

Molecular Dynamics with Restrictions Derived from Optical Spectra

JAKUB KAMINSKÝ, JIŘÍ ŠEBEK, PETR BOUŘ

*Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo nám. 2,
16610, Prague 6, Czech Republic*

Received 13 June 2008; Revised 21 July 2008; Accepted 12 August 2008

DOI 10.1002/jcc.21123

Published online 24 September 2008 in Wiley InterScience (www.interscience.wiley.com).

Abstract: The information about molecular structure coded in the optical spectra must often be deciphered by complicated computational procedures. A combination of spectral modeling with the molecular dynamic simulations makes the process simpler, by implicit accounting for the inhomogeneous band broadening and Boltzmann averaging of many conformations. Ideally, geometries of studied systems can be deduced by a direct confrontation of such modeling with the experiment. In this work, the comparison is enhanced by restrictions to molecular dynamics propagations based on the Raman and Raman optical activity spectra. The methodology is introduced and tested on model systems comprising idealized H_2O_2 , H_2O_3 molecules, and the alanine zwitterion. An additional gradient term based on the spectral overlap smoothed by Fourier transformation is constructed and added to the molecular energy during the molecular dynamics run. For systems with one prevalent conformation the method did allow to enrich the Boltzmann ensemble by a spectroscopically favored structure. For systems with multiconformational equilibria families preferential conformations can be selected. An alternative algorithm based on the comparison of the averaged spectra with the reference enabling iterative updates of the conformer probabilities provided even more distinct distributions in shorter times. It also accounts for multiconformer equilibria and provided realistic spectra and conformer distribution for the alanine.

© 2008 Wiley Periodicals, Inc. J Comput Chem 30: 983–991, 2009

Key words: molecular dynamics; vibrational spectroscopy; Raman optical activity; *ab initio*

Introduction

Optical spectroscopy methods have always provided relatively simple and extremely flexible means for investigation of molecular structure.¹ Because the wavelength of visible or infrared light is much longer than usual molecular sizes, the techniques do not capture every structural detail and provide a lower resolution than the X-ray crystallography.² However, an increased conformational and enantiomer sensitivity can be achieved with some polarized methods, particularly with those using different absorption or scattering of circularly polarized light on optically active (chiral) systems. Thus the optical rotatory dispersion (ORD) or circular dichroism (CD) became standard analytical tools in chemistry and molecular biology.^{3,4} The extension of the optical activity techniques into the vibrational domain opened further possibilities, as the vibrational transitions are numerous and better resolved than the electronic ones. The vibrational circular dichroism (VCD)⁵ and the Raman optical activity (ROA)⁶ gave unique information about structure of many organic compounds, proteins, or nucleic acids.^{7,8} Other important ways to increase effective resolution of the optical spectra include site-specific labeling with stable isotopes⁹ and two-

dimensional multiphoton spectra sampling the chromophore coupling.¹⁰

Simulations by modern quantum-chemical methods represent powerful tools for the spectra interpretations.^{7,11,12} By comparison of the computed and experimental Raman and ROA spectral shapes not only the equilibrium structure but also an important knowledge about molecular flexibility and multiconformational states could be acquired.^{13–15} Even more objective information can be gained from a numerical spectra decomposition,¹⁶ as a complement to the traditional visual comparison of the theoretical spectra with the experiment.

Yet for complicated systems containing multiple conformations, with many overlapping bands or unknown equilibrium structures, the methods available today cannot be applied. A more complete dynamical molecular picture is needed, which

Correspondence to: P. Bouř; e-mail: bour@uochb.cas.cz

Contract/grant sponsor: Grant Agency of the Czech Republic; contract/grant numbers: 203/06/0420, 202/07/0732

Contract/grant sponsor: Grant Agency of the Academy of Sciences; contract/grant number: A400550702

can be, at least in principle, obtained by molecular dynamics (MD) techniques. Therefore, in this work, we explore the link between MD and the spectra, suggesting a possibility to introduce optical spectral restrains into the MD simulations. Similar restrictions are widely used in the NMR spectroscopy,¹⁷ but according to our knowledge they were never attempted in the optical spectroscopy. As shown below, the results document that a stable spectral-restricted MD propagation algorithm can be found, and that a multiscale QM/MM scheme for the spectral interpretation can potentially be used for a better structural refinement of molecules.

Theory

Spectra-Restricted Molecular Dynamics

Let us suppose that we can generate an average or instantaneous spectrum $S_{i,\lambda}$ at each time step i of the MD propagation, λ is the wavelength. This might not be accurate due to the quantum effects and coupling of different excitations. However, for many systems such a quasi-classical approach has been well-justified within the instantaneous mode approximations and other computational schemes.^{18–20} Typically, the vibrational and electronic excitations observable in the spectra occur in a much faster time scale than common conformation motions and the coupling is minimized.

The theoretical ($S_{i,\lambda}$) and experimental (s_λ) spectra are compared with the aid of a difference function, in our implementation chosen as a difference of the absolute values,

$$d(S_i, s) = \sum_{\lambda} |S_{i,\lambda} - s_\lambda|. \quad (1)$$

This term is added to the unperturbed system energy E_0 via an adjustable coupling parameter, α , so that the equation of motion are derived from the target function E ,

$$E = E_0 + \alpha d. \quad (2)$$

As the new potential term $V(\mathbf{X}) = \alpha d(\mathbf{X})$ is dependent on the nuclear coordinates \mathbf{X} , corresponding gradient can be used for the force estimation and coordinate propagation within the MD numerical scheme, such as that of Verlet.²¹

Implementation

The spectra-restricted molecular dynamics (SRMD) propagation was implemented within the Tinker MD suite of programs.²² The Raman and ROA spectra were generated at each time step from a predefined library of conformations using the Cartesian coordinate tensor transfer (CCT) techniques.²³ By default, normalized ($\int |S(\lambda)| d\lambda = 1$) simulated and experimental spectral curves were compared so that d in eq. (2) is dimensionless and α had the units of energy (kcal/mol). As shown below, the spectral restriction is efficient for a reasonable range of α , as far as its magnitude is high enough to be comparable with the numerical noise and the temperature Boltzmann quantum ($kT \sim 0.6$

kcal/mol), but it must be smaller than energies associated with unrealistic molecular bond and angle deformations.

