

Simulation Protocols for Coherent Femtosecond Vibrational Spectra of Peptides

Wei Zhuang, Darius Abramavicius, Tomoyuki Hayashi, and Shaul Mukamel*

Department of Chemistry, University of California at Irvine, Irvine, California 92697-2025

Received: October 11, 2005; In Final Form: December 2, 2005

Two algorithms for simulating the response of peptides to sequences of IR pulses are developed and applied to *N*-methyl acetamide (NMA) and a 17 residue α -helical peptide (YKKKH17) in D₂O. A fluctuating vibrational-exciton Hamiltonian for the amide I mode is constructed from molecular dynamics trajectories. Coupling with the environment is described using a density functional theory electrostatic map. The cumulant expansion of Gaussian fluctuation incorporates motional narrowing due to fast frequency fluctuations and is adequate for NMA and for isotopically labeled bands in large peptides. Real-space truncation of the scattering matrix of the nonlinear exciton equations significantly reduces the computational cost, making it particularly attractive for slow fluctuations in large globular proteins.

I. Introduction

Probing the structure and folding dynamics of proteins is one of the most fundamental problems in biophysics and has been the subject of intensive effort.^{1–16} IR spectroscopy provides a valuable tool in these studies.^{17–24} Significant progress has been made over the past decade in developing coherent two-dimensional infrared (2DIR) techniques.^{25–33} In a 2DIR experiment, three incoming pulses with wavevectors \mathbf{k}_1 , \mathbf{k}_2 , and \mathbf{k}_3 interact with the peptide to generate a coherent signal in one of the directions $\mathbf{k}_s = \pm\mathbf{k}_1 \pm \mathbf{k}_2 \pm \mathbf{k}_3$. The pulse sequence and time delays (t_1 , t_2 , and t_3) are shown in Figure 1. 2D correlation plots of the signals plotted as Fourier transforms with respect to two of these delay periods reveal new types of information with enhanced spectral resolution. Diagonal peaks show the fundamental transitions, while cross-peaks and their line shapes probe the fluctuations of the correlations among various structural elements.²⁶

2D NMR techniques originally introduced in the 1970s³⁴ had turned it into a useful structural tool.³⁵ Extracting information from NMR data requires extensive simulations. The “direct inversion” of spectra to structures is based on constrained fits guided by simulations.^{36–38} A recent study of human ubiquitin in solution³⁷ combined NMR experiments with molecular dynamics (MD) simulations to directly determine the entire ensembles of protein conformations, rather than merely the average structure: Both the native structure and its associated dynamics are thus obtained simultaneously. Since IR spectra are more congested, and anharmonic vibrational Hamiltonians are much more complex than spin Hamiltonians,^{39,40} efficient simulation strategies are essential for the interpretation and analysis of 2DIR signals. Thanks to the ultrafast time scale, the simulation of 2DIR signals with atomic level details only requires subnanosecond trajectories, which are readily available.^{41–44}

Standard tools are available for computing two-point correlation functions which are commonly used in the analysis of fluctuations. Nonlinear response functions require the efficient and accurate sampling of multipoint correlation functions; this

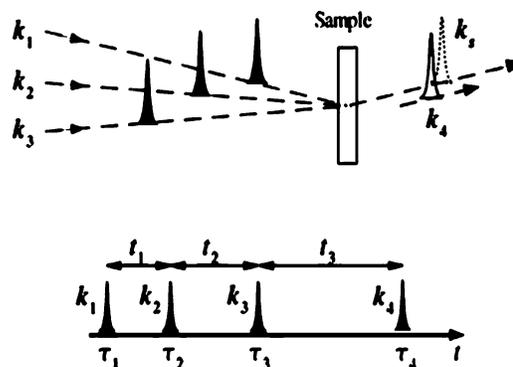


Figure 1. (top) Pulse configuration for heterodyne four wave mixing. \mathbf{k}_1 , \mathbf{k}_2 , and \mathbf{k}_3 are the input pulses, \mathbf{k}_s is the signal generated, which is in the same direction as the detection pulse \mathbf{k}_4 . (bottom) The pulse sequence for coherent 2D experiments. t_j are the time intervals between pulses centered at τ_i .

article addresses some of the simulation challenges involved in predicting and interpreting multidimensional measurements.

We focus on the amide vibrational modes which are localized along the peptide backbone. Their couplings (~ 10 cm⁻¹), which depend on the secondary structure and its fluctuations, are typically much smaller than the mode frequencies (~ 1600 cm⁻¹). The spectrum thus consists of well-separated groups of energy levels representing single excitations $|e\rangle$, double excitations $|f\rangle$, and so forth (Figure 4, part vi). The laser pulses probe the system by inducing transitions among these manifolds. Linear (single-pulse) techniques only access the lowest (single-exciton) manifold, whereas doubly excited (two-exciton) states can be effectively monitored by third-order spectroscopies.

Spectra of small peptides may be calculated using sum-over-states (SOS) expressions⁴⁵ in conjunction with simulated structural trajectories. The optical response is described in terms of transitions among eigenstates. For short peptides, in which level crossings among eigenstates can be ignored, the cumulant expansion of Gaussian fluctuation (CGF)⁴⁵ can be adopted to account for bath fluctuations of arbitrary time scale, ranging from motional narrowing (fast), to the inhomogeneous broadening regime (slow). For several reasons, a different simulation protocol should be adopted for large peptides: The SOS

* To whom correspondence should be addressed. E-mail: smukamel@uci.edu.

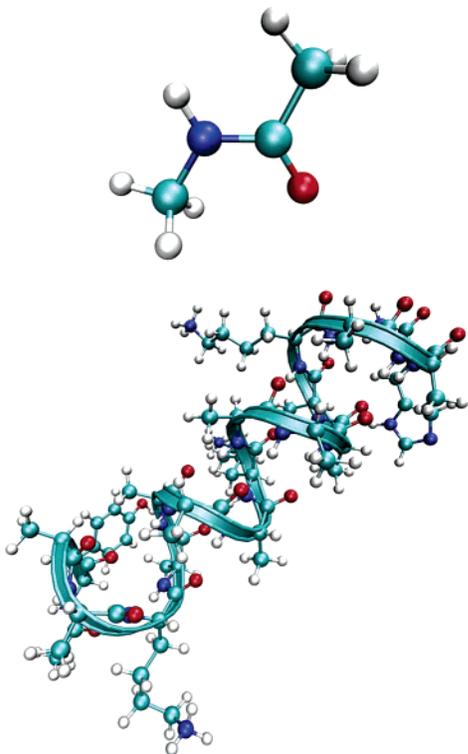


Figure 2. Structure of NMA (top) and YKKKH17 (bottom).

computational cost of third-order coherent signals scales as $\sim N^4$,⁴⁵ where N is the number of coupled oscillators and this calculation becomes prohibitively expensive as N is increased. In addition, as the peptide size is increased, the energy levels become more dense and the adiabatic approximation underlying the CGF no longer holds. Finally, in large peptides, the spectra are often dominated by slow large-scale structural fluctuations. Line broadening may then be simulated by averaging the homogeneous response over ensembles of system configurations.⁴⁶

The nonlinear exciton equations (NEEs)^{45,47–49} provide a practical alternative for large proteins. These equations use the single-exciton basis and view the nonlinear response in terms of scattering among one-exciton quasiparticles. Two-exciton resonances enter through the exciton scattering matrix, totally avoiding the expensive computation of multiple-exciton states. The apparent NEE scaling of the computational cost for third-order signals is also $\sim N^4$, similar to the SOS. However, when the short-range nature of exciton couplings is exploited, the cost may be reduced considerably and could even scale linearly with protein size.

We explore the applicability and test the limits of the CGF and NEE formalisms. Linear absorption and several third-order signals are predicted and discussed for the amide I band line shape of NMA and a 17 residue α -helical peptide Ac-YAAKAAAKAAAKAAH-NH₂ (named YKKKH17 since YKKKH are its nonalanine (lysine) residues) in D₂O (lower panel of Figure 2). We demonstrate that the CGF is suitable for NMA and for narrow isotopically labeled bands whereas the NEE with inhomogeneous averaging yields accurate spectra for large peptides.

A brief survey of coherent multidimensional signals is given in section II, more details can be found in refs 29 and 50. An accurate vibrational-exciton Hamiltonian is crucial for the simulation of the spectra. An electrostatic DFT map is constructed for the fundamental frequencies and anharmonicities and used to generate the Hamiltonian of NMA.⁵¹ The vibrational

Hamiltonian is introduced in section III, and the simulation protocols are outlined in section IV and presented in detail in sections V (CGF) and VI (NEE). These are then applied to NMA (section VII) and YKKKH17 (section VIII). We finally discuss our results and future applications in section IX.

II. Coherent Multidimensional Signals of coupled vibrations

We consider third-order signals induced by three resonant femtosecond pulses with the electric field

$$\mathbf{E}(\mathbf{r}, \tau) = \frac{1}{2} \sum_{j=1}^3 E_{\nu_j}^{(j)} \delta(\tau - \tau_j) \exp(i\mathbf{k}_j \mathbf{r} - i\omega_j \tau) + c.c. \quad (1)$$

where $E^{(j)}$ is the amplitude of the j th pulse and $\nu_j = x, y, z$ denotes its polarization direction.

We assume an ideal impulsive experiment with very short pulses where the desired time ordering of the various interactions is enforced. Finite-pulse envelopes can be included,⁵⁰ but the experiment is best understood in this impulsive limit. The response of an ensemble of molecules is determined by the nonlinear polarization,⁵⁰ \mathbf{P} , which serves as the source for the \mathbf{k}_s signal in the Maxwell equations

$$\mathbf{P}_{\nu_4}^{\mathbf{k}_s}(t_3, t_2, t_1) = \frac{1}{2^3} \sum_{\nu_3 \nu_2 \nu_1} S_{\nu_4 \nu_3 \nu_2 \nu_1}^{\mathbf{k}_s}(t_3, t_2, t_1) E_{\nu_3}^{(3)} E_{\nu_2}^{(2)} E_{\nu_1}^{(1)} \quad (2)$$

The system is initially at equilibrium in the ground state, and the response function depends parameterically on the three time intervals between pulses t_j (Figure 1). Due to interference of the induced polarization of the various molecules, the signal is only generated along specific (phase-matching) directions \mathbf{k}_s . There are four possible signals with wavevectors: $\mathbf{k}_I = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$, $\mathbf{k}_{II} = +\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3$, $\mathbf{k}_{III} = +\mathbf{k}_1 + \mathbf{k}_2 - \mathbf{k}_3$, and $\mathbf{k}_{IV} = +\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$. These techniques can be understood by using Feynman diagrams⁵⁰ which depict the evolution of the vibrational density matrix in the course of the nonlinear process. (Figure 4). Time goes from the bottom to the top, and the two vertical lines in the diagram represent the ket and the bra of the density matrix, while arrows represent interactions with the laser pulses.

