Published in final edited form as:

J Phys Chem B. 2011 April 7; 115(13): 3713–3724. doi:10.1021/jp200745r.

Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides

L. Wang, C. T. Middleton, M. T. Zanni, and J. L. Skinner*

Department of Chemistry, University of Wisconsin, Madison, Wisconsin 53706, United States

Abstract

Infrared (IR) spectroscopy of the amide I band has been widely utilized for the analysis of peptides and proteins. Theoretical modeling of IR spectra of proteins requires an accurate and efficient description of the amide I frequencies. In this paper, amide I frequency maps for protein backbone and side chain groups are developed from experimental spectra and vibrational lifetimes of *N*-methylacetamide and acetamide in different solvents. The frequency maps, along with established nearest-neighbor frequency shift and coupling schemes, are then applied to a variety of peptides in aqueous solution and reproduce experimental spectra well. The frequency maps are designed to be transferable to different environments; therefore, they can be used for heterogeneous systems, such as membrane proteins.

I. INTRODUCTION

Proteins play essential roles in biological processes, functioning as structural supporters, catalysts, signal transmitters, and so on.^{1–3} Their biological functions are determined by their three-dimensional structures, which undergo constant motions over a wide range of time and length scales.^{4,5} The structure and dynamics of proteins can be probed by a variety of experimental techniques. X-ray crystallography is used to determine the complete three-dimensional structure of proteins at high resolution,^{6,7} as long as the proteins can be crystallized. Nuclear magnetic resonance spectroscopy measures protein structure in solution at the atomic level and has greatly enriched our understanding of how proteins fold and interact with other molecules.^{5,8} Other techniques, such as Raman, circular dichroism, electron paramagnetic resonance, and ultraviolet absorption and fluorescence spectroscopy are also commonly used for the analysis of proteins.^{9–13}

Infrared (IR) spectroscopy is another sensitive and versatile technique for studying the structure and dynamics of proteins under a variety of sample conditions, such as in aqueous solution or lipid membranes.^{14,15} In IR experiments, the amide I vibrational mode, primarily associated with the peptide bond carbonyl stretch,¹⁴ is often utilized for structural analysis. Each local amide I mode is called a chromophore, and the many chromophores in a peptide or protein couple together to form the amide I band. The frequencies at which the amide I band occur, along with the band widths and intensities, depend on different patterns of intra- and intermolecular couplings. Thus, different protein secondary structures have characteristic spectral features in the amide I region (1600–1700 cm⁻¹).^{16–18} For example, α -helices absorb near 1650 cm⁻¹, whereas β -sheets have two bands at ~1620 and 1675 cm⁻¹. Recent development of time-resolved two-dimensional infrared (2DIR) spectroscopy has enabled the measurement of molecular dynamics over picosecond time scales and also provides a more detailed picture of the couplings among chromophores.^{19–25} Moreover, residue-specific information can be revealed by the isotope-labeling technique. For example, ¹³C and ¹³C=¹⁸O isotope labels lower the amide I frequency by ~40 and 70 cm⁻¹, respectively.^{26–40} Carbonyls with isotope labels are essentially shifted out of the main amide I band and can be resolved individually.

The detailed interpretation of experimental spectra is facilitated by theoretical calculations. Modern calculations of vibrational spectra for the amide I region of proteins in complex environments treat the amide stretches quantum mechanically and all other nuclear degrees of freedom, classically.^{41–77} For IR absorption spectra, which involve only one quantum of amide vibrational excitation, a central quantity is the one-exciton Hamiltonian and its dependence on the classical degrees of freedom (which evolve in time according to a classical molecular dynamics (MD) simulation). The diagonal terms of this Hamiltonian correspond to the transition frequencies of the local amide chromophores, whereas the off-diagonal terms involve the couplings between pairs of chromophores.

For each configuration of the classical coordinates (each time step in the MD simulation), the exciton Hamiltonian can, in principle, be determined from ab initio electronic structure calculations,^{78–80} but in practice, it is impossible to perform accurate enough calculations for peptides in solution or membranes for the very large number of configurations required for computing line shapes. This has led to the development of ab initio-based coupling and frequency “maps”. These are most conveniently discussed by first considering model peptides in the gas phase. Electronic structure calculations are performed for a grid of the relevant dihedral angles. For each set of dihedral angles, the exciton Hamiltonian is determined from the normal modes through the Hessian matrix reconstruction approach.^{42,45,47,50,52,67,72,81,82} This then leads to nearest-neighbor frequency shift (NNFS) maps,⁶⁷ which show explicitly how the local chromophore’s frequency depends on the neighboring dihedral angles. It also leads to coupling maps for both nearest neighbors (the nearest-neighbor coupling (NNC) maps)^{42,67} and chromophores that are more distant (through the transition dipole coupling (TDC) approach).^{14,41,42,45,62,64,67,73,74,83–86}

The second step is to include the effects of environmental perturbations, from water, ions, lipids, etc. on the local chromophore frequencies. These are also conveniently expressed as frequency maps.^{46,49,55–58,60,64,66,68,69} Numerous maps have been developed for the amide I vibrational frequencies of protein backbone chromophores in condensed phase, mostly by considering *N*-methylacetamide (NMA), a molecule with a single amide I chromophore, in water. Experimentally, heavy water (D₂O) is commonly used instead of water to avoid the interference of the water bend mode (~1640 cm⁻¹) with the amide I band. The major isotopic form of NMA in D₂O is *N*-deuterated NMA (NMAD), as shown in Figure 1a. By performing electronic structure calculations on clusters of NMAD surrounded by D₂O, frequency maps have been developed that relate the amide I frequency to the electric

potential^{46,49,55,56,64,66} or electric field^{57,58,60,68,69} on different atom sites from the surrounding solvent molecules.

One example of a field map was developed by Schmidt et al.⁵⁷ In their development, classical MD simulations of NMAD in D₂O were performed, and representative NMAD/D₂O clusters were extracted. For each cluster, ab initio calculations were carried out to obtain the amide I frequency of NMAD. The frequency was assumed to be linear in the electric fields on the NMAD atoms, and the coefficients were determined by multiple linear regression to the ab initio frequencies. The resulting frequency map includes electric fields on the C, O, N, and D atoms, with the largest coefficients on the C and N atoms in the C=O bond direction. Following the same procedure, Lin et al.⁵⁸ simplified the map by fixing the zero-field intercept at the gas-phase absorption frequency of NMAD and incorporating electrostatic contributions only on the C and N atoms in the C=O bond direction. This map is termed as S2 in the following. Both maps reproduced the experimental IR absorption spectrum of NMAD in D₂O fairly well.

Nonetheless, discrepancies between theoretical and experimental spectra reveal problems associated with ab initio-based maps. For example, the level of theory and size of basis set will influence the accuracy of the maps, and the cluster sizes chosen are rather limited and do not account for the entire solvent effect. To circumvent these problems, Lin et al.⁵⁸ switched to an empirical approach. Assuming the same form of S2, they reoptimized the coefficients using the experimental spectrum of NMAD in D₂O. This empirical map was applied to a transmembrane peptide, CD3 ζ ,⁵⁸ the M2 proton channel of the influenza A virus,³⁹ and an antimicrobial peptide ovispirin⁴⁰ and reproduced the experimental 2DIR diagonal line widths quite well. This approach is appealing in that although its simple functional form is based on ab initio calculations, the parameters are chosen to ensure that it is the best frequency map compared with experiment for a given force field. A drawback of this map is that it is not necessarily transferable to different environments, mainly because it was parametrized in a single solvent (D₂O).

To ameliorate this situation, we present here a new empirical frequency map for protein backbone modes. Again assuming the S2 form, but relaxing the constraint that the zero-field intercept reproduce the gas-phase frequency, the coefficients are optimized from experimental spectra and vibrational lifetimes of NMAD in the three solvents D₂O, DMSO, and CHCl₃, each representing a different electrostatic environment. Using the dielectric constant ϵ as a rough measure of the solvent's polarity, these three solvents range from a highly polar environment (water, $\epsilon = 78.5$); to a polar but aprotic environment (DMSO, $\epsilon = 47.2$); and finally, to a nearly nonpolar environment (chloroform, $\epsilon = 4.8$).

Asparagine (Asn) and glutamine (Gln) have amide groups in their side chains, and their absorption spectra contribute to the amide I region. Therefore, for peptides containing these residues, it is also important to include them in the exciton Hamiltonian and provide maps for them. These side chain groups have two hydrogen-bonding sites at the amide hydrogen atoms, so distributions of surrounding solvent molecules are quite different from those around backbone amide I chromophores. Therefore, the backbone map will not necessarily describe these side chains well. We have chosen N-deuterated acetamide (ACED, as shown in Figure 1b) as a model system and use spectra from it to develop a side chain map. Note that as these new backbone and side chain maps are designed to be transferable to different environments; they are potentially applicable to heterogeneous systems, such as membrane proteins.

