N-Acetylated amino sugars: the dependence of NMR $^3J_{(H^NH^2)}$-couplings on conformation, dynamics and solvent†

Mehdi Mobli and Andrew Almond*

Received 16th April 2007, Accepted 11th May 2007
First published as an Advance Article on the web 31st May 2007
DOI: 10.1039/b705761j

N-Acetylated amino sugars are essential components of living organisms, but their dynamic conformational properties are poorly understood due to a lack of suitable experimental methodologies. Nuclear magnetic resonance (NMR) is ideally suited to these conformational studies, but accurate equations relating the conformation of key substituents (*e.g.*, the acetamido group) to NMR observables are unavailable. To address this, density functional theory (DFT) methods have been used to calculate vicinal coupling constants in *N*-acetylated amino sugars and derive empirical Karplus equations for $^3J_{(H^NH^2)}$ of *N*-acetyl-D-glucosamine (GlcNAc) and *N*-acetyl-D-galactosamine (GalNAc). The fitted Karplus parameters were found to be similar to those previously derived for peptide amide groups, but are consistently larger in magnitude. Local intramolecular interactions had a small effect on the calculated *J*-couplings and comparison with experimental data suggested that DFT slightly overestimated them. An implicit solvation model consistently lowered the magnitude of the calculated values, improving the agreement with the experimental data. However, an explicit solvent model, while having a small effect, worsened the agreement with experimental data. The largest contributor to experimentally-determined $^3J_{(H^NH^2)}$-couplings is proposed to be librations of the amide group, which are well approximated by a Gaussian distribution about a mean dihedral angle. Exemplifying the usefulness of our derived Karplus equations, the libration of the amide group could be estimated in amino sugars from experimental data. The dynamical spread of the acetamido group in free α-GlcNAc, β-GlcNAc and α-GalNAc was estimated to be 32°, 42° and 20°, with corresponding mean dihedral angles of 160°, 180° and 146°, respectively.

1. Introduction

N-Acetylated amino sugars are essential components of living organisms from bacteria to plants and animals.[1–3] For example, chitin, a polymer of *N*-acetyl-D-glucosamine (GlcNAc), is the main constituent of the arthropod exoskeleton.[4] Amino sugars are also part of the oligosaccharide decorations to glycoproteins in humans, which allow the body to recognize itself (GlcNAc is characteristic of the A blood-group antigen), among other functions.[1,5] Furthermore, amino sugar derivatives comprise a large mass-fraction of the vertebrate extracellular matrix, being the major constituent of glycosaminoglycans (GAGs).[6,7]

The GAGs are a family of carbohydrate polymers, named according to their basic repeating disaccharide, which comprises an amino sugar and a single uronic acid sugar. For instance, the GAGs hyaluronan and heparan sulfate contain GlcNAc, while chondroitin sulfate contains *N*-acetyl-D-galactosamine (GalNAc), see Fig. 1. They are found in all vertebrate tissue and although it is known that GAGs partake in cell signaling, development and growth,[2,6,8] many of their functions are not presently understood. One of the major reasons for this poor appreciation of GAG function is the current lack of detailed information about their dynamic 3D-organization in solution.[7,9] Among experimental techniques, vibrational and nuclear magnetic resonance (NMR) spectroscopy are applicable to these investigations in aqueous solution. While vibrational spectroscopy can yield some of this information, extensive theoretical analysis is currently required to deconvolute the resultant spectra.[10] A more suitable technique is NMR, which can determine 3D-structural information at atomic resolution in relatively complex molecules. Unfortunately, the repeating nature of GAGs results in little chemical shift dispersion, which has reduced the effectiveness of NMR in this area.[11,12] However, recent advances in NMR instrumentation and experiments, coupled with chemical and biological methods for derivation, purification and labeling, are finally allowing the

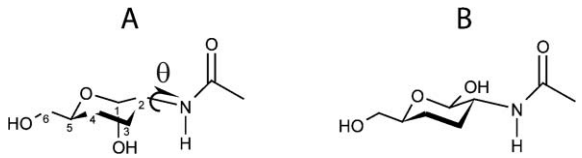

Fig. 1 Structure and ring numbering of the α- (A) and β- (B) anomers of *N*-acetylated amino sugars; only the anomeric-ring hydroxyl group is shown. The two hydrogen atoms involved in the $^3J_{(H^NH^2)}$-coupling are shown together with the corresponding angle θ. In GlcNAc (**1**) all ring hydroxyl groups are equatorial; GalNAc (**2**) differs in that the hydroxyl group at the C^4-position is axial.

Manchester Interdisciplinary Biocentre, 131 Princess Street, University of Manchester, Manchester, UK M1 7DN. E-mail: Andrew.Almond@manchester.ac.uk; Fax: +44 161 30 68918; Tel: +44 161 30 64199

† Electronic supplementary information (ESI) available: Calculated energies, relative populations and calculated coupling constants of GlcNAc. See DOI: 10.1039/b705761j

3D solution structures of these complex carbohydrates to be investigated with unprecedented atomic accuracy.[11,13–15]

Structural information is usually obtained from NMR experiments using the nuclear Overhauser effect (NOE) or residual dipolar couplings (RDC) induced by weak alignment. Furthermore, the scalar coupling (also referred to as spin–spin or *J*-coupling) between nuclei can be used to provide structural information if the relationship between bond geometry and coupling is known. This approach has been used to determine the conformational dependence of several saccharides,[16–19] but in many cases, such as around the acetamido moiety of the important amino sugars, the geometry–coupling relationship has not been investigated. In general, this relationship can be described by the Karplus equation, an empirical relationship relating coupling constants to bond dihedral angle and a second order expansion in the cosine of this angle, which is regularly used in chemistry to obtain geometrical information. The parameters needed to define a Karplus relationship are traditionally determined using coupling constants of molecules with known geometries and are then used to analyze systems with unknown geometries.[20] However, one may calculate coupling constants using theoretical quantum-chemistry methods for any arbitrary geometry. This approach has been unpopular because significant computational effort is required to produce accurate results.

