Conformational dynamics of carbohydrates : raman optical activity of D-glucuronic acid and N-acetyl-D-glucosamine using a combined molecular dynamics and quantum chemical approach

Journal Name

RSCPublishing

ARTICLE

Conformational dynamics of carbohydrates: Raman optical activity of D-glucuronic acid and N-acetyl-D-glucosamine using a combined molecular dynamics and quantum chemical approach

Cite this: DOI: 10.1039/xoxxoooooox

Received ooth January 2012,
Accepted ooth January 2012

DOI: 10.1039/xoxxoooooox

www.rsc.org/

Shaun T. Mutter,[a] François Zielinski,[b] James R. Cheeseman,[c] Christian Johannessen,[d] Paul L. A. Popelier,[b] and Ewan W. Blanch[a*]

As two biologically and medically relevant monosaccharides, the constituents of hyaluronic acid, D-glucuronic acid and N-acetyl-D-glucosamine, constitute perfect test cases for the development of carbohydrate-specific structural methods. These two molecules have been analysed by Raman Optical Activity (ROA), a spectroscopic technique exhibiting exquisite sensitivity to stereochemistry. We show that it is possible to support the experiment with a simulation approach combining density functional theory (DFT) and molecular dynamics (MD), both using explicit solvation. Thus, we have gained new insight into the crucial hydration effects that contribute to the conformational dynamics of carbohydrates and managed to characterize in detail the poorly understood vibrational nature of this class of biomolecules. Experimental and calculated ROA spectra of these two molecules are reported and excellent agreement has been found. More specifically, comparison has been made with the more commonly used gas phase and implicitly solvated calculation approaches, which offer poor or zero modelling of solvent interactions. The calculated spectra have been used to resolve the structural origins of the observed bands, a current challenge in the study of carbohydrates due to a lack of definitive vibrational assignments. We report and analyse major features in the fingerprint region of the ROA spectra, with recurrent structural and spectral features between the two monosaccharides observed.

Introduction

Raman optical activity (ROA) spectroscopy is a powerful analytical technique that measures small differences in the Raman scattering of chiral molecules using circularly polarized light.[1-3] Since its inception, this chiroptical approach has developed into an important tool for the treatment of a variety of biological and chemical problems, including assignment of absolute configuration, conformational analysis, and structure elucidation of biomolecules. ROA is also particularly useful when the more traditional methods of structure determination are not applicable.[4-9] Whilst advances in instrumentation have led to the development of ROA, it has also greatly benefited from computational modelling. There is a mutually beneficial complementarity between experiment and theory, as it is routinely possible to calculate the ROA spectra of small-to-medium sized systems with modern *ab initio* methods. In turn, comparison of a modelled spectrum to experiment can reveal structural features and offer insight into conformational dynamics, whilst it is not always obvious from the experimental data.[10-12] The reverse is also true, as the high sensitivity of ROA to changes in molecular geometry allows it to be used as a gold standard in guiding force field design and solvation modelling.

Unlike conventional Raman spectroscopy, which arises solely from the interaction of the incident light with the electric polarizability, ROA also considers interactions with the magnetic polarizability and electric quadrupole. As such, in the simplest description of ROA, where the excitation wavelength is far from any electronic resonance, the intensity difference is dependent on the three types of scattering tensor: the electric dipole-electric dipole polarizability tensor, the electric dipole-magnetic dipole optical activity tensor, and the electric dipole-electric quadrupole optical activity tensor. These tensors can be averaged over all possible molecular orientations to derive the equations for the Raman and ROA invariants, which can be used to write intensity expressions covering all possible scattering geometries and polarization modulations.[13]

One of the areas of application that has generated the most interest in ROA spectroscopy is that of protein structure, with many recent experimental and computational based publications on peptides and amino acids.[6,14-20] Comparatively few ROA studies of carbohydrates have been reported in the literature, while even fewer have considered the theoretical

treatment of carbohydrate ROA spectra. One of the earliest was reported by Macleod *et al.*[21], examining D-lactose, D-glucose, and D-galactose and highlighting the importance of conformational averaging and inclusion of a solvent model. This last point was later explored by Luber and Reiher in their study on an anhydro derivative of D-glucose, by comparing calculated ROA spectra using implicit COSMO solvent models and a small number of explicit solvent molecules.[22] They demonstrated that explicit hydration has a significant effect on the calculated spectra and noted that better agreement with experiment could be achieved by incorporating molecular dynamics (MD) simulations with quantum mechanics / molecular mechanics (QM/MM) calculations. Kaminský *et al.*[23] used MD simulations with *ab initio* calculations to show the importance of conformational averaging for ROA simulations on carbohydrates by studying the highly flexible linear gluconic acid anion. Using a tensor transfer approach to create a large library of spectra, they were able to reproduce most of the observable experimental features.

More recently, Cheeseman *et al.*[24] obtained excellent agreement between the calculated and experimental spectra of methyl-β-D-glucose, by incorporating a full MD simulation of hydration effects and including a large number of explicit solvent molecules in the QM/MM calculation of ROA. In this paper we present a significant development to this approach, with the computational modelling of ROA spectra of two monosaccharides, D-glucuronic acid (GlcA) and N-acetyl-D-glucosamine (GlcNAc). GlcA and GlcNAc are molecules or particular interest as they make up the repeating disaccharide subunit of hyaluronan, an important glycosaminoglycan that performs vital regulatory and structural roles in animal tissues as well as being a highly important healthcare product.[9] Despite its undoubted importance, there is still dispute over the three dimensional structure of hyaluronan, given the several conflicting descriptions reported in the literature.[25-28] These two molecules also offer test cases for ROA calculations of this nature, with GlcA existing in an anionic form in aqueous solution and GlcNAc exhibiting much greater conformational flexibility than methyl-β-D-glucose or any other carbohydrate previously investigated, due to the presence of an unconstrained acetylamine moiety. It should be noted that for these cases unconstrained simulations have been used, contrary to the case studied in reference 24, where known stable conformers were used to constrain MD sampling. Therefore this method is more generally applicable, as no prior conformational information is required. Ball-and-stick representations of the structures of GlcA and GlcNAc, including a numbering scheme, are shown in figure 1.

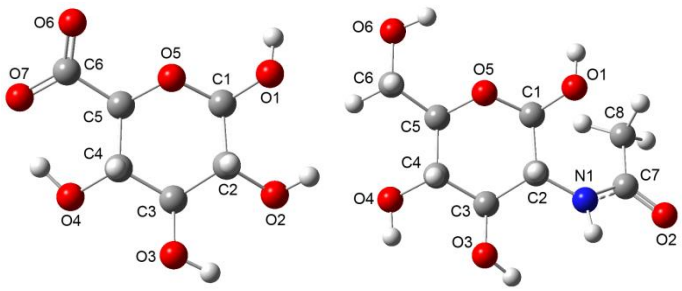


Fig. 1 Ball and stick representations of GlcA (left) and GlcNAc (right) with the numbering scheme used herein. Carbon, hydrogen, oxygen, and nitrogen atoms are shown in grey, white, red, and blue, respectively.

