PCCP

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Simulation of Raman optical activity

of multi-component monosaccharide samples

Adéla Melcrová,[a, b] Jiří Kessler,[a] Petr Bouř,[a,*] and Jakub Kaminský[a,*]

[a] *Institute of Organic Chemistry and Biochemistry, AS CR, Flemingovo nám. 2, 166 10 Prague, Czech Republic, email: kaminsky@uochb.cas.cz, bour@uochb.cas.cz, Fax: +420-220183-579*

[b] *J. Heyrovský Institute of Physical Chemistry, AS CR, Dolejškova 2155, 182 23 Prague, Czech Republic*

Abstract

Determination of saccharide structure in solution is a laborious process that can be significantly enhanced by optical spectroscopies. Raman optical activity (ROA) spectra are particularly sensitive to the chirality and conformation. However, the interpretation of them is largely dependent on computational tools providing a limited precision only. To understand the limitations and the link between spectral shapes and the structure, in the present study we measured and interpreted using a combination of molecular dynamics (MD) and density functional theory (DFT) Raman and ROA spectra of glucose and mannose solutions. Factors important for analyses of mixtures of conformers, anomers, and different monosaccharides are discussed as well. The accuracy of the simulations was found to be strongly dependent on the quality of the hydration model; the dielectric continuum solvent model provided lower accuracy than averaging of many solvent-solute clusters. This was due to different conformer weighting rather than direct involvement of water molecules in scattering recorded as ROA. However, the cluster-based simulations also failed to correctly reproduce the ratios of principal monosaccharide forms. The best results were obtained by a combined MD/DFT simulation, with the ratio of α- and β- anomers and the $-CH_2OH$ group rotamers determined experimentally by NMR. Then a decomposition of experimental spectra into calculated subspectra provided realistic results even for the glucose and mannose mixtures. Raman spectra decomposition provided a better overall accuracy ($\sim$5%) than ROA ($\sim$10%). The combination of vibrational spectroscopy with theoretical simulations represents a powerful tool for analysing saccharide structure. Conversely, the ROA and Raman data can be used to verify the quality of MD force fields and other parameters of computational modeling.

Introduction

Raman optical activity (ROA), a small difference in scattering of right- and left-circularly polarized light, is a useful probe of structure, applicable to a wide range of chiral molecules. Since its discovery[1] it has been used for small molecules, proteins, nucleic acids and whole viruses.[2-4] The spectra are often complex and their interpretation not trivial at all. Fortunately, a steady development of theoretical chemistry in the last decades provided useful interpretation tools and significantly stimulated the experimental field. The possibility to simulate Raman and ROA intensities by available quantum chemical codes[5, 6] helps to shed light on the behavior of a wide range of molecular systems. [7-16]

Carbohydrates are known to support cellular structures (e.g. bacterial or plant cell walls) and participate in numerous events in living organisms. Long-chain glycosaminoglycans fill the extracellular space, influence angiogenesis, immunosuppression, cell differentiation, ovulation, embryogenesis, and regeneration. Small sugars act as anti-inflammatory and angiogenic agents, and support immune and signaling systems.[17-19] However, current knowledge about their three-dimensional structure and its relation to the function is far from complete ad standard methods for structural characterization of biomolecules often fails for carbohydrates.[20] Some saccharides are hard to crystallize or provide poor NMR spectra because of peak overlaps and broadening due to fast relaxation. Sugars also typically lack active chromophores needed for absorption or electronic circular dichroism (ECD) measurements in the usual ultraviolet and visible regions (~190-800 nm). Such chromophores may be introduced synthetically, but this may perturb the saccharide geometry. Saccharides are typically studied in an aqueous environment causing many troubles in infrared techniques. In this context, ROA spectroscopy appears to be a very convenient alternative, as yields relevant data on carbohydrate structure and interactions.[21-33] For example, early works on small saccharides[21-25, 34] provided an insight into the anomeric configuration, relative configurations of the

3

hydroxyl, ring puckering and the glycosidic linkage. In polysaccharides and glycoproteins, ROA was used to track the secondary and tertiary structure.[20, 27, 32, 35-38]

Simulations of saccharide ROA spectra are, however, rather complicated due to molecular flexibility and strong polar interactions with the solvent.[29, 39, 40] Realistic description of carbohydrate-water interactions determining the conformation of the ring and hydroxyl groups is needed for the results to be useful.[29-31, 39, 41] Chiroptical methods are generally very sensitive to the solvent-solute interactions, and comparable solvent-induced features have also been observed in ECD, vibrational circular dichroism or optical rotation of several systems.[42-44 45, 46] In our opinion, further quantitative estimates of the solvent and conformational factors are thus needed to make the spectroscopic studies of saccharides more reliable and to extract maximum information from the spectra. We show, for example, that a relatively small number of conformers is sufficient to reproduce the spectra, and that water-sugar interactions are important for conformer weighting (less so for water spectra contributions).

To understand spectra of complex polysaccharides and their conjugates with proteins, individual parts have to be carefully controlled. We thus consider increasing the number of components and investigated samples (e.g., α and β anomers, as well as glucose/mannose mixture) as the first step in this direction. Anomers are stereoisomers often occuring in cyclic saccharides, differing only in the configuration of the hemiacetal (anomeric) carbon. Individual anomers of reducing sugars transform to their inverse form via the process of anomerization (mutarotation), and in solution an equilibrium of the two anomeric forms is usually established quite quickly. However, the anomer ratio ("α" / "β") is specific for each sugar. As the two anomers differ in reducing configuration at one chiral center only, their ROA spectra are not mere "mirror images"; the anomerization ($\alpha \leftrightarrow \beta$) significantly changes the spectral pattern.

Another objective is to estimate as to how the ROA spectra can provide feedback on the quality of molecular dynamics (MD) force fields, accuracy of density functional theory (DFT), and

solvent models. For example, available computational tools do not appear to provide completely satisfactory estimates of the anomeric and –CH₂OH conformational species; experimental ratios taken from previous NMR studies led to better results. As proven by earlier studies[30, 41, 47] the combination of quantum and molecular dynamics is used as the best tool to tackle the flexible and hydrated sugar molecules. The sensitivity of ROA to monosaccharide structure was tested for arbitrary monosaccharide mixtures decomposed into both calculated and experimental subspectra.

Methods

Raman and ROA Spectra Measurement. D and L glucose and mannose (Fig. 1) were purchased from Carbosynth and dissolved in milli-Q water to concentrations of 0.6 mg/100 μL. The samples were left undisturbed for several hours to equilibrate various sugar forms; required equilibration times were estimated by NMR. Backscattered Raman and scattered circular polarization (SCP) ROA spectra were recorded on a ChiralRAMAN-2X (Biotools Inc.) spectrometer equipped with Opus diode-pumped solid-state laser emitting 532 nm green light. Laser power at the head was about 350 mW (approximately 2/3 reached the sample), total acquisition times were ~ 20 h for each sample contained in a fused silica cell (3 mm optical path, 60 μL sample volume). Fluorescence was reduced by removing sample impurities by active carbon or quenched by leaving the sample in the laser beam for an hour before measurement. Raman spectra are presented with solvent signal and polynomial baseline subtracted. Luminescence standard material (SRM 2243) was used for ROA and Raman intensity calibration.

Simulations of the Spectra. Programs Gaussian 09,[48] MacroModel,[49] and Amber12[50] were used for the quantum and molecular dynamics simulations. Only the pyranose forms of glucose and mannose were considered as they are prevalent in solutions (> 99%, see Refs. [51] and [52]). We used two models of sugar geometries, a systematic conformer search coupled with a DFT calculation, and averaging of MD solvent-solute clusters.

The systematic conformer search was done within MacroModel using a hybrid search method, employing 10 000 cycles of simulated annealing followed by 10 000 cycles of large-scale low mode search steps. The MMFF94s force field, temperature of 1 000 K, 0.5 ps annealing time, 3 fs time step, the GB/SA water solvation model, 100 kcal/mol energy window and 0.5 Å cutoff for eliminating redundant conformers were used. Further computations were done in Gaussian. Produced geometries (348 for α-**DG**, 350 for β-**DG**, 423 for α-**DM**, and 396 for β-**DM**) were optimized at the HF/6-31G**/CPCM[53] level, and the lowest-energy conformers (relative energies

6

below 5 kcal/mol) were re-optimized at the B3LYP/6-311++G**/CPCM approximation level with the D2 dispersion correction.[54] Raman and ROA spectra were generated at the same B3LYP/6-311++G**/CPCM level for conformers with populations higher than 1%. As recommended previously,[55, 56] Boltzmann populations were also calculated from the CCSD(T)/6-311++G**/CPCM single-point energies; however, the CCSD(T) populations were nearly the same, and thus only B3LYP populations are presented and were used for spectra generations.

