THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

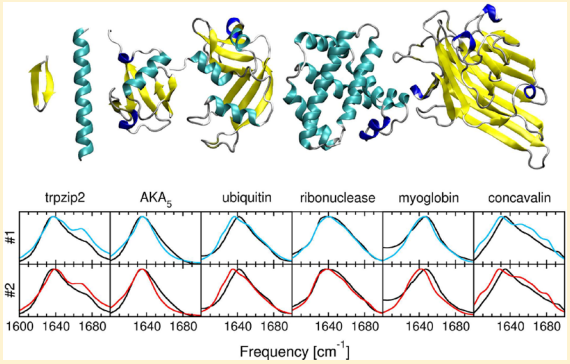
Empirical Maps For The Calculation of Amide I Vibrational Spectra of Proteins From Classical Molecular Dynamics Simulations

Edyta Małolepsza and John E. Straub*

Department of Chemistry, Boston University, 590 Commonwealth Avenue, Boston, Massachusetts 02215, United States

Supporting Information

ABSTRACT: New sets of parameters (maps) for calculating amide I vibrational spectra for proteins through a vibrational exciton model are proposed. The maps are calculated as a function of electric field and van der Waals forces on the atoms of peptide bonds, taking into account the full interaction between peptide bonds and the surrounding environment. The maps are designed to be employed using data obtained from standard all-atom molecular simulations without any additional constraints on the system. Six proteins representing a wide range of sizes and secondary structure complexity were chosen as a test set. Spectra calculated for these proteins reproduce experimental data both qualitatively and quantitatively. The proposed maps lead to spectra that capture the weak second peak observed in proteins containing β-sheets, allowing for clear distinction between α-helical and β-sheet proteins. While the parametrization is specific to the CHARMM force field, the methodology presented can be readily applied to any empirical force field.



1. INTRODUCTION

The amide I vibration arises predominantly from the stretching vibration of the carbonyl group in the peptide bond with small contributions from CN and HN bond vibrations and CCN angle bending. This vibration is notable due to its substantial oscillator strength and dependence on backbone secondary structure, while being relatively insensitive to the geometries of the amino acid's side chain. This vibration is often used to probe and examine the secondary structure of proteins, both experimentally and in computational studies.[1−8]

A number of observations have been made for relating amide I vibrational shifts to the local conformation of the peptide backbone. These "fingerprinting" approaches are based on relative simple rules[7] for assigning the secondary structure based on an amide I vibration spectrum. (i) α-helix has a peak typically located near 1655 cm^{-1}, which shifts toward lower frequencies with increasing helix length; this peak does not appear for peptides shorter than six amino acids. (ii) β-sheet exhibits a strong peak near 1630 cm^{-1} and a weak peak near 1685 cm^{-1}. (iii) Parallel β-sheets have a main peak at slightly higher wavenumbers than antiparallel β-sheets, but the difference does not typically allow one to distinguish between them in many experimental studies. Detailed discussions of the amide I vibration can be found elsewhere.[4,7,9]

A number of refined methods for computing amide I vibrational spectra of globular proteins have been developed. These methods typically make use of structural data for the protein derived from molecular dynamics simulation. The models are often based on the vibrational exciton approximation,[5,6,10−13] where the local amide I vibrations have frequencies dependent on the environment (or isotopic labeling) and are coupled through dipole−dipole interactions with other amide I vibrations. These techniques have proven highly successful in the interpretation of spectra of globular proteins in aqueous environments. Discussion of other methods used to compute amide I spectra can be found elsewhere.[14,15]

In recent years, there has been increasing interest in the interpretation of amide I spectra of peptide systems other than globular proteins, including (1) smaller peptides[11,13,16−20] and (2) peptides and proteins in more heterogeneous environments.[10,21−25] For peptides in reverse micelle environments, direct interaction with the surfactant and nonpolar phase, as well as interaction with aqueous solvent, must be addressed.[26] For peptides in membrane environments, there may be widely varying local environments, including the lipid tail region of a bilayer, for transmembrane peptides, or the headgroup region, which may include direct interactions with mono- and multivalent ions. Methods for the calculation of amide I spectra in these highly heterogeneous environments must account for the substantial changes to the local amide I frequencies and their coupling induced by the more heterogeneous solvation surroundings.

This work presents an extension of the vibrational exciton based models for the calculation of amide I vibrational spectra

Special Issue: James L. Skinner Festschrift

Received: December 31, 2013
Revised: March 15, 2014
Published: March 22, 2014

of peptides and proteins designed to be applicable to peptides and proteins in heterogeneous environments. The method is parametrized based on simulations of a small molecule, *N*-methylacetamide, in a variety of solvents, including water, DMSO, and chloroform, and applied to the calculation of spectra for a set of proteins using the CHARMM22 force field.[27] A comparison of the resulting spectra with experimental data suggests that the method is robust in capturing essential features of spectra in a variety of systems.

2. METHODS

Theory of Linear Spectroscopy. Ab initio calculations of IR spectra including the amide I mode contribution are computationally demanding for large proteins making it necessary to find methodologies that can simplify such calculations from data generated in standard molecular dynamics (MD) simulations of a protein. One such method employs the vibrational exciton model.[12,16] The protein is represented by a set of oscillators, also called chromophores, one for each peptide bond. It is assumed that the amide I modes do not couple with other types of vibrations in the system, only to the corresponding amide I mode of another peptide bond. The Hamiltonian of the system is expressed in terms of creation and annihilation operators, $b_i^\dagger$ and b_i, as

$$H(t) = \sum_i^N \omega_i(t) b_i^\dagger b_i + \sum_{i,j}^N \beta_{ij}(t) b_i^\dagger b_j + \sum_i^N \vec{\mu}_i(t)\vec{E}(t)[b_i^\dagger + b_i] \quad (1)$$

The diagonal element, $\omega_i = (E_{i,1} - E_{i,0})/\hbar$, is a frequency of transition between the vibrational ground and first excited states of the i-th oscillator. This fundamental frequency corresponds to the frequency of an oscillator unperturbed by interactions with other oscillators. Off-diagonal elements, β_{ij}, represent coupling between oscillators i and j, and μ_i is the fluctuating transition dipole moment of oscillator i interacting with applied electric field $\vec{E}(t)$.

