Population distribution of flexible molecules from maximum entropy analysis using different priors as background information: application to the ϕ, ψ-conformational space of the α-(1→2)-linked mannose disaccharide present in *N*- and *O*-linked glycoproteins

Elin Säwén,[a] Tariq Massad,[b] Clas Landersjö,[a] Peter Damberg*[c] and Göran Widmalm*[a]

Received 9th March 2010, Accepted 18th May 2010
First published as an Advance Article on the web 24th June 2010
DOI: 10.1039/c003958f

The conformational space available to the flexible molecule α-D-Man*p*-(1→2)-α-D-Man*p*-OMe, a model for the α-(1→2)-linked mannose disaccharide in *N*- or *O*-linked glycoproteins, is determined using experimental data and molecular simulation combined with a maximum entropy approach that leads to a converged population distribution utilizing different input information. A database survey of the Protein Data Bank where structures having the constituent disaccharide were retrieved resulted in an ensemble with >200 structures. Subsequent filtering removed erroneous structures and gave the database (DB) ensemble having three classes of mannose-containing compounds, *viz*., *N*- and *O*-linked structures, and ligands to proteins. A molecular dynamics (MD) simulation of the disaccharide revealed a two-state equilibrium with a major and a minor conformational state, *i.e.*, the MD ensemble. These two different conformation ensembles of the disaccharide were compared to measured experimental spectroscopic data for the molecule in water solution. However, neither of the two populations were compatible with experimental data from optical rotation, NMR ^{1}H,^{1}H cross-relaxation rates as well as homo- and heteronuclear ^{3}J couplings. The conformational distributions were subsequently used as background information to generate priors that were used in a maximum entropy analysis. The resulting posteriors, *i.e.*, the population distributions after the application of the maximum entropy analysis, still showed notable deviations that were not anticipated based on the prior information. Therefore, reparameterization of homo- and heteronuclear Karplus relationships for the glycosidic torsion angles ϕ and ψ were carried out in which the importance of electronegative substituents on the coupling pathway was deemed essential resulting in four derived equations, two $^{3}J_{COCC}$ and two $^{3}J_{COCH}$ being different for the ϕ and ψ torsions, respectively. These Karplus relationships are denoted JCX/SU09. Reapplication of the maximum entropy analysis gave excellent agreement between the MD- and DB-posteriors. The information entropies show that the current reparametrization of the Karplus relationships constitutes a significant improvement. The ϕ_H torsion angle of the disaccharide is governed by the *exo*-anomeric effect and for the dominating conformation $\phi_H = -40°$ and $\psi_H = 33°$. The minor conformational state has a negative ψ_H torsion angle; the relative populations of the major and the minor states are ~3 : 1. It is anticipated that application of the methodology will be useful to flexible molecules ranging from small organic molecules to large biomolecules.

Introduction

Protein glycosylation and regulation require a highly sophisticated machinery and several glycosylated amino acids have been identified together with a dozen different monosaccharides as constituents of glycoproteins (GPs).[1,2] The main forms of glycosylation of proteins are either as *N*- or *O*-linked GPs. The *N*-linked forms are usually attached *via* an asparagine residue and are present as high-mannose, hybrid or complex glycans.[3] The *O*-linked ones are typically glycosidically linked *via* serine or threonine.[4,5] In these GPs D-mannose residues are present to a large extent in either α-(1→2)-, α-(1→3)-, or α-(1→6)-linkages and the structural element α-D-Man*p*-(1→2)-α-D-Man*p* in these *N*-linked glycans has been shown to be part of the carbohydrate structure recognized by antibodies and lectins.[6–13]

In characterization of biomolecules with respect to structure it is well-known that also flexibility and dynamics should be considered in order to describe their structure adequately.[14] In proteins there are regions that can be classified by secondary structure, namely α-helices and β-sheets, but a similar standard description for glycans is not in use. Although helical structures can be identified in glycans they refer directly to the three-dimenstional structure of the chain of sugar residues, *e.g.*, in pectin which consists of α-(1→4)-linked D-galacturonic acid residues[15] and in a GP with an *O*-linked heptasaccharide, containing five

[a]*Department of Organic Chemistry, Arrhenius Laboratory, Stockholm University, S-106 91, Stockholm, Sweden. E-mail: peter.damberg@ownit.nu, gw@organ.su.se*
[b]*Department of Biochemistry and Biophysics, Arrhenius Laboratory, Stockholm University, S-106 91, Stockholm, Sweden*
[c]*Tallinn University of Technology, Technomedicum, Ehitajate tee 5, 19086, Tallinn, Estonia*

N-acetyl-α-D-galactosamine residues, that can be described as having a well-defined structure,[16] implying limited glycosidic flexibility. In proteins and nucleic acids the use of generalized order parameters from NMR relaxation data [17] results in the description of amplitudes of motion on the pico- to nanosecond time scales and this methodology has also been used for oligosaccharides.[18,19] Since oligosaccharides are indeed flexible[20–22] and the population of two or more conformational states should be anticipated one needs to consider an ensemble of states as well as the spatial extensions in each populated state. However, to date a good deal of information is available in the literature on preferred and disallowed conformations. The conformational preferences are based on knowledge of low populations of eclipsed conformations, steric interactions leading to increase in potential energy, attractive van der Waal's interactions, hydrogen bonding, C–H/π- and charge/π-interactions. For carbohydrates, additional knowledge on stereochemistry and conformational preferences has been accumulated and described as the anomeric effect, the *gauche* effect, the Hassel-Ottar effect, and in particular for the conformational description of oligosaccharides the *exo*-anomeric effect,[23] which describes the preference of the ϕ torsion angle at the glycosidic linkage. This information has to a large extent been incorporated into molecular mechanics force fields that are used for molecular simulations, such as Monte Carlo, molecular dynamics (MD) or Langevin dynamics. However, the results from the simulations using different force fields[24,25] are, at best, similar but can also differ significantly and it may be difficult to assess which agreement between experimental observables or data derived thereof and simulation is better when one has various data, *e.g.* from solution-state NMR experiments, to compare to. An alternative approach to obtain information on conformational preferences is to utilize databases of experimentally determined structures, such as those determined in the solid state by X-ray diffraction (XRD) techniques and compiled in the Protein Data Bank (PDB) or the Cambridge Structural Database. It has been suggested that through filtering out elements of secondary structure, the conformational preferences for different amino acid types can be extracted from the PDB.[26] The corresponding approach for carbohydrates, to estimate conformational preferences for certain glycosidic linkages of samples from the database,[27] has not been extensively used. This collection of structures is anticipated to describe conformations that have a low potential energy. However, the selection of populated conformations may be limited or biased due to the evolutionary origin of the structures, but still it offers a set of experimentally determined structures that are independent from those from an empirically developed molecular mechanics force field.

Molecular spectroscopy offers a wide range of observables that can be obtained by different experimental techniques in order to attain information on structure and dynamics. These techniques are *e.g.* infrared spectroscopy, optical rotation, circular dichroism, Raman optical activity, fluorescence depolarization and NMR spectroscopy. For carbohydrates the use of circular dichroism has given information of stereochemistry[28] and infrared spectroscopy as well as Raman optical activity have resulted in descriptions of conformational preferences.[29,30] Likewise, population distributions can be assigned from optical rotation of oligosaccharides.[31,32] NMR spectroscopy has for decades been a major tool in the elucidation of conformation, flexibility and dynamics of carbohydrates, ranging from monosaccharides to polysaccharides.[33] To address the population distribution of the glycosidic torsion angles in an oligosaccharide NMR observables are measured, in particular, ^{1}H,^{1}H cross-relaxation rates from which effective interproton distances can be obtained,[34,35] transglycosidic $^3J_{C,H}$ and $^3J_{C,C}$ that can be interpreted *via* Karplus-type relationships,[36–38] and residual dipolar couplings: $d_{C,H}$, $d_{H,H}$, and $d_{C,C}$. The interpretation of the latter is more cumbersome, mainly due to the difficulty in describing the different orientational preferences for different conformations, but information on orientations between residues not directly joined by a glycosidic linkage is possible to obtain.[39–41]

Experimental data can be compared to the expectations from a molecular simulation or a database survey, in order to assess if the model is adequate. It is also possible to formulate constraints or restraints biasing a simulation to be in agreement with the experimental data. For flexible molecules the ensemble character needs to be taken into account, which has been performed either through time-averaged restrained MD simulations[42,43] or using restrained ensemble calculations.[44,45] However, the practical and correct treatment of the ensemble character is not trivial. The maximum entropy (ME) formalism provides an alternative approach to the detailed interpretation of experimental data, which may be preferred as it is an optimal interpretation in the sense that over-interpretation is systematically avoided. In the ME-analysis the information from a molecular simulation or a database survey may be considered as background information, which can be used to formulate a hypothesis about the conformational preferences, *i.e.* a *prior*, before considering the experimental data. When considering the experimental data it is often found that the hypothesis needs improvement, which can be achieved using the ME-formalism. The model after considering the experimental data is referred to as the *posterior*. The ME-formalism yields the most conservative interpretation by selecting, among the distributions consistent with the experimental data, the one which requires the smallest amount of information relative to the prior. Thereby, only information that is necessary to explain the experimental data is inferred and over-interpretation is avoided.

