Raman optical activity of tetra-alanine in the poly(L-proline) II type peptide conformation†

Masakazu Furuta, a Tomotumi Fujisawa, a Hiroyasu Urago, a Takahiro Eguchi, a Takahito Shingae, a Satoshi Takahashi, b Ewan W. Blanch c and Masashi Unno* a

The poly(L-proline) II (PPII) helix is considered to be a major conformation in disordered polypeptides and unfolded proteins in aqueous solution. The PPII conformation can be identified by using Raman optical activity (ROA), which measures the different intensities of right- and left-circularly polarized Raman scattered light from chiral molecules and provides information on stereochemistry associated with vibrational motions. In the present study, we used tetra-alanine (Ala₄) as a model system, since its central amide bond adopts the PPII conformation. The predominance of the PPII conformation was supported by 11 ns molecular dynamics (MD) simulations at 300 K. The MD snapshots were used for subsequent quantum mechanical/molecular mechanical (QM/MM) calculations to compute the Raman and ROA spectra. The present MD + QM/MM analysis leads to a good agreement between the observed and simulated spectra, allowing us to assign most of the spectral features including the ROA band near 1320 cm⁻¹, which has been used as a marker for the PPII conformation. This positive ROA band has three components. The lower frequency component near 1310 cm⁻¹ arises from an internal peptide bond, whereas the higher frequency components around 1320–1335 cm⁻¹ appear due to N- and C-terminal peptide groups. The MD + QM/MM calculations also reproduced the electronic circular dichroism spectra of Ala₄. The present results provide a satisfactory framework for future investigations of unfolded/disordered proteins as well as peptides in solutions by chiral spectroscopic methods.

Introduction

A poly(L-proline) II (PPII) helix was originally defined for the conformation adopted by polymers of L-proline. The PPII helix, however, has been found in amino acid sequences other than those based on L-proline and is now recognized as a common structural motif of disordered polypeptides and unfolded proteins in aqueous solution.¹⁻³ The PPII helix is also seen in loops for the X-ray crystal structures of many folded proteins.⁴⁻⁵ Previous studies have shown that the PPII helix is important for the functional machinery of peptides and proteins.¹¹,¹²,¹³ Convertible helix folds are also observed in the following conformation: a left-handed helix with 3-fold rotational and 75°, and +145°, respectively. Because of its extended character, the PPII helix does not form intrachain hydrogen bonds. The PPII conformation can be observed by several spectroscopic methods such as circular dichroism (CD), vibrational circular dichroism (VCD), and Raman optical activity (ROA).⁹⁻¹² NMR spectroscopy is also useful to assess PPII fraction in peptides.¹³ Among these methods, ROA is a powerful technique that is applicable to both proteins and peptides. ROA measures the different intensity of right- and left-circularly polarized Raman scattered light from chiral molecules and provides information on dynamic changes in stereochemistry associated with vibrational modes.¹⁴⁻¹⁶ Since most biological samples including peptides are chiral, this technique is considered an ideal tool for the study of biological molecules in aqueous solution.¹¹,¹⁴ In fact, the ROA spectra of peptides and proteins that have significant levels of PPII conformation have been previously reported, and the strong positive ROA band at ~1320 cm⁻¹ and the weaker positive band at ~1677 cm⁻¹ have been used as characteristic marker bands for the PPII helix.¹¹,¹⁷,¹⁸ Although the PPII helix is considered to be a common motif for unfolded proteins and disordered peptides, the characterization of its ROA spectral
details is still incomplete. For example, vibrational analysis based on quantum chemical calculations for poly-\(\ell\)-glutamic acid, which is a typical peptide known to support the PPII conformation,\(^2,14,19\) is very difficult because of its large molecular size and its disordered nature.

Thus, in the present study, we focus on a short peptide, tetra-\(\ell\)-alanine (Ala\(_4\)), whose chemical structure is illustrated in Fig. 1. Previous Raman, FTIR, VCD, NMR, and ROA studies have revealed that cationic Ala\(_4\) adopts predominantly the PPII conformation in aqueous solution.\(^17,20–22\) Both experimental and theoretical studies on similar short peptides also indicated that the conformation of the peptide bond is predominantly PPII-like.\(^23–27\) This indicates that Ala\(_4\) is an ideal system for the analysis of Raman and ROA spectra for the PPII conformation. We, therefore, measured the Raman and ROA spectra of cationic Ala\(_4\), and the observed spectra were analysed using quantum mechanical/molecular mechanical (QM/MM) calculations combined with molecular dynamics (MD) simulations.\(^28–31\)

The present spectral simulation for Ala\(_4\) provides a good level of agreement between the observed and calculated spectra. This allows us to analyse in greater detail the positive ROA band around 1320 cm\(^{-1}\), which is a characteristic spectral feature for the PPII conformation.\(^16,17,32–34\) We also simulated the electronic circular dichroism (ECD) spectra of Ala\(_4\), and the calculated spectra well reproduce the experimental data.