The linear spectra were calculated as the Fourier transform of the correlation function of the transition dipole moments with the phenomenologically added term $e^{-t/2T_1}$ to account for the vibrational lifetime, according to the method presented by Jansen and Knoester[28]

$$I(\omega) = \mathbb{R}\int_0^\infty dt \frac{1}{\hbar}\sum_i^N \langle \mu_i(t)\mathbf{U}(t,0)\mu_i(0)\rangle e^{-i\omega t} e^{-t/2T_1} \quad (2)$$

where T_1 is the lifetime of the singly excited states, $\mathbf{U}(t, 0)$ represents time evolution operator, described in more details in Supporting Information.

Standard Model. Couplings between the i-th and j-th peptide bonds can be calculated in the transition dipole coupling approximation.[29] The electrostatic interaction between two peptide bonds is modeled by the interaction of the transition dipole moments, $\vec{\mu}_i$ and $\vec{\mu}_j$, associated with these two peptide bonds, as

$$\beta_{ij} = \frac{1}{4\pi\varepsilon_0}\left(\frac{\vec{\mu}_i\cdot\vec{\mu}_j}{r_{ij}^3} - 3\frac{(\vec{\mu}_i\cdot\vec{r}_{ij})(\vec{\mu}_j\cdot\vec{r}_{ij})}{r_{ij}^5}\right)$$

This approximation is assumed to be reasonable if the separation between peptide bonds, r_{ij}, is greater than their dimension. It should not be applied to model interaction between the nearest neighboring peptide bonds. In works by Jansen et al.,[11,30] Gorbunov et al.,[6] and Torii and Tasumi,[31] couplings between the nearest neighboring peptide bonds were calculated ab initio and represented as functions of Ramachandran angles, ψ and ϕ.

There are several methods for determining values of the fundamental frequencies and two methods are briefly presented below. One approach utilizes only geometrical properties of the protein independent of the environment.

(i) The fundamental frequency is represented as a sum of a constant, ω_{const}, and frequency shift, $\delta\omega$, reflecting when the oxygen or nitrogen atoms from the peptide bond form a hydrogen bond. The value of ω_{const} is usually chosen from between 1655 and 1710 cm^{-1}.[5,12,13,16,32] $\delta\omega$ depends on the length of the hydrogen bond and is typically taken to be −20 and −10 cm^{-1} for oxygen or nitrogen atoms, respectively. An isotopic shift can also be included in the model as a constant value added to ω_{const}.[22]

The second approach employs the electric field or electrostatic potential at the peptide bond atoms and therefore introduces a dependence on the surrounding environment.

(ii) The fundamental frequency for each peptide bond is calculated as a function of the electrostatic potential[33−36] or electric field[19,25,36] or all of them[13] experienced by atoms of the peptide bond, which may be approximated in terms of the partial charges of the surrounding solvent molecules and the electrostatic potential

$$\omega = \omega_0 + \sum_i c_i\phi_i,\ \phi_i = \sum_j \frac{q_j}{r_{ij}}$$

or the electric field

$$\omega = \omega_0 + \sum_{i\alpha} c_{i\alpha}\vec{E}_{i\alpha},\ \vec{E}_i = \sum_j \frac{q_j}{r_{ij}^2}\hat{\mathbf{r}}_{ij}$$

where i runs over atoms of the peptide bond, C, O, N, H, $\alpha = \{x,y,z\}$, and j runs over point charges surrounding the i-th peptide bond. Coefficients c_i and $c_{i\alpha}$ are determined by fitting the frequency to the electrostatic potential or electric field on a test system such as *N*-methylacetamide (NMA) to reproduce either ab initio calculations for NMA−water clusters, containing up to four water molecules,[36] or the experimentally measured transition frequencies for NMA. The intercept, ω_0, is often chosen to be the gas-phase amide I frequency of NMA equal to 1717 cm^{-1}. Therefore, the value of the shift in the fundamental frequency, $\omega - \omega_0$, can be understood as the first order perturbation correction to the frequency in vacuum due to interactions with solvent molecules. This model reminds one of the Stark effect in which each component of the electric field on each atom may have a different proportionality coefficient. ω_0 can also be treated as an additional parameter in the fitting procedure.

Proposed Extended Model. Work presented in this paper is focused on improving the second method presented above through the use of data obtained from molecular dynamics simulations. There are five aspects to be considered in the improvement of the fits.

(i) If the amide I frequency of NMA in vacuum is chosen as the reference (i.e., unperturbed) frequency, to first order the

dx.doi.org/10.1021/jp412827s | J. Phys. Chem. B 2014, 118, 7848−7855

corrections to the frequency arise from interactions of NMA with solvent molecules. There are two types of such nonbonded interactions, electrostatic and van der Waals, and both are included in the frequency shift.

(ii) In existing methods, the electrostatic potential and electric field employed in the fits are typically derived from partial charges on selected solvent molecules surrounding the NMA molecule. Such a field or potential does not fully represent an actual field or potential on the NMA molecule during molecular dynamics simulations, where periodic boundary conditions are employed and electrostatic interactions are calculated using the Ewald summation. In this work, all fits presented were obtained based on the total electrostatic and van der Waals forces exerted on the atoms of peptide bonds during simulations.

