
Computations of the Raman Optical Activity via the Sum-Over-States Expansions

PETR BOUŘ

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

Received 26 July 2000; accepted 25 September 2000

ABSTRACT: A computationally convenient and reasonably accurate scheme of computation of the Raman Optical Activity (ROA) is presented and tested on model examples. Electromagnetic tensors were obtained using the sum-over-states (SOS) methodology, while their nuclear derivatives were estimated through numerical differentiation. An origin dependence of the results was overcome by a distributed origin gauge transformation. Becke-3LYP functional and corresponding Kohn-Sham orbitals are used for the excited states. The method was compared to a benchmark coupled-perturbed (CP) calculation on formamide and a standard ROA spectral simulation and experiment for α -pinene. Spectra of four standard peptide conformations (α -helix, 3_{10} -helix, coil, and β -sheet) were simulated with smaller fragments and compared to previous experimental observations. © 2001 John Wiley & Sons, Inc. *J Comput Chem* 22: 426–435, 2001

Keywords: Raman optical activity; sum over states; *ab initio*; molecular property tensors; peptides

Introduction

Raman optical activity (ROA) has been receiving attention as a convenient probe for conformational studies of biologically interesting systems, especially in aqueous environment.^{1–3} Unlike conventional Raman spectroscopy, it monitors

differences in scattering for left and right circularly polarized, and thus is sensitive to individual enantiomers. Capabilities of ROA spectrometers have been significantly improved in the last three decades⁴ so that interesting molecules related to the living matter can be studied in water and their spectra recorded in a wide range of wavelengths. Such a flexibility often cannot be achieved by the complementary vibrational circular dichroism (VCD) technique.⁵ Finally, advantages with respect to the NMR spectroscopy should be mentioned, namely the intrinsically fast response of the light to confor-

Correspondence to: P. Bouř; e-mail: bour@uochb.cas.cz
Contract/grant sponsor: Grant Agency of the Czech Republic; contract/grant number: 203/97/P002

mational changes and a relatively low experimental cost.

On the other hand, interpretation of ROA is predominantly dependent on *ab initio* simulations and thus limited in size of the systems studied. In addition to the usual limitations of quantum mechanics, an origin dependence of calculated intensities and a lengthy finite-difference computation of the optical activity tensor complicates such calculations.^{6–8}

Many obstacles were overcome in the past. The two group and bond polarizability simplified models enabled first *ab initio* computations of ROA.⁹ The origin dependence was ultimately removed through an introduction of magnetic field dependent atomic orbitals,¹⁰ as was similar for the enumeration of NMR shielding¹¹ or VCD intensities.¹² The polar model⁸ and Cartesian tensor transfer techniques¹³ enabled further extensions towards bigger systems. Nevertheless, accurate computation of ROA intensities is still rather a difficult computational task. For example, ROA spectra for systems with more than about 300 basis functions cannot be routinely calculated with the common Dalton¹⁴ or Cadpac¹⁵ programs due to intrinsic code limits and excessive demands on computer time and memory.

A fast alternative based entirely on sum-over-state (SOS) expansions was proposed to the usual coupled-perturbed (CP) calculations of the tensors.^{16,17} In a limit of exact wave functions, the same results should be obtained by both procedures. However, the fully analytical implementation of the SOS scheme did not provide sufficiently accurate results for ROA.¹⁷ Double sum over excited states and the SOS approximation to nuclear derivatives were not suitable for approximate wave functions that could be obtained from most quantum chemical programs. In this work, modified SOS computational scheme mixed with finite differentiation is proposed, which better balances accuracy with demands on computer resources. As discussed below, it is applicable for most systems studied experimentally, and does not impose special limits on their size.