Armed with these new transferable backbone and side chain frequency maps and existing protocols for NNFS and NNC maps,⁶⁷ as well as TDC for longer range interactions,⁴² we

next validate these maps by comparing to experimental spectra for a number of quite different model peptides. Specifically, we first consider simple molecules with one backbone and one side chain chromophore. Next, we consider the α -helical peptide AKA, with and without isotope labels;^{32,87} the Trpzip2 peptide, with and without isotope labels;^{88,89} and the rat and human versions of islet amyloid polypeptide (IAPP).^{90,91} In all cases, agreement between theory and experiment, without any ad hoc frequency shifts or further adjustments, is very good. The paper is organized as follows: In Section II, we describe the mixed quantum/classical treatment of the line shape theory and give the formula used for line shape calculations in this work. We also describe experimental details for spectra and lifetime measurements. Section III presents the development of the frequency maps. Validations of the maps are discussed in Sections IV and V. Finally, conclusions are drawn in Section VI.

II. MATERIALS AND METHODS

A. Line Shape Theory

The absorption line shape can be written as the Fourier transform of the quantum dipole time-correlation function.⁹² If the electric field of the excitation light is polarized in the $\hat{\epsilon}$ direction, the linear absorption line shape is

$$I(\omega) \sim \text{Re} \int_0^\infty dt e^{-i\omega t} \langle \hat{\epsilon} \cdot \vec{\mu}(0) \vec{\mu}(t) \cdot \hat{\epsilon} \rangle \quad (1)$$

where $\vec{\mu}$ is the dipole operator of the system. The brackets indicate a quantum equilibrium statistical mechanical average, which is impossible to evaluate for proteins in a condensed phase.

As discussed before, one practical approximation is to treat the amide I subspace quantum mechanically, ignore other high-frequency modes, and treat the low-frequency degrees of freedom (translations, rotations, and torsions) classically. Within such a mixed quantum/classical approach, the IR line shape for a multichromophore system is^{93,94}

$$I(\omega) \sim \text{Re} \int_0^\infty dt e^{-i\omega t} \sum_{ij} \langle m_i(0) F_{ij}(t) m_j(t) \rangle e^{-t/2T_1} \quad (2)$$

In the above equation, i and j index the amide I vibrational chromophore; $m_i(t) = \vec{m}_i(t) \cdot \hat{\epsilon}$, where $\vec{m}_i(t)$ is the transition dipole of the i th chromophore at time t . The Hamiltonian (divided by $\hbar$) in the amide I subspace is

$$\kappa_{ij}(t) = \omega_i(t) \delta_{ij} + \omega_{ij}(t) (1 - \delta_{ij}) \quad (3)$$

where $\omega_i(t)$ are the local-mode transition frequencies and $\omega_{ij}(t)$ are the vibrational couplings. The matrix $F(t)$ is the time-evolution operator for the one-exciton Hamiltonian,

$$\dot{F}(t) = i F(t) \kappa(t) \quad (4)$$

and it satisfies the initial condition $F_{ij}(0) = \delta_{ij}$. The angular brackets now indicate a classical equilibrium statistical mechanical average. T_1 is the lifetime of the first excited state of an isolated amide I vibration, and the term $e^{-t/2T_1}$ is added phenomenologically to incorporate

lifetime broadening. Equation 2 will be used to calculate theoretical line shapes for the different peptides.

In the map development, both model systems NMAD and ACED contain a single chromophore. For such single-chromophore systems, eq 2 becomes⁹³

$$I(\omega) \sim \text{Re} \int_0^\infty dt e^{-i\omega t} \left\langle mk(0) m(t) e^{i \int_0^t \omega(t') dt'} \right\rangle e^{-t/2T_1} \quad (5)$$

If the frequency fluctuation from its average, $\delta\omega(t) = \omega(t) - \langle\omega\rangle$, follows Gaussian statistics, the absorption line shape can be related to the frequency fluctuation time-correlation function, $C(t) = \langle\delta\omega(t) \delta\omega(0)\rangle$. If one makes the cumulant and Condon approximations and neglects orientational dynamics, the IR line shape can be simplified as^{58,95,96}

$$I(\omega) \sim \text{Re} \int_0^\infty dt e^{-i(\omega - \langle\omega\rangle)t} e^{-g(t)} e^{-t/2T_1} \quad (6)$$

where

$$g(t) \equiv \int_0^t dt' (t - t') C(t') \quad (7)$$

This approximation is particularly useful for our map development, since $g(t)$ can be calculated easily for different map parameters.

Actually, the experiments show that vibrational population relaxation is not single-exponential, but rather, is biexponential. Therefore, we modify the above appropriately,⁹⁷

$$I(\omega) \sim \text{Re} \int_0^\infty dt e^{-i(\omega - \langle\omega\rangle)t} e^{-g(t)} \{f e^{-t/2T_s} + (1 - f) e^{-t/2T_l}\} \quad (8)$$

where T_s and T_l are the “short” and “long” lifetimes, respectively, and f is the amplitude of the fast contribution.

B. Experimental IR Spectra and Pump–Probe Measurements

NMA and acetamide were purchased from Aldrich. NMAD and ACED were obtained by dissolving the samples in D₂O, removing D₂O by speed vacuum concentration, and repeating this procedure three times. FTIR spectra in the three different solvents were taken in an IR cell, with CaF₂ windows separated by a 100 μ m Teflon spacer. The concentrations of NMAD and ACED for FTIR measurements were ~20 mM.

Lifetimes were determined through mid-IR pump–probe transient absorption measurements. The femtosecond laser system and optical parametric amplifier used to generate 60 fs mid-IR pulses has been described previously.⁹⁸ The mid-IR pulses were split into pump and probe pulses with 1 μ J and 0.2 μ J pulses energies, respectively, at the sample. The pump pulse polarization was set at 54.7° (magic angle) relative to the probe pulse to eliminate rotational dynamics from the measured kinetics. The pump pulses were chopped at 500 Hz using a mid-IR pulse shaper.⁹⁹ Measurements were made at a series of sample concentrations and checked for consistency to eliminate effects due to self-association. The pump–probe spectrum was recorded as an average over 20–30 scans of the time delay between pump and probe pulses. Transient kinetics at the bleach maxima were fit to a biexponential function convoluted with a Gaussian instrument response function using a

nonlinear least-squares fitting routine with the Levenberg–Marquardt algorithm. The standard deviation at each time delay was used as a weighting factor in the fitting routine, and the resulting uncertainties in the fit parameters are reported as twice the standard deviation.

In Table 1, we report the experimental lifetimes and amplitudes, for NMAD in DMSO and chloroform and ACED in DMSO and D₂O (but not in chloroform; see below). The results for NMAD in D₂O are taken from DeCamp et al.⁹⁷ DeCamp et al. also studied NMA in DMSO-*d*₆ and CDCl₃, but concluded that H/D exchange had not fully occurred, so they were actually studying NMA (rather NMAD). In Table 2, we report the experimental peak frequencies, ω , and full-width-half-maxima, Γ .

III. DEVELOPMENT OF THE EMPIRICAL AMIDE I FREQUENCY MAPS

The GROMOS96 53a6 force field^{100–102} has been extensively used in biological simulations.^{90,91,103,104} In this work, transferable frequency maps for protein backbone and side chain amide chromophores are developed using this force field. We assume the same form of the ab initio-based S2 map,^{57,58}

$$\omega_i = \omega_0 + aE_{Ci} + bE_{Ni} \quad (9)$$

ω_i is the instantaneous frequency for the *i*th chromophore, E_{Ci} is the electric field on the C atom in the *i*th chromophore along the C=O bond direction, and E_{Ni} is that on the N atom. Parameters ω_0 , a , and b are obtained such that they optimally reproduce the experimental IR absorption spectra of NMAD or ACED in different solvents. Note that the intercept ω_0 is allowed to vary instead of being fixed to the gas-phase absorption frequency. ω_0 is the amide I frequency with no external field. Because we are mainly interested in condensed-phase systems, such a zero-field environment usually represents a nonpolar environment in which nonelectrostatic forces, such as dispersion, dominate. In the GROMOS96 53a6 force field, examples include alkanes, CCl₄, and the hydrophobic tails of lipids. Thus, we allow ω_0 to vary to take those nonelectrostatic effects into account, and thus, ideally the map will reproduce the experimental frequency of NMAD or ACED in those environments.

A. Backbone Frequency Map

NMAD is used as a model compound for the protein backbone amide linkage. Its structure is shown in Figure 1a. NVT MD simulations of NMAD in the solvents D₂O, DMSO, or CHCl₃ were performed at 298 K using the GROMACS simulation package,^{105–109} with the GROMOS96 53a6 force field.^{100–102} In each case, the experimental solvent density determined the simulation volume. The LINCS algorithm¹¹⁰ was used to constrain all bonds in the NMAD/D₂O and NMAD/CHCl₃ systems, and the SHAKE algorithm¹¹¹ was used for NMAD/DMSO. Particle-mesh Ewald (PME)^{112,113} was used to calculate the long-range Coulombic interactions. After equilibration, 2 ns production runs with a 2 fs time step were performed. Electric fields on the C and N atoms from solvent molecules within 20 Å were calculated. The line shapes were obtained from eq 8, using the experimental lifetime parameters for NMAD in the three solvents.