(iii) The proposed method uses data obtained from standard all-atom molecular simulations carried out with the NAMD[37] package where the SHAKE[38] algorithm was applied to fix the length of bonds to hydrogen atoms. However, it can be applied to results from any other program. This stands in contrast to other approaches in which additional restrictions are imposed in the calculation of the spectrum, such as fixing the length of all other bonds using the LINKS[39] algorithm (which is not implemented at this time in CHARMM or in NAMD). (Heavy atom bonds vibrate with a period of at least 20 fs that is equal to the shortest period of angle vibrations for C−O−H. Therefore, constraining all bonds can take full advantage of the LINKS algorithm only when those angles are also constrained.)

(iv) NMA provides a model of an isolated peptide bond. Results based solely on MD simulations of NMA in water may not be representative of more heterogeneous environments experienced by a peptide bond in a protein. Reliable fits should also incorporate the hydrophobic surroundings of an amino acid buried inside a protein, interactions with lipophilic regions of a membrane, or contact with monovalent ions. This can be accomplished by fitting data from MD simulations of NMA in water and in solvents representing a range of dielectric constants including DMSO and chloroform.[19]

(v) Finally, the maps developed in this work are designed to highlight the difference between the amide I frequency of an α-helix and of a β-sheet. Such fits are especially useful in studying aggregation processes where proteins commonly undergo changes in secondary structure.[40−43]

In the next sections, new maps are presented together with the methodology employed in their calculation. The amide I spectra of a test set of proteins are presented to validate the quality of the proposed maps.

3. MODEL PARAMETRIZATION FOR NMA

NMA Simulations. All fits presented below were calculated based on the results of MD simulations performed using the CHARMM22 force field and the NAMD package. A single molecule of NMA was placed in the center of a pre-equilibrated cubic box ($70 \times 70 \times 70$ Å^3) of solvent and equilibrated at 300 K. Three solvents were considered, water, DMSO, and chloroform. Water was represented by the TIP3P[44] model, parameters for DMSO were taken from the work by Strader and Feller,[45] and parameters for chloroform were derived from the work of Cieplak et al.[46] Periodic boundary conditions and particle mesh Ewald summation were employed, nonbonded interactions were switched to zero over the window 10 to 12 Å, the time step was set to 1 fs, geometries were saved every 100 fs, and all bonds with hydrogen atoms were constrained using the SHAKE algorithm.

Force Calculations. All forces on the C, O, N, and H atoms of NMA were computed under the same conditions used during the production MD runs, including the same cutoffs, PME parameters, and box size. All force vectors were aligned in such a way that the *XY*-plane was defined by the C, O, and N atoms of a peptide bond with the CO bond oriented along the *Y*-axis. Finally, forces were converted from kcal/mol/Å to atomic units by a factor 0.0008432983.

Fit Formula. New maps for the fundamental frequency proposed in this work include not only the electric field vectors on the atoms of the peptide bonds, but also van der Waals forces. Initially, the fitting function contained 25 parameters, 12 for components of electric fields, $\vec{E}_i$, on C, O, N, and H atoms of NMA, 12 parameters corresponding to vectors from van der Waals forces, $\vec{F}_i$, and one parameter for the intercept

$$\omega = \omega_0 + \sum_{i\alpha} c_{i\alpha} E_{i\alpha} + \sum_{i\alpha} d_{i\alpha} F_{i\alpha} \tag{3}$$

where frequency is in units of wave numbers.

Fitting Procedure. Frequency functions were fitted to six experimental values including positions of the peak maximum, ω_{exp}, and full width at half-maximum, Γ_{exp}, of the amide I frequency of NMA in water, DMSO, and chloroform. Several forms of the penalty function were employed in the fitting process (for details see Supporting Information).

Table 1. Experimental Values of Peak Maximum and Full Width at Half Maximum in cm^{-1} for Amide I Spectra of NMA in Water, DMSO, and Chloroform

	ω_{exp}	Γ_{exp}
water[16]	1624.0	30.5
DMSO[16]	1659.5	25.0
chloroform[19]	1665.0	26.0

Some field or force components make greater contributions to the fundamental frequency than others, and their analysis can help in identifying the most important parameters. Figure 1 shows histograms of the field and force components on the NMA peptide bond atoms derived from MD simulations in three solvents. We observed that on average only 9 out of 24 components made a greater than 0.001 au contribution. The

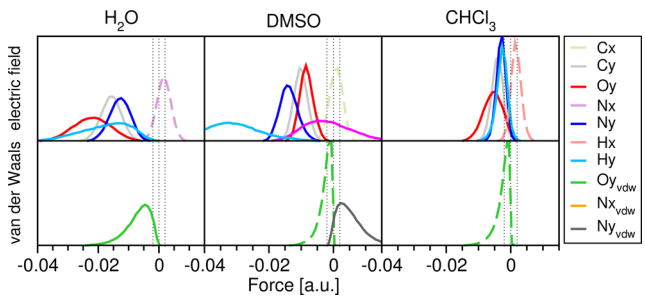


Figure 1. Histograms of field and force components on NMA peptide bond atoms in water (left), DMSO (center), and chloroform (right). Top panel refers to electric field components, while bottom shows van der Waals forces. Dotted lines show positions of 0, ± 0.002 au. Only components with absolute mean value greater than 0.001 au are shown. If the absolute mean value is in the range 0.001 and 0.002 au, its histogram is shown as a dashed line.

dx.doi.org/10.1021/jp412827s | J. Phys. Chem. B 2014, 118, 7848−7855

Table 2. Coefficients of the Two Best Fits for the Electric Field and van der Waals Force Components on Atoms of the Peptide Bond for Use in Equation 3

	ω_0	C_y	O_y	N_x	N_y	H_x	H_y	O_y^{vdw}	N_x^{vdw}	N_y^{vdw}
#1	1674.55	−103.6	1320.4	−632.7	3932.5	−1390.8	−1740.4		−7905.1	−858.3
#2	1674.64	−374.3	1787.4	−286.1	3643.2	−1598.1	−1412.3	−705.9	−8233.2	

next fit included only those nine dominant components. As the amide I frequency consists mostly of vibrations of the carbonyl group in peptide bonds, it is not surprising that force components in the direction of the CO bond (which defines the *Y*-axis) are the most dominant contributions. There is only one other component, $\vec{E}_{Hx}$, with a peak further from zero than 0.002 au. Out of 12 parameters for van der Waals forces, only the *y*-components on oxygen and nitrogen atoms appear to be significant.