With the advance of computational methods and power it is now feasible to pursue the theoretical approach more routinely. Recently, it has been shown that density function theory (DFT) can calculate scalar couplings very accurately by employing basis-sets that describe the electronic environment near the nucleus well.[21,22] It has also been shown that the computational cost can be greatly reduced by calculating the dominating Fermi contact (FC) term of the scalar coupling using a larger basis-set and the computationally more demanding but smaller remaining contributions from the paramagnetic spin–orbit (PSO), diamagnetic spin–orbit (DSO) and spin dipolar (SD) terms using a smaller basis-set.[21–23] Modern DFT calculations hold the promise of unprecedented accuracy in geometry–coupling estimation because all dynamics are frozen-out in calculations. This is not possible directly from experimental measurements of even the most rigid molecules, which generally report lower values than theoretical methods due to local librations.[24] Furthermore, DFT calculations that include solvent effects show improved accuracy, giving results in close agreement with the experimental NMR measurements.[25,26]

Hydrogen-bonding interactions (intramolecular and intermolecular) are essential for defining carbohydrate 3D-structure in aqueous solution.[27] The theorized role of the amide group as both a hydrogen-bond donor and acceptor in solvated GAGs makes detailed analysis of this group particularly important.[28] In *N*-acetylated 2-deoxyamino sugars, the H^2 proton is fixed within a pyranose ring and its coupling to the H^N proton provides detailed information of the rotation of the entire acetamido group. Thus, we have employed the recent developments in DFT calculations to derive Karplus equations suitable for use in the amino sugars found in GAGs (GlcNAc and GalNAc). The change in the derived Karplus parameters with a change in the anomeric form of the amino sugar was investigated in GlcNAc, where both the α- and β-anomers were considered. Also, the effects of including an implicit or an explicit solvent model was explored, the latter with the aid of data from molecular dynamics (MD) simulations. Since the interaction with water has been found to be important for determining carbohydrate conformation,[29] we employed classical MD simulations with explicit water molecules, rather than more complex *ab initio* MD methods which are generally used *in vacuo*.[30] Information from the MD simulations was also used to model dynamical effects and an equation was derived to estimate the dynamic conformational spread of the acetamido moiety directly from a derived Karplus equation and NMR measurements.

2. Methods

2.1 Density functional theory (DFT) quantum methods

In all cases, the D-pyranose rings of the *N*-acetylated amino sugars were fixed in the $^4C_1$-chair conformation, in which the O^1-hydroxyl is axial and equatorial in the α- and β-anomers, respectively. For the α-anomer of GlcNAc, SD, PSO and DSO spin–spin coupling terms were included using DFT (B3LYP[23]) with the IGLOO-III basis-set (11s,7p,2d/6s,2p) [7s,6p,2d/4s,2p].[31] The FC term (used for all coupling calculations) was calculated using the HIIIsu3 basis-set,[22] where for first and second row atoms the s-orbital functions of the IGLOO-III basis-set are decontracted and the resulting basis is augmented by three tight s-orbital functions in an even-tempered manner (14s,7p,2d/9s,2p) [14s,6p,2d/9s,2p]. Using this approach, calculations could be performed with a reasonable amount of computer time for these purposes and gave results in good agreement with the basis-set limit results published previously.[21]

All coupling constants were calculated using structures optimized at the B3LYP/6-31G(d,p) level of theory. The calculations were performed using the Gaussian 03 software[23] rev. D (where the FC calculation can easily be calculated separately from the spin–spin coupling terms). The implicit PCM method implemented in the G03 software was used where specified.

2.2 Derivation of the Karplus and related equations

The Karplus equation is an empirical relationship that relates a three-bond dihedral angle (θ) to the corresponding scalar coupling. It is a truncated cosine expansion, shown in eqn (1), with a parameter (φ) that reflects the phase shift between the cosine modulation and the measured dihedral angle. In eqn (1), the predicted scalar coupling is *J* and *A*, *B*, and *C* are empirical constants fitted to either calculated or experimental data.

$$J = A\cos^2(\theta + \varphi) + B\cos(\theta + \varphi) + C \quad (1)$$

In this case, the angle θ was defined to be the H^2–C^2–N^H–H^N torsion (Fig. 1), which by symmetry of the electronic orbitals allows φ to be zero for the three-bond coupling. This torsion was rotated through 360° in 30° degree increments, and at each point minimization was performed, followed by spin–spin coupling constant calculation. A Karplus-type equation was derived by non-linear least-squares fitting the resultant spin–spin couplings to the general eqn (1). In this procedure, implicit solvation (when used) was applied during both minimization and spin–spin coupling calculation.

Although eqn (1) only allows calculation of spin–spin coupling for a static structure, it can be generalized to the case of harmonic motion about a mean angle θ by integration. If the motion is

assumed to be Gaussian with a standard deviation of σ, then the A, B and C parameters in eqn (1) can be replaced by the modifications shown by eqn (2), as described previously.[32]

$$A' = Ae^{(-2\sigma^2)};\ B' = Be^{\left(-\frac{\sigma^2}{2}\right)};\ C' = C + \frac{A}{2}\left\{1 - e^{(-2\sigma^2)}\right\} \qquad (2)$$

Assuming that the libration (σ) is small, it is possible to simplify eqn (2) considerably assuming $B' \cong B$, since the exponential term in B' is very close to 1 for small angles. Furthermore, by substitution of these approximated equations into eqn (1), it is possible to derive a relationship between the dynamic spread, σ, the measured coupling, J, and the original Karplus coefficients from eqn (1), which is shown in eqn (3). Such an equation can be used to calculate the angular spread for librations around the *cis* or *trans* conformation (by suitable substitution of $\bar{\theta}$), from an experimentally measured coupling and derived Karplus parameters. Due to the assumption made for B', the result of eqn (3) will always be slightly larger than the correct value, but will be negligible for small angles.

$$\sigma^2 \approx \frac{1}{2}\log_e \frac{A(2\cos^2\bar{\theta} - 1)}{2J - 2B\cos\bar{\theta} - 2C - A} \qquad (3)$$

2.3 Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed using the CHARMm[33] package with explicit solvent and using a force-field suitable for carbohydrates. Temperature was kept constant at 298 K using weak coupling to a heat bath and long-range electrostatics were treated using the particle-mesh Ewald summation method, as described previously.[13] Following equilibration, each simulation was performed for a total time of 5 ns and coordinate snapshots were saved at 0.05 ps intervals for subsequent analysis.

Models of the sugars that included several important explicit water molecules were constructed from a representative snapshot of the MD simulation (selected using statistical analysis of water interactions during the complete MD simulation) by deletion of non-essential water molecules and minimization using a molecular mechanics force field (MMX[34] in PCModel[35]), which includes a hydrogen-bonding potential. The resulting structures were then minimized without any restraints using the aforementioned DFT method. This procedure was iterated until a structure was found that both satisfied the MD data and converged to a low-energy minimum according to DFT.