The parameters ω_0 , a , and b are determined by the Fletcher–Reeves–Polak–Ribiere method,¹¹⁴ which globally minimizes

$$\xi^2 = \sum_{\alpha=1}^3 \left\{ \frac{(\Delta\omega_{\alpha} - \Delta\omega_{\alpha,\text{exp}})^2}{\Delta\omega_{\alpha,\text{exp}}^2} + \frac{(\Gamma_{\alpha} - \Gamma_{\alpha,\text{exp}})^2}{\Gamma_{\alpha,\text{exp}}^2} \right\} \quad (10)$$

“1”, “2”, and “3” denote D₂O, DMSO, and CHCl₃, respectively. $\Delta\omega_{\alpha} = \omega_{\alpha} - \omega_{\text{g}}$, which is the frequency shift from the gas phase value of NMAD (1717 cm⁻¹).¹¹⁵ As noted earlier, the experimental line shape parameters are shown in Table 2. The amide I peak frequency of NMAD varies by ~40 cm⁻¹ between D₂O and CHCl₃, showing that the solvents chosen represent different electrostatic environments and the amide I mode is sensitive enough to distinguish among them.

The optimized backbone map is

$$\omega_i = 1684 + 7729E_{\text{Ci}} - 3576E_{\text{Ni}} \quad (11)$$

In the above equation, ω_i is in cm⁻¹ and E_{Ci} and E_{Ni} are in atomic units. The calculated IR line shapes for NMAD using this map (shown in Table 2 and Figure 2a) are in good agreement with experiment. Note, in particular, that the deviation between experiment and theory for the peak frequency is, at most, 3 cm⁻¹.

The map has coefficients a and b with opposite signs, which we can understand as follows. It is generally accepted that the NMAD molecule has three hydrogen-bonding sites and that hydrogen bonds lead to red shifts in local frequencies.^{46,49,50,55,57} The two water molecules hydrogen-bonded to the amide oxygen produce negative (opposite to the CO direction) electric fields on the C atom, and thus, the term $+7729E_{\text{Ci}}$ corresponds to a red shift due to these hydrogen bonds. Similarly, the water molecule hydrogen-bonded to the amide deuterium causes a positive electric field on the N atom, and from eq 11, it also leads to a red shift.

In a nonpolar environment, the map predicts a frequency of 1684 cm⁻¹. Experimentally, spectra for NMA in the nonpolar solvents *n*-hexane and CCl₄ peak at 1697 and 1688 cm⁻¹, respectively.^{116–118} Because NMAD is expected to have a similar (but slightly red-shifted) absorption compared with NMA,¹¹⁷ the map gives a reasonable result for nonpolar solvents.

B. Side Chain Map

We develop a side chain map from ACED following similar procedures. The minimization function is chosen to be

$$\xi^2 = (\omega_1 - \omega_{1,\text{exp}})^2 + (\Gamma_1 - \Gamma_{1,\text{exp}})^2 + (\omega_2 - \omega_{2,\text{exp}})^2 + (\Gamma_2 - \Gamma_{2,\text{exp}})^2 + (\omega_3 - \omega_{3,\text{exp}})^2 \quad (12)$$

Absolute peak positions are used in this case because no experimental gas-phase absorption spectrum of ACED is available. The line width in CHCl₃ is not included in the above equation because of self-association of ACED in this solvent. ACED shows two peaks in chloroform, similar to its undeuterated counterpart.^{119,120} The strongly overlapped peaks are attributed to monomeric and associated ACED species, with the former absorbing at ~1700 cm⁻¹, and the line width is not well resolved.

The optimized side chain map is

$$\omega_i^S = 1714 + 2154E_{Ci} + 3071E_{Ni} \quad (13)$$

and the line shape results are shown in Table 2 and Figure 2b. As seen, the experimental and theoretical spectra are in excellent agreement. Note, however, that the experimental spectrum for ACED in D₂O shows a shoulder on the red side not captured by theory. Calculations without making the cumulant approximation (not shown) also do not produce this shoulder. Also note that the calculated peak frequency in chloroform is 1697 cm⁻¹, in good agreement with the monomer absorption discussed above. Unlike the backbone map, for this side chain map, *a* and *b* have the same sign. Solvent molecules distribute differently around ACED compared with NMAD. In ACED, the -ND₂ group can form two hydrogen bonds, and the resulting electric field on the N atom depends on the hydrogen-bond angles; thus, *E*_{Ni} is not necessarily positive. The zero-field intercept is checked by comparing with spectra for undeuterated acetamide (ACE) in a nonpolar solvent. ACE absorbs at 1714 cm⁻¹ in CCl₄,^{121,122} and it is expected that ACED will have similar behavior, which corresponds well to the zero-field intercept of 1714 cm⁻¹.

IV. VALIDATION OF BACKBONE AND SIDE CHAIN MAPS: ONE- AND TWO-CHROMOPHORE SYSTEMS

A. Temperature Dependence of the Amide I Peak Frequency of NMAD in D₂O

Experimentally, it has been shown by Amunson and Kubelka¹²³ that the NMAD amide I peak frequency in D₂O is a very weak function of temperature, increasing by only 6 cm⁻¹ as the temperature varies between 0 and 85 °C. As a check of the map and probably even more of the GROMOS 53a6 force field (which uses the SPC potential¹²⁴ for water), it is interesting to see what our map predicts for the frequency shift with temperature. To this end, 1 ns NVT MD simulations of NMAD in D₂O were performed at temperatures ranging from 5 to 85 °C. The experimental density of heavy water at each temperature was used to determine the simulation box size. Theoretical and experimental¹²³ amide I peak frequencies at the lowest and highest temperature are shown in Table 3. As seen, the theoretical temperature dependence of the peak frequency is also very weak, changing by only 2 cm⁻¹. This rough agreement between theory and simulation is reassuring and suggests the possible application of our maps to temperature-jump transient IR experiments that study protein folding processes.

B. Two-Chromophore Systems: AAO and AGO

Side chains often play crucial roles in biological processes, such as protein folding, misfolding, and aggregation. Therefore, tracking their spectroscopic changes can provide important information. Researchers have long been interested in the amide I absorptions of Asn and Gln side chains in proteins,^{15,125–130} which, as mentioned earlier, contribute to the overall amide I band. Their absorptions in D₂O are deduced experimentally from two model peptides:¹²⁵ *N*-acetyl asparagine methyl ester (Ac-Asn-OMe, or AAO) and *N*-acetyl glutamine methyl ester (Ac-Gln-OMe, or AGO). We have chosen (deuterated) AAO and AGO as test cases for our maps.

Both AAO and AGO contain two amide I chromophores (see Figure 1), and the interactions between them modify their local frequencies and create couplings. Because the two chromophores are relatively far apart in space, their interactions are mainly electrostatic. Local frequencies of the backbone chromophore are modeled using eq 11, with the electric fields coming from all side-chain atoms and surrounding water molecules. Similarly, local frequencies of the side-chain chromophore are calculated from eq 13, with the electrostatic contributions from the backbone atoms and water molecules.

Coupling constants are calculated using the TDC scheme:^{14,41,42,74,83–86}

$$\omega_{ij} = \frac{A}{\epsilon} \left\{ \frac{\vec{m}_i \cdot \vec{m}_j}{r_{ij}^3} - 3 \frac{(\vec{m}_i \cdot \vec{r}_{ij})(\vec{m}_j \cdot \vec{r}_{ij})}{r_{ij}^5} \right\} \quad (14)$$

In the above equation, the vector $\vec{r}_{ij}$ (in Å) connects the two transition dipoles $\vec{m}_i$ and $\vec{m}_j$, which are in the units of $\text{D Å}^{-1} u^{-1/2}$ (u is atomic mass unit). ϵ is the dielectric constant and is taken to be 1. The conversion factor, $A = 0.1 \times 848\,619/1650$, gives the coupling frequencies in cm^{-1} .^{14,61} The TDC parameters are taken from ab initio calculations by Torii and Tasumi.⁴² Specifically, the transition dipoles have the magnitude of $2.73 \text{ D Å}^{-1} u^{-1/2}$ and are oriented 10.0° away from the C=O bond, pointing toward the N atom. Their origins are at $\vec{r}_C + 0.665\hat{n}_{\text{CO}} + 0.258\hat{n}_{\text{CN}}$ (in units of Å), where $\vec{r}_C$ is the location of the carbonyl carbon atom, $\hat{n}_{\text{CO}} = (\vec{r}_O - \vec{r}_C)/|\vec{r}_O - \vec{r}_C|$ and $\hat{n}_{\text{CN}} = (\vec{r}_N - \vec{r}_C)/|\vec{r}_N - \vec{r}_C|$.

Note that although D_2O is used as the solvent in experiment, we have instead used H_2O in the simulations. Because H_2O has a smaller mass than D_2O , its dynamics are slightly faster and lead to a slightly larger motional narrowing effect on the spectrum. However, because all potentials involving H_2O are the same as D_2O , the ensembles of configurations in the two cases are identical, so the influence on the IR spectra calculation is expected to be nearly negligible.

