Solvent and conformation dependence of amide I vibrations in peptides and proteins containing proline

Solvent and Conformation Dependence of Amide I Vibrations in Peptides and Proteins Containing Proline

Santanu Roy[1], Joshua Lessing[2], Georg Meisl[2], Ziad Ganim[3],
Andrei Tokmakoff[2], Jasper Knoester[1], and Thomas L. C. Jansen[1]
[1] *Center for Theoretical Physics and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands.*
[2] *Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.*
[3] *Physik-Department E22, Technische Universität München, Germany.*

We present a mixed quantum-classical model for studying the amide I vibrational dynamics (predominantly CO stretching) in peptides and proteins containing proline. There are existing models developed for determining frequencies of and couplings between the secondary amide units. However, these are not applicable to proline because this amino acid has a tertiary amide unit. Therefore, a new parametrization is required for infrared-spectroscopic studies of proteins that contain proline, such as collagen, the most abundant protein in humans and animals. Here, we construct the electrostatic and dihedral maps accounting for solvent and conformation effects on frequency and coupling for the proline unit. We examine the quality and the applicability of these maps by carrying out spectral simulations of a number of peptides with proline in D_2O and compare with experimental observations.

I. INTRODUCTION

Investigating the three dimensional structure and dynamics of proteins is crucial to understand their biological functions. Numerous experimental techniques allow the examination of the complex structures of proteins and their conformational changes over a wide range of time scales. For example, nuclear magnetic resonance (NMR [1, 2]) is extensively used for atomistic structure determination in solution and probing folding or unfolding dynamics on microseconds and longer time scales [3, 4]. Other techniques, such as small angle X-ray scatering [5, 6], circular dichroism [7], fluorescence [8], Raman [9], Fourier transform infrared (FTIR) and two dimensional infrared (2DIR) [10–13] spectroscopy are excellent complementary tools that provide insights on local and global structure and dynamics.

Infrared spectroscopy is a great tool for studying local secondary structures of proteins, i.e. α-helix and β-sheet structures [14], which are stabilized by the hydrogen bonding between carbonyl groups and amide hydrogens. The stretching vibrations of these carbonyl groups form the amide I

band (1600-1700 cm^{-1}) which exhibits strong infrared absorption. The amide band for an α-helix is located between 1640 cm^{-1} and 1660 cm^{-1} with two subbands (A and E_1) separated by 10 cm^{-1}, while for a β-sheet structure two peaks are observed at 1640 and 1680 cm^{-1} [15–17]. The 2DIR technique distinguishes β-sheet structures by producing Z shape like spectra [18, 19] when this particular structural motif is present. In the past decades a number of 2DIR studies have been reported of peptides and proteins in solution [16, 20–33] or confined in membranes [34–38], revealing structural details and conformational changes from femtosecond (fs) to nanosecond (ns) time scales, and the nature of dynamic environments. For the structure determination of peptides that are in gas phase or micro-solvated (surrounded by few solvent molecules) the mid-infrared spectroscopic technique has become a promising tool [39–42].

Interpreting experimental 2D spectra is a great challenge and modeling of the amide I band is required to understand the spectra. Exisiting mixed-quantum classical models which describe the frequency and interaction of amide vibrations located at each peptide site of a protein are applicable to all proteinogenic amino acids except proline. Here, we present a model to describe the amide I vibrational dynamics of proline containing proteins. Proline is a unique amino acid because the side chain binds to its backbone nitrogen forming a tertiary amide unit, which results in a lower vibrational amide I frequency than the secondary amide I unit found in other amino acids. Proline is typically located in γ- and β-turns of proteins and plays an important role in determining folding rates of proteins [43, 44]. Proline is an integral amino acid in proteins in connective tissue such as elastin and collagen. Proline doesn't typically contribute to the standard delocalized amide I vibrational modes, because of its site frequency shift, and therefore it is generally considered to be more of a local mode. To study the amide I vibrations in proteins with proline we have constructed a Density Functional Theory (DFT) based parametrization for the vibrational properties of the tertiary amide unit.

For the secondary amide units there are already existing frequency and coupling maps [45–58]. N-methyl acetamide (NMA), a simple model system for a single secondary amide unit, was considered to construct these maps. For example, for the Cho4 map [47] restricted Hartree-Fock (RHF) calculations were carried out on NMA-D in water molecules with the CHELPG charges [59] and the frequency was correlated with the electrostatic potential at C, O, N, and D atoms. The construction of the Skinner map [49] is based on DFT calculation at B3LYP/6-3111G** level of theory on the deuterated NMA (NMA-D) surrounded by TIP3P water molecules [60]. The frequency was considered to be a linear function of the electric field at the C, O, N, and D atoms due to charges from the water molecules. An electrostatic map was constructed where the

frequency was correlated with the electric field and electric field gradients at C, O, N, and D atoms of the NMA-D molecule [53]. This map may be considered transferable because general point charge environments were considered instead of a particular solvent force field. The transferability was tested by simulating NMA-D in a number of solvents and comparing with experiments. The mapping was found to work well in polar solvent, but fail in non-polar solvent where the direct electrostatic effects are less dominant.

To describe the interaction between the amide I modes, the electrostatic map was later combined with maps describing the nearest neighbor interaction (Ramachandran angle based nearest neighbor frequency shifts (NNFS) and nearest neighbor coupling (NNC)) as well as the long range through space transition charge coupling (TCC) [54]. In predicting spectral line shapes this combined parametrization works in excellent agreement with the experiments, apart from a systematic shift in the peak position [29, 32, 56, 61]. Recently, empirical maps [57] have been developed for the amide I vibrations in the backbone and side chains of proteins. This was done using the experimental spectra and the vibrational life time of NMA and acetamide in both polar and non-polar solvents. However, all the maps reported till now are applicable to a secondary amide unit, and thus not to the tertiary amide unit in proline. Here, we construct the electrostatic map for the solvent effect, nearest neighbor maps, and long range TCC map for proline using the mapping methods essentially identical to those reported in Ref. 53, 54. We can use this parametrization to simulate infrared spectra for proteins with proline to reveal the structural and dynamical information, especially related with the hydrogen bonding. Applying these maps we carry out spectral simulations of a number of proline containing peptides: AcP (N-Acetyl-LPro-COOH), VP (N-Acetyl-Val-LPro-COOH), PG (N-Acetyl-LPro-Gly-COOH) and the β-hairpin peptide, PG12 (Ac-RYVEV-DPro-GKKILQ-NH2) [27]. We compare the simulated FTIR and 2DIR spectra with the experimental ones to examine the frequencies and line shapes and determine the quality of the maps and to what extent these can be applied to other proteins and peptides containing proline.

The remainder of this paper is outlined as follows. In Section II we introduce the theory and the methods used to construct the map. In Section III we compare between simulated and experimental spectra and discuss the applicability of the maps. Finally, we draw conclusions in Section IV.

II. THEORY AND METHODS

Modeling of the amide I band of a peptide requires treatment of the electrostatic interaction between amide I oscillators and solvent molecules as well as interactions between the oscillators

of the peptide. The electrostatic interaction within the peptide can be broken down into two parts: nearest neighbor and long range interactions. Based on these interactions frequencies of and couplings between the amide I oscillators are calculated. The following floating oscillator Hamiltonian is generally employed to represent a system of N amide I oscillators [10].

$$H(t) = \sum_{i=1}^{N} \left[\omega_i(t) b_i^\dagger b_i - \frac{\Delta_i(t)}{2} b_i^\dagger b_i^\dagger b_i b_i \right] + \sum_{i,j}^{N} J_{ij}(t) b_i^\dagger b_j + \sum_{i=1}^{N} \vec{\mu_i}(t) \cdot \vec{E}(t) \left[b_i^\dagger + b_i \right] \tag{1}$$

Here, i labels the amide I units counting in the direction from the N-terminus to the C-terminus of a peptide. $b_i^\dagger$ and b_i are the Bosonic creation and annihilation operators for the amide I vibration on site i. $\omega_i(t)$ and $\Delta_i(t)$ are the fluctuating frequency and anharmonicity, respectively, of the i^{th} site, $J_{ij}(t)$ is the fluctuating coupling between the i^{th} and the j^{th} site. The external electric field $\vec{E}(t)$ interacts with the system through its transition dipole $\vec{\mu_i}(t)$. The site frequency $\omega_i(t)$ is expressed as

$$\omega_i(t) = \omega_i^{\text{gas}} + \delta\omega_i^{\text{solvent}}(t) + \delta\omega_i^{\text{NN}}(t) \tag{2}$$

ω_i^{gas} is the frequency of the i^{th} amide I oscillator without the influence from surroundings and non-nearest neighbor protein units. $\delta\omega_i^{\text{solvent}}(t)$ is the frequency shift due to the interaction between the oscillator and the solvent. $\delta\omega_i^{\text{NN}}(t)$ is the frequency shift due to the influence from the nearest neighbors (site $i-1$ and $i+1$). Except the nearest neighbors the peptide is treated as a part of the solvent. We predict the solvent shift and the nearest neighbor shift using an electrostatic map [53] and a dihedral angle based map [54], respectively. The coupling, J_{ij} has two parts: nearest neighbor coupling for $j = i \pm 1$, and long range coupling for other j, treated with the TCC scheme [54, 62].