For model H_2O_2 and H_2O_3 molecules sets of 36 and $36 \times 36 = 1296$ conformations were generated by scanning the torsion angles with 10° step within the $0 \dots 360^\circ$ range. Similarly, 50 random conformers of the alanine zwitterion were obtained from the MD simulation described below (Tinker, Amber 99 force field). For each geometry the remaining coordinates were allowed to relax and the Raman and ROA spectral parameters computed *ab initio* by the Gaussian program.²⁴ The parameters included the harmonic force field and nuclear derivatives of electric dipole–electric dipole (α), electric dipole–magnetic dipole (\mathbf{G}'), and electric dipole–electric quadrupole (\mathbf{A}) polarizability derivatives.²⁵

For the MD simulations the periodic boundary conditions were used by default, with the Amber99 force field,²⁶ 1 fs integration time step and an NpT ensemble (pressure $p = 1$ atm, temperature $T = 298$ K). The solute MD structures were compared atom-by-atom with these precalculated geometries so that the best local overlap could be found and the spectra generated from the transferred parameters. After the transfer on the MD structures, theoretical spectra for backscattered intensities were simulated by the usual procedures^{16,27,28} using Lorentzian bands with 20 cm^{-1} width at half maxima, and compared to the reference spectrum according to eq. (1).

The spectral comparison was also implemented within the Tinker program. In eq. (2) and the MD propagation, Fourier-smoothed energy term $\alpha d(\mathbf{X})$ was used, obtained from the dependence of the spectral error on selected internal coordinates $\tilde{d}(\mathbf{q})$ (in our case torsion angles q_j , $j = 1 \dots N$). The Cartesian-internal coordinate transformation can be found, for example, in Supporting Information of ref. 29. The internal error surface was constructed from the Fourier expansion coefficients c_{i_1, i_2, \dots, i_N} as

$$\tilde{d}(\mathbf{q}) = \sum_{i_1=1 \dots M} \sum_{i_2=1 \dots M} \dots \sum_{i_N=1 \dots M} c_{i_1, i_2, \dots, i_N} g_{i_1}(q_1) g_{i_2}(q_2) \dots g_{i_N}(q_N), \quad (3)$$

where $M = 7$ and the basis functions were chosen as $g_1(q) = 1/\sqrt{2\pi}$, $g_2(q) = \cos(kq)/\sqrt{\pi}$, $g_3(q) = \sin(kq)/\sqrt{\pi}$, $g_4(q) = \cos(2kq)/\sqrt{\pi}$, $g_5(q) = \sin(2kq)/\sqrt{\pi}$, $g_6(q) = \cos(3kq)/\sqrt{\pi}$, and $g_7(q) = \sin(3kq)/\sqrt{\pi}$, $k = 2\pi/(q_{\max} - q_{\min})$. Typically, for the torsion angles, $q_{\max} = 180^\circ$ and $q_{\min} = -180^\circ$. The smoothing basis function size was limited, because for large M the number of the coefficients in eq. (3) becomes excessive and cannot be unambiguously determined even by long MD simulations, especially for multidimensional systems ($N > 1$). On the other hand, the current limit to seven functions is the minimum that enable us to accommodate the common threefold torsional potentials, e.g. in molecules containing the sp^3 hybridized carbon atoms.

The coefficients c_{i_1, i_2, \dots, i_N} were obtained by Fourier transformation of the average error $\bar{d}(\mathbf{q})$,

$$c_{i_1, i_2, \dots, i_N} = \int_{q_{1,\min}}^{q_{1,\max}} dq_1 \int_{q_{2,\min}}^{q_{2,\max}} dq_2 \dots \int_{q_{N,\min}}^{q_{N,\max}} dq_N \bar{d}(\mathbf{q}) g_{i_1}(q_1) g_{i_2}(q_2) \dots g_{i_N}(q_N). \quad (4)$$

The error surfaces were kept in memory for an internal coordinate grid (with the $(q_{\max} - q_{\min})/L$ increment, $L = 100$) and were instantaneously updated each point of the MD run as $\bar{d}(q_i)' = (d(q_i) + \bar{d}(q_i)(N_i - 1))/N_i$, where $d(q_i)$, $\bar{d}(q_i)'$ and $\bar{d}(q_i)$ are the respective immediate, averaged new, and old error at the grid point i corresponding to the current coordinate, N_i is the total number of updates in this point.

Direct Probability Generation

In the case of a multiconformational equilibrium, the direct comparison of an immediate MD spectrum with a reference by eq. (1) is problematic, unless one conformation is dominant and provides distinct spectra. However, we can always define an average spectrum as

$$S_{A,\lambda} = \int p(\mathbf{q}) S_\lambda(\mathbf{q}) dq_1 dq_2 \dots dq_N = \sum_j p_j S_\lambda(q_j), \quad (5)$$

where $p(\mathbf{q})dq_1dq_2\dots dq_N$ is the probability of finding the molecule in the coordinate-volume element $dq_1dq_2\dots dq_N$. The volume is in the second part of eq. (5) replaced by discrete probabilities p_j which could be kept in computer memory on a coordinate grid. As a modified agreement criterion, we could choose to minimize the expression

$$d'(\mathbf{p}) = \sum_\lambda \left(\sum_j p_j S_\lambda(q_j) - s_\lambda \right)^2 = \left(\sum_j p_j \mathbf{S}(q_j) - \mathbf{s} \right)^2, \quad (6)$$

which has, in contrary to the sum of absolute values in eq. (1), smooth derivatives. With the condition $\sum_j p_j = 1$ and the aid of the Lagrange multipliers, the minimization of (6) leads to a problem exactly solvable by an inversion of the spectra-correlation matrix.¹⁶

However, the inversion is an unstable numerical procedure. Additionally, the dimension of the matrix and the computer memory required for the storage of the spectra sharply increases for multidimensional problems ($N > 1$). Therefore, instead of the inversion, we iteratively update the probability p_k associated with current geometry q_k by a correction Δ_k . At each MD step we obtain a new probability $p_k' = p_k + \beta \Delta_k$; β is an adjustable parameter. For the stationary condition from eq. (6) we get

$$\frac{\partial d'}{\partial p_k} = 2 \left(\sum_j p_j \mathbf{S}(q_j) - \mathbf{s} \right) \cdot \mathbf{S}(q_k) = 0. \quad (7)$$

To fulfill eq. (7), the new probability can be chosen as

$$p_k' = p_k + \beta \frac{\mathbf{s} \cdot \mathbf{S}(q_k) - \mathbf{S}_A \cdot \mathbf{S}(q_k)}{\mathbf{S}(q_k) \cdot \mathbf{S}(q_k)} \quad (8)$$

with $\beta = 1$, so that $\Delta_k = \frac{\mathbf{s} \cdot \mathbf{S}(q_k) - \mathbf{S}_A \cdot \mathbf{S}(q_k)}{\mathbf{S}(q_k) \cdot \mathbf{S}(q_k)}$.