**Experimental section**

**Raman, ROA and ECD spectroscopy**

Ala\(_4\) was purchased from Bachem (Torrance, CA, USA) and used as received. The peptide was dissolved in 1 M HCl or DCI solution with a final concentration of 0.1 M. D\(_2\)O (Wako Pure Chemical Industries Inc.; 99.9 atom% D) and DCI (Cambridge Isotope Laboratories Inc.; 99.5 atom% D) were used to prepare 1 M DCI solution. The solution was filtered into a 3 × 3 × 48 mm quartz cuvette. The Raman and ROA spectra of a cationic Ala\(_4\) were recorded on an incident circular polarization ROA spectrometer described elsewhere.\(^30,35\) The 532 nm light from a diode-pumped solid-state laser excited the sample. The laser power at the sample was ~650 mW, and acquisition times were about 69 h. The solvent signal was subtracted from the Raman spectra. ECD spectra were measured on a Jasco J-720 spectropolarimeter. The ~1 mM sample was loaded into a 3 × 3 × 48 mm quartz cuvette.

The spectral measurements were taken at room temperature (~25 °C).

**MD calculations**

Both the initial setup and the MD runs were performed with the AMBER11 program\(^36\) using an explicit representation of solvent molecules and the ff99SB all-atom force field\(^37\) as described previously.\(^30\) Water was modeled by the TIP3P potential,\(^38\) and a peptide molecule was surrounded by a periodic octahedral box of TIP3P water molecules [1161 waters]. All Lennard-Jones interactions were cut off at 8 Å and a particle mesh Ewald method\(^39\) was applied to calculate the long-range electrostatic interactions. The MD run was set up using the following protocol. First, cationic Ala\(_4\) with a PPII conformation (\(\psi = -75^\circ\), \(\psi = 145^\circ\)) was used as an initial structure, and the system was subjected to 1000 steps of minimization to remove close van der Waals contacts and to allow the formation of hydrogen bonds between solvent molecules and the peptide. In this stage, we kept the peptide molecule fixed and simply minimized the positions of the water. In the second step, 2500 steps of minimization were performed without the restraints, i.e., the entire system was minimized. As the third step, the system was then gradually heated to 300 K over 20 ps of constant volume dynamics. The temperature was controlled via Langevin dynamics\(^40\) using a collision frequency of 1.0 ps\(^{-1}\). The SHAKE algorithm\(^41\) was used to constrain bonds between hydrogen and heavy atoms, and the time integration step was set to 2 fs. After heating, the obtained system was simulated for 11 ns at 300 K and 1 atm, with the time integration step being set to 2 fs.

**Conformational analysis**

Seven dihedral angles, \(\psi_1, \phi_2, \psi_2, \phi_3, \psi_3, \phi_4\), and \(\psi_4\), were selected to classify the conformation of Ala\(_4\). The values spanned by each dihedral angle were divided into 2 or 3 broad domains. The \(\psi_1\), \(\psi_2\), and \(\psi_3\) values are classified into \([-180, -135]\) U \([58, 180]\) and \([-135, 58]\), while the \(\phi_2\) and \(\phi_3\) values are grouped by \([-110, 0]\) and \([-180, -110]\) U \([0, 180]\). In a C-terminal moiety, the \(\phi_4\) value is divided into \([-110, 0]\), \([-180, -110]\), and \([0, 180]\), whereas three angle regions of \([-180, -135]\) U \([118, 180]\), \([24, 118]\), and \([-135, 24]\) were used for \(\psi_4\). Thus, there are \(2 \times 2 \times 2 \times 2 \times 3 \times 3 = 288\) possible conformers for Ala\(_4\). We used these classifications to analyze the structures from the MD snapshots.

