Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Phys Chem B. Author manuscript; available in PMC 2017 June 23.

Published in final edited form as:

J Phys Chem B. 2016 June 23; 120(24): 5325–5339. doi:10.1021/acs.jpcb.6b02608.

Structure of Penta-Alanine Investigated by Two-Dimensional Infrared Spectroscopy and Molecular Dynamics Simulation

Yuan Feng[1], **Jing Huang**[2], **Seongheun Kim**[1], **Ji Hyun Shim**[2], **Alexander D. MacKerell Jr.**[2], and **Nien-Hui Ge**[1,*]

[1]Department of Chemistry, University of California at Irvine, Irvine, CA 92697-2025, USA

[2]Department of Pharmaceutical Science, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA

Abstract

We have studied the structure of $(Ala)_5$, a model unfolded peptide, using a combination of 2D IR spectroscopy and molecular dynamics (MD) simulation. Two different isotopomers, each bis-labeled with $^{13}C{=}O$ and $^{13}C{=}^{18}O$, were strategically designed to shift individual site frequencies and uncouple neighboring amide-I′ modes. 2D IR spectra taken under the double-crossed $\langle \pi/4, -\pi/4, Y, Z \rangle$ polarization show that the labeled four-oscillator systems can be approximated by three two-oscillator systems. By utilizing the different polarization dependence of diagonal and cross peaks, we extracted the coupling constants and angles between three pairs of amide-I transition dipoles through spectral fitting. These parameters were related to the peptide backbone dihedral angles through DFT calculated maps. The derived dihedral angles are all located in the polyproline-II (ppII) region of the Ramachandran plot. These results were compared to the conformations sampled by Hamiltonian replica-exchange MD simulations with three different CHARMM force fields. The C36 force field predicted that ppII is the dominant conformation, consistent with the experimental findings, whereas C22/CMAP predicted similar population for α_+, β, and ppII, and the polarizable Drude-2013 predicted dominating β structure. Spectral simulation based on MD representative conformations and structure ensembles demonstrated the need to include multiple 2D spectral features, especially the cross-peak intensity ratio and shape, in structure determination. Using 2D reference spectra defined by the C36 structure ensemble, the best spectral simulation is achieved with nearly 100% ppII population, although the agreement with the experimental cross-peak intensity ratio is still insufficient. The dependence of population determination on the choice of reference structures/spectra and the current limitations on theoretical modeling relating peptide structures to spectral parameters are discussed. Compared with the previous results on alanine based oligopeptides, the dihedral angles of our fitted structure and the most populated ppII structure from the C36 simulation are in good agreement with those suggesting a major ppII population. Our results provide further support for the importance of ppII conformation in the ensemble of unfolded peptides.

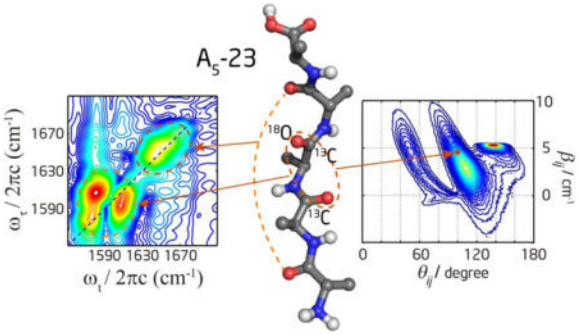
*To whom correspondence should be addressed. nhge@uci.edu. Phone: 949-824-1263. FAX: 949-824-8571.

Supporting Information Available

Details on the nearest-neighbor coupling map for amide-I modes, transition dipoles of amide-I local modes, isotope effects on the transition dipole, 2D spectral fitting, effects of multiple conformers and residual diagonal peaks, and vertical slices of 2D spectra in Figure 4 are available in the Supporting Information. This information is available free of charge via the Internet at http://pubs.acs.org

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

TOC image



1. Introduction

The solution structure of unfolded proteins and peptides has long been an intriguing problem. The unfolded state is common in peptides, intrinsically disordered proteins and even folded proteins. Structural knowledge of the unfolded state is needed to understand the conformational propensities of sequences and protein unfolding. In the last two decades, the notion that the unfolded state is much more restricted in the Ramachandran space compared to the random coil model has gained increasing experimental support.[1–22] These studies implicate polyproline II (ppII) as an important conformation in the ensemble of disordered states of peptides and proteins. This view, if confirmed, could impact the prediction of protein folding, and merit further experimental verification.

Alanine (Ala) based oligopeptides are simple yet important model systems to understand the intrinsic propensity of protein folding because Ala is abundant in proteins. However, determining the conformational sampling of these short peptides is challenging because ensembles of structures rapidly interconvert in solution.[19–20, 23–24] Multiple spectroscopic techniques have been used to investigate this problem, from the simplest systems of trialanine[10–12, 16–19, 25–26] and Ac-Ala-NHMe (Ac: acetyl)[27–29] with only two peptide bonds, to longer alanine-rich peptides including $(Ala)_n$ (n = 4 – 7)[6, 13–14, 20] and Ac-$X_2A_7O_2$-NH_2 (XAO, here X = diaminobutyric acid, and O = ornithine),[5, 15, 21–22] and host–guest peptides like GGXGG.[7–8] Kim et al. suggested ppII-like conformation for Ac-Ala-NHMe in aqueous solution, employing 2D IR in combination with DFT calculation.[28] Woutersen et al. determined (ϕ, ψ) on the central residue of $(Ala)_3$ to be ppII-like (−60°, 140°) using 2D IR spectral fitting.[17–18] They then estimated the ppII population to be 80% from the spectral diffusion experiment assuming the other conformer is α-helix.[19] Schweitzer-Stenner et al. investigated $(Ala)_3$ using a joint fitting procedure combining FTIR, isotropic and anisotropic Raman and vibrational CD spectra.[10] They suggested that the conformation is a 50:50 mixture of ppII and an extended β-strand-like structure,[11–12] which was also supported by their temperature-dependent CD study.[11–12] Later, by considering the conformational distribution explicitly, they updated the ppII population to 80%.[16] NMR can extract residue-specific information. Kallenbach and coworkers first showed that the seven alanine residues in XAO mainly adopted ppII at 2°C, and the β strand content increased by 10% at 55°C.[5] Later a joint NMR and MD simulation study of $(Ala)_n$ (n = 3 – 7) peptides

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

showed that the population distribution is almost independent from their chain length, with 90% ppII, 10% β, and no detectable population of the α conformation.[20] However, Hummer and coworkers studied $(Ala)_5$ conformations using MD simulations, and argued that the NMR data can be consistent with force fields that give a small α population and do not require exclusive formation of the ppII structure.[30] Cho and coworkers developed a self-consistent singular value decomposition based method to fit CD spectra and NMR J-coupling constant simultaneously, and determined the ppII population of the central residue in $(Ala)_3$ is 66%.[31]

Many studies on the structural propensity of alanine oligopeptides suggested a high ppII population, but the quantitative results depend on different molecular environments and spectroscopic methods utilized. Most, if not all, of the residue-specific studies on peptides containing more than two alanine peptide units were conducted by NMR spectroscopy. Although Schweitzer-Stenner et al. extended their joint fitting procedure to oligopeptides as long as XAO based on an excitonic coupling model,[13–14, 32] it still did not offer direct interrogation of a single residue like NMR. These studies also showed that it is essential to invoke a large set of spectral constraints to determine peptide conformation distribution. In NMR, although it has been argued that different Karplus parameters give different quantitative results,[30] several different J-coupling constants can be measured independently for each residue,[20] which makes structural determination more robust. Similarly, in vibrational spectroscopy, it is necessary to combine data from FTIR, isotropic and anisotropic Raman scattering, and vibrational CD to obtain results consistent with all techniques.[10–11, 13–16]

In this work, we investigated the conformation of $(Ala)_5$ by 2D IR spectroscopy to provide multiple spectral constraints and by isotope labeling to achieve residue specificity. This peptide has four peptide units and three pairs of dihedral angles describing the backbone conformation (Figure 1). To determine these angles in a site specific manner, we designed two bis-labeled pentapeptides: A_5-23 has ^{13}C and $^{13}C{=}^{18}O$ at the second and third peptide units, respectively, and A_5-43 has ^{13}C and $^{13}C{=}^{18}O$ at the fourth and third peptide units, respectively. This design enabled us to differentiate the four amide-I′ modes and strategically probe the interaction between specific pairs of vibrators. Spectral fitting was performed to determine the vibrational coupling and angles between transition dipoles. These spectral parameters were connected to structural parameters though the DFT-calculated transition dipole orientation and coupling map. The effects of multiple conformations on spectral features were explored. Previous studies have already demonstrated that 2D IR spectroscopy can be very useful for testing and validation of force fields.[24, 33–35] Here we performed MD simulation with three different CHARMM force fields and compared the results with the experimental 2D spectra and cross peak intensities. The relative efficacies of different force fields in generating realistic structural ensembles were analyzed and discussed. We also attempted to extract conformational population based on reference spectra defined by the MD simulated structural ensembles. The results were discussed in the context of the literature.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Experimental and Computation

2.1 Experimental details

Unlabeled (L-Ala)$_5$ was purchased from Sigma-Aldrich. Ala-(^{13}C)Ala-(^{13}C,^{18}O)Ala-Ala-Ala (A_5-23) and Ala-Ala-(^{13}C,^{18}O)Ala-(^{13}C)Ala-Ala (A_5-43) were purchased from CPC Scientific, Inc. The uncapped peptides were used without further purification. The isotope purity of the labeled peptides was examined with mass spectrometry. The ^{13}C labeling is almost 100%, and ^{18}O labeling purity is about 92% by comparing the signal intensity of (m +2)/z and m/z peaks.

Peptides were lyophilized in 5 mM DCl/D_2O solution at least three times before use in order to remove trifluoroacetic acid and deuterate the NH groups.[36] The peptide was then dissolved in pD = 1 DCl/D_2O solution to convert the carboxyl groups into acidic form to minimize the spectral overlap with amide-I′ modes. For FTIR and 2D IR measurements, the concentration of peptides solution was ~30 mM and the sample solution was held in between two CaF_2 windows with a 25 μm spacer. FTIR measurements were performed using a Nicolet magna-IR 860 spectrometer with a 2-cm^{-1} resolution and averaged over 32 scans. The solvent background and water vapor peaks were removed by programs written in MATLAB. Concentration dependent measurements were performed to confirm that no aggregation occurred.

The principles of 2D IR spectroscopy and our experimental setup for obtaining 2D IR spectra have been described in detail elsewhere.[37–39] Experimental parameters relevant to the current study are briefly described below. Ultrafast broadband mid-IR pulses were generated at 1 kHz with a pulse width of ~110 fs and spectral bandwidth of ~180 cm^{-1}. The mid-IR pulses were always tuned to be resonant with the center of peptide amide-I′ modes. For isotope-labeled peptides, the mid-IR wavelength was centered at around 1625 cm^{-1}, which is in the middle of the labeled and unlabeled regions. The effect of finite spectral width can be taken into account by a simple convolution procedure.[40] The 2D IR experiments were performed by focusing the excitation beams with wavevectors $\boldsymbol{k}_a$, $\boldsymbol{k}_b$, and $\boldsymbol{k}_c$ onto the samples in boxcar geometry. The nonlinear signal emitted in the phase-matching direction ($\boldsymbol{k}_d = -\boldsymbol{k}_a + \boldsymbol{k}_b + \boldsymbol{k}_c$) was spatially overlapped with the local oscillator pulse to do spectral interferometry with a 4-cm^{-1} resolution. The polarizations of all beams were controlled individually and denoted as $\langle a, b, c, d\rangle$. The signal was recorded as a function of two time intervals, τ (time interval between first and second pulses) and T (time interval between second and third pulses), and the detection frequency ω_t. The data was processed using the double Fourier transform methods[28, 41] to produce the 2D IR spectrum $S(\omega_\tau, \omega_t; T)$. We employed rephasing (R, a–b–c) and nonrephasing (NR, b–a–c) pulse sequences. The time intervals were controlled by moving the translation stages. τ was scanned from 0 to 3.6 ps for the rephasing sequence, and 0 to 3.0 ps for the nonrephasing sequence, with a step of 4.5 fs. T was set to 0 for the double-crossed $\langle \pi/4, -\pi/4, \mathrm{Y}, \mathrm{Z}\rangle$ polarization spectra, and 300 fs for the $\langle \mathrm{ZZZZ}\rangle$ and $\langle \mathrm{YYZZ}\rangle$ polarization spectra to minimize nonresonant solvent response. The local oscillator was set to precede the signal by 1.0 ps. Absorptive 2D IR spectra were obtained by summing the real part of rephasing and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

nonrephasing spectra that were properly phased against a dispersed broadband pump-probe spectrum measured at the same waiting time.[28, 42–43]

2.2 MD simulation

$(Ala)_5$ was solvated in a cubic water box of ~31.5 Å per side containing 988 water molecules and one Cl^- ion. MD simulations were carried out using three different force fields, including two additive (CHARMM22/CMAP[44–45] and its recent refinement CHARMM36[46]) force fields, and a fully polarizable Drude-2013 force field[47] based on the classical Drude oscillator model.