Also the probability update was implemented within the Tinker dynamics. At the beginning of the simulations the probability was distributed equally, $p_k = \prod_{i=1..N} N_{ig}^{-1} \forall k$, where N_{ig} is

the number of the grid points for coordinate i , and the average spectrum \mathbf{S}_A was set to that corresponding to the initial structure. During a free MD propagation eq. (8) was used for the probability updates. At each step, the probabilities were also renormalized to obey $\sum_j p_j = 1$. With the renormalized probability, a new average spectrum was obtained as $\mathbf{S}_A' = \mathbf{S}_A(1 - p_k) + p_k \mathbf{S}(q_k)$. To account for the dispersion of the theoretical spectra (spectrum $\mathbf{S}(\mathbf{q})$ is not exquisite function of the selected subset of the internal coordinates \mathbf{q}) and for a better convergence of the resultant probability patterns a cumulative update with $\beta = 0.01$ by default was used. Typically, for large values of this parameter ($\beta \sim 1$) the probabilities p_k would oscillate around an average distribution and the information from the earlier stages of MD simulation would be lost. The period of the oscillations also somewhat depended on the number of the grid points (N_g); as indicated above 100 points were chosen to cover the 180° angular interval. On the contrary, small values ($\beta = 0.01 \dots 0.001$) lead to a smooth convergence of the probabilities within the usual (1–10 ns) simulation time.

Results and Discussion

One-Dimensional Model

Because of the enormous time required for calculation of ROA spectra, we test the method on an abstract example, the H_2O_2 molecule. The Raman and ROA spectra depend on the $q_1 = \phi = \angle \text{HOOH}$ dihedral angle ($N = 1$). For four randomly selected angles we can document two important features of the vibrational optical activity on the spectra plotted in Figure 1: the polarized spectra (ROA) are more sensitive to the conformational changes than the (unpolarized) Raman scattering. The ROA is in principle able to distinguish between enantiomers (e.g. with $\phi = -90^\circ$ and $+90^\circ$) that provide the same Raman signal.³⁰ As a reference spectrum we have chosen that $\phi = 90^\circ$. Obviously, an experimental ROA spectrum of this conformer cannot be obtained because of the fast enantiomer ($-90^\circ \leftrightarrow +90^\circ$) exchange.

The spectra generated for the free-rotating MD H_2O_2 model depend strongly on the ϕ angle, although the error from the reference spectrum (plotted in Fig. 2) exhibits a dispersion because of the additional dependence on other coordinates (bond lengths and angles) changing within the MD propagation. An additional error stems from the generation of the spectra from the fragments sampled at discrete angular increments ($\Delta\phi = 10^\circ$). Nevertheless, the most convenient conformation can be clearly identified at $\phi \sim 90^\circ$, and also the Fourier smoothed error (the red line in Fig. 2) reasonably well describes the main features of the dependence.

When the spectral energy correction [eq. (2)] is applied, the SRMD propagation does favor the conformations with angles close to the reference value of 90° . The angular probability distribution obtained after 1×10^6 propagation steps is plotted in Figure 3. As can be seen, small values of the coupling parameter ($\alpha < \sim 2$) provide negligible enrichment by the desired conformer. Also for large coupling ($\alpha > \sim 25$) the energy becomes unrealistic and the MD propagation breaks down. However, in a relatively wide interval the effect of the spectral restriction is

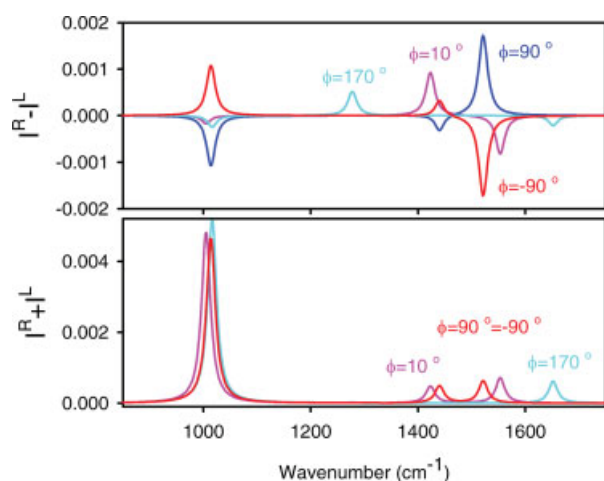


Figure 1. Example of the ROA (top) and Raman (bottom) HF/6-31G spectra of four H₂O₂ conformers.

clearly reproducible, with a maximum efficiency for $\alpha \sim 20$. The empirical SRMD scheme has thus at least a limited ability to select the spectrally favored conformation.

Also the direct generation of the probability distribution during the MD run based on eqs. (5)–(8) provides the desired conformation (left hand side of Fig. 4). The distribution suffers from the discrete coordinate grid and exhibits some noise. Nevertheless, this method provides even sharper distribution (peaked exactly at 90°) than the SRMD based on eq. (2) within a shorter time (cf. Fig. 4 with the SRMD in Fig. 3).

To mimic a multiconformer situation, we also used the averaged Raman and ROA spectrum from the conformers with $\phi = 90^\circ$ and $\phi = 150^\circ$ as the reference. As apparent from the right hand side of Figure 4 the algorithm is able to distinguish both of them, providing maximal probabilities at the right positions. In both the one- and two-conformer reference a secondary maximum appears at $\phi \sim -45^\circ$ because of an accidental similarity of the spectral shapes.

Two-Angle Dependence

For the H₂O₃ model, the spectra primarily depend on the two angles defined in Figure 5. In the same figure, the *ab initio* (HF/

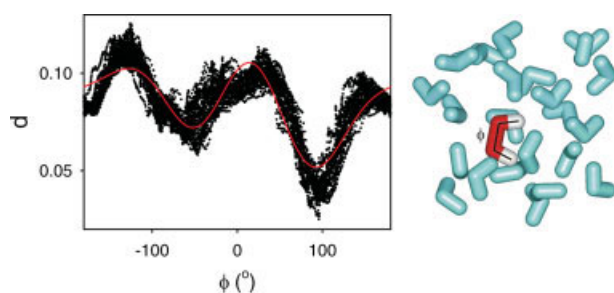


Figure 2. The errors between the reference ($\phi = 90^\circ$) and actual MD spectra (Raman and ROA sum) within a 10 ps MD run (black dots) and the Fourier-smoothed average (red curve), for a hydrated H₂O₂ molecule displayed on the right.