**DFT calculations**

All quantum chemical calculations were performed using the program Gaussian09.\(^42\) A two-layer ONIOM\(^43\) method was used to perform the QM/MM calculations of Ala\(_4\) surrounded by explicit water molecules. The initial geometries were obtained from snapshots of the MD simulation. The number of surrounding water molecules was reduced to 100, and a counter anion Cl\(^-\), which was needed for the MD calculation, was removed. In this process, water molecules closer to the solute molecule were selected by the custom-written software. The QM region consisted of Ala\(_4\), and the MM region consisted of all of the water molecules. All MM calculations were performed using the
weaker positive amide I ROA band at (1677/1646 cm$^{-1}$) in the amide I region. The positive/negative couplet $N_{\text{ro}}$ in the case of ROA spectra, the results.\textsuperscript{17,46} The positive features in the ROA spectrum also agree with the previous $E_{\text{ro}}$ scheme,\textsuperscript{45} which incorporates the partial charges of the $\text{C}^{\alpha}$ parameters for water. The QM part of the system was computed at the B3LYP/6-31+G** level of theory. An electronic embedding (EE) scheme,\textsuperscript{45} which does not include the polarization of the QM wave function by the MM environment system, was used. For geometry optimizations, the positions of the MM water molecules were frozen. Harmonic frequencies as well as Raman and ROA intensities with 532 nm excitation were computed, and the simulated spectra were generated assuming a Gaussian band shape with a half-width of 10 cm$^{-1}$. The calculated vibrational frequencies were uniformly scaled using a factor of 0.984.

We also calculated the electronic absorption and ECD spectra. Transition energies, dipole and rotational strengths have been calculated by TD-DFT at the B3LYP/6-31+G** level. The number of computed excited states was 20. For comparison with experiment, theoretical spectra were simulated as sums of Gaussian bands with empirical half-widths of 3500 cm$^{-1}$.

Results and discussion

Experimental Raman and ROA spectra

The upper two traces in Fig. 2 show the observed Raman and ROA spectra of Ala$_4$ in 1 M HCl or DCl with 532 nm excitation. The data for Ala$_4$ in HCl and DCl are also displayed separately in Fig. S1 and S2 in the ESI,\textsuperscript{†} respectively. The Raman spectrum (trace a) is characterized by bands at 1677, 1461, 1264, 1104, 905, 868, and 836 cm$^{-1}$, and these spectral features are consistent with those reported previously.\textsuperscript{17,46} The figure also displays the ROA spectrum, which exhibits some positive bands at 1677, 1336, 1311, and 1120 cm$^{-1}$, whereas negative bands are observed at 1467, 1405, 1377, 1264, 1163, and 868 cm$^{-1}$. These features in the ROA spectrum also agree with the previous results.\textsuperscript{17,46} The positive $\sim 1320$ cm$^{-1}$ ROA band together with a weaker positive amide I ROA band at $\sim 1677$ cm$^{-1}$ have been used as the main ROA signatures of the PPII conformation.\textsuperscript{17,18,32–34} Here we note the broad nature of the $\sim 1320$ cm$^{-1}$ band. Although its exact band shape varies among different samples, peptide molecules that adopt the PPII conformation exhibit a clear shoulder near 1330 cm$^{-1}$.\textsuperscript{17,18,34}

The figure also displays the Raman and ROA spectra of Ala$_4$ in 1 M DCl (traces b and d), and many of the observed bands are affected by the H/D exchange. A sharp Raman band at 1467 cm$^{-1}$ for Ala$_4$ in HCl exhibits a clear shoulder near 1479 cm$^{-1}$ for the N-deuterated derivative. Another notable H/D exchange effect is seen in the extended amide III region around 1200–1350 cm$^{-1}$, and a main Raman band at 1264 cm$^{-1}$ for Ala$_4$ in HCl almost disappears upon deuteration. The H/D exchange also causes intensity changes and/or frequency shifts for almost all the Raman bands in a lower frequency region ($<1000$ cm$^{-1}$). In the case of ROA spectra, the N-deuteration affects the band in the amide I region. The positive/negative couplet (1677/1646 cm$^{-1}$) for Ala$_4$ in HCl shows only a positive band in DCl at 1670 cm$^{-1}$. A similar D$_2$O-induced spectral change was also reported for di-L-alanine (Ala$_2$) and tri-L-alanine (Ala$_3$).\textsuperscript{46}

We observed that a new positive band appears at 1479 cm$^{-1}$ for the N-deuterated derivative. This new ROA band correlates with the above-mentioned shoulder in the Raman spectrum. The main ROA features in the extended amide III region also show the spectral changes upon the H/D exchange.