A. Electrostatic map for proline

To construct the electrostatic map we chose the N-N-Dimethyl acetamide (DMA) molecule which is obtained by truncating the proline ring (see Fig. 1a). This was done because the truncation of the ring hardly affects the gas phase amide I frequency except for a small systematic shift. Using the smaller molecule reduces computational costs. However, the full ring was included while considering the nearest neighbor map, where it is important to account for the rigidity of the proline ring.

Density functional theory (DFT) based geometry optimization of DMA was done with the ORCA package [63] using the Def2-TZVP basis sets [64, 65] and the revised Perdew-Burke-Ernzerhof (RPBE) exchange correlation functional [66, 67]. At the same level of theory, normal

modes and frequencies were obtained constructing the Hessian calculated numerically by distorting the geometry from its equilibrium. In this procedure all the methyl groups were frozen and the position of the carboxyl carbon and oxygen as well as the nitrogen were varied. The methyl groups were fixed to make sure that the same amide normal mode can be used in further calculations. To construct the amide I Hamiltonian the following procedure was followed. A potential energy surface ($V(r)$) was obtained as a function of the distortion from the equilibrium geometry with a step size of 0.06 Å using the amide I normal mode. $V(r)$ was expanded in a Taylor series upto 6^{th} order.

$$V(r) = V_0 + \sum_{n=1}^{6} V_n r^n \tag{3}$$

Using a least-square fit the coefficients were extracted. Then, the Hamiltonian ($H = \frac{P^2}{2m} + V(r)$) was obtained by transforming the coordinate (r) and the momentum (P) to b and $b^\dagger$ [68, 69]. This Hamiltonian was then expanded using the lowest 21 harmonic basis functions. Diagonalizing the Hamiltonian we obtained the eigen frequencies and vectors for the ground state (g) and the singly (e) and doubly (f) excited states. Using the first order derivative of the dipole moments obtained from DFT calculations and the dipole operator, $\vec{\mu} = \vec{\mu_0} + \frac{\partial \vec{\mu}}{\partial r}\sqrt{\frac{\hbar}{2m\omega}}(b^\dagger + b)$, the transition dipole moments (μ_{eg}, μ_{fe}, and μ_{fg}) were calculated.

In order to model the solvent shift we followed a Stark shift based method, where, the change in a molecular quantity ($\delta\Omega$) (frequency shift, transition dipole moment, etc) can be expressed as

$$\delta\Omega = \langle C_\Omega | E \rangle. \tag{4}$$

Here, $|E\rangle$ is a vector with electric field parameters and $|C_\Omega\rangle$ is the map vector. We considered the x and y components of the electric field and the zz and xy components of the electric field gradient ($\frac{\partial E_z}{\partial z}$ and $\frac{\partial E_y}{\partial x}$) on the C, O, N, and C_δ atoms of DMA (Fig. 1a) to parametrize the map. Here, x, y, and z refer to the axis system fixed to the molecular frame as shown in Fig. 1b. For constructing the map, the fundamental (ω_{eg}) and overtone (ω_{fg}) frequencies and transition dipole moments μ_{eg}, μ_{fe}, and μ_{fg} were calculated in the gas phase as well as in 74 different point charge environments to make sure that all the electric field and gradient components were sufficiently sampled. We applied both the least squares fit (LSF) method and the singular value decomposition (SVD) method to fit these molecular properties to the electric field and gradient components. The purpose of using two different fitting methods is to check if there exist any linear dependencies of the fit coefficients [53]. In case of linear dependencies the fit constants will get large values in the LSF method and smaller values in the SVD method. We found almost identical fit coefficients ($|C_\Omega\rangle$), which shows

that there is no linear dependency. The LSF coefficients (listed in Table I and Table II) will be used later as the map coefficients.

We note here that the experimental gas phase frequency of DMA (1690 cm^{-1} [70]) was underestimated in the DFT calculation by a factor of 1.0157. Hence, this factor was used as a scaling factor for the calculated frequency to reproduce the gas phase experimental frequency. After this scaling, the frequency for the doubly excited state was estimated to be 3366.09 cm^{-1}. The corresponding anharmonicity, Δ is then 13.9 cm^{-1}. We have made a comparison between frequencies obtained from the DFT calculations ($\omega_{\rm DFT}$) and frequencies calculated using these map coefficients ($\omega^{\rm tert}_{\rm Map}$) for the 74 charge environments (Fig. 2a). We find a strong correlation between the frequencies calculated with both methods, demonstrating that the DFT frequencies are well represented by the map. We compared this map with the map developed for the secondary amide I unit [53], but with a shifted gas phase frequency from 1717 cm^{-1} (experimental gas phase frequency of NMA-D [71]) to 1690 cm^{-1}. The frequencies estimated from the map for the secondary amide I unit ($\omega^{\rm sec}_{\rm Map}$) are compared with $\omega_{\rm DFT}$ in Fig. 2b. While the overall behavior is quite well represented by such mappings, more outliers are found compared to the new map. The slope of the line passing through these points is 0.91 and $\omega^{\rm sec}_{\rm Map}$ is still correlated with $\omega_{\rm DFT}$. Thus, applying a 27 cm^{-1} shift of the original map and scaling the frequency shift could be a working approximation. Of course, having the more accurate new map it is preferable to apply that.

B. Dihedral map for proline

To construct the dihedral map for proline we systematically chose Gly-Pro and Pro-Gly configurations. The trans configurations of these peptides are depicted in Fig. 3a and Fig. 3b. For both the peptides the Ramachandran angles ($\Phi = C - N - C_\alpha - C$, $\Psi = N - C_\alpha - C - N$ [72]) were varied from -180 to +180 with a separation of 30 degree between every two conformations and each conformation was optimized by the DFT method for a fixed Φ and a fixed Ψ. The optimized energy landscapes for the *trans*-Gly-Pro and *trans*-Pro-Gly are given in the Fig. 3c and 3d. The energy minimum is set to zero. It turns out that for *trans*-Pro-Gly the region 90°<Φ <180°is energetically unfavorable due to strain in the proline ring and the corresponding conformations are unlikely to occur. Therefore, these conformations were neglected for the coupling and frequency calculations.

The normal mode frequencies and vectors were calculated for all the optimized conformations. Hessian reconstruction [48] was employed to obtain site frequencies and couplings between the neighboring sites. The magnitude of the CO stretch vibration was used to obtain the eigenvector

matrices needed in the reconstruction [50]. All frequencies for the secondary amide unit (site i-1 and i+1 in Fig. 3a and 3b) and the tertiary amide unit (site i in Fig. 3a and 3b) were scaled with a factor of 1.02157 and 1.01864, respectively. This was done because in the DFT calculations the NMA and DMA gas phase frequencies (here methyl groups were not fixed) were underestimated by these factors compared to the experimental ones. The couplings were scaled with the average factor (1.020105). **Although the change in the couplings due to this scaling is not significant, we scaled them for consistency.** The couplings between site i-1 and i and between site i and i+1 are depicted in Fig. 3e and Fig. 3f, respectively. The frequency shifts of the oscillators i-1 and i+1 were calculated by subtracting the NMA gas phase frequency (Fig. 3g and 3j) and for the frequency shift of site i the DMA gas phase frequency was subtracted (Fig. 3i and 3h). The couplings and frequency shifts for the *cis*-configurations were also obtained. These can be found in the supplementary material. The couplings are stronger for the *trans*-configurations than the *cis*-configurations. These dihedral maps are for L-proline and can be converted to the maps for D-proline by $\Phi \rightarrow -\Phi$ and $\Psi \rightarrow -\Psi$ transformations.

