Visualization and Characterization of the Infrared Active Amide I Vibrations of Proteins

Hoi Sung Chung and Andrei Tokmakoff*

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received: July 18, 2005; In Final Form: October 17, 2005

To facilitate the analysis of frequency−structure correlations in the amide I vibrational spectroscopy of proteins, we investigate visualization methods and spatial correlation functions that describe delocalized vibrations of proteins and protein secondary structures. To study those vibrational modes revealed in infrared spectroscopy, we characterize frequency-dependent bright states obtained from doorway mode analysis. Our visualization methods pictorially color code amplitude and phase of each oscillator within the structure to reveal spatially varying patterns characteristic of excitations within sheets and helices. Spatial correlation functions in the amplitude and phase of amide I oscillators quantitatively address the extent of delocalization and the α helical and β sheet character of these modes. Specifically, we investigate the vibrations of idealized antiparallel β sheets and α helices and perform case studies on three proteins: concanavalin A, myoglobin, and ubiquitin.

I. Introduction

Amide I vibrational spectroscopy of proteins has long been used to describe secondary structure content and is increasingly targeted for investigations into the dynamics of protein structures. Of the many possible vibrations, the amide I mode (1600−1700 cm^{-1}) is of particular interest since it is intrinsic to all proteins, and additionally is localized on the peptide backbone and not strongly influenced by side chains.[1] The amide I vibration of a single peptide unit is primarily carbonyl stretching, but its vibrational frequency in a protein is sensitive to the type and amount of secondary structures. Strong electrostatic interactions between the many neighboring amide I oscillators of the protein lead to delocalized (excitonic) states, whose absorption frequencies reflect the underlying structural arrangement of oscillators. Well-established empirical frequency−structure correlations find that the β sheet contributes both at low (1630−1640 cm^{-1}) and high frequencies (above 1680 cm^{-1}), and the α helix and random coil structure are located at 1650−1660 cm^{-1} and 1640−1650 cm^{-1}, respectively.[2,3] However, the broad amide I line width arising from structural heterogeneity of the protein hides much of the underlying structure needed for peak assignment. To overcome this problem in infrared spectra, several numerical treatments have been developed such as Fourier self-deconvolution,[3] second derivative technique,[4] and factor analysis.[5,6] More recently, two-dimensional infrared spectroscopy of peptides has offered a new way of dissecting congested amide I line shapes and quantitatively describing the structural origins of the vibrational spectroscopy.[7−11] The 2D IR spectroscopy of proteins is now being used to characterize structure and molecular dynamics of proteins through the delocalized amide I states.[12−14]

The quantitative description of FTIR and 2D IR spectra requires a molecular description of the origin of spectral features. A key advance in the description of amide I spectroscopy came from Torii and Tasumi,[15] who calculated the amide I spectrum of several proteins by diagonalizing a Hamiltonian constructed in the basis of each peptide unit with couplings calculated from transition dipole interactions between sites. More recent improvements used ab initio calculations to get a more realistic coupling model between the adjacent residues.[16−18] This relatively simple model, the local amide I Hamiltonian, has had considerable success reproducing the band shape of amide I spectrum.

The eigenstates from such a calculation form the basis for trying to understand the correlations between amide I frequency and protein structure characterized by IR spectra. In particular, they can be used to answer a number of questions. What are the vibrations that we observe in an experiment at a given frequency? To what degree are these states delocalized? To what degree do they give us information on secondary and tertiary structure? Are the vibrations of secondary structures a good basis set for decomposing protein IR spectra?

Previous studies have addressed some of these questions. The characterization of the phase of various amide I vibrations has been started for the ideal model of the antiparallel β sheet and the α helix with perfect symmetries.[19−25] The role of energetic disorder in mode symmetry and localization has recently been addressed.[26] Analysis of several proteins indicates that the lowest frequencies observed in the IR spectrum of β proteins are localized to the sheet, and that resonance frequencies for α helical excitations are size dependent.[27] The visualization of the phases and amplitudes of several vibrational eigenstates having large transition dipole moments for two proteins, myoglobin and flavodoxin, has been reported,[28] indicating that vibrations of different eigenstates are localized either in the α helix or β sheet. Even with the considerable progress, more quantitative analysis methods are needed to interpret the structural origin of vibrational resonance frequencies within the amide I spectrum, particularly in the presence of structural heterogeneity found in proteins. The large number of eigenstates complicates the characterization of spectra, since analysis of a small number of eigenstates with large transition dipole intensity may not be representative of the whole. Also, sequence dependent characterizations of normal modes or eigenstates are not particularly

* Corresponding author. E-mail: tokmakof@mit.edu. Fax: 617-253-7030.

10.1021/jp053956a CCC: $33.50 © 2006 American Chemical Society
Published on Web 01/13/2006

helpful when structural correlations must be characterized in three spatial dimensions.

In this paper, we present methods that can be used to address frequency−structure correlations in infrared spectra of proteins. Drawing on the eigenstates of the local amide I Hamiltonian, the structural origin of protein amide I infrared spectroscopy can be characterized using visualization methods that identify the amplitude and phase relationship of amide I oscillators in one, two, and three dimensions. Vibrational amplitudes and phases are obtained from doorway modes[27,29] rather than eigenstate vibrations to find the average nature of the vibrations in a particular frequency region. Doorway modes represent a transformation of a subset of eigenstates into a bright state characteristic of a part of the IR spectrum. The characterization is performed for the six split frequency ranges of the amide I band to find frequency-dependent vibrational characters. For the visualization, the vibrational amplitudes and phases or individual peptide oscillators are color-coded directly onto the crystal structure and then projected onto a two-dimensional grid for β sheets.[14,26] These methods can then be quantitatively analyzed with spatial correlation functions for the oscillators within delocalized vibrations, as previously demonstrated in electronic systems.[30] One and two-dimensional analyses are performed for the α helix and β sheet, respectively. These correlation analyses effectively present the phase relationship between adjacent vibrations through bond and through space. To quantify the specific vibrational mode character associated with the idealized vibrations of β sheets and α helices, a mode character f is developed based on a phase-associated correlation length. As a reference, the characterizations are first performed for the ideal α helix and β sheet. Then, detailed case studies for real systems are presented for three proteins: concanavalin A, an all β sheet containing protein; ubiquitin, a mixed α/β protein; and myoglobin, an all α protein.

II. Methods

A. Calculation of Amide I Eigenstates and IR Spectrum. To characterize frequency−structure relationships in the amide I spectrum, we calculate the IR spectrum using a sum over amide I eigenstates and then analyze frequency-dependent eigenstate characteristics using doorway mode analysis for a subset of eigenstates within a given frequency range. The amide I eigenstates and energy eigenvalues are obtained by diagonalizing a local amide I Hamiltonian (LAH) **H**,[14] an approach pioneered by Krimm[1] and Torii[15] and more recently extended by Hamm and Hochstrasser.[7,31] The LAH is constructed in the site basis of the amide I vibrations for the individual peptide units, drawing on the coordinates from a crystal structure. The energies of the individual sites (diagonal terms) are assigned based on hydrogen bonding criteria, and the vibrational couplings (off-diagonal terms) include through-space and through-bond effects. IR spectra are calculated by a sum over amide I vibrational eigenstates weighted by the transition dipole matrix elements squared and convoluted with a Lorentzian line shape

$$A(\omega) = \mathrm{Im}\sum_q |\langle q|\mathbf{m}|0\rangle|^2/(\omega - \omega_{q0} - i\Gamma) \qquad (1)$$

To account for natural small variations in structure or solvation, site energies are averaged over a distribution.

The diagonal elements of the LAH, H_{ii}, consist of site frequencies ω_i for the individual amide I vibrations of the peptide units. The site frequencies are red shifted $\omega_i = \omega_0 + \delta\omega$ from an unperturbed frequency ($\omega_0 = 1688$ cm^{-1}) to reflect hydrogen bonding interactions with other residues or solvent molecules. Three criteria are used to determine the changes of site frequency: (*i*) hydrogen bonding to the carbonyl oxygen by hydrogen of the NH group of another residue, (*ii*) hydrogen bonding to the carbonyl oxygen by hydrogen of surrounding water molecules, and (*iii*) hydrogen bonding to the NH group by another residue or solvent molecules. The empirical formula used to calculate the effect of (*i*) is $\delta\omega = a(2.6 - r_{OH})$, where r_{OH} is the O···H hydrogen bond distance in Ångstroms and $a = -30$ cm^{-1}.[7] The formation of hydrogen bond is assumed when $r_{OH} < 2.6$ Å and the angle formed by O···H and H−N is less than 35°. Because of the insufficient number of water molecules in the crystal structure, the uniform shifts of site frequency of -20 cm^{-1} and -10 cm^{-1} are used for (*ii*) and (*iii*), respectively.[32]

Off diagonal elements, except those of the nearest neighbor pairs, are constructed using the empirically parametrized transition dipole coupling model:[1,15]

$$H_{ij} = \frac{A}{r_{ij}^3}(\hat{\mu}_i\cdot\hat{\mu}_j - 3(\hat{\mu}_i\cdot\hat{\mathrm{r}}_{ij})(\hat{\mu}_j\cdot\hat{\mathrm{r}}_{ij})) \qquad (2)$$