Molecular dynamics simulations

We performed MD calculations of Ala$_4$ to examine the structure in an aqueous environment. We placed a cationic Ala$_4$ molecule in an 18.8 Å octahedral box containing TIP3P water molecules, and an 11 ns MD run was executed using the AMBER11 MD package.\textsuperscript{16} We analysed the trajectory that consists of a total of 22,000 snapshots with a 0.5 ps interval, and Fig. S3 in the ESI displays the time dependence of the energies and the backbone (N, C$_\alpha$, and C) root-mean-square deviation (RMSd). The data shown in this figure demonstrate stability over nanosecond timescales, indicating that the system has reached equilibrium. Thus 250 snapshots of the last 10 ns simulations were taken every 40 ps for the subsequent QM/MM calculations. The conformational distribution of Ala$_4$ for the selected 250 snapshots
is represented as a Ramachandran plot in Fig. 3. The probability distributions for the dihedral angles $\psi$ and $\phi$ are also shown in Fig. S4 in the ESI.† As illustrated in these figures, the three residues mainly adopt values with $\phi$ and $\psi$ angles around $-70^\circ$ and $+145$–$160^\circ$, which are typical values for the PPII conformation. This result from the MD calculations is consistent with previous experimental and theoretical studies showing a predominant PPII conformation in short alanine peptides such as Ala$_4$. Although most of the peptide groups are characterized by the PPII conformation, a small fraction of the C-terminal moiety exhibits structural parameters with $\phi_4$ and $\psi_4$ angles around $+50^\circ$ and $+70^\circ$, respectively.

**Calculated Raman and ROA spectra**

To compute the Raman and ROA spectra, hydrated clusters with 100 water molecules were extracted from the 250 MD snapshots. For quantum chemical calculations, we used the QM/MM method to simulate the Raman and ROA spectra. The QM part comprises the solute molecule, whereas the water molecules were treated as the MM region. The positions of the MM water molecules were frozen during geometry optimizations of the solute. On the basis of the seven dihedral angles ($\psi_1$, $\phi_2$, $\psi_2$, $\phi_3$, $\psi_3$, $\phi_4$, and $\psi_4$), the optimized structures of Ala$_4$ from the QM/MM calculations were classified into 288 conformations. In Table 1, we have selected the five most populated conformers, whose populations are each larger than 5%, and compare the averaged values for each dihedral angle. The most stable conformation, which is denoted as conformation 1, exhibits $\phi$ and $\psi$ angles around $-76^\circ$ and $+140^\circ$ for three peptide groups. These are typical values for the PPII conformation, and the population of conformation 1 is 28%. This result indicates that the PPII helix is a major motif for Ala$_4$. Fig. 4 displays a typical structure for conformation 1, which is based on a MD snapshot at 4.58 ns. This structure was selected because its seven dihedral angles $\psi_1$, $\phi_2$, $\psi_2$, $\phi_3$, $\psi_3$, $\phi_4$, and $\psi_4$ exhibit the minimum root-mean-square deviation from the corresponding averaged values for conformation 1 (see Table 1).

In this study, we calculated the Raman and ROA intensities of Ala$_4$ with B3LYP/6-31+G** using the EE scheme. Adding a diffused function to hydrogen atoms (i.e., B3LYP/6-31+G**) affects the intensities little (see Fig. S5 in the ESI†). The Raman and ROA spectra calculated on the basis of 250 hydrated clusters were averaged, and Fig. 2 compares the simulated and observed spectra. The figure also includes the calculated spectra of a deuterated sample, in which exchangeable protons were deuterated (N1–D, N2–D, N3–D, N4–D, C4(=O)OD). As shown in the comparison in Fig. 2, the predicted Raman and ROA spectra are very similar to their corresponding experimental spectra. In addition, the observed H/D exchange effects on the spectra are well reproduced in the simulated spectra. As mentioned above, we performed an 11 ns MD simulation for the analysis. Although this simulation time might not be long enough to fully explore all possible conformations of Ala$_4$, the good agreement between the observed and calculated spectra implies that the MD simulation captures the main conformational features of Ala$_4$.

