This item is the archived peer-reviewed author-version of:

The influence of the amino acid side chains on the Raman Optical Activity spectra of proteins

Reference:

Mensch Carl, Johannessen Christian.- The influence of the amino acid side chains on the Raman Optical Activity spectra of proteins
ChemPhysChem: a European journal of chemical physics and physical chemistry - ISSN 1439-4235 - 20:1(2019), p. 42-54
Full text (Publisher's DOI): <https://doi.org/10.1002/CPHC.201800924>
To cite this reference: <https://hdl.handle.net/10067/1565340151162165141>

The influence of the amino acid side-chains on the Raman optical activity spectra of proteins

Carl Mensch,^[a,b] and Christian Johannessen^{*[a]}

Abstract: The Raman optical activity (ROA) spectra of proteins show distinct patterns arising from the secondary structure. It is generally believed that the spectral contributions of the side-chains largely cancel out because of their flexibility and the occurrence of many side-chains with different conformations. Yet, the influence of the side-chains on the ROA patterns assigned to different secondary structures is unknown. Here, the first systematic study of the influence of all amino acid side-chains on the ROA patterns is presented based on density functional theory (DFT) calculations of an extensive collection of peptide models that include many different side-chain and secondary structure conformations. It was shown that the contributions of the side-chains to a large extent average out with conformational flexibility. However, specific side-chain conformations can have significant contributions to the ROA patterns. It was also shown that α -helical structure is very sensitive to both the exact backbone conformation and the side-chain conformation. Side-chains with $\chi_1 \approx -60^\circ$ generate ROA patterns alike those in experiment. Aromatic side-chains strongly influence the amide III ROA patterns. Because of the huge structural sensitivity of ROA, the spectral patterns of proteins arise from extensive conformational averaging of both the backbone and the side-chains. The averaging results in the fine spectral details and relative intensity differences observed in experimental spectra.

1. Introduction

Raman optical activity (ROA) is a powerful tool to study the conformational propensities of peptides and proteins in solution.^[1–4] The parent technique, Raman spectroscopy, gives very rich spectra of proteins arising from the many molecular vibrations in both the backbone of the protein and its side-chains. The aromatic groups in the side-chains of the residues Phe, Tyr, Trp and His, for example, generate strong Raman bands.^[2,5] ROA is the chiroptical variant of Raman spectroscopy, since it is measured as the difference in the right- (I_R) and left-handed (I_L) circularly polarized components in Raman scattered light.^[1] In the ROA spectra of proteins, the most intense bands are observed in spectral regions that arise from the backbone of the protein such as vibrations of the amide groups.^[1,6] Therefore, the patterns observed in the ROA spectra of proteins reflect the secondary structure of the protein as only the most rigid

structural components give strong signals.^[6]

Many papers that report on the ROA of peptides and proteins in scientific literature, introduce the strength of the technique by stating that the contributions of the side-chains cancel out because of conformational flexibility and disorder, which results in the sensitivity of the technique to secondary structure. However, as for example suggested in the cornerstone review on ROA by Barron that was published in 2000, the side-chains could still affect the exact spectral patterns.^[6]

The past decade, the scientific field of ROA has set its focus largely on the use of quantum chemical calculations to generate a more thorough understanding of the experimental spectra.^[4,7–9] Following the development of analytical derivative procedures to calculate the property tensors necessary to simulate Raman and ROA spectra,^[10–13] and the implementation of these in quantum chemical programmes,^[14–16] many studies have used computational approaches to create a better understanding of the protein structure-spectrum relation.^[4,7]

Since ROA spectra mainly contain signals in specific spectral regions assigned to amide (amide I, amide III modes) and skeletal vibrational modes (C-C, C-N stretching), numerous ROA studies, including by our research group, have used density functional theory (DFT) calculations of poly-L-alanine model systems in specific conformations to compute the ROA patterns arising from the secondary structure, as this is the smallest chiral amino-acid (AA) and the side-chain contributions do not need to be averaged.^[8,9,17,18] Using such simple models, often exceptional agreement with experimental ROA spectra was achieved, also for experimental spectra of peptides that do not contain Ala, as demonstrated by our group.^[9] This raises questions, why such simple systems give a good agreement with experiment, and whether the side-chain contributions are effectively cancelled out or that they do affect the exact spectral patterns somehow.

Recent reports using DFT calculations indeed suggest that the AA side-chains could affect the ROA spectral patterns of proteins and that specific side-chains such as those with aromatic groups might influence the spectral patterns that are used to assign the secondary structure of peptides and proteins.^[8,19–23] A recent study by our group, showed that for different types of β -turn and other secondary structure conformations, the AA side-chains have important contributions to the ROA patterns assigned to secondary structure motifs, even after extensive conformational averaging.^[22]

We here report the first systematic study of the influence of all the side-chains to the secondary structure patterns of proteins in ROA spectra. As the ROA patterns are very dependent on the backbone conformation of the molecule,^[9] multiple secondary structure models were included in this study. Furthermore, the side-chains of the 20 AAs (except Ala and Gly) can adopt different conformations.^[24] As we specifically want to consider the conformational averaging of the side-chains, many of the possible and relevant conformations of those were included in this study. The combination of multiple secondary structures, 20

[a] Carl Mensch, Prof. Dr. Christian Johannessen
Department of Chemistry,
University of Antwerp
Groenenborgerlaan 171, 2020 Antwerp (Belgium).
E-mail: christian.johannessen@uantwerpen.be

[b] Carl Mensch
Department of Chemistry
Ghent University
Krijgslaan 281 (S3), 9000 Ghent, Belgium

Supporting information for this article is given via a link at the end of the document.

AAs and many side-chain conformations of those, means a large amount of peptide structures were needed to be generated and the corresponding spectra calculated. To this end, the ROA spectra of specific tripeptide structures were calculated. Such systems are small enough to be computationally feasible for a study of this magnitude, but their computed spectra systems show patterns similar to those in experimental spectra (see below). By comparing these tripeptide spectra to that of trialanine reference structures with the same backbone geometry, the effect of the side-chain was evaluated.