Following Krimm, we take $A = 580$ cm^{-1} Å^3, $\hat{\mu}_i$ is the unit vector of the transition dipole moment associated with the amide I vibrational coordinate of the ith peptide site, $\hat{r}_{ij}$ is unit vector from site i to j, and r_{ij} is the separation. The transition dipole of each peptide group is assumed to be located on the C═O bond axis, 0.868 Å from the carbon atom to the oxygen with an orientation of 20° off from the C═O bond axis toward the nitrogen.[15,21] For interactions between the bonded pairs of adjacent peptide units ($i = j \pm 1$), amide I couplings determined from ab initio calculations on the diglycine molecule are used.[17]

With these elements of the model, the amide I eigenstates are obtained by diagonalizing the LAH

$$\mathbf{H} = \mathbf{TET}^T \qquad (3)$$

where **T** is the transformation matrix formed by the eigenvectors. The x, y, and z components of the transition dipoles connecting eigenstates can be obtained from the $N \times 3$ transition dipole matrix **d** representing the x, y, and z components of transition dipoles for the N peptide sites. $\mathbf{d}^E = \langle q|\mathbf{m}|0\rangle$ in eq 1 is obtained from $\mathbf{T}_q^T\cdot\mathbf{d}$ where $\mathbf{T}_q$ is the qth column of **T**. IR spectra are calculated from **H** using eq 1 with a Lorentzian line width of $\Gamma = 5$ cm^{-1}. To account for broadening of the spectrum as a result of disorder of the local mode frequency, 10 000 spectra are averaged in which the site energies are chosen from a Gaussian random distribution with width $\sigma = 6$ cm^{-1} centered at the nominal site energy ω_i.

B. Doorway State Calculation. Due to the high density of states, the LAH eigenstates are of limited use for describing and visualizing the vibrations contributing to different frequency ranges within the amide I infrared spectrum. Given a frequency range, one cannot generally select a representative vibrational eigenstate by its strength in the IR spectrum in the presence of several equivalent transition dipole moments. This raises questions about how to average the contribution of several states without overestimating the vibrational character of states that do not strongly contribute to the IR spectrum. We have chosen to apply the doorway state analysis used previously by Torii and Tasumi.[27,29] We obtain a set of three representative bright states for a specific subset of eigenvectors grouped by frequency through an orthogonal transformation. For analysis of structure−frequency correlations, we divide the amide I spectrum into six frequency ranges. The region between 1640 and 1680 cm^{-1} is

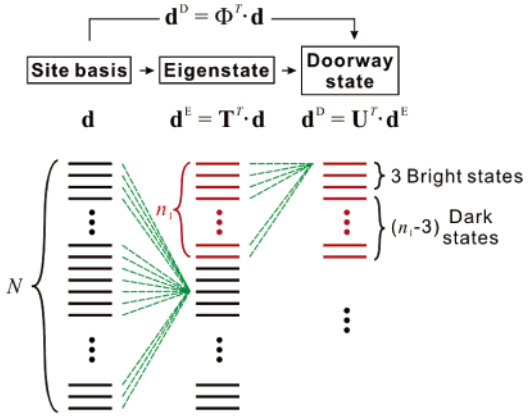

Figure 1. Transformation from site basis to doorway states. Diagonalization of the local amide I Hamiltonian (LAH) of N amide I oscillators in the site basis gives N eigenstates. The transform matrix $\mathbf{T}$ consists of the eigenvectors. Each eigenstate is the linear combination of the states in site basis as shown with green dashed lines. Doorway transformation of the subset of n_1 eigenstates (red) in the first frequency region R_1 gives 3 bright states and $n_1 - 3$ dark states. The transformations for the following frequency regions also give three bright states for each frequency region. Note that all the doorway states in the same frequency region are not associated with a definite energy. The orthogonal set of bright states is obtained by SVD and the first doorway mode is used for the visualization and correlation analyses.

divided into 10 cm^{-1} intervals. Additionally, the modes below 1640 cm^{-1} and above 1680 cm^{-1} form two groups. All the eigenstates within each frequency region constitute the vibrational subset for doorway analysis. The frequency range could be divided in another manner guided by the density of eigenstates and the desired frequency resolution, and retaining a minimum of three eigenstates in each region. As the number of eigenstates per interval decreases, the character of the doorway mode approaches that of the eigenstates. If the region is too wide, various vibrational mode characters are overlapped and we lose the information on the frequency−structure relationship. We found the appropriate frequency splitting is 5−10 cm^{-1}. In this paper, we show results of 10 cm^{-1} splitting, but the analysis with 5 cm^{-1} splitting gives similar results.

The relationship between the site basis, eigenstates, and doorway states is depicted in Figure 1. In the transformation for n_i eigenstates within a particular frequency range R_i, three bright states (nonzero transition dipole moment) and $n_i - 3$ dark states (zero transition dipole moment) that are orthogonal to the bright states are expressed as a linear combination of eigenstates. (The number of the bright states is determined by the dimensionality of the dipole moment.) The first, second, and third doorway modes are labeled by decreasing transition dipole strength. Although any linear combination of the three bright states is orthogonal to the dark states, choosing one of the orthogonal bright states eliminates interference between them.[29]

We use singular value decomposition (SVD) to find the transformation $\mathbf{U}_i$ that generates the three orthogonal bright states from the eignestates in the ith frequency region R_i. The $n_i \times 3$ eigenstate dipole matrix, $\mathbf{d}^E{}_i$ is formed as

$$\mathbf{d}^E{}_i = \mathbf{T}_i{}^T \mathbf{d} \tag{4}$$

$\mathbf{T}_i$ is the subset of the transformation matrix $\mathbf{T}$ (eq 3) for the eigenstates within R_i. The doorway dipole matrix, $\mathbf{d}^D{}_i$ can be obtained by the orthonormal transformation $\mathbf{U}_i$

$$\mathbf{d}^D{}_i = \mathbf{U}_i{}^T \mathbf{d}^E{}_i \tag{5}$$

SVD decomposes $\mathbf{d}^E{}_i$ into three matrices,

$$\mathbf{d}^E{}_i = \mathbf{U}_i' \mathbf{S}_i \mathbf{V}_i{}^T \tag{6}$$

$\mathbf{U}_i'$ and $\mathbf{V}_i$ are $n \times n$ and 3×3 orthonormal matrices, respectively. All elements of $\mathbf{S}_i$ ($n \times 3$) are zero except the three diagonal singular values (s_{11}, s_{22}, s_{33}). Since the transformation to the three orthogonal bright states is unique, eqs 5 and 6 are equivalent. Therefore, $\mathbf{U}_i' = \mathbf{U}_i$ and $\mathbf{d}^D{}_i = \mathbf{S}_i \mathbf{V}_i{}^T$ and the first three columns of $\mathbf{U}_i$ are the transformation vectors from eigenstates to the three orthogonal doorway modes. The three-column vectors of $\mathbf{V}_i$ consist of unit vectors of the three representative transition dipoles, and three nonzero singular values of $\mathbf{S}_i$ are their transition dipole moments.

The transform matrix Φ_i from the site basis to the doorway states in the frequency region R_i is given by rearranging eqs 4 and 5 as

$$\Phi_i = \mathbf{T}_i \mathbf{U}_i \tag{7}$$

Among the three doorway modes, the first mode is chosen to represent the frequency region R because it has the most intense dipole strength, although it does not totally dominate the other two modes. In most cases, the dipole moment square of the first doorway mode exceeds 50% of the total. The elements of ϕ_1, the first column of Φ_i, give the amplitudes and phases of local amide I vibrations that constitute the first doorway mode. ϕ_1 is used for the analyses on the frequency-dependent vibrational characteristics: the visualization of vibrational phases and calculation of the correlation functions, correlation lengths, and secondary structure contents. All calculations are performed averaging over 10 000 disorder realizations ($\sigma = 6$ cm^{-1}). Visualization of the vibrational amplitude and phase is presented for the crystal structure without disorder. The assignments of a residue to a particular secondary structure are taken from the data in the PDB file. Since the square of each element of ϕ_1 is the fraction of each local mode vibration, the secondary structure contents are calculated by the summation of the squares of ϕ_1 over the secondary structure indices.

