Empirical modeling of the peptide amide I band IR intensity in water solution

Petr Bouř [a)]
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo náměstí 2, 16610, Praha 6, Czech Republic

Timothy A. Keiderling
Department of Chemistry (M/C 111), University of Illinois at Chicago, Chicago, Illinois 60607-7061

(Received 29 July 2003; accepted 5 September 2003)

An empirical correction to amide group vacuum force fields is proposed in order to account for the influence of the aqueous environment on the C═O stretching vibration (amide I). The dependence of the vibrational absorption spectral intensities on the geometry is studied with density functional theory methods at the BPW91/6-31G** level for N-methyl acetamide interacting with a variety of of water molecule clusters hydrogen bonded to it. These cluster results are then generalized to form an empirical correction for the force field and dipole intensity of the amide I (C═O stretch) mode. As an example of its extension, the method is applied to a larger (β-turn model) peptide molecule and its IR spectrum is simulated. The method provides realistic bandwidths for the amide I bands if the spectra are generated from the *ab initio* force field corrected by perturbation from an ensemble of solvent geometries obtained using molecular dynamic simulations. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1622384]

I. INTRODUCTION

Models for the influence of solvent are becoming increasingly popular in computational chemistry, since they reflect the actual environment of most experimental results for which an interpretive basis is sought, and since including solvent effects often brings a substantial increase in the accuracy of the simulations.[1–4] Vibrational spectra of the peptide and protein amide I band provide a typical and very important example where such an improvement is desirable. The amide I occurs in a spectral region of little overlap with other modes and is a relatively local motion (amide C═O stretch), whose spectral intensity dominates the IR spectrum, is not mixed with other modes, and whose band shape reflects molecular conformation.[5,6] However, in solution the C═O bond stretching is strongly affected by solvent, particularly formation of hydrogen bonds with water, the common solvent for biological samples.[4,7–9] Apart from the absorption, faithful reproduction of frequency shifts and fine vibrational mode splitting in solvated systems is also desirable for interpreting a wide range of analytical methods, including vibrational circular dichroism (VCD),[10] Raman or Raman optical activity (ROA) spectra.[11,12]

N-methyl acetamide (NMA), a single amide ''blocked'' with methyl groups, has traditionally been used for accurate *ab initio* computations as a relatively faithful model for the properties of the peptide linkage.[13–17] Because of the importance of the amide I band for vibrational spectroscopy, in this study we explore how the vibrational frequencies and IR intensities are influenced by the number and geometry of solvent molecules explicitly included in an *ab initio* calculation using DFT (density functional theory) methods. Particularly, we extend the results of Ham *et al.*[14] so that an empirical force field correction based on the electrostatic field perturbation of this mode by the solvent can also be applied to DFT computations for larger molecules. This manuscript consists of two conceptual parts, DFT computations on NMA/water clusters and development of an empirical model suitable for correcting amide I calculations for oligomeric peptide systems. For the *ab initio* computations, the more precise DFT methods at the BPW91/6-31G** level is used instead of the HF approximation that was used by previous workers.[8,14] Additionally, instead of using an *ad hoc* cluster construction we derive the solvent geometries used for these tests from molecular dynamic (MD) simulation configurations and thus vary their structures in a more systematic way. As a model oligopeptide calculation, we simulate IR spectra for a pentapeptide constrained to a β-turn geometry, chosen in particular as a secondary structure component for which solvation by water could lead to significant effects.[18,19]

II. CLUSTER CALCULATION

A. MD snapshot clusters

With the aid of the TINKER molecular software package,[20] the NMA molecule was placed into a box of water (of a size of 18.56 Å) and the geometry minimized under periodic boundary conditions with the AMBER[21] force field. Subsequently, a molecular dynamic simulation was run. After equilibrium was achieved, ten configurations were generated, separated by 1000 1-femtosecond steps. Using a homemade graphical interface, MCM95,[22] approximately spherical clusters were selected for each MD-derived configuration. These used a common cutoff of 6.8 Å, based on

[a)]Electronic mail: bour@uochb.cas.cz

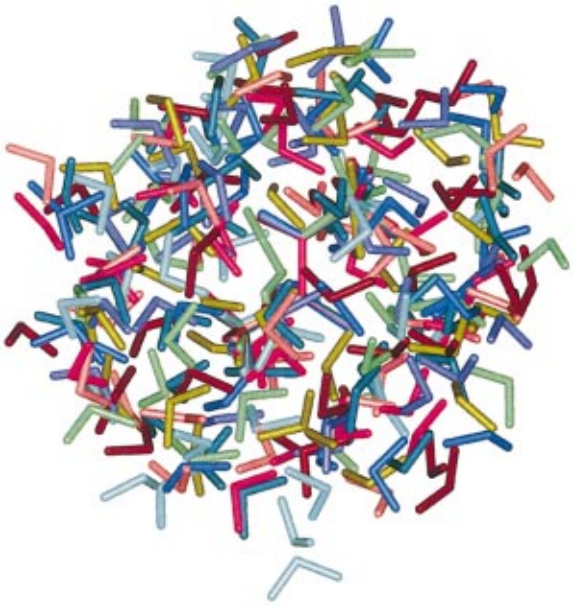

FIG. 1. (Color) An overlap of the ten NMA/water clusters, each having a different color, obtained from the MD simulation (the NMA molecules are aligned in the middle).

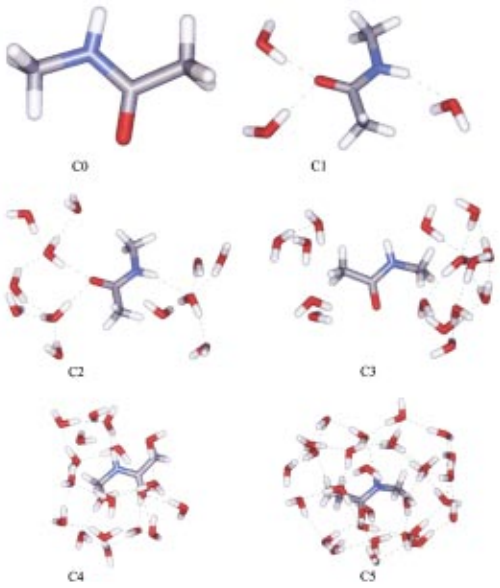

FIG. 2. The variable-size NMA/water clusters (as described in Table I).

the distances of the water from the NMA C(═O) atom. Retaining only complete water molecules falling within the cutoff resulted in a series of clusters containing 23, 28, 30, 31, 28, 27, 25, 30, 31, and 29 H_2O molecules, respectively. These geometries are shown overlapped, but distinguished by color, in Fig. 1 to give a sense of how the water positions varied for the 10 MD-derived structures. *Ab initio* density functional theory (DFT) computation of frequencies and dipole intensities were carried out for each of these clusters with the GAUSSIAN program[23] at the BPW91/6-31G** level.[24,25] To maintain the MD conformation but relax the higher frequency vibrational modes (particularly the amide I), so they would be calculated near their energy minimum, the geometry of the clusters was optimized prior to the frequency calculation (at the BPW91/6-31G** level) using the normal mode optimization method[26] (modes with $|\omega| < 300\ \mathrm{cm}^{-1}$ held fixed).