The effects of the spectral averaging on the Raman and ROA spectra are examined in Fig. 5, where all the calculated spectra of 250 conformations as well as the averaged spectra are compared to the experimental spectra. Black lines in traces c and d

![Fig. 3 Plot of the \( \psi \) and \( \phi \) angles (indicated as dots) for 250 MD snapshots of cationic Ala$_4$, superimposed on a Ramachandran plot. The data for residue 2 (\( \psi_2, \phi_2 \), red), 3 (\( \psi_3, \phi_3 \), green), and 4 (\( \psi_4, \phi_4 \), blue) are shown in different colors. L stands for the left-handed \( \alpha \) helix.](image)

![Fig. 4 A representative optimized geometry of cationic Ala$_4$, which is explicitly hydrated with 100 water molecules. This structure is based on a MD snapshot at 4.58 ns. The QM region of the QM/MM calculation is illustrated by a ball and stick model. Black, blue, and red represent carbon, nitrogen, and oxygen atoms, respectively.](image)

<table>
<thead>
<tr>
<th>Conformer</th>
<th>( \psi_1^a )</th>
<th>( \phi_2^a )</th>
<th>( \psi_2^a )</th>
<th>( \phi_3^a )</th>
<th>( \psi_3^a )</th>
<th>( \phi_4^a )</th>
<th>( \psi_4^a )</th>
<th>( P^b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>139.0</td>
<td>-74.8</td>
<td>136.0</td>
<td>-76.5</td>
<td>144.2</td>
<td>-77.1</td>
<td>130.4</td>
<td>28.0</td>
</tr>
<tr>
<td>2</td>
<td>142.1</td>
<td>-137.0</td>
<td>137.3</td>
<td>-80.1</td>
<td>156.4</td>
<td>-76.2</td>
<td>145.6</td>
<td>8.8</td>
</tr>
<tr>
<td>3</td>
<td>150.9</td>
<td>-68.6</td>
<td>160.6</td>
<td>-73.7</td>
<td>139.1</td>
<td>55.7</td>
<td>59.8</td>
<td>6.0</td>
</tr>
<tr>
<td>4</td>
<td>156.8</td>
<td>-80.7</td>
<td>148.3</td>
<td>-137.4</td>
<td>143.4</td>
<td>-77.4</td>
<td>137.0</td>
<td>5.6</td>
</tr>
<tr>
<td>5</td>
<td>147.8</td>
<td>-74.4</td>
<td>141.9</td>
<td>-73.8</td>
<td>148.0</td>
<td>-136.1</td>
<td>72.5</td>
<td>5.2</td>
</tr>
</tbody>
</table>

*a Definitions in Fig. 1. * Relative populations (%).

**Table 1 Selected dihedral angles (deg) and populations for major conformations of tetra-L-alanine**
Assignments and structural interpretations

(A) Extended amide III region. A cationic Ala4 contains 126 fundamental vibrational modes, which are denoted as $\nu_i$ in the present study. The good level of agreement between the experimental and calculated spectra allows us to assign the Raman and ROA bands of Ala4 as summarized in Table 2. Fig. S6 in the ESI† presents atomic displacements of selected normal modes for conformation 1 based on a MD snapshot at 4.58 ns. First we note the band near 1320 cm$^{-1}$ that is characteristic of the PPII conformation. The 1250–1400 cm$^{-1}$ range in the vibrational spectra of peptides and proteins is referred to as the extended amide III region. The bands appearing in this region are due to the coupling of the classical amide III vibration, which is an in-phase combination of the N–H bending and the C–N stretching vibration of the amide group, with C–H bending vibrations. The C$\equiv$H–H bending vibrations can be classified into two modes, i.e., the C$\equiv$H–H bending (I) modes are mainly due to the vibration of the C$\equiv$H–H group along the direction of the N–C$\equiv$H bond, while the C$\equiv$H–H bending (II) mode is due to the vibration of the C$\equiv$H–H group that is approximately perpendicular to the N–C$\equiv$H bond. Because of couplings among these modes, the vibrational spectra in the extended amide III region have been shown to be sensitive to changes in the secondary structure. The methine bending regions of VCD spectra are also useful to study the structures and conformations of amino acids and peptides.

