

Simulation of Vibrational Spectra of Large Molecules by Arbitrary Time Propagation

Jan Kubelka*[†] and Petr Bouř*[‡]

Department of Chemistry, University of Wyoming, 1000 East University Avenue, Laramie, Wyoming 82071, and Institute of Organic Chemistry and Biochemistry, Academy of Sciences, Flemingovo nám. 2, 16610, Praha 6, Czech Republic

Received July 25, 2008

Abstract: Modern ab initio and multiscale methods enable the simulation of vibrational properties of very large molecules. Within the harmonic approximation, the traditional generation of the spectra based on the force field diagonalization can become inefficient due to the excessive demands on computer time and memory. The present study proposes to avoid completely the matrix diagonalization with a direct generation of the spectral shapes. For infrared absorption (IR) and vibrational circular dichroism (VCD) electric and magnetic dipole moments are propagated in a fictitious time and spectral intensities are obtained by Fourier transformation. The algorithm scales quasi-linearly, and for model polypeptide molecules the method was found numerically stable and faithfully reproduced exact transition frequencies and relative intensities.

1. Introduction

Modern vibrational spectroscopy provides a powerful means for studying the structure of polymeric materials and biological molecules.¹ In particular, the infrared absorption (IR) and vibrational circular dichroism (VCD) techniques shed light on structure and structural transitions in peptides and proteins,^{2,3} nucleic acids,^{4,5} biological membranes,⁶ and on enzymatic function.⁷ Kinetic and nonlinear response enhancements are possible owing to the time-dependent⁸ and two-dimensional (two-photon) techniques.⁹

Ultimately, the information about the structure and dynamics of studied systems can be verified by comparison of the measured spectra with simulations. Unfortunately, structural interpretation of the vibrational spectra is often far from being straightforward, especially for large molecules with a plethora of overlapping spectral transitions. Although the molecular size remains the main limitation for quantum mechanical calculations, the exciting possibility of obtaining accurate vibrational force fields for the whole molecules appears realistic owing to the latest advances in quantum chemistry.^{10–12} In the meantime, the Cartesian transfer tensor

techniques¹³ obtained spectroscopically accurate fields for large molecules from quantum-chemical computations on smaller fragments.^{3,14} In principle, such fragmentation approaches would enable calculation of the vibrational spectra for arbitrarily large structures. However, for systems with thousands of atoms a new problem arises: the force field matrices become too large for direct diagonalization, which is necessary to obtain the vibrational frequencies and intensities. For such very large matrices special computational procedures have to be applied.¹¹

Traditionally, the Householder transformation and the following complete diagonalization routine based on the QL or QR algorithm provide the fastest in-memory procedure for the determination of the eigenvectors and eigenvalues of a real symmetric matrix.¹⁵ Similar routines exist for a partial diagonalization, yielding a given number of the lowest or highest eigenvalues, but these algorithms are less efficient except for cases when the number of the vectors of interest is small.¹⁶ For matrices that cannot be stored in computer memory or elements of which are created on the fly, the power iteration methods appear as a better alternative.^{11,17} The eigenvalues and vectors are built from the smallest or the largest value iteratively, using a limited number of trial vectors (“Krylov space”).¹⁸ This is very convenient, for example, for the configuration interaction electron computations, when only a limited number of the low energy

* Corresponding author e-mail: jkubelka@uwyo.edu (J.K.), bour@uochb.cas.cz (P.B.).

[†] University of Wyoming.

[‡] Academy of Sciences.

electronic states (often just the ground state) is needed. However, for the vibrational problem it is usually not sufficient to find a few largest or smallest eigenvalues. Rather, a complete spectrum, or at least a broad frequency range, is desired, which is rarely limited to either the high or low frequency side. For such a task the iterative (often called Davidson) methods quickly become impractical. As each new vector has to be normalized to the rest, the process slows down progressively. Additionally, for semidegenerate eigenvalues, the method can become numerically unstable.¹¹

Fortunately, in some cases, the force field matrix diagonalization can be completely avoided by Fourier techniques and time propagations of molecular properties. An interesting alternative is offered by the normal mode tracking,¹⁹ where the orthogonality of the harmonic normal modes is used to construct an incomplete Hamiltonian, and, in a final effect, a preselected part of the vibrational spectrum is obtained. The decision as to which mode to follow has to be done already at the ab initio stage of computation of the force field, which somewhat limits possible applications. The Fourier procedures pursued in the presented work were also successfully used in the past, e.g. for generation of the infrared absorption (IR) and optical activity (vibrational circular dichroism, VCD) in connection with empirical¹² or ab initio²⁰ based force fields. In this work, we investigate a variant of these approaches which differs from the previous schemes in that the time propagation is based directly on the force field matrix instead of the harmonic Hamiltonian. The spectra are not collected in a real frequency (ω) space, but in the space of the force field eigenvalues λ ($\lambda = \omega^2$). This approach brings a tremendous simplification of the process, because it eliminates the problems associated with molecular dynamics, such as the temperature definition, normal mode energy-redistribution,²⁰ and the need to hold the large force field matrix in memory. For giant molecules this method yields accurate relative absorption and VCD intensities and the exact IR/VCD ratio in a fraction of computer time needed for the direct diagonalization routines. The principal drawback is that the exact relative intensity distribution is somewhat dependent on the initial guess, but this can be circumvented by averaging of more trajectories. While the absolute intensities retain a small amount of error, this does not represent a major problem in the practical applications.

In this report, we briefly review the process of generating the theoretical IR and VCD spectra within the harmonic approximation. Most of the Method section is devoted to the detailed description of the new algorithm. Finally, the numerical stability and convergence tests and comparisons with other diagonalization methods for model polypeptide IR and VCD spectra simulations are presented and discussed.

