Empirical solvent correction for multiple amide group vibrational modes

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

David Michalík
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610, Praha 6, Czech Republic and Department of Analytical Chemistry, Institute of Chemical Technology, Technická 5, 166 28 Prague 6, Czech Republic

Josef Kapitán
Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 16610, Praha 6, Czech Republic

(Received 7 December 2004; accepted 28 January 2005; published online 8 April 2005)

Previously proposed solvent correction to the amide *I* peptide vibration was extended so that it can be applied to a general solvated chromophore. The combined molecular and quantum mechanics (MM/QM) method is based on a linear dependence of harmonic force field and intensity tensor components of the solute on solvent electrostatic field. For N-methylacetamide, realistic solvent frequency and intensity changes as well as inhomogeneous band widths were obtained for amide *A*, *I*, *II*, and *III* modes. A rather anomalous basis set size dependence was observed for the amide *A* and *I* vibrations, when bigger basis lead to narrowing of spectral bands and lesser molecular sensibility to the environment. For a model α-helical peptide, a *W*-shape of the vibrational circular dichroism signal observed in deuterated solvent for the amide *I* band was reproduced correctly, unlike with previous vacuum models. © *2005 American Institute of Physics*. [DOI: 10.1063/1.1877272]

I. INTRODUCTION

Combined molecular and quantum mechanics (MM/QM) methods became a popular choice for simulating of solvent effects.[1–11] Since interpretation of the infrared (IR) spectroscopic data is to a large extend dependent on their theoretical simulation involving detailed description of studied systems, realistic solvent models are particularly needed for polar, strongly hydrated biopolymers, such as peptides[8,12–15] and nucleic acids.[16,17] For peptides and proteins, the strong absorption bands of the amide group chromophore provide indispensable probes of the secondary structure, both in linear and differential (circular dichroism) absorption spectroscopy.[18–22] Although good qualitative agreement with the experimental signal can be often achieved with models based on *ab initio* computations on vacuum peptide fragments,[23–26] more accurate models are needed especially for nonperiodic inhomogeneously hydrated structures, such as β-hairpins.[27–29] In this work, we expand the electrostatic correction proposed for the amide *I* band previously,[8] so that it can be applied to any chromophore and arbitrary number of vibrational modes.

The role of the solvent electrostatics on solute properties been recognized quite early[30] and became utilized in many continuum solvent models.[4,31–35] Since these cannot discriminate among individual solvent molecules, more explicit approaches are looked for, providing better link between classical and quantum dynamics. For vibrational spectroscopy, models based on solvent electrostatic potential or its gradient (intensity of the electric field) sampled on solute bonds or atoms appear particularly promising.[7,36,37] For example, solvent frequency shifts of water O–H stretching vibrations were found to be approximately linearly proportional to water electrostatic field projected along the O–H bond, and thus could be used in MM/QM simulations of the absorption band profiles.[6,38] Similarly for N-methylacetamide (NMA), linear relation between the C═O bond length, stretching frequency, and solvent electrostatic potential at NMA atoms has been found.[39–41] Thus empirical corrections could be developed in order to model realistic amide *I* (C═O stretch) frequencies and inhomogeneous band widths.[7,13,41] For the NMA–water system, alternative approach was proposed with fitting based on geometrical parameters instead of the potential, providing encouraging results even for amide *II* and *III* modes.[42] We attempted to generalize the electrostatic procedure for multiple-amide systems and used the NMA fitting parameters in simulations of peptide amide *I* IR and vibrational circular dichroism (VCD) spectra.[8] Although the amide *I* mode is of primary importance in peptide vibrational spectroscopy, other vibrational bands of the amide group are also affected by the solvent in a similar manner.[43] Thus it appears natural to extend the model further, in order to capture the influence of the solvent on the whole amide group. In fact, the approach presented below is not dependent on particular chemical species and can be extended to any solvated chromophore. Also, rather minor attention in the previous models has been paid to the role of the approximation level and basis

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

set selection. Thus we test these dependencies more extensively. Rather surprising effect of increasing the basis set size was found for some modes, leading to a narrowing of the vibrational bands, i.e., the NMA molecules becomes less susceptible to external perturbations.

The manuscript is organized as follows. In the method section, the empirical correction to the force constants and dipole derivatives of a vibrating chromophore is introduced, based on the solvent electrostatic field modeled by fixed partial charges. In the next section, the model is calibrated against *ab initio* computations on small NMA–water clusters. Then simulated spectra in the aqueous environment are compared with experiment, in terms of frequency shifts and inhomogeneous band broadening. Finally, the model is applied to a α-helical 19-amide peptide and relation of simulated absorption and vibrational circular dichroism spectra to experiment is discussed.

II. METHOD

We wish to find dependence of vibrational frequencies on external (solvent) electrostatic field for any molecular part where vibrations are localized. For brevity, we talk about the amide moiety, but obviously the first order correction can be easily generalized for any other chemically distinct group. In the harmonic approximation the observable frequencies are given by the harmonic force field.[44] In order to eliminate the coordinate-system dependence and the redundancies in the force field, we diagonalize a partial field matrix. Particularly, for a set of the force constants $\{f_{\lambda\alpha,\mu\beta}\}$ comprising M atoms of interest (λ, $\mu=1\cdots M$; α, $\beta=x, y, z$), the eigenvalues $\{\Lambda_I\}$, $I=1\cdot\cdot 3\times M$, are given by

$$f_{\lambda\alpha,\mu\beta}=\sum_{I=1}^{3M} s_{\lambda\alpha,I}\Lambda_I s_{\mu\beta,I}, \tag{1}$$

with an orthogonal transformation

$$\sum_{I=1}^{3M} s_{\lambda\alpha,I}s_{\mu\beta,I}=\delta_{\lambda,\mu}\delta_{\alpha,\beta} \text{ and } \sum_{\alpha=1}^{3}\sum_{\lambda=1}^{M} s_{\lambda\alpha,I}s_{\lambda\alpha,J}=\delta_{I,J}, \tag{2}$$

where δ is the Kronecker δ-symbol. In order to obtain eigenvalues more relevant to experimental frequencies, we consider the mass-weighted force field $f_{\lambda\alpha,\mu\beta}=F_{\lambda\alpha,\mu\beta}/\sqrt{m_\lambda m_\mu}$, where $\boldsymbol{F}$ is the usual second-derivative matrix (Hessian) and m_μ atomic mass of an atom μ.

