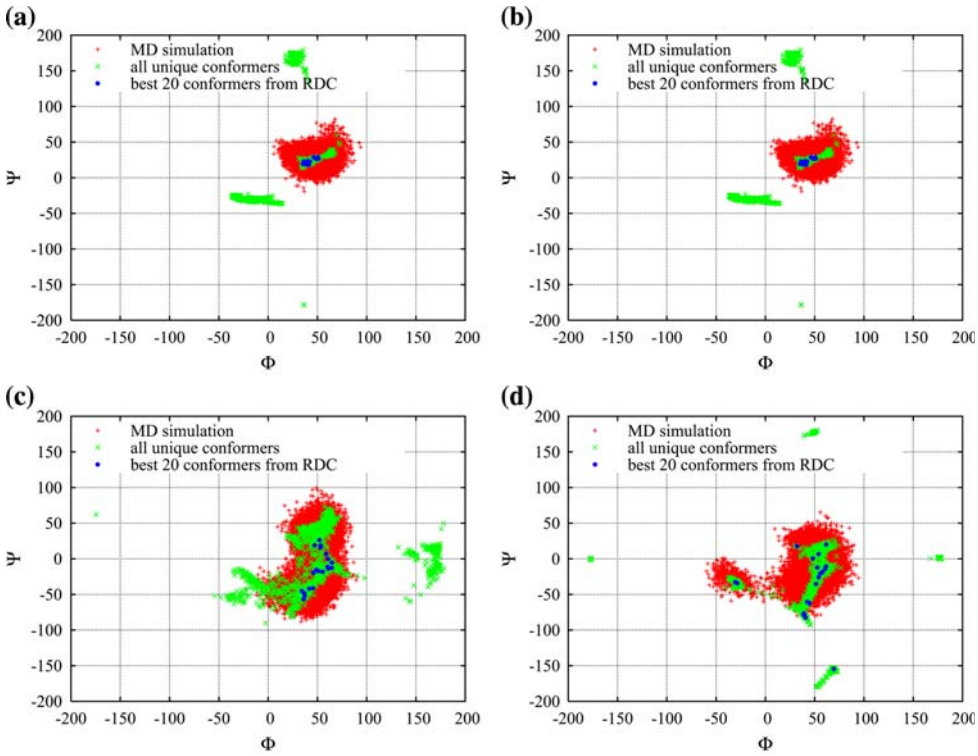

Fig. 5 Same as in Fig. 3 but for LNF-2. **a** Link 1, α-L-Fuc-(1→4)-β-D-GlcNAc, **b** Link 2, β-D-Gal-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc

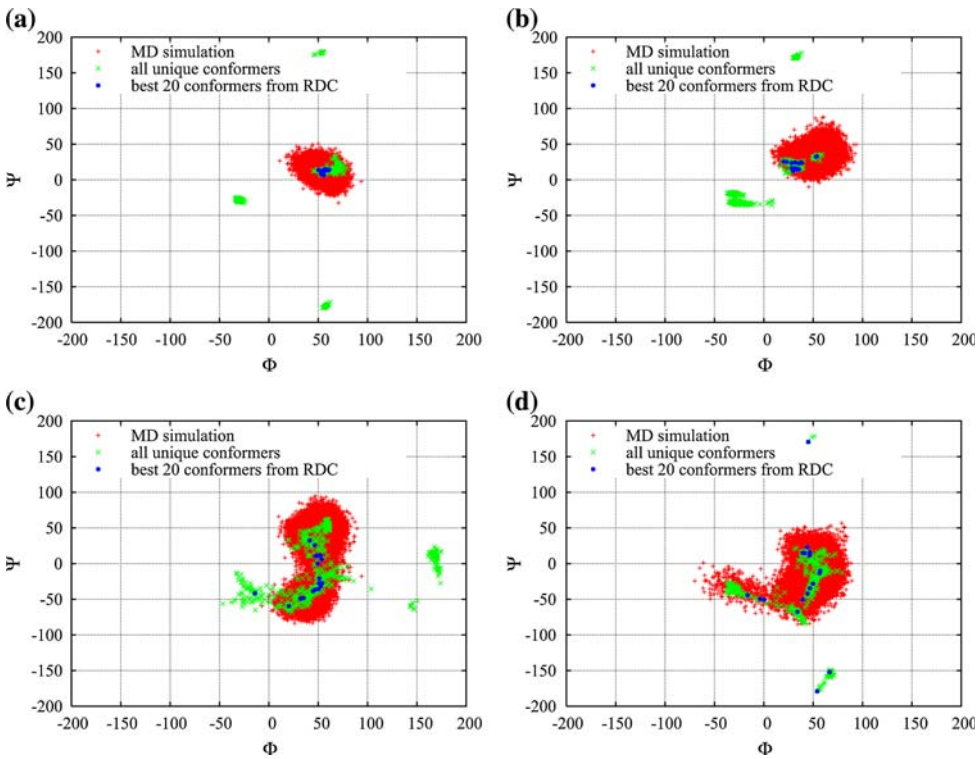

Fig. 6 Same as in Fig. 3 but for LNF-3. **a** Link 1, β-D-Gal-(1→4)-β-D-GlcNAc, **b** Link 2, α-L-Fuc-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc

LND-1, LNF-2, LNF-3, LNnT, and LNT respectively. These figures also show the best 20 conformers generated by an RDC ranking of these FSPS generated structures as well as the 40,000 conformations generated from our explicit solvent MD simulations. In all cases, the green points correspond to the local energy minima (unique