In the following sections, we first outline the construction of this database of calculated spectra. In the subsequent sections, the general effect of the side-chains is evaluated by looking at how the spectra of specific combinations of a backbone and a side-chain conformation differ and furthermore whether the effect of the side-chain on the ROA spectrum averages out with conformational flexibility. Then, more specific sections on the effect of the side-chains on a few ongoing topics in ROA research are described as examples of applications of the database. We look into the effect of the side-chains on α -helical patterns as the spectral patterns of this type of secondary structure are matter of debate in scientific literature^[9] and the effect of the aromatic side-chains specifically.^[8,20,25,26]

2. Database construction

To study the influence of the side-chains on the ROA patterns, model tripeptide structures with the general formula HCO-Ala-X-Ala-NH₂ were created with, X being one of the 20 AAs, and capped with a formyl (HCO) and amide group (NH₂) at the N- and C-terminus, respectively. The backbone torsion angles of each of the three residues in the model peptide were set to a single pair of torsion angles (ϕ, ψ), thereby creating regular backbone conformations of a specific secondary structure type. Ten different combinations of (ϕ, ψ) angles that are of interest in this study were selected across the Ramachandran plot as listed and depicted in Figure 1. For both Ala and Gly, ten HCO-Ala-X-Ala-NH₂ (X= Ala or Gly) model structures were created by setting the backbone angles to the same ten selected backbone conformations. For the other 18 AAs, for each of the ten backbone geometries, model structures with sterically allowed combinations of the χ angles were created. A recent analysis of a representative subset of high-quality structures in the Protein Data Bank (PDB) by Hintze *et al.* showed that the protein side-chains adopt preferred conformations, referred to as rotamers, defined by a set of side-chain dihedral angles (χ).^[24] The individual torsion angles of the rotamers preferentially take gauche (g) or trans (t) angles with values around -60° (g(-)), 60° (g(+)) or 180° (t), labelled m, p or t by Hintze *et al.*, respectively. Figure 1 shows histograms of the preferred χ_1 and χ_2 torsion angles of Trp in the database by Hintze *et al.* as an example. It demonstrates the preferred side-chain angles adopted by the Trp side-chain. The red lines indicate the values we have manually selected for this study based on the histograms. For the other AAs, similar histograms are given in the supporting information (Figure S1).

Based on the preferred χ values for each residue, we created model peptides (HCO-Ala-X-Ala-NH₂) by making all combinations of these preferred χ angles (see the supporting information sections S1 and S2 for more details).

Furthermore, for the charged residues (Arg, Asp, Glu, His, Lys), all conformations, both with charged and neutral side-chains, were calculated. Unless stated otherwise, the structures containing Arg and Lys with a positively, His with a neutral and Asp and Glu with a negatively charged side-chain were considered.

In the final database, 5775 peptide models with specific backbone and side-chain conformations with corresponding Raman and ROA spectra were retained after validating the connectivity and stereochemistry of the DFT optimized model peptides and removal of duplicate structures (see supporting information section S 2). By using a partial optimization in normal coordinates,^[27,28] the desired backbone and side-chain conformation are largely retained in the optimized geometries.^[9,22] The latter approach, combined with the C-PCM/B3PW91/6-31++G(d,p) level of theory to compute the Raman and ROA spectra, produces very good spectra in comparison with experiment (see methodology) and was therefore selected for the evaluation of the side-chain contributions here. Full geometry optimization leads to large deviations in the backbone conformation (see e.g. ref ^[22]) and therefore renders the calculated spectra useless for mutual comparison and to evaluate the side-chain contributions to the spectral characteristics of the secondary structure.

3. Spectral contributions of the side-chain

The goal of this study is to evaluate the influence of the side-chains on the ROA spectrum. This influence can either arise as a specific band in a specific region of the spectrum (a so-called marker band) or by influencing a spectral region for example sensitive to the secondary structure (amide bands, CH, CH₂, CH₃ deformation bands).

Since there are too many combinations of all 20 AAs with the different secondary structure elements and multiple side-chain conformations to discuss in detail, first, two residues are discussed here as examples. To this end, Trp and Arg, were selected. The first one has the most atoms in its side-chain, the latter has the second most atoms in its side-chain but is much more flexible. Furthermore, Trp has been reported before in scientific literature because it displays a conformationally sensitive ROA marker band around 1550 cm⁻¹.^[25,26]

In order to create a better understanding of how much the side-chains affect the ROA spectrum, the computed spectra of a single conformation of the HCO-Ala-X-Ala-NH₂ models with X= Trp or Arg are compared to that of the all-alanine model HCO-Ala₃-NH₂ with the same backbone conformation in Figure 2.

As shown in Figure 2, the Raman and ROA spectra with Trp as the central residue are completely different when compared to the Ala reference. The Raman spectrum is dominated by the bands arising from the indole in the Trp side-chain. The most intense band for example, computed at 1587 cm⁻¹, arises from

the W3 type vibration of the indole ring, which experimentally is observed around 1550 cm^{-1} .^[5,25] Also in the ROA spectrum, some of the bands arising from the aromatic side-chain are observed, but this is discussed in more detail below. The Ala reference spectrum (orange line in Figure 2) has a -/+ pattern at $1267/1304/1343\text{ cm}^{-1}$ in the amide III region. The two positive bands at 1304 cm^{-1} and 1343 cm^{-1} arise from the coupling of the amide modes with $\text{C}\alpha\text{-H}$ bending perpendicular or parallel to the $\text{C}\alpha\text{-N}$ bond, respectively.^[29] The band at 1343 cm^{-1} is retained in the ROA spectrum with the Trp side-chain while there is no positive band around 1303 cm^{-1} .

On the other hand, the Raman and ROA spectra of the peptide with the Arg side-chain overall show patterns that are similar to that of the reference (Figure 2). In the Raman spectrum of the peptide with the Arg side-chain, a few bands are observed that can be attributed to the side-chain itself. While the overall ROA pattern is similar to the reference spectrum, the backbone modes (amide and skeletal modes) also seem to be affected by the side-chain. For example the relative intensity of the positive amide III bands is again strongly affected by the Arg side-chain. Visualization of the spectral modes reveals that the amide III modes ($\text{C}\alpha\text{-H}$ bending coupled with amide III mode) are coupled with C-H bending motions in the side-chain. As a result, the relative intensities in this region are affected. Furthermore, the negative portion of the amide III region ($1250\text{-}1280\text{ cm}^{-1}$) shifts down in wavenumber in the spectrum with the Arg side-chain.

These two examples show that the Raman and ROA spectra are very sensitive to the side-chain conformation of the residues. This raises the question how conformationally sensitive this influence is and furthermore whether it cancels out by conformational averaging. Except for e.g. the tryptophan band around 1550 cm^{-1} (see further), it is generally believed that most of the spectral contributions that arise from the side-chains are cancelled out in ROA spectra and that only signals from the most rigid structural components, i.e. the backbone, are retained in the spectra.^[6] Therefore, in the next section, the spectra of multiple side-chain conformations are considered and how the patterns average with side-chain flexibility.

4. Side-chain conformational averaging

To demonstrate the effect of the individual side-chain conformations, the separate Raman and ROA spectra of the $\text{g}(-)\text{ttt}, \text{g}(+)\text{ttt}$ and tttt rotamers of $\text{HCO-Ala-Arg-Ala-NH}_2$ in an α -helical conformation are shown in Figure 3. Only the χ_1 angle is thus varied over three values, while the other χ angles are kept in the trans conformation ($\sim 180^\circ$). The variation in the Raman is limited, while in the ROA there is a lot more variation arising from the different side-chain conformations. The amide III region is again clearly affected by the orientation of the side-chain. Since these are only three conformations of the side-chain, we next consider the averaging over a larger number of side-chain conformations and also vary the other χ angles in the Arg side-chain.

