

The effect of protein backbone hydration on the amide vibrations in Raman and Raman optical activity spectra

Carl Mensch,^{a,b} Patrick Bultinck^b and Christian Johannessen^{*a}

Raman and specifically Raman optical activity (ROA) spectroscopy are very sensitive to the solution structure and conformation of biomolecules. Because of this strong conformational sensitivity, density functional theory (DFT) calculations are often used to get a better understanding of the experimentally observed spectral patterns. While e.g. for carbohydrate structure the water molecules that surround the solute have been demonstrated to be of vital importance to get accurate modelled ROA spectra, the effect of explicit water molecules on the calculated ROA patterns of peptides and proteins is less well studied. Here, the effect of protein backbone hydration was studied using DFT calculations of HCO-(L-Ala)₅-NH₂ in specific secondary structure conformations with different treatments of the solvation. The effect of the explicit water molecules on the calculated spectra mainly arises from the formation of hydrogen bonds with the amide C=O and N-H groups. Hydrogen bonding of water with the C=O group determines the shape and position of the amide I band. The C=O bond length increases upon formation of C=O...H₂O hydrogen bonds. The effect of the explicit water molecules on the amide III vibrations arises from hydrogen bonding of the solvent with both the C=O and N-H group, but their contributions to this spectral region differs: Geometrically, the formation of a C=O...H₂O bond decreases the C-N bond length, while upon forming a N-H...H₂O hydrogen bond, the N-H bond length increases.

Introduction

Raman and Raman optical activity (ROA) spectroscopy are regularly used to study the solution structure of biomolecules.^{1,2} Raman spectroscopy probes the many molecular vibrations of a molecule and as such provides very rich spectra containing contributions of all the molecular components of the analyte. ROA is measured as the tiny intensity difference in the right-handed (I_R) and left-handed (I_L) circularly polarized components of Raman scattered light from chiral molecules.³ Because of its chiroptical nature, ROA is much more sensitive to the solution conformation of biomolecules compared to its parent technique Raman spectroscopy and has therefore been applied to study the conformational propensities of peptides,⁴⁻⁶ proteins,^{2,7} glycoproteins,^{8,9} intrinsically disordered proteins (IDPs),^{10,11} nucleic acids,¹² carbohydrates^{13,14} and viruses^{15,16}. Traditionally, experimental ROA spectra were interpreted empirically based on the mutual comparison of different spectra supplemented by structural information available by other techniques.⁷ Since the first report of the ROA spectrum of a protein in 1990,¹⁷ many more studies have been reported focussed on the spectroscopic study of protein structure in solution.² More advanced interpretation of the experimental ROA spectra of proteins is based on statistical analysis, such as principal component analysis^{18,19} or non-linear mapping²⁰ of collections of spectra.² Such analyses suggest that ROA is very sensitive to protein structure in solution. Inspired by this huge structural sensitivity, in many experimental studies tentative empirical assignments were reported, relating specific patterns in the experimental ROA spectra to structural elements of the protein in solution.⁷

Although ROA was first described in 1969 and further developed in the early 1970s,^{21,22} it is only due to more recent theoretical developments and furthermore because of the

substantial gain in computing power in the past decade, that ROA spectral patterns can now be routinely simulated using density functional theory (DFT).²³⁻²⁷

Because of these advances, a considerable number of recent research efforts adopt computational approaches in comparison with experiment to evaluate the empirically made spectral assignments and to generate a more thorough understanding of that very sensitive relation between the solution structure of a protein and the corresponding spectral components.^{1,27} Recent DFT studies of different structural elements such as the α -helix,^{6,28-30} β -sheet,³¹ β -turn,³¹⁻³³ poly-L-proline type II conformation (PPII),^{6,10,34} and larger peptides^{4,35} and proteins^{36,37}, indeed demonstrated that detailed spectral assignments in the ROA spectra of proteins must be made with caution and showed that some earlier empirically made assignments had to be revised.

One aspect in such computational studies that is much less studied is the influence of the solvent on the calculated spectral patterns. In computational studies, the interaction of the solute (the peptide) with water is often taken into account by means of an implicit solvent model. Such approaches are very useful as they do not entail a large additional computational cost, and are a good way to approximate the effect of the solvent if the solvation can be considered mainly as a minor perturbation of the gas phase solute.^{38,39} In an implicit solvent model, the solute is placed in a cavity in a polarizable continuum that represents the solvent and the necessary effects are included in the Hamiltonian.³⁹ Implicit solvent models are routinely used in theoretical simulations of vibrational spectra in solution.³⁹ However, the effect of the solvent may become so important that a continuum model no longer suffices and the solvent molecules interact so strongly that they should in fact be considered part of the solute. In such cases, explicit solvent models are used in which water molecules are explicitly placed around the solute. In this way,

direct interactions, such as hydrogen bonding between solvent and solute, are taken into account.

The amide vibrations in peptides are affected by hydrogen bonding with water molecules.⁴⁰ Therefore, here the effect of hydrogen bonding of water molecules to the peptide backbone is further explored, which has been suggested to be important in understanding the experimental ROA patterns of peptides and proteins.^{7,31,33}

The effect of hydration on calculated Raman and ROA patterns of specifically peptides and proteins has been documented in scientific literature only to a very small extent.^{30,31,41} By explicitly including water molecules, the system (solute+explicit water molecules) increases in the number of atoms, but also in the degrees of freedom. If both conformational averaging as well as averaging over multiple water configurations have to be considered, the computational demands increase drastically. Therefore, the study of the aqueous environment on the ROA spectrum has mainly been the subject of research for smaller molecules. There have been reports on the influence of hydration on the ROA spectra of molecules such as the histidine zwitterion,⁴² lactamide,^{43,44} methyl lactate,^{45,46} glycidol,⁴⁵ 2-aminopropanol,⁴⁴ methyloxirane,⁴⁶ methyl- β -D-glucose,¹³ 1,6-anhydro- β -D-glucopyranose,⁴⁷ D-glucuronic acid,⁴⁸ β -D-xylose,⁴⁹ and N-acetyl-D-glucosamine⁴⁸. For these molecules, it was demonstrated that the explicit water molecules are often needed to get a good comparison between the calculated spectrum and the experiment. For example, sugars contain many hydroxyl groups and accurate modelling requires explicit water molecules to be included.^{13,47,48,50}

Although methodologies for calculating the spectra of large peptides and proteins have been proposed,^{36,51,52} understanding the effect of water - the lubricant of life⁵³ - on the calculated spectra needs to be further explored to improve the comparison between theoretically predicted spectra and experiment. There are only a few reports of how explicit solvation of a polypeptide model affects the ROA spectrum. Mainly the explicit solvation of smaller systems such as alanine dipeptides (Ala)₂ have been reported.^{41,54} Weymuth *et al.* included two water molecules in a β -sheet model consisting of two parallel β -strands of (Ala)₈, and found only a minor effect on the calculated ROA spectrum for such a small number of water molecules.³¹ Yamamoto *et al.* looked into the detailed assignment of the amide III region of α -helical (Ala)₁₈ by including 18 dichloroacetic acid molecules in the ROA calculations using the Cartesian coordinate tensor transfer method (CTT).²⁹ To the best of our knowledge, the largest peptide model for which ROA calculations were performed that included explicit water molecules, was reported by Lubert.³⁰ She studied (Ala)₂₀ in a single α -helical conformation, and included a water molecule hydrogen bonded to each carbonyl group and two water molecules at the termini. In that study, the geometry was fully optimised, which therefore changed the conformation of the peptide away from the initial α -helical structure.³⁰ Furthermore, only a single hydration configuration (a single set of positions of the water molecules) was considered. As shown by our group, ROA is very sensitive

to the exact peptide backbone conformation, specifically in the case of α -helical structure.^{6,32} Therefore, we here study the effect of hydration on multiple peptide models with different secondary structures and multiple hydration configurations. Furthermore, ROA has in the past been proposed to be sensitive to the hydration of helices.⁵⁵ However, the spectral assignments associated with that sensitivity were later questioned and shown to be inaccurate.^{6,30,36} This thus raised the question how the hydration of α -helical structure affects the ROA patterns.

Methodology

Creation of peptide conformations

To study the effect of solvation on computed Raman and ROA spectra, we first created different conformations of HCO-(L-Ala)₅-NH₂, which subsequently were solvated. In Fig. 1, the workflow of the generation of the peptide models is illustrated. The different HCO-(L-Ala)₅-NH₂ conformations were created using the PeptideBuilder by Tien *et al.* that generates the geometries based on a set of ϕ and ψ angles (step 1 and 2 in Fig. 1).⁵⁶ In analogy to our earlier study of the secondary-structure-spectrum relationship,⁶ each conformation was characterised by a single pair of ϕ and ψ angles. In such regular conformations, all ϕ angles are thus the same and all ψ angles are the same.

Creation of hydration shells

After creating the different conformations of HCO-(L-Ala)₅-NH₂, 30 different solvent configurations (placements of the explicit water molecules around the solute) were created with the program AMBER14⁵⁷ using the TIP3P water model⁵⁸ (step 3 in Fig. 1) for each peptide conformation. The force field describing the peptide was the general amber force field (GAFF)⁵⁹. The geometry of the peptide itself was restrained in all steps. By using such a positional restraint, the peptide was not completely frozen, but the torsion angles remained very close to the originally chosen angles (see below).

As shown in step 3 in Fig. 1, the peptide was embedded in a solvent box consisting of about 1600 water molecules, followed by a stepwise equilibration procedure under (1) NVE (constant number of particles, constant volume and constant energy), (2) NVT (constant number of particles, constant volume and constant temperature) and (3) NPT (constant number of particles, constant pressure, constant temperature) conditions. After equilibration, 30 snapshots were taken with different water configurations but with the same backbone conformation. To obtain hydrated pentapeptide structures for the Raman and ROA calculations, we included all water molecules with the water oxygen atom within 3 Å of any of the atoms of the peptide (step 4 in Fig. 1). The typical distance of the water oxygen to the amide nitrogen of hydrogen bonded water molecules is 2.8 Å,⁶⁰ and as such a first layer of solvation was included. Larger distance limits, resulted in systems (explicit water + solute) that were too large to be computationally feasible.

DFT calculations: geometry optimization and Raman and ROA spectra

The geometries of the molecular systems (the solute or the solute with explicit water molecules) were partially optimised in normal coordinates by constraining the normal modes from $i300\text{ cm}^{-1}$ (imaginary) to 300 cm^{-1} (step 5 in Fig. 1).^{43,61} Using this approach by Bouř and Keiderling, the backbone conformation of the peptide and the positions of the water molecules are mostly retained, while the modes of spectroscopic interest are fully relaxed.^{43,61} The B3PW91 functional was used with the 6-31G(d,p) basis set for geometry optimization, while the spectra were calculated at the B3PW91/6-31++G(d,p) level of theory. This combination of functional and basis set with the partial optimization in normal coordinates is often used in computational studies of polypeptides and often gives very good agreement between experiment and theory.^{4-6,32,51,62} Realistic line shapes were simulated using one Lorentzian function for each normal mode with a full-width at half maximum of 20 cm^{-1} . All DFT calculations were carried out using the Gaussian16 rev. A.03 programme.⁶³ Implicit solvent corrections were included using the conductor-like polarizable continuum model (C-PCM).^{64,65} The default C-PCM parameters of water as implemented in Gaussian16 were used.⁶³ Explicit solvation calculations were performed by including water molecules to the system and treating them at the same level of theory as the solute.

To account for spectral averaging of multiple solvation configurations, the arithmetic mean spectrum of 30 Raman and ROA spectra was calculated without using weighting of the individual contributions.

Experimental spectra

To verify the computational results with experiment, the Raman and ROA spectra reported before of the AK21 (Ac-AAKAAAAKAAAAKAAK-AGY-NH₂), and XAO (Ac-X₂A₇O₂-NH₂) peptides are used here.⁶ Both peptides were dissolved to a concentration of 50 mg/mL in MilliQ water and the Raman (I_{R+L}) and ROA (I_{R-L}) were measured on a ChiralRAMAN (Biotools, Inc.) instrument^{66,67} as described in ref 6.