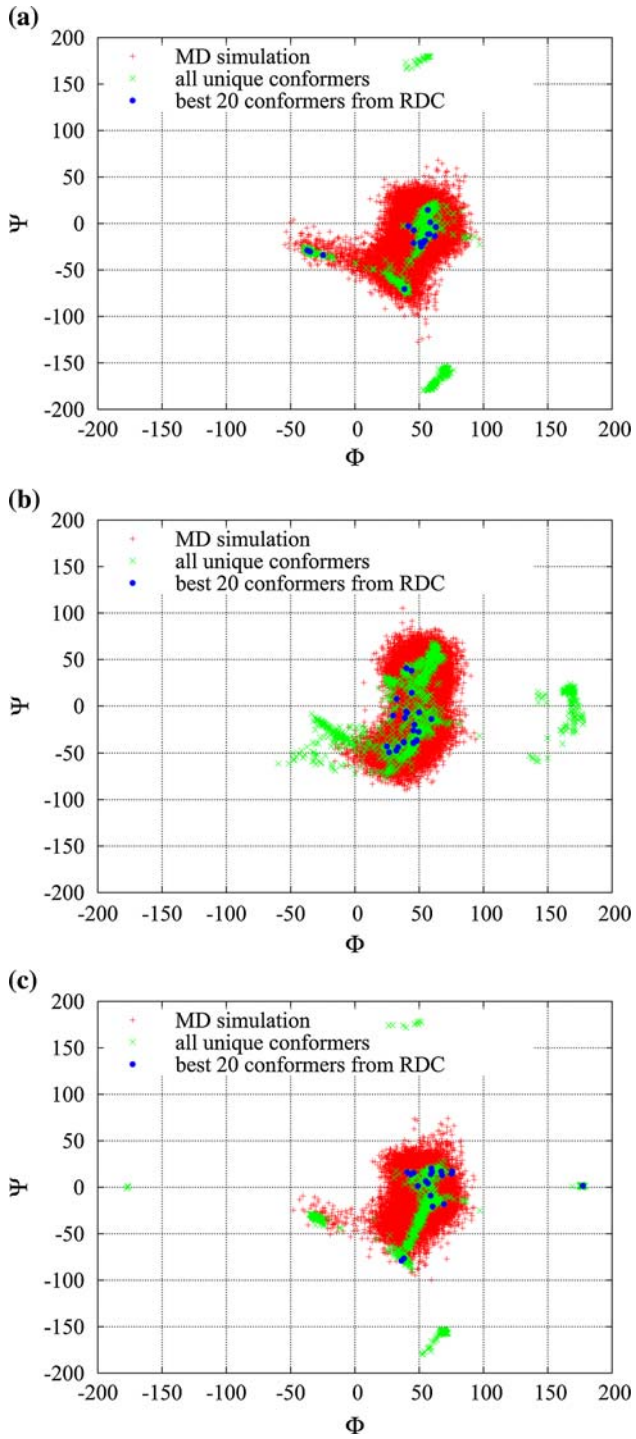

Fig. 7 Same as in Fig. 3 but for LNnT. **a** Link 1, β-D-Gal-(1→4)-β-D-GlcNAc, **b** Link 2, β-D-GlcNAc-(1→3)-β-D-Gal, **c** Link 3, β-D-Gal-(1→4)-β-D-Glc

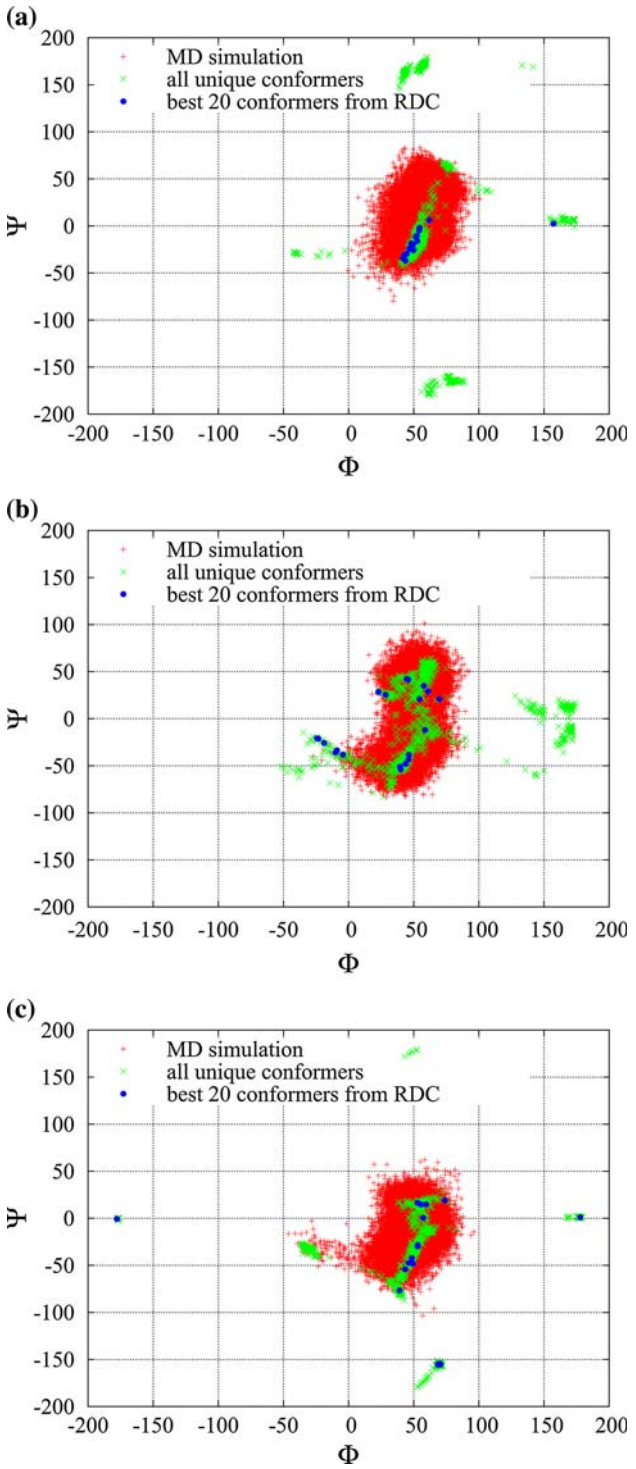

Fig. 8 Same as Fig. 3 but for LNT. **a** Link 1, β-D-Gal-(1→3)-β-D-GlcNAc, **b** Link 2, β-D-GlcNAc-(1→3)-β-D-Gal, **c** Link 3, β-D-Gal-(1→4)-β-D-Glc

structures) obtained from the FSPS as described in section "Coarse-graining systematic search" and the blue points are the best 20 unique conformers obtained from our RDC ranking with respect to the experimental data of Martin Pastor and coworkers (Martin-Pastor and Bush 2000a). The RDCs of the best FSPS and MD refined structures and their

Table 1 Residual dipolar couplings for selected spin-spin pairs in different monosaccharides (a, b, c, d, e, and f) of human milk sugars LNF-1, LNF-2, LNF-3, LND-1, LNnT and LNT as defined in Fig. 1