2.4 NMR experiments

N-Acetyl-D-galactosamine was obtained from a commercial source (SigmaAldrich) and used without further purification. The NMR sample was made up in H_2O with 10% D_2O for deuterium locking and the pH was adjusted to 6.0 using NaOH and HCl. The NMR spectra were recorded at 600 MHz and the ^3J-coupling constant was determined using a simple 1D ^{1}H spectrum with WATERGATE solvent suppression; 128 scans of 32k complex datapoints were collected (interscan delay of 2s) and the resulting free-induction decay was processed using NMRPipe.[36]

3. Results

Theoretical density functional theory (DFT) quantum calculations of spin–spin couplings were performed on isolated molecules and in the presence of implicit and explicit solvent to determine the relative importance of intramolecular and intermolecular effects. Results obtained by considering each of the two effects separately are reported in the following sections, followed by an investigation of the importance of internal dynamics.

3.1 Intramolecular effects

The DFT calculation of the scalar coupling constants of α-GlcNAc was carried out by including the Fermi contact (FC) term using the larger HIIISu3 basis-set and the remaining terms (PSO, DSO and SD) using the smaller IGLOO-III basis-set. This calculation was performed on 12 static (DFT optimized) structures that differed by successive 30° rotations of the amide group and the results are shown in Fig. 2 and 3. Eqn (1) was fitted to these calculated coupling constants *via* the parameters A, B and C. At the maximum of the Karplus curve (*i.e.*, when the H^2–C^2–N^H–H^N angle, θ, is 180°) the calculated non-FC terms are: SD: −0.03 Hz, PSO: 3.04 Hz and DSO: −2.95 Hz giving a net total of 0.06 Hz contribution, compared with a FC contribution of 11.90 Hz. This result suggests that the FC term is by far the dominant contribution to this three-bond coupling. Indeed, when the fitting procedure was repeated on calculations performed using only the Fermi contact term (FC-only), identical values for the A, B and C parameters were obtained within fitting error. These results are illustrated in Fig. 2, which shows that the PSO and DSO terms nearly cancel and that the SD term is sine-modulated, since it exhibits a correlation to the inter-nuclear distance, rather than orbital overlap. Based on these findings, the calculations on β-GlcNAc and α-GalNAc (β-GalNAc was not considered) were performed using the same method, but included only the FC term. The results are presented in Fig. 3 and the fitted parameters are given in Table 1.

Each geometry minimization was first performed using an empirical force-field followed by DFT at fixed angle θ. In the latter high-level minimizations, only local minima could be found and hence the rotamer states of hydroxyl moieties did not change. For some values of θ, the demand that this angle should remain fixed led to deformation of the amide plane during DFT minimization (see Fig. 4), which resulted from a steric clash between the carbonyl oxygen and the O^3 hydroxyl oxygen. Consequently, in

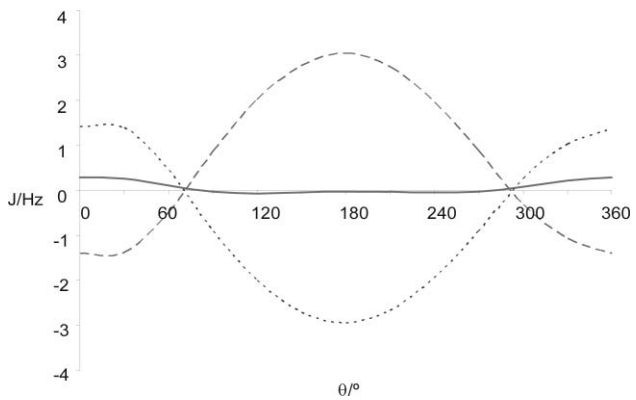

Fig. 2 Calculated non-FC (SD = '—', PSO = '---', DSO = '···') contributions to the density functional theory (DFT) calculated coupling constants in α-GlcNAc as a function of the intervening dihedral angle θ.

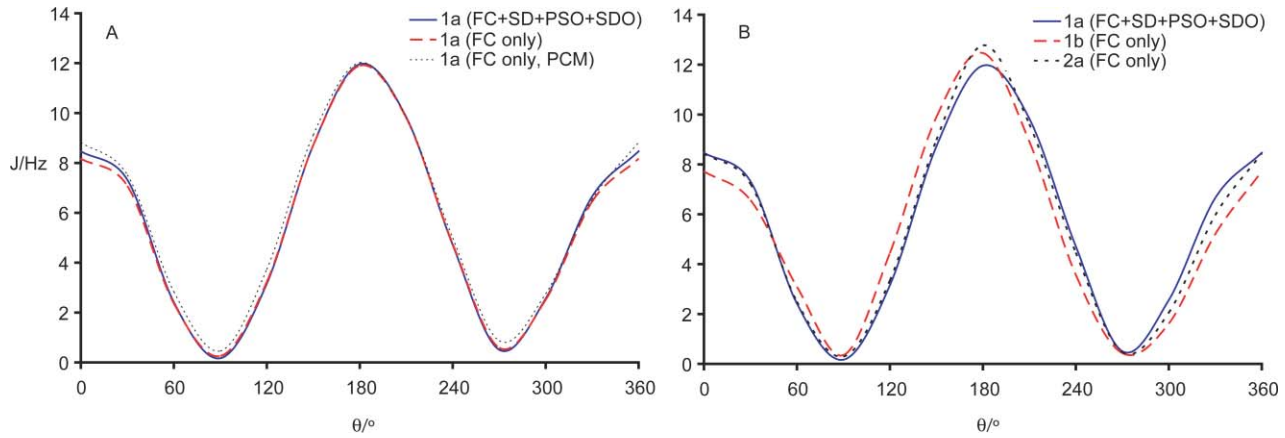

Fig. 3 Calculated $^3J_{(H^NH^2)}$ plotted as a function of the dihedral angle θ (note arbitrary line-fit). A) DFT calculated scalar coupling constants for α-GlcNAc. Full calculation (SD + PSO + DSO + FC) = '—', FC-only = '---', FC-only including implicit solvent effects (PCM) = '···'. B) α-GlcNAc (full) = '—', β-GlcNAc (FC-only) = '---', α-GalNAc (FC-only) = '···'.

Table 1 Karplus coefficients fitted to DFT-calculated scalar coupling constants. Full refers to DFT calculation, including all terms of the coupling constant (FC, DSO, PSO, SD). FC refers to DFT calculations including only the FC-term of the coupling constant. FC [PCM] refers to DFT calculations of the coupling constant including only the FC-term and including the implicit PCM for solvation. Error refers to the fitting error to the Karplus equation (1)

	A	Error	B	Error	C	Error
[1] 1α-Full	9.81	±0.23	−1.51	±0.12	0.62	±0.14
[2] 1α-FC	9.56	±0.23	−1.62	±0.11	0.69	±0.14
[3] 1α-FC[PCM]	9.60	±0.20	−1.51	±0.10	0.99	±0.12
[4] 1β-FC	9.45	±0.26	−2.08	±0.13	0.63	±0.16
[5] 2α-FC	10.02	±0.21	−1.79	±0.10	0.49	±0.13
Ref. 24[a]	9.44	—	−1.53	—	0.07	—
Ref. 24[b]	9.14	—	−2.28	—	−0.29	—

[a] Values derived using Ace–Ala–NMe. [b] Values derived using Ala–Ala–NH_2.