In the current work the conformational preferences of α-(1→2)-linked mannopyranoses, being constituents of many GPs, are investigated in aqueous solution. As a model we have chosen α-D-Man*p*-(1→2)-α-D-Man*p*-OMe (M2M) shown in Fig. 1. Information on the conformational space available to the molecule can be obtained form MD simulations that use a molecular mechanics force field and from structures present in the solid state determined by XRD crystallography. To this end we have carried out MD simulations on M2M with explicit water molecules as solvent and performed a database survey of available α-D-Man*p*-(1→2)-α-D-Man*p* structures compiled from XRD data and NMR studies of protein-ligand interactions from which well-defined structures were obtained. The conformational preferences were assessed through ME-analyses, where either the results of the database survey or the MD simulation were used to formulate priors. In the process it was found that the resulting population distributions would benefit from improved Karplus-type relationships. These developments were carried out for $^3J_{C,H}$ and $^3J_{C,C}$ for the ϕ and ψ torsion angles at glycosidic linkages in oligosaccharides, *i.e.*, in total four novel parameterizations are proposed. The ϕ/ψ population distribution of M2M was subsequently determined

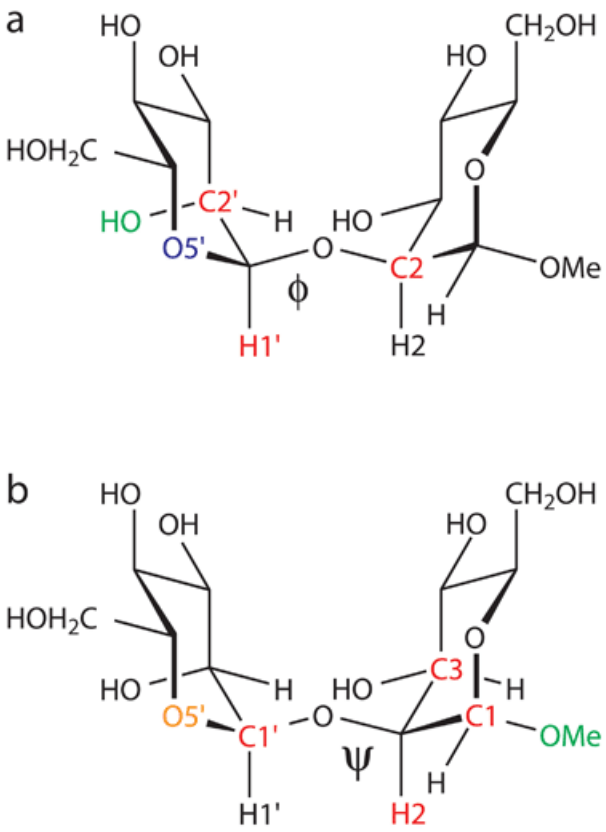

Fig. 1 Schematic of α-D-Man*p*-(1→2)-α-D-Man*p*-OMe (M2M): (a) torsion angle denoted ϕ with pertinent atoms in red that can be related *via* transglycosidic ^3J coupling constants; (b) the corresponding atoms for the torsion angle denoted ψ. The specific torsion angles are defined as follows: ϕ_H = H1′–C1′–O2–C2, $\phi_{O5'}$ = O5′–C1′–O2–C2, $\phi_{C2'}$ = C2′–C1′–O2–C2, ψ_H = H2–C2–O2–C1′, and ψ_{C1} = C1–C2–O2–C1′. The O5′ atom which decreases the ^3J coupling constants related to ϕ, *via* an inner oxygen substituent (IOS) effect, is shown in blue and electronegative groups resulting in enhanced ^3J coupling constants, *via* a constant in-plane (CIP) effect, are shown in green. A variable in-plane (VIP) effect due to O5′, highlighted in orange color, is present for the ψ torsion angle when the $\phi_{O5'}$ torsion angle is close to *anti*-periplanar (180°) resulting in enhanced $^3J(\psi)$ coupling constants. The relationships in M2M between different torsion angles are given by: $\phi_H = \phi_{O5'} - 120°$ and $\psi_H = \psi_{C1} + 120°$.

using the novel Karplus-type relationships using the two *different* priors. The resulting posterior distributions reveal the conformational preferences of M2M and the information entropies of the posteriors are used to select the one that most adequately describes ϕ/ψ population distribution.

Theory

All the spectroscopic observations, *i.e.* *J*-couplings, distance estimates from cross-relaxation and optical rotation, correspond to ensemble averages. The expectation values for a particular observation, $\langle obs \rangle$, are calculated as the integral of the state-resolved response function, *R*, describing the response from each member of the ensemble, multiplied by the population of each member of the ensemble:

$$\langle obs \rangle = \int R(\phi, \psi)\, \rho(\phi, \psi)\, d\phi d\psi \qquad (1)$$

In this formulation it is apparent that ensemble and distribution are equivalent.

In the current work three different types of data are used, *viz*., distances from NOE measurements, *J*-couplings, and optical rotation. The state-resolved response function for each of them is described below.

For NOE-derived effective distances the state-resolved response, *R*, describes the inverse distance to the power of six for each potential member of the ensemble. The expectation value of $\langle r^{-6} \rangle$ is subsequently taken to the power of −1/6 to obtain the effective distance. The functions $r^{-6}(\phi,\psi)$ were initially estimated by assuming a rigid structure except for the dihedral angles of the glycosidic linkage. The bond lengths and mannopyranose conformations were fixed at the averages from the MD-simulation (vide infra). When comparing averaged r^{-6}-distances for small regions of the Ramachandran space it was found that this approximation needed further considerations since vibrational and librational modes make the distances variable. The r^{-6}-weighting results in a larger relative impact from shorter distances. Hence, it is found that for many combinations of ϕ and ψ, the r^{-6}-averaged distance is shorter in the MD simulation than the rigid unit model predicts. In certain situations the opposite trend is observed. For $\phi,\psi \approx 0°$ the rigid unit model predicts that the H1′,H2 distance is shorter than what is observed in the MD simulation. Direct steric repulsion between H1′ and H2 tend to push the two atoms away from each other resulting in somewhat distorted bond angles for $\phi,\psi \approx 0°$. Hence, the rigid unit model may have a bias in either direction.

To obtain better averaging estimates the snapshots from the MD simulation were grouped according to ϕ,ψ torsion angles. For each 5°×5°-segment the average of r^{-6} was calculated. The difference to the rigid unit model was fitted by a cubic spline function. For non-populated regions the difference was set to zero. The correction terms were subsequently added to the rigid unit model. The resulting r^{-6}-maps were used in the analysis.

For ^3J, *R* corresponds to Karplus-type relationships:[46,47]

$$^3J = A\cos^2(\theta + \delta) + B\cos(\theta + \delta) + C \qquad (2)$$

where θ is the intervening dihedral angle and δ is a potential phase shift, while *A*, *B* and *C* are empirical constants. Initially, a parameterization derived from quantum chemical calculations for model compounds was used[48] where $A = 6.17$ Hz, $B = -0.51$ Hz, $C = 0.30$ Hz and $\delta = 0°$ for $^3J_{COCC}$ and $A = 7.49$ Hz, $B = -0.96$ Hz, $C = 0.15$ Hz and $\delta = 0°$ for $^3J_{COCH}$.

The state-resolved response function for optical rotation for M2M has been calculated by semi-empirical methods.[49] In order to cover the part of Ramachandran space potentially populated the map was extended somewhat to higher ψ_H torsion angles through linear extrapolation.

While it is straightforward to calculate the expectation values (eqn (1)) from an ensemble it is more challenging to estimate the ensemble from the observations, as many different ensembles may fit the observations. Any reasonable distribution must explain the experimental observations such that residual deviation between experiments and expectation values can be explained as experimental uncertainty and/or shortcomings in *R*. However, many distributions may pass these criteria. The maximum entropy argument is often used to select the best distribution. Out of the distributions which are consistent with the experimental observations and other known facts, in particular the distribution

should be normalized; the one which is least informative is selected to avoid over-interpretation of the data. As the least informative distribution is selected no conclusion that is not necessary to explain observations is inferred, and thereby over-interpretation is avoided. The maximum entropy approach provides a systematic way of approaching the correct distribution as new or better information becomes available. The information entropy is used as the quantitative measure of how informative a distribution is. For continuous distributions the information entropy is defined relative to a prior, P, in order to avoid an invariance problem.[50]

$$S = -\int\int \rho(\phi,\psi) \log \frac{\rho(\phi,\psi)}{P(\phi,\psi)} d\phi d\psi \quad (3)$$

When the base two logarithm is used the information entropy is a direct measure of how many bits of information one needs in order to transform the prior, $P(\phi,\psi)$, into the distribution, $\rho(\phi,\psi)$. The information entropy is zero if $\rho(\phi,\psi)$ is identical to $P(\phi,\psi)$ and negative otherwise. The prior, $P(\phi,\psi)$, should express the investigators background information, or expectation. If the investigator has no expectations a flat prior which assigns a constant probability to all conformations should be used. For M2M a flat prior is naïve or ignorant. It is very well established that bonds between sp^3-hybridized atoms preferentially populate staggered over eclipsed conformations. Steric clashes and electrostatic interactions can also be predicted. The result of a state of the art force field simulation should therefore provide a more realistic prior. By including such information as background information (prior) together with experimental data in the ME calculations, more realistic distributions of the M2M glycocidic linkage will be produced.

An alternative to continuous distribution functions is to assign weights to discrete conformations such that the ensemble explains the experimental observations. With a finite number, N, of discrete conformations the information entropy is defined as:

$$S = -\sum_{k=1}^{N} \rho_k \log \frac{\rho_k}{P_k} \quad (4)$$

where P_k is the *a priori* probability of conformation k, generally $P_k = \frac{1}{N}$.

The criterion that the distribution must be consistent with the experimental observations is formulated as one or several constraints. The observations where point estimates are available are combined into a reduced sum of squared errors, χ^2:

$$\chi^2 = \sum_{l=1}^{n} \left(\frac{\langle obs_l \rangle - obs_l}{\sigma_l} \right)^2 \quad (5)$$

where l is the pertinent index. Here, χ^2 is the sum of squared deviations between expectation values and observations divided by an uncertainty which should include both experimental uncertainty and errors from less than ideal state-resolved response functions. For the J-couplings these uncertainties are estimated to 0.6 Hz, for the cross-relaxation rates they are estimated to 30% of the experimental cross-relaxation rates and for the optical rotation it is estimated to 70 deg cm^{-2} mol^{-1}. To be consistent with observations with respect to χ^2, the distribution should correspond to a χ^2 lower than or equal to the number of observations used in the summation (*cf.* the chi-squared distribution). For certain observations, *i.e.* $^3J_{C1,C1'}$ and $r_{H2,H2'}$ only upper or lower limits are obtained. These observations therefore correspond to inequality constraints. The distribution must be normalized, which corresponds to one equality constraint. The distribution which maximizes the information entropy subject to the constraints (so called posterior) can be found by Lagrange's method as described in the data analysis section under materials and methods. The posterior is the target distribution that is used in this work to analyze the conformational preferences at the glycosidic linkage of M2M in solution.