Spin–spin pair	LNF-1			LNF-2			LNF-3			LND-1			LNnT			LNT		
	exp	FSPS	MD	exp	FSPS	MD	exp	FSPS	MD	exp	FSPS	MD	exp	FSPS	MD	exp	FSPS	MD
H1-C1a	−18.4	−16.6	−15.7	1.8	−2.4	2.2				−26.4	−18.4	−18.6				8.6	8.4	8.0
H2-C2a	−6.7	−5.1	−8.0	−1.3	−0.3	−1.7	10.8	5.6	9.2	−2.3	−3.5	−0.9	22.8	20.3	22.0			
H3-C3a	−5.5	−2.7	−5.2	−0.8	−1.0	−0.7	9.3	6.0	8.2	−1.4	−2.1	−2.3	18.6	19.5	18.3	10.0	9.0	12.2
H4-C4a	−16.2	−17.1	−16.8	0.3	0.4	0.2	−17.2	−2.7	−15.4	−26.5	−18.5	−21.5	−21.6	−19.3	−24.4	−2.7	−2.3	−2.6
H5-C5a	−3.2	−3.9	−3.5							−0.9	−1.5	−0.4	21.0	20.5	19.9	9.1	8.0	10.0
H1-C1b				−5.5	−3.6	−5.9										19.2	16.3	17.2
H2-C2b	5.9	6.9	4.6	−3.2	−3.8	−2.5	11.1	0.4	1.1	−2.8	−3.1	−3.5				19.0	18.5	17.6
H3-C3b	5.5	7.3	6.5	−4.8	−3.9	−2.3	10.0	4.7	9.5	−3.1	−3.3	−3.8	29.4	26.6	25.1	18.5	19.1	17.5
H4-C4b	9.6	7.7	6.8	5.8	5.1	5.5	−5.6	−4.8	−5.7	19.0	10.5	14.1	26.6	24.7	26.4	17.7	15.4	15.8
H5-C5b	6.0	4.0	5.3	−5.8	−6.3	−5.6				−4.5	−2.8	−2.5	26.6	25.1	26.4	19.8	17.8	17.0
H1-C1c																16.6	17.7	17.6
H2-C2c	12.1	7.7	6.5	12.1	11.9	12.2	17.2	15.1	17.5									
H3-C3c	10.6	8.9	7.6	11.9	11.8	12.6	17.2	15.3	16.6	19.2	10.1	11.8				16.4	18.1	17.4
H4-C4c	8.7	6.3	8.5	10.3	11.4	10.4	15.6	16.1	14.1	16.3	9.3	9.3	−9.2	−9.5	−9.1	2.2	2.8	1.6
H5-C5c	9.2	6.2	9.8	11.5	11.4	11.6	17.4	15.2	17.2	18.8	9.7	10.8	23.3	27.2	25.4	14.5	18.2	17.6
H1-C1d	10.2	10.3	10.0	12.5	11.2	12.7				15.2	8.5	10.1						
H2-C2d	10.5	10.3	11.1	13.7	10.8	11.9	18.5	15.6	16.8	20.0	8.7	11.2						
H3-C3d				15.3	11.1	10.7							18.6	16.8	17.0	9.7	9.1	8.6
H4-C4d	7.5	7.9	8.2	−2.6	−3.0	−2.6	−4.2	−5.2	−4.2	0.9	2.0	1.5	16.7	18.6	17.5			
H5-C5d	9.9	8.8	7.4	12.4	11.5	12.5	17.4	16.2	17.5	16.3	8.2	10.6	15.8	17.3	16.5	11.6	9.7	12.0
H3-C3e	4.9	6.3	4.8	6.3	12.8	11.7	10.8	14.2	12.7	13.4	8.6	8.1						
H4-C4e	6.7	6.9	10.8	8.2	11.7	9.8	9.2	13.1	11.0									
H5-C5e	5.0	6.3	5.1				10.0	11.6	11.6									
H1-C1f										9.1	0.8	5.7						
H2-C2f										7.0	2.1	7.2						
H4-C4f										6.7	4.4	8.0						
RMSD		0.24	0.22		0.63	0.28		0.41	0.25		0.54	0.41		0.09	0.07		0.14	0.13

"exp" represents experimental RDCs from reference; (Martin-Pastor and Bush 2000a) "FSPS" denotes the RDCs of our best candidates obtained from ranking all unique structures using the FSPS algorithm; "MD" corresponds to the RDCs of the best MD refined structures obtained from analyzing 40,000 snapshots of our 10 ns explicit solvent MD simulations which had as initial conditions the conformations from the best ranked structures derived by the FSPS algorithm

Table 2 Comparison between the $\phi - \psi$ angles of the best conformational candidates in solution for LNF-1, LNF-2, LNF-3, LND-1, LNnT and LNT obtained from an RDC rank using the FSPS algorithm and those from our MD refined structures

		Link 1	Link 2	Link 3	Link 4	link 5
LNF-1	FSPS	(42.3, 18.5)	(51.1, −5.9)	(61.0, 57.4)	(56.1, 20.2)	
	MD	(37.4, 27.1)	(47.7, 8.1)	(65.0, 56.7)	(63.5, −11.7)	
LNF-2	FSPS	(50.8, 26.7)	(49.6, 11.9)	(60.7, −0.3)	(39.1, −76.5)	
	MD	(34.0, 35.4)	(48.2, 31.1)	(54.1, −28.6)	(48.5, −19.1)	
LNF-3	FSPS	(55.1, 9.7)	(31.1, 16.0)	(47.6, 10.7)	(49.9, −28.3)	
	MD	(54.7, 10.6)	(44.7, 20.8)	(40.8, −39.9)	(61.3, −13.9)	
LND-1	FSPS	(38.6, 13.1)	(52.5, 15.0)	(60.5, 14.4)	(53.5, −26.2)	(51.3, 27.4)
	MD	(50.9, 21.0)	(62.6, 18.1)	(54.1, 14.0)	(57.4, 10.6)	(45.0, 30.4)
LNnT	FSPS	(62.9, −3.7)	(39.9, −5.9)	(67.1, 16.9)		
	MD	(55.7, −9.0)	(41.2, −42.2)	(55.4, 26.7)		
LNT	FSPS	(48.7, −21.4)	(54.5, 21.0)	(59.5, 14.8)		
	MD	(63.2, −1.8)	(49.2, 18.1)	(58.6, 21.8)		

corresponding $\phi - \psi$ angular values are reported in Tables 1 and 2 respectively. For example in the case of LNF-1 the best structure from MD has most angles identical to the structure selected by the FSPS except that the ψ value of link 4 is slightly different. It is clear in all cases that MD simulations only visit certain limited $\phi - \psi$ regions, but since initial conditions were chosen to be those predicted by the FSPS, these regions are the ones that are relevant to the global free energy minimum. Since these systems are not ergodic on the time scale of our simulations, it is to be expected that the regions visited will depend on initial conditions. Even though MD is not a reliable sampling tool, these simulations are very useful in order to gauge the size of the accessible global free energy basin at room temperature. In general, we find that angular fluctuations appear to be within the range ±50°. It is interesting that the number of local minima that one finds when studying LND-1 is smaller than that in the case of LNF-1. These two sugars are identical except for an extra linkage in the case of LND-1. In general we find that the more branched an oligosaccharide is the less total local minima one finds.