Similarity between experiment and calculation

A numerical comparison between a calculated Raman or ROA spectrum f and an experimental spectrum g was performed using the normalised overlap integral S_{fg} (eqn 1) that has a value between 0 % (no overlap) to 100 % (for identical Raman spectra) and between -100 % ("identical" mirror ROA spectra) and 100 % for identical ROA spectra.

$$S_{fg} = 100\% \cdot \frac{\int f(\sigma\tilde{\nu})g(\tilde{\nu})d\tilde{\nu}}{\sqrt{\int f(\sigma\tilde{\nu})^2d\tilde{\nu} \int g(\tilde{\nu})^2d\tilde{\nu}}} \quad (1)$$

Because of i.a. the harmonic approximation, the wavenumbers of the calculated spectra are typically overestimated with the level of theory used here.⁶ Therefore, the calculated spectrum

first has to be shifted in the wavenumber dimension with a scaling factor σ . Since the wavenumbers of the amide I region typically are even more overestimated in the calculated spectra, this region requires a higher wavenumber shift factor compared to the remainder of the spectrum.⁶ Therefore, first, the optimal scaling factor was determined for each calculated spectrum only for the amide I modes and a global scaling factor of 0.987 is applied to the remainder of the spectrum to obtain a good overlap between the experimental and calculated spectra (see the ESI,† for details).

Results and Discussion

Geometry optimization: backbone conformation and hydration

The backbone conformation of a peptide is affected by hydration. As depicted in Fig. 2, the hydrogen bonding of a water molecule to a carbonyl group (C=O) in the backbone, results in the tilting of the C=O outwards from the helix axis, towards the solvent. This was reported by Blundell *et al.* based on a set of crystal structures.⁶⁸ They showed that the backbone torsion angles ϕ_{i+1} and ψ_i (the two angles around a peptide bond between residue i and $i+1$) in hydrophilic or hydrophobic environment are slightly different. Sundaralingam *et al.* used a larger set of crystal structures to show that the torsion angles in hydrophilic environment have $\phi_{i+1};\psi_i$ values close to $-66^\circ;-41^\circ$ and those in hydrophobic environment close to $-59^\circ;-44^\circ$ (see Fig. 2).⁶⁹ Therefore, the average of the α -helical backbone torsion angles falls in between these two points in the Ramachandran plot (see the "general" case in Fig. 2 (c)). Specifically for α -helical structures, we have previously showed that ROA is sensitive to such a conformational differences of the backbone.⁶ In this study, however, we want to investigate what the contributions caused by the presence of the water molecules to the spectra are and not due to any conformational change.

To study the effect of backbone hydration specifically, we used the partial optimization in normal coordinates developed by Bouř and Keiderling, which was especially developed to retain peptide backbone conformations and as a suitable strategy for the geometry optimization of water clusters.^{43,61} First, we assessed that the conformations we originally selected were not altered too much because of the MD simulations and subsequent geometry optimization. For different regular backbone conformations of the peptide, we created 30 different water configurations surrounding the peptide. After the geometry optimization, the averages of the five pairs of torsion angles of each of the 30 peptide-solvent clusters were calculated. In Table 2, these mean torsion angles with the corresponding standard deviations are compared to the originally chosen set of backbone angles. The mean backbone angles are indeed very close to the values we initially selected, which demonstrates the usefulness of the geometry optimization scheme. In a recent study, we showed that such small standard deviations of the backbone torsion angles do not have a significant impact on the averaged ROA patterns.³⁴ Therefore, the hydrated systems that were created can be

used to study the impact of the explicit water molecules on the calculated spectra. In the remainder of the text, we will refer to each of the eight conformations representing five different secondary structure elements by using the initially chosen angles as listed in Table 2.

Effect of implicit and explicit solvation

Before we evaluate the effect of the explicit solvent molecules for the ensembles of 30 hydration configurations, we first consider the effect of both the implicit and explicit solvent corrections for a single hydration configuration. To that end, for HCO-(L-Ala)₅-NH₂ in an α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$), the effect of the different types of hydration on the Raman and ROA spectra is demonstrated in Fig. 3. The top spectrum (a) was calculated in the gas phase, without any solvent correction. In (b), the C-PCM implicit solvent model was included. Although the overall patterns in the Raman and ROA spectra of (a) and (b) are very similar, the amide I and amide III spectral regions display differences as relative spectral intensities and the position of the bands are shifted. E.g. the amide I band shifts down upon including the implicit solvent model. In the spectra c-f, explicit water molecules were included. To evaluate what spectral regions are influenced by the presence of the explicit water molecules, only a single hydration configuration was taken into account. As mentioned above, the hydrogen bonding of the water molecules to the amide groups affects the vibrational modes associated with this group.^{40,44} Therefore, the entire solvent configuration (f) was also compared to that when only the hydrogen-bonded water molecules are considered.

In the Raman and ROA spectra in panels (c), only the water molecules hydrogen-bonded to a carbonyl group were included and in (d) only the water molecule bonded to the terminal N-H group was kept. Finally, in (e) both the C=O and N-H hydrogen bonded water molecules were included. For each of these systems (solute+explicit water molecules), the partial geometry optimization was performed prior to calculating the Raman and ROA spectra. By not only including the C-PCM correction but also a few water molecules that are hydrogen bonded to the C=O groups (c), the biggest spectral change compared to both the Raman and ROA in (a) and (b) is observed in the amide I region (1600-1800 cm⁻¹). The presence of the water molecules results in a reduced maximal intensity in the amide I region in both the Raman and ROA. In the ROA, the -/+ couplet that is characteristic for α -helical structure is lost upon including a single configuration of explicit water molecules and becomes completely positive. Also a small impact on the shape of the amide III region (1240-1400 cm⁻¹) is observed in (c).

In this particular explicit solvation configuration, only a single N-H hydrogen bonded water molecule was observed (d). Therefore the effect on the ROA patterns by this single H₂O as shown in (d) is very small. The amide I is unchanged compared to (b) and in the amide III there is only a small difference. Including both the C=O and N-H hydrogen bonded water molecules in (e) reveals a Raman and ROA spectrum that is

very similar to the one in (d). Comparing all these spectra in a-e with the spectra calculated when all water molecules in this single hydration configuration were included (f), shows that the changes in the amide I mainly arise from the water molecules that are hydrogen bonded with the C=O groups.

In Fig. 4, the spectral differences discussed above are shown in more detail. The most distinct effect of the explicit water molecules, is the shift of the amide I region towards lower wavenumbers. While in the gas phase calculation the amide I ROA couplet is centred at 1788 cm⁻¹, by including the C-PCM correction, the position of this band is lowered to 1736 cm⁻¹. For the explicitly solvated calculations, the vibrational modes are shifted further to lower values (see both the Raman and ROA in Fig. 4). In the amide III region, both the positions of negative and positive bands as well as the relative ratio of the positive bands are affected because of the presence of the explicit water molecules. To further understand the spectral shapes, the effect of averaging over multiple hydration configurations needs to be considered (see below).

Myshakina *et al.* reported a computational study of the effect of the formation of a hydrogen bond between a water molecule and the C=O and N-H groups of N-methylacetamide.⁴⁰ They interpreted the spectral changes by looking at the specific vibrational modes that are affected by the hydrogen bond formation. Besides a conformational change (see above), the water molecule can geometrically result in a change of the amide bond lengths and charge distribution as demonstrated by the amide resonance model (Fig. 2).^{40,70} While the amide-resonance model is very simplistic,⁷⁰ it does provide an explanation of how the geometry can be affected by the presence of the water molecules. In the amide-resonance model, the planarity of the amide bond, arises from the resonance structure [O=C-NH] as shown on the left-hand side of Fig. 5 and the charged resonance structure [O=C-NH⁺], as shown on the right-hand side. Hydrogen bonding of water with the amide group stabilizes the charged structure.⁴⁰ Because of this stabilization, the C-N bond length decreases and the C=O bond becomes longer. As shown in Table S 1 in the ESI,[†] the C=O bond lengths are indeed elongated upon hydration. Concomitant with that geometrical change, the C=O becomes less stiff (less rigid), as shown in the shift of the Raman and ROA amide I bands to lower wavelengths that mainly consists of the stretching vibration of the carbonyl group.⁴⁰

The amide III vibrational modes are more complex as these comprise the coupling of C-N stretching with N-H bending and C α -H bending vibrations.^{40,71} Compared to the gas phase calculation, a few geometrical changes occur that involve these bonds by including the C-PCM correction or by adding the explicit water molecules. Compared to the gas phase geometry, the C-N bond length decreases upon including the C-PCM correction and it becomes even shorter upon including explicit water molecules as shown in table S 2 in the ESI,[†]. The N-H bond length in most cases increases (table S 3 in the ESI,[†]) compared to the gas phase, but this particular hydration configuration only contains a single water molecule hydrogen

bonded to a N-H group so this bond length will therefore be discussed in more detail below. The negative lobe of the amide III region in Fig. 4 shows a small shift to higher wavenumbers compared to the gas phase upon including the solvent.

Spectral averaging over multiple water configurations

So far, only a single hydration configuration was considered. A more realistic representation of the solvation can be obtained by averaging the calculated Raman and ROA spectra over multiple hydration configurations. Here, we created 30 different such configurations to study the average effect of the explicit solvent. As an example, in Fig. 6, 8 out of the 30 water configurations surrounding the α -helical conformation with average $\phi; \psi \approx -66^\circ; -41^\circ$, demonstrate the differences in the positions of the water molecules surrounding the peptide.

In Fig. 7, the effect of the averaging of the calculated Raman spectra over the 30 water configurations of the α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$) is demonstrated. Within each of the 4 groups of 30 individual spectra (left-hand side in Fig. 7), there is only a modest effect of the different water positions as the individual spectra are very similar. For the single hydration configuration discussed in the previous section, it was noted that the spectral effect of the layer of 3 Å of water molecules mostly arises from water molecules that are hydrogen-bonded to the solute. Since the hydrogen bonding of a water molecule to either the C=O or the N-H of the amide group can differently impact the vibrational modes,⁴⁰ we here also first evaluate their effects separately as shown in (a) and (b) in Fig. 7, respectively. In (a), the 30 individual spectra (shown on the left-hand side) are very similar although some variation can be witnessed in the region above 1600 cm⁻¹. The band around 1630 cm⁻¹ arises from the bending modes of the water molecules. Also the amide I region shows a small variation because of the hydrogen bonding of the C=O group with the solvent. If only hydrogen bonding of water with the N-H group is considered (b), the amide I region shows hardly any variation. The individual spectra in (b) do show a slight variation in the lower wavenumber portion of the amide III region. The amide III vibrations also have an important contribution of N-H in-plane bending,^{40,71} which could explain the variation in this spectral region because of hydrogen bonding of water with the N-H group. If all water molecules that form a hydrogen bond with either the C=O or the N-H group are considered (c) or when the entire 3 Å layer is considered (d), we also observed variations in both the amide I and amide III region. Because of that, the four mean Raman spectra (right-hand side) are very similar but they do show slight differences in these respective spectral regions (see also below).

The ROA patterns are overall also not drastically affected by the presence of the explicit water molecules as shown in Fig. 8. Again, in the four groups of individual ROA spectra (on the left), the amide I and III regions show small variations which lead to spectral differences in the mean spectra in both these spectral regions (right-hand side). Compared to the gas phase and C-PCM ROA spectra (see Fig. 3), the relative ratio of the

positive bands in the amide III region is sensitive to the hydration.