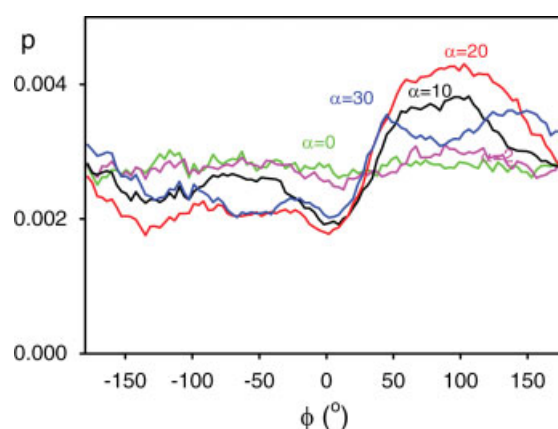


Figure 3. The angular probability distribution for the hydrated H₂O₂ molecule obtained from one million SRMD steps with five different values of the coupling parameter.

3-21G) potential energy surface and the errors from the reference ($\phi_1 = \phi_2 = 100^\circ$) Raman and ROA spectra are plotted. Similarly as for H₂O₂ we see that the H₂O₃ ROA spectrum is more discriminatory than the Raman scattering; for example the ($\phi_1 = \phi_2 = 100^\circ$) and ($\phi_1 = \phi_2 = -100^\circ$) enantiomers can be distinguished only by ROA. For the Raman not only the enantiomer spectra are the same, but also the spectra of conformers with ($\phi_1 = 100^\circ, \phi_2 = -100^\circ$) and ($\phi_1 = -100^\circ, \phi_2 = 100^\circ$) are quite similar to them.

Within the SRMD the error maps can be faithfully generated. However, as it is apparent from Figure 6 where the Raman and total Fourier-smoothed error surfaces are plotted, a relatively long simulation time is needed for reproduction of all details. The longer times are needed to sample the complete (ϕ_1, ϕ_2) plane during the MD run. The resolution is somewhat limited also by the Fourier basis [cf. eq. (3)] that smoothes some details of the *ab initio* surfaces in Figure 5.

Addition of the energy term corresponding to the total Raman and ROA error in eq. (2) causes conformer discrimination, as can be seen in Figure 7 where the probability densities are plotted for five coupling parameters. For a free dynamics ($\alpha = 0$) the angles are more or less randomly distributed, while for $\alpha = 1$ the molecule preferentially adopts conformations provid-

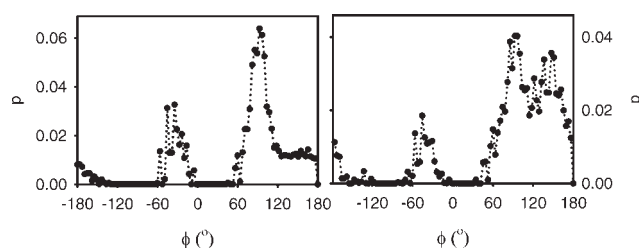


Figure 4. The angular probability distribution for the hydrated H₂O₂ molecule obtained by the direct algorithm [eqs. (5)–(8)] after 1×10^5 MD steps, for the one-conformer (left) and two-conformer (right) reference spectra.

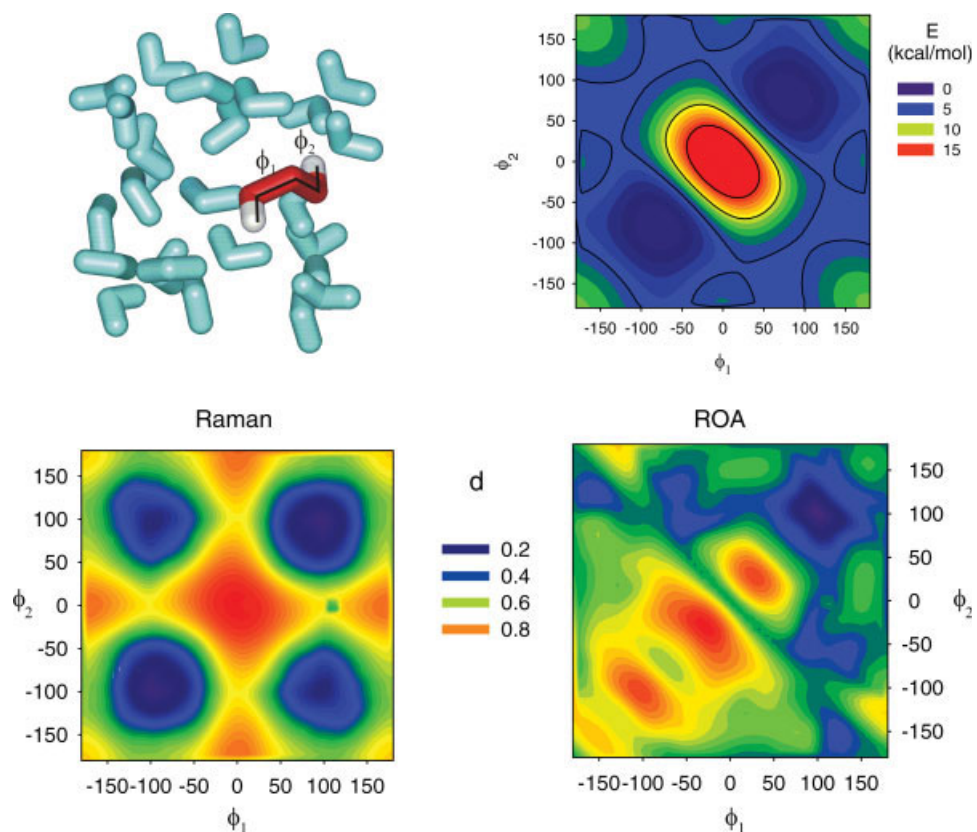


Figure 5. The hydrated H_2O_3 model molecule, and *ab initio* (HF/3-21G) maps of the relative conformer energy (E) and deviations (d) from the reference ($\phi_1 = 100^\circ$, $\phi_2 = 100^\circ$) Raman and ROA spectra.