Based on previous works,[8,39–41] we model the influence of the solvent by a linear fit

$$\Lambda_I=\Lambda_{0I}+\sum_{j=1,N} b_{Ij}\varphi_j, \tag{3}$$

where Λ_{0I} is the force constant in vacuum, and φ_j is the electrostatic potential of the solvent at the nucleus j of the solute. In the present combined MM/QM scheme the potential is modeled by atomic partial charges of water. Note that different atoms may be included in Eqs. (1) and (3) and generally $M\neq N$; for example, in the peptide group, in order to investigate force constants for the HNCO atoms ($M=4$), we may include additionally the two neighboring α-carbons ($N=6$) in order to map better the surrounding electrostatic field.

When the potential is constant ($\varphi_1=\varphi_2=\cdots=\varphi_N$), we expect nearly zero intensity of the electric field and consequently no frequency shifts. This is automatically ensured with a conservation conditions (for each I)

$$\sum_{j=1,N} b_{Ij}=0. \tag{4}$$

The condition was already proposed for the amide I mode modeling, on the basis of charge conservation in a local atomic current model.[40,41]

The coefficients b_{Ij} can be obtained from *ab initio* computations on a "learning set" of m solute–solvent clusters (in this work 11 NMA–water clusters), via minimizing the root mean square differences between the computed and fitted $\{\Lambda_{iI}, i=1\cdot\cdot m, I=1..M\}$ force constants

$$\sum_{i=1\cdot m}\left(\Lambda_{iI}-\Lambda_{0I}-\sum_{j=1,N} b_{Ij}\varphi_{ij}\right)^2\rightarrow \min. \tag{5}$$

Now the potential φ_{ij} thus has a cluster (i) and atomic (j) index. The method of Lagrange multipliers for (4) and (5) leads to a set of linear equations for the fitting coefficients and a multiplier λ. For each force constant I we thus obtain the coefficients by matrix inversion:

$$\begin{pmatrix} \sum_{i=1..m}\varphi_{i1}\varphi_{i1} & \cdots & \sum_{i=1..m}\varphi_{iN}\varphi_{i1} & 1 \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1..m}\varphi_{i1}\varphi_{iN} & \cdots & \sum_{i=1..m}\varphi_{iN}\varphi_{iN} & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix}\begin{pmatrix} b_{I1} \\ \cdots \\ b_{IN} \\ \lambda \end{pmatrix}$$
$$=\begin{pmatrix} \sum_{i=1..m}(\Lambda_{iI}-\Lambda_{0I})\varphi_{i1} \\ \cdots \\ \sum_{i=1..m}(\Lambda_{iI}-\Lambda_{0I})\varphi_{iN} \\ 0 \end{pmatrix},$$

$$\mathbf{A}\cdot\mathbf{b}=\mathbf{c}\Rightarrow\mathbf{b}=\mathbf{A}^{-1}\cdot\mathbf{c}. \tag{6}$$

Providing that the NMA force field is transferable, the same fitting coefficients can be used in a solvent correction for any amide group in a peptide. New Cartesian force constants are then obtained from the vacuum values as

$$f_{\lambda\alpha,\mu\beta}=f^{(0)}{}_{\lambda\alpha,\mu\beta}+\sum_{I=1}^{3M} s_{\lambda\alpha,I}s_{\mu\beta,I}\sum_{j=1,N} b^I{}_j\varphi_j, \tag{7}$$

where the **s**-matrices are calculated for each chromophore separately within the first-order approach.

An influence of the solvent electrostatic field on the absorption and Raman intensities is treated in a similar way. In this case, however, working in the local normal mode space does not bring any convenience in terms of reduction of number of variables, because of the character of the electromagnetic tensor derivatives. Also, phases of the normal

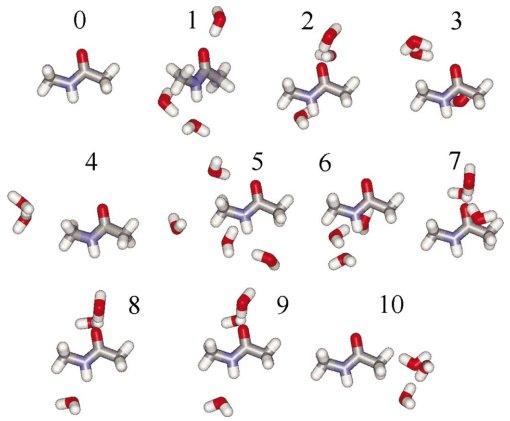

FIG. 1. (Color) Ten clusters (1–10) of NMA (0) and three water molecules used for the fit.

modes (signs of the **s**-vectors) are arbitrary, which would make the fitting ill-defined. Thus Cartesian components of the dipole derivatives (atomic polar tensor, $P_{\mu\alpha,\beta}$) and polarizability derivatives ($\alpha_{\mu\alpha,\beta,\gamma}$) are fitted directly as expressed in a local Cartesian coordinate system attached to the CON amide group atoms

$$P_{\mu\alpha,\beta} = P^{(0)}{}_{\mu\alpha,\beta} + \sum_{j=1,N} c_{j,\mu\alpha,\beta}\varphi_j, \tag{8}$$

$$\alpha_{\mu\alpha,\beta\gamma} = \alpha^{(0)}{}_{\mu\alpha,\beta\gamma} + \sum_{j=1,N} d_{j,\mu\alpha,\beta\gamma}\varphi_j, \tag{9}$$

where the α-index belongs to atom μ, and β,γ denote the electric dipole components. The fitting coefficients $\{c_{j,\mu\alpha,\beta}, d_{j,\mu\alpha,\beta\gamma}\}$ were obtained using an analogous procedure as for the force constants. Note, that the same matrix **A** in Eq. (6) can be used for the force field as well as for the tensor components.

Computations

For the *ab initio* calculations, in addition to the common hybrid Becke3LYP[45] functional, the BPW91 general gradient approximation was used,[46,47] since the latter has been extensively used for simulation of peptide vibrational spectra in the past.[22] GAUSSIAN programs[48] were used for computation of *ab initio* energies, gradients and vibrational frequencies and intensities, with standard basis sets (6-31G, 6-31G**, and 6-311++G**) as specified below. Homemade QGRAD program[49] was applied for the normal mode optimization routine. The molecular-dynamics simulations were run with the TINKER program package.[54] Peak fits in experimental and simulated spectra were done with the SPECTRACALC[50] software. Particularly, the center transition frequencies, absorption band areas, and band widths (according to common convention given as full peak widths at their half heights, FWHH) could be obtained with the latter program by a curve fitting into the mixed Gaussian–Lorentzian bands.