C. Visualization of Amide I Vibrations. The amide I doorway modes are visualized using a red-to-blue color map that encodes the amplitude and phase of each of the amide I sites. Red is positive, blue is negative, and the amplitude scales with the depth of the color. The direct color-coding of the amplitudes on the protein structure is performed by color change of the amino acid residues using the software Pymol (DeLano Scientific LLC). This scheme is effective to show the behavior of vibrational correlations under structural variances in the real proteins. It also provides a unique visualization method for α helix that is not easily projected onto two dimensions. We colored the amino acid residues, for convenience, instead of the actual oscillators, which lie in the middle of the two residues forming peptide groups. In other words, the amplitude of the amide oscillator linking residues n and $n + 1$ is recorded on the residue n as shown in Figure 2a. Although this results in the shift of the location of the oscillator by a half residue, the color-coding is straightforward for α helix and parallel β sheet because every oscillator is shifted to the same direction in case of unidirectional growth. This shift induces a problem in the antiparallel β sheet. As seen in Figure 2a, the shift of the oscillator 1 and 2 to the left and the shift of the residue 3 and 4 to the right results in the mismatch by one residue. To avoid this mismatch, the oscillators of the every other strand in the antiparallel β sheet are shifted, i.e., if the amplitudes of the

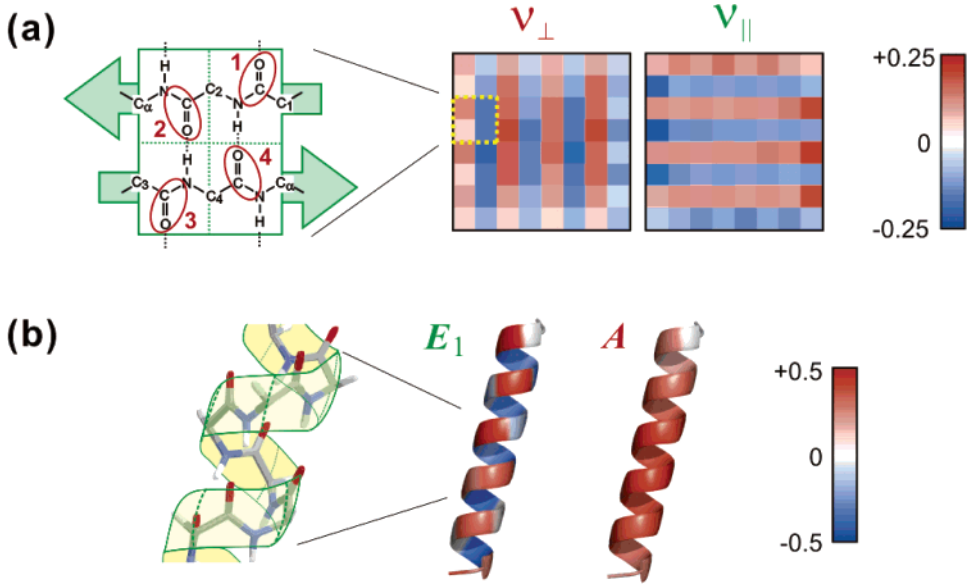

Figure 2. (a) Projection of the antiparallel β sheet on a plane. (left) Four peptide groups in the unit cell are shown in two antiparallel β strands. Peptide units are numbered with red figures and amino acids are numbered in C_α. (right) Projection of the amplitudes and phases oscillators in the $\nu_\perp$ and $\nu_{||}$ vibrations of the ideal β sheet formed by 8 × 8 oscillators. (b) Amplitudes and phases of oscillators in the A and E_1 vibrations of the ideal 21-residue α helix.

oscillator 1 and 2 are colored on the residue 1 and 2, the amplitudes of the oscillator 3 and 4 are colored on the residue 4 and 5.

III. Result and Discussion

A. Characterization of the Vibrations of Secondary Structure. As a starting point for the interpretation of protein amide I vibrations, we characterize the doorway modes of idealized secondary structures. Here we consider the antiparallel β sheet and α helix. The empirical frequency-structure relationships indicate that antiparallel β sheets are characterized by a two-peak structure in the infrared absorption spectrum at 1630−1640 cm^{-1} and >1680 cm^{-1}, and that α helix and random coil structure are located at ∼1650−1660 cm^{-1} and 1640−1650 cm^{-1}, respectively.[2,3]

The split two peak spectrum of the antiparallel β sheet has been explained by the long-range transition dipole coupling and the symmetry of the four oscillator unit cell for an idealized sheet shown in Figure 2a.[1,20,21] These studies indicate that there are four characteristic vibrations for the antiparallel β sheet that differ in the phase relationships between adjacent oscillators, of which two are observed in the IR spectrum. The lowest frequency transition carries most of the oscillator strength because it involves in-phase oscillation of the aligned amide I transition dipoles on adjacent strands but alternating phase between bonded amide units. Vibrations of this symmetry will be designated $\nu_\perp$ since the in phase oscillations and the transition moments lie perpendicular to the strands of the β sheet. The other high-frequency mode, designated $\nu_{||}$, is weaker because the vibrations of adjacent oscillators are in-phase for bonded units and alternating for oscillators between strands. We use these notations instead of the $a+$ and $a-$ notation used for idealized structures[21,26] because this symmetry is broken for the real β sheet structure and these assignments also apply to parallel β sheets and the mixed structure of parallel and antiparallel sheets such as ubiquitin. The vibrational frequency of the $\nu_\perp$ mode is sensitive to the number of the strands because the vertically in-phase relationship of the amide I vibration is energetically more favorable in the addition of strands while the $\nu_{||}$ mode is insensitive if at least two strands are present.[21] Therefore, the splitting between the two modes is larger and the $\nu_\perp$ transition becomes stronger in the bigger β sheet, and thus the IR spectrum of the β sheet containing protein becomes more asymmetric.

The symmetry of the $\nu_\perp$ and $\nu_{||}$ modes can be visualized by projecting the β-sheet structure onto a two-dimensional grid that represents the spatial placement of nearest neighbors to each amide I oscillator.[14,26] The horizontal dimension represents the direction of a given strand within the sheet. Each box is color coded to indicate the amplitude and phase of the oscillator in a particular mode. Figure 2a shows a visualization of totally delocalized amide I "doorway states" with $\nu_\perp$ and $\nu_{||}$ character for an idealized antiparallel sheet of 8 × 8 oscillators with degenerate site energies. The vibrational amplitude and phase relationships between oscillators of the two modes are shown with varying depth of red (positive) and blue (negative) colors. These diagrams emphasize that the character of the two modes expected for the sheet differ by either columns or rows retaining constant phase and the other dimension exhibiting alternating phase.

A similar visualization strategy can be adopted for helices. For the amide I vibrations of an ideal α helix, A and E_1 modes are expected.[11,19,20,22,25,33] The vibrational phases and amplitudes of these two modes for ideal 21-residue α helices are shown on the structure in Figure 2b by projecting the amplitude and phase of the amide I vibrations onto a ribbon diagram of the helix. All of the amide I oscillators are in phase for the A mode. For the E_1 mode, the phase relationship between adjacent oscillators in the primary sequence is such that oscillators every 3.6 residues are in phase, reflecting the fact that there are 18 residues for 5 turns in the unit cell of the helix.

B. Spatial Correlation Functions. To characterize the frequency-structure relationships for the doorway modes of secondary structures and proteins as distinguished by the phase relationship between oscillators within the structure, we consider spatial correlation functions that characterize the amplitude and phase relationships between the amide I oscillators and the degree of delocalization. A one-dimensional correlation function

for the vibrational amplitude A between the amide I sites as a function of their separation (n) is defined[30]

$$C(n) = \langle \sum_{j=1}^{N-n} A_j A_{j+n} \rangle / \langle \sum_{j=1}^{N} A_j^2 \rangle \tag{8}$$

Here N is the number of oscillators, n takes on values from 0 to $N-1$, and A_j is the ϕ_1 element of the jth oscillator. The brackets represent the averaging performed over disorder in the site energies. Additionally, we can characterize the phase relationship with displacement between oscillators through a discrete Fourier transform

$$\tilde{C}(k) = \sum_{n=0}^{N-1} e^{2\pi i k n} C(n) \tag{9}$$

where the wavevector $k = 0$ to 1/2. Since the correlation function is the sum of the cosine functions that peak at $n = 0$, only the real part of the Fourier transform is meaningful.

In the case of an α helix, the index j in eq 8 is chosen to be the oscillator index along the peptide backbone. As pictured in Figure 3a, the amide I A mode will be characterized by constant phase for all oscillators, and its correlation function $C_A(n)$ decays monotonically with oscillator separation for a finite helix. The correlation function for the E_1 mode $C_{E_1}(n)$ oscillates at a periodicity given by the pitch of the helix. These modes are clearly distinguished in the Fourier transform $\tilde{C}(k)$, where $\tilde{C}_A$ and $\tilde{C}_{E_1}$ are seen to peak sharply at $k_A = 0$ and $k_{E_1} = 5/18$, respectively.

For the characterization of the delocalization of a particular mode, a correlation length can be defined as

$$L = \frac{2}{N+1} \sum_{n=0}^{N-1} C(n) = \frac{2}{N+1} \tilde{C}(0) \tag{10}$$

The factor $2/(N+1)$ normalizes to the number of oscillators and dictates that the value of L varies from 1 for totally delocalized states to 0 for large chains with uncorrelated vibrational excitations. Generally, this expression gives an accurate estimate only for oscillators that are entirely in-phase, as for the A mode. In the case of a periodically varying phase, such as the E_1 mode, the correlation length is more properly evaluated from the amplitude of the correlation function at the mode wavevector k_{E_1}

$$L_{E_1} \simeq \frac{2}{N+1} \mathrm{Re}[\sum_{n=0}^{N-1} 2e^{-2\pi i k_{E_1} n} C(n)] = \frac{4}{N+1} \mathrm{Re}[\tilde{C}(k_{E_1})] \tag{11}$$

Since we use a symmetric correlation function $C(n) = C(-n)$, the normalization factor is twice compared to eq 10 to account for the symmetric peak amplitude at k_{E_1} and $-k_{E_1}$ in $\tilde{C}(k)$.

