

2D-IR Experiments and Simulations of the Coupling between Amide-I and Ionizable Side Chains in Proteins: Application to the Villin Headpiece

Sayan Bagchi,^{†,‡} Cyril Falvo,^{‡,§} Shaul Mukamel,[§] and Robin M. Hochstrasser^{*,†}

Department of Chemistry, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6323, and Department of Chemistry, University of California Irvine, Irvine, California 92697-2025

Received: January 10, 2009; Revised Manuscript Received: May 1, 2009

The carboxylate side chains of Asp and Glu have significant coupling with the amide states of the backbone of the Villin headpiece. In two-dimensional spectroscopy, cross peaks are observed between these side chains and the main amide-I band. To model the absorption of the side chains, the electric field variations of vibrational frequencies of a carboxylic acid group (neutral form, $\text{CH}_3\text{—COOH}$) and a carboxylate group (ionized form, $\text{CH}_3\text{—COO}^-$) are parametrized by means of density functional theory calculations. Simulations indicate that the side chains significantly couple to only one or two amide-I modes out of all of the amino acid residues which makes them useful as spectroscopic markers, providing information about the local structural behavior of the protein. Both experiment and simulations show that the cross peaks between the carboxylate and the amide-I bands are significantly diminished above the melting temperature.

1. Introduction

It has been recognized for some time that the vibrational modes of the peptide units of proteins are sensitive indicators of secondary structure. In particular, the amide-I mode, which is approximately a backbone carbonyl vibration, is highly degenerate and delocalized in polypeptides such that important aspects of the topology of the structure are often clearly manifested in the infrared spectrum. The delocalization is commonly understood to arise from electrostatic interactions that have sufficient range to sense secondary structure over useful distances. The introduction of two-dimensional infrared (2D-IR) methods has permitted the direct experimental examination of these presumed electrostatic couplings between amide units, thereby enlarging the scope and sharpening interpretations of vibrational spectroscopy as a structural indicator. The delocalization of the vibrations in proteins critically depends on the near degeneracy of the backbone modes at different residues. However, there are also carbonyl-like vibrations associated with side chains which are nearby in frequency and are therefore expected to participate in the delocalization and to extend the plausible interpretations of the spectra. In this paper we directly examine by means of 2D-IR methods and theory the side chain carboxylic acid modes and their coupling to the backbone states of a small protein, Villin headpiece (HP35).

Since the 2D-IR method has intrinsically broad bandwidth, it promises a new view of the time-dependent changes in local and global structural features that occur during folding or unfolding. The Villin headpiece was chosen for the present study because it is a small, 35 residue, naturally occurring protein.^{1–3} In addition, it belongs in the group having the fastest known folding kinetics,^{4,5} and it has been suggested to utilize one of the simplest protein folding mechanisms.⁶ Its conformational dynamics is grounded in many types of experiments and theoretical models.⁶ Therefore, Villin presents an excellent

opportunity for 2D spectroscopy to contribute to the knowledge of protein structural evolution, especially since its relative simplicity suggests that the experiments could be directly simulated by molecular dynamics (MD). The HP35 contains three small helices and a hydrophobic core.⁷ The sequence contains two aspartic acid (Asp, D) and two glutamic acid (Glu, E) residues. The pK_a of these ionizable side chains is greatly influenced by their interaction with the surrounding environment.⁸ The 2D-IR spectra of HP35 including these ionizable residues and the interaction of their vibrations with the backbone amide-I states are the focus of the present work.

Isotope editing techniques combined with 2D-IR, where the amide-I transition frequency is modified by substituting $^{12}\text{C}=^{16}\text{O}$ by $^{13}\text{C}=^{16}\text{O}$ or $^{13}\text{C}=^{18}\text{O}$, have often been used to simplify complex systems with many overlapping transitions.^{9–11} This approach permits spectral selection of particular residues and provides site specific information regarding protein backbone states. However, isotope edited transition frequencies of the backbone occur at frequencies that are very similar to those of the carboxylate side chain vibrations. The present 2D-IR work investigates the interaction of the carboxylate side chains with the backbone amide states of HP35, focusing on the effect of these ionizable groups on the spectra. The results show that the interactions between the side chains and the amide-I band provide information about the local structure of the protein and suggest the use of these carboxylate side chains as naturally occurring spectroscopic probes with which to study the protein conformational dynamics. However, the carboxylate infrared transitions are not as well understood as amide-I modes, so a detailed theoretical study of them is also needed to establish their utility as probes.

Experimental and theoretical studies of the infrared absorption of the aliphatic monoacids and diacids,^{12,13} salicylic acid and its derivatives,¹⁴ cyanoacetic acid,¹⁵ sodium acetate,¹⁶ and several other acids¹⁷ have been reported. Many experimental studies^{12–14,17} have verified that there is a pH dependent infrared absorption. At low pH, one absorption band is in the range of 1700–1730 cm^{-1} (C=O stretch region), and another is in the range of 1200–1250 cm^{-1} (C—OH bending). At high pH, the IR

* Corresponding author. E-mail: hochstra@sas.upenn.edu. Tel: 215-898-8203. Fax: 215-898-0590.

[†] University of Pennsylvania.

[‡] These authors contributed equally to this work.

[§] University of California Irvine.

absorption is dramatically changed. One band is observed in the 1550–1650 cm⁻¹ frequency range, corresponding to the asymmetric COO⁻ vibration. The symmetric stretch of the carboxylate is observed in the range of 1300–1450 cm⁻¹. A linear correlation between the asymmetric vibration frequency and the pK_a of the acids has been reported.^{12,13} To model the absorption of the side chains, we have parametrized the vibrational frequencies of the carboxylic acid group (neutral form, R—COOH) and the carboxylate group (ionized form, R—COO⁻) by means of density functional theory (DFT) calculations on acetic acid, CH₃—COOH. MD simulations of the Villin headpiece in folded and unfolded states are then carried out with the results being used to simulate their IR and 2D-IR spectra including both the amide-I backbone and the carboxylic acid vibrational states.

2. Materials and Methods

Sample Preparation. The headpiece (HP35) of the 76 residue protein, Villin, was synthesized by Biomer Technologies (Hayward, CA). HP35 consists of 35 amino acids (from amino acid 42 to amino acid 76 of Villin) to have the sequence LSDEDFFKAVFGMTRSAFANLPLWKQQNLKKEKGLF. A double labeled peptide isotopomer, containing the isotopic amide carbonyl label ¹³C= ¹⁶O at Ala57 and ¹³C= ¹⁸O at Ala59, was synthesized along with the unlabeled HP35. The peptides were dissolved in D₂O to obtain a concentration of approximately 10 mM. HCl (NaOH) was added to a phosphate buffer solution (pH = 5.7) to attain the acidic (basic) pH used in the experiments. Finally these buffer solutions were lyophilized and dissolved in D₂O. The sample contained some residual trifluoroacetic acid (TFA) from the peptide purification. No buffer was used for the Fourier transform infrared (FTIR) and 2D-IR experiments of HP35 in neat D₂O for which the pD is estimated at about five. The circular dichroism spectrum of the sample solution containing residual TFA in neat D₂O showed a melting temperature (*T*_m) of 70 °C which is within the experimental error of the reported *T*_m of 70.5 °C for the Villin headpiece.¹⁸

Linear IR Spectroscopy. The FTIR absorption spectra were recorded at a 1 cm⁻¹ resolution on a Perkin-Elmer 2000 Explorer spectrometer in a temperature controllable cell with CaF₂ windows and a path length of 56 μm.

Two-Dimensional IR Spectroscopy. The multidimensional IR spectra were obtained using heterodyned spectral interferometry with the laser arrangement described previously.^{6–8} Three Fourier transform limited 75 fs pulses (wave-vectors \mathbf{k}_1 , \mathbf{k}_2 , and \mathbf{k}_3) with an energy of about 400 nJ each were incident on the sample. The signal in the phase matching direction $-\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$ was detected by heterodyning it with a local oscillator (LO) pulse, which preceded the signal pulse by ~1500 fs. The time interval between the first and the second pulse is denoted as the coherence time τ , that between the second and the third pulse is denoted as the waiting time T , and that between the third pulse and the detected signal is denoted as t . The rephasing and nonrephasing sequences correspond to \mathbf{k}_1 arriving earlier or later than \mathbf{k}_2 , respectively. The signal and the LO were combined at the focus of the monochromator (with a groove density 50 lines/mm), and the heterodyned signal created was detected by a 64 element HgCdTe array detector. Each detector element is 200 μm in width and 1 mm in height. All of the experiments reported in this manuscript were performed for the $\langle zzzz \rangle$ condition, where the polarization of \mathbf{k}_1 , \mathbf{k}_2 , \mathbf{k}_3 , and LO beams are parallel to one another.

Data Processing. The experimental raw data collected from the 2D-IR experiments was in the form of a two-dimensional

array of time (in 2 fs steps along τ) and wavelength (in ~15 nm steps). The 2 fs time step provided ~10 points per amide-I cycle and prevented aliasing due to undersampling. By performing an inverse Fourier transform to the time domain followed by a Fourier transform to the frequency domain, we converted $\tilde{S}(\tau, T, \lambda_k)$ into $\tilde{S}(\tau, T, \omega_m)$, where λ_k represents corresponding wavelength of the k th ($k = 1\dots64$) detector element and ω_m represents m th of 400 evenly spaced frequencies. The two Fourier transforms were performed as follows: the inverse Fourier transform was

$$S(\tau, T, t) = \sum_{k=1}^{64} \tilde{S}(\tau, T, \lambda_k) \cdot \exp(i\omega_k t) \cdot \Delta\omega_k \\ t = 0, 1, 2, \dots, 5000 \text{ fs} \quad (1)$$

where $\omega_k = 2\pi c/\lambda_k = 2\pi c/(\Delta\lambda(k-32) + \lambda_{32})$ and $\Delta\omega_k = 2\pi c/(\Delta\lambda(k-32 - 1/2) + \lambda_{32}) - 2\pi c/(\Delta\lambda(k-32 + 1/2) + \lambda_{32})$. $\Delta\lambda$ is the wavelength difference between two adjacent detector elements and is a constant in a limited spectral range. λ_{32} is the wavelength observed by the 32nd element of the detector. The $S(\tau, T, t)$ was Fourier transformed to generate $\tilde{S}(\tau, T, \omega_m)$.

$$\tilde{S}(\tau, T, \omega_m) = \sum_m S(\tau, T, t) \cdot \exp(-i\omega_m t) \cdot \Delta t \quad (2)$$

At this point, $\tilde{S}(\tau, T, \omega_m)$ is in the form $\tilde{S}(\tau, T, \omega_t)$. The spectra were obtained as combined frequency time data sets defined as $\tilde{S}_R(\tau, T, \omega_t)$ and $\tilde{S}_{NR}(\tau, T, \omega_t)$, where R and NR refer to the rephasing and nonrephasing configurations. The complex 2D-IR spectra $\tilde{S}_R(-\omega_t, T, \omega_t)$ and $\tilde{S}_{NR}(\omega_t, T, \omega_t)$ were obtained by Fourier transformation along τ . To obtain the absorptive or correlation spectra, rephasing and nonrephasing 2D frequency spectra were added. The detailed protocols of data processing have been described previously.¹⁹