The MD cluster averaging was based on the separate MD simulations of α and β anomers of **DG** and **DM**. One molecule of sugar was placed into a cubic box (23 Å a side) and surrounded by 400 water molecules. For the initial geometry only the ${}^4C_1$ chair conformation was considered as other pyranose ring conformations are too high in energy.[57-59] The GLYCAM06[60] (sugars) and TIP3P[61] (water) force fields were used. A short heating of the system from 0 to 300 K for 20 ps under NVT conditions and 100 ps NpT equilibration preceded the production run, performed for 10 ns under the NpT conditions at the temperature of 300 K, pressure of 1 atm, and 1 fs integration step. Snapshot geometries were saved during the production run each 1 ps and the parallel variable method[62] was used to reduce the number of the snapshots to 70, which were then used to generate the spectra.

In the snapshots, only *s*olvent molecules in the first solvation sphere were retained. This was achieved with a 2 Å cutoff distance between the any solute and solvent atoms. The solute-solvent clusters were generated using our ("XSHELL") software.[63] Geometries of the resultant clusters comprised 5-11 water molecules and were partially optimized using the normal mode vibrational coordinates with the normal mode frequency limit[64] ω_{max} of 300 cm^{-1}. The normal mode optimizations was found previously to be excellent tool for optimization of MD snapshots.[65, 66, 67] It has been in particular developed[64, 68, 69] for computation of vibrational spectra, as it allows to relax the higher-frequency modes of interest, whereas the MD geometry dispersion reflecting non-zero temperature is largely conserved.

The Raman and ROA intensities were then calculated at the B3LYP/6-311++G**/CPCM level (same as for the systematic conformer search). Polarizability derivatives of water atoms were deleted before the spectral simulations, which significantly minimizes computational ROA artifacts resulting from incomplete averaging of solvent orientation.[65] (Note also that Raman experimental spectra are presented with the water background subtracted.) Backscattered Raman and SCP ROA intensities were convoluted with Lorentzian peaks of full width of half-maximum of 20 cm^{-1}, and taking into account the Boltzmann temperature factor at 300 K. As a rough indicator of the cluster averaging, a relative error $\delta(N)$ was estimated as:

$$\delta = \int_{\omega_{min}}^{\omega_{max}} |S_N(\omega) - S_M(\omega)| d\omega \Big/ \int_{\omega_{min}}^{\omega_{max}} |S_M(\omega)| d\omega, \tag{1}$$

, where S_N is a spectrum from N averaged clusters $(1 \leq N \leq 70)$, M = 70, ω_{min} = 200 cm^{-1}, and ω_{max} = 2000 cm^{-1}. Resultant spectra were obtained by equal weighting of all clusters, as well as selective weighting of –CH$_2$OH rotamers according to NMR data. For the latter method, the clusters were divided into three groups according to their –CH$_2$OH orientation, the subspectra generated with equal cluster weighting for each group separately, and resultant spectra of rotamers were mixed acoording to experimental NMR ratios taken from Ref. [70] [71].

For further tests selected experimental spectra $S(\omega)$ were decomposed into calculated or experimental subspectra $s_i(\omega)$ as:

$$S(\omega) = \sum_{i=1}^{n} c_i s_i(\omega), \tag{2}$$

where the coefficients were determined by minimization of $\int_{\omega_1}^{\omega_2} (S(\omega) - \sum_{i=1}^{n} c_i s_i(\omega))^2 d\omega + \alpha \sum_{i=1}^{n} (c_i - 1/n)^2 \to \min$, with ω_1 = 100 cm^{-1} and ω_2 = 1600 cm^{-1}. The parameter α = 0.00001 was introduced to prevent negative coefficient; the coefficients were normalized after the decomposition to one ($\sum_{i=1}^{n} c_i = 1$). The n = 2, because the experimental spectra

were decomposed[16] into calculated spectra of free **DG** and **DM** simulated as described above. Similarly, the experimental spectra of mixtures were decomposed to experimental spectra of individual components in order to establish the accuracy of the decomposition method.

In the reviewing process, it was suggested to compare our simulations to the "QM/MM" approach pursued in references.[30, 41, 47] For this purpose, the α-**DG** spectra were simulated with various models, as for this anomer experimental data are available in Ref. [72]. As a simplest solvent model, the explicit water molecules in our clusters were deleted and the solute geometries fully optimized at the B3LYP/6-311++G**/CPCM level. To follow the "QM/MM" approach form Refs. [30, 41, 47], the Amber TIP3P[61] water force field was used for the solvent molecules in all 70 clusters, the clusters were fully optimized using the ONIOM method[73-76] and the B3LYP/6-311++G** level for the solute, and Raman and ROA obtained at the same QM/MM level. All spectra were compared with those obtained by our default QM/QM procedure.

Results

Experimental Spectra. Experimental Raman and ROA spectra are presented in Fig. 2. They are consistent with previous data;[21, 24] our results additionally include the region below 600 cm[-1]. The main features of mannose also resemble those in α-D-mannoside.[24] The ROA baseline below 200 cm[-1] is affected by instrumental artifacts, the region of 200-600 cm[-1], however, contains distinct ROA features in intensity comparable with the higher-frequency bands.

As discussed before,[24] monosaccharide vibrational spectrum can be approximately divided into a low wavenumber (< 600 cm[-1]), anomeric (600-950 cm[-1]), fingerprint (950-1200 cm[-1]) and CH_2 and -COH deformation (> 1200 cm[-1]) regions. This can be complemented by calculated potential energy distributions (PED), some of them presented in Fig. S1 (Electronic Supplementary Information). For example, the OH bending motion is relatively restricted to the vicinity of 1200 cm[-1], whereas most other coordinates contribute throughout the vibrational spectrum. Below 600 cm[-1] torsional-deformational modes of the six-member ring prevail.

In the low wavenumber range (< 600 cm[-1]), Raman and ROA bands sensitive to the anomeric forms often comprise exo- and endocyclic torsion and bending deformations coupled with exo- and endocyclic deformations. The "true" anomeric region (600-950 cm[-1]) typically comprises C-C and C-O stretching, C-O-C bending, and CH_2 rocking. The experimental bands around 850 cm[-1] were found to be characteristic for the α anomeric hydroxyl.[22-25, 34] The 848 cm[-1] **DG** Raman band appears at 880 cm[-1] for **DM** (Fig. 2). ROA spectra exhibit even greater sensitivity to the anomeric configuration in this region.[23-25, 34] β-anomer sugars have very small ROA signal between ~700-800 cm[-1], while there are several strong signals for the α-anomers.[23-25, 34] Also the ROA signal of mannose (predominantly α) is stronger than for glucose in this region.

ROA intensities within 800-950 cm[-1] are weak (Fig. 2). For **DG,** the 848 and 890 cm[-1] negative bands correspond to axial and equatorial positions of the anomeric hydroxyl, respectively.

DM has a very different ROA sign pattern here. In the fingerprint region there are two intense Raman bands (1062 and 1125 cm^{-1}) of glucose and three (1068, 1106 and 1139 cm^{-1}) of mannose, and rich ROA features. As discussed before[39] the conformation of the hydroxymethyl (–CH$_2$OH) group significantly influences Raman and ROA bands intensities in this region, but *ribo-* and *xylo-* saccharides provide similar patterns. Similarly, the signal in the highest wavenumber region is quite complex, and detailed vibrational analysis can be found elsewhere.[23, 24] For example, COH bending and CH$_2$OH twisting motions participate on the ROA intensity pattern, and the ROA signal at 1260 cm^{-1} is characteristic for β-D-glucose.[25]

Experimental vs. Computed Anomeric Ratios. Although the unbiased conformational scans including systematic conformer search and molecular dynamics provide a good basis for estimation of the prevalent saccharide conformers, accurate modeling and interpretation of the spectra require a more careful consideration of the anomeric equilibrium. As previously determined by NMR (cf. Table 1, Ref. [51] and [52]) the α/β ratios in glucose and mannose significantly differ. This was explained by mutually competing anomeric and 1,3-diaxial effects.[77-79] Experimentally, the populations of the α-anomer in glucose and mannose water solutions are 37% and 67%, respectively. Our computed Boltzmann populations (46% and 56%, Table 1) based on anomer relative energies follow the experimental trend with an error similar to that observed in other works.[80, 81] However, such small inconsistencies in the anomeric ratios have a bigger effect on the spectra as documented for **DG** in Fig. 3, where Raman and ROA spectra of α- and β-forms of **DG** are presented as calculated for the prevalent $^4C_1$ main chain conformer. For example, the characteristic glucose and mannose Raman and ROA patterns around 300 cm^{-1} largely stemming from the different α- and β-content are to some extent caused by deformation vibrations in the vicinity of the anomeric carbon. Therefore, as another option, we investigate weighting of the computed spectra using the experimental α/β form ratios.