B. Variable-size clusters

For one of these particular MD configurations, the size of the cutoff sphere was varied in a wider range, so that various parts of the NMA hydration shell were present, as summarized in Table I. Clusters C2 and C3 were constructed so that water molecules approximately surround hydrophilic and hydrophobic molecular parts, respectively; see also Fig. 2. These selected ''subclusters'' allow one to obtain a better feel for the importance of different solvent interactions on the spectral parameters. It can be seen that different arrangements of the water molecules, particularly when resulting in H-bonded as opposed to ''free'' C═O groups (which only occur for isolated NMA or specially configured solvent molecules), can cause dramatic changes (reductions) in the amide I frequency. The amide II frequency rises by $\sim 30\ \mathrm{cm}^{-1}$ for the hydrogen-bond clusters as compared to isolated (vacuum) NMA, consistent with previous results.[7,17] Interestingly, the amide II frequency for the ''hydrophobic'' cluster where the water molecules are assembled around the methyl groups is shifted down to $1515\ \mathrm{cm}^{-1}$, which is lower

TABLE I. Variable-size NMA/water clusters.

Cluster	Number of water molecules	ω_{AI}[a]	ω_{AII}[b]	d(C═O) (Å)	Remark
C0	0	1732	1522	1.235	Vacuum
C1/n32q	3	1701	1554	1.250	First (H-bonding) hydration shell
C2/nmaq	12	1677	1544	1.254	First and second hydration shell
C3/n15	15	1704	1515	1.245	Only waters around —CH_3 groups ''Hydrophobic hydration''
C4/n19	19	1668	1552	1.253	Extended second hydr. shell
C5/n35	30	1658	1557	1.258	

[a]ω_{AI} amide I frequency for deuterated compound, in cm^{-1}.
[b]ω_{AII} amide II frequency for natural (all ^1H) compound, in cm^{-1}.

than the vacuum level. Surprisingly, in comparison to previous explicit water calculations, adding a second hydration shell continues to have a significant impact on the amide I frequency; however, frequencies obtained with larger shells seem to converge. This is consistent with our and other lab's findings that just computing three explicit waters H-bonded to NMA did not sufficiently correct the amide I frequency. Experimentally comparable frequencies were only obtained by adding a reaction field or using the polarized continuum model (PCM).[9,17] Analogous behavior was observed for hydrogen bonding to other NMA molecules instead of water; in this case, however, the amide II frequency shift was predicted to be bigger and amide I shift smaller than for H_2O.[8] There is an approximately linear relation between the C═O bond length and amide I frequency, but the regression relation is slightly worse than found in the previous study,[14] presumably because of the wider range of geometries investigated here.

III. EMPIRICAL CORRECTIONS

A. Amide I frequency correction

We follow the recent work[14] suggesting that the amide I frequency (ω) is proportional to electrostatic potential (φ) from water molecules at the six nuclei which form the core of the amide group

$$\Delta\omega=\omega-\omega_0=u\sum_{i=1}^{6} l_i\varphi_i, \tag{1}$$

where ω_0 is the frequency (or wave number) in vacuum and the unit factor $u=116\,080$ was introduced in order to comply with the original meaning[14] of the parameters l_i. Then, the frequency has units of wave number (cm^{-1}) and the potentials can be calculated as

$$\varphi_i=\sum_{j=1}^{j}\frac{q_i}{|r_i-r_j|},$$

where the partial atomic charges for water atoms are $q_H=0.412$ and $q_O=-0.814$, and the distances are measured in Å. This relation was originally derived based on a consideration of the anharmonicity of the C═O vibration and effective vibrational transition charges (derivative of the partial atomic charge q_i with respect to the normal mode Q ($\partial q_i/\partial Q$)) within the amide group atoms.[14,27] A conservation condition for the parameters, i.e., $\Sigma_{i=1}^{6} l_i=0$, was thereby developed. The values of the l_i parameters were originally determined on the basis of fitting to the vibrational frequencies obtained with HF/6-311++G** calculations for relatively small NMA–water clusters.[14]

In this work, we refit the frequency shift [Eq. (1)] using the DFT BPW91/6-31G** frequencies obtained for the (altogether 16) clusters described above. We included also the smaller clusters (with 0–15 water molecules) because we were interested in testing the validity of the formula (1) for any configuration of the water molecules; obviously, their influence to the fit was rather minor. We include the partially hydrated examples of Fig. 2 to exemplify how the works some extreme perturbations. Additionally, we found that only a minor error ($0.7\ cm^{-1}$ rms) arises when the number of the parameters is reduced to four (three independent, because of the conservation condition) parameters, corresponding to C, O, and the N and C_α bound to the C(═O). In fact, if only two-parameter fit for the two C═O group atoms was applied, most of the solvent amide I frequency shift was still recovered. But, in this case the error (rms) rose further by about $2\ cm^{-1}$. This would probably be acceptable for the training set for NMA, but may lead to bigger spectral distortions for more complicated systems. Additionally, because of the conservation condition, such a fit with only one independent parameter would not be sensitive to detailed geometry, e.g., to the rotation of the water shell around the C═O bond, and a principal advantage of the model would be lost. Thus, we feel that the four-parameter fit is a reasonable compromise for the ensemble of the clusters described above, with balanced contribution of the atoms connected to the carbonyl group. Similarly, we introduce an analogous fit of a set of parameters to the dipole strengths (D, units of debyes[2], are used here). All calculations presented here were done for NMA in D_2O with N-deuterated NMA, since the amide I vibration cannot be clearly separated from H_2O vibrations in the protonated form. Thus, our starting equations were

$$\Delta\omega=\omega-\omega_0=u\sum_{i=1}^{4} p_i\varphi_i, \tag{2}$$

$$\Delta D=D-D_0=\sum_{i=1}^{4} d_i\varphi_i,$$

$$\sum_{i=1}^{4} p_i=\sum_{i=1}^{4} d_i=0,$$

with the parameters given in Table II. Figure 3 shows the agreement of the fitted values with the *ab initio* (BPW91/6-31G**) results for the 16 clusters. The values are somewhat more scattered than those in the original work of Ham,[14] which we attribute to the higher diversity encompassed in our clusters, and the fact that we also include clusters (e.g., with H_2O only around the CH_3 groups) that would not be physically reasonable for aqueous solution. Indeed, for the MD snapshot clusters, whose overall geometries change but whose numbers of water and H bonds are relatively consistent, the dipolar strengths vary only by about 0.1 debye[2].