In Figure 4, the individual ROA spectra (the Raman are shown in figure S4) of the different combinations of the Arg and Trp side-

chain χ angles are superimposed (top halves of the panels) for two types of secondary structure (α -helix ($-64^\circ; -47^\circ$) and PPII ($-75^\circ; 145^\circ$) secondary structure^[30]). In these spectra, the backbone has the same conformation, while the side-chain conformation is varied. Both for Trp and Arg, there is a lot of variation because of the different conformations of the side-chains. In the bottom panels, the average of the separate spectra is compared to the Ala model with the same backbone conformation. Considering the good resemblance of the averaged ROA spectra with the reference model, these graphs support the generally accepted view that the contribution of the side-chains is mostly cancelled out in the experimental spectra. This can for example be seen by comparing the individual ROA spectra of Trp, that display sharp bands across the entire spectrum, yet these bands cancel out in the average spectrum because they occur with either a positive or negative sign depending on the conformation. In proteins, because of the occurrence of multiple residues with the same side-chain with different conformations (structural disorder) and the structural flexibility of the side-chains (dynamics) the contribution of the side-chain can thus largely cancel out. In Figure 4, it is however shown that both for Arg and Trp, the relative intensities in the extended amide III region are affected. As was shown by our group before, the ratio of the two positive bands in the amide III region (around 1300 cm^{-1} and 1345 cm^{-1}) is very sensitive to the backbone conformation of the peptide or protein.^[9] Here, it is shown that this relative ratio can furthermore be affected depending on the specific side-chain, its conformation and its flexibility. Also the negative portion of the amide III region is affected by the side-chain; the specific bands are shifted and broadened compared to the reference.

Since the extended amide III region is very sensitive to α -helical secondary structure,^[9] in Figure 4 also the ROA spectra demonstrating the effect of the Arg and Trp side-chain are shown for a different secondary structure backbone, namely the PPII extended conformation. In the case of Arg, the side-chain adds a lot of variation to the ROA spectrum, although, the spectrum averages out to the same overall pattern as the Ala reference. In the case of Trp, however, there is bigger difference between the averaged spectrum and the reference spectrum, yet the overall pattern is again similar to that of the reference.

As discussed e.g. by Barron et al., in the ROA spectra of peptides and proteins small signals are observed in the spectral region $1400\text{-}1480\text{ cm}^{-1}$ that arise from CH_2 and CH_3 deformations in the side-chains.^[6] As shown in Figure 4 (and the figures discussed further in the text), the sign, magnitude and wavenumber of spectral features in this spectral window depend on the conformation of the side-chain and the backbone conformation. After conformational averaging, in the case of an α -helical backbone, a small negative signal remains discernible. This is in agreement with experimental ROA spectra of many α -helical peptides and proteins that show such a negative band, sometimes with a small positive component at higher wavenumbers.^[6,9] In the case of a PPII backbone conformation, the signal in this spectral region is much smaller after extensive conformational averaging of the side-chain (Figure 4). When the amide N-H protons are exchanged for deuterons, the spectral

features in this region of peptides and proteins become much more intense.^[6,9] In future studies, extending this database to include the spectra after H/D exchange could help in elucidating these spectral patterns.

The examples of Arg and Trp thus show that the side-chains can have a large influence on the ROA spectrum. The magnitude of this effect depends on the backbone conformation, the specific residue, the side-chain conformation and the side-chain flexibility. Therefore, also the effect of the other side-chains on the secondary structure patterns has to be considered. Instead of showing all spectra separately, we used similarity calculations to compare the influence of the side-chains on the spectra with the reference spectra. The similarity measure we used to this end, is the generalized cosine measure that conveniently expresses how similar two ROA spectra are from -100% (mirror image spectrum is identical to the reference) to 100% (identical spectra).^[9,31] The formula is given in the methodology below and calculates the similarity S_{fg} of a spectrum f with a spectrum g as the overlap integral of the two spectra.

As shown in Figure 4, the mean ROA spectrum is very similar to the reference spectrum. In Figure 5 (a), the spectral similarities are displayed as bar graphs for α -helical structure for all residues (see section S4 in the supporting information for the other secondary structures). First, the similarity of the mean ROA spectrum for each residue (one backbone conformation and the ROA averaged over all side-chain conformations) with the Ala reference spectrum with the same backbone conformation is shown as the top bar graph (a). To get an idea of how much each spectral region contributes to the total similarity, the bars are vertically subdivided in partial similarities (see supporting information section S5 for details). The total similarity is the sum of these separate similarities. For example, the similarity of the mean ROA of an α -helical backbone geometry averaged over multiple Arg side-chain conformations in Figure 4, is very similar to the reference spectrum and indeed shows a S_{fg} value of 84% in Figure 5 (a).

As shown in Figure 3 and Figure 4, the spectra of each individual rotamer show some overlap with the reference spectrum, yet they are considerably different from the reference spectrum. Therefore, the similarity of each individual ROA spectrum (one backbone conformation and one side-chain conformation) was calculated compared to the Ala reference spectrum (HCO-Ala₃-NH₂ with the same backbone conformation). By taking the average of all these individual similarity values, an indication of how much the individual side-chain conformations affect the ROA pattern is obtained. The bars in Figure 5 (b) show this averaged similarity for each residue. For example, for Arg the similarity of each individual spectrum with the reference is on average about 46 %. This is much lower than the 84% similarity of the mean ROA, which is as expected as the individual spectra are much more different from the reference than the mean ROA (see e.g. Figure 3).

In general, for the other residues the same trends are observed. About half of the residues show similarity values of about 80% or higher when comparing the mean ROA to the reference spectrum (Figure 5 (a)), which shows that the side-chain contributions to a large extent cancel out. On the left-hand side

of the bar graph, the lowest total similarities are shown. Both Gly and Pro display very low similarities, which can be explained by the fact that they are different from the other residues; Gly does not have a C β atom and Pro does not have an amide N-H hydrogen, which both are involved in the backbone (amide) modes. Furthermore, the aromatic residues His, Tyr, Phe and Trp show lower similarity values because their rigid aromatic side-chains generate specific spectral bands and furthermore affect the backbone modes (see above and below).

To conclude, these results demonstrate that the side-chains indeed have a strong effect on the ROA patterns (Figure 5 (b)), but that if the side-chain is flexible, its contribution to a large extent cancels out (Figure 5 (a)). These results must be interpreted as trends, as the spectral averaging is arbitrary and is performed specifically to average out the patterns using a large set of side-chain conformations. Therefore, in the next section, the effect of specific side-chain conformations is considered.

5. The ROA patterns are determined by the preferred side-chain conformations

Hintze *et al.* used a large collection of crystal structures (see also the histograms in Figure 1 and Figure S1) to show that the side-chains have preferred conformations; *i.e.* specific combinations of χ angles.^[24,32] Therefore, for the flexible side-chains of the residues Arg, Gln, Glu, Lys and Met, we assessed the ROA spectra of the preferred conformations. As demonstrated by Hintze *et al.*, these residues preferentially adopt g(-)t (also called mt) rotamers, which means χ_1 takes a value near -60° and χ_2 near 180°.^[24] Since we above demonstrated that specifically the α -helical patterns are sensitive to the side-chain conformations and averaging, this type of secondary structure is considered here as an example. In Figure 6, the spectra of the g(-)t conformers are displayed (top) in comparison to the spectra of all side-chain conformations (bottom) for α -helical structure. The top panels show the individual spectra for these five residues.

