PCCP

Accepted Manuscript

This article can be cited before page numbers have been issued, to do this please use: C. Mensch, L. Barron and C. Johannessen, *Phys. Chem. Chem. Phys.*, 2016, DOI: 10.1039/C6CP05862K.

ROYAL SOCIETY OF CHEMISTRY

rsc.li/pccp

View Article Online
DOI: 10.1039/C6CP05862K

Journal Name

ARTICLE

Ramachandran mapping of peptide conformation using a large database of computed Raman and Raman optical activity spectra

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Carl Mensch,[a,b] Laurence D. Barron[c] and Christian Johannessen*[a]

The past decades, Raman optical activity (ROA) spectroscopy has been shown to be very sensitive to the solution structure of peptides and proteins. A major and urgent challenge remains the need to make detailed assignments of experimental ROA patterns and relate those to the solution structure adopted by the protein. The past years, theoretical developments and implementations of ROA theory have made it possible to use quantum chemical methods to compute ROA spectra of peptides. In this work, a large database of ROA spectra of peptide model structures describing the allowed backbone conformations of proteins were systematically calculated and used to make unprecedented detailed assignments of experimental ROA patterns to the conformational elements of the peptide in solution. By using a similarity index to compare an experimental spectrum to the database spectra (2902 theoretical spectra), the conformational preference of the peptide in solution can be assigned to a very specific region in the Ramachandran space. For six (poly)peptides this approach was validated and gives excellent agreement between experiment and theory. Additionally, hydrogen/deuterium exchanged structures and the conformational dependence of the amide modes in Raman spectra can be analysed using the new database. The excellent agreement between experiment and theory demonstrates the power of the newly developed database as a tool to study Raman and ROA patterns of peptides and proteins. The interpretation of experimental ROA patterns of different proteins published in scientific literature is discussed based on the spectral trends observed in the database.

Introduction

The study of protein structure and function has been at the core of biochemistry for decades. X-ray crystallography and multidimensional NMR methods are applied to study protein structure at atomic resolution, complemented by information obtained from low-resolution methods such as small-angle X-ray scattering (SAXS), Förster resonance energy transfer (FRET) spectroscopy and vibrational spectroscopy.[1–3] While X-ray crystallography can be applied to study complex and large biomolecular systems, the necessary crystallisation of the sample is not trivial and often not feasible. Furthermore, the high-resolution structure provides a static picture of the protein and does not describe the solution-phase behaviour. NMR spectroscopy can on the other hand be applied to the peptide and protein solution structure and dynamics, but is limited to study smaller systems (routinely up to 30 kDa for atomic resolution).[4] Because of the huge importance of understanding protein structure and function in biochemistry and the limitations and challenges of the above-mentioned techniques, new methods need to be developed to characterise protein structure and behaviour to provide complementary structural information, in particular in cases where the conventional techniques struggle, such as intrinsically disordered proteins and glycoproteins.

Vibrational spectroscopic methods such as Fourier transform infrared spectroscopy (FTIR) and Raman spectroscopy have been widely employed to study the structure and function of biomolecules including proteins.[5,6] These techniques are specifically apt at analysing the secondary structure of proteins.[2,7] Especially the chiroptical variant of Raman spectroscopy, Raman optical activity (ROA), is very sensitive to the three-dimensional solution structure and dynamics of biomolecules.[8,9] It has been applied to study the solution structure of peptides[10–12], proteins[13–15], glycoproteins[16–18], complex sugars[19] and even intact viruses[20–22]. While the unique conformational sensitivity of ROA is recognised, the biggest challenge lies in the detailed interpretation of the experimental ROA patterns. Clustering methods such as non-linear mapping[23] and principal component analysis[22] are capable of delineating experimental ROA spectra in different structural classes, thereby demonstrating the huge potential of solution-phase ROA measurements. Nonetheless, to make detailed spectral assignments, theoretical interpretations using

[a] Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, B-2020 Antwerp, Belgium. E-mail: christian.johannessen@uantwerpen.be.
[b] Department of Inorganic and Physical Chemistry, Ghent University, Krijgslaan 281 (S3), B-9000 Ghent, Belgium.
[c] School of Chemistry, University of Glasgow, Joseph Black Building, Glasgow G12 8QQ, United Kingdom.
†Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

density function theory (DFT) must be employed. Together with increasing computer power and with the development and implementation of the analytical derivatives of the ROA tensors[24–26] and a general Kohn-Sham density functional response-theory scheme for evaluating the property tensors[27], the calculation of ROA patterns has been a major focus in the field. While metallothionein (31 amino acids) is the largest peptide for which *ab initio* calculations have been performed for the whole molecule[28], fragmentation methods allowed the simulation of larger proteins such as the insulin dimer (51 amino acids)[29] and even globular proteins up to 585 amino acids such as human serum albumin[14]. Nonetheless, because of the huge structural sensitivity of ROA, these calculations are very dependent on the initial structures used in the simulations and to conformational averaging. The starting geometries are often obtained from X-ray crystallography or molecular dynamics simulations, which provide a limited source of information on the solution structure of proteins. Therefore, it is very challenging to obtain a good agreement between theory and experiment and elaborate computational modelling is required.

Here, we present the first large-scale systematic study on the conformational sensitivity of ROA to create a thorough understanding of the relation between the solution structure of a protein and its experimental ROA spectral patterns. Protein conformation is essentially determined by the two dihedral angles ϕ and ψ in the backbone of a protein. As can be seen from the Ramachandran plot in Figure 1, these torsion angles take preferential values, which depend on steric interactions and hydrogen bonding. By constructing a database of model structures all across the allowed regions in the Ramachandran plot and calculating their respective spectra, a very detailed analysis of the ROA spectral components can be performed. In this work, the experimental ROA spectra of different peptides with known conformational preference were studied using the newly developed database to validate the simulated spectra and investigate the experimental ROA patterns of proteins in much detail.

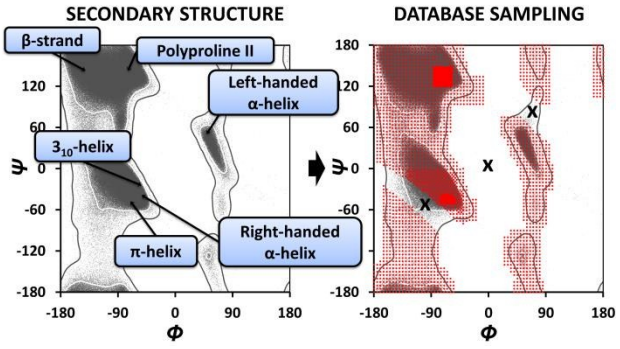

Figure 1 Spectral database creation: Ramachandran plot showing the backbone dihedral angles Φ and Ψ for 8000 high-resolution protein crystal structures adapted from Richardson *et al.* (left).[33] Conformational sampling of 3035 model structures across the relevant and allowed Ramachandran regions indicated by red dots (right). The black crosses indicate three regions where $(Ala)_{11}$ peptides cannot be created using repetitive angles; these geometries overlap internally. For a larger version see ESI† Figure S 1.

Experimental

Computational methodology: database construction View Article Online
DOI: 10.1039/C6CP05862K

A structural database consisting of regular poly-L-alanine HCO-$(Ala)_{11}$-$CONH_2$ peptides defined by a single pair of ϕ and ψ torsion angles for each peptide was created using Biopython[30,31] with the peptide builder by Tien *et al.*[32] The database was initially constructed using 3035 of such regular structures, by selecting possible combinations of ϕ and ψ torsion angles across the entire Ramachandran plot. The ϕ and ψ grid is created with a spacing of 5° in both directions as can be seen in Figure 1. Within the polyproline type II conformation (PPII) and right-handed (RH) α-helical regions, a 2° spacing was used. This grid was manually selected based on the Ramachandran plot by Richardson *et al*. as can be seen in Figure 1.[33] The regions for which regular $(Ala)_{11}$ peptides cannot be created were not populated.