Materials and Methods

NMR Spectroscopy and molecular simulation

NMR experiments were carried out at 315 K and a concentration of 60 mM (Shigemi tube BMS-005TV, Shigemi, Allison Park, PA, USA) in D_2O containing 20 mM phosphate buffer (pD_c = 7.1)[51] on a 600 MHz Varian Inova spectrometer equipped with a 5 mm PFG triple-resonance probe. Proton-proton cross-relaxation rates in the disaccharide M2M were measured using buildup curves from 1D ^1H,^1H DPFGSE T-ROESY experiments[52] or a 2D ^1H,^1H DPFGSE T-ROESY experiment[53] with a mixing time of 200 ms. Measurements of the transglycosidic carbon–proton coupling constants were performed with Hadamard excitations of the selected ^{13}C resonances as described previously.[54] Additional experimental NMR data were available from the literature.[55,56] Molecular dynamics simulations were performed at 315 K using CHARMM[57] and the PARM22/SU01 force field essentially as described.[58] The production period used for analysis was 21.6 ns.

Data analysis

The maximum entropy distribution, subject to the constraints is found by Lagrange's method. The Lagrangian becomes:

$$L(\rho(\phi,\psi)) = S(\rho(\phi,\psi)) + \lambda_0\left(1 - \int \rho(\phi,\psi) d\phi d\psi\right) + \lambda_1\left(\chi^2(\rho(\phi,\psi)) - n\right) + \lambda_2\left(1.3 - {}^3J^{pred}_{C1,C1'}(\rho(\phi,\psi))\right) + \lambda_3\left(3.5 - r^{pred}_{H2,H2'}\right) \quad (6)$$

The optimal distribution minimizes the Lagrangian for the correct Lagrange multipliers, λ. No closed form solution is known, but finding a numerical solution is straightforward. It may be noted that the Lagrangian is an everywhere convex function, *i.e.*, the second derivative of the Lagrangian with respect to the probability density at every combination of ϕ and ψ is positive ($\lambda_1 \geq 0$). Hence, only one minimum exists and this minimum can be found by following the gradient in the downhill direction without encountering problems with local minima.

In the current work the Lagrangian is minimized for a trial set of Lagrange multipliers. The constraints are evaluated for the resulting distribution and the Lagrange multipliers are changed accordingly. The Lagrangian is minimized for the new set of Lagrange multipliers and the procedure is repeated until the correct Lagrange multipliers are found. It should be noted that when inequality constraints are used one should check whether certain constraints are satisfied trivially, *i.e.*, when the constraint is not used explicitly the resulting distribution may anyway satisfy the constraint. This is achieved by omitting inequality constraints

from the Lagrangian by setting the corresponding Lagrange multiplier to zero.

The distribution and prior are expressed on the same grid. For the modulation of the force field simulation result a 128 × 128 grid is used to cover Ramachandran space. It may be noted that less than 10% of Ramachandran space is actually populated. When modulating the database prior a weight is assigned to each particular conformation. With the definition of the information entropy (eqn (4), *vide supra*) the information entropies are directly comparable and represent how many bits of information are needed when modifying the prior to explain the experimental data.

Results and Discussion

The glycosidic torsion angles ϕ and ψ of M2M (Fig. 1) constitute the two major degrees of freedom that alter the shape of the molecule. For the hydroxymethyl groups flexibility is indeed present since at least two rotameric states are populated for the ω torsion angle,[59] defined by O5–C5–C6–O6 atoms, in each of the two mannopyranosyl residues. In addition, some flexibility may also be present at the *O*-methyl group. In M2M the major conformation for the ϕ torsion angle is anticipated to be populated as a result of the *exo*-anomeric effect leading to $\phi_H \approx -40°$. A minor conformational state may also be populated with $\phi_H \approx +40°$, referred to as the non-*exo*-anomeric conformation. To a first approximation $\psi_H \approx 0°$ since a strong ^{1}H,^{1}H-NOE is typically observed across glycosidic linkages of oligosaccharides. A more detailed analysis reveals that two conformational states may be populated in this region for molecules similar to M2M where ψ_H is slightly positive or ψ_H is slightly negative.[60] In addition, for some oligosaccharides also an *anti*-ψ_H conformational state may be populated where $\psi_H \approx 180°$,[21] but for M2M it is anticipated that this state is populated to <1% based on the structural similarity to an α-(1→2)-linked L-rhamnose containing disaccharide.[61]

The conformational preferences of M2M were first investigated by MD simulations with explicit water molecules as solvent. Three significantly populated regions (Fig. 2a) are identified, *viz*., (i) where ϕ_H is present in the *exo*-anomeric conformation and ψ_H is positive (second quadrant), (ii) where ϕ_H has the *exo*-anomeric conformation and ψ_H is negative (third quadrant), and (iii) where ϕ_H has the non-*exo*-anomeric conformation and ψ_H is positive (first quadrant). This population distribution from the MD simulation will be used as a starting point in the subsequent analysis that will be carried out, *i.e.*, it will be one of the priors, referred to as the MD-prior.

The disaccharide element α-D-Man*p*-(1→2)-α-D-Man*p* is present in *N*- and *O*-linked glycans for which there are X-ray crystal structures as well as ligand conformation information in the PDB. We have previously shown that the conformational population distribution for the ω torsion angle is similar in the solid state and in water solution for another glycosidic linkage in *N*-linked glycans, *viz*., α-D-Man*p*-(1→6)-D-Man*p*, and the constituent disaccharide.[62] This information indicated that PDB data should be useful in generating a second prior. However, since there are inconsistencies in some of the reported structures a protocol was developed (Fig. 3) based on GLYTORSION and PDBCARE,[63–65] as well as visual inspection in order to extract structures suitable for a second prior. This procedure resulted in 11 *O*-linked glycans and 40 ligands (Fig. 2b) and 136 *N*-linked glycans (Fig. 2c) containing the α-D-Man*p*-(1→2)-α-D-Man*p* disaccharide element. In addition to the population distribution resulting from the database the crystal structure of M2M[66] was included in the generation of the second prior, referred to as the DB-prior.

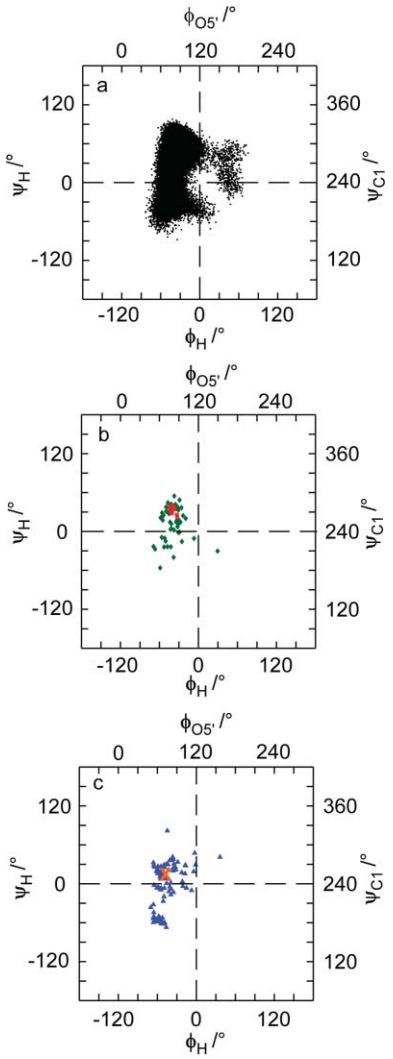

Fig. 2 Scatter plots of ϕ *vs.* ψ (in degrees) for M2M: (a) MD simulation; (b) crystal structures of *O*-linked glycans (red) and ligands (green) to proteins having the structural element α-D-Man*p*-(1→2)-α-D-Man*p*; (c) crystal structures of *N*-linked glycans having the structural element α-D-Man*p*-(1→2)-α-D-Man*p*. The conformation of M2M in a crystal structure (corrected for its absolute configuration) is denoted by an orange cross (×) in panel c ($\phi_{O5'} = 64°$ and $\psi_{C1} = -105°$). In the first quadrant $\phi_H > 0°$ and $\psi_H > 0°$; in the second quadrant $\phi_H < 0°$ and $\psi_H > 0°$; in the third quadrant $\phi_H < 0°$ and $\psi_H < 0°$; in the fourth quadrant $\phi_H > 0°$ and $\psi_H < 0°$.

The distribution of conformations in the MD simulation and in the database of determined structures are reminiscent with all conformations in the $-90° < \phi_H < 80°$ and $-90° < \psi_H < 100°$. A dominating population is present in the second quadrant and a lower population is observed in the third quadrant while the first and fourth quadrants are only sparsely populated. However, the details differ in that the mode of the distribution appears at $\phi,\psi \approx -40°,60°$ in the MD-prior, while the DB-prior displays the

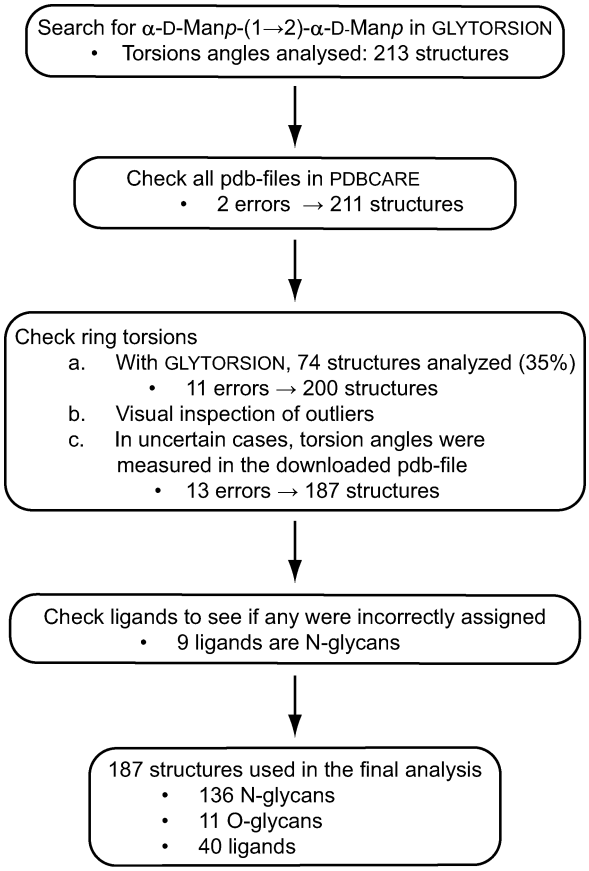

Fig. 3 Flowchart for evaluation of the structural elements α-D-Man*p*-(1→2)-α-D-Man*p* in the Protein Data Bank. The 187 selected fragments are used for generation of a database (DB) prior.

major mode at $\phi,\psi \approx -40°,30°$. In the third quadrant the DB-prior displays a distinctly populated mode at $\phi,\psi \approx -60°,-60°$ while the MD-prior rather tails towards negative ψ_H torsion angles, resulting in a significantly lower population of the third quadrant. In the MD-prior the first quadrant is more populated than the fourth while the very limited sampling in the database prohibits meaningful comparisons of relative populations of the first and fourth quadrant for the DB-prior.