In Fig. 9, the mean Raman and ROA spectra are plotted on top of each other to allow a better comparison in the amide I and amide III region. The effect on the amide I region in both the Raman and ROA of the fully hydrated structure, arises almost exclusively from the formation of hydrogen bonds with the C=O groups, which affects both the position and shape of the amide I band. The lower wavenumber portion of the amide III region is sensitive to the formation of hydrogen bonds of the solvent with both the N-H and C=O groups (compared to the gas phase and only implicit solvent; see Fig. 3). However, the formation of hydrogen bonds of water with the N-H groups seems to have a larger impact compared to the effect of C=O...H₂O hydrogen bonds. Geometrically, the C=O...H₂O hydrogen bonds result in longer C=O and shorter C-N bond lengths, compared to when only N-H...H₂O hydrogen bonds are considered (see Fig. S 2 in the ESI,†). On the other hand, N-H...H₂O hydrogen bonds determine the N-H bond length. Since the amide groups in α -helical structure are involved in intramolecular hydrogen bonds, these geometric effects are also discussed below for other secondary structure elements. First, the effect of the implicit solvent is considered.

Effect of C-PCM

In computational studies that include explicit solvent molecules, the bulk solvent is often represented by a continuum model such as C-PCM. Since C-PCM was included on all explicitly solvated systems discussed above, we here address the contribution of the C-PCM on the fully hydrated system. As shown in Fig. 10, if the 30 water configurations are calculated with or without the C-PCM to represent the bulk solvent, the Raman and the ROA spectra are very similar. However, the amide I position is still affected by the C-PCM background. As shown in Fig. S 2 in the ESI,† because of the geometry optimization of the 30 water configurations in the gas phase or with C-PCM to represent the bulk solvent, the C=O bond length is shorter if the C-PCM is included.

The amide I ROA couplet characterising α -helical structure is experimentally observed centred around 1635 cm⁻¹.⁶ If the C-PCM background is included with the explicit solvation layer, the calculated amide I couplet is centred at 1722 cm⁻¹ (without a wavenumber scaling factor). This is already much closer to the experiment compared to the couplet being centred at 1757 cm⁻¹ when only the explicit solvation is included without the C-PCM background (without a wavenumber scaling factor). Further improvement of the band position could e.g. be accomplished using a larger basis set as shown in Fig. S 5 in the ESI,†. While most spectral features are very similar when using a 6-311G(2d,p) basis set for the geometry optimization, the amide I couplet in the ROA is calculated at 1694 cm⁻¹ (see Fig. S5 in the ESI,†).

Comparison with experiment: α -helix

The results discussed above give a qualitative idea of how the explicit water molecules affect the amide modes of HCO-(L-

Ala)₅-NH₂, but it must be assessed how the spectral influence of the explicit hydration relates to experimental spectral patterns. In Fig. 11, the different calculated spectra are compared to that of the peptide AK21, since this peptide is predominantly α -helical in aqueous solution and mainly consists of alanine.⁷² The ROA spectrum of AK21 can be very well simulated using a poly-L-alanine model peptide in α -helical conformation, as we demonstrated before.⁶

To get a detailed and objective comparison of the spectra, the similarity S_{fig} of each predicted ROA spectrum with the experimental ROA spectrum of AK21 was calculated based on the same approach we used before⁶ (see eq. 1 and further details in the ESI,†). Because of *i.a.* the harmonic approximation, the calculated spectra at this level of theory are typically overestimated in the wavenumber dimension and therefore the calculated spectrum is scaled to align it with the experimental spectrum by multiplying the wavenumbers by a scaling factor. For the level of theory used here, we have shown a global scaling factor of 0.987 works very well.⁶ The amide I region is typically even more overestimated in the calculated spectra, and thus requires a bigger shift (lower scaling factor) compared to the remainder of the spectrum.⁶ Therefore, the optimal scaling factor was used for the amide I vibrational modes (see details in the ESI,†). The final alignment of the wavenumber scaled calculated spectra with the experimental spectra is shown in Fig. 11. While all calculated ROA spectra are very similar to the experiment, the Raman spectra do show some differences since there are also contributions from the non-alanine side-chains in the experiment, which are not included in the calculations.

In the similarity analysis in Fig. 11, it is shown that the gas phase calculation has the lowest similarity with experiment and requires the biggest shift of the calculated amide I region in comparison with experiment (Fig. S 6 in the ESI,†). This demonstrates the importance of including the solvent in the calculations. However, most of the differences between the similarity values of the ROA spectra arise from the low wavenumber region (orange bar), which must not be overinterpreted because of the partial geometry optimization used here. The differences in the amide III region (green bar) are small, yet show that all the models that include hydrogen bonds of water molecules with the N-H groups have a lower similarity value (see a, b and d), compared to when only hydrogen bonds of water with C=O groups were included (c). In the actual spectra in Fig. 11, it can be observed that the reduced intensity of the positive amide III band around 1300 cm⁻¹ causes the lowering of the similarity values. The HCO-(L-Ala)₅-NH₂ model that was calculated is significantly smaller than the actual AK21 peptide. As a result, in the 30 different hydration configurations, only the N-H groups that do not form an intramolecular hydrogen bond with a C=O group were found to form hydrogen bonds with water molecules (see Table S 4 in the ESI,†). Therefore, the contribution of these hydrated N-H groups does not simulate the spectral behaviour of AK21 if it adopts a longer α -helical backbone conformation in solution. In that case, hydrogen bonds between N-H groups and water molecules would not form if most of the N-H groups

form intramolecular hydrogen bonds with the C=O groups. On the other hand, the C=O groups in the HCO-(L-Ala)₅-NH₂ in an α -helical conformation, were found to be able to form hydrogen bonds with the solvent (Table S 5 in the ESI,†). As shown here, such hydrogen bonds (C=O...H₂O) do not affect the ratio of the positive amide III bands to the same extent the hydration of the N-H groups does. Therefore, in the next section, the effect on multiple secondary structures and different α -helical conformations is considered to further evaluate the effect of the explicit water molecules on the amide modes.

Effect of explicit solvation on secondary structure motifs

Since ROA is very sensitive to secondary structure,⁶ in this section, the effect of the hydration on the Raman and ROA spectra of different secondary structures is evaluated. In Fig. 12, the Raman spectra are shown for each of the eight different backbone conformations. Each spectrum was obtained an arithmetic mean from the 30 hydrated configurations for the relevant conformation. On the left, for each of the eight different peptide conformations, the 30 individual spectra are shown and their respective averages on the right-hand side. The most distinctive difference due to the explicit compared to the implicit solvation in the computed Raman spectra is the shift of 15-20 cm⁻¹ of the amide I region to lower wavenumbers and the broadening of that band, hence improving its position in comparison with experiment. This is in agreement with the amide resonance model as the carbonyl bond length increases, as discussed above. Besides the effect on the amide I region, most of the Raman spectra do not significantly change because of the explicit solvation. Only the extended amide III regions of the peptide models with PPII and β -strand conformations are distinctly altered upon backbone hydration. The Raman band around 1250 cm⁻¹ shifts to 1285 cm⁻¹ upon including explicit water molecules surrounding the PPII conformation and the amide III Raman band around 1222 cm⁻¹ shifts to about 1278 cm⁻¹ in the case of the β -strand conformation (wavenumbers without scaling factor). In contrast to the 3_{10} - and α -helical structures where the N-H groups are shielded from hydration because of intramolecular hydrogen bonds with the carbonyl groups, the N-H groups of both the PPII and β -strand conformation are almost all forming hydrogen bonds with the explicit water molecules (Table S 4 in the ESI,†). This explains why the amide III regions of these two types of secondary structure are much more affected by the explicit hydration.

In the ROA spectra in Fig. 13, it is also clear that the amide I shape, position and relative intensity are affected by the explicit water molecules. In the amide III region of the α -helical conformations (top 4 panels), the ratio of the positive α -helical bands is altered by the presence of the water molecules (see e.g. panel 3 and 4). As was demonstrated above, both the hydration of the C=O and N-H groups affect the amide III patterns, yet their effect on the geometry and the spectra is different. Furthermore, in these small helical systems, the effect arises from the hydration of the N-H groups that are not

hydrogen bonded to a carbonyl group. The effect of the separate hydration of the C=O and N-H groups of the α -helical conformations with $\phi; \psi \approx -59^\circ; -44^\circ$ and $-77^\circ; -34^\circ$ also support these observations (see Fig. S 7 and Fig. S 8 in the ESI,†). Therefore, the distinct lowering of the maximum intensity of the band around 1300 cm^{-1} does not correspond to the effect of hydration on longer α -helical structures, where the N-H groups do not form hydrogen bonds with the solvent, but only with the carbonyl groups.

On the other hand, the carbonyl groups do form hydrogen bonds with the solvent (see Table S 5 in the ESI,†), but this does not have such a pronounced effect on the amide III region as that of the hydrogen bonding of water molecules to the N-H groups.

Based on the comparison of a set of experimental ROA spectra with protein crystal structures and hydrogen-deuterium exchange experiments, Blanch *et al.* assigned the band around 1300 cm^{-1} to be dominant for α -helical structure in a hydrophobic environment and that around 1345 cm^{-1} to be dominant in α -helical conformation in a hydrophilic environment.^{55,73} Based on a few computational studies, this assignment was later questioned and shown to be inaccurate.^{6,29,30} As shown in Fig. 13 when comparing the ROA spectra in orange colour of α -helical structure with backbone torsion angles around $\phi = -59^\circ$ and $\psi = -44^\circ$ (hydrophobic environment^{68,69}) (see Fig. 2) with the conformation with $\phi = -66^\circ$ and $\psi = -41^\circ$, the band around 1300 cm^{-1} has an increased maximal intensity relative to the band around 1345 cm^{-1} for the latter conformation which contradicts the traditional assignment. As discussed above, the lowering of the band around 1300 cm^{-1} upon hydration of the α -helical structure with $\phi = -66^\circ$ and $\psi = -41^\circ$ (compare blue and orange spectra in Fig. 13), arises mostly from the hydration of the free N-H groups in the peptide model studied here. The calculated spectra that are shown here thus demonstrate that the conformational change is more important in the assignment of this band than the hydrogen bonding of the explicit water molecules and further supports our earlier conclusions that the band around 1345 cm^{-1} is a conservative marker of regular α -helical structure and that the band around 1300 cm^{-1} is much more conformationally sensitive.

While the amide III region in the Raman spectra of the PPII and β -strand conformation are strongly affected by hydration, the amide III region in the ROA spectrum of PPII also does not change much due to the presence of the explicit water molecules. For the β -strand conformation, the amide III $-/+$ signal at $1236/1253\text{ cm}^{-1}$ (without scaling factor) disappears and the negative band around 1368 cm^{-1} shifts to around 1390 cm^{-1} when including explicit water molecules. Similar to the comparison above for α -helical structure, in the next section the effect of hydrogen bonding of water molecules to either the C=O and N-H groups of the PPII conformations is evaluated in comparison with experiment, as suitable experimental spectra for this backbone conformation are available.