We emphasize that the actual energies obtained from minimization, either in the gas phase or in implicit solvent, are usually not useful to accurately rank solution conformations. Our minimizations are therefore used only to group structures into families and not for predictive purposes. It is very important however, that the force field used for minimizations (either in the gas phase or in implicit solvent) yield local minima that have glycosidic angles similar to those of the global free energy minimum in solution even when these local minima may not be of lowest energy under that particular force field. If this does not occur, i.e. the force field does not produce the global free energy minimum structure as one of its local energy minima, predictions are doomed to be poor. A comparison of the performance of different available force fields is given as Supplementary material.

It is of course a possibility that NMR observables do not correlate with any single structure or set of similar structures, but instead represent information averaged over a range of widely different sets of conformations, none of them being dominant. In this case the barriers in the free energy profile in solution would necessarily have to be low

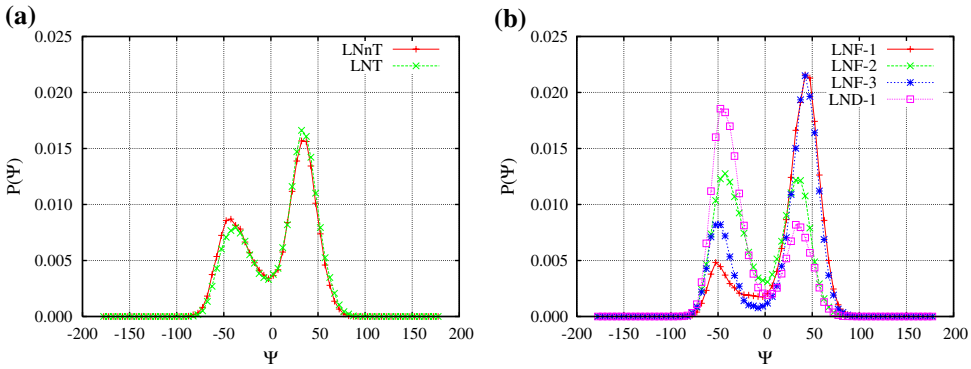

Fig. 9 Probability distributions, $P(\psi)$ (the most relevant angle providing conformational variability), for the common glycosidic linkage, β-D-GlcNAc-(1→3) -β-D-Gal, for different human milk sugars: **a** LNnT and LNT, **b** LNF-1, LNF-2, LNF-3, and LND-1. All dihedral values of ψ were extracted from 10ns MD simulations in explicit SPC water using the OPLS-AA force field

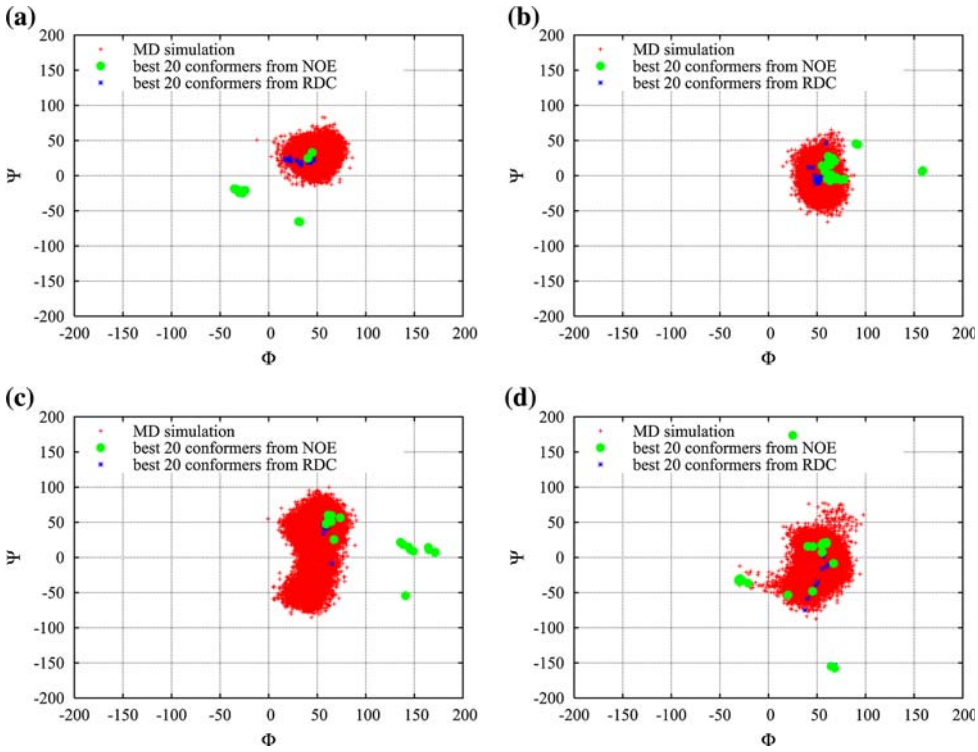

Fig. 10 Comparison between the best 20 unique conformations of LNF-1 from NOE and RDC ranks. The green points represent the best 20 unique structures from the NOE rank which analyzes our structures in comparison to the NOE data in reference (Almond et al. 2004). The blue points correspond to the best 20 conformers from the RDC rank. In this case the experimental data comes from reference (Martin-Pastor and Bush 2000a). The red points are conformations from our MD simulation using as initial structure the best FSPS prediction based on RDCs. **a** Link 1, α-L-Fuc-(1→2)-β-D-Gal, **b** Link 2, β-D-Gal-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc

rendering the free energy profile flat when compared to KT (K is the Boltzmann constant and T is the temperature). In that case, none of our unique structures could match all experimental RDCs and NOEs. In all our studies of different milk sugars, we have never found this to be the case. This goes to the core of the question of whether sugars fold or not. By this we mean whether complex oligosaccharides have a global free energy minimum that is well separated from other free energy minima by energies larger than KT. The answer to this question will of course depend on the identity of the sugar in question. Just as in the case of peptides where small oligopeptides have higher flexibility than proteins the same is likely to be true in the case of oligosaccharides. One very important thing distinguishing oligopeptides from oligosaccharides is that peptides can only form linear chains while sugars can branch. Branching causes crowding, and crowding makes dihedral rotations highly coupled and less likely to occur, increasing free energy barriers not because of energetic constraints but because of entropy. The effect of crowding is obviously also present in proteins; any particular peptidic bond can easily rotate in a dipeptide, but this is not the case in the context of a protein. Our FSPS predicts and population analysis from our MD simulations confirms that all of these sugars exist almost exclusively in two different conformational states determined by their common glycosidic linkage, β-D-GlcNAc-(1→3)-β-D-Gal.