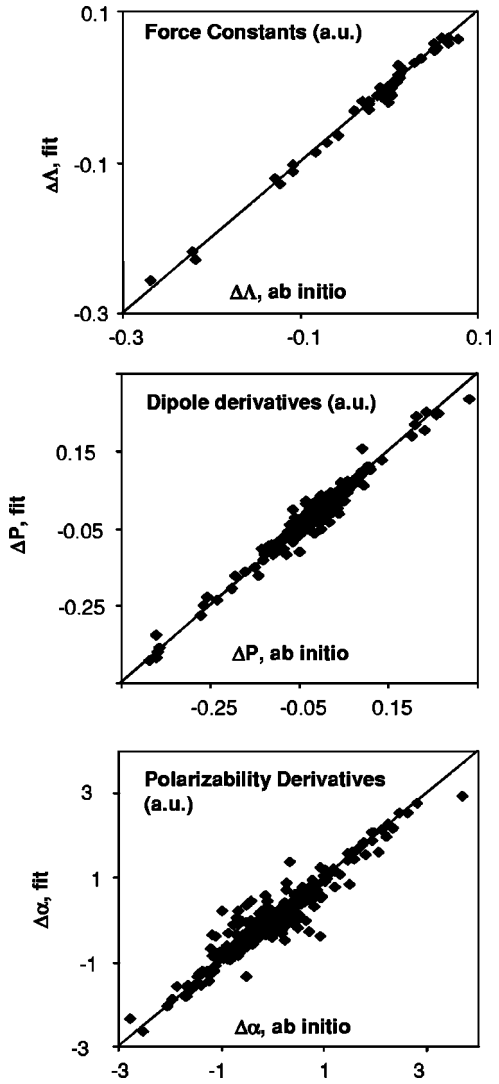

FIG. 2. Comparison of fitted and *ab initio* values of solvent shifts of diagonal force constants (up), atomic polar tensor (middle), and polarizability (bottom) components in local coordinate system. The tensors comprise six amide group atoms and all the clusters displayed in Fig. 1.

III. RESULTS AND DISCUSSION

A. Vibrations of the amide group

Supposedly, the properties of the four atoms directly involved in the partial bond rearrangement

$$O = C - N - H \leftrightarrow O^- - C = N^+ - H,$$

are the most influenced by the solvent, because the polar environment stabilizes the charged form. Thus if $M=4$ in Eq. (1), most of the solvent shifts of the higher-frequency vibrational modes of interest should be comprised in the model. For an isolated OCNH group, six of the $3\times4=12$ eigenvalues Λ_I would correspond to translations and rotations of zero frequencies. If built in a peptide, two of the translational modes roughly match $^{\alpha}C-C(O)$ and $N-{}^{\alpha}C$ stretchings of medium frequencies, and the rest corresponds to lower-frequency bending and torsional modes. Additionally, because of the local planar (C_s) symmetry, we may expect one lower-frequency out of plane and five in-plane intrinsic HNCO vibrations. The latter five higher frequency modes

TABLE I. Calculated force field fitting parameters [b_{Ij} in Eq. (3)]. For the 6-atom fit ($M=6$), with force field (**F**) in atomic units, atom masses (m) in g/mol, potential of solvent atoms (μ) at a site j, $\varphi_j=\Sigma_\mu(Z_\mu/r_{\mu j})$, is used with Z_μ in atomic units and the distances $r_{\mu j}$ in Å.

Amide mode (I)	j:H(N)	N	C(O)	O	$^{\alpha}$C(N)	$^{\alpha}$C(CO)
	BPW91/6-31G**					
A	8.4	−8.6	−16.8	5.1	5.6	6.3
I	−0.1	2.5	−5.0	0.4	−0.4	2.5
II	0.2	−1.3	−0.5	0.7	−0.2	1.1
III	−0.2	−0.3	0.2	0.4	−0.5	0.4
	BPW91/6-311++G**					
A	8.7	−14.4	0.7	0.8	4.4	-0.2
I	0.0	1.6	−1.9	−0.4	−0.4	1.1
II	−0.6	2.8	−8.9	2.6	−0.6	4.0
III	−0.5	1.5	−3.5	1.3	−0.5	1.7
	Becke3LYP/6-311++G**					
A	6.4	−7.7	−0.7	3.4	3.2	5.4
I	0.1	0.9	−0.3	−0.9	−0.2	0.4
II	−0.9	3.8	−9.5	2.9	−0.7	4.5
III	−0.6	1.8	−4.1	1.5	−0.5	2.0

involve the analytically important amide *A* (N–H stretch), amide *I* (C═O stretch), amide *II* and *III* (N–H bend and C–N stretch) vibrations. The fifths in-plane mode can be imagined as a N–C═O bending, but cannot be easily assigned in the spectrum due to coupling with other motions. Thus, as far as we are interested in the high- and mid-IR spectral region ($\sim>1000$ cm^{-1}), only 4–7 highest diagonal elements could be left in the fit [Eq. (3)]. However, inclusion of the full number (12) was used since it does not bring any further significant computer cost. As shown below, in order to account for the coupling of the amide group vibrations with the $^{\alpha}$C–C(O) and N–$^{\alpha}$C stretching modes, it may be desirable to involve also the $^{\alpha}$C carbons in the fit and use a local force field matrix with $M=6$ and $3\times6=18$ eigenvalues.

B. Training cluster set

The constants needed for the empirical corrections of the amide group force field and tensor derivatives were calibrated on a set of N-methylacetamide molecule and ten [$m=11$ in Eq. (5)] clusters with three water molecules shown in Fig. 1. It has been shown previously that a relatively small number of water molecules in the first hydration shell (solvent molecules closest to the solute chromophore) causes most of the vibrational spectral changes and relatively small number of clusters is needed in order to reproduce an average solvent influence.[51–53] This can be also partially justified by the fact that water molecules in the vicinity of the polar N–H and C═O groups are strongly oriented due to the formation of the hydrogen bonds.[43] In an extreme case, two geometries would be sufficient for providing the slope and intercept for the linear model. Nevertheless, in order to obtain statistically significant dependence, positions of the water molecules in the training clusters were arbitrarily varied in order to produce a wide range of solvent electrostatic fields, from the vacuum case, over the situation where the waters were left only around the hydrophobic (CH_3) parts of the molecule (clusters 4 and 10 in Fig. 1) up to the "full hydrated" case (clusters 8 and 9) where all the three possible hydrogen bonds on the N–H and C═O groups were saturated. All the geometries were generated from configurations obtained by molecular-dynamics (MD) simulation with the TINKER program package.[54] The solvent electrostatic field was generated with the same partial charges (−0.84 for water oxygen and 0.42 for hydrogen) as for the MD simulation.

The normal mode optimization was applied so that the higher-frequency modes ($\nu>300$ cm^{-1}) could be relaxed in the clusters. The quality of the fit for the diagonal force constants, dipole and polarizability derivatives (linear dependencies 3, 8, and 9, $M=6$) can be seen in Fig. 2. The maximal variation of the force constants under the solvent influence is about 15%, while the deviation from the linear dependence does not usually exceed 1%. The dispersion of the electromagnetic tensor derivatives deviates more from the linearity; however, we find it satisfactory for the purpose of the empirical model.