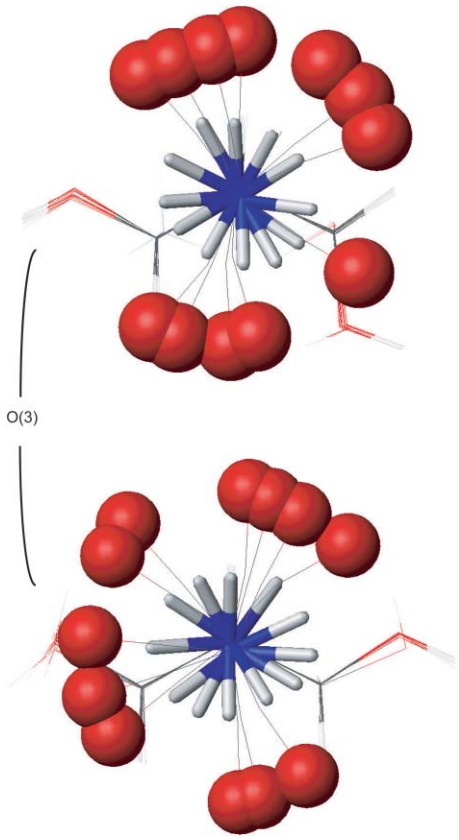

Fig. 4 Overlay of 12 DFT minimized structures of GlcNAc (top α, bottom β), with 30° incremental rotation of the H^N–H^2 dihedral angle θ. The carboxyl oxygen is represented by a sphere and the N^H–H^2 bond by a cylinder.

successive minimizations the O^3 hydroxyl group could be found either pointing towards or away from the carbonyl group. These two orientations for the O^3 hydroxyl group (referred to as T and A, respectively) resulted in differing levels of amide deformation (Fig. 4). However, when the coupling constant was calculated for both the A and T orientations at $\theta = 180°$, the calculated difference in coupling was less than 0.2 Hz (Table 2). Furthermore, although the extent of deformation was greater when θ was fixed at around 90° or 270°, the calculated vicinal coupling constants, and subsequently any errors due to deformation, were found to be very small. To investigate this further, the FC term was calculated for α-GlcNAc with θ fixed at 180°, and the out-of-plane bending of the amide nitrogen was varied between 0° and 30° in 10° increments. The result was a change of the vicinal coupling of 0.28 Hz (*i.e.*, less than 3% of the coupling constant), further suggesting that these amide deformations have little effect on the accuracy of the scalar coupling constants.

One may use the potential energy surface calculated during the geometry optimization to estimate an average *J*-coupling. The 12 static conformers generated were Boltzmann weighted according to eqn (4), where conformer a has internal energy E_a and a probability of observation of p_a (k_B is the Boltzmann constant and T is the temperature).

$$p_a = \frac{e^{-E_a/k_B T}}{\sum_a e^{-E_a/k_B T}} \quad (4)$$

The average *J*-coupling is then estimated from the relative populations of the 12 conformers and their calculated couplings. This was performed on GlcNAc and the resulting average values

Table 2 DFT-calculated scalar coupling constants (using only the FC contribution) for possible structures and solvent models of GlcNAc

	Solvent	θ/°	OH config.	Calc. $^3J_{(H^NH^2)}$/Hz	Pred. $^3J_{(H^NH^2)}$/Hz
α-GlcNAc					
	Vacuo	180	A	11.9	11.9
	Vacuo	180	T	12.1	11.9
	PCM	180	A	11.7	11.9
	PCM	180	T	12.0	11.9
	Explicit	168	T	11.2	11.4
β-GlcNAc					
	Vacuo	180	T	12.5	12.2
	Vacuo	180	A	12.5	12.2
	PCM	180	T	11.9	12.2
	Explicit	163	A	11.8	11.3

were 5.09 Hz and 1.04 Hz for the α- and β-anomers respectively (*cf.* experimental values of 8.9 Hz and 9.1 Hz, respectively). The very poor agreement of the calculated average couplings of the β-anomer with the experimental data is due to the potential energy surface of this anomer showing a minimum energy value around 270°, where a stabilizing intramolecular hydrogen bond is formed. Similarly for the α-anomer, the minimum energy is found in the *cis* orientation (near 60°), where a hydrogen bond is formed between the carbonyl group of the amide and the anomeric hydroxyl group. Similar results were found when including an implicit solvent model (data only calculated for the α-anomer). Again the stabilizing intramolecular hydrogen bond in the *cis* orientation (near 30°) yielded the most stable conformation giving an average J of 7.29 Hz. These results further support the need for explicit water molecules to accurately predict the populations of individual conformers.

3.2 Intermolecular effects and internal dynamics

In aqueous solution, complex carbohydrates form numerous hydrogen bonds to surrounding water molecules, which may be regarded as an integral part of the carbohydrate 3D-solution structure. The simplest method for including solvent effects in DFT calculations is by the use of an implicit solvent model (such as the PCM used here), where the solute molecule is placed in a cavity and polarized according to the properties of the solvent. Using this model the FC contribution of the scalar coupling was calculated (for GlcNAc), while forcing the dihedral angle θ to 180° during *in vacuo* DFT optimization. This was done to facilitate direct comparison of the different methods, and since at this angle the highest value for the scalar coupling is obtained (Table 2). Including implicit solvent has a small effect and appears to decrease the magnitude of the scalar coupling. The effect is larger in the β-anomer (0.6 Hz) of GlcNAc compared to that in the α-anomer (0.2 Hz).

While the implicit solvation model is computationally efficient, it ignores important electronic effects that may result from local hydrogen-bonding interactions to solvent water molecules. Therefore, in order to investigate the effect of these specific interactions on coupling constants, explicit models of water interaction were produced from 5 ns molecular dynamics (MD) simulations of GlcNAc using a molecular mechanics force-field. The mean orientation of various rotatable bonds and their standard deviations were calculated from the simulations (Table 3). From the table, it can be seen that the H^2–C^2–N^H–H^N torsion is predominantly in the *trans* orientation in both cases, with mean torsion angles of 161° and 180° for the α- and β-anomers, respectively. Furthermore, the MD simulations were used to identify the positions of key water molecules in GlcNAc (Table 4). Those water molecules found to be involved in hydrogen bonds and water bridges to the amide were suitably positioned in a static model, which was subsequently optimized using DFT. The two optimized structures for α- and β-GlcNAc, with attendant explicit water molecules, are shown in Fig. 5. These geometries were used as a basis for calculating the FC contribution to the three-bond scalar coupling. These results are also given in Table 2, which includes calculations of the two anomers of GlcNAc (for both the A and T orientations) excluding any solvent effects. Furthermore, the scalar couplings have also been calculated by using the Karplus relationship derived *in vacuo* (see above).