It needs to be commented that the cross peak regions of the 2D-IR spectra of HP35 (Figure 4c) superficially resemble the types of signals obtained from Fourier transform artifacts that are common in the processing of signal waveforms. The 2D-IR data sets were routinely and systematically processed for different coherence time ranges in the rephasing direction. When the coherence time is scanned from 0 to 1 ps, a distinct, repeating pattern corresponding to the often observed Fourier transform artifact from truncated free induction decay (FID) and aliasing shows up. This pattern disappears when data over coherence times of 2 ps are processed with appropriate sampling, which is again common in 2D-IR data processing but seldom mentioned. The cross peaks were most clearly spectrally resolved and had definitive repeatable frequencies when an even longer coherence time scan over 4 ps in 2 fs steps was used: the latter conditions are those presented in this paper. Further control processing tests were performed to ensure that an appropriately sampled rephasing scan of 4 ps does not truncate the FID. To ensure apodization along τ , $\tilde{S}(\tau, T, \omega_t)$ was convoluted to the Hann window function of the form:

$$w(m) = \frac{1}{2}(1 - \cos(2\pi m/N - 1)) \quad (3)$$

where N is the total number of data points along τ . The convoluted signal $\tilde{S}_{conv}(\tau, T, \omega_t)$ showed a difference of 2% from $\tilde{S}(\tau, T, \omega_t)$ at $\tau = 4$ ps, which is comparable with the noise level (amplitude of the nonabsorbing frequency region). The introduc-

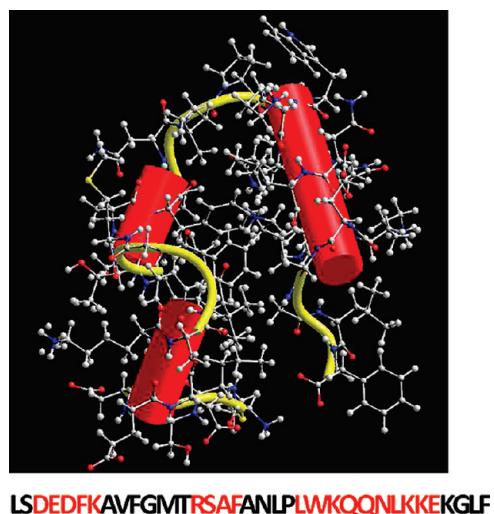


Figure 1. Structure of the Villin headpiece (HP35). The residues forming the helices are marked in red, and four amino acid residues with ionizable side chains are underlined (data from Protein Data Bank, identification number 1VII).

tion of the window function did not change the cross peak positions along ω_r . When $\tilde{S}(\tau, T, \omega_r)$ was convoluted with the window function for the shorter τ scan of 1 ps, the earlier mentioned repeating pattern was not observed, but the peak shapes and frequencies were distorted in the corresponding 2D-IR spectrum. The convoluted signal $\tilde{S}_{\text{conv}}(\tau, T, \omega_r)$ showed a difference of 20% from $\tilde{S}(\tau, T, \omega_r)$ at $\tau = 1$ ps, conclusive of the fact that a rephasing scan of 1 ps without a window function truncates the FID. These data and tests were consistently reproducible for many samples and conditions such as temperature and optical setup.

MD Simulation of Villin Headpiece Subdomain. Simulations were carried out using the available structure of recombinant form of HP-36^{1,20} which contains a supplementary Met residue on the N-terminal side. We verified that this amino acid does not influence the overall structure of the peptide. The amide vibration close to the N-terminus of the peptide (located between the Met and the Leu residues) is not included in the 2D-IR simulation. Two series of MD simulations were conducted: one for the folded and one for the unfolded state. They were performed using the NAMD 2.6 program²¹ with the CHARMM27 force field²² and periodic boundary conditions with a 2 fs time step. Long-range electrostatic interactions were computed using particle-mesh Ewald (PME),^{23,24} and a real space cutoff of 12 Å was used for nonbonded interactions. Langevin dynamics with a 1 ps damping coefficient were used to achieve a constant temperature. A 1 atm constant pressure was maintained using a Nose-Hoover Langevin piston^{25,26} with a decay period of 200 fs and a 100 fs damping time, when pressure regulation was employed.

Folded State Simulation. The folded-ionized (FI) simulation was initiated from the HP-36 NMR minimized average structure² (Protein Data Bank identification number 1VII; Figure 1). The peptide, with all side chains ionized, is embedded in a 52 Å box containing 4650 water molecules and two chloride ions to neutralize the box. After minimization and a 5 ns equilibration using the normal pressure and temperature ensemble at the temperature of 300 K, a normal volume and temperature simulation was performed, with snapshots recorded every 2 fs for 250 ps.

For comparison we have generated a similar simulation starting from the mutant HP-35 X-ray structure⁷ (Protein Data

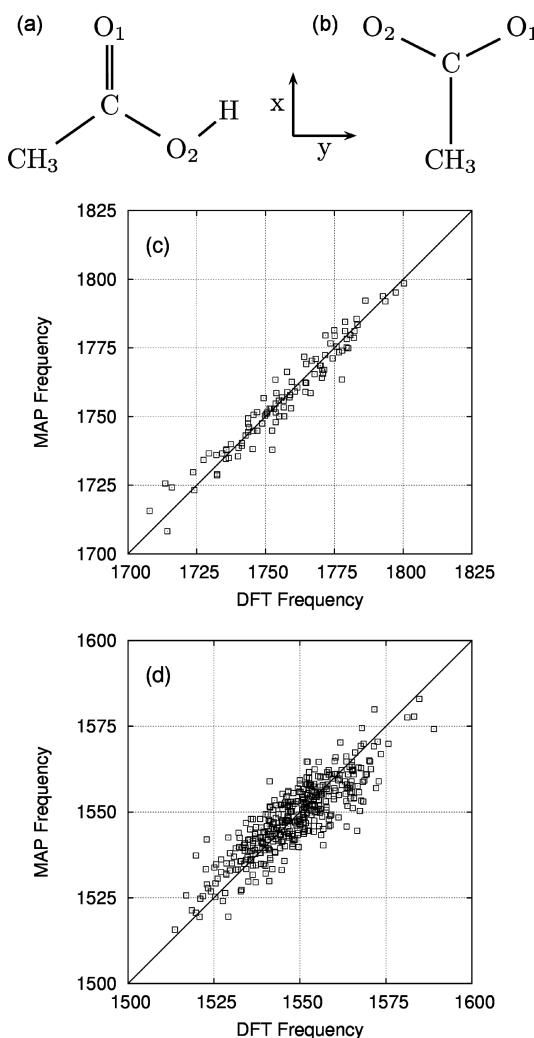


Figure 2. Molecular frames and correlation maps: (a) acetic acid molecules and (b) acetate ions and the corresponding molecular frame used in the DFT calculations. DFT/MAP correlation plots for (c) acetic acid and (d) acetate.

Bank identification number 1yrf). In this mutant, the Asn residue N68 is replaced by His. The HP35 and HP36 sequences both converge to a similar structure. The average C_α root mean square deviation from the NMR structure including residues 42–76 is 1.95 Å for HP-35 and 1.88 Å for HP-36.

In a second simulation, the folded-neutral state (denoted herein by FN) initiated from the NMR structure, the side chains of the glutamic and aspartic acid residues D44, E45, D46, and E72 and the terminal carboxylate group I (F76) were replaced by their neutral acid forms. The peptide was embedded in a 52 Å box containing 4691 water molecules and seven chloride ions. After minimization and equilibration, an MD simulation was performed for 250 ps.

Unfolded State Simulation. The unfolded-ionized (UI) simulation was initiated using the FI trajectory. We have performed a high-temperature simulation of the Villin headpiece by heating to 800 K for 5 ns. Ten snapshots were selected along this trajectory at 500 ps intervals, and each were equilibrated at 358 K during 1 ns to obtain 10 trajectories of 250 ps length of the unfolded state denoted here as UI.

Electronic Structure Computations. DFT calculations were performed at the B3LYP/aug-cc-pVDZ level using the Gaussian 03 software package.²⁷ Calculations employed the molecular frame defined in Figure 2. The coordinates are centered at the

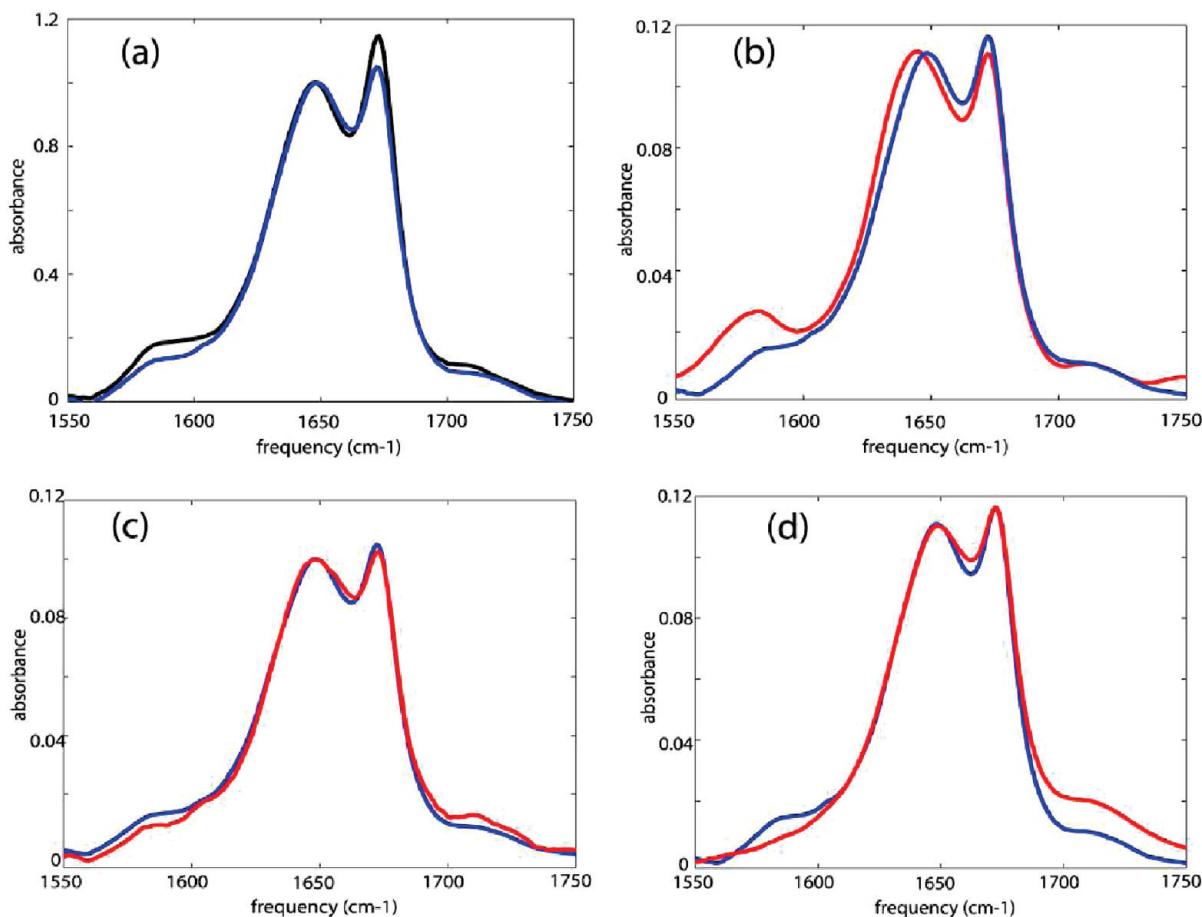


Figure 3. Experimental linear IR spectra. Linear IR spectra of (a) unlabeled HP35 (blue) and double labeled HP35 (black) in neat D₂O, (b) unlabeled HP35 in D₂O (blue) and at pD \sim 11.0 (red), (c) unlabeled HP35 in D₂O (blue) and at pD \sim 3.0 (red), and (d) unlabeled HP35 in D₂O (blue) and at pD \sim 1.0 (red).

carbonyl carbon atom; the *X* axis is defined along the C=O bond, and the *XY* plane coincides with the COO plane. After optimization of the structure, the carbonyl harmonic frequency of acetic acid was found to be 1808 cm⁻¹. This value is 20 cm⁻¹ higher than the observed value (1788 cm⁻¹).²⁰ It is acceptable considering the anharmonic shifts of similar vibrational modes (amide-I, -16 cm⁻¹; gas-phase CO molecule, -27 cm⁻¹).