2. The Method

Vibrational Analysis. In the harmonic approximation²¹ the vibrational Hamiltonian can be written in terms of nuclear momenta P_i and displacements from equilibrium positions of N atoms ΔR_i ($i = 1..N$) as

$$H = \frac{1}{2} \left(\sum_{i=1}^{3N} \frac{P_i^2}{M_i} + \sum_{i=1}^{3N} \sum_{j=1}^{3N} F_{ij} \Delta R_i \Delta R_j \right) = \frac{1}{2} (\mathbf{p}' \cdot \mathbf{p} + \mathbf{q}' \cdot \mathbf{f} \cdot \mathbf{q}) \quad (1)$$

where M_i are nuclear masses, $p_i = P_i/(M_i)^{1/2}$, $q_i = (M_i)^{1/2} \Delta R_i$, and $f_{ij} = F_{ij}/(M_i M_j)^{1/2}$ are the respective mass-weighted momenta, coordinates, and force field. The harmonic force field (referred to also as the Hessian or Cartesian force constant matrix) is formed by the second energy (ϵ) derivatives

$$F_{ij} = \frac{\partial^2 \epsilon}{\partial R_i \partial R_j} \quad (2)$$

The multidimensional problem is solved by a transformation into the normal coordinates, $Q_k = \sum_{j=1}^{3N} s_{kj}^{-1} q_j$, and to corresponding momenta $\Pi_k = \sum_{j=1}^{3N} s_{kj}^{-1} p_j$, so that $\mathbf{s}' \cdot \mathbf{s} = \mathbf{E}$, where \mathbf{E} is the identity matrix and the transformed force field becomes diagonal

$$s_{ki} f_{ki} s_{ij} = \omega_i^2 \delta_{ij} \quad (3)$$

where δ_{ij} is the Kronecker symbol ($\delta_{ij} = 1$ for $i = j$, $\delta_{ij} = 0$ for $i \neq j$), and ω_i are the normal mode angular frequencies. The vibrational Hamiltonian then becomes a sum of one-dimensional harmonic oscillator Hamiltonians h_i

$$H = \frac{1}{2} \sum_{i=1}^{3N} (\Pi_i^2 + \omega_i^2 Q_i^2) = \sum_{i=1}^{3N} h_i(Q_i) \quad (4)$$

Consequently, the equations of motion for the nuclei reduce to a set of $3N$ uncoupled one-dimensional Schrödinger equations, energies, and wave functions of which can be obtained analytically.²¹ (Equation 4 comprises also the translational and rotational modes not visible in the spectrum.) In particular, frequencies of the fundamental vibrational bands are equal to ω_i . Solving the vibrational problem therefore reduces to the diagonalization of the force field matrix \mathbf{f} (eq 3).

The intensities of the absorption and VCD bands of a fundamental transition $0 \rightarrow v$ are proportional, respectively, to the dipolar (D) and rotational (R) strength defined as $D_{0v} = \boldsymbol{\mu}_{0v} \cdot \boldsymbol{\mu}_{v0}$ and $R_{0v} = \text{Im} \boldsymbol{\mu}_{0v} \cdot \mathbf{m}_{v0}$, where $\boldsymbol{\mu}_{0v}$ is the electric and the \mathbf{m}_{0v} is the magnetic transition dipole moment, $v = 1..3N$. These are obtained as²²

$$\boldsymbol{\mu}_{0v} = \left(\frac{\hbar}{2\omega_v} \right)^{1/2} \sum_{i=1}^{3N} \mathbf{p}_i S_{vi} \quad (5a)$$

$$\mathbf{m}_{0v} = -\sqrt{2\hbar^3 \omega_v} \sum_{i=1}^{3N} \mathbf{a}_i S_{vi} \quad (5b)$$

where \mathbf{p}_i is the atomic polar tensor (APT), and \mathbf{a}_i is the atomic axial tensor (AAT) with Cartesian ($\beta = x, y, z$) components

$$p_{i\beta} = \frac{\partial \mu_\beta}{\partial R_i} \quad (6a)$$

$$a_{i\beta} = \frac{\sqrt{-1} M_i \partial m_\beta}{2\hbar \partial P_i} \quad (6b)$$

$S_{vi} = \partial R_i / \partial Q_v = s_{vi} / \sqrt{M_i}$ is the direct (not mass-weighted) Cartesian-normal mode transformation matrix, and \hbar is the

Planck constant.²³ For calculation of the infrared absorption and VCD intensities it is therefore necessary to compute the derivatives of the molecular electric ($\boldsymbol{\mu}$) and magnetic (\mathbf{m}) dipole moments.

The Algorithm. As stated in the Introduction, for very large molecules the direct or iterative procedures for finding the eigenvalues and eigenfunctions of the force field matrix \mathbf{f} (eq 3) become impractical. An alternative means is to use the force field matrix to propagate an arbitrary test vector in time. The process is started by a generation of a random set of M unit vectors $\boldsymbol{\sigma}_1 \dots \boldsymbol{\sigma}_M$ at the Cartesian coordinate $3N$ -dimensional space. These can be thought of as linear combinations of the eigenvectors of the force field matrix, $\boldsymbol{\sigma}_i = \sum_{j=1}^{3N} c_{ij} \mathbf{s}_j$. Then we introduce time-dependent vectors $\boldsymbol{\sigma}_i(t) = \sum_j c_{ij} \mathbf{s}_j e^{i\lambda_j t}$. Note that t referred to as the fictitious “time” has units of second² (in this work corresponding atomic units are used) since λ_j is not a frequency, but $\lambda_j = \Lambda_{jj} = \omega_j^2$. Using eq 3 we get $f_{\beta\alpha} s_{j\alpha} = \lambda_j s_{j\beta}$, and the propagation of $\boldsymbol{\sigma}_i$ can be approximated for small time intervals dt as

$$\boldsymbol{\sigma}_i(t+dt) \cong \boldsymbol{\sigma}_i(t) + \frac{d\boldsymbol{\sigma}_i(t)}{dt} dt + \frac{1}{2} \frac{d^2\boldsymbol{\sigma}_i(t)}{dt^2} dt^2 = \boldsymbol{\sigma}_i(t) + \sum_j ic_{ij} \lambda_j \mathbf{s}_j e^{i\lambda_j t} dt + \boldsymbol{\sigma}_i^{(2)} = \boldsymbol{\sigma}_i(t) + i\mathbf{f} \cdot \boldsymbol{\sigma}_i(t) dt + \boldsymbol{\sigma}_i^{(2)} \quad (7)$$