For the correlation function of the finite idealized α helix of 21 residues (Figure 2b), we find $L = 0.95$ for the A mode and $L(k = 1/3.6) = 0.98$ for the E_1 mode, both of which are close to the maximum correlation length. These correlation lengths also indicate that the degree of α helix mode character for the A and E_1 modes can be defined as

$$f_A = \bar{L}(k_A)$$
$$f_{E_1} = \bar{L}(k_{E_1}) \tag{12}$$

Here, $\bar{L}$ indicates an average over several helices in one protein.

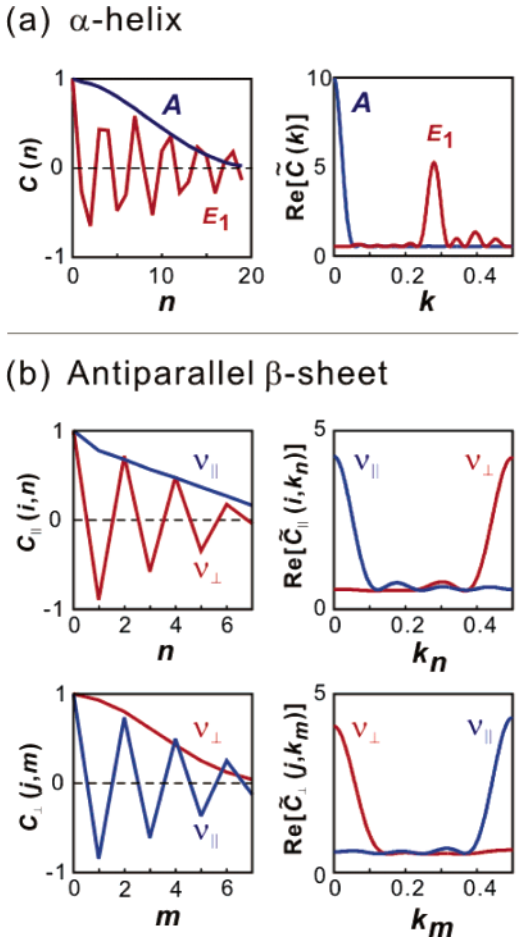

Figure 3. Amide I oscillator correlation functions (left) and their Fourier transforms (right). (a) The correlation function of the A mode of the ideal α helix of Figure 2b decreases smoothly while that of E_1 mode oscillates with the period of 3.6 residues as shown in the Fourier transform. A and E_1 modes are found in the doorway vibrations of the frequency ranges of 1645 and 1655 cm^{-1}, respectively. (b) Correlation functions for the ideal 8 × 8 β sheet (Figure 2a). Row correlations (top) and column correlations (bottom) of the fourth row and column of show the in-phase and out of phase natures of the two vibrations ($\nu_\perp$ and $\nu_\parallel$) between the adjacent residues and strands. $\nu_\perp$ and $\nu_\parallel$ modes are the doorway vibrations of <1640 cm^{-1} and >1680 cm^{-1}, respectively.

In the case of β sheets, a correlation function in one dimension is inadequate to characterize the symmetry of the vibration, since $\nu_\perp$ and $\nu_\parallel$ character is distinguished by the phase relationship in two dimensions. Drawing on the grid representation of the β sheet doorway modes shown in Figure 2a, we can investigate the spatial correlation function as a function of separation along rows (intrastrand) and columns (interstrand). If the amplitude of an oscillator in a particular row and column is $A_{i,j}$, then the correlation function for a row i parallel to the β strands is

$$C_\parallel(i,n) = \langle \sum_{j=1}^{N_i-n} A_{i,j} A_{i,j+n} \rangle / \langle \sum_{j=1}^{N_i} A_{i,j}^2 \rangle \tag{13}$$

where N_i is the number of oscillators in the row and n is the separation. A similar correlation function $C_\perp(j,m)$ can be defined for a column j perpendicular to the strands. Figure 3b shows the correlation functions for the $\nu_\perp$ and $\nu_\parallel$ modes of a large sheet of 64 amino acids (8 × 8 oscillators). Their Fourier transforms are peaked at $k_n = 0.5$ and $k_m = 0$ for $\nu_\perp$ and $k_n = 0$ and $k_m = 0.5$ for $\nu_\parallel$.

In analogy to eq 10, we also define the correlation length L_i for a given row i, and the correlation length averaged over M

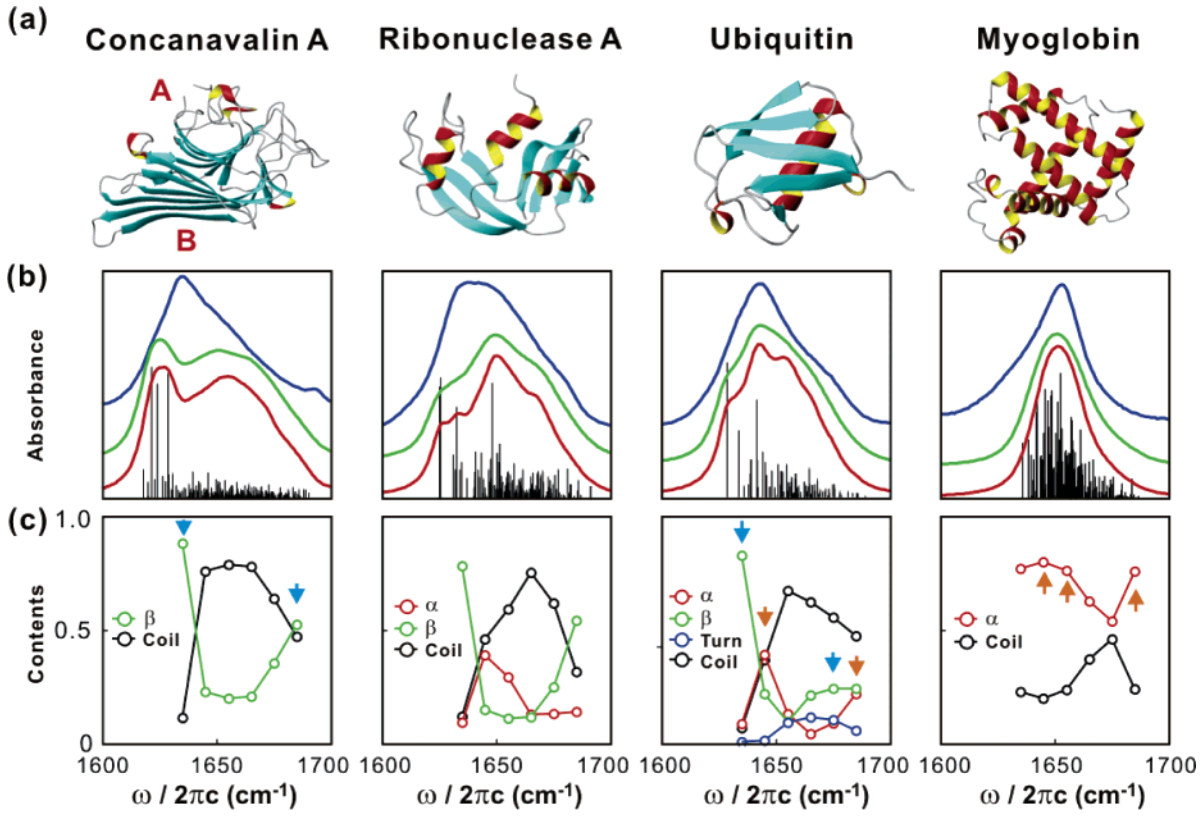

Figure 4. (a) Ribbon diagrams of the structures for the proteins investigated. (b) Experimental (blue) and calculated (red, green) infrared spectra of proteins. Averaged spectra with diagonal disorder calculations (green) show smoother features compared to those calculated without disorder (red). Black bars are the square of the dipole strength of amide I eigenstates. (c) Secondary structure content to the doorway modes. β sheet has high contents in the red and blue side of the spectrum. The turn and coil regions lie in the middle and α helix show high probability in the red side between random coil region and β sheet region. Blue and orange arrows indicate the doorway modes where the vibrational phases of β sheets and α helices are analyzed.

rows

$$\bar{L}_{\parallel}(k) = \frac{1}{M}\sum_{i=1}^{M} L_i(k) \tag{14}$$

Here the correlation length $L_i(k) = 2\tilde{C}(k)/(N+1)$ is evaluated at the appropriate mode wavevector. A similar definition for $\bar{L}_{\perp}$ can be obtained from the average over the column correlation lengths of N columns.

As a working definition, the fractional $\nu_{\perp}$ and $\nu_{\parallel}$ character can be obtained by multiplying the average row and column correlation lengths with mode wavevectors of $k = 0$ and 0.5 to account for the delocalization and phase information along both the row and column direction:

$$f_{\perp} = \bar{L}_{\parallel}(k=0.5)\cdot\bar{L}_{\perp}(k=0)$$
$$f_{\parallel} = \bar{L}_{\parallel}(k=0)\cdot\bar{L}_{\perp}(k=0.5) \tag{15}$$

Finally, while our analysis was performed for antiparallel β sheets, the assignment of modes and spatial correlation functions for $\nu_{\perp}$ and $\nu_{\parallel}$ character apply equally well to the analysis of IR spectra of parallel β sheets.