Table 3 Conformer population of rotatable bonds in GlcNAc from MD simulations, together with average torsion angle ($\bar{\theta}$) and standard deviation (σ). The rotamer states $g+$, $g-$ and t represent torsion angles of 0° to 120°, −120° to 0° and 120° to −120°, respectively

	α-GlcNAc			β-GlcNAc		
Rotamer	$g+$	$g-$	t	$g+$	$g-$	t
O^5–C^1–O^1–H						
Population (%)	94	2	4	29	64	7
$\bar{\theta}$	45	341	157	36	315	201
σ	20	22	20	21	23	20
C^2–C^3–O^3–H						
Population (%)	13	57	29	9	52	39
$\bar{\theta}$	56	287	196	63	283	195
σ	21	22	23	20	21	23
C^3–C^4–O^4–H						
Population (%)	78	13	9	76	12	12
$\bar{\theta}$	65	310	144	66	307	146
σ	23	21	17	23	21	18
C^4–C^5–C^6–O^6						
Population (%)	57	2	41	53	2	45
$\bar{\theta}$	61	282	187	60	284	186
σ	10	17	13	11	17	13
C^5–C^6–O^6–H						
Population (%)	32	21	47	32	23	45
$\bar{\theta}$	66	292	180	67	288	180
σ	23	25	25	22	24	25
H–N–C^2–H						
Population (%)	—	0	100	—	13	87
$\bar{\theta}$	—	—	161	—	3	180
σ	—	—	29	—	57	21

[a] For H^N–N^H–C^2–H^2 the *cis* orientation is given under '$g-$'.

Table 4 Important water bridges and hydrogen-bonding interactions identified from MD simulations of GlcNAc; those shown in bold were found to be important for the conformation of the amide moiety

	α-GlcNAc			β-GlcNAc	
	H_2O (%)	$(H_2O)_2$ (%)		H_2O (%)	$(H_2O)_2$ (%)
N^2–O^1	13	**15**	$1O^5$–$1O^1$	18	**14**
O^3–N^2	12	**15**	$1O^4$–$1O^3$	**22**	26
O^4–O^3	19	**24**	$1O^6$–$1O^5$	**15**	11
O^6–O^5	19	16	$1O^3$–$1N^2$	2	12
O^6–O^1	6	17	$1O^6$–$1O^1$	3	**40**
O^6–O^5	19	16	$1O^6$–$1O^5$	**15**	11
O^6–O^4	2	11	$1O^6$–$1O^4$	3	12

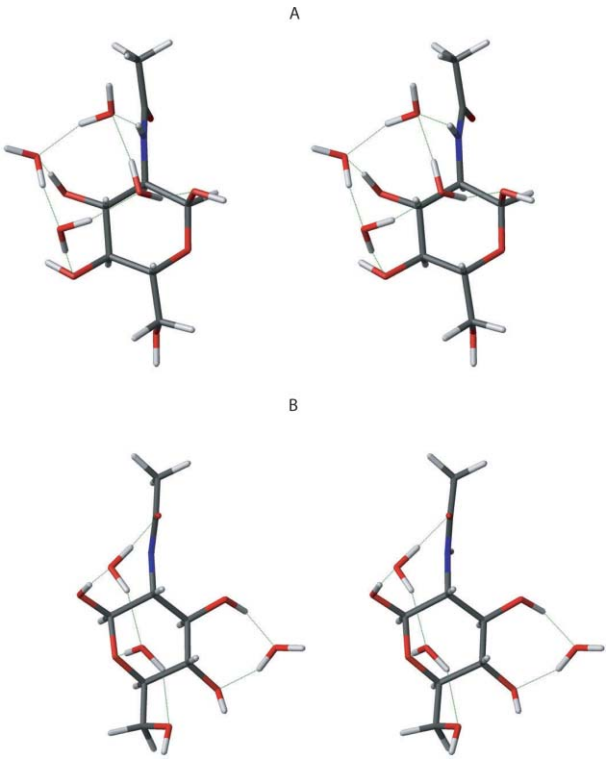

Fig. 5 Stereo view of the *ab initio* optimized structure of GlcNAc, including water bridges identified from MD simulations (A = α, B = β). The dashed lines show intermolecular hydrogen bonds.

The change in the scalar coupling on inclusion of explicit water molecules does not follow a definite trend, which is expected since the solvent molecules are directly interacting with the solute and thus the change in electronic configuration is complex. It is noted that although only the α-anomer model has direct interactions between the amide proton and water molecules (Fig. 5), it is the β-anomer that shows the greatest deviation from the derived Karplus relationship defined above. Therefore, from the data in Table 2 it is concluded that the deviation from the $^3J_{(H^NH^2)}$ due to solvent effects is less than 0.6 Hz (*i.e.*, less than 5%).

Using the H^2–C^2–N^H–H^N dihedral angle data from the MD simulations and the Karplus parameters (defined earlier), an average value for the scalar coupling constant could be determined for α/β-GlcNAc. The coupling constants were calculated by integrating the Karplus equation, as described in eqns (1) and (2), about an angle $\bar{\theta}$ with a Gaussian distribution (standard deviation σ) The average angles and standard deviations extracted from the MD simulations were used (Table 2). This latter method was repeated for α-GalNAc, using the same $\bar{\theta}$ and σ parameters as for α-GlcNAc. The results are presented in Table 5 together with the experimentally-measured coupling constants.

4. Discussion

The use of isotopic ^{15}N-labeling has greatly benefited NMR studies of proteins, and brought about a cascade of experimental and theoretical developments. In particular, the $^3J_{(H^NH^\alpha)}$-coupling can be measured residue-specifically by multidimensional NMR techniques, which is a useful measure of protein and peptide conformation. Based on these methodological successes, it is natural to apply them to other, similar, molecules. *N*-Acetylated amino sugars are ideal candidates because although they are carbohydrates, they also have amide-containing side-chains. Furthermore, the acetamido side-chains play an important role in determining the conformation of vertebrate and bacterial polysaccharides and N- and O-linked glycans. However, there are both experimental and theoretical limitations to using ^3J-couplings in the acetamido side-chain for conformational analysis. First, a method is required to measure these couplings in oligosaccharides. Recent work has shown how ^{15}N-labeling can be used to resolve individual amide protons in defined-length hyaluronan oligosaccharides, overcoming the first limitation.[13] Secondly, while there are equations

