Structural classification of the amide I sites of a beta-hairpin with isotope label 2DIR spectroscopy
Roy, Santanu; Jansen, Thomas L. C.; Knoester, Jasper

Published in:
Physical Chemistry Chemical Physics

DOI:
10.1039/b925645h

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Structural classification of the amide I sites of a β-hairpin with isotope label 2DIR spectroscopy†

Santanu Roy, Thomas L. C. Jansen* and Jasper Knoester

Received 7th December 2009, Accepted 23rd April 2010
DOI: 10.1039/b925645h

We present a theoretical study of the possibility to use isotope label two-dimensional infrared (2DIR) spectroscopy to obtain site specific structural information in trpzip2. This small β-hairpin peptide was designed as a model system for studying protein folding in β-sheet structures. In order to unravel the folding mechanism, the surroundings of local sites should be characterized, which in principle is possible by using 2DIR in combination with isotope labeling. This requires a classification that correlates local structures to two-dimensional spectra. To this end, we provide the first spectral simulation of the isotope label spectra of all the amide I sites in trpzip2. We find that the anti-diagonal width of the 2DIR peak associated with a labelled site is a good measure of solvent exposure and the key parameter to distinguish between solvent exposed and internal sites. The diagonal widths are not particularly sensitive to this, but they do reveal the presence of slowly interconverting turn structures.

1. Introduction

The structure of proteins determine their function and one needs to know this structure to understand the function. Proteins that have folded into a wrong structure are involved in numerous common diseases, such as Alzheimer’s, Parkinson’s disease,1 and diabetes.2 Therefore, the protein folding problem achieved a considerable amount of attention from the scientific community and there have been numerous studies, as for example, the folding at home programme3 in the effort to unravel the underlying mechanism. However, the problem is still largely unsolved even for small peptides. Protein structure and dynamics can be followed with different experimental methods, such as nuclear magnetic resonance (NMR),5,6 small angle X-ray scattering (SAXS),7,8 fluorescence spectroscopy,9 circular dichroism (CD) spectroscopy,10 Fourier transform infrared (FTIR) and two dimensional infrared (2DIR)11–13 correlation spectroscopy. Using SAXS, it is very challenging to track the intra-molecular structural changes during the folding process. The NMR technique is suitable for studying folding and unfolding dynamics of proteins,14 but is limited to slow dynamics on microsecond and longer time scales.15 Fluorescence is fast enough to catch folding dynamics, but limited by the need of fluorescent chromophores to be located in the protein and even when those are present gives a limited spatial resolution. Likewise, CD spectroscopy is only sensitive to overall structural changes and provides little site specific information. FTIR spectroscopy coupled with isotope labeling can provide information on local structural changes.16 2DIR spectroscopy, which is analogous to 2D-NMR spectroscopy, can resolve dynamics down to the picosecond and femtosecond time scale. Isotope label 2DIR spectroscopy has already been proven useful for extracting the details of the local site specific structural information in membrane proteins17,18 as well as for investigating the amyloid fibril formation mechanism.19

Interpreting any complex isotope label 2DIR spectra and understanding the origin of the lineshapes are real challenges which require theoretical studies of the IR lineshapes in isotope labeled peptides. Here, we report the theoretical description of site-specific vibrational dynamics of a β-hairpin peptide and investigate to what extent it is possible to classify the peptide units based on their signature in isotope labeled 2DIR spectra. The effect of local distributions of conformations and dynamics on the site frequencies will be addressed. This will provide useful insight into the protein structure and folding studies using 2DIR spectroscopy with isotope labeling. We choose trpzip2 (Fig. 1A) as a model system, a synthesized artificial β-hairpin with four fluorescent tryptophan units with the amino acid sequence SWTENGKWTKW.20 This peptide was designed especially for fluorescence and CD spectroscopic studies of protein folding and has been subjected to several folding studies.21–26

For simple β-hairpin peptides mainly four different folding mechanisms have been suggested. These are the zip-out,27,28 hydrophobic collapse,29,30 reptation,31 and the hybrid zipper model.32 In the zip-out model folding begins with formation of the turn, and the inter-strand hydrogen bonds then form sequentially towards the terminals. The hydrophobic core forms at the end. In the hydrophobic collapse mechanism the hydrophobic core forms first, then the inter-strand hydrogen bond formation propagates in both directions. The reptation mechanism describes the sliding motion of one strand with respect to another strand of the β-sheet, initially formed with non-native hydrogen bonds, until it reaches its native state. In
folding kinetics of these peptides shows that they have although these peptides only differ by their turns, a study of the local environment on their spectra have been addressed.33,36

Individual solvent exposed sites, W2 and G7, and the influence of units. These zippers were found to have type II (trpzip1 and trpzip3) were designed, differing only in the turn sequence in the turn. Simultaneously two other zippers (trpzip2 and trpzip2B), where the serine (S1) and threonine (T10) residues were replaced by alanine, have been subjected to (un)folding study with time resolved temperature jump (T-jump) IR spectroscopy.41 (Un)folding time constants were found to be a few microseconds.

T-jump experiments with only these few isotope labeled sites of trpzip2 will be insufficient to distinguish among the proposed folding mechanisms. It is furthermore impossible to draw any solid conclusions from these studies on the general trends about the relation between the lineshape parameters and the local structure and dynamics. Hence, in this paper, we will focus on all the amide I sites of this peptide and investigate the variation of the site frequency and its standard deviation, full width at half maximum (FWHM) of the FTIR peak, the diagonal and anti-diagonal linewidth of 2DIR peaks as a function of the amide I oscillator’s position in the peptide. We will examine how these spectral lineshape parameters can be useful for classifying the amide I sites into distinguishable groups in trpzip2 and discuss the transferability of these results to other peptides or proteins as well as the possibility of using this classification in protein folding studies. We will address the structural heterogeneity of the turn using a clustering method42 to distinguish different configurations and examine the frequency distributions of some turn sites in these configurations.