For each conformation of the peptide, the Raman and ROA spectra were calculated using the Cartesian coordinate tensor transfer method (CTTM) by quantum mechanically (QM) calculating the Raman and ROA tensors for computationally manageable molecular fragments of the full peptide and transferring these properties to the full size peptide.[34] Firstly, the geometry of each $(Ala)_{11}$ peptide was partially optimized using the normal mode optimization procedure.[35] By locking the normal modes below 300 cm^{-1} in the optimization, the ϕ and ψ torsion angles are retained, while the modes of spectroscopic interest are fully relaxed. From each $(Ala)_{11}$ peptide the central $(Ala)_5$ fragment was selected and capped with a formyl and an amide functionality at the N- and C-termini, respectively. This molecular fragment was also partially optimized using the normal mode optimization by constraining the normal modes below 300 cm^{-1}, after which the Hessian matrix and property tensors were calculated and transferred to the $(Ala)_{11}$ peptide using CTTM. A fragment was only assigned if all the atom pairs with the corresponding section of the parent $(Ala)_{11}$ structure are within a 0.4 Å distance. For this reason, the 3035 structures that were manually selected as shown in Figure 1, were reduced to 2902 after optimization and pairing with the parent structure (see ESI† S1). The final backscattered Raman and ROA spectra were constructed by using a Lorentzian function for each normal mode with a full width at half height of 20 cm^{-1} to mimic the physical line broadening in the experimental spectra. A Boltzmann distribution factor was used to correct for temperature (300 K) (see ESI† S2). The normal mode optimizations were performed using density function theory (DFT) calculations using the B3PW91/6-31G(d,p) level of theory and the Hessian and Raman and ROA tensors were calculated using B3PW91/6-31++G(d,p). Water was taken into account as the solvent using the self-consistent reaction field model (scrf). For all QM calculations, the Gaussian 09 (rev D.01) program was used.[36]

In chiroptical spectroscopy, numerical measures are commonly used to express the similarity of a calculated spectrum with an experimental spectrum. Many sorts of (di)similarity measures have been defined but a particularly often used one in chiroptical spectroscopy is the generalised

cosine measure.[37–41] This measure is used here to compare a theoretical spectrum *f* from the database with an experimental spectrum *g*, expressed as the similarity S_{fg} of *f* with *g*, calculated using the following formula for a specific wavenumber ($\tilde{\nu}$) range (eqn (1), see ESI† S3.3 for more details).

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

The numerator of this similarity measure normalises the spectral overlap of *f* with *g*, hence S_{fg} conveniently takes values from 0 to 100% for Raman spectra and from -100% to 100% for ROA spectra. The closer the similarity value is to 100%, the more the simulated spectrum resembles the experimental spectrum. Since a ROA spectrum has both positive and negative intensity, a similarity value lower than 0, indicates the mirror spectrum *-f* is more similar to *g* than the original spectrum *f*. By calculating the S_{fg} for all spectra in the database compared to an experimental spectrum, the database spectra that are the most similar to the experiment are identified. The similarity of the computed spectrum for each peptide conformation in the database can thus be assessed with respect to a specific experimental spectrum. By mapping the similarities of the experimental spectrum with the computed spectra, each with known ϕ and ψ angles in a Ramachandran plot, the conformational preferences of the peptide can be visualised and analysed. To create such a similarity map, the Ramachandran plot from -180° to 180° for both ϕ and ψ is divided in 360 by 360 cells where each cell is coloured according to the similarity of the nearest database point with a distance limit of 10. The coloured contour map is shown ranging from 40% in blue to 80% similarity in brown/red. The colour coding range is chosen to be fixed, allowing a direct comparison of the different similarity maps.

Since the wavenumbers in the QM calculated spectra are overestimated due to *inter alia* the harmonic approximation, a scaling factor σ has to be applied to align the two spectra to be able to calculate the similarity with experiment. Here, a global scaling factor of 0.987 was used for the wavenumber range up to 1580 cm^{-1} and an optimized scaling factor between 0.930–0.960 was used for the 1580–1750 cm^{-1} wavenumber range, since this amide I region is known to be even more overestimated compared to the rest of the spectrum (see ESI† S3 for details).[42]

Due to its sensitivity to chirality, ROA spectra are dominated by signals arising from the most chiral structural components of the peptide; signals from flexible side-chains generally cancel out.[8] For this reason, the use of L-alanine as building block for the peptides is justified. A few distinct regions do show signals from the side-chains such as the tryptophan band around 1550 cm^{-1} and signals around 1480 cm^{-1} arising from CH_2 and CH_3 deformations.[8,43,44] The use of poly-L-alanine models will however reduce the computational cost considerably as the simulation of the spectra for these models does not require conformational averaging of the side-chains and alanine is the smallest chiral amino acid. The similarities can still be calculated for the full range 450-1800 cm^{-1}, since none of the peptides in this study contains tryptophan. As the region below 300 cm^{-1} is locked in the geometry optimization and because the region below 450 cm^{-1} shows large intensities, the region below 450 cm^{-1} is not included in the similarity calculation. As illustrated in ESI† S4.2, the large intensity in this region would dominate the similarity values, while the ROA regions of specific interest are mainly located within 850-1700 cm^{-1}. For the spectra acquired in chloroform or dichloroacetic acid, the similarities are calculated for 800-1800 cm^{-1}. To account for small local shifts in the spectra, triangular weighting is applied to the calculated and experimental spectra before calculating the similarity, where for each spectral point a contribution is added from the neighbouring points.[40] The effect of the triangular weighting is depicted in Figure S 9 of the ESI†. The similarities generally increase, while the main secondary structure assignments remain the same.

Materials and spectral acquisition

The peptide Ac-AAKAAAAKAAAAKAAAAKAGY-NH_2 (AK21) was synthesized by GeneCust at 95% purity and used without further purification. The peptide Ac-XXAAAAAAAOO-NH_2 (XAO; "X" denotes diaminobutyric acid and "O" ornithine) was synthesized and kindly donated by the Institute of Organic and Biochemistry at the Academy of Sciences of the Czech Republic and was used without further purification.

Poly-L-alanine, (PLA MW: ~21400 Da) Poly-L-glutamic acid (PLGA MW: ~100000 Da), poly-(β-benzyl)-L-aspartate (PBLA MW: ~26600 Da) and poly-(γ-benzyl)-L-glutamate (PBLG MW: ~28000 Da) were obtained from Sigma-Aldrich and used without further purification. The peptides were dissolved in MilliQ water (AK21, XAO, PLGA), HPLC grade chloroform ($CHCl_3$, Sigma, PBLA, PBLG) or dichloroacetic acid (DCA, Sigma, PLA) in concentrations between 50 and 75 mg/mL. For PLGA, the pH was adjusted to 4.6 in order to ensure α-helical conformation. For the hydrogen/deuterium (H/D) exchange studies, the peptides were dissolved in high purity D_2O (Sigma), lyophilized and re-dissolved in D_2O in order to insure full H/D exchange.

The Raman and ROA spectra were measured at ambient conditions using the previously described ChiralRAMAN scattered circular polarization (SCP) ROA instrument (BioTools, Inc).[45] The Raman spectra are displayed as the circular intensity sums ($I_R + I_L$) and the ROA spectra as the circular intensity differences ($I_R - I_L$) with I_R and I_L denoting the scattered Raman intensities with right- and left-circular polarization, respectively. The instrument excitation wavelength was 532 nm; laser power at the source was in the range of 200-800 mW, depending on the sample; spectral resolution of 7 cm^{-1}; and an acquisition time of 12-35 hours. Solvent spectra were subtracted from the Raman spectra after which the baseline correction procedure by Boelens *et al.* was applied.[46] The ROA spectra were smoothed using a 2rd order, 5-point Savitzky-Golay filter.

Results and discussion

The experimental spectra of the three peptides AK21, XAO and PBLA were interpreted using the new spectral database. In

Figure 2, the similarity maps of these three examples are shown with the region with the highest similarities indicated by a red rectangle; the spectra with the highest intensity (red/brown) in this region were averaged (see ESI† S4) and shown at the bottom of Figure 2. The similarity maps assign the conformational preference of each peptide in a completely different region of the Ramachandran plot: AK21 shows the highest similarity with the database spectra in the RH α-helical region of the Ramachandran space, the XAO peptide shows a conformational preference for PPII structure and PBLA displays the highest similarities in the left-handed (LH) α-helical region. For each peptide this is in agreement with other experimental techniques that were previously used to classify the secondary structure of these particular peptides. AK21 was for example shown to be helical by UV circular dichroism (CD) and was used before as a model for the assignment of ROA patterns for α-helical structure.[47] Karlson *et al.* showed that PBLA adopts a LH helix in chloroform solution by using optical rotatory dispersion.[48,49] The structural propensities of the XAO peptide have been the matter of debate in scientific literature as reviewed by Adzhubei *et al.*, but the model of a flexible structure not restricted by a regular pattern of hydrogen bonds and capable of fast conformational changes within the PPII region in ϕ/ψ space seems to be corroborated by all experiments.[50–52] While SAXS shows that the radius of gyration is much smaller than it would be for a fully extended structure[53], spectroscopic methods indicate a high content of PPII conformation.[54,55] Although the ROA spectrum of a very similar peptide Ac-OO-A_7-OO-NH_2 has been published before[11,55], the ROA spectrum of the XAO peptide is, to the best knowledge of the authors, reported and assigned here for the first time. Such assignments are relevant, as spectroscopic methods seem to be specifically sensitive to this kind of secondary structure, which is an important structural element of both globular as well as the important class of intrinsically disordered proteins.[50,55]

Some regions in the Ramachandran plots also show similarity with the experiment because of coincidental spectral overlap between the database spectra and the experiment. Since ROA spectra mainly contain signals from the amide and skeletal stretch vibrational modes, intensity will appear in the same spectral regions and there could hence be considerable coincidental overlap. Visual inspection of the spectra immediately reveals these coincidental overlaps; these spectra do not resemble the experimental spectra, but show relatively high similarity values (see ESI† S5). For example, if a very intense calculated band coincides with positive intensity in the experimental spectrum, the S_{fg} can be high. Nonetheless, the S_{fg} remains a good similarity measure for the reasons described above. Furthermore, as AK21 and XAO are likely to be flexible in solution, other secondary structure elements contribute to the experimental ROA spectrum. The region with the highest similarity in the Ramachandran graphs corresponds to the most important structural element.