We shall focus on \mathbf{k}_I and \mathbf{k}_{III} . The three diagrams contributing to \mathbf{k}_I are shown in Figure 4 (top). In all the diagrams, the density matrix is in a single-quantum coherence $|g\rangle\langle e|$ between the ground state and the singly excited state during t_1 . During t_3 , it is either in the conjugate coherence $|e\rangle\langle g|$ ((i) and (ii)) or in a coherence between the one- and two-exciton manifolds $|f\rangle\langle e|$ (iii). The \mathbf{k}_I signal is displayed by performing a double Fourier transform with respect to the first and the third time delays ($F(\Omega) \equiv \int_0^\infty dt \exp(i\Omega t) F(t)$)

$$\sigma_{\nu_4}^I(\Omega_3, t_2, \Omega_1) = \frac{1}{2^3} \sum_{\nu_3 \nu_2 \nu_1} S_{\nu_4 \nu_3 \nu_2 \nu_1}^{\mathbf{k}_s}(\Omega_3, t_2, \Omega_1) E_{\nu_3}^{(3)} E_{\nu_2}^{(2)} E_{\nu_1}^{(1)} \quad (3)$$

where Ω_1 and Ω_3 are the Fourier conjugates to t_1 and t_3 , respectively. In all calculations, we set $t_2 = 0$.

\mathbf{k}_{III} is similarly described by the two Feynman diagrams (iv) and (v) and shows double-quantum coherences between the ground state and the two-exciton band $|f\rangle\langle g|$ during the t_2 interval. During t_3 , it has a single-quantum coherence $|e'\rangle\langle g|$ (iv) and $|f\rangle\langle e'|$ (v). \mathbf{k}_I , known as the photon echo, can improve the resolution by eliminating certain types of inhomogeneous broadening. The spectral bandwidth is doubled in \mathbf{k}_{III}

which carries direct information regarding the coherence between the two-exciton states and the ground state (double-quantum coherence).²⁸ The system is transferred to a coherence by the first pulse. The second pulse takes the system either to a population or to a coherence between two excitonic states. Then the population and coherence evolution can be probed by holding the second delay time, t_2 (often referred to as population waiting time) fixed. The third pulse creates coherences either between the ground and one-exciton states or between the one- and two-exciton states.

By performing a Fourier transforms with respect to the second and the third time delays (Ω_2 and Ω_3 are the Fourier conjugates to t_2 and t_3 , respectively), we can observe double- (single-) quantum coherence along Ω_2 (Ω_3). The \mathbf{k}_{III} signal is

$$\sigma_{\nu_s}^{\text{III}}(\Omega_3, \Omega_2, T_1) = \frac{1}{2^3} \sum_{\nu_3 \nu_2 \nu_1} S_{\nu_3 \nu_2 \nu_1}^{\mathbf{k}_{\text{III}}}(\Omega_3, \Omega_2, t_1) E_{\nu_3}^{(3)} E_{\nu_2}^{(2)} E_{\nu_1}^{(1)} \quad (4)$$

In all the calculations, we have set t_1 to zero.

The linear response is similarly related to the linear response function

$$\mathbf{P}_{\nu_2}^{\mathbf{k}_1}(t) = \frac{1}{2} \sum_{\nu_1} S_{\nu_2 \nu_1}^{\mathbf{k}_1}(t) E^{(1)} \quad (5)$$

whose Fourier transform gives the absorption spectrum

$$\sigma_A(\Omega) = \frac{1}{2} \text{Im} \sum_{\nu_2 \nu_1} S_{\nu_2 \nu_1}^{\mathbf{k}_1}(\Omega) E^{(1)} \quad (6)$$

III. Fluctuating Vibrational-Exciton Hamiltonian

The construction of a fluctuating Hamiltonian for the primary vibrations and their coupling with a bath is the first step in the simulation of vibrational line shapes. Nonlinear signals are sensitive to fine details (e.g., anharmonicities, overtone transitions, etc.) and require a high-level Hamiltonian.⁵² Vibrational motions and spectra are commonly described by normal modes. These collective coordinates provide a convenient zero-order approximation for the vibrational eigenstates and frequencies. However, a normal-mode analysis is often too expensive for large polypeptides. Instead, since the amide I vibrations are localized on the backbone peptide bonds, it is desirable to trace the structural origin of spectral features to local vibrational coordinates.⁵³ A peptide can be viewed as a chain of beads, each containing one amide residue (O=C–N–H). The Frenkel exciton model has long been used^{17,19,54,55} to represent the vibrational Hamiltonian in this localized basis. Diagonal elements of the Hamiltonian matrix give the zero-order local mode frequencies while off-diagonal elements represent their couplings. For large, globular proteins, the Hamiltonian may be constructed using parameters obtained from quantum chemistry calculations performed on small segments.

Quantum calculations of 96 NMA–water clusters were recently combined with a fitting procedure to obtain a linear relation between the frequency of NMA, a simplest model system for the peptide bond, and the instant external electric potential.⁵⁶ A similar procedure was adopted to obtain a relation between NMA frequency and the external electric fields from 200 cluster calculations.⁵⁷ Both approaches work very well for the absorption line width (27 cm^{-1} , compared with the experimental width of 29 cm^{-1}) and give a reasonable solvent peak shift ($\sim 60 \text{ cm}^{-1}$ compared with the experimental shift of 90 cm^{-1}). By considering some of the extreme configurations

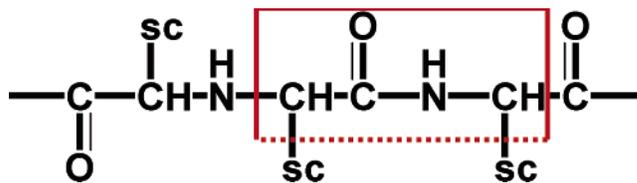


Figure 3. Amide I vibrational chromophore (red rectangle). “SC” stands for “side chain”. The dotted line at the bottom indicates that some atoms of the side chain may be included in the chromophore depending on the “group” defined in the Charmm force field.

of NMA–water clusters, Keiderling reproduced the solvent shift even better.⁵⁸

The cluster fitting for specific solvents is highly accurate but may be too expensive for applications which require repeated calculations for a large number of configurations of complex polypeptides. We thus adopt a different approach, starting with the vibrational-exciton Hamiltonian

$$\hat{H} = \hat{H}_S + \hat{H}_F \quad (7)$$

where

$$\hat{H}_S = \sum_m \epsilon_m \hat{B}_m^\dagger \hat{B}_m + \sum_{mn}^{m \neq n} J_{mn} \hat{B}_m^\dagger \hat{B}_n - \frac{1}{2} \sum_m \Delta_m \hat{B}_m^\dagger \hat{B}_m^\dagger \hat{B}_m \hat{B}_m \quad (8)$$

is the system Hamiltonian and \hat{H}_F is the interaction with the optical field, $E(t)$

$$\hat{H}_F = -\mathbf{E}(t) \sum_m \mu_m (\hat{B}_m^\dagger + \hat{B}_m) \quad (9)$$

\hat{B}_m^\dagger (\hat{B}_m) is the creation (annihilation) operator for the m th amide I mode, localized within the amide unit (O=C–N–H), with frequency ϵ_m , anharmonicity Δ_m , and transition dipole moment μ_m . These operators satisfy the Bose commutation relations $[\hat{B}_m, \hat{B}_n^\dagger] = \delta_{mn}$. J_{mn} are the harmonic intermode couplings. All parameters of \hat{H}_S fluctuate due to conformational changes of the backbone, the solvent, and side-chain dynamics.

To describe the m th localized amide I mode, it must be separated from all other modes: The segment made of that amide residue and two neighboring neutral groups (according to CHARMM27 force field,⁴⁴ which we use in electrostatic interaction calculation) containing the α carbons is defined as the chromophore (Figure 3). The rest of the protein, that is, all atoms other than this chromophore, and the solvent molecules are treated as a bath, whose effect on the chromophore will be described by a fluctuating electrostatic field.

The frequency of mode m subjected to a time-dependent electric field is given by $\epsilon_m(\mathbf{E}) = \epsilon - \delta\epsilon_m(\mathbf{E})$. $\epsilon = 1717 \text{ cm}^{-1}$ is the frequency of an isolated NMA in the gas phase.⁵⁸ The spectral shift $\delta\epsilon_m(\mathbf{E})$ and the anharmonicity $\Delta_m(\mathbf{E})$ are obtained from an ab initio DFT map which relates the electric field and its derivatives at a reference point to the fundamental and overtone frequencies of the amide I mode.⁵¹ Geometry optimization was performed on a gas-phase NMA reference structure at the BPW91/6-31g(d,p) level of DFT using Gaussian 03.⁵⁹ This functional is known to give accurate vibrational frequencies.⁵⁸ During optimization, the origin is set to be the reference point, which is in the middle of the line connecting the oxygen and hydrogen atoms of this NMA molecule and the x axis goes from O to H. O=C–N–H then defines the x – y plane, where the y axis points toward the N atom.

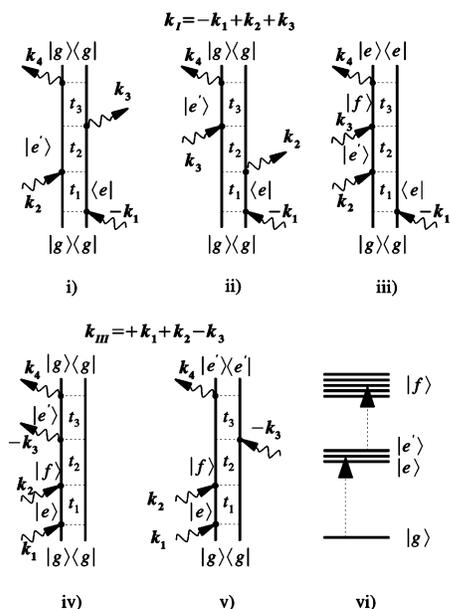


Figure 4. Double-sided Feynman diagrams for the k_1 and k_m techniques showing the pulse sequence and the vibrational density matrix during each time interval.

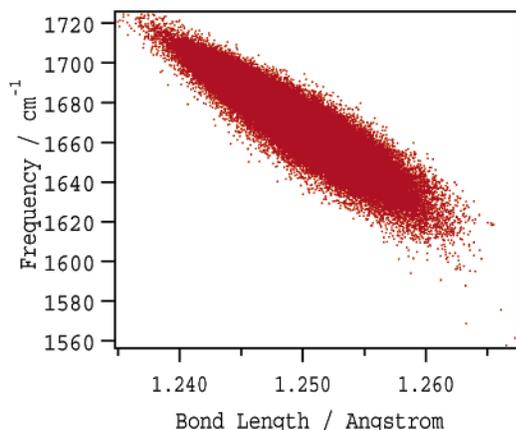


Figure 5. Scatter plot of the amide I fundamental frequency vs the C=O bond length.