C. IR Spectra. Infrared absorption spectra were calculated for four proteins with varying secondary structural content, depicted in Figure 4a. The LAH was constructed drawing on structures from the Protein Data Bank. Concanavalin A (Con A, PDB id 1JBC[34]) is a β protein with two large extended β sheets. Ribonuclease A (RNase A, 7RSA[35]) and ubiquitin (1UBQ[36]) are α/β proteins. RNase A has two antiparallel β sheets, whereas ubiquitin has one β sheet with mixed parallel/antiparallel hydrogen-bonding registry. Myoglobin (1WLA[37]) is an α protein.

Figure 4b shows calculated IR spectra for the four proteins. Black bars at the bottom of the each spectrum are dipole moment squares of eigenstates calculated from the crystal structures. Each eigenstate is broadened homogeneously and added to construct the IR spectrum (red curves). The green curves are spectra averaged over 10 000 disorder realizations. Although the calculated spectra do not show quantitative agreement with the experimental spectra (blue curves), the relatively simple models used in this paper reproduce many of the trends in peak position, line width, and line shape successfully. Perhaps the biggest deviation is the low-frequency region of Con A, for which a distinct resonance in the calculation is not observed in the experiment. This results from an overestimation of short-range TDC couplings in β sheets. Changes to the LAH model that would improve the comparison are possible, and the parametrization of site frequencies, couplings, and influence of structural or dynamic disorder is presently a topic of considerable research.[18,38−40]

For Con A, because of the large content and size of β sheets, the spectrum is asymmetric and the splitting between $\nu_{\perp}$ and $\nu_{\parallel}$ modes is large as explained in III A. Several eigenstates with very large dipole moments in the red side have the $\nu_{\perp}$ vibration character. For the mixed α/β proteins, RNase A and ubiquitin spectra show similar asymmetric shape with less splitting between the two modes, and the more dipole strength in the middle of the amide I band originated from α helix, turn, and random coil structures. The small splittings between $\nu_{\perp}$ and $\nu_{\parallel}$ modes arise from the small size of the β sheet. The myoglobin

spectrum shows a more symmetric one-peak structure with a much narrower width.

The secondary structure contents for the doorway modes are shown in Figure 4c. These calculations confirm many of the empirical relations between amide I frequency and secondary structure content. The β sheet content is particularly high to the red of the absorption (<1640 cm^{-1}, $\nu_\perp$). It also peaks to the blue side (>1680 cm^{-1}, $\nu_\parallel$), although various structural elements also contribute. The intermediate frequency region is dominated by turn and random coil structures. The α helix region is known to be overlapped with the random coil region; here we find that it is inclined to the red side ($1640-1650$ cm^{-1}) of the amide I region. The relationships between secondary structure, amide I absorption frequency, and the character of doorway modes is described in detail for three proteins below.

D. β-Protein: Concanavalin A. The visualization and analysis of amide I doorway modes for the β-protein Con A reveals many similarities to the idealized β sheet. Based on the structure assignments, we expect that doorway modes for the lowest (<1640 cm^{-1}) and highest (>1680 cm^{-1}) frequency range, marked with the blue arrows in Figure 4c, correspond to amide I modes with $\nu_\perp$ and $\nu_\parallel$ symmetry. Figure 5a shows a visualization of these doorway modes on the crystal structure and the two-dimensional projection of the amide I oscillators within the larger β-sheet B. We note the constant phase relationship perpendicular and parallel to the strands of both sheets. We here focus on the analysis of oscillators within sheet B, but the results are similar for sheet A. These visualizations confirm the assignment of the two modes. For the $\nu_\perp$ mode, the projections of the vibrations show in-phase correlation across the strands (column direction) and alternating phase relation along the strands (row direction) clearly. For the $\nu_\parallel$ mode, the in-phase and alternating phase relationships hold for the row and column directions, respectively. The $\nu_\perp$ mode is clearly delocalized over the entirety of both sheets, whereas the amplitude of oscillators in the $\nu_\parallel$ mode appears patchier and concentrated on the first four strands of sheet B and much of sheet A.

Figure 5a also shows the oscillator correlation functions of sheet B for the $\nu_\perp$ and $\nu_\parallel$ modes and their Fourier transforms. For the $\nu_\perp$ mode, the row correlation function, $C_\parallel$ oscillates with every other residue spacing, whereas the column correlation function, $C_\perp$ decreases monotonically. For the $\nu_\parallel$ mode, the row correlation function oscillates and the column correlation function decreases monotonically. Therefore, $C_\perp$ and $C_\parallel$ of the $\nu_\parallel$ mode resemble $C_\parallel$ and $C_\perp$ of the $\nu_\perp$ mode, respectively. The overall effect is seen in the averages of the real part of the Fourier transform of rows ($\tilde{C}_\parallel$) and columns ($\tilde{C}_\perp$), for which peaks at $k = 0$ and 0.5 represent in-phase and oscillating features of vibrations, respectively.

From eq 14, we see that the magnitude of $\tilde{C}_\parallel$ and $\tilde{C}_\perp$ at $k = 0$ and 0.5 is proportional to the average correlation length along rows or columns. As can be seen in Figure 5a, the column and row correlation lengths of $\nu_\perp$ mode are longer than those of the $\nu_\parallel$ mode. The results for the ideal β sheet shown in Figure 3b give a similar correlation length both for the $\nu_\perp$ mode and for the $\nu_\parallel$ mode, implying that an irregular phase relationship presents itself more in the $\nu_\parallel$ mode than that of $\nu_\perp$ mode. The origin of the shorter correlation length for $\nu_\parallel$ is the preferential localization of this mode in the upper left part of the phase map of sheet B. The amplitude of the $\nu_\parallel$ mode is low in the lower right and the phase relationship is broken while the amplitude and the phase correlation is almost perfect in $\nu_\perp$ mode, which results in the fast decays of the envelope of $C_\perp$ of columns 10, 11, and 12 of the $\nu_\parallel$ mode. This selective disruption of the vibrational phase results mostly from the mixture of β-sheet oscillators with those in coiled regions in the $\nu_\parallel$ transition. This is not the case for the $\nu_\perp$ mode, which is located in the far red side of the spectrum and is localized to the β sheet as indicated by the high β content in Figure 4c. In addition, the structure of the lower right region is bent (see Figure 5a) and almost perpendicular to the center of the sheet. The modification of the couplings between the oscillators of this varied structure also disrupts the phase relationship.

The row and column correlation lengths provide measures of the $\nu_\perp$ and $\nu_\parallel$ character of the β-sheet vibration. Figure 5b shows these as a function of amide I frequency for β sheet B. As expected, the row and column correlation lengths reach maxima on the blue (>1680 cm^{-1}) and the red side (<1640 cm^{-1}) of the spectrum which are dominated by the $\nu_\parallel$ and $\nu_\perp$ mode, respectively. The irregularity of the phases in the lower right region of the sheet leads to short row correlation lengths for strands IV$-$VI.

Characterization of the mode character and degree of delocalization of β-sheet vibrations requires a phase-sensitive characterization in two dimensions. To account for the constant phase in one dimension and alternating phase in the other dimension for $\nu_\perp$ and $\nu_\parallel$ modes, we draw on the average correlation length $\bar{L}$ parallel and perpendicular to the strands evaluated at the wavevectors $k = 0$ and $k = 0.5$. These are shown as a function of amide I doorway mode frequency in Figure 5c. As expected, $\bar{L}_\perp$ ($k = 0$) and $\bar{L}_\parallel$ ($k = 0.5$) decrease and $\bar{L}_\parallel$ ($k = 0$) and $\bar{L}_\perp$ ($k = 0.5$) increase as the frequency becomes higher. The resemblance of $\bar{L}_\perp$ ($k = 0$) and $\bar{L}_\parallel$ ($k = 0$) with $\bar{L}_\parallel$ ($k = 0.5$) and $\bar{L}_\perp$ ($k = 0.5$) indicates the similar degree of the correlation of the vibrations along both the row and the column direction. Equation 15 relates the product of $\bar{L}_\parallel$ and $\bar{L}_\perp$, evaluated at the appropriate wavevector, to the β-sheet character of the doorway mode f. $f_\perp$ and $f_\parallel$ are shown in Figure 5d. It is clear that the two β sheet vibrational characters are separated to the red and the blue side. On the other hand, the middle of the spectrum does not show any of these characters. The larger value of $f_\perp$ (<1640 cm^{-1}) than $f_\parallel$ (>1680 cm^{-1}) implies that $\nu_\perp$ mode is more preserved than $\nu_\parallel$ mode as indicated by other phase properties.