Proline can be found in an α-helical or β-sheet structure forming a turn, where the corresponding Ramachandran angles [73] are approximately, Φ = -61°, Ψ = -35° and Φ = -65°, Ψ = 150°, respectively. For another frequently occurring structure, prolyproline II (PP_{II} [74]) the Ramachandran angles are around Φ = -75° and Φ = 145°. We have extracted the couplings between two nearest neighbor sites and their frequency shifts ($\delta\omega_C$: shift of the first one induced by the second one, and $\delta\omega_N$: shift of the second one induced by the first one) from the dihedral maps ($NN_{Pro\text{-}Gly}$) and $NN_{Gly\text{-}Pro}$) as presented in Fig. 3(c-j) and also from the dihedral maps developed on glycine dipeptide ($NN_{Gly\text{-}Gly}$[54]). From Table III it appears that the couplings extracted from both maps are found to be within 0.9 cm^{-1} of each other. The difference in frequency shifts is larger, typically with a significant red-shift of the amide I frequency of proline. **In the experimental 2DIR study of the alanine dipeptide [75] the coupling between two amide units was reported as 1.5 ± 0.5 cm−1. The peptide is thought to exist in the PP_{II} conformation for which our mapping predict a coupling of 3 cm^{-1}. In reality of course in solution a broad distribution of conformations exist and the exact weight of those is difficult to predict [76]. Therefore, verifying the map against gas phase measurements [39–42] where the conformation is well known will be more reliable.**

C. TCC map

The transition charge coupling model [54, 62] is based on the electrostatic interaction between the charge clouds in the vicinity the site oscillators. A point charge approximation of the charge clouds is made, i.e. the point charges are positioned on the atomic sites. Therefore, the charge density in the vicinity of a site oscillator i, can be expressed as

$$\rho(x_i, \mathbf{r}_i) = \sum_j \delta(\mathbf{r}_i - \mathbf{r}_j(x_i))(Q_j + \delta Q_j). \tag{5}$$

Here, $r_j(x_i)$ is the position vector of the j^{th} atom while x_i, Q_j, and δQ_j are oscillator displacement, partial charge of the j^{th} atom, and the derivative of the partial charge with respect to the oscillator coordinate, respectively. The position vector is evaluated as $r_j(x_i) = r_j(0) + x_i v_j^i$, where v_j^i is the normal mode coordinate for the j^{th} atom of the i^{th} oscillator. Now, the coupling between the non-nearest neighbor sites, site i and site j is

$$J_{ij}^{\mathrm{TCC}} = \frac{1}{4\pi\epsilon_0}\left[\frac{\partial^2}{\partial x_i \partial x_j}\sum_{k,l}\frac{(Q_k + \delta Q_k x_i)(Q_l + \delta Q_l x_j)}{|r_k(x_i) - r_l(x_j)|}\right]_{x_i=0, x_j=0}. \tag{6}$$

For the TCC parametrization of proline amide unit, we used the Mulliken charge analysis of the DMA molecule to estimate the partial charges of its atoms. The partial charges of the atoms of each methyl group were summed up and assigned to the carbon atom to which these are covalently connected. All methyl groups were kept frozen while calculating the normal modes required for this model. The TCC parameters (transition charges and normal mode coordinates) are listed in Table IV. **In the original TCC model [62] a parametrization for NMA-D was obtained with the B3LYP exchange correlation functional. The present parameters obtained here for DMA with the RPBE functional are quite different even in the partial charges because of the replacement of the deuterium with a methyl group. This also results in differences in the normal mode coordinates. The actual numbers are compared in the supplementary material (Table VI).**

III. VALIDATION OF THE MAP

A. Experiments

FTIR and 2DIR experiments were performed on four peptides: AcP, VP, PG, and PG12. AcP was synthesized and purified by Sigma-Aldrich, Saint Louis MO. VP and PG were synthesized

and purified by NeoBioScience, Cambridge MA. PG12 was synthesized and purified by the MIT Biopolymers Laboratory, Cambridge MA. All samples were lyophilized repeatedly against a solution of DCl in D_2O, Cambridge Isotope Laboratories, Andover MA, to remove trifluoroacetic acid and acidic protons. Finally the AcP, VP, PG and PG12 samples were dissolved in pH = 1.0 DCl in D_2O at a concentration of 15 mg/ml, pH = 1.0 DCl in D_2O at a concentration of 15 mg/ml, and pH 3.85 deuterated acetate buffer at a concentration of 10 mM, respectively.

The sample cell used for all measurements consisted of a brass sample holder that held two 1 mm thick CaF_2 windows that were placed on either side of a 50 micrometer Teflon spacer. The brass sample holder was mounted inside a brass cooling jacket that was thermally regulated to T = 10° C by a recirculating water chiller. Infrared absorption spectra were collected using a Thermo Scientific Nicolet 380 FTIR Spectrometer. The spectra for the deuterated AcP, VP, and PG were acquired at a spectral resolution of 2.0 cm^{-1} and averaged over 64 one-second scans. The AcP, VP, and PG 2DIR data were collected as described previously [77], using a rephasing evolution time of $t_1 = 3.1$ ps and a nonrephasing evolution time of $t_1 = 2.5$ ps in 4 fs time steps. The PG12 data was collected as described in Ref. 27.

B. Simulations

The structures of AcP (Fig. 4a), VP (Fig. 7a) and PG (Fig. 8a) were built using the Gabedit software [78] and then optimized with the DFT method using the same basis sets and exchange correlation function as used for constructing the proline map. These optimized structures were used as the initial structures of MD simulations. For PG12 (Fig. 10a), the initial structure was taken from a set of NMR structures provided by Prof. Samuel Gellman at the University of Wisconsin, Madison. All three peptides were solvated in D_2O to mimic the experiments and all the acidic protons were replaced with deuteriums. PG12 has two lysine, one arginine and one glutamic acid, which have charged side chains under the experimental conditions (pD 4.25). The total formal charge of the system is +2. Two chlorine ions were added to keep the system neutral in the simulation. All the solvated systems were subjected to MD simulations performed with Gromacs-4.0.7 [79]. The all atoms OPLS force field [80] was used to describe the peptides and for the water the SPC/E force field [81] was used. The MD simulations were done with 2 fs time steps using the LINCS algorithm [82] to constrain all bond lengths. For the long range electrostatic interactions a cut-off of 1.4 nm was employed with the reaction field method. The trajectories of all three peptides were obtained in the canonical (NVT) ensemble at 300 K. The Nośe-Hoover

thermostat [83, 84] was used to couple the system to a constant temperature bath (300 K). The length of each trajectory was 10 ns, where the last 8 ns was used for the spectral simulations. For all trajectories snapshots were saved every 20 fs and for every snapshot the Hamiltonian (Eq. 1) was constructed. For a secondary amide unit the parametrization reported in Ref. 53, 54 was used, while the new parametrization developed for proline was used for a tertiary amide unit. The time dependent Hamiltonian was used in spectra calculations with the *Numerical Integration of the Schrödinger Equation* (NISE) scheme [85, 86]. In this scheme the Hamiltonian is considered to be constant for short time intervals (20 fs) during which the time independent Schrödinger equation can be solved. The time evolution for the total time is determined by the successive propagation during the short time intervals. Linear and nonlinear response functions are then calculated to obtain FTIR and 2DIR spectra. An *ad hoc* vibrational lifetime (1 ps) [87] is accounted for in the spectral simulations.

The line shape analysis of FTIR or 2DIR spectra give a quantitative comparison between simulations and experiments [88–90]. Here, we analyze the Full Width at Half Maximum (FWHM) of FTIR peaks, diagonal (DW) and *anti*-diagonal width (ADW) of 2D peaks, and nodal slopes. We define DW as the FWHM of the slice through the maximum intensity of the diagonal peak of a 2D spectrum. ADW is the FWHM of the slice through the diagonal peak in the *anti*-diagonal direction centered at a particular ω_1. The nodal slope is the slope of the nodal line passing between the diagonal and overtone peak, where the intensity is zero.

C. Results and Discussion

AcP contains a single amide unit and can be used to validate only the electrostatic map for proline. The simulated FTIR spectrum of this peptide is compared with the experimental one in Fig. 4b. The experimental spectrum has two peaks: the first one centered at 1607 cm^{-1} corresponds to the proline amide I vibration and the second one centered at 1711 cm^{-1} corresponds to the CO vibration in the acid carboxyl group. The later is not included in our spectral simulation and we get a single peak due to the proline amide unit. The simulated spectrum is red shifted by 31.7 cm^{-1} to match with the experimental peak position for the proline amide unit. Apart from this shift and slightly smaller asymmetry than the experimental spectrum, the line shape of the simulated spectrum (FWHM = 29.5 cm^{-1}) is in good agreement with the experimental one (FWHM = 30.1 cm^{-1}). We note here that the FTIR peak for NMA-D in D_2O is almost as broad (FWHM = 27.8 cm^{-1} [53]) as for AcP. **The origin of the discrepancy between the experimental and**

simulated frequencies observed for AcP and for the other peptides described below is discussed in detail at the end of this section.

The 2DIR spectra of AcP at different waiting times (t_2 = 0 fs, t_2 = 300 fs, and t_2 = 600 fs) are depicted in Fig. 5. The diagonal peak (red) of a 2D spectrum comes from the stimulated emission and ground state bleach, whereas the overtone peak (blue) appears due to the excited state absorption. Here, the shape of the 2D peaks in the simulated spectra as a function of waiting time at first glance appear to be in very good agreement with the experimental one. To be more quantitative, we made a nodal slope analysis of these 2D spectra. The nodal lines are plotted on the 2D spectra in Fig. 5. We estimated the slope of the nodal lines in the region, $1600 < \omega_1 < 1620$, making sure that the nodal line is straight in this region. We find an exponential decay of the nodal slope (Fig. 6) with a time scale $\sim$ 1.5 ps for both experiment and simulation. **The small differences between the simulated and experimental nodal slopes at the early time regime might be due to slightly slower dynamics of the solvent in the MD simulation than in reality or by the simple fact that the laser pulses have a finite pulse width, while they are taken to be instantaneous in the spectral simulation.**

In the spectral simulation of the dipeptides we can combine the electrostatic and the nearest neighbor dihedral maps. For the VP dipeptide, the experimental FTIR spectrum (Fig. 7b) contains two peaks. The peak at 1714 cm^{-1} originates from the terminal acid carboxyl. Both of the two amide units contribute to the first peak (1615 cm^{-1}). This is also seen in the simulated spectrum which has only one peak. Like for AcP, here, similar frequency shifts of both units (28.7 cm^{-1}) are made to match to the experimental peak at 1615 cm^{-1}. After this shift the frequencies of the two sites are found to be 1609.7 cm^{-1} and 1614.1 cm^{-1}. The simulated linewidth (FWHM = 31.9 cm^{-1}) is close to that found in experiment (FWHM = 30.3 cm^{-1}). One possible reason of the same shift for both AcP and VP can be attributed to the presence of the C-terminal carboxyl group. This argument will be justified later when we will discuss the spectra for PG12, where this terminus is absent. The 2DIR spectra of VP at t_2 = 0 fs and t_2 = 600 fs are shown in Fig. 7c-f. DW and ADW extracted from these spectra are listed in Table V. The simulated DWs are within $\sim$2-4 cm^{-1} of the experimental ones. The simulated and experimental ADW deviate less than 1 cm^{-1} from each other.