Apart from analyzing the field and force components on NMA peptide bond atoms, one can test the components together with corresponding fit coefficients to determine which coefficient provides the largest contribution to the fundamental frequency shift. (A sample of results for some 25 parameter fits is shown in Figure 7 in Supporting Information.) In each case, a different set of fit parameters was found to be important in determining the frequency shift including components that were neglected based solely on the field and force histograms. This indicates that an analysis of fit parameters alone or forces alone might be misleading. Each set was chosen as input for another fitting procedure and this process was repeated until the most significant fit parameters were determined.

4. THE BEST MAPS

The number of variables in the fit formula led to the generation of many maps by a multivariable fitting procedure. The choice of the best maps was based on reproducing amide I spectra of two proteins, an alanine-based peptide with sequence $(AAAAK)_4AAAAY$ (referred to as AKA_5) and trpzip2 (1LE1). AKA_5 consists of only 25 residues and forms one well-defined α-helix, while even smaller trpzip2 is an example of β-hairpin protein. AKA_5 and trpzip2 make a complementary set to study the assignment of secondary structure based on amide I vibrational spectra. Technical details of spectra calculations are presented in the next section.

The two best fits for the amide I frequency are presented in Table 2. Both fits consist of nine components.

From the analysis of fields and forces on the atoms of NMA presented above, only two components of van der Waals forces were found to be significant, the *y*-component on the oxygen and nitrogen atoms, O_y^{vdw} and N_y^{vdw}, respectively. Only these two components were expected to make a significant contribution to the frequency shift. Surprisingly, for both fits the coefficient corresponding to the *x*-component of the van der Waals force on the nitrogen atom, N_x^{vdw}, is the largest of all. However, analysis of the frequency shift caused by this component in the case of the NMA spectrum (presented in Figure 2) shows that despite the value of this coefficient, its contribution is less than 5 cm^{-1} for the spectrum of NMA in DMSO.

The main contributions to the frequency for all fits come from three components of the electric field, O_y, N_y and H_y. The contribution from the carbon atom is found to be negligible even though the amide I vibration consists primarily of the CO vibration. Interestingly, for map #1 the van der Waals forces contribute more than 1 cm^{-1} only in the case of spectra in DMSO. For map #2 the van der Waals forces contribute more than 1 cm^{-1} for spectra in water and DMSO. For NMA in chloroform, van der Waals forces make negligible contributions for both maps.

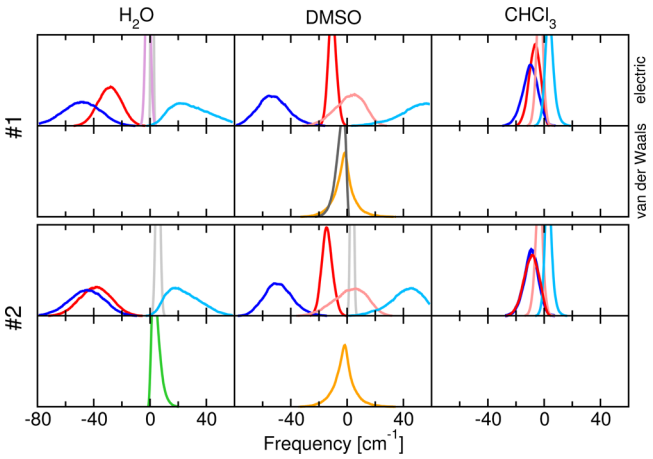


Figure 2. Histograms of the main contributions to the NMA amide I frequency shift in water, DMSO, and chloroform for the two best fits. Histograms were smoothed using running averaging over 10 frames. Only contributions greater than 1 cm^{-1} are shown. Colors as in Figure 1.

Figure 3 presents the comparison of experimental and computational amide I spectra for NMA in three considered solvents. Calculated spectra agree well with experimental spectra for both proposed maps.

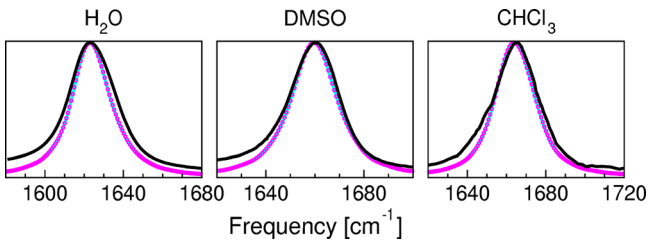


Figure 3. Amide I spectra of NMA in water, DMSO, and chloroform. Black lines represent experimental data, while cyan lines and magenta circles show spectra calculated using the #1 and #2 proposed maps, respectively. Experimental data (digitized and interpolated) were derived for NMA in water and DMSO from Hamm et al.[16] and for NMA in chloroform from Wang et al.[19]

5. VALIDATION OF THE NEW MAPS AND SPECTRA FOR PROTEINS

To validate the new maps, amide I frequency spectra of four additional proteins were calculated and compared with experimental results. For benchmark calculations, a series of proteins were chosen including ubiquitin (1UBQ), ribonuclease (7RSA), myoglobin (1WLA), and concavalin (1JBC). The set of six proteins include a wide range of structures and sizes as shown in Figure 4 from the 12-residue β-hairpin trpzip2,

dx.doi.org/10.1021/jp412827s | J. Phys. Chem. B 2014, 118, 7848−7855

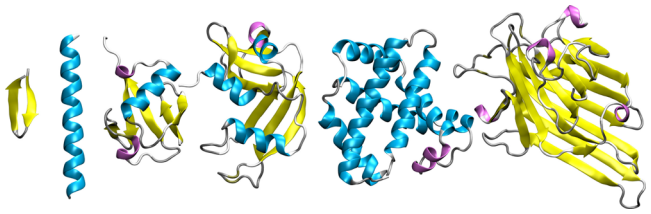


Figure 4. Ribbon structures of the six proteins in the test set showing β-strands (yellow), α-helices (cyan), and 3_{10}-helices (pink). The structures were visualized with VMD.[47]

through the 153-residue myoglobin (formed almost exclusively of α-helices), to the 237-residue concavalin (including two large β-sheet surfaces).