E. α-Protein: Myoglobin. The vibrations of an infinite idealized α helix will have A and E_1 symmetry.[19,20,33,41] Since all local vibrations are in-phase for the A mode, the transition dipole moment is parallel to the helical axis. The degenerate E_1 mode consists of linear combinations of vibrations with periodic phases of $2\pi n/m$, where m is the number of oscillators per turn (3.6 for α helix) and n is the index of the oscillator. For a helix of finite length, though, the degeneracy of the E_1 mode is broken and higher order of E type vibrations become IR active, the shape of the spectrum is determined by the most intense A mode.[33] Previously, a doorway mode analysis of all amide I eigenstates was performed for the α helices of myoglobin,[27] and the result shows that the first and second doorway modes had A and E_1 mode character, respectively.

For the α-helical protein myoglobin, we analyzed the six helices (I, II, V$-$VIII) with more than 12 residues to investigate the frequency$-$structure correlation for amide I doorway modes. The results are shown in Figure 6 for vibrations in the three frequency regions as marked with orange arrows in Figure 4c. For the 1645 cm^{-1} doorway mode visualized on the structure of Figure 6a, oscillators are in phase, although one can observe a slight periodic amplitude variation in the helix in almost every three residues in helix I, V, and VI. Similar periodicity can be

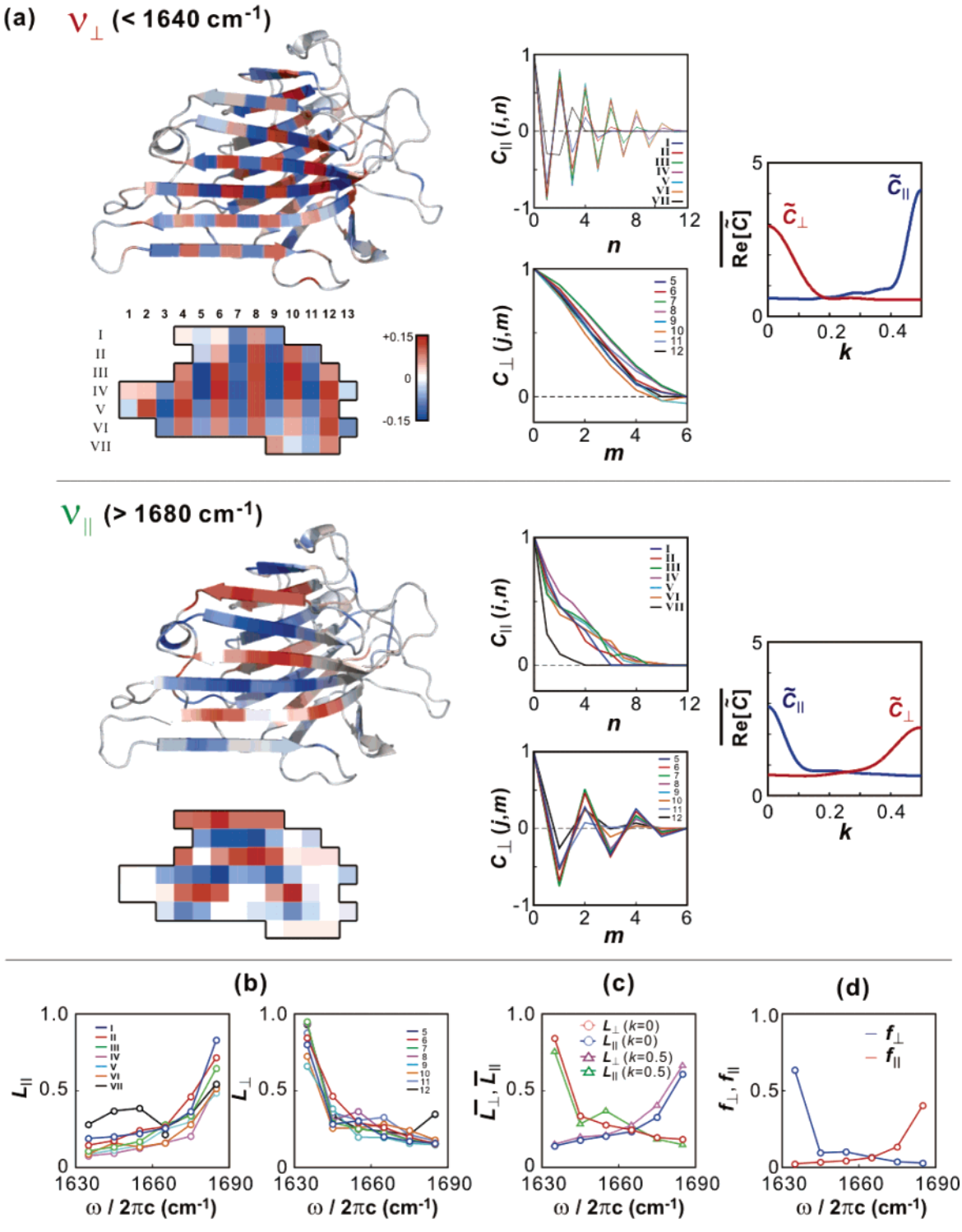

Figure 5. Visualization and characterization of the $\nu_\perp$ and $\nu_\parallel$ doorway modes for concanavalin A. (a) Vibrational amplitudes and phases of the $\nu_\perp$ and $\nu_\parallel$ modes visualized on the crystal structure. The contribution of sheet B is projected onto a two-dimensional grid representing the spatial placement of oscillators within the sheet. Row (m) and column (n) correlation functions, and the Fourier transform of the average are presented. (b) Row and column correlation lengths, (d) average correlation lengths, and (e) fractional $\nu_\perp$ and $\nu_\parallel$ mode characters are shown as a function of doorway mode frequency.

found in other helices. This periodic feature characteristic of the E_1 mode appears in the correlation function (Figure 6b) atop an overall monotonic decrease characteristic of the A mode. The mixed mode character is highlighted in the Fourier transforms shown in Figure 6c, which shows a high $k = 0$ peak and weaker broad peak around $k \sim 0.33$. As a result, the dipole moment of this frequency region is relatively parallel to the helical axis. The opposite sign of vibrations of the helix V arise from the opposite growing direction to those of helix I and VI. Therefore, the transition dipole moments of the three helices are relatively parallel.

The vibrations of the 1655 cm^{-1} region at first glance do not look very different from those at 1645 cm^{-1}, but more residues of alternating sign are present in the same helix. The oscillations of the correlation functions are more pronounced and the period is longer than 3. Fourier transform of the correlation functions

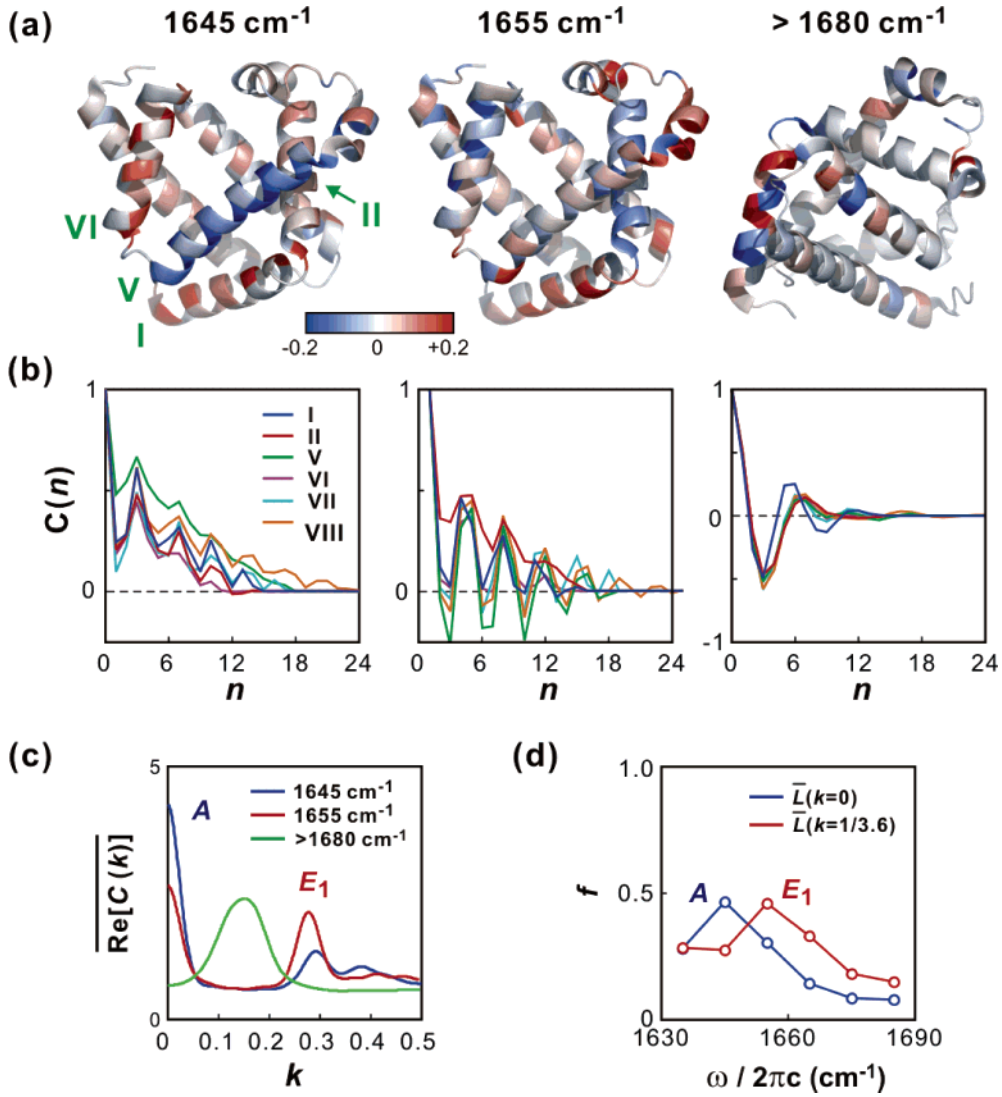

Figure 6. (a) Visualization and characterization of three doorway modes for myoglobin (1645 cm^{-1}, 1655 cm^{-1}, and >1680 cm^{-1}, marked with orange arrows in Figure 4). Amplitudes and phases of oscillators are encoded on the structure, with helices numbered. (b) Helix correlation functions for amide I oscillators along the various helices. (c) The average of the Fourier transformed correlation functions for six helices are shown. (d) The fractional A and E_1 mode character as a function of amide I frequency given by the average correlation length at $k = 0$ and $k = 1/3.6$, respectively.

shows the period is close to $k = 1/3.6$. Though the peak at $k = 0$ is higher than the peak at $k = k_{E_1}$, the extra factor of 2 for $k \neq 0,0.5$ (eq 11) indicates that the contribution of E_1 mode character is greater than that of A mode.