Detailed assignments in the α-helical region

Besides being capable of delineating the spectra in different conformational regions of the Ramachandran surface, the similarity maps in Figure 2 show there is a very high similarity between the experimental spectrum and the database spectra in a specific region of the Ramachandran surface, but the similarity rapidly decreases further away from that region. This indicates that ROA is very sensitive to the backbone conformation and that the database can be used to make very detailed structural assignments. In Figure 3, the similarity map of PLA is compared to that of PBLG and PLGA. These peptides

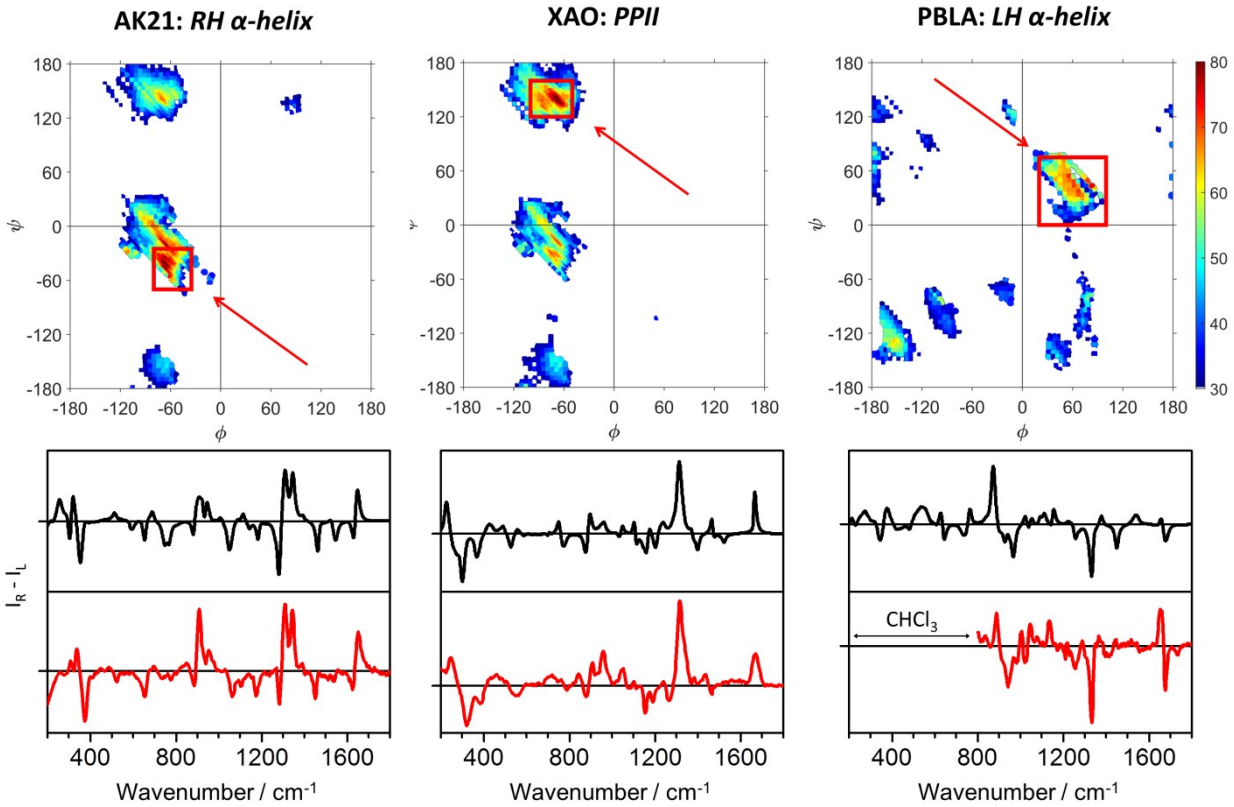

Figure 2 Similarity maps of AK21, XAO and PBLA (top). Comparison (bottom) of the experimental ROA spectrum (red) to the averaged database spectrum (black).The region with the highest similarity is indicated by the red rectangle and arrow. The low-wavenumber region of PBLA is shown while this region in the experimental spectrum is hampered by artefacts arising from the chloroform ($CHCl_3$) solvent. See details of the Ramachandran maps and the spectra contributing to the average in ESI† S4.

are known to adopt α-helical conformations in solution under specific conditions (*vide infra*), but the experimental ROA spectra show distinctly different patterns.

The similarity maps of PBLG and PLGA show slightly lower intensity compared to AK21 and PLA, most likely because the couplet at 1090-1140 cm^{-1} in the experimental spectra of PBLG and PLGA is more dispersed in the simulated spectra, hence lowering the overlap and thus the S_{fg}.

While PBLG and PLGA are also correctly assigned to be α-helical, the similarity maps highlight a slightly different region of the Ramachandran surface compared to AK21 and PLA. Traditionally, a few important ROA spectral characteristics were assigned to α-helical structure. The most prominent are the -/+ couplet, negative at low and positive at high wavenumbers, centred at 1650 cm^{-1} in the amide I region and two positive ROA bands around 1300 cm^{-1} and 1340 cm^{-1} in the amide III region.[8,56] Also positive intensity in the range 870-950 cm^{-1} and a -/+ couplet centred at 1100 cm^{-1} have been tentatively put as ROA characteristics of α-helical structure.[8] Especially the two positive bands in the amide III region, for example in the PLA and AK21 spectra, have been of interest in the field of ROA.[8,14,57,58] Based on the comparison with crystal structures and H/D exchange experiments (*vide infra*), the ROA band around 1300 cm^{-1} was assigned to unhydrated α-helical structure in a hydrophobic environment and the band around 1340-1345 cm^{-1} was assigned to hydrated α-helical conformations.[8,56,59] In hydrated conformations, the backbone torsion angles ϕ_{i+1} and ψ_i take mean values of -66° and -41°, respectively, tilting the peptide carbonyls and allowing hydrogen bonding with water.[60] For the unhydrated conformation, mean torsion angles ϕ_{i+1} and ψ_i are -59°