Hamiltonian replica exchange molecular dynamics (H-REMD) simulations were performed in the NPT ensemble at 293 K and 1 atm. The Hamiltonian was biased using a CMAP potential applied to all backbone ϕ and ψ dihedrals. For the zeroth (unbiased) replica the unperturbed CMAP from the corresponding CHARMM protein force field was used, whereas for the highest replica a biasing potential CMAP (bpCMAP) generated by inverting the 2D ϕ and ψ potential energy surface of Ac-Ala-NHMe was applied. Eight replicas were used for each H-REMD simulation using the additive force fields with linear scaling between the adjacent bpCMAPs, while twelve replicas were used for the Drude-2013 force field.

Periodic boundary conditions were applied and Lennard-Jones interactions were truncated at 12 Å with a smoothing function (force switch smoothing for the C22/CMAP and C36 simulations and switch smoothing for the Drude-2013) from 10 to 12 Å. The non-bonded interaction lists were generated with a distance cutoff of 16 Å and updated heuristically. Electrostatic interactions were calculated using the particle mesh Ewald method[48] with a real space cutoff of 12 Å on an approximately 1 Å grid with a sixth-order spline. Covalent bonds to hydrogen atoms were constrained by SHAKE.[49] Temperature control was performed based on the Nose-Hoover thermostats[50–51] and pressure control performed with the Andersen barostats.[52] For the polarizable simulations, the Drude particles were kept at low temperature (1 K) using a dual thermostat Nose-Hoover algorithm,[53] imposing the electronic degrees of freedom to approach the adiabatic SCF limit.

The integration time step was set to 2 fs for the additive simulations and 1 fs for the polarizable simulations. Exchanges were attempted every 1 ps, and coordinates were saved every 10 ps. The acceptance ratio for exchange between two replicas ranged from 20% to 50%. H-REMD simulations were run for 100 ns for the additive force fields and 60 ns for the polarizable Drude-2013 force field. Analysis was performed to check that the populations of the $(Ala)_5$ conformations had converged in the H-REMD simulation. The coordinates and velocities at every 100 ps during the last 50 ns H-REMD simulations were used to start NVT simulations that were run for 40 ps with the coordinates saved every 20 fs. This leads to 500 short MD trajectories for spectral calculations. All the simulations were carried out using CHARMM.[54] We checked the conformational distribution of the 500 short trajectories, and found that they gave very similar results to the last 50 ns of REMD simulation. The difference between the two is about 1–2%. All further analysis and calculations below were based on the 500 short trajectories.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3. Results and discussion

3.1 Experimental results of unlabeled $(Ala)_5$

As shown in Figure 2a, the four amide-I′ modes in $(Ala)_5$ are congested in a single FTIR band centered at 1650 cm^{-1}, with a small shoulder around 1670 cm^{-1}. The band at 1720 cm^{-1} is due to the carboxylic acid CO stretch.[55] The shoulder around 1670 cm^{-1} can be attributed to the N-terminal amide-I′ mode. It is blue shifted from the other amide-I′ modes due to the presence of the $-ND_3^+$ group, as suggested in previous studies.[17–18, 55] The broad and featureless FTIR spectrum does not intuitively provide useful information about the structures.

The 2D IR spectra of $(Ala)_5$ shown in Figure 2b–d also do not provide significant structural information. The rephasing spectrum under the ⟨ZZZZ⟩ polarization has only one broad peak. The nonrephasing spectrum has a higher resolving power for closely spaced spectral features,[37, 39, 56–57] and reveals two peaks at 1673 and 1648 cm^{-1}. The higher frequency peak corresponds to the first amide-I′ mode and the lower frequency peak corresponds to the second-fourth amide-I′ modes. There appears to be some cross peaks between the two peaks, but no internal spectral features can be distinguished within the broad lower frequency peak. We then measured the 2D IR spectra under the double-crossed $\langle \pi/4, -\pi/4, Y, Z \rangle$ polarization since it has been shown theoretically and experimentally that this polarization[33, 37–38, 58–60] and its variations[37, 61–62] can suppress the strong diagonal peaks and retain only the cross peaks. As shown in Figure 2d, the double-crossed spectrum exhibits a cross-peak pattern that has an intense and broad peak above the diagonal line and tails into two regions below the diagonal. This pattern is different from the doublet and multiple peak patterns observed in the 3_{10}- and α-helical structures, respectively.[38, 59] To model this four-vibrator system, a 4×4 matrix is needed to describe the one-exciton Hamiltonian which contains six independent coupling terms describing the interactions between amide-I′ local modes, four local frequencies, and some dynamics parameters. The poor peak separation and lack of spectral features in the unlabeled peptide spectra do not allow unique determination of these parameters.

3.2 Experimental results of labeled $(Ala)_5$

To resolve separate peaks, we designed two different isotope-edited peptides: A_5-23 and A_5-43. The first amide-I′ mode is close to the positively charged N-terminal, so it always has the highest frequency. The third amide-I′ mode is always double-labeled with $^{13}C{=}^{18}O$, so it has the lowest frequency. As shown in Figure 3, the isotope-labeled and unlabeled regions are well separated in FTIR spectra. However, within each amide-I′ band, two peaks are only partially resolved because the frequency separation is comparable to the linewidth. The frequency separation is about 20 cm^{-1} between two peaks within the same band, and about 45 cm^{-1} between the lower frequency peak in the unlabeled region and the higher frequency peak in the labeled region. The frequency ordering (from high to low wavenumber) for the four amide-I′ modes is 1-4-2-3 for A_5-23, and 1-2-4-3 for A_5-43, as labeled in Figure 3.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 4a–d and f–i show the rephasing and nonrephasing spectra collected under the ⟨ZZZZ⟩ and ⟨YYZZ⟩ polarization for the labeled $(Ala)_5$ peptides. The rephasing spectra are similar to the FTIR spectra, where the isotope labeled band is well separated from the unlabeled band, but within each band the two peaks are not well resolved. The nonrephasing spectra have better resolution, so all four peaks are resolved in both A_5-23 and A_5-43. In all 2D spectra the unlabeled band is stronger than the labeled band, which is also consistent with the intensity profile of the FTIR spectra. Since the mid-IR spectrum is centered at 1625 cm^{-1} (between the labeled and unlabeled regions), the difference should not come from the spectral bandwidth convolution effect. Previous studies[63–64] and our ab initio calculation on deuterated *N*-methylacetamide (see Section 3 in the Supporting Information) both showed that isotope labeled amide-I′ modes have smaller transition dipole strengths μ. Therefore, the difference is most possibly due to weaker transition dipole strength for the labeled amide-I′ modes. Because 2D IR intensity is proportional to μ^4, the contrast is more apparent compared to linear IR where the intensity is proportional to μ^2.

The cross peaks in 2D IR reflect vibrational couplings between different vibrational modes, which can be related to the dihedral angles describing the polypeptide conformation. The intensity of cross peaks, in the weak coupling limit, is proportional to the off-diagonal anharmonicity Δ_{ij}[65]

$$\Delta_{ij}=4\beta_{ij}{}^2\Delta/(\omega_i-\omega_j)^2 \quad (1)$$

where β_{ij} is the coupling constant between the local modes *i* and *j*, Δ is the diagonal anharmonicity of the local modes, and ω_i and ω_j are local mode frequencies. To show strong cross peaks, β_{ij} needs to be large whereas the frequency separation needs to be small. Our strategy was to isotope labeled some amide C=O bonds to introduce a larger frequency separation to simplify the cross peak patterns, so that we can use a set of smaller Hamiltonians to model the peptide instead of using the full Hamiltonians. In this labeling scheme, for the six pairs of nearest-neighbor amide-I′ modes in two isotopomers, only three pairs (1–2 pair and 3–4 pair in A_5-43, and 2–3 pair in A_5-23) have a frequency separation of ~20 cm^{-1}, whereas the other three pairs (2–3 pair in A_5-43, 1–2 pair and 3–4 pair in A_5-23) are separated by ~65 cm^{-1}. For (*i*, *i*+2) pairs, the frequency separation between the 1–3 pair of amide-I′ modes is always ~85 cm^{-1}, and that between the 2–4 pair is always ~45 cm^{-1}.

In the rephasing spectra in Figure 4, the cross peaks are buried under the strong diagonal peaks. In the nonrephasing spectra, the cross peaks are more visible but still overlap with strong diagonal peaks. In the 2D absorptive spectra plotted in the top row of Figure 5, the cross peaks are less clearly discerned from the strong diagonal peaks compared to the nonrephasing spectra. In both the absolute magnitude and absorptive spectra, cross peaks are more pronounced in the ⟨YYZZ⟩ polarization than the ⟨ZZZZ⟩ polarization. When comparing the nonrephasing spectra at different polarizations, the peak patterns between the 1–2 pair in the unlabeled region and the 3–4 pair in the labeled region of A_5-43 change significantly between the ⟨ZZZZ⟩ and ⟨YYZZ⟩ polarizations, so do the pattern between the 2–3 pair in the labeled region of A_5-23, indicating the presence of strong cross peaks between these nearest-neighbor amide-I′ modes. In contrast, the peak patterns between the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

1–4 pair in the unlabeled region of A_5-23 are very similar in the ⟨ZZZZ⟩ and ⟨YYZZ⟩ polarizations, suggesting that the cross peaks between these two amide-I′ modes are quite weak.

The existence and relative strength of cross peaks can be further examined by the double-crossed rephasing spectra. Figure 4e and j show three pairs of strong cross-peak doublets together with residual diagonal peaks: the 2–3 pair in the labeled region of A_5-23, the 1–2 pair in the unlabeled region and the 3–4 pair in the labeled region of A_5-43. Between the 1–4 pair in A_5-23, only the residual diagonal peaks are observed which excludes the possibility of strong coupling. Except for the 1–4 pair, the cross peaks between two modes within the same region (labeled or unlabeled) are stronger, whereas the cross peaks between a mode from the labeled region and another mode from the unlabeled region are much weaker. Figure S3 in the Supporting Information displays slices of the double-crossed spectra to more clearly illustrate the different strengths in the cross peaks. It is clear from these spectra that the three pairs of amide-I′ modes that show strong cross peaks are nearest neighbors and have about 20-cm^{-1} frequency separation. The other pairs exhibit much weaker peaks either due to a large frequency separation, or a small β_{ij}, or a combination of two factors. Figure S3 also includes slices of the nonrephasing spectra to illustrate how cross-peak intensities change with polarization. Large changes are seen for the cross peaks between nearest-neighbor modes within the same region, but not for those from different regions. This behavior is consistent with the results from the double-crossed polarization spectra.

Although there are still diagonal peaks remaining in the double-crossed polarization spectra that are not fully suppressed, they are relatively weak compared to the cross peaks. The residual diagonal peaks may originate from several reasons, including the non-collinear beam geometry, some nonrephasing contribution from smaller τ when pulses overlap, and internal rotation of peptide molecules. In our setup, the diagonal peak can be suppressed to 0.6–1% with the double-crossed polarization for a single amide-I mode.[38] For $(Ala)_5$, the intensity of the cross peaks under the double-crossed polarization are about 2% of the intensity of the diagonal peaks under the ⟨ZZZZ⟩ polarization, about 2–3 times the intensity of the residual diagonal peaks.