In the current implementation the second-derivative correction $\boldsymbol{\sigma}_i^{(2)}$ is not calculated directly, as this would involve a computationally impractical formation of the square matrix \mathbf{f}^2 . Instead, we use an approximate formula, $\boldsymbol{\sigma}_i^{(2)}(t) \cong 0.5(\bar{\boldsymbol{\sigma}}_i(t-2dt) + \bar{\boldsymbol{\sigma}}_i(t) - 2\bar{\boldsymbol{\sigma}}_i(t-dt))$, where $\bar{\boldsymbol{\sigma}}_i$ are the vectors propagated according to eq 7 with finite time steps dt and normalized afterward. The involvement of $\boldsymbol{\sigma}^{(2)}$ appeared very beneficial for the quality of the results by allowing for a significant increase of dt . Only a minor influence of a third-derivative correction introduced in a similar manner was observed; therefore, the third derivative was not used by default.

In principle, we can construct autocorrelation functions

$$C_i(t) = \boldsymbol{\sigma}_i(t) \cdot \boldsymbol{\sigma}_i = \sum_j c_{ij}^2 e^{i\lambda_j t} \quad (8)$$

and obtain the eigenvalues as peak positions of its Fourier transform $I_i(\lambda)$

$$I_i(\lambda) = \int C_i(t) e^{-i\lambda t} dt = 2\pi \sum_j c_{ij}^2 \delta(\lambda_j - \lambda) = \sum_j I_{ij}(\lambda) \quad (9)$$

In a similar way, we can define M time-dependent electric and magnetic dipole moments as

$$\boldsymbol{\mu}_{i\beta}(t) = \sum_{j=1}^{3N} p_{j\beta} \boldsymbol{\sigma}_{ij}(t) \text{ and } m_{i\beta}(t) = \sum_{j=1}^{3N} a_{j\beta} \boldsymbol{\sigma}_{ij}(t) \quad (10)$$

with corresponding Fourier transforms

$$\boldsymbol{\mu}_i(\lambda) = \int \boldsymbol{\mu}_i(t) e^{-i\lambda t} dt \text{ and } \mathbf{m}_i(\lambda) = \int \mathbf{m}_i(t) e^{-i\lambda t} dt \quad (11)$$

For a finite-time propagation the δ -functions in (9) can be thought of as finite-width peaks. In fact, in the actual simulations the transforms (eq 11) were calculated for a discrete set of vibrational frequencies ω_i ($\omega_i = \lambda_i^2$), typically spaced by 1 cm^{-1} from 100 to 4000 cm^{-1} . Instead of

the usual fast Fourier transformation the Fourier sums (9, 11) were incremented at each step, which did not significantly increase the total time but enabled to avoid storage of the entire trajectory in computer memory. To avoid too narrow peaks, at each time step the transforms were additionally convoluted with a dispersion function in the form of either Lorentzian, $(2/\pi\Delta)/\{1+[2(\omega-\omega_0)/\Delta]^2\}$, or Gaussian, $2(\ln(2))^{1/2} \exp\{-[2(\ln(2))^{1/2}(\omega-\omega_0)/\Delta]^2\}/(\Delta\sqrt{\pi})$, peaks; Δ is the full width at half-maximum.

Performing the integration (11), from (8) and (10) we obtain for the electric dipole $\boldsymbol{\mu}_i(\lambda) = 2\pi \sum_v c_{iv} \sum_k \mathbf{p}_k \cdot \mathbf{s}_{vk} \delta(\lambda_v - \lambda)$. Thus for each resolved transition v the integration of $\boldsymbol{\mu}_i(\lambda)$ would yield a spectral intensity proportional to the sum $\sum_k \mathbf{p}_k \cdot \mathbf{s}_{vk}$ (cf. eq 5a). But the expansion coefficients c_{iv} are unknown. They can in principle be obtained from eq 9, but this is impractical as unrealistically long accumulation times are required to resolve modes that are very close in frequency. Instead, we use a numerically more stable procedure and directly calculate the dipolar strengths averaged over the M $\boldsymbol{\sigma}$ -vectors as

$$\begin{aligned} \langle D(\lambda) \rangle &= M^{-1} \sum_{i=1, M} \text{Re} \boldsymbol{\mu}_i(\lambda) \cdot \text{Re} \boldsymbol{\mu}_i(\lambda) \\ &= (2\pi)^2 M^{-1} \sum_{i=1, M} \sum_{v, v'} \langle c_{iv} c_{iv'} \rangle \sum_{k, k'} \mathbf{p}_k \cdot \mathbf{p}_{k'} \cdot \mathbf{s}_{v'k'} \cdot \mathbf{s}_{vk} \delta(\lambda_v - \lambda) \delta(\lambda_{v'} - \lambda) \\ &\cong (3N)^{-1} K (2\pi)^2 M^{-1} \sum_{i=1, M} \sum_v \sum_{k, k'} \mathbf{p}_k \cdot \mathbf{p}_{k'} \cdot \mathbf{s}_{vk'} \cdot \mathbf{s}_{vk} \delta(\lambda_v - \lambda) \quad (12a) \end{aligned}$$