MD Simulations. Each peptide was placed in a pre-equilibrated water box, neutralized, energy minimized, heated, and equilibrated at 300 K. For AKA_5, trpzip2, ubiquitin, ribonuclease, and myoglobin, a cubic box of 70 × 70 × 70 Å^3 was used, while for concavalin an orthorhombic box of size 70 × 70 × 85 Å^3 was employed. The simulations were run under the same conditions as described for NMA.

Force Calculations. All fields and forces were calculated according to the protocol used for NMA with one exception. Off-diagonal couplings between the nearest neighbor peptide bonds are taken from Jansen's maps.[11,30] As such, partial charges of all backbone atoms of the neighboring residues should not contribute to the fundamental frequencies. Therefore, those atoms were excluded from the field and force calculations. Partial charges on all remaining atoms of the protein were included in the calculations.

Construction of the Hamiltonian Matrix and Transition Dipole Vectors. Transition dipoles were constructed according to the work by Torii and Tasumi.[31] Each vector's length was set to 2.73 D with the vector displaced 10° from the CO bond in the CON plane. Its origin was located at $\vec{r}_C$ + 0.665 $\hat{n}_{CO}$ + 0.258 $\hat{n}_{CN}$, where $\vec{r}_C$ is the position of the carbon atom of the peptide bond, while $\hat{n}_{CO}$ and $\hat{n}_{CN}$ are normalized vectors in the directions of the CO and CN bonds, respectively. The magnitudes of all vectors are in angstroms.

The fundamental frequency for each peptide bond (the diagonal values of the Hamiltonian matrix) was calculated as a sum of three terms: (1) the unperturbed oscillator frequency calculated from the frequency fit, (2) the shift caused by the closest neighboring peptide bond on the N-site (for the i-th oscillator that shift is a function of two Ramachandran angles $\phi_{i,i-1}$ and $\psi_{i,i-1}$), and (3) the corresponding shift caused on the C-side of the peptide bond (function of angles $\phi_{i,i+1}$ and $\psi_{i,i+1}$). Both types of shifts were taken from tables calculated by Jansen and co-workers.[11,30] The only exceptions were made for the fundamental frequencies of the amide groups in the side chains and for the peptide bond associated with proline, where, due to the lack of the hydrogen atom in the peptide bond, the frequency fit cannot be used. Both cases have been treated explicitly in the recent papers[19,48] using approaches that impose constraints on all bonds, and as such could not be applied in the present work. Hence, the amide groups located on the side chains were not taken into account, and values of 1620 or 1670 cm^{-1} were used for proline. The chosen value makes no significant difference to the final line shape.

The off-diagonal couplings for the nearest neighbors as functions of the Ramachandran angles between adjacent peptide bonds were taken from maps provided by Jansen.[11,30] The remaining off-diagonal couplings were calculated in the transition dipole coupling approximation.[29]

Spectra Calculations. Spectra were calculated according to eq 2 with the moving window approach skipping 10 frames between windows. MD simulation data were collected every 100 fs and this time step was chosen to be the integration step in the eq 2. To validate the 100 fs time step employed, line shapes were also computed using a time step of 10 fs for NMA in three solvents and the two smallest and most mobile proteins, see Supporting Information in Figure 8. The window size was set to 256 frames. The homogeneous dephasing time T_1 was set to 1.0 ps for proteins, and to 0.45 ps for NMA. Frequencies and couplings were converted from cm^{-1} to ps^{-1} using the factor $2\pi c$ before integration, where c is the speed of light. Calculations were performed using in-house Python scripts, written using the MDAnalysis package,[49] and can be shared upon request.

Results. Amide I spectra calculated using maps #1 and #2 for proteins in the test set are presented in Figure 5 together with experimental data. Spectra were normalized so that the maximum value was 1.

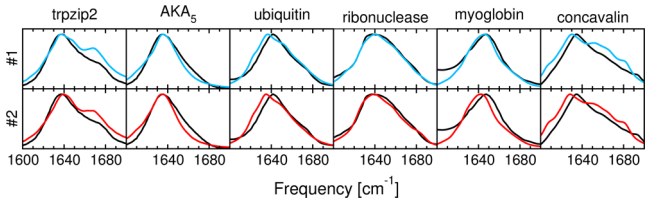


Figure 5. Comparison of the experimental and calculated spectra for the amide I vibration for the six test proteins. Black lines indicate experimental data, while blue and red lines show spectra calculated using the #1 and #2 proposed maps, respectively. Experimental data (digitized and interpolated) were derived for trpzip2 from the work of Smith and Tokmakoff,[17] for AKA_5 from Barber-Armstrong et al.,[50] for ubiquitin, ribonuclease, and myoglobin from Ganim et al.,[18] and for concavalin from Karjalainen et al.[51]

The two smallest proteins in the benchmark set, trpzip2 and AKA_5, are of special interest. As shown in Figure 4, trpzip2 consists of only one β-hairpin, while AKA_5 forms one α-helix. Because of that substantial difference in the secondary structures, these two peptides serve as excellent benchmarks for methods proposed for the determination of amide I spectra. Successful capture of the discrepancies between the amide I spectra of α-helical and β-type proteins is a validation of the usefulness of the fits in distinguishing between these two types of secondary structure.