Fig. 6A displays the extended amide III region of the observed (a and b) and calculated (c and d) Raman and ROA spectra of Ala4. Important normal modes ($\nu_{51}$, $\nu_{52}$, $\nu_{53}$, and $\nu_{56}$) are illustrated in Fig. 6B–E. The positive band characteristic of PPII is mainly ascribed to $\nu_{53}$ calculated at 1300 cm$^{-1}$. As illustrated in Fig. 6D, the $\nu_{53}$ mode is a combination of the C$\equiv$H–H bending (II) and C$\equiv$H–H bending (II) vibrations. The positive ROA band near 1320–1340 cm$^{-1}$ contains high frequency components, and these are assigned to C$\equiv$H–H bending modes around N– ( $\nu_{51}$) and C– ( $\nu_{52}$) terminal moieties.

In the 1250–1400 cm$^{-1}$ region, there are several negative ROA bands, and the band at 1405 cm$^{-1}$ originates from methyl symmetric deformation (umbrella) modes ($\nu_{43}$–$\nu_{46}$). Since this band is due to a methyl group, alanine-peptides
substitutions of exchangeable protons. The simulated spectra shown in Fig. 2 reproduce these observations, and the band is assigned to the N-terminal NH$_3^+$ scissor mode $\nu_{39}$.

The sharp Raman and ROA bands near 1465 cm$^{-1}$ are assigned to coupled CH$_3$ deformation (scissor) modes ($\nu_{34}$-$\nu_{11}$), and a similar assignment was made for Ala$_2$,$^{22,23}$ Since these modes are partially mixed with N–H bending vibrations, both Raman and ROA spectra exhibit clear H/D exchange effects.

(C) Low frequency region: 800–1200 cm$^{-1}$. The negative and positive ROA bands observed at 1163 and 1120 cm$^{-1}$ can be correlated with the calculated bands at 1163 and 1110 cm$^{-1}$, respectively (Fig. 2). The negative Raman band at 1163 cm$^{-1}$ mainly involves the N–C$^\alpha$ stretching coordinate ($\nu_6$-$\nu_5$), whereas both of the N–C$^\beta$ and C$^\alpha$-CH$_3$ stretching motions ($\nu_{4a}$-$\nu_{47}$) contribute to the positive ROA band at 1120 cm$^{-1}$. Because of the nature of the normal modes, we could expect that these ROA bands reflect the PPII-like conformation. In fact, these bands were also observed for various alanine peptides such as Ala$_x$Ala$_y$Ala$_z$ and acetyl-OOAAAAAAOO-amide (OAO), where O represents ornithine.$^{17,46}$ OAO has been also shown to adopt a predominantly PPII conformation.$^{13}$ However, because of a contribution of deformation motions of the side chain methyl group, it would be difficult to use the positive ROA band near 1120 cm$^{-1}$ as a general structural marker for the PPII conformation.

A negative Raman band at 1009 cm$^{-1}$, along with its corresponding weak Raman band calculated at 1003 cm$^{-1}$, appears due to a coupled N-terminal NH$_3$ and CH$_3$ deformation mode, $\nu_{77}$. Similar Raman and ROA bands were also observed in cationic Ala$_2$ and Ala$_3$,$^{17,46}$ and the assignment is confirmed by the absence of this band in anionic Ala$_2$ where the N-terminal amide moiety is deprotonated.$^{46}$ The present assignment is also consistent with the findings of a previous study on Ala$_2$.$^{23}$

A positive ROA band at 962 cm$^{-1}$ is observed in both conformers 1 and 4 and is assigned to the N-terminal C$^\alpha$-C$^\beta$ and C$^\alpha$-CH$_3$ stretching vibrations calculated at 950 cm$^{-1}$ ($\nu_{40}$). This mode is relatively localized in the N-terminal two Ala residues, and a similar ROA band was also observed for Ala$_3$ and Ala$_5$.$^{17}$

The positive ROA band at 909 cm$^{-1}$ is a delocalized skeletal vibration that involves C=C/C=N stretching and C–H bending motions. Because of the delocalized nature, its intensity and frequency depend on the molecular structure. In Ala$_4$, this mode is predicted to appear at 920 cm$^{-1}$ with a weak Raman intensity ($\nu_{77}$). In the Raman spectra of alanine-based peptides (e.g., Ala$_2$, Ala$_3$, Ala$_4$, Ala$_5$, OAO), there is a sharp Raman band around 900 cm$^{-1}$. We assign the measured 905 cm$^{-1}$ Raman band for Ala$_4$ to the calculated 895 cm$^{-1}$ band, which contains a C$^\alpha$-CH$_3$ stretching motion coupled with methyl C–H bending motions ($\nu_{77}$). Thus, this sharp Raman band is only seen for alanine-rich peptides.$^{17}$