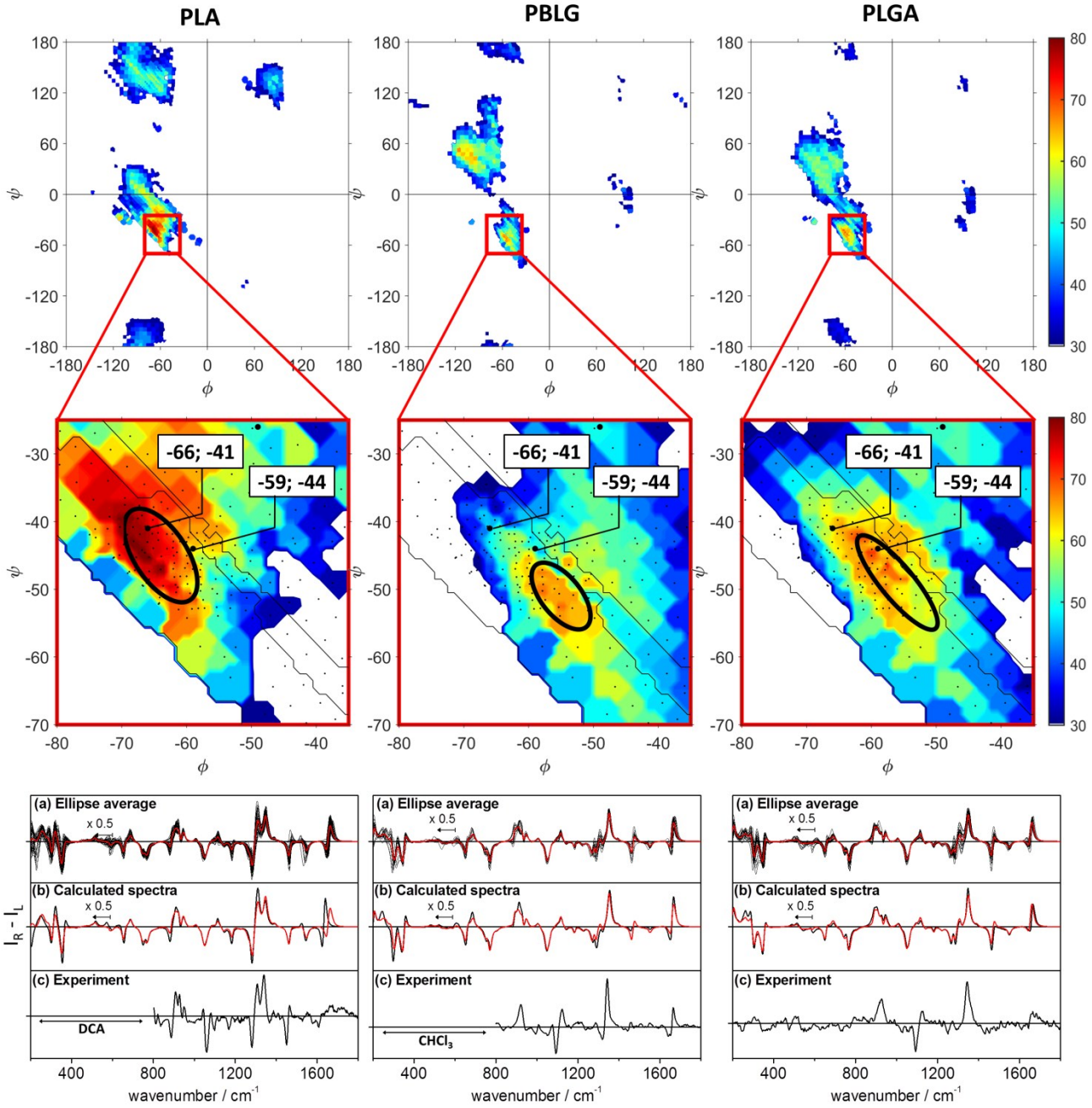

Figure 3 Detailed structural assignments of PLA, PBLG and PLGA: Similarity maps (top) with details of the highest intensity regions (middle) and the comparison of (a) the spectra within the bold black ellipse in the maps, (b, in black) the spectrum with the highest S_{fg}, (b, in red) the average of these spectra with (c) the experimental spectrum, (bottom). Since PLA and PBLG were measured in DCA and $CHCl_3$, respectively, the region below 800 cm^{-1} is not shown in the experimental spectra because of artefacts arising from these solvents. The black contour lines in the zoomed similarity maps indicate the α-helix and 3_{10} helix regions in the database (see ESI† S6)

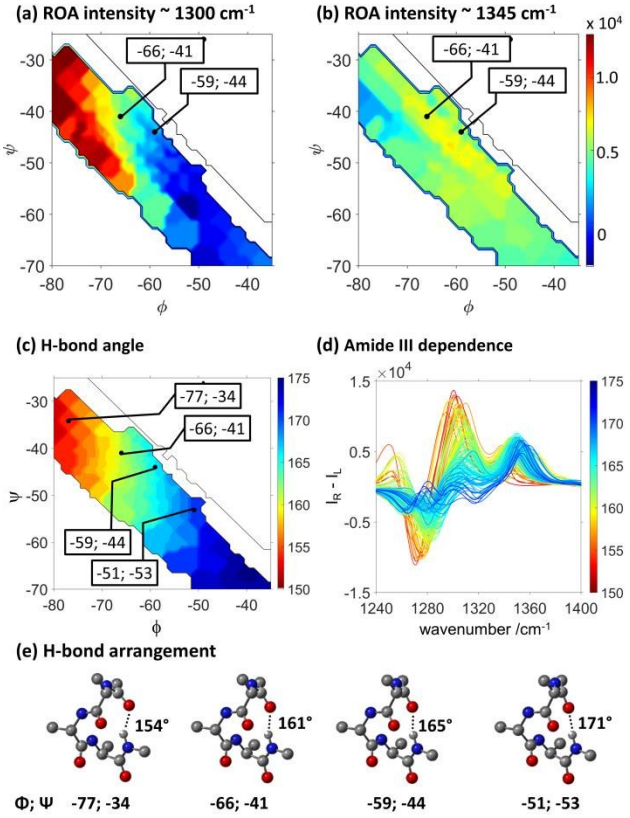

Figure 4 (a) ROA peak intensity in the spectral region 1290-1325 cm^{-1} and (b) in the region 1330-1365 cm^{-1} shown with the same colour coding. (c) Tilting of the C=O group compared to the H-N group in the hydrogen bond. The angle between the CO and NH vectors is given in the contour plot from a 150° angle (tilted, red) to a 175° angle (linear, blue). (d) The database spectra in this Ramachandran region and coloured according to H-bond angles depicted in (c). (e) For four representative data points in the database the four central residues making up an α-helical hydrogen bond are shown at the bottom (see the respective spectra in the ESI† Figure S 17). The black contour lines indicate the α-helix region (see ESI† S6).

and -44°, respectively, resulting in a more linear arrangement of the C=O···H-N bond (see Figure 4 (e)). Recently however, Yamamoto *et al.* reported that the conformational difference between hydrated and unhydrated α-helix conformations does not account for the intensity ratio between the two bands in the amide III region of α-helical PLA in DCA solution.[58] PLA adopts an α-helical conformation in DCA solution and shows two prominent positive ROA bands at 1304 cm^{-1} and 1338 cm^{-1}, with the latter band with a slightly higher relative intensity (see Figure 3). When PLA is measured in $CHCl_3$/DCA in a 7/3 ratio (v/v), the intensity of the 1300 cm^{-1} band increases relative to the 1338 cm^{-1} band. Yamamoto *et al.* concluded from their DFT calculations that the influence of the conformation is minor compared to the change of the dielectric constant of the solvent. The database spectra however show there is a very distinct effect of slight changes in the conformation on the relative intensity of the two α-helical bands.

The results for AK21, PLA, PBLG and PLGA in Figure 2 and Figure 3 discussed above demonstrate that the classical assignment of the hydrated compared to unhydrated α-helix to the two positive amide III bands is ambiguous. The similarity maps of PLA and AK21 show that spectra in the Ramachandran region around -66°; -41° such as in hydrated helical structure fit best with experiment, while the ROA spectra in this region in the database show both positive bands around 1300 cm^{-1} and 1340 cm^{-1}. The similarity maps of PLGA and PBLG, however, show a conformational preference closer to conformations such as in an unhydrated helix (-59°; -44°) or even further to the bottom-right of the α-helical region, while the spectrum only shows the amide III band around 1340 cm^{-1}, which traditionally was assigned to hydrated conformations.[56] The neutral side-chains of PBLG and PLGA stabilise regular α-helical structure in solution,[61,62] hence indicating the 1340 cm^{-1} band to rather be a marker of unhydrated or regular α-helical structure. For PLA the ellipse delineating the database spectra that are averaged as shown in Figure 3 is chosen slightly more to the bottom right of the α-helical region compared to the AK21 ellipse (see ESI† Figure S 9). As a result, the intensity ratio of the 1300/1340 cm^{-1} bands compare well with the experimental ratios (see Figure 2 and Figure 3).

Using the database, the conformational sensitivity of the intensity ratio of these two amide III bands can be explored in more detail. In Figure 4 (a) and (b), the maximal peak intensity of the two bands is shown for the α-helical structures in the database. For the band around 1300 cm^{-1}, the intensity starts around zero at the bottom-right of the α-helical region and increases continuously to the top-left of the α-helical region; this band is very sensitive to conformational changes. The band around 1340 cm^{-1} is more conservative as it shows much less variation for different conformations. This is also apparent in the spectra that contribute to the averages in Figure 3. Depending on the torsion angles, the hydrogen bond goes from a linear to a more tilted arrangement, and the intensity of the ROA band around 1300 cm^{-1} goes from slightly negative to twice the intensity of the 1340 cm^{-1} band, as can be seen in Figure 4.

As a result, the intensity ratio of the two bands is extremely sensitive to slight conformational changes. In peptides and proteins in solution, structural heterogeneity and flexibility in α-helical sequences, make that the torsion angles explore a range of angles within the α-helical region of the Ramachandran surface resulting in an averaged spectral profile.