TABLE II. Fitted parameters for the empirical amide I frequency and dipole strength correction as obtained from the BPW91/6-31G** calculation.

i	p_i	d_i
1 (O)	−0.032 5	0.970
2(C(═O))	−0.006 24	−1.842
3 (C(CO))	0.002 69	0.209
4(N)	0.006 80	0.663

B. Force field and dipole derivative correction

In order to extend this method to oligomer molecules, we approximate the amide I mode by a local C═O stretch-

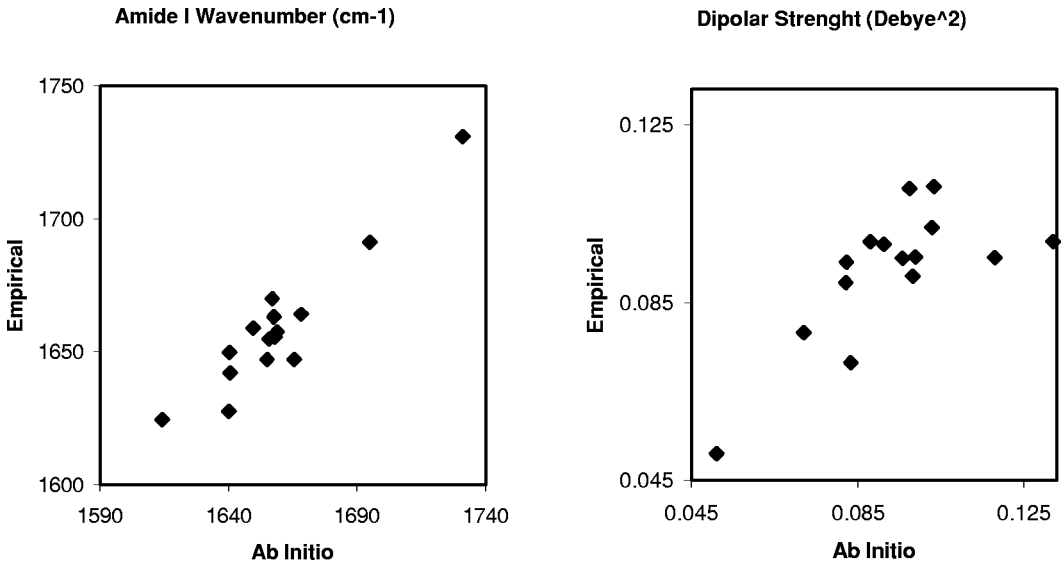

FIG. 3. Comparison of the amide I frequencies and dipolar strengths of the 16 NMA/water clusters (Figs 1 and 2) obtained *ab initio* (BPW91/6-31G**) and using the empirical correction [Eq. (2)]. The rms error was 8.7 cm^{-1} and 0.01 $debye^2$ for the left and right graph, respectively.

ing vibration, whose bond strength and frequency can vary as influenced by the solvent. Note that coupling terms responsible for interactions among individual C═O groups in a peptide are not affected within the model. The corresponding local harmonic oscillator potential can be represented as

$$\frac{kd^2}{2}=\frac{m\omega^2 d^2}{2}, \tag{3}$$

where m is a reduced mass, and the change of the force constant k upon hydration is

$$\Delta k \approx 2\frac{\Delta\omega}{\omega}k. \tag{4}$$

We wish to use these relationships to "repair" the vacuum force field (f_{ij}, second derivatives of the energy with respect to Cartesian atomic coordinates, x_i), so that might reflect the influence of the water potential according to Eq. (2) for the C═O stretching motion. Because of the normal mode (Q) to Cartesian transformation relations

$$x_i=S_{i,a}Q_a, \quad Q_a=S^{-1}_{a,i}x_i, \tag{5}$$

the vacuum C═O bond force constant ($d \approx Q$) is given by

$$k=\frac{\partial^2 E}{\partial d^2}=S_{i,d}\frac{\partial^2 E}{\partial x_i \partial x_j}S_{j,d}. \tag{6}$$

The Einstein summation convention (sums run over repeated indices occurring twice in a product) is used in this and the following formulas. Upon hydration, the C═O stretching force constant, k, changes according to Eq. (4), and for the resulting hydrated complex the Cartesian force field $\{f_{ij}\}$ we obtain results from the following working equation:

$$f'_{ij}=f_{ij}+p_f[S^{-1}_{i,d}S_{i',d}f_{i'j'}S_{j',d}S^{-1}_{j,d}]\frac{2\Delta\omega}{\omega}. \tag{7}$$

In Eq. (7), the $S^{-1}_{i,d}$ elements are equal to components of a unit vector **e** pointing along the C═O bond (e.g., positive for oxygen, negative for carbon), as are the elements of $S_{i,d}$ (but with a weighting factor given by the atomic masses, $S_{i,d}=\frac{16}{28}e_i$ for O, and $S_{i,d}=-\frac{12}{28}e_i$ for C, $i=x,y,z$), and $p_f=1.119$ is an empirical parameter accounting for the fact that the normal mode is not a pure C═O (carbon monoxide) motion. Setting $\omega=\omega_0=1732\ cm^{-1}$, which corresponded to the amide I frequency for N-deuterated NMA in vacuum calculated at the BPW91/6-31G** level, appears to be a reasonable approximation in Eq. (7).

In the same manner as the Cartesian force field, atomic polar tensors (derivatives of permanent molecular electric dipole moment μ^0 with respect to atomic coordinates) are changed upon vibration. As a first approximation the amide I transition dipole, μ, points along the C═O bond. The dipole strength (D) is defined as

$$D=\boldsymbol{\mu}^2, \tag{8}$$

where

$$\mu=|\boldsymbol{\mu}|\sim\omega^{-1/2}\frac{\partial\mu^0}{\partial x_i}S_{i,d}=\omega^{-1/2}P_{d,i}S_{i,d}. \tag{9}$$

The dipole derivative matrix $P_{d,i}$ is usually referred to as the atomic polar tensor. For the first-order changes upon hydration, similarly as for the for constant, we get

$$\Delta D \approx 2\mu\Delta\mu, \tag{10}$$

$$\frac{\Delta\mu}{\mu}\approx-\frac{\Delta\omega}{2\omega}+\frac{\Delta P_i S_{d,i}}{P_i S_{d,i}}, \tag{11}$$

and, using similar logic as for the force field, we get a working equation for the polar tensor correction

$$P'_{\alpha\alpha i}=P_{\alpha\alpha i}+p_d\left(\frac{\Delta D}{2D}+\frac{\Delta\omega}{2\omega}\right)S^{-1}_{i,\alpha}S_{i',d}P_{d,i'}. \tag{12}$$

Here, the approximation is somewhat more crude than that for frequencies, since we omit the dipole moment direction deviation, as well as the contribution of the off-diagonal **P**-tensor components. The p_d correction constant, which

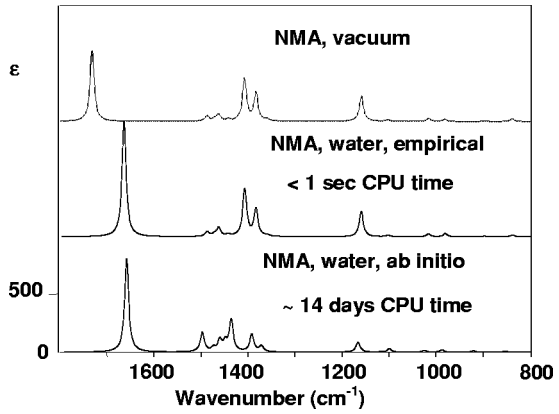

FIG. 4. Absorption spectra, comparison of the empirical and *ab initio* amide I frequency correction for a snapshot of one MD configuration, NMA plus 30 water molecules, taken into the *ab initio* (BPW91/6-31G**) computation.

should account for some of these simplifications, was set to 0.72. Note that the scaling constants p_d and p_f, which had to be introduced to extend the NMA results on general amides, even though they were obtained by a least-square fit, originate in the intrinsic nature of the amide I vibration in the *trans*-amide group.