In Table I, the calculated force field fitting parameters [b_{Ij} in Eq. (3)] are listed for the four highest eigenvalues Λ_I, which approximately correspond to the amide A, I, II, and III vibrations. For the amide *I* (C═O stretching) vibrations we can compare them qualitatively to those in Refs. 7,8, although in the previous works field dependence of the stretching frequency and not the force constants was fit. However, we obtained the same sign pattern except for the C(O) atom as in Refs. 7 and same pattern as in Ref. 8. However, it may be difficult to assign a simple physical meaning to these parameters and they should always be considered together with the selected mathematical model. For example, the values seem to be spread relatively evenly over the six selected atoms, while the electrostatic field in the vicinity of the C═O bond only should be most important for this mode. Indeed, the number of atoms was previously[8] found rather artificial and is perhaps somewhat redundant for this localized mode. Similarly for the amide *A* (N–H stretch), we see

TABLE II. Frequencies of absorption maxima (in cm^{-1}) for amide modes of N-methylacetamide.

	Calc/Vacuum				Calc/Water[a]				
Normal mode	BPW91 6-31G	BPW91 6-31G**	BPW91 6-311++G**	Becke3LYP 6-311++G**	BPW91 6-31G	BPW91 6-31G**	BPW91 6-311++G**	Becke3LYP 6-311++G**	Exp/Water
ν(N–H), amide *A*	3563	3573	3561	3647	3434	3468	3460	3584	(3300)[b]
ν(C=O), amide *I*	1673	1741	1699	1747	1613	1678	1644	1681	1630
Amide *II*	1534	1507	1497	1545	1535	1512	1520	1563	1576
Amide *III*	1279	1242	1236	1277	1325	1288	1259	1301	1314

[a]6-parameter fit, 68- water molecule periodic cage.
[b]Cannot be reliably estimated because of the overlap with the water absorption band, the peak center at 3325 cm^{-1} was obtained by Raman scattering measurement (backscattering, spectral resolution 5.5 cm^{-1}, volume ratio NMA:H_2O=0.05).

that biggest parameters belong to the H and N atoms (as expected), but also to the C(O) atom for the BPW91/6-31G** calculation. In general, similar patterns were obtained with the three *ab initio* methods listed in Table I (and also at the BPW91/6-31G level which is not shown), with the calculations differing in basis set (6-31G** vs 6-311++G**) more variant then when only the functionals (BPW91 vs Becke3LYP) were different; this dependence on the *ab initio* level, however, is significantly bigger than for physically observable variables as frequencies and intensities given in Tables II and III below. Bigger absolute values of the fitting coefficients b_{Ij} for the higher-frequency modes result from the fitting of the absolute force field (diagonal) elements and do not indicate that these modes would be more sensitive to the solvent. From this approach [Eqs. (3), (8), and (9)] one can imagine fitting of relative values; we see, however, no advantage of the latter within the present first-order scheme. We restrain ourselves from the analysis of the fitting coefficients for the dipole and polarizability derivative components [Eqs. (8) and (9)], because of their large number and even fuzzier physical meaning.

Perhaps more interesting is the accuracy with which the amide *A*, *I*, *II*, and *III* vibrational frequencies of NMA can be reproduced by the empirical correction. This is illustrated in Fig. 3, where the model values are compared to the *ab initio* results for all the clusters described above. Apparently, the model reproduces trends in the solvent shifts for all the clusters, all the amide vibrational modes and thus within the entire frequency range. Occasional deviations of up to $\sim$10 cm^{-1} from the *ab initio* values occur, a price which we find acceptable for the simplicity of the simple model.

C. Vibrational spectra of hydrated N-methylacetamide

In order to simulate the inhomogeneous band widths of all the amide bands, the NMA molecule was placed in a water box of 68 H_2O molecules and 7000 MD configurations (see Fig. 4) were generated by an MD run. The simulation was performed with the Amber force field of TINKER,[54] periodic boundary conditions, and NpT ensemble using atmospheric pressure (1 atm) and room temperature (298 K). After 10 ps of equilibration the geometries were recorded separated by 1000 MD steps (of 1 fs) so that they can be considered independent. For each configuration the perturbation to the NMA force field and tensor derivatives were obtained based on the water electrostatic potential, and harmonic vibrational frequencies and intensities calculated. Note, that only one (vacuum) *ab initio* computation of the NMA molecule was needed.

The resultant vacuum and solvent absorption and Raman intensities calculated at the Becke3LYP/6-311++G** level are compared in Fig. 5. Lorentzian bands 7 cm^{-1} wide (FWHH) were used for the simulation of the spectra. This bandwidth allows to smooth the discrete distribution, but is small enough not to change the inhomogeneous broadening. In accordance with the experiment[55] the amide *A* band is broadened most by the solvent interaction, with the calculated bandwidth (FWHH) of about 150 cm^{-1}. This contrasts with the C–H stretching signal, which is not dispersed. Although the hydrogen constants were not taken explicitly in the fit, we do not expect any significant effect of solvent on these hydrophobic groups. The lower-frequency region (below 2000 cm^{-1}), however, comprises again amide group

TABLE III. Absorption band widths (cm^{-1}) [full widths at half heights (FWHH) of absorption peaks]/dipolar strengths (debye^2) for N-methylacetamide. The dipolar strengths (*D*) were obtained by fitting of the spectral intensities (ε) by mixed Gaussian–Lorentzian peaks and integration over the frequencies (ν), using the formula $D(\text{debye}^2)=0.009\,184\times\int\varepsilon(\text{Lmol}^{-1}\,\text{cm}^{-1})d\nu/\nu$.

	Calc(Vacuum)				Calc(Water)[a]				
Normal mode	BPW91 6-31G	BPW91 6-31G**	BPW91 6-311++G**	Becke3LYP 6-311++G**	BPW91 6-31G	BPW91 6-31G**	BPW91 6-311++G**	Becke3LYP 6-311++G**	Exp/Water
ν(N–H), amide *A*	0/0.001	0/0.002	0/0.002	0/0.003	179/0.035	175/0.025	157/0.035	125/0.031	(>60/0.01)[b]
ν(C=O), amide *I*	0/0.039	0/0.047	0/0.065	0/0.070	200/0.101	89/0.098	53/0.111	43/0.117	45/0.065
Amide *II*	0/0.028	0/0.038	0/0.041	0/0.049	70/0.017	42/0.072	49/0.043	52/0.048	46/0.050
Amide *III*	0/0.025	0/0.023	0/0.025	0/0.028	155/0.063	45/0.024	57/0.028	55/0.029	32/0.011

[a]6-parameter fit, 68- water molecule periodic cage.
[b]Cannot be reliably estimated because of overlap with the water absorption band, width of 99 cm^{-1} was determined by the Raman scattering measurement specified in Table II.