Deuterated AK21 and XAO

An important motivation why the band around 1340-1345 cm^{-1} was initially assigned to a hydrated form of α-helix, was that in ROA measurements of filamentous bacteriophages Pf1 and M13 in D_2O this band disappears, while it is prominently present in H_2O solution.[59] In H_2O solution, the two α-helical bands are present for both Pf1 and M13, whereas in D_2O positive intensity is retained only near 1300 cm^{-1}. The disappearing of the band was considered to indicate the exposure to the solvent of the molecular part involved in the vibrational mode around 1340 cm^{-1}, *i.e.* the amide N-H proton in a hydrated conformation, hence allowing H/D exchange. However, the extended amide III region (1230-1340 cm^{-1})[63] involves in-phase combination of N-H in-plane deformations with C-N stretch vibrations. It is because of the coupling between N-H and Cα-H deformations that this region is so sensitive to geometry changes. Jacob *et al.* used a localised mode procedure to show that the 1340 cm^{-1} band

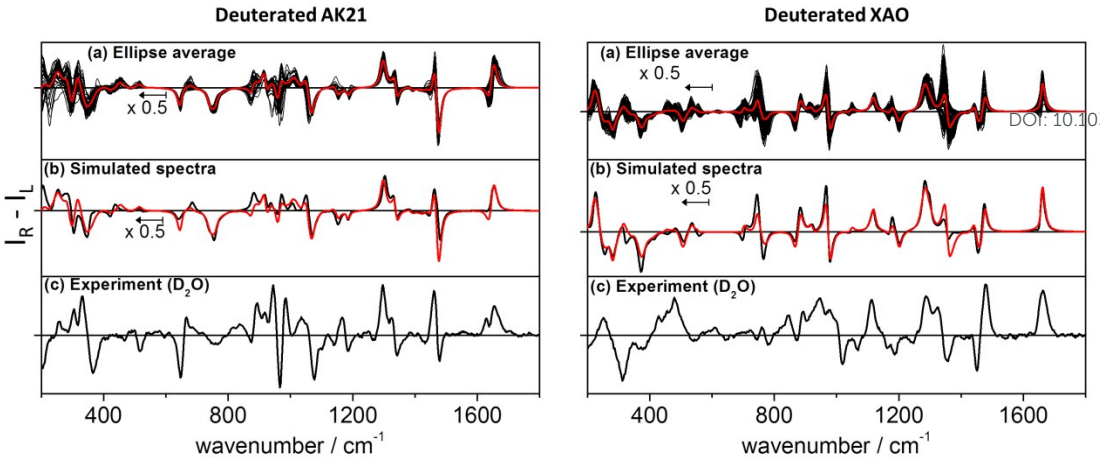

Figure 5 Comparison of the ROA spectra of H/D exchanged AK21 (left) and XAO (right). (a) Comparison of the spectra in the same region identified in Figure 2 and (b,red) the average of these spectra and (b, black) the spectrum with the highest S_{fg} to (c) the experimental spectrum in D_2O.

arises from mainly Cα-H deformation in the direction of the Cα-N bond, called a CαH bend (I) mode, while in the 1300 cm^{-1} band, Cα-H bending is perpendicular to the Cα-N bond, called a CαH bend (II).[64] For this reason, exchanging the N-H proton by a deuteron, has a large effect on the entire amide III region and not only the CαH bend (I) mode around 1340 cm^{-1}.

To study the effect of the amide H/D exchange, the model conformations identified in Figure 2 for the AK21 and the XAO peptides in H_2O were used to calculate the ROA spectra with all the amide protons replaced by deuterons. The respective similarity maps are given in ESI† Figure S 9 and also show the same secondary structure assignment: α-helical structure for AK21 and PPII conformation for XAO. The similarity maps however show lower values compared to the non-deuterated analogues and are more sensitive to the triangular weighting.

In Figure 5, the simulated spectra are compared to the respective experimental ROA spectra measured in D_2O. The experimental spectra collected for the H/D exchanged samples measured in D_2O are remarkably different from the spectra in H_2O. For example, the 1230–1340 cm^{-1} region for both samples is completely altered compared to the respective spectra in H_2O. Interestingly, this region is very comparable for both AK21 (N-D) and XAO (N-D) in D_2O, while this wavenumber region in H_2O holds the most structural information. The modelled spectra again reproduce the experimental patterns quite well. The +/- pattern at 1326 cm^{-1} and 1358 cm^{-1} in the experimental spectrum of XAO (N-D) is however calculated at too high wavenumbers, thereby decreasing the overlap of experiment and theory. An important observation is that the 1420-1520 cm^{-1} region is entirely different for both peptides, indicating that this is a key region for structural elucidation using H/D exchange in ROA. The ROA spectrum of disordered poly-L-lysine in D_2O shows a -/+ couplet similar to that from the XAO peptide, while a -/+/- signature like that of AK21 in D_2O is observed for poly-L-lysine in an α-helical conformation.[65] The -/+/- signatures of the latter two peptides however show different ratio's that might again stem from differences in the distribution of the dihedral angles in the α-helical region.

In H_2O, this spectral region arises from CH_2 and CH_3 deformations. A dynamic visualisation of the normal modes of the deuterated models using the localised mode procedure implemented in the Pyvib2 program[66], reveals that the CH_3 deformations couple with the Cα-N and to lesser extent with the in-phase N-D in plane bending (amide II' mode), giving rise to a conformation sensitive region. In H_2O, the N-H is strongly involved in the amide III and amide II modes. The deuteration decouples the N-D from these modes and gives rise to amide II' (the prime signifies the N-D deuteration) modes.[67,68] The deuteration seems to have a large effect on most of the ROA spectrum. The amide I and amide I' are quite similar, only for AK21 the small negative lobe becomes positive upon deuteration. The backbone skeletal stretch region (870–1150 cm^{-1}), mainly involving the Cα-C, Cα-Cβ and Cα-N stretches, is also altered after the deuteration. This wavenumber region is very complex, making the comparison between the simulated and the experimental spectra difficult, but it indicates the richness of this region in structural information. This is for example visible in the simulated ROA spectra in Figure 5 (a) of AK21, which seem very sensitive to small conformational changes. These results indicate the high value of using H/D exchange in ROA analyses of proteins and peptides, providing a supplementary source of information next to the measurements without the exchange. The effect of H/D exchange on the ROA patterns of proteins has been studied in the past, however careful experimental protocols must be used to achieve full H/D exchange for full-size proteins.[8,59,65,69,70] If a protein is dissolved in D_2O, only the solvent-accessible N-H protons will be rapidly exchanged and the spectral properties of these molecular regions will change. This could be very informative to identify the solvent-exposed parts, however the partial exchange immediately complicates the analysis. If, for example, only the amide protons in the solvent exposed part of an α-helical sequence are exchanged and the hydrophobic part remains unaltered, the effect on the spectral patterns is complex, which ,for example, is the case in the ROA spectra of human serum albumin (HSA), Pf1 and M13 measured in D_2O.[8,59] This is illustrated in section S8 of the ESI† by partially exchanging the N-H protons of database structures by deuterons. If four hydrogens are exchanged on one side of an α-helical model structure, most of the ROA spectrum remains unchanged, while the intensity around 1340 cm^{-1} is reduced by ~ 50% and the intensity of the band around 1300 cm^{-1} is retained. This is in accordance with the

observation of the large decrease of the ROA intensity around 1340 cm^{-1} in Pf1, M13 and HSA when measured in D_2O.[8,59] Also the spectral change of AK21 upon full deuteration supports this as intensity around 1300 cm^{-1} is present both with and without the deuteration, while the ROA intensity around 1340 cm^{-1} disappears upon full H/D exchange. The disappearing of the 1340 cm^{-1} α-helical band in D_2O does hence not support this band being a marker for hydrated helices.