The priors were tested for how well spectroscopic observations can be predicted from the distributions. Experimental data were acquired by 1D and 2D ^{1}H,^{1}H T-ROESY NMR experiments from which proton-proton cross-relaxation rates were obtained and interpreted as effective distances. Transglycosidic ^{1}H,^{13}C-coupling constants were determined as devised by Nishida *et al.*;[54] additional NMR data used in the analysis were available from the literature.[55,56] As a complement optical rotation data for M2M was included.[49,67] The experimental data are compiled in Table 1. It is worth noting that the reported value for the $^3J_{C1,C1'}$ coupling constant is indeed very small and that the *J*-coupling may appear smaller as a result of T_1-relaxation of its coupling partner as described by Harbison.[68]

It is straightforward to calculate expectation values for the priors derived from the MD simulation and the database, respectively (Table 1). However, it is evident that the agreement between the expectation values from either prior and the experimental data is less than ideal. Through the maximum entropy approach described

Table 1 NMR experimentally derived effective proton-proton distances, transglycosidic hetero- and homonuclear *J* coupling constants, and optical rotation data for M2M together with the corresponding ones calculated from the MD simulation and the database analysis

Interaction	Experiment	MD-prior	DB-prior
$r_{H1',H2}$ (Å)	$2.2 \pm 0.1^{a,b}$	2.32^g	2.30^g
$r_{H1',H1}$ (Å)	$3.0 \pm 0.2^{a,c}$	2.60^g	3.42^g
$r_{H2',H2}$ (Å)	$\geq 3.5 \pm 0.5^{a,d}$	3.71^g	3.97^g
$^3J_{H1',C2}$ (Hz)	4.1 ± 0.2	4.9^h	3.6^h
$^3J_{C1',H2}$ (Hz)	4.6 ± 0.2	2.7^h	4.7^h
$^3J_{C2',C2}$ (Hz)	3.7 ± 0.2^e	6.3^i	6.5^i
$^3J_{C1',C1}$ (Hz)	$\leq 0.7 \pm 0.2^e$	1.5^i	2.5^i
[M] (deg cm^2 dmol^{-1})	241^f	425^j	323^j

[a] Effective distance calculated using the isolated spin-pair approximation (ISPA): $r_{ij} = r_{ref}(\sigma_{ref}/\sigma_{ij})^{1/6}$. [b] Cross-relaxation rates from 1D ^{1}H,^{1}H T-ROESY NMR experiments: $\sigma_{H1',H2} = 0.148$ s^{-1} and $\sigma_{H1',H2'} = 0.068$ s^{-1}. Reference distance from MD simulation $r_{H1',H2'} = 2.51$ Å. [c] From reference 56 [d] From ISPA analysis using a 2D ^{1}H,^{1}H T-ROESY NMR experiment. [e] From reference 55. [f] Molar rotation from reference 67. [g] $r = < r^{-6} >^{-1/6}$. [h] $^3J_{C,H} = 7.49 \cos^2\theta - 0.96 \cos\theta + 0.15$; from reference 48. [i] $^3J_{C,C} = 6.17 \cos^2\theta - 0.51 \cos\theta + 0.30$; from reference 48. [j] Calculated using information from the ϕ/ψ map of reference 49.

Table 2 Populations (%) in the four quadrants from different methods

Method	1st	2nd	3rd	4th
MD-prior	0.5	91.0	8.3	0.2
DB-prior	0.5	61.0	38.0	0.5
MD-posterior	15.1	69.2	1.2	14.6
DB-posterior	0.1	83.0	2.7	14.2
MD-posterior-JCX/SU09[a]	1.1	80.9	17.1	0.9
DB-posterior-JCX/SU09	1.4	74.4	24.1	0.3

[a] JCX/SU09 denotes the herein modified Karplus-type relationships for $^3J_{C,X}$.

in the theory section the posterior distributions which are least committal to the prior and consistent with the experimental observations were determined. The posterior distributions resulting from either the MD-prior or the DB-prior are shown in Fig. 4. In the posterior distributions the population of non-*exo*-anomeric conformations, with ϕ far from –40°, is significantly increased, also for negative values of ψ. In Table 2 the populations of the four quadrants are compared for the priors and posteriors.

The population distributions of the posteriors deviated between themselves but also significantly to the priors, which indicated that convergent description based on posteriors for the conformational distribution of M2M needed further improvements in the analysis procedure. Results of a small survey of transglycosidic $^3J_{COCH}$ couplings constants related to ϕ and ψ in oligosaccharides[69] and the effect of internal electronegative substituents on transglycosidic $^3J_{COCC}$ Karplus curves[70] in combination with the discrepancy between empirical parameterizations[37,38,71,72] indicate that an investment in the improvements of the Karplus-type relationships for $^3J_{COCC}$ and $^3J_{COCH}$ is warranted.

Below two effects which may influence empirical Karplus parameterization are considered. First uncertainty and flexibility in torsion angles of the reference molecules are treated and subsequently the effects of electronegative substituents, which is further subdivided into the in-plane effect and inner substituent effect (IOS). Uncertainty and flexibility in the torsion angles of the reference molecules results in a compressed Karplus curve,[73] which

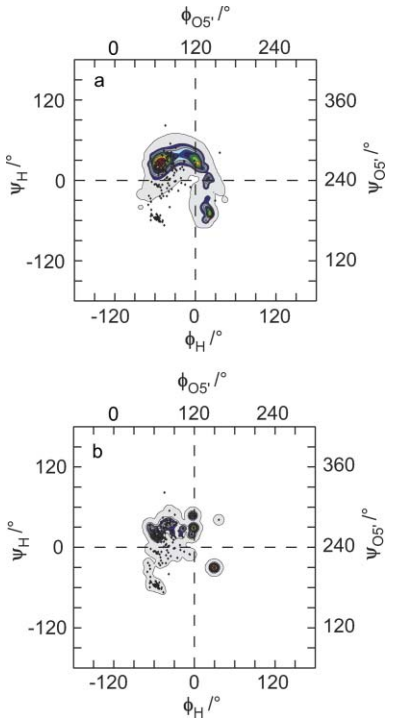

Fig. 4 Posteriors obtained from the ME-analysis of the spectroscopic data using either the prior from the MD simulation (a) or the prior from the database (b). In the analysis the following Karplus relationships[48] were used: $^3J_{C,H} = 7.49\ \cos^2\theta - 0.96\ \cos\theta + 0.15$ and $^3J_{C,C} = 6.17\ \cos^2\theta - 0.51\cos\theta + 0.30$. Contour levels are given in increments of 10% where outermost contour covers the population up 90%. Glycosidic torsion angle relationships from PDB data having the structural element α-D-Man*p*-(1→2)-α-D-Man*p* are denoted by black diamonds.

is observed when comparing empirical and quantum chemical Karplus parameterizations for transglycosidic coupling constants. For normally distributed errors in the torsion angle for the reference compounds or intrinsic flexibility resulting in normally distributed instantaneous torsion angles, it is straightforward to quantitatively predict the bias in the Karplus coefficients. Simultaneously as the uncertainty in the torsion angles will cause a bias in the empirical Karplus coefficients the uncertainty in the resulting coefficients will be increased. In order to scrutinize the Karplus parameters, simulations were performed where the bias and uncertainty in the resulting Karplus parameters are estimated. In Fig. 5 the results are shown from repeated simulations aiming at visualizing the influence from uncertainty in the torsion angles.

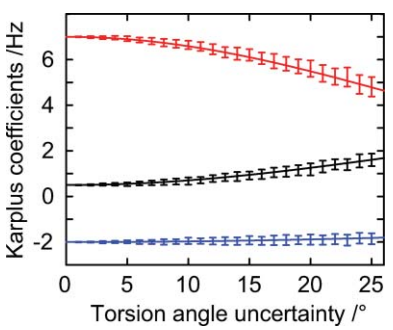

Fig. 5 The effect of the uncertainty in the torsion angles on the resulting Karplus coefficients: *A* (red), *B* (blue), and *C* (black).

To generate the results in Fig. 5, 72 *J*-couplings corresponding to evenly distributed torsion angles (5° increments) were calculated using Karplus coefficients of $A = 7$ Hz, $B = -2$ Hz and $C = 0.5$ Hz. New Karplus coefficients were subsequently fitted to the *J*-couplings after adding normally distributed errors to the 72 torsion angles. For each level of uncertainty the simulation was repeated 100 times and standard deviations in the Karplus coefficients were calculated. With increasing errors in the torsions, the Karplus coefficients are increasingly biased in a predictable way, as expected.[73] The random error decreases with increasing number of data. It is concluded that uncertainty and flexibility in the torsion angles contribute to the discrepancy between quantum chemical and empirical Karplus parameterizations relevant to the glycosidic linkage.