Comparison with experiment: PPII secondary structure

We demonstrated before that the experimental Raman and ROA spectrum of the XAO peptide can very well be reproduced by calculating the spectra of a poly-L-alanine peptide in a regular PPII conformation.⁶ Therefore, here, the calculated Raman and ROA spectra of the PPII conformation including explicit water molecules are compared to the experimental spectra of XAO. In Fig. 14, the effect of hydration is demonstrated similarly to the results discussed above for AK21. The gas phase calculation (a) is compared to when no explicit water molecules are included but only C-PCM is used to represent the solvent (b). The effect of only including the water molecules that are hydrogen bonded to the carbonyl groups (c) or the N-H groups (d) or both (e) and finally when the entire 3 \AA hydration shell is included (f). Similar as above, for the comparison with experiment, the similarity index S_{fig} was calculated for all spectra as shown in the bar graphs in Fig. 14. The spectral differences in both the Raman and ROA are small, which is also reflected in the minor differences in the S_{fig} values. The most prominent amide III band in the experimental Raman spectrum of XAO has a maximum at 1260 cm^{-1} . When only C-PCM is used to correct for the solvent, the most intense band in this spectral region is calculated at 1237 cm^{-1} (the calculated wavenumbers in Fig. 14 include a global wavenumber scaling factor of 0.987). In both (c) and (d) the most intense amide III band is observed at 1250 cm^{-1} . By including the entire 3 \AA hydration shell (f), this band has a maximum calculated at 1266 cm^{-1} . In comparison with experiment, the explicitly hydrated structures give thus a better amide III region than when only C-PCM is included. The effect of the explicit water shell on the amide III region, seems to be a combination of the formation of hydrogen bonds of both C=O and N-H with water molecules. These types of hydrogen bond interactions affect the amide modes in a different way.⁴⁰ Geometrically, the formation of a hydrogen bond of a water molecule with the oxygen atoms of the amide groups, shortens the C-N bond length and elongates the C=O bond length and thus follows the amide resonance model (see bond lengths in Fig. S 3 in the ESI,†). Hydrogen bonding of the solvent with the N-H groups, does not have such a pronounced effect on the C-N and C=O bond lengths (Fig. S 3 in the ESI,†). On the other hand, this type of interaction does have a distinct effect on the N-H bond length. If a N-H \cdots H₂O hydrogen bond is formed, the N-H bond length increases, which therefore affects the amide III vibrational modes as these involve N-H bending. Also the calculated ROA spectra are affected by the presence of the explicit water molecules. Because of small differences in the relative intensities and in the relative band positions, the comparison of the experimental ROA spectrum of XAO with the calculated spectra slightly improves when the explicit water molecules are included (Fig. 14).

The above sections demonstrate that the effect of the explicit water molecules on the Raman and ROA patterns is relatively small. Residues on the exterior of globular proteins are exposed to the solvent and their hydration is therefore different from that of residues on the interior of the protein. As the latter are shielded from the solvent. The results here show that eventual spectral differences are more likely to arise from

differences in the backbone conformation (possibly mediated by the solvent) than from the hydrogen bonding interactions themselves. However in some cases, such as for PPII structure that is not stabilized by intramolecular hydrogen bonds, the hydration has a pronounced effect on the amide vibrations and therefore the spectra. Recent studies show that it is very challenging to understand the structure-spectrum relation in much detail.^{6,32} It will therefore be very interesting to see further development and more applications of specific methods for solvation^{43,45,46,74} and further examine how the side-chains,^{32,62} conformational disorder³⁴ and hydrogen bonding with the solvent all collectively contribute to the calculated spectra of proteins.

Conclusions

The effect of protein backbone hydration on the secondary structure patterns observed in calculated Raman and ROA spectra was evaluated using DFT calculations of specific conformations of HCO-(L-Ala)₅-NH₂ in comparison with a few selected experimental spectra. It was demonstrated that for different secondary structure conformations (α -helical, 3_{10} -helix, PPII, β -strand) the conformation is much more important in determining the spectral patterns than the explicit water molecules. The spectral differences that were observed upon hydration of the structures, arise from hydrogen bonding of water molecules with the backbone C=O and N-H groups, while more distal and non-hydrogen bonded water molecules did not have a significant impact on the spectra. The shape and position of the calculated Raman and ROA amide I band is determined by the formation of hydrogen bonds of water molecules with the C=O group, while the impact of the hydration on the amide III region arises from hydrogen bonding of water molecules with both the C=O and N-H group. The effect of hydrogen bonding with the C=O group was shown to be consistent with the amide resonance model as the C=O bond length increases and the amide C-N bond length decreases. Hydrogen bonding of water molecules with the C=O groups, results in an increase of the C=O bond length and a shift of the amide I bands to lower wavenumbers, while the decrease in C-N bond length shifts the amide III bands to higher wavenumbers. Hydrogen bonding with the amide N-H group does not have such a pronounced effect on the C=O and C-N bond lengths, but rather results in an increase of the N-H bond length and affects the amide III modes (C-N stretching, N-H and C α -H bending) via the bending modes.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The Flemish Supercomputing Centre (VSC) is acknowledged for providing computational resources and support. Financial support by the BOF concerted research action '4D protein

structure' is acknowledged. C.M. thanks prof. dr. W. Herrebout for helpful discussions.

Notes and references

- 1 S. Ostovar pour, L. D. Barron, S. T. Mutter and E. W. Blanch, in *Chiral Analysis*, Elsevier, 2018, pp. 249–291.
- 2 L. D. Barron, *Biomed. Spectrosc. Imaging*, 2015, **4**, 223–253.
- 3 L. D. Barron, *Molecular light scattering and optical activity*, Cambridge University Press, Cambridge, second ed., 2004.
- 4 S. Yamamoto, H. Watarai and P. Bouř, *ChemPhysChem*, 2011, **12**, 1509–1518.
- 5 S. Yamamoto, M. Straka, H. Watarai and P. Bouř, *Phys. Chem. Chem. Phys.*, 2010, **12**, 11021.
- 6 C. Mensch, L. D. Barron and C. Johannessen, *Phys. Chem. Chem. Phys.*, 2016, **18**, 31757–31768.
- 7 L. D. Barron, L. Hecht, E. W. Blanch and A. F. Bell, *Prog. Biophys. Mol. Biol.*, 2000, **73**, 1–49.
- 8 C. Johannessen, R. Pendrill, G. Widmalm, L. Hecht and L. D. Barron, *Angew. Chemie Int. Ed.*, 2011, **50**, 5349–5351.
- 9 C. Mensch, R. Pendrill, G. Widmalm and C. Johannessen, *ChemPhysChem*, 2014, **15**, 2252–2254.
- 10 F. Zhu, J. Kapitan, G. E. Tranter, P. D. A. Pudney, N. W. Isaacs, L. Hecht and L. D. Barron, *Proteins Struct. Funct. Bioinforma.*, 2007, **70**, 823–833.
- 11 C. Mensch, A. Konijnenberg, R. Van Elzen, A.-M. Lambeir, F. Sobott and C. Johannessen, *J. Raman Spectrosc.*, 2017, **48**, 910–918.
- 12 M. Gasior-Glogowska, K. Malek, G. Zajac and M. Baranska, *Analyst*, 2016, 291–296.
- 13 J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier and E. W. Blanch, *J. Am. Chem. Soc.*, 2011, **133**, 4991–4997.
- 14 F. Zhu, N. W. Isaacs, L. Hecht, G. E. Tranter and L. D. Barron, *Chirality*, 2006, **18**, 103–115.
- 15 E. W. Blanch, L. Hecht, C. D. Syme, V. Volpetti, G. P. Lomonosoff, K. Nielsen and L. D. Barron, *J. Gen. Virol.*, 2002, **83**, 2593–600.
- 16 E. W. Blanch, D. J. Robinson, L. Hecht, C. D. Syme, K. Nielsen and L. D. Barron, *J. Gen. Virol.*, 2002, **83**, 241–246.
- 17 L. D. Barron, A. R. Gargaro and Z. Q. Wen, *J. Chem. Soc. Chem. Commun.*, 1990, **114**, 1034.
- 18 F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *Structure*, 2005, **13**, 1409–1419.
- 19 E. W. Blanch, D. D. Kasarda, L. Hecht, K. Nielsen and L. D. Barron, *Biochemistry*, 2003, **42**, 5665–73.
- 20 F. Zhu, G. E. Tranter, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 2006, **363**, 19–26.
- 21 P. W. Atkins and L. D. Barron, *Mol. Phys.*, 1969, **16**, 453–466.
- 22 L. D. Barron and A. D. Buckingham, *Mol. Phys.*, 1971, **20**, 1111–1119.
- 23 O. Quinet and B. Champagne, *J. Chem. Phys.*, 2001, **115**, 6293–6299.
- 24 O. Quinet, V. Liégeois and B. Champagne, *J. Chem. Theory Comput.*, 2005, **1**, 444–452.
- 25 V. Liégeois, K. Ruud and B. Champagne, *J. Chem. Phys.*,

- 2007, **127**, 204105–204105.
- 26 A. J. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen and S. Coriani, *J. Chem. Phys.*, 2008, **129**, 214108.
- 27 C. Johannessen and E. W. Blanch, *Curr. Phys. Chem.*, 2013, **3**, 140–150.
- 28 K. V. J. Jose and K. Raghavachari, *Chirality*, 2016, **28**, 755–768.
- 29 S. Yamamoto, T. Furukawa, P. Bouř and Y. Ozaki, *J. Phys. Chem. A*, 2014, **118**, 3655–3662.
- 30 S. Lubner, *J. Phys. Chem. A*, 2013, **117**, 2760–2770.
- 31 T. Weymuth and M. Reiher, *J. Phys. Chem. B*, 2013, **117**, 11943–11953.
- 32 C. Mensch and C. Johannessen, *Chemphyschem*, DOI:10.1002/cphc.201800678.
- 33 T. Weymuth, C. R. Jacob and M. Reiher, *ChemPhysChem*, 2011, **12**, 1165–1175.
- 34 C. Mensch, P. Bultinck and C. Johannessen, *ACS Omega*, 2018, **3**, 12944–12955.
- 35 S. Lubner and M. Reiher, *J. Phys. Chem. B*, 2010, **114**, 1057–1063.
- 36 J. Kessler, J. Kapitán and P. Bouř, *J. Phys. Chem. Lett.*, 2015, **6**, 3314–3319.
- 37 S. Yamamoto and P. Bouř, in *Frontiers of Quantum Chemistry*, Springer Singapore, Singapore, 2018, pp. 181–197.
- 38 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3093.
- 39 B. Mennucci, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 386–404.
- 40 N. S. Myshakina, Z. Ahmed and S. A. Asher, *J. Phys. Chem. B*, 2008, **112**, 11873–11877.
- 41 P. Mukhopadhyay, G. Zuber and D. N. Beratan, *Biophys. J.*, 2008, **95**, 5574–5586.
- 42 S. Cardamone, B. A. Caine, E. Blanch, M. G. Lizio and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2016, **18**, 27377–27389.
- 43 J. Hudecová, K. H. Hopmann and P. Bouř, *J. Phys. Chem. B*, 2012, **116**, 336–342.
- 44 K. H. Hopmann, K. Ruud, M. Pecul, A. Kudelski, M. Dračinský and P. Bouř, *J. Phys. Chem. B*, 2011, **115**, 4128–4137.
- 45 T. Giovannini, G. Del Frate, P. Lafiosca and C. Cappelli, *Phys. Chem. Chem. Phys.*, 2018, **20**, 9181–9197.
- 46 T. Giovannini, M. Olszówka, F. Egidi, J. R. Cheeseman, G. Scalmani and C. Cappelli, *J. Chem. Theory Comput.*, 2017, **13**, 4421–4435.
- 47 S. Lubner and M. Reiher, *J. Phys. Chem. A*, 2009, **113**, 8268–8277.
- 48 S. T. Mutter, F. Zielinski, J. R. Cheeseman, C. Johannessen, P. L. A. Popelier and E. W. Blanch, *Phys. Chem. Chem. Phys.*, 2015, **17**, 6016–6027.
- 49 F. Zielinski, S. T. Mutter, C. Johannessen, E. W. Blanch and P. L. A. Popelier, *Phys. Chem. Chem. Phys.*, 2015, **17**, 21799–21809.
- 50 S. T. Mutter, F. Zielinski, C. Johannessen, P. L. A. Popelier and E. W. Blanch, *J. Phys. Chem. A*, 2016, **120**, 1908–1916.
- 51 S. Yamamoto, J. Kaminský and P. Bouř, *Anal. Chem.*, 2012, **84**, 2440–2451.
- 52 K. V. J. Jose and K. Raghavachari, *J. Chem. Theory Comput.*, 2016, **12**, 585–594.
- 53 L. D. Barron, L. Hecht and G. Wilson, *Biochemistry*, 1997, **36**, 13143–13147.
- 54 K. J. Jalkanen, I. M. Degtyarenko, R. M. Nieminen, X. Cao, L. A. Nafie, F. Zhu and L. D. Barron, *Theor. Chem. Acc.*, 2008, **119**, 191–210.
- 55 I. H. McColl, E. W. Blanch, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2004, **126**, 8181–8188.
- 56 M. Z. Tien, D. K. Sydykova, A. G. Meyer and C. O. Wilke, *PeerJ*, 2013, **1**, e80.
- 57 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. C. III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, AMBER14, University of California, San Francisco, 2014.
- 58 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 59 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157–1174.
- 60 P. Almeida, *Proteins: Concepts in Biochemistry*, Garland Science, Taylor & Francis Group, New York, 2016.
- 61 P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126–4132.
- 62 J. Hudecová, J. Kapitán, V. Baumruk, R. P. Hammer, T. A. Keiderling and P. Bouř, *J. Phys. Chem. A*, 2010, **114**, 7642–7651.
- 63 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, Gaussian16 (rev. A.03), Gaussian, Inc., Wallingford CT, 2016.
- 64 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- 65 M. Cossi, N. Rega, G. Scalmani and V. Barone, *J. Comput. Chem.*, 2003, **24**, 669–681.
- 66 Y. He, B. Wang, R. K. Dukor and L. A. Nafie, *Appl. Spectrosc.*, 2011, **65**, 699–723.
- 67 W. Hug, *Appl. Spectrosc.*, 2003, **57**, 1–13.