3.3 Conformational analysis by fitting the 2D spectra

As discussed above, only three pairs of amide-I′ modes (2–3 pair in A_5-23, 1–2 pair and 3–4 pair in A_5-43) have strong cross peaks. As a first order approximation, the four-oscillator system can then be effectively reduced into three sets of two-oscillator systems in the two isotopomers. The cross peaks between them contain information of dihedral angles that are pertinent to the middle three residues. Because the frequency separation is only about 20 cm^{-1} which is comparable to the diagonal peak linewidth, there is a lot of diagonal peak contribution at the cross peak positions. This renders direct extraction of structural information from cross peaks impossible. We fit the experimental 2D spectra using a vibrational exciton model and a sum-over-states approach to extract the structural parameters (using the same equations as reported in ref 38 to calculate the R and NR spectra). One-exciton Hamiltonians and the transition dipole operator were constructed in the site basis with ω_i as the diagonal elements, β_{ij} as the off-diagonal elements in the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Hamiltonian, and θ_{ij} as the angle between two transition dipoles. Assuming Bloch dynamics, the spectral line shapes were broadened by homogeneous and inhomogeneous contributions. The homogeneous broadening was described by γ and γ', the dephasing rate for the v = 0→1 and 1→2 transitions, respectively. The static inhomogeneous distribution of the site energies were approximated by 2000 sets of uncorrelated normally distributed random frequencies centered at ω_i and ω_j with standard deviations of σ_i and σ_j, respectively. The two-exciton Hamiltonian was then built from the one-exciton Hamiltonian with the harmonic approximation. Diagonalization of the Hamiltonians provides the eigen-frequencies and the corresponding transition dipoles in the exciton basis. 2D spectra were calculated with the exciton basis in the frequency domain, averaged over the inhomogeneous distribution and finally convoluted with the mid-IR spectrum. There are eight free parameters: center frequencies ω_i and ω_j, coupling constant β_{ij} and angle θ_{ij} between two transition dipoles, inhomogeneous widths σ_i and σ_j, and dephasing rates γ and γ'. Because the ^{18}O isotope purity is 92%, the spectral fitting also included 8% isotopomers bis-labeled with ^{13}C=O.

To maximize the information content extracted from the experimental spectra, we fit the model to either the absolute magnitude spectra or the absorptive spectra and check their consistency. The absolute magnitude spectra have individual rephasing and nonrephasing components, but do not have phase information. The absorptive spectra have phase information, but rephasing and nonrephasing components are combined. The spectra under the $\langle$ZZZZ$\rangle$ and $\langle$YYZZ$\rangle$ polarization were given the same weighting in the fitting. Because we used a two-oscillator model, the off-diagonal regions between isotope labeled and unlabeled bands were not included in the fitting. The spectral regions for fitting are inside the blue square boxes, as illustrated in Figure 5. Either the spectrum inside each box was fitted individually with a single two-oscillator model, or the spectra inside both boxes were fitted together by adding up contributions from two two-oscillator models. We fitted both ways to check the consistency because the isotope labeled and unlabeled bands partially overlap and interfere with each other, which mainly affects the diagonal peaks rather than the cross peak region. When fitting the spectrum inside each box individually, multiple initial guesses were used to find the global minimum. When fitting the spectra inside both boxes together, the fitted parameters from individual box fitting were used as the initial guess. The Levenberg–Marquardt algorithm was used to do least-square fitting, and the final fitted parameters (more details are provided in Section 4 of the Supporting Information) are β_{12} = 4.3 +1.3/−1.2 cm^{-1} and θ_{12} = 101° +10°/−8°, β_{23} = 4.6 +2.0/−2.6 cm^{-1} and θ_{23} = 100° +9°/−8°, and β_{34} = 4.4 +1.8/−2.4 cm^{-1} and θ_{34} = 100° +10°/−9°. The estimation of the confidence interval was performed by changing the parameters until the fitting error was doubled, similar to the method adopted in literatures.[66] The fitted parameters are very similar across different residues.

3.4 Conformational distributions from MD simulations

We analyzed the conformation distributions of MD trajectories based on three secondary structure definitions commonly used in the MD community,[30, 46]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\begin{aligned}
\alpha_+ &: \ -160^\circ<\phi<-20^\circ \text{ and } -120^\circ<\psi<50^\circ \\
\beta &: \ -180^\circ<\phi<-90^\circ \text{ and } 50^\circ<\psi<180^\circ \text{ or} \\
&\quad -180^\circ<\phi<-90^\circ \text{ and } -180^\circ<\psi<-120^\circ \text{ or} \\
&\quad 160^\circ<\phi<180^\circ \text{ and } 110^\circ<\psi<180^\circ \\
\text{ppII} &: -90^\circ<\phi<-20^\circ \text{ and } 50^\circ<\psi<180^\circ \text{ or} \\
&\quad -90^\circ<\phi<-20^\circ \text{ and } -180^\circ<\psi<-120^\circ
\end{aligned} \tag{2}$$

Here (ϕ, ψ) are the peptide backbone dihedral angles, α_+ is a broader definition of α helix often used in the force-field related discussion.[46] Although $(Ala)_5$ has five residues, only the middle three residues have both well-defined ϕ and ψ based on which secondary structures are defined. The population fractions of each secondary structure for the middle three residues are listed in Table 1. From the table, it is clear that different force fields predict very different conformational distributions. C22/CMAP force fields predict almost equal population for the three secondary structures, C36 force fields predict more ppII and less α_+ compared to C22/CMAP, whereas the Drude-2013 force fields predict predominately β structure. The conformational distributions across the three residues are similar in the C36 force fields, but there are some variations in the C22/CMAP and Drude-2013 force fields.

3.5 Comparison of fitting results with MD

The spectral parameters obtained by fitting experimental spectra are nearest-neighbor coupling constant β_{ij} and angle θ_{ij}, whereas the MD simulation gives conformational distributions. To convert conformational parameters into spectral parameters, we invoke the mapping shown in Fig. S1 of Supporting Information. For every MD snapshot, β_{ij} was determined using the nearest-neighbor coupling map (Table S1), and θ_{ij} was calculated from the directions of transition dipoles on neighboring amide units. In this way, each single conformation in the Ramachandran plot can also be represented by a point in the coupling-angle plot which is directly related to the observables from experiment. Figure 6 shows the overall conformation distribution for all three middle residues predicted by different force fields, in both the Ramachandran plot (top row) and β_{ij}–θ_{ij} plot (second row). To more clearly illustrate how the different regions in the β_{ij}–θ_{ij} plot correspond to the secondary structures defined in eq (2), the points within each colored box in the top row are plotted in the corresponding β_{ij}–θ_{ij} plot in the third row of Figure 6. It is clear that the mapping between two plots is very complex. The relatively narrow distribution of ppII in the Ramachandran plot corresponds to a relatively broad distribution in the β_{ij}–θ_{ij} plot, whereas the relatively broad distribution of β structure in the Ramachandran plot translates to a very narrow distribution in the β_{ij}–θ_{ij} plot. The latter is the reason why the β structure always gives rise to the strongest peak intensity in Figure 6d–f even when its total population is not the highest.

On the basis of above discussion, we can invert the problem and find the corresponding Ramachandran angles from the (β_{ij}, θ_{ij}) parameters obtained from spectral fitting. In general the inversion problem is neither unique nor robust since a single set of (β_{ij}, θ_{ij}) can correspond to more than one set of (ϕ, ψ) (see, for example, Figure 4 in ref[17]), and a small change in (β_{ij}, θ_{ij}) might lead to a big change in (ϕ, ψ). However, based on the realistic

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

conformational distribution sampled by MD simulation, we can find the unique distribution of (ϕ, ψ) that corresponds to the confidence interval of ($\beta_{ij} \pm \Delta\beta$, $\theta_{ij} \pm \Delta\theta$) from fitting. The results are plotted in the fourth row of Figure 6. From the plots, it is clear that the distribution of (ϕ, ψ) allowed by the confidence interval of fitting parameters are mostly within the definition of ppII for all three residues. The allowed (ϕ, ψ) distribution is very asymmetric due to the nonlinear nature of mapping between (β_{ij}, θ_{ij}) and (ϕ, ψ). The center of the allowed (ϕ, ψ) distribution is then the solution of the inversion problem. Taking for example the fitted parameters β_{12} = 4.3 cm^{-1} and θ_{12} = 101°, denoted as the white dot in Figure 6e, the corresponding (ϕ_2, ψ_2) = (−60°, 150°). It is only (1°, 5°) off from the most populated point (−61°, 145°) in the C36 MD simulation. The dihedral angles for other two residues are only 1~2 degrees off from (ϕ_2, ψ_2), which is a natural consequence of similar fitted parameters. It should also be noted that, in the spectral fitting, (β_{ij}, θ_{ij}) and ($-\beta_{ij}$, 180° $-\theta_{ij}$) give exactly the same results. The fitted parameters β_{ij} = 4.3 cm^{-1} and θ_{ij} = 101° can also be β_{ij} = −4.3 cm^{-1} and θ_{ij} = 79°. However, the latter set of parameters corresponds to a structure that is physically unrealizable and is very far from any distribution in the β_{ij}–θ_{ij} plot. Therefore, comparing the results from spectral fitting and MD simulation, we can conclude that the C36 force field yields the best agreement with the experimental data for $(Ala)_5$ by predicating the most ppII population (Table 1), while the Drude-2013 and C22/CMAP force fields are in poorer agreement.

Further examination of conformational distribution in the β_{ij}–θ_{ij} plot shows that these two spectral parameters play different roles in the determination of secondary structures. Figure 7 shows the histogram of β_{ij} and θ_{ij} taken from the C36 simulation in Figure 6e. It is apparent that the distributions of coupling constants from different secondary structures significantly overlap with each other, whereas the distributions of angles can be better distinguished from each other.

To check the validity of two-oscillator model, we also calculated the coupling constant and angle between transition dipoles for non-nearest-neighbor amide-I modes using the C36 MD simulation trajectories.[67] The coupling constant was calculated by the transition charge coupling method.[68] The distribution of the coupling constant for 1–4 pair is most populated at 0.5 cm^{-1} and they can be considered as totally uncoupled. The distributions of coupling constants for 1–3 and 2–4 pairs are most populated at about −2.2 cm^{-1}, smaller than the nearest-neighbor coupling. Because the frequency separations between the (i, i+2) pairs are 2–4 times larger than that between the nearest neighbors, the effect of their coupling can be neglected. These results are consistent with the observed cross peak patterns in our double-crossed spectra in Figure 4(e, j), and also the slices shown in Figure S3.

3.6 Spectral signatures of different conformers

To demonstrate the dependence of linear and 2D IR spectra on peptide conformations, we performed spectral calculation for the 1–2 pair of A_5-43 using a single representative conformation of ppII, α_+ and β structure. Figure 8 shows the comparison with the measured and fitted spectra. The representative conformation was chosen as the most populated position in the Ramachandran plots in Figure 6(g–i). The parameters for ppII are (ϕ, ψ) = (−61°, 145°) and (β_{ij}, θ_{ij}) = (3.0 cm^{-1}, 109°). The parameters for β are (ϕ, ψ) = (−156°,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

156°) and (β_{ij}, θ_{ij}) = (5.3 cm^{-1}, 139°). The parameters for α_+ are (ϕ, ψ) = (−65°, −42°) and (β_{ij}, θ_{ij}) = (5.5 cm^{-1}, 46°). For other spectral parameters, we used the fitted values listed in Table S2 in the Supporting Information. The results clearly show that ppII conformation produces spectra that closely resemble the experimental and fitted spectra, but β (α_+) conformation produces spectra that are too intense on the low (high) frequency side.