In addition to the uncertainty in the torsion angles for the model compounds in empirical parameterizations several lines of evidence point to other important factors influencing *J*-couplings. In particular the effects of electronegative substituents have attracted considerable attention.[74–81] For three-bond couplings involving ^{13}C spins it is convenient to distinguish between the effects from electronegative substituents directly bonded to the ^{13}C and substituents attached to either of the two central atoms in the three bond pathway, *i.e.* outer and inner substituents. For the outer substituents, Serianni and coworkers have described an in-plane effect, which is particularly important in carbohydrates.[72] The in-plane effect is an increased coupling between C and Z for the fragment O–C–X–Y–Z when the dihedral angle O–C–X–Y is *anti*-periplanar. For carbohydrates the ubiquitous hydroxyl groups have provided ample empirical evidence for the in-plane effect.[47] The empirical findings are furthermore corroborated by quantum chemical calculations.[82] In situations when the O–C–X–Y torsion angle is variable the quantum chemical calculations indicates that the coupling is largest when the O–C bond is in an *anti*-periplanar relationship to the X–Y bond. Hence, there is a dependence of the $^3J_{C,Z}$ on the O–C–X–Y torsion angle in addition to the dominating contribution form the intervening torsion angle. Interestingly, our close examination of the supplementary information of data for the protein GB3[83] shows that for serine and threonine residues density functional theory calculations indicate that ϕ-dependent couplings to C^β in these residues may be elevated when the C^β-O^γ is in an *anti*-periplanar relationship to the C^α–N bond, in exact agreement with the in-plane effect found in carbohydrates. For $^3J_{C1,C1'}$ and $^3J_{C1',H2}$ related to ψ in M2M the O5′ atom will be in the plane defined by C1′–O2–C2 when $\phi_{O5'}$ is 180°. Hence, we propose that a variable in-plane effect (VIP), related to the ϕ torsion angle, contributes to the Karplus relations mainly related to the ψ torsion angle in M2M.

For the inner substituents, it is well established, in particular for $^3J_{H,H}$, that electronegative substituents directly bonded to the central atoms in a linear sequence of four atoms generally decreases the magnitude of three-bond couplings and at the same time induce an asymmetry in the Karplus curve. In the early work of Pachler[84] it was shown that the maxima in the Karplus curve were not present at the *syn*-clinal and *anti*-periplanar relationships for $^3J_{H,H}$ in fluoroethane but shifted towards a conformation in which the electronegative substituent and the coupling partner on the vicinal carbon atom were closer to 90° for the H–C–C–F torsion. For $^3J_{H,H}$ Altona and coworkers[75–77] have developed sophisticated relations backed up by a significant

number of experimental ^3J couplings. For couplings involving heteronuclear spins the investigations are less mature. In the case of transglycosidic couplings the ring-oxygen atom at the anomeric carbon atom is to be considered an inner substituent for both ϕ-dependent coupling pathways while the ψ-dependent coupling pathways are free from inner electronegative substituents. For transglycosidic couplings empirical data show that the ϕ-dependent $^3J_{COCH}$ tends to be smaller on average compared to the ψ-dependent $^3J_{COCH}$.[69] For transglycosidic $^3J_{COCC}$ a recent quantum chemical study indicates a significant difference between the ϕ- and ψ-dependent Karplus relations.[70] Hence, it is motivated to entertain the possibility that different Karplus relations should be used for the ϕ- and ψ-dependent couplings. In order to investigate transferable patterns, the literature on ^3J for different coupling pathways was surveyed. For several coupling pathways, involving different elements, a reduced magnitude of the coupling is reported in the presence of an electronegative substituent.[78–80,85,86] When a phase-shifted Karplus-like relationship has been considered, the direction of the phase shift is consistent for all considered coupling pathways, involving different elements, except for the $^3J_{C,P}$ where the authors advocate caution.[85] The consistent pattern is that the maximum magnitude of the J-coupling is shifted from the conformation when the J-coupled nuclei have an *anti*-periplanar relationship and the electronegative substituent has a *gauche* relationship towards the conformation where the electronegative substituent is clinal, *i.e.*, shifted from approximately ±60° towards ±90°. In the literature three cases were found where dihedral angle dependent $^3J_{C,C}$-couplings have been calculated for corresponding systems in the presence and absence of inner oxygen substituents (IOS). Those are $^3J_{C',Cγ}$ for valine and threonine,[86] substituted ethyl methyl ether compounds[70] and substituted cyclohexane *vs.* substituted tetrahydropyran (compounds **5** *vs*. **2b**) of Zhao *et al.*[70] Phase-shifted Karplus-like curves were fitted to the four series from Zhao *et al.*[70] and compared to the relations for threonine and valine. It is found that in all three cases the A-coefficient is reduced by between 10 and 19%, accompanied by a phase difference of between 9 and 15°, while the changes in the B and C coefficients are smaller and less systematic. It appears that empirical parameterizations of transglycosidic ^3J couplings may be improved through consideration of i) uncertainty and flexibility in the torsion angles of the model compounds ii) in-plane effects and iii) inner oxygen substitutents (IOS).

In order to propose empirical relations for the transglycosidic $^3J_{C,C}$, where the identified perturbing factors are considered, the empirical two-parameter equation proposed by Bose *et al.* was used as a starting point.[72] This equation was obtained for selected reference molecules absent of in-plane effects, but ignoring IOS and uncertainty in the torsion angle in the model compounds. By first accounting for IOS through the use of the average change found for the three series described above, through a phase shift of 12° accompanied by a decrease of A by 7% in the presence of IOS and an increase of A of 7% in the absence of IOS, followed by accounting for uncertainty in the torsion angles of the reference compounds as described by Case *et al.*[73] using an estimated uncertainty of 15° and last accounting for potential in-plane effects, by an addition of 0.6 Hz per in-plane oxygen atom, an approximate empirical relation, where the three identified perturbing factors are accounted for, is obtained. For the case of a variable in-plane effect (VIP) where rotation around a sigma bond can position an oxygen atom in-plane, *i.e.* $^3J_{C1,C1'}$ related to ψ in M2M where O5′ would be an in-plane oxygen for $\phi_{O5'}$ close to 180°, the in-plane effect is treated by a scaled von Mise's distribution; a circular normal distribution with the variance parameter κ set to 8. The explicit formulas for the resulting relations are:

$$^3J_{C,C}\,(\phi_{C2'}) = 3.72\cos^2(\phi_{C2'} + \Delta) - 0.08 + \mathrm{CIP} \quad (7)$$

$$^3J_{C,C}\,(\psi_{C1'}) = 4.28\cos^2(\psi_{C1'}) - 0.11 + 0.6\exp(\kappa\cos(\phi_{O5'} - 180))/\exp(\kappa) + \mathrm{CIP} \quad (8)$$

The phase shift, Δ, is dependent on the stereochemistry of the sugar residue with $\Delta = -12°$ for α-D- and β-L-hexopyranosides and $\Delta = +12°$ for β-D- and α-L-hexopyranosides. The constant in-plane (CIP) effect is when present given by CIP = 0.6 Hz. It accounts for a potential in-plane effect if a terminal in-plane oxygen substituent is present, and is zero otherwise. The variable in-plane (VIP) effect is implemented by setting $\kappa = 8$ for which the in-plane effect is reduced to 50% of the full effect when the oxygen atom being terminal in the $\phi_{O5'}$ torsion angle is rotated 24° from the *anti*-periplanar relationship. The transglycosidic $^3J_{C,C}$ for M2M are plotted in Fig. 6a.

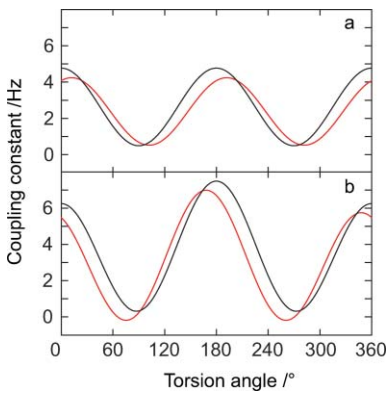

Fig. 6 Karplus relationships (JCX/SU09) for M2M: (a) $^3J_{C,C}$ for $\phi_{C2'}$ (red) and ψ_{C1} (black) with $\phi_{O5'} = 240°$; (b) $^3J_{C,H}$ for ϕ_H (red) and ψ_H (black) with $\phi_{O5'} = 240°$.

In order to propose corresponding equations for $^3J_{C,H}$ the empirical relation for $^3J_{COCH}$ published by Tvaroška *et al.*[71] was used as a starting point ($A = 5.7$ Hz, $B = -0.6$ Hz and $C = 0.5$ Hz). For the $^3J_{C,H}$ no corresponding DFT-calculations in the presence and absence of IOS could be found. Several reports hint towards an IOS-effect also for $^3J_{C,H}$. For transglycosidic $^3J_{C,H}$ couplings a limited survey[69] still revealed that on average couplings related to ϕ are smaller than couplings related to ψ, consistent with an IOS-effect. In the report by Perez *et al.*,[80] where substituent effects for homo- and heteronuclear three-bond couplings related to χ^1 dihedral angle in proteins were analyzed, a tendency for smaller $^3J_{C',Hβ}$ and $^3J_{Cγ,Hβ}$ for threonine and serine residues were found. In that report couplings related to χ^1 in flavodoxin were analyzed and the effect of substitutents was accommodated through the C-coefficient. In a more recent report[87]potentially asymmetric relations were considered, albeit without considering potential phase shifts. In analogy with Perez *et al.*,[80] the reduced magnitude of $^3J_{C,H}$ in the presence of IOS is accommodated *via* the C-coefficient in the current work. In the absence of clear evidence against a phase shift for $^3J_{C,H}$ analogy is made to $^3J_{H,H}$, $^3J_{C,C}$ and

$^3J_{F,H}$ where consistent phase shifts have been found. Furthermore, the recent study on the effect of electronegative substituents on $^3J_{C,H}$ coupling constants as a function of torsion angle[88] is consistent with our reasoning. The equation derived by Tvaroška *et al.*[71] is modified to account for the IOS by increasing the *C*-coefficient by 0.25 Hz in the absence of IOS and decreasing the *C*-coefficient by 0.25 Hz, accompanied by a 12° phase shift, in the presence of IOS. Thereafter, uncertainty in the torsion angles of the reference compounds and potential in-plane effects are accounted for as described above for $^3J_{C,C}$ The explicit relations are:

$$^3J_{C,H}(\phi_H) = 6.54\cos^2(\phi_H - \Delta) - 0.62\cos(\phi_H - \Delta) - 0.17 \quad (9)$$

$$^3J_{C,H}(\psi_H) = 6.54\cos^2(\psi_H) - 0.62\cos(\psi_H) + 0.33 + 0.6\exp(\kappa\cos(\phi_{O5'} - 180))/\exp(\kappa) \quad (10)$$

Relative to the coupling partner, *i.e.*, H1′ or C2′, the O5′ atom appears at alternate prochiral positions. Hence, the signs of the phase shift are opposite for $^3J_{C,H}$ and $^3J_{C,C}$ related to ϕ, as indicated in equations 7 and 9, respectively. The Karplus relationships formulated above are denoted JCX/SU09 hereafter. It may be noted that compared to the original empirical parameterizations by Bose *et al.*[72] and Tvaroška *et al.*[71] the JCX/SU09 equations fit the DFT-data of Cloran *et al.*[48] significantly better (data not shown). Hence, it may be argued that the currently proposed modifications to the empirical parameterization reduce the gaps between empirical and quantum chemical parameterizations. At the same time when accounting for the IOS as described above the fit to the original empirical data is somewhat better (data not shown). The transglycosidic $^3J_{C,H}$ for M2M are plotted in Fig. 6b.