The effect of –CH₂OH rotamers. A similar situation occurs with the hydroxymethyl group

conformations (*gg*/*gt*/*tg*, see Fig. 4, top, for definition). Their calculated ratios (Table 2) differ by up

to 30% from the experimental ones. The sole effect of the *gg*/*gt*/*tg* conformer ratio on the Raman

and ROA spectra can be documented for glucose, for which experimental spectra of pure β-anomer

were previously obtained by W. Hug et al.[72] In Fig. 4, lower part, we compare spectra obtained by

direct averaging of 70 MD clusters (top) with those corrected for the experimental *gg*/*gt*/*tg* ratios

from Table 2 (middle) and experiment (bottom). Although the –CH₂OH group rotation do not

change resultant Raman and ROA spectral patterns so much as the anomerization (cf. Fig. 3),

several spectral features are sensitive to it. Incorporation of the experimental weights leads to a

better reproduction of the relative Raman intensities of the 500/520 cm^{-1} bands, and more realistic

ROA signal around 420 cm^{-1} and within 800-1400 cm^{-1}. As expected, the ROA spectra are more

sensitive to the conformational changes than the Raman scattering.

The ROA spectroscopy thus offers itself for future development and quality assessments of

the MD saccharide force fields. At the same time, force field amelioration enabling to include the

fine rotational effects associated with small energy differences appear as a complex problem,

because of the solvent-solute interactions, etc. On the other hand, the six membered ring

conformation seems to be rigid and relatively reliably determined by MD or the DFT/CPCM-based

conformational scan; e.g. $^1C_4$ chairs of **DG** were higher in energy by about 12 kJ/mol (for α

anomer) and 19 kJ/mol (for β anomer) if compared to the $^4C_1$ conformers. This is in agreement with

DFT results of Ionescu for penta-*O*-methyl-D-glucopyranose.[59] Similar energetic differences (18

and 20 kJ/mol for α-**DM** and β-**DM**, resp.) were observed for mannose.

Simulated spectra. Raman and ROA glucose and mannose spectra simulated with the

DFT/PCM and cluster/MD models, without and with the NMR conformer correction, are plotted in

Fig. 5 and 6. As expected, the Raman spectra (Fig. 5) are less influenced by the computational

methodology, and all simulations reproduced the Raman main features observed in experiment. In

comparison with the DFT/CPCM conformer averaging, the MD model better reproduces several

positions and relative Raman band intensities within the entire spectral range. The anomeric and –

CH_2OH rotational NMR correction has almost no effect on mannose Raman intensities, but causes

minor changes in the glucose spectra, such as improvement of relative intensities of the 1062/1125

cm^{-1} bands, and the shape of the 520 cm^{-1} peak (numbers relate to experiment).

For the ROA spectra (Fig. 6) the DFT/CPCM model gives a rather poor agreement with the

experiment, in particular in the low-frequency (< 600 cm^{-1}) wavenumber region. This can be

explained by direct participation of water molecules in the low-wavenumber vibrations,[30, 41, 82]

which cannot be realistically simulated with the dielectric CPCM solvent model. At the same time,

OH torsional modes (cf. Fig. S1) contribute heavily here, conformation of which is very dependent

on the interactions with water environment as well. Within 800-1800 cm^{-1}, the simulated

DFT/CPCM spectra follow the experiment more; however, although a band-to-band comparison is

rarely possible. The NMR corrections qualitatively do not improve the overall quality of the

simulation.

On the other hand, the MD calculations provide more realistic ROA patterns, almost to the

same extend as for the Raman intensities. Also, the NMR correction appears more useful here than

for the CPCM results. For example, the 344 cm^{-1} positive and 1112 cm^{-1} negative ROA **DG** bands

are obtained with it, easily assignable to their experimental counterparts. Similarly, for **DM**, the

positive 1060/1090 cm^{-1} double band or the 1320 cm^{-1} negative band could not be obtained with the

plain MD averaging. Experimentally as well as in theory, **DG** and **DM** ROA patterns are rather

similar within 200-600 cm^{-1} (a broad "+ - +" w-shape signal), whereas there are more characteristic

features above 1000 cm^{-1}.

The comparison (Fig. 7) reveals that the involvement of water molecules explicitly in the

Raman/ROA computation brings a minor effect only; the CPCM and QM/QM spectral shape are

quite similar, although minor spectral features, e.g. around 1400 cm^{-1}, may be better reproduced by

the more advanced QM/QM procedure. This is consistent with data obtained on similar systems previously.[65] However, the "explicit" water environment is necessary to properly weight the sugar conformers; indeed, relying on the continuum solvent model only (cf. spectra DFT and DFT', Figure 6) provides much inferior results.

The QM/MM methodology for α-**DG** provided very good Raman shapes, better than QM/QM; however, the latter method led to more realistic ROA profile (especially in the region below 400 cm^{-1}, as well as around 1400 cm^{-1}). This may be caused by a limited cluster size as discussed in the previous works.[30, 41, 47] Indeed, the QM/MM methodology enables in work with much larger clusters than QM/QM, containing up to several hundredths of water molecules.[47] We can thus see that the combination of quantum and molecular dynamics provides many ways of spectral simulations, and their performance should be carefully tested, because of their semi-empirical character.

MD versus CPCM geometries. The differences in the description of the sugar structures obtained by the DFT/CPCM (with DFT-based conformer distribution) and MD modeling can be seen when histograms of individual torsion angles are compared, as exemplified for **DG** in Fig. 8. (cf. Fig S3 for **DM**). This reveals, for example, the fundamental difference in the dispersion of the torsion angles, much more limited for DFT/CPCM. This is explicable by the inadequate description of the hydrogen bonds by the CPCM model, not able to describe their local and partially covalent character.[83] In addition, MD provides conformers (angle values) that are not supported by DFT/CPCM within low relative energies. There are also minor differences in equilibrium torsion angles of individual conformers, usually not exceeding 10-20°.

Glucose-mannose mixtures. Analysis of mixtures of different sugars via Raman spectroscopy is a convenient analytical method that may be used, for example, in the food industry.[84-86] [87] A decomposition of an experimental spectrum into theoretical spectra also provides an important feedback on the accuracy of the modeling, and widens the application range of ROA

spectroscopy. The present accuracy can be estimated in Fig. 9, where the **DG** content in **DG/DM** mixtures as obtained from the Raman and ROA spectra is plotted against the exact ratios. Apparently, the decomposition provides the ratios with an error of 5-20%, similar when done with the experimental and calculated sub-spectra. This result suggests that the simulation has approached a reasonable precision, and it also justifies the use of the Raman and ROA spectroscopy for determination of different sugar forms. In particular, the error in calculated spectral intensities does not seem to have a dramatic effect on the resultant precision of the whole decomposition process; the precision appears limited more by experimental error of the concentrations, Raman and ROA spectral noise, and baseline drift. At the present, Raman spectra provide better precision of the decomposition (~5-10% error in conformer ratios) than ROA (~10-20%).

How is the **DG**:**DM** ratio in a mixture reflected by spectral changes in different wavenumber regions can be seen in Fig. S4 with experimental Raman and ROA spectra of 1:0, 3:1, 1:1, 2:3, 1:3 and 0:1 mixtures. For the 1:3 and 3:1 **DG**:**DM** mixtures the experimental and calculated Raman and ROA spectra are plotted in Fig. 10. Clearly, the spectroscopy is able to discriminate between various concentration ratios, and the theory is able to capture this. For example, the intensity of the 340-520 cm^{-1} **DG** Raman bands drops almost by 25% for higher **DM** content, for which a new band appears at 668 cm^{-1}. Also the Raman bands at 800-900 cm^{-1} and 1100-1200 cm^{-1} are very useful indicators, varying in intensity by 20-50%. As expected, the ROA spectra are sensitive even more, with many bands even changing sign according to actual sugar composition.

Conclusions

We explored several computational technologies with respect to their performance to simulate Raman and ROA spectra of monosaccharides and their mixtures (α/β anomers, *gg*/*gt*/*tg* conformers and the glucose/mannose species). We found this important for better understanding of molecular behavior as well as for future applications of the spectroscopies. Simulations based on the systematic conformational scan with the polarizable solvent model (DFT/CPCM) provided notably worse results than the combined MD/DFT cluster methodology, which better reflects the strong interaction between the sugars and water molecules. The MD/DFT approach was found to be limited by the inaccuracy of MD force field, which predicted the anomeric and $-CH_2OH$ form ratios associated with tiny energy differences with a significant error. The simulations then profited from experimental NMR data, which to some extend spoiled the "a priory" parameter-less approach, but enabled to better understand the link between the spectra and the structure. The results also indicate that Raman and the ROA spectroscopy can be used as excellent tools for future improvements of saccharide force fields (although this goes beyond the scope of present work), providing feedback on tiny structural changes. The decomposition of glucose/mannose mixtures provided conformer ratios with an error of 5-20%, and the error was similar for experimental and calculated sub-spectra, indicating.

Acknowledgements

Notes and references

Electronic Supplementary Information (ESI) available: Relative potential energy distributions calculated at the B3LYP/6-311++G**/CPCM for **DG**; converg Bouř ence of the cluster averaging for α-**DG**; comparison of MD and DFT geometry distributions for α-**DG**, β-**DG**, α-**DM** and β-**DM**; and experimental spectra of **DG** and **DM** mixtures.