The blue side doorway mode, >1680 cm^{-1}, in Figure 6c has a different periodicity in that the sign of vibration is alternating in every three residues. The total transition dipole moment is small because of the cancellation, which results in the low transition moment in the blue side of the spectrum despite the high α helical content (see Figure 4c). The Fourier transform of the correlation function has a peak at $k \approx (1/2)k_{E_1}$.

The average correlation lengths representing the A and E_1 mode character are shown in Figure 6d. The vibrations of 1645 and 1655 cm^{-1} show the highest A and E_1 mode characters, respectively. The 10 cm^{-1} difference between these peaks matches well with the predicted and observed value of 10 cm^{-1} or less.[1,11,42] However, compared to the smooth and oscillating correlation functions of the ideal A and E_1 modes in Figure 3a, the correlation functions of 1645 and 1655 cm^{-1} regions look like the mixtures of the two ideal correlation functions. We thus see a relatively low peak mode character of ∼0.5 compared to those of the ideal helix that exceed 0.9. This is partially a result of the variation of partial A and E_1 mode character between different helices for a given frequency range. For example, the correlation function of helix II in 1655 cm^{-1} shows more A-like character than most helices at 1645 cm^{-1}. The origin of this can be understood in terms of the near-perpendicular alignment of dipoles within helix II with helices I and V. Therefore, E_1 and, to a lesser extent, A character are spread over all frequencies even though the A and E_1 mode characters peak sharply at 1645 cm^{-1} and 1655 cm^{-1} respectively. Torii and Cho have found that E mode character peaks at lower amide I frequency than does A character.[27,28] This discrepancy is a result of the LAH model; both studies employed degenerate site energies and transition dipole coupling.

F. α/β-Protein: Ubiquitin. In evaluating the structure−frequency correlation for the mixed α/β protein ubiquitin, we analyzed the helix and sheet character of the amide I doorway modes. For sheets, Figure 7a shows the vibrational amplitude and phase correlations for $\nu_\perp$ and $\nu_\parallel$ doorway modes found at the red (<1640 cm^{-1}) and blue (1675 cm^{-1}) sides of the amide I spectrum. We note that in ubiquitin the β sheet is not purely antiparallel, with strands I and V lying parallel to each other, but the phase relationships characteristic of antiparallel β sheets still hold. Vibrations are in-phase for oscillators aligned across strands and alternating for those along the backbone for the $\nu_\perp$

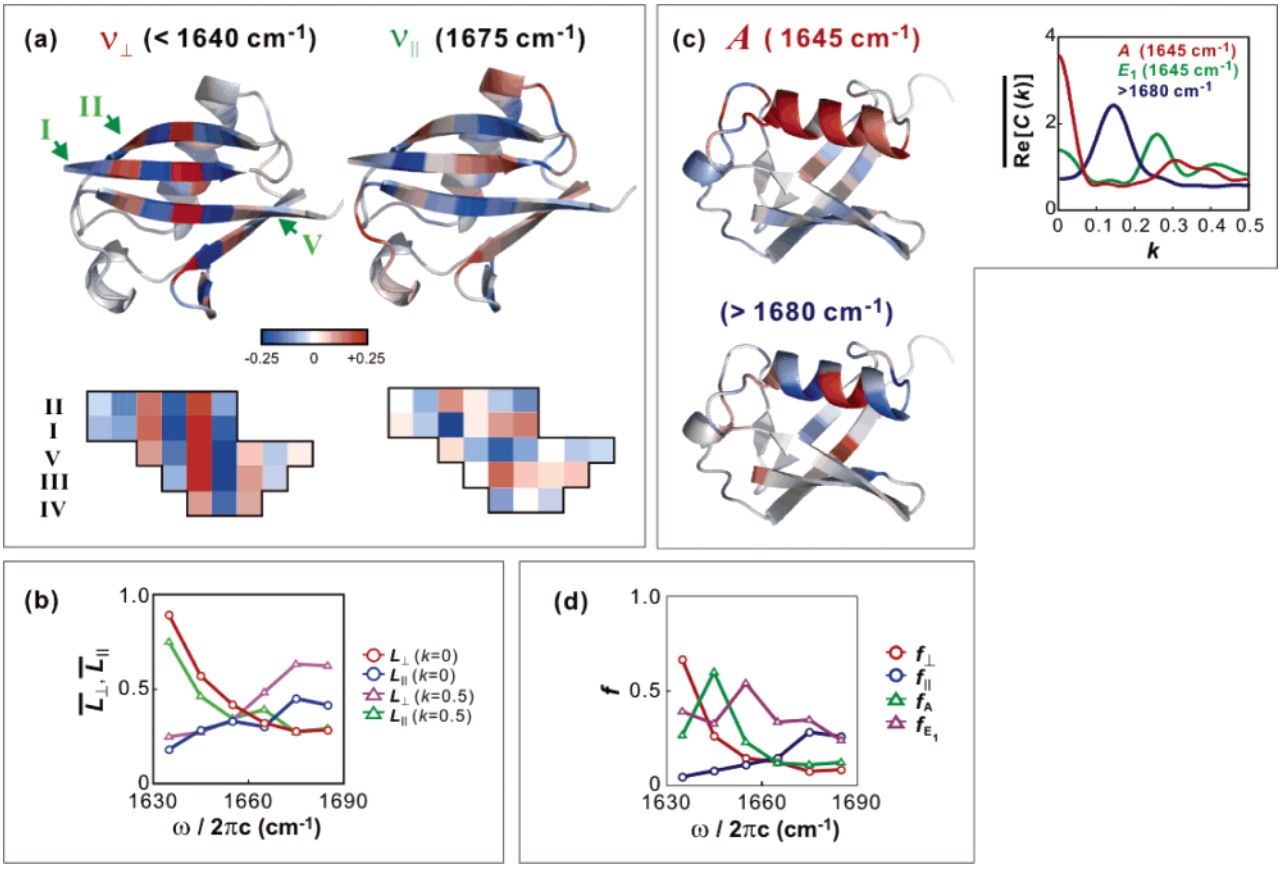

Figure 7. Phase correlation of vibrations of ubiquitin. (a) Amplitudes and phases of vibrations are encoded on the crystal structure and projected on a two-dimensional plane for $\nu_\perp$ and $\nu_{||}$ modes of β sheet of ubiquitin. (b) Averages of the normalized correlation lengths for the β sheet are shown. (c) Vibrational amplitudes of A mode and the mode of >1680 cm^{-1} of α helix are encoded in the crystal structure. The Fourier transforms show a sharp peak at $k = 0$ and close to $1/(2 \times 3.6)$, respectively. (d) $\nu_\perp$ and $\nu_{||}$ character for the β sheet and A and E_1 mode character for the α helix.

symmetry mode, and vice versa for the $\nu_{||}$ mode. Also, the degree of phase correlation of $\nu_{||}$ mode is weaker than that of $\nu_\perp$ mode as a result of mixing with other structural components. Figure 7a indicates that the <1640 cm^{-1} doorway mode is mostly $\nu_\perp$ localized on the sheet, with weak mixing with helix vibrations, whereas the 1675 cm^{-1} mode contains not only the $\nu_{||}$ sheet vibration but also the influence of the helix and coils. In Figure 7b, average correlation lengths are shown for the β sheet, showing similar features to Con A for the $k = 0$ and $k = 0.5$ correlation lengths. Both $\bar{L}_\perp$ ($k = 0$) and $\bar{L}_{||}$ ($k = 0.5$) for the red side $\nu_\perp$ vibration show similar correlation lengths with that of Con A. The β-sheet correlations are distinctly weaker on the blue side of the line. This again is due to mixing and interference of β-sheet oscillators with other structures and the relatively small β sheet content of ubiquitin.