Electronic circular dichroism spectra and effects of water environments

Fig. 7A illustrates the ECD spectrum of cationic Ala$_4$, and this spectrum is consistent with that reported previously.$^{20}$ The ECD spectra of Ala$_4$ showed strong negative 195 nm and weaker
positive 220 nm ECD bands,\textsuperscript{20} which are characteristic of PPII conformations.\textsuperscript{2,12} In this study, the 250 QM/MM optimized structures that are used for the calculations of the Raman and ROA spectra are also utilized to compute the ECD spectra (see Fig. S7 in the ESI\textsuperscript{†}). Fig. 7B displays the averaged ECD spectrum (black), which nicely reproduces negative (-190 nm) and weaker positive (-200 nm) features that are characteristic of the PPII conformation. We also show a spectrum for conformation 1 based on a MD snapshot at 4.58 ns (magenta) in Fig. 7B. This spectrum from a single structural model accounts for the PPII helix.\textsuperscript{11,17,18} In the case of Ala\textsubscript{4}, this band is broad and exhibits a clear shoulder near 1336 cm\textsuperscript{-1}. As discussed above, the high frequency component is mainly ascribed to C=O\textsubscript{p}–H bending modes around N- and C-terminal peptide groups. This assignment is consistent with a previous study on alanine oligopeptides (Ala\textsubscript{2}–Ala\textsubscript{5}) by McColl \textit{et al.,} who found that the ROA intensities of the high frequency components become relatively less pronounced in longer oligopeptides. The present assignment is also consistent with a previous study\textsuperscript{17} which showed that a relatively sharp ROA band is observed at 1319 cm\textsuperscript{-1} for an alanine-rich peptide OAO and poly(l-glutamic acid).

Finally, we have examined the effects of surrounding water environments on the Raman, ROA, and ECD spectra of Ala\textsubscript{4}. Here we used the structure based on a MD snapshot at 4.58 ns as a model. The simulated spectra presented so far were obtained by the QM/MM method using an EE scheme. In contrast, the data shown in panel B of Fig. S8 (ESI\textsuperscript{†}) shows the computed spectra using a ME scheme, in which the polarization of the QM wave function by the MM environment system (i.e., surrounding water molecules) is not included. Table S1 in the ESI\textsuperscript{†} compares the dihedral angles of the optimized structures between EE and ME schemes, and both of the structures correspond to conformation 1. In panel A of Fig. S8 (ESI\textsuperscript{†}), we compare the calculated Raman and ROA spectra between the EE and ME schemes. The figure demonstrates that the calculated spectra are similar between the two schemes except for a high frequency region around 1600–1800 cm\textsuperscript{-1}. The difference in the high frequency region can be interpreted in terms of a lack of hydrogen bonding interactions between a solute (QM part) and water molecules (MM part) in the ME scheme, since a solute–solvent hydrogen bond is expressed as an electrostatic interaction between the QM and MM layers in a QM/MM calculation. Thus the Raman and ROA spectra below 1600 cm\textsuperscript{-1} are relatively insensitive to solvent environments, implying that these vibrational spectra can be used as a good structural marker.

In contrast, the data shown in panel B of Fig. S8 (ESI\textsuperscript{†}) indicate that the water environment distinctly affects the ECD spectrum of Ala\textsubscript{4}. In this case, we used the same optimized structures obtained by an EE scheme to calculate the ECD spectra. As illustrated in the figure, the computed spectrum using a ME scheme exhibits a negative ECD band around 260 nm, which is very small in the observed spectrum of Ala\textsubscript{4} (Fig. 7A).\textsuperscript{20} This observation indicates that solvent environments significantly affect the ECD spectrum on Ala\textsubscript{4}, even if the structural changes are absent. We can therefore conclude that vibrational spectra such as Raman and ROA are more suitable to discuss peptide structures than ECD, which is sensitive to both molecular structures and electrostatic effects of solvent environments.