- 68 T. Blundell, D. Barlow, N. Borkakoti and J. Thornton, *Nature*, 1983, **306**, 281–283.
- 69 M. Sundaralingam and Y. C. Sekharudu, *Science*, 1989, **244**, 1333–1337.
- 70 E. J. Milner-White, *Protein Sci.*, 1997, **6**, 2477–2482.
- 71 T. Weymuth, C. R. Jacob and M. Reiher, *J. Phys. Chem. B*, 2010, **114**, 10649–10660.
- 72 E. W. Blanch, L. A. Morozova-Roche, D. A. E. Cochran, A. J. Doig, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 2000, **301**, 553–563.
- 73 E. W. Blanch, A. F. Bell, L. Hecht, L. A. Day and L. D. Barron, *J. Mol. Biol.*, 1999, **290**, 1–7.
- 74 J. Kessler, V. Andrushchenko, J. Kapitán and P. Bouř, *Phys. Chem. Chem. Phys.*, 2018, **20**, 4926–4935.

Table 1 Initially chosen ϕ (°) and ψ (°) angles before geometry optimization.

Conformation	Set angles		Secondary structure
	ϕ	ψ	
1	-64	-47	RH α -helix*
2	-59	-44	RH α -helix (hydrophobic)**
3	-77	-34	RH α -helix with large tilt**
4	-66	-41	RH α -helix (hydrophilic)**
5	-71	-18	3_{10} -helix
6	66	41	LH α -helix
7	-75	145	PPII
8	-125	150	β -strand

*The calculation of the regular structure with these backbone torsion angles was found to reproduce the experimental ROA spectrum of the α -helical peptide AK21 very well (see ref 6). **See Fig. 2.

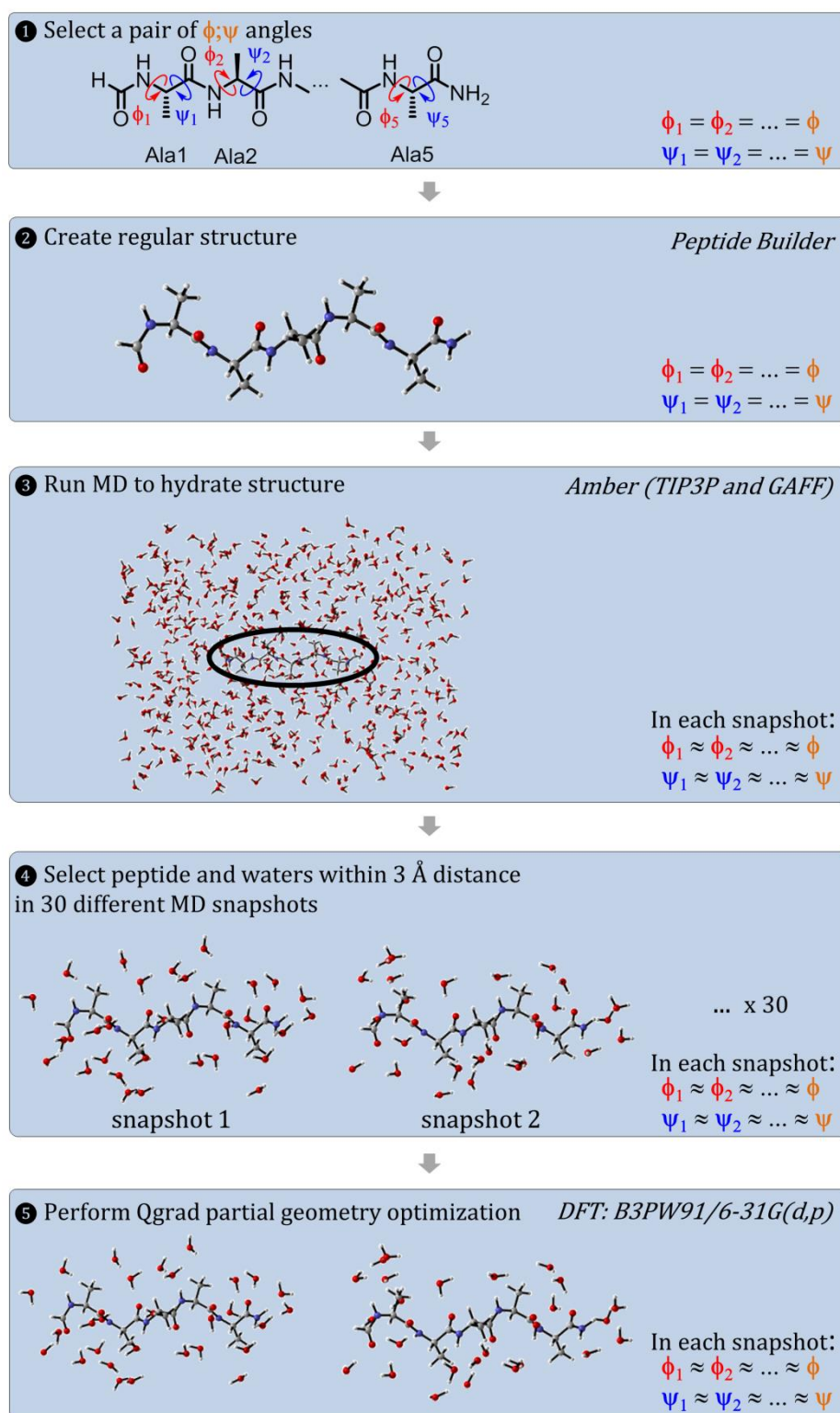
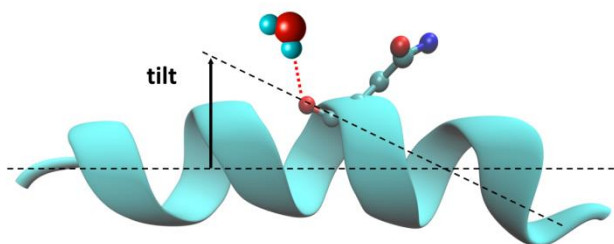
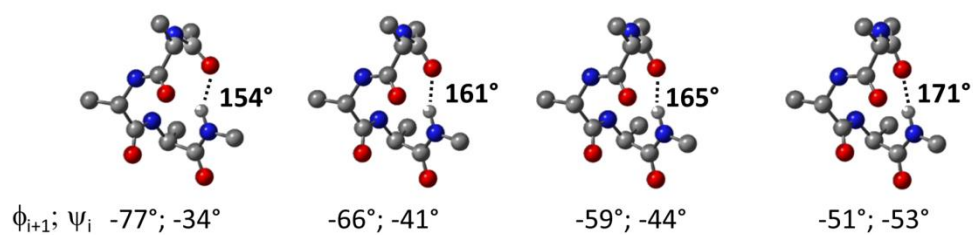


Fig. 1 Description of the workflow used to generate 30 different solvent clusters surrounding each conformation of HCO-(L-Ala)₅-NH₂. The conformation of the peptides are defined by a single pair of (ϕ , ψ) angles (see Table 1).

(a) Hydration of α -helical structure

(b) Carbonyl tilting



(c) Ramachandran plots

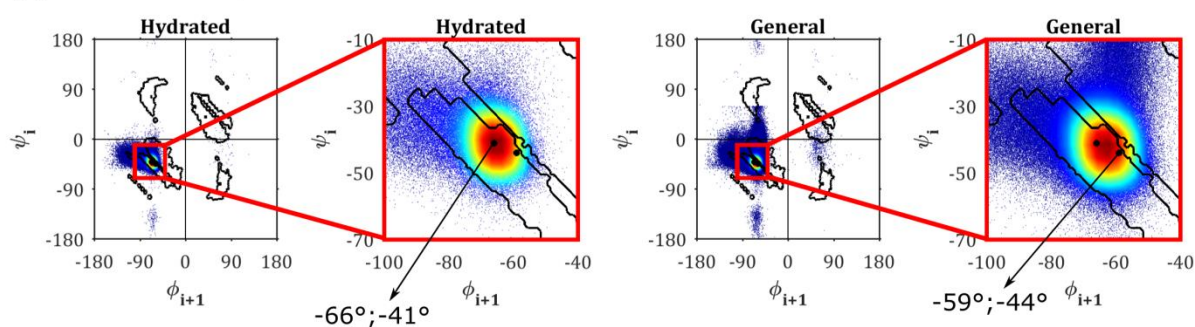


Fig. 2 The effect of hydrogen bonding of water to the backbone angles of α -helical structure: (a) A water molecule binding to Asn93 in a crystal structure of lysozyme (PDB id. 1IEE) affecting the (b) tilt of the carbonyl group outwards of the helix axis. (c) Ramachandran plots of all α -helical segments (stride) in the Top8000 database^{41,42} "general" case and only the ones with a hydrogen bonded water molecule to the carbonyl group.

Table 2 Deviation of the initially chosen ϕ (°) and ψ (°) angles after the MD simulations and geometry optimization at DFT level. For each ensemble of 30 structures, the mean torsion angles were calculated as the average of each angle of each residue of each structure with standard deviation σ .