IV. RESULTS

In the previous section we developed an empirical correction for the molecular force field and dipole derivatives that gives, for the 16 NMA clusters, the same amide I frequencies and intensities as obtained by Eq. (2) (in practice with a frequency error of ~1%). However, the method as formulated above is now applicable to systems with an arbitrary number of amide groups. Given the enormous difference in the computational effort for the empirical (fraction of second) and *ab initio* (days or weeks of the CPU time) procedure, the appeal of such an approximation is obvious. The advantage as well as a drawback of the method can be illustrated in Fig. 4, where *ab initio* and empirical computations of the absorption spectra for one of the snapshot clusters are compared. The clear advantage is that empirical calculation reproduces the frequency shift and a part of the intensity change for the amide I band in a fraction of the computer time that is required for the *ab initio* result. On the other hand, the influence of the solvent on the other modes, particularly the *ab initio* derived frequency and intensity changes in the CH bend/C–N stretching region ($1350–1550\ \mathrm{cm}^{-1}$), could not be reproduced empirically. This limitation results solely because, as formulated, our correction is for only one local mode, the C═O stretch. Analogous parametrization for nonlocal mode, such as amide II, would be more problematic. Nevertheless, because of the paramount importance of the amide I signal for analytical vibrational spectroscopy, and because of its exceptional sensitivity to solvent interactions, we feel it is useful to explore the possibilities of this proposed model further.

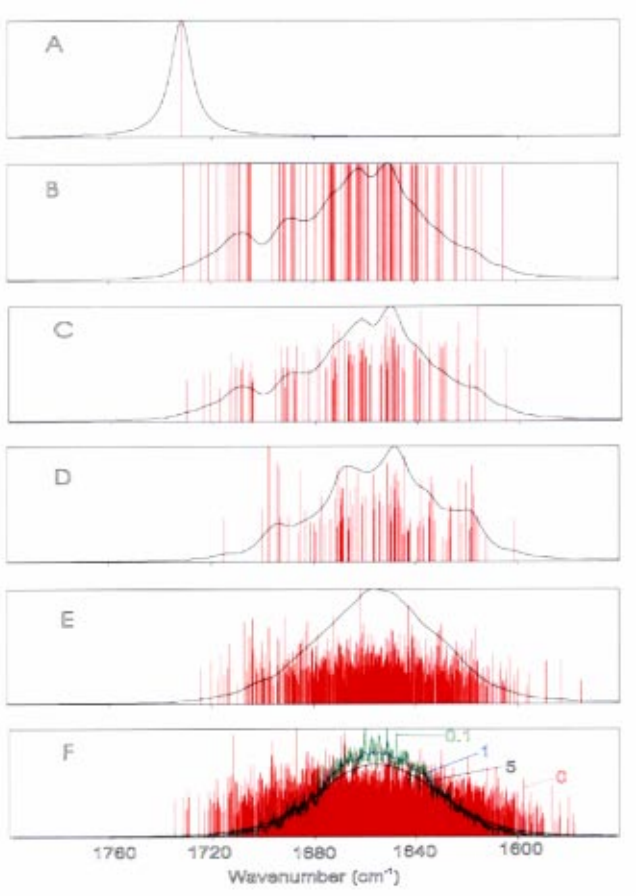

FIG. 5. Models of the amide I absorption band in NMA: (A) vacuum; (B)–(F) water clusters. (B) force field correction only; (C) force field and dipole derivatives correction applied. For (D)–(F) periodic bondary conditions were additionally applied. For (B)–(D), (E), and (F) 100, 1000, and 10 000 configurations (corresponding to the red lines) were averaged, respectively. For (F), the spectrum was simulated with four different Lorentzian bandwiths as indicated (units of halfwidth in cm^{-1}).

A. MD simulation of NMA in water solution

In order to test the capabilities of our model for simulation of vibrational spectra, we performed an MD simulation for NMA in a small water cage with periodic boundary conditions. The dimension of the cage was 13.08 Å and it contained 1 NMA and 69 waters for a total of 219 atoms. With the TINKER programs, the structure was energy minimized with periodic boundary conditions, then an MD simulation was initiated (as an NpT ensemble: constant number of particles, temperature of 300 K, and a pressure of 1 atmosphere; with 1 fs steps). After about 2000 steps equilibrium was achieved (according to temperature and energy floating averages), and configuration samples were then taken after each additional 1000 Newtonian steps. Altogether, 10 000 geometries were generated, which could be used to empirically adjust the vacuum (DFT/BPW91/6-31G**) NMA force field and dipole derivatives, using Eqs. (7) and (12).

These calculational outcomes are summarized by the spectral simulations in Fig. 5. The spectra A–E are represented as vertical (red) lines as well as by assigning arbitrary Lorentzian bands of $7.5\ \mathrm{cm}^{-1}$ band width (full width at half height, FWHH); for the last spectrum (F) the width is varied

in the interval $0-5\ \mathrm{cm}^{-1}$ as indicated in the figure. Clearly, except for the vacuum calculation, the width of the Lorentzian band assumed does not determine the resultant width of the spectral signal, since it is much smaller than the heterogeneous broadening ($\sim 50\ \mathrm{cm}^{-1}$). Nevertheless, for the smaller bandwidths of 1 and $0.1\ \mathrm{cm}^{-1}$ statistical fluctuations are still visible even when the full number of 10 000 configurations is averaged. A best fit to the simulated band [Fig. 5(F), with the width of $5\ \mathrm{cm}^{-1}$] was achieved with a mixed Gaussian (84%)–Lorentzian (16%) shape. We speculate that the ''Lorentzian part'' of the band indirectly reflects an influence of the molecular shape on the absorption profile, e.g., deviations from spherical geometry, both for the solvent and the solute.