As can be seen, the g(-)t conformation of these five residues gives rise to very similar ROA spectra with the typical -/+ pattern in the amide III region characteristic for α -helical structure. By including all other side-chain conformations (right-hand side in Figure 6), the amide III region does not display the experimental -/+ amide III pattern as it is observed in experimental ROA spectra of α -helical peptides and proteins.

Further analysis revealed that the effect observed in Figure 6 arises from the χ_1 angle. This is demonstrated in Figure 7 by classifying the α -helical structures of Arg, Gln, Glu, Lys and Met according to the χ_1 angle taking either a g(-), g(+) or t angle (torsion angle near -60°; +60° or 180°, respectively). The g(-) conformations are very similar to the spectra shown for the g(-)t rotamers in Figure 6. The amide III region of g(-) differs distinctly from that of the g(+) and t conformations. Since the amide III region arises from the combination of C-N stretching and N-H and C α -H bending motions, it is affected by coupling to the CH and CH₂ vibrational motions in the side-chain. The orientation of

the C β H₂ group to the main-chain atoms is determined by the χ_1 angle and, as demonstrated here, determines the α -helical ROA pattern. As mentioned before, ROA is very sensitive to α -helical conformation. Therefore, this side-chain database was assembled for four different α -helical backbone conformations. For all these α -helical conformations, the behaviour as depicted in Figure 7 is the same: the χ_1 adopting g(-) angles displays an amide III region that is very different from those with χ_1 being g(+) or t (see supporting information section S4.5). For the other secondary structure elements included in the database developed here, this χ_1 dependence was not observed, definitely not as distinct as for helical structure.

6. How are the ROA patterns of α -helical secondary structure influenced?

As noted above, the ROA spectra of the database peptides with χ_1 angles near -60° (g(-)) give rise to the two positive bands commonly observed in experimental ROA spectra of α -helical peptides and proteins. As was demonstrated by our group before, the ratio of these two bands strongly depends on the actual backbone conformation.^[9] Therefore, this dependence is evaluated in more detail here by including the side-chains. In Figure 8, the ratio of the two α -helical bands is compared for four different α -helical backbone conformations. Each individual bar gives the ratio of the maximum intensity of the two ROA bands of a specific residue. For example, the first bar (dark blue) of each of the four groups gives the ratio of the two bands in the ROA spectrum of Arg averaged over all side-chain conformations. The 20th bar (the yellow bar at the right of each group) is that of Ala and is used as a reference. The bottom panel of Figure 8, gives the difference of each bar with the respective Ala reference. A few observations can be made in Figure 8: (1) generally, the ratio strongly depends on the backbone conformation and follows the same trend as we observed in our previous work using alanine peptides.^[9] Going from the bottom right of the α -helical region in the Ramachandran plot to the top left, the ratio increases.^[9] This can be observed in Figure 8, as the ratios increase from left to right. (2) Overall the side-chains result in a lower ratio compared to Ala. (3) The ratio is very dependent on the specific side-chain, even after averaging over multiple side-chain conformations. For example Asn, Ser and Gly result in quite high ratios, while Pro results in very low ratios and does not follow the trend with the backbone. For Gly and Pro this is very reasonable, since these two do not have a C α -H as the other residues do. On the other hand, Asn and Ser are examples that the exact ratio of the two helical bands depends on the actual side-chain.

In Figure S17, the ratio of the two α -helical bands is shown only based on all structures having a χ_1 angle near -60° (g(-)). Since the relative bars in Figure 8 and Figure S17 differ, it can be concluded that the ratio also depends on the exact side-chain conformation and conformational averaging. As was e.g. shown in Figure 3 above for Arg, the ROA in the amide III region depends on the χ_1 angle, probably because this angle

determines the orientation of the side-chain CH₂ groups to the C α -H in the backbone.

To conclude the sections above, it was shown that the side-chains can have a significant effect on the secondary structure patterns of proteins. Specific residues, such as those with an aromatic group (His, Phe, Trp and Tyr), and Gly and Pro were shown to have a pronounced effect on the spectral patterns. So far, the effect of the side-chains was studied using tripeptides. Since the individual spectra of Trp have such a large impact on the ROA spectra, in the following section, it is evaluated whether this effect is still retained when the peptide is considerably larger.

7. ROA of longer peptides: aromatic residues

To study whether the side-chains still affect the ROA patterns of larger peptides, the spectra of longer α -helical model peptides (HCO-Ala₅-X-Ala₅-NH₂) were calculated. Since the calculation of these larger peptides is much more computationally demanding than that of the tripeptides, we have limited this set of calculations to one α -helical backbone conformation as it was shown that ROA is very sensitive to α -helical structure. Furthermore, to not limit the analysis only to the aromatic residues, also Gly, Thr and Ser and Val were included in this set of model peptides as these have small side-chains and do not need extensive conformational averaging and are thus computationally feasible. While for Gly, Thr, Ser and Val it was noticed that the side-chain almost has no effect on the ROA pattern for these longer peptide models (see figure S18), the aromatic residues (His, Phe, Trp and Tyr) with a specific conformation of the side-chain still have a major impact on the ROA pattern (see the top panels of the plots in Figure 9). The aromatic residues have a distinct influence on the amide III region depending on the specific side-chain conformation. Upon averaging over the different side-chain conformations, however, the spectra become again very similar to the spectrum of a HCO-Ala₁₁-NH₂ reference model with the same backbone conformation (see the bottom panels of the plots in Figure 9).

Based on the comparison of the spectra in Figure 9 and the spectra in the database with characteristic Raman bands of the aromatic residues in Table 1, not many marker bands in the ROA can be identified. The Raman bands of His are much weaker compared to the other aromatic residues and furthermore depend on its protonation state.^[33] The most prominent Raman bands of e.g. Phe (around 1000 cm⁻¹ and 1035 cm⁻¹) do not show corresponding ROA intensity. Based on the strong ROA signals of 1-phenylethanol,^[34] Zhu *et al.* suggested that the ring breathing mode of Phe around 1000 cm⁻¹ might also contribute to the ROA spectra of proteins.^[35] The absence of this ROA band in the DFT calculations discussed here contradicts that statement. The direct environment of a residue such as Phe embedded in the protein might induce ROA signals, but that is outside of the scope of this study. The Raman bands, experimentally observed around 1610-1630 cm⁻¹, of Phe, Tyr and Trp, thus just below the amide I region, show minor ROA intensity in the DFT calculations, depending on the geometry. This is in accordance with the observation of ROA

bands in that region in the spectra of e.g. a few Phe containing peptides^[23], or also for example the positive band around 1610 cm^{-1} in the ROA spectrum of poly(β -benzyl L-aspartate)^[36]. In protein ROA spectra, however, such signals are often very small or indistinguishable from amide I modes.^[2,37] The most distinct ROA marker band is the W3 band of Trp, which experimentally is observed around 1550 cm^{-1} .^[25] As was shown both experimentally^[25] and computationally^[21,26] before, the sign of the ROA band is determined by the $\chi_{2,1}$ side-chain angle of Trp. In a few cyclic dipeptides, a small negative W3 band was observed by Hudecová *et al.*^[21] To the best knowledge of the authors, only a single published experimental ROA spectrum of a protein shows a negative ROA W3 band of a protein, while in other cases the Trp band is either positive or very weak.^[25] In a few experimental spectra, such as that of hen lysozyme^[6,36,38] and α -lactalbumin,^[6,36,39] which have multiple Trp residues in their sequences, the spectral contribution of Trp is observed as a couplet in the with a positive band around 1550 cm^{-1} and a small negative component around 1580 cm^{-1} . However, as shown in Figure 2, 4 and 9, such -/+ or +/- couplets can arise from one specific Trp conformation and do not necessarily reflect the conformation of multiple Trp residues. The above observations hence prompted us to further investigate this specific ROA marker band.