The importance of solvation effects to the conformational behaviour of biological molecules is well-recognized and as such it is highly recommended that ROA calculations on biomolecules include a solvent model. Implicit solvent models, in which the solute is placed in a cavity in a structureless continuum with a dielectric charge to represent the solvent, such as the polarizable continuum model (PCM), are widely used and are effective in studies on peptides.[29-31] However, as was shown by Cheeseman *et al.*[24] and Luber *et al.*,[22] this approach is not generally suitable for carbohydrates, which contain multiple solvated hydroxyl groups that are only correctly modelled when explicit solvent molecules are present. Here, we included the full explicit modelling of hydration effects within the MD simulations and QM/MM calculations.

Conformational analysis is another important facet of ROA calculations that needs to be carefully considered for accurate computations. Indeed, it is fundamentally assumed that the experimental spectra can be recovered by a time-average over all the low-energy conformer geometries. Such information can be trivially extracted from the MD trajectories used to find the various conformers. However, two problems can arise: sampling bias if the simulation length is too short and the accuracy and transferability of MD force fields cannot always guarantee reliable conformer ratios. The optimal approach is then to use a Boltzmann average of all the energy minimized conformers to weight the calculated spectrum, as it prevents any sampling bias and can be based upon non-dynamic computations at higher levels of theory. This approach is suitable for gas phase calculations and calculations using implicit solvent models but is not necessarily applicable to QM/MM calculations with large numbers of water molecules treated with force field methods. Maintaining a consistent solvent cutoff distance across multiple MD configurations, resulting in fixed diameter spherical snapshots, can lead to slight fluctuations in the number of solvent molecules, resulting in large variances in calculated energies and inaccurate Boltzmann weightings. Within this work, we have resorted to using the MD conformer populations for the weighting of spectra from QM/MM calculations. Very long MD simulations were then carried out to ensure extensive sampling. Furthermore, we are handling common and isolated monosaccharides with a carbohydrate-specific force field, in cases simple enough that classical MD can still be considered a reasonable source of quantitative conformational information.

Conformational flexibility of these molecules is an important concern to address and, as such, likely conformers have been identified. However, the consequences of the rotation of hydroxyl groups have not been analysed in depth here because inclusion of multiple snapshots should present a time-averaged view of preferred C-O-H positions.

The flexible GlcNAc molecule has two sites that will likely give rise to different conformers. Rotation of the O5-C5-C6-O6 dihedral angle (noted as ω in the following) to the positions -60˚, +60˚, and 180˚ gives the three rotamers gauche-gauche (gg), gauche-trans (gt), and trans-gauche (tg), respectively. Several experimental and computational studies on glucopyranosides have shown that the tg conformers exists in negligible amounts and that there is a 60:40 relationship for gg:gt.[32-34] The other potential site for exhibiting different conformers of GlcNAc is the acetylamine group, which can exist in several conformers rotating around the dihedral angle represented by H-N1-C2-H (noted γ in the following) .The simpler GlcA molecule, on the other hand, exhibits only a single geometrical feature that will likely result in different

rotamers, namely the freely rotating anionic CO_2 group represented by the dihedral angle O5-C5-C6-O6 (also denoted as ω in the following, it should be noted however that it is not associated here with a potential glycosidic linkage as with GlcNAc).

Whereas ROA spectra of proteins typically enjoy well-assigned spectral features relating to structural parameters, ROA spectra of carbohydrates lack of definitive peak assignments. The vibrational modes observed in carbohydrates are complex and are frequently delocalized across most, if not all, of the molecule, making traditional assignment approaches difficult to apply. The early experimental ROA spectra of carbohydrates were reported by Barron and coworkers,[35,36] who also carried out assignments based on studies on different anomers of several closely related monosaccharides. Three main spectral regions of importance were identified, with 750-950 cm^{-1} relating to anomeric configurations, 950-1200 cm^{-1} providing information about ring structure and the pattern of substituents, and the region above 1200 cm^{-1} being dominated by CH_2 and COH deformations. Cheeseman *et al.*[24] assigned the Raman and ROA bands for the two observed rotamers of methyl-β-D-glucose based on gas phase calculations. However, due to the complexity of the vibrational motions involved, definitive assignments of most carbohydrate modes are yet be made.

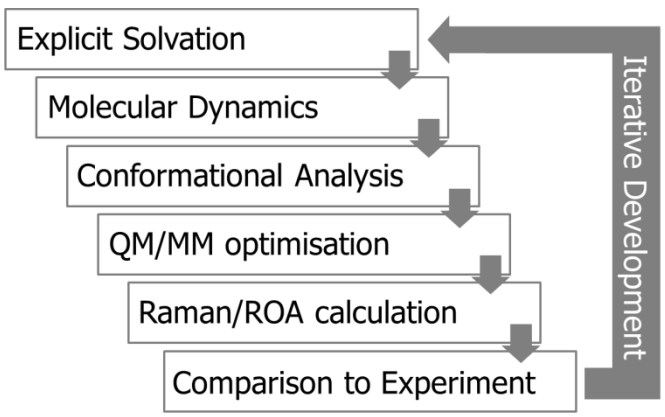


Fig. 2. Schematic representation of the steps involved in our ROA spectral calculation.

The approach used in this work for the calculation of ROA spectra can be expressed by the schematic representation shown in figure 2. The process starts with explicit solvation of the target molecule, followed by MD simulations to fully model hydration effects. Analysis of the trajectories then follows, in order to give an insight into conformational dynamics. These initial steps provide a sound basis for the subsequent quantum chemical calculations and have been adapted to best recreate the experimental spectra. This representation also shows the potential for development, as carbohydrate stereochemistry makes accurate calculation of spectra difficult to achieve and not all spectral features will be expected to be reproduced exactly. The observation of incorrectly modelled features can then lead to improvements of the simulation approach, in an attempt to improve the accuracy of results, reinforcing the complementary nature between ROA experiment and theory.

Technical Details

Computational Methods

All *ab initio* calculations were carried out using Gaussian09,[37] while semi-empirical calculations were carried out using MOPAC2009,[38] and MD simulations using DL_POLY v4.[39]

Isolated and Implicitly Hydrated Molecules

A PM6 relaxed scan around the ω angle of the anionic form of GlcA was carried out resulting in one low-energy conformer, with a value of ω = 17.5°. For this conformer further geometry optimization and force field calculations were carried out at the B3LYP level of density functional theory (DFT), utilizing the 6-31G(d) basis set.