The spectroscopic data are reinterpreted using the herein suggested Karplus relations (Eqn (7)–(10)) using the maximum entropy formalism with the information entropy defined relative to either the MD- or DB-prior. The resulting posteriors are displayed in Fig. 7 and the projections along ϕ_H and ψ_H are displayed in Fig. 8. When comparing the prior and the posterior the impact of the constraints from the spectroscopic observations is apparent. The posterior derived from the MD prior displays the major mode at $\phi_H,\psi_H = -40°,33°$ compared to $\phi_H,\psi_H \approx -40°,60°$ in the prior. Hence, the mode is shifted to a position similar to the mode in the database prior. At the same time the population of large ψ torsion angles where $\psi_H > 45°$, is significantly decreased in the posterior. It is noteworthy that conformations with $\psi_H > 45°$ are rare in the database with only 3 out of the 187 structures. Hence, both the spectroscopic data and the database survey indicate that the population having large ψ torsion angles is overestimated by the force field. The population in the third quadrant increased to 17%, to be intermediate between the populations in the MD (8%) and DB (38%) priors. However, the mode at $\phi_H,\psi_H = -40°,-60°$ distinguishable in the database survey is not apparent. We conclude that such details are beyond the limits of the current investigation. The agreement between the prior and posterior is generally better for the ϕ_H torsion angle distribution than for the ψ_H torsion angle distribution. The mode is shifted only slightly and the population of the non-*exo*-conformations with $\phi_H > 0°$ is 0.7 and 2% for the prior and posterior, respectively. Notably, the used CHARMM22/SU01 force field is an improved CHARMM force field where the potentials related to the ϕ torsion angle were specifically tuned for glycosidic linkages. The agreement further validates the modifications of the force field,

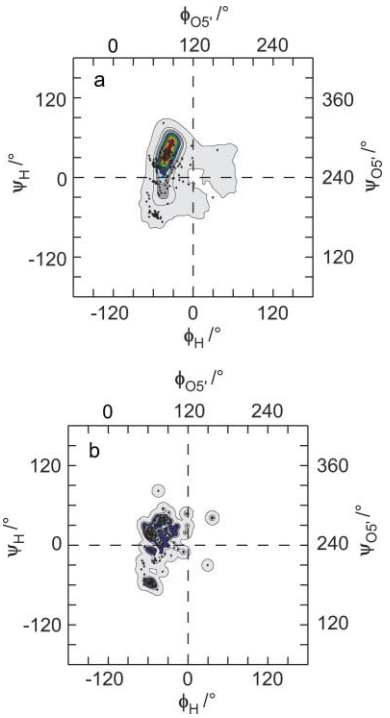

Fig. 7 Posteriors obtained from the ME-analysis of the spectroscopic data using either the prior from the MD simulation (a) or the prior from the database (b). In the analysis the modified Karplus relationships (JCX/SU09) parameterized in the current work were used (Eqn (7)–(10)). Contour levels are given in increments of 10% where outermost contour covers the population up 90%. Glycosidic torsion angle relationships from PDB data having the structural element α-D-Man*p*-(1→2)-α-D-Man*p* are denoted by black diamonds.

while the changes along the ψ torsion angle indicate the direction of future improvements of the force field. When comparing the posterior obtained from using the JCX/SU09 Karplus relations to the posterior obtained using the DFT-parameterized Karplus relations suggested by Cloran *et al.*[48] the most striking difference is the absence of a large population in the fourth quadrant. The difference illustrates the impact and importance of work aiming at improved parameterization of Karplus relations as such improvements enable the assignment of structural preferences.

The posterior resulting from the database prior and the JCX/SU09 Karplus curves is reminiscent of the posterior resulting from the MD-prior, with strikingly similar populations in the four quadrants. The population in the third quadrant is reduced to become intermediate between the two priors. The minor mode at $\phi_H,\psi_H \approx -40°,-60°$ remains in the same position but the population is significantly reduced compared to the database prior. Conformations with ψ_H close to –60° correspond to large $^3J_{C1',C1}$ as the carbon atoms adopt an *anti*-periplanar conformation. The experimental observation[55] is that of a small *J*-coupling < 0.7 Hz and hence a large population of conformations with ψ_H close to –60° can be excluded. It may be noted that the empirical $^3J_{C,C}$ used by Bose *et al.*[72] for parameterizing of the transglycosidic Karplus relationships may indicate that these couplings have the unusual property of a global maximum for *syn*-clinal relationship and a lower local maximum for the *anti*-periplanar relationship, *i.e.*, a positive prefactor for the cosθ term. Under the assumption

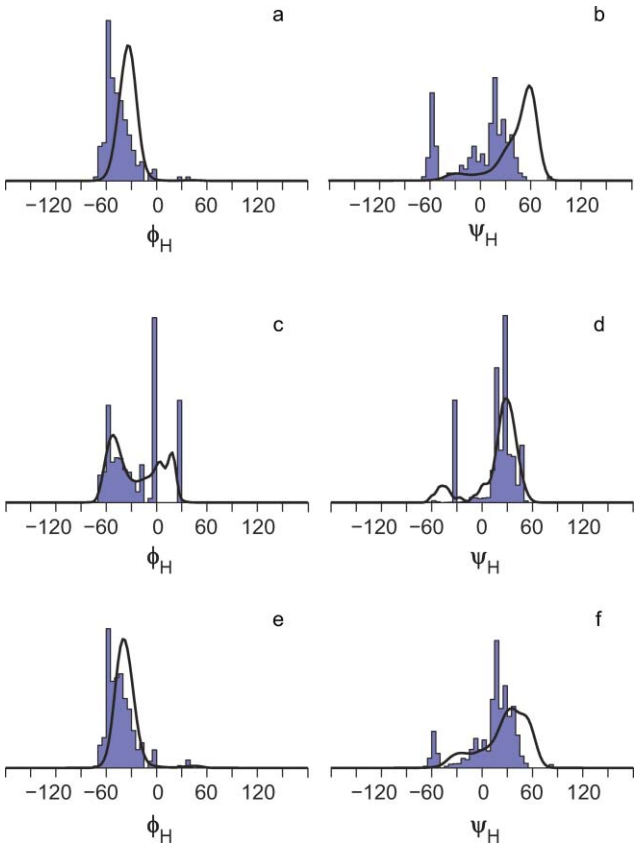

Fig. 8 Projections from the priors (a and b), from the posteriors obtained using the Karplus parameters given in the legend to Fig. 4 (c and d) and obtained when using the herein modified (JCX/SU09) Karplus relationships (e and f). Distributions from MD and DB are represented by continuous lines and histograms, respectively.

of a positive *B*-coefficient the population of conformations with $\psi_H \approx -60°$, *i.e.* $\psi_{Cl} \approx 180°$, may be somewhat increased. The occurrence of non-*exo*-anomeric conformations is similar in the DB-prior and in the posterior. The relative population of the non-*exo*-anomeric conformations is somewhat shifted towards the non-*exo*-anomeric conformation having a positive value for its ψ_H torsion angle, in agreement with the MD simulation as well as the posterior resulting from the MD-prior when the JCX/SU09 parameterization is employed. However, relative changes in populations at the level of 1% are less reliable in the current analysis. In comparison to the posterior obtained using the DFT-parameterized Karplus relations the most striking difference is the absence of a large population in the non-expected fourth quadrant, in agreement with the results based on the MD-prior. In general the JCX/SU09 parameters results in posteriors which are significantly less committal to either of the two priors, as evident from the significantly higher information entropy (Table 3). The posteriors resulting from the two different priors are in significantly closer agreement than their corresponding priors, giving credibility to the result. The small difference between the two posterior distributions reflects residual uncertainty.

In the present study of M2M the most probable conformation has its glycosidic torsion angles $\phi_H = -40°$ and $\psi_H = 33°$ (Fig. 9) as derived from experimental data, molecular simulations and database information. Notably, the high probability region has

Table 3 Expectation values and information entropies resulting from the different posteriors

Interaction	MD	DB	MD-JCX/SU09	DB-JCX/SU09
$r_{H1',H2}$ (Å)	2.2	2.2	2.3	2.2
$r_{H1',H1}$ (Å)	3.0	3.0	2.9	3.2
$r_{H2',H2}$ (Å)	3.5	3.5	3.7	3.7
$^3J_{H1',C2}$ (Hz)	4.7	4.2	4.4[a]	3.8[a]
$^3J_{C1',H2}$ (Hz)	4.6	4.9	4.1[a]	5.0[a]
$^3J_{C2',C2}$ (Hz)	4.9	4.9	4.0[a]	4.1[a]
$^3J_{C1',Cl}$ (Hz)	1.3	1.3	1.3[a]	1.3[a]
[M] (deg cm^2 dmol^{-1})	316	315	382	356
S (bit)	−4.2	−1.8	−0.4	−0.1

[a] Calculated using the herein modified Karplus relationships for $^3J_{C,X}$ referred to as JCX/SU09.