The relatively high vacuum frequency [$\sim 1730\ \mathrm{cm}^{-1}$, Fig. 5(A)] drops to $\sim 1650\ \mathrm{cm}^{-1}$ for the amide I when averaged over all the water ''clusters'' in the box. This value is much closer to that which is experimentally observed,[17] but there is a substantial variation in C═O frequencies computed for each MD configuration sampled [Fig. 5(B)]. A relatively minor correction to the band shape results from the influence of the perturbation field on the dipole strengths in each cluster when averaged to give an absorption profile [compare Figs. 5(B) and 5(C)]. There is a significant difference between the vacuum and all the other values of the dipole strengths in Fig. 3, suggesting that, once the hydrogen bonds to water are formed, the C═O absorption intensity (dipole strength) varies little with added solvation. Application of periodic boundary conditions for calculation of the potential causes a slight narrowing of the absorption band and increases its symmetry, as can be seen by comparison of [Figs. 5(C) and 5(D)]. Similarly, when the periodic boundary condition was omitted completely in the simulation, an additional higher-frequency wing of the absorption band around $1710\ \mathrm{cm}^{-1}$ was observed (the spectrum is not shown here). In this test, the NMA molecule was expelled from the inside to the surface of the water drop during MD simulation, due to its relative hydrophobicity. Nevertheless, when a greater number of MD configurations are properly averaged [Figs. 5(E) and 5(F)] the band profile becomes smooth and the shape symmetric.

B. Solvent correction of the vacuum computation on an oligopeptide amide I band

In this section, absorption and vibrational circular dichroism spectra of a penta-peptide, Ala-Ala-Aib-D-Ala-Ala-$NHCH_3$, are simulated; Ala = L-alanine, Aib = aminoisobutyric acid (or α,α-dimethylglycine). Its geometry is shown in Fig. 6. The secondary structure of this peptide fragment corresponds to a β-hairpin type I′ loop of a longer peptide, the structure of which was determined by x-ray crystallography.[28] According to our theoretical simulations and recent experimental experience[29] such -Aib-D-Ala- turn motifs should be quite stable, so these results have potential direct application. Thus, the x-ray structure may be maintained at some level in solution, and this structure may thus be relevant for simulations including solvent, partially justifying our lack of a full search for the solvated peptide itself.

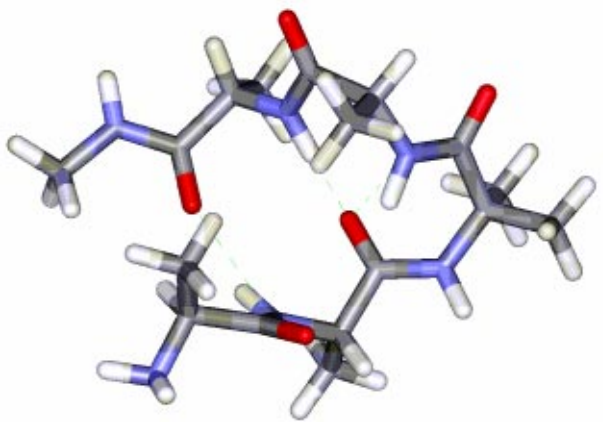

FIG. 6. Structure of the β-hairpin type I′ pentapeptide fragment, Ala-Ala-Aib-(D-Ala)-Ala-$NHCH_3$.

The geometry of the fragment was generated with the TINKER protein builder, but with the torsion angles restricted to the x-ray determined values. Then, the molecule was solvated in a cubic cage of water, 18.56 Å on a side. The system was allowed to minimize for 50 steps in order to relax the highest-energy geometric parameters (which effectively repairs inadequacies in the structure as obtained from the builder). Then, the positions of the peptide atoms were restrained, and the MD simulation was run for only motions of the water molecules. We feel this was a reasonable approach for this test calculation since the AMBER force field used here could have difficulty predicting realistic changes of the peptide secondary structure and its coupling with the solvent arrangement. This is especially true since the side-chain interactions present for a real peptide are eliminated in this model pentapeptide (i.e., only methyls are used, since only Ala and Aib form this sequence) for purposes of simplifying *ab initio* calculations. In fact, variation of the peptide structure is not the goal of this simulation, but rather the impact on the spectrum of the variation in the water structure is what we seek to simulate. The conditions of the MD run were analogous to those used for the NMA cluster simulations. After equilibrium was achieved (~ 1000 steps at 300 K) 10 000 Newtonian steps were obtained and the computed geometry was recorded every 100 steps, so that 100 configurations were averaged. In this case further increase of the number of configurations did not have significant impact on the resultant spectrum.

In a parallel computation, the vacuum spectral frequencies and intensities were calculated for the fragment at the DFT/BPW91/6-31G** level. Prior to the force field calculation, normal modes with $|\omega| > 300\ \mathrm{cm}^{-1}$ were allowed at this level to relax the peptide structure from the x-ray derived geometry, which caused a negligible change in secondary structure. The force field and dipole tensors for this pentapeptide in vacuum were then calculated and the result (N-deuterated) is shown in Fig. 7 (top) for the IR of the amide I band. Based on the distributions of the water molecules

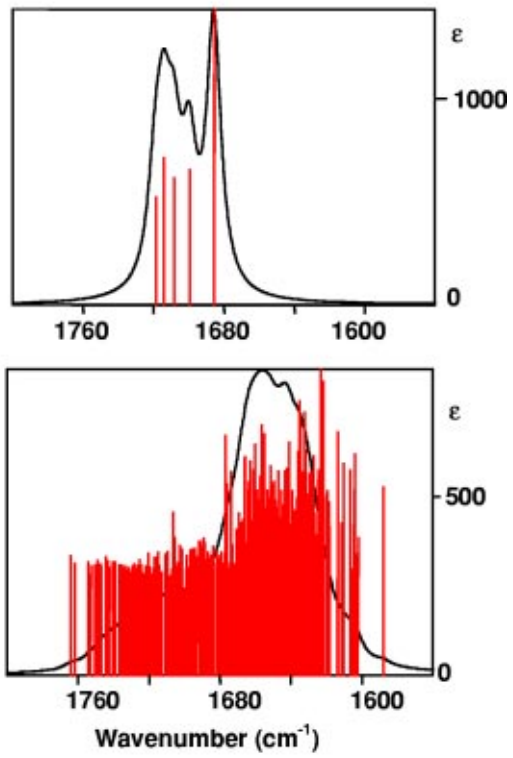

FIG. 7. The influence of the hydration on the absorption spectra of the pentapeptide. On the top the BPW91/6-31G** amide I IR simulation in vacuum is presented; the bottom spectrum was obtained by simulating the influence on the amide I of 168 water molecules in a periodic cage, summed over 100 configurations obtained from an MD simulation.