A DFT map was constructed which relates $\delta\epsilon_m(\mathbf{E})$ and $\Delta_m(\mathbf{E})$ to a 19 component vector $\mathbf{C}^T \equiv (E_x, E_y, E_z, E_{xx}, \dots, E_{xxx}, \dots)$ representing all the independent components of the electric field, its gradients, and second derivatives at the origin point

$$\delta\epsilon_m(\mathbf{E}) = \sum_{\nu} K_{\nu} C_{\nu} + \frac{1}{2} \sum_{\nu\nu'} K_{\nu\nu'} C_{\nu} C_{\nu'} \quad (10)$$

The coefficients K_{ν} and $K_{\nu\nu'}$ are obtained from vibrational eigenstate calculations for the sixth-order DFT anharmonic vibrational potential of a single NMA chromophore in a spatially nonuniform electric field.⁵¹ $\Delta_m(\mathbf{E})$ was expanded in a similar fashion. A modified Gaussian 03 code was used to generate the anharmonic potential and the ARNOLDI algorithm was employed for the eigenstates calculations. To trace the origin of the electric field influence on the amide I band, the C=O bond length obtained by energy minimization for the various field values was also parametrized in terms of \mathbf{C} .⁵¹

The instantaneous C=O bond length was calculated from the DFT map. The scatter plot of the amide I frequency vs C=O bond length is shown in Figure 5. The strong positive correlation suggests that electric field fluctuation causes geometrical

changes of the amide unit due to polarization, which is the reason for the amide I frequency fluctuation. This agrees with the conclusions from previous studies.^{51,61–63}

The components of \mathbf{C}^T were determined at each time point by a least-squares fit to the electric field sampled at 67 points. Spanning the spatial region of the electric field relevant in the amide vibrations, to decide this region we have expanded the molecular charge density $\rho(\mathbf{r})$ in the vibrational modes Q_i

$$\rho(\mathbf{r}) = \rho^{(0)}(\mathbf{r}) + \sum_i \left(\frac{\partial \rho(\mathbf{r})}{\partial Q_i} \right) Q_i + \dots \quad (11)$$

The transition charge density (TCD) $\partial \rho / \partial Q_i$ represents the electronic structure change due to the Q_i vibration. We expect the electrostatic potential generated by the solvent in the region where the TCD is large to dominate the optical response of that vibration. Three electric field components were sampled at each of these 67 points and used for the fitting calculated for each bath configuration. The \mathbf{C}^T calculated using this protocol represents the electric field distribution across the TCD region.

The couplings of different amide modes were assumed to depend solely on the peptide backbone structure. For the nearest covalently bonded modes, we used Torii and Tasumi's ab initio map.²⁰ All other couplings were calculated by making the dipole approximation for each amide unit^{17,19} and using the electrostatic TDC model

$$J_{m,n} = \frac{0.1A (\mu_m \cdot \mu_n) - 3[\mu_m \cdot \mathbf{e}_{mn}][\mu_n \cdot \mathbf{e}_{mn}]}{\epsilon r_{mn}^3} \quad (12)$$

where μ_m is the transition dipole in ($\text{D} \text{ \AA}^{-1} \text{ u}^{-1/2}$) units, r_{mn} is the distance between dipoles (in \AA), \mathbf{e}_{mn} is the unit vector connecting m and n , and $\epsilon = 1$ is the dielectric constant. The angle between the transition dipole and C=O bond is 10° . The conversion factor $A = 848619/1650$ gives the coupling energy in cm^{-1} .

IV. Simulation Protocols

The CGF modeling of the coherent vibrational response involves the following steps:

1. A MD simulation generates a sequence of protein and solvent configurations for NMA and YKKKH17.

2. A fluctuating vibrational-exciton Hamiltonian, $\hat{H}_S(t)$, and the transition dipole matrix, $\mu(t)$, in the one-exciton and two-exciton bases are constructed for each configuration. The fundamental frequencies and the diagonal anharmonicities are generated by our electrostatic DFT map.

3. The reference exciton Hamiltonian \bar{H}_S is obtained as the time average of $\hat{H}_S(t)$. The eigenvalues, $\bar{\omega}_n$, and eigenvectors, $\bar{\psi}_n$, of this Hamiltonian form a reference basis. The trajectory of $\hat{H}_S(t)$ and $\mu(t)$ is transformed to the reference basis set, creating a fluctuating Hamiltonian matrix with $\delta\omega_n(t) = \omega_n(t) - \bar{\omega}_n$ (diagonal fluctuations) and off-diagonal fluctuations of couplings, $\delta J_{mn}(t) \equiv J_{mn}(t)$. The line broadening functions (eq 15) were generated using this Hamiltonian.

4. The necessary products of four transition dipoles are calculated. Both orientational averaging⁶⁴ and time averaging were performed to account for fluctuations of the transition dipoles.

5. The reference Hamiltonian, line broadening functions, and average amplitudes of the Liouville space pathways were

combined to calculate the response functions. The experimental lifetime broadening is added.⁶⁵

The NEE approach requires the following steps:

1. An MD or a Monte Carlo simulation generates an ensemble of system configurations.

2. The fluctuating vibrational-exciton Hamiltonian, $\hat{H}_S(t)$, and transition dipole matrix, $\mu(t)$, were constructed in a fashion similar to the CGF, except that only the one-exciton block is calculated.

3. At each configuration, we used $\hat{H}_S(t)$ to calculate the exciton scattering matrix in the one-exciton basis.

4. The products of four transition dipoles⁵⁰ were calculated and ensemble averaged.

5. The response functions were calculated at each configuration using the Green function expressions⁶⁴ and then averaged over configurations.

V. CGF simulations of Multipoint Response Functions

Simulations were performed on the basis of eigenstates ψ_n of the reference Hamiltonian \bar{H}_s with energies $\bar{\epsilon}_n$. All elements of the Hamiltonian fluctuate due to coupling with the bath: diagonal frequency fluctuations, $\delta\epsilon_n$, are primarily responsible for the line shapes, while off-diagonal fluctuations of couplings, δJ_{mn} , cause population transport and lifetime broadening. An adiabatic approximation is used by assuming that all fluctuation amplitudes are much smaller than the average energy spacings between states so that curve crossing is negligible. Thus, the CGF is suitable for calculating equilibrium fluctuations of small systems with well-separated energy levels.

The one-exciton states were obtained by solving the eigenvalue problem

$$\sum_n [\delta_{mn}\epsilon_m + (1 - \delta_{mn})J_{mn}]\psi_{en} = \epsilon_e \psi_{em} \quad (13)$$

where the summation runs over the chromophores. This yields the one-exciton eigenfunctions, ψ_{em} , and energies, ϵ_e . Two-exciton states were obtained by diagonalizing the two-exciton block. The response function was calculated using the second-order cumulant expansion,⁶⁶ which is exact for Gaussian fluctuation statistics. A detailed description of this method and closed correlation function expressions for the signals can be found elsewhere.^{45,50}

The linear response function is given by a sum over transitions from the ground to all first excited states e , whose energies lie within the laser bandwidth

$$S_{v_2, v_1}^{k_1}(t) = i\theta(t) \sum_e \langle \langle \mu_{ge}^{v_2} \mu_{eg}^{v_1} \rangle \rangle_t \exp(-i\epsilon_e t - \tilde{g}_{eg}(t) - \bar{\gamma}_{eg} t) \quad (14)$$

where $\theta(t) = 0$ ($\theta(t) = 1$) for $t < 0$ ($t > 0$) is the Heaviside step function, μ_{eg} is the transition dipole element between the ground state and the e th one-exciton state, and $\langle \langle \dots \rangle \rangle_t$ represents orientational averaging and $\langle \langle \dots \rangle \rangle_t$ represents time averaging. The line broadening functions \tilde{g}_{ab} for the transition between the two states are given by

$$\tilde{g}_{ab}(t) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{1 - \cos(\omega t)}{\omega^2} \coth(\hbar\omega/(2k_B T)) C''_{ab}(\omega) + i \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{\sin(\omega t) - \omega t}{\omega^2} C''_{ab}(\omega) \quad (15)$$

where a and b can be states in either the exciton basis e, f or the local basis n, m . $C''_{ab}(\omega)$ is the spectral density of the

fluctuation of the transition frequency $U(t) \equiv \epsilon_a - \epsilon_b - (\bar{\epsilon}_a - \bar{\epsilon}_b)$

$$C''_{ab}(\omega) = \frac{1}{2\tau\hbar^2} \frac{\hbar\omega}{2k_B T} |\tilde{U}_{ab}(\omega)|^2 \quad (16)$$

$\tilde{U}(\omega)$ is the Fourier transform of $U(t)$. In the present simulations, we calculated $\tilde{g}_{ab}(t)$ in the exciton basis, but one can also calculate it first in the local basis and then transform to the exciton basis.⁶⁷

We have added lifetime broadening, $\bar{\gamma}_{eg}$, which originates from fluctuations of the couplings

$$\bar{\gamma}_{ab} = \frac{1}{2} (K_{aa} + K_{bb}) \quad (17)$$

K_{aa} is the diagonal element of the vibrational relaxation matrix \mathbf{K} representing the total relaxation rate from state a to all the other states. Off-diagonal elements, K_{ab} , give the relaxation rate from state a to b . Conservation of probability implies that the sum of the elements of the relaxation matrix in each column is zero

$$K_{bb} = - \sum_a K_{ba} \quad (18)$$

The vibrational relaxation rates are directly related to the correlation functions of fluctuations of couplings

$$K_{ab} = - \int_0^{\infty} d\tau [e^{i\omega_{ba}\tau} M_{ab}(\tau) + e^{-i\omega_{ba}\tau} M_{ab}(-\tau)] \quad (19)$$

where

$$M_{ab}(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} M''_{ab}(\omega) [\cos(\omega\tau) \coth(\hbar\omega/(2k_B T)) - i \sin(\omega\tau)] \quad (20)$$

and $M''_{ab}(\omega)$ is a spectral density of J_{ab}

$$M''_{ab}(\omega) = \frac{1}{2\tau\hbar^2} \frac{\hbar\omega}{2k_B T} |J_{ab}(\omega)|^2 \quad (21)$$

The third-order response function is given by the four point generalization of eq 6 (eqs 5.22–5.23 in ref 45). Equation 5.23 in ref 45 neglects vibrational relaxation. This can be included by adding the Doorway–Window (DW) expressions⁶⁸ (eqs D6–D9 of section 5.2 in ref 45) which require the numerical solution of master equations (eq 5.31 in ref 45). The response functions were calculated by the procedure described in section IV.