The 1300/1340 cm^{-1} ratio: Interpretation of protein ROA

To conclude the results of the previous two sections, whereas the ratio of the two α-helical bands in the amide III region is very dependent on the conformation as shown by the database spectra, the previous assignment of the 1300 cm^{-1} and the 1340 cm^{-1} bands being markers of unhydrated and hydrated α-helix, respectively, is invalidated. The database spectra even indicate that the opposite is more likely: That the 1340 cm^{-1} band is a conservative indicator of regular α-helical structure with a geometry such as in unhydrated conformations and the 1300 cm^{-1} band is an indicator of conformations allowing hydrated helices. This is supported by the experimental ROA spectra of for example Fd[21], Pf1[59], M13[59], rabbit aldolase[8], rabbit calcyclin[21], HSA[8] and bovine serum albumin[71] (BSA), which all show a dominant positive band around 1340-1345 cm^{-1} and which all have structures with long and regular α-helical sequences. The proteins HSA and BSA, for example, have long hydrophobic pockets as binding sites in their mainly α-helical structure, which supports the assignment of the ROA band around 1340 cm^{-1} to regular or hydrophobic/unhydrated helical structure. In aqueous solution, these proteins are at least partially exposed to the water, hence also positive ROA intensity around 1300 cm^{-1} is observed. The database RH α-helical structures of which the backbone carbonyls tilt outwards from the structure, therefore allowing hydrogen bonding with water or side-chains,[60] show dominant intensity around 1300 cm^{-1} as demonstrated in Figure 4 and Figure S 16. Also the database spectra in the RH 3_{10} helical region between α-helix and towards structures with dihedral angles -71°; -18° have dominant positive intensity around 1300-1315 cm^{-1} as shown in Figure S 13 and Figure S 14. In smaller proteins with shorter α-helical portions in their crystal structures such as insulin[8,70], hen egg white lysozyme[69], human lysozyme[47], bovine α-lactalbumin[8,72,73] and equine lysozyme[73] a strong positive band around 1300 cm^{-1} is distinctly observed.[8,70,72] In the first three proteins of these proteins, the intensity around 1300 cm^{-1} is even distinctly higher than that around 1340 cm^{-1}. Also the spectra of Fd, Pf1 and M13 show a separate band around 1300 cm^{-1}. Both the database spectra as well as the seven latter mentioned experimental ROA spectra suggest that positive intensity around 1300 cm^{-1}, which seems very dependent on the backbone conformation (*vide supra*), is associated with 3_{10} helix and α-helical conformations allowing hydrogen bonding with water. The solution structural ensemble of human insulin as determined by NMR (PDB id. 2mvc, low pH) and the crystal structure of bovine insulin (PDB id. 2a3g), for example, show mainly backbone angles corresponding to conformations with carbonyls tilting outwards from the helix and angles between α-helical and 3_{10} structure (see ESI† S9). This might explain that the ROA spectrum of bovine and

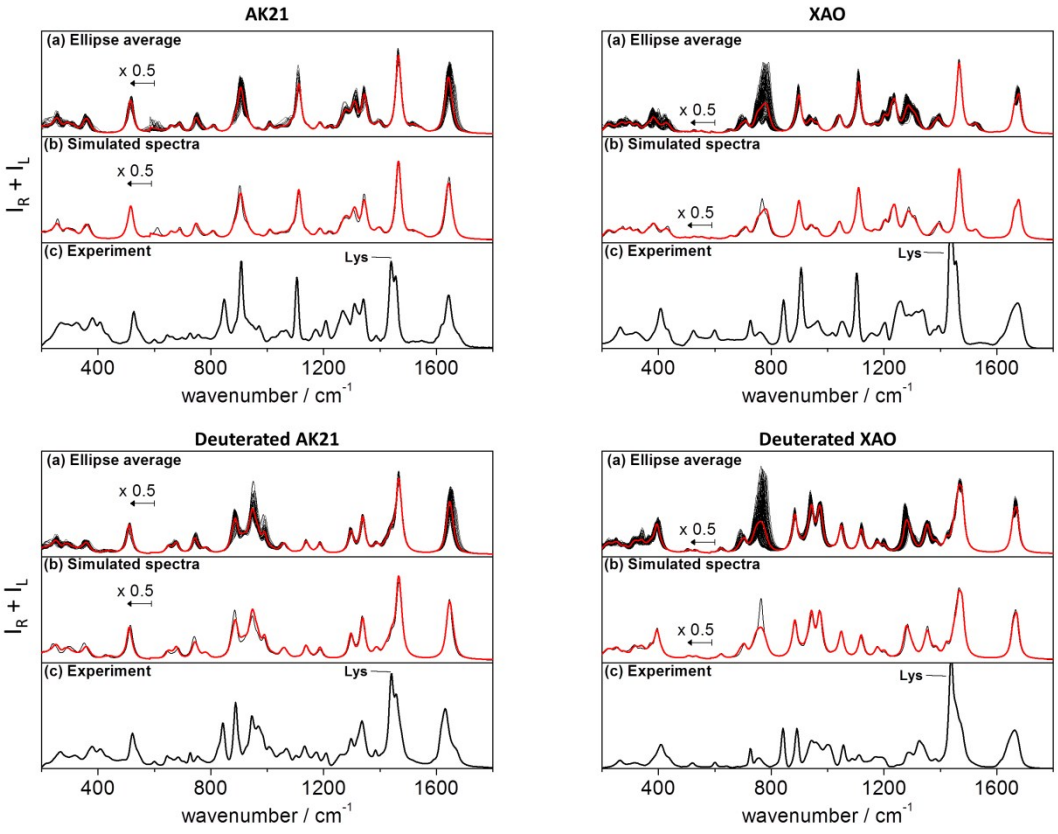

Figure 6 Raman spectra of the AK21 (left) and the XAO (right) peptides in H_2O (top) and in D_2O (bottom): (a) average of the Raman spectra of the model structures identified in Figure 2; (b) the average spectrum (red) and the spectrum with the highest S_{fg} in black; (c) the experimental spectrum.

porcine insulin at low pH show dominant positive intensity around 1307 cm^{-1} with almost no contribution around 1340 cm^{-1}.[8,29,70] Yamamoto *et al.* also noted that 3_{10} helical structure contributes significantly to the positive amide III band in the experimental spectrum of insulin.[29] The interpretation of differences in the 1300/1340 cm^{-1} intensity ratio of native α-lactalbumin and equine, hen and human lysozyme has received considerable attention in scientific literature as these proteins have very similar folds but show different 1300/1340 cm^{-1} intensity ratios.[8,14,72,73] Hen and human lysozyme have a dominant 1300 cm^{-1} band while for α-lactalbumin and equine lysozyme the 1340 cm^{-1} band is relatively more intense. The presence of the 1340 cm^{-1} band indicates there is at least a proportion of regular α-helical structure in all these proteins. Based on the discussion above, the dominant intensity around 1300 cm^{-1} in the hen and human lysozyme spectra could suggest these proteins have a higher content of 3_{10} helix in solution or a different distribution of the dihedral angles in the helical portions. This in turn might be caused by a differential stability or hydration of one of the four helices in the lysozyme/α-lactalbumin folds or of the C-terminal 3_{10} helix, for example. A short stretch of 3_{10} helix or α-helix that is not present could cause the 1300 cm^{-1} band to be of lower intensity. Caution must however be exerted in making such assignments as the database spectra only represent the conformational sensitivity of the spectral patterns. The database does not directly include the effect of explicit hydration which has only limitedly been studied theoretically[74] nor is the effect of different amino acid compositions included, which has been proposed by Kessler *et al.* to affect the intensity ratio of the two α-helical bands[14]. Kessler and co-workers suggested that the difference in the intensity ratio might be explained by a difference in the content of aromatic residues as these couple to main chain Cα-H bending.[14] However, the ROA spectrum of human lysozyme reported in their analysis is notably different from the spectrum of native human lysozyme published by Blanch *et al*.[47] The experimental spectrum reported by Kessler *et al.* and compared to their DFT calculations is much more similar to that of equine lysozyme, with the 1340 cm^{-1} band dominating over the 1300 cm^{-1} band. It is currently unclear why the experimental spectra differ. Kessler *et al.* used an aqueous solution without the use of a buffer of a commercial source of recombinant human lysozyme expressed in rice (*Oryza sativa*). In the study by Blanch *et al.* human lysozyme was purified from human milk and dissolved in an acetate buffer.[47] The recombinant form might be different from the wild-type variant. However, if the sequences of both variants are identical, the 1300/1340 cm^{-1} ratio cannot be described by a differential content of (aromatic) residues, rather by a conformational difference or a combined effect of a conformational change in a specific molecular part containing an aromatic residue, for example. It must be noted that α-lactalbumin and the lysozymes are very sensitive to the experimental conditions such as temperature and pH in solution[72,75,76], which is also important to consider when comparing the spectral patterns with structures deposited in the PDB. Identifying the differences in the ROA spectra of these proteins might be very valuable as they have different folding mechanisms and some of them, such as α-lactalbumin and equine lysozyme, can form stable molten globule states, while other lysozymes do not show such behaviour despite their homologous structure.[77–79] The database provides important conformational information that could aid in the understanding of the spectral patterns of said proteins in the future.

Raman spectra of AK21 and XAO

As shown above, ROA is extremely sensitive to protein conformation. An important reason why ROA is better at capturing the solution structure and conformation of proteins compared to conventional Raman spectroscopy, is that ROA mainly shows signals from the backbone conformation of the protein. Due to the chiroptical character, ROA signals of flexible side-chains of the amino acid residues generally cancel out, while Raman spectra on the other hand are dominated by these signals. However, the peptides AK21 and XAO mainly consist of alanine residues, so their Raman spectra can suitably be described by the database containing alanine models as well as the H/D exchanged models.