Even though only one helix is present, the α helical character of the ubiquitin doorway mode vibrations shows striking similarities to myoglobin. A high fraction of A and E_1 mode character (f_A and f_{E_1}) is observed at 1645 and 1655 cm^{-1}, respectively. The doorway modes for these frequency ranges are predominately localized on the helix. This is illustrated for the 1645 cm^{-1} doorway mode in Figure 7c. The excitation is largely localized to the helix, although some mixing with the $\nu_\perp$ vibration of the β sheet is apparent. The Fourier transform of the correlation function for amide I oscillators along the peptide sequence is strongly peaked at $k = 0$ indicating the predominant A symmetry to the mode. Analyses of the remaining doorway modes also show similar trends and correlation functions to those seen in myoglobin. The doorway mode of the blue side (>1680 cm^{-1}) shown in Figure 7c contains a $k = (1/2)k_{E_1}$ α-helix excitation similar to that observed in myoglobin.

From the analysis for ubiquitin, we find that β sheet and α helical characters that are found in β-sheet protein (conA) and α-helical protein (myoglobin) are preserved even though ubiquitin is an α/β mixed protein. For the lower frequency transitions, the calculated doorway modes appear to be mainly localized to secondary structures and separated by frequency. This independence may result from the small interference because of the long distance between β sheet and α helix that is separated by the hydrophobic core. Even so, the highest frequencies of the amide I band appear to be far more mixed between the sheet, helix, and random coil regions.

IV. Conclusions

We investigated the relationship between amide I absorption frequency and the vibrations of secondary structural elements through the spatial amplitude and phase correlations between amide I oscillators in proteins. The phases and amplitudes are extracted from a bright doorway mode, rather than the eigenstate vibrations, to more generally describe the vibrations contributing to the IR spectrum in six frequency regions. The mode character f, defined with respect to a correlation length, is useful to summarize the relation of specific vibrational characters with frequency regions. Generally, the vibrational characters are not strongly localized at a particular frequency, though different modes peak at characteristic frequencies. The β sheet absorption is split into the extreme high- and low-frequency sides of the amide I band. The $\nu_\perp$ mode on the red side (<1640 cm^{-1}) is

strongly localized within the β sheet, while the blue side $\nu_{||}$ mode is mixed with other structural components. The A and E_1 mode of the α helix are located in the 1645 and 1655 cm^{-1} regions, respectively, with a fair amount of mixing with themselves and with other components. The result for the α/β protein ubiquitin showed the preservation of the same vibrational characteristic for all the four modes. The red side is dominated by the $\nu_{\perp}$ mode of the β sheet structure and other regions result from the superposition of various secondary structure vibrational modes. As shown in the case of the helix II of the myoglobin the interference between adjacent secondary structures can shift the location of the mode characters.

Generally speaking, we find that doorway mode vibrations and thus the IR spectra involve vibrations that are largely localized on secondary structures at lower amide I frequencies but involve mixtures of various secondary structure modes on the high-frequency side of the line. It appears that within the constraints of our current coupling model, the four characteristic vibrational modes ($\nu_{\perp}$, $\nu_{||}$, A, E_1) plus a random coil contribution may be an effective basis set for describing the amide I vibrations of proteins. We find that the simple superposition of the vibrational characters works well for the protein having well-separated secondary structures such as ubiquitin, but the interference effect may not be negligible in the presence of the strong tertiary contacts.

The calculation model used in this paper is relatively simple, and several avenues for improvement exist. The unperturbed amide I frequency was chosen for the best match to the experimental spectrum, and the solvent interaction is considered uniformly in the red shift of the site frequency. The ongoing work of many groups is targeted at parametrizing a quantitatively accurate LAH. Various improved models are being developed to get more accurate site frequency by considering the local electric field[38] or electrostatic potential[39,40] drawing on ab initio calculations. For the coupling model, transition charge coupling[18] by oscillating charges located at the atoms of peptide groups can be used to account for the short-range interactions. In addition, in the case of the real protein, dynamics are certain to be important in the spectroscopy, and the static average implied by eq 1 may not adequately characterize the spectrum. Femtosecond-to-picosecond fluctuations in the solvent, side chains, and periphery of the protein could lead to a number of effects, including narrowing of the statically calculated spectrum, motional narrowing of subsets of the eigenstates we use here, and exciton relaxation within the amide I band. IR spectra calculated directly from the Fourier transform of a dipole correlation function for the LAH model, although computationally taxing, would allow nonadiabatic effects to be characterized.

Despite the needs for the improvement of the model, we believe that the correlation characteristics of this paper are still applicable for calculation based in the local vibrational mode basis. The correlation between spectral changes observed in the experiments and the changes of calculated parameters with a proper model for structural dynamics will provide a quantitative way of interpretation of the structural changes on the molecular level. The visualization and spectral characterization techniques will find considerable applicability in the interpretation of infrared and 2D IR data of proteins, and well as in the comparison of experimental and simulation studies of folding, unfolding or binding of proteins.

Acknowledgment. We thank Arend Dijkstra and Jasper Knoester for stimulating discussions and suggestions on several aspects of characterizing delocalized states. This work was supported by the National Science Foundation (Grant CHE-0316736) and the Petroleum Research Fund of the American Chemical Society. A.T. thanks the David and Lucile Packard Foundation for their fellowship support.

References and Notes

(1) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
(2) Jackson, M.; Mantsch, H. H. *Crit. Rev. Bichem. Mol.* **1995**, *30*, 95.
(3) Byler, D. M.; Susi, H. *Biopolymers* **1986**, *25*, 469.
(4) Dong, A.; Huang, P.; Caughey, W. S. *Biochemistry* **1990**, *29*, 3303.
(5) Lee, D. C.; Haris, P. I.; Chapman, D.; Mitchell, R. C. *Biochemistry* **1990**, *29*, 9185.
(6) Baumruk, V.; Pancoska, P.; Keiderling, T. A. *J. Mol. Biol.* **1996**, *259*, 774.
(7) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.
(8) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2036.
(9) Woutersen, S.; Hamm, P. *J. Phys. Chem. B* **2000**, *104*, 11316.
(10) Woutersen, S.; Mu, Y.; Stock, G.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11254.
(11) Woutersen, S.; Hamm, P. *J. Chem. Phys.* **2001**, *115*, 7737.
(12) Demirdöven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126*, 7981.
(13) Chung, H. S.; Khalil, M.; Tokmakoff, A. *J. Phys. Chem. B* **2004**, *108*, 15332.
(14) Chung, H. S.; Khalil, M.; Smith, A. W.; Ganim, Z.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 612.
(15) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.
(16) Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29*, 81.
(17) Ham, S.; Cho, M. *J. Chem. Phys.* **2003**, *118*, 6915.
(18) Hamm, P.; Woutersen, S. *Bull. Chem. Soc. Jpn.* **2002**, *75*, 985.
(19) Higgs, P. W. *Proc. R. Soc. London, Ser. A* **1953**, *133*, 472.
(20) Miyazawa, T. *J. Chem. Phys.* **1960**, *32*, 1647.
(21) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.
(22) Ham, S.; Hahn, S.; Lee, C.; Kim, T.-K.; Kwak, K.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 9333.
(23) Lee, C.; Cho, M. *J. Phys. Chem. B* **2004**, *108*, 20397.
(24) Abramavicius, D.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. B* **2004**, *108*, 18034.
(25) Wang, J.; Hochstrasser, R. M. *Chem. Phys.* **2004**, *297*, 195.
(26) Dijkstra, A. G.; Knoester, J. *J. Phys. Chem. B* **2005**, *109*, 9787.
(27) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *97*, 92.
(28) Choi, J. H.; Ham, S.; Cho, M. *J. Chem. Phys.* **2002**, *117*, 6821.
(29) Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *97*, 86.
(30) Didraga, C.; Knoester, J. *J. Chem. Phys.* **2004**, *121*, 10687.
(31) Woutersen, S.; Hamm, P. *J. Phys.: Condens. Matter* **2002**, *14*, R1035.
(32) Torii, H.; Tatsumi, T.; Tasumi, M. *Mikrochim. Acta [Suppl.]* **1997**, *14*, 531.
(33) Nevskaya, N. A.; Chirgadze, Y. N. *Biopolymers* **1976**, *15*, 637.
(34) Parkin, S.; Rupp, B.; Hope, H. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **1996**, *52*, 1161.
(35) Wlodawer, A.; Svensson, L. A.; Sjolin, L.; Gilliland, G. L. *Biochemistry* **1988**, *27*, 2705.
(36) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531.
(37) Maurus, R.; Overall, C. M.; Bogumil, R.; Luo, Y.; Mauk, A. G.; Smith, M.; Brayer, G. D. *Biochim. Biophys. Acta* **1997**, *1341*, 1.
(38) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. *J. Chem. Phys.* **2004**, *121*, 8887.
(39) Kwac, K.; Cho, M. *J. Chem. Phys.* **2003**, *119*, 2247.
(40) Bour, P.; Keiderling, T. A. *J. Chem. Phys.* **2003**, *119*, 11253.
(41) Barth, A.; Zscherp, C. *Q. Rev. Biophys* **2002**, *35*, 369.
(42) Lee, S.-H.; Krimm, S. *Biopolymers* **1998**, *46*, 283.