VI. NEE Simulations of Multipoint Response Functions

The NEEs were derived for a vibrational Hamiltonian which conserves the number of excitons (see, e.g., eq 8). Response functions are solely expressed using the one-exciton states. Doubly excited resonances are obtained from the scattering of one-exciton states. The scattering is then the source of the nonlinear response (excitations in a linear system are noninteracting bosons and their response is strictly linear). The scattering matrix contains all necessary information regarding the two-exciton resonances, and the calculation of two-exciton eigenstates is avoided.

The NEE response is expressed in terms of one-exciton Greens functions, which represent the time evolution of coherences between the ground and the one-exciton states. The general Green function expressions for the response are given in ref 68. The simplified expressions given below neglect

population relaxation and therefore hold only for short second time delays t_2 . The \mathbf{k}_I signal is

$$\mathcal{J}_{\nu_4\nu_3\nu_2\nu_1}^{\mathbf{k}_I}(\Omega_3, t_2 = 0, \Omega_1) = 2i \sum_{e_4 \dots e_1} \langle \mu_{e_4}^{\nu_4} \mu_{e_3}^{\nu_3} \mu_{e_2}^{\nu_2} \mu_{e_1}^{\nu_1} \rangle_0 \times \\ I_{e_1}^*(-\Omega_1) I_{e_4}(\Omega_3) \Gamma_{e_4 e_1 e_3 e_2}(\Omega_3 + \epsilon_{e_1} + i\gamma_{e_1}) \mathcal{J}_{e_3 e_2}(\Omega_3 + \epsilon_{e_1} + i\gamma_{e_1}) \quad (22)$$

and for the \mathbf{k}_{III} signal we get

$$\mathcal{J}_{\nu_4\nu_3\nu_2\nu_1}^{\mathbf{k}_{III}}(\Omega_3, \Omega_2, t_1 = 0) = 2i \sum_{e_4 \dots e_1} \langle \mu_{e_4}^{\nu_4} \mu_{e_3}^{\nu_3} \mu_{e_2}^{\nu_2} \mu_{e_1}^{\nu_1} \rangle_0 \times \\ I_{e_4}(\Omega_3) I_{e_3}^*(\Omega_2 - \Omega_3) [\Gamma_{e_4 e_3 e_2 e_1}(\Omega_2) \mathcal{J}_{e_2 e_1}(\Omega_2) - \Gamma_{e_4 e_3 e_2 e_1}(\Omega_3 + \epsilon_{e_3} + i\gamma_{e_3}) \mathcal{J}_{e_2 e_1}(\Omega_3 + \epsilon_{e_3} + i\gamma_{e_3})] \quad (23)$$

where

$$I_e(\Omega) = \frac{i}{\Omega - \epsilon_e + i\gamma_e} \quad (24)$$

is the one-exciton Green's function and

$$\mathcal{J}_{ee'}(\omega) = \frac{i}{\omega - \epsilon_e - \epsilon_{e'} + i(\gamma_e + \gamma_{e'})} \quad (25)$$

is the two-exciton Green's function.

The broadening γ_e can be obtained either from simulations or from experiment.

The exciton scattering matrix $\Gamma_{e_4 e_3 e_2 e_1}(\Omega)^{64}$ expresses the two-exciton resonances in terms of one-exciton Green's functions. The four indices represent the two incoming (e_1 and e_2) and two outgoing (e_3 and e_4) excitons (we denote this the scattering configuration). The matrix is frequency dependent and shows the two-exciton resonances. The magnitude of the scattering matrix for some scattering configuration reflects its contribution to the spectrum. We note that $\Gamma_{e_4 e_3 e_2 e_1} = \Gamma_{e_4 e_3 e_1 e_2} = \Gamma_{e_3 e_4 e_2 e_1}$, due to permutation symmetry of excitons.

The apparent $\sim N^4$ scaling of the computational effort for NEE is the same as in the SOS, however, the number of terms can be greatly reduced when excitons are localized and the anharmonicities are local, resulting in a much more favorable scaling. To that end, we define the exciton overlap factor

$$\eta_{ee'}^{(1)} = \sum_m |\psi_{e,m}| |\psi_{e',m}| \quad (26)$$

This parameter relates two excitons in real space: For $e = e'$, we have $\eta_{ee'}^{(1)} \equiv 1$. For uncoupled chromophores, $J_{mm} = 0$ and $\eta_{ee'}^{(1)} = \delta_{ee'}$, indicating that the excitons do not interact. Assuming that the excitons only scatter provided they spatially overlap, we can quantify the probability of this event by assuming that exciton pairs ($e_1 e_2$) scatter only provided their overlap is larger than a certain cutoff η_C , $\eta_{e_1 e_2}^{(1)} > \eta_C$. This criterion may also be used for pairs of final outgoing states ($e_4 e_3$). It restricts the distance between two initial excitons and between two final excitons in the scattering matrix due to the local nature of the anharmonicity. When this cutoff is applied, the number of relevant scattering matrix elements should scale as N^2 rather than N^4 .

An additional constraint comes from the exciton-exciton scattering radius. This is related to how far two excitons can travel during their interaction and restricts the distance between

initial and final excitons $e_3 e_2$. We introduce a second overlap parameter

$$\eta_{ee'}^{(2)} = \sum_{e_1, m, n} |\psi_{e_1, m}| |\psi_{e_1, n}| |\psi_{e, m}| |\psi_{e', n}| \quad (27)$$

$\eta^{(2)}$ is the amplitude of a path going from e to e' through all possible intermediate states e_1 . We will use $\eta_{ee'}^{(2)}$ to select the dominant $e_3 e_2$ pairs in the scattering matrix.

Using the cutoff parameters $\eta_C^{(1)}$ and $\eta_C^{(2)}$, we retained only those scattering matrix elements $\Gamma_{e_4 e_3 e_2 e_1}$ which satisfy $\eta_{e_2 e_1}^{(1)} > \eta_C^{(1)}$, $\eta_{e_4 e_3}^{(1)} > \eta_C^{(1)}$, $\eta_{e_3 e_2}^{(1)} > \eta_C^{(2)}$, $\eta_{e_3 e_1}^{(1)} > \eta_C^{(2)}$, and $\eta_{e_4 e_1}^{(1)} > \eta_C^{(2)}$. The scaling of the NEE effort with system size reduces to $\sim N$ for large systems with localized excitons. The efficiency of this truncation stems from two reasons: (i) We can identify the important exciton states *before* calculating the scattering matrix itself. Their number will be much smaller than N^4 . The scattering matrix is then calculated only for the selected set of scattering combinations. (ii) The numerical effort required for calculating the signal using multiple summations is greatly reduced when the scattering matrix is sparse.

A different truncation procedure can be obtained by calculating the complete scattering matrix (which scales as $\sim N^4$) and neglecting all matrix elements smaller than some cutoff $\Gamma^{(C)}$. This allows us to identify the important scattering configurations. The scattering matrix values for all relevant frequencies must be added for each selected scattering configuration. This method is more straightforward to apply and does not require physical arguments regarding exciton localizations in space. However, unlike the above truncation, it still requires the calculation of the entire scattering matrix. Both truncation schemes should be optimized for specific applications.

VII. Vibrational Response of NMA

The amide I band of NMA has one fundamental and one overtone frequency and no coupling. The NMA structure obtained from ref 51 was placed in a simulation box with 1000 water molecules. The CHARMM package⁴⁴ was then used to carry out the MD simulation with the NMA molecule constrained as a rigid body using the "SHAPE" command.⁴⁴ The Charmm27 force field⁴⁴ is used for NMA with TIP3 water. A 4 fs time step 1 ns trajectory was then generated. All water molecules were included in the bath. The ground state (0 cm^{-1}), the fundamental ($\sim 1650 \text{ cm}^{-1}$), and the overtone ($\sim 3300 \text{ cm}^{-1}$) are well separated. The frequency fluctuation ($\sim 15 \text{ cm}^{-1}$) is much smaller than these frequencies, making CGF the method of choice. Interestingly, we found that the fluctuations of the amide I mode anharmonicity and its fundamental frequency are uncorrelated.⁵¹

Figure 6 shows the simulated CGF linear absorption spectra of the amide I band. Neglecting vibrational relaxation (green line), the fwhm is 19 cm^{-1} , after adding the experimental lifetime broadening 450 fs⁶⁵ (red line), the fwhm becomes 30 cm^{-1} , which is remarkably close to the experiment (29 cm^{-1}).

The simulated \mathbf{k}_I spectra are presented in Figure 7 both excluding (upper panel) and including (middle panel) vibrational relaxation. The finite vibrational lifetime is significant in this case. For comparison, we have carried out an NEE simulation. Both linear absorption (Figure 6, black) and the \mathbf{k}_I signal (Figure 7, lower panel) are broader than the CGF. The inhomogeneous averaging protocol assumes that the energy fluctuations are much slower than their inverse magnitude. Obviously, this assumption is not justified for NMA where motional narrowing, included in the CGF, should be taken into account.

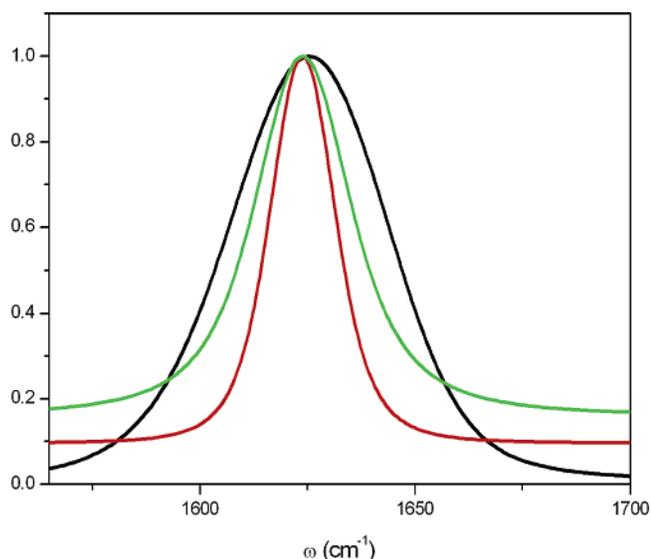


Figure 6. Simulated absorption spectra of the amide I band of NMA. CGF, excluding lifetime broadening (green), fwhm (19 cm^{-1}), with lifetime broadening (red), the fwhm becomes 30 cm^{-1} , which is very close to the experiment (29 cm^{-1}).⁶⁵ The inhomogeneous NEE simulated signal (black) overestimates the line width (52 cm^{-1}).