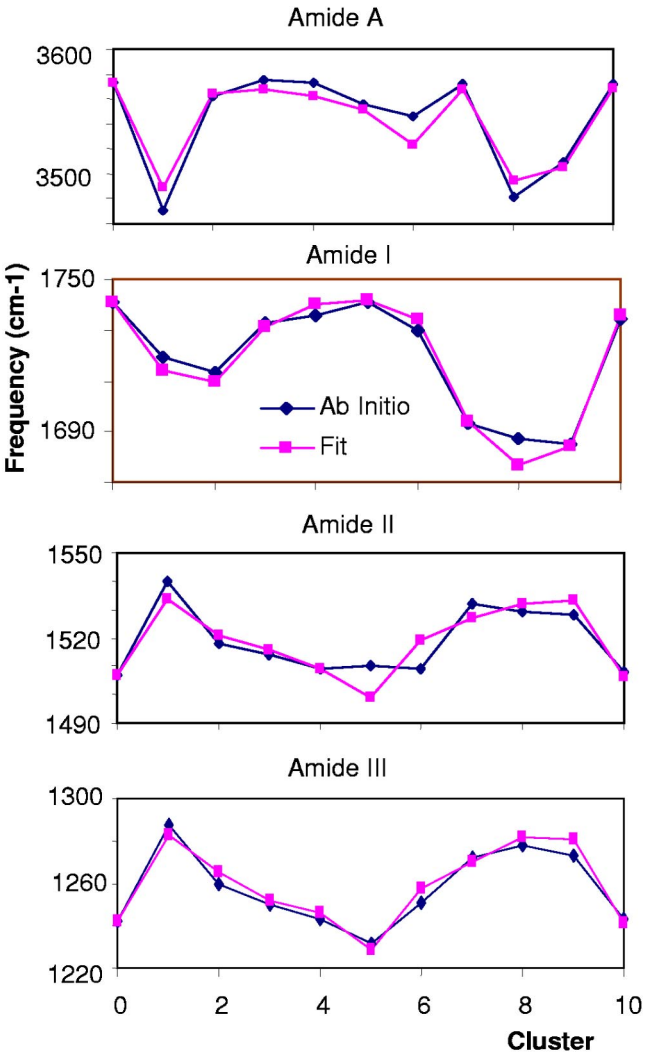

FIG. 3. Comparison of the amide *A* (N–H stretch), *I* (C═O stretch), *II* and *III* (N–C stretch and N–H bend) frequencies for the clusters from Fig. 1 as obtained *ab initio* and using the empirical correction to the vacuum force field [Eq. (7)].

modes most effected by the solvent. Even the frequencies of the C–H bending modes ($\sim$1550–1400 cm^{-1}) are indirectly dispersed via coupling with other modes. The overall frequency shifts are summarized in Table II. Amide *A* modes shifts down by $\sim$100 cm^{-1} and the calculated value is rather independent on the basis set size. The remaining difference to the experiment (since the model was calibrated against *ab initio* computations) can probably be accounted for by anharmonic potential of the N–H bond.[56] Also the amide *I* frequency shifts significantly (by $\sim$70 cm^{-1}); the shift, however, is more dependent on the basis set and functional chosen. For example, the BPW91 frequency (1644 cm^{-1}) is closer to experiment (1630 cm^{-1}) than the Becke3LYP value (1681) for the 6-311++G** basis set. In accord with the simple conjugation model where the solvent strengthens the C–N bond, calculated amide *II* and *III* frequencies shift up. Similarly as for the amide *I* mode, the shift (5–40 cm^{-1}) is quite dependent on the basis set size and the functional, with the Becke3LYP/6-311++G** level providing the best agreement with the experiment (calculated frequencies 1563

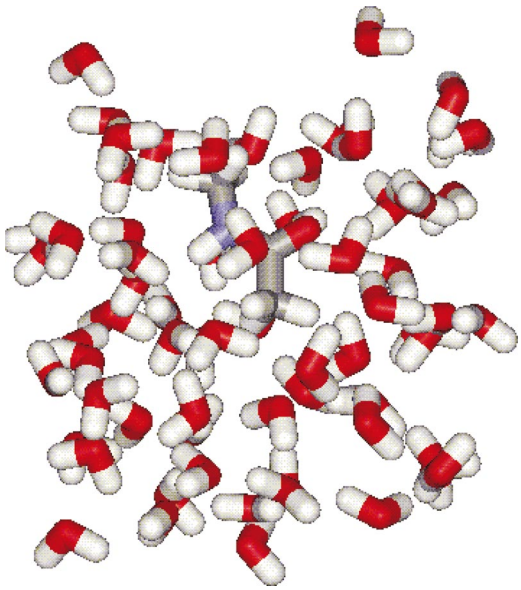

FIG. 4. (Color) Periodic box with NMA and 68 water molecules used for the MD simulation.

and 1301 cm^{-1}, experimentally 1576 and 1314 cm^{-1}, respectively).

In Table III absorption bandwidths and dipolar strengths are compared to experimental values, as obtained by least-square fitting of the spectra by mixed Gaussian–Lorentzian bands. As for the solvent frequency shifts, the amide *I–III* band widths are most affected by the *ab initio* approximation. Particularly the values obtained with the smallest basis set (6-31G) are quite unrealistic. Surprisingly, bigger basis sets lead to narrowing of the amide *A* and *I* vibrational bands (by 50% for amide *I*!) and these motions become less susceptible to external perturbations, while the widths slightly increase in the bigger basis for amide *II* and *III*. Little difference can be observed between the performance of the BPW91 and Becke3LYP methods, the latter yielding a slightly better agreement with experimental bandwidths (calculated 43, 52, and 55 cm^{-1}, experimentally 45, 46, and

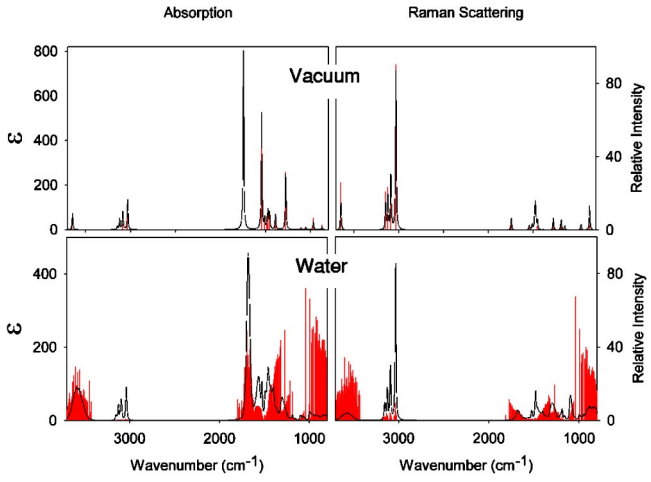

FIG. 5. Calculated IR absorption (left) and Raman (right) spectra of the NMA molecule in vacuum (top) and in water (bottom), at the Becke3LYP/6-311++G** level. For the latter, 7000 MD configurations were averaged for the box displayed in Fig. 4.