So far we fit the experimental spectra with a single set of β_{ij} and θ_{ij}, and compare the experimental spectra with the simulated spectra of some representative conformations which are also described by a single set of β_{ij} and θ_{ij}. However, real peptides can adopt multiple conformations, and the measured spectra will be a population-weighted sum of contributions from different conformations. In that case, the fitted β_{ij} and θ_{ij} may not necessarily correspond to any real structure, although it is quite close to ppII here. Therefore, getting reasonable results from single conformation fitting is not sufficient. As shown in Figure 8(d, e), the β conformation produces spectra with the intensity mainly located at the low frequency, and α_+ conformation produces spectra with the intensity mainly located at the high frequency. Therefore, a question is raised: whether the peptide can adopt a mixture of β and α_+ conformations rather than mainly ppII; and whether there are some spectral observables of ppII that cannot be a weighted-average of the other two conformations.

As discussed in Section 3.4 and illustrated in Figure 7, the three conformations are better distinguished by θ_{ij}, and hence we first focus on the spectral observables as a function of θ_{ij}.[60] One useful spectral observable is the cross-peak anisotropy, which is given by

$$R_{cross}=\frac{I_{ZZZZ}-I_{YYZZ}}{I_{ZZZZ}+2I_{YYZZ}} =\frac{1}{5}\left(\frac{13P_2+5}{2P_2+7}\right)=\frac{13\cos^2\theta'_{ij}-1}{10\cos^2\theta'_{ij}+20} \quad (3)$$

in the weak coupling limit for the experimental configuration used in this study. Here I_{ZZZZ} and I_{YYZZ} are the cross-peak intensity in absorptive 2D IR spectra taken with the ⟨ZZZZ⟩ and ⟨YYZZ⟩ polarization, respectively. $P_2 = P_2(\cos\theta'_{ij})$ is the second order Legendre polynormial. θ'_{ij} is the angle between transition dipoles in the eigen basis. In the weak coupling limit, θ'_{ij} is very close to θ_{ij} or 180°− θ_{ij}, (this ambiguity comes from the fact that eigenvectors are indifferent to sign), and hence R_{cross} exhibits basically the same dependence on θ'_{ij} and θ_{ij}. R_{cross} is at the lowest value at 90° which is close to the ppII conformation (around 110°). Because both α_+ (around 45°) and β (around 140°) give higher values, they cannot be mixed to get the same value as ppII. When the frequency separation between the diagonal peaks is large enough, R_{cross} can be determined directly from measured cross peak intensities.[17] Unfortunately, in the current system the intensity of cross peaks cannot be obtained directly because they are in close proximity to the strong and broad diagonal peaks. They can only be obtained by fitting over the entire range including the strong diagonal peaks. Therefore, the result on R_{cross} is nevertheless obscured.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Another choice is to use the double-crossed polarization by which the diagonal peaks are mostly removed. We did not include the double-crossed polarization spectra in our fitting due to the incomplete suppression of the diagonal peaks. Using the fitted parameters, we simulated the double-crossed polarization spectra and checked that the consistency with experimental spectra in the cross peak region is reasonable (Figure 8b). For comparison, we also calculated the double-crossed spectra of representative conformations in Figure 8, and included another spectra with (β_{ij}, θ_{ij}) = (3.0 cm^{-1}, 46°) which has the same θ_{ij} as that of the representative α_+ conformation, but the same β_{ij} as that of the representative ppII conformation. Comparing Figure 8d with e, and Figure 8c with f, we found that the shape of double-crossed spectra is invariant with different θ_{ij}. However, comparing Figure 8e with f, the shape of double-crossed spectra is very sensitive to the change in β_{ij}. This different dependence of spectral shape on θ_{ij} and β_{ij} has been further confirmed by simulating the double-crossed spectra with many different combinations of (β_{ij}, θ_{ij}). This behavior seems to be a special case for two-vibrator systems only, because it has been previously shown that the shape of double-crossed spectra for many-vibrator systems are very sensitive to structures.[33, 38–39] Compared to the experimental spectrum (Figure 8a), with the cross peaks above the diagonal line being normalized to the same intensity, the cross peaks below the diagonal line in the ppII and (β_{ij}, θ_{ij}) = (3.0 cm^{-1}, 46°) spectra (Figure 8c and f, both have β_{ij} = 3.0 cm^{-1}) are marginally stronger than the experiment, but they are weaker and smeared in the α_+ and β spectra (Figure 8d and e, β_{ij} ~ 5.4 cm^{-1}). To show a clear doublet cross-peak feature, β_{ij} needs to be smaller than 5 cm^{-1}, although this condition alone is not sufficient to rule out any conformations because the distributions of β_{ij} are very broad and significantly overlap (see Figure 7b).

On the other hand, the intensities of the cross peaks are sensitive to both θ_{ij} and β_{ij}. In the weak coupling limit, the orientation factor for cross peaks under the double-crossed polarization is $\left(\sin^2\theta'_{ij}\right)/6$ which peaks at $\theta'_{ij} \sim \theta_{ij}=90°$. Let us define a quantity, $R_{\text{cross/diag}}$, that is dependent on both θ_{ij} and β_{ij}:

$$R_{\text{cross/diag}}=\frac{I_{\text{cross},\ \langle \frac{\pi}{4},\ -\frac{\pi}{4}\ ,\ \text{Y},\ \text{Z}\rangle}}{I_{\text{diag},\langle \text{ZZZZ}\rangle}} \quad (4)$$

Here the numerator is the maximal intensity of the cross peaks in the absolute magnitude $\langle \pi/4, -\pi/4, \text{Y}, \text{Z}\rangle$ rephasing spectrum, and the denominator is the maximal intensity of the diagonal peaks in the absolute magnitude $\langle \text{ZZZZ}\rangle$ rephasing spectrum. The experimental $R_{\text{cross/diag}}$ ratio is 2.2 ± 0.2% (The double-crossed polarization spectrum was taken at T = 0, but the parallel polarization spectrum was taken at T = 300 fs. The fitted and simulated values reported below are corrected by a factor of 0.74 to account for the ~ 1-ps population decay). Using the fitted parameters (β_{ij} = 4.3 cm^{-1}, θ_{ij} = 101°), we obtained $R_{\text{cross/diag}}$ = 2.4%, comparable to the experiment. Although the double-crossed spectrum contains residual diagonal peaks, their effect on the $R_{\text{cross/diag}}$ ratio is estimated to be less than 10% (see Section 5 in Supporting Information).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

If β_{ij} is fixed at the fitted value and θ_{ij} takes the values of the representative conformations, $R_{\text{cross/diag}}$ is 2.0% for ppII (θ_{ij} = 109°), close to the experimental result, but it becomes 1.2% for α_+ (θ_{ij} = 46°) and 0.77% for β (θ_{ij} = 139°), much smaller than the experiment. If β_{ij} is allowed to arbitrarily vary, it needs to be 6.1 cm^{-1} for α_+, and 9.2 cm^{-1} for β in order to reach $R_{\text{cross/diag}}$ = 2.0%, but the doublet feature in the double-crossed spectra will completely disappear with these large values of β_{ij}. To see whether it is possible to mix α_+ and β conformers in different ratios to increase $R_{\text{cross/diag}}$, we tested many mixing ratios under the constraints of maintaining the experimental diagonal profile of ⟨ZZZZ⟩ and ⟨YYZZ⟩ spectra and the double-crossed spectral shape. It should be noted that when mixing α_+ and β conformers, the resultant $R_{\text{cross/diag}}$ is not a simple population-weighted average of their respective values because the maximal diagonal peaks of the two conformers are located at different frequency positions. We found that it is impossible to find a mixing ratio that gives a higher $R_{\text{cross/diag}}$ than that of the individual conformer (see Section 5 in Supporting Information), and hence the $R_{\text{cross/diag}}$ of the mixture is always much lower than the experimental value. Therefore, based on the double-crossed spectral shape and $R_{\text{cross/diag}}$, we can rule out the possibility that the spectral features well-represented by the fitted ppII conformation actually resulted from a weighted average of α_+ and β conformations.

3.7 Spectral simulation with MD trajectories

In this section, we will simulate 2D IR spectra by the sum-over-states method using the fluctuating conformations (ϕ, ψ) from the MD simulation trajectories. Each structure corresponds to one (β_{ij}, θ_{ij}) with normally distributed random local mode frequencies and other dynamic parameters taken from the fitting. The results are shown in Figure 8g–i. Compared to the simulation above for single representative conformations, the calculation here incorporates the conformational heterogeneity sampled in the MD simulation. The spectra simulated using Drude-2013 force fields MD trajectories are very similar to those of the β representative conformation, and are not consistent with the experimental spectra. This is expected because Drude-2013 force field predicts 86% β conformation. The spectra simulated by the additive C22/CMAP and C36 force fields are very similar and both consistent with the experimental and fitted spectra, even though they predict different conformational distributions. However, the ensemble-averaged $R_{\text{cross/diag}}$ is 1.59% for the C36 simulation, 1.57% for C22/CMAP, and 1.22% for Drude-2013, all of which are quite off from the experimental and fitted values. These results suggest that none of the three MD conformational ensembles is an ideal representation of the actual conformational distribution.

When comparing the simulation results between C22/CMAP and C36, it is somewhat unexpected that they predict very similar spectral shape as well as $R_{\text{cross/diag}}$ values. As we discussed in the previous section, the ppII spectra under the parallel and perpendicular polarizations can be somewhat approximated by a weighted average of α_+ and β spectra, which rationalizes the similarity in spectral shape between the C22/CMAP and C36 simulations. Although C22/CMAP has a lower percentage of ppII than C36, the higher percentages of α_+ and β combined in a way that gives spectral features similar to C36. However, it is not obvious why $R_{\text{cross/diag}}$ values are also very similar. To elucidate the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reason, we simulated the spectra averaged over the MD simulated sub-ensemble of each conformer (within the range inside the boxes in Figure 6a–c). The spectra are shown in Figure 8j–l. We found the ratio to be 1.8% for ppII, 1.6% for α_+, and 1.0% for β. Compared to the ratios given previously for representative conformations at a fixed β_{ij} = 4.3 cm^{-1}, the sub-ensemble-averaged ratio is smaller for ppII, but larger for α_+ and β. The behavior is mainly due to the difference in the calculated β_{ij}. As shown in Figure 6, the ppII conformation is most populated around β_{ij} = 3.0 cm^{-1}, which is smaller than the fitted value and will result in weaker cross-peak intensities. The most populated β_{ij} is around 5.5 cm^{-1} for α_+ and 5.3 cm^{-1} for β conformer, both of which are larger than the fitted value and will result in stronger cross-peak intensities. Therefore, the ensemble-averaged $R_{cross/diag}$ values are quite similar even though population distributions are quite different from the C22/CMAP and C36 force fields.

To quantitatively estimate the population distribution of different conformers, we took the sub-ensemble-averaged 2D spectra of each conformer, as shown in Figure 8j–l, as the reference spectra and mixed them in different population ratios to find the combination that can reproduce the experimental spectra and $R_{cross/diag}$. Our result shows that for any possible population distributions constrained by the diagonal peak intensity profile, the $R_{cross/diag}$ value can never be larger than 1.8%. Decreasing the population of ppII reduces the ratio. Therefore, 100% ppII population produces the closest spectra (Figure 8j) and $R_{cross/diag}$ to the experiment, within the accuracy of our MD simulation and current model.