Positions of the experimental main peaks for both proteins are separated by only 5 cm^{-1} with 1636 cm^{-1} for trpzip2[17] and 1631 cm^{-1} for AKA_5.[50] This separation is too small for reliable assignment of the secondary structure based on positions of the main peaks. However, the experimental spectrum of trpzip2 also exhibits a weak second peak near 1673 cm^{-1}.[17] Consequently, the secondary structure can be determined by the presence of the weak second peak in the spectrum. For this reason, all presented fits somewhat overestimate the weak peak in the spectrum of trpzip2.

In the case of larger proteins from the test set with more complex tertiary structure (ubiquitin, ribonuclease, myoglobin, and concavalin) good agreement between the experimental and the computed spectra is observed. The positions of the main peaks are reproduced as well as the line widths.

The two proposed maps presented here do not utilize the same set of parameters for the contribution of van der Waals forces, but the most important aspects of the spectra are captured in both fits. There are three main contributions to the frequency shift. Two come from the *y*-component of the electric field on the oxygen and nitrogen atoms and with opposing sign from the *y*-component of the electric field on the hydrogen atom (shown in Figure 6 with black, blue, and cyan for both maps, respectively). Similar behavior was observed for the case of the NMA molecule, especially in water and DMSO.

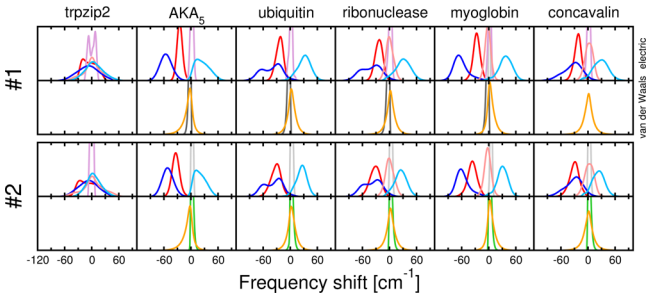


Figure 6. Histograms of the contributions to the frequency shift for six proteins in the test set for the maps #1 and #2. Panels show contributions from the electric field and van der Waals force components, respectively. Histograms were smoothed using running averaging over 10 frames. Only contributions greater than 1 cm^{-1} are shown. Colors as in Figure 1.

van der Waals forces on the oxygen atom appear only in map #2, while map #1 utilizes the *x*-component on the nitrogen atom. However, the *y*-component of the van der Waals forces on the nitrogen atom makes the dominant contribution of all van der Waals force components. For both maps, this component plays a significant role in spectra calculations for each protein in the benchmark set (except the smallest protein trpzip2, where only electric field components contribute to the fundamental frequency).

Figures 9 and 10 in the Supporting Information present comparison of correlations between electric and van der Waals contributions to the fundamental frequency for both maps. For small proteins, trpzip2 and AKA_5, we observe minor correlation between components, such as between the *y*-component of the electric field on the hydrogen atom, and between the *y*-component of the van der Waals field on the nitrogen atom for trpzip2, or between *x*-component of the electric field on the hydrogen atom and *x*-component of the van der Waals field on the nitrogen atom. However, for larger proteins correlations become negligible and there is no example of correlation that is present in any of the studied proteins.

van der Waals forces make smaller contributions to the fundamental frequency shift than the electric field forces. However, this contribution is not negligible except in the case of trpzip2. Therefore, while van der Waals forces are relatively less important in the calculation of spectra, maps without those components are less accurate.

6. CONCLUSIONS

This paper presents a new approach to the evaluation of fundamental vibrational frequencies and spectra using input from standard classical molecular dynamics simulations. We have presented a revised methodology for the calculation of infrared spectra of amide I vibrations for proteins in solution. The spectra are calculated using the vibrational exciton model. The couplings between oscillators corresponding to peptide bonds are calculated in the transition dipole coupling approximation. The fundamental frequencies are obtained using the total electric field and van der Waals forces on atoms of all peptide bonds. Addition of the van der Waals forces is a novel approach, as in earlier works only the electric field was utilized. Use of van der Waals interactions together with the electrostatic contributions improves the accuracy and robustness of the method.

To validate the new methodology over a wide range of secondary and tertiary structures of proteins, a set of six test proteins with varying content of secondary and tertiary structure was tested. In each case, the calculated spectra satisfactorily reproduced the experimental results, which validates the quality of the proposed maps for amide I frequency calculations and the methodology for the calculation of spectra. Reproduction of the weak second peak in the spectra for β-sheet rich proteins is a sensitive measure of the accuracy of any proposed approach for the calculations of amide I spectra of proteins, and the proposed methodology successfully addresses that challenge and sets a standard for other proposed methods.

The amide I vibration is a strong indicator of protein structure. Even linear spectra may be used to gain insight into the predominance of α-helices and β-strands. The development of new approaches to the computation of amide I spectra continues to be a highly active field of research [see refs 5, 6, 8, 11, 13, 15−20, 25, 30−36, 51−54]. The goal of this work was to develop a robust method for computing amide I spectra for proteins in heterogeneous environments. The method presented may be combined with the widely used CHARMM22 force field to allow computation of IR spectra as a function of protein conformation and solvent environment. The methodology has an advantage over existing methods in that it does not require additional constraints of bonds in the system under study or special "filtering" of forces. Standard results of molecular dynamics simulations can be used to compute the spectra in a straightforward manner. The application of this methodology to the calculation of spectra using molecular dynamics other than CHARMM22 should be straightforward.

■ ASSOCIATED CONTENT

Supporting Information

Supporting Information contain details of the applied theory of linear spectroscopy and associated fitting procedure, examples of shift contributions to the NMA frequency for 6 randomly chosen 25-parameter fits, comparison of spectra for NMA, trpzip2, and AKA_5 calculated with two time steps, mean values and standard deviation values of contributions to the fundamental frequency shift from electric and van der Waals terms for the proposed maps, and correlation of contributions to the fundamental frequency from the electric field and van der Waals forces for the proposed map. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: straub@bu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

E.M. would like to thank Professor David Coker, Dr. Thomas C. Jansen, and Dr. Marek Orzechowski for valuable discussions. We thank Professor James L. Skinner for valuable comments on an earlier version of this manuscript. The authors gratefully acknowledge the support of a grant from the National Institutes of Health (NIH: RO1GM076688) and a grant from the National Science Foundation (CHE-1114676).