In Figure 6 the database Raman spectra of the model structures identified by the similarity maps in Figure 6 are compared to the experimental Raman spectra of AK21 and XAO. Besides a few lysine bands that are not included in the database, most spectral bands are very well reproduced (see also the Raman spectra assigned to PLA in ESI† Figure S 24). This implicates the use of the database to study Raman spectral characteristics not only of alanine peptides, but also of for example the amide III and amide I region that are informative on the conformation of proteins in solution. For AK21 both in H_2O and in D_2O, the experimental and theoretical patterns match very well. Upon deuteration mainly the amide III and skeletal stretch region change. The region 800-1000 cm^{-1} for the deuterated AK21 shows the most variation with geometric changes. For XAO, the comparison is also very good. Here the region 700-800 cm^{-1} shows the most variation with varying structure. The amide I region does not change much because of the deuteration, which is to be expected as these modes mainly involve carbonyl stretching with little or no contribution from the amide proton or deuteron. The amide I and amide I' of XAO are asymmetric both in experiment and simulation. The spectrum with the highest similarity also shows this asymmetry, thus the asymmetry does not stem from different conformational families, but rather from dispersion of the normal modes.[2]

Conclusions

A large database of computed Raman and Raman optical activity (ROA) spectra has been constructed as a tool to analyse the experimental spectra of peptides and proteins. For six model peptides with left- or right-handed α-helical and PPII structure (AK21, XAO, PBLG, PLA, PBLA and PLGA) the database assigns the secondary structure correctly by creating a similarity map on the Ramachandran plot that compares the experimental ROA spectrum to the database spectra. The

experimental spectra are assigned to very specific regions of the Ramachandran plot, which shows the high sensitivity of ROA to the backbone conformation of peptides and proteins; the database spectra change distinctly with a slight change in the conformation. The spectral assignments of the amide III region of α-helical proteins have been the matter of debate in scientific literature, but the database shows that the positive band around 1300 cm^{-1} is specifically sensitive to the conformation of the helix and thus the orientation of the intramolecular hydrogen bonds and hydration. Also 3_{10} helix with dihedral angles close to α-helix contributes to intensity around 1300-1315 cm^{-1}. The band around 1340 cm^{-1} is much more conservative for α-helical structure and seems a reliable marker for regular RH α-helical structure.

Averaging of the database spectra in specific regions of the Ramachandran plot, gives spectra that compare very well with the experiment and thus gives representative spectra for specific secondary structure types. Both the analysis of specific vibrational modes or spectral regions and the creation of average spectra of different secondary structure types are informative ways of studying the spectra of proteins by using the database. Although the Ramachandran mapping of peptides with a specific conformational propensity works very well, the spectra of proteins contain contributions from all the different secondary structure elements or domains of that protein. Therefore, we are currently developing suitable approaches to use this database to study the solution structure of proteins in detail as it was shown that analysing the spectral trends in the database provides very valuable information on the spectral patterns not only of peptides but also of proteins. In short, this new tool will be valuable to understand the solution structure of peptides and proteins in future Raman and ROA studies as it provides the relation between the secondary structure elements and their spectral components. This is important and timely as for example the ROA spectra of molten globule states of proteins and those of intrinsically disordered proteins show rich ROA patterns which are difficult to assign in detail.[55,72,73,80] Yet, the detailed interpretation of such ROA spectra might complement structural information accessible by other techniques. These states and types of proteins are being studied intensively out of general interest to understand protein folding and misfolding pathways and to understand the molecular origins of neurodegenerative diseases such as Parkinson's or Alzheimer's, for example.[78,81–83]

Acknowledgements

The authors acknowledge the financial support provided by the Flemish Community, the University of Antwerp (BOF-NOI) for the pre-doctoral scholarship of C.M., the University of Ghent (IOF Advanced TT) for the purchase of the ChiralRAMAN spectrometer and the Flemish Super Computing Centre (VSC) financed by the Flemish government and the Hercules Foundation. The authors thank Dr. Vincent Liégeois and Prof. Dr. Patrick Bultinck for useful discussions and Dr. Jaroslav Šebestík for kindly donating the XAO peptide sample.

Notes and references

1 N. P. Cowieson, B. Kobe and J. L. Martin, *Curr. Opin. Struct. Biol.*, 2008, **18**, 617–622.
2 R. Schweitzer-Stenner, *Vib. Spectrosc.*, 2006, **42**, 98–117.
3 C. K. Fisher and C. M. Stultz, *Curr. Opin. Struct. Biol.*, 2011, **21**, 426–431.
4 Ē. Kupče, in *Modern NMR Methodology*, eds. H. Henrike and M. Stephen, 2011, pp. 71–96.
5 A. Barth, *Biochim. Biophys. Acta - Bioenerg.*, 2007, **1767**, 1073–1101.
6 A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk and M. Baranska, *J. Raman Spectrosc.*, 2013, **44**, 1061–1076.
7 J. Kong and S. Yu, *Acta Biochim. Biophys. Sin. (Shanghai).*, 2007, **39**, 549–59.
8 L. D. Barron, L. Hecht, E. W. Blanch and A. F. Bell, *Prog. Biophys. Mol. Biol.*, 2000, **73**, 1–49.
9 L. D. Barron, *Biomed. Spectrosc. Imaging*, 2015, **4**, 223–253.
10 P. Mukhopadhyay, G. Zuber and D. N. Beratan, *Biophys. J.*, 2008, **95**, 5574–5586.
11 I. H. McColl, E. W. Blanch, L. Hecht, N. R. Kallenbach and L. D. Barron, *J. Am. Chem. Soc.*, 2004, **126**, 5076–5077.
12 V. Parchaňský, J. Kapitán, J. Kaminský, J. Šebestík and P. Bouř, *J. Phys. Chem. Lett.*, 2013, **4**, 2763–2768.
13 F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *Structure*, 2005, **13**, 1409–1419.
14 J. Kessler, J. Kapitán and P. Bouř, *J. Phys. Chem. Lett.*, 2015, **6**, 3314–3319.
15 G. Thiagarajan, E. Widjaja, J. H. Heo, J. K. Cheung, B. Wabuyele, X. Mou and M. Shameem, *J. Raman Spectrosc.*, 2015, **46**, 531–536.
16 C. Johannessen, R. Pendrill, G. Widmalm, L. Hecht and L. D. Barron, *Angew. Chemie Int. Ed.*, 2011, **50**, 5349–5351.
17 F. Zhu, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2005, **127**, 6142–6143.
18 C. Mensch, R. Pendrill, G. Widmalm and C. Johannessen, *ChemPhysChem*, 2014, **15**, 2252–2254.
19 N. R. Yaffe, A. Almond and E. W. Blanch, *J. Am. Chem. Soc.*, 2010, **132**, 10654–10655.
20 E. W. Blanch, L. Hecht and L. D. Barron, *Methods*, 2003, **29**, 196–209.
21 E. W. Blanch, L. Hecht, C. D. Syme, V. Volpetti, G. P. Lomonossoff, K. Nielsen and L. D. Barron, *J. Gen. Virol.*, 2002, **83**, 2593–2600.
22 E. W. Blanch, D. J. Robinson, L. Hecht, C. D. Syme, K. Nielsen and L. D. Barron, *J. Gen. Virol.*, 2002, **83**, 241–246.
23 F. Zhu, G. E. Tranter, N. W. Isaacs, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 2006, **363**, 19–26.
24 O. Quinet and B. Champagne, *J. Chem. Phys.*, 2001, **115**, 6293–6299.
25 V. Liégeois, K. Ruud and B. Champagne, *J. Chem. Phys.*, 2007, **127**, 204105–204105.
26 O. Quinet, V. Liégeois and B. Champagne, *J. Chem. Theory Comput.*, 2005, **1**, 444–452.
27 A. J. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen and S. Coriani, *J. Chem. Phys.*, 2008, **129**, 214108.
28 S. Luber and M. Reiher, *J. Phys. Chem. B*, 2010, **114**, 1057–