The above spectral simulation based on the C36 structure ensemble suggested an almost exclusive population of ppII. This conclusion supports the validity of single conformation fitting. However, the most populated ppII conformation from the MD simulation has different (ϕ, ψ) and (β_{ij}, θ_{ij}) from those of the fitted conformation. Also, the 1.8% ratio is still much smaller than the measured ratio, indicating that there are still some problems with the current spectral simulation using MD structures. Several reasons can contribute to the discrepancy. First, let us focus on the double-crossed spectra because they are more sensitive to the small changes of (β_{ij}, θ_{ij}). As we showed in Figure 8, 1~2 cm^{-1} of increase in β_{ij} can erase the doublet feature of cross peaks. We can also calculate how sensitive $R_{cross/diag}$ is around the fitted (β_{ij}, θ_{ij}) = (4.3 cm^{-1}, 101°). At θ_{ij} =101°, $R_{cross/diag}$ changes from 2.4% at β_{ij} =4.3 cm^{-1}, to 1.5% at β_{ij} =3.0 cm^{-1}. At β_{ij} =4.3 cm^{-1}, $R_{cross/diag}$ changes from 2.4% at θ_{ij} =101° to 2.0% at θ_{ij} = 109°. If we calculate the dependence of (β_{ij}, θ_{ij}) on (ψ, ϕ) around (−61°, 145°), β_{ij} has a slope of 0.2 cm^{-1}/degree against ψ, and 0.1 cm^{-1}/degree against ϕ, whereas θ_{ij} has a slope of −0.75 degree/degree against ψ or ϕ. Therefore, even a small difference of ~5° in dihedral angles can introduce remarkable difference in both the shape and intensity of double-crossed spectra, and affect the calculated $R_{cross/diag}$ value. A deviation of several degrees is quite normal for the accuracy of MD force fields, especially for a highly flexible peptide where the distribution of ψ can range by ±25° or more for the ppII conformation, as seen in Figure 6a–c.

Another uncertainty comes from the current nearest-neighbor coupling map. Depending on the calculation details, the coupling maps from different references[68–71] are quite different in magnitude, although they have similar shape. For example, at (ϕ, ψ) = (−61°, 145°), Torri's map gives 6.2 cm^{-1},[69] Cho's map gives 1.8 cm^{-1},[72] Jansen's map gives 3.3 cm^{-1},[68] Stock's

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

map gives 3.0 cm^{-1},[70] and our map (Figure S1) gives 2.5 cm^{-1}. The more recent maps (including ours) are closer to one another compared to the earlier map[69] used in the $(Ala)_3$ and other studies.[17–18, 26] However, it has been shown that Torri's map predicts β_{ij} closer to the experimental value for the fully extended conformation.[62] At this point, it is still unclear which map is the best due to the lack of systematic experimental validation. Another contributing factor is the direction of transition dipoles. We calculated the angle between transition dipole and C=O bond to be 15.8°, whereas the literature values vary between 10° and 25°,[28, 69, 73] depending on both the calculation details and specific model peptides used. Instead of using a fixed angle, it may be beneficial to introduce a model which can account for how this angle varies with local environments.

3.8 Comparison with literatures

In this section we summarize some results from previous studies on alanine-based oligopeptides (Table 2), and compare their conformational distribution or single reference conformations to our fitted and MD simulated conformations. From the table, the reference structures of ppII (including the canonical conformation from the protein structural database[3, 74]) cited in these papers vary by 10–15°. Both our fitted (ϕ, ψ) = (−60°, 150°) and simulated ppII (ϕ, ψ) = (−61°, 145°) are within or close to this range. Additionally, our residue-specific fitting result shows that the conformation is insensitive to the residue position which is consistent with the NMR study.[20]

The ppII population predicted from C36 MD simulation is 59%. On the basis of spectral calculation where the C36 structure ensemble was used to define the reference spectra, we determined the ppII population for $(Ala)_5$ to be almost 100%. Therefore, our result is consistent with these previous studies that concluded ppII is the predominating conformation in aqueous solution. In this sense, the C36 force field is improved over C22/CMAP by predicting more ppII conformation, but there is still too much sampling of the α_+ and β conformers. The Drude-2013 force fields significantly over samples the β conformer and sampling of α_+ is also observed; efforts towards improving the Drude force field are ongoing. It should be noted that the quantitative results of population distribution are highly dependent on the way reference structures are defined, and also on the theoretical modeling of the relationship between peptide structure and spectra. The latter factor has already been discussed in Section 3.7. The effect of the former factor has been shown in previous studies on alanine based oligopeptide by Schweitzer-Stenner et al,[11–12, 14, 16] where they obtained reference structures of each conformer from protein libraries or to reproduce the NMR *J*-coupling constants, and then fit the vibrational spectra for the population distribution. Different choices of reference structures yielded different population distribution,[11–12, 14, 16] and sometimes it was necessary to modify the coordinates of reference structure to obtain satisfactory fitting.[14] When more conformers were included in the fitting, the population distribution also changed.[16]

The explicit inclusion of conformational distribution also makes a big difference in population determination.[16] For example, if we had chosen the representative ppII structure at the most populated position as our reference structure, it could affect the conclusion because its $R_{cross/diag}$ is only 1.4%, much smaller than 1.8% obtained for the whole ppII sub-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ensemble. For a given conformer, spectral calculation using its averaged or most populated structure cannot in general reproduce the 2D spectra averaged over its sub-ensemble because both the mapping from (ϕ, ψ) to (β_{ij}, θ_{ij}), and the relationship between (β_{ij}, θ_{ij}) and 2D spectra are quite complex. Therefore, it is important to use sub-ensemble-averaged 2D spectra as the reference spectra. If different structure ensembles from a different force field are used in the simulation analysis, they will give rise to different reference spectra, and hence result in a different population distribution. Cho and co-workers showed that the average ppII population over two peptide bonds in $(Ala)_3$ changes from 88% to 65% by simply changing the reference structure ensembles from Gromos 43A1 to AMBER ff03 but using the same experimental J-coupling constants and Karplus equations.[31] Such dependence on force fields demonstrates the need for accurate MD simulations and validation.

4. Conclusion

We have determined the conformation of $(Ala)_5$ using a combination of 2D IR spectroscopy, isotope editing, and MD simulation. Fitting the experimental 2D IR spectra under the $\langle ZZZZ \rangle$ and $\langle YYZZ \rangle$ polarizations to a simple two-vibrator exciton model allowed us to extract the coupling constants and angles between transition dipoles for the 1–2, 2–3 and 3–4 peptide unit pairs. DFT-based maps were developed to relate the fitted parameters to the Ramachandran angles on residue 2, 3 and 4. All three residues adopt very similar structures located within the range of ppII conformation. Because 2D IR spectra under the $\langle ZZZZ \rangle$ and $\langle YYZZ \rangle$ polarizations are insufficient to distinguish between whether single or multiple conformations coexist in solution, we explored other spectral parameters and found that both the spectral shape under the double-crossed polarization and the $R_{cross/diag}$ ratio are good spectral indicators to check the validity of fitting and spectral simulation. The spectral shape mainly provides a constraint for β_{ij}, whereas $R_{cross/diag}$ provides a constraint for both β_{ij} and θ_{ij} for the case of two coupled vibrators. For $(Ala)_5$, both indicators in the fitted double-crossed spectra are consistent with the experimental spectra, suggesting the validity of single conformation fitting.

Spectral simulations based on the C22/CMAP, C36, and Drude-2013 conformational ensembles show that none can fully reproduce the experimental $R_{cross/diag}$ ratio. Defining the 2D reference spectra based on the C36 conformational ensemble, the simulated $R_{cross/diag}$ ratio is the closest to the experiment when the ppII population is maximized to 100%. This finding provided further validation for single conformation fitting. Compared to the literature, the dihedral angles of our fitted structure and the most populated ppII structure from the C36 simulation are within the same range as previous studies on alanine based oligopeptides, although our work is the only study to suggest exclusive ppII population. We can conclude that ppII is the dominating conformation for $(Ala)_5$ at ambient temperature, but the quantitative results on conformer population are subjected to the definition of reference structures and accuracy in theoretical modeling. More reliable spectral simulation methods and MD force fields are important to elucidate peptide structures, and require further theoretical development.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants from the U.S. National Science Foundation (CHE-0450045, CHE-1013071 and CHE-1310693) to N.-H. Ge and the National Institutes of Health (GM072558) to ADM Jr. We are grateful to Hiroaki Maekawa for helpful discussion. We acknowledge the use of the UC Irvine Laser Spectroscopy Facility and Modeling Facility, and shared instruments supported by an NSF grant (CHE-0802913) to N.-H. Ge.

References

1. Rucker AL, Creamer TP. Polyproline II Helical Structure in Protein Unfolded States: Lysine Peptides Revisited. Protein Sci. 2002; 11:980–985. [PubMed: 11910041]
2. Rath A, Davidson AR, Deber CM. The Structure of “Unstructured” Regions in Peptides and Proteins: Role of the Polyproline II Helix in Protein Folding and Recognition. Pept Sci. 2005; 80:179–185.
3. Adzhubei AA, Sternberg MJE, Makarov AA. Polyproline-II Helix in Proteins: Structure and Function. J Mol Biol. 2013; 425:2100–2132. [PubMed: 23507311]
4. Shi Z, Woody RW, Kallenbach NR. Is Polyproline II a Major Backbone Conformation in Unfolded Proteins? Adv Protein Chem. 2002; 62:163–240. [PubMed: 12418104]
5. Shi Z, Olson CA, Rose GD, Baldwin RL, Kallenbach NR. Polyproline II Structure in a Sequence of Seven Alanine Residues. Proc Natl Acad Sci USA. 2002; 99:9190–9195. [PubMed: 12091708]
6. McColl IH, Blanch EW, Hecht L, Kallenbach NR, Barron LD. Vibrational Raman Optical Activity Characterization of Poly(L-proline) II Helix in Alanine Oligopeptides. J Am Chem Soc. 2004; 126:5076–5077. [PubMed: 15099084]
7. Chen K, Liu Z, Kallenbach NR. The Polyproline II Conformation in Short Alanine Peptides Is Noncooperative. Proc Natl Acad Sci USA. 2004; 101:15352–15357. [PubMed: 15489268]
8. Shi Z, Chen K, Liu Z, Ng A, Bracken WC, Kallenbach NR. Polyproline II Propensities from GGXGG Peptides Reveal an Anticorrelation with β-Sheet Scales. Proc Natl Acad Sci USA. 2005; 102:17964–17968. [PubMed: 16330763]
9. Shi ZS, Chen K, Liu ZG, Kallenbach NR. Conformation of the Backbone in Unfolded Proteins. Chem Rev. 2006; 106:1877–1897. [PubMed: 16683759]
10. Schweitzer-Stenner R. Dihedral Angles of Tripeptides in Solution Directly Determined by Polarized Raman and FTIR Spectroscopy. Biophys J. 2002; 83:523–532. [PubMed: 12080139]
11. Eker F, Cao X, Nafie L, Schweitzer-Stenner R. Tripeptides Adopt Stable Structures in Water. A Combined Polarized Visible Raman, FTIR, and VCD Spectroscopy Study. J Am Chem Soc. 2002; 124:14330–14341. [PubMed: 12452707]
12. Eker F, Griebenow K, Schweitzer-Stenner R. Stable Conformations of Tripeptides in Aqueous Solution Studied by UV Circular Dichroism Spectroscopy. J Am Chem Soc. 2003; 125:8178–8185. [PubMed: 12837087]
13. Schweitzer-Stenner R, Eker F, Griebenow K, Cao X, Nafie LA. The Conformation of Tetraalanine in Water Determined by Polarized Raman, FT-IR, and VCD Spectroscopy. J Am Chem Soc. 2004; 126:2768–2776. [PubMed: 14995194]
14. Schweitzer-Stenner R, Measey T, Kakalis L, Jordan F, Pizzanelli S, Forte C, Griebenow K. Conformations of Alanine-Based Peptides in Water Probed by FTIR, Raman, Vibrational Circular Dichroism, Electronic Circular Dichroism, and NMR Spectroscopy. Biochemistry. 2007; 46:1587–1596. [PubMed: 17279623]
15. Schweitzer-Stenner R, Measey TJ. The Alanine-Rich XAO Peptide Adopts a Heterogeneous Population, Including Turn-Like and Polyproline II Conformations. Proc Natl Acad Sci USA. 2007; 104:6649–6654. [PubMed: 17416675]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