VIII. Vibrational Response of YKKKH17

Our MD simulation of the α -helical peptide used the velocity verlet algorithm and includes all the atoms of one YKKKH17 molecule and 4330 water molecules. Four chloride ions were added into the simulation box to make the system neutral. The initial YKKKH17 structure was obtained from the maestro package.⁶⁹ The Charmm27 force field⁴⁴ was employed for all interactions with a cutoff of 12 \AA for the nonbonded interactions. Long-range electrostatic interactions were calculated using the Ewald sum.⁷⁰ The simulation was carried using the CHARMM package.⁴⁴ The structure was first refined in a vacuum using a 4000 step energy minimization procedure with the adopted basis Newton–Raphson method (ABNR).⁴⁴ The molecule was then embedded in a cubic unit cell of TIP3 water with box length of 52 \AA . To release the internal tension, a 10 000 steps adopted basis Newton–Raphson energy minimization was performed.⁴⁴ The system was equilibrated under NPT ensemble⁷¹ with 1 fs time step for 2 ns to obtain the right system density and box size, the extended system method was used to keep the temperature and pressure constant, and the final box length was 50.19 \AA . This was followed by an NVE equilibration with 2 fs time step for 10 ns. After the equilibration phase, a 1 ns trajectory was recorded by applying the NVE ensemble with a 1 fs time step. The structure was saved in 4 fs increments, giving a total of 2.5×10^5 sample points. The peptide unit ($\text{O}=\text{C}=\text{N}-\text{H}$) is first aligned into the geometry optimized NMA structure of.⁵¹

The fluctuating vibrational Hamiltonian was created for all snapshots along the trajectory. All water molecules were included in the bath.

In the NEE simulations, 100 snapshots with 20 ps time intervals are selected from the 2 ns trajectory and used for the inhomogeneous averaging (sampling was tested using a 4 ns trajectory which gave very similar inhomogeneously averaged linear absorption). A 5.5 cm^{-1} homogeneous dephasing rate γ was added.⁷² Unlike NMA, the inhomogeneously averaged linear absorption is much broader and is not sensitive to the precise value of γ . The simulated linear absorption is compared with experiment⁷³ in Figure 8. Both the line width and the weak shoulder at 1655 cm^{-1} are well reproduced.

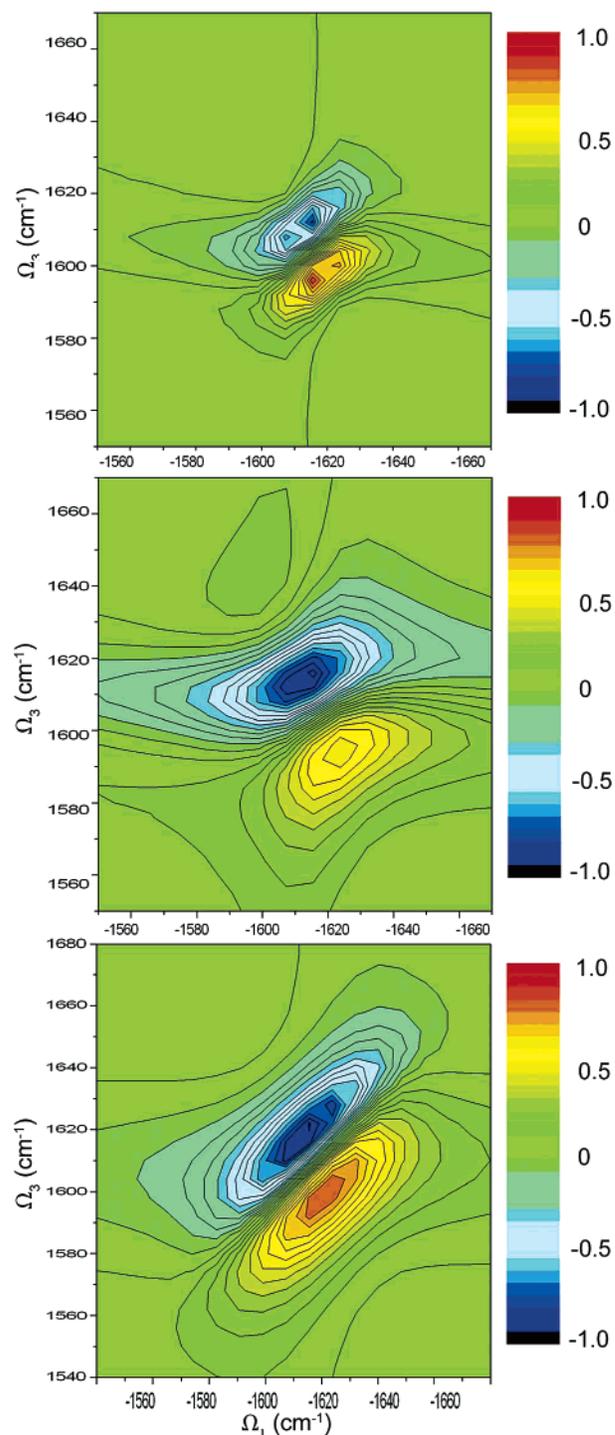


Figure 7. CGF simulated \mathbf{k}_I spectrum of NMA (imaginary plot) without (top) and with (middle) vibrational relaxation. NEE simulation (bottom).

The bottom panel of Figure 9 gives the distribution of eigenstate energies (density of states). A comparison with the linear absorption spectrum (top) shows that the mid-band states have the largest transition dipoles.

The simulated \mathbf{k}_I spectrum presented in the left column of Figure 10 is elongated along the diagonal, this is most clearly seen in the imaginary part (left middle); the diagonal (anti-diagonal) fwhm of the absolute magnitude of the signal (left top) is around 40 cm^{-1} (25 cm^{-1}). This is close to the experimental observation in the amide I spectra of a similar 25 residue α -helical peptide.⁷² The simulated \mathbf{k}_{III} spectrum is presented in the right column. Taking $\Omega_2 = 1/2\Omega_1$ as the

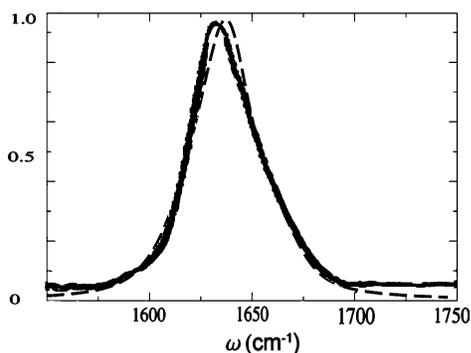


Figure 8. NEE simulated (dashed line) and experimental⁷³ (solid line) absorption spectrum of YK K K K H 17. The experimental (simulation) peak position is 1633 cm⁻¹ (1643 cm⁻¹), fwhm 36 cm⁻¹ (36 cm⁻¹). The simulated peak was red shifted by 6 cm⁻¹ for a better comparison of the line shape.

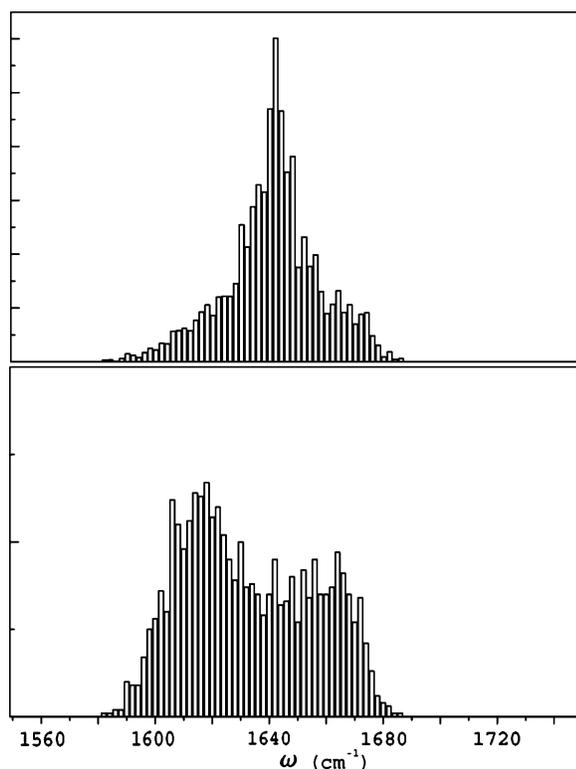


Figure 9. Absorption spectrum (top) and the density of states (bottom) of YK K K K H 17. Comparing the two panels shows that the mid-band states have the strongest oscillator strength.

diagonal line, we obtain ~ 40 cm⁻¹ for the diagonal and “antidiagonal” width in the amplitude plot (right top). \mathbf{k}_I is expected to have a smaller antidiagonal width, since it eliminates some inhomogeneous broadening; this is one major advantage of \mathbf{k}_I , known as the photon echo. \mathbf{k}_{III} spreads the signal along Ω in a region twice as large as that in \mathbf{k}_I which improves the resolution.

A nonfluctuating experimental anharmonicity -16 cm⁻¹⁷² has been assumed in many earlier studies of amide I spectra. To estimate the role of anharmonicity fluctuations, we have calculated the relative difference (RD), defined as the difference between the fluctuating anharmonicity and fixed-anharmonicity signals divided by the fixed-anharmonicity signals. The largest RD is 2% for \mathbf{k}_I and 5% for \mathbf{k}_{III} , indicating that \mathbf{k}_{III} is more sensitive to anharmonicity fluctuations.

Isotopic substitutions are commonly used in the interpretation of vibrational spectra and their relationship to structure.^{72–74}

Of particular interest to coherent IR applications is the use of isotopes to shift frequencies into regions where the behavior of that local unit can be measured, free from interference with other modes. For the amide I mode, the approximate -67 cm⁻¹ shift by ¹³C=¹⁸O substitution is large enough to displace the substituted amide group frequencies beyond the range of the natural distribution of frequencies found in most secondary structures. Isolated bands can also be obtained by using other chromophores. For example, the nitrile group (C≡N) attached to the side chains provides a distinct band at ~ 2235 cm⁻¹. It can be used as IR environmental probes, due to the sensitivity of the CN stretching vibration to hydration as well as other factors.⁷⁵

The participation ratio⁷⁶ of the exciton state e

$$W_e \equiv \frac{1}{\sum_m |\Psi_{em}|^4} \quad (28)$$

provides a convenient measure for exciton localization which can be used to quantify how much the isotope band is separated from the main band. For a localized exciton, $W_e = 1$, whereas when e is completely delocalized, and has equal contributions from all the 17 local modes, $|\Psi_{em}|$ will equal $1/(17)^{1/2}$ and $W_e = 17$.