■ REFERENCES

(1) Surewicz, W. K.; Mantsch, H. H. New Insight into Protein Secondary Structure from Resolution-Enhanced Infrared Spectra. *Biochim. Biophys. Acta* **1988**, *952*, 115−130.
(2) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. The Two-Dimensional IR Nonlinear Spectroscopy of a Cyclic Penta-Peptide in Relation to Its Three-Dimensional Structure. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 2036−2041.
(3) Barron, L. D.; Hecht, L.; Blanch, E. W.; Bell, A. F. Solution Structure and Dynamics of Biomolecules from Raman Optical Activity. *Prog. Biophys. Mol. Biol.* **2000**, *73*, 1−49.
(4) Barth, A.; Zscherp, C. What Vibrations Tell Us About Proteins. *Q. Rev. Biophys.* **2002**, *35*, 369−430.
(5) Abramavicius, D.; Zhuang, W.; Mukamel, S. Peptide Secondary Structure Determination by Three-Pulse Coherent Vibrational Spectroscopies: A Simulation Study. *J. Phys. Chem. B* **2004**, *108*, 18034−18045.
(6) Gorbunov, R. D.; Kosov, D. S.; Stock, G. Ab Initio-Based Exciton Model of Amide I Vibrations in Peptides: Definition, Conformational Dependence, and Transferability. *J. Chem. Phys.* **2005**, *122*, 224904.
(7) Barth, A. Infrared Spectroscopy of Proteins. *Biochim. Biophys. Acta* **2007**, *1767*, 1073−1101.
(8) Jeon, J.; Yang, S.; Choi, J.-H.; Cho, M. Computational Vibrational Spectroscopy of Peptides and Proteins in One and Two Dimensions. *Acc. Chem. Res.* **2009**, *42*, 1280−1289.
(9) Uversky, V. N. *Methods in Protein Structure and Stability Analysis: Vibrational Spectroscopy*; Nova Publishers: Hauppauge, NY, 2007.
(10) Chernyak, V.; Zhang, W. M.; Mukamel, S. Multidimensional Femtosecond Spectroscopies of Molecular Aggregates and Semiconductor Nanostructures: The Nonlinear Exciton Equations. *J. Chem. Phys.* **1998**, *109*, 9587−9601.
(11) Jansen, T. L. C.; Dijkstra, A. G.; Watson, T. W.; Hirst, D.; Knoester, J. Modeling the Amide I Bands of Small Peptides. *J. Chem. Phys.* **2006**, *125*, 044312.
(12) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: New York, 2011.
(13) Reppert, M.; Tokmakoff, A. Electrostatic Frequency Shifts in Amide I Vibrational Spectra: Direct Parameterization Against Experiment. *J. Chem. Phys.* **2013**, *138*, 134116.
(14) Surewicz, W. K.; Mantsch, H. H.; Chapman, D. Determination of Protein Secondary Structure by Fourier Transform Infrared Spectroscopy: A Critical Assessment. *Biochemistry* **1993**, *32*, 389−394.
(15) Schweitzer-Stenner, R. Advances in Vibrational Spectroscopy As a Sensitive Probe of Peptide and Protein Structure: A Critical Review. *Vib. Spectrosc.* **2006**, *42*, 98−117.
(16) Hamm, P.; Lim, M.; Hochstrasser, R. M. Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy. *J. Phys. Chem. B* **1998**, *102*, 6123−6138.
(17) Smith, A. W.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of β- Hairpin Peptides. *J. Chem. Phys.* **2007**, *126*, 045109.
(18) Ganim, Z.; Chung, H. S.; Smith, A. W.; DeFlores, L. P.; Jones, K. C.; Tokmakoff, A. Amide I Two-Dimensional Infrared Spectroscopy of Proteins. *Acc. Chem. Res.* **2008**, *41*, 432−441.
(19) Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides. *J. Phys. Chem. B* **2011**, *115*, 3713−3724.
(20) Kaminsky, J.; Bour, P.; Kubelka, J. Simulations of the Temperature Dependence of Amide I Vibration. *J. Phys. Chem. A* **2013**, *115*, 30−34.
(21) Tretiak, S.; Middleton, C.; Chernyak, V.; Mukamel, S. Exciton Hamiltonian for the Bacteriochlorophyll System in the LH2 Antenna Complex of Purple Bacteria. *J. Phys. Chem. B* **2000**, *104*, 4519−4528.
(22) Paul, C.; Wang, J.; Wimley, W. C.; Hochstrasser, R. M.; Axelsen, P. H. Vibrational Coupling, Isotopic Editing, and β-Sheet Structure in a Membrane-Bound Polypeptide. *J. Am. Chem. Soc.* **2004**, *126*, 5843−5850.
(23) Fang, C.; Senes, A.; Cristian, L.; DeGrado, W. F.; Hochstrasser, R. M. Amide Vibrations Are Delocalized Across the Hydrophobic Interface of a Transmembrane Helix Dimer. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 16740−16745.
(24) Mukherjee, P.; Kass, I.; Arkin, I. T.; Zanni, M. T. Structural Disorder of the CD3 Transmembrane Domain Studied with 2D IR Spectroscopy and Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 24740−24749.
(25) Lin, Y.-S.; Shorb, J. M.; Mukherjee, P.; Zanni, M. T.; Skinner, J. L. Empirical Amide I Vibrational Frequency Map: Application to 2D-IR Line Shapes for Isotope-Edited Membrane Peptide Bundles. *J. Phys. Chem. B* **2009**, *113*, 593−602.
(26) Mukherjee, S.; Chowdhury, P.; Gai, F. Infrared Study of the Effect of Hydration on the Amide I Band and Aggregation Properties of Helical Peptides. *J. Phys. Chem. B* **2007**, *111*, 4596−4602.
(27) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545−1615.
(28) Jansen, T. L. C.; Knoester, J. Nonadiabatic Effects in the Two-Dimentional Infrared Spectra Od Peptides: Applications to Alanine Dipeptide. *J. Phys. Chem. B* **2006**, *110*, 22910−22916.
(29) Krimm, S.; Abe, Y. Intermolecular Interaction Effects in the Amide I Vibrations of β Polypeptides. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 2788−2792.
(30) Jansen, T. L. C.; Dijkstra, A. G.; Watson, T. W.; Hirst, D.; Knoester, J. Erratum: Modeling the Amide I Bands of Small Peptides. *J. Chem. Phys.* **2012**, *136*, 209901.
(31) Torii, H.; Tasumi, M. Ab Initio Molecular Orbital Study of the Amide I Vibrational Interactions Between the Peptide Groups in Di- and Tripeptides and Considerations on the Conformation of the Extended Helix. *J. Raman. Spectrosc.* **1998**, *29*, 81−86.
(32) Chung, H. S.; Tokmakoff, A. Visualization and Characterization of the Infrared Active Amide I Vibrations of Proteins. *J. Phys. Chem. B* **2006**, *110*, 2888−2898.
(33) Ham, S.; Kim, J.-H.; Lee, H.; Cho, M. Correlation Between Electronic and Molecular Structure Distortions and Vibrational Properties. II. Amide I Modes of NMA−N D_2O Complexes. *J. Chem. Phys.* **2003**, *118*, 3491−3498.
(34) Choi, J.-H.; Ham, S.; Cho, M. Local Amide I Mode Frequencies and Coupling Constants in Polypeptides. *J. Phys. Chem. B* **2003**, *107*, 9132−9138.
(35) Bour, P.; Keiderling, T. A. Empirical Modeling of the Peptide Amide I Band IR Intensity in Water Solution. *J. Chem. Phys.* **2003**, *119*, 11253−11262.
(36) Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. Ultrafast Vibrational Spectroscopy of Water and Aqueous N-Methylacetamide: Comparison of Different Electronic Structure/Molecular Dynamics Approaches. *J. Chem. Phys.* **2004**, *121*, 8887−8896.
(37) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(39) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.