16. Schweitzer-Stenner R. Distribution of Conformations Sampled by the Central Amino Acid Residue in Tripeptides Inferred from Amide I Band Profiles and NMR Scalar Coupling Constants. J Phys Chem B. 2009; 113:2922–2932. [PubMed: 19243204]
17. Woutersen S, Hamm P. Structure Determination of Trialanine in Water Using Polarization Sensitive Two-Dimensional Vibrational Spectroscopy. J Phys Chem B. 2000; 104:11316–11320.
18. Woutersen S, Hamm P. Isotope-Edited Two-Dimensional Vibrational Spectroscopy of Trialanine in Aqueous Solution. J Chem Phys. 2001; 114:2727–2737.
19. Woutersen S, Pfister R, Hamm P, Mu Y, Kosov DS, Stock G. Peptide Conformational Heterogeneity Revealed from Nonlinear Vibrational Spectroscopy and Molecular-Dynamics Simulations. J Chem Phys. 2002; 117:6833–6840.
20. Graf J, Nguyen PH, Stock G, Schwalbe H. Structure and Dynamics of the Homologous Series of Alanine Peptides: A Joint Molecular Dynamics/NMR Study. J Am Chem Soc. 2007; 129:1179–1189. [PubMed: 17263399]
21. Asher SA, Mikhonin AV, Bykov S. UV Raman Demonstrates That α-Helical Polyalanine Peptides Melt to Polyproline II Conformations. J Am Chem Soc. 2004; 126:8433–8440. [PubMed: 15238000]
22. Makowska J, Rodziewicz-Motowidło S, Bagińska K, Vila JA, Liwo A, Chmurzyński L, Scheraga HA. Polyproline II Conformation Is One of Many Local Conformational States and Is Not an Overall Conformation of Unfolded Peptides and Proteins. Proc Natl Acad Sci USA. 2006; 103:1744–1749. [PubMed: 16446433]
23. Mu Y, Stock G. Conformational Dynamics of Trialanine in Water: A Molecular Dynamics Study. J Phys Chem B. 2002; 106:5294–5301.
24. Mu Y, Kosov DS, Stock G. Conformational Dynamics of Trialanine in Water. 2. Comparison of Amber, CHARMM, GROMOS, and OPLS Force Fields to NMR and Infrared Experiments. J Phys Chem B. 2003; 107:5064–5073.
25. Woutersen S, Mu Y, Stock G, Hamm P. Subpicosecond Conformational Dynamics of Small Peptides Probed by Two-Dimensional Vibrational Spectroscopy. Proc Natl Acad Sci USA. 2001; 98:11254–11258. [PubMed: 11553784]
26. Schweitzer-Stenner R, Eker F, Huang Q, Griebenow K. Dihedral Angles of Trialanine in D_2O Determined by Combining FTIR and Polarized Visible Raman Spectroscopy. J Am Chem Soc. 2001; 123:9628–9633. [PubMed: 11572684]
27. Kim YS, Hochstrasser RM. Dynamics of Amide-I Modes of the Alanine Dipeptide in D_2O. J Phys Chem B. 2005; 109:6884–6891. [PubMed: 16851775]
28. Kim YS, Wang JP, Hochstrasser RM. Two-Dimensional Infrared Spectroscopy of the Alanine Dipeptide in Aqueous Solution. J Phys Chem B. 2005; 109:7511–7521. [PubMed: 16851862]
29. Parchaňský V, Kapitán J, Kaminský J, Šebestík J, Bouř P. Ramachandran Plot for Alanine Dipeptide as Determined from Raman Optical Activity. J Phys Chem Lett. 2013; 4:2763–2768. [PubMed: 26706714]
30. Best RB, Buchete N-V, Hummer G. Are Current Molecular Dynamics Force Fields Too Helical? Biophys J. 2008; 95:L07–L09. [PubMed: 18456823]
31. Oh K-I, Lee K-K, Park E-K, Yoo D-G, Hwang G-S, Cho M. Circular Dichroism Eigenspectra of Polyproline II and β-Strand Conformers of Trialanine in Water: Singular Value Decomposition Analysis. Chirality. 2010; 22:E186–E201. [PubMed: 21038390]
32. Schweitzer-Stenner R. Secondary Structure Analysis of Polypeptides Based on an Excitonic Coupling Model to Describe the Band Profile of Amide I' of IR, Raman, and Vibrational Circular Dichroism Spectra. J Phys Chem B. 2004; 108:16965–16975.
33. Sengupta N, Maekawa H, Zhuang W, Toniolo C, Mukamel S, Tobias DJ, Ge N-H. Sensitivity of 2D IR Spectra to Peptide Helicity: A Concerted Experimental and Simulation Study of an Octapeptide. J Phys Chem B. 2009; 113:12037–12049. [PubMed: 19496555]
34. Baiz CR, Lin Y-S, Peng CS, Beauchamp KA, Voelz VA, Pande VS, Tokmakoff A. A Molecular Interpretation of 2D IR Protein Folding Experiments with Markov State Models. Biophys J. 2014; 106:1359–1370. [PubMed: 24655511]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

35. Lee MW, Carr JK, Göllner M, Hamm P, Meuwly M. 2D IR Spectra of Cyanide in Water Investigated by Molecular Dynamics Simulations. J Chem Phys. 2013; 139:054506. [PubMed: 23927269]
36. Andrushchenko VV, Vogel HJ, Prenner EJ. Optimization of the Hydrochloric Acid Concentration Used for Trifluoroacetate Removal from Synthetic Peptides. J Pept Sci. 2007; 13:37–43. [PubMed: 17031869]
37. Sul S, Karaiskaj D, Jiang Y, Ge N-H. Conformations of N-Acetyl-L-Prolinamide by Two-Dimensional Infrared Spectroscopy. J Phys Chem B. 2006; 110:19891–19905. [PubMed: 17020375]
38. Maekawa H, Toniolo C, Broxterman QB, Ge N-H. Two-Dimensional Infrared Spectral Signatures of 3_{10}- and α-Helical Peptides. J Phys Chem B. 2007; 111:3222–3235. [PubMed: 17388471]
39. Maekawa H, Formaggio F, Toniolo C, Ge N-H. Onset of 3_{10}-Helical Secondary Structure in Aib Oligopeptides Probed by Coherent 2D IR Spectroscopy. J Am Chem Soc. 2008; 130:6556–6566. [PubMed: 18444622]
40. Asbury JB, Steinel T, Kwak K, Corcelli SA, Lawrence CP, Skinner JL, Fayer MD. Dynamics of Water Probed with Vibrational Echo Correlation Spectroscopy. J Chem Phys. 2004; 121:12431–12446. [PubMed: 15606264]
41. Dorrer C, Belabas N, Likforman J-P, Joffre M. Spectral Resolution and Sampling Issues in Fourier-Transform Spectral Interferometry. J Opt Soc Am B. 2000; 17:1795–1802.
42. Gallagher Faeder SM, Jonas DM. Two-Dimensional Electronic Correlation and Relaxation Spectra: Theory and Model Calculations. J Phys Chem A. 1999; 103:10489–10505.
43. Khalil M, Demirdöven N, Tokmakoff A. Obtaining Absorptive Line Shapes in Two-Dimensional Infrared Vibrational Correlation Spectra. Phys Rev Lett. 2003; 90:047401. [PubMed: 12570457]
44. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. J Phys Chem B. 1998; 102:3586–3616. [PubMed: 24889800]
45. Mackerell AD, Feig M, Brooks CL. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. J Comput Chem. 2004; 25:1400–1415. [PubMed: 15185334]
46. Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, Mackerell AD Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone φ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. J Chem Theory Comput. 2012; 8:3257–3273. [PubMed: 23341755]
47. Lopes PEM, Huang J, Shim J, Luo Y, Li H, Roux B, MacKerell AD. Polarizable Force Field for Peptides and Proteins Based on the Classical Drude Oscillator. J Chem Theory Comput. 2013; 9:5430–5449. [PubMed: 24459460]
48. Darden T, York D, Pedersen L. Particle Mesh Ewald: An N·Log(N) Method for Ewald Sums in Large Systems. J Chem Phys. 1993; 98:10089–10092.
49. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. J Comput Phys. 1977; 23:327–341.
50. Nosé S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. J Chem Phys. 1984; 81:511–519.
51. Hoover WG. Canonical Dynamics: Equilibrium Phase-Space Distributions. Phys Rev A. 1985; 31:1695–1697.
52. Andersen HC. Molecular Dynamics Simulations at Constant Pressure and/or Temperature. J Chem Phys. 1980; 72:2384–2393.
53. Lamoureux G, Roux B. Modeling Induced Polarization with Classical Drude Oscillators: Theory and Molecular Dynamics Simulation Algorithm. J Chem Phys. 2003; 119:3025–3039.
54. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, et al. CHARMM: The Biomolecular Simulation Program. J Comput Chem. 2009; 30:1545–1614. [PubMed: 19444816]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

55. Sieler G, Schweitzer-Stenner R, Holtz JSW, Pajcini V, Asher SA. Different Conformers and Protonation States of Dipeptides Probed by Polarized Raman, UV−Resonance Raman, and FTIR Spectroscopy. J Phys Chem B. 1998; 103:372–384.
56. Ge N-H, Zanni MT, Hochstrasser RM. Effects of Vibrational Frequency Correlations on Two-Dimensional Infrared Spectra. J Phys Chem A. 2001; 106:962–972.
57. Maekawa H, Sul S, Ge N-H. Vibrational Correlation between Conjugated Carbonyl and Diazo Modes Studied by Single- and Dual-Frequency Two-Dimensional Infrared Spectroscopy. Chem Phys. 2013; 422:22–30.
58. Zanni MT, Ge N-H, Kim YS, Hochstrasser RM. Two-Dimensional IR Spectroscopy Can Be Designed to Eliminate the Diagonal Peaks and Expose Only the Crosspeaks Needed for Structure Determination. Proc Natl Acad Sci U S A. 2001; 98:11265–11270. [PubMed: 11562493]
59. Maekawa H, Toniolo C, Moretto A, Broxterman QB, Ge N-H. Different Spectral Signatures of Octapeptide 3_{10}- and α-Helices Revealed by Two-Dimensional Infrared Spectroscopy. J Phys Chem B. 2006; 110:5834–5837. [PubMed: 16553386]
60. Hochstrasser RM. Two-Dimensional IR-Spectroscopy: Polarization Anisotropy Effects. Chem Phys. 2001; 266:273–284.
61. Fulmer EC, Mukherjee P, Krummel AT, Zanni MT. A Pulse Sequence for Directly Measuring the Anharmonicities of Coupled Vibrations: Two-Quantum Two-Dimensional Infrared Spectroscopy. J Chem Phys. 2004; 120:8067–8078.
62. Maekawa H, Ballano G, Toniolo C, Ge N-H. Linear and Two-Dimensional Infrared Spectroscopic Study of the Amide I and II Modes in Fully Extended Peptide Chains. J Phys Chem B. 2011; 115:5168–5182. [PubMed: 20845957]
63. Maekawa H, Poli MD, Moretto A, Toniolo C, Ge N-H. Toward Detecting the Formation of a Single Helical Turn by 2D IR Cross Peaks between the Amide-I and -II Modes. J Phys Chem B. 2009; 113:11775–11786. [PubMed: 19642666]
64. Maekawa H, Ballano G, Formaggio F, Toniolo C, Ge N-H. $^{13}C{=}^{18}O/^{15}N$ Isotope Dependence of the Amide-I/II 2D IR Cross Peaks for the Fully Extended Peptides. J Phys Chem C. 2014; 118:29448–29457.
65. Hamm P, Lim M, DeGrado WF, Hochstrasser RM. The Two-Dimensional IR Nonlinear Spectroscopy of a Cyclic Penta-Peptide in Relation to Its Three-Dimensional Structure. Proc Natl Acad Sci U S A. 1999; 96:2036–2041. [PubMed: 10051590]
66. Huerta-Viga A, Shaw DJ, Woutersen S. pH Dependence of the Conformation of Small Peptides Investigated with Two-Dimensional Vibrational Spectroscopy. J Phys Chem B. 2010; 114:15212–15220. [PubMed: 20977228]
67. Feng, Y. PhD Dissertation. University of California; Irvine, CA: 2014. Two-Dimensional Infrared Spectroscopy of Model Peptides.
68. la Cour Jansen T, Dijkstra AG, Watson TM, Hirst JD, Knoester J. Modeling the Amide I Bands of Small Peptides. J Chem Phys. 2006; 125:044312.
69. Torii H, Tasumi M. Ab Initio Molecular Orbital Study of the Amide I Vibrational Interactions between the Peptide Groups in Di- and Tripeptides and Considerations on the Conformation of the Extended Helix. J Raman Spectrosc. 1998; 29:81–86.
70. Gorbunov RD, Kosov DS, Stock G. Ab Initio-Based Exciton Model of Amide I Vibrations in Peptides: Definition, Conformational Dependence, and Transferability. J Chem Phys. 2005; 122:224904. [PubMed: 15974713]
71. Ham S, Cho M. Amide I Modes in the *N*-Methylacetamide Dimer and Glycine Dipeptide Analog: Diagonal Force Constants. J Chem Phys. 2003; 118:6915–6922.
72. Choi J-H, Ham S, Cho M. Local Amide I Mode Frequencies and Coupling Constants in Polypeptides. J Phys Chem B. 2003; 107:9132–9138.
73. Torii H, Tasumi M. Model Calculations on the Amide-I Infrared Bands of Globular Proteins. J Chem Phys. 1992; 96:3379–3387.
74. Bochicchio B, Tamburro AM. Polyproline II Structure in Proteins: Identification by Chiroptical Spectroscopies, Stability, and Functions. Chirality. 2002; 14:782–792. [PubMed: 12395395]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