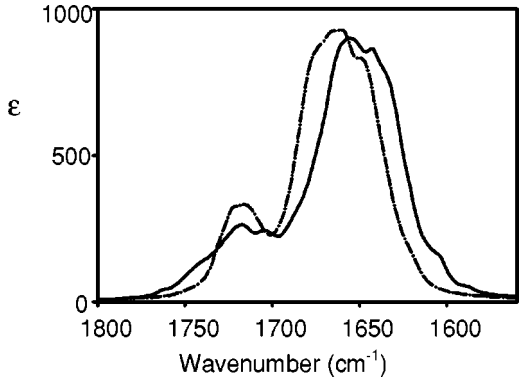

FIG. 8. Simulated absorption spectra for the pentapeptide and 168 water molecules for two temperatures (300 K solid line, 380 K dot-dashed line), summed as in Fig. 7.

around the peptide as obtained in the sampled configurations from the MD simulation, the empirical corrections [Eqs. (7) and (12)] to the *ab initio* force field and dipole derivatives were applied. The resulting solvent-corrected amide I′ spectrum for this peptide is in Fig. 7 (bottom). Both the vacuum and solvated peptide absorption spectrum of the amide I band are more complicated than for the NMA molecule (compare with Fig. 5). Additionally, due to the number of peptide groups in the peptide and their compact turn geometry, some internal hydrogen bonds are already formed in vacuum. In particularly, the inner loop carbonyl oxygen forms a bifurcated hydrogen bond to two other amide groups, which is reflected in the vacuum absorption spectrum as the most intense, and lowest frequency, IR band at 1685 cm^{-1}. This mode is distinct from the absorbances of nonbonded or weakly H-bonded groups. This pattern is significantly disturbed by the empirical water correction, where the center of the amide I absorption band shifts to approximately 1650 cm^{-1} and a significant shoulder/sideband develops to higher wave number. Clearly, despite the general broadening, its nonsymmetric shape still reflects the different carbonyl groups.

Additional efforts were made to simulate the VCD of the pentapeptide with water. Obviously, the VCD of NMA would be zero because of its effective planar symmetry. The hydration also causes large changes in the simulated VCD spectrum. Both the vacuum and solvent model predict a net negative signal for the pentapeptide turn amide I mode and a ± couplet from high to low wave numbers. The empirical solvent correction, however, leads to a more negatively biased VCD and gives a rise to an extra positive peak around 1620 cm^{-1}. Thus, the correction can potentially improve VCD simulations as well. However, we feel that detailed assessment of its performance for VCD should be left for the future, when more comparisons with experimental data are available.

In order to document further possible application of our model, we repeated the pentapeptide simulation for a different temperature of 380 K. The change in spectral frequencies and intensities with temperature can be seen in Fig. 8. The strongest intensity shifts to higher frequency, which is in accord with trivial expectation, since hydrogen bonds are less favored at elevated temperatures. Our thermal unfolding data for a number of β-hairpins do show a steady increase in temperature for the main absorption band; however, it is difficult to quantitatively separate the salvation effect from a conformational change (fraying of the hairpin strands).[30] On the basis of our experimental experience we suppose that the model overestimates the temperature influence on the amide I band, probably due to the limitations of the AMBER force field. Similarly, the model did not qualitatively reproduce the differences in bandwidths for H_2O and D_2O NMA solutions.[17] This bandwidth can be partially rationalized due to the resonant coupling between the amide I and H_2O bending modes, which does not exist for D_2O. In test computations (not shown here) we ran unrestricted MD simulations for the whole system, with the peptide allowed to move as well, and obtained qualitatively similar results, suggesting that this turn structure corresponds to a well-defined minimum on the molecular potential energy surface.

So far, we have used the empirical correction for averaging effects of solvent interaction, which is most appropriate for a large system. Nevertheless, it is interesting to test the extent to which a particular solvent configuration can be modeled. This is tested on *ab initio* computations for the peptide surrounded by 9 water molecules, the geometry of which can be seen in Fig. 9. This particular geometry was randomly selected from one of the MD configurations described above; however, to simplify the computation, only water molecules making hydrogen bonds to the peptide were

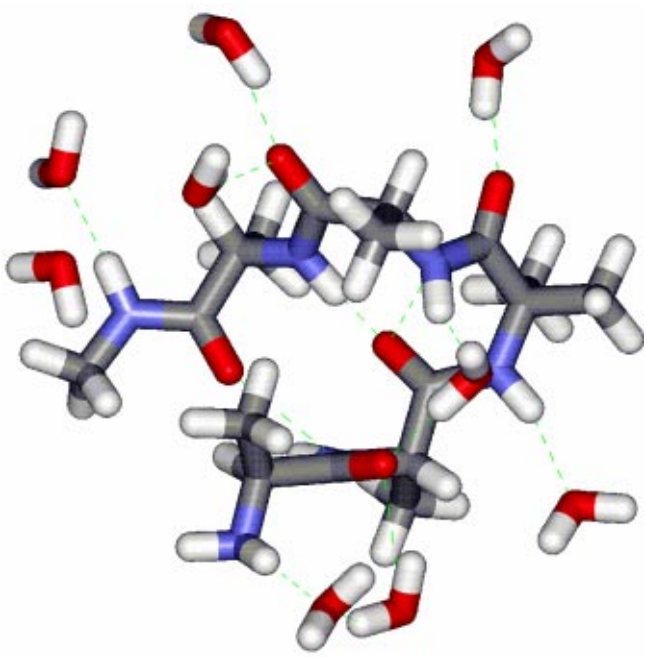

FIG. 9. Structure of the minimally solvated pentapeptide fragment hydrated by nine water molecules.

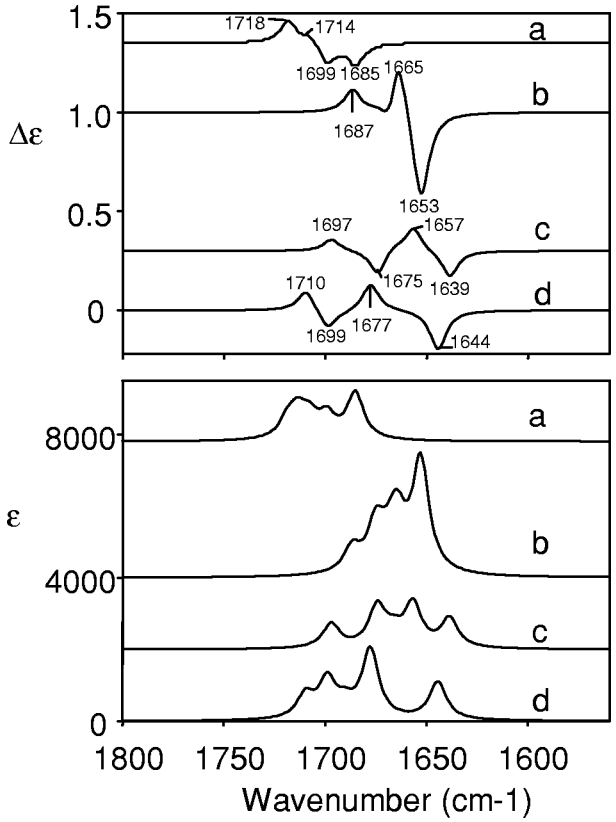

FIG. 10. Calculated (BPW91/6-31G**) absorption (ε) and VCD ($\Delta\varepsilon$) spectra of the pentapeptide: (a) for vacuum; (b) hydration modeled by the PCM continuum model alone; (c) empirical solvent correction [Eqs. (7) and (12)], and (d) full *ab initio* calculation. See Figs. 6 and 9 for the geometries for (a) and (b), and (c) and (d), respectively.