Theoretical [using eq 2, a lifetime of 600 fs (a reasonable compromise between the complicated biexponential decays for NMAD and ACED in D_2O , and for peptides in D_2O ³¹), and 5 ns trajectories] and experimental¹²⁵ spectra of AAO and AGO in aqueous D_2O solutions are shown in Figure 3. Theoretical spectra correctly capture the blue shift of AAO compared to AGO. The average coupling constants for both peptides are $\sim 1 \text{ cm}^{-1}$ and are not expected to influence the spectra much. Assuming no coupling between the two modes, we are able to obtain spectra for the backbone and side-chain chromophores individually. Theoretical backbone and side-chain absorptions peak at 1631 and 1640 cm^{-1} , respectively, for AAO and 1626 and 1636 cm^{-1} for AGO. Those values are close to experimentally extracted results.¹²⁵ The blue shift of AAO is mainly due to the higher frequencies of both chromophores, which are determined by its structure. The two chromophores in AAO are closer in space compared with AGO, and this leads to a larger electrostatic repulsion and a blue shift of the local frequencies.

V. VALIDATION OF THE AMIDE I FREQUENCY MAPS: APPLICATIONS TO MODEL PEPTIDES

A. General Protocol for Protein Line Shape Calculations in the Amide I Region

As discussed earlier, the frequency maps are developed from model compounds that contain only one chromophore. In the last section, we validated the maps for simple molecules with two (one backbone and one side-chain) chromophores. Further systematic validations are needed before using the maps with confidence for peptides and proteins.

For a chromophore in a protein backbone, the nearby peptide units can be divided into two categories: nearest neighbors and others. It has been shown that nearest-neighbor peptide units affect the local frequency in a through-bond manner,⁶⁷ and we have adopted the NNFS map of Jansen et al.⁶⁷ to account for these effects. On the basis of DFT calculations, the NNFS map relates the frequency shift to (ϕ, ψ) angles between adjacent peptide units. All other peptide units are treated as charged sites and influence the local frequency in a purely electrostatic way. Therefore, the diagonal terms of the Hamiltonian are calculated as

follows: For the i th backbone chromophore, contributions from the $(i - 1)$ th and $(i + 1)$ th residues (termed as $\Delta\omega_N$ and $\Delta\omega_C$) are determined by the corresponding (φ, ψ) angles. Electric fields from the rest of the backbone atoms, side chains, solvents and counterions within 20 Å of the i th chromophore are used to calculate ω_i from eq 11. The resulting local frequency is

$$\omega_i^b = \omega_i + \Delta\omega_N(\varphi_{i-1}, \psi_{i-1}) + \Delta\omega_C(\varphi_{i+1}, \psi_{i+1}) \quad (15)$$

The amino acid proline is a special case because its nitrogen atom is bound to two carbons and its φ angle is constrained. Herein, we do not treat proline differently from the other amino acids, but a more refined map should make this distinction.

The frequency of a side-chain chromophore is calculated from eq 13, with electric fields from all atoms within 20 Å. Note that when the C-terminus of the peptide is amidated, this creates an additional chromophore with the same form as the Asn and Gln side chains. Thus, for these chromophores, we use eq 13, but also add the NNFS contribution $\Delta\omega_N$. Also note that for both the backbone and the side chain, electric fields from the C, O, N, and H atoms within the chromophore itself are excluded.

Couplings between adjacent peptide units are mainly due to overlapping charge densities,^{45,67,74} and are determined from the NNC map.⁶⁷ Couplings between all other peptide units are primarily long-distance electrostatic interactions,^{45,67,74} which are approximated using TDC,^{14,41,42,74,83–86} as in eq 14.

After constructing the system Hamiltonian, eq 2 is used for the line shape calculation. T_1 is set to be 600 fs.³¹ Note that the term $m_i(t) = \vec{m}_i(t) \cdot \hat{\epsilon}$ describes rotations of transition dipoles with respect to the electric field unit vector, $\hat{\epsilon}$, of the excitation light. On the computer simulation time scale, the peptide cannot rotate sufficiently fast to sample all possible orientations, so one needs instead to average over all possible orientations of $\hat{\epsilon}$ with respect to the lab-fixed axes. This is equivalent to averaging over the three cases where $\hat{\epsilon}$ is $\hat{x}$, $\hat{y}$, and $\hat{z}$, which is what we do here.

A general protocol for the calculation of protein line shapes in the amide I region is as follows:

1. Generate configuration trajectories from MD simulations. At each time step, the configuration can be used to calculate backbone (φ, ψ) angles and electric fields on each chromophore.
2. Calculate ω_i and ω_i^s from eqs 11 and 13, respectively.
3. Determine NNFS and NNC from the corresponding (φ, ψ) angles.⁶⁷
4. ω_i^b is then calculated by adding ω_i and NNFS, as in eq 15.
5. Couplings between all non-nearest-neighbor amide groups are calculated from TDC.⁴²
6. Construct the fluctuating Hamiltonian matrix κ and propagate the F matrix.^{131,132}
7. Calculate the absorption line shape using eq 2.

Below, we carry out systematic studies on model peptides with well-defined secondary structures. These model systems represent the most abundant protein secondary structure motifs and are used to illustrate how well the frequency maps can work for larger systems.

B. α -Helical Model System: The AKA Peptide

The AKA peptide has been chosen to represent an α -helical configuration. AKA has a sequence motif of (AAAAK) $_n$, which has been shown to form α -helical structures, especially in the central region.^{133–139} The *N*-acetyl capping further increases the helical stability at the N-terminus.¹³⁷ The sequence and configuration of AKA, with $n = 4$, are shown in Figures 4 and 5a. Note that acetylation produces a backbone chromophore at the N-terminus, and amidation produces a “side-chain” chromophore at the C-terminus, as discussed above.

The initial structure of AKA was constructed by assuming a perfect α -helix with all (ϕ, ψ) angles set to $(-57^\circ, -47^\circ)$. It was solvated in enough H₂O molecules to ensure a distance of 1 nm from all box edges. Side chains of lysines are positively charged, and chloride ions were added to neutralize them. The system was then energy-minimized and equilibrated with constant temperature for 20 ps with the positions of protein atoms frozen.

Unconstrained simulations were then carried out to equilibrate the system, first in the NVT ensemble for 100 ps and then in the NpT ensemble for 5 ns. Properties such as temperature, pressure, density, and root-mean-square-deviations were monitored to ensure equilibration. An MD run for 10 ns was then performed at a time step of 2 fs. The LINCS algorithm was used to constrain all bonds, and PME was used to calculate the long-range Coulombic interactions. Berendsen coupling schemes for temperature and pressure¹⁴⁰ were used in the simulation. Temperature was set to be 275 K, in accordance with experiment.^{30,32,87} The GROMACS source code was modified to report local frequencies on the fly using eqs 11 and 13. Frequencies and configurations were saved every 10 fs for spectra calculations. Throughout the 10 ns simulation, the central part of the peptide stays α -helical, while the C-terminus frays, as described in experimental studies.^{134,136,137,141}

Theoretical and experimental⁸⁷ spectra of unlabeled (but suitably deuterated) AKA peptide in D₂O are shown in Figure 6a. The theoretical spectrum peaks at 1627 cm^{-1} , differing from experiment by only 4 cm^{-1} . This peak position is determined by both the diagonal and off-diagonal elements of the Hamiltonian.¹⁴² On average, local frequencies of individual chromophores, at least in the central region of the helix (see below), are around 1640 cm^{-1} , which provides a central frequency that will be modified by couplings. Coupling constants between peptide units that are adjacent to each other (β_{12}), one residue apart (β_{13}), and two residues apart (β_{14}) have the largest magnitude and, thus, have the largest impact on the overall spectrum. β_{12} 's are positive for α -helices, as predicted by the NNC map and Ham et al.⁴⁷ The overall effect of positive β_{12} 's is a blue shift of the peak frequency. On the other hand, β_{13} , β_{14} , and other couplings between nonadjacent neighbors are negative, which counteract the effect of β_{12} and lead to a final red shift of the peak position. The average β_{12} , β_{13} , and β_{14} for residues 2–20 are shown in Table 4. Experimentally, coupling constants were extracted by systematically isotope-labeling two residues zero, one, and two residues apart.³⁰ Compared with experiment, the signs and relative magnitude of the three coupling constants are correct.

We also carried out calculations on isotope-edited AKA peptides. The frequency shift of $^{13}\text{C}=^{18}\text{O}$ is assumed to be -70 cm^{-1} , consistent with previous experiments and ab initio calculations.^{27–29,32,47} One $^{13}\text{C}=^{18}\text{O}$ label was placed on residue 12 (denoted as [12]), and the full spectrum was calculated by shifting the frequency of the labeled residue and retaining all the couplings. This procedure was repeated for residues [13], [14], and [15]. All four spectra appear quite similar. The [12] spectrum is shown in Figure 6b along with the experimental result.³² Note the good agreement of the labeled features at about 1570 cm^{-1} . To compare further the isotope-labeled feature, we assume *each* residue is $^{13}\text{C}=^{18}\text{O}$ -labeled, one at a time, and treat each one of them as an isolated chromophore and neglect couplings. Isotope-labeled peak frequencies as a function of residue are shown in Figure 7. Peaks in the

central region stay more or less constant, consistent with a homogeneous α -helical configuration. Experimental results³² for residues [12], [13], [14], and [15], which differ from theory by only a few cm^{-1} , confirm our combined electrostatic and NNFS⁶⁷ scheme for calculating backbone frequencies.