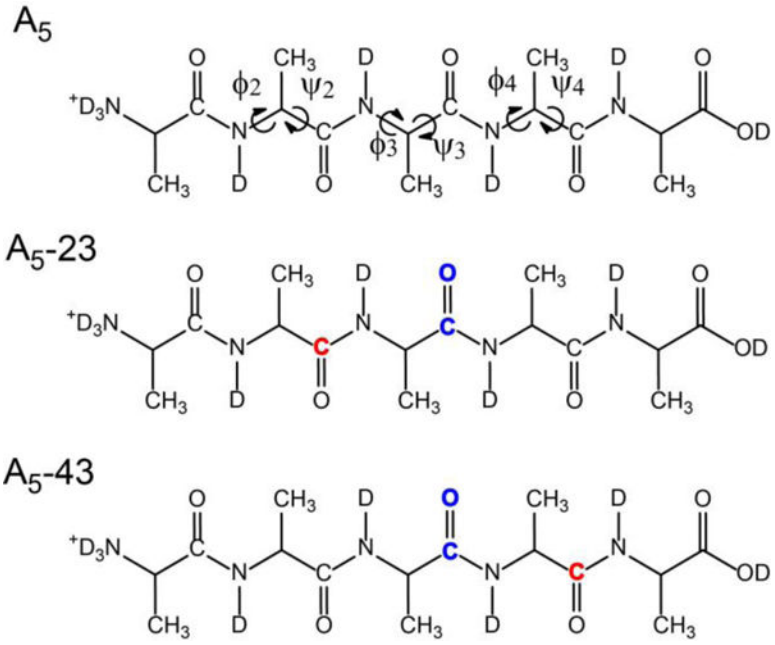


Figure 1.

Structure formula of $(Ala)_5$ in D_2O with pD = 1. Three different isotopomers were studied: unlabeled (A_5); bis-labeled with $^{13}C{=}O$ and $^{13}C{=}^{18}O$ at the second and third peptide units, respectively (A_5-23); and bis-labeled with $^{13}C{=}O$ and $^{13}C{=}^{18}O$ at the fourth and third peptide units, respectively (A_5-43). ^{13}C and $^{13}C{=}^{18}O$ are colored in red and blue, respectively. The middle three pairs of dihedral angles are indicated by the arrows.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

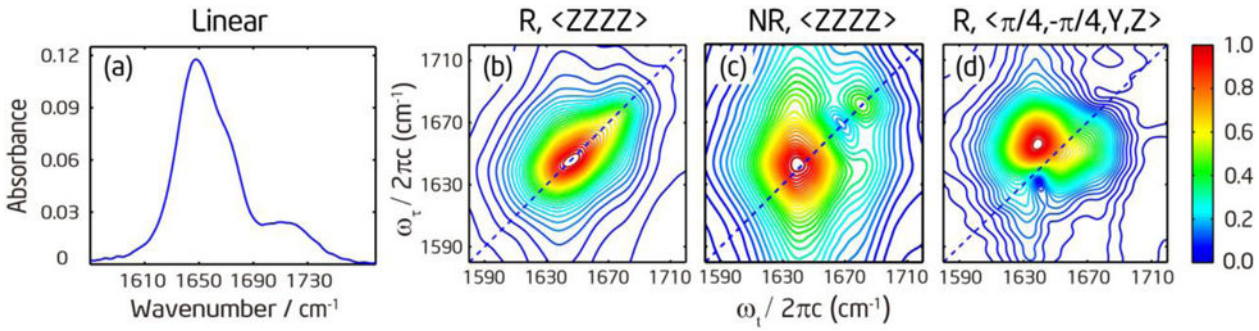


Figure 2.

(a) Linear absorption spectra of deuterated $(Ala)_5$ in D_2O at pD = 1. (b) and (c): Rephasing and nonrephasing spectra under the $\langle ZZZZ \rangle$ polarization at $T = 300$ fs, respectively. (d) Rephasing spectrum under the double-crossed $\langle \pi/4, -\pi/4, Y, Z \rangle$ polarization at $T = 0$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

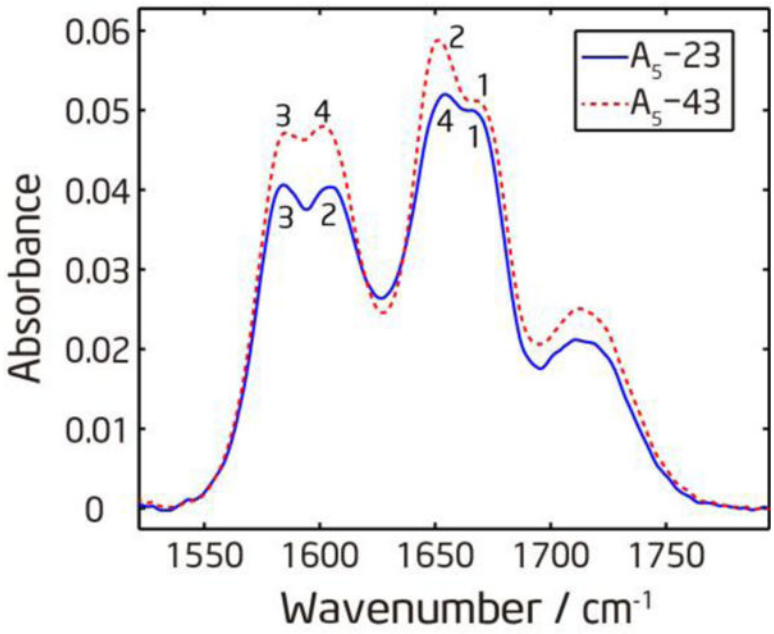


Figure 3.

Linear absorption spectra of isotope labeled $(Ala)_5$ in D_2O at pD=1. The numbers labeled on the peaks indicate which amide-I′ mode the peak belongs to.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

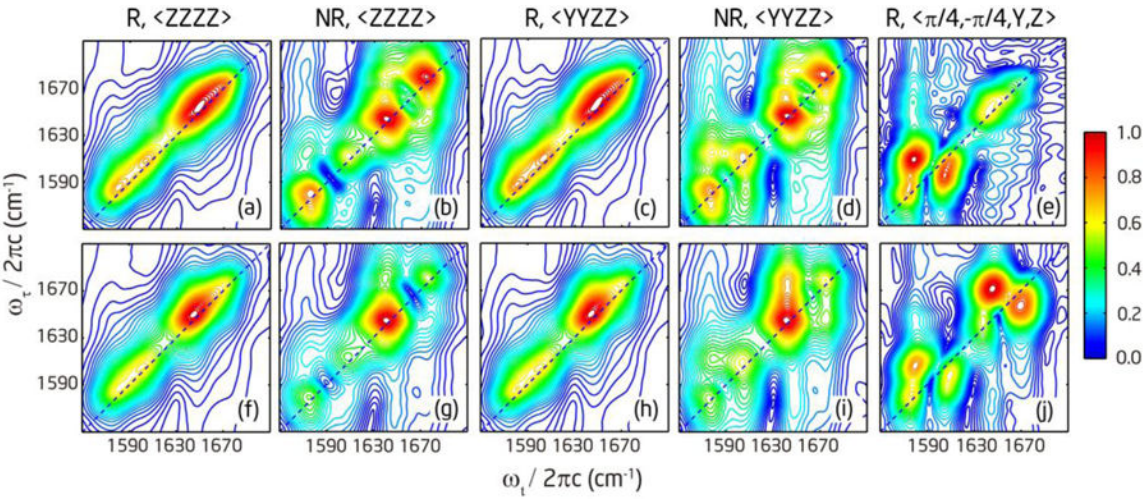


Figure 4.

2D IR absolute magnitude spectra of isotope labeled $(Ala)_5$ in D_2O at pD=1. (a) and (b): Rephasing and nonrephasing spectra of A_5-23 under the ⟨ZZZZ⟩ polarization at T = 300 fs, respectively. (c) and (d): Rephasing and nonrephasing spectra of A_5-23 under the ⟨YYZZ⟩ polarization at T = 300 fs, respectively. (e) Rephasing spectra of A_5-23 under the ⟨**π**/4, −**π**/4, Y, Z⟩ polarization at T = 0 fs. The corresponding spectra for A_5-43 are plotted in (f) – (j).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

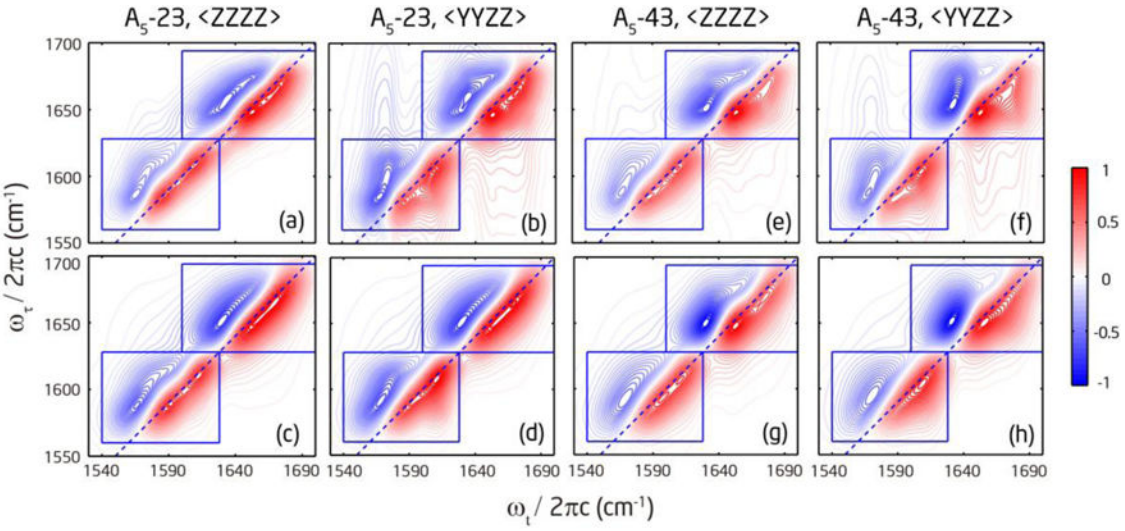


Figure 5.