3. Experimental Results

Linear IR Spectra. The FTIR spectrum of the unlabeled Villin headpiece (HP35) in D₂O shows an amide-I band in the 1620–1660 cm⁻¹ region and another sharp peak at 1674 cm⁻¹ due to TFA (see Sample Preparation section). Two broad but less intense bands were observed as shoulders in the 1560–1600 cm⁻¹ and 1700–1730 cm⁻¹ regions (see Figure 3). The line shape parameters of the peak at 1674 cm⁻¹ are obtained separately by fitting the linear IR spectrum of neat TFA in D₂O. Fitting the linear spectrum of HP35 in D₂O with three Gaussian profiles along with the line shape parameters obtained for the TFA peak yields the relative cross sections of 0.2 and 0.1 for the bands at 1560–1600 cm⁻¹ and 1700–1730 cm⁻¹, relative to the amide-I band. The FTIR spectrum of the isotopically double labeled HP35 shows a 20% increase in cross section in the 1560–1600 cm⁻¹ region compared with that of the unlabeled compound, but the isotope regions are not clearly separated.

The linear spectrum of the unlabeled HP35 at pD \sim 11.0 showed an increase in cross section at 1560–1600 cm⁻¹ as

compared with the solution in neat D₂O (Figure 3). At pD \sim 3.0, there was a slight weakening of the lower frequency broad shoulder and a small increase in absorption above 1700 cm⁻¹ as compared with the solution in neat D₂O. This broad band at 1560–1600 cm⁻¹ was less intense for the unlabeled compound at pD \sim 1.0 as compared with that of the spectrum in neat D₂O. At pD \sim 1.0 an increased cross section at \sim 1700–1730 cm⁻¹ was observed. The broad band at 1560–1600 cm⁻¹ was one-third of the cross section at pD \sim 1.0 compared with that of the neat D₂O.

Two-Dimensional IR Spectra. The 2D-IR spectrum of unlabeled HP35 in D₂O at room temperature, shown in Figure 4a, has pairs of diagonal peaks with each consisting of a $\nu = 0 \rightarrow \nu = 1$ transition on the diagonal and a $\nu = 1 \rightarrow \nu = 2$ transition shifted along the horizontal axis by the diagonal anharmonicity. The amide-I transition region consists of two clearly distinguishable bands, at 1630 and 1643 cm⁻¹. A diagonal peak at 1674 cm⁻¹ corresponds to the sharp band from TFA. Less intense (\sim 20% of that of the peak at 1643 cm⁻¹) diagonal peaks were observed around 1660 and 1615 cm⁻¹. Overlapping transitions were observed along the diagonal outside the amide-I region in the 1560–1590 cm⁻¹ frequency region. An enlarged view of these diagonal peaks is shown in Figure 4b: cross peaks between them and the amide-I band are observed. The occurrence and the shapes of these cross peaks are shown in Figure 4c. As seen in Figure 4c, the cross peaks are clearly distinguishable, though the diagonal peaks in the same region overlap on one another (Figure 4b). A main

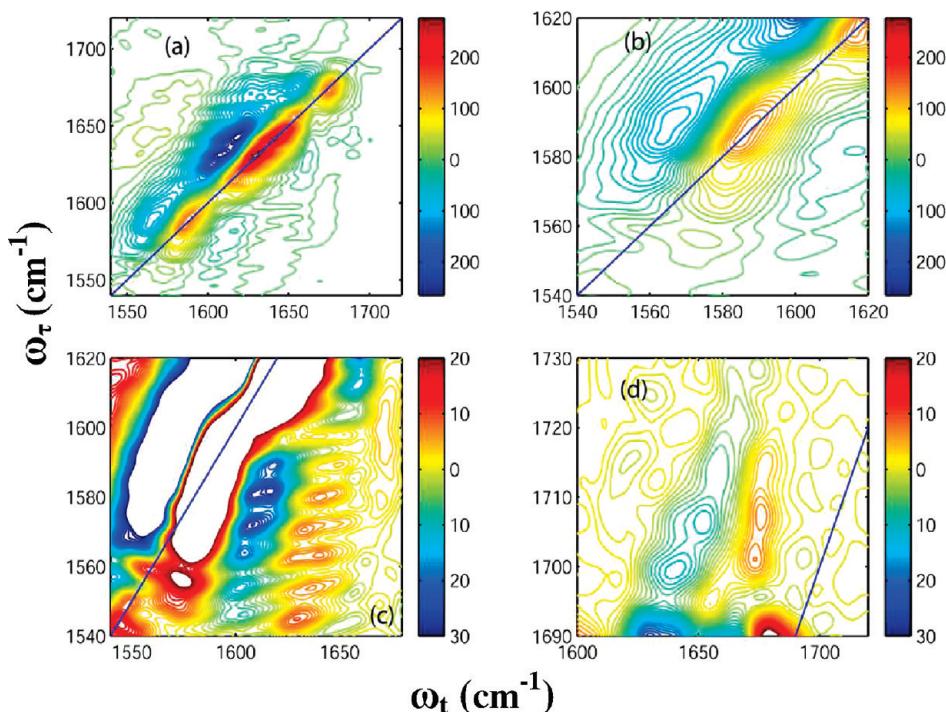


Figure 4. 2D-IR spectra of the Villin headpiece at 23 °C. Real part of the absorptive 2D-IR spectrum of HP35 in neat D₂O ($T = 0$) at 23 °C focusing on (a) the whole frequency range for the carbonyl stretches, (b) the diagonal peaks from Asp and Glu ionizable side chains, (c) the cross peaks arising due to the coupling between the ionizable side chains and the main amide-I bands, and (d) the cross peaks arising between the COOH side chain of Asp and Glu and the TFA.

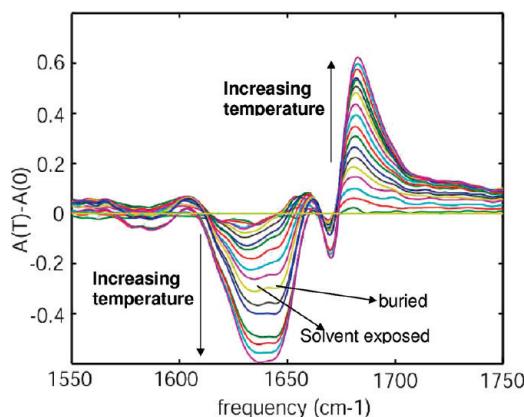


Figure 5. Difference infrared spectra of Villin headpiece (HP35). Difference FTIR spectra of unlabeled HP35 as a function of temperature, in the range of 0–90 °C in 5° increments.

objective of the theory will be to augment the interpretation of the cross peaks shown in Figure 4c. Cross peaks were also observed between the side chain COOH and the TFA as shown in Figure 4d.

Temperature Dependence of FTIR. The difference FTIR spectra of unlabeled HP35 (Figure 5) at each temperature were obtained by subtraction of the absorbance at the lowest temperature (0 °C). Figure 5 shows a decrease in cross section in the range of 1610–1655 cm⁻¹ and an increase in cross section in the range of 1655–1750 cm⁻¹ with the increase in temperature. The decrease in cross section at 1655–1750 cm⁻¹ identifies two underlying components as marked in Figure 5.

Figure 6a shows the 2D-IR spectrum of HP35 at 85 °C. The temperature increase causes a 60% decrease in the signal amplitude for the transitions at 1630 and 1643 cm⁻¹. New pairs of diagonal peaks are observed at 1680 and 1690 cm⁻¹ (Figure

6a) at the higher temperature. The ratio of the signal amplitude of the peak at 1680 cm⁻¹ to that at 1690 cm⁻¹ is 0.33. A cross peak can also be clearly seen between these two transitions. The intensity of the peak at 1660 cm⁻¹ increases with temperature (it becomes ~50% of that of the peak at 1643 cm⁻¹ at 85 °C). Details of different regions of this spectrum are displayed in Figure 6 for comparison with the room temperature 2D-IR spectra as shown in Figure 4. The intensity of the overlapping diagonal peaks in the frequency region at 1560–1590 cm⁻¹ shown in Figure 4b was decreased to 50% of that at room temperature by heating to 85 °C. The cross peaks originating from the carboxylate side chains and amide-I show a ~50% decrease in intensity at 85 °C compared with those of the room temperature spectrum (see Figure 4c). A 50% increase in intensity is observed for the cross peaks between the side chain COOH and the TFA at around 1700–1730 cm⁻¹ (Figure 4d), and the diagonal peak at 1674 cm⁻¹ from residual TFA decreases by 30% in signal amplitude on heating.

4. Discussion of Experimental Results

The amide-I band is primarily composed of two components, one from the carbonyl groups buried in the hydrophobic core (~1643 cm⁻¹) and the other from the carbonyls that are solvent exposed (~1630 cm⁻¹). The presence of these two bands composing the main amide-I peak were confirmed from the temperature dependent difference FTIR spectrum, which shows the growth of two distinct negative difference peaks in the range of 1610–1655 cm⁻¹ with an increase in temperature as seen in Figure 5. This result is comparable to that obtained by Dyer and co-workers.²⁸ No secondary structural units in peptides and proteins are reported to have amide-I transitions below 1600 cm⁻¹. Thus, the broad absorption at ~1560–1600 cm⁻¹ was considered to be associated with side chain transitions of the amino acids. The carboxylate groups of aspartic and glutamic

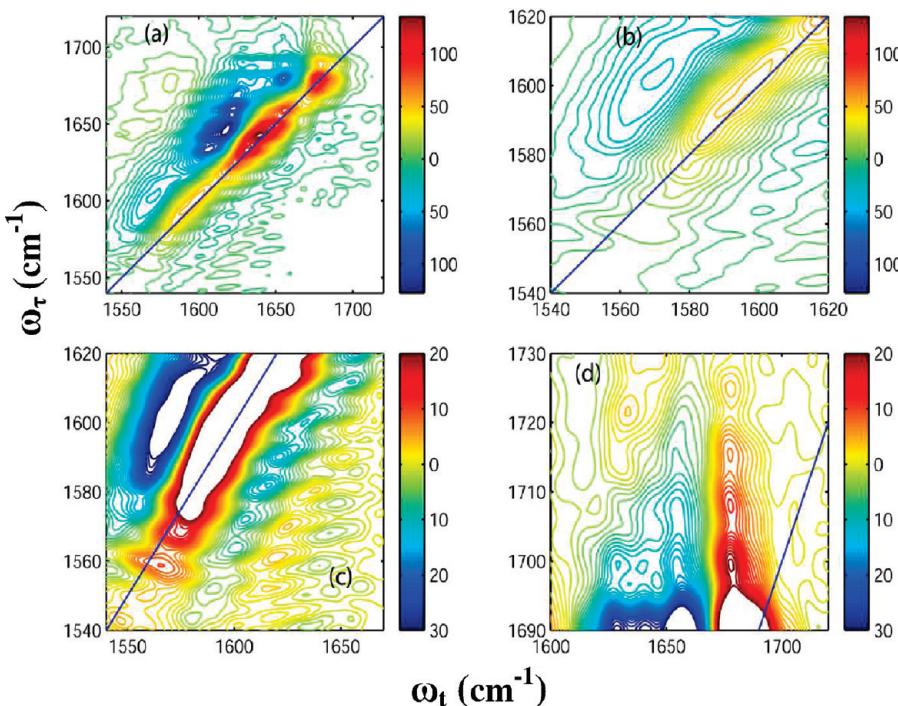


Figure 6. 2D-IR spectra of the Villin headpiece at 85 °C. Real part of the absorptive 2D-IR spectrum of HP35 in neat D₂O ($T = 0$) at 85 °C focusing on (a) the whole frequency range for the carbonyl stretches, (b) the diagonal peaks from Asp and Glu ionizable side chains, (c) the cross peaks arising because of the coupling between the ionizable side chains and the main amide-I bands, and (d) the cross peaks arising between the COOH side chain of Asp and Glu and the TFA.