C. β -Hairpin Model System: The Trpzip2 Peptide

Tryptophan zippers (Trpzips) are stable, monomeric β -hairpins in aqueous solution at room temperature.¹⁴³ Trpzip2 is one of the most stable molecules in the Trpzip family and has been extensively studied.^{70,71,88,89,144–146} The sequence and configuration of Trpzip2 are shown in Figures 4 and 5b. Amidated Trpzip2 contains 11 backbone chromophores, 1 “side-chain” chromophore at the C-terminus, and 1 side-chain chromophore on Asn6.

The initial configuration of Trpzip2 was taken from the Protein Data Bank (PDB ID 1LE1). MD simulations were run with procedures similar to those for AKA. The simulation temperature was 298 K. Theoretical and experimental⁸⁸ spectra are shown in Figure 8a, both of which have a two-peak feature. The two peaks are presumably due, in part, to collective motions of chromophores that give rise to a^- and a^+ type peaks.^{88,145–148} Agreement between theory and experiment, while not perfect, is still quite good (in that the theoretical positions and intensities of both peaks are about right).

Apart from couplings, inhomogeneity of diagonal frequencies also contribute to the spectral features. Assuming that the ^{13}C label shifts the frequency by -43 cm^{-1} ,^{26,34,47,89} and assuming that labeled chromophores are isolated, the labeled peak frequencies as a function of backbone residue number are plotted in Figure 9. The frequencies show a zigzag pattern due to alternating hydrogen-bonding environments, in agreement with previous simulation results.^{70,145} Carbonyls facing inside the hairpin form a single intramolecular hydrogen bond, whereas carbonyls facing outside the hairpin form two hydrogen bonds with surrounding water molecules. Correspondingly, the latter have lower frequencies. The nondegenerate local frequencies fall into two regions, corresponding to the two peaks in the spectrum. ^{13}C -labeling experiments^{88,89} confirm this inhomogeneity because the frequencies of residues 3 and 10 are $\sim 25 \text{ cm}^{-1}$ higher than that of residue 2, which is, in turn, 10 cm^{-1} higher than the frequency of residue 7. These experimental frequencies are also shown in Figure 9. They are qualitatively in agreement with theory; the lack of quantitative agreement is most likely due (in part) to the fact that ^{13}C -labeled residues are not fully localized, as was assumed in these particular theoretical calculations.

Theoretical and experimental spectra of Trpzip2 doubly labeled by ^{13}C at positions 3 and 10 (denoted as T3*-T10*) are shown in Figure 8b. In accordance with experiment,⁸⁸ the theoretical spectrum shows a less prominent shoulder around 1680 cm^{-1} , with a new shoulder appearing around 1605 cm^{-1} . The former is easy to understand, since the unlabeled local frequencies of residues 3 and 10 are both about 1680 cm^{-1} (from the results in Figure 9 plus 43 cm^{-1}), when these residues are labeled the intensity near 1680 cm^{-1} drops. When residues 3 and 10 are labeled, they mix with other lower-frequency residues, causing the theoretical shift of the main peak and appearance of the low-frequency shoulder.

The lack of quantitative agreement between theory and experiment for the Trpzip2 problem may result from our use of the TDC approximation for nearby chromophores interacting across the β -hairpin, which is questionable.¹⁴⁹

D. IAPP

Islet amyloid polypeptide (IAPP) is a 37-residue peptide. Amyloid fibrils of human IAPP (hIAPP) are a common pathological characteristic of type 2 diabetes.^{150,151} The aggregation process and cell disruption mechanisms are not clear due to its fast aggregation kinetics. Rat

IAPP (rIAPP), which differs from hIAPP at only 6 residues, does not form aggregates. The sequences of hIAPP and rIAPP are shown in Figure 4. Three proline mutations at positions 25, 28, and 29 are believed to prevent rIAPP from forming aggregates. Investigation of the monomer conformational states of both peptides in solution is a first step toward understanding the aggregation process. Note that rIAPP and hIAPP each have six Asn residues (each with a side-chain chromophore), one Gln residue with its side-chain chromophore, and an additional “side-chain” chromophore at the C-terminus as a result of amidation. For both peptides, Cys2 and Cys7 are connected by disulfide bonds.

Reddy et al. presented combined replica exchange umbrella sampling and MD simulations on rIAPP monomer.⁹⁰ Both “folded” and random-coil conformations were observed for rIAPP in aqueous solution, with relative populations of 55% and 45%, respectively.⁹⁰ Representative configurations of the two conformers are shown in Figure 10. Theoretical IR absorption spectra of rIAPP were presented previously⁹⁰ using the improved S2 map.⁵⁸ In the present paper, the same line shape calculations have been performed with the new frequency maps. The results of each conformer, weighted by their relative populations, and the total spectrum are shown in Figure 11. The theoretical spectrum is considerably improved compared with the previous result⁹⁰ and is in good agreement with experiment.⁹⁰ (Note that the slight disagreement with experiment around 1600 cm^{-1} might be ameliorated with an improved map for proline because there are three such residues in rIAPP.) In part, the improvement between theory and experiment is due to the new side-chain map (rIAPP has eight side-chain chromophores); the previous calculation used a backbone map for all (including side-chain) chromophores. One notices that both conformers absorb at about 1645 cm^{-1} , and the random-coil conformer has a wider line width due to its flexibility. The similarity in spectral features is due to the similarity in the three-dimensional structures. The folded conformer has an α -helical segment consisting of only residues 7–17, with the remaining part being mainly random coil. The α -helical segment leads to a red shift of the spectrum, but because 70% of the structure is random coil, its spectroscopic feature is quite similar to the random-coil conformer.

Following similar procedures, the hIAPP monomer is found to adopt α -helical, β -hairpin (“misfolded”), and random-coil conformations, with relative populations of 31%, 40%, and 29%, respectively.⁹¹ Representative snapshots along the simulations are shown in Figure 12. Theoretical and experimental spectra are shown in Figure 13.⁹¹ In this case, the α -helical and random-coil conformers have distinct spectral features. The α -helical conformer has residues 9–17 forming an α -helical structure, with a short antiparallel β -sheet at the C-terminus.^{91,152} This more-ordered structure absorbs at 1644 cm^{-1} , 10 cm^{-1} red-shifted compared with the random-coil conformer. The misfolded conformer has an antiparallel β -hairpin structure,^{91,152} whose amide I peaks appear at 1642 and 1670 cm^{-1} . The total spectrum is in good agreement with experiment, suggesting the existence of the β -hairpin structure in the solution phase. The fiber of aggregated hIAPP molecules is composed of parallel β -sheet structures, and the misfolded conformer might be a precursor for aggregation.^{91,152}

VI. CONCLUSIONS

In this paper, we have developed frequency maps for protein backbone and side chain amide I absorptions, from experimental line shapes and vibrational lifetimes on model compounds in different solvents. We have validated the maps by applying them, together with previously developed NNFS⁶⁷ and coupling maps,^{42,67} to absorption spectra of model compounds with a pair of chromophores, and unlabeled and (isotope)-labeled peptides with well-defined α -helical and β -hairpin secondary structures. Our theoretical results are in very good agreement with experiment, without any ad hoc frequency shifts or further adjustments

(typical discrepancies between theory and experiment are on the order of a few wavenumbers). The fact that the frequency maps, applied with the GROMOS96 53a6 force field, are able to capture key features in the experimental spectra for a variety of different systems provides a way to bridge computer simulations and IR experiments. As computer power increases, simulations for large proteins for biologically relevant time scales will soon be possible. This and other theoretical frameworks that can accurately and efficiently calculate IR spectra, combined with experimental approaches, will enable the understanding of protein structure and dynamics at the molecular level.

The advancement of 2DIR opens a wide area for applications of these maps. 2DIR, with its superior time and frequency resolution, has become a powerful tool for probing protein structure and dynamics.^{20–25,37,153} The frequency maps can thus be utilized to study protein folding, hydrogen-bonding dynamics, protein aggregation, and so forth in aqueous solutions. Moreover, because the maps are designed (by considering different solvents) to be transferable, they are potentially applicable to more heterogeneous, including hydrophobic, environments. One possible application is the study of protein structure and dynamics in the presence of membranes. Membrane peptides are usually hard to crystallize and, thus, remain challenging for techniques such as X-ray crystallography, whereas IR spectroscopy can be utilized to study structure and in situ kinetics. We are carrying out further studies to validate the frequency maps using 2DIR experiments both in the presence and absence of lipid membranes.

For example, our ability to reproduce experimental absorption spectra of rIAPP and hIAPP monomers in solution is promising. 2DIR experiments, coupled with $^{13}\text{C}=^{18}\text{O}$ isotope labeling techniques, have been applied to the aggregation study of hIAPP in aqueous solution and in lipid vesicles.^{37,153} It has been widely accepted that the aggregation process of hIAPP is greatly accelerated in the presence of lipid membranes.^{154–156} In turn, intermediates along the aggregation pathway disrupt the membrane structure and possibly lead to cell death.^{157–160} Line shape calculations, combined with extensive computer simulations, can provide details on oligomer and aggregate formation and explain changes of spectroscopic features observed in experiment along the aggregation pathway, which may shed light on the aggregation mechanism and, ultimately, on the medical treatment of type 2 diabetes.