For GlcNAc PM6 relaxed scans were carried out around the ω torsion angle and the acetylamine moiety γ with two low-energy conformers observed, corresponding to gauche-gauche (gg) and gauche-trans (gt). For both conformers, the values of the γ angle are close to zero. Geometry optimization and force field calculations were carried out for both conformers at the B3LYP/6-31G(d) level of theory. As well as gas phase calculations the effect of water as a solvent was also modelled using the CPCM polarizable conductor calculation model.

Explicitly Hydrated Molecules

The QM/MM method was used to model the explicitly solvated systems of the two monosaccharides. The two-layer ONIOM method was used,[40] as implemented in GAUSSIAN09, with the electronic embedding scheme, which incorporates MM partial charges in the QM Hamiltonian.[41] The initial geometries for calculations were taken from snapshots from the MD simulations, with the high layer/QM region for the monosaccharide and the low layer/MM region for all of the water molecules. The high layer was treated at the B3LYP/6-31G(d) level of theory and the low layer treated with AMBER force field (parm96), including TIP3P parameters for the water molecules.

Despite having geometries obtained from MD simulations, optimization still needed to be carried out, as vibrational mode calculations are only valid for minima on the potential energy surface. For the geometry optimization and harmonic frequency calculations two approaches were used: (i) the water molecules were frozen in the MD snapshot geometry and only the sugar optimized (henceforth referred to as OPTGlcA and OPTGlcNAc), and (ii) the second where the entire system was optimized (called OPTALL).

Optimization of systems for both approaches was carried out in Cartesian coordinates. However, the OPTALL scheme was problematic due to systems of this size having relatively flat potential energy surfaces, making full optimization difficult. To circumvent this, systems were considered optimized when the electronic energies for each optimization step changed in an oscillating fashion but overall did not become lower, as well as when the maximum force and RMS force values converged below a threshold (0.000450 au and 0.000300 au, respectively). The OPTGlcA and OPTGlcNAc schemes were much less computationally intensive and full optimization was reached in both cases.

Calculation of ROA Intensities

Frequency dependent ROA tensors were calculated using magnetic field dependent basis functions (GIAOs),[42,43] using the fully analytic two-step procedure, the (n+1) algorithm.[12] The two-step procedure calculates the ROA tensors at a level of theory different to that used for the optimization and harmonic

frequency calculation, with the ROA tensors calculated at the B3LYP/rDPS level of theory, as recommended by Cheeseman and Frisch.[44] rDPS is a rarefied basis set, constructed from 3-21++G with semi-diffuse *p* functions on all hydrogens with an exponent of 0.2. Benchmark studies have shown that this basis set returns ROA intensities at a similar level of accuracy compared to those calculated with much larger basis sets.[45] The excitation wavelength for all of these calculations was set at the experimental value of 532 nm.

Scattered circular polarization backscattered (SCP-180) ROA intensities at far-from-resonance conditions were obtained from combinations of the appropriate invariants and included ν^4 and Boltzmann factors for comparison against experimental spectra. Calculated spectra were generated using a Lorentzian bandshape with a peak half-width of 10 cm^{-1}. All calculated spectra are shown in arbitrary units.

Conformational Averaging

All reported calculated spectra have been weighted to account for the conformational dynamics present in the experimental systems. The isolated and implicitly hydrated molecules were averaged based upon Boltzmann weightings, calculated from the free energies of individual conformers, except for GlcA which only exhibited one major conformer. The OPTALL, OPTGlcA, and OPTGlcNAc calculated spectra were weighted based upon the conformational populations observed in the MD simulations, for the relevant major conformers.

Spectra of individual conformers, used in the weightings, were constructed from averaging the calculated spectra for the six selected snapshots. Averaging of spectra was carried out using the lineshape form, i.e. the spectra were generated for each snapshot and the curves averaged (instead of generating a spectrum from average ROA intensities at individual frequencies).

Molecular Dynamics Simulations

Gas-phase optimized structures of the considered monosaccharides were obtained as pdb files from the GLYCAM online tool. The required input files were generated with DL_FIELD v3.1,[46] using the GLYCAM force field parameters. Each configuration was explicitly solvated within a 30 Å cubic cell filled with TIP3P water molecules, suitable for periodic boundary condition. A sodium counterion was included in the GlcA system for completion of the simulation. Electrostatic interactions were handled by DL_POLY's parallel implementation of the Smooth Particle-Mesh Ewald method.[47] According to a number of tests and benchmarks a 7 Å cutoff ($>2\sigma_{max}$) was chosen for the real-space Ewald Sum and the van der Waals interactions, in order to maximize computational efficiency. Such a cutoff has been considered sufficient for the present work.

Flexible water molecules were used for the first processed monomer, GlcA, resulting in snapshots difficult to optimize during the subsequent QM/MM treatment, because of the large kinetic distortions encountered by some water molecules. Consequently, the water molecules were handled as rigid bodies for the MD simulations of GlcNAc, in order to expedite the process and reduce the extra computational cost associated with multiple conformers.

To comply with the experimental conditions of ROA analysis, the snapshots have been extracted from NPT dynamic trajectories (integration time step 0.5 fs), regulated by a Berendsen coupling set up at 298 K and 1 atm (time constants: 0.1 and 1.0 ps, respectively). Simulation timeframes of 30 ns and 50 ns for GlcA and GlcNAc, respectively, have been deemed necessary to sample adequately the conformational space of the considered monosaccharides.[34]

The equilibration processes were differently set up for each molecule. Indeed, the coarse solvation process used to generate the initial configurations leads to a system lying far from equilibrium. In the case of GlcA, the initial state of the system was equilibrated over 1 ps, with a force capping of 1000 kT/Å, a timestep of 0.1 fs, and a temperature scaling every 5 steps. A subsequent check with a run in NVT ensemble (Berendsen thermostat, a similar set up as for the NPT run) showed a sudden rise in temperature and Coulombic energy linked to the transition from equilibration to full dynamics. However, the system properties show quick stabilization and the proper sugar structure conserved after 2 ns. The final configuration was then retained as the NPT run starting point. Regarding GlcNAc, the same process proved unsuccessful: the transition from equilibration to MD generated extreme distortions leading to significant structural transformations. In order to obtain a properly equilibrated system, it was necessary to use a force capping of 10,000 kT/Å over a 3 ps equilibration run. The NPT starting configuration was in this case extracted after an extra 3 ps of NVT check run.