The outline of the remainder of this paper is as follows. The methods used in MD simulations, clustering, and spectra calculations are described in section 2. In section 3, the results are presented and discussed. We make our final remarks and draw our conclusion in section 4.

2. Methods

MD simulations were performed at both neutral pH (7) where the glutamic acid residue is deprotonated and low pH (2.5) where it is protonated. We used the GROMACS-3.3.3 program43 with the implemented all atoms OPLS force field44 for the b-hairpin peptide and the SPC force field for water.45 The initial trpzip2 structure was taken from the Protein Database (PDB 1LE1). The N-terminal and C-terminal were capped with an acetyl group (ACE) and an amine group (NH2), respectively. These are the actual terminals used in the 2DIR experiments presented in ref. 34. The formal charge of the system is +1 at pH 7 and +2 at pH 2.5. Chlorine counter ions were added to keep the systems neutral. The structure (at both pH values) was solvated in heavy water and all acidic protons were replaced by deuterium. This is done because the bend vibration of normal water overlaps with the amide I region and the experiments are performed in heavy water to eliminate this spectral overlap. All MD simulations were performed with 2 fs time steps using the LINCS algorithm46 in order to constrain all bond lengths. The reaction
field method was applied for the treatment of long range electrostatic interactions with a cut-off of 1.4 nm. The structure was equilibrated for 2 ns at 300 K using the Berendsen method for temperature and pressure coupling. A production run for 20 ns was performed to obtain the trajectory for spectral simulation. During the production run snapshots were saved every 20 fs.

At every snapshot the vibrational Hamiltonian was constructed. We employ a Hamiltonian of the commonly used form:11

\[
H(t) = \sum_{i=1}^{N} \left[ \omega_i(t) b_i^\dagger b_i - \frac{\Delta_i(t)}{2} b_i^\dagger b_i b_i^\dagger b_i \right] + \sum_{ij} J_{ij}(t) b_i^\dagger b_j
\]

(1)

Here, \(b_i^\dagger\) and \(b_i\) are the Bosonic creation and annihilation operators for the amide I vibration on the site \(i\). \(\omega_i(t)\) and \(\Delta_i(t)\) are the fluctuating frequency and anharmonicity, respectively, of \(i\)th site, \(J_{ij}(t)\) is the fluctuating coupling between the \(i\)th and the \(j\)th site. The last term describes the interaction of the applied infrared laser field \(E(t)\) with the amide I site through the transition dipoles \(\mu_i\). The summations run over the \((N = 12)\) amide I sites in trpzip2.

The site frequencies \((\omega_i)\) were found with a Stark effect based approach using the electric field and its gradient on the C, O, N and D atoms. The frequencies were corrected for through bond effects with neighboring units, using a Ramachandran angle based nearest-neighbor frequency shift (NNFS) map.49 The anharmonicity needed for describing doubly excited states was set to 16 cm\(^{-1}\) on all units.11 The couplings between different units were found using the transition charge coupling (TCC) scheme.49 For the nearest-neighbor couplings the Ramachandran angle based NNC scheme49 was used. Parameterizations similar to the one we employed are widely used.50–55

After the construction of the time dependent Hamiltonian, the spectra were calculated using the numerical integration of Schrödinger equation scheme (NISE).56 In this scheme the Hamiltonian is considered to be constant for short time intervals. During each of these time intervals the time independent Schrödinger equation can then be solved. The time evolution for longer times is determined by the successive propagation in the short time intervals. Linear and nonlinear response functions are then calculated to obtain FTIR and 2DIR spectra.

For the 2DIR spectra we used a waiting time of 0 fs. This is the time between the pump and the probe at frequencies \(\omega_1\) and \(\omega_3\), respectively. An \(ad\ hoc\) vibrational lifetime (1 ps) is accounted for.57 For analysis purposes, the spectral calculation was performed in three different cases: (A) considering the frequency shift from the total electric field and the NNFS map, (B) considering the frequency shift from the electric field generated by only the peptide and the NNFS map, and (C) considering the frequency shift from the electric field generated by only the water. In all cases the electric field generated by charges within a radius of 20 Å from the amide I site was accounted for. The frequency shift was found to have converged for this distance.

For simplicity, when considering isotope labeling, we calculated the spectra of isolated sites, which means that the couplings between the isolated site and the other sites are neglected. This is a good approximation when the signal from the labeled unit is well separated from the main band. In the calculated spectra, we shifted the frequency by 41 cm\(^{-1}\) to the red, which corresponds to a \(^{13}\)C\(^{16}\)O label. Experimentally it is unfortunately not always possible to separate completely the absorption of the isotope labeled site from the main band. In the ESI (Fig. 1S)† we have displayed a comparison between the 2DIR spectrum obtained for the isolated labeled site T10 and the full spectrum with the site T10 labeled with \(^{13}\)C\(^{16}\)O at pH 2.5. In this case the spectral shape of the labeled site is only little affected by the presence of the main band. In case of an overlap between the \(^{13}\)C\(^{16}\)O labeled peak and the main peak, a \(^{13}\)C\(^{18}\)O label can be used to shift the frequency of the labeled site by an additional 19 cm\(^{-1}\). The difference between isolated and coupled isotope labeled sites has also been discussed in ref. 58.