where we multiplied the finite-width “ δ -functions” as $\delta(\lambda_v - \lambda) \delta(\lambda_{v'} - \lambda) = K \delta_{vv'} \delta(\lambda_v - \lambda)$, where K is a constant dependent on the actual shape of δ . For the Gaussian shapes, for example, $K = 1/(\Delta\sqrt{2\pi})$, $\Delta = 2\omega d$, where d is the bandwidth. For a random initial distribution of the coefficients $\langle c_{iv}^2 \rangle = (3N)^{-1}$, where N is the number of atoms. Analogously, we get for the rotational strength

$$\langle R(\lambda) \rangle = M^{-1} \sum_{i=1, M} \text{Re} \boldsymbol{\mu}_i(\lambda) \cdot \text{Re} \mathbf{m}_i(\lambda) \cong (3N)^{-1} K (2\pi)^2 M^{-1} \times \sum_{i=1, M} \sum_v \sum_{k, k'} \mathbf{p}_k \cdot \mathbf{a}_{k'} \cdot \mathbf{s}_{vk'} \cdot \mathbf{s}_{vk} \delta(\lambda_v - \lambda) \quad (12b)$$

Thus, if we allow for a small inaccuracy of the absolute intensity scale given by the variance of the peak shapes (“ δ -functions”, which may deviate from the Gaussian peaks), we obtain precise relative absorption and circular dichroism peak intensities with correct ABS/CD ratios from eqs 12a-b.

Matrix Storage. The storage of the Hessian matrix \mathbf{F} requires $(3N)^2$ words of memory. Although this is usually a minor problem for modern computers, the handling of the matrix may still become impractical for systems with several thousands of atoms. Moreover, because of the locality of vibrational interactions,²⁴ an overwhelming majority of the matrix elements is negligible. Particularly if the force field of a “big” molecule is constructed from smaller fragments, as in the Cartesian tensor transfer (CCT) method,¹³ for linear biopolymers (nucleic acids, peptides) only a diagonal-like band of the matrix is formed, and the required amount of memory scales linearly with N . Therefore, in the current implementation, for each line l of the matrix we store only

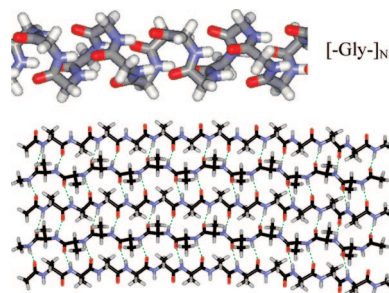


Figure 1. The model systems, variable-length polyglycine in the α -helical conformation (top) and the 5-strand antiparallel β -sheet with Ac-(Ala)₁₂-Me strands (Ac=acetyl, Ala=alanine, Me=methyl, bottom).

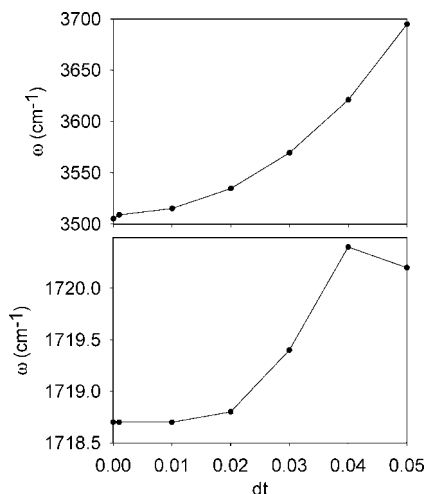


Figure 2. The dependence of the frequency on the integration time step (fictitious time in atomic units, see text), for a N–H stretching (high frequency - top) and C=O stretching (low frequency -bottom) vibrational bands of Gly₁₃₀.

n_l nonzero nonredundant elements F_{lk} , $k = r_1 \dots r_{n_l}$ (using $F_{ij}=F_{ji}$) and the corresponding row-indices r_i . For convenience, diagonal elements F_{ii} are divided by two; then multiplication of a vector \mathbf{A} by \mathbf{F} , $\mathbf{B} = \mathbf{F}\mathbf{A}$, can be realized by a simple algorithm when only one line of \mathbf{F} at a time is read from the disk and stored in computer memory:

1. Zero-out \mathbf{B} , $l = 1$.
2. Read-in l -line of \mathbf{F} .
3. For all $j \in (r_1 \dots r_{n_l})$, update $B_j = B_j + F_{lj}A_l$ and $B_l = B_l + F_{lj}A_j$.
4. Increment $l = l + 1$, if $l < 3N$ go to 2.

The storage required for the atomic axial and polar tensors cannot be reduced; however, this scales linearly with N and thus does not limit computations of even considerably large molecules.

Model Systems. The method was tested on model periodic systems, the polyglycine peptide in an α -helical conformation, and a five-strand antiparallel β -sheet $5 \times$ Ac-(Ala)₁₂-Me, shown in Figure 1. The periodicity facilitated the construction of a reasonably accurate force field based on ab initio computation of smaller fragments. For the α -helix a peptide heptamer (containing 7 peptide bonds), Ac-[Gly]₆-Me, in the same conformation was subjected to the normal mode optimization routine²⁵ with modes within -300.300

cm^{-1} fixed. Thus the higher-frequency mode visible in the spectra could be relaxed under a minimal change of the geometry. Similarly for the β -sheet, a smaller $3 \times$ Ac-(Ala)₂-Me fragment²⁶ was used as a source of the force field and the intensity tensors, the geometry of which was optimized with fixed main chain torsion angles. The BPW91 level of approximation²⁷ and 6-31+G** (α -helix) and 6-31G** (β -sheet) bases were used for the calculation with the aid of the Gaussian program package.²⁸ For the optimized geometries, the harmonic force field and intensity tensors were computed at the same level by Gaussian and transferred on the polymer by the Cartesian tensor transfer techniques.¹³