8. ROA of the tryptophan side-chain

Jacob *et al.* reported DFT computed spectra of N-acetyl-(S)-Trp-N'-methylamide with different orientations of the side-chain to study the effect of the χ_1 and $\chi_{2,1}$ side-chain torsion angles.^[26] Since there are two C δ carbons in the Trp side-chain, the "2,1" subscript denotes the carbon next to the nitrogen atom in imidazole (see Figure 1 (b)). They used six different side-chain conformations to show that the sign of the W3 ROA band depends on the $\chi_{2,1}$ angle. The calculated spectra of structures with a positive $\chi_{2,1}$ angle gave rise to a positive ROA band and the negative $\chi_{2,1}$ angle resulted in a negative ROA band. It is however unknown whether this sign-relationship depends on other geometrical properties such as the secondary structure. Here, we extended their study to tripeptides adopting various backbone conformations and many more side-chain orientations. First, the effect of the backbone conformation on the W3 ROA band was probed by comparing structures with a different backbone conformation but with the same side-chain conformation. For most side-chain conformations, the W3 ROA bands of structures with a different backbone conformation were shown to have the same sign and only minor differences in intensity. This is illustrated with one example in the top plot in Figure 10, which shows the comparison of the Raman and ROA spectrum of a model peptide with an α -helical backbone to the corresponding spectra with a PPII backbone with the same side-chain conformation of Trp. However, as is illustrated in the bottom plot of Figure 10, for small absolute values of the $\chi_{2,1}$ angle, distinct differences are observed for the sign and intensity of the W3 ROA band.

Therefore, all the ROA spectra of the HCO-Ala-Trp-Ala-NH₂ structures in the database (with 10 different backbone conformations) are superimposed, based on the classification of the χ_1 and $\chi_{2,1}$ angles being either positive or negative, in Figure 11. The plot shows that there is no correlation between the sign of the W3 ROA band and the χ_1 angle but that the sign of the band depends on the $\chi_{2,1}$ angle, which is in accordance with the earlier work by Blanch *et al.* and Jacob *et al.*^[25,26] There are, however, a few spectra in Figure 11 that show an inverse sign relationship (cfr. bottom plot in Figure 10) which would suggest that the rule does not strictly hold true.

To study the sign dependence in more detail, a larger set of side-chain angles was selected. Therefore, a set of N-acetyl-(S)-Trp-N'-methylamide structures was constructed by setting the χ_1 and $\chi_{2,1}$ angles to all combinations of -180° to 180° with a step size of 30° (-180°, -150°, -120°, ..., 150, 180°). This leads to 13 x 13 structures of which only 145 structures were retained that were found to be geometrically viable (no steric clashes). Next, we fully optimized the geometry of all structures, which led to only a small set of distinct conformations and side-chain orientations. In Figure 12 (a), the W3 ROA band intensity (y-axis) is plotted in function of the $\chi_{2,1}$ angle (x-axis). Each point in the plot gives the maximal (or minimal) intensity of the W3 ROA band. The coloring of the data points is based on the χ_1 angle. For this set of fully relaxed structures, the $\chi_{2,1}$ sign rule of the W3 band holds true. Furthermore, the plot also supports that there is no distinct relation of the W3 ROA band with χ_1 .

If the initial geometries of the structures were partially optimized in normal coordinates, the chosen side-chain torsion angles do not deviate as much as with the full geometry optimization (Figure 12 (b)). For the partially optimized geometries, the relation between the $\chi_{2,1}$ angle and the sign of the W3 ROA band is shown to be mostly true, while only a few $\chi_{2,1}$ angles close to 0° and 180° show also an inverse sign relation. We therefore conclude that the $\chi_{2,1}$ angle is the most important geometrical parameter in determining the sign and intensity of the W3 ROA band.

When Blanch *et al.* reported the relation of the W3 ROA band with the $\chi_{2,1}$ angle based on their experimental measurements, they furthermore suggested a possible $\sin(\chi_{2,1})$ dependence of the W3 ROA band. Jacob *et al.* however concluded based on DFT calculations, that there was no sine relation. Yet, our larger set of conformations does not contradict such a $\sin(\chi_{2,1})$ relation as demonstrated by the grey line in Figure 12.

To conclude this section, our data on the W3 ROA band shows that (1) the sign of this band depends on the $\chi_{2,1}$ and not the χ_1 side-chain angle, (2) the intensity and sign of this band relates with $\sin(\chi_{2,1})$ and (3) there is only a minor or no effect of the secondary structure on the intensity and sign of this band.

9. Conclusions

We have created an elaborate database of DFT calculated Raman and ROA spectra of HCO-Ala-X-Ala-NH₂ peptide structures, with X being one of the 20 amino acids. Based on this database, we have shown that each side-chain in a certain

conformation has a specific and significant effect on the ROA pattern compared to HCO-Ala₃-NH₂ reference with the same backbone conformation. When the ROA spectrum is averaged over multiple side-chain conformations, however, the mean ROA more closely resembles that of the reference as bands that specifically arise from the side-chains are cancelled out. This observation supports the general conception that the experimental ROA spectra of proteins mostly arise from the backbone of the protein and that the contribution of the side-chain cancels out if the side-chains are flexible or if multiple side-chain conformations are present.

The calculated ROA spectra in the database of the peptide models an α -helical backbone were shown to be very sensitive to the side-chain conformation. The characteristic amide III -/+ pattern that is experimentally observed around 1240-1280/1300/1345 cm⁻¹ is reproduced by averaging the ROA spectra in the database that have a g(-) side-chain conformation (χ_1 close to -60°). This observation shows that the spectral contributions of the side-chains not merely cancel out for this secondary structure, but that the pattern depends on the preferred side-chain conformations.