When the first residue is isotopically labeled, W_e of the lower frequency (1570 cm⁻¹) eigenstates for 1000 configurations has an average value of 1.078 with a standard deviation 0.067. $|\Psi_{e1}|^2$ has an average value of 0.96, with standard deviation 0.003, indicating that the dominant contribution for this state is from the isotopically labeled unit. A similar trend is seen when the 10th residue is isotopically labeled. Figure 11 shows the isotope region of the inhomogeneously averaged simulated linear absorption of the full helix where the first unit is isotopically labeled (red, the isotope peak can be fitted to a Gaussian (black dash) with 28.5 cm⁻¹ fwhm), the NEE simulated linear absorption of the labeled mode is given by the green line, with fwhm 26 cm⁻¹, and the CGF simulated linear absorption of isotope labeled mode is given by the blue line (14.3 cm⁻¹). Recent experiment on a similar helical peptide⁷² gives a 15.2 cm⁻¹ fwhm for the C¹³O¹⁸ labeled peak, in close agreement with the CGF simulation.

Figure 12 shows the isotope region of the NEE simulated \mathbf{k}_I signal of the full helix (top panel) and of the isotopically labeled mode (NEE, middle and CGF, bottom). Because the isotope band is spectrally isolated, the NEE simulated isotope mode signal mainly reproduces the width and shape of the isotope region of the simulated full helix spectrum, while the CGF simulated isotope mode signal is much narrower (motional narrowing).

Isotope labeling also affects the main peak line shape, in a way that depends on the labeled site. In the upper panel of Figure 13, we show the linear absorption of native, first residue labeled (green) and 10th residue labeled (red) helices simulated by NEE. Labeling the 10th residue significantly changes the line shape, while labeling the first hardly affected it. This is more clearly demonstrated using the relative difference of linear absorption between the isotopomer and the unlabeled helix, shown in the lower panel of Figure 13.

IX. Discussion

We have developed first-principles protocols for simulating the coherent IR spectra of the amide I band of peptides. Two methods, the CGF and NEE, which make different approxima-

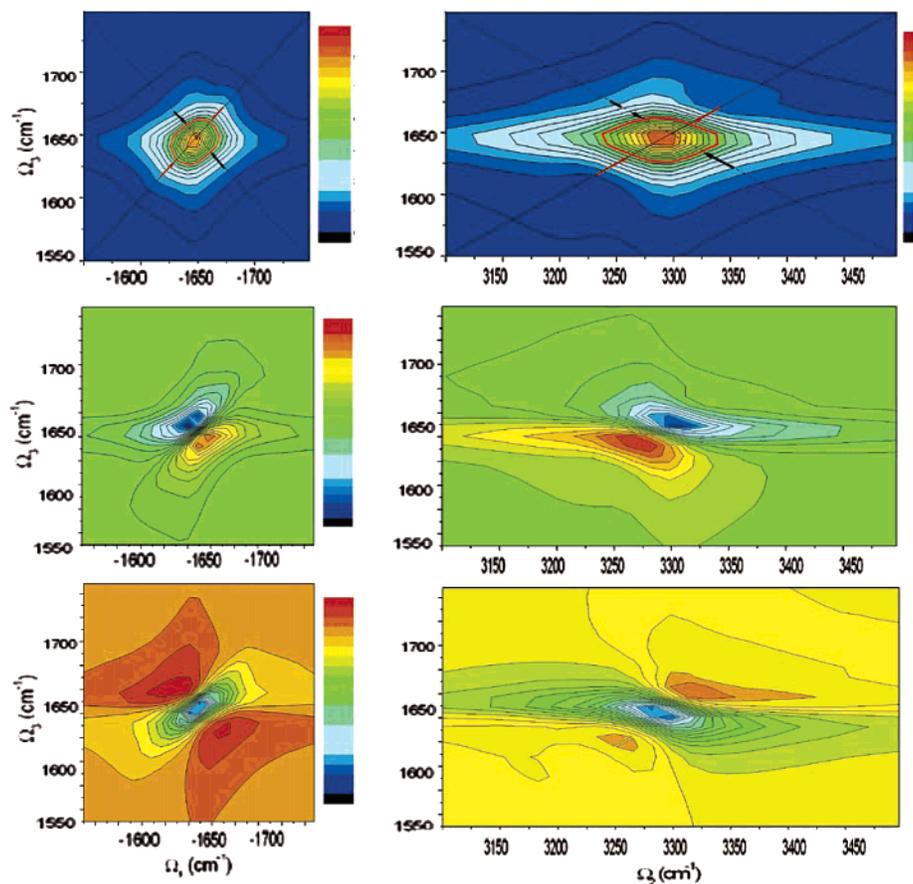


Figure 10. Left column, the NEE simulated \mathbf{k}_I spectra of YKKKH17 (eq 3): (top) absolute magnitude, the diagonal (solid) and anti-diagonal (dashed) lines are marked. The half-maxima contour is shown in red. Black (red) arrows mark the diagonal (anti-diagonal) width. The middle represents the imaginary part and the bottom the real part. The right column repeats these calculations for \mathbf{k}_{III} .

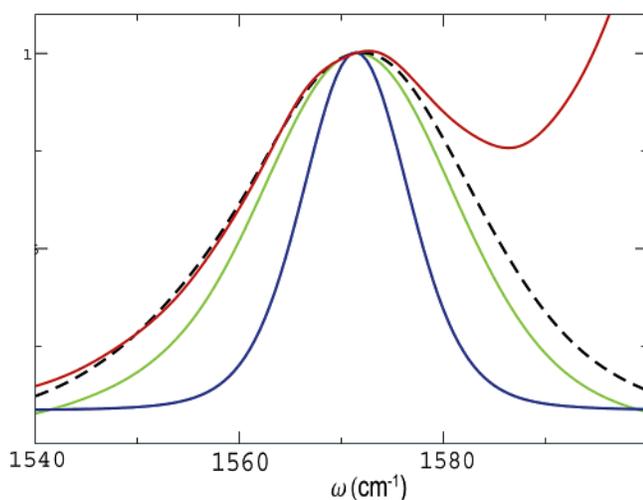


Figure 11. Isotope region of the NEE simulated linear absorption of the full YKKKH17 helix (red line) and the isotope peak fitted to a Gaussian (black dashed line) with 28.5 cm^{-1} fwhm. The NEE simulated linear absorption of the isotope labeled mode is represented by the green line (fwhm 26 cm^{-1}), and the CGF simulated linear absorption of the isotope labeled mode is represented by the blue line (14 cm^{-1}). The NEE simulated isotope mode reproduces the width and shape of the isotope region of the simulated full helix spectrum. The CGF simulated isotope mode signal is much narrower.

tions and reproduce a broad range of properties were compared. These are the first attempts to systematically address the issue of simulating the coherent nonlinear signals in large peptides.

The NEE approach only uses the one-exciton basis and attributes the nonlinear response to exciton–exciton scattering.

Homogeneous broadening is explicitly incorporated in the equations of motion, and inhomogeneous contributions are added by ensemble averaging. Classifying the broadening mechanisms as either homogeneous or inhomogeneous is not always possible and does not apply for intermediate fluctuation time scales as required for short peptides. The complete scattering matrix size and the computational effort scales to the fourth power in size. Predicting the dominant scattering configurations ($e_4e_3e_2e_1$) in the scattering matrix based on the exciton overlap parameters $\eta^{(2)}$ and $\eta^{(1)}$ leads to a truncation scheme based on the assumption that excitons cannot scatter unless they spatially overlap. This truncation applies for anharmonicities with a finite interaction range. It identifies the possible scattering configurations *before* calculating the scattering matrix itself. The number of scattering configurations and therefore the size of the effective scattering matrix are greatly reduced. Linear scaling of the simulation cost with system size for systems larger than the exciton coherence size makes them particularly attractive for large peptides.

We next illustrate the reduction in computational cost obtained using $\eta_{ee'}^{(1)}$ in eq 26. Figure 14 compares the calculated \mathbf{k}_I spectrum with all scattering configurations (left), $\eta^{(1)} = 0.3$ (middle), and $\eta^{(1)} = 0.5$ truncation (right) (we used $\eta_{ee'}^{(1)} = \eta_{ee'}^{(2)}$). The red circles in the top plot give the half-maximum contour. The left spectrum took 35 h on an AMD Opteron 244 CPU, compared with 2 h for the middle spectrum. The right spectrum (20 min) reproduces the main pattern. The diagonal and anti-diagonal fwhm's obtained from the truncated schemes coincide with the complete scattering matrix, however, some

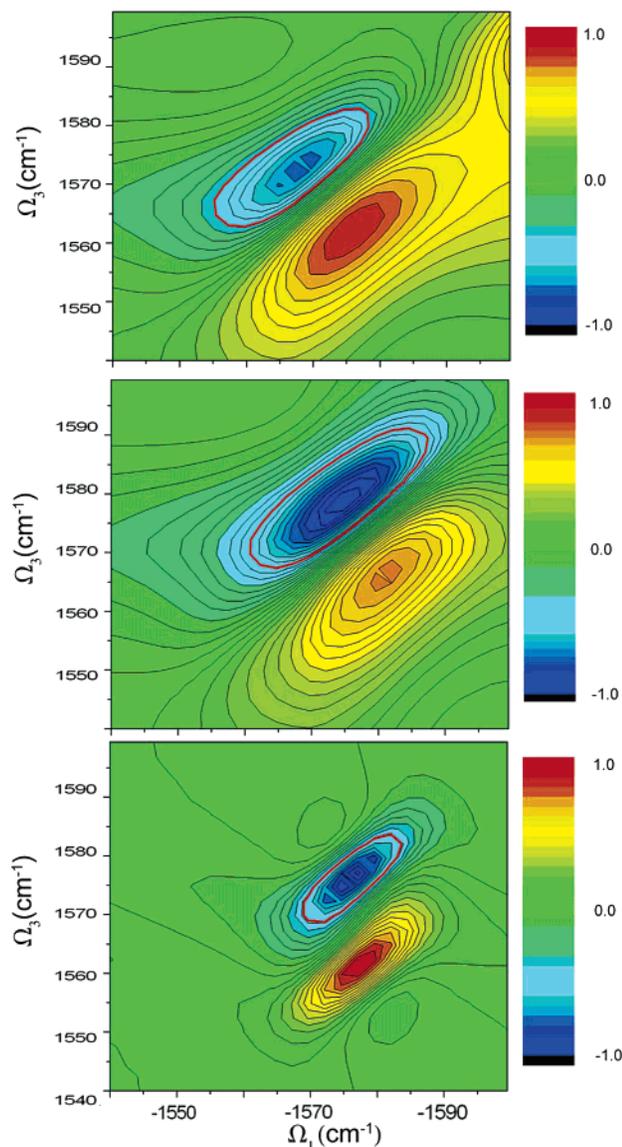


Figure 12. Isotope region of the NEE simulated k_1 signal (imaginary plot) of the full YKKKH17 helix (upper panel). Simulation of the isotopically labeled mode alone (NEE, middle panel and CGF, lower panel).

of the details are lost. The effect of this favorable scaling should be more dramatic for larger proteins.