3. Results and Discussion

Time Step Dependence. First we have investigated the sensitivity of the method to the size of the integration time step. Previous simulation of the vibrational spectra using classical molecular dynamics trajectories revealed a significant dependence of the Fourier-transformed frequencies on the integration steps.²⁰ This is true also for the current method, as can be seen in Figure 2 where the dependence is plotted for a low (carbonyl stretching) and high (N–H stretching) frequency band of the (Gly)₁₃₀ system. The “time” variable, the force field, and the eigenvalues Λ_{ii} are given in atomic units; then we can obtain the angular frequency ω (wavenumber) in cm^{-1} as $\omega = 1302.8\sqrt{\Lambda_{ii}}$. We can see that the dependence is much sharper for the N–H stretching band, where the time step dt of 0.05 (au) introduces a huge error of $\sim 200 \text{ cm}^{-1}$. However, this can be easily fixed by keeping the step small as with $dt \sim 0.01$ an acceptable error is obtained even for the higher-frequency band, while the propagation gives virtually the exact value for the carbonyl stretch. Note that harmonic frequencies of the higher-frequency hydrogen stretching transitions computed by quantum chemical methods are regularly by up to 10% larger than the experimental transitions,²⁹ which makes a 1% inaccuracy ($\sim 10\text{--}30 \text{ cm}^{-1}$) in the vibrational frequencies acceptable. Shorter steps are not desirable as they make the computation unnecessarily longer.

Band Width Convergence. As pointed out in the Method section, the consequence of the finite interval of time propagation is that spectral peaks are not infinitely sharp but have finite widths. Therefore, it is necessary to establish the number of the time propagation steps needed to obtain sufficiently narrow bands. The dependence of the absorption and VCD spectra on the number of the propagation steps plotted in Figure 3 reflects the obvious behavior of the Fourier transformation where the peak width is inversely proportional to the evolution time. However, as the square of the frequency is transformed, we can observe that the higher-frequency bands become narrower than the lower-frequency ones. This is shown quantitatively in Figure 4, where the spectral line widths for the C=O (1718 cm^{-1}) and N–H (3505 cm^{-1}) stretching bands are plotted as functions of the number of propagation steps. Needless to say, for any given region of the spectrum, an arbitrarily narrow width can be obtained by a sufficiently long simulation time. Overall, already at the early stages of the propagation correct relative absorption intensities are obtained. The converged

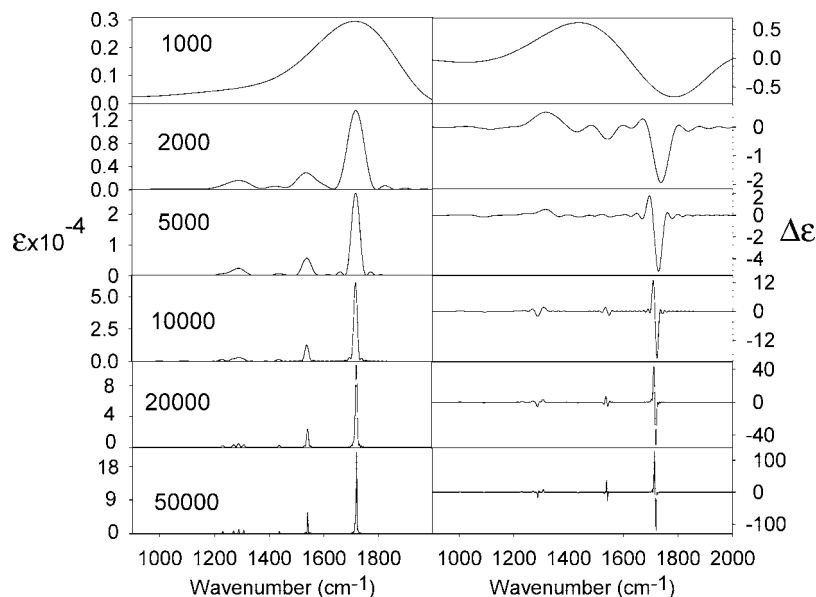


Figure 3. The dependence of the absorption (left) and VCD (right) spectra of Gly₁₃₀ on the number of the propagation steps (indicated in the left panel, for 50 average spectra).

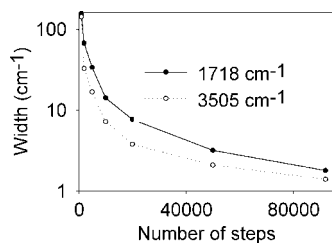


Figure 4. The dependence of the full widths at half-maximum of the two selected peaks, the N–H stretch (3505 cm⁻¹) and C=O stretch (1718 cm⁻¹) in Gly₁₃₀ on the number of the propagation steps. Note, that the y-scale is logarithmic.

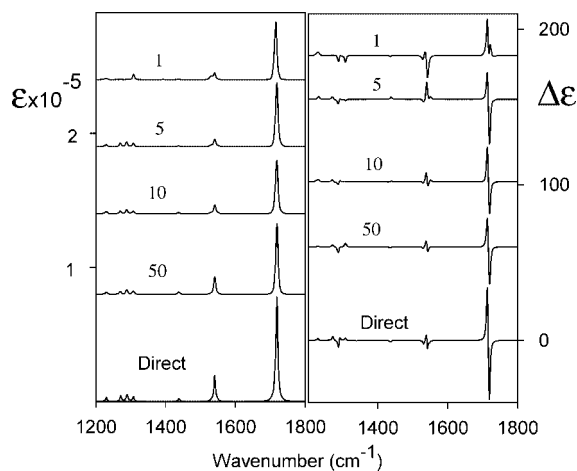


Figure 5. Absorption (left) and VCD (right) spectra simulated (from top, for 2000 propagation points) for 1, 5, 10, and 50 random vector averages as compared to the exact results obtained by the direct diagonalization (bottom traces), for Gly₁₃₀.