Table 5 Experimental $^3J_{(H^NH^2)}$ scalar coupling and corresponding calculated coupling using appropriate Karplus equations derived here. Calculations include averaging over MD populations and by assuming a Gaussian distribution of conformers (σ) about a *trans* mean dihedral angle ($\bar{\theta}$)

	α-GlcNAc/Hz	β-GlcNAc/Hz	α-GalNAc/Hz
Experimental	8.88	9.07	8.42
MD average[a] (θ both *cis* and *trans*)	8.93[b]	10.39[e]	—
	8.91[c]	—	—
	9.13[d]	—	—
Gaussian average (θ exclusively *trans*)	9.10[f]	10.91[i]	9.35[j]
	9.07[g]	—	—
	9.29[h]	—	—

[a] Using corresponding parameters from Table 1. [b] Parameters from [1] in Table 1. [c] Parameters from [2] in Table 1. [d] Parameters from [3] in Table 1. [e] Parameters from [4] in Table 1. [f] Modified parameters from [1] in Table 1; $A' = 5.9$, $B' = -1.3$, $C' = 0.6$; $\bar{\theta} = 159$, $\sigma = 29$. [g] Modified parameters from [2] in Table 1; $A' = 5.7$, $B' = -1.4$, $C' = 0.7$; $\bar{\theta} = 159$, $\sigma = 29$. [h] Modified parameters from [3] in Table 1; $A' = 5.7$, $B' = -1.3$, $C' = 2.9$; $\bar{\theta} = 159$, $\sigma = 29$. [i] Modified parameters from [4] in Table 1; $A' = 7.2$, $B' = -1.4$, $C' = 1.7$; $\bar{\theta} = 180$, $\sigma = 21$. [j] Modified parameters from [5] in Table 1; $A' = 6.0$, $B' = -1.6$, $C' = 2.5$; $\bar{\theta} = 159$, $\sigma = 29$.

that relate measured $^3J_{(H^NH^\alpha)}$-couplings to conformation in proteins,[20,37–39] no corresponding equations exist for these important amino sugars, restricting the usefulness of accurate experimental measurements. The investigation undertaken here was designed to determine the sensitivity of the $^3J_{(H^NH^2)}$ vicinal coupling in *N*-acetylated amino sugars to conformation, within their unique chemical environment. This involved the analysis of electronic structure, solvent and dynamics.

4.1 Scalar couplings in the pyranose ring

A few observations can be made from the ^3J-data involving the H^2 atom, shown in Table 6. Firstly, in all cases the calculated ^3J-coupling constants are larger in magnitude than the corresponding experimental values. This is corrected to some extent by inclusion of the implicit solvent model, but the values remain consistently larger than the experimental values. Secondly, the discrepancies are not fixed amounts and vary depending on the chemical environment of the coupling. It is therefore suggested that these observations can be explained by internal molecular dynamics. Furthermore, it would suggest that the pyranose ring itself undergoes slight changes in puckering, which should be quantifiable by detailed analysis of accurate coupling-constant measurements. From the data in Table 6 it is observed that the larger *trans* H^2–H^3 coupling constant differs more from the experimental data than the smaller *trans* H^1–H^2 coupling constant in the β-anomer. This observation can be explained by the fact that the larger coupling constant will decrease by a larger amount than a smaller one, given they have the same ranges of motion and average angles. Similarly for the α-anomer it is observed that the larger H^2–H^3 coupling differs from the experimental data more than the smaller H^1–H^2 coupling constant. Although one may intuitively also expect this discrepancy to be due to dynamical effects, the libration is not about a local maximum on a Karplus-type curve, in fact here the dihedral angle is *gauche* (51°) and the librations would lead to a very small change (increase or decrease depending on the Karplus parameters) in the coupling constant. The size of this overestimation may be approximated from a set of Karplus parameters derived for cyclic H–C–C–H vicinal couplings ($A = 11.16$, $B = -1.28$, $C = 0.77$).[40] Using this equation at the *gauche* angle it was determined that a standard deviation spread of angle of 10° (the MD simulation predicted that the spread would be 8° in the pyranose ring) leads to a maximum change in coupling of 0.1 Hz. This suggests that the overestimation will be in the range of 0.4–0.6 Hz.

Table 6 Vicinal coupling constants involving H^2 of GlcNAc, together with the corresponding DFT-calculated dihedral angles

		$^3J_{H^1H^2}$/Hz	$^3J_{H^2H^3}$/Hz	$^3J_{H^2H^N}$/Hz
Experimental	α	3.60	10.74	8.88
	β	8.46	10.40	9.07
In vacuo	α	4.12	11.70	11.90
	β	8.63	10.98	12.46
Implicit solvent	α	4.08	11.56	11.69
	β	8.79	10.98	11.94
Explicit solvent	α	4.62	12.25	11.21
	β	9.27	11.02	11.82

Based on these results it may be concluded that the calculated values for the H^1–H^2 scalar coupling in α-GlcNAc are slightly overestimated in the DFT calculations. One may speculate that it is the solvent effects that are poorly modeled. However, experimental results[41] on a form of GlcNAc where all the hydroxyl groups are acetylated shows that there is no change in this coupling constant (the data are given to 0.1 Hz) when going from H_2O to $CDCl_3$, suggesting that solvent effects are likely to be small. The remaining sources for this overestimation are either due to the basis-set or the DFT method itself. A larger basis-set is not expected to improve the current calculations significantly, based on a previous study.[22] Furthermore, the same study found that overestimation of coupling constants beyond motional dynamics is an intrinsic property of the DFT methodology.

It is also noted that there are no simple factors that can be used to correct for errors resulting from DFT calculations, suggesting that the extent of this overestimation is coupling-specific. For the *trans* three-bond H^2–H^3 coupling the error is likely to be less than 10%, as any greater decrease would lead to calculated values lower than those measured experimentally. It is therefore concluded that although this inherent systematic error is small, one must be vigilant when using Karplus parameters fitted to such calculations as they are likely to result in either overestimation of the bond librations or inaccurate mean dihedral angles.