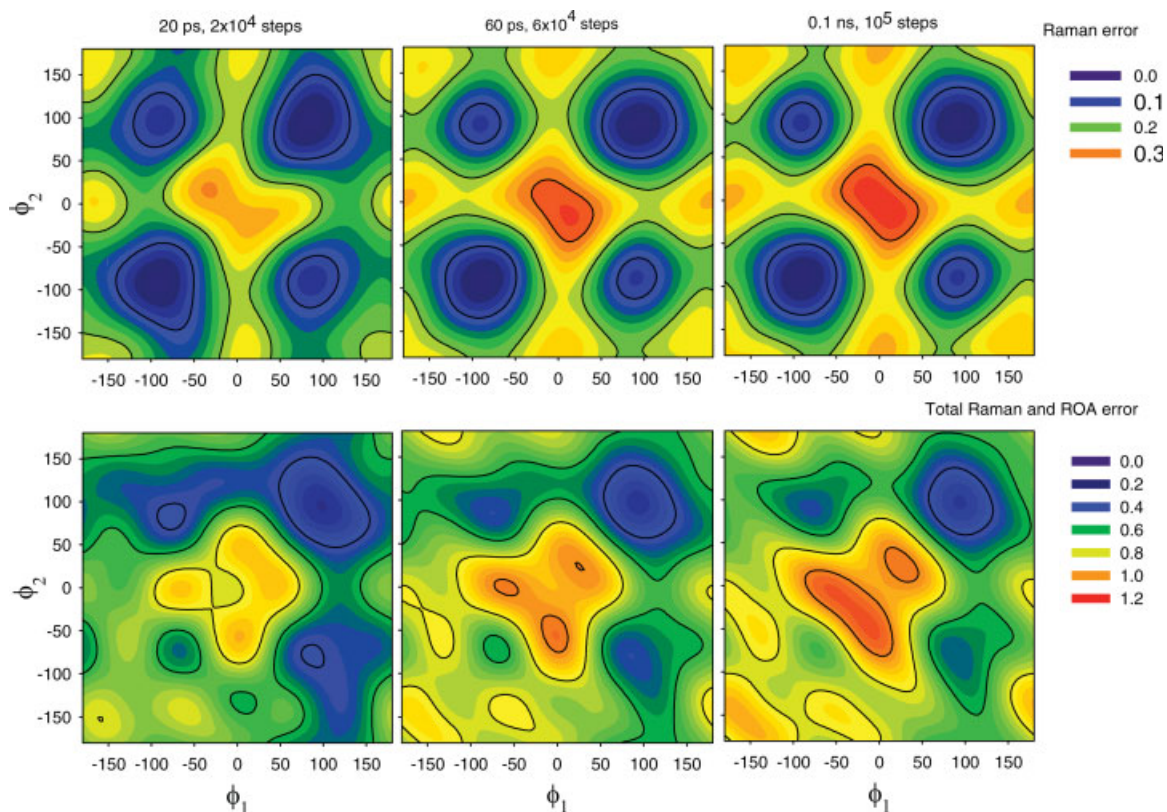


Figure 6. Fourier error maps of the differences between the actual and reference spectra of the H_2O_3 molecule constructed at different MD times.

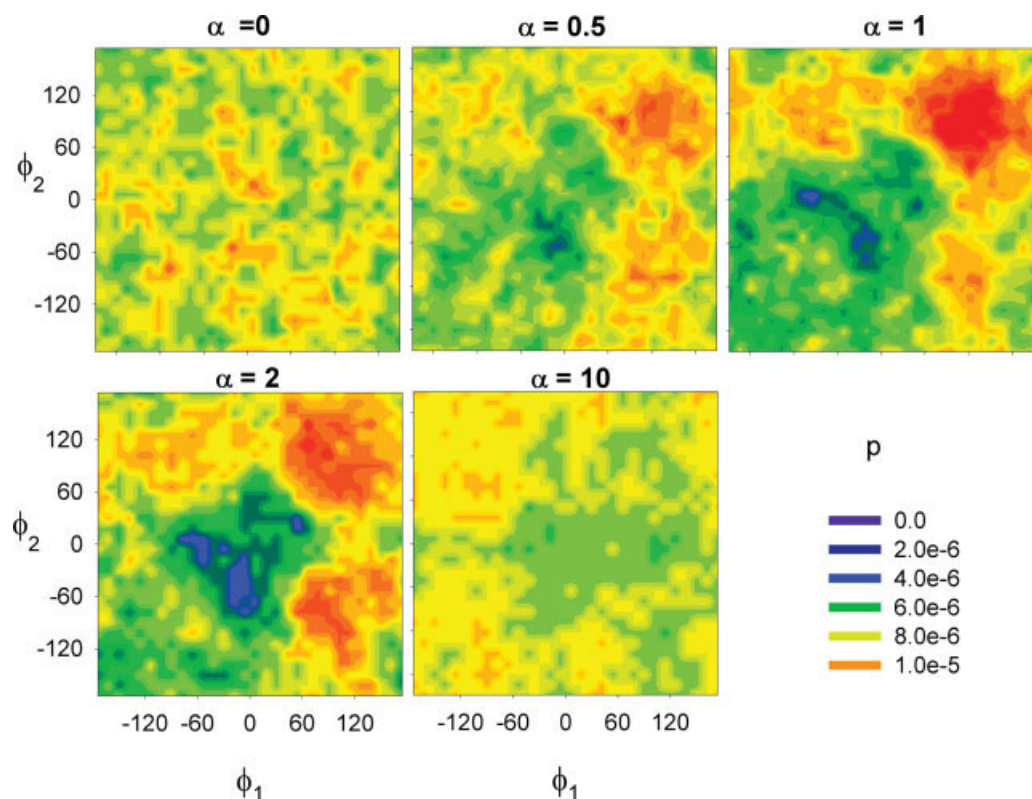


Figure 7. Angular probability distributions for the H_2O_3 conformers obtained after 1×10^6 MD steps with five different coupling parameters. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