The FTIR and 2DIR spectra of PG are given in Fig. 8b-d. A two peak structure is observed in spectra: the Pro peak (1607.6 cm^{-1}) corresponding to proline amide unit and the Gly peak (1650 cm^{-1}) corresponding to the amide unit between proline and glycine. The respective frequency red-shifts for these sites made in simulation to reproduce the experimental ones are 26.6 cm^{-1}

and 12.7 cm^{-1}. One interesting point to note here is that both experimental and simulated 2D spectra for PG show ridge like cross peaks around 1613 cm^{-1} which originate from the coupling of the proline amide unit to its C-terminal neighbor, i.e. glycine. The two peak structure of PG and single peak structure of VP may lead to the conclusion that the coupling of the proline amide unit to the C-terminal unit is stronger than the N-terminal unit. To investigate that we have compared between these couplings estimated from simulations. We find that the coupling in PG has a broader distribution than that in VP ((Fig. 9)), though both are centered at the same position (4.4 cm^{-1}). From the line shape analysis of 2D spectra for PG (Table VI) we find that the experimental Gly peak is well reproduced by the simulation. The simulated DW of the Pro peak is 4-5 cm^{-1} narrower than the experimental one, while ADW is reproduced within 1 cm^{-1}.

Finally, for PG12 with 12 backbone amide I oscillators, all the maps i.e. electrostatic map, dihedral map as well as TCC map can be applied. The charged side chain of glutamic acid and arginine also contribute to the amide I band. In the spectral simulation these two side chain amide I oscillators and the side chain as well as backbone amide I oscillators of glutamine are not considered. The analysis for the 5^{th} amide unit (proline amide unit, refered to as V5) in PG12 will help us understand whether the shifts of the simulated proline frequency in VP, PG, and AcP with respect to the experimental ones are due to the presence of the terminal or not. A comparison between the experimental and simulated FTIR spectra is represented in Fig. 10b-e. It turns out that a systematic shift of 14 cm^{-1} is required for all the amide I site frequencies except for the proline amide unit to match with the experimental peak (Fig. 10b). The required shift for the proline amide unit is negligible (only 3 cm^{-1}). These shifts arise because the electrostatic map was developed only considering the amide I group, hence, the boundary between the amide I groups and the surrounding peptide is not well defined. This type of shift was previously observed in other studies as well [29, 31, 32, 56, 91]. When we take the difference between the experimental and the simulated FTIR spectrum (Fig. 10b), we find two pronounced peaks at 1590 and 1610 cm^{-1} which originate from the symmetric and asymmetric CN stretching vibrations of the deuterated arginine side chain [92]. To investigate this, we have done a simple modeling of these stretching vibrations by fitting a linear combination of two Gaussian functions (Eq. 7) to the difference spectrum (green curve in Fig. 10b).

$$f(\omega) = \alpha_{\rm s} \exp\Big[-\frac{(\omega_0^{\rm s}-\omega)^2}{2\sigma_{\rm s}^2}\Big] + \alpha_{\rm a} \exp\Big[-\frac{(\omega_0^{\rm a}-\omega)^2}{2\sigma_{a}^2}\Big] \tag{7}$$

Here, s and a stand for symmetric and and anti-symmetric stretch vibration. From the fit parameters (listed in Table VII) a FTIR spectrum for the CN stretching vibrations is obtained and

added to the simulated FTIR spectrum of PG12, then the total simulated spectrum reproduces the experimental spectrum in good agreement (Fig. 10c).

The FTIR and 2DIR spectra for PG12 with the V5 isotope labeled with $^{13}C^{16}O$ are given in Fig. 10d and Fig. 10f-g, respectively. In the simulated FTIR and 2DIR spectra, the peak corresponding to V5 (41 cm^{-1} red shifted with respect to its original peak position due to the labeling) is clearly isolated from the main band. **In the main band of the 2D spectra a Z-shape like structure is observed, which is consistent with the β-hairpin structure of PG12. This Z-shape like structure is much less pronounced than observed for the ideal β-sheet structure [18, 19]. Such features were also found in the studies of other β-hairpin peptides [24, 29].** In the 2D spectra in Fig. 10f-g ridge like cross peaks appear ($\omega_1 \sim 1625$ cm^{-1}, $\omega_3 \sim 1675$ cm^{-1}) due to couplings between the amide I oscillators. The peaks between the V5 peak and the main band in the experimental 2D spectrum are again due to the CN stretching vibration of the arginine side chain, which is absent in the simulated spectrum. Both the position and the shape of the simulated 2D spectrum for the labeled V5 unit fit nicely with the experimental observations. In the experimental isotope labeled FTIR spectrum for V5, the peak corresponding to V5 is overlapping slightly with one of the arginine side chain peaks, and not completely comparable with the simulated FTIR spectrum. We have therefore obtained the difference spectrum for V5 by subtracting the unlabeled FTIR spectrum from the V5 labeled FTIR spectrum, that isolates the spectrum for V5. We have done this for both the experimental and simulated spectra and compared in Fig. 10e. In the simulation, we predict the line shape of the linear difference spectrum for the proline amide unit in good agreement with the experiment.

We will now revisit the origin of the site frequency shifts in simulation of the peptides while compared to experiments. The electrostatic map was developed on DMA and the dihedral map was developed on the Pro-Gly and Gly-Pro dipeptides where the C-terminal -COOD unit was absent. But Acp, VP, and PG contain the C-terminal -COOD unit, which can affect the site frequencies. The proline frequencies were red-shifted by a similar amount ($\sim$30 cm^{-1}) in these peptides. In PG, the frequency shifts corresponding to the Gly peak in simulation is similar to the systematic shift found for PG12 and has the same explanation for the shift. To overcome the problems with studying a single peptide unit or dipeptides arising due the C-terminal, future study with explicit modeling of the C-terminal -COOD unit is recommended. Otherwise, for a dipeptide, a systematic shift (as reported for VP) of both sites can be a working approximation when proline is close to C-terminal -COOD group. When proline is close the N-terminal in a dipeptide, the shifts observed in the PG simulation can be used. We suggest this because, apart from these frequency shifts

the line shape agreement between simulations and experiments are good. In the case of PG12, proline is situated in the turn and its simulated frequency is almost the same as the experimental one. Like PG12, our maps can be applied to large peptides with proline located in the turn, just making a systematic 14 cm^{-1} frequency shift of the secondary amide I units, where no shift of the proline frequency is required. These maps can also be applied to the gas phase mid-infrared study of bare or micro-solvated peptides to determine amide I vibrational frequencies and corresponding backbone and hydrogen bonding structures.

IV. CONCLUSIONS

We have developed a parametrization applicable for studying structure and dynamics of proline containing proteins. More precisely, the modeling addresses probing the amide I vibration of the tertiary amide unit. To build up the parametrization, DFT calculations were used. The frequency of an amide I unit was modeled by accounting for the electrostatic interaction between the solvent and the amide I unit in addition to treating the electrostatic effects due to the conformational changes within the peptide. The coupling between two nearest neighbor amide I units is treated with an NNC map, for longer-range interactions the TCC map is employed. We have tested our model by simulating infrared spectra for the single amide I unit in AcP, two amide I units in the VP and PG dipeptides and finally twelve amide I units in PG12 and then comparing with experiments. It turns out that in all three cases, the line shapes of the simulated spectra are in good agreement with the experimental observations. In the dipeptide case, it turns out that proline is more strongly coupled to the C-terminal neighbor than the N-terminal neighbor in the amino acid sequence. The largest shortcoming of the model is that the presence of C-terminal -COOD shifts the proline frequency in simulation by a large amount ($\sim$ 30 cm^{-1}) with respect to the experimental one. But when proline is not a neighbor of the terminal (like in PG12), the shift of the simulated proline frequency is negligible. Proline is mostly found in the turn of large peptides or proteins and hardly ever located next to a terminal. Therefore, our model will be a good tool for studying amide I vibrational dynamics of proline containing peptides and proteins.

V. SUPPLEMENTARY INFORMATION

All the dihedral maps can be found in the supplementary material.