Structural flexibility

From the results predicted by the FSPS and our MD simulations we see that structural differences mainly originate from changes in the $\phi-\psi$ values of the common linkage, β-D-GlcNAc-(1→3)-β-D-Gal (linkage 3 for the first four sugars in Fig. 1 and linkage 2 for the last two sugars). Other linkages constitute the histo-blood group epitopes: H type 3 for LNF-1, Lewis^a for LNF-2, Lewis^x for LNF-3, and Lewis^b for LND-1. The distribution of conformations obtained via MD simulations for these first four sugars (see Figs. 3–6) shows that the histo-blood group epitopes are quite rigid. This is in agreement with the results from several research groups (Imberty and Perez 2000; Martin-Pastor and Bush 2000a, b), that suggest that the histo-blood group epitopes can be represented by single well-defined conformations and not by an ensemble of conformations.

Our analysis of MD trajectories of all sugars in this paper show that the common linkage fluctuates between two distinct conformations, namely, $\psi^- \approx -50°$ and $\psi^+ \approx 50°$. Fundamentally important is that both structures (ψ^+ and ψ^-) have been predicted to be within the best 20 candidates by the FSPS as shown in Figs. 3c, 4c, 5c, 6c, 7b and 8b respectively. Our best candidate in Table 2 corresponds to one of these two conformers. This conformational variability is consistent with recent findings (Landersjo et al. 2005) based on measured RDCs and

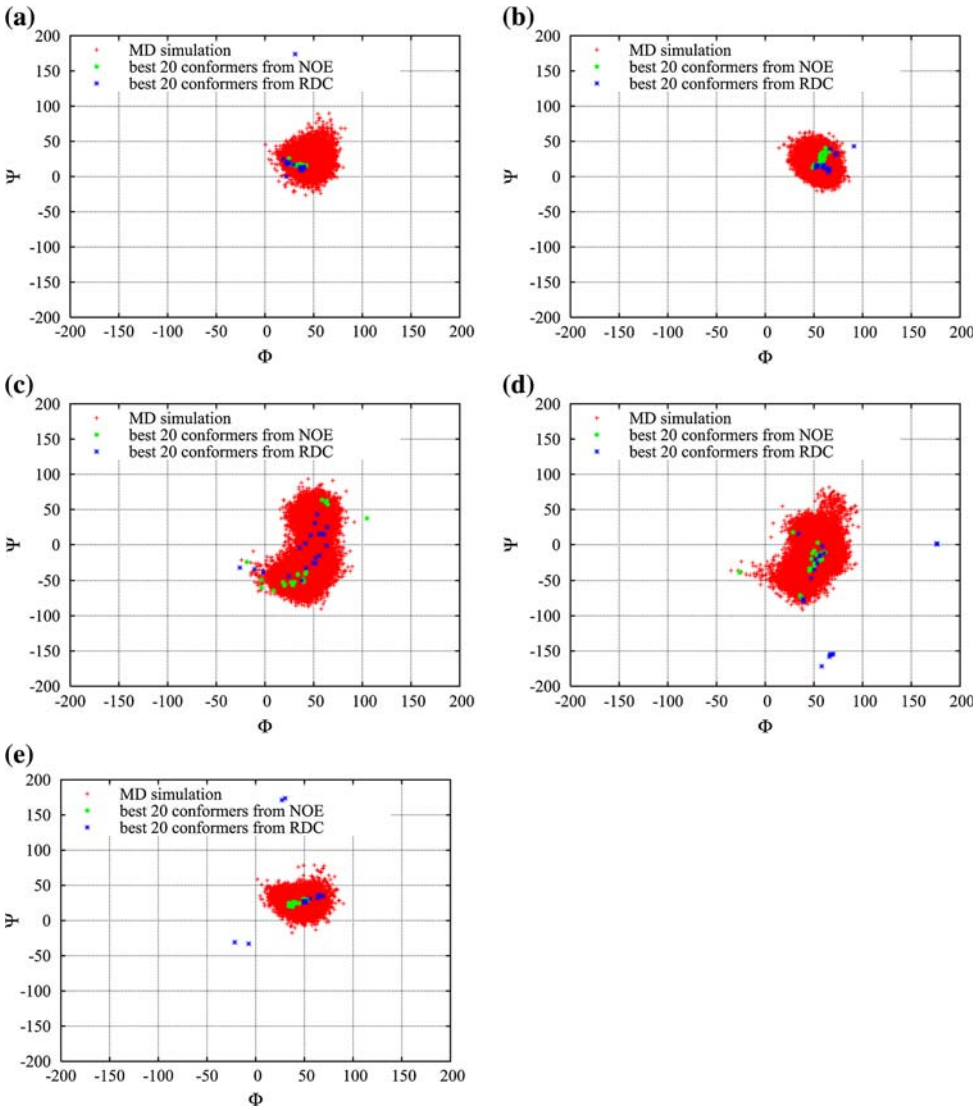

Fig. 11 Same as Fig. 10 in the case of LND-1. Link 1 through 4 are the same as those in LNF-1; Link 5 is the only additional branch. **a** Link 1, α-L-Fuc-(1→2)-β-D-Gal, **b** Link 2, β-D-Gal-(1→3)-β-D-GlcNAc, **c** Link 3, β-D-GlcNAc-(1→3)-β-D-Gal, **d** Link 4, β-D-Gal-(1→4)-β-D-Glc

molecular simulations using the CHARMM22 force field (MacKerell et al. 1998) modified for carbohydrates (Ekklund and Widmalm 2003).