Theory

ROA spectroscopy exists in many modifications, and complete theory for various experimental settings can be found elsewhere.^{9,18,19} Typically, recorded intensity depends on nuclear derivatives of three molecular property tensors: electric dipole–electric dipole (α), electric dipole–electric quadrupole (\mathbf{A}), and electric dipole–magnetic dipole (“optical activity tensor,” \mathbf{G}') polarizability. They can be

written for a real wave function in a general form as⁹

$$T_{xy} = \sum_{j \neq n} f(\omega_j, \omega_n, \omega) \langle n|x|j \rangle \langle j|y|n \rangle, \quad (1)$$

where x and y are components of electric dipole, electric quadrupole, or imaginary part of the magnetic dipole operators, $|n\rangle$ and $|j\rangle$ are electronic ground and excited states, respectively. The function of energies

$$f = \frac{2\omega_{jn}}{\hbar(\omega_{jn}^2 - \omega^2)}$$

for α and \mathbf{A} , and

$$f = \frac{-2\omega}{\hbar(\omega_{jn}^2 - \omega^2)}$$

for \mathbf{G}' , ω is the laser excitation frequency. The sum over excited states is replaced by a perturbed ground-state function in a CP calculation, based usually on the analytic derivative approach or on the response theory.²⁰ As a simplest example, the static limit ($\omega \rightarrow 0$) of α becomes,

$$\begin{aligned} \alpha_{\alpha\beta} &= \sum_{j \neq n} \omega_{jn}^{-1} \langle n|\mu_\alpha|j \rangle \langle j|\mu_\beta|n \rangle \\ &= 2 \left\langle n|\mu_\alpha| \frac{\partial n}{\partial E_\beta} \right\rangle = \frac{\partial \langle \mu_\alpha \rangle}{\partial E_\beta}, \end{aligned}$$

where μ is electric dipole and \mathbf{E} the electric intensity.

In the SOS scheme, eq. (1) is used directly with simplified excited states. Thus, reasonably accurate polarizabilities and magnetic susceptibilities could be obtained using one-determinantal wave functions.¹⁶ Especially convenient for the SOS computation appears the exploitation of the Kohn–Sham orbital energies, relatively closely related to singlet and triplet excitation energies.^{21–23} Analytical nuclear derivatives of the tensors can also be obtained within the SOS scheme, but the double sums over excited states used in this case lead to an unpleasant drop of accuracy for calculated ROA intensities.¹⁷ Therefore, the dependence on nuclear coordinates is still evaluated by the numerical differentiation here.

In restricted basis sets the SOS results generally become origin dependent because of the transformation properties of the \mathbf{G}' tensor. Unfortunately, the magnetic field-dependent orbitals [known also as the London (LAO) or gauge-invariant atomic orbitals (GIAO)¹¹] cannot be used in the SOS scheme formally lacking an explicit dependence on the magnetic field, unlike for CP. This inconvenience can be

overcome with a distributed origin gauge (DOG) for the tensor \mathbf{G}' :

$$\frac{\partial G'_{\alpha\beta}}{\partial r_\varepsilon^\lambda}(\text{DOG}) = \frac{\partial G'_{\alpha\beta}}{\partial r_\varepsilon^\lambda}(\text{COM}) + \frac{i\omega}{2} \varepsilon_{\beta\gamma\delta} r_\gamma^\lambda \left(\frac{\partial \alpha_{\alpha\delta}^{(v)}}{\partial r_\varepsilon^\lambda} - \frac{\partial \alpha_{\alpha\delta}}{\partial r_\varepsilon^\lambda} \right). \quad (2)$$

The tensor calculated in common origin (COM) corresponds to eq. (1), $\alpha^{(v)}$ is the dipole-dipole polarizability calculated in the mixed-length velocity form, r_ε^λ is ε -coordinate of the nucleus λ . The procedure is analogous to the DOG transformation commonly used in computations of the axial atomic tensor²⁴ needed for VCD intensities. Furthermore, computed by a CP scheme can be conveniently used in eq. (2), as the scheme is more accurate than SOS, and can be performed analytically by many available programs, unlike for the \mathbf{G}' tensor. Using a similar origin transformation, the tensor \mathbf{A} can be improved with better α as

$$\frac{\partial A_{\alpha\beta\gamma}}{\partial r_\varepsilon^\lambda} \rightarrow \frac{\partial A_{\alpha\beta\gamma}}{\partial r_\varepsilon^\lambda}(\text{SOS}) - \frac{3}{2} \left(r_\beta^\lambda \frac{\partial \Delta \alpha_{\alpha\gamma}}{\partial r_\varepsilon^\lambda} + r_\gamma^\lambda \frac{\partial \Delta \alpha_{\alpha\beta}}{\partial r_\varepsilon^\lambda} \right) + r_\delta^\lambda \Delta \alpha_{\alpha\delta} \delta_{\beta\gamma}, \quad (3)$$

where $\Delta \alpha = \alpha(\text{SOS}) - \alpha(\text{CP})$.