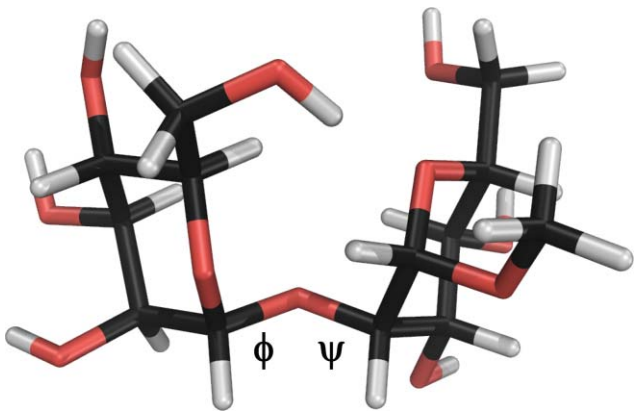

Fig. 9 Molecular model of M2M in the most probable conformation (*cf.* MD-posterior in Fig. 7a) with glycosidic torsion angles $\phi_H = -40°$ and $\psi_H = 33°$.

a positive ψ_H torsion angle. The molecule is indeed flexible with a minor conformational state having a negative ψ_H torsion angle; the population distribution between the major and the minor conformational states is approximately 3 : 1. Early conformational studies of M2M employing rigid sugar residues (HSEA approach) in the conformational analysis including Ramachandran maps identified the global energy minimum at $\phi_H = -50°$ and $\psi_H = -20°$,[67,89] where the latter torsion has a negative value. This conformation differs from that determined by XRD having $\phi_H \approx -56°$ and $\psi_H \approx 15°$, in which the latter torsion in the solid state has a positive value. Subsequent analysis of the α-(1→2)-linked disaccharide used the MM3(92) force field in which the mannosyl rings were allowed to relax for the generation of the adiabatic ϕ/ψ conformational map.[90] Two regions similar in energy were identified as being important to the solution conformation behavior of the disaccharide. In both regions the *exo*-anomeric effect prevails, *i.e.*, $\phi_H = -26°$, $\psi_H = 60°$, and $\phi_H = -42°$, $\psi_H = -21°$. In contrast to the HSEA approach the global energy minimum is now shifted significantly by 80° to a positive ψ_H torsion angle, but the potential energy difference between the wells is small, only 0.7 kcal mol^{-1}. The conformation with $\psi_H = 60°$ is stabilized by an intramolecular hydrogen bond between the hydroxymethyl groups of each mannose residue, but it was judged not to be important in aqueous solution. However, it has been suggested to be transiently present from MD-simulations with explicit solvent molecules,[56] although a persistent hydrogen

bond is not supported by NMR studies in a water:acetone mixture.[91] A recent MD-simulation of the disaccharide using the OPLS-AA force field together with flexible SPC water molecules identified two populated conformational regions, a major one centered approximately at $\phi_H = -45°$, $\psi_H = 40°$, and a minor one at $\phi_H = -45°$, $\psi_H = -20°$, with a free energy difference of about 1 kcal mol^{-1}.[92] Thus, the force field developments and the inclusion of explicit water molecules in the modeling procedures, *i.e.*, nowadays MD simulations, have changed the description of the conformational flexibility of M2M from a molecule with limited flexibility populating a single well, with a negative ψ_H torsion angle, to a major/minor population distribution in which the most highly populated state has a positive ψ_H torsion angle. The latter description of conformational flexibility is confirmed herein based on experimental data.

In our analysis we included available experimental data for M2M, *i.e.*, from NMR, XRD and optical rotation. At the time when the conformation of M2M was analyzed by optical rotation the resulting conformational preference agreed with that from the molecular mechanics HSEA approach, and the optical rotation approach was judged viable. The predicted optical rotations for the MD- and DB-posteriors deviate significantly from experimental optical rotation data although they give a converged result and the posteriors themselves only differ by a few percent from each other in their population distributions. Notably, the calculated molar rotation for M2M is quite sensitive not only to the conformational distribution of the ϕ and ψ torsion angles at the glycosidic linkage between the two sugar residues but also to the conformations of the two hydroxymethyl groups and the *O*-methyl group.[49] The hydroxymethyl groups of mannopyranosyl residues are known to populate three staggered conformations[62] and for the *O*-methyl group the non-*exo*-anomeric conformation may be populated besides the *exo*-anomeric conformation. Thus, in predicting the molar rotation for M2M a very good description of the conformational population distribution for the additional degrees of freedom is required, *i.e.*, $3 \times 3 \times 2 = 18$ conformational states may need to be considered for each glycosidic conformation. The discrepancy between the predicted molar rotation and that experimentally determined[67] may be due in part to deviations in the populations of the additional degrees of freedom of the optical rotation map[49] used in the analysis. As optical rotation provides another independent experimental parameter that may be used in conformational analysis[93] and further developments have been carried out[32] we hope that future studies on oligosaccharides similar to M2M or on the molecule itself can reconcile the current discrepancy compared to the NMR and XRD data that together with MD simulations result in a convergent description of the conformational equilibrium of this disaccharide which is a constituent of many N-linked glycans.

Conclusions

The methodology developed herein and exemplified for an α-(1→2)-linked disaccharide provides a general approach to obtain conformation population distributions of flexible molecules. The approach makes use of available background information, MD simulations and XRD data, to generate reasonable starting points for further analysis, *i.e.*, the priors. Experimental data, *viz.*, ^{1}H,^{1}H cross-relaxation rates, homo- and heteronuclear ^3J couplings and optical rotation, were used for the disaccharide in solution. Application of maximum entropy methodology in which prior information is used leads to converged population distributions from the priors and the small differences originate from uncertainties in experimental data, Karplus relationships and the optical rotation map used. It is demonstrated that a conformational equilibrium exists for the disaccharide in which the preference of the ϕ_H torsion angle is governed by the *exo*-anomeric effect, the major state has a positive ψ_H torsion angle and the minor conformational state has a negative ψ_H torsion angle. The relative populations of these states are ~3 : 1, respectively. Furthermore, it is shown that the $^3J_{C,C}$ and $^3J_{C,H}$ Karplus relationships for the ϕ and ψ torsion angles developed herein lead to significantly improved descriptions of the conformational space at the glycosidic linkage since the difference between the priors and the posteriors is very low, *i.e.*, the information entropy is close to zero. Not only did the currently developed methodology give an excellent description of the population distribution of the disaccharide, but the deviations between the MD simulation and the MD posterior may now facilitate further developments of the force field used in simulations of oligosaccharides. We foresee that the approach will be applicable to biomolecules in general and future investigations will address flexible molecules of higher complexity with a larger number of major degrees of freedom.

Acknowledgements

This work was supported by grants from the Swedish Research Council (VR), The Knut and Alice Wallenberg Foundation, Carl Tryggers Stiftelse för Vetenskaplig Forskning, Magn. Bergvalls Stiftelse, and the European Social Fund. Computing resources were kindly provided by the Center for Parallel Computers (PDC), Stockholm, Sweden.

References

1 M. R. Wormald and R. A. Dwek, *Structure*, 1999, **7**, R155–R160.
2 B. Winchester, *Glycobiology*, 2005, **15**, 1R–15R.
3 S. Thobhani, C.-T. Yuen, M. J. A. Bailey and C. Jones, *Glycobiology*, 2009, **19**, 201–211.
4 P. Gellerfors, K. Axelsson, A. Helander, S. Johansson, L. Kenne, S. Lindqvist, B. Pavlu, A. Skottner and L. Fryklund, *J. Biol. Chem.*, 1989, **264**, 11444–11449.
5 H. J. Hülsmeier, P. Deplazes, S. Naem, N. Nonaka, T. Hennet and P. Köhler, *Glycobiology*, 2010, **20**, 127–135.
6 D. N. Moothoo, B. Canan, R. A. Field and J. H. Naismith, *Glycobiology*, 1999, **9**, 539–545.
7 D. A. R. Sanders, D. N. Moothoo, J. Raftery, A. J. Howard, J. R. Helliwell and J. H. Naismith, *J. Mol. Biol.*, 2001, **310**, 875–884.
8 C. A. Bewley, *Structure*, 2001, **9**, 931–940.
9 I. Botos, B. R. O'Keefe, S. R. Shenoy, L. K. Cartner, D. M. Ratner, P. H. Seeberger, M. R. Boyd and A. Wlodawer, *J. Biol. Chem.*, 2002, **277**, 34336–34342.
10 D. A. Calarese, C. N. Scanlan, M. B. Zwick, S. Deechongkit, Y. Mimura, R. Kunert, P. Zhu, M. R. Wormald, R. Stanfield, K. H. Roux, J. W. Kelly, P. M. Rudd, R. A. Dwek, H. Katinger, D. R. Burton and I. A. Wilson, *Science*, 2003, **300**, 2065–2071.
11 D. A. Calarese, H.-K. Lee, C.-I. Huang, M. D. Best, R. D. Astronomo, R. Stanfield, H. Katinger, D. R. Burton, C.-H. Wong and I. A. Wilson, *Proc. Natl. Acad. Sci.*, 2005, **102**, 13372–13377.
12 L. Buts, A. Garcia-Pino, L. Wyns and R. Loris, *Glycobiology*, 2006, **16**, 635–640.
13 A. Garcia-Pino, L. Buts, L. Wyns, A. Imberty and R. Loris, *Plant Physiol.*, 2007, **144**, 1733–1741.
14 H.-J. Gabius, *Biochem. Soc. Trans.*, 2008, **36**, 1491–1496.