acids are reported²⁹ to have transitions at 1584 and 1567 cm⁻¹, respectively, in D₂O. A symmetric stretching mode (CN₃H₅⁺) of arginine (R) has been reported to be near 1581–1586 cm⁻¹. These are the only reports that we have located concerning the transitions of amino acid side chains in the range of 1560–1600 cm⁻¹. The increase in cross section of the peak in the 1560–1600 cm⁻¹ region of HP35 at pD ∼ 11.0 confirmed the fact that this broad band is not likely to be associated with arginine. A small decrease in cross section in the region at 1560–1600 cm⁻¹ at pD ∼ 3.0, along with a small increase in the region above 1700 cm⁻¹ (the carbonyl stretch of the carboxylic acid side chain), occurred on lowering the pD to 3.0. This result is consistent with the transitions at 1560–1600 cm⁻¹ being from the carboxylate side chains. An ionizable side chain in a protein usually adopts a charge determined by its standard pK_a (3.9 and 4.3 for Asp and Glu, respectively). However, the pK_a of an ionizable group in a protein is greatly influenced by its geometry and its interaction with the surrounding environment. When the pD was further decreased to 1.0, a distinct decrease in cross section was observed at 1560–1600 cm⁻¹, confirming the fact that the broad transition in the region at 1560–1600 cm⁻¹ is associated with the carboxylate side chains of Asp and Glu. Moreover, the increase in cross section in the region at ∼1700–1730 cm⁻¹ is consistent with the shift of the equilibrium toward the acid forms. This assignment is consistent with the calculations of Scherega et al.³⁰ who found that Asp44, Glu45, and Asp46 of the Villin headpiece (HP36) displayed a significantly different average degree of charge from that predicted from the standard pK_a . This broad IR absorption arising from the carboxylate groups of the amino acid side chains has not been reported previously. Its frequency range overlaps with that expected for the amide-I transition region of the ¹³C= ¹⁶O and ¹³C= ¹⁸O isotope edited amino acids, such as reported in reference 28.

The carboxylate transition region of the difference FTIR spectrum appears to be composed of overlapping bands, so it

was natural to use 2D-IR spectroscopy in an attempt to simplify their interpretation. The diagonal peak at 1660 cm⁻¹ most probably represents the amide-I transitions of the amino acids from nonhelical regions of the protein. These residues are fewer in number compared with the helical residues accounting for their intensity of the diagonal peak at 1660 cm⁻¹ being only 20% of that of the main band in the 2D-IR spectrum. Diagonal peaks in the region at 1700–1730 cm⁻¹ are not at all prominent, though their cross peaks (Figure 4d) can be seen. This result suggests that a minor fraction of the acid form is also present in the sample solution in neat D₂O. The helices of HP35 become more disordered as the temperature is raised and the protein unfolds. The new diagonal peak at 1680 cm⁻¹ is a signature of these disordered structures, and the peak at 1660 cm⁻¹ becomes more pronounced for similar reasons. At elevated temperatures, the main band transitions at 1643 and 1630 cm⁻¹ at 85 °C are decreased to 55% and 40% of their initial amplitude at room temperature, and the cross section of the carboxylate side chain band decreases. There is a corresponding increase in the region at ∼1700–1730 cm⁻¹ that we suggest arises from increased exposure to the solvent of ionizable carboxylate groups, thereby shifting the equilibrium to the acid side.

The 2D-IR spectrum of unlabeled HP35 shows cross peaks between the carboxylate side chains and the main amide-I bands (see Figure 4). The positive cross peaks (arising from $\nu = 0 \rightarrow \nu = 1$ transitions) and negative cross peaks (arising from $\nu = 1 \rightarrow \nu = 2$ transitions) are found to be interfering with one another and have intensities of 5% and 10%, respectively, as compared with the amide-I band of the protein. The frequencies along ω_r of the string of cross peaks near $\omega_t = 1640$ cm⁻¹ indicate that they are arising from four diagonal peaks at frequencies at 1589, 1581, 1574, and 1564 cm⁻¹. These peaks are not seen directly in the diagonal traces. Four out of six cross peak features in Figure 4c were chosen as they correspond to the cross peaks arising from the four carboxylate peaks. The remaining two cross peaks are from the amide-II band which is around 1520–1555 cm⁻¹. The frequencies

along ω_r of the four cross peaks correspond to the diagonal peak frequencies arising from carboxylate side chains. Careful data analysis confirms that the cross peaks between the carboxylates and the amide-I band are from coupling. These cross peaks between the carboxylates and the main amide-I band are clearly visible in both the absorptive and the rephasing spectra.

The cross peak locations are shifted to lower frequencies by about 2 cm^{-1} in the rephasing spectra as a result of line shape distortion but otherwise form an identical pattern to that seen in the absorptive spectra and shown in Figure 4c. The ionizable side chains appear to be coupled to only a few amide-I modes as is evidenced by the cross peaks being narrower than the diagonal peaks of the amide-I band comprising the amide-I modes. As reported by Barth and Zscherp,²⁹ the carboxylate side chain transitions from Asp and Glu are expected at 1584 and 1567 cm^{-1} , respectively, in D_2O . The peak frequencies depend strongly on the environment of the individual amino acids. Theoretical calculations by Scherega et al.³⁰ showed the ionizable groups of the two Asp's in HP35 to have an average charge of ~ -1.0 at $\text{pD} = 3.7$, whereas the charges on Glu45 and Glu72 are ~ -0.5 and ~ -0.2 , respectively. The 2D-IR spectrum shows the diagonal peaks at 1589 and 1581 cm^{-1} to be more intense than those at 1574 and 1564 cm^{-1} , the former high frequency pair being attributed to the ionizable carboxylate side chain of Asp and the latter to that of Glu. This is further supported by the data shown in Figure 4d where the lower frequency cross peaks between the TFA and the side chain COOH are more intense than the higher frequency ones. As reported by Barth et al.,²⁹ the COOH side chain transitions from Asp and Glu appear at 1713 and 1706 cm^{-1} , respectively, in D_2O . The TFA absorption is shifted by ~ 75 – 80 cm^{-1} from the absorption of the carboxylate side chain and is considered not to affect the cross peak region we are looking at in the experiments of this paper.

Although the cross peaks between the carboxylate groups and the main amide-I band are clearly observed in the 2D-IR spectrum, their assignment to specific residues requires a quantitative description of the carboxylate vibrational spectra as elaborated in the next section. The simulation of the cross peaks arising from interactions between the carboxylates and the amide-I modes, typified by the results in Figure 4c, represents a main objective of this paper. The theoretical development of the vibrational spectra are presented in sections 5–8. In section 5 the vibrational Hamiltonian for Villin HP35 is given in terms of the amide-I modes represented as vibrational excitons interacting with light pulses. Section 6 describes the development of a new electrostatic map that allows the determination of the acid and carboxylate ion vibrational frequency distributions from the charges obtained from the MD simulations. The computed infrared response is then presented and discussed in section 7.

5. Vibrational Hamiltonian for Villin HP36

Simulations of the infrared response of the Villin headpiece subdomain were performed using an effective vibrational Hamiltonian that includes 39 vibrational modes: the 34 amide-I modes incorporated between the residues Leu42 and Phe76, the four carboxylic/carboxylate modes corresponding to the side chains of the residues Asp44, Glu45, Asp46, and Glu72, and the C-terminal carboxylic/carboxylate Phe76. Using the boson creation and annihilation operator of a vibrational exciton B_i^\dagger and B_i , the model Hamiltonian is

$$H(t) = \sum_i \hbar \omega_i(t) B_i^\dagger B_i + \sum_{ij} \hbar J_{ij}(t) B_i^\dagger B_j - \sum_i \hbar \frac{\Delta_i}{2} B_i^{\dagger 2} B_i^2 \quad (4)$$

where $\omega_i(t)$ is the fundamental frequency of the local mode i , J_{ij} is the coupling between the mode i and the mode j , and Δ_i is the anharmonicity. The indices $i = 1, \dots, 34$ denote the 34 amide-I vibrations, and $i = 35, \dots, 39$ designates the five carboxylic/carboxylate modes. The coupling of the excitons with the electric field $\mathbf{E}(t)$ of the laser pulses is

$$H'(t) = \sum_i \mu_i(t) \mathbf{E}(t) (B_i + B_i^\dagger) \quad (5)$$

with $\mu_i(t)$ as the transition dipole of the i th vibrational mode.

Several models have been used in the past to simulate the vibrational parameters and their corresponding fluctuations. To establish a link between the structure fluctuations and the Hamiltonian, DFT maps were recently introduced to connect ab initio calculations and MD coordinates.^{31–39} The electrostatic parameters represent the adiabatic interaction between intramolecular high frequency vibrations and slow structural motions. Many simulations using several parametrization schemes have been employed for the amide-I mode.^{31–37} Cho and co-workers parametrized the amide-I vibration by identifying the electrostatic potential at four coordinates corresponding to the atoms C, O, N, and H of the amide bond.³¹ An anharmonic vibrational Hamiltonian for the amide I, II, III, and A modes of nuclear microanalysis has been recast in terms of 19 components of an external electric field, and its first and second derivative tensors have been evaluated at a single point.³⁴ Other parametrizations have used the electrostatic field and its gradients at several points.^{32,35,37} In all approaches the electrostatic potential is sampled in the neighborhood of the atoms involved in the vibrational mode. Experimental tests of such maps have been reported for larger peptides.⁴⁰ Similar electrostatic maps have been developed for the OH stretch of H_2O and the HOD molecule.^{38,39}

We have used the CHO4 electrostatic map to simulate the fluctuations of the fundamental frequency of the amide-I modes of the Villin headpiece subdomain.^{41,42} The amide-I transition dipole and the intraband coupling are used as defined by the Torii and Tasumi transition-dipole coupling (TDC) model.⁴³ The coupling between nearest-neighbor amide-I modes is given by the Torii and Tasumi dihedral angles map.⁴³ The amide-I mode anharmonicity is fixed to the measured value of -16 cm^{-1} .⁴⁴

A parametrization comparable with that developed for amide-I modes is not available for the carboxylic/carboxylate vibrational transitions, so a new electrostatic map for these fundamental frequencies and transition dipoles needed to be developed, the details of which are described below. The computed transition dipoles are used to obtain the carboxylic and carboxylate intraband couplings and the carboxylic/amide-I and carboxylate/amide-I interband couplings by the TDC model.

6. Parameterization of the Carboxylic Acid and Carboxylate Vibrational Modes

In the following simulations we used acetic acid in its ionized and acid forms as models for the vibrational frequency parameters of the ionizable side chains of the Villin headpiece subdomain. The acetic acid carbonyl diagonal anharmonicity is fixed at -16 cm^{-1} and that of the acetate carboxylate asymmetric stretch at -24 cm^{-1} .