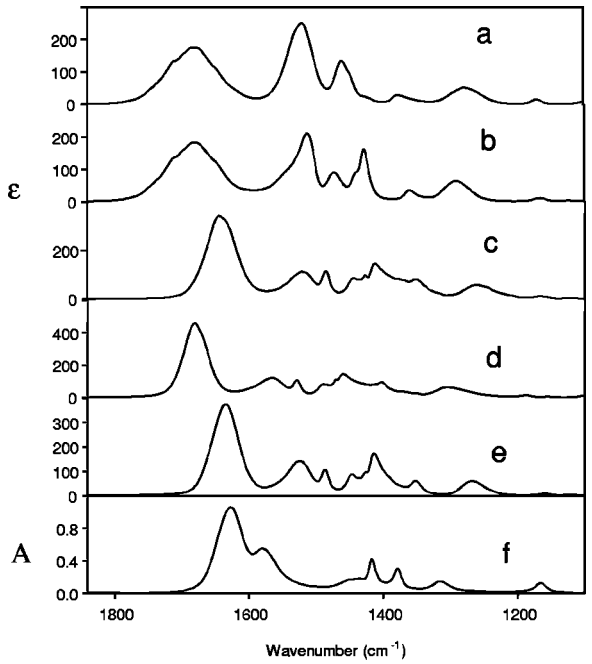

FIG. 6. Effect of the basis set and box size on the spectra (a) 4-atom ($M=4$) fit, (b)–(e) 6-atom ($M=6$) fit, (a)–(d) small box (13.4 Å), (e) big box (20 Å), (a) and (b) 6-31G** basis set, (c)–(e) 6-311++G** basis set, (a)–(c) and (e) BPW91 functional, (d) Becke3LYP functional, (f) the experimental spectrum [2.74 M solution of NMA in water, FTIR Bruker IFS-66/S spectrometer, 10 μm CaF_2 cell (Biotools), 298 K, baseline subtracted].

32 cm^{-1}, for the amide *I*, *II*, and *II* modes, respectively). The dipolar strengths are also significantly dependent of the basis set, namely for the amide *II* mode, while the two functionals provide similar values. Computed values agree reasonably well with the experiment, given the expected error originating in the water base line subtraction. Calculated bandwidth and dipole strength for the amide *III* mode exhibit biggest deviation from the experiment. This can be explained by coupling of this mode with other lower-frequency motions, such as methyl-wagging, which could not be included in the model.

In Fig. 6, the dependence of the simulated spectra on the number of atoms included (M), box width, functional and basis set size on simulated absorption spectra of NMA can be seen for the mid-IR region. If a force field of 6 instead of 4 atoms is included in the fit [Fig. 6(a) vs Fig. 6(b)], namely the middle part of the spectra (1400–1500 cm^{-1}) comprising methyl-group vibration is influenced; nevertheless frequency of the amide *III* bands also somewhat increases, because of the inter-mode couplings. As expected, the amide *I* mode is not influenced by this change, since it is localized primarily at the two C=O atoms. As discussed above for Tables II and III, the bigger basis set [Fig. 6(b) vs Fig. 6(c)] is necessary in order to obtain more realistic amide *I* frequency and bandwidth. Namely the relative intensity ratio of the amide *II* and *I* bands is affected by the basis set increase, in favor to the experiment. The application of the Becke3LYP functional [Fig. 6(d) vs Fig. 6(c)] leads to improvement of the amide *II* and *III* frequencies, while the amide *I* band shifts away from the experimental value, probably because of internal anharmonicity of the C=O vibration[57] not included in the model. Finally, the box size [Fig. 6(c) vs Fig. 6(e)] influences details of the absorption profile, nevertheless most spectral characteristics are obtained already with the smaller box and shifts of peak maxima do not exceed 10 cm^{-1}, which confirms the dominant role of the first hydration sphere. We have also investigated the effect of various water–peptide force force fields (not shown) used in the MD simulation, but did not find any significant influence. Overall, we may conclude that the model provides realistic frequency, intensity and bandwidth changes, although proper *ab initio* level must be found similarly as for the vacuum computation. With regards to the modeling of vibrational spectra of peptides, the sensitivity of the amide *I* and *II* bands to anharmonic interactions and the *ab initio* level appears somewhat unpleasant, because this region is most accessible and usually provides the strongest experimental signal.[19,20]

D. Hydrated peptide α-helix

In order to estimate the performance of the solvent correction to the amide group vibration for a more realistic systems, we simulate vibrational circular dichroism (VCD) and absorption spectra of hydrated 19-amide L-alanine-based peptide, $CH_3-CO-[NHCH(CH_3)CO]_{17}-NH-CH_3$. Vacuum dipole derivatives and harmonic molecular force field of this 19-mer (with respect to the number of amide groups) were calculated by transferring Cartesian molecular tensors[22,24] obtained for a 7-amide fragment at the BPW91/6-31G** level. Standard helical geometry[58] was chosen ($\varphi=-57$, $\psi=-47°$) and the normal mode optimization method[49] used in order to relax the higher-frequency motions in the 7-amide. Then the 19-mer was placed in a water box ($45\times20\times20$ Å, with 192 peptide and 1587 water atoms) displayed in Fig. 7. Under periodic boundary conditions molecular dynamics was run with the TINKER software[54] for an NpT ensemble ($p=1$ atm, $T=300$ K), using the Amber 94 force field, similarly as for the NMA system. Only water molecules were allowed to move, while positions of peptide atoms were fixed. In order to get a representative distribution of the solvent, after an equilibration stage (1000 steps of 1 fs) 900 configurations separated by 1000 MD steps (1 ns) were generated. For each configuration solvent corrections to the vacuum *ab initio* force field and dipole derivatives of the 19-mer were applied, based on the immediate electrostatic field of the solvent and the fitting parameters obtained from the NMA–water system. Thus the 900 corrected force fields could be diagonalized and resultant normal mode frequencies and absorption intensities obtained for each mode. Note, that in this way the amide–amide interactions obtained for vacuum are not perturbed by the solvent, which disperses only the localized "diagonal" components of the 19-mer force field.