1063.
29 S. Yamamoto, J. Kaminský and P. Bouř, *Anal. Chem.*, 2012, **84**, 2440–2451.
30 P. J. A. Cock, T. Antao, J. T. Chang, B. A. Chapman, C. J. Cox, A. Dalke, I. Friedberg, T. Hamelryck, F. Kauff, B. Wilczynski and M. J. L. de Hoon, *Bioinformatics*, 2009, **25**, 1422–1423.
31 T. Hamelryck and B. Manderick, *Bioinformatics*, 2003, **19**, 2308–2310.
32 M. Z. Tien, D. K. Sydykova, A. G. Meyer and C. O. Wilke, *PeerJ*, 2013, **1**, 1:e80; DOI 10.7717/peerj.80.
33 J. S. Richardson, D. A. Keedy and D. C. Richardson, in *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*, eds. M. Bansai and N. Srinivasan, World Scientific, Singapore, 2013, pp. 46–61.
34 P. Bouř, J. Sopková, L. Bednárová, P. Maloň and T. A. Keiderling, *J. Comput. Chem.*, 1997, **18**, 646–659.
35 P. Bouř and T. A. Keiderling, *J. Chem. Phys.*, 2002, **117**, 4126.
36 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. Montgomery, J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian Inc Wallingford CT*, 2009, 34, Wallingford CT.
37 T. Kuppens, W. Langenaeker, J. P. Tollenaere and P. Bultinck, *J. Phys. Chem. A*, 2003, **107**, 542–553.
38 T. Kuppens, K. Vandyck, J. Van der Eycken, W. Herrebout, B. J. van der Veken and P. Bultinck, *J. Org. Chem.*, 2005, **70**, 9103–9114.
39 T. Kuppens, K. Vandyck, J. van der Eycken, W. Herrebout, B. van der Veken and P. Bultinck, *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.*, 2007, **67**, 402–411.
40 E. Debie, E. De Gussem, R. K. Dukor, W. Herrebout, L. A. Nafie and P. Bultinck, *ChemPhysChem*, 2011, **12**, 1542–1549.
41 C. L. Covington and P. L. Polavarapu, *J. Phys. Chem. A*, 2013, **117**, 3377–3386.
42 J. Kubelka and T. A. Keiderling, *J. Phys. Chem. A*, 2001, **105**, 10922–10928.
43 E. W. Blanch, L. Hecht, L. A. Day, D. M. Pederson and L. D. Barron, *J. Am. Chem. Soc.*, 2001, **123**, 4863–4864.
44 C. R. Jacob, S. Luber and M. Reiher, *ChemPhysChem*, 2008, **9**, 2177–2180.
45 L. D. Barron, F. Zhu, L. Hecht, G. E. Tranter and N. W. Isaacs, *J. Mol. Struct.*, 2007, **834-836**, 7–16.
46 H. F. M. Boelens, R. J. Dijkstra, P. H. C. Eilers, F. Fitzpatrick and J. A. Westerhuis, *J. Chromatogr. A*, 2004, **1057**, 21–30.
47 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.
48 E. R. Blout and R. H. Karlson, *J. Am. Chem. Soc.*, 1958, **80**, 1259–1260.
49 R. H. Karlson, K. S. Norland, G. D. Fasman and E. R. Blout, *J. Am. Chem. Soc.*, 1960, **82**, 2268–2275.
50 A. A. Adzhubei, M. J. E. Sternberg and A. A. Makarov, *J. Mol. Biol.*, 2013, **425**, 2100–2132.
51 J. Makowska, S. Rodziewicz-Motowidło, K. Bagińska, J. A. Vila, A. Liwo, L. Chmurzyński and H. A. Scheraga, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1744–1749.
52 R. Schweitzer-Stenner and T. J. Measey, *Proc. Natl. Acad. Sci.*, 2007, **104**, 6649–6654.
53 B. Zagrovic, J. Lipfert, E. J. Sorin, I. S. Millett, W. F. van Gunsteren, S. Doniach and V. S. Pande, *Proc. Natl. Acad. Sci.*, 2005, **102**, 11698–11703.
54 Z. Shi, K. Chen, Z. Liu and N. R. Kallenbach, *Chem. Rev.*, 2006, **106**, 1877–1897.
55 F. Zhu, J. Kapitán, G. E. Tranter, P. D. A. Pudney, N. W. Isaacs, L. Hecht and L. D. Barron, *Proteins Struct. Funct. Bioinforma.*, 2007, **70**, 823–833.
56 I. H. McColl, E. W. Blanch, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 2004, **126**, 8181–8188.
57 L. Hecht, L. D. Barron, E. W. Blanch, A. F. Bell and L. A. Day, *J. Raman Spectrosc.*, 1999, **30**, 815–825.
58 S. Yamamoto, T. Furukawa, P. Bouř and Y. Ozaki, *J. Phys. Chem. A*, 2014, **118**, 3655–3662.
59 E. W. Blanch, A. F. Bell, L. Hecht, L. A. Day and L. D. Barron, *J. Mol. Biol.*, 1999, **290**, 1–7.
60 T. Blundell, D. Barlow, N. Borkakoti and J. Thornton, *Nature*, 1983, **306**, 281–283.
61 K. Tsuji, H. Ohe and H. Watanabe, *Polym. J.*, 1973, **4**, 553–559.
62 P. R. Bergethon, *The Physical Basis of Biochemistry: The Foundations of Molecular Physics*, Springer, New York, 1998.
63 M. Diem, *Introduction to modern vibrational spectroscopy*, 1993, vol. 543.
64 C. R. Jacob and M. Reiher, *J. Chem. Phys.*, 2009, **130**, 084106.
65 I. H. McColl, E. W. Blanch, A. C. Gill, A. G. O. Rhie, M. A. Ritchie, L. Hecht, K. Nielsen and L. D. Barron, *J. Am. Chem. Soc.*, 2003, **125**, 10019–10026.
66 V. Liégeois and B. Champagne, *Theor. Chem. Acc.*, 2012, **131**, 1284.
67 G. Mix, R. Schweitzer-Stenner and S. A. Asher, *J. Am. Chem. Soc.*, 2000, **122**, 9028–9029.
68 X. G. Chen, S. A. Asher, R. Schweitzer-Stenner, N. G. Mirkin and S. Krimm, *J. Am. Chem. Soc.*, 1995, **117**, 2884–2895.
69 S. J. Ford, A. Cooper, L. Hecht, G. Wilson and L. D. Barron, *J. Chem. Soc., Faraday Trans.*, 1995, **91**, 2087–2093.
70 E. Smyth, *Raman optical activity of proteins and glycoproteins. PhD thesis.*, Faculty of Science, University of Glasgow, 2000.
71 Z. Q. Wen, L. Hecht and L. D. Barron, *J. Am. Chem. Soc.*, 1994, **116**, 443–445.

72 G. Wilson, S. J. Ford, A. Cooper, L. Hecht, Z. Q. Wen and L. D. Barron, *J. Mol. Biol.*, 1995, **254**, 747–760.

73 E. W. Blanch, L. a. Morozova-Roche, L. Hecht, W. Noppe and L. D. Barron, *Biopolymers*, 2000, **57**, 235–248.

74 S. Luber, *J. Phys. Chem. A*, 2013, **117**, 2760–2770.

75 G. Wilson, L. Hecht and L. D. Barron, *J. Mol. Biol.*, 1996, **261**, 341–347.

76 H. Kumeta, A. Miura, Y. Kobashigawa, K. Miura, C. Oka, N. Nemoto, K. Nitta and S. Tsuda, *Biochemistry*, 2003, **42**, 1209–1216.

77 M. Demura, in *Modern Magnetic Resonance*, Springer Netherlands, Dordrecht, 2006, pp. 497–501.

78 L. A. Morozova-Roche, *FEBS Lett.*, 2007, **581**, 2587–2592.

79 H. I. Rösner and C. Redfield, *J. Mol. Biol.*, 2009, **394**, 351–362.

80 L. D. Barron, E. W. Blanch and L. Hecht, in *Unfolded Proteins*, ed. G. D. Rose, Academic Press, 2002, vol. Volume 62, pp. 51–90.

View Article Online
DOI: 10.1039/C6CP05862K

81 K. H. Kim, S. Yun, K. H. Mok and E. K. Lee, *J. Mol. Recognit.*, 2016, DOI: 10.1002/jmr.2543.

82 V. N. Uversky, V. Davé, L. M. Iakoucheva, P. Malaney, S. J. Metallo, R. R. Pathak and A. C. Joerger, *Chem. Rev.*, 2014, **114**, 6844–6879.

83 G. Tóth, S. J. Gardai, W. Zago, C. W. Bertoncini, N. Cremades, S. L. Roy, M. a Tambe, J.-C. Rochet, C. Galvagnion, G. Skibinski, S. Finkbeiner, M. Bova, K. Regnstrom, S.-S. Chiou, J. Johnston, K. Callaway, J. P. Anderson, M. F. Jobling, A. K. Buell, T. a Yednock, T. P. J. Knowles, M. Vendruscolo, J. Christodoulou, C. M. Dobson, D. Schenk and L. McConlogue, *PLoS One*, 2014, **9**, e87133.