TABLE 1: Acetic Acid Carbonyl Vibration Electrostatic Map Coefficients

$\omega_{01}^{(0)}$	1794.99	$\mu_{01}^{x(0)}$	-0.273784	$\mu_{01}^{y(0)}$	0.034589	$\mu_{01}^{z(0)}$	0.001108
$C_{x,C}$	7061.95	$D_{x,C}^x$	1.073082	$D_{x,C}^y$	-0.216676	$D_{x,C}^z$	-0.332509
$C_{y,C}$	-161.19	$D_{y,C}^x$	-0.188739	$D_{y,C}^y$	-0.544033	$D_{y,C}^z$	0.559513
$C_{xx,C}$	4401.51	$D_{xx,C}^x$	1.66013	$D_{xx,C}^y$	0.912052	$D_{xx,C}^z$	0.582458
$C_{yy,C}$	-8319.04	$D_{yy,C}^x$	-0.313005	$D_{yy,C}^y$	1.131204	$D_{yy,C}^z$	1.665374
$C_{zz,C}$	-2686.97	$D_{zz,C}^x$	0.739752	$D_{zz,C}^y$	1.130211	$D_{zz,C}^z$	1.525385
$C_{xy,C}$	577.7	$D_{xy,C}^x$	-0.597171	$D_{xy,C}^y$	-0.72804	$D_{xy,C}^z$	0.929922
C_{x,O_1}	-3084.59	D_{x,O_1}^x	0.411037	D_{x,O_1}^y	0.200399	D_{x,O_1}^z	0.475283
C_{y,O_1}	313.7	D_{y,O_1}^x	0.307642	D_{y,O_1}^y	0.805647	D_{y,O_1}^z	-0.573463
C_{xx,O_1}	703.1	D_{xx,O_1}^x	1.078939	D_{xx,O_1}^y	0.920538	D_{xx,O_1}^z	0.896748
C_{yy,O_1}	-224.45	D_{yy,O_1}^x	0.873689	D_{yy,O_1}^y	1.061465	D_{yy,O_1}^z	1.091423
C_{zz,O_1}	-1322.81	D_{zz,O_1}^x	1.015866	D_{zz,O_1}^y	1.090813	D_{zz,O_1}^z	1.099215
C_{xy,O_1}	-286.06	D_{xy,O_1}^x	-0.250912	D_{xy,O_1}^y	-0.280752	D_{xy,O_1}^z	0.528865

Frequency and Transition-Dipole Parametrization of the Acetic Acid Vibrations. To describe the anharmonic acetic acid vibrational dynamics we assumed that the acid carbonyl normal mode in solution is close to the gas-phase normal mode, and we neglected its anharmonic couplings to other normal modes. As a result, the Hamiltonian describing the carbonyl vibration is expanded to fourth-order in the single carbonyl normal mode coordinate Q :

$$H = \frac{P^2}{2M} + \sum_{k=0}^4 \frac{1}{k!} V_k [\mathbf{E}(\mathbf{r})] Q^k \quad (6)$$

The dipole is similarly expanded as

$$\boldsymbol{\mu} = \boldsymbol{\mu}_0 [\mathbf{E}(\mathbf{r})] + \boldsymbol{\mu}_1 [\mathbf{E}(\mathbf{r})] Q + \dots \quad (7)$$

Here M and P correspond respectively to the reduced mass and conjugate momentum of the normal mode Q . All potential parameters depend parametrically on the electrostatic field $\mathbf{E}(\mathbf{r})$ generated by the surrounding molecules. To obtain the dependence on the parameters of the electrostatic field, we used MD simulations of 100 water clusters, each containing one acetic acid molecule. The MD structure of acetic acid is replaced by the minimized DFT gas-phase structure and the water molecules by point charges (TIP3P model). Electronic structure calculations were performed by taking 25 random <0.1 Å displacements along Q . The intramolecular parameters V_0, V_1, V_2, V_3 , and V_4 were then extracted using a linear least-squares fitting. The same methodology was applied to obtain the permanent dipole $\boldsymbol{\mu}_0$ and the transition dipole $\boldsymbol{\mu}_1$. For each molecular cluster, the energies of the lowest quantum number states were found by numerical diagonalization of H including levels up to $n = 20$. In this manner we obtained the frequency difference $\omega_{10}[\mathbf{E}(\mathbf{r})]$ between the ground state and the first excited state and the corresponding transition dipole $\mu_{10}[\mathbf{E}(\mathbf{r})]$. The fundamental frequency and the corresponding transition dipole were then expanded linearly in terms of a finite number of electrostatic parameters. For the acetic acid vibration, we choose 12 parameters, $E_x, E_y, E_{xx}, E_{yy}, E_{zz}$, and E_{xy} calculated at the location of both the C and the O₁ atoms, with E_a being the electrostatic field in the a direction and E_{ab} its derivative with respect to b .

$$\begin{aligned} \omega_{10}[\mathbf{E}(\mathbf{r})] &= \omega_{10}^{(0)} + \sum_{ij} C_{ij} E_i(r_j) \\ \mu_{10}[\mathbf{E}(\mathbf{r})] &= \mu_{10}^{(0)} + \sum_{ij} D_{ij} E_i(r_j) \end{aligned} \quad (8)$$

with $i = x, y, z, xx, yy, zz, xy$ and $j = C, O_1$. The parameters $\omega_{01}^{(0)}$, C_{ij} , $\mu_{01}^{(0)}$, and D_{ij} , reported in Table 1, were obtained by a linear least-squares fitting on the 100 configurations. Equation 8 gives a very good correlation between the DFT frequency and the electrostatic map (see Figure 2a) (in Tables 1 and 2 the electrostatic field is in atomic units (au), the frequency coefficients are in cm⁻¹/au, and the dipole coefficients in D/au).

Frequency and Transition-Dipole Parametrization of the Acetate Vibration. The acetate ion contains two strongly coupled CO vibrations, yielding a symmetric, Q_s , and an antisymmetric, Q_a , stretching mode. We found the harmonic gas-phase frequency for the symmetric vibration to be 1348 cm⁻¹ and the antisymmetric vibration to be 1635 cm⁻¹. To obtain the vibrational frequency fluctuations, a simulation of the antisymmetric vibration region must take into account its coupling with the symmetric mode. We therefore used a vibrational Hamiltonian that contains both modes:

$$\begin{aligned} H = & \sum_i \frac{P_i^2}{2M_i} + \sum_i V_i [\mathbf{E}(\mathbf{r})] Q_i + \frac{1}{2} \sum_{ij} V_{ij} [\mathbf{E}(\mathbf{r})] Q_i Q_j + \\ & \frac{1}{3!} \sum_{ijk} V_{ijk} [\mathbf{E}(\mathbf{r})] Q_i Q_j Q_k + \frac{1}{4!} \sum_{ijkl} V_{ijkl} [\mathbf{E}(\mathbf{r})] Q_i Q_j Q_k Q_l + \dots \end{aligned} \quad (9)$$

where $i, j, k, l = a, s$. To compute the intramolecular potential, 50 random <0.1 Å deformations are generated in the xy plane. The 15 parameters in eq 9 are extracted by linear least-squares fitting. For each of the 400 water cluster configurations, numerical diagonalization of the Hamiltonian leads to the antisymmetric stretch fundamental frequencies and the corresponding transition dipoles. To parametrize both quantities we expand them using 33 electrostatic parameters: $E_x, E_{xx}, E_{yy}, E_{zz}, E_{yz}, E_{xy}, E_{xz}, E_{yz}$, and E_{xy} , computed on the atom C, and $E_x, E_y, E_{xx}, E_{yy}, E_{zz}, E_{xy}, E_{xz}, E_{yy}, E_{xy}, E_{xz}$, and E_{yz} computed on each of the atoms O₁ and O₂. A similar expansion to eq 8 was used, and the corresponding coefficients are reported in Table 2. This parametrization leads to the DFT/MAP correlation shown in Figure 2b.

7. Simulation of the Infrared Response

The Direct Nonlinear Exciton Propagation approach was used to simulate the linear and nonlinear IR spectra.⁴⁵ Similar to other approaches,⁴⁶ the excitons are propagated numerically under the time-dependent Hamiltonian $H(t)$ given in eq 1. However, the quasi-particle Nonlinear Exciton Equation^{47,48} picture is retained. Unlike the sum-overstates picture, it is unnecessary to compute all the Liouville pathways.^{46,48} A strong cancellation between

TABLE 2: Acetate Molecule Asymmetric Carbonyl Vibration Electrostatic Map Coefficients