We performed cluster analysis42 of the pH 7 MD trajectory to obtain structurally distinct sub-ensembles of the turn. This was done calculating the root mean square deviation (RMSD) of the positions of all atoms in the main chain of the turn (N, H, C\(_a\), C and O of E5, N6, G7 and K8). First, the RMSD of the chosen atoms between all pairs of conformations is calculated. For every conformation the number of neighbors which are conformations, for which the RMSD is less than or equal to a chosen cutoff (0.035 nm), is determined. The conformation which has the maximum number of neighbors is denoted the central structure of cluster 1 and that cluster is formed by this structure and all its neighbors. The conformations of the cluster 1 are then eliminated from the whole set of conformations and the procedure is repeated until the whole set is empty. In this way one obtains a series of structurally distinct clusters. More advanced clustering methods exist,59 but the present method is sufficient for our needs.

3. Results and discussion

In Fig. 2A a comparison between the experimental and simulated FTIR spectra of unlabeled trpzip2 is depicted. The simulated FTIR curves are red-shifted by 20 cm\(^{-1}\) to fit the experimental peak position. The main origin of this shift is a systematic overestimation of the amide I site frequencies, as will be explained below. Apart from this shift an excellent agreement between theory and experiment is found. The measured34 and simulated 2DIR spectra at pH 7 with parallel and perpendicular polarization of trpzip2 are given in Fig. 2B–E. The simulated cross peaks are somewhat too strong, indicating slight overestimation of the couplings among the sites, which gives rise to the off-diagonal anharmonicity resulting the cross peaks between the \(a^-\) and \(a^+\) states. These spectra support the conclusion that the overall structure of trpzip2 is the \(\beta\)-hairpin structure found by NMR.20 For site specific information we turn to isotope labeling.

We divide the 12 amide I sites of trpzip2 into four groups based on their location in the NMR structure20 of the hairpin-peptide (Fig. 1A): terminal (ACE), turn (N6 and G7), internal (S1, T3, E5, K8, and T10), and external (W2, W4, W9, and W11).
The internal sites are characterized by the fact that the carbonyl groups are pointing inwards and are involved in hydrogen bonding with an NH group on the opposite strand. Similarly, the external sites have the carbonyl groups pointing outwards and are involved in hydrogen bonding with the solvent water. The carbonyl groups of the terminal and turn sites are also exposed to the water solvent. From the MD trajectory at pH 7 we have calculated the actual average number of hydrogen bonds between the amide I CO for each site and the solvent ($H_{\text{solvent}}$) and between the amide I CO for each site and the rest of the peptide ($H_{\text{peptide}}$). These are presented in Fig. 3. It turns out that E5 in the MD simulation is behaving partially as an external site, with frequent hydrogen bonding to the solvent.

The average frequency and the standard deviation calculated for each isolates site (unlabeled) are given in Table 1 and plotted in Fig. 1B. The experimental assignments of the frequencies of T3, T10, W2, and G7 at pH 7 have been reported in the literature. Compared to these experimental values, our calculations predict frequencies higher by 13.3 and 12.1 cm$^{-1}$ for T3 and T10, respectively, and 17.5 and 13.2 cm$^{-1}$ for W2 and G7, respectively. The main origin of this discrepancy is systematic overestimation of the amide I site frequencies. This is due to the boundary between the amide group for which the electrostatic map was developed and the surrounding peptide is not well defined. Previous studies also found a systematic frequency shift for all sites with a magnitude depending on the force field charges.

Classification of the residues into four groups (terminal, internal, external and turn) makes sense when we look at the standard deviations of the site frequencies reported in Table 1. The terminal residue ACE has the largest standard deviation ($22.5$ cm$^{-1}$ at pH 2.5 or $26.3$ cm$^{-1}$ at pH 7), whereas internal residues except E5 have the lowest standard deviations, around 9–15 cm$^{-1}$ at both pH values. The high value for the E5 site is attributed to its position in the $\beta$-turn. The

![Fig. 2](A) Experimental and simulated linear absorption spectra of trpzip2 in D$_2$O. (B,D) and (C,E) are the experimental and simulated 2DIR spectra, respectively. (B,C) are for parallel and (D,E) are for perpendicular polarization. Simulated spectra are red shifted by 20 cm$^{-1}$ to fit the experimental ones. In the 2DIR spectra, a total of 20 equally spaced contours are plotted between 0 and 20% of the most intense peak. Blue (negative) contours correspond to simulated emission and bleaching, while red (positive) contours indicate excited state absorption.

![Fig. 3](Average number of hydrogen bonds between the amide I CO of the different sites and the solvent, $H_{\text{solvent}}$ (black) and between the amide I CO of the different sites and the peptide, $H_{\text{peptide}}$ (red). We have used the standard hydrogen bond criteria with angle cutoff for acceptor–donor–hydrogen of $30^\circ$ and distance cutoff for donor–acceptor distance of 0.35 nm.)

13.2 cm$^{-1}$ for W2 and G7, respectively. The main origin of this discrepancy is systematic overestimation of the amide I site frequencies. This is due to the boundary between the amide group for which the electrostatic map was developed and the surrounding peptide is not well defined. Previous studies also found a systematic frequency shift for all sites with a magnitude depending on the force field charges.