Experimental Measurements

N-acetyl-D-glucosamine (GlcNAc) and the sodium salt of D-glucuronic acid (GlcA) were purchased from Sigma-Aldrich and used without further purification. The ROA spectra were measured at ambient temperature in water using the previously described ChiralRAMAN instrument (BioTools, Inc.),[48] which employs the scattered circular polarization (SCP) measurement strategy. The ROA spectra are presented as circular intensity differences (I_R-I_L) with I_R and I_L denoting the Raman intensities with right- and left-circular polarization, respectively. The sample concentrations were ~150 mM for each sample, with pH values of 7.0 and 6.3 for GlcA and GlcNAc, respectively. Experimental conditions: laser wavelength 532 nm; laser power at the sample 350 mW; spectral resolution 7 cm^{-1}; acquisition times 17-18 hrs. Cosmic ray spikes were removed from the ROA spectra by means of a median filter, after which the spectra were smoothed using a second-order Savitzky-Golay filter.

Results

A consideration when using explicit solvent models based around MD simulations for ROA computation, is the number of snapshots needed to give good agreement with experiment. Cheeseman *et al.*[24] used sixteen snapshots for each of the two conformers of methyl-β-D-glucose, taken at intervals of 25 ps, which were sufficient to provide excellent agreement with the experimental spectrum. Recently, Urago *et al.*[19] carried out similar MD and QM/MM calculations on a cyclic dipeptide and found significantly more snapshots were required for satisfactory agreement, especially in the high frequency regions. It should be noted that the authors in reference 19 used a scheme comparable to OPTGlcA and OPTGlcNAc and not the arguably more accurate OPTALL approach.

Selecting a statistically relevant amount of MD snapshots for the subsequent QM/MM step by random sampling, would result in a large set of computationally expensive calculations for such explicitly solvated sugars. As such our objective is to select a minimal amount of snapshots encompassing the

dominant conformers, which are not known for solvated GlcA and GlcNAc. It is necessary to monitor the evolution, along the MD trajectories, of the geometric features susceptible to present different occurring forms. In this study, we focused on the significant dihedral angles for each monomer and assessed their distributions by counting the values as per intervals of 5 degrees. For the sake of visual clarity, these distributions are not plotted as histograms in figures 3 and 4, but as lines linking the binned data points.

For both considered monosaccharides, the cycle's puckering was analysed by a pair of complementary dihedral angles abbreviated as α and β, which correspond to O5-C1-C2-C3 and O5-C5-C4-C3, respectively, as shown in figure 1. In both cases, fairly rigid chair conformations are observed, without any sign of secondary conformers. The considered dihedral angles show similar values: for GlcA, with average values of -52.5°/48.7° (standard deviations of 6.6°/7.5°) for angles α and β respectively, while average values are -57.7°/49.8° (standard deviations 7.0°/6.7°) in the case of GlcNAc. In the simple case of GlcA, we considered a single other dihedral, ω, linked to the orientation of the carboxylate group with respect to the cycle. Along the whole trajectory the dihedral's absolute values were monitored since both carboxylic oxygens are equivalent. As shown in figure 3 only a single conformer can be observed, although the large standard deviation around its average value implies significant flexibility.

In the case of GlcNAc, both the acetylamine moiety γ and the ω torsion angle were monitored, since distinct conformers can be expected to emerge from some energetically favoured arrangement of their atoms. For the ω torsion angle, we observed two significant conformations: the gg population amounting to 54% whereas the gt conformation occurs in 42% of the trajectory frames. The orientation of the hydroxymethyl terminal hydrogen atoms showed preferred orientations as well, but these were not considered and thus discarded from the snapshot selection. In the case of the large acetylamine group, the methyl end-group and the internal geometry of the substituent were not involved in the selection process. Indeed we observed that the former is freely rotating and that the latter stays planar (average -0.3°, standard dev. 9.8°). However, as shown in figure 4, two preferred orientations of the substituent were observed, with angles of circa -180° and 0° (populations 18% and 82%, respectively). Although the minor orientation is more than 4 times less frequent than the major one, according to the MD sampling, we consider this conformer significant enough to be accounted for in the snapshots selection.

For each significant conformer observed by monitoring the dihedral angles, six snapshots were selected from a table of MD frames featuring the values of the monitored angles (including cyclic dihedrals). By means of usual table functions, candidate frames were extracted when exhibiting values lying within target intervals for each angle. That is to say, the GlcA processing resorted to only 6 snapshots, whereas 24 were used for GlcNAc and its 4 identified conformers. Typically (4 snapshots out of the 6 for a single conformer), these intervals were centred on the monitored angles' average values, as observed on the trajectory histograms, with a +/-1° spread (this condition was progressively relaxed to find a match, when needed). Furthermore, in order to provide a diversity of starting structures to the QM/MM process, slight distortions were applied to 2 snapshots, by centring the target intervals onto +/-σ off the average values. As seen in figure 3, the snapshots selected for GlcA are slightly off the ω angle's peak maximum value. Since a single significant conformer could be observed from the raw trajectory plot of the ω angle values, the global average has been used to centre the snapshots. To improve centring, and to deal with the multiple conformers of GlcNAc, a preliminary analysis of the trajectory histograms was introduced in the method. The angle value averaging could then be restricted to each conformer's domains, i.e. intervals limited to the corresponding peak width.

As a general guideline, the snapshots were chosen from different parts of the trajectories, in order to explore the diversity of the solvent layer, and not before the volume could properly equilibrate. From the selected frames, only the atoms located within a 12 Å sphere centred on the sugar were extracted. In the case of GlcA, frames were selected where the sodium counterion was at a minimum distance of 10 Å from the sugar, so as to have no bearing on the results of the vibrational calculations.

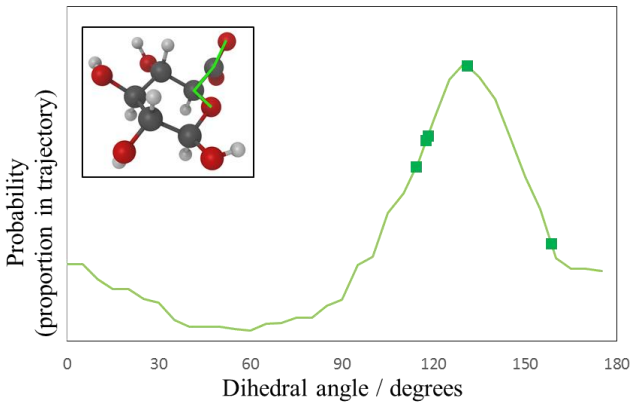


Fig 3. The distribution of the ω torsion angle in GlcA. The squares mark the values assumed by this angle at the selected snapshots.