To demonstrate the use of the database, we furthermore analyzed the contributions of the ROA bands of the aromatic residues in detail. Our data shows that only Trp has a distinct ROA marker band around 1550 cm⁻¹. In accordance with earlier studies, the database shows that the sign of this band is determined by the $\chi_{2,1}$ side-chain angle, which determines the orientation of the indole group. Furthermore, our data supports the suggestion by Blanch *et al.* that the intensity of this ROA band is related to the side-chain angle via a sine relation $\sin(\chi_{2,1})$. Altogether, the database was shown to provide relevant data to elucidate the experimental ROA spectra of proteins and the results described here add to the understanding of the structure-spectrum relation. Together with other recent computational studies, the findings presented here show that care must be taken in making very detailed assignments in the ROA spectra of proteins. Because of the huge sensitivity of ROA to e.g. the backbone conformation and the exact side-chain conformations, the spectral patterns that are observed experimentally are the result of extensive conformational averaging.

Computational details

Model peptides were generated by using the Peptide Builder developed by Tien *et al.* that generates the models based on a set of ϕ and ψ torsion angles to define the backbone conformation and a set of χ torsion angles to define the side-chain conformation.^[40] The side-chain χ angles were selected by using all combinations of the most common χ angles as determined from the Top8000 database (see supporting information section S1).^[24,41] Most side-chains preferentially take a value near -60°, +60° or 180°.^[24] Unless stated otherwise, the molecular geometries of the model peptides were optimized using the partial optimization in normal coordinates by Bouř and Keiderling by fixing the normal modes between i300 cm⁻¹ (imaginary) and 300 cm⁻¹, thereby retaining the overall conformation and relaxing the vibrational modes of interest (Qgrad program).^[27,28] In this geometry optimization scheme, none of the Cartesian coordinates are frozen, rather delocalized geometry deformations characterized by vibrational modes in the wavenumber

range i300-300 cm⁻¹ are restricted in the geometry optimization.^[9,27,28,42] The Raman and ROA spectra were calculated at the B3PW91/6-31++G(d,p) level of theory for an incident laser wavelength of 532 nm. The combination of the partial geometry optimization in normal coordinates with this DFT functional and basis set for the calculation of the Raman and ROA spectra has been used in many different studies of peptides and proteins and produces spectra that are very good in comparison with experiment.^[8,9,20,22,28] The rDPS basis set is often used for ROA calculations since it is small, yet we found that it gives very comparable results to the level of theory used here as discussed in ref.^[22] Line broadening was modelled using a Lorentzian band shape with a full width at half height of 20 cm⁻¹ and a Boltzmann intensity correction was applied for a temperature of 300 K (see e.g. Cheeseman *et al.*).^[9,43] Since, the calculated wavenumbers are typically slightly overestimated due to *inter alia* the harmonic approximation, the calculated spectra were scaled in the wavenumber dimension by a factor of 0.987 which was found by our group to be appropriate for this level of theory to allow comparison with experimentally observed band positions.^[9] The amide I vibrational modes require a different scaling factor at this level of theory, but as this factor depends on e.g. the secondary structure^[9], basis set^[22] and solvation,^[44,45] as discussed in ref.^[9], a separate scaling factor was not included. For all DFT calculations, the C-PCM implicit solvent model was used to account for solvent-solute interactions with water. All DFT calculations were performed using the Gaussian16 rev. A.03 programme.^[16] Normal modes were visualised in GaussView 6.0.16 (Gaussian, Inc.). Optimized geometries were visualised using CylView v.1.0 beta.

Spectral similarities between two spectra f and g that represent the spectral intensities at each wavenumber $\tilde{\nu}$ were calculated based on the following overlap integral, i.e.

$$S_{fg} = \frac{\int f(\tilde{\nu})g(\tilde{\nu})d\tilde{\nu}}{\sqrt{\int f(\tilde{\nu})^2d\tilde{\nu}}\sqrt{\int g(\tilde{\nu})^2d\tilde{\nu}}}$$

Identical spectra have a similarity of 100%, while identical mirror image spectra (e.g. $g = -1 \times f$) have a S_{fg} of -100%. The integral of the numerator of S_{fg} was integrated over different wavenumber regions to account for separate spectral regions (see the supporting information section S5).

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.

Keywords: protein side-chains • Raman optical activity • DFT • Computational chemistry • Secondary Structure

- [1] L. D. Barron, *Biomed. Spectrosc. Imaging* **2015**, *4*, 223–253.
- [2] C. Mensch, A. Konijnenberg, R. Van Elzen, A.-M. Lambeir, F. Sobott, C. Johannessen, *J. Raman Spectrosc.* **2017**, *48*, 910–918.
- [3] C. Mensch, R. Pendrill, G. Widmalm, C. Johannessen,

- ChemPhysChem* **2014**, *15*, 2252–2254.
- [4] S. Ostovar pour, L. D. Barron, S. T. Mutter, E. W. Blanch, in *Chiral Anal.*, Elsevier, **2018**, pp. 249–291.
- [5] A. Rygula, K. Majzner, K. M. Marzec, A. Kaczor, M. Pilarczyk, M. Baranska, *J. Raman Spectrosc.* **2013**, *44*, 1061–1076.
- [6] L. D. Barron, L. Hecht, E. W. Blanch, A. F. Bell, *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1–49.
- [7] C. Johannessen, E. W. Blanch, *Curr. Phys. Chem.* **2013**, *3*, 140–150.
- [8] J. Kessler, J. Kapitán, P. Bouř, *J. Phys. Chem. Lett.* **2015**, *6*, 3314–3319.
- [9] C. Mensch, L. D. Barron, C. Johannessen, *Phys. Chem. Chem. Phys.* **2016**, *18*, 31757–31768.
- [10] O. Quinet, B. Champagne, *J. Chem. Phys.* **2001**, *115*, 6293–6299.
- [11] O. Quinet, V. Liégeois, B. Champagne, *J. Chem. Theory Comput.* **2005**, *1*, 444–452.
- [12] V. Liégeois, K. Ruud, B. Champagne, *J. Chem. Phys.* **2007**, *127*, 204105–204105.
- [13] A. J. Thorvaldsen, K. Ruud, K. Kristensen, P. Jørgensen, S. Coriani, *J. Chem. Phys.* **2008**, *129*, 214108.
- [14] K. Aidas, C. Angeli, K. L. Bak, V. Bakken, R. Bast, L. Boman, O. Christiansen, R. Cimiraglia, S. Coriani, P. Dahle, E. K. Dalskov, U. Ekström, T. Enevoldsen, J. J. Eriksen, P. Ettenhuber, B. Fernández, L. Ferrighi, H. Fliegl, L. Frediani, K. Hald, A. Halkier, C. Hättig, H. Heiberg, T. Helgaker, A. C. Hennum, H. Hettema, E. Hjertenæs, S. Høst, I.-M. Høyvik, M. F. Iozzi, B. Jansík, H. J. A. Jensen, D. Jonsson, P. Jørgensen, J. Kauczor, S. Kirpekar, T. Kjærgaard, W. Klopper, S. Knecht, R. Kobayashi, H. Koch, J. Kongsted, A. Krapp, K. Kristensen, A. Ligabue, O. B. Lutnæs, J. I. Melo, K. V. Mikkelsen, R. H. Myhre, C. Neiss, C. B. Nielsen, P. Norman, J. Olsen, J. M. H. Olsen, A. Osted, M. J. Packer, F. Pawłowski, T. B. Pedersen, P. F. Provasi, S. Reine, Z. Rinkevicius, T. A. Ruden, K. Ruud, V. V. Rybkin, P. Salek, C. C. M. Samson, A. S. de Merás, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. Sneskov, A. H. Steindal, K. O. Sylvester-Hvid, P. R. Taylor, A. M. Teale, E. I. Tellgren, D. P. Tew, A. J. Thorvaldsen, L. Thøgersen, O. Vahtras, M. A. Watson, D. J. D. Wilson, M. Ziolkowski, H. Agren, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014**, *4*, 269–284.
- [15] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. J. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian, Inc., Wallingford CT*, **2009**, *34*, Wallingford CT.
- [16] 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, D. J. Fox, *Gaussian, Inc., Wallingford CT*, **2016**.
- [17] F. Zhu, J. Kapitan, G. E. Tranter, P. D. A. Pudney, N. W. Isaacs, L. Hecht, L. D. Barron, *Proteins Struct. Funct. Bioinforma.* **2007**, *70*, 823–833.
- [18] S. Luber, *J. Phys. Chem. A* **2013**, *117*, 2760–2770.
- [19] T. Weymuth, C. R. Jacob, M. Reiher, *ChemPhysChem* **2011**, *12*, 1165–1175.
- [20] J. Kessler, S. Yamamoto, P. Bouř, *Phys. Chem. Chem. Phys.* **2017**, *19*, 13614–13621.
- [21] J. Hudecová, J. Horníček, M. Buděšínský, J. Šebestík, M. Šafařík, G. Zhang, T. A. Keiderling, P. Bouř, *ChemPhysChem* **2012**, *13*, 2748–2760.