Conformation	Set angles		Secondary Structure	Angles after geometry optimization			
	ϕ	ψ		mean ϕ	σ	mean ψ	σ
1	-64	-47	RH α -helix	-63.1	2.2	-48.2	1.2
2	-59	-44	RH α -helix	-58.1	2.1	-44.9	1.7
3	-77	-34	RH α -helix	-75.8	2.1	-35.1	1.4
4	-66	-41	RH α -helix	-65.1	1.9	-41.9	1.5
5	-71	-18	3_{10} -helix	-69.8	2.4	-19.2	1.5
6	66	41	LH α -helix	68.0	2.1	39.2	1.6
7	-75	145	PPII	-75.1	1.8	146.0	1.4
8	-125	150	β -strand	-125.7	1.6	150.8	1.5

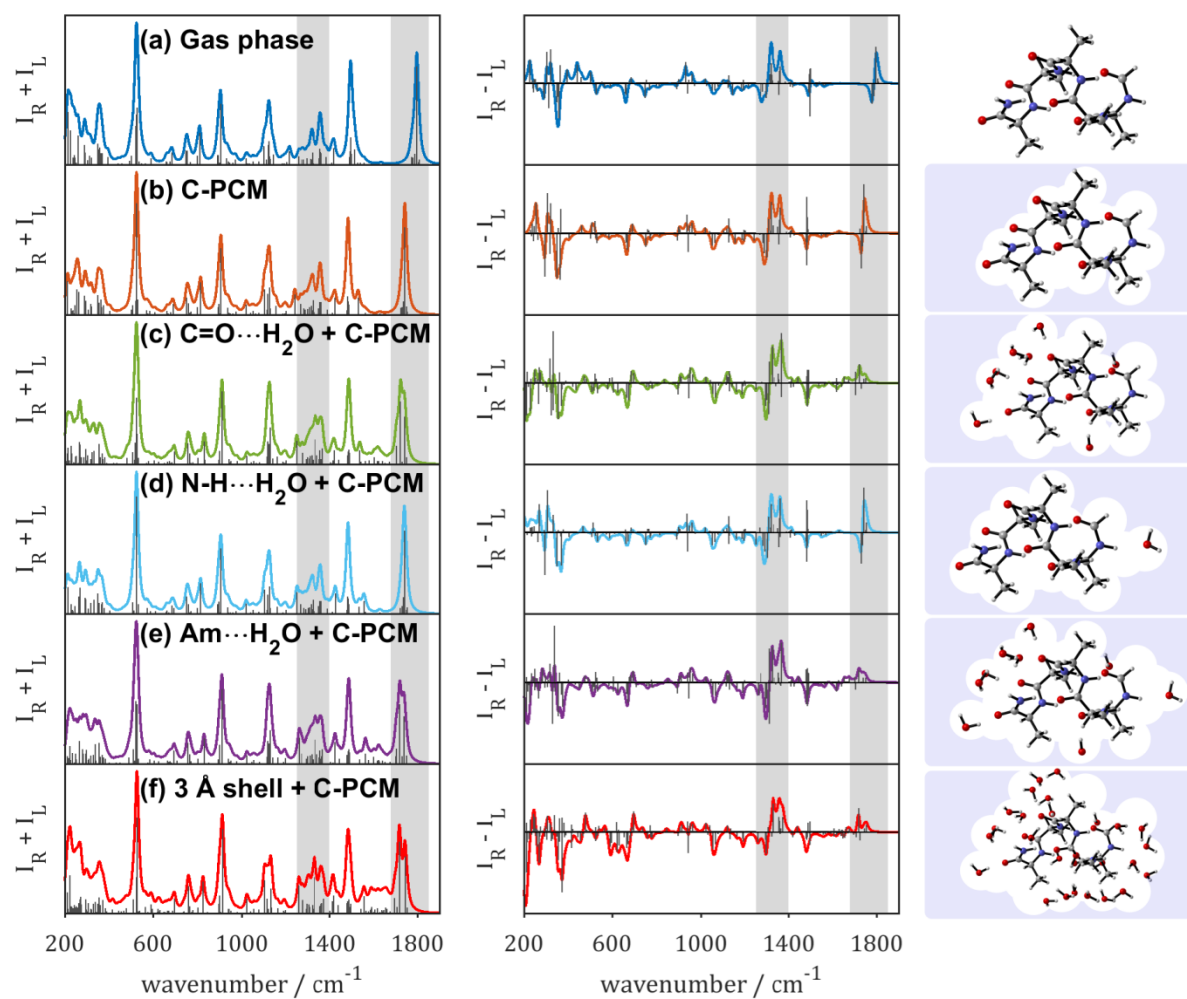


Fig. 3 Raman ($I_R + I_L$) and ROA ($I_R - I_L$) of HCO-(L-Ala)₅-NH₂ in an α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$): (a) gas phase calculation; (b) solute with C-PCM solvent correction; (c) solute with all water molecules hydrogen bonding to the solute's carbonyl groups (C=O) with C-PCM background; (d) solute with all water molecules that form a hydrogen bond to the amide N-H groups with C-PCM background; (e) solute including all water molecules that form a hydrogen bond with the amide (Am) groups of the solute with C-PCM background (f) a single MD snapshot including all water molecules within a 3 Å distance of the solute with C-PCM background. Each system was partially optimised in normal coordinates prior to calculating the spectra.

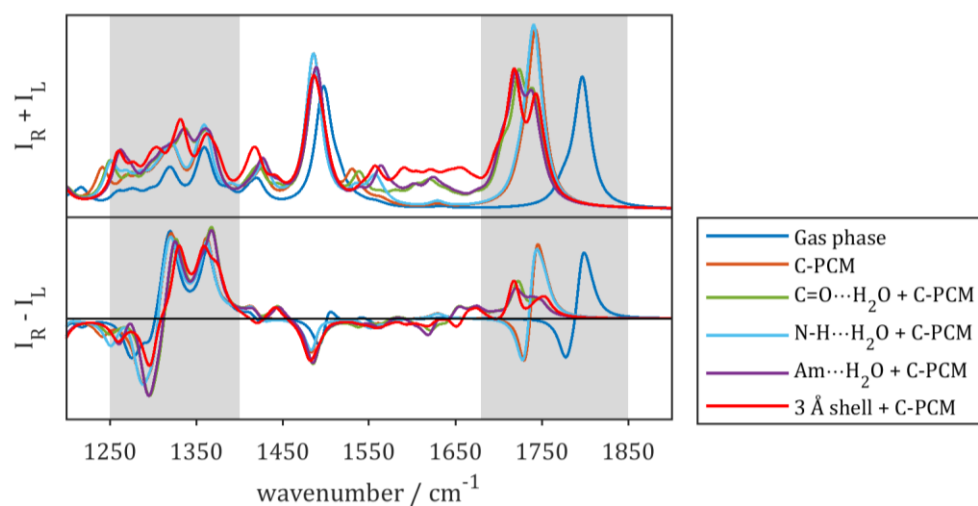


Fig. 4 Superposition of the Raman ($I_R + I_L$) and ROA ($I_R - I_L$) shown in Fig. 3 using the same colouring code.

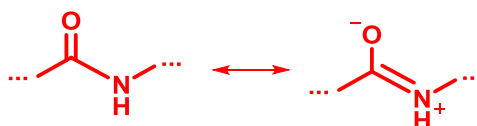


Fig. 5 The amide resonance model.^{4,43} Hydrogen bonding of solvent molecules to amide groups affects the amide geometry by stabilisation of the partially charged oxygen and nitrogen atoms. This stabilisation of the charged resonance structure (right) compared to the uncharged amide resonance structure (left) leads to an increased C=O bond length and decrease in C-N bond length.

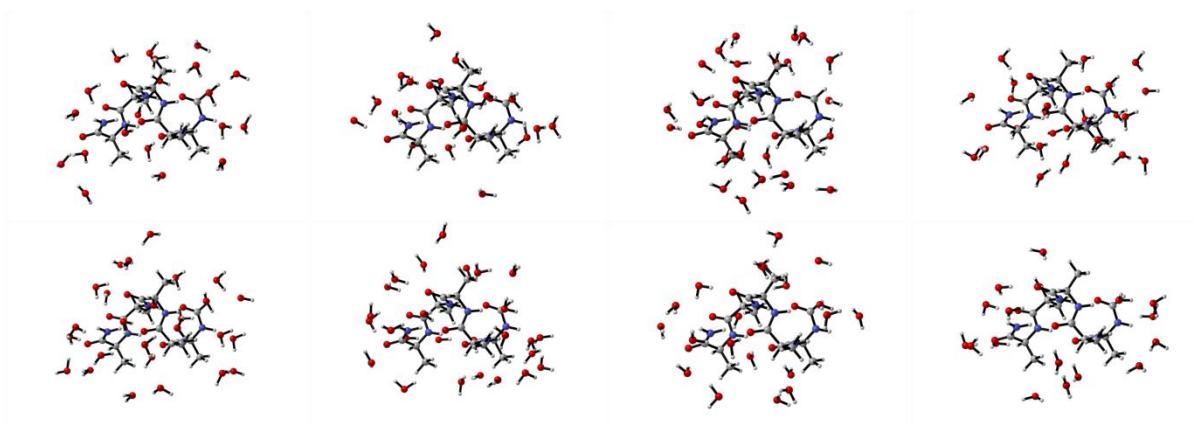


Fig. 6: Eight examples of water configurations surrounding the solute. All water molecules that are within a 3 Å distance (water oxygen distance to any atom of the solute). Fig. was created with CylView v1.0 beta.

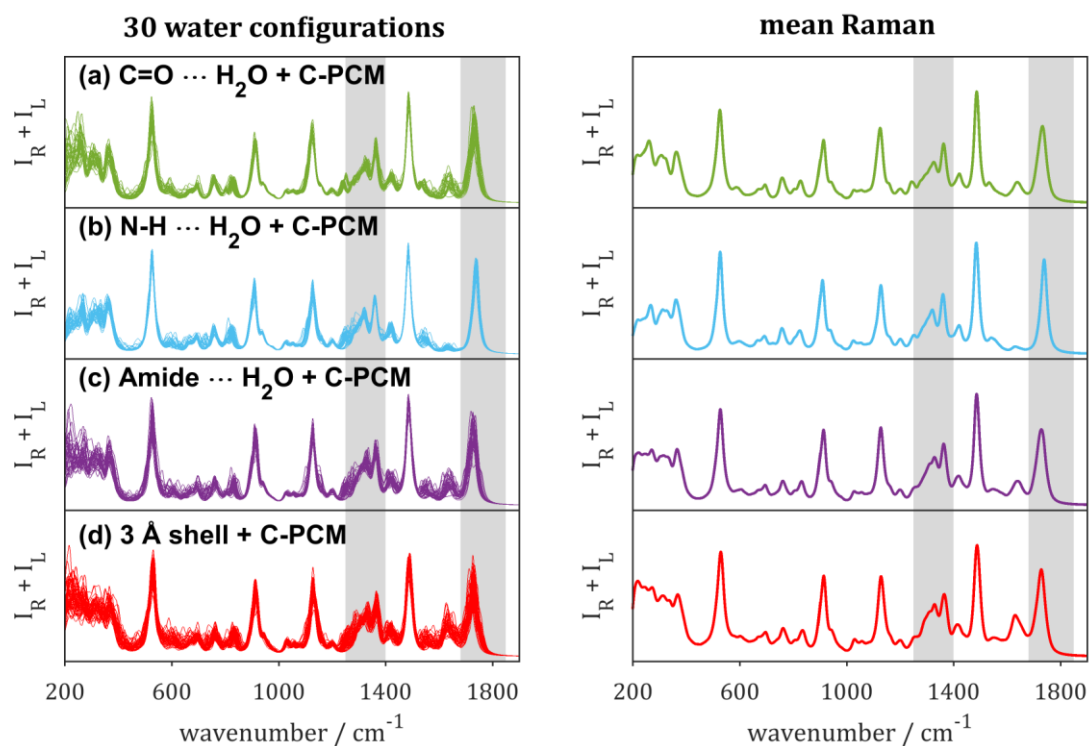


Fig. 7 Averaging over 30 water configurations of HCO-(L-Ala)₅-NH₂ in an α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$): individual Raman ($I_R + I_L$) spectra of each of the 30 solvent configurations (left) and the corresponding mean Raman spectra (right). 30 hydration configurations surrounding the solute are considered including (a) all water molecules that are hydrogen bonded to the solute's carbonyl groups (C=O) with C-PCM background; (b) all water molecules that form a hydrogen bond to the amide N-H groups with C-PCM background; (c) all water molecules that form a hydrogen bond with the amide (Am; both N-H and C=O) groups of the solute with C-PCM background (d) all water molecules within 3 Å of the solute (distance between the oxygen atom of water and any one of the solute's atoms) with C-PCM background.