4.2 Scalar couplings in the acetamido group, dynamics and solvent effects

The experimental coupling constants for α- and β-GlcNAc are 8.88 Hz and 9.07 Hz for the two anomers respectively. If dynamical effects were frozen out completely these experimental values would correspond to a rotation of 30° from *trans* in both cases (using our Karplus equation). This result is neither consistent with X-ray diffraction data (which suggest a dihedral angle of 180° for α-GlcNAc)[42] nor MD simulations (where the average angles are 161° and 180° for α- and β-GlcNAc, respectively). Therefore, dynamical effects must explain the large deviation between the experimental and calculated coupling constants for the flexible $^3J_{(H^NH^2)}$ scalar coupling. If it is assumed that the θ dihedral is exclusively *trans* in β-GlcNAc, eqn (3) predicts that the standard deviation spread (σ) on the dihedral angle is 42°. Similar calculations for α-GlcNAc result in a value for σ of 32°. This indicates that the β-anomer of GlcNAc is more dynamic than the α-anomer, which is supported by ^{15}N-NMR relaxation studies of hyaluronan oligosaccharides, where the β-anomer is found to have a lower order parameter than the α-anomer.[13]

It should be noted that the MD simulation of β-GlcNAc shows a bimodal distribution of acetamido angles (while the *trans* orientation of the θ dihedral is dominant, *cis* is also populated), which cannot be taken into account by eqn (3). Averaging only the *trans* conformations from the MD simulation, *via* fitting to eqn (3), results in a larger predicted coupling and a worse agreement with experimental data than averaging the whole simulation (Table 5). Therefore, if this dihedral angle populates both conformers in reality, the assumption that θ is exclusively *trans* for free GlcNAc will result in an overestimation of the spread (σ) of dihedral angles at the acetamido group. Alternatively, the MD simulation both overestimates the relative population of the

cis conformer and underestimates the dynamic spread around the *trans* conformation.

In Table 5 the experimental and calculated coupling constants (using the various approaches) are shown. The calculated coupling constant using various Karplus equations derived for α-GlcNAc differ little. The coupling constant calculated including an implicit solvent model (PCM) is surprisingly higher than when a solvent model is excluded. This difference appears as the geometry optimization including the implicit solvent model favors the A orientation of the O^3 atom, and in this orientation the coupling constant is slightly higher than for the T orientation favored in the *in vacuo* optimizations. It is interesting to note that local geometrical changes have a similar (but slightly larger) effect on the coupling constants than the calculated solvent effect. Similar results are obtained when using the modified Karplus parameters compared to averaging over the MD population, suggesting that the distribution of conformers is Gaussian in the MD simulation.

In α-GalNAc the experimental coupling constant is 8.42 Hz, which is in disagreement with the calculated value reported in Table 5 (9.35 Hz). This calculation is made using the Karplus parameters derived specifically for α-GalNAc, but the θ and σ values of eqn (1) and (2)are those determined for α-GlcNAc. Thus it would appear that the mean dihedral angle or the extent of libration is different in these two molecules. The crystal structure of α-GalNAc is known and the dihedral angle of interest is reported to be 146°.[43] Inserting this value into eqn (3), a dynamical spread of 20° is found for this molecule. This suggests that both the dynamics and the structure of α-GalNAc are different from those of α-GlcNAc. This procedure can easily be repeated for more complex systems such as in hyaluronan where the $^3J_{(H^NH^2)}$ coupling constants for each amino sugar have been measured in the hexasaccharide.[13] We find by the methods presented here that the amide group libration of the central ring (γ) in this molecule is 34°, suggesting that this group is less dynamical than the corresponding one in the free amino sugar.

In general, inclusion of the explicit solvent molecules worsens the agreement between the experimental and calculated coupling constants (see Tables 4 and 5). This may be due to the fact that the intermolecular interactions with water are very dynamic and are thus difficult to simulate using static DFT calculations. However, the models that included explicit solvent molecules are much more conformationally realistic than those without the solvent model. This is evident when the α-GlcNAc structure is extracted and superimposed with the available crystal structure (Fig. 6). The positioning of the ring hydroxyl groups that form hydrogen bonds with water molecules is nearly identical to those determined by X-ray crystallography. The only differences between the two structures are a deviation of the amide plane from planarity by 20° in the crystal structure (*cf.* planar in the *ab initio* calculation) and a deviation of the H^N–H^2 dihedral angle by a compensating amount in the *ab initio* calculated structure (*cf. trans* in the crystal structure). These discrepancies aside, this method does appear to be able to generate realistic 3D-structures for these highly-solvated carbohydrates *in silico*. Furthermore, one may argue from these data that the α-anomer is involved in more hydrogen bonds, see Fig. 6, which may explain the higher stability of this anomer in aqueous solution than predicted by *in vacuo* calculations (where the more sterically-relaxed β-anomer has the lower energy).

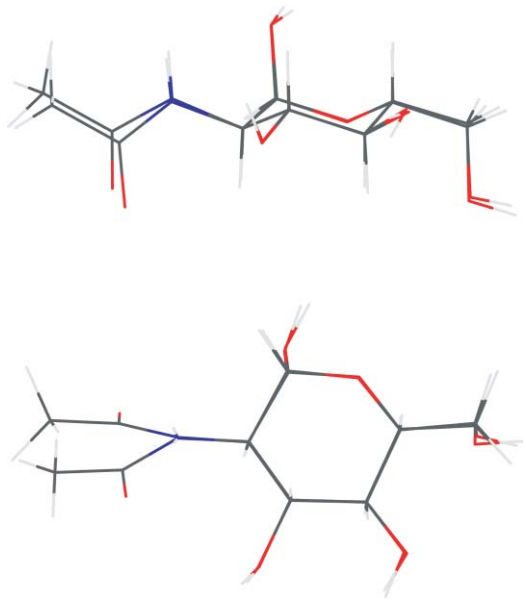

Fig. 6 Overlay of the crystal structure and *ab initio* refined average MD-structure of α-GlcNAc.

5. Conclusions

Density functional theory (DFT) quantum calculations were performed on *N*-acetylated amino sugars to make predictions of ^3J-couplings and the results were used to derive empirical Karplus equations for $^3J_{(H^NH^2)}$. The Karplus equations derived for these sugars are sensitive to conformation and while the trend is similar to those derived previously for proteins, the magnitude of the couplings in sugars is found to be consistently larger due to differences in chemical environment. It is therefore suggested that Karplus equations derived for proteins are no longer used for these important sugars. However, detailed comparison with experimental data indicates that these extensive calculations using the B3LYP method (even with a very large basis-set) tend to slightly overestimate three-bond scalar couplings, which cannot be corrected by any simple regularization process.

Calculated implicit solvent effects are small and lower the DFT-calculated coupling constants, which improves the agreement with the experimental data. The effect of local geometrical parameters, excluding the dependence on the intervening dihedral angle, is of the same order as implicit solvent effects. Due to the modest contributions from these effects it is concluded that the $^3J_{(H^NH^2)}$ and thus the Karplus parameters derived here are not sensitive to these effects. Inclusion of explicit solvent molecules did not improve the coupling constant calculations, but it was shown that these are important to maintain a realistic structure for the sugars during DFT minimization. For non-rigid bonds the largest effect on the scalar coupling constant is due to dynamical effects (bond libration). By integration of the Karplus equations, a relation is presented that can estimate the range of libration of the acetamido group from the DFT-derived Karplus parameters presented here and an experimental $^3J_{(H^NH^2)}$. Application of this method to the sugars studied here suggests that the dynamical spreads at the acetamido groups of α-GlcNAc, β-GlcNAc and α-GalNAc are 32°, 42° and 20°, respectively.