- [22] C. Mensch, C. Johannessen, *ChemPhysChem* **2018**, DOI 10.1002/cphc.201800678.
- [23] J. Hudecová, J. Kapitán, V. Baumruk, R. P. Hammer, T. A. Keiderling, P. Bouř, *J. Phys. Chem. A* **2010**, *114*, 7642–7651.
- [24] B. J. Hintze, S. M. Lewis, J. S. Richardson, D. C. Richardson, *Proteins Struct. Funct. Bioinforma.* **2016**, *84*, 1177–1189.
- [25] E. W. Blanch, L. Hecht, L. A. Day, D. M. Pederson, L. D. Barron, *J. Am. Chem. Soc.* **2001**, *123*, 4863–4864.
- [26] C. R. Jacob, S. Luber, M. Reiher, *ChemPhysChem* **2008**, *9*, 2177–2180.
- [27] P. Bouř, T. A. Keiderling, *J. Chem. Phys.* **2002**, *117*, 4126–4132.
- [28] J. Hudecová, K. H. Hopmann, P. Bouř, *J. Phys. Chem. B* **2012**, *116*, 336–342.
- [29] C. R. Jacob, M. Reiher, *J. Chem. Phys.* **2009**, *130*, 084106.
- [30] A. A. Adzhubei, M. J. E. Sternberg, A. A. Makarov, *J. Mol. Biol.* **2013**, *425*, 2100–2132.
- [31] T. Kuppens, K. Vandyck, J. Van der Eycken, W. Herrebout, B. J. van der Veken, P. Bultinck, *J. Org. Chem.* **2005**, *70*, 9103–9114.
- [32] J. W. Ponder, F. M. Richards, *J. Mol. Biol.* **1987**, *193*, 775–791.
- [33] J. M. Benevides, S. A. Overman, G. J. Thomas, in *Curr. Protoc. Protein Sci.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, **2003**, p. 17.8.1-17.8.35.
- [34] J. Kapitán, C. Johannessen, P. Bouř, L. Hecht, L. D. Barron, *Chirality* **2009**, *21*, E4–E12.
- [35] F. Zhu, N. W. Isaacs, L. Hecht, L. D. Barron, *Structure* **2005**, *13*, 1409–1419.
- [36] I. H. McColl, E. W. Blanch, L. Hecht, L. D. Barron, *J. Am. Chem. Soc.* **2004**, *126*, 8181–8188.
- [37] L. Ashton, P. D. A. Pudney, E. W. Blanch, G. E. Yakubov, *Adv. Colloid Interface Sci.* **2013**, *199–200*, 66–77.
- [38] E. W. Blanch, L. A. Morozova-Roche, D. A. E. Cochran, A. J. Doig, L. Hecht, L. D. Barron, *J. Mol. Biol.* **2000**, *301*, 553–563.
- [39] A. N. L. Batista, J. M. Batista, V. S. Bolzani, M. Furlan, E. W. Blanch, *Phys. Chem. Chem. Phys.* **2013**, *15*, 20147–52.
- [40] M. Z. Tien, D. K. Sydykova, A. G. Meyer, C. O. Wilke, *PeerJ* **2013**, *1*, e80.
- [41] J. S. Richardson, D. A. Keedy, D. C. Richardson, in *Biomol. Forms Funct. A Celebr. 50 Years Ramachandran Map* (Eds.: M. Bansal, N. Srinivasan), World Scientific, Singapore, **2013**, pp. 46–61.
- [42] S. Yamamoto, P. Bouř, in *Front. Quantum Chem.*, Springer Singapore, Singapore, **2018**, pp. 181–197.
- [43] J. R. Cheeseman, M. S. Shaik, P. L. A. Popelier, E. W. Blanch, *J. Am. Chem. Soc.* **2011**, *133*, 4991–4997.
- [44] N. S. Myshakina, Z. Ahmed, S. A. Asher, *J. Phys. Chem. B* **2008**, *112*, 11873–11877.
- [45] J. Kubelka, T. A. Keiderling, *J. Phys. Chem. A* **2001**, *105*, 10922–10928.

Figures

Figure 1. (a) Ramachandran plot showing the pairs of (ϕ ; ψ) angles that were used to generate model structures in this study. Right-handed α -helix: (-64°;-47°), (-59°;-44°), (-66°;-41°), (-77°;-34°); right-handed 3_{10} -helix: (-62°;-25°), (-71°;-18°); left-handed PPII-helix: (-75°;145°), (-65°;146°); β -strand: (-132°;158°) and left-handed α -helix: (66°;41°) were included. The contour lines delineate secondary structure regions as adopted from our previous study.^[9] (b) Histograms of χ_1 and χ_2 Trp in the Top8000 database. Since there are two C δ carbons in the Trp side-chain, the χ_2 indicated here as a general example is the $\chi_{2,1}$ angle of Trp. The "2,1" subscript denotes the carbon next to the nitrogen atom in imidazole. The red lines indicate the χ -angles that were selected in this study. The definition of the χ_1 (N-C α -C β -C γ) and χ_2 (C α -C β -C γ -C δ) angles is given in the structure of Trp.