Acknowledgments

The authors thank Juan de Pablo and Thomas Jansen for helpful discussions. This work was supported in part by a CRC grant from the NSF (CHE-0832584) to J.L.S. and M.T.Z. J.L.S. also thanks NSF for support of this work through Grant CHE-1058752 and NIH, through 1R01DK088184. M.T.Z. also thanks NIH for support through 1R01DK079895.

References

1. Dickerson, RE.; Geis, I. The structure and action of proteins. W. A. Benjamin, Inc; Menlo Park, CA: 1969.
2. Nelson, DL.; Cox, MM. Lehninger Principles of Biochemistry. W. H. Freeman and Company; New York: 2005.
3. Berg, JM.; Tymoczko, JL.; Stryer, L. Biochemistry. W. H. Freeman and Company; New York: 2007.
4. McCammon, JA.; Harvey, SC. Dynamics of Proteins and Nucleic Acids. Cambridge University Press; Cambridge: 1987.
5. Boehr DD, Dyson HJ, Wright PE. Chem Rev. 2006; 106:3055. [PubMed: 16895318]
6. Kraut J. Annu Rev Biochem. 1965; 34:247. [PubMed: 14321170]
7. Cartailier J-P, Luecke H. Annu Rev Biophys Biomol Struct. 2003; 32:285. [PubMed: 12598369]

8. Clore, GM.; Gronenborn, AM., editors. NMR of Proteins. CRC Press Inc; Boca Raton, FL: 1993.
9. Thomas GJ Jr. Annu Rev Biophys Biomol Struct. 1999; 28:1. [PubMed: 10410793]
10. Kelly SM, Price NC. Biochim Biophys Acta. 1997; 1338:161. [PubMed: 9128135]
11. Margittai M, Langen R. Q Rev Biophys. 2008; 41:265. [PubMed: 19079806]
12. Weng L, Baker GM. Biochemistry. 1991; 30:5727. [PubMed: 1645999]
13. Yan Y, Marriott G. Curr Opin Chem Biol. 2003; 7:635. [PubMed: 14580569]
14. Krimm S, Bandekar J. Adv Protein Chem. 1986; 38:181. [PubMed: 3541539]
15. Barth A, Zscherp C. Q Rev Biophys. 2002; 35:369. [PubMed: 12621861]
16. Susi H, Byler DM. Methods Enzymol. 1986; 130:290. [PubMed: 3773736]
17. Haris PI, Chapman D. Trends Biochem Sci. 1992; 17:328. [PubMed: 1412707]
18. Surewicz WK, Mantsch HH, Chapman D. Biochemistry. 1993; 32:389. [PubMed: 8422346]
19. Hamm P, Lim M, Hochstrasser RM. J Phys Chem B. 1998; 102:6123.
20. Zanni MT, Hochstrasser RM. Curr Opin Struct Biol. 2001; 11:516. [PubMed: 11785750]
21. Hochstrasser RM. Proc Natl Acad Sci USA. 2007; 104:14190. [PubMed: 17664429]
22. Park S, Kwak K, Fayer MD. Laser Phys Lett. 2007; 4:704.
23. Cho M. Chem Rev. 2008; 108:1331. [PubMed: 18363410]
24. Ganim Z, Chung HS, Smith AW, DeFlores LP, Jones KC, Tokmakoff A. Acc Chem Res. 2008; 41:432. [PubMed: 18288813]
25. Strasfeld DB, Ling YL, Shim S-H, Zanni MT. J Am Chem Soc. 2008; 130:6698. [PubMed: 18459774]
26. Tadesse L, Nazarbaghi R, Walters L. J Am Chem Soc. 1991; 113:7036.
27. Torres J, Adams PD, Arkin IT. J Mol Biol. 2000; 300:677. [PubMed: 10891262]
28. Torres J, Kukol A, Goodman JM, Arkin IT. Biopolymers. 2001; 59:396. [PubMed: 11598874]
29. Fang C, Wang J, Charnley AK, Barber-Armstrong W, Smith AB, Decatur SM, Hochstrasser RM. Chem Phys Lett. 2003; 382:586.
30. Fang C, Wang J, Kim YS, Charnley AK, Barber-Armstrong W, Smith AB, Decatur SM, Hochstrasser RM. J Phys Chem B. 2004; 108:10415.
31. Mukherjee P, Krummel AT, Fulmer EC, Kass I, Arkin IT, Zanni MT. J Chem Phys. 2004; 120:10215. [PubMed: 15268045]
32. Fang C, Hochstrasser RM. J Phys Chem B. 2005; 109:18652. [PubMed: 16853400]
33. Arkin IT. Curr Opin Chem Biol. 2006; 10:394. [PubMed: 16935550]
34. Decatur SM. Acc Chem Res. 2006; 39:169. [PubMed: 16548505]
35. Mukherjee P, Kass I, Arkin I, Zanni MT. Proc Natl Acad Sci USA. 2006; 103:3528. [PubMed: 16505377]
36. Mukherjee P, Kass I, Arkin IT, Zanni MT. J Phys Chem B. 2006; 110:24740. [PubMed: 17134238]
37. Shim S-H, Gupta R, Ling YL, Strasfeld DB, Raleigh DP, Zanni MT. Proc Natl Acad Sci USA. 2009; 106:6614. [PubMed: 19346479]
38. Strasfeld DB, Ling YL, Gupta R, Raleigh DP, Zanni MT. J Phys Chem B. 2009; 113:15679. [PubMed: 19883093]
39. Manor J, Mukherjee P, Lin Y-S, Leonov H, Skinner JL, Zanni MT, Arkin IT. Structure. 2009; 17:247. [PubMed: 19217395]
40. Woys AM, Lin Y-S, Reddy AS, Xiong W, de Pablo JJ, Skinner JL, Zanni MT. J Am Chem Soc. 2010; 132:2832. [PubMed: 20136132]
41. Torii H, Tasumi M. J Chem Phys. 1992; 96:3379.
42. Torii H, Tasumi M. J Raman Spectrosc. 1998; 29:81.
43. Torii H. J Phys Chem B. 2007; 111:5434. [PubMed: 17441760]
44. Torii H. Mol Phys. 2009; 107:1855.
45. Hamm P, Woutersen S. Bull Chem Soc Jpn. 2002; 75:985.
46. Ham S, Kim J-H, Lee H, Cho M. J Chem Phys. 2003; 118:3491.
47. Ham S, Cha S, Choi J-H, Cho M. J Chem Phys. 2003; 119:1451.

48. Ham S, Cho M. *J Chem Phys.* 2003; 118:6915.
49. Kwac K, Cho M. *J Chem Phys.* 2003; 119:2247.
50. Choi J-H, Ham S, Cho M. *J Phys Chem B.* 2003; 107:9132.
51. Hahn S, Ham S, Cho M. *J Phys Chem B.* 2005; 109:11789. [PubMed: 16852448]
52. Choi J-H, Hahn S, Cho M. *Int J Quantum Chem.* 2005; 104:616.
53. Choi J-H, Lee H, Lee K-K, Hahn S, Cho M. *J Chem Phys.* 2007; 126:045102. [PubMed: 17286512]
54. Jeon J, Yang S, Choi J-H, Cho M. *Acc Chem Res.* 2009; 42:1280. [PubMed: 19456096]
55. Bouř P, Keiderling T. *J Chem Phys.* 2003; 119:11253.
56. Bouř P, Michalík D, Kapitán J. *J Chem Phys.* 2005; 122:144501. [PubMed: 15847539]
57. Schmidt JR, Corcelli SA, Skinner JL. *J Chem Phys.* 2004; 121:8887. [PubMed: 15527353]
58. Lin Y-S, Shorb JM, Mukherjee P, Zanni MT, Skinner JL. *J Phys Chem B.* 2009; 113:592. [PubMed: 19053670]
59. Hochstrasser RM. *Adv Chem Phys.* 2006; 132:1.
60. Hayashi T, Zhuang W, Mukamel S. *J Phys Chem A.* 2005; 109:9747. [PubMed: 16833288]
61. Zhuang W, Abramavicius D, Hayashi T, Mukamel S. *J Phys Chem B.* 2006; 110:3362. [PubMed: 16494351]
62. Sengupta N, Maekawa H, Zhuang W, Toniolo C, Mukamel S, Tobias DJ, Ge N-H. *J Phys Chem B.* 2009; 113:12037. [PubMed: 19496555]
63. Zhuang W, Hayashi T, Mukamel S. *Angew Chem, Int Ed.* 2009; 48:3750.
64. Maekawa H, Ge N-H. *J Phys Chem B.* 2010; 114:1434. [PubMed: 20050636]
65. Ganim Z, Tokmakoff A. *Biophys J.* 2006; 91:2636. [PubMed: 16844758]
66. Watson TM, Hirst JD. *Mol Phys.* 2005; 103:1531.
67. la Cour Jansen T, Dijkstra AG, Watson TM, Hirst JD, Knoester J. *J Chem Phys.* 2006; 125:044312.
68. la Cour Jansen T, Knoester J. *J Chem Phys.* 2006; 124:044502. [PubMed: 16460180]
69. Bloem R, Dijkstra AG, la Cour Jansen T, Knoester J. *J Chem Phys.* 2008; 129:055101. [PubMed: 18698926]
70. Roy S, la Cour Jansen T, Knoester J. *Phys Chem Chem Phys.* 2010; 12:9347. [PubMed: 20596553]
71. Smith AW, Lessing J, Ganim Z, Peng CS, Tokmakoff A, Roy S, la Cour Jansen T, Knoester J. *J Phys Chem B.* 2010; 114:10913. [PubMed: 20690697]
72. Gorbunov RD, Kosov DS, Stock G. *J Chem Phys.* 2005; 122:224904. [PubMed: 15974713]
73. Gorbunov RD, Nguyen PH, Kobus M, Stock G. *J Chem Phys.* 2007; 126:054509. [PubMed: 17302487]
74. Gorbunov RD, Stock G. *Chem Phys Lett.* 2007; 437:272.
75. Kobus M, Gorbunov RD, Nguyen PH, Stock G. *Chem Phys.* 2008; 347:208.
76. Cai K, Han C, Wang J. *Phys Chem Chem Phys.* 2009; 11:9149. [PubMed: 19812835]
77. Paschek D, Pühse M, Perez-Goicochea A, Gnanakaran S, Garcia AE, Winter R, Geiger A. *ChemPhysChem.* 2008; 9:2742. [PubMed: 19035605]
78. Kim J, Huang R, Kubelka J, Bouř P, Keiderling TA. *J Phys Chem B.* 2006; 110:23590. [PubMed: 17107215]
79. Grahnen JA, Amunson KE, Kubelka J. *J Phys Chem B.* 2010; 114:13011. [PubMed: 20857992]
80. Zhao J, Wang J. *J Phys Chem B.* 2010; 114:16011. [PubMed: 21070067]
81. Choi J-H, Cho M. *J Chem Phys.* 2004; 120:4383. [PubMed: 15268607]
82. Choi J-H, Cho M. *Chem Phys.* 2009; 361:168.
83. Krimm S, Abe Y. *Proc Natl Acad Sci USA.* 1972; 69:2788. [PubMed: 4507602]
84. Moore WH, Krimm S. *Proc Natl Acad Sci USA.* 1975; 72:4933. [PubMed: 16592297]
85. Cheam TC, Krimm S. *Chem Phys Lett.* 1984; 107:613.
86. Cheam TC, Krimm S. *J Chem Phys.* 1985; 82:1631.