Calculated vacuum and solvent-corrected absorption and VCD spectra of the 19-amide can be seen in Fig. 8. Similarly as shown in previous studies[22,23,59] the vacuum approximation provides qualitatively correct relative intensities and signs of the VCD bands. Particularly, the (+/−) sign pattern of the big VCD couplet belonging to the amide *I* vibration calculated around 1740 cm^{-1}, and the overall negative inten-

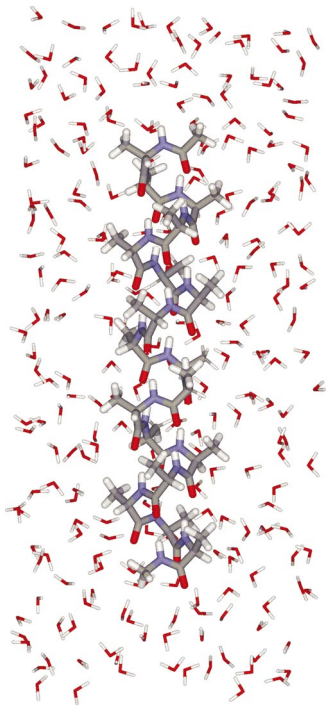

FIG. 7. (Color) The model α-helical alanine-based 19-amide in a box (45 $\times 20 \times 20$ Å) of water. Solvent molecules in front of the peptide are not visible in the figure.

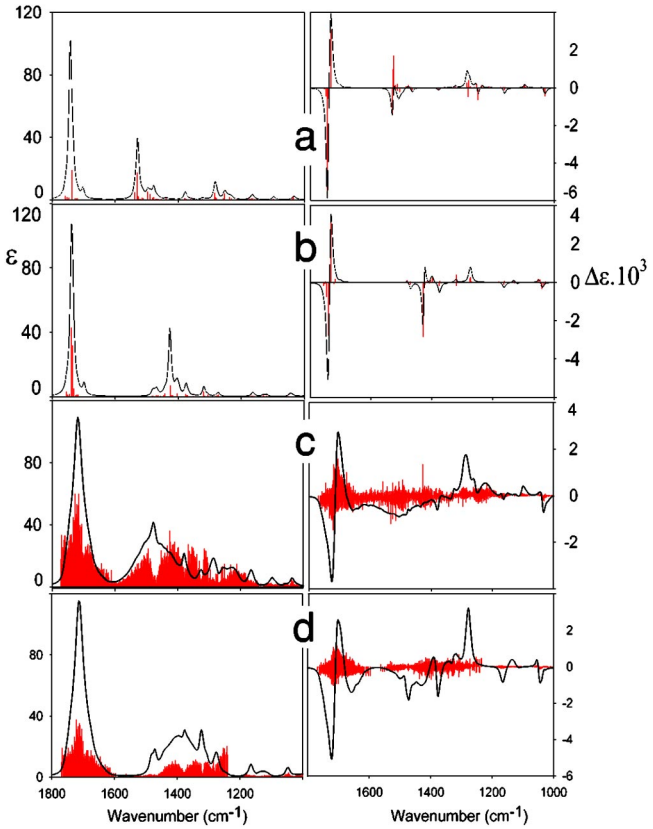

FIG. 8. Simulated absorption (left) and VCD (right) spectra of the α-helical 19-amide, (a) and (b) in vacuum and (c) and (d) with the MM/QM correction. Spectra for (a) and (c) natural and (b) and (d) N-deuterated peptide.

sity of the amide *II* motion with a minimum at 1534 cm^{-1} are in agreement with experimental spectra of α-helical peptides and proteins, where the sharp amide *I* couplet occurs at 1615–1620 cm^{-1} and the amide *II* signal appears as a broad negative band centered around 1550 cm^{-1}.[18,60–62] The solvent-corrected VCD and absorption spectra reproduce correctly this difference in the dirspersion of the amide *I* and *II* frequencies. Unlike for the N-methylacetamide, the amide *II* mode can couple with other amide groups and skeletal α-helix vibrations, which further increases the frequency dispersion. Experimentally, measurement in D_2O solutions is preferred, because it allows to better subtract the environmental baseline for the most intensive amide *I* signal.[18] While the deuteration (N–D) for the vacuum does not cause any qualitative change in spectral shapes, except of a minor downshift of the amide *I* band (by 6 cm^{-1}) and a more significant redshift of the amide *II* signal (from 1534 to 1430 cm^{-1}), it produces a new strong negative VCD band at 1657 cm^{-1}. In experiment, the negative lobe occurs at $\sim$1622 cm^{-1}, and although such change of amide *I* couplet to a W-shape has been described many times for deuterated α-helical systems.[18,22,60] It could never be explained by studies based on vacuum modeling.

Obviously, we realize the limitations of the QM/MM approach, for example, the neglect of the polarization effects proven indispensable for explanation of many solvent properties.[63] Also the distribution of the solvent molecules and electron-charge effects would be described better by more advanced techniques, as the quantum or Car–Parrinello molecular dynamics.[64] These, however, are currently impossible for big peptides, while the presented approximation can be implemented quite easily and in fact provides a better description of the solvent effect than computationally more expensive methods, particularly the continuum solvent models.[43] Thus we find it viable for reproducing the main features of the vibrational solvent effects on amide group vibrations in peptides.

IV. CONCLUSION

The empirical correction realistically models solvent effects in vibrational spectra of the amide group, in terms of intensity changes, frequency shifts and inhomogeneous vibrational band broadenings. For N-methylacetamide, the changes caused by the solvent were found to be significantly dependent on the approximation level and basis set size used, similarly as for vacuum computations. When applied to the α-helical peptide, the model significantly improved the simulated spectra a explained many features observed experimentally, including the isotopic deuterium-exchange effects.

ACKNOWLEDGMENT

This work was supported by the Grant Agency of the Academy of Sciences (A4055104).

[1] J. Noell and K. Morokuma, Chem. Phys. Lett. **36**, 465 (1975).
[2] U. C. Singh and P. A. Kollman, J. Comput. Chem. **7**, 718 (1986).
[3] M. A. Thomson, J. Phys. Chem. **100**, 14492 (1996).
[4] J. Florián and A. Warshel, J. Phys. Chem. B **101** 5583 (1997).
[5] P. Bouř, Chem. Phys. Lett. **365**, 82 (2002).
[6] S. A. Corcelli, C. P. Lawrence, and J. L. Skinner, J. Chem. Phys. **120**, 8107 (2004).