1.　　L. D. Barron, M. P. Bogaard and A. D. Buckingham, *J. Am. Chem. Soc.*, 1973, **95**, 603-605.
2.　　L. D. Barron, L. Hecht, I. H. McColl and E. W. Blanch, *Mol. Phys.*, 2004, **102**, 731-744.
3.　　L. D. Barron, E. W. Blanch, I. H. McColl, C. D. Syme, L. Hecht and K. Nielsen, *Spec.-Int. J.*, 2003, **17**, 101-126.
4.　　E. W. Blanch, L. Hecht and L. D. Barron, *Methods*, 2003, **29**, 196-209.
5.　　R. D. Amos, *Chem. Phys. Lett.*, 1982, 87, 23.
6.　　K. Ruud, T. Helgaker and P. Bouř, *J. Phys. Chem. A*, 2002, **106**, 7448-7455.
7.　　J. Šebestík and P. Bouř, *J. Phys. Chem. Lett.*, 2011, **2**, 498–502.

8.   J. Šebestík and P. Bouř, *Angew. Chem. Int. Ed.*, 2014, **53**, 9236 –9239.
9.   S. Yamamoto and P. Bouř, *Angew. Chem. Int. Ed.*, 2013, **51**, 11058-11061.
10.   S. Yamamoto, H. Watarai and P. Bouř, *Chem. Phys. Chem.*, 2011, **12**, 1509-1518.
11.   P. Bouř, *J. Comput. Chem.*, 2001, **22**, 426-435
12.   S. Yamamoto and P. Bouř, *J. Comput. Chem.*, 2013, **34**, 2152–2158.
13.   S. Yamamoto, J. Kaminský and P. Bouř, *Anal. Chem.*, 2012, **84**, 2440–2451.
14.   V. Parchaňský, J. Kapitán and P. Bouř, *RSC Adv.*, 2014, **4**, 57125–57136.
15.   V. Parchaňský, J. Kapitán, J. Kaminský, J. Šebestík and P. Bouř, *J. Phys. Chem. Lett.*, 2013, **4**, 2763-2768.
16.   L. Benda, P. Štěpánek, J. Kaminský and P. Bouř, in *Comprehensive Chirality*, eds. E. M. Carreira and H. Yamamoto, Elsevier, 2012, vol. 8, ch. Spectroscopic Analysis: Ab initio Calculation of Chiroptical Spectra, pp. 520-544.
17.   R. Stern, A. A. Asari and K. N. Sugahara, *Eur. J. Cell Biol.*, 2006, **85**, 699-715.
18.   T. Feizi and W. Chai, *Nature Rev. Mol. Cell Biol.*, 2004, **5**, 582-588.
19.   *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2nd. edn., 2009.
20.   N. R. Yaffe, A. Almond and E. W. Blanch, *J. Am. Chem. Soc.*, 2010, **132**, 10654-10655.
21.   L. D. Barron, A. R. Gargaro and Z. Q. Wen, *Carbohydr. Res.*, 1991, **210**, 39-49.
22.   A. F. Bell, L. Hecht and L. D. Barron, *J. Raman Spectrosc.*, 1993, **24**, 633-635.
23.   A. F. Bell, L. Hecht and L. D. Barron, *Spectrochim. Acta A*, 1995, **51**, 1367-1378.
24.   Z. Q. Wen, L. D. Barron and L. Hecht, *J. Am. Chem. Soc.*, 1993, **115**, 285-292.
25.   A. F. Bell, L. D. Barron and L. Hecht, *Carbohydr. Res.*, 1994, **257**, 11-24.
26.   A. F. Bell, L. Hecht and L. D. Barron, *Chem. Eur. J.*, 1997, **3**, 1292-1298.
27.   T. R. Rudd, R. Hussain, G. Siligardi and E. A. Yates, *Chem. Commun.*, 2010, **46**, 4124–4126.
28.   F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2005, **127**, 6142-6143.
29.   J. Kaminsky, J. Kapitan, V. Baumruk, L. Bednarova and P. Bouř, *J. Phys. Chem. A*, 2009, **113**, 3594–3601.
30.   J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier and E. W. Blanch, *J. Am. Chem. Soc.*, 2011, **133**, 4991-4997.
31.   S. Luber and M. Reiher, *J. Phys. Chem. A*, 2009, **113**, 8268-8277.
32.   C. Johannessen, R. Pendrill, G. Widmalm, L. Hecht and L. D. Barron, *Angew. Chem. Int. Ed.*, 2011, **50**, 5349-5351.
33.   L. Ashton, P. D. Pudney, E. W. Blanch and G. E. Yakubov, *Adv. Colloid Interface Sci.*, 2013, **66**, 166-200.
34.   A. F. Bell, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 1994, **116**, 5155-5161.
35.   C. Mensch, R. Pendrill, G. Widmalm and C. Johannessen, *Chem. Phys. Chem.*, 2014, **15**, 2252–2254.
36.   L. D. Barron, A. R. Gargaro, Z. Q. Wen, D. D. MacNicol and C. Butters, *Tetrahedron: Assym.*, 1990, **1**, 513-516.
37.   A. F. Bell, L. Hecht and L. D. Barron, *Chem. Eur. J.*, 1997, **3**, 1292-1298.
38.   A. F. Bell, L. Hecht and L. D. Barron, *J. Raman Spectrosc.*, 1995, **26**, 1071-1074.
39.   N. A. Macleod, C. Johannessen, L. Hecht, L. D. Barron and J. P. Simons, *Int. J. Mass Spectrom.*, 2006, **253**, 193-200.
40.   M. Pecul, A. Rizzo and J. Leszczynski, *J. Phys. Chem. A*, 2002, **106**, 11008–11016.
41.   S. T. Mutter, F. Zelinski, J. R. Cheeseman, C. Johanessen, P. L. A. Popelier and E. W. Blanch, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6016-6027.
42.   J. Hilario, J. Kubelka and T. A. Keiderling, *J. Am. Chem. Soc.*, 2003, **125**, 7562-7574.
43.   V. W. Jürgensen and K. Jalkanen, *Phys. Biol.* , 2006, **3**, S63-S79.
44.   M. Losada, H. Tran and Y. Xu, *J. Chem. Phys.*, 2008, **128**, 014508.

45.   M. Losada, P. Nguyen and Y. Xu, *J. Phys. Chem. A*, 2008, **112**, 5621-5627.

46.   M. R. Poopari, Z. Dezhahang, G. Yang and Y. Xu, *Chem. Phys. Chem.*, 2012, **13**, 2310-2321.

47.   F. Zielinski, S. T. Mutter, C. Johanessen, E. W. Blanch and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21799-21809.

48.   Gaussian 09, Revision D.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. J. A. Montgomery, 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, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian, Inc., Wallingford CT, 2009.

49.   a) F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W. C. Still, J. Comput. Chem., 1990, 11, 440. b) MacroModel, version 9.1, Schrödinger, LLC, 2005, New York, NY.

50.   AMBER 12, D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, University of California, San Francisco, 2012.

51.   S. J. Angyal, *Angew. Chem. Int. Ed.*, 1969, **8**, 157-166.

52.   S. J. Angyal, *Carbohydr. Res.*, 1994, **263**, 1-11.

53.   A. Klamt, in *The Encyclopedia of Computational Chemistry*, eds. P. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer III and P. R. Schreiner, John Wiley & Sons, Chichester, 1998, vol. 1, pp. 604-615.

54.   S. Grimme, *J. Comput. Chem.*, 2006, 27 1787-1799.

55.   G. I. Csonka and J. Kaminský, *J. Chem. Theory Comput.*, 2011, **7**, 988-997.

56.   W. M. C. Sameera and D. A. Pantazis, *J. Chem. Theory Comput.*, 2012, **8**, 2630-2645.

57.   B. J. Smith, *J. Am. Chem. Soc.*, 1997, **119**, 2699-2706.

58.   X. Biarnés, A. Ardevol, A. Planas, C. Rovira, A. Laio and M. Parrinello, *J. Am. Chem. Soc.*, 2007, **129**, 10686-10693.

59.   A. R. Ionescu, A. Berces, M. Z. Zgierski, D. M. Whitfield and T. Nukada, *J. Phys. Chem. A*, 2005, **109**, 8096-8105.

60.   K. N. Kirschner, A. B. Yongye, S. M. Tschampel, C. R. Daniels, B. L. Foley and R. J. Woods, *J. Comput. Chem.*, 2008, **29**, 622-655.

61.   W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926-935.

62.   J. Kessler, M. Dračínský and P. Bouř, *J. Comput. Chem.*, 2012, **34**, 366-371.

63.   XSHELL, P. Bouř (http://hanicka.uochb.cas.cz/~bour/programs/F/xshell.f), Praha, 2015.

64.   J. Hudecová, K. Hopmann and P. Bouř, *J. Phys. Chem. B*, 2012, 116, 336-342.

65.   K. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dračínský and P. Bouř, *J. Phys. Chem. B*, 2011, 115, 4128-4137.