**Implications**

In the present study, we have successfully assigned most of the Raman and ROA bands for cationic Ala\textsubscript{4}. The comprehensive assignment allows us to characterize useful marker bands for the PPII conformation. The strong positive ROA band at \(\sim 1320\) cm\textsuperscript{-1} has been used as the most characteristic feature for the PPII helix.\textsuperscript{11,17,18} In the case of Ala\textsubscript{4}, this band is broad and exhibits a clear shoulder near 1336 cm\textsuperscript{-1}. As discussed above, the high frequency component is mainly ascribed to C=O\textsubscript{p}–H bending modes around N- and C-terminal peptide groups. This assignment is consistent with a previous study on alanine oligopeptides (Ala\textsubscript{2}–Ala\textsubscript{4}) by McColl \textit{et al.,} who found that the ROA intensities of the high frequency components become relatively less pronounced in longer oligopeptides. The present assignment is also consistent with a previous study\textsuperscript{17} which showed that a relatively sharp ROA band is observed at 1319 cm\textsuperscript{-1} for an alanine-rich peptide OAO and poly(l-glutamic acid).

Zhu \textit{et al.}\textsuperscript{56} used a multivariate analysis of a set of ROA data and obtained an averaged spectrum for disordered/irregular proteins. Since the PPII conformation is a common structural motif of disordered polypeptides and unfolded proteins, the averaged spectrum represents a typical ROA spectrum for the PPII conformation. In this spectrum, a small negative feature is seen at 1263 cm\textsuperscript{-1} in addition to a strong positive band at 1318 cm\textsuperscript{-1}. The ROA spectrum of Ala\textsubscript{4} shows a similar negative band at 1264 cm\textsuperscript{-1} (Fig. 2), which is assigned to a classical amide III vibration. Thus a strong positive band at \(\sim 1320\) cm\textsuperscript{-1} with a small negative feature at \(\sim 1265\) cm\textsuperscript{-1} acts as a set of markers for the PPII conformation. It should be noted, however, that the
negative feature may not be clear, if the other conformational elements such as the \( z \)-helix or \( \beta \)-sheet co-exist.

Another marker for the PPII conformation is a positive ROA band at \( \sim 1675 \text{ cm}^{-1} \), which is due to the amide I mode. The ROA spectrum of \( \text{Ala}_4 \) exhibits a similar positive band at \( 1677 \text{ cm}^{-1} \) and accompanies a negative feature at \( 1646 \text{ cm}^{-1} \) (Fig. 2). This negative feature is due to N-terminal NH\( _3 \) deformations, so that this band does not work as a marker for the PPII conformation. In fact, the typical ROA spectrum for PPII determined by Zhu et al.\(^{16} \) does not show a negative band in the 1500–1700 cm\(^{-1} \) region. It has been considered that disordered poly(\(-\text{glutamic acid}\)) contains large amounts of PPII.\(^{1,19} \) Its reported ROA spectrum\(^{16} \) shows a negative feature near 1640 cm\(^{-1} \). Because only the positive ROA band is expected in this region for PPII, the presence of a small negative feature suggests the presence of the other conformational elements like an \( \alpha \)-helix and \( \beta \)-sheet, which exhibit a negative band near 1632 and 1654 cm\(^{-1} \), respectively.\(^{36} \)

Conclusions

We have measured the Raman and ROA spectra of cationic \( \text{Ala}_4 \), and the observed spectra were analysed by QM/MM calculations combined with MD simulation in order to incorporate an explicit water environment as well as conformational flexibility. As previously suggested,\(^{13,17,20,21} \) the present MD simulations demonstrate that PPII is a main conformational element in \( \text{Ala}_4 \). We used 250 MD snapshots for the subsequent QM/MM calculations, and the average of the computed spectra well reproduces most of the observed features including the ROA band near 1320 cm\(^{-1} \), which is a marker for the PPII conformation.\(^{17,18,32–34} \) This characteristic ROA band around 1320 cm\(^{-1} \) has three components, the low frequency component near 1310 cm\(^{-1} \) arises from an internal peptide bond, whereas higher frequency components around 1320–1335 cm\(^{-1} \) are due to N- and C-terminal peptide groups. In addition, we have assigned most of the features in the Raman and ROA spectra of \( \text{Ala}_4 \). These results will provide useful information on future investigations of unfolded/disordered proteins as well as peptides in solutions.

Acknowledgements

This work was supported by JSPS KAKENHI Grant to M. U. (26410017) and to T. F. (16K17859) and UK Engineering and Physical Sciences Research Council [EP/J019623/1] to E. B. A portion of the computations was performed at the Research Center for Computational Science, Okazaki, Japan.

Notes and references