Computationally feasible excited electronic states [those used in eq. (1)] were successfully approximated by Kohn–Sham Slater determinants in the past.^{16, 17, 21, 22, 25, 26} In the simplest approximations the energy differences are approximated by

$$\hbar\omega_{jn} = \varepsilon_j - \varepsilon_k \quad (4)$$

or

$$\hbar\omega_{jn} = \varepsilon_j - \varepsilon_k + 2K_{JK} - J_{JK}, \quad (5)$$

where ε_i denotes orbital energies, and K_{JK} and J_{JK} are the usual exchange and Coulomb integrals, respectively.

Computational Aspects

For actual computation of ROA intensities, the Gaussian²⁷ program was used together with home-made “Roa” and other supplementary programs. With Gaussian single-point (SP) energies and orbitals were calculated for each nuclear displacement. The options “nosymm” and “SCF = tight” together with a one-step (0.01 Å) differentiation provided sufficient numerical stability of the results. Tensors α , $\alpha^{(v)}$, \mathbf{A} , and \mathbf{G}' were then calculated for each point using the SOS scheme and, consequently, their nuclear derivative could be obtained numerically. Static limits of the tensors ($\omega \rightarrow 0$) were

considered, as no significant dependence on excitation frequency in the nonresonance region was observed for systems studied.

The SP computations were the slowest step of the procedure, because the time needed scales as N^2 – N^3 for DFT or HF/SCF methods on larger systems, N being the number of atoms. This results in a reasonable scaling of N^3 – N^4 for ROA intensities. Thus, their calculation does not impose a special limit on size of investigated systems, as the same scaling applies for harmonic force fields needed also for simulations of ROA.

The CP computations were done with Gaussian (for α) and Dalton (for α , \mathbf{G}' , and \mathbf{A}) programs. Alternatively, calculation of \mathbf{G}' and \mathbf{A} tensors is implemented in Cadpac. The current version of Dalton, however, was preferred, because it was faster, numerically more stable, and enabled the use of LAO and frequency-dependent computations.¹⁰ From calculated ROA ICP backscattering intensities spectra were simulated using Lorentzian band shapes ($\propto [(x - x_0)^2 + \Delta^2]^{-1}$), with a bandwidth Δ of 5 cm^{-1} . Differential Raman scattering cross sections²⁸ were calculated for a temperature of 295 K and excitation light with a wavelength of 574 nm.

Results and Discussion

FORMAMIDE

The formamide molecule, which can be thought of as a simplistic model for the peptide amide group, was used as a benchmark. Dependence on the basis set size of the SOS results as well as the comparison with CP computations can be seen in Table I. For SOS, single-point energies and orbitals were generated with Becke3-LYP (B3LYP) DFT functional^{29, 30} with the use of eq. (4) for excitation energies. CP results were obtained by Dalton at the HF/SCF level. In the table, three “random” components of each tensor are selected, in an effort to save paper space. All components, however, were compared to the last reference CP HF/AUG-cc-pVTZ calculation via a linear fit $y_i = ax_i$, with a dispersion

$$\delta = \sqrt{n^{-1} \sum_i (ax_i - y_i)^2}.$$

Atomic units were used in the table.

A relatively mild dependence of α on the size of basis set can be observed. The SOS calculations provide the tensor overestimated by almost 100%

TABLE I. Comparison of the Polarizability Tensors of Formamide as Calculated by the SOS and CP Schemes.