66.   J. Kapitán, M. Dračínský, J. Kaminský, L. Benda and P. Bouř, *J. Phys. Chem. B*, 2010, **114**, 3574–3582.

67.   M. Dračínský and P. Bouř, *J. Chem. Theory Comput.*, 2010, **6**, 288-299.
68.   P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126-4132.
69.   P. Bouř, *Collect. Czech. Chem. Commun.*, 2005, **70**, 1315-1340.
70.   Y. Nishida, H. Ohrui and H. Meguro, *Tetrahedron Lett.*, 1984, **25**, 1575–1578.
71.   K. Bock and J. O. Duus, *J. Carbohydr. Chem.*, 1994, **13**, 513-543.
72.   W. Hug, in *Encyclopedia of Spectroscopy and Spectrometry, 2nd Edition*, ed. J. Linfon, Academic Press, 2010, pp. 2387-2405.
73.   S. Dapprich, I. Komáromi, K. S. Byun, K. Morokuma and M. J. Frisch, *J. Mol. Struct.*, 1999, **462**, 1-21.
74.   T. Vreven, K. S. Byun, I. Komáromi, S. Dapprich, J. A. Montgomery Jr., K. Morokuma and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 815-826.
75.   F. Clemente, T. Vreven and M. J. Frisch, *Quantum Biochemistry*, Wiley VCH, Weinheim, 2010.
76.   T. Vreven and K. Morokuma, *Continuum Solvation Models in Chemical Physics: From Theory to Applications*, Wiley, 2008.
77.   J. T. Edward, *Chem. Ind.*, 1955, 1102-1104.
78.   R. U. Lemieux and G. Huber, *J. Am. Chem. Soc.*, 1956, **78**, 4117-4119.
79.   R. U. Lemieux and P. Chu, presented in part at the Abstracts of Papers; 133rd National Meeting of the American Chemical Society;, Washington D.C., 1958.
80.   C. O. da Silva, B. Mennucci and T. Vreven, *J. Org. Chem.*, 2004, **69**, 8161-8164.
81.   S. E. Barrows, J. W. Storer, C. J. Cramer, A. D. French and D. G. Truhlar, *J. Comput. Chem.*, 1998, **19**, 1111-1204.
82.   J. Kapitán, V. Baumruk, V. Kopecký, Jr., R. Pohl and P. Bouř, *J. Am. Chem. Soc.*, 2006, **128**, 13451-13462.
83.   P. Bouř, D. Michalík and J. Kapitán, *J. Chem. Phys.*, 2005, **122**, 144501.
84.   B. Özbalci, I. H. Boyaci, A. Topcu, C. Kadılar and U. Tamer, *Food Chem.*, 2013, **136**, 1444-1452.
85.   M. M. Paradkar and J. Irudayaraj, *Food Chem.*, 2001, **76**, 231-239.
86.   M. M. Paradkar, S. Sakhamuri and J. Irudayaraj, *J. Food Sci.*, 2002, **67**, 2009-2015.
87.   K. Ilaslan, I. H. Boyaci and A. Topcu, *Food Control*, 2014, **48**, 56-61.
88.   M. Appell, G. Strati, J. L. Willet and F. A. Momany, *Carbohydr. Res.*, 2004, **339**, 537-551.
89.   M. Appell, J. L. Willet and F. A. Momany, *Carbohydr. Res.*, 2005, **340**, 459–468.

Table 1 Calculated and experimental populations (%) of α-anomeric forms of glucose and mannose

	Calculation		Experiment[d]
	α-sum		α-sum
DG	46[a]	44[b] (39[c])	37
DM	56[a]		67

[a] based on B3LYP/6-311++G**/CPCM Gibbs energies.

[b] Ref. [81]

[c] Ref. [80]

[d] Ref. [51, 52]

Table 2 Experimental and calculated –CH$_2$OH rotamer populations (*gg*/*gt*/*tg*, in %)

	DFT[a]	MD-GLYCAM06	DFT (Ref.[88, 89])	Exp.[70, 71]
DG	39 / 47 / 14	24 / 76 / 0	38 / 41 / 21	54 / 44 / 2
DM	41 / 24 / 35	26 / 74 / 0	59 / 3 / 38	52 / 43 / 5

[a] B3LYP/6-311++G**/CPCM Gibbs energy.
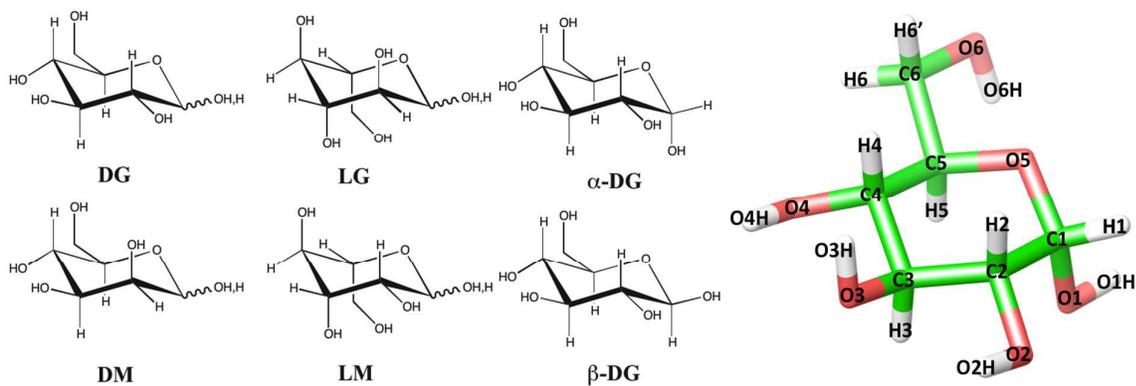
Fig. 1 Studied compounds, D- and L-glucose (**DG** and **LG**), D- and L-mannose (**DM** and **LM**),

example of the α- and β-anomers for **DG**, and used atom numbering shown for α-**DG**.
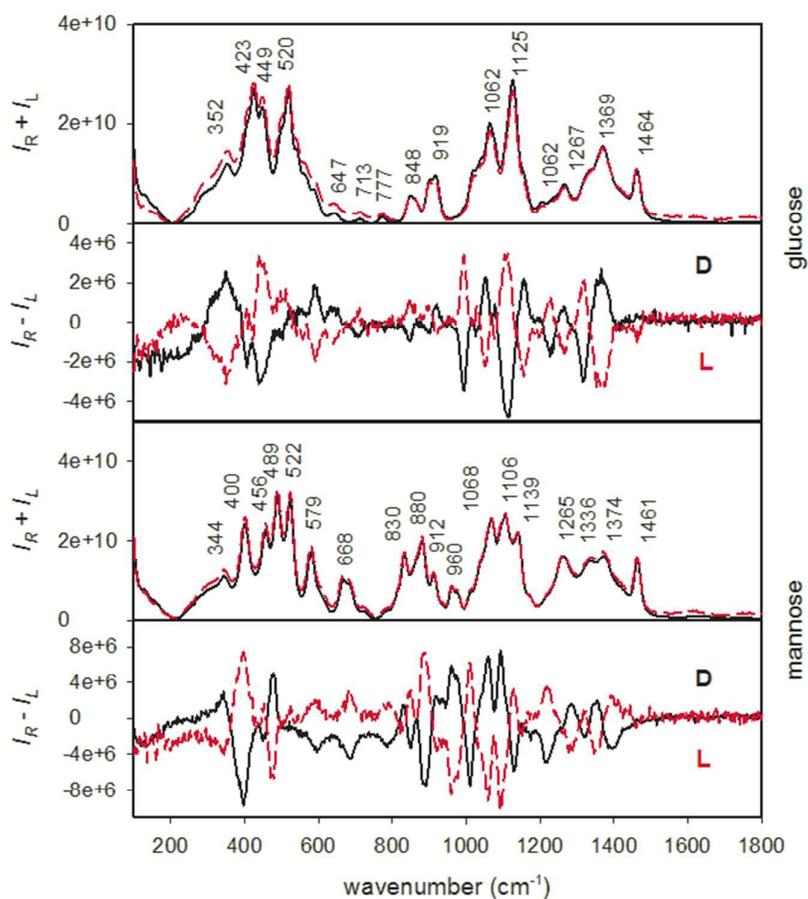
Fig. 2 Experimental Raman ($I_R + I_L$) and ROA (($I_R - I_L$)) spectra of D- and L- glucose and mannose enantiomers.
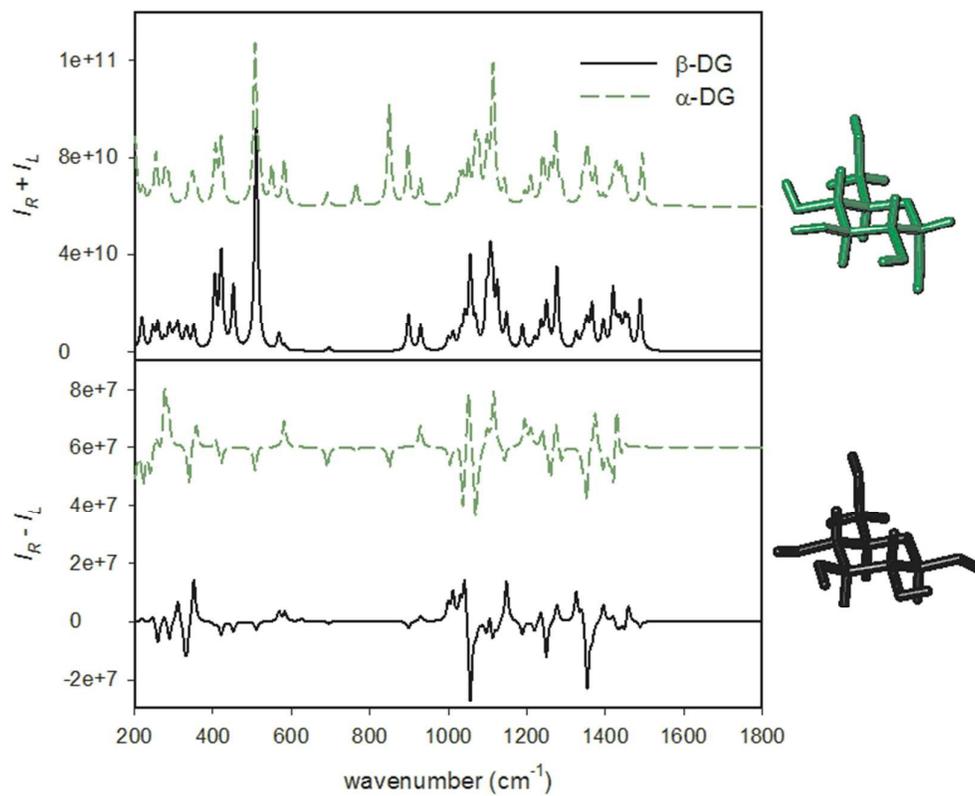
Fig. 3 Calculated (B3LYP/6-311++G**/CPCM) Raman and ROA spectra of α- and β-**DG,** for a