Figure 2. Different side-chains: Raman ($I_R + I_L$) and ROA ($I_R - I_L$) of HCO-Ala-Trp-Ala-NH₂ and HCO-Ala-Arg-Ala-NH₂ (both in blue) compared to HCO-Ala₃-NH₂ (orange), all three with an α -helical backbone structure with $(\phi; \psi) \approx (-59^\circ; -44^\circ)$. Each spectrum was calculated for a single model with only one single backbone and one side-chain conformation. The Arg side-chain is in the all-trans conformation, while the Trp model has $\chi_1 \approx -67^\circ$, $\chi_2 \approx -96^\circ$, as an example. The optimized geometries are depicted on the right.

Figure 3. Side-chain conformation: Raman ($I_R + I_L$) and ROA ($I_R - I_L$) of HCO-Ala-Arg-Ala-NH₂ with an α -helical backbone with $(\phi; \psi)$ set to $(-64^\circ; -47^\circ)$. The side-chain conformation χ_1 is varied from 179° , -62° to 68° . The remaining χ -angles are in a trans conformation ($\sim 180^\circ$). The torsion angles were taken from the optimized geometries which are depicted on the right.

Figure 4. Side-chain conformational averaging: The effect of the Arg (left) and Trp (right) side-chain on the ROA patterns of a HCO-Ala-X-Ala-NH₂ (X=Arg or Trp) peptide with (top) an α -helical backbone conformation and (bottom) a PPII conformation. The top halves of the plots show the individual spectra arising from the different combinations of the side-chain angles (χ). The bottom panels show the average of the separate ROA spectra in blue compared to the HCO-Ala₃-NH₂ reference having the same backbone conformation in orange.

Figure 5. Similarity S_{fg} of the HCO-Ala-X-Ala-NH₂ peptides compared to the HCO-Ala₃-NH₂ peptide with the same backbone conformation ($\phi;\psi$)=(-64°;-47°) for (a) the S_{fg} of the mean ROA spectrum averaged over the side-chain conformations or (b) the mean S_{fg} of the individual ROA spectra. The S_{fg} was calculated for the ROA spectrum of each side-chain rotamer separately and the bar graph here shows the mean S_{fg} per residue.

Figure 6. ROA of preferred side-chain conformations: The calculated ROA spectra of the α -helical model peptides with X= Arg, Gln, Gly, Lys and Met adopting the preferential g(-)t conformation (χ_1 near -60° and χ_2 near 180°; top plot) are compared to the ROA spectra with all side-chain-conformations (bottom plot). The top panels show all the separate ROA spectra and the bottom panels show the average of the mean spectra per residue. All geometries have an α -helical backbone (-64°;-47°).

Figure 7. The amide III region depends on the χ_1 angle: (a) For the residues Arg, Gln, Glu, Lys and Met, the α -helical structures (-64° , -47°) were classified based on the χ_1 torsion angle (N-C α -C β -C γ) conformation (gauche g(+), g(-) or trans t) and the corresponding ROA spectra were averaged for each residue as plotted in the top three panels. The bottom panel displays the average of each of the upper three panels. (b) The χ_1 torsion angle; (c) detail of the bottom panel of a in the amide III region.

Figure 8. Ratio of the two positive amide III band marking α -helical structure (experimentally around 1300 cm⁻¹ and 1340 cm⁻¹). Each bar shows the ratio for the ROA spectrum averaged over all side-chain conformations of each residue. The four bar groups each show the ratio for a different backbone conformation. From blue to yellow we see the bar of: Arg, Asn, Asp, Cys, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Ser, Thr, Trp, Tyr, Val, Pro, Gly and Ala. The bottom panel shows the difference of each ratio with the Ala reference ratio.

Table 1. Characteristic experimental Raman bands of aromatic residues.^[5,33]

Wavenumber (cm ⁻¹)	Aromatic residue	ROA
640-650	Tyr	no
742-795	Trp	no
830-850	Tyr	no
850-890	Tyr	no
870-890	Trp	no
960-980	Trp	no
1000-1010	Phe	no
1010	Trp	no
1030-1055	Phe	no
1330-1360	Trp	(yes)*
1400-1420	His	no
1550-1560	Trp	yes
1600-1630	Tyr, Trp, Phe	small

* The ROA in the amide III region is affected by the Trp conformation, but this is not a separate marker band.

Figure 9. ROA spectra of HCO-Ala₅-X-Ala₅-NH₂ model peptides (X = His, Phe, Trp, Tyr) in an α -helical conformation (-66° ; -41°). The top part of each plot displays the variation with different side-chain torsion angles. The bottom part shows the average ROA spectrum (blue) compared to the reference spectrum of a HCO-Ala₁₁-NH₂ model with the same backbone conformation. The His imidazole was calculated as neutral (Hid). The molecular structures are that of HCO-Ala₅-X-Ala₅-NH₂ with a single side-chain conformation of X=Trp (left) and X=Tyr (right).

Figure 10. Is the W3 band of Trp sensitive to the secondary structure? Raman ($I_R + I_L$) and ROA ($I_R - I_L$) of the database HCO-Ala-Trp-Ala-NH₂ peptides in an α -helical conformation (blue; $-59^\circ; -44^\circ$) compared to that adopting a PPII conformation (orange, $-75^\circ; 145^\circ$). The side-chain χ angles in the optimized geometries are $\chi_1 \sim 180^\circ$; $\chi_{2,1} \sim -69^\circ$ and $\chi_1 \sim 180^\circ$; $\chi_{2,1} \sim 15^\circ$.

Figure 11. ROA sign dependence of the Trp W3 band: ROA W3 band of all structures in the database having a Trp with a positive (top) or negative (bottom) χ_1 angle. Similarly, for the $\chi_{2,1}$ angle being either positive or negative (bottom). For the sake of comparison the y-axis limits are the same in the four panels.

(a) full geometry optimization

(b) partial geometry optimization

Figure 12. Dependence of the W3 ROA band on the Trp side-chain conformation of a set of N-acetyl-(S)-Trp-N'-methylamide structures with 145 different combinations of χ_1 and $\chi_{2,1}$ angles. On the x-axis the $\chi_{2,1}$ angle is shown, while the χ_1 being trans ($150 < |\chi| \leq 180$), positive or negative is given in red, blue or green, respectively. The y-axis gives ROA intensity of the minimum or maximum of the W3 band. (a) Full optimization of the geometries in two steps: (1) geometry optimization by locking the χ_1 and $\chi_{2,1}$ angles followed by (2) full relaxation of the geometry without geometry constraints. (b) The geometries were partially optimized in normal coordinates using Qgrad.^[27]

Entry for the Table of Contents (Please choose one layout)

Layout 1:

ARTICLE

Specific side-chain conformations give strong contributions to the Raman and Raman optical activity (ROA) spectra of proteins. Because of the chiroptical nature of ROA, the spectral contributions of the side-chains to a large extent cancel out due to extensive conformational averaging. The detailed ROA patterns that are used to study protein secondary structure, are however still affected by the side-chains, even after extensive averaging.

Carl Mensch and Christian
Johannessen*

Page No. – Page No.
The influence of the amino acid side-chains on the Raman optical activity spectra of proteins