VCD shapes require a longer time; for example, the conservative carbonyl stretching split-band at ~ 1718 cm⁻¹ stabilizes at ~ 20000 steps. For a typical IR or VCD experiment where the inhomogeneous band widths seldom become smaller than ~ 10 cm⁻¹ a propagation with ~ 20000

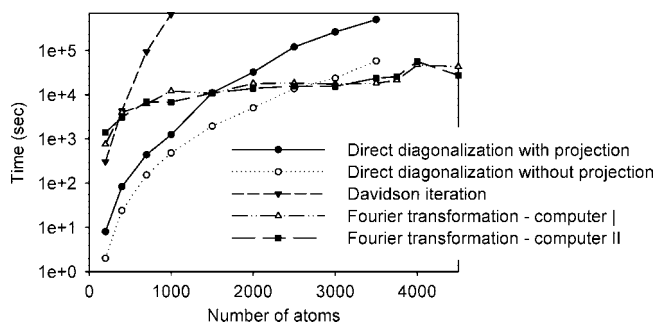


Figure 6. The dependence of the time (in seconds, on a logarithmic scale) needed to diagonalize the force field matrix on the number of atoms in the (Gly)_N polymer. (Detailed description of each method is given in the text.)

steps thus already provides the desired spectral shape in the C=O stretching region.

Trial Vector Averaging. With a sufficiently small integration step the procedure yields the correct vibrational frequencies. However, the spectral intensities resulting from a particular propagation run differ as a consequence of the randomly chosen initial vectors σ_i . To obtain a stable solution, independent of the choice of the initial conditions, it is necessary to average several such runs. The convergence of the relative IR and VCD band intensities on the number of the initial vectors σ_i is explored in Figure 5. An instantaneous convolution with the Lorentzian bands was performed during the simulation in order to achieve a constant bandwidth of 10 cm⁻¹. The resultant spectra of (Gly)₁₃₀ for $M = 1, 5, 10,$ and 50 are compared with the exact result based on the direct Householder diagonalization. While a randomly selected vector ($M = 1$) provides unrealistic relative intensities, even with some wrongly predicted VCD signs, the spectral profile quickly stabilizes, and for $M = 50$ the spectra are practically indistinguishable from the exact intensities. As explained above, the minor underestimation of the absolute intensities can be explained

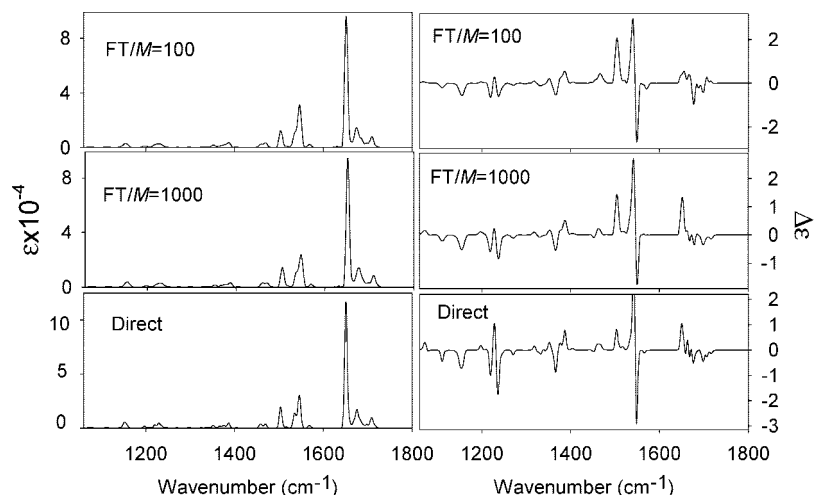


Figure 7. Absorption (left) and VCD (right) spectra of the 660 atom β -sheet simulated by the time propagation (FT, for $M = 100$ and 1000 trial vectors) as compared to the exact results obtained by the direct diagonalization (bottom trace); the 5 cm^{-1} bandwidth was used in the simulations.

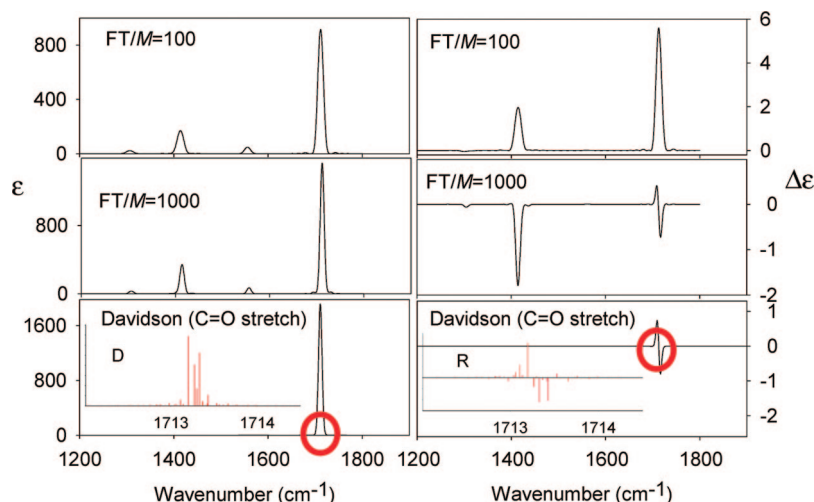


Figure 8. Absorption (left) and VCD (right) spectra of the 14000 atom poly glycine α -helix simulated by the time propagation (for $M = 100$ and 1000 trial vectors, 8 h and 3.5 days of the computer time was needed, respectively) as compared to the exact result obtained by the Davidson method (finished in 9 days and 15 h, only the C=O stretching was included). The spectra were normalized to one amide, and 5 cm^{-1} bandwidth was used. For the Davidson method, positions of the dipolar and rotatory strengths of individual C=O stretching transitions are indicated in the inset graphs.

by the approximation for the averaging and the δ -function multiplication.