<table>
<thead>
<tr>
<th>Residues</th>
<th>pH 2.5</th>
<th>pH 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal</td>
<td>1647.2</td>
<td>1633.7</td>
</tr>
<tr>
<td>ACE</td>
<td>1617.1</td>
<td>1626.3</td>
</tr>
<tr>
<td>S1</td>
<td>1669.3</td>
<td>1663.4</td>
</tr>
<tr>
<td>T3</td>
<td>1674.3</td>
<td>1679.3</td>
</tr>
<tr>
<td>E5</td>
<td>1668.6</td>
<td>1672.8</td>
</tr>
<tr>
<td>K8</td>
<td>1672.4</td>
<td>1674.9</td>
</tr>
<tr>
<td>T10</td>
<td>1676.4</td>
<td>1680.1</td>
</tr>
<tr>
<td>External</td>
<td>1675.2</td>
<td>1678.5</td>
</tr>
<tr>
<td>W2</td>
<td>1675.3</td>
<td>1676.3</td>
</tr>
<tr>
<td>W4</td>
<td>1675.2</td>
<td>1671.1</td>
</tr>
<tr>
<td>W9</td>
<td>1691.4</td>
<td>1696.5</td>
</tr>
<tr>
<td>Turn</td>
<td>1676.4</td>
<td>1674.1</td>
</tr>
<tr>
<td>G7</td>
<td>1663.1</td>
<td>1658.2</td>
</tr>
</tbody>
</table>

The average frequency and the standard deviation calculated for each isolates site (unlabeled) are given in Table 1 and plotted in Fig. 1B. The experimental assignments of the frequencies of T3, T10, W2, and G7 at pH 7 have been reported in the literature. Compared to these experimental values, our calculations predict frequencies higher by 13.3 and 12.1 cm$^{-1}$ for T3 and T10, respectively, and 17.5 and 13.2 cm$^{-1}$ for W2 and G7, respectively. The main origin of this discrepancy is systematic overestimation of the amide I site frequencies. This is due to the boundary between the amide group for which the electrostatic map was developed and the surrounding peptide is not well defined. Previous studies also found a systematic frequency shift for all sites with a magnitude depending on the force field charges.

Classification of the residues into four groups (terminal, internal, external and turn) makes sense when we look at the standard deviations of the site frequencies reported in Table 1. The terminal residue ACE has the largest standard deviation ($22.5$ cm$^{-1}$ at pH 2.5 or $26.3$ cm$^{-1}$ at pH 7), whereas internal residues except E5 have the lowest standard deviations, around 9–15 cm$^{-1}$ at both pH values. The high value for the E5 site is attributed to its position in the $\beta$-turn. The
standard deviations for the external residues and the turn residues are more or less the same, around 17 cm\(^{-1}\). While those standard deviations are not directly observable experimentally, they may be compared with the measured lineshape parameters, such as width of the FTIR spectra, which are influenced by the time-dependence of the frequency fluctuations.

We have investigated the time scales of the frequency fluctuations in the system by calculating the normalized auto-correlation functions of the site frequencies \(C(t) = \frac{\langle \delta a_i(t) \delta a_i(0) \rangle}{\sigma_i \sigma_j}\) (\(\sigma_i\) is the standard deviation of the site frequency). These are plotted for the pH 7 trajectory in Fig. 4. We find that they can be fitted very well by a triexponential function:

\[
C(t) = x_1 \exp(-t/\tau_1) + x_2 \exp(-t/\tau_2) + x_3 \exp(-t/\tau_3) \tag{2}
\]

Here, the amplitudes \((x_i)\) are constrained to sum up to one. These and the time scales \((\tau_i)\) are listed in Table 2. Here three fundamentally different time scales are found: (67–124) fs, (2–13) ps and (223–2098) ps. In the remainder of this paper we will denote these three time scales as the fs, ps, and ns time scales, respectively. The internal sites, especially E5, S1 and K8 exhibit ns time scale dynamics with significant amplitudes (30, 32 and 26%, respectively), which can be attributed to slow local structural changes such as backbone conformational changes. For the external sites and the turn sites the fs time scale dynamics dominates. This dynamics can be mainly attributed to the hydrogen bond dynamics. The relative importance of the dynamics at the intermediate speed is similar for all sites: \(\tau_2\) varies from 0.14 to 0.32. We have found similar results at pH 2.5 (not shown).

The zero time cross correlations between site frequencies, defined as \(\frac{\langle \delta a_i(0) \delta a_j(0) \rangle}{\sigma_i \sigma_j}\), were calculated (Fig. 2S in the ESI). Here \(\sigma_i\) and \(\sigma_j\) are the standard deviations of \(i\)th and \(j\)th site frequencies, respectively. The frequencies of the nearest neighbor sites are slightly anticorrelated and the corresponding correlation coefficient is on average −0.11. The lack of correlation between the frequencies of different sites shows that these are not sensitive to the global structure, but mostly determined by the local hydrogen bonding environment. This can justify treating the site frequencies as largely independent in the models used in the simulations, which has frequently been done.\(^{39}\)

The average couplings were calculated from the Hamiltonian trajectory at pH 7 (Fig. 3S in the ESI).\(^{\dagger}\) The strongest coupling is found between the nearest neighbors along the strand and the units close to each other across the strand. These large cross-strand couplings are typical of the antiparallel \(\beta\)-sheet structure.