Abbreviations

GlcNAc: *N*-acetyl-D-glucosamine; GalNAc: *N*-acetyl-D-galactosamine; GAG: glucosaminoglycan; DFT: density functional theory; FC: Fermi contact; DSO: diamagnetic spin orbit; PSO: paramagnetic spin orbit; SD: spin dipolar; MD: molecular dynamics; PCM: polarizable continuum model; NOE: nuclear Overhauser effect; RDC: residual dipolar coupling.

Acknowledgements

The authors would like to thank Drs Ola Lutnæs and Torgeir Ruden for providing the HIIIsu3 basis-set and Dr Charles Blundell for helpful discussions and the experimental coupling constants of **1**α and **1**β. This work was funded by a Wellcome Trust project grant, reference number 075489.

References

1 A. Varki, *Cell*, 2006, **126**(5), 841–845.
2 P. L. DeAngelis, *Anat. Rec.*, 2002, **268**, 317–326.
3 I. S. Roberts, *Annu. Rev. Microbiol.*, 1996, **50**, 285–315.
4 P. R. Austin, C. J. Brine, J. E. Castle and J. P. Zikakis, *Science*, 1985, **212**, 749–753.
5 E. Weerapana and B. Imperiali, *Glycobiology*, 2006, **16**(6), 91R–101R.
6 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, ed. S. Gibbs, Garland Science, New York, NY, 4th edn, 2002.
7 M. Fukuda and O. Hindsgaul, *Molecular Glycobiology*, ed. B. D. Hames and D. M. Glover, IRL Press at Oxford University Press, Oxford, 1994.
8 R. Sasisekharan and J. R. Myette, *Am. Sci.*, 2003, **91**, 432–441.
9 A. Almond, *Carbohydr. Res.*, 2005, **340**, 907–920.
10 K. J. Jalkanen, V. W. Jurgensen, R. C. Wade, F. Nardi, C. Jung, I. M. Degtyarenko, R. M. Nieminen, F. Herrmann, M. Knapp-Mohammady, T. A. Niehaus, K. Frimand and S. Suhai, *Int. J. Quantum Chem.*, 2006, **106**, 1160–1198.
11 C. A. Bush, M. Martin-Pastor and A. Imberty, *Annu. Rev. Biophys. Biomol. Struct.*, 1999, **28**, 269–293.
12 T. Peters and B. M. Pinto, *Curr. Opin. Struct. Biol.*, 1996, **6**, 710–720.
13 A. Almond, P. L. DeAngelis and C. D. Blundell, *J. Mol. Biol.*, 2006, **358**(5), 1256–1269.
14 C. D. Blundell, P. L. DeAngelis, A. J. Day and A. Almond, *Glycobiology*, 2004, **14**(11), 999–1009.
15 J. H. Prestegard, H. M. Al-Hashimi and J. R. Tolman, *Q. Rev. Biophys.*, 2000, **33**(4), 371–424.
16 T. E. Klepach, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 2005, **127**, 9781–9793.
17 R. Stenutz, I. Carmichael, G. Widmalm and A. S. Serianni, *J. Org. Chem.*, 2002, **67**, 949–958.
18 F. Cloran, I. Carmichael and A. S. Serianni, *J. Am. Chem. Soc.*, 2000, **122**, 396–397.
19 O. L. Malkina, M. Hrivovini, F. Bizik and V. G. Malkin, *J. Phys. Chem. A*, 2001, **105**, 9188–9195.
20 V. F. Bystrov, *Prog. Nucl. Magn. Reson. Spectrosc.*, 1976, **10**, 41–81.
21 W. Deng, J. R. Cheeseman and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**(4), 1028–1037.
22 O. B. Lutnæs, T. A. Ruden and T. Helgaker, *Magn. Reson. Chem.*, 2004, **42**(Special issue), 117–127.
23 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochtersk, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, *Gaussian 03, D*, Gaussian Inc., Wallingford, CT, 2006.
24 D. A. Case, C. Scheurer and R. Bruchweiler, *J. Am. Chem. Soc.*, 2000, **122**, 10390–10397.
25 D. Zaccari, V. Barone, J. E. Peralta, R. H. Contreras, O. E. Taurian, E. Diez and A. Esteban, *Int. J. Mol. Sci.*, 2003, **4**, 93–106.
26 V. Sychorvsky, B. Schneider, P. Hobza, P. Zidek and V. Sklenar, *Phys. Chem. Chem. Phys.*, 2003, **5**, 734–739.
27 F. Corzana, J. H. Busto, G. Jiminez-Oses, J. L. Asensio, J. Jimenez-Barbero, J. M. Peregrina and A. Avenoza, *J. Am. Chem. Soc.*, 2006, **128**, 14640–14648.
28 C. D. Blundell, P. L. DeAngelis and A. Almond, *Biochem. J.*, 2006, **396**(3), 487–498.
29 K. N. Kirschner and R. J. Woods, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10541–10545.
30 D. Wei, H. Guo and D. R. Salahub, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2001, **64**, 011907.
31 W. Kutzelnigg, U. Fleischer and M. Schindler, in *NMR, Basic Principles and Progress*, ed. P. Diehl, E. Fluck, H. Gunther, R. Kosfeld and J. Seelig, Springer-Verlag, Berlin, 1991, vol. 23, pp. 165–262.
32 R. Bruschweiler and D. A. Case, *J. Am. Chem. Soc.*, 1994, **116**, 11199–11200.
33 B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**(2), 187–217.
34 T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 616–641.
35 K. Gilbert, *PCMODEL, 9.0*, Serena Software, Bloomington, IN, 2005.
36 F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pheifer and A. Bax, *J. Biomol. NMR*, 1995, **6**, 277–293.
37 M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870–2871.
38 G. W. Vuister and A. Bax, *J. Am. Chem. Soc.*, 1993, **115**, 7772–7777.
39 A. C. Wang and A. Bax, *J. Am. Chem. Soc.*, 1996, **118**, 2483–2494.
40 R. J. Abraham and R. Koniutou, *Magn. Reson. Chem.*, 2003, **41**, 1000–1008.
41 B. Coxon, *Carbohydr. Res.*, 2005, **340**, 1714–1721.
42 L. N. Johnson, *Acta Crystallogr.*, 1966, **21**, 885.
43 R. D. Gilardi and J. L. Flippen, *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.*, 1974, **B30**, 2931–2932.