$\omega_{01}^{(0)}$	1641.13	$\mu_{01}^{x(0)}$	0.011625	$\mu_{01}^{y(0)}$	0.382309	$\mu_{01}^{z(0)}$	0.000523
$C_{x,C}$	-9369.28	$D_{x,C}^x$	-0.282794	$D_{x,C}^y$	1.985792	$D_{x,C}^z$	0.610575
$C_{xx,C}$	-9346.53	$D_{xx,C}^x$	-5.243465	$D_{xx,C}^y$	4.936492	$D_{xx,C}^z$	-0.191843
$C_{yy,C}$	1711.93	$D_{yy,C}^x$	4.891628	$D_{yy,C}^y$	-1.0319	$D_{yy,C}^z$	1.909363
$C_{zz,C}$	592.38	$D_{zz,C}^x$	-4.335272	$D_{zz,C}^y$	3.484305	$D_{zz,C}^z$	-0.365647
$C_{xy,C}$	-83.13	$D_{xy,C}^x$	0.143076	$D_{xy,C}^y$	0.056743	$D_{xy,C}^z$	-1.3309
$C_{xxx,C}$	4630.01	$D_{xxx,C}^x$	0.208477	$D_{xxx,C}^y$	1.058037	$D_{xxx,C}^z$	0.416076
$C_{xyy,C}$	-13330.66	$D_{xyy,C}^x$	3.030343	$D_{xyy,C}^y$	1.239544	$D_{xyy,C}^z$	2.565149
$C_{xzz,C}$	2966.15	$D_{xzz,C}^x$	0.711041	$D_{xzz,C}^y$	0.680399	$D_{xzz,C}^z$	0.727763
$C_{xyz,C}$	196.62	$D_{xyz,C}^x$	0.022373	$D_{xyz,C}^y$	-0.158278	$D_{xyz,C}^z$	-0.792099
C_{x,O_1}	6434.73	D_{x,O_1}^x	0.274243	D_{x,O_1}^y	-1.168619	D_{x,O_1}^z	-0.300579
C_{y,O_1}	160.83	D_{y,O_1}^x	-3.480055	D_{y,O_1}^y	1.241321	D_{y,O_1}^z	-0.938871
C_{xx,O_1}	-2172.37	D_{xx,O_1}^x	0.059112	D_{xx,O_1}^y	1.64456	D_{xx,O_1}^z	0.897237
C_{yy,O_1}	1557.63	D_{yy,O_1}^x	3.471973	D_{yy,O_1}^y	-0.272564	D_{yy,O_1}^z	1.558936
C_{zz,O_1}	1494.8	D_{zz,O_1}^x	0.114831	D_{zz,O_1}^y	1.247511	D_{zz,O_1}^z	0.656936
C_{xy,O_1}	-3920.65	D_{xy,O_1}^x	1.471899	D_{xy,O_1}^y	-0.209534	D_{xy,O_1}^z	0.580431
C_{xxx,O_1}	144.34	D_{xxx,O_1}^x	1.343968	D_{xxx,O_1}^y	0.924385	D_{xxx,O_1}^z	1.077781
C_{yyy,O_1}	-290.91	D_{yyy,O_1}^x	0.531879	D_{yyy,O_1}^y	1.290794	D_{yyy,O_1}^z	0.875888
C_{xyy,O_1}	61.73	D_{xyy,O_1}^x	0.452601	D_{xyy,O_1}^y	1.152418	D_{xyy,O_1}^z	0.869541
C_{xyy,O_1}	990.11	D_{xyy,O_1}^x	0.980082	D_{xyy,O_1}^y	0.88304	D_{xyy,O_1}^z	0.957939
C_{xzz,O_2}	-351.33	D_{xzz,O_2}^x	1.406842	D_{xzz,O_2}^y	0.853096	D_{xzz,O_2}^z	1.103958
C_{yzz,O_2}	-214.98	D_{yzz,O_2}^x	1.457727	D_{yzz,O_2}^y	0.780095	D_{yzz,O_2}^z	1.13968
C_{x,O_2}	6434.73	D_{x,O_2}^x	0.274243	D_{x,O_2}^y	-1.168619	D_{x,O_2}^z	-0.300579
C_{y,O_2}	-160.83	D_{y,O_2}^x	3.480055	D_{y,O_2}^y	-1.241321	D_{y,O_2}^z	0.938871
C_{xx,O_2}	-2172.37	D_{xx,O_2}^x	0.059112	D_{xx,O_2}^y	1.64456	D_{xx,O_2}^z	0.897237
C_{yy,O_2}	1557.63	D_{yy,O_2}^x	3.471973	D_{yy,O_2}^y	-0.272564	D_{yy,O_2}^z	1.558936
C_{zz,O_2}	1494.8	D_{zz,O_2}^x	0.114831	D_{zz,O_2}^y	1.247511	D_{zz,O_2}^z	0.656936
C_{xy,O_2}	-3920.65	D_{xy,O_2}^x	-1.471899	D_{xy,O_2}^y	0.209534	D_{xy,O_2}^z	-0.580431
C_{xxx,O_2}	144.34	D_{xxx,O_2}^x	1.343968	D_{xxx,O_2}^y	0.924385	D_{xxx,O_2}^z	1.077781
C_{yyy,O_2}	290.91	D_{yyy,O_2}^x	-0.531879	D_{yyy,O_2}^y	-1.290794	D_{yyy,O_2}^z	-0.875888
C_{xyy,O_2}	61.73	D_{xyy,O_2}^x	0.452601	D_{xyy,O_2}^y	1.152418	D_{xyy,O_2}^z	0.869541
C_{xyy,O_2}	-990.11	D_{xyy,O_2}^x	-0.980082	D_{xyy,O_2}^y	-0.88304	D_{xyy,O_2}^z	-0.957939
C_{xzz,O_2}	-351.33	D_{xzz,O_2}^x	1.406842	D_{xzz,O_2}^y	0.853096	D_{xzz,O_2}^z	1.103958
C_{yzz,O_2}	214.98	D_{yzz,O_2}^x	-1.457727	D_{yzz,O_2}^y	-0.780095	D_{yzz,O_2}^z	-1.13968

the different pathways is naturally built in to the theory, and the signal is computed directly. Using this methodology we have computed separately the rephasing and the nonrephasing signals. The final signal corresponds to the sum of these contributions. To allow meaningful comparisons between theory and experiment and to compensate for errors in the DFT calculations of the vibrational frequencies, five parameters were adjusted in our simulations: the central frequencies of the carboxylate, carboxylic acid, and amide-I modes, an overall scale factor $0 < \eta < 1$ for the couplings $J_{ij}(t)$, which can be considered as incorporating the dielectric effect of the surrounding water, and a dephasing parameter Γ which represents an additional exponential relaxation process not included explicitly in the model.

Calculation of the Linear Absorption Spectra. To make the comparisons with experiments clearer we introduced a representation of the TFA (see Sample Preparation section) IR absorption into the simulation results by adding a Lorentzian of 8.4 cm^{-1} full width at half-maximum centered around 1673 cm^{-1} to the computed spectrum at 27°C . The simulated absorption spectrum for the FI and the FN trajectories are shown in Figure 7a. The open circles correspond to the experiment in neat D_2O . The experimental linear spectra in Figure 3 show both the carboxylate and the carboxylic acid transitions at lower and higher frequencies than the amide-I region, respectively, indicating that the peptide has significant populations of those different ionization states in equilibrium under all conditions. Since the simulations assume that the peptide has just two states, one where the carboxylate groups are fully ionized (indicated by I) and the other a fully neutral (N) state, statistical mixtures of FI and FN, labeled as FI + FN in Figure 7a, chosen with an

FI/FN ratio of 2:1, were used to compare with the experiment. The FI infrared absorption shows three main peaks: the main amide I band and two bands corresponding to the carboxylate vibrations located at 1648 cm^{-1} , 1581 cm^{-1} , and 1608 cm^{-1} , respectively. The experimental spectrum shows bands that are broader than those computed, which is consistent with the protein having relaxation processes that are not included in the simulation.

To trace the origin of the two peaks in the carboxylate absorption region we computed the projected density of states $\rho_i(\omega)$ which shows the contributions of a site to different spectral regions

$$\rho_i(\omega) \equiv \sum_{\lambda} \langle |\Psi_{i,\lambda}|^2 \delta(\omega - \omega_{\lambda}) \rangle \quad (10)$$

where $\Psi_{i,\lambda}$ is the component of the eigenstate λ on the site i and ω_{λ} are the corresponding eigenfrequencies of the Hamiltonian (eq 1). The $\rho_i(\omega)$ displayed in Figure 8 show that the two absorption bands correspond to different amino acids with the low frequency band characterizing the glutamate and the high frequency band corresponding to both the aspartate and the C-terminal carboxylate.

Figure 7b,c shows that the linear infrared spectrum is insensitive to the dipole-dipole coupling scaling factor η . However, the nonlinear photon echo signal is much more sensitive to η as shown in the next section. When the homogeneous contribution is fixed at $\Gamma = 5.5 \text{ cm}^{-1}$, the agreement between theory and experiment is optimized as illustrated in the IR absorption computed from the UI trajectory,

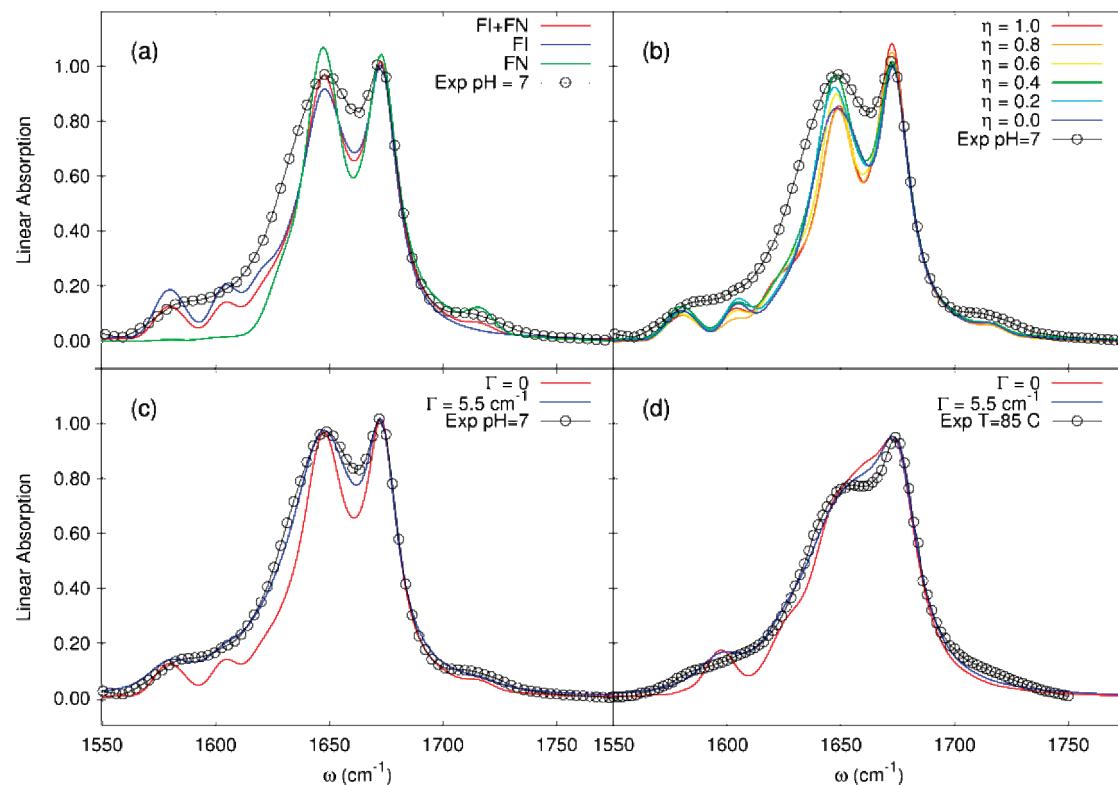


Figure 7. Simulated linear absorption spectrum of the Villin headpiece (HP36). (a) Simulated linear absorption of FI (blue), FN (green), and the (FI + FN) mixture (red) with $\eta = 0.4$ and $\Gamma = 0$ compared with the experiment at 27 °C (open circles). (b) Spectra of the (FN + FI) mixture at various indicated values of η at $\Gamma = 0$ compared with experiments as in (a) (open circles). (c) Spectra of the (FI + FN) mixture at various indicated values of Γ at $\eta = 0.4$ compared with the experiment as in (a). (d) The simulated UI spectrum with $\eta = 0.4$ compared with the experiment at 85 °C (open circles).

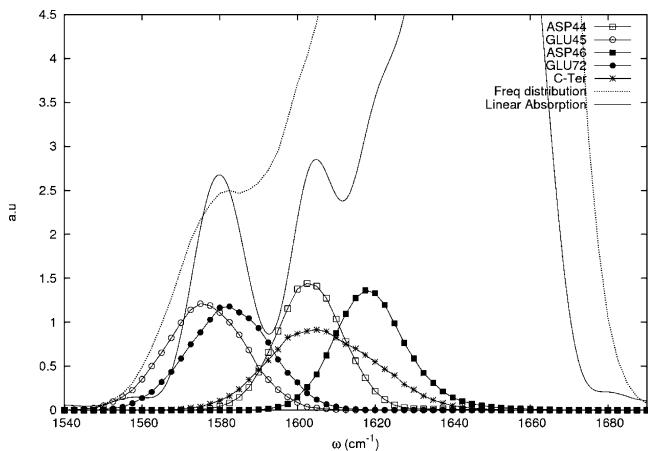


Figure 8. Density of states. Projected density of states of the carboxylate side chains. The total density of states is the dotted line.

displayed in Figure 7d. For the 85° simulation, also shown in Figure 7, the TFA Lorentzian line width was increased to 10.6 cm^{-1} to mimic the variation with temperature indicated by the experimental spectra.

2D-IR Photon Echo Simulations. The 2D-IR photon echo simulations were carried out for FI, FN, and UI. In contrast with the linear IR absorption, there are strong variations in 2D-IR when the dipole–dipole scaling factor is varied from $\eta = 0$ to $\eta = 1$ (Figure 9). For $\eta = 1$, the diagonal of the 2D signal shows three well-resolved bands corresponding to two carboxylates and the amide-I band. For smaller coupling ($\eta \leq 0.4$) the higher frequency carboxylate bands tend to merge with the amide-I transitions. The case $\eta = 0.4$ provides the best agreement with the experiments.