87. Barber-Armstrong W, Donaldson T, Wijesooriya H, Silva RAGD, Decatur SM. *J Am Chem Soc.* 2004; 126:2339. [PubMed: 14982437]
88. Smith AW, Tokmakoff A. *J Chem Phys.* 2007; 126:045109. [PubMed: 17286519]
89. Wang J, Chen J, Hochstrasser RM. *J Phys Chem B.* 2006; 110:7545. [PubMed: 16599536]
90. Reddy AS, Wang L, Lin Y-S, Ling Y, Chopra M, Zanni MT, Skinner JL, de Pablo JJ. *Biophys J.* 2010; 98:443. [PubMed: 20141758]
91. Reddy AS, Wang L, Singh S, Ling YL, Buchanan L, Zanni MT, Skinner JL, de Pablo JJ. *Biophys J.* 2010; 99:2208. [PubMed: 20923655]
92. McQuarrie, DA. *Statistical Mechanics.* Harper and Row; New York: 1976.
93. Auer BM, Skinner JL. *J Chem Phys.* 2007; 127:104105. [PubMed: 17867735]
94. Auer BM, Skinner JL. *J Chem Phys.* 2008; 128:224511. [PubMed: 18554033]
95. Mukamel, S. *Principles of Nonlinear Optical Spectroscopy.* Oxford; New York: 1995.
96. Schmidt JR, Corcelli SA, Skinner JL. *J Chem Phys.* 2005; 123:044513. [PubMed: 16095375]
97. DeCamp MF, DeFlores L, McCracken JM, Tokmakoff A, Kwac K, Cho M. *J Phys Chem B.* 2005; 109:11016. [PubMed: 16852342]
98. Shim S-H, Strasfeld DB, Zanni MT. *Opt Express.* 2006; 14:13120. [PubMed: 19532209]
99. Shim S-H, Strasfeld DB, Fulmer EC, Zanni MT. *Opt Lett.* 2006; 31:838. [PubMed: 16544641]
100. van Gunsteren, WF.; Billeter, SR.; Eising, AA.; Hünenberger, PH.; Krüger, P.; Mark, AE.; Scott, WRP.; Tironi, IG. *Biomolecular Simulation: The GROMOS96 manual and user guide.* Hochschuleverlag AG an der ETH Zürich; Zürich, Switzerland: 1996.
101. Scott WRP, Hünenberger PH, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF. *J Phys Chem A.* 1999; 103:3596.
102. Oostenbrink C, Villa A, Mark AE, van Gunsteren WF. *J Comput Chem.* 2004; 25:1656. [PubMed: 15264259]
103. Zagrovic B, van Gunsteren WF. *Proteins.* 2006; 63:210. [PubMed: 16425239]
104. Zhou Y, Oostenbrink C, Jongejan A, van Gunsteren WF, Hagen WR, de Leeuw SW, Jongejan JA. *J Comput Chem.* 2006; 27:857. [PubMed: 16541426]
105. Bekker, H.; Berendsen, HJC.; Dijkstra, EJ.; Achterop, S.; van Drunen, R.; van der Spoel, D.; Sijbers, A.; Keegstra, H.; Reitsma, B.; Renardus, MKR. *Physics Computing '92.* de Groot, RA.; Nadrchal, J., editors. World Scientific; Singapore: 1993.
106. Berendsen HJC, van der Spoel D, van Drunen R. *Comput Phys Commun.* 1995; 91:43.
107. Lindahl E, Hess B, van der Spoel D. *J Mol Model.* 2001; 7:306.
108. van der Spoel, D.; Lindahl, E.; Hess, B.; van Buuren, AR.; Apol, E.; Meulenhoff, PJ.; Tieleman, DP.; Sijbers, ALTM.; Feenstra, KA.; van Drunen, R.; Berendsen, HJC. *Gromacs User Manual version 3.3.* 2005. www.gromacs.org
109. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. *J Comput Chem.* 2005; 26:1701. [PubMed: 16211538]
110. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. *J Comput Chem.* 1997; 18:1463.
111. Ryckaert J-P, Ciccotti G, Berendsen HJC. *J Comput Phys.* 1977; 23:327.
112. Darden T, York D, Pedersen L. *J Chem Phys.* 1993; 98:10089.
113. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. *J Chem Phys.* 1995; 103:8577.
114. Press, WH.; Flannery, BP.; Teukolsky, SA.; Vetterling, WT. *Numerical Recipes.* Cambridge University Press; Cambridge: 1986.
115. Mayne LC, Hudson B. *J Phys Chem.* 1991; 95:2962.
116. Eaton G, Symons MCR, Rastogi PP. *J Chem Soc, Faraday Trans 1.* 1989; 85:3257.
117. Kubelka J, Keiderling TA. *J Phys Chem A.* 2001; 105:10922.
118. Nyquist RA. *Spectrochim Acta.* 1963; 19:509.
119. Davies M, Hallam HE. *Trans Faraday Soc.* 1951; 47:1170.
120. Brown TL, Regan JF, Schuetz RD, Sternberg JC. *J Phys Chem.* 1959; 63:1324.
121. Titov EV, Poddubnaya MV, Kapkan LM. *Theor Exp Chem.* 1973; 6:421.