[7]K. Kwac and M. Cho, J. Chem. Phys. **119**, 2247 (2003).
[8]P. Bouř and T. A. Keiderling, J. Chem. Phys. **119**, 11253 (2003).
[9]S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, J. Chem. Phys. **120**, 10896 (2004).
[10]C. H. Langley and N. L. Allinger, J. Phys. Chem. A **107**, 5208 (2003).
[11]J. Kongsted, A. Osted, K. Mikkelsen, and O. Christiansen, J. Phys. Chem. A **107**, 2578 (2003).
[12]S. Y. Ham, S. Hahn, C. Lee, T. K. Kim, K. Kwak, and M. Cho, J. Phys. Chem. B **108**, 9333 (2004).
[13]J. H. Choi, S. Y. Ham, and M. Cho, J. Phys. Chem. B **107**, 9132 (2003).
[14]H. Huang, S. Malkov, M. Coleman, and P. Painter, J. Phys. Chem. A **107**, 7697 (2003).
[15]S. Krimm and J. Bandekar, Adv. Protein Chem. **38**, 181 (1986).
[16]V. Andrushchenko, H. Wieser, and P. Bouř, J. Phys. Chem. B **106**, 12623 (2002).
[17]V. Andrushchenko, H. Wieser, and P. Bouř, J. Phys. Chem. B **108**, 3899 (2004).
[18]T. A. Kedierling, in *Circular Dichroism: Principles and Applications*, edited by N. Berova, K. Nakanishi, and R. W. Woody (Wiley, New York, 2000), pp. 621–666.
[19]*Spectroscopic Methods for Determining Protein Structure in Solution*, edited by H. Havel (VCH, New York, 1995).
[20]P. L. Polavarapu, Vib. Spectrosc. **17B**, 319 (1989).
[21]T. A. Keiderling, R. A. G. D. Silva, G. Yoder, and R. K. Dukor, Bioorg Med. Chem. **7**, 133 (1999).
[22]R. A. G. D. Silva, J. Kubelka, P. Bouř, S. M. Decatur, and T. A. Keiderling, Proc. Natl. Acad. Sci. U.S.A. **97**, 8318 (2000).
[23]P. Bouř and T. A Keiderling, J. Am. Chem. Soc. **15**, 9602 (1993).
[24]P. Bouř, J. Sopková, L. Bednárová, P. Maloň, and T. A. Keiderling, J. Comput. Chem. **18**, 646 (1997).
[25]P. Bouř and T. A. Keiderling, J. Mol. Struct.: THEOCHEM **675**, 95 (2004).
[26]J. Kubelka and T. A. Keiderling, J. Am. Chem. Soc. **123**, 12048 (2001).
[27]J. Hilario, J. Kubelka, F. A. Syud, S. H. Gellman, and T. A. Keiderling, Biopolymers **67**, 233 (2002).
[28]A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, Proc. Natl. Acad. Sci. U.S.A. **98**, 5578 (2001).
[29]S. H. Gellman, Curr. Opin. Struct. Biol. **2**, 717 (1998).
[30]L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).
[31]M. W. Wong, M. J. Frisch, and K. B. Wiberg, J. Am. Chem. Soc. **113**, 4776 (1991).
[32]V. Barone, M. Cossi, and J. Tomasi, J. Comput. Chem. **19**, 404 (1998).
[33]F. Eckert and A. Klamt, AIChE J. **48**, 369 (2002).
[34]H. Sato and S. Sakaki, J. Phys. Chem. A **108**, 1629 (2004).
[35]S. Tenno, F. Hirata, and S. Kato, J. Chem. Phys. **100**, 7443 (1994).
[36]J. R. Schmidt, S. A. Corcelli, and J. L. Skinner, J. Chem. Phys. **121**, 8887 (2004).
[37]J. Choi and M. Cho, J. Chem. Phys. **120**, 4383 (2004).
[38]T. Hayashi, T. la C. Jansen, W. Zhuang, and S. Mukamel, J. Phys. Chem. A **109**, 64 (2005).
[39]H. Torii, J. Phys. Chem. A **108**, 7272 (2004).
[40]S. Ham, J. H. Kim, H. Kochan, and M. Cho, J. Chem. Phys. **118**, 3491 (2003).
[41]M. Cho, J. Chem. Phys. **118**, 3480 (2003).
[42]N. A. Besley, J. Phys. Chem. A **108**, 10794 (2004).
[43]P. Bouř, J. Chem. Phys. **121**, 7545 (2004).
[44]D. Papoušek and M. R. Aliev, *Molecular Vibrational Rotational Spectra* (Academia, Prague, 1982).
[45]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).
[46]A. D. Becke, in *Modern Electronic Structure Theory*, edited by D. R. Yarkony (World Scientific, Singapore, 1995), Vol. 2, pp. 1022–1046.
[47]A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[48]M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria *et al.*, GAUSSIAN 03, Revision A.1, Gaussian, Inc., Pittsburgh PA, 2003.
[49]P. Bouř and T. A. Keiderling, J. Chem. Phys. **117**, 4126 (2002).
[50]SPECTRACALC (Galactic Industries Corp., 1990).
[51]H. Torii, T. Tatsumi, and M. Tasumi, J. Raman Spectrosc. **29**, 537 (1998).
[52]H. Guo and M. Karplus, J. Phys. Chem. **96**, 7273 (1992).
[53]J. Kubelka and T. A. Kedierling, J. Am. Chem. Soc. **105**, 10922 (2001).
[54]R. V. Pappu, R. K. Hart, and J. W. Ponder, J. Phys. Chem. B **102**, 9725 (1998).
[55]R. Zhang, H. Li, Y. Lei, and S. Han, J. Mol. Struct. **693**, 17 (2004).
[56]S. K. Gregurick, G. M. Chaban, and R. B. Gerber, J. Phys. Chem. **106**, 8696 (2002).
[57]M. T. Zanni, M. C. Asplund, and R. M. Hochstrasser, J. Chem. Phys. **114**, 4579 (2001).
[58]G. D. Fasman, *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum, New York, 1989).
[59]P. Bouř, J. Kubelka, and T. A. Keiderling, Biopolymers **65**, 45 (2002).
[60]J. Kubelka, R. A. G. D. Silva, P. Bouř, S. M. Decatur, and T. A. Keiderling, *Chirality: Physical Chemistry*, ACS Symp. Ser. 810, edited by J. M. Hicks (American Chemical Society, Washington DC, 2002), Vol. 810, pp. 50–64.
[61]P. Maloň, R. Kobrinskaya, and T. A. Keiderling, Biopolymers **23**, 733 (1988).
[62]V. Baumruk and T. A. Keiderling, J. Am. Chem. Soc. **115**, 6939 (1993).
[63]H. Yu, T. Hansson, and W. F. van Gunsteren, J. Chem. Phys. **118**, 221 (2003).
[64]D. Marx and J. Hutter, *Modern Methods and Algorithms of Quantum Chemistry* (Forschungszentrum Juelich, NIC Series, 2000), Vol. 1, pp. 301–449.