### 3.1 Site specific spectra

In the following we examine how the frequency fluctuations are reflected in the spectra of the individual sites and investigate how the spectra can be used to characterize the structure. As was noted above already, for several specific quantities the pH 2.5 and pH 7 results show little difference. Therefore, we focus on the spectral calculations at pH 7. In Fig. 5 the 2DIR spectra with parallel polarization calculated for all isolated sites at pH 7 are shown. These spectra were calculated for the case A introduced in section 2. Following the backbone we see a trend of alternation between diagonally broad and narrow spectra, whenever there is a switching between an external site and an internal site with the exception of E5. This is because the internal sites are more rigid as their carbonyl groups are involved in the hydrogen bonds in the interior of the peptide, which renders the local conformation relatively stable. The external sites interact with the solvent water molecules and have a broad frequency distribution due to the hydrogen bond fluctuations. In order to quantify the lineshapes we cut along the diagonal and \(anti\)-diagonal of the negative peak in the 2DIR spectra for all the sites and measure the full width at half maximum (FWHM) for each slice. A pictorial way of comparing these peak widths along with the FWHM of the FTIR spectra and the previously discussed standard deviations of the site frequencies is given in Fig. 6. Exact values of these widths are listed in Table 3. In the following subsections we will discuss these spectral features in detail for the individual groups we have defined for trpzip2.

<table>
<thead>
<tr>
<th>Residues</th>
<th>(x_1)</th>
<th>(\tau_1/)ps</th>
<th>(x_2)</th>
<th>(\tau_2/)ps</th>
<th>(x_3)</th>
<th>(\tau_3/)fs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terminal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE</td>
<td>0.08</td>
<td>137</td>
<td>0.20</td>
<td>3.7</td>
<td>0.72</td>
<td>98</td>
</tr>
<tr>
<td>T10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>0.30</td>
<td>2098</td>
<td>0.17</td>
<td>7.4</td>
<td>0.53</td>
<td>103</td>
</tr>
<tr>
<td>T3</td>
<td>0.21</td>
<td>829</td>
<td>0.14</td>
<td>5.3</td>
<td>0.65</td>
<td>67</td>
</tr>
<tr>
<td>E5</td>
<td>0.32</td>
<td>312</td>
<td>0.22</td>
<td>5.4</td>
<td>0.46</td>
<td>124</td>
</tr>
<tr>
<td>K8</td>
<td>0.26</td>
<td>370</td>
<td>0.19</td>
<td>13.2</td>
<td>0.55</td>
<td>98</td>
</tr>
<tr>
<td>W4</td>
<td>0.09</td>
<td>223</td>
<td>0.26</td>
<td>3.7</td>
<td>0.66</td>
<td>118</td>
</tr>
<tr>
<td>W9</td>
<td>0.07</td>
<td>354</td>
<td>0.27</td>
<td>2.5</td>
<td>0.66</td>
<td>100</td>
</tr>
<tr>
<td>W11</td>
<td>0.10</td>
<td>234</td>
<td>0.27</td>
<td>3.2</td>
<td>0.63</td>
<td>120</td>
</tr>
<tr>
<td>Turn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>0.08</td>
<td>367</td>
<td>0.30</td>
<td>2.1</td>
<td>0.62</td>
<td>90</td>
</tr>
<tr>
<td>W1</td>
<td>0.09</td>
<td>223</td>
<td>0.26</td>
<td>3.7</td>
<td>0.66</td>
<td>118</td>
</tr>
<tr>
<td>W5</td>
<td>0.07</td>
<td>354</td>
<td>0.27</td>
<td>2.5</td>
<td>0.66</td>
<td>100</td>
</tr>
<tr>
<td>W8</td>
<td>0.10</td>
<td>234</td>
<td>0.27</td>
<td>3.2</td>
<td>0.63</td>
<td>120</td>
</tr>
<tr>
<td>Turn</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W6</td>
<td>0.19</td>
<td>356</td>
<td>0.22</td>
<td>2.2</td>
<td>0.59</td>
<td>112</td>
</tr>
<tr>
<td>G7</td>
<td>0.04</td>
<td>244</td>
<td>0.32</td>
<td>1.6</td>
<td>0.64</td>
<td>96</td>
</tr>
</tbody>
</table>

*Fig. 4* Normalized auto-correlation functions for the site frequencies.
3.2 Spectral behavior of the internal sites

For the internal sites the diagonal widths of the corresponding 2DIR peaks lie in a wide range (between 15 and 27 cm\(^{-1}\)). The anti-diagonal widths are similar for all internal sites (between 7 and 8 cm\(^{-1}\)), apart from E5 where it is slightly larger (8.9 cm\(^{-1}\)). The different behavior observed for E5 is due to the fact that the hydrogen bond between the CO of this site and the NH group on the opposite strand is frequently broken and hydrogen bonding with the solvent water takes over making this site partially solvent exposed (see Fig. 3).

T3 and T10 show similar spectral behavior (Fig. 5). We choose T10 for the 2DIR spectra calculation considering the cases B and C introduced in section 2. In both cases, the 2DIR spectra for T10 are narrow along the diagonal (Fig. 7). This is because the T10 site is involved in the stable and rigid inter-strand hydrogen bond and does not get exposed to the solvent water molecules. Hence, T10 (and T3) shows the spectral signature of a true internal site.

E5 and K8 are the special internal sites which belong to the \(\beta\)-turn region and the strong influence of the turn conformational change on their spectra is obvious. The diagonal widths of the 2DIR spectra in Fig. 5 for E5 and K8 are visibly larger than that of a true internal site (T3). This comparison is made in more detail in the ESI (Fig. 4S). We will discuss E5 and K8 further in sub-section 3.4, where the spectra for the turn sites are discussed.