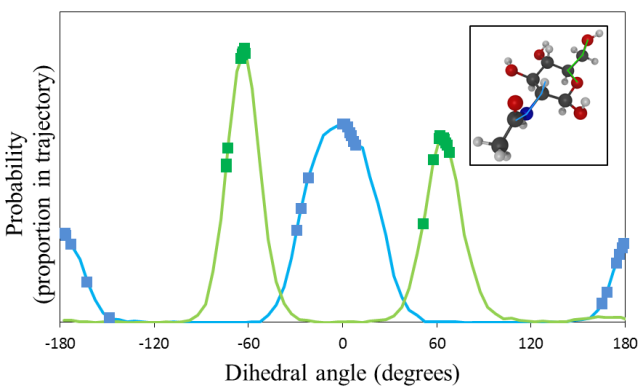


Fig 4. The distribution of selected dihedral angles in GlcNAc. Squares mark the values assumed by this angle at the selected snapshots. Green corresponds to the ω torsion angle and blue to the acetylamine moiety γ.

The calculated and experimental ROA spectra of GlcA are presented in figure 5. Panels A and B represent the gas phase and PCM calculated spectra, respectively. Panels C and D represent the two combined MD-QM/MM approaches, OPTGlcA and OPTALL, respectively, with panel E showing the experimentally measured ROA spectrum. These spectra reveal strong agreement between panels D and E. This shows that the OPTALL approach can reproduce the experimental spectrum for this monosaccharide very well. Most band positions and signs are correctly predicted and most relative

intensities are close to those observed in the experiment. However, it should be noted that there are some features that are not as well reproduced. The strong positive peak at 1350 cm^{-1} is offset by 60 cm^{-1}, appearing at 1410 cm^{-1} in the unshifted OPTALL spectrum, as well as exhibiting a broader peak width and comprising of several bands whilst the experimental peak is one dominant band with a single shoulder at lower wavenumber. Also, the weakly positive singlet peak at 1042 cm^{-1} in panel E appears as a weakly positive doublet in D. Despite these minor differences the bandshape, peak intensities, and peak wavenumbers are in excellent agreement, especially compared to the spectra calculated using the other three approaches.

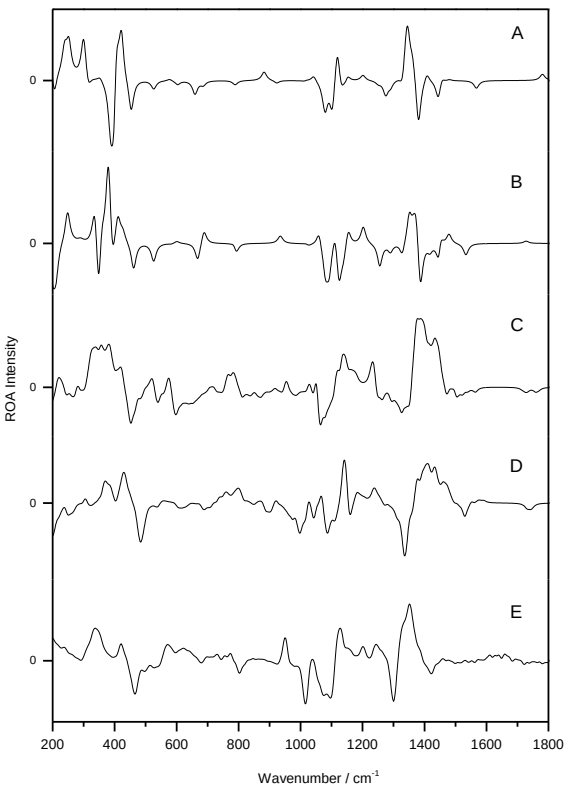


Fig. 5 ROA spectrum of GlcA. Calculated gas phase spectrum (A), calculated spectrum using PCM (B), OPTGlcA spectrum (C), OPTALL spectrum (D), and experimental (E).

The importance of correctly modelling solvent interactions has already been noted, and is reinforced here as the gas phase and PCM ROA calculations are in very poor agreement with the experimental spectrum, as shown in figure 5. This is particularly evident at low wavenumbers; although as vibrational modes in this region are strongly affected by the solvent this is not unexpected for gas phase and implicitly solvated systems. The OPTGlcA scheme shown in panel C offers a significant improvement over the gas phase and PCM calculations, although some of the finer details shown by the OPTALL approach are lost in this case. This is exemplified in the low wavenumber region as OPTALL reproduces the +ve/+ve/-ve profile at 300-500 cm^{-1}, whilst OPTGlcA shows a broad +ve/-ve profile. This difference is likely due to the unconstrained movement of water during the optimization step of OPTALL, allowing better modelling of the vibrational modes that couple with solvent molecules. Comparison of the calculated spectra in figure 5 shows that as we progress through panels A to D there is an improvement in the level of agreement with the experimental spectrum. This also correlates with a more accurate modelling of the solvent environment, confirming that accurate treatment of solvent is essential for these calculations.

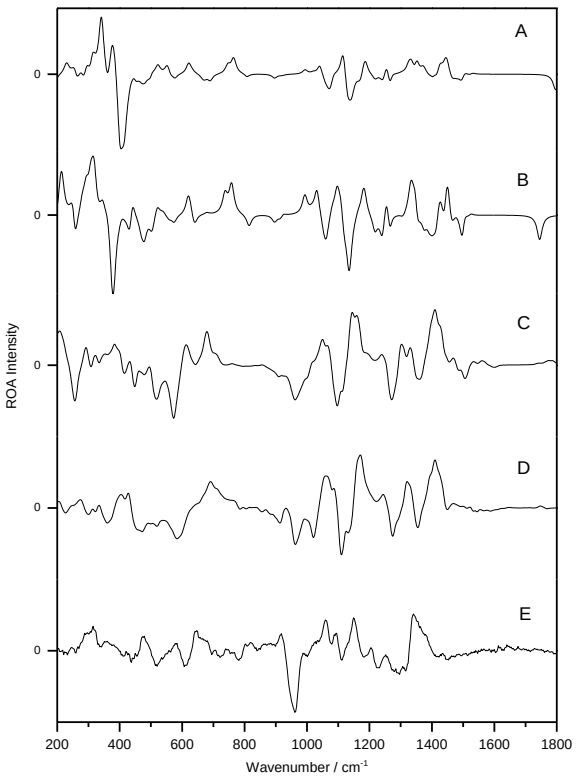


Fig 6. ROA spectrum of GlcNAc. Calculated gas phase spectrum (A), calculated spectrum using PCM (B), OPTGlcNAc spectrum (C), OPTALL spectrum (D), and experimental (E).