(40) Cheatum, C. M.; Tokmakoff, A.; Knoester, J. Signatures of β-Sheet Secondary Structures in Linear and Two-Dimensional Infrared Spectroscopy. *J. Chem. Phys.* **2004**, *120*, 8201.
(41) Karjaleinen, E.-L.; Ravi, H. K.; Barth, A. Simulation of the Amide I Absorption of Stacked β-Sheets. *J. Phys. Chem. B* **2011**, *115*, 749−757.
(42) Schweitzer-Stenner, R. Simulated IR, Isotropic and Anisotropic Raman, and Vibrational Circular Dichroism Amide I Band Profiles of Stacked β-Sheets. *J. Phys. Chem. B* **2012**, *116*, 4141−4153.
(43) Welch, W. R. W.; Keiderling, T. A.; Kubelka, J. Structural Analyses of Experimental ^{13}C Edited Amide I′ IR and VCD for Peptide β-Sheet Aggregates and Fibrils Using DFT-Based Spectral Simulations. *J. Phys. Chem. B* **2013**, *117*, 10359−10369.
(44) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys. 79*, 926−935.
(45) Strader, M. L.; Feller, S. E. A Flexible All-Atom Model of Dimethyl Sulfoxide for Molecular Dynamics Simulations. *J. Phys. Chem.* **2002**, *106*, 1074−1080.
(46) Cieplak, P.; Caldwell, J. W.; Kollman, P. A. Molecular Mechanical Models for Organic and Biological Systems Going Beyond the Atom Centered Two Body Additive Approximation: Aqueous Solution Free Energies of Methanol and N-Methyl Acetamide, Nucleic Acid Base, and Amide Hydrogen Bonding and Chloroform/Water Partition Coefficients of the Nucleic Acid Bases. *J. Comput. Chem.* **2001**, *22*, 1048−1057.
(47) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(48) Roy, S.; Lessing, J.; Meisl, G.; Ganim, Z.; Tokmakoff, A.; Knoester, J.; Jansen, T. L. C. Solvent and Conformation Dependence of Amide I Vibrations in Peptides and Proteins Containing Proline. *J. Chem. Phys.* **2011**, *135*, 234507.
(49) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319−2327.
(50) Barber-Armstrong, W.; Donaldson, T.; Wijesooriya, H.; Silva, R. A. G. D.; Decatur, S. M. Empirical Relationships Between Isotope-Edited IR Spectra and Helix Geometry in Model Peptides. *J. Am. Chem. Soc.* **2004**, *126*, 2339−2345.
(51) Karjalainen, E.-L.; Ersmark, T.; Barth, A. Optimization of Model Parameters for Describing the Amide I Spectrum of a Large Set of Proteins. *J. Phys. Chem. B* **2012**, *116*, 4831−4842.
(52) Cai, K.; Wang, J. Molecular Mechanics Force Field-Based Map for Peptide Amide-I Mode in Solution and Its Application to Alanine Di- and Tripeptides. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9149−9159.
(53) Cai, K.; Su, T.; Lin, S.; Zheng, R. Molecular Mechanics Force Field-Based General Map for the Solvation Effect on Amide I Probe of Peptide in Different Micro-Environments. *Spectrochim. Acta Mol. Biomol. Spectros.* **2014**, *117*, 548−556.
(54) Jansen, T. L. C. The linear Absorption and Two-Dimensional Infrared Spectra of N-Methylacetamide in Chloroform Revisited: Polarizability and Multipole Effects. *J. Phys. Chem. B* **2014**, DOI: 10.1021/jp5012445.