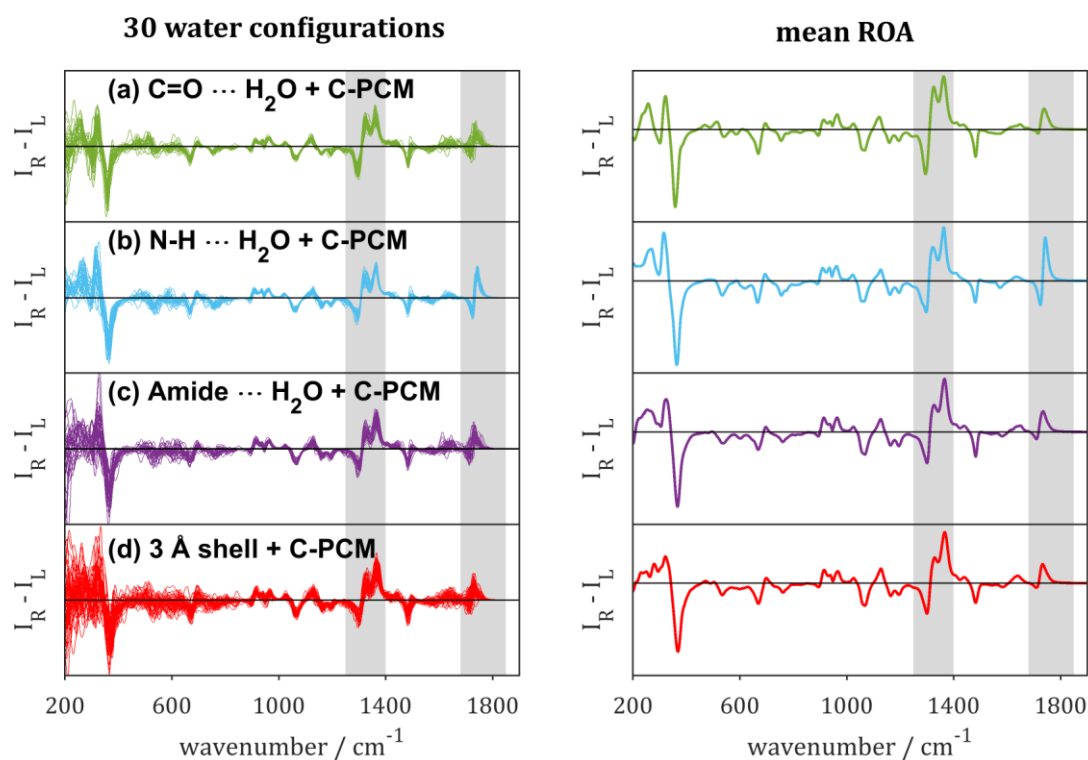


Fig. 8 Averaging over 30 water configurations of HCO-(L-Ala)5-NH2 in an α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$): individual ROA (I_R+I_L) spectra of each of the 30 solvent configurations (left) and the corresponding mean ROA spectra (right). 30 hydration configurations surrounding the solute are considered including (a) all water molecules that are hydrogen bonded to the solute's carbonyl groups (C=O) with C-PCM background; (b) all water molecules that form a hydrogen bond to the amide N-H groups with C-PCM background; (c) all water molecules that form a hydrogen bond with the amide (both N-H and C=O) groups of the solute with C-PCM background (d) all water molecules within 3 Å of the solute (distance between the oxygen atom of water and any one of the solute's atoms) with C-PCM background.

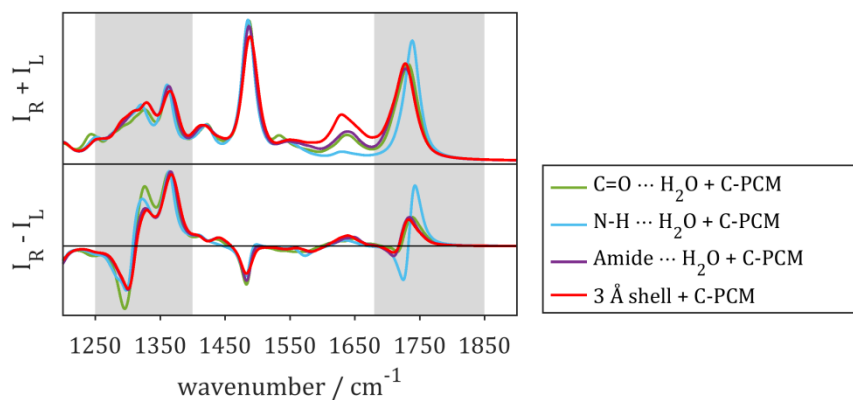


Fig. 9: Superposition of the Raman (I_R+I_L) and ROA (I_R-I_L) spectra shown in Fig. 7 and Fig. 8, respectively (same colouring).

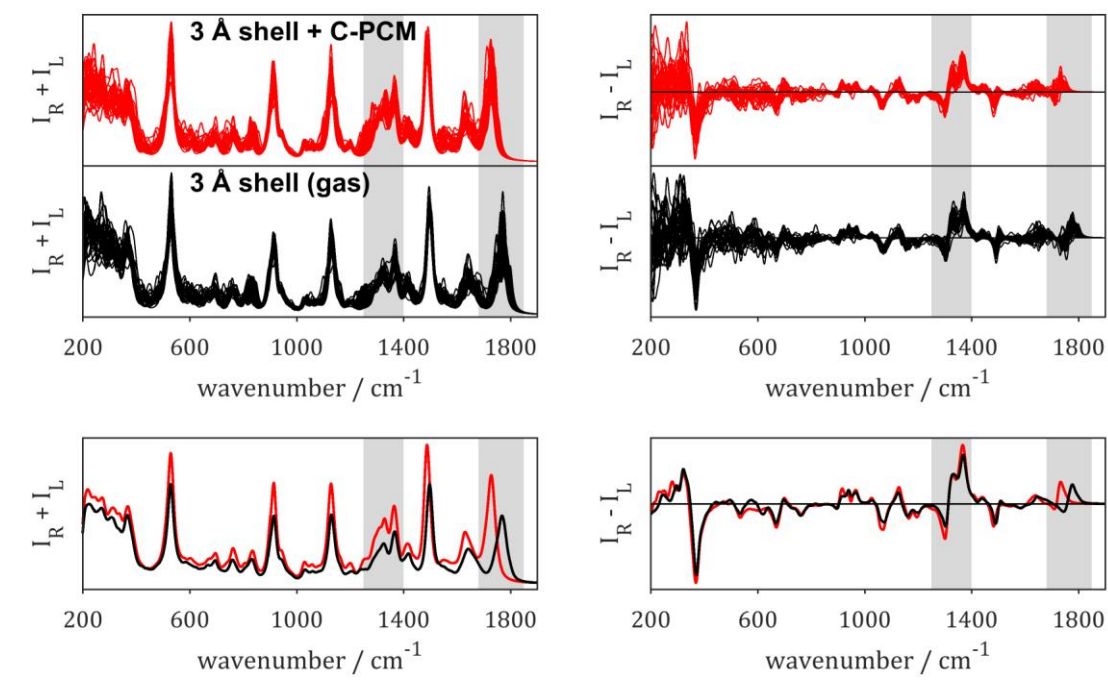


Fig. 10 Raman ($I_R + I_L$) and ROA ($I_R - I_L$) of 30 hydration configurations of HCO-(L-Ala)₅-NH₂ in an α -helical conformation ($\phi; \psi \approx -66^\circ; -41^\circ$) with (red) and without (black) C-PCM correction. The top two rows show the 30 individual spectra. The bottom row shows the corresponding mean Raman and ROA.

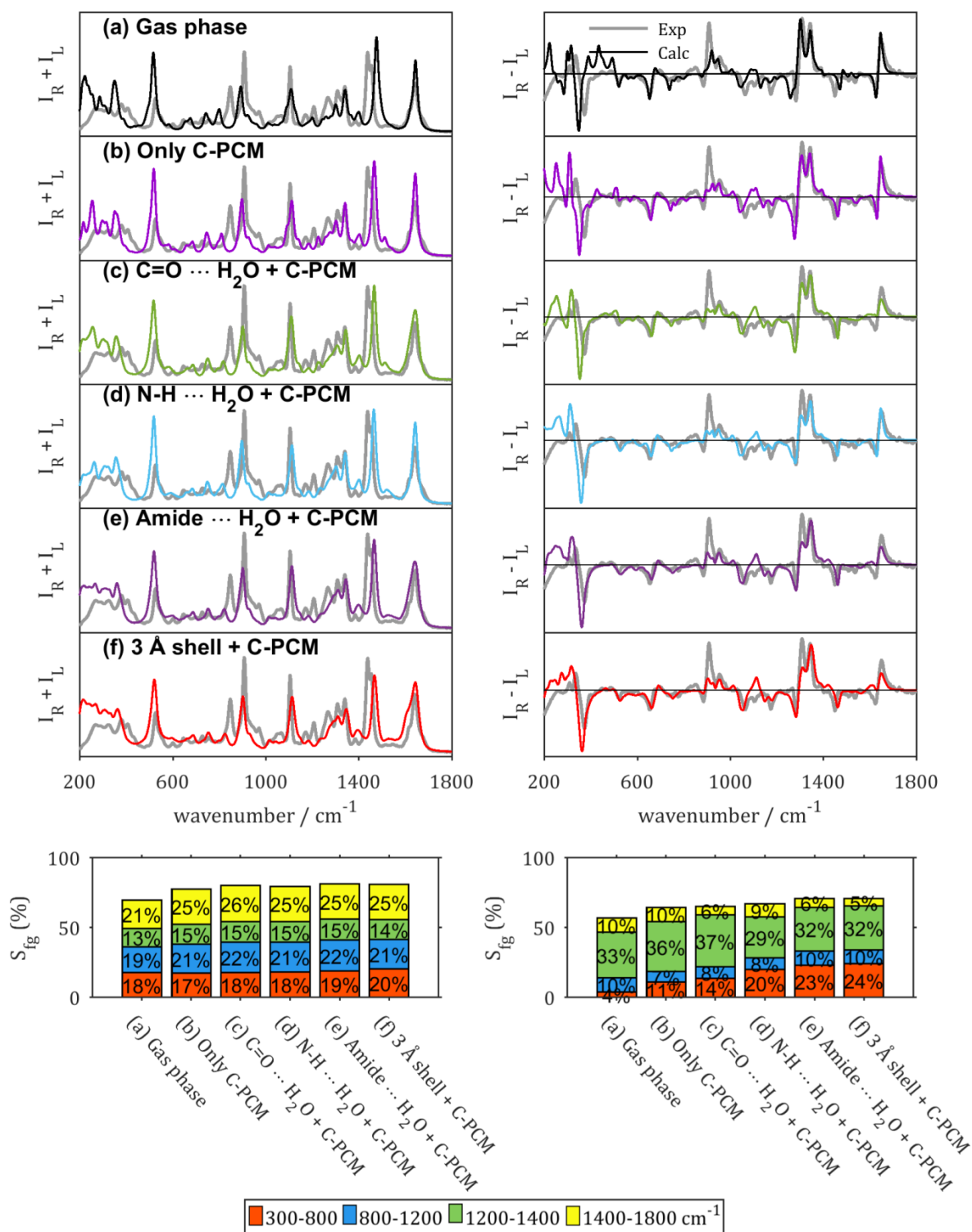


Fig. 11 Raman (I_R+I_L) and ROA (I_R-I_L) of the calculated spectra in comparison with the experimental spectra of AK21. (a) The gas phase spectrum does not include explicit water molecules nor C-PCM and (b) only includes C-PCM and no explicit water molecules. The explicitly solvated systems (c-f) are averaged over 30 water configurations. The amide I region was scaled using the optimal scaling factor in that region and the remainder of the spectrum was scaled using a global scaling factor of 0.987 (see the ESI,[†]). The bar graphs display the similarity (S_{fig}) of all Raman (left) and ROA (right) spectra compared to experiment.