The calculated and experimental ROA spectra of GlcNAc are shown in figure 6, with the same panel representations as GlcA in figure 5. Comparison of panels D and E, OPTALL and experimental spectra, respectively, show good agreement in respect to bandshape and peak wavenumbers, once again demonstrating the power of this approach for reproducing experimental spectra of monosaccharides. In this case the intensities of the experimental peaks have not been completely reproduced, which is especially apparent for the strong negative peak at 957 cm^{-1} in panel E which is calculated as a lower intensity peak using the OPTGlcNAc and OPTALL approaches. The weakly negative minimum at 1252 cm^{-1} in the experimental spectra is also poorly modelled with a moderately

intense positive signal in panels C and D. Possible explanations for these differences are explored below in the discussion of the vibrational modes of the calculated spectra The calculated OPTALL and OPTGlcNAc spectra also exhibit the same slight shift to a higher frequency exhibited by the calculations of the GlcA spectra.

Despite the more notable differences between the experimental and OPTALL spectra of GlcNAc the agreement is still very good especially in light of this molecule's higher conformational flexibility. The importance of correct modelling of the solvent for this system is also clear, with panels A and B (gas phase and PCM calculated spectra, respectively) giving very poor representations of the experimental spectrum, especially in the low wavenumber region. Unlike the MD approaches, which generated the calculated spectra from conformer populations, the gas phase and implicitly solvated spectra were weighted using Boltzmann averages. For the spectra shown in panels A and B in figure 6, the gas phase spectrum (A) had a ratio of 63:37 for gg:gt and the PCM calculated spectrum (B) had a ratio of 56:44 for gg:gt. Moving from the gas phase and implicitly solvated models to the explicitly solvated approaches yields a significant improvement. Unsurprisingly, the OPTALL approach performs better than OPTGlcNAc. Panel C, OPTGlcNAc, particularly shows worse agreement with experiment in the low wavenumber region, as well as predicting several peaks in the fingerprint region that do not exhibit shoulders or coalesce to single peaks when compared to OPTALL, the more powerful approach.

A current challenge for the study of carbohydrate conformations using ROA spectroscopy is the resolution of the structural origins of observed bands. Other biomolecules, such as peptides and proteins, have well-defined spectral regions that correspond to primary and secondary structural parameters, whilst there is relatively little detailed information in the literature with regards to carbohydrate peak assignments. Calculations of ROA spectra, such as those presented in this article, can offer insights not available from experiment by allowing the visualization of vibrational modes. However, the conformational flexibility of carbohydrates makes it difficult to assign definitive vibrational modes of monosaccharides, as their vibrations are typically diffuse and often spread over most of the considered molecule. As such, the only reported results that extensively assign vibrational modes from calculated ROA spectra of carbohydrates were those of Cheeseman *et al.*[24] on methyl-β-D-glucose and these show very complicated assignments, with multiple stretches, bends, or deformations for modes within the fingerprint region. To our knowledge, no one has previously attempted to base peak assignments on spectra calculated from approaches using MD and QM/MM, which include a large number of explicit solvent molecules. This is likely due to the fact that visualization of individual vibrations can be greatly obscured by the relatively large system size and to accurately assign peaks to a spectrum formed from weighted spectra each individual snapshot needs to be analysed.

Assignment of vibrational modes has been carried out to further analyse the ROA spectra of GlcA and GlcNAc, enabling interpretation of bands in the calculated spectra. Table 1 shows the ROA band assignments for GlcA based on the spectrum calculated using the OPTALL approach. The assignments have been determined by identifying the main features in panel D in figure 5 and analysing the vibrations that contribute to those peaks from the 6 spectra used to generate it from the 6 snapshots. By correlation of visual analysis with atomic displacement values from the calculation output, it was possible to corroborate the inclusion of particular atoms in more localized vibrations. This allows the multiple diffuse vibrations to be reduced to the most significant contributions of the observed peaks.

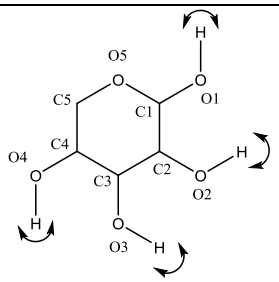
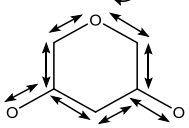
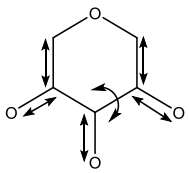
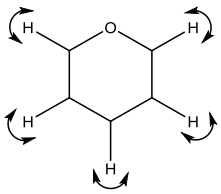
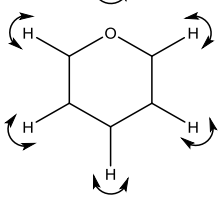
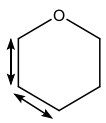
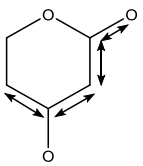
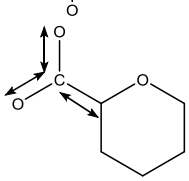
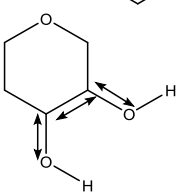
Within table 1 the peak assignments for the significant features for 1000 cm^{-1} and above are reported alongside a graphical representation of the vibrational mode. It should be noted that despite this new approach for visualizing particular modes by combining visual analysis with atomic displacement over multiple snapshots, many atoms within the molecule were often involved in diffuse vibrations, only the most intense and recurrent contributions are then shown. Where we say that vibrations are related to the plane of the molecule, this refers to an approximate molecular plane based on the six-membered ring. Features in the lower wavenumber region have not been explored because of their lower energy, coupled to the librational nature of these vibrations with strong contributions from the solvent. This, generally results in most atoms of the molecule being involved in these vibrations, thereby making the modes diffuse and their descriptions nondistinctive.

Table 2 shows the ROA band assignments for GlcNAc based on the OPTALL approach. Band assignments were obtained in the same fashion as for GlcA. However, because in this case 6 spectra were used for each conformer to form the weighted spectrum, all 24 snapshots were analysed to form the visualizations of the vibrational modes.

The previous experimental work by Barron *et al.*[35,36] on monosaccharide peak assignments revealed three regions of interest: 750-950 cm^{-1} showing anomeric configurations, 950-1200 cm^{-1} showing ring structure, and 1200 cm^{-1} and above showing CH_2 and COH deformations. The region relating to anomeric configurations has not been analysed further here. Since GlcA and GlcNAc adapt the same anomeric conformation one would not expect to observe significant differences in this region. For the vibrations above 1200 cm^{-1} there are strong occurrences of C-H wagging motions for both GlcA and GlcNAc, although both molecules exhibit some bond stretches within the rings and of the substituents. The modes calculated for the 950-1200 cm^{-1} region prove to be dominated by stretches relating to ring structure and some substituents. The most significant difference between these results and the reported experimental assignments is that calculated spectra for both molecules have peaks close to 1000 cm^{-1} exhibiting C-O-H deformations, downshifted from the region of 1200 cm^{-1} and above.