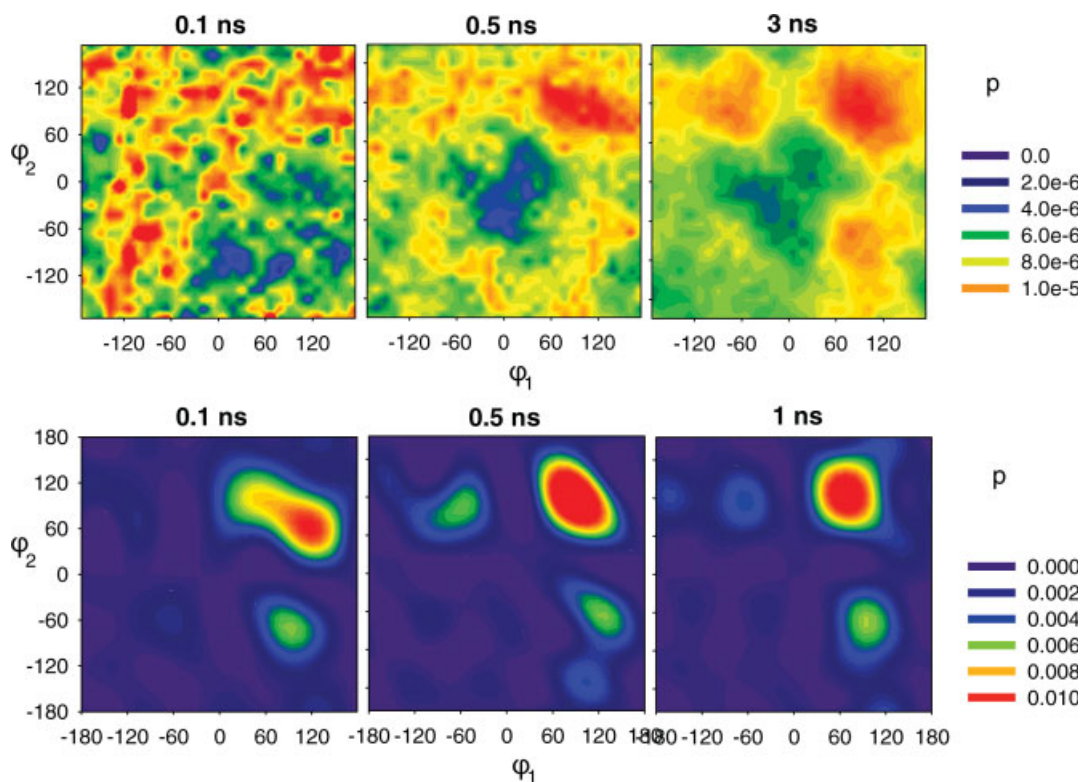


Figure 8. Angular probability distributions of the H_2O_3 conformers obtained at different times. Top: the SRMD dynamic [eq. (2)] with the coupling parameter $\alpha = 1$. Bottom: the direct probability algorithm [eq. (8)] gathered within a free MD propagation. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

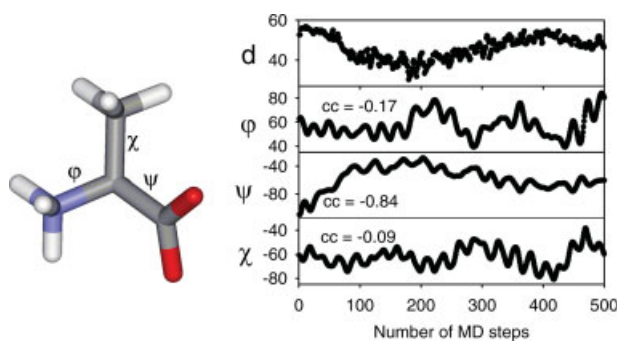


Figure 9. The dihedral angles in the alanine zwitterion (left) $\phi = \angle \text{HNCC(O)}$, $\psi = \angle \text{NCCO}$, $\chi = \angle \text{C(O)CCH}$, and spectral error, the angles, and the correlation coefficients [cc, eq. (9)] obtained during 500 MD steps (right). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ing favorable spectra. Similarly as for H_2O_2 larger coupling parameters ($\alpha \geq 10$) lead to an unstable dynamics and the angular distribution becomes erratic.

For realistic conformer distributions obtained by SRMD (Fig. 8, top) somewhat longer times are needed than for the error maps (cf. Fig. 6). This can be expected as the error (energy) surface must be constructed first before it can influence the probability distribution. For example, for 0.1 ns neither clear probability maximum nor the proper symmetry is apparent in Figure 8; these, however are quickly established at times longer than ~ 0.5 ns. For 3.0 ns we can see a reasonably converged pattern reflecting the molecular symmetry.

The Fourier-smoothed maps generated by the direct algorithm [eq. (8)] displayed at the bottom of Figure 8 appear even more practical: the preferential conformation can be recognized already at a shorter time (0.1 ns) and, similarly as for the one-dimensional H_2O_2 problem, the distribution is much sharper than that obtained by SRMD (upper part of Fig. 8). For longer times, a very distinct peak appears at the right position ($\phi_1 = \phi_2 = 100^\circ$) although minor asymmetry, caused by MD fluctuations and a finite coordinate grid (30×30 points), is still apparent.

Alanine Zwitterion

As a more applied example, we investigate the dependence of the alanine Raman and ROA spectra on the ϕ , ψ , and χ -angles driving the rotations of the NH_3^+ , CO_2^- , and CH_3 groups (see Fig. 9). Previous studies indicate that the rotations are associated with relatively small energy differences and, consequently, the large-amplitude motion of these groups strongly influences the spectral shapes.^{14,28} The continuous generation of the spectra during the MD run implemented in this work reveals further relations between the spectra and molecular structure. For example, the correlation coefficients between the spectral error and the angles,

$$cc = \frac{N \sum_i x_i d_i - \sum_i x_i \sum_i d_i}{\sqrt{N \sum_i x_i^2 - \sum_i x_i \sum_i x_i} \sqrt{N \sum_i d_i^2 - \sum_i d_i \sum_i d_i}},$$

$$x = \phi, \psi, \text{ and } \chi, i = 1..N, \quad (9)$$

obtained from the dependencies plotted at the right-hand side of Figure 9 indicate rather weak correlation between the spectra and the NH_3^+ ($cc = -0.17$) and CH_3 ($cc = -0.09$) rotations. On the other hand, the correlation for the CO_2^- group is much stronger ($cc = -0.84$) and is already evident from the visual comparison of the two dependencies (d , ψ in Fig. 9).

This completes the earlier observations¹⁴ where all the three motions were found important. The current model additionally suggests that not all geometry parameters can be separated, and that the consideration of the coupling provides a more accurate relation between the spectral shapes and molecular structure. The weak one-dimensional correlation between some angles and the spectra, however, does not mean that the geometry parameters cannot be determined. In this case the SRMD was not attempted due to excessive computational demands. The two-dimensional projections of the 3D probability distributions obtained by the second method [eq. (8)] are shown in Figure 10. The distributions are relatively sharp, narrower than those obtained from the unconstrained MD run (bottom part of Fig. 10). This can be explained by the force field inaccuracy and possibly also by some quantum effects not included in the MD averaging.³¹

The effect of the averaging on the ROA and Raman alanine spectra can be seen in Figure 11. Both the equilibrium and averaged spectra well-describe the main ROA and Raman intensity features within $600\text{--}1600\text{ cm}^{-1}$. It should be noted that the spectra can be affected by other molecular motions, interaction with the solvent, hydrogen bond formation,¹⁴ and anharmonic