Basis	SOS					CP				
	Tensor Components ^a			Fit ^b		Tensor Components ^a			Fit ^b	
				a	δ				a	δ
6-31G										
α	-22.0	6.7	-0.8	1.97	2.0	-7.3	5.1	-0.4	0.94	0.7
G^{COM}	-6.1	-4.1	-0.8	0.66	2.2	-3.7	-1.0	-1.8	0.54	1.4
$G^{\text{DOG/LAO}}$	-9.5	-4.7	-0.8	0.80	2.4	-5.6	-1.9	-2.2	0.71	1.7
A	-45.7	12.9	3.5	1.31	1.22	-16.9	7.2	2.3	0.94	0.7
6-31G**										
α	-23.1	6.1	-0.81	1.91	1.8	-7.1	4.9	-0.4	0.92	0.6
G^{COM}	-7.4	-4.1	-0.5	0.73	2.4	-4.7	-0.8	-1.0	0.59	1.3
$G^{\text{DOG/LAO}}$	-10.2	-4.6	-0.5	0.83	2.6	-5.5	-0.7	-1.2	0.67	1.4
A	-45.6	12.5	2.4	1.11	6.1	-17.6	8.0	0.2	0.70	3.4
6-31++G**										
α	-23.1	5.0	1.6	1.98	1.5	-9.2	4.4	0.8	1.04	0.3
G^{COM}	-7.8	-0.3	1.8	1.00	2.2	-6.4	2.4	1.7	0.90	0.7
$G^{\text{DOG/LAO}}$	-10.6	-0.7	1.5	1.11	2.2	-7.5	2.6	1.5	0.97	0.8
A	-48.1	16.2	-23.0	1.36	4.4	-21.8	11.2	-24.4	1.04	0.3
6-311++G**										
α	-23.4	5.4	1.9	1.97	1.6	-9.3	4.5	1.2	1.03	0.3
G^{COM}	-7.9	0.1	1.5	1.02	2.2	-7.7	3.1	1.1	0.97	0.7
$G^{\text{DOG/LAO}}$	-10.8	-0.3	1.2	1.12	2.2	-6.4	2.8	1.3	0.97	0.7
A	-49.6	16.7	-21.8	1.36	1.15	-23.0	11.8	-24.4	0.93	0.3
AUG-cc-pVTZ										
α	-22.5	4.6	1.0	1.96	1.3	-8.7	3.9	0.5	1	0
G^{COM}	-9.3	4.5	1.7	1.14	2.0	-7.5	7.3	1.0	1	0
$G^{\text{DOG/LAO}}$	-11.8	4.1	1.6	1.23	2.0	-7.4	7.4	1.0	1	0
A	-56.7	9.7	-33.2	1.48	3.5	-29.3	7.5	-32.0	1	1

^a α : xx,yy,zz; $\omega^{-1}G'$ (static limit): xz,yz,zx; A: xxx,yyy,zzz; derivatives with respect to the x-coordinate of the first atom.

^b A linear fit $y = ax$ for all tensor components with respect to the CP/AUG reference values; δ is the mean square deviation from the fit.

A BPW91/6-31G** optimized geometry was used, coordinates in Å: C 0, 0.422121, 0; O 1.211444, 0.234052, 0; N -0.947918, -0.565064, 0; H -0.465609, 1.438620, 0; H -1.939109, -0.351977, 0; H -0.651409, -1.536330, 0.

(c.f. coefficient $a = 1.91, \dots, 1.98$), while the CP results reproduce always at least 92% of the reference computation on average. Despite the high SOS values of the tensor (e.g., the first component,

$$\frac{\partial \alpha_{xx}}{\partial r_x^1},$$

is overestimated by 160% for the AUG-cc-pVTZ basis, see the values of -22.5 and -8.7, signs and relative magnitudes copy the CP basis set dependence reasonably well. This indicates that a reasonable estimate of Raman intensities can be obtained with this SOS scheme, which was confirmed by calculations on bigger systems not included in this

study. However, the intensities are proportional to the square of the tensor components, which magnifies the systematic error.

On the other hand, the second order (with respect to multipole electromagnetic expansion) tensors G' and A are more sensitive to the basis set size than α , and their SOS and CP values are closer. For example, common origin values of G' if obtained in the 6-31G basis reproduce on average 66 and 54% of the reference for SOS and CP computations, respectively. Signs of

$$\frac{\partial G'_{yz}}{\partial r_x^1} \quad \text{and} \quad \frac{\partial G'_{zx}}{\partial r_x^1}$$

components are wrong if calculated with 6-31G or 6-31G** bases for either method. An inclusion of the diffuse functions (++) improves the results remarkably. The quadrupole polarizability behaves similarly as \mathbf{G}' , for example, the

$$\frac{\partial A_{zzx}}{\partial r_x^1}$$

component is obtained as a small positive number if calculated in 6-31G and 6-31G** bases, while in bigger sets approaches a huge negative value of -33.0 . In the limit of infinite basis the SOS components of the tensors will be probably still systematically higher than the CP results, but the two approaches can be considered equivalent for real systems where the smallest bases are the only choice. As expected, the removal of the origin dependence of \mathbf{G}' via LAO (for CP) or DOG [for SOS, see eq. (2)] improves the results namely for the two smallest basis sets.