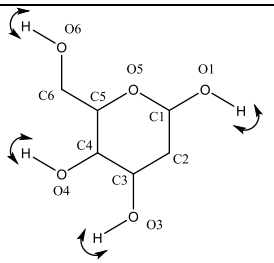
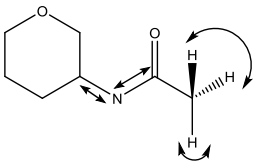
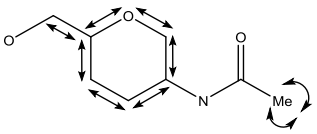
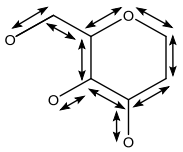
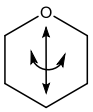
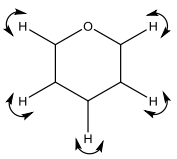
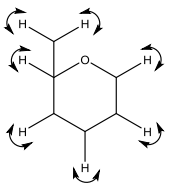
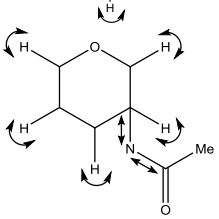
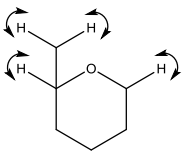
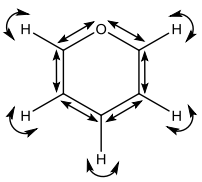
The data in table 2 can be used to examine the regions of the calculated spectra in figure 6 that are not as well reproduced and offer insights into any deficiencies of the current GlcNAc calculations. The strong negative peak in the experimental spectrum at 957 cm^{-1} correlates with the lower intensity peak assigned with O-H wags at 960 cm^{-1} for the OPTALL spectrum. The intensity difference between the two may be due to coalescence of this feature with the vibrations relating to the acetylamine moiety at 1020 cm^{-1}. These two vibrations appear as a distinct doublet in panel D in figure 6, which is not observed in panel E, the experimental spectrum. For the strong negative peak in panel E there is a damped shoulder at 1000 cm^{-1}, which could correlate with the calculated peak at 1020 cm^{-1} so resulting in the larger intensity observed in the experiment compared to the calculation.

Table 1 ROA band assignments for GlcA based on OPTALL scheme.

Wavenumber (cm^{-1})	Assignments	Visualisation of Mode[1]
1000	O-H in-plane wag	
1090	Coupled asymmetric C-C with C-O stretches	
1145	C1-C2-O2 & O4-C4-C5 asymmetric stretches C3-O3 stretch C2-C3-C4 bond deformation	
1240	Coupled C-H wag	
1335	Coupled C1-H C5-H wags C2-H, C3-H, C4-H wags Weak C-C stretches	
1360-1400	C-C stretches C-H wags	
1400-1460	C-C stretches C-O stretches C-H wag	
1460-1500	Coupled O-C-O symmetric & C5-C6 asymmetric stretch	
1530	C2-O2-H & C3-O3-H bend O2-C2-C3 asymmetric stretch	

[1] Numbering convention used for the visualization of vibrations is the same as shown in figure 1.

Table 2 ROA band assignments for GlcNAc based on the OPTALL scheme.

Wavenumber (cm^{-1})	Assignments	Visualisation of Mode[1]
960	O-H in plane wags	
1020	Methyl deformation C2-N1 stretch N1-C7 stretch	
1030-1090	Various methyl deformations Coupled asymmetric C-C with C-O stretches	
1100-1140	Coupled asymmetric C-C with C-O stretches	
1170	Ring deformations	
1245	C-H wags	
1275	C-H wags CH_2 twist	
1320	C-H wags Asymmetric C2-N1-C7 stretch	
1355	C1-H & C5-H wags coupled with CH_2 twist	
1410	Various C-C stretches, C-H wags coupled with CH_2 deformations	

[1] Numbering convention used for the visualization of vibrations is the same as shown in figure 1.

Use of peak assignments, such as those shown in tables 1 and 2, and comparisons of spectra of similar monosaccharides can identify key spectral regions, with the possibility to act as structural markers for future studies on more complex carbohydrates. Comparison of panel D in figures 5 and 6 shows bands that are consistent for the two monosaccharides. The most striking of these is the broad positive feature, centred around 1410 cm^{-1} for GlcA and GlcNAc, with the peak assignments revealing complex diffuse vibrations involving various C-C ring stretches and C-H wags. The peak is much broader in the GlcA spectra, due to the contribution of the strong CO_2 symmetric stretch in the same region. A similar feature is observed in both experimental spectra, albeit appearing at a slightly lower frequency in both cases.

There are also calculated peaks appearing in both OPTALL spectra for O-H wags, at 1000 cm^{-1} and 960 cm^{-1}, for GlcA and GlcNAc, respectively. The slight shift in frequency is likely related to the difference in structure around C2 and C5, and both peaks appear in the experimental spectra. However, GlcNAc exhibits a much higher intensity peak than GlcA, as discussed above. There are weak positive peaks at 1240 cm^{-1} and 1245 cm^{-1} respectively for GlcA and GlcNAc, which arise from C-H wagging motions across the entire ring. The strong negative peak at 1335 cm^{-1} for GlcA, assigned to coupled C-H wags across the entire ring, also appears for GlcNAc but at a lower frequency of 1320 cm^{-1}. In this case, however, the sign has changed with the latter being of positive intensity. For both molecules these peaks are associated with C-H wags but for GlcNAc there is also a contribution from an asymmetric C2-N1-C7 stretch, apparently causing the switch of sign.

As discussed above there are many similarities when comparing the spectral profiles and the vibrational motions of GlcA and GlcNAc. However, there are also important differences indicating the sensitivity of ROA to carbohydrate structural details. Even though GlcA and GlcNAc have very similar backbone structures, generating the strong similarities between their ROA spectra, there are still regions with markedly different profiles.

Conclusions

Our understanding of the structural constraints responsible for the diverse functionality of complex carbohydrates lags behind that for proteins and nucleic acids, because of the difficulties in obtaining detailed structural information on sugars. By combining ROA's inherent sensitivity to stereochemistry with the diagnostic capabilities of QM/MM modelling and of MD simulations for investigating the crucial contributions of hydration interactions on conformational dynamics, we successfully probed the conformational behavior of the two monosaccharide components of hyaluronan. GlcA and GlcNAc are also important constituents of other glycosaminoglycans and glycans and the approaches presented here provide the means to investigate the dynamic structure of carbohydrates in detail.

Acknowledgements

We are grateful to Dr Steve Liem and Dr Chin Yong for their help with the simulations. EWB and PLAP would like to thank the UK Engineering and Physical Sciences Research Council for funding (EP/J019623/1).