VI. ACKNOWLEDGEMENTS

T.L.C.J. acknowledges the Netherlands Organization for Scientific Research (NWO) for support through a VIDI grant. We thank Sam Gellman for providing the NMR structure of PG12. This work was supported by the National Science Foundation (CHE-0911107).

---

[1] M. Pfuhl and P. C. Driscoll, Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences **358**, 513 (2000).

[2] J. Balbach, J. Am. Chem. Soc. **122**, 5887 (2000).

[3] H. J. Dyson and P. E. Wright, Methods in Enzymology **394**, 299 (2005).

[4] A. G. Palmer III, C. D. Kroenke, and J. P. Loria, Methods in Enzymology **339**, 204 (2001).

[5] D. J. Segel, A. Bachmann, J. Hofrichter, K. O. Hodgson, S. Doniach, and T. Kiefhaber, J. Mol. Bio. **288**, 489 (1999).

[6] S. Arai and M. Hirai, Biophysical J. **76**, 2192 (1999), article.

[7] N. J. Greenfield, Nature Protocols **1**, 2876 (2006).

[8] C. A. Royer, Biophys. J. **68**, 1191 (1995).

[9] G. J. Thomas, J. Annu. Rev. Biophys. Biomol. Struct. **28**, 1 (1999).

[10] P. Hamm, M. H. Lim, and R. M. Hochstrasser, J. Phys. Chem. B **102**, 6123 (1998).

[11] M. Cho, Chem. Rev. **108**, 1331 (2008).

[12] S. Mukamel, Phys. Rev. A **61**, 021804R (2000).

[13] P. Hamm and M. Zanni, *Concepts and Methods of 2d Infrared Spectroscopy* (Cambridge Univ. Pr., New York, 2011).

[14] S. Krimm and J. Bandekar, Adv. Protein Chem. **38**, 181 (1986).

[15] M. Jackson, P. I. Haris, and D. Chapman, J. Mol. Struc. **214**, 329 (1989).

[16] Z. Ganim, H. S. Chung, A. W. Smith, L. DeFlores, K. C. Jones, and A. Tokmakoff, Acc. Chem. Res. **41**, 432 (2008).

[17] S. Woutersen and P. Hamm, J. Chem. Phys. **115**, 7737 (2001).

[18] N. Demirdöven, C. M. Cheatum, H. S. Chung, M. Khalil, J. Knoester, and A. Tokmakoff, J. Am. Chem. Soc. **126**, 7981 (2004).

[19] A. G. Dijkstra and J. Knoester, J. Phys. Chem. B **109**, 9787 (2005).

[20] M. T. Zanni, S. Gnanakaran, J. Stenger, and R. M. Hochstrasser, J. Phys. Chem. B **105**, 6520 (2001).

[21] M. T. Zanni, M. C. Asplund, and R. M. Hochstrasser, J. Chem. Phys. **114**, 4579 (2001).

[22] M. T. Zanni, N. H. Ge, Y. S. Kim, S. M. Decatur, and R. M. Hochstrasser, Biophys. J. **82**, 66 (2002).

[23] A. W. Smith, H. S. Chung, Z. Ganim, and A. Tokmakoff, J. Phys. Chem. B **109**, 17025 (2005).

[24] Z. Ganim and A. Tokmakoff, Biophys. J. **91**, 2636 (2006).

[25] J. Wang, J. Chen, and R. M. Hochstrasser, J. Phys. Chem. B **110**, 7545 (2006).

[26] H. S. Chung, Z. Ganim, K. C. Jones, and A. Tokmakoff, P. Nat. Acad. Sci. **104**, 14237 (2007).

[27] A. W. Smith and A. Tokmakoff, J. Chem. Phys. **126**, 045109 (2007).

[28] A. W. Smith and A. Tokmakoff, Angew. Chem. Int. Ed. **46**, 7984 (2007).

[29] T. L. C. Jansen and J. Knoester, Biophys. J. **94**, 1818 (2008).

[30] J. Wang, W. Zhuang, S. Mukamel, and R. M. Hochstrasser, J. Phys. Chem. B **112**, 5930 (2008).

[31] N. Sengupta, H. Maekawa, W. Zhuang, C. Toniolo, S. Mukamel, D. J. Tobias, and N. H. Ge, J. Phys. Chem. B **113**, 12037 (2009).

[32] S. Roy, T. l. C. Jansen, and J. Knoester, Phys. Chem. Chem. Phys. **12**, 9347 (2010).

[33] A. W. Smith, J. Lessing, Z. Ganim, C. S. Peng, A. Tokmakoff, S. Roy, T. l. C. Jansen, and J. Knoester, J. Phys. Chem. B **114**, 10913 (2010).

[34] P. Mukherjee, A. T. Krummel, E. C. Fulmer, I. Kass, I. T. Arkin, and M. T. Zanni, J. Chem. Phys. **120**, 10215 (2004).

[35] P. Mukherjee, I. Kass, I. T. Arkin, and M. T. Zanni, P. Nat. Acad. Sci. **103**, 3528 (2006).

[36] J. Manor, P. Mukherjee, Y.-S. Lin, H. Leonov, J. L. Skinner, M. T. Zanni, and I. T. Arkin, Structure **17**, 247 (2009).

[37] A. M. Woys, Y.-S. Lin, A. S. Reddy, W. Xiong, J. J. de Pablo, J. L. Skinner, and M. T. Zanni, J. Am. Chem. Soc. **132**, 2832 (2010).

[38] C. Liang, T. L. C. Jansen, and J. Knoester, J. Chem. Phys. **134**, 044502 (2011).

[39] I. Compagnon, J. Oomens, and G. v. Helden, J. Am. Chem. Soc. **128**, 3592 (2006).

[40] H. Zhu, M. Blom, I. Compagnon, A. M. Rijs, S. Roy, G. v. Helden, and B. Schmidt, Phys. Chem. Chem. Phys. **12**, 3415 (2010).

[41] P. Kupser, K. Pagel, J. Oomens, N. Polfer, B. Koksch, G. Meijer, and G. v. Helden, J. Am. Chem. Soc. **132**, 2085 (2010).

[42] W. H. James III, E. G. Buchanan, C. W. Mueller, T. S. Zwier, S. H. Choi, L. Guo, and S. H. Gellman, Abstracts Of Papers Of the J. Am. Chem. Soc. **241**, 25 (2011).

[43] W. J. Wedemeyer, E. Welker, and H. A. Scheraga, Biochemistry **41**, 14637 (2002).

[44] S. Osváth and M. Gruebele, Biophys. J. **85**, 1215 (2003).

[45] H. Torii and M. Tasumi, J. Raman Spec. **29**, 81 (1998).

[46] P. Bour and T. A. Keiderling, J. Chem. Phys. **119**, 11253 (2003).

[47] S. Ham, J. H. Kim, H. Lee, and M. H. Cho, J. Chem. Phys. **118**, 3491 (2003).

[48] J. H. Choi, S. Y. Ham, and M. Cho, J. Phys. Chem. B **107**, 9132 (2003).

[49] J. R. Schmidt, S. A. Corcelli, and J. L. Skinner, J. Chem. Phys. **121**, 8887 (2004).

[50] T. M. Watson and J. D. Hirst, Mol. Phys. **103**, 1531 (2005).

[51] T. Hayashi, W. Zhuang, and S. Mukamel, J. Phys. Chem. A **109**, 9747 (2005).

[52] R. D. Gorbunov and G. Stock, Chem. Phys. Lett. **437**, 272 (2007).

[53] T. L. C. Jansen and J. Knoester, J. Chem. Phys. **124**, 044502 (2006).

[54] T. L. C. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, and J. Knoester, J. Chem. Phys. **125**, 044312 (2006).

[55] R. Bloem, A. G. Dijkstra, T. L. C. Jansen, and J. Knoester, J. Chem. Phys. **129**, 055101 (2008).

[56] H. Maekawa and N. H. Ge, J. Phys. Chem. B **114**, 1434 (2010).

[57] L. Wang, C. T. Middleton, M. T. Zanni, and J. L. Skinner, J. Phys. Chem. B **115**, 3713 (2011).

[58] K. Cai, C. Han, and J. Wang, Phys. Chem. Chem. Phys. **11**, 9149 (2009).

[59] M. W. Wong, M. Frisch, and K. B. Wiberg, J. Am. Chem. Soc. **113**, 4776 (1991).

[60] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[61] Z. Ganim, K. C. Jones, and A. Tokmakoff, Phys. Chem. Chem. Phys. **12**, 3579 (2010).

[62] P. Hamm and S. Woutersen, Bull. Chem. Soc. Jpn. **75**, 985 (2002).

[63] F. Neese, ORCA an ab initio, Density Functional and Semiempirical program package, Version 2.6. University of Bonn, (2008).

[64] A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).

[65] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[66] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[67] B. Hammer, L. B. Hansen, and J. K. Norskov, Phys. Rev. B **59**, 7413 (1999).

[68] A. M. Moran, J. Dreyer, and S. Mukamel, J. Chem. Phys. **118**, 1347 (2003).

[69] J. Dreyer, J. Chem. Phys. **122**, 184306 (2005).