S1 is the last internal site. It is the closest one to the C-terminal and its carbonyl oxygen forms a hydrogen bond with the amide hydrogen of K12. The fluctuation of this hydrogen bond length is expected to have a strong influence on the spectra of S1. Fraying of the terminal causes an increase of the distance \(d_{S1-K12}\) between the carbonyl oxygen of S1 and the amide hydrogen of K12. The distribution of \(d_{S1-K12}\) from the 20 ns MD trajectories at both low and neutral pH is calculated and is shown in Fig. 8A. At pH 7, \(d_{S1-K12}\) is mostly distributed around 0.2 nm, but has a small population around 0.4 nm as well. At pH 2.5, this distance is entirely distributed around 0.2 nm. The partial fraying of the terminals seen in the pH 7 trajectory explains the broader spectrum of S1 at pH 7 (Fig. 8B–E). It is unclear if the difference between the two trajectories on this point is really due to the pH or simply due to the fact that the dynamics between the partially frayed structures and the strongly bound ones is slow and our trajectories are not long enough to sample this.

![Fig. 5 Simulated 2DIR spectra (parallel polarization) of all the sites of trpzip2 at pH 7. A total of 18 equally spaced contours are plotted between \(\pm 10\%\) and \(\pm 90\%\) of the most intense peak.](image)

![Fig. 6 Comparison between standard deviations and FWHMs for the 12 sites in trpzip2. The black circles are the scaled standard deviation \(2.355\sigma\). The red squares are the FWHM of the FTIR peaks. The blue diamonds and the green crosses represent the diagonal and the anti-diagonal widths of the negative 2DIR peaks, respectively.](image)

<table>
<thead>
<tr>
<th>Residues</th>
<th>2DIR</th>
<th>FTIR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DW</td>
<td>AW</td>
</tr>
<tr>
<td>Terminal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACE</td>
<td>27.1</td>
<td>14.1</td>
</tr>
<tr>
<td>Internal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>26.8</td>
<td>8.1</td>
</tr>
<tr>
<td>T3</td>
<td>18.5</td>
<td>6.9</td>
</tr>
<tr>
<td>E5</td>
<td>20.9</td>
<td>8.9</td>
</tr>
<tr>
<td>K8</td>
<td>23.3</td>
<td>7.4</td>
</tr>
<tr>
<td>T10</td>
<td>15.8</td>
<td>7.2</td>
</tr>
<tr>
<td>External</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2</td>
<td>16.1</td>
<td>9.3</td>
</tr>
<tr>
<td>W4</td>
<td>27.1</td>
<td>11.1</td>
</tr>
<tr>
<td>W9</td>
<td>16.9</td>
<td>9.4</td>
</tr>
<tr>
<td>W11</td>
<td>26.3</td>
<td>10.5</td>
</tr>
<tr>
<td>Turn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N6</td>
<td>26.3</td>
<td>11.0</td>
</tr>
<tr>
<td>G7</td>
<td>23.3</td>
<td>11.4</td>
</tr>
</tbody>
</table>

The different behavior observed for E5 is due to the fact that the hydrogen bond between the CO of this site and the NH group on the opposite strand is frequently broken and hydrogen bonding with the solvent water takes over making this site partially solvent exposed (see Fig. 3).

T3 and T10 show similar spectral behavior (Fig. 5). We choose T10 for the 2DIR spectra calculation considering the cases B and C introduced in section 2. In both cases, the 2DIR spectra for T10 are narrow along the diagonal (Fig. 7). This is because the T10 site is involved in the stable and rigid inter-strand hydrogen bond and does not get exposed to the solvent water molecules. Hence, T10 (and T3) shows the spectral signature of a true internal site.
3.3 Spectral behavior of the terminal and external sites

The terminal site ACE has the largest diagonal (27.1 cm\(^{-1}\)) and anti-diagonal widths (14.1 cm\(^{-1}\)). Hence, ACE has the widest frequency distribution, which results from the combination of solvent exposure and flexibility of the terminal. For the external sites the diagonal and anti-diagonal widths vary between 16 and 27 cm\(^{-1}\) and between 9 and 11 cm\(^{-1}\), respectively.

From the standard deviation, diagonal width, anti-diagonal width, and FWHM of the FTIR peaks for the external sites, we find that these spectral lineshape parameters have larger values for W4 and W11 than those for W2 and W9. The explanation for this difference is found in the number of hydrogen bonds between the carbonyl group of all external sites (CO) and solvent molecules (D\(_2\)O). The distribution of the number of hydrogen bonds is depicted in Fig. 9. W4 and W11 are forming two hydrogen bonds with the solvent more frequently than W2 and W9. This results in the broader spectra for W4 and W11.

Let us now consider the site W11 in more detail. Its 2DIR spectrum resulting from only the water electric field is much broader than that for only the peptide contribution (Fig. 7). This is because the carbonyl group is exposed to water and the interaction with the water molecules gives the dominant contribution to the frequency distribution. Fast frequency fluctuation resulting from the interaction of the external sites with the solvent water leads to the large values of the anti-diagonal widths.

3.4 Spectral behavior of the turn sites

The turn sites, N6 and G7 have almost identical spectral behavior. The diagonal and anti-diagonal widths are around 24 and 11 cm\(^{-1}\), respectively. For further analysis we choose G7 as representative of this group.