2D IR absorptive spectra of isotope labeled $(Ala)_5$ at $T = 300$ fs. (a) and (b): Measured absorptive spectra of A_5-23 under the $\langle ZZZZ \rangle$ and $\langle YYZZ \rangle$ polarization, respectively. (c) and (d): Fitted absorptive spectra of A_5-23 under the $\langle ZZZZ \rangle$ and $\langle YYZZ \rangle$ polarization, respectively. The corresponding spectra for A_5-43 are plotted in (e) (f) (g) (h). The blue square boxes indicate the areas for fitting.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

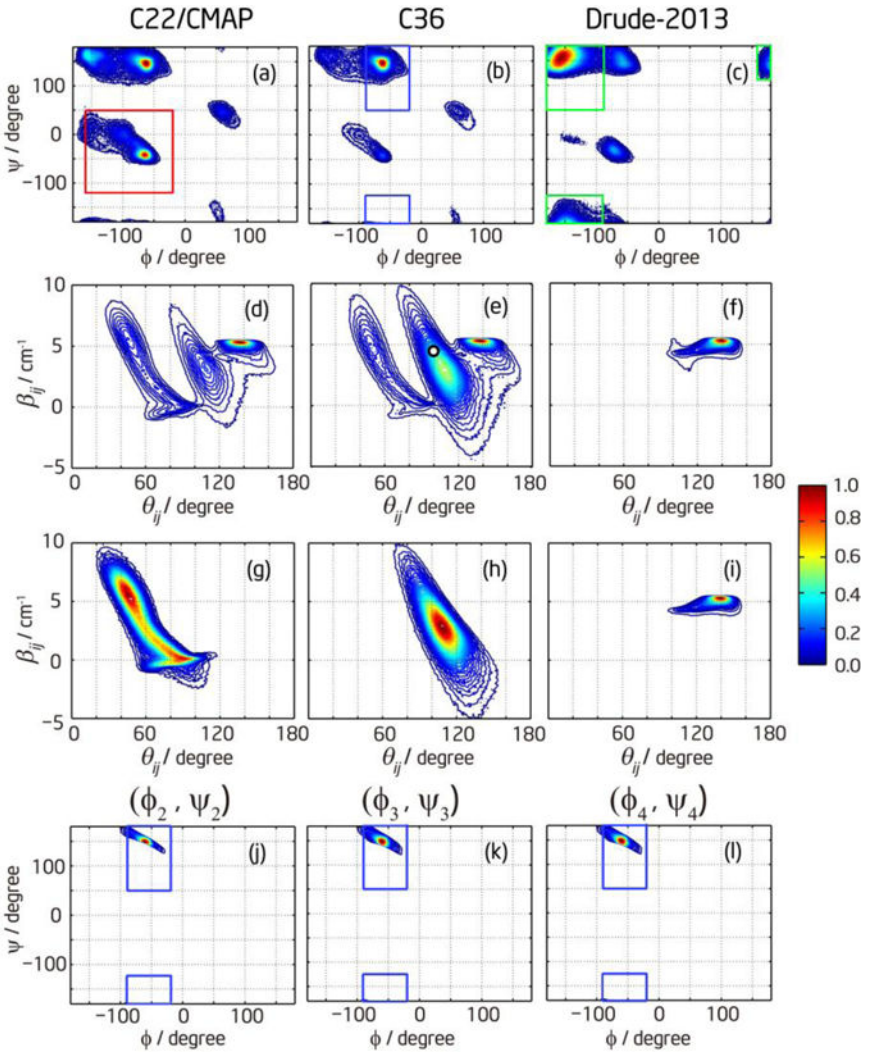


Figure 6.

The conformational distributions of $(Ala)_5$ from the MD simulations. The first row shows the Ramachandran plots for different force fields including all conformers, and the corresponding β_{ij}-θ_{ij} plots are in the second row. (a, d) are for C22/CMAP, (b, e) are for C36, and (c, f) are for Drude-2013. The third row is the β_{ij}–θ_{ij} plot for each conformer. The red box in (a) shows the definition of α_+, and the corresponding distribution in the β_{ij}–θ_{ij} plot is shown in (g). The blue boxes in (b) and the corresponding distribution in (h) are for ppII. The green boxes in (c) and the corresponding distribution in (i) are for β structure. (a)–(i) include statistics from all three middle residues. The fourth row shows the region of Ramachandran plot in (b) allowed by the confidence interval from spectral fitting for each residue: (j) is for (ϕ_2, ψ_2), (k) is for (ϕ_3, ψ_3), and (l) is for (ϕ_4, ψ_4). The blue boxes show the definition of ppII.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

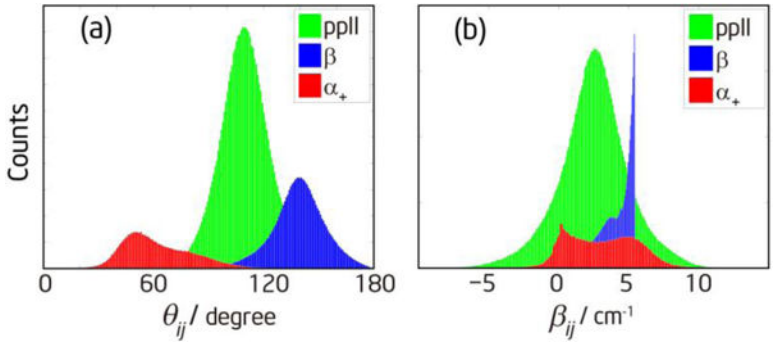


Figure 7.

(a) Histogram of the angles between two transition dipoles on nearest-neighbor amide-I modes for different conformations. (b) Histogram of the coupling constants between nearest-neighbor amide-I modes for different conformations. This distribution is calculated for the C36 force field.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

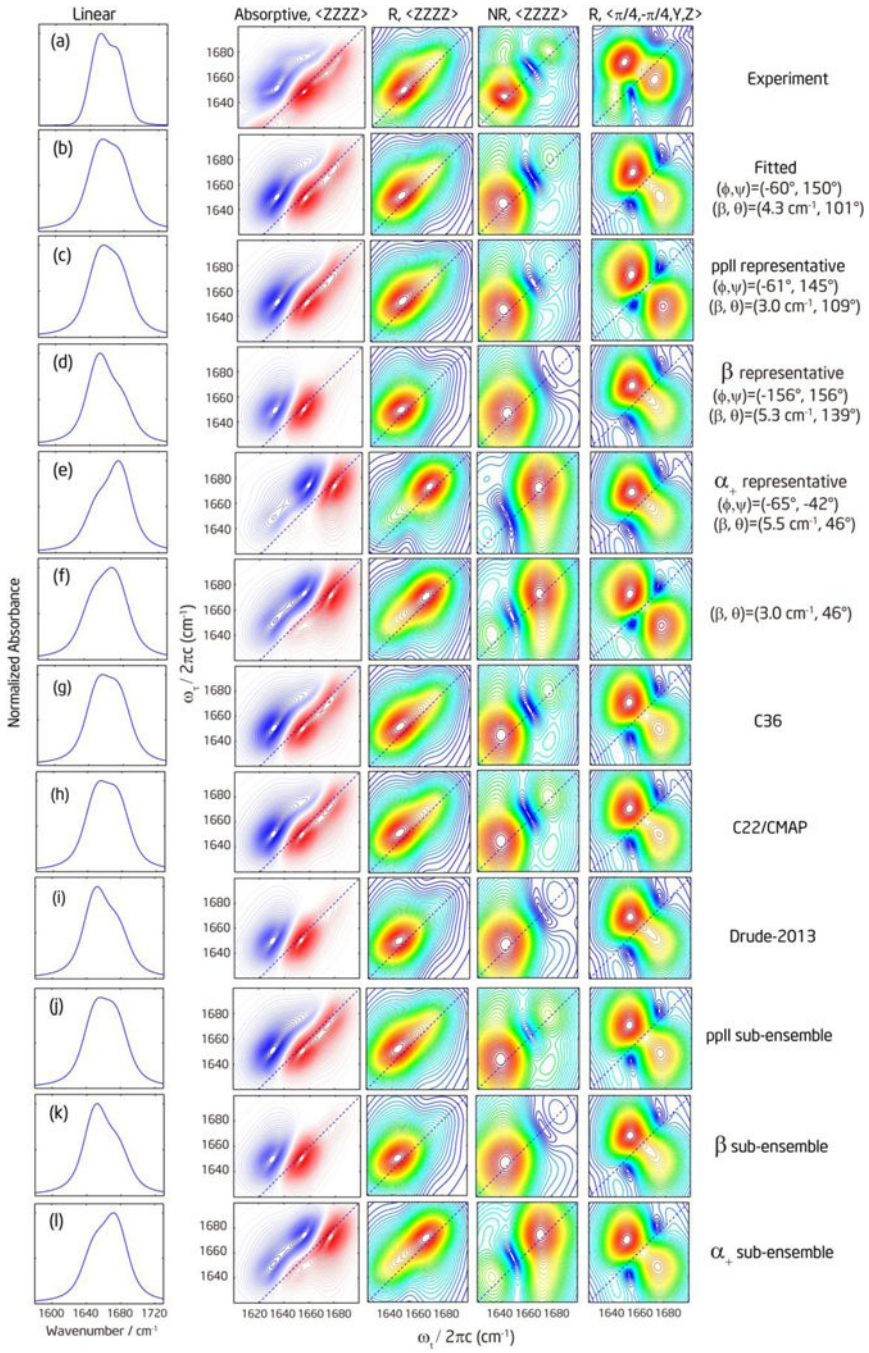


Figure 8.

Comparison of experimental, fitted, and simulated spectra of representative conformations and different MD simulation trajectories. Each row represents (a) Experimental spectra for the 1–2 pair of A_5-43. The linear spectrum of the unlabeled band was obtained from fitting the FTIR spectra of A_5-43 with multiple Voigt functions. (b) Fitted spectra. (c–f) Simulated spectra for representative conformations: (c) ppII; (d) β; (e) α_+; (f) a conformation with (β_{ij}, θ_{ij}) = (3 cm^{-1}, 46°). (g–i) Simulated spectra using MD simulation trajectories with different force fields: (g) C36; (h) C22/CMAP; (i) Drude-2013. (j–l) Simulated spectra averaged over

the MD C36 simulated distribution of each conformer: (j) ppII; (k) β; (l) α_+ distribution. The color scales are the same as those used in Figure 2 and Figure 5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Population fractions of different conformations at each residue under different force fields.

Force Field	Conformation	(ϕ_2, ψ_2)	(ϕ_3, ψ_3)	(ϕ_4, ψ_4)
C22/CMAP	α+	0.26	0.39	0.30
	β	0.32	0.23	0.32
	ppII	0.36	0.31	0.32
C36	α+	0.11	0.13	0.12
	β	0.24	0.21	0.21
	ppII	0.58	0.60	0.59
Drude-2013	α+	0.02	0.08	0.10
	β	0.86	0.75	0.77
	ppII	0.12	0.16	0.13

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Summary of reference results on ppII. The red As represent the studied residues.

Reference	System	Methodology	(ϕ, ψ) of ppII	ppII population	other population
Tamburro[3, 74]		PDB	(−75°, 145°)		
Hamm[17–19]	A AA	2D IR	(−60°, 140°)	80%	α = 20%
Schweitzer-Stenner[11–12]	A AA	Polarized Raman, FTIR, VCD and ECD	(−60°, 150°)	50%	β = 50%
Schweitzer-Stenner[16]	A AA	Polarized Raman, FTIR, and VCD	(−70° ± 20°, 150° ± 20°)	84%	β = 16%
	A AA		(−69° ± 10°, 140° ± 10°)	84%	β = 8%, α = 4%, γ = 5%
Schweitzer-Stenner[14]	A AKA	Polarized Raman, FTIR, VCD, ECD and NMR	(−65°, 150°)	63%	β = 7%, α = 30%
	A AAA		(−65°, 150°)	66%	β = 14%, α = 20%
	AA AA		(−70°, 140°)	73%	β = 17%, α = 10%
Cho[31]	Averaged A AA	NMR, used $^3J(H_N,H_\alpha)$ from Ref[20]	AMBER ff03	65%	β = 35%
			Gromos 43A1	88%	β = 12%
Graf[20]	A AAAA	NMR	Gromos 43A1	84%	β = 16%
This work	A AAAA	Fitting	(−60°, 150°)		
		C36 prediction	(−61°, 145°)	59%	β = 22%, α = 12%
		Simulation based on C36 ensemble	(−61°, 145°)	~100%	