[70] NIST Chemistry WebBook, http://webbook.nist.gov/chemistry/, Origin: Sadtler Research Labs Under US-EPA Contract (2009), data compilation copyright by the U.S. Secretary of Commerce on behalf of the U.S.A.

[71] L. C. Mayne and B. Hudson, J. Phys. Chem. **95**, 2962 (1991).

[72] J. S. Richardson, Advances in Protein Chemistry **34**, 167 (1981).

[73] M. W. MacArthur and J. M. Thornton, J. Mol. Biol. **218**, 397 (1991).

[74] B. J. Stapley and T. P. Creamer, Protein Science **8**, 587 (1999).

[75] Y. S. Kim, J. Wang, and R. M. Hochstrasser, J. Phys. Chem. B **109**, 7511 (2005).

[76] Y. Mu, D. S. Kosov, and G. Stock, J. Phys. Chem. B **107**, 5064 (2003).

[77] M. Khalil, N. Demirdöven, and A. Tokmakoff, J. Phys. Chem. A **107**, 5258 (2003).

[78] A.-R. Allouche, J. Comp. Chem. **32**, 174 (2011).

[79] B. Hess, C. Kutzner, D. v. d. Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).

[80] W. L. Jorgensen and J. Tirado-Rives, J. Am. Chem. Soc. **110**, 1657 (1988).

[81] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[82] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comp. Chem. **18**, 1463 (1997).

[83] S. Nose, J. Chem. Phys. **81**, 511 (1984).

[84] W. G. Hoover, Phys. Rev. A **31**, 1695 (1985).

[85] T. L. C. Jansen and J. Knoester, Acc. Chem. Res. **42**, 1405 (2009).

[86] T. L. C. Jansen and J. Knoester, J. Phys. Chem. B **110**, 22910 (2006).

[87] K. Kwac, H. Lee, and M. Cho, J. Chem. Phys. **120**, 1477 (2004).

[88] S. Roy, M. S. Pshenichnikov, and T. l. C. Jansen, J. Phys. Chem. B **115**, 5431 (2011).

[89] S. T. Roberts, J. J. Loparo, and A. Tokmakoff, J. Chem. Phys. **125**, 084502 (2006).

[90] K. Lazonder, M. S. Pshenichnikov, and D. A. Wiersma, Optics Lett. **31**, 3354 (2006).

[91] Y.-S. Lin, J. M. Shorb, P. Mukherjee, M. T. Zanni, and J. L. Skinner, J. Phys. Chem. B **113**, 592 (2009).

[92] Y. N. Chirgadze, O. V. Fedorov, and N. P. Trushina, Biopolymers **14**, 679 (1975).

Atom	Field	$c(\omega_{ge})$	$c(\omega_{fg})$
	Gas phase	1690	3366.09
C	E_x	4178.94	8655.21
	E_y	1045.99	2204.88
	E_{zz}	506.529	1279.07
	E_{xy}	2079.24	5696.18
O	E_x	416.453	924.028
	E_y	-1448.31	-3573.18
	E_{zz}	-2.61955	58.2984
	E_{xy}	1458.73	3493.46
N	E_x	-812.651	-1953.89
	E_y	-2351.57	-4550.64
	E_{zz}	-475.445	-849.318
	E_{xy}	1141.32	3314.71
C_δ	E_x	1530.94	3669.27
	E_y	733.508	1769.54
	E_{zz}	-595.518	-1371.38
	E_{xy}	2963.12	7584.79

TABLE I: The electrostatic map parameters for the oscillator frequency given in $\mathrm{cm}^{-1}/\frac{E_h/\mathrm{Bohr}}{e}$ for the electric field components and in $\mathrm{cm}^{-1}/\frac{E_h/\mathrm{Bohr}^2}{e}$ for gradient components.

Atom	Field	$c(\mu_{ge:x})$	$c(\mu_{ge:y})$	$c(\mu_{ef:x})$	$c(\mu_{ef:y})$	$c(\mu_{gf:x})$	$c(\mu_{gf:y})$
	Gas phase	-0.268549	0.086947	-0.380947	0.123338	0.014298	-0.004629
C	E_x	0.741719	-0.148044	1.067028	-0.214632	-0.073204	0.018972
	E_y	0.495945	0.474987	0.709203	0.671606	-0.016181	-0.027262
	E_{zz}	-0.184370	-0.303594	-0.246652	-0.435630	-0.017385	0.024871
	E_{xy}	-0.246449	-0.238569	-0.262681	-0.367332	-0.099153	0.050682
O	E_x	1.076179	-0.274115	1.532203	-0.390849	-0.066240	0.017435
	E_y	0.019542	0.384490	-0.010210	0.558547	0.035273	-0.034126
	E_{zz}	0.307355	-0.173502	0.439860	-0.247577	-0.021485	0.011150
	E_{xy}	-0.077275	-0.030841	-0.077959	-0.054687	-0.025553	0.012669
N	E_x	-0.102316	0.128855	-0.163625	0.188791	0.037652	-0.017356
	E_y	-0.644828	.041822	-0.903825	0.055686	0.017238	0.003255
	E_{zz}	-0.142988	0.190637	-0.196934	0.268595	-0.004258	-0.006745
	E_{xy}	-0.806273	-1.259109	-1.090176	-1.803962	-0.023614	0.089243
C_δ	E_x	-0.308426	-0.540875	-0.403768	-0.778496	-0.031965	0.045087
	E_y	-0.253393	-0.318652	-0.343475	-0.457396	-0.006693	0.023688
	E_{zz}	0.023141	-0.064134	0.022248	-0.087555	0.014464	-0.001921
	E_{xy}	-0.982255	-0.421342	-1.303537	-0.627561	-0.084108	0.067477

TABLE II: The electrostatic map parameters for the transition dipole moments given in D/$\frac{E_h/\text{Bohr}}{e}$ for the electric field components and in D/$\frac{E_h/\text{Bohr}^2}{e}$ for gradient components.

	$NN_{Gly\text{-}Gly}$	$NN_{Gly\text{-}Pro}$	$NN_{Pro\text{-}Gly}$
α_{helix}			
J_{NN}	2.77	2.90	2.60
$\delta\omega_C$	8.97	0.48	-2.71
$\delta\omega_N$	13.24	-14.3	4.87
β_{Sheet}			
J_{NN}	4.13	3.60	4.26
$\delta\omega_C$	5.15	-2.07	-11.99
$\delta\omega_N$	11.69	-11.57	1.86
PP_{II}			
J_{NN}	3.28	2.70	3.15
$\delta\omega_C$	2.58	-3.80	-12.55
$\delta\omega_N$	7.96	-14.84	0.90

TABLE III: Comparison between the nearest neighbor dihedral maps developed on glycine dipeptide and proline containing dipeptide, respectively. The couplings and the energy shifts are given in cm^{-1}.

Atom	$\frac{Q}{e}$	$\frac{dQ}{e}$	v_x	v_y	v_z
C_α(C)	0.015998	0.011574	0	0	0
C	0.159744	0.010457	-0.836	-0.004	0
O	-0.347870	-0.044243	0.537	0.023	0
N	0.024241	0.013641	0.099	-0.018	0
C_δ(N)	0.061036	0.003277	0	0	0
C_α(N)	0.086851	0.005295	0	0	0

TABLE IV: Parameters for the TCC model Eq. 6. See Fig. 1a for C_α(C), C_α(N), and C_δ(N) atoms. The normal mode coordinates are given in units of the vibrational amplitude 0.0280896 Å.

	Exp.	Sim.
$t_2 = 0$ fs		
DW	29.7	27.5
ADW	10.4	10.7
$t_2 = 600$ fs		
DW	31.7	27.6
ADW	12.0	12.3

TABLE V: Diagonal width (DW) and *anti*-diagonal width (ADW) in cm^{-1} of 2D spectra for VP. ADW were extracted at $\omega_1 = 1614$ cm^{-1}.

	Exp.	Sim.
Pro peak		
DW	29.2	24.6
ADW	12.4	11.7
Gly peak		
DW	34.2	34.9
ADW	11.7	12.1

TABLE VI: Diagonal width (DW) and *anti*-diagonal width (ADW) in cm^{-1} of 2D spectra for PG: experiment at $t_2 = 150$ fs and simulation at $t_2 = 140$ fs. ADW values were extracted at $\omega_1 = 1610$ cm^{-1} and $\omega_1 = 1645$ cm^{-1}, respectively, for the Pro peak and the Gly peak.

Parameter	Value
α_{s}	0.13
α_{a}	0.18
ω_0^{s}	1590.8
ω_0^{a}	1608.7
σ_{s}	10.1
σ_{a}	5.6

TABLE VII: Parameters for modeling the CN stretch vibration in the arginine side chain (Eq. 7). ω_0^{s}, ω_0^{a}, σ_{s}, and σ_{a} are given in cm^{-1}.

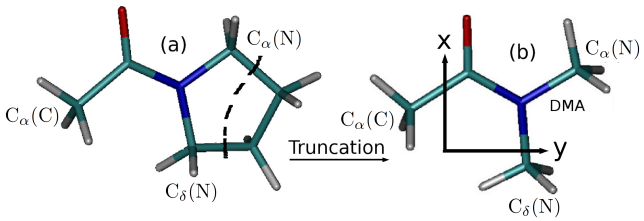

FIG. 1: The DMA molecule: truncation of the proline ring (a to b). CO is pointing along the x axis.