We have computed the probability ($P(\psi)$) in the case of the flexible linkage for all the oligosaccharides studied. It is apparent from Fig. 9a that there is no significant difference between the angular distribution $P(\psi)$ in the case of LNT and LNnT. For both oligosaccharides, ψ^+ is the most likely conformation in solution. This result is in agreement with recent NMR and MD results obtained in the case of LNnT (Landersjo et al. 2005). A population analysis for the LNF-1, LNF-2, LNF-3, and LND-1 milk sugars is shown in Fig. 9b. It is clear from this plot that populations vary significantly across the family depending on the identity of the epitope. In the case of LNF-1 and LNF-3, the most likely state is still ψ^+ though relative populations are different than in the case of LNT and LNnT. LND-1 appears to show different behavior, with ψ^- being most likely. ψ^+ and ψ^- are almost equally likely in the case of LNF-2. Similar structural transitions for LNF-1 and LND-1 have been observed in the experiments and simulations of Almond and coworkers (Almond et al. 2004).

RDC rank versus NOE rank

While RDCs report of relatively long distance correlations, NOEs only appear when nuclei are at close distance, making these two techniques highly complementary.

In a previous article (Xia et al. 2007b), we generated an NOE ranking obtained from our FSPS algorithm for LNF-1 and LND-1. It is very instructive to compare these results with the ones derived from the RDC ranking since the two

Table 3 Distances of inter-residue proton pairs in different monosaccharides (a, b, c, d, e, and f) of human milk sugars LNF-1, LNF-2, LNF-3, LND-1, LNnT and LNT as defined in Fig. 1

Spin–spin pair	exp	FSPS	MD	Spin–spin pair	exp	FSPS	MD
	LNT				LNnT		
H1a-H3b	2.56	2.25	2.61	H1a-H4b	2.62	2.5	2.27
H1b-H3c	2.48	2.53	2.56	H1b-H3c	2.46	2.28	2.36
H1c-H4d	2.33	2.52	2.53	H1c-H4d	2.27	2.42	2.45
RMSD		0.09	0.05	RMSD		0.06	0.09
	LNF-1				LNF-2		
H1a-H2b	2.47	2.34	2.36	H1a-H4c	2.57	2.54	2.47
H5a-H2c	2.38	2.33	2.55	H5a-H2b	2.25	2.34	2.58
H1b-H3c	2.61	2.30	2.24	H1b-H3c	2.64	2.39	2.85
NHc-H1b	2.38	2.32	2.54	H1c-H3d	2.19	2.35	2.17
H1c-H3d	2.50	3.05	3.11	H1c-H4d	3.10	2.69	2.82
H1d-H4e	2.85	2.52	2.39	H1d-H4e	2.30	2.48	2.33
RMSD		0.11	0.13	RMSD		0.08	0.12
	LNF-3				LND-1		
H1b-H3c	3.08	2.21	2.50	H1a-H2b	2.36	2.21	2.44
H5b-H2a	2.81	2.41	2.72	H5a-H2c	2.97	2.44	2.62
H1a-H4c	2.43	2.39	2.36	H1b-H3c	2.53	2.44	2.52
H1a-H6c	2.89	2.42	2.19	H1f-H4c	2.53	2.64	2.64
H1c-H3d	2.22	2.34	2.43	H5f-H2b	2.40	2.32	2.14
H1d-H4e	2.69	2.26	2.22	H1c-H3d	2.44	2.54	2.51
				H1c-H4d	2.38	2.25	2.21
				H1d-H4e	2.25	2.25	2.43
RMSD		0.16	0.15	RMSD		0.08	0.12

"exp" represents distances derived from experimental NOEs in reference; (Martin-Pastor and Bush 2000a) "FSPS" denotes the distances of our best candidates obtained from ranking all unique structures using the FSPS algorithm; "MD" corresponds to the distances of the best MD refined structures obtained from analyzing 40,000 snapshots of our 10 ns explicit solvent MD simulations

techniques provide complementary long and short range information and can be used to validate the results obtained from the FSPS (Martin-Pastor and Bush 2000a, b).

Figure 10 displays conformations of the best 20 LNF-1 unique structures obtained from the RDC and NOE ranks respectively. Similar information is provided in Fig. 11 for LND-1. The experimental NOE data is from Almond and coworkers (Almond et al. 2004) while the RDCs are from the work of Martin Pastor and coworkers (Martin-Pastor and Bush 2000a). For the most part RDC and NOE predictions generate similar conformers although some deviations exist. Even though many of the best ranked structures based on NOEs appear to be very similar to those found using the RDC criterion, it is clear by looking at Figs. 10 and 11 that the overlap between the 20 best conformers from NOEs and RDCs is largest in the case of LND-1 when compared to LNF-1. This is due to the fact that LND-1 is more rigid because of the extra branch which generates a crowded linkage. Our results are consistent with recent findings (Martin-Pastor and Bush 2000a, b) in the literature which show that just a few NOEs values can be very helpful in screening structures determined from RDC fittings.

As previously discussed in this article, if oligosaccharides had a free energy landscape that was mostly flat when compared to KT it would be highly unlikely that one could find a single structure that would closely match all experimental RDCs and NOEs for each nuclei pair involved. The fact that one can find one or two such structures matching all experimental RDCs and NOEs, indicates that human milk sugar oligosaccharides have a free energy landscape with clearly defined free energy minima that are significantly deep when compared to KT.

For completion, in Table 3 we show inter-residue H–H distances in the case of LNF-1, LNF-2, LNF-3, LND-1, LNnT and LNT obtained from ranking all unique structures using the FSPS algorithm based on RDCs, as well as from our 10 ns explicit solvent MD refinement simulations. Clearly the predictions based only on RDCs match very well the distances derived from independent NOE experiments and validate our approach.

Conclusions

In this paper, we have extended our FSPS software tool to produce predictions of sugar structures not only based on energy or NOE rankings but also using RDCs. This is an important improvement because we can now make predictions based on short and long range structural order. We have tested our scheme in the prediction of the structure of a set of human milk sugars: LNF-1, LND-1, LNF-2, LNF-3, LNnT, and LNT. Predictions from NOEs and RDCs produce consistent and complementary data but more variability in the results is found for sugars that are more flexible. MD simulations in explicit solvent were used as a tool for structural refinement in which effects such as water mediated hydrogen bonding can be taken into account. By using as starting conditions the best candidates from the RDC or NOE rankings we avoid being trapped in local energy minima that are not relevant to the global free energy minimum region.