retained. In Fig. 10, the *ab initio* (BPW91/6-31G**, obtained using the normal mode constrained optimization) vacuum and hydrated IR and VCD spectra are compared to those calculated by applying our empirical correction to this smaller cluster as well as to the outcome of the IEF-PCM continuum model.[23,31] We here reference the more sensitive VCD pattern as a probe of the normal mode ordering. The vacuum computation (a) predicts high frequencies (a peak at 1718 and a shoulder at 1714 cm^{-1}), as is normal for DFT peptide frequency calculations since hydrogen bonds are not formed for the loop C═O residues. The inner loop C═O group provides the lowest frequency (1685 cm^{-1} VCD) band, because of the extensive bifurcated H bonding. The 1699 cm^{-1} VCD band arises from the ends of the molecule, which is virtually a short, flat, antiparallel β-sheet hairpin segment. However, in the solvent environment it is the β-sheet component that provides the lowest frequency peak, at 1639 and 1644 cm^{-1} for spectra c, d in Fig. 10, respectively. Unlike in vacuum, due to H bonding to the solvent, the inner and outer C═O groups provide modes at similar frequencies, providing a positive band for the inner (1677 cm^{-1} *ab initio*, 1657 empirical) and a negative signal (1699 cm^{-1} *ab initio*, 1675 empirical) for the outer C═O group vibrations. The inner β-sheet C═O groups are shielded from the solvent, and thus it is not surprising that their net frequency is not changed much by the solvent (to 1710 for *ab initio* and 1699 cm^{-1} for the empirical model), but the inversion of the VCD sign for these modes is quite surprising. The PCM model without explicit waters (b) generally improves the vacuum frequencies, but does not reproduce the VCD sign pattern and normal mode ordering. In particular, it predicts that the outer β-sheet C═O group vibrates with higher frequency (1665 cm^{-1}) than the inner loop group (1653 cm^{-1}), as might be expected for a continuum model, but in disagreement with the results from the explicit models (c, d). The comparison in Fig. 10 may be affected by a slight relaxation of the geometry in the *ab initio* computation. Nevertheless, we can draw two important conclusions: (i) mode reordering in the solvent can happen; and (ii) the empirical model reproduces qualitative patterns found in the full *ab initio* simulation, although more detailed assessment of the apparent frequency errors should be the topic of a separate study. Note that this empirical scheme is the only method, apart from the *ab initio* computations, that can reproduce the extended band splitting needed for qualitatively correct VCD and ROA spectra simulation.

V. DISCUSSION

In this work on NMA solvated with a cluster of waters we have extended previous knowledge about hydration of the amide group by addressing these particular problems: what is the necessary size of an explicit water environment needed for a computation to obtain a realistic frequency shift, what is the influence of that water environment on spectral intensities, and how can these results be generalized for application to other systems? We extended the results of Ham and co-workers,[14] taking into account more general clusters and using the more accurate, especially for vibrational frequencies, DFT theory as a basis for the computations, instead of the HF approximation.

Our solvent modeling is focused on the amide I vibration and thus cannot compete fully with more general solvent models, such as continuum polarization models[32–34] or Car–Parrinello molecular dynamics simulations[35] or direct *ab initio* cluster calculations.[17] However, these methods computationally become extremely resource intensive. Furthermore, for peptide conformational analysis with IR spectroscopy the amide I vibration provides both the strongest IR signal and the most useful diagnostic tool, while at the same time it is strongly influenced by hydrogen bonds to water. Therefore, any vacuum calculations of frequencies are in need of significant correction. *Ab initio* frequencies have been corrected by scaling in many previous studies, but such a scaling factor for missing the H-bond effect on the amide C═O would be very large (compared to the 5% or so reduction usually required for DFT frequencies). Such a large perturbation may change the nature of the modes, so that using just a simple scaling could miss that important impact of the H bonds to solvent. This has been shown by many tests, e.g., where one can include water explicitly in the DFT computations.[17,36] The empirical method presented here allows us to make such corrections to the FF on an arbitrarily large system. An example was given in Fig. 10, where for the amide I mode we can, unlike with previous modeling, with minimal computational effort obtain realistic frequencies sensitive to actual water orientations and, in addition, obtain heterogeneous band broadening via molecular dynamics averaging. The spectral bandwidth has been a point not addressed in conventional spectral simulations, and is usually circumvented by an introduction of arbitrary empirical constant chosen to mimic the experimental bandwidth.

From the computation on the pentapeptide turn model, we can see that the broadening caused by the solvent is comparable in magnitude with the dispersion caused by the intramolecular interactions. A significant part of the dispersion in the vacuum calculations is due to the difference of ''free'' and H-bonded C═O groups, whereas in aqueous solution all C═O groups are likely to be H-bonded, unless they are in a particularly protected globular fold. The NMA result provides a control with regard to the amide I dispersion due to (ϕ,ψ) variation, as there is only one amide and the spectral width must come primarily from the heterogeneous broadening due to the water environment. In this case it would seem that the water dispersion effect might correspond to the experimental value of $50\,\mathrm{cm}^{-1}$ found for H_2O solution (FWHH), but in that case mixing with and correcting for solvent modes is problematic.[17] Consequently, we simulated the spectra for N-deuterated NMA in D_2O, where the observed width somewhat narrower ($\sim 30\,\mathrm{cm}^{-1}$). Thus, while our method yields the useful idea of the heterogeneous broadening, it overestimates the effect at this time.

Also, the frequencies obtained by this correction are indeed much closer to experiment than by vacuum calculation or even by just using just a single layer of explicit, H-bonded waters in the DFT calculations.[9,11,17] Only with several layers can one get a converging frequency correction. The same sort of observation was made previously when it was reported that with a single layer of water H-bonded to NMA, a polarized continuum model or reaction field was needed to get reasonably accurate frequencies.[9,17] The resulting frequencies are still high by a few cm^{-1}, but this just reflects the DFT results, since the empirical method was calibrated against the set of DFT (BPW91/6-31G** level) calculations done for the various clusters used to develop the fitting parameters. The correction can be no better than the *ab initio* results on which it is based, but it offers convenience and applicability to large systems. Furthermore, it could be easily further repaired by introduction of a scaling factor, which in this case would be a minor correction, not likely to distort the modes significantly. At this point, however, scaling would not extend physical insight into the amide I vibration. It may be also worthy to model the water electrostatic potential differently than from the fixed partial atomic charges, e.g., taking into account water polarizability or using proper electrostatic field-fitted charges obtained at the given (BPW91/6-31G**) method. This would, however, make the method more robust. Because of the relatively good fit of the frequency shifts obtained by the charges given in the original reference (0.412 and -0.814), we rather rely on the fitting parameters that may incorporate also inaccuracies in the determination of the atomic charges. This parametrization is also close to that used in the AMBER force field (0.417 and -0.834).