CPU Time Scaling. As the number of nonzero matrix elements is approximately proportional to the number of atoms N , and the matrix-vector multiplication in eq 7 required for the propagation σ_i involves a number of multiplications proportional to N^2 , we can expect practical scaling between N^1 and N^2 . This is confirmed by the time dependence on the number of atoms for $(\text{Gly})_K$, $K = 28 \dots 0.642$ plotted in Figure 6. The detailed scaling of the disk-based time-propagation procedure is dependent on the disk fragmentation and detailed memory management; therefore, two computers with the same processors (“computer I and II” in the Figure 6, AMD64 1 GHz) provided slightly different dependencies. Curiously, a larger force field matrix can occasionally even yield a shorter simulation time. We can see that the Fourier method scales much more favorably than the other procedures. For the comparison we used our own implementation

of the iteration (Davidson) method based on the Mitin’s modification of the algorithm,²⁶ while a standard code was used for the direct (Householder) diagonalization, expanded by the possibility of correcting the ab initio force fields for the translational and rotational invariance.¹⁵ In Figure 6 we can see that the complete Davidson diagonalization is faster than the time propagation up to $N \sim 500$, and then the time quickly grows to immeasurable values. The direct Householder diagonalization is more efficient up to $N \sim 1500$, which can be somewhat improved by skipping the projection of the zero-vibrational modes (rotations and translations) from the force field. However, because of the intrinsic N^3 dependence and huge memory requirements of the direct method, the propagation in the fictitious time becomes the only method usable for $N > 3000$.

Large Peptide Systems. We have applied the Hessian propagation to simulations of the IR and VCD spectra for the β -sheet poly alanine segment (660 atoms) and the poly

glycine α -helix containing 14000 atoms (Figure 1). The absorption and VCD spectra for these two protein models simulated for 20000 time steps are plotted in Figures 7 and 8. The β -sheet spectrum could be compared to that obtained by the Householder diagonalization, while the reference α -helical spectrum was calculated using a limited Davidson diagonalization. The Davidson method provided 2000 vibrational C=O stretching mode frequencies, starting from the highest value, when hydrogen masses were arbitrarily increased to 10 g/mol, not to interfere with the carbonyl vibrations. For both systems, we can see that the time propagation method faithfully reproduces the relative IR intensity and VCD sign patterns, although a relatively large number of the trial vectors is needed for accurate results. Particularly, for the α -helix, the 100-vector average (Figure 8, top) is not sufficient for the VCD simulation as it provides a net positive signal instead of the split carbonyl band. The results nicely confirm the ability of the IR and VCD spectroscopies to distinguish various peptide and protein secondary structures: the high-frequency IR carbonyl stretch components (1660–1740 cm^{-1} , Figure 8) and the associated weak VCD signal are characteristic for the β -sheet forms,³⁰ while a single IR carbonyl band and an intense, positive VCD couplet (i.e., positive to negative, from low to high frequency) correspond well to the experimental data from helix-rich proteins.³

The absorption and VCD spectra of regular structures, particularly the poly glycine α -helix, originate in a few semidegenerate transitions. The positions of the most intense peaks are shown in the insets in the bottom panels of Figure 8. Most of the intensities come from a frequency interval about 0.5 cm^{-1} wide. As discussed previously, this accumulation of spectral intensities reflects the translational symmetry of the structure and phonon-like delocalized normal modes.⁵ Therefore, the propagation method is capable of correctly reproducing the overall spectral shape arising from the specific intensity distribution also among nearly degenerate modes. However, the degeneracy of the C=O stretching transitions requires a relatively large number of the trial vectors (~ 1000) to be averaged, which leads to longer computational time.

Another limitation of the present approach is the restriction to the harmonic force field, although diagonal anharmonic effects and Fermi resonances might be important for many spectroscopic phenomena.^{21,31} In general, the wave function propagation and Fourier transformation of an autocorrelation function is suitable for obtaining transition energies of any Hamiltonian;³² the implementation for large molecules, however, exceeds the scope of the present work.

4. Summary

The force-field mediated time propagation of trial vectors proved to be a stable algorithm for the generation of vibrational spectra. It can offer important computational advantages over standard matrix diagonalization methods for very large systems. The method yielded accurate IR and VCD intensities for model systems, provided that the time propagation was averaged over a sufficient number of the trial initial vectors. As a drawback, the procedure provides

only approximate absolute spectral intensities. However, this is seldom a problem in most applications, since the absolute IR and VCD intensities are difficult to measure, and the structural properties are deduced mostly from the spectral shapes. Because of the modest computer time and memory requirements the time propagation thus represents a convenient means for modeling of vibrational properties of large molecular systems.

Acknowledgment. The work was supported by Faculty Grant-in-Aid and Basic Research Grant programs of the University of Wyoming, by the Grant Agency of the Czech Republic (grants 203/06/0420, 202/07/0732), and the Grant Agency of the Academy of Sciences (A400550702).