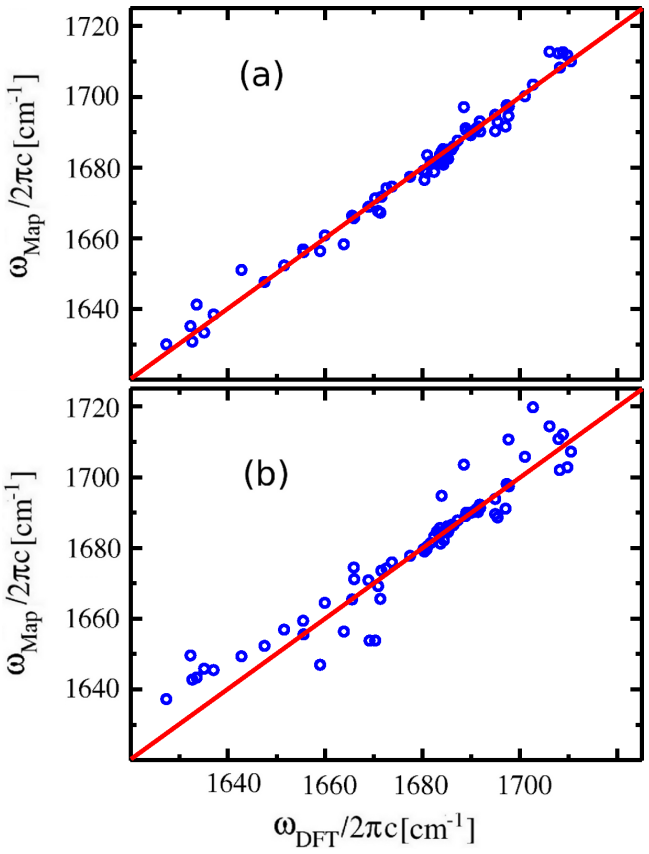

FIG. 2: Correlation plots between frequencies calculated with the DFT method and frequencies calculated with the electrostatic maps: (a) using the map parameters developed for the proline unit, and (b) using the electrostatic map reported for the secondary amide unit in Ref. 53.

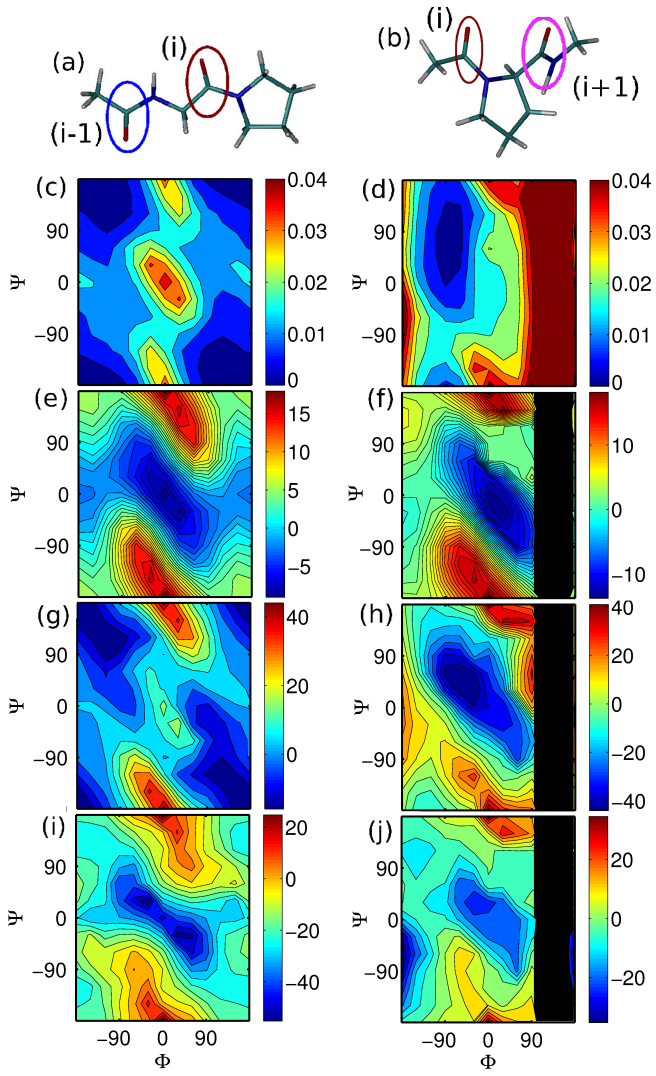

FIG. 3: The Gly-Pro (left) and Pro-Gly map (right) for the *trans*-configuration: structure and energy landscape (Hartree) of *trans*-Gly-Pro (a and c) and *trans*-Pro-Gly (b and d). Couplings (cm^{-1}) between site i and site $i-1$ (e) and between site i and site $i+1$ (f). Frequency shift (cm^{-1}) of site $i-1$ from the NMA gas phase frequency due to the site i (g). The frequency shift (cm^{-1}) of the site i from the DMA gas phase frequency due to the site $i+1$ (h) and site $i-1$ (i). The frequency shift of site $i+1$ from the NMA gas phase frequency due to site i (j). The black shaded region (f, h and j) is the energetically unfavorable zone which is unlikely to be encountered in simulations.

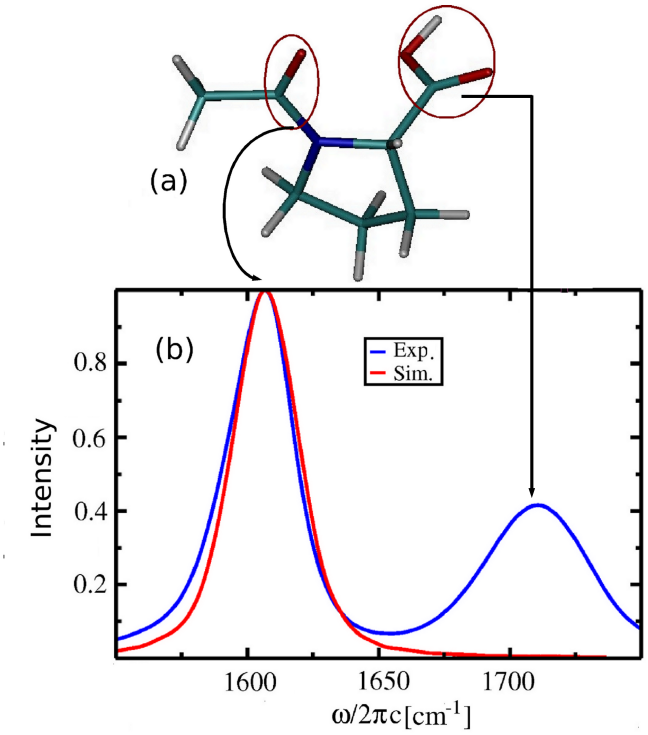

FIG. 4: *Trans*-AcP (a) and its linear absorption spectra at low pH(b). The arrows indicate the origin of the absorption peaks.

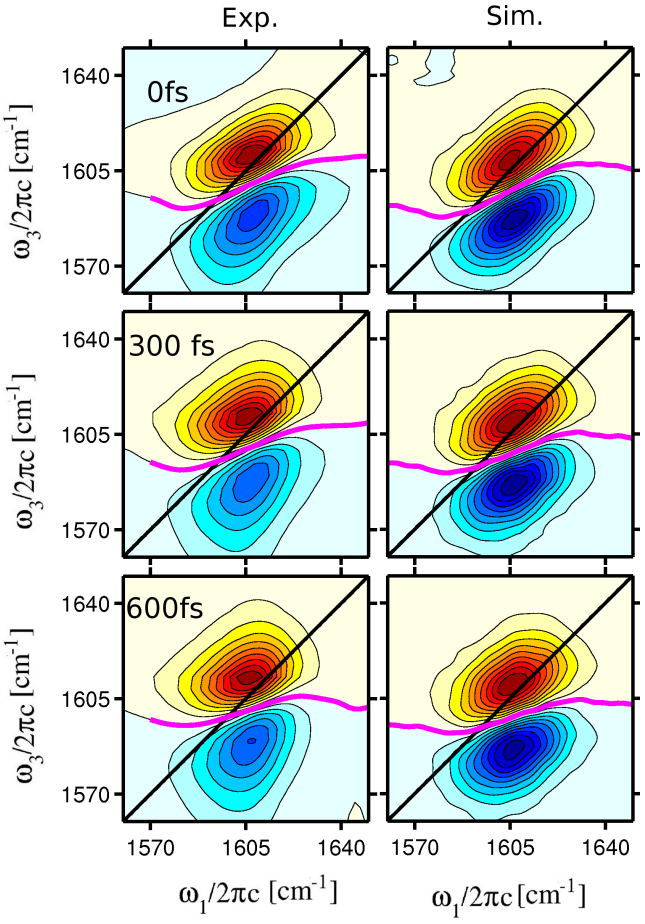

FIG. 5: 2DIR spectra (parallel polarization) of *trans*-AcP and nodal lines (red). 20 equally spaced contours are plotted from -100 % to 100 % of the maximum intensity.