Notes and references

[a] Manchester Institute of Biotechnology and Faculty of Life Sciences, University of Manchester, 131 Princess Street, Manchester, M1 7DN, U.K..

[b] Manchester Institute of Biotechnology and School of Chemistry, University of Manchester, 131 Princess Street, Manchester, M1 7DN, U.K..

[c] Gaussian, Inc, 340 Quinnipiac Street, Building 40, Wallingford, Connecticut 06492, United States

[d] Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

† (ESI) available: Calculated ROA spectra for the 4 individual GlcNAc conformers using the OPTALL and OPTGlcNAc approaches. See DOI: 10.1039/b000000x/

1 L. D. Barron and A. D. Buckingham, *Mol. Phys.*, 1971, **20**, 1111.
2 L. D. Barron, M. P. Bogaard and A. D. Buckingham, *J. Am. Chem. Soc.*, 1973, **95**, 603.
3 L. D. Barron, L. Hecht, I. H. McColl and E. W. Blanch, *Mol. Phys.*, 2004, **102**, 731.
4 J. Haesler, I. Schindelholz, E. Riguet, C. G. Bochet and W. Hug, *Nature*, 2007, **446**, 526.
5 Y. He, B. Wang, R. K. Dukor, and L. A. Nafie, *Appl. Spectrosc.*, 2011, **65**,699.
6 S. Yamamoto, *Anal. Bioanal. Chem.*, 2012, **403**, 2203.
7 F. Zhu, G. E. Tranter, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Mol. Biol*., 2006, **363**, 19.
8 L. D. Barron, L. Hecht, E. W. Blanch and A. F. Bell, *Prog. Biophys. Mol. Biol*., 2000, **73**, 1.
9 N. R. Yaffe, A. Almond and E. W. Blanch, *J. Am. Chem. Soc.*, 2010, **132**, 10654.
10 M. Pecul and K. Ruud, *Int. J. Quant. Chem.*, 2005, **104**, 816.
11 M. Pecul, *Chirality*, 2009, **21**, E98-E104.
12 K. Ruud and A. J. Thorvaldsen, *Chirality*, 2009, **21**, E54-E67.
13 L. D. Barron and A. D. Buckingham, *Chem. Phys, Lett.*, 2010, **492**, 199.
14 P. Mukhopadhyay, G. Zuber and D. N. Beratan, *Biophys. J*., 2008, **95**, 5574.
15 J. Kapitan, F. Zhu, L. Hecht, J. Gardiner, D. Seebach and L. D. Barron, *Angew. Chem. Int. Ed.* 2008, **47**, 6392.
16 C. R. Jacob, S. Luber and M. Reiher, *Chem.—Eur. J*., 2009, **15**, 13491.
17 M. Kaminski, A. Kudelski and M. Pecul, *J. Phys. Chem. B*, 2012, **116**, 4976.
18 V. Parchansky, J. Kapitan, J. Kaminsky, J. Sebestik, P. Bour, *J. Phys. Chem. Lett.*, 2013, **4**, 2763.
19 H. Urago, T. Suga, T. Hirata, H. Kodama, M. Unno, *J. Phys. Chem. B*, 2014, **118**, 6767.
20 S. Yamamoto, T. Furukawa, P. Bour, and Y. Ozaki, *J. Phys. Chem. A*, 2014, **118**, 3655.
21 N. A. Macleod, C. Johannessen, L. Hecht, L. D. Barron and J. P. Simons, *Int. J. Mass Spectrom.*, 2006, **253**, 193.
22 S. Luber and M. Reiher, *J. Phys. Chem. A*, 2009, **113**, 8268.
23 J. Kaminsky, J. Kapitan, V. Baumruk, L. Bednarova and P. Bour, *J. Phys. Chem. A*, 2009, **113**, 3594.
24 J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier and E. W. Blanch, *J. Am. Chem. Soc.*, 2011, **133**, 4991.
25 J. E. Scott, F. Heatley and W. E. Hull, *Biochem. J*., 1984, **220**, 197.
26 J. Kaufmann, K. Mohle, J.-G. Hofman and K. Arnold, *J. Mol. Struct.*, 1998, **422**, 109.
27 J. E. Scott and F. Heatley, *Biomacromolecules*, 2002, **3**, 547.
28 C. D. Blundell, P. L. DeAngelis and A. Almond, *Biochem. J*., 2006, **396**, 487.
29 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999.

30 M. Pecul, C. Deillon, A. J. Thorvaldsen and K. Ruud, *J Raman Spectrosc.*, 2010, **41**, 1200.

31 J. Sebek, J. Kapitan, J. Sebestik, V. Baumruk and P. Bour, *J. Chem. Phys. A*, 2009, **113**, 7760.

32 J. Gonzalez-Outeirino, K. N. Kirschner, S. Thobhani and R. J. Woods, *Can. J. Chem.*, 2006, **84**, 569.

33 C. S. Pereira, D. Kony, R. Baron, M. Muller, W. F. van Gunsteren and P. H. Hunenberger, *Biophys. J.*, 2006, **90**, 4337.

34 K. N. Kirschner and R. J. Woods, *Proc. Natl. Acad. Sci. USA* 2001, **98**, 10541.

35 L. D. Barron, A. R. Gargaro and Z. Q. Wen, *Carbohydr. Res.*, 1991, **20**, 39.

36 A. F. Bell, L. D. Barron and L. Hecht, *Carbohydr. Res.*, 1994, **257**, 11.

37 Gaussian 09, Revision B.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, 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. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

38 MOPAC2009, J. J.P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net 2008.

39 I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove, *J. Mater. Chem.*, 2006, **16**, 1911.

40 S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma and M. J. Frisch, *J. Mol. Struct. (Theochem)*, 1999, **462**, 1.

41 T. Vreven, K. S. Byun, I. Komaromi, S. Dapprich, J. A. Montgomery Jr, K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815.

42 F. London, *J. Phys. Radium*, 1937, **8**, 397.

43 R. Ditchfield, *Mol. Phys.*, 1974, **27**, 789.

44 J. R. Cheeseman and M. J. Frisch, *J. Chem. Theory Comput.*, 2011, **7**, 3323.

45 G. Zuber and W. Hug, *J. Phys. Chem. A*, 2004, **108**, 2108.

46 C. W. Yong, 'DL_FIELD - A force field and model development tool for DL_POLY', R Blake Ed., CSE Frontier, STFC Computational Science and Engineering, Daresbury Laboratory, p38-40, 2010.

47 I. J. Bush, I. T. Todorov and W. Smith, *Comput. Phys. Commun.*, 2006, **175**, 323.

48 L. D. Barron, F. Zhu, L. Hecht, G. E. Tranter and N. W. Isaacs, *J. Mol. Struct.*, 2007, **834**, 7.