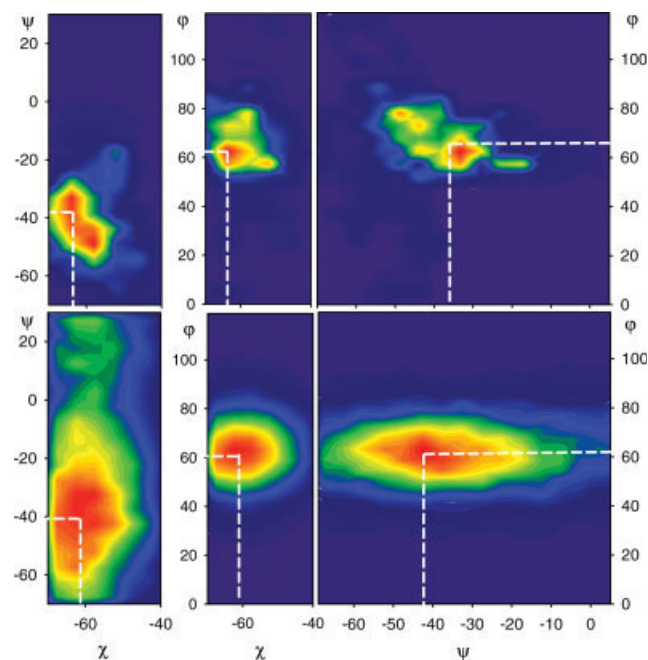


Figure 10. Angular probability distributions for the alanine zwitterion obtained from the free molecular dynamics (bottom) and from the ROA and Raman spectra by the iterative probability update [eq. (8), top]. Approximate centers of the distributions are marked by the dashed white lines. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

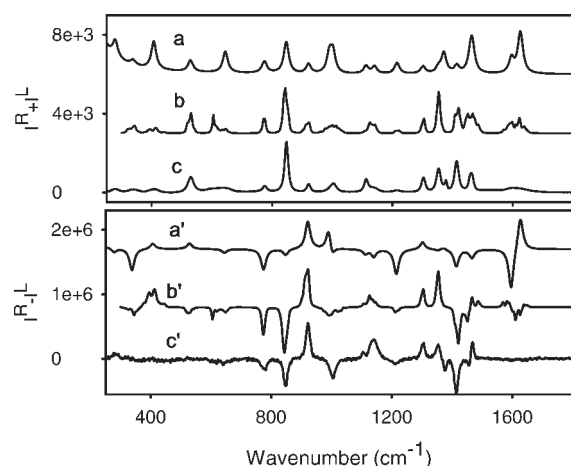


Figure 11. The ROA (a–c) and Raman (a'–c') alanine spectra calculated *ab initio* at the B3LYP/6-311++G**/CPCM level for the lowest-energy conformer (a, a'), MD averaged spectra (b, b'), and the experiment (c, c').

effects²⁸ not included in the present study. In spite of these uncertainties many positive effects of the averaging are evident, such as the broadening of the Raman and intensity-loss of the ROA bands below 450 cm^{-1} and above 1550 cm^{-1} , frequency shift and broadening of the band at $\sim 660 \text{ cm}^{-1}$. The ROA signs for peaks at 984 and 1132 cm^{-1} , and the relative 768/835 cm^{-1} peak intensities are predicted erroneously for the lowest-energy structure, unlike for the averaged spectrum.

In fact, the *ab initio* methods themselves had problems in converging to an equilibrium alanine structure.¹⁴ For example, the current wider ψ -angle MD distribution in agreement with the spectral probability seems to be more realistic than that obtained previously¹⁴ with DFT (with $\psi \sim \pm 5^\circ$), as the hydrate carbonyl oxygen atom is strongly hydrated. The hydrogen-bound water molecules, not included in the DFT-continuum solvent model, hinder too low values of the ψ -angle. For the same reason, positive values of ψ (e.g. $+40^\circ$) are not probable, as the oxygen pointing out of the NH_3^+ group would interfere with the CH_3 residue (see Fig. 12). A definite determination of the alanine hydrated structure would require further studies going beyond the focus of this work. However, the complementing multidimensional probability distributions obtained on the basis of the experimental spectra well illustrate possibilities of the method.

The conformational behavior of alanine is relatively simple and in such a case the spectra can also be interpreted by a direct comparison to simulated spectra of individual conformations. For the alanine previously performed Boltzmann averaging¹⁴ provided Raman and ROA spectra very close to the MD average in Figure 11. Similarly, a least-square spectra decomposition led to realistic estimates of conformer population of a β -hairpin forming tetrapeptide.¹⁶ The two methods presented in this study thus provide complementary means for the available spectral interpretations, allowing to include the effects of the environment, dynamics, and coupling of various molecular motions more consistently. None of the procedures can probably be applied universally. For example, the restricted MD [based on eq. (2)] may lead

to a straightforward determination of the dominant conformer in simple systems, but fail for multiconformer equilibria. The coupling of MD with the direct probability generation [eq. (8)] seems to provide sharper conformer probabilities in a shorter time also for multiconformer systems (cf. Figs. 3, 4, and 8), but might be dependent on inaccurate MD force fields.

We saw that the molecular spectral response could be used as a constraint or a reference in a MD run. In the case of the model systems the preferential conformers could unambiguously be determined. This can be generalized only when the conformer spectra are unique and only one combination matches the experiment. In a strict mathematical sense, a spectrum is viewed as a vector dimensioned by the number of recorded points, and optical activity spectra of different conformers are almost always unique. However, experimental noise and limited accuracy of the simulation severely restrict practical spectral interpretations, namely for larger molecules. As the larger system also provide more vibrational bands that can be compared, we can speculate about an optimal range of molecular sizes suitable for the statistical analysis of the spectra, including the comparison with the MD simulations. The possibility of using electronic CD spectra for the dynamics restrictions remains to be determined in future studies, as they may not contain enough information about peptide conformational preference.³²

The current implementation of the method potentially bears also other limitations: the process of the on-the-fly Raman and ROA spectral generation slows down the dynamical run and the CCT fragment method itself introduces some error in the spectra. Because of the orbital delocalization, similar semiempirical approach cannot be used for the generation of the electronic spectra at all.³³ For most applied problems a more direct *ab initio* computation of the spectrum for each MD geometry is not feasible. For example, *ab initio* computation (HF/6-31G) of one ROA spectrum for the H_2O_3 molecule on an Intel Duo 2.66 GHz processor takes about 2 min; for a 1 ns MD run and with 1×10^6 geometries 3.8 years of CPU time would be needed. Needless to say that for a more advanced approximation levels and bigger systems the computer time rises sharply.³⁴ On the other hand, owing to the locality of the vibrational interactions,

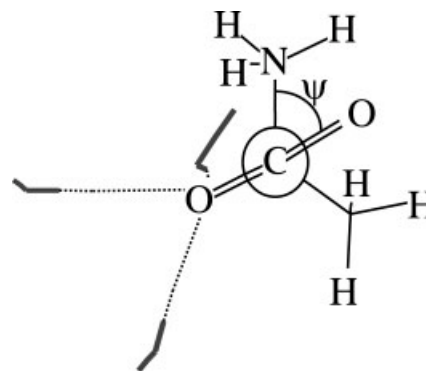


Figure 12. An alanine-water cluster (with $\psi \sim -60^\circ$) randomly selected from MD. The water molecules bound to the carbonyl oxygen are repelled by the CH_3 group and indirectly influence the solute conformation.