MD simulations in explicit solvent confirmed that all sugars studied have flexibility at the common linkage (β-D-GlcNAc-(1→3)-β-D-Gal) and the barrier between the two conformers ($\psi^+ \approx 50$ and $\psi^- \approx 50$) is small compared to KT since these two readily interconvert on a time scale of 10 ns at room temperature. The relative population of the two conformers depends strongly on the nature of the oligosaccharide epitope.

Acknowledgment This research was funded by Grant#05-2182 from the Roy J. Carver Charitable Trust awarded to CJM.

References

Allinger NL, Yuh YH, Lii JH (1989) Molecular mechanics—the mm3 force-field for hydrocarbons.1. J Am Chem Soc 111(23): 8551–8566

Allinger NL, Chen KS, Rahman M, Pathiaseril A (1991) Molecular mechanics (mm3) calculations on aldehydes and ketones. J Am Chem Soc 113(12):4505–4517

Almond A (2005) Towards understanding the interaction between oligosaccharides and water molecules. Carbohydr Res 340(5):907–920

Almond A, Axelsen JB (2002) Physical interpretation of residual dipolar couplings in neutral aligned media. J Am Chem Soc 124(34):9986–9987

Almond A, Duus JO (2001) Quantitative conformational analysis of the core region of n-glycans using residual dipolar couplings, aqueous molecular dynamics, and steric alignment. J Biomol NMR 20(4):351–363

Almond A, Bunkenborg J, Franch T, Gotfredsen CH, Duus JO (2001) Comparison of aqueous molecular dynamics with nmr relaxation and residual dipolar couplings favors internal motion in a mannose oligosaccharide. J Am Chem Soc 123(20):4792–4802

Almond A, Petersen BO, Duus JO (2004) Oligosaccharides implicated in recognition are predicted to have relatively ordered structures. Biochemistry 43(19):5853–5863

Azurmendi HF, Bush CA (2002) Tracking alignment from the moment of inertia tensor (tramite) of biomolecules in neutral dilute liquid crystal solutions. J Am Chem Soc 124(11): 2426–2427

Basma M, Sundara S, Calgan D, Vernali T, Woods RJ (2001) Solvated ensemble averaging in the calculation of partial atomic charges. J Comput Chem 22(11):1125–1137

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) In: Pullman B (ed) Intermolecular forces. Reidel, Dordrecht, Holland, pp 331–342

Berendsen HJC, Postma JPM, Vangunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690

Bohne A, Lang E, von der Lieth CW (1998) W3-sweet: carbohydrate modeling by internet. J Mol Model 4(1):33–43

Brady JW, Schmidt RK (1993) The role of hydrogen-bonding in carbohydrates—molecular-dynamics simulations of maltose in aqueous-solution. J Phys Chem 97(4):958–966

Bush CA, Martin-Pastor M, Imberty A (1999) Structure and conformation of complex carbohydrates of glycoproteins, glycolipids, and bacterial polysaccharides. Annu Rev Biophys Biomol Struct 28:269–293

Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The amber biomolecular simulation programs. J Comput Chem 26(16):1668–1688

Case D, Darden T, Cheatham T III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker R, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) Amber 9

Cheatham TE, Young MA (2000) Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. Biopolymers 56(4):232–256

Cumming DA, Carver JP (1987a) Reevaluation of rotamer populations for 1,6 linkages—reconciliation with potential-energy calculations. Biochemistry 26(21):6676–6683

Cumming DA, Carver JP (1987b) Virtual and solution conformations of oligosaccharides. Biochemistry 26(21):6664–6676

Damm W, Frontera A, TiradoRives J, Jorgensen WL (1997) Opls all-atom force field for carbohydrates. J Comput Chem 18(16): 1955–1970

Duus JO, Gotfredsen CH, Bock K (2000) Carbohydrate structural determination by nmr spectroscopy: modern methods and limitations. Chem Rev 100(12):4589–4614

Dwek RA (1996) Glycobiology: toward understanding the function of sugars. Chem Rev 96(2):683–720

Eklund R, Widmalm G (2003) Molecular dynamics simulations of an oligosaccharide using a force field modified for carbohydrates. Carbohydr Res 338(5):393–398

Engelsen SB, Cros S, Mackie W, Perez S (1996) A molecular builder for carbohydrates: application to polysaccharides and complex carbohydrates. Biopolymers 39(3):417–433

French AD, Brady JW (eds) (1990) Computer modelling of carbohydrate molecules, ACS symposium series. American Chemical Society, Washington

Galonic DP, Gin DY (2007) Chemical glycosylation in the synthesis of glycoconjugate antitumour vaccines. Nature 446(7139): 1000–1007

Hawkins GD, Cramer CJ, Truhlar DG (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. J Phys Chem 100(51):19824–19839

Hoover WG (1985) Canonical dynamics—equilibrium phase-space distributions. Phys Rev A 31(3):1695–1697

Imberty A, Perez S (2000) Structure, conformation, and dynamics of bioactive oligosaccharides: theoretical approaches and experimental validations. Chem Rev 100(12):4567–4588

Imberty A, Gerber S, Tran V, Perez S (1990a) Data-bank of 3-dimensional structures of disaccharides, a tool to build 3-d structures of oligosaccharides.1. oligo-mannose type n-glycans. Glycoconj J 7(1):27–54

Imberty A, Tran V, Perez S (1990b) Relaxed potential-energy surfaces of n-linked oligosaccharides—the mannose-alpha(1-3)-mannose case. J Comput Chem 11(2):205–216

Imberty A, Delage MM, Bourne Y, Cambillau C, Perez S (1991) Data-bank of 3-dimensional structures of disaccharides. 2. n-acetyllactosaminic type n-glycans—comparison with the crystal-structure of a biantennary octasaccharide. J Comput Chem 8(6):456–483

Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236

Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105(28):6474–6487

Kiddle GR, Homans SW (1998) Residual dipolar couplings as new conformational restraints in isotopically c-13-enriched oligosaccharides. FEBS Lett 436(1):128–130

Kirschner KN, Woods RJ (2001) Solvent interactions determine carbohydrate conformation. Proc Natl Acad Sci USA 98(19):10541–10545

Klyosov AA, Witczak ZJ, Platt D (eds) (2006) Carbohydrate drug design, ACS symposium series. American Chemical Society, Washington

Koca J (1998) Travelling through conformational space: an approach for analyzing the conformational behaviour of flexible molecules. Prog Biophys Mol Biol 70(2):137–173