The carbonyl group of G7 is in the turn exposed to the solvent water. From the analysis of the 2DIR spectra accounting for the water and the peptide contributions separately (Fig. 7), it is seen that the water is responsible for the majority of the line broadening. It is clear from the spectrum for only the
peptide contribution that the frequency of this site is hardly affected by the structure of the turn.

As mentioned in subsection 3.2, here we will discuss the spectra of the special internal sites K8 and E5. The 2DIR spectrum of the K8 site (Fig. 5) is highly inhomogeneous and the calculated normalized auto-correlation function shows that the ns time scale dynamics has a considerable amplitude (0.26). This shows that there are multiple slowly interchanging states. In Fig. 7, the 2DIR spectra of the K8 site for only the peptide and only the water contributions, respectively, are presented and the corresponding simulated FTIR spectra are depicted in panel A of Fig. 10. Due to only the peptide contribution, three diagonal peaks (two intense peaks and one small peak) are observed in the 2DIR spectrum, whereas in the FTIR spectrum two peaks and one shoulder are seen. This indeed shows that multiple states exist for the K8 site. This peak structure is not seen in the 2DIR spectrum when only the water electric field is accounted for. The 2DIR spectrum of the E5 site at pH 7 (Fig. 5) has a two peak structure along the diagonal. The peptide effect splits the E5 site frequency into three peaks, whereas the water field generates an inhomogeneous spectrum (Fig. 7). The corresponding FTIR spectra are plotted in Fig. 10B. From these results we conclude that it is the peptide’s conformational change that is responsible for the multiple states of the K8 and E5 sites. We have analyzed the effect of the structural changes in the peptide on the these site frequencies calculated for the total contribution from the whole system. Both K8 and E5 are involved in the β-turn region, therefore we expect that structural changes in the turn influence their spectra and we will investigate that using the cluster analysis of the pH 7 MD trajectory.

From the cluster analysis on the turn region performed as described in section 2, 32 clusters are obtained. We only consider the four dominant clusters: cluster 1 (29%), cluster 2 (25%), cluster 3 (25%), and cluster 4 (8%). The configurations of the turn region for the central structures of these clusters are shown in Fig. 11. Each cluster is characterized by the Ramachandran angles of the turn residues. A β-turn isformed by four amino acid residues and characterized by the Ramachandran angles of the 2nd and 3rd residues in the sequence of these four residues. For the central and average structures of the first four clusters these Ramachandran angles are given in Table 4. We have listed the ideal values for the Ramachandran angles for the most common types of the β-turn in Table 5. The Ramachandran angles for other standard types of the β-turn can be found in ref. 61. The turns in cluster 1 and cluster 3 are close to type I’, but with high deviation from the ideal value of $\psi_{N6}$ and $\phi_{G7}$. A type II turn is observed in cluster 4, but, there are also large deviations from the ideal values of $\psi_{N6}$ and $\phi_{G7}$. The turn in cluster 2 does not match with any of the standard types of β-turns discussed in ref. 61. We saw that the E5 CO exhibits both hydrogen bonding with solvent water and the peptide backbone (see Fig. 3). It is a tempting thought that the hydrogen bonding pattern is strongly correlated with the clusters and water exposure is only present in some of the clusters. This is, however, not the case. Clusters 1, 2, and 3 all have around 0.75 hydrogen bonds with water, while cluster 4 is less water exposed than the other clusters with an average of 0.13 hydrogen bond to water. Clusters 2 and 4 both have around 1 hydrogen bond with the peptide on average, while clusters 1 and 3 on average have 0.24 and 0.4 hydrogen bond with the peptide, respectively.

In Fig. 12 we show the frequency distribution of K8 and E5 in the first four clusters. The clusters 1, 2, and 3 have different distributions of the K8 site frequency. Cluster 4 has the same K8 site frequency as cluster 3. The E5 site frequency is different in these clusters. Therefore the slow structural transitions among the clusters are responsible for the inhomogeneity of the K8 or E5 isotope labeled peaks. However, the clusters found here do not satisfactorily explain the double peak structure of the K8 site observed in experiment, as the measured frequency splitting in the double peak structure is not resolved by us; the peaks that are well resolved experimentally are hardly resolved in the simulation, when the total contribution is accounted for (Case A).

3.5 Discussion

From the discussions on the spectral behavior of each group we find that anti-diagonal widths for the sites with solvent exposed carbonyls are significantly larger than that of the
The first four clusters obtained from the pH 7 MD trajectory. Black, red, green, and blue indicate cluster 1, cluster 2, cluster 3, and cluster 4, respectively.

**Fig. 11** Configurations of the turn for the central structures of the first four clusters obtained from the pH 7 MD trajectory.

**Table 4** \((\Phi_{N6}, \Psi_{G7})\) angles of the central structures and the average \((\Phi_{N6}, \Psi_{G7})\) angles of the first four clusters. These values are in degree.