A more serious limitation is the lack of a similar correction for the amide II mode. This mode is more coupled to other vibrations, and the same sort of scaled correction as used here for the amide I may not be possible without consideration of the influence on other parts of the spectrum, and certainly it could not have the same formalistic basis. The amide II frequency is shifted up in frequency by $\sim 30\,\mathrm{cm}^{-1}$ upon hydration. This is in the opposite direction (due to H-bonding stiffening the bending motion) but is of less magnitude than for amide I. The experimental significance of the amide II has not been so well developed;[37,38] nevertheless, it is clear that measurements in H_2O are of growing importance for IR and Raman so that future modeling would have to deal either with solvent correction to the amide II mode or include the influence of solvent onto all vibrational degrees of freedom more consistently. Other mid-IR modes are obviously uncorrected as well, but these are mostly well represented by vacuum computations. As a full *ab initio* approach is still impossible for larger systems, extensions of this empirical method or a combination of explicit and reaction field solvent models may provide a solution to consideration of solvent effects.[9,17,39]

VI. CONCLUSIONS

These cluster computations indicated that, for an explicit solvent model, positions of water molecules are as important as their number, if the influence on the amide I frequency is to be reproduced correctly. The empirical correction presented here takes into account this geometry dependence and speeds up the computations considerably when compared with *ab initio* methods. Thus, realistic amide I spectral frequencies could be calculated for a model pentapeptide and heterogeneous broadening and temperature dependence

could be simulated when the correction was averaged for an ensemble of geometries obtained from a molecular dynamics simulation.

ACKNOWLEDGMENTS

This work was supported by grants to P.B. from the Grant Agency of the Academy of Sciences (A4055104), Grant Agency of the Czech Republic (Grant 203/01/0031), and in part by a grant to T.A.K. from the donors of the Petroleum Research Fund, administered by the American Chemical Society and from the National Science Foundation (CHE03-16014).

[1] M. W. Wong, M. J. Frisch, and K. B. Wiberg, J. Am. Chem. Soc. **113**, 4776 (1991).
[2] S. Miertus, E. Scrocco, and J. Tomasi, Chem. Phys. **55**, 117 (1981).
[3] M. Cossi, G. Scalmani, N. Rega, and V. Barone, J. Chem. Phys. **117**, 43 (2002).
[4] P. Bouř, Chem. Phys. Lett. **365**, 82 (2002).
[5] K. Kato, T. Matsui, and S. Tanaka, Appl. Spectrosc. **41**, 861 (1987).
[6] P. Pančoška, J. Kubelka, and T. A. Keiderling, Appl. Spectrosc. **53**, 655 (1999).
[7] M. Knapp-Mohammady, K. J. Jalkanen, F. Nardi, R. C. Wade, and S. Suhai, Chem. Phys. **63**, 240 (1999).
[8] R. Ludwig, O. Reis, R. Winter, F. Weinhold, and T. C. Farrar, J. Phys. Chem. B **102**, 9312 (1998).
[9] K. J. Jalkanen KJ, R. M. Nieminen, M. Knapp-Mohammady, and S. Suhai S, Int. J. Quantum Chem. **92**, 239 (2003).
[10] T. Keiderling, in *Practical Fourier Transform Infrared Spectroscopy*, edited by J. R. Ferraro and K. Krishnan (Academic, San Diego, 1990), pp. 203–284.
[11] K. J. Jalkanen, R. M. Nieminen, K. Frimand, J. Bohr, H. Bohr, R. C. Wade, E. Tajkhorshid, and S. Suhai, Chem. Phys. **265**, 125 (2001).
[12] P. Bouř, J. Kapitán, and V. Baumruk, J. Phys. Chem. A **105**, 6362 (2001).
[13] S. Ham and M. Cho, J. Chem. Phys. **118**, 6915 (2003).
[14] S. Ham, J. H. Kim, H. Kochan, and M. Cho, J. Chem. Phys. **118**, 3491 (2003).
[15] X. G. Chen, R. S. Stenner, S. A. Asher, N. G. Mirkin, and S. Krimm, J. Phys. Chem. **99**, 3074 (1995).
[16] S. Iuchi, A. Morita, and S. Kato, J. Phys. Chem. B **106**, 3466 (2002).
[17] J. Kubelka and T. A. Keiderling, J. Phys. Chem. A **105**, 10922 (2001).
[18] T. A. Keiderling and Q. Xu, Adv. Protein Chem. **62**, 155 (2002).
[19] B. Imperiali, S. L. Fisher, R. A. Moats, and T. J. Prins, J. Am. Chem. Soc. **114**, 3182 (1992).
[20] R. V. Pappu, R. K. Hart, and J. W. Ponder, J. Phys. Chem. B **102**, 9725 (1998).
[21] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).
[22] Program can be downloaded at http://hanicka.uochb.cas.cz/~bour/mcm95.zip
[23] M. J. Frisch, G. W. Trucks, H. B. Schlegel, *et al.*, GAUSSIAN 03, Revision A.1, (Gaussian, Inc., Pittsburgh, PA, 2003).
[24] A. Becke, Phys. Rev. A **38**, 3098 (1988).
[25] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[26] P. Bouř and T. A. Keiderling, J. Chem. Phys. **117**, 4126 (2002).
[27] M. Cho, J. Chem. Phys. **118**, 3480 (2003).
[28] S. Aravinda, N. Shamala, R. Rajkishore, H. N. Gopi, and P. Balaram, Angew. Chem., Int. Ed. Engl. **41**, 3863 (2002).
[29] V. Setnička, R. P. Hammer, T. A. Keiderling (unpublished).
[30] J. Hilario, J. Kubelka, T. A. Keiderling, J. Am. Chem. Soc. **125**, 7562 (2003).
[31] M. Cossi, V. Barone, B. Mennucci, and J. Tomasi, Chem. Phys. Lett. **286**, 253 (1998).
[32] L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).
[33] A. Klamt and G. Schürman, J. Chem. Soc., Perkin Trans. 2 **2**, 799 (1993).
[34] R. Gammi, C. Capelli, S. Corni, and J. Tomasi, J. Phys. Chem. A **104**, 9874 (2000).
[35] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[36] J. Kubelka, Ph.D. thesis (University of Illinois at Chicago, 2002).
[37] V. Baumruk, P. Pančoška, T. A. Keiderling, J. Mol. Biol. **259**, 774 (1996).
[38] V. P. Gupta, T. A. Keiderling, Biopolymers **32**, 239 (1992).
[39] T. M. Watson and J. D. Hirst, J. Phys. Chem. A **107**, 6843 (2003).