The 2D-IR photon echo signals for the FI + FN sample are displayed in Figure 10a. The expanded scale of the cross peak region between the carboxylate and the amide-I bands, displayed in Figure 10b, agrees well with the experimental identification of cross peaks in this region. The simulated 2D-IR of the UI trajectory is shown in Figure 10d, and the close-up of the carboxylate/amide-I cross peak region is displayed in Figure 10e. In the unfolded state, both the carboxylate and the amide-I bands are broader, and the signal amplitude in the carboxylate/amide-I cross peak region becomes much smaller than in the folded state, again in qualitative agreement with the experimental spectra in Figure 6.

The delocalization of the vibrational modes for the folded FI and the unfolded UI configurations is made clear from the maps of the average coupling shown in Figure 11. In the amide-I/amide-I coupling (the amide-I modes are 1–34), the nearest neighbor coupling is dominant for both the folded and the unfolded configurations. Figure 11a shows the signature of the three α -helices of the Villin headpiece subdomain through their $n/(n + 3)$ coupling which is known to favor the delocalization of the excitons along a single spine of α -helices.⁴⁹ For both FI and UI configurations the carboxylate/carboxylate coupling (modes 35–39) is much weaker compared with the amide-I/amide-I coupling, indicating that the carboxylate vibrations are more isolated.

8. Further Comparison between Experiments and Simulations

Both 2D-IR experiments and simulations show cross peaks between the carboxylate side chains and the amide-I states of the Villin headpiece. The presence of these cross peaks in the

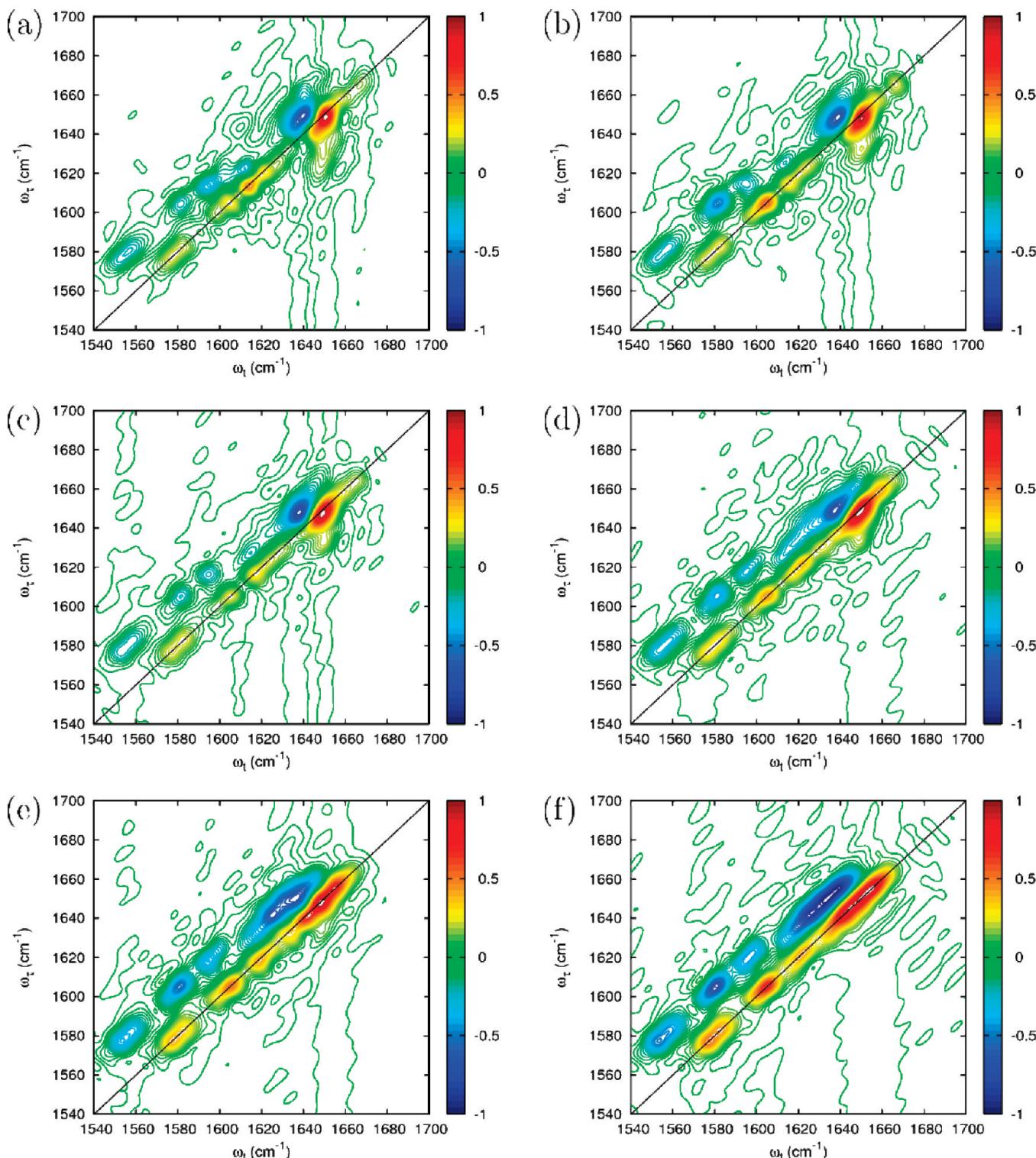


Figure 9. Coupling dependence of the simulated 2D-IR spectrum at 27 °C. Simulated photon echo using the folded FI configuration at $\Gamma = 0$: (a) $\eta = 1$, (b) $\eta = 0.8$, (c) $\eta = 0.6$, (d) $\eta = 0.4$, (e) $\eta = 0.2$, and (f) $\eta = 0.0$.

experiments proves unequivocally that there is measurable coupling between the carboxylates and the backbone amide states. This was not previously known or considered in the amide-I band analysis of proteins. The relative magnitudes of the couplings are estimable from a comparison of the 2D-IR experiments with theory based on dipole–dipole coupling; it would not be possible to establish these conclusions from a linear spectrum. The coupling between the carboxylate and the amide-I vibration depends strongly on the peptide structure and

the side chain. In the folded state, the C-terminal carboxylate group (no. 39 in Figure 11) is coupled to the nearest amide-I unit located at the end of the peptide (no. 35). The side chain Glu72 (no. 38) is mainly coupled to the amide-I unit located between the amino acids Glu72 and Lys73. The carboxylate side chain of Glu45 (no. 36) shows the weakest coupling to amide-I units. The side chain Asp44 (no. 35) is mainly coupled to the amide-I units located between Asp44 and Glu45 and between Thr54 and Arg55, which is a tertiary structure contact.

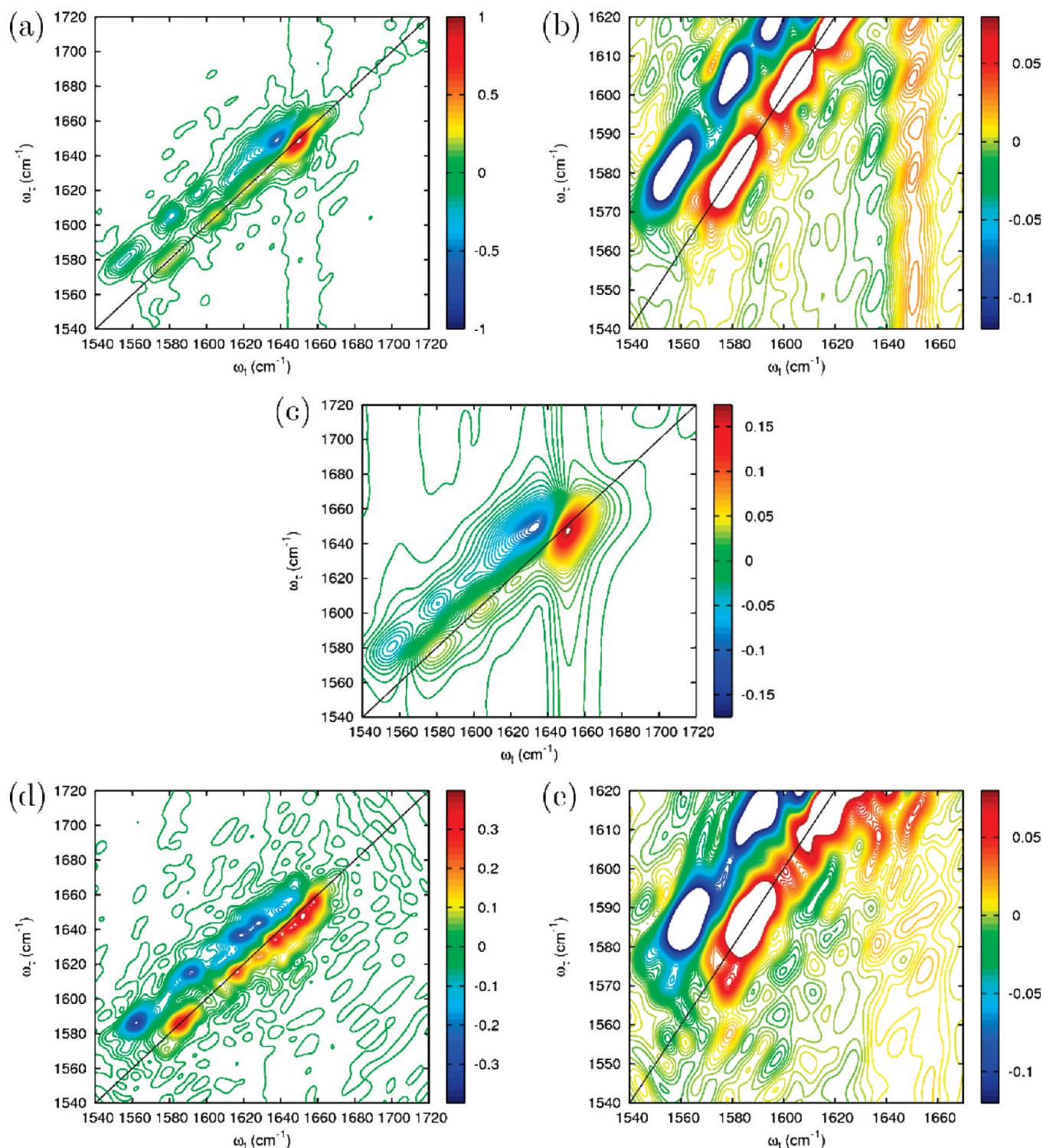


Figure 10. Dependence of the simulated 2D-IR spectrum on temperature and dephasing. Simulated photon echo at 27 °C of a 2:1 FI/FN mixture at $\eta = 0.4$: (a) full frequency range, $\Gamma = 0$, (b) close-up of the carboxylate/amide-I cross peak region, $\Gamma = 0$, and (c) full frequency range, $\Gamma = 5.5 \text{ cm}^{-1}$. Simulated photon echo at 85 °C of the UI configuration at $\eta = 0.4$ and $\Gamma = 0$: (d) full frequency range and (e) close-up of the carboxylate/amide-I cross peak region.

Finally the side chain (Asp46) is coupled to several amide-I units ranging from Leu42 to Phe47. Generally the carboxylates are significantly coupled only to the immediately adjacent backbone states, although in one case there is predicted to be a tertiary interaction. Both the experiment and the theory show that the aspartic and glutamic carboxylic acids show their carbonyl absorptions at a higher frequency than the amide-I.