15 S. Cros, C. Garnier, M. A. V. Axelos, A. Imberty and S. Pérez, *Biopolymers*, 1996, **39**, 339–352.
16 V. Slynko, M. Schubert, S. Numao, M. Kowarik, M. Aebi and F. H.-T. Alain, *J. Am. Chem. Soc.*, 2009, **131**, 1274–1281.
17 C. K. Wang, H. J. Schirra and D. J. Craik, *PLoS ONE*, 2008, **3**, e3820-1-7.
18 A. Kjellberg, T. Rundlöf, J. Kowalewski and G. Widmalm, *J. Phys. Chem. B*, 1998, **102**, 1013–1020.
19 J. Angulo, M. Hricovíni, M. Gairi, M. Guerrini, J. L. de Paz, R. Ojeda, M. Martín-Lomas and P. M. Nieto, *Glycobiology*, 2005, **15**, 1008–1015.
20 L. Poppe, *J. Am.Chem. Soc.*, 1993, **115**, 8421–8426.
21 J. Dabrowski, T. Kožár, H. Grosskurth and N. E. Nifant'ev, *J. Am. Chem. Soc.*, 1995, **117**, 5534–5539.
22 C. Landersjö, R. Stenutz and G. Widmalm, *J. Am. Chem. Soc.*, 1997, **119**, 8695–8698.
23 H. Thøgersen, R. U. Lemieux, K. Bock and B. Meyer, *Can. J. Chem.*, 1982, **60**, 44–57.
24 K. N. Kirschner, A. B. Yongve, S. M. Tschampel, J. González-Outeriño, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2007, **29**, 622–655.
25 O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, Jr., *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.
26 M. B. Swindells, M. W. MacArthur and J. M. Thornton, *Nat. Struct. Biol.*, 1995, **2**, 596–603.
27 A. D. French, G. P. Johnson, A.-M. Kelterer, M. K. Dowd and C. J. Cramer, *Int. J. Quant. Chem.*, 2001, **84**, 416–425.
28 M. Andersson, L. Kenne, R. Stenutz and G. Widmalm, *Carbohydr. Res.*, 1994, **254**, 35–41.
29 P. Çarçabal, I. Hünig, D. P. Gamblin, B. Liu, R. A. Jockusch, R. T. Kroemer, L. C. Snoek, A. J. Fairbanks, B. G. Davis and J. P. Simons, *J. Am. Chem. Soc.*, 2006, **128**, 1976–1981.
30 F. Zhu, N. W. Isaacs, L. Hecht, G. E. Tranter and L. D. Barron, *Chirality*, 2006, **18**, 103–115.
31 C. A. Duda and S. S. Stevens, *J. Am. Chem. Soc.*, 1990, **112**, 7406–7407.
32 E. P. Stroyan and S. S. Stevens, *Carbohydr. Res.*, 2000, **327**, 447–453.
33 *NMR spectroscopy of glycoconjugates*, J. Jiménez-Barbero and T. Peters eds., Wiley-VCH, Weinheim, 2003.
34 J. L. Asensio, F. J. Cãnada, X. Cheng, N. Khan, D. R. Mootoo and J. Jiménez-Barbero, *Chem. Eur. J.*, 2000, **6**, 1035–1041.
35 A. M. Dixon, G. Widmalm and T. E. Bull, *J. Magn. Reson.*, 2000, **147**, 266–272.
36 B. Mulloy, T. A. Frenkiel and D. B. Davies, *Carbohydr. Res.*, 1988, **184**, 39–46.
37 M. J. Milton, R. Harris, M. A. Probert, R. A. Field and S. W. Homans, *Glycobiology*, 1998, **8**, 147–153.
38 Q. Xu and C. A. Bush, *Carbohydr. Res.*, 1998, **306**, 335–339.
39 P. J. Bolon and J. H. Prestegard, *J. Am. Chem. Soc.*, 1998, **120**, 9366–9367.
40 M. Staaf, C. Höög, A. Maliniak and G. Widmalm, *Biochemistry*, 2001, **40**, 3623–3628.
41 D. I. Freedberg, *J. Am. Chem. Soc.*, 2002, **124**, 2358–2362.
42 W. R. P. Scott, A. E. Mark and W. F. van Gunsteren, *J. Biomol. NMR*, 1998, **12**, 501–508.
43 Z. Gattin, J. Schwartz, R. I. Mathad, B. Jaun and W. F. van Gunsteren, *Chem. Eur. J.*, 2009, **15**, 6389–6398.
44 G. M. Clore and C. D. Schweiters, *J. Am. Chem. Soc*, 2004, **126**, 2923–2938.
45 A. De Simone, B. Richter, X. Salvatella and M. Vendruscolo, *J. Am. Chem. Soc*, 2009, **131**, 3810–3811.
46 M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870–2871.
47 B. Coxon, *Adv. Carbohydr. Chem. Biochem.*, 2009, **62**, 17–82.
48 F. Cloran, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 1999, **121**, 9843–9851.
49 E. S. Stevens, *Biopolymers*, 1994, **34**, 1403–1407.
50 T. Massad, J. Jarvet, R. Tanner, K. Tomson, J. Smirnova, P. Palumaa, M. Sugai, T. Kohno, K. Vanatalu and P. Damberg, *J. Biomol. NMR*, 2007, **38**, 107–123.
51 P. K. Glasoe and F. A. Long, *J. Phys. Chem.*, 1960, **64**, 188–190.
52 A. Kjellberg and G. Widmalm, *Biopolymers*, 1999, **50**, 391–399.
53 T.-L. Hwang, M. Kadkhodaei, M. Mehran and A. J. Shaka, *Magn. Reson. Chem.*, 1992, **30**, S24–S34.
54 T. Nishida, G. Widmalm and P. Sándor, *Magn. Reson. Chem.*, 1996, **34**, 377–382.
55 T. E. Klepach, C. Thibaudeau, S. Zhao, I. Carmichael and A. S. Serianni, *Biochemistry*, 2002, **41**, 8970.
56 K. Lycknert, A. Helander, S. Oscarson, L. Kenne and G. Widmalm, *Carbohydr. Res.*, 2004, **339**, 1331–1338.
57 B. R. Brooks, C. L. Brooks III, A. D. MacKerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.
58 R. Eklund and G. Widmalm, *Carbohydr. Res.*, 2003, **338**, 393–398.
59 R. Stenutz, I. Carmichael, G. Widmalm and A. S. Serianni, *J. Org. Chem.*, 2002, **67**, 949–958.
60 C. Landersjö, B. Stevensson, R. Eklund, J. Östervall, P. Söderman, G. Widmalm and A. Maliniak, *J. Biomol. NMR*, 2006, **35**, 89–101.
61 G. Widmalm, R. A. Byrd and W. Egan, *Carbohydr. Res.*, 1992, **229**, 195–211.
62 U. Olsson, E. Säwén, R. Stenutz and G. Widmalm, *Chem. Eur. J.*, 2009, **15**, 8886–8894.
63 T. Lütteke, M. Frank and C.-W. von der Lieth, *Carbohydr. Res.*, 2004, **339**, 1015–1020.
64 T. Lütteke, M. Frank and C.-W. von der Lieth, *Nucleic Acids Res.*, 2005, **33**, D242–D246.
65 http://www.glycosciences.de/tools/glytorsion/(accessed September 2008).
66 T. Srikrishnan, M. S. Chowdhary and K. L. G. Matta, *Carbohydr. Res.*, 1989, **186**, 167–175.
67 T. Peters, *Liebigs Ann. Chem.*, 1991, 135–141.
68 G. S. Harbison, *J. Am. Chem. Soc.*, 1993, **115**, 3026–3027.
69 T. Rundlöf and G. Widmalm, *Magn. Reson. Chem.*, 2001, **39**, 381–385.
70 H. Zhao, I. Carmichael and A. S. Serianni, *J. Org. Chem*, 2008, **73**, 3255–3257.
71 I. Tvaroška, M. Hricovíni and E. Petráková, *Carbohydr. Res.*, 1989, **189**, 359–362.
72 B. Bose, S. Zhao, R. Stenutz, F. Cloran, P. B. Bondo, G. Bondo, B. Hertz, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 1998, **120**, 11158–11173.
73 D. A. Case, C. Scheurer and R. Brüschweiler, *J. Am. Chem. Soc.*, 2000, **122**, 10390–10397.
74 K. G. R. Pachler, *Tetrahedron*, 1971, **27**, 187–199.
75 C. A. G. Haasnoot, F. A. A. M. de Leeuw and C. Altona, *Tetraheron*, 1980, **36**, 2783–2792.
76 L. A. Donders, F. A. A. M. de Leeuw and C. Altona, *Magn. Reson. Chem.*, 1989, **27**, 556–563.
77 C. Altona, J. H. Ippel, A. J. A. W. Hoekzema, C. Erkelens, M. Groesbeek, F. A. A. M. de Leeuw and L. A. Donders, *Magn. Reson. Chem.*, 1989, **27**, 564–576.
78 A. A. van Beuzekom,, F. A. A. M. de Leeuw and C. Altona, *Magn. Reson. Chem.*, 1990, **28**, 68–74.
79 C. Thibaudeau, J. Plavec and J. Chattopadhyaya, *J. Org. Chem.*, 1998, **63**, 4967–4984.
80 C. Pérez, F. Löhr, H. Rüterjans and J. M. Schmidt, *J. Am. Chem. Soc.*, 2001, **123**, 7081–7093.
81 B. Bose-Basu, T. E. Klepach, G. Bondo, P. B. Bondo, W. Zhang, I. Carmichael and A. S. Serianni, *J. Org. Chem.*, 2007, **72**, 7511–7522.
82 C. Thibaudeau, R. Stenutz, B. Hertz, T. Klepach, S. Zhao, Q. Wu, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 2004, **126**, 15668–15685.
83 B. Vögeli, J. Ying, A. Grishaev and A. Bax, *J. Am. Chem. Soc.*, 2007, **129**, 9377–9385.
84 K. G. R. Pachler, *Tetrahed. Lett.*, 1970, **22**, 1955–1958.
85 J. Plavec and J. Chattopadhyaya, *Tetrahed. Lett.*, 1995, **36**, 1949–1952.
86 J. J. Chou, D. A. Case and A. Bax, *J. Am. Chem. Soc.*, 2003, **125**, 8959–8966.
87 J. M. Schmidt, *J. Biomol. NMR*, 2007, **37**, 287–301.
88 G. Palermo, R. Riccio and G. Bifulco, *J. Org. Chem.*, 2010, **75**, 1982–1991.
89 P.-E. Jansson and G. Widmalm, *J. Chem. Soc., Perkin Trans. 2*, 1992, 1085–1090.
90 M. K. Dowd, A. D. French and P. J. Reilly, *J. Carbohydr. Chem.*, 1995, **14**, 589–600.
91 B. Hakkarainen, L. Kenne, M. Lahmann, S. Oscarson and C. Sandström, *Magn. Reson. Chem.*, 2009, **45**, 1076–1080.
92 C. J. Margulis, *J. Phys. Chem. B*, 2005, **109**, 3639–3647.
93 B. J. Hardy, S. Bystricky, P. Kovac and G. Widmalm, *Biopolymers*, 1997, **41**, 83–96.