$^4C_1$ –CH$_2$OH conformer.
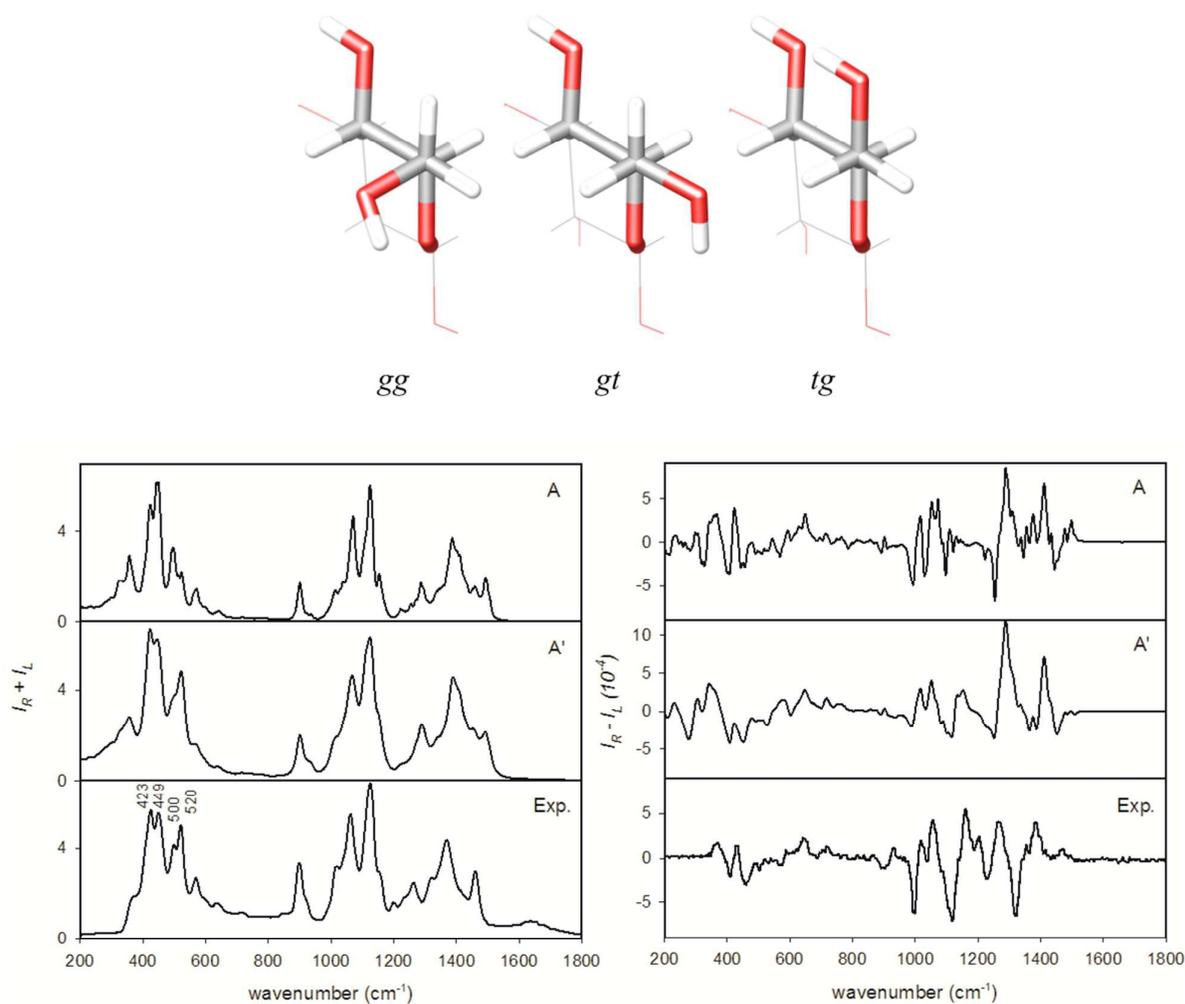
Fig. 4 Raman (left) and ROA (right) spectra of the β anomer of **DG** calculated by averaging using the equal weighting of 70 MD clusters without (A) and with (A') the *gg/gt/tg* ratio adjusted according to NMR data (**Table 2**), and experiment from Ref. [72].
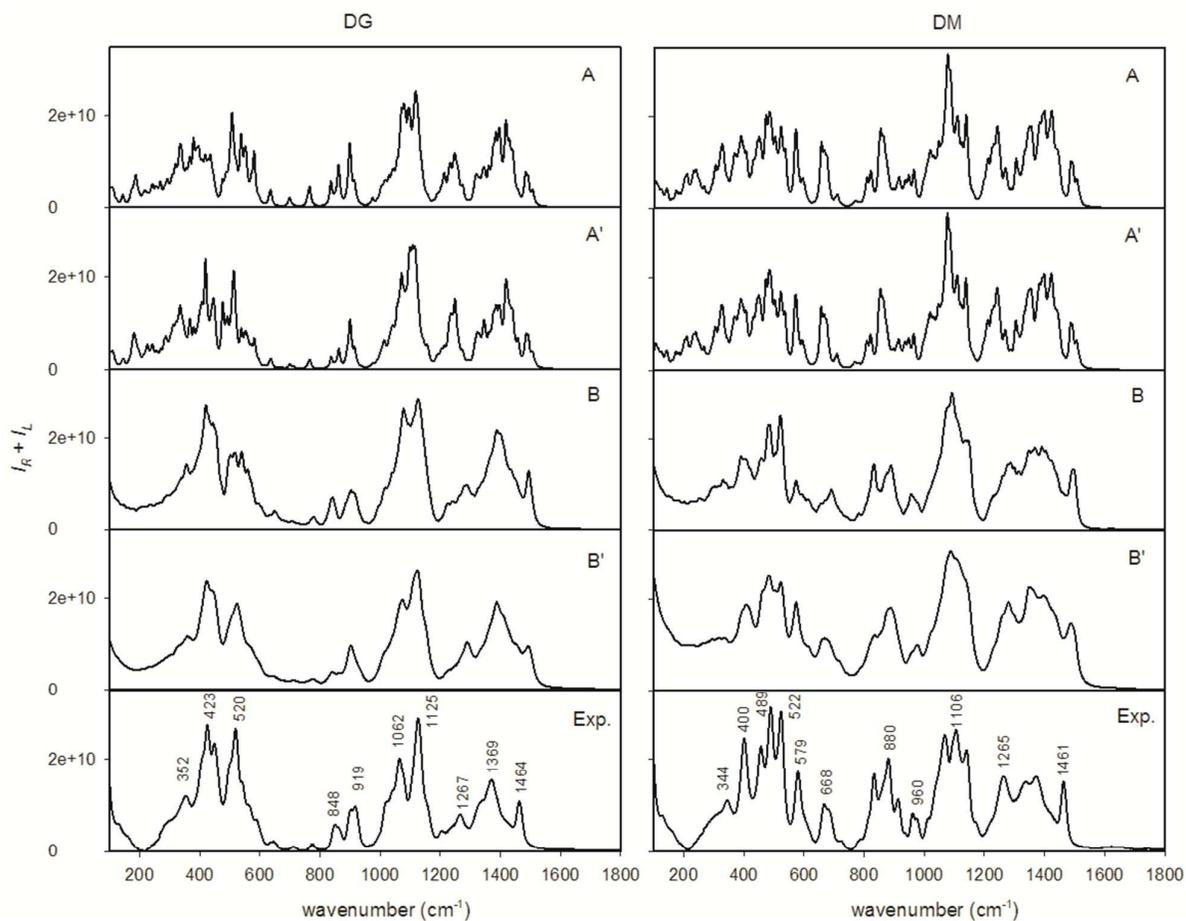
Fig. 5 Raman DG and DM spectra calculated by the plain (A) DFT/PCM and MD (B) computations, and with the α/β and gg/gt form populations adjusted according to the experiment [51, 52] (A' and B'), and the experiment.
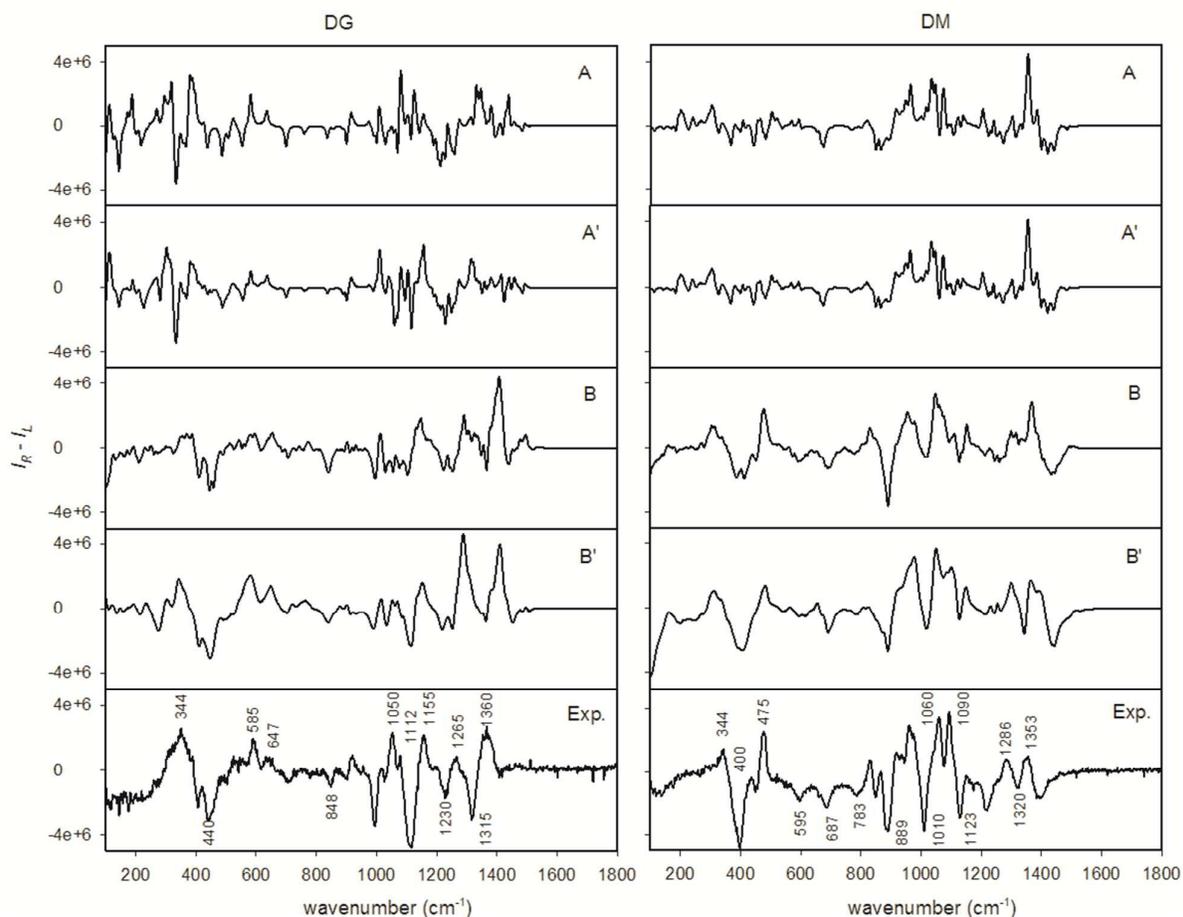
Fig. 6 ROA **DG** and **DM** spectra calculated by the plain (A) DFT/PCM and MD (B) computations, and with the α/β and *gg/gt* form populations adjusted according to the experiment[51, 52] (A' and B'), and the experiment.