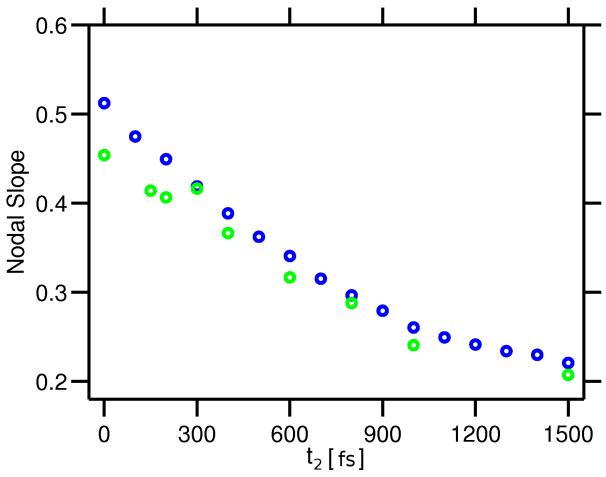

FIG. 6: Waiting time dynamics in the 2DIR spectra of AcP from the nodal slope: green and blue circles are from experiments and simulations, respectively.

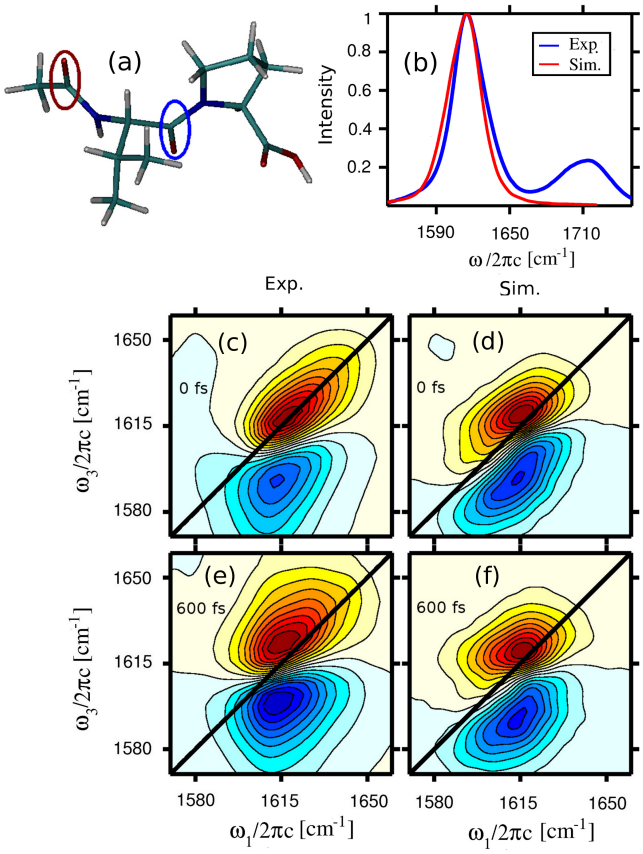

FIG. 7: The molecular structure of the *trans*-VP dipeptide (a) and its linear absorption (b) and 2DIR spectra (parallel polarization) (c-f). Contour lines in the 2D spectra are plotted as in Fig. 4.

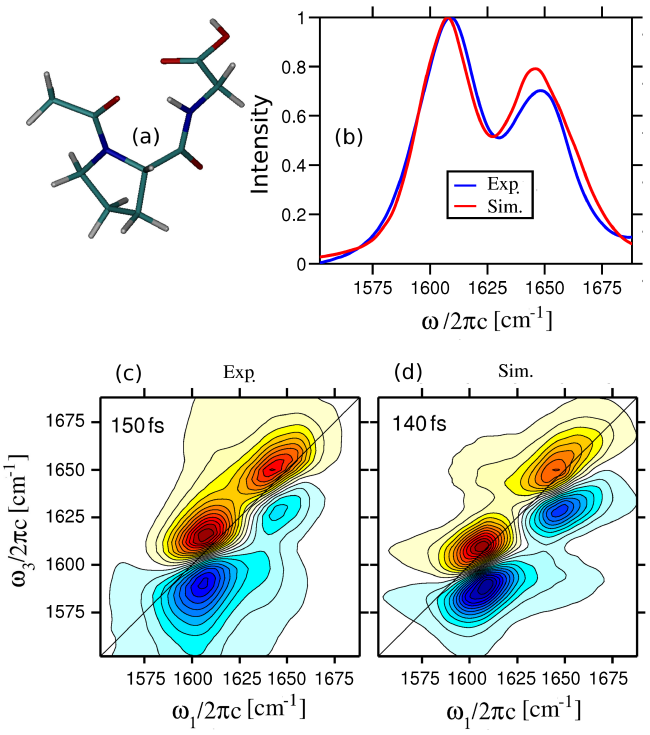

FIG. 8: The molecular structure of the *trans*-PG dipeptide (a) and its linear absorption (b) and 2DIR spectra (perpendicular polarization) (c-d). In the 2D spectra the contours are plotted from -100 % to 100 % of the maximum intensity with a spacing of 8% between the contours.

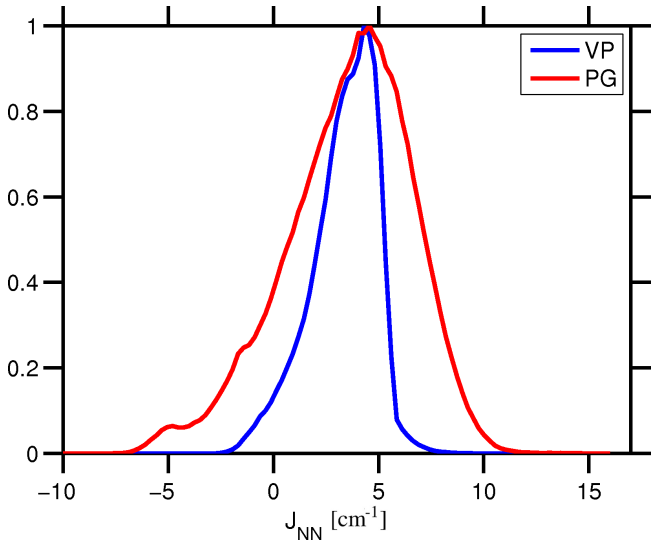

FIG. 9: Distribution of nearest neighbor couplings in VP and PG.

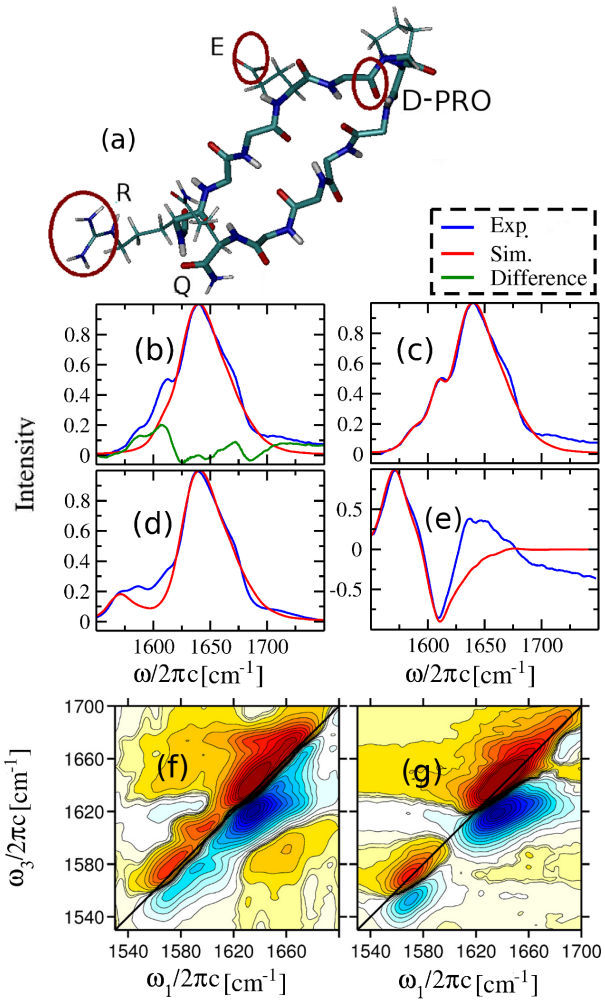

FIG. 10: Main chain of PG12 along with the arginine (R), glutamic acid (E) and glutamine (Q) side chains (a). Comparison between the experimental and simulated linear absorption spectra of PG12: arginine side chain vibration is not included in the simulation (b), while it is included in the model in (c), the V5 unit is labeled with $^{13}C^{16}O$ (d), difference spectra for V5 (e). 2DIR spectra (perpendicular polarization) for PG12 with the isotope labeled V5 unit (f and g). In the 2D spectra, the square root of the intensity is considered and the contours are plotted from -100 % to 70 % of the maximum intensity with a spacing of 6% between the contours.