Four cross peaks between the carboxylate groups and the main amide-I band are clearly observed in the experimental 2D-IR spectrum near $\omega_t = 1640 \text{ cm}^{-1}$ and $\omega_r = 1560\text{--}1600 \text{ cm}^{-1}$, whereas the diagonal peaks from the carboxylates are not clearly distinguishable. Vibrational response simulations of a 2:1 F1/FN mixture predicted two separated diagonal bands in the frequency range of 1560–1600 cm^{-1} ; the higher

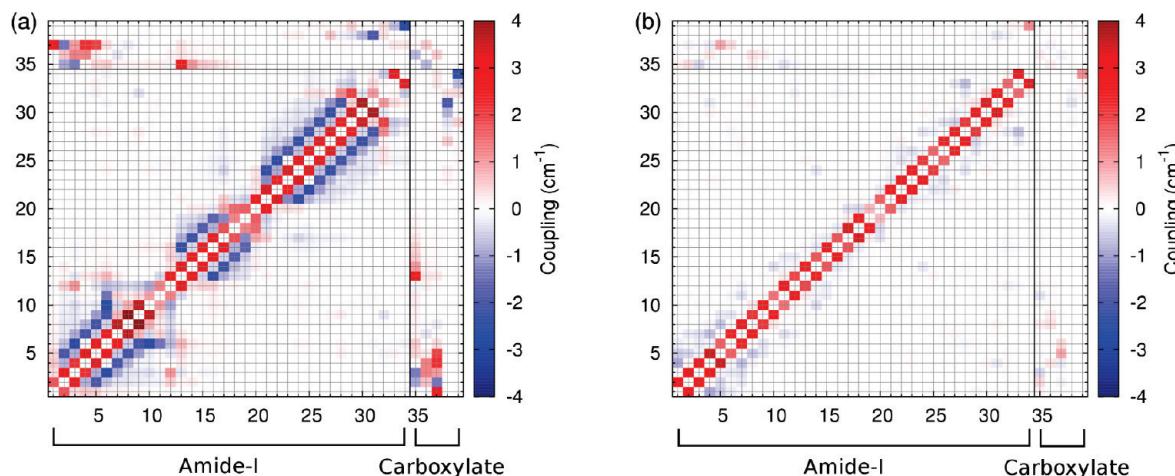


Figure 11. Average coupling among amide-I and carboxylate ion vibrational modes for: (a) the folded FI and (b) the unfolded UI configurations. The backbone residues are labeled 1–34 and the carboxylates 35–39.

frequency band corresponds to the aspartates and the C-terminal carboxylate, while the lower frequency band corresponds to the glutamates. The two diagonal bands give rise to cross peaks with the amide-I band near 1640 cm^{-1} , similar to what is seen in the experiment (compare Figures 4d and 10b). The density of states calculations predict that the Asp44 and Asp46 have a higher mean frequency compared with Glu45 and Glu72, respectively. The experimental 2D-IR spectrum shows a 50% decrease in the signal amplitudes of these cross peaks at 85°C ; however, the four cross peaks can still be clearly distinguished. In contrast, this cross-peak region essentially disappears at 85°C in the simulated photon echo spectrum (see Figure 10e). However, the diagonal amide-I features of the experimental 2D spectrum at 85°C (Figure 6) contain a 40% residual signal contribution from the folded states, whereas the simulation at 85°C was performed on the unfolded protein, equilibrated at 527°C . The residual folded states present in the experiment are considered to be the source of the distinguishable cross peaks between the carboxylate side chains and the amide-I band in the 2D-IR spectrum at higher temperatures in the experiment.

When the homogeneous line width in the simulation is increased to 5.5 cm^{-1} , the long vertical and horizontal tails in the computed 2D-IR spectra (Figure 10c) that are characteristic of an exponential decay overlap with the carboxylate/amide-I cross peak region and mask the corresponding signal amplitude. Those tails are not evident in the experimental spectra. This discrepancy between theory and experiment arises because the amide-I mode relaxation is mainly Gaussian and not dominated by the exponential dephasing process assumed in the simulation.

9. Conclusions

The carboxylate side chains were found to have significant coupling with the main amide-I band and give rise to distinct cross peaks with amide-I transitions. Simulations of the 2D-IR optical response of the Villin headpiece subdomain showed that these side chains significantly couple to only a few amide-I modes among all of the amino acid residues, which makes them useful as spectroscopic markers, providing information about the local structural behavior of the protein. The carboxylate absorption range of the Villin headpiece is composed of two separated bands corresponding respectively to the two amino acids, Asp and Glu. It was necessary to use a scaling factor of $\eta = 0.4$ for the transition dipole–dipole

coupling to reproduce the experimental results. The bands at higher frequency than the amide-I band corresponding to the aspartic acid and the C-terminal carboxylate are both significantly mixed with the amide-I band. Non-exponential relaxation process not included in the simulation will be required to explain the observed broadening. Both experiment and simulation find that the folded state of the Villin headpiece subdomain has cross peaks between the carboxylate and the amide-I bands that are significantly diminished by unfolding.

The negatively charged carboxylate side-chain vibrational transitions have significant spectral overlap with the $^{13}\text{C}=^{16}\text{O}$ and $^{13}\text{C}=^{18}\text{O}$ isotopically edited amide-I transition frequency region. Thus, careful analysis and spectral assignment is needed in interpreting the infrared spectra of isotopically edited amide-I transitions of proteins containing Asp and Glu.

Acknowledgment. The research was supported by the NIH Grant GM12592, by NSF, and by instrumentation through the Resource NIH P41 RR001348. S.M. greatly acknowledges the support of The National Institute of Health, Grant GM59230, and the National Science Foundation, Grant CHE-0745892. C.F. and S.M. would like to thank Prof. D. Tobias, Dr. D. Abramavicius, and Dr. W. Zhuang for helpful discussions.

References and Notes

- (1) McKnight, C. J.; Doering, D. S.; Matsudaira, P. T.; Kim, P. S. *J. Mol. Biol.* **1996**, *260*, 126–134.
- (2) McKnight, C. J.; Matsudaira, P. T.; Kim, P. S. *Nat. Struct. Biol.* **1997**, *4*, 180–184.
- (3) Frank, B. S.; Vardar, D.; Buckley, D. A.; McKnight, C. J. *Protein Sci.* **2002**, *11*, 680–687.
- (4) Kubelka, J.; Eaton, W. A.; Hofrichter, J. *J. J. Mol. Biol.* **2003**, *329*, 625–630.
- (5) Kubelka, J.; Hofrichter, J.; Eaton, W. A. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76–88.
- (6) Kubelka, J.; Henry, E. R.; Cellmer, T.; Hofrichter, J.; Eaton, W. A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 18655–18662.
- (7) Chiu, T. K.; Kubelka, J.; Herbst-Irmer, R.; Eaton, W. A.; Hofrichter, J.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7517–7522.
- (8) Laskowski, M.; Scheraga, H. A. *J. Am. Chem. Soc.* **1954**, *76*, 6305–6319.
- (9) Hamm, P.; Hochstrasser, R. M. *Ultrafast Infrared and Raman Spectroscopy*; CRC Press: Boca Raton, FL, 2001; p 273.
- (10) Ge, N. H.; Zanni, M. T.; Hochstrasser, R. M. *J. Phys. Chem. A* **2002**, *106*, 962–972.
- (11) Fang, C.; Wang, J.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B.; Decatur, S. M.; Hochstrasser, R. M. *Chem. Phys. Lett.* **2003**, *382*, 586–592.

- (12) Cabaniss, S. E.; McVey, I. F. *Spectrochim. Acta, Part A* **1995**, *51*, 2385–2395.
- (13) Cabaniss, S. E.; Leenheer, J. A.; McVey, I. F. *Spectrochim. Acta, Part A* **1998**, *54*, 449–458.
- (14) Humbert, B.; Alnot, M.; Quiles, F. *Spectrochim. Acta, Part A* **1998**, *54*, 465–476.
- (15) Binev, I. G.; Vassileva-Boyadjieva, P.; Binev, Y. I. *J. Mol. Struct.* **1998**, *447*, 235–246.
- (16) Kereszty, G.; Istvan, K.; Sundius, T. *J. Phys. Chem. A* **2005**, *109*, 7938–7945.
- (17) Max, J. J.; Chapados, C. *J. Phys. Chem. A* **2004**, *108*, 3324–3337.
- (18) Tang, Y. F.; Grey, M. J.; McKnight, J.; Palmer, A. G.; Raleigh, D. P. *J. Mol. Biol.* **2006**, *355*, 1066–1077.
- (19) Kim, Y. S.; Wang, J.; Hochstrasser, R. M. *J. Phys. Chem. B* **2005**, *109*, 7511–7521.
- (20) Wei, D. Q.; Guo, H.; Salahub, D. R. *Phys. Rev. E* **2001**, *64*, 016101.
- (21) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinotsaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.
- (22) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. J. Comput. Chem.* **1983**, *4*, 187.
- (23) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (24) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (25) Tu, K.; Tobias, D. J.; Klein, M. L. *Biophys. J.* **1995**, *69*, 2558–2562.
- (26) Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (27) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2004.
- (28) Brewer, S. H.; Song, B. B.; Raleigh, D. P.; Dyer, R. B. *Biochemistry* **2007**, *46*, 3279–3285.
- (29) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2003**, *35*, 369–430.
- (30) Ripoll, D. R.; Vila, J. A.; Scheraga, H. A. *J. Mol. Biol.* **2004**, *339*, 915–925.
- (31) Ham, S.; Kim, J. H.; Lee, H.; Cho, M. H. *J. Chem. Phys.* **2003**, *118*, 3491–3498.
- (32) Bour, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253–11262.
- (33) Jansen, T. L.; Knoester, J. *J. Chem. Phys.* **2006**, *124*.
- (34) Hayashi, T.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. A* **2005**, *109*, 9747–9759.
- (35) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. *J. Chem. Phys.* **2004**, *121*, 8887–8896.
- (36) Watson, T. M.; Hirst, J. D. *Mol. Phys.* **2005**, *103*, 1531–1546.
- (37) Bloem, R.; Dijkstra, A. G.; Jansen, T. L. C.; Knoester, J. *J. Chem. Phys.* **2008**, *129*.
- (38) Paarmann, A.; Hayashi, T.; Mukamel, S.; Miller, R. J. D. *J. Chem. Phys.* **2008**, *128*.
- (39) Auer, B. M.; Skinner, J. L. *J. Chem. Phys.* **2008**, *128*.
- (40) Wang, J. P.; Zhuang, W.; Mukamel, S.; Hochstrasser, R. *J. Phys. Chem. B* **2008**, *112*, 5930–5937.
- (41) Kwac, K.; Cho, M. H. *J. Chem. Phys.* **2003**, *119*, 2256–2263.
- (42) Kwac, K.; Cho, M. H. *J. Chem. Phys.* **2003**, *119*, 2247–2255.
- (43) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81–86.
- (44) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *12*, 6123–6138.
- (45) Falvo, C.; Palmieri, B.; Mukamel, S. *J. Chem. Phys.* **2009**, *130*, 184501.
- (46) Jansen, T. L.; Knoester, J. *J. Phys. Chem. B* **2006**, *110*, 22910–22916.
- (47) Mukamel, S.; Abramavicius, D. *Chem. Rev.* **2004**, *104*, 2073–2098.
- (48) Abramavicius, D.; Palmieri, B.; Yang, L.; Voronine, D.; Mukamel, S. *Chem. Rev.* **2008**, in press.
- (49) Falvo, C.; Pouthier, V. *J. Chem. Phys.* **2005**, *123*.

JP900245S