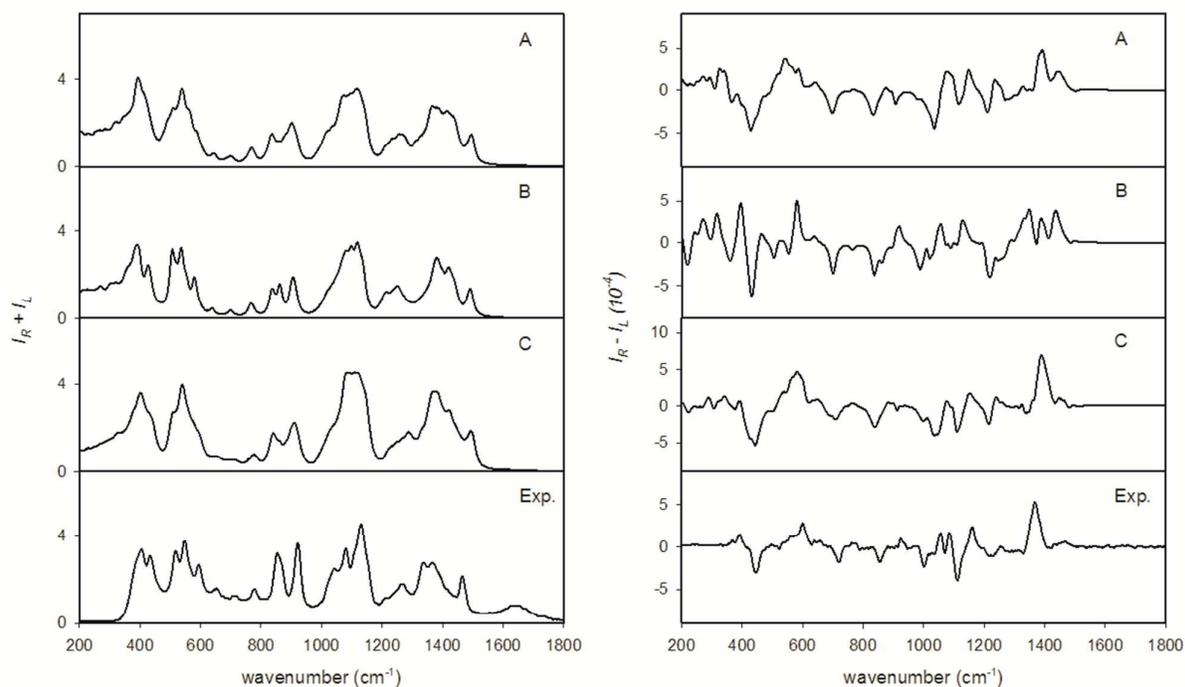
Fig. 7 Different solvent models: Raman and ROA spectra of the α-**DG** calculated by averaging of 70 MD cluster geometries using the equal weighting, with all water molecules deleted (A), using the Amber MM level for the waters (B), using the DFT level also for the waters (C), and the experiment. The *gg/gt/tg* ratio was adjusted according to NMR data (**Table 2**), experimental spectra of the α-anomer are redrawn from Ref. [72].
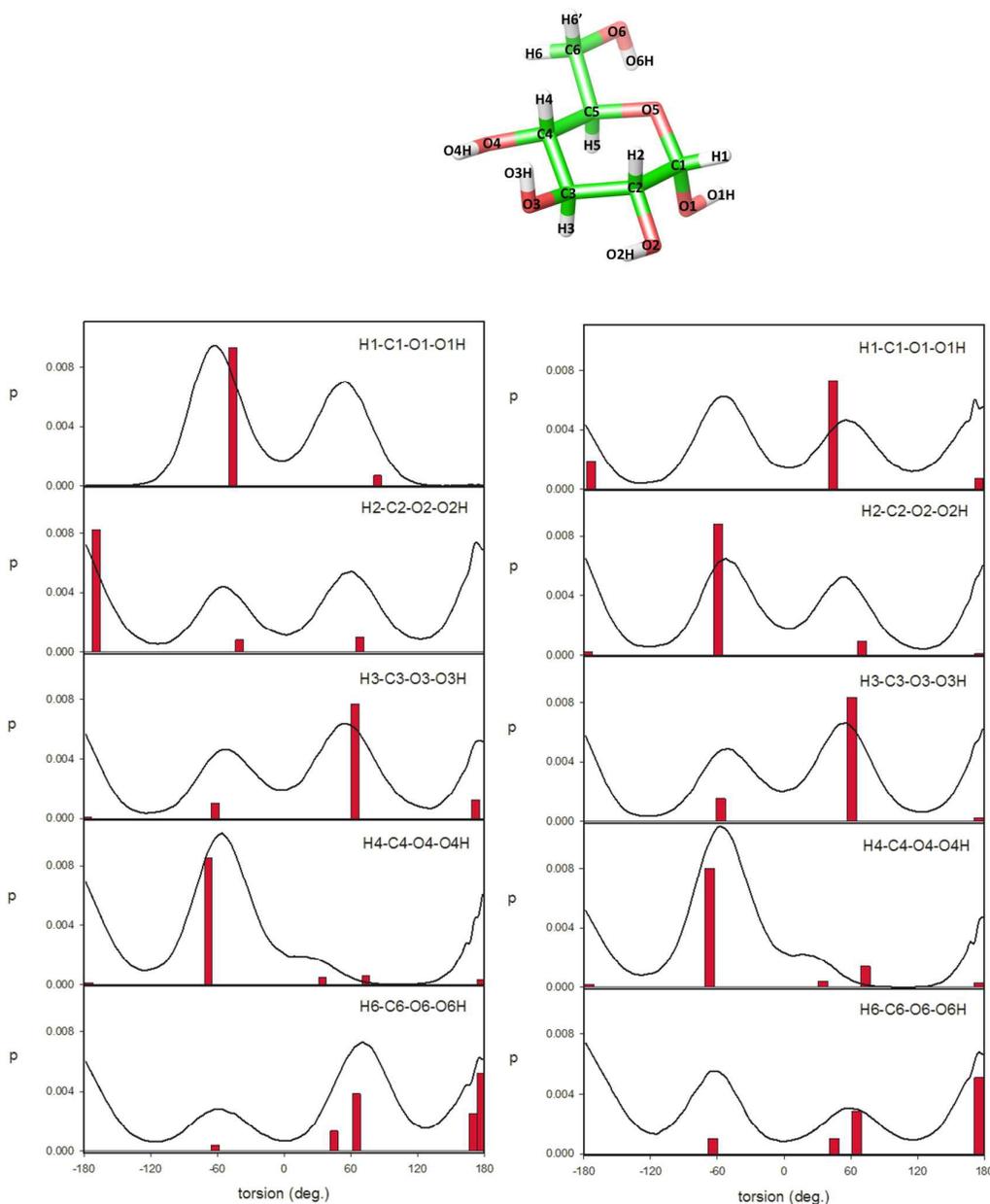
Fig. 8. Selected torsional angles as obtained by MD and DFT computations, for α-**DG** (left) and β-**DG** (right). MD probability histograms (black line) were obtained during 10 ns simulations with the GLYCAM06 force field. The DFT distribution bars (red, in arbitrary y-scale) were obtained from the B3LYP/6-311++G**/CPCM conformational scan and Boltzmann weighting.
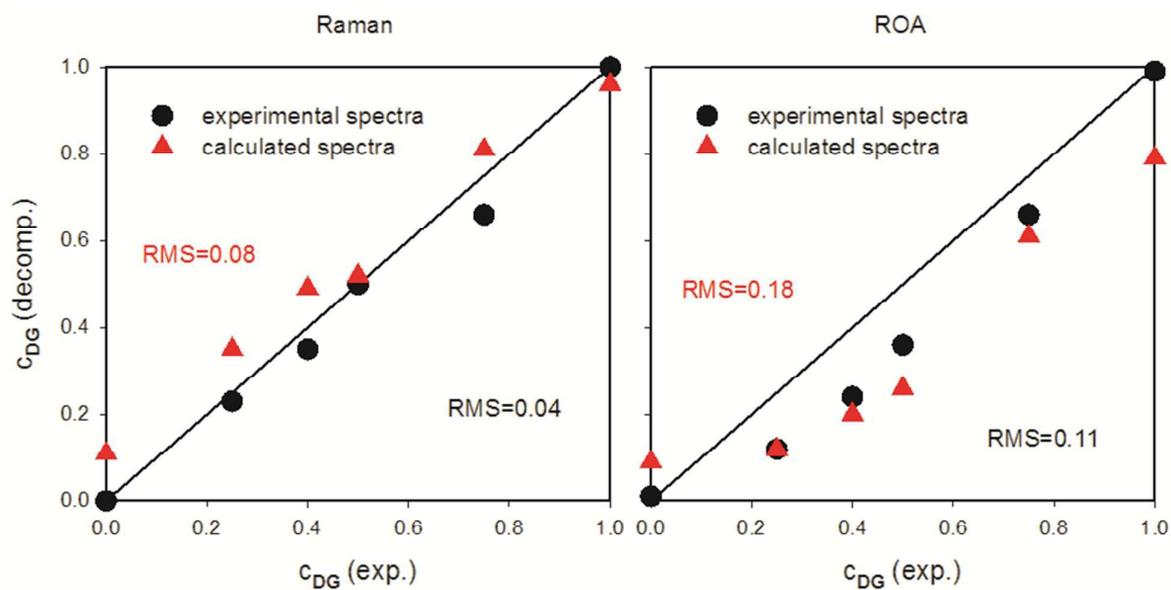
Fig. 9. DG content in **DG/DM** mixtures as obtained by decomposition of their Raman (left) and ROA (right) spectra into experimental (black circles) and calculated (red triangles) sub-spectra of individual components, see the Methods (Eq. 2) for the decomposition algorithm.
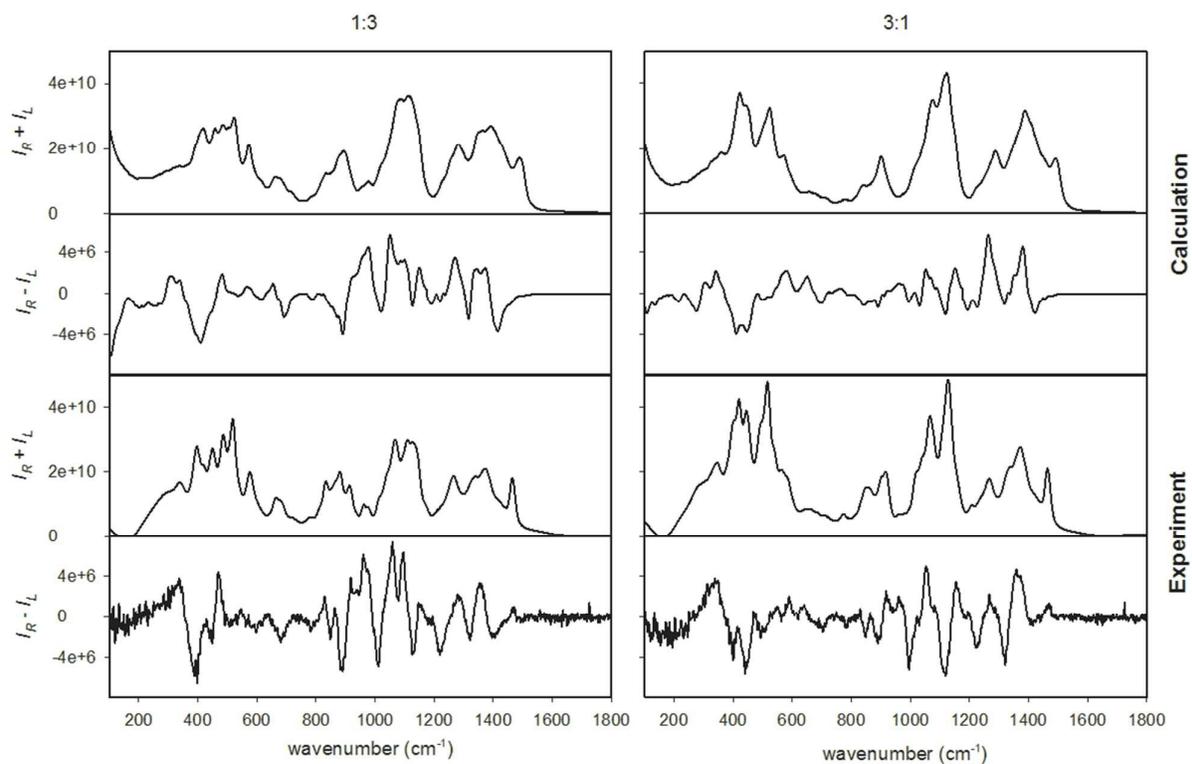
Fig. 10. Simulated and experimental spectra of 1:3 (left) and 3:1 (right) **DG/DM** mixtures. The adjusted MD spectra of **DG** and **DM** were used for the simulation.