References

- (1) (a) Gregoriou, V. G.; Braiman, M. S. *Vibrational spectroscopy of biological and polymeric materials*; CRC: Boca Raton, FL, 2005. (b) Gremlich, H. U.; Yan, B. *Infrared and Raman Spectroscopy of Biological Materials*; CRC: Boca Raton, FL, 2005.
- (2) (a) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35*, 369–430. (b) Haris, P. I. Fourier Transform Infrared Spectroscopic Studies of Peptides: Potentials and Pitfalls. In *Infrared Analysis of Peptides and Proteins: Principles and Applications*. ACS Symposium Series; Ram Singh, B., Ed.; ACS: Washington, DC, 2000; pp 54–95. (c) Braiman, M. S.; Rothschild, K. J. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 541–570.
- (3) Keiderling, T. A.; Kubelka, J.; Hilario, J., Vibrational circular dichroism of biopolymers. Summary of methods and applications. In *Vibrational spectroscopy of polymers and biological systems*; Braiman, M., Gregoriou, V., Eds.; CRC Press: Boca Raton, 2006; pp 253–324.
- (4) Andrushchenko, V.; Bouř, P. *J. Phys. Chem. A* **2007**, *111*, 9714–9723.
- (5) Andrushchenko, V.; Bouř, P. *J. Comput. Chem.* **2008**, *29*, 2693–2703.
- (6) (a) Dluhy, R. A.; Stephens, S. M.; Widayati, S.; Williams, A. D. *Spectrochim. Acta, Part A* **1995**, *51*, 1413–1447. (b) Goormaghtigh, E.; Raussens, V.; Ruysschaert, J. M. *Biochim. Biophys. Acta* **1999**, *1422*, 105–185.
- (7) (a) Barth, A.; Zscherp, C. *FEBS Lett.* **2000**, *477*, 151–156. (b) Pandyrá, A. A.; Yamniuk, A. P.; Andrushchenko, V.; Wieser, H.; Vogel, H. *Biopolymers* **2005**, *79*, 231–237.
- (8) (a) Callender, R.; Dyer, R. B. *Chem. Rev.* **2006**, *106*, 3031–3042. (b) Callender, R. H.; Dyer, B. R.; Gilmanishin, R.; Woodruff, W. H. *Annu. Rev. Phys. Chem.* **1998**, *49*, 173–202.
- (9) (a) Hamm, P.; Lim, M.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123–6138. (b) Zanni, M. T.; Hochstrasser, R. M. *Curr. Opin. Struct. Biol.* **2001**, *11*, 516–522. (c) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.* **2008**, *41*, 432–441.
- (10) Besley, N. A. *Philos. Trans. R. Soc., A* **2007**, *365*, 2799–2812.
- (11) Kaledin, A. L.; Kaledin, M.; Bowman, J. M. *J. Chem. Theory Comput.* **2006**, *2*, 166–174.
- (12) Choi, J. H.; Lee, H.; Lee, K. K.; Hahn, S.; Cho, M. *J. Chem. Phys.* **2007**, *126*, 045102.

- (13) Bouř, P.; Sopková, J.; Bednářová, L.; Maloň, P.; Keiderling, T. A. *J. Comput. Chem.* **1997**, *18*, 646–659.
- (14) Silva, R. A. G. D.; Kubelka, J.; Decatur, S. M.; Bouř, P.; Keiderling, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 8318–8323.
- (15) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in Fortran*, 2nd ed.; Cambridge University Press: New York, 1992.
- (16) Grotendorst, J. *Modern methods and algorithms of quantum chemistry*; John von Neumann Institute for Computing: Jülich, 2000; Vol. 1.
- (17) (a) Davidson, E. R. *J. Comput. Phys.* **1975**, *17*, 87–94. (b) Mitin, A. V. *J. Comput. Chem.* **1994**, *15*, 747–751. (c) Martins, J. L.; Cohen, M. L. *Phys. Rev. B* **1988**, *37*, 6134–6138.
- (18) vanderVorst, H. A. *Iterative Krylov Methods for Large Linear Systems*; Cambridge University Press: Cambridge, 2003.
- (19) (a) Reiher, M.; Neugebauer, J. *J. Chem. Phys.* **2003**, *118*, 1634–1641. (b) Neugebauer, J.; Reiher, M. *J. Phys. Chem. A* **2004**, *108*, 2053–2061.
- (20) Horníček, J.; Kaprálová, P.; Bouř, P. *J. Chem. Phys.* **2007**, *127*, 084502.
- (21) Papoušek, D.; Aliev, M. R. *Molecular Vibrational/Rotational Spectra*; Academia: Prague, 1982.
- (22) Stephens, P. J.; Devlin, F. J.; Ashvar, C. S.; Chabalowski, C. F.; Frisch, M. J. *Faraday Discuss.* **1994**, *99*, 103–119.
- (23) (a) Stephens, P. J. *J. Phys. Chem.* **1985**, *89*, 748–752. (b) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623–11627. (c) Polavarapu, P. L. *Vibrational spectra: principles and applications with emphasis on optical activity*; Elsevier: Amsterdam, 1998; Vol. 85. (d) Nafie, L. A.; Freedman, T. B. Vibrational optical activity theory. In *Circular Dichroism. Principles and Applications*, 2nd ed.; Berova, N., Nakanishi, K., Woody, R. W., Eds.; Wiley-VCH: New York, 2000; pp 97–131.
- (24) (a) Andrushchenko, V.; Wieser, H.; Bouř, P. *J. Phys. Chem. B* **2003**, *91*, 12623–12634. (b) Bouř, P.; Keiderling, T. A. *J. Phys. Chem. B* **2005**, *109*, 23687–23697. (c) Bouř, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* **2002**, *65*, 45–69.
- (25) (a) Bouř, P.; Keiderling, T. A. *J. Chem. Phys.* **2002**, *117*, 4126–4132. (b) Bouř, P. *Collect. Czech. Chem. Commun.* **2005**, *70*, 1315–1340.
- (26) Kubelka, J.; Keiderling, T. A. *J. Am. Chem. Soc.* **2001**, *123*, 6142–6150.
- (27) Becke, A. *Phys. Rev. A* **1988**, *38*, 3098–3100.
- (28) 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, Revision C.02*; Gaussian, Inc.: Wallingford, CT, 2004.
- (29) Pulay, P. *J. Phys. Chem.* **1995**, *99*, 3093–3100.
- (30) (a) Setnička, V.; Huang, R.; Thomas, C. L.; Etienne, M. A.; Kubelka, J.; Hammer, R. P.; Keiderling, T. A. *J. Am. Chem. Soc.* **2005**, *127*, 4992–4993. (b) Torii, H. *J. Phys. Chem. B* **2008**, *112*, 8737–8743.
- (31) Palmer, R. A. 2-D Infrared-Spectroscopy *Chem. Eng. News* 1990; Mar 26, pp 6175–6185.
- (32) Žd'ánská, P.; Moiseyev, N. *J. Chem. Phys.* **2004**, *121*, 6175–6185.

CT800298N