This truncation cannot be applied in SOS calculations which require the complete set of two-exciton states. Multiple interferences between different Liouville space pathways then determine the contributions of the various eigenstates. This is especially important for weakly anharmonic systems. The nonlinear response of the system of harmonic oscillators vanishes identically. In the NEE, this interference is naturally built in whereas in the SOS it is obtained only when all possible contributions are combined. The SOS simulation time including numerical matrix diagonalization scales as $\sim N^3$ with system size.

Periodic systems (such as J aggregates and molecular crystals) have strict selection rules. A few transition dipoles may be very large, and many dark states can be eliminated before calculating the signals. This could be used to speed up both NEE and SOS simulation. However, the spectra of disordered proteins have no obvious selection rules; they depend on many contributions with similar transition dipole amplitudes. The CGF, which is based on the SOS, describes fluctuations with arbitrary time scales and relies on the adiabatic approximation of energy levels.

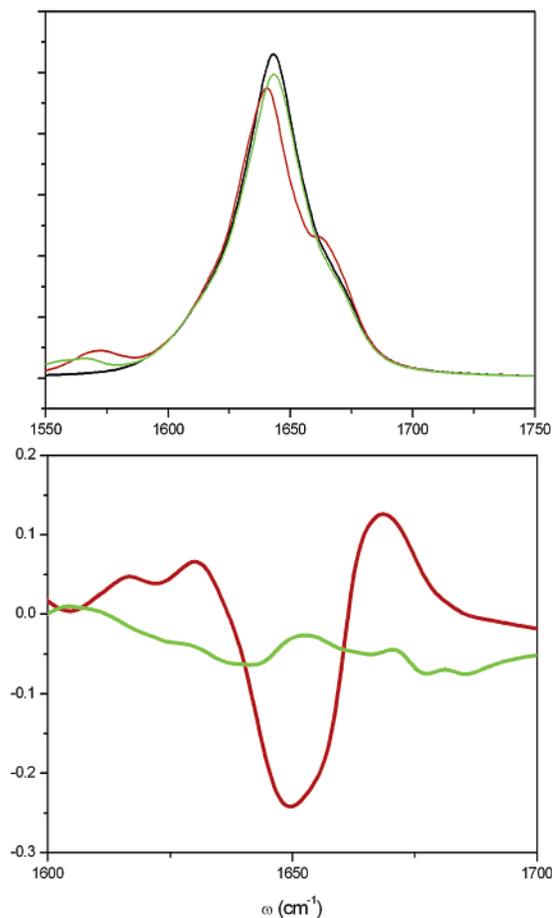


Figure 13. (top) NEE simulated absorption spectrum of the native helix (black) and the central 10th residue $C^{13}O^{18}$ isotope labeled (red) and first residue labeled (green) helices. The isotopic effect is simulated by adding a -67 cm^{-1} shift to the local frequency $\epsilon_m(Q)$ in eq 8. (bottom) The relative difference of linear absorption signal between isotopomer and native helix. The 10th residue is labeled red, and first residue is labeled green. Labeling the central residues has a stronger effect on the line shape.

It is, thus, most suitable for small localized modes such as in isotope substitutions and artificially implemented chromophores such as nitrile groups.⁷⁵

Isotope substitutions in vibrational spectroscopy provide an ideal local probe for structure and dynamics. This is conceptually similar to mutations used extensively in studies of biomolecular complexes. By labeling specific sites, isotopes can be used to explore local amide environments and coupling dependence on distances and orientations of amide groups. The CGF is suitable for simulating systems with well-separated energy levels and reproduces the motional narrowing effect caused by fast fluctuations. Inhomogeneous averaging of slow bath motions can be combined with CGF for intermediate size systems. However, due to its $\sim N^3$ scaling, this procedure is expensive. A detailed analysis of the line widths as a function of residue along the helix and the variation of the motional narrowing with the environment will be of interest for a future study.

The zero-order frequencies in our Hamiltonian were obtained from an electrostatic DFT map, which describes the frequency shift of the amide I mode in response to a spatially nonuniform external electric field. This map which allows the simulations of any peptide and solvent neglects covalent interactions of the amide unit with the surrounding molecules. This may account for the lower accuracy of the frequency shift compared with the cluster methods (56 cm^{-1} vs $\sim 62 \text{ cm}^{-1}$, the experimental

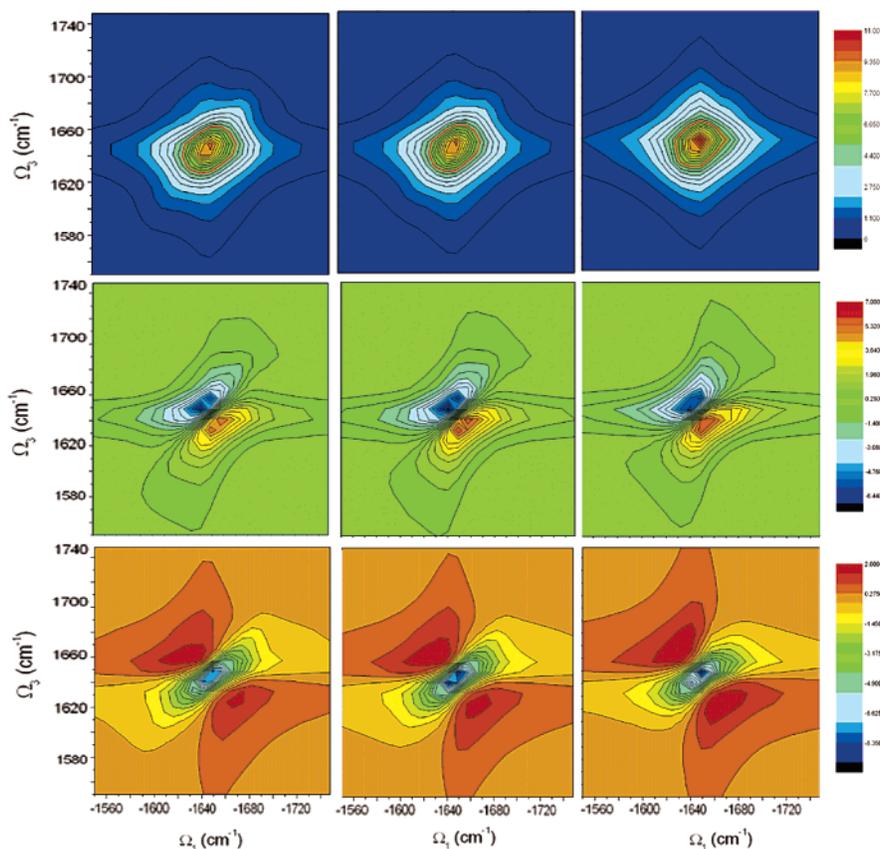


Figure 14. Simulated NEE \mathbf{k}_I spectra of YKKKH17 with the full scattering matrix (left) and with a truncated scattering matrix with $\eta_c = 0.3$ (middle) and $\eta_c = 0.5$ (right): top (absolute magnitude), middle (imaginary part), and bottom (real part).

value is 91 cm^{-1}). Nevertheless, the electrostatic fluctuations model seems to adequately reproduce the dominant features of the vibrational spectra. We have used a 19 component vector C^T to describe the electric field property at one amide unit. This vector is obtained by doing a least-squares fitting of the electric field components of 67 points in the TCD region. In a previous work,⁵¹ different sampling schemes were tested for the calculation of linear spectra of all amide bands (III, II, I, and A) of NMA. For amide I and A stretching motions, C^T can be obtained by fitting the electric field at the amide atom positions (C, O, N, and H), however, the global sampling were found to be crucial for the bending modes (III and II). Overall the global sampling works best.

Lifetime broadening is significant in NMA, however, CGF simulated isotope peaks which neglect lifetime broadening agree with experiment, implying that vibrational relaxation is less important in the vibrational line shapes of α -helical peptides. Predicting the 2DIR signatures of vibrational relaxation and transport behavior of different peptide structures is an interesting open problem. These are expected to show up in the t_2 variation of photon echo or pump probe experiments.

The present protocols should help the interpretation of nonlinear vibrational signals from a broad range of biological systems. The vibrational relaxation of N–H stretching in nucleic acids^{83–85} is believed to play an important role in the photo-physics of DNA and RNA. Time-resolved IR spectra can be simulated. Another promising example is carbohydrates, whose dynamics are faster than the time resolution of NMR but may be studied in the IR.⁸⁶ One potential application is to amyloid fibrils, which form in Alzheimer's disease and whose structure and dynamics are under extensive study. Recently, the fibrils have been crystalized and the structure has been resolved by X-ray.⁷⁷ However, the early stage of the fibril formation, during

which the randomly coiled or helical peptides collide in the solution, change into hairpins, start to form the oligomers, and finally become the fibrils,^{78–80} is still unresolved. Because of its sensitivity to H-bonding structure, backbone angles, and electrostatic environment, multidimensional IR spectroscopy should help resolve this problem. Recent modeling of the \mathbf{k}_I and \mathbf{k}_{III} spectra of ideal β -sheets and helices⁶⁴ show that by tuning the pulse polarizations, significant difference can be observed in the \mathbf{k}_{III} spectra between helices and sheets due to differences in local chirality. Time-domain chiral techniques can be used to detect the early dynamics of amyloid fibrils formation.

Most effort in IR spectroscopy of peptides has so far been focused on the amide I band and its cross-peaks. The cross-peak pattern of other amide bands such as amide II and A can provide additional structural information. A DFT map for all amide bands has been constructed.⁵¹

The CGF and NEE approaches were originally developed for electronic spectroscopy⁸⁷ and have been used to predict multidimensional electronic spectroscopy in conjugated polymers and photosynthetic aggregates. Electronic spectra are more complex than vibrational spectra, and their simulations pose additional challenges. 2D measurements of photosynthetic complexes have been reported.^{88,89}

Both CGF and NEE methods hold for equilibrated fluctuations or slow nonequilibrated adiabatic dynamics. For faster fluctuations, the direct simulation involves multiple-level crossings and becomes much more expensive. In this case, an alternative approach will be to include explicitly the relevant collective bath modes and work in an extended phase space using the stochastic Liouville equations.^{81,82}

Acknowledgment. This work was supported by the National Institutes of Health Grant 2RO1-GM59230-05 and the National

Science Foundation Grant CHE-0446555. We thank Professor Hajime Torii for providing numerical values of the ab initio map of amide I mode couplings. We thank Dr. Ravindra Venkatramani and Dr. Thomas la Cour Jansen for many helpful discussions. W.D. and D.A. contributed equally to this work.