in most cases the CCT technique does account for the dependence of the spectra on molecular conformation.^{11,35}

Also, the present implementation is limited to a few torsion angles only. When many torsion angles are considered, the problem might become mathematically ill-conditioned and would require extensive computer memory. For the minimal planar waves basis used (containing 7 functions) the number of the Fourier coefficients [eq. (4)] is equal to 7^N . For a modest angular resolution of $3.6^\circ = 360^\circ/K$, $K = 100$, and K^N energy/error values and probabilities must be stored. Nevertheless, with suitably chosen characteristic torsion angles or other geometry parameters, the method can supposedly provide useful information about larger systems, too. Preliminary results for a glucose sugar derivative (not shown in this study) indicate that the multiscale spectral generation and analysis dramatically improve the agreement with the experiment and might enable to make structural predictions also for systems with a large number of rotating bonds.

Conclusions

The dependence of the Raman and ROA spectra on molecular conformation at the medium wavenumber region ($\sim 200\text{--}2000\text{ cm}^{-1}$) could be used as an MD constraint that influenced the resultant conformer distribution for medium-sized molecules. The restraint was added in the form of the penalty function to the MD potential and averaged by Fourier transformation, so that it provided smooth gradient that could be used in the coordinate propagation. Alternatively, the conformer probability density was obtained by an iterative update on a coordinate grid. Both methods enabled to determine the favored conformers on the basis of the optical spectra. The iterative probability update appeared more convenient as it provided sharper distributions in shorter computer times. It could also be used for multiconformer problems. In spite of some size limitations the methods thus provide powerful tools for analysis and interpretation of the optical spectra. Particularly, they help to elucidate the relations between the spectral shapes, molecular structures, and dynamics.

Acknowledgments

The authors thank Ladislav Benda for the suggestions to the manuscript.

References

1. Cantor, C. R.; Timasheff, S. N. In *The Proteins*; Academic Press: New York, 1982; p 145.
2. Drenth, J. *Principles of Protein X-ray Crystallography*; Springer-Verlag: Berlin, 1994.
3. Greenfield, N.; Davidson, B.; Fasman, G. D. *Biochemistry* 1967, 6, 1630.
4. Keiderling, T. A. In *Circular Dichroism: Principles and Applications*; Berova, N.; Nakanishi, K.; Woody, R. W., Eds.; Wiley: New York, 2000; p 621.
5. Keiderling, T. A.; Kubelka, J.; Hilario, J. In *Vibrational Spectroscopy of Polymers and Biological Systems*; Braiman, M.; Gregoriou, V., Eds.; CRC Press: Boca Raton 2006; pp 253–324.
6. Barron, L. D.; Hecht, L.; McColl, I. H.; Blanch, E. W. *Mol Phys* 2004, 102, 731.
7. Haesler, J.; Schindelholtz, I.; Riguet, E.; Bochet, C. G.; Hug, W. *Nature* 2007, 446, 526.
8. Andrushchenko, V.; Bouř, P. *J Phys Chem A* 2007, 111, 9714.
9. Silva, R. A.; Kubelka, J.; Decatur, S. M.; Bouř, P.; Keiderling, T. A. *Proc Natl Acad Sci USA* 2000, 97, 8318.
10. Chung, H. S.; Ganim, Z.; Jones, K. C.; Tokmakoff, A. *Proc Natl Acad Sci USA* 2007, 104, 14237.
11. Bouř, P.; Keiderling, T. A. *J Phys Chem B* 2005, 109, 23687.
12. Kapitán, J.; Baumruk, V.; Kopecký, V., Jr.; Bouř, P. *J Am Chem Soc* 2006, 128, 2438.
13. Kapitán, J.; Baumruk, V.; Kopecký, V., Jr.; Pohl, R.; Bouř, P. *J Am Chem Soc* 2006, 128, 13451.
14. Kapitán, J.; Baumruk, V.; Kopecký, V., Jr.; Bouř, P. *J Phys Chem A* 2006, 110, 4689.
15. Bouř, P.; Kapitán, J.; Baumruk, V. *J Phys Chem A* 2001, 105, 6362.
16. Buděšínský, M.; Šebestík, J.; Bednárová, L.; Baumruk, V.; Šafařík, M.; Bouř, P. *J Org Chem* 2008, 73, 1481.
17. Laub, P. B.; Khorasanizadeh, S.; Roder, H. *Protein Sci* 1995, 4, 973.
18. Yang, S.; Cho, M. *J Chem Phys* 2005, 123, 134503.
19. Bouř, P. *Chem Phys Lett* 2002, 365, 82.
20. Šebek, J.; Bouř, P. *J Phys Chem A* 2008, 112, 2920.
21. Verlet, L. *Phys Rev* 1967, 159, 201.
22. Ponder, J.; Tinker 4.2; W. Washington University School of Medicine: Saint Louis, 2000.
23. Bouř, P.; Sopková, J.; Bednárová, L.; Maloň, P.; Keiderling, T. A. *J Comput Chem* 1997, 18, 646.
24. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; Gaussian 03, Rev. A. 1; Gaussian, Inc.; Wallingford, CT, 2004.
25. Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 2004.
26. Kamiya, N.; Watanabe, Y. S.; Ono, S.; Higo, J. *Chem Phys Lett* 2005, 401, 312.
27. Polavarapu, P. L. *Vibrational Spectra: Principles and Applications with Emphasis on Optical Activity*; Elsevier: Amsterdam, 1998.
28. Daněček, P.; Bouř, P. *J Chem Phys* 2007, 126, 224513.
29. Bouř, P. *Collect Czech Chem Commun* 2005, 70, 1315.
30. Barron, L. D.; Zhu, F.; Hecht, L. *Vib Spectrosc* 2006, 42, 15.
31. Kaprálová, P.; Hanák, F.; Lazebník, J. *J Chem Phys* 2008, 128, 154316.
32. Glättli, A.; Daura, X.; Seebach, D.; van Gunsteren, W. F. *J Am Chem Soc* 2002, 124, 12972.
33. Šebek, J.; Kejík, Z.; Bouř, P. *J Phys Chem A* 2006, 110, 4702.
34. Ruud, K.; Helgaker, T.; Bouř, P. *J Phys Chem A* 2002, 106, 7448.
35. Kim, J.; Huang, R.; Kubelka, J.; Bouř, P.; Keiderling, T. A. *J Phys Chem B* 2006, 110, 23590.