122. Ruggirello A, Liveri VT. *Colloid Polym Sci.* 2003; 281:1062.
123. Amunson KE, Kubelka J. *J Phys Chem B.* 2007; 111:9993. [PubMed: 17676791]
124. Berendsen, HJC.; Postma, JPM.; van Gunsteren, WF.; Hermans, J. *Intermolecular Forces.* Pullman, B., editor. Reidel; Dordrecht: 1981.
125. Chirgadze YN, Fedorov OV, Trushina NP. *Biopolymers.* 1975; 14:679. [PubMed: 1156632]
126. Venyaminov SY, Kalnin NN. *Biopolymers.* 1990; 30:1243. [PubMed: 2085660]
127. Rahmelow K, Hübner W, Ackermann T. *Anal Biochem.* 1998; 257:1. [PubMed: 9512765]
128. Barth A. *Prog Biophys Mol Biol.* 2000; 74:141. [PubMed: 11226511]
129. Wolpert M, Hellwig P. *Spectrochim Acta, Part A.* 2006; 64:987.
130. Barth A. *Biochim Biophys Acta.* 2007; 1767:1073. [PubMed: 17692815]
131. la Cour Jansen T, Knoester J. *J Phys Chem B.* 2006; 110:22910. [PubMed: 17092043]
132. Yang M, Skinner JL. *Phys Chem Chem Phys.* 2010; 12:982. [PubMed: 20066383]
133. Marqusee S, Robbins VH, Baldwin RL. *Proc Natl Acad Sci USA.* 1989; 86:5286. [PubMed: 2748584]
134. Miick SM, Casteel KM, Millhauser GL. *Biochemistry.* 1993; 32:8014. [PubMed: 8394121]
135. Fiori WR, Miick SM, Millhauser GL. *Biochemistry.* 1993; 32:11957. [PubMed: 8218270]
136. Decatur SM, Antonic J. *J Am Chem Soc.* 1999; 121:11914.
137. Decatur SM. *Biopolymers.* 2000; 54:180. [PubMed: 10861379]
138. Silva RAGD, Nguyen JY, Decatur SM. *Biochemistry.* 2002; 41:15296. [PubMed: 12484768]
139. Huang R, Kubelka J, Barber-Armstrong W, Silva RAGD, Decatur SM, Keiderling TA. *J Am Chem Soc.* 2004; 126:2346. [PubMed: 14982438]
140. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. *J Chem Phys.* 1984; 81:3684.
141. Silva RAGD, Kubelka J, Bou P, Decatur SM, Keiderling TA. *Proc Natl Acad Sci USA.* 2000; 97:8318. [PubMed: 10880566]
142. Hamm, P.; Zanni, MT. *Concepts and Methods in 2D IR Spectroscopy.* Cambridge University Press; Cambridge, UK: 2011.
143. Cochran AG, Skelton NJ, Starovasnik MA. *Proc Natl Acad Sci USA.* 2001; 98:5578. [PubMed: 11331745]
144. Smith AW, Chung HS, Ganim Z, Tokmakoff A. *J Phys Chem B.* 2005; 109:17025. [PubMed: 16853169]
145. Wang J, Zhuang W, Mukamel S, Hochstrasser RM. *J Phys Chem B.* 2008; 112:5930. [PubMed: 18078331]
146. la Cour Jansen T, Knoester J. *Biophys J.* 2008; 94:1818. [PubMed: 17981904]
147. Cheatum CM, Tokmakoff A, Knoester J. *J Chem Phys.* 2004; 120:8201. [PubMed: 15267740]
148. Dijkstra AG, Knoester J. *J Phys Chem B.* 2005; 109:9787. [PubMed: 16852179]
149. Moran A, Mukamel S. *Proc Natl Acad Sci USA.* 2004; 101:506. [PubMed: 14704267]
150. Clark A, Cooper GJS, Lewis CE, Morris JF, Willis AC, Reid KBM, Turner RC. *The Lancet.* 1987; 2:231.
151. Lorenzo A, Razzaboni B, Weir GC, Yankner BA. *Nature.* 1994; 368:756. [PubMed: 8152488]
152. Dupuis NF, Wu C, Shea J-E, Bowers MT. *J Am Chem Soc.* 2009; 131:18283. [PubMed: 19950949]
153. Ling YL, Strasfeld DB, Shim S-H, Raleigh DP, Zanni MT. *J Phys Chem B.* 2009; 113:2498. [PubMed: 19182939]
154. Knight JD, Miranker AD. *J Mol Biol.* 2004; 341:1175. [PubMed: 15321714]
155. Jayasinghe SA, Langen R. *Biochemistry.* 2005; 44:12113. [PubMed: 16142909]
156. Knight JD, Hebda JA, Miranker AD. *Biochemistry.* 2006; 45:9496. [PubMed: 16878984]
157. Jayasinghe SA, Langen R. *Biochim Biophys Acta.* 2007; 1768:2002. [PubMed: 17349968]
158. Brender JR, Lee EL, Cavitt MA, Gafni A, Steel DG, Ramamoorthy A. *J Am Chem Soc.* 2008; 130:6424. [PubMed: 18444645]

159. Engel MFM, Khemtémourian L, Kleijer CC, Meeldijk HJD, Jacobs J, Verkleij AJ, de Kruijff B, Killian JA, Höppener JWM. Proc Natl Acad Sci USA. 2008; 105:6033. [PubMed: 18408164]
160. Smith PES, Brender JR, Ramamoorthy A. J Am Chem Soc. 2009; 131:4470. [PubMed: 19278224]

Figure 1.
Molecular structures of (a) NMAD, (b) ACED, (c) AAO, and (d) AGO.

Figure 2. Simulated and experimental absorption spectra of (a) NMAD and (b) ACED in different solvents. The experimental spectrum of NMAD in D₂O is from DeCamp et al.⁹⁷ The experimental spectrum of ACED in CHCl₃ has two peaks and is not shown in the figure.

Figure 3. Theoretical and experimental¹²⁵ IR absorption spectra of AAO and AGO in aqueous D₂O solutions.

AKA	Ac-AAAAKAAAAKAAAAKAAAAKAAAAAY-NH ₂
Trpzip2	⁺ H ₃ N-SWTWENGKWTWK-NH ₂
rIAPP	⁺ H ₃ N-KCNTATCATQRLANFLV RSSNNLGPVLPPT NVGSNTY-NH ₂
hIAPP	⁺ H ₃ N-KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY-NH ₂

Figure 4.

Sequences of peptides in this study: the AKA peptide, Trpzip2, rIAPP, and hIAPP. Different residues between rIAPP and hIAPP are shown in red in the rIAPP sequence. In both IAPP proteins, Cys2 and Cys7 are connected with disulfide bonds.

Figure 5.
Representative snapshots of (a) AKA and (b) Trpzip2.

Figure 6. Theoretical and experimental IR absorption spectra of (a) unlabeled⁸⁷ and (b) [12] labeled³² AKA peptide.

Figure 7. $^{13}\text{C}=^{18}\text{O}$ labeled peak frequencies as a function of residue number for the AKA peptide. The theoretical isotope shift is taken to be -70 cm^{-1} . Experimental 32 isotope-labeled peak frequencies are shown for residues 12–15.

Figure 8. Theoretical and experimental⁸⁸ IR absorption spectra of (a) unlabeled, and (b) T3*–T10* labeled Trpzip2 in D₂O.

Figure 9.

¹³C-labeled peak frequencies as a function of residue number for Trpzip2. Residue 13 is the side chain chromophore on Asn6. The theoretical isotope shift is taken to be -43 cm^{-1} . Also shown are experimental^{88,89} isotope-labeled peak frequencies at residues 2, 3, 7, and 10.

Figure 10. Representative snapshots of rIAPP monomer in (a) folded, and (b) random coil conformation.⁹⁰.

Figure 11. Theoretical (weighted average) and experimental⁹⁰ IR line shapes in the amide I stretch region of rIAPP in D₂O. Also shown are line shapes for the folded and random coil states, weighted by their relative probabilities.

Figure 12. Representative snapshots of hIAPP monomer in (a) α -helical, (b) β -hairpin, and (c) random coil conformation.⁹¹

Figure 13. Theoretical (weighted average) and experimental⁹¹ IR line shapes in the amide I stretch region of hIAPP in D₂O. Also shown are line shapes for the α -helical, β -hairpin, and random coil states, weighted by their relative probabilities.

Table 1

Biexponential Fit Parameters from the Pump-Probe Measurements of NMAD and ACED in Different Solvents

solute	solvent	<i>f</i>	<i>T_s</i> (ps)	<i>T_l</i> (ps)
NMAD	D ₂ O ⁹⁷	0.55	0.20 ± 0.02	0.86 ± 0.09
NMAD	DMSO	0.778	0.66 ± 0.09	3.1 ± 0.7
NMAD	CHCl ₃	0.769	1.06 ± 0.10	5.2 ± 0.7
ACED	D ₂ O	0.753	0.13 ± 0.05	1.01 ± 0.11
ACED	DMSO	0.651	0.39 ± 0.05	2.27 ± 0.12

Table 2
Summary of Simulated and Experimental IR Line Shape Parameters for NMAD and ACED in Different Solvents^a

	ω_{D_2O}	Γ_{D_2O}	ω_{DMSO}	Γ_{DMSO}	ω_{CHCl_3}	Γ_{CHCl_3}
sim (NMAD)	1621	34	1656	22	1668	16
exptl (NMAD)	1623 ⁹⁷	28 ⁹⁷	1659	23	1665	26
sim (ACED)	1634	40	1666	22	1697	
exptl (ACED)	1633	37	1664	22	1700	

^a All quantities are in units of cm^{-1} .

Table 3Temperature Dependence of Theoretical and Experimental¹²³ Peak Frequencies of NMAD in D₂O^a

$T_{\text{sim}}/^{\circ}\text{C}$	$\omega_{\text{sim}}/\text{cm}^{-1}$	$T_{\text{exptl}}/^{\circ}\text{C}$	$\omega_{\text{exptl}}/\text{cm}^{-1}$
5.0	1621.4	0.7	1621.4
85.0	1623.5	83.7	1627.7

^aPeak frequencies at the lowest and highest temperatures studied are listed.

Table 4Theoretical and Experimental³⁰ Coupling Constants for the AKA Peptide^a

	β_{12}	β_{13}	β_{14}
theory	5.2 ± 0.3	-1.5 ± 0.2	-5.6 ± 0.2
experiment	8.5 ± 1.8	-5.4 ± 1.0	-6.6 ± 0.8

^aTheoretical values are averaged over residues 2–20. All coupling constants are in cm^{-1} .