Landersjo C, Hoog C, Maliniak A, Widmalm G (2000) Nmr investigation of a tetrasaccharide using residual dipolar couplings in dilute liquid crystalline media: effect of the environment. J Phys Chem B 104(23):5618–5624

Landersjo C, Jansson JLM, Maliniak A, Widmalm G (2005) Conformational analysis of a tetrasaccharide based on nmr spectroscopy and molecular dynamics simulations. J Phys Chem B 109(36):17320–17326

Lindahl E, Hess B, van der Spoel D (2001) Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7(8):306–317

Liu Q, Schmidt RK, Teo B, Karplus PA, Brady JW (1997) Molecular dynamics studies of the hydration of alpha,alpha-trehalose. J Am Chem Soc 119(33):7851–7862

Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25(13):1584–1604

MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616

Martin-Pastor M, Bush CA (2000a) Conformational studies of human milk oligosaccharides using h-1-c-13 one-bond nmr residual dipolar couplings. Biochemistry 39(16):4674–4683

Martin-Pastor M, Bush CA (2000b) The use of nmr residual dipolar couplings in aqueous dilute liquid crystalline medium for conformational studies of complex oligosaccharides. Carbohydr Res 323(1–4):147–155

Martin-Pastor M, Canales A, Corzana F, Asensio JL, Jimenez-Barbero J (2005) Limited flexibility of lactose detected from residual dipolar couplings using molecular dynamics simulations and steric alignment methods. J Am Chem Soc 127(10):3589–3595

Nahmany A, Strino F, Rosen J, Kemp GJL, Nyholm PG (2005) The use of a genetic algorithm search for molecular mechanics (mm3)-based conformational analysis of oligosaccharides. Carbohydr Res 340(5):1059–1064

Naidoo KJ, Brady JW (1999) Calculation of the ramachandran potential of mean force for a disaccharide in aqueous solution. J Am Chem Soc 121(10):2244–2252

Neuhaus D, Williamson M (1989) The Nuclear Overhauser Effect in structural and conformational analysis. VCH Publishers, INC, New York

Newburg DS (2005) Innate immunity and human milk. J Nutr 135(5):1308–1312

Newburg DS, Ruiz-Palacios GM, Morrow AL (2005) Human milk glycans protect infants against enteric pathogens. Annu Rev Nutr 25:37–58

Nose S (1984) A molecular-dynamics method for simulations in the canonical ensemble. Mol Phys 52(2):255–268

Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, Debolt S, Ferguson D, Seibel G, Kollman P (1995) Amber, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. Comput Phys Commun 91(1–3):1–41

Peters T, Pinto BM (1996) Structure and dynamics of oligosaccharides: nmr and modeling studies. Curr Opin Struct Biol 6(5):710–720

Peters T, Meyer B, Stuikeprill R, Somorjai R, Brisson JR (1993) A monte-carlo method for conformational-analysis of saccharides. Carbohydr Res 238:49–73

Ponder JW (2004) Tinker: Software tools for molecular design

Ponder JW, Case DA (2003) Force fields for protein simulations. Adv Protein Chem 66:27–85

Ponder JW, Richards FM (1987) An efficient newton-like method for molecular mechanics energy minimization of large molecules. J Comput Chem 8(7):1016–1024

Prestegard JH, Al-Hashimi HM, Tolman JR (2000) Nmr structures of biomolecules using field oriented media and residual dipolar couplings. Q Rev Biophys 33(4):371–424

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104(8):3519–3540

Ren P, Ponder JW (2003) Polarizable atomic multipole water model for molecular mechanics simulation. J Phys Chem B 107(24):5933–5947

Rudd PM, Elliott T, Cresswell P, Wilson IA, Dwek RA (2001) Glycosylation and the immune system. Science 291(5512):2370–2376

Seeberger PH, Werz DB (2005) Automated synthesis of oligosaccharides as a basis for drug discovery. Nat Rev Drug Discov 4(9):751–763

Strino F, Nahmany A, Rosen J, Kemp GJL, Sa-correia I, Nyholm PG (2005) Conformation of the exopolysaccharide of burkholderia cepacia predicted with molecular mechanics (mm3) using genetic algorithm search. Carbohydr Res 340(5):1019–1024

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by nmr in a dilute liquid crystalline medium. Science 278(5340):1111–1114

Tolman JR, Ruan K (2006) Nmr residual dipolar couplings as probes of biomolecular dynamics. Chem Rev 106(5):1720–1736

Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) Gromacs: fast, flexible, and free. J Comput Chem 26(16):1701–1718

Van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC (2006) Gromacs: Groningen machine for chemical simulations

Veluraja K, Margulis CJ (2005) Conformational dynamics of sialyl lewis(x) in aqueous solution and its interaction with selectine. A study by molecular dynamics. J Biomol Struct Dyn 23(1):101–111

Vliegenthart JFG, Woods RJ (eds) (2006) NMR spectroscopy and computer modeling of carbohydrates: recent advances, ACS symposium series. American Chemical Society, Washington

Wong CH (ed) (2003) Carbohydrate-based drug discovery. Wiley-VCH, Weinheim

Woods RJ (1996) The application of molecular modeling techniques to the determination of oligosaccharide solution conformations. In: Lipkowitz kB, Boyd DB (eds) Review in computational chemistry, vol 9. VCH Publishers, New York, pp 129–165

Woods RJ (1998) Computational carbohydrate chemistry: what theoretical methods can tell us. Glycoconj J 15(3):209–216

Woods RJ, Chappelle R (2000) Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates. J Mol Struct Theochem 527:149–156

Woods RJ, Dwek RA, Edge CJ, Fraserreid B (1995) Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. 1. glycam-93 parameter development. J Phys Chem 99(11):3832–3846

Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Conformational studies of oligosaccharides and glycopeptides: complementarity of nmr, x-ray crystallography, and molecular modelling. Chem Rev 102(2):371–386

Xia JC, Daly RP, Chuang FC, Parker L, Jensen JH, Margulis CJ (2007a) Sugar folding: a novel structural prediction tool for oligosaccharides and polysaccharides. J Chem Theory Comput 3(4):1620–1628

Xia JC, Daly RP, Chuang FC, Parker L, Jensen JH, Margulis CJ (2007b) Sugar folding: a novel structural prediction tool for oligosaccharides and polysaccharides. J Chem Theory Comput 3(4):1629–1643