<table>
<thead>
<tr>
<th>Clusters</th>
<th>(\Phi_{N6})</th>
<th>(\Psi_{N6})</th>
<th>(\Phi_{G7})</th>
<th>(\Psi_{G7})</th>
<th>(\Phi_{N6})</th>
<th>(\Psi_{N6})</th>
<th>(\Phi_{G7})</th>
<th>(\Psi_{G7})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>54.3</td>
<td>56.9</td>
<td>119.9</td>
<td>-8.4</td>
<td>63.3</td>
<td>52.9</td>
<td>115.2</td>
<td>-8.1</td>
</tr>
<tr>
<td>2nd</td>
<td>76.8</td>
<td>-58.7</td>
<td>-131.6</td>
<td>1.4</td>
<td>73.9</td>
<td>-53.3</td>
<td>-125.0</td>
<td>-5.9</td>
</tr>
<tr>
<td>3rd</td>
<td>76.5</td>
<td>10.7</td>
<td>108.4</td>
<td>9.3</td>
<td>71.9</td>
<td>13.9</td>
<td>112.3</td>
<td>15.9</td>
</tr>
<tr>
<td>4th</td>
<td>-64.0</td>
<td>89.2</td>
<td>115.7</td>
<td>-8.5</td>
<td>-56.3</td>
<td>86.3</td>
<td>110.2</td>
<td>-3.2</td>
</tr>
</tbody>
</table>

**Table 5** The most common of the standard \(\beta\)-turns: ideal Ramachandran angles (in degree) of second and third of four amino acid residues forming the turn region.

<table>
<thead>
<tr>
<th>Types of (\beta)-turns</th>
<th>(\Phi_{i+1})</th>
<th>(\Psi_{i+1})</th>
<th>(\Phi_{i+2})</th>
<th>(\Psi_{i+2})</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>-60</td>
<td>-30</td>
<td>-90</td>
<td>0</td>
</tr>
<tr>
<td>I’</td>
<td>60</td>
<td>30</td>
<td>90</td>
<td>0</td>
</tr>
<tr>
<td>II</td>
<td>-60</td>
<td>120</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>II’</td>
<td>60</td>
<td>-120</td>
<td>-80</td>
<td>0</td>
</tr>
</tbody>
</table>

**Fig. 12** Distribution of the K8 (left) and E5 (right) site frequencies in the first four clusters. Black, red, green, and blue indicate cluster 1, cluster 2, cluster 3, and cluster 4, respectively.
between the magnitude of the B-factors and the observables in the 2D IR spectra. One reason for the incomplete correlation is that the solvent fluctuations play a large role in the spectra and do not contribute directly to the B-factor. Still our conclusion about the presence of large flexibility of the terminal and turn residues based on the 2D IR analysis is supported by the large B-factors for those sites.

Two sources exist for the deviation of the simulation results from the experimental observation. First, we used an empirical classical force field for MD simulations. A study of comparing four different classical force fields\(^{53}\) in a 20 ns long MD simulation of trialanine shows that the Ramachandran angle (\(\psi, \phi\)) distribution of trialanine is substantially different for different force fields. The results of MD simulations are sensitive to which particular force field is used. Second, we see from the frequency auto-correlation plots that internal sites exhibit slow dynamics ranging from the picosecond to the nanosecond time scale, with significant amplitudes. This shows that for tracking the slow local conformational changes and investigating their influence on the spectral behavior the simulations need to be on this time scale and preferably even longer. As for the dynamics reported in Fig. 4 (and Table 2) the reported time scales are 2 ns or faster, which is one order of magnitude shorter than the integration time, which should be sufficient. It is impossible to guarantee that we have sampled the complete relevant configuration space. In order to improve this we would need longer simulation times combined with replica exchange molecular dynamics.\(^{59,64,65}\) Finally, we note that we started the MD simulation from the NMR structure and sample in the neighborhood of this structure. We therefore expect that our results are representative for the most commonly occurring natural structure.

4. Conclusions

By studying the trpzip2 hairpin, we have investigated how isotope label 2DIR spectroscopy can be used as a tool to reveal the structure and dynamics of peptides. The amide I sites were classified into four groups according to their location in the hairpin: terminal, internal, external and turn. These groups could be characterized by the standard deviation of their amide I frequency fluctuation, which relates to the diagonal and the anti-diagonal widths of the peak of an isolated site in the 2DIR spectrum. Broadening of the spectra for the water exposed sites is mainly caused by the electric field generated by water on the amide I site. Relatively narrow spectra are seen for the sites with the carbonyl groups buried in the interior of the peptide, where they are protected from interacting with the solvent water. Quantifying the lineshapes by means of the spectral linewidths shows that the anti-diagonal widths make a clear distinction between the group of internal sites and the group of solvent exposed sites. We find a significant contribution to the frequency fluctuations on the nanosecond time scale for the internal sites. On the other hand the frequency fluctuations of the solvent exposed sites are dominated by the femtosecond timescale dynamics stemming from the water motion. The type I' turn originally observed in the NMR structure of trpzip2 is not preserved throughout the MD simulation. From the cluster analysis we find that it switches to a different \(\beta\)-turn and a non-standard turn resulting in multiple states that should be observable in the isotope label spectra for the E5 and K8 sites.

Partial fraying of the terminals leads to a broader 2DIR spectrum seen for S1 at pH 7 as compared to when the terminals do not fray at pH 2.5. Based on this phenomenon together with the spectral behavior of the internal and external sites we draw the conclusion that the (anti-)diagonal linewidth of isotope label 2DIR spectra will be an excellent tool for studying the folding dynamics in trpzip2 and other peptides. When the peptide unfolds, carbonyl groups buried in the interior of the peptide get exposed to the solvent, which will lead to broadening of the spectra in the anti-diagonal direction. One should therefore be able to distinguish experimentally between the suggested zip-out, hybrid zipper, hydrophobic collapse, and reptation mechanisms by T-jump experiments on trpzip2 with different internal carbonyl sites labeled.

Acknowledgements

We would like to acknowledge the Tokmakoff group in MIT for supplying the experimental data and useful discussions. T.L.C.I. acknowledges the Netherlands Organization for Scientific Research (NWO) for support through a VIDI grant.

References