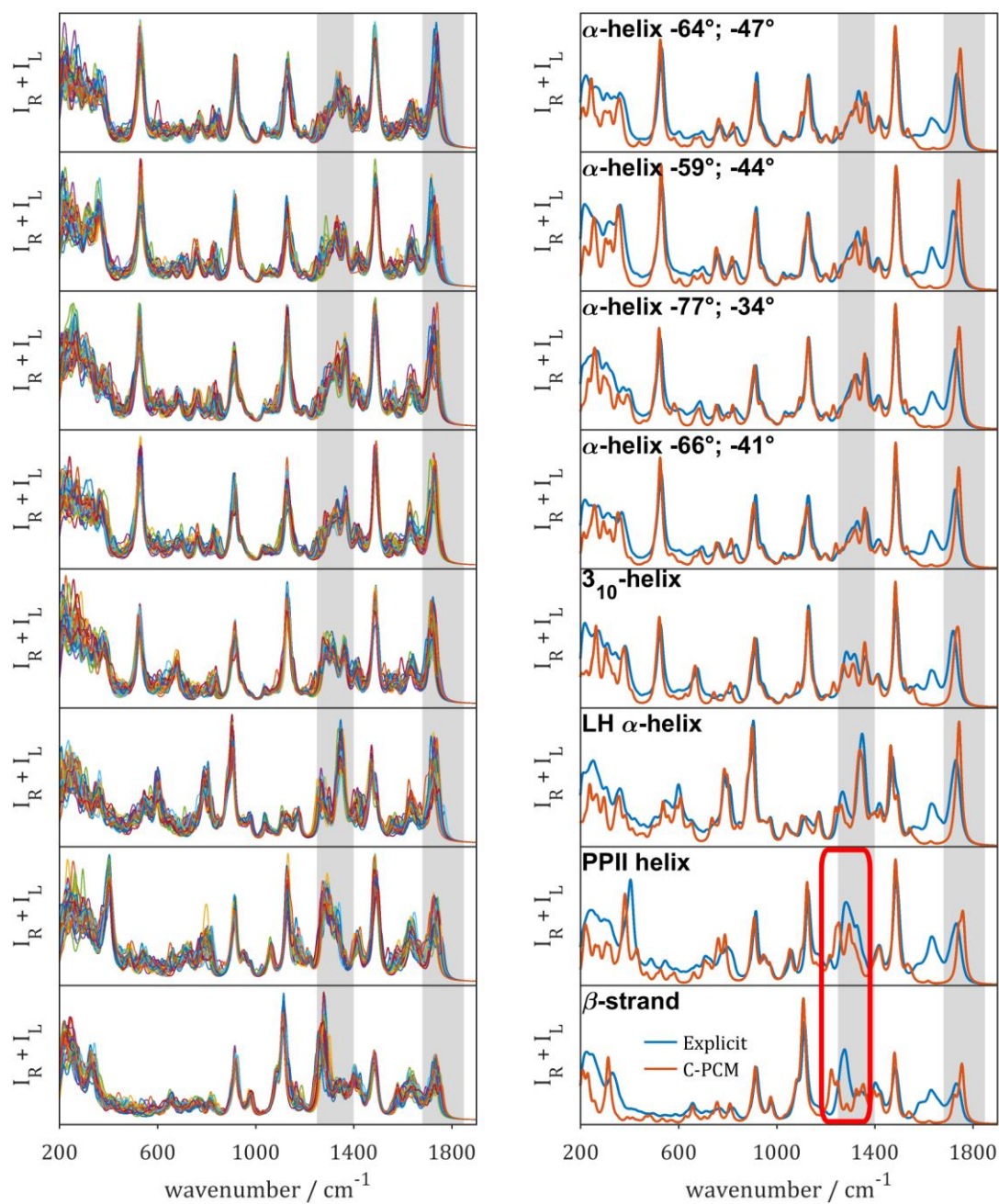


Fig. 12 Raman spectra ($I_R + I_L$) of HCO-(L-Ala)₅-NH₂ in different regular secondary structure conformations. Each conformation was calculated with 30 different explicit water configurations (individual spectra on the left; 3 Å water cluster; including C-PCM background). The mean spectrum (blue; right) is compared to the same backbone conformation only including C-PCM and no explicit waters. The bands around 1630 cm⁻¹ arise from the bending vibrational modes of the water molecules.

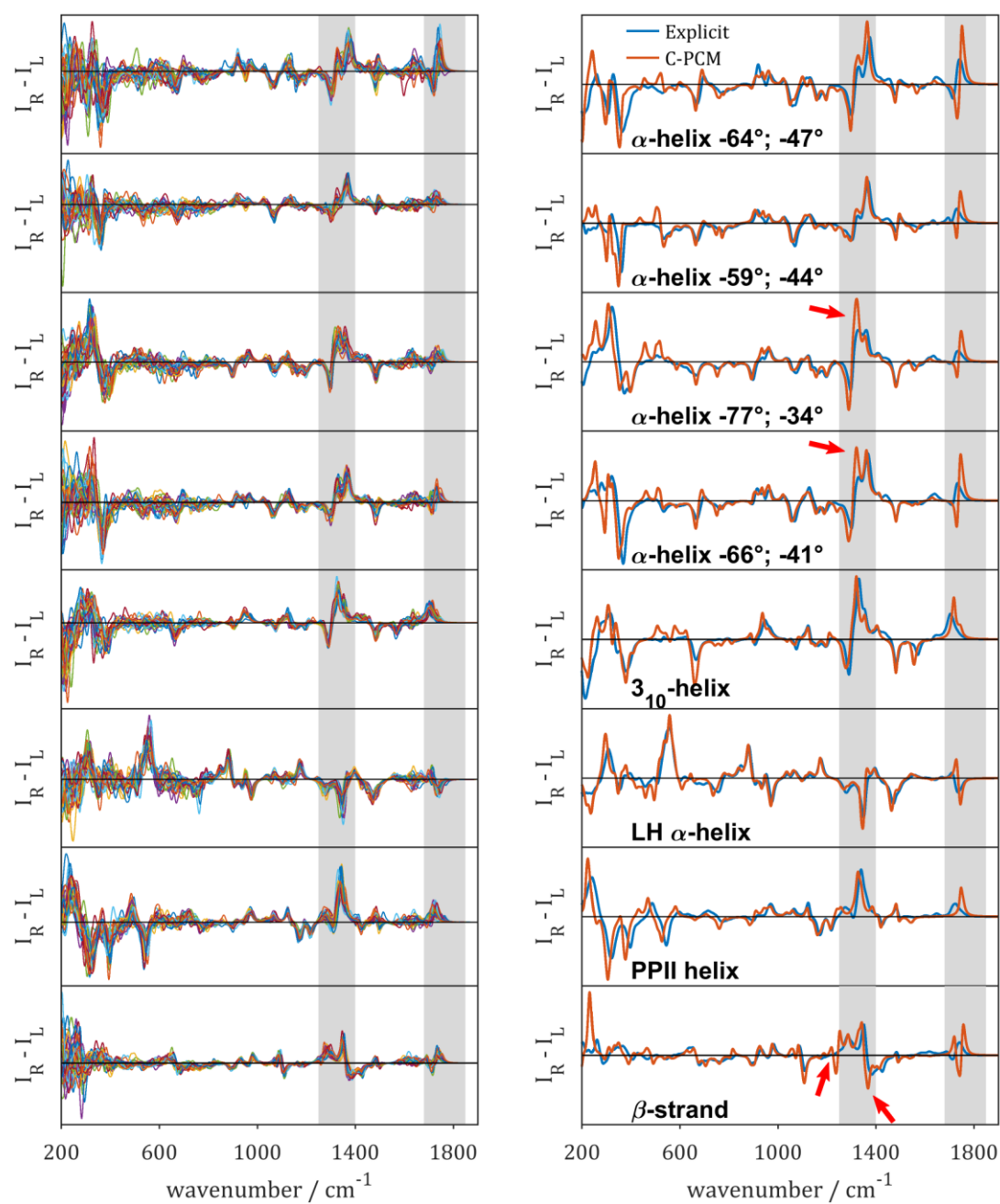


Fig. 13 ROA ($I_R - I_L$) spectra of HCO-(L-Ala)₅-NH₂ in different regular secondary structure conformations. Each conformation was calculated with 30 different explicit water configurations (individual spectra on the left; 3 Å water cluster; including C-PCM background). The mean spectrum (blue; right) is compared to the same backbone conformation only including C-PCM.

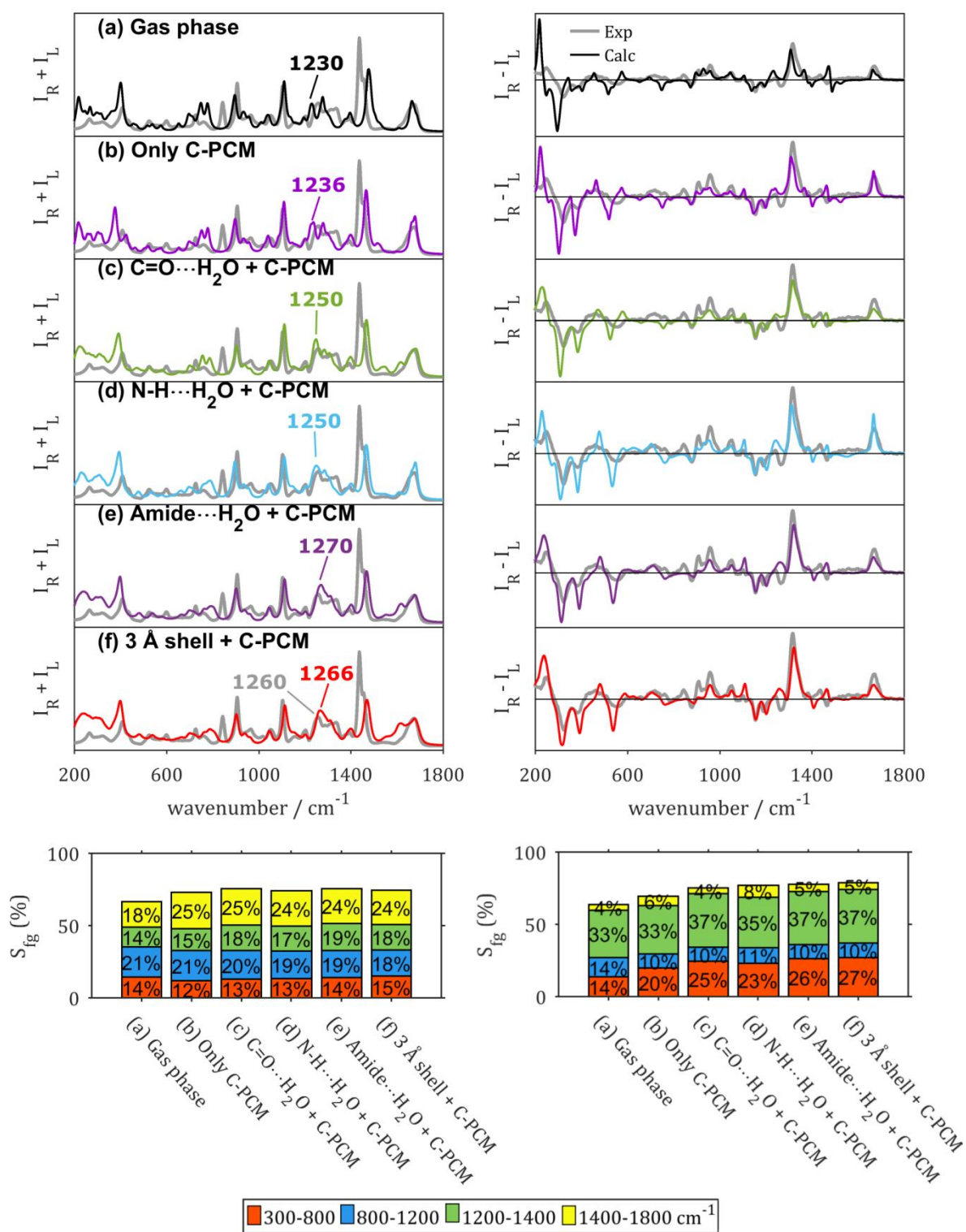


Fig. 14 Raman (I_R+I_L) and ROA (I_R-I_L) of the calculated spectra in comparison with the experimental spectra of the XAO peptide. (a) The gas phase spectrum does not include explicit water molecules nor C-PCM and (b) only includes C-PCM and no explicit water molecules. The explicitly solvated systems (c-f) are averaged over 30 water configurations. The amide I region was scaled using the optimal scaling factor in that region and the remainder of the spectrum was scaled using a global scaling factor of 0.987 (see the ESI,†). The bar graphs display the similarity (S_{fg}) of all Raman (left) and ROA (right) spectra compared to experiment.