References and Notes

- Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254* (5038), 1598–1603.
- Shakhnovich, E. I. *Curr. Opin. Struct. Biol.* **1997**, *7* (1), 29–40.
- Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652.
- McCammon, J.; Harvey, S. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, U.K., 1987.
- Onuchic, J. N.; Luthey Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48*, 545–600.
- Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G. *Proteins: Struct., Funct., Genet.* **1995**, *21* (3), 167–195.
- Wolynes, P. G.; Onuchic, J. N.; Thirumalai, D. *Science* **1995**, *267* (5204), 1619–1620.
- Kubelka, J.; Hofrichter, J.; Eaton, W. *Curr. Opin. Struct. Biol.* **2004**, *14*, 76–88.
- Ballow, R.; Sabelko, J.; Gruebele, M. *Nat. Struct. Biol.* **1996**, *3*, 923.
- Williams, S.; Causgrove, T.; Gilmanshin, R.; Fang, K.; Callender, R.; Woodruff, W.; Dyer, R. *Biochemistry* **1996**, *35*, 691.
- Munoz, V.; Thompson, P.; Hofrichter, J.; Eaton, W. *Nature* **1997**, *390*, 196.
- Werner, J.; Dyer, R.; Fesinmeyer, R.; Andersen, N. *J. Phys. Chem. B* **2002**, *106*, 487.
- Huang, C.; Getahun, Z.; Zhu, Y.; Klemke, J.; DeGrado, W.; Gai, F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99* (5), 2788.
- Duan, Y.; Kollman, P. *Science* **1998**, *282*, 740.
- Daura, X.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. *J. Mol. Biol.* **1998**, *280*, 925.
- Zhou, Y.; Karplus, M. *Nature* **1999**, *401*, 400.
- Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181.
- Decius, W. E. B.; Cross, J. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; McGraw-Hill: New York, 1955.
- Torii, H.; Tasumi, M. *J. Chem. Phys.* **1992**, *96*, 3379.
- Torii, H.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29* (1), 81–86.
- Baumruk, V.; Pancoska, P.; Keiderling, T. A. *J. Mol. Biol.* **1996**, *259* (4), 774–791.
- Cheatum, C. M.; Tokmakoff, A.; Knoester, J. *J. Chem. Phys.* **2004**, *120*, 8201.
- Huang, C. Y.; He, S.; DeGrado, W. F.; McCafferty, D. G.; Gai, F. *J. Am. Chem. Soc.* **2002**, *124* (43), 12674–12675.
- Mantsch, H. H.; Chapman, D., Eds. *Infrared Spectroscopy of Biomolecules*; Wiley-Liss: New York, 1996.
- Tanimura, Y.; Mukamel, S. *J. Chem. Phys.* **1993**, *99* (12), 9496–9511.
- Abramavicius, D.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. B* **2004**, *108*, 18034.
- Demirdoven, N.; Cheatum, C. M.; Chung, H. S.; Khalil, M.; Knoester, J.; Tokmakoff, A. *J. Am. Chem. Soc.* **2004**, *126* (25), 7981–7990.
- Zhuang, W.; Abramavicius, D.; Mukamel, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102* (21), 7443–7448.
- Mukamel, S. *Annu. Rev. Phys. Chem.* **2000**, *51*, 691–729.
- Bredenbeck, J.; Helbing, J.; Behrendt, R.; Renner, C.; Moroder, L.; Wachtveitl, J.; Hamm, P. *J. Phys. Chem. B* **2003**, *107* (33), 8654–8660.
- Asplund, M. C.; Zanni, M. T.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97* (15), 8219–8224.
- Golonzka, O.; Khalil, M.; Demirdoven, N.; Tokmakoff, A. *Phys. Rev. Lett.* **2001**, *86* (10), 2154–2157.
- Larsen, O.; Bodis, P.; Buma, W.; Hannam, J.; Leigh, D.; Woutersen, S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, in press.
- Ernst, R. R.; Bodenhausen, G.; Wokaun, A. *Principles of nuclear magnetic resonance in one and two dimensions*; Oxford University Press: New York, 1995.
- Wuthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley-Interscience: 1986.
- Henry, E. R.; Szabo, A. *J. Chem. Phys.* **1985**, *82* (11), 4753.
- Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433* (7022), 128–132.
- Tjandra, N.; Feller, S.; Paster, R.; Bax, A. *J. Am. Chem. Soc.* **1995**, *117*, 12562–12566.
- Scheurer, C.; Mukamel, S. *Bull. Chem. Soc. Jpn.* **2002**, *75* (5), 989–999.
- Scheurer, C.; Mukamel, S. *J. Chem. Phys.* **2002**, *116* (15), 6803–6816.
- Elber, R. *Curr. Opin. Struct. Biol.* **2005**, *15* (2), 151–156.
- Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151* (1), 283–312.
- Van gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem.* **1990**, *29* (9), 992–1023.
- Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4* (2), 187–217.
- Mukamel, S.; Abramavicius, D. *Chem. Rev.* **2004**, *104* (4), 2073–2098.
- Fried, L. E.; Mukamel, S. *Adv. Chem. Phys.* **1993**, *84*, 435.
- Chernyak, V.; Zhang, W. M.; Mukamel, S. *J. Chem. Phys.* **1998**, *109*, 9587.
- Piryatinski, A.; Chernyak, V.; Mukamel, S. *Chem. Phys.* **2001**, *266*, 285.
- Mukamel, S. In *Molecular Nonlinear Optics*; Zyss, J., Ed.; Academic Press: New York, 1994; pp 1–46.
- Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford University Press: Oxford, U.K., 1995.
- Hayashi, T.; Zhuang, W.; Mukamel, S. *J. Phys. Chem.*, in press.
- Peyerimhoff, S. *Spectroscopy: Computational Methods, Encyclopedia of Computational Chemistry*; John Wiley: New York, 1998; Vol. 1.
- Dreyer, J.; Moran, A. M.; Mukamel, S. *J. Phys. Chem. B* **2003**, *107* (24), 5967–5985.
- Kopelman, R. In *Excited States*; Lim, E., Ed.; Academic: New York, 1975.
- Scott, A. C. *Phys. Rep.* **1992**, *217*, 1.
- Kwac, K.; Cho, M. H. *J. Chem. Phys.* **2003**, *119* (4), 2247–2255.
- Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. *J. Chem. Phys.* **2004**, *121* (18), 8887–8896.
- Kubelka, J.; Keiderling, T. A. *J. Phys. Chem. A* **2001**, *105* (48), 10922–10928.
- Frisch, M. J.; et al. *Gaussian 03*, revision c.01; Technical report, 2003.
- Hayashi, T.; Jansen, T. L.; Zhuang, W.; Mukamel, S. *J. Phys. Chem. A* **2005**, *109* (1), 64–82.
- Guo, H.; Karplus, M. *J. Phys. Chem.* **1992**, *96* (18), 7273–7287.
- Torii, H.; Tatsumi, T.; Tasumi, M. *J. Raman Spectrosc.* **1998**, *29* (6), 537–546.
- Ham, S.; Kim, J. H.; Lee, H.; Cho, M. H. *J. Chem. Phys.* **2003**, *118* (8), 3491–3498.
- Abramavicius, D.; Mukamel, S. *J. Chem. Phys.* **2005**, *122* (13), 134101.
- Zanni, M. T.; Asplund, M. C.; Hochstrasser, R. M. *J. Chem. Phys.* **2001**, *114* (10), 4579–4590.
- Mukamel, S. *Phys. Rev. A* **1983**, *28*, 3480.
- Venkatramani, R.; Mukamel, S. *J. Chem. Phys.* **2002**, *117* (24), 11089–11101.
- Zhang, W. M.; Meier, T.; Chernyak, V.; Mukamel, S. *J. Chem. Phys.* **1998**, *108* (18), 7763–7774.
- Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Cauffield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11* (4), 440–467.
- Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593.
- Hoover, W. G. *Phys. Rev. A* **1985**, *31* (3), 1695–1697.
- Fang, C.; Wang, J.; Kim, Y. S.; Charnley, A. K.; Barber-Armstrong, W.; Smith, A. B.; Decatur, S. M.; Hochstrasser, R. M. *J. Phys. Chem. B* **2004**, *108* (29), 10415–10427.
- Decatur, S. M.; Antonic, J. *J. Am. Chem. Soc.* **1999**, *121*, 11914–11915.
- Moran, A. M.; Park, S. M.; Dreyer, J.; Mukamel, S. *J. Chem. Phys.* **2003**, *118* (8), 3651–3659.
- Tucker, M. J.; Getahun, Z.; Nanda, V.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2004**, *126* (16), 5078–5079.
- Thouless, D. *Phys. Rep.* **1974**, *13*, 93.
- Ritter, C.; Maddelein, M. L.; Siemer, A. B.; Luhrs, T.; Ernst, M.; Meier, B. H.; Saupé, S. J.; Riek, R. *Nature* **2005**, *435* (7043), 844–848.
- Kirschner, D. A.; Inouye, H.; Duffy, L. K.; Sinclair, A.; Lind, M.; Selkoe, D. J. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84* (19), 6953–6957.
- Inouye, H.; Fraser, P. E.; Kirschner, D. A. *Biophys. J.* **1993**, *64* (2), 502–519.
- Serpell, L. C.; Blake, C. C. F.; Fraser, P. E. *Biochemistry* **2000**, *39* (43), 13269–13275.
- Jansen, T. I. C.; Zhuang, W.; Mukamel, S. *J. Chem. Phys.* **2004**, *121*, 10577–10598.
- Gamliel, D.; Levanon, H. *Stochastic Processes in Magnetic Resonance*; World Scientific: River Edge, NJ, 1995.
- Woutersen, S.; Cristalli, G. *J. Chem. Phys.* **2004**, *121* (11), 5381–5386.

- (84) Kyogoku, Y.; Lord, R. C.; Rich, A. *Science* **1966**, *154* (3748), 518–532.
(85) Hamlin, R. M.; Lord, R. C.; Rich, A. *Science* **1965**, *148* (3678), 1734–1748.
(86) Liquier, J.; Letellier, R.; Dagneaux, C.; Ouali, M.; Morvan, F.; Raynier, B.; Imbach, J. L.; Taillandier, E. *Biochemistry* **1993**, *32*, 2 (40), 10591–10598.

- (87) Chernyak, V.; Wang, N.; Mukamel, S. *Phys. Rep.* **1995**, *263*, 213–309.
(88) Brixner, T.; Stenger, J.; Vaswani, H. M.; Cho, M.; Blankenship, R. E.; Fleming, G. R. *Nature* **2005**, *434* (7033), 625–628.
(89) Sundstrom, V.; Pullerits, T.; van Grondelle, R. *J. Phys. Chem. B* **1999**, *103* (13), 2327–2346.