The SOS calculation presented in Table I can be thought of either as an uncoupled approximation used in perturbed Kohn–Sham equations, or an explicit estimation of the rigorous formula (1) with crude excited states. The former approach can explain the exceptionally good convergence of the \mathbf{G}' tensor, because in this case SOS and CP formulae accidentally become equivalent for pure functionals with electronic density independent on magnetic field. The latter excitation approach “justifies” calculation of the other two tensors α and \mathbf{A} , because, in principle, the accuracy of such an SOS scheme can be improved up to the Schrödinger limit. A close (and still rather unknown) relation among Kohn–Sham and excitation energies^{21–23} further favors the DFT approach in the SOS theories.

As follows from the results in Table I, the values of the tensors are rather overestimated if obtained by the present SOS implementation, and thus probably not suitable for benchmark calculations

of absolute magnitudes. Note, however, that in the scheme described above the SOS contribution to ROA intensities is minor, due to the dominance of the polar part. Its error thus results in an overall smaller relative inaccuracy. Furthermore, the local part (ROA intensity originating in a circular motion of electrons around atoms during their vibration) is consistently included in this approach, unlike in the polar model.

As indicated above, the SOS method should be more economic in the demands for computer resources. Indeed, as shown in Table II, computational times grow more rapidly for the CP method, and are accompanied with a sharp rise of the disk usage. It should be noted, however, that a different implementation of any computational method could lead to quite a different scaling with molecular size.

To explore the dependence of the SOS method on the model of excited electronic states and energies, five models are compared in Table III: B3LYP frozen orbitals and energies [eq. (4)], analogous BPW91³¹ and HF energies, multiplicity-corrected HF orbital energies [eq. (5)], and complete singly excited states (CIS). All calculations were done with 6-31G** basis and compared (c.f. coefficients a and δ) to the same reference as in Table I. Values obtained with B3LYP typically lie between those calculated with the HF and BPW91 functionals. Indeed, the former significantly underestimates tensor values ($a = 0.28$ for \mathbf{G}') while the latter is overestimating them (up to 280% on average for α). This is in agreement with observation for the relation between Kohn–Sham (Hartree–Fock) orbital energies and electronic excitations for these functionals.^{21, 22}

The inclusion of the multiplicity spin correction [eq. (5)] significantly improves the uncorrected HF results, and the use of proper CIS states in eq. (1) further brings the SOS results to the CP limit. However, as these corrections lead to a dramatic increase

TABLE II. Computational Demands for the Calculations in Table I.

Basis	Number of Functions	Time (min)		Disk Space (MB)	
		SOS	CP	SOS	CP
6-31G	33	12	3	10	13
6-31G**	60	29	16	10	56
6-31++G**	75	51	31	10	122
6-311++G**	90	68	53	10	240
AUG-cc-pVTZ	240	1615	2090	34	6684

Times related to one Intel PIII 500 MHz processor; the last calculation was done on a supercomputer with R10000/195 MHz processors of approximately same power.

TABLE III. Comparison of the Polarizability Tensors of Formamide as Calculated with Various SOS Schemes.

Scheme	Tensor Components			a	δ
B3LYP/eq. (4)					
α	-23.1	6.1	-0.81	1.91	1.8
G^{COM}	-7.4	-4.1	-0.5	0.73	2.4
A	-45.6	12.5	2.4	1.11	6.1
BPW91/eq. (4)					
α	-36.8	7.7	-0.9	2.8	2.8
G^{COM}	-15.9	-7.4	0.1	1.3	4.3
A	-64.1	18.5	2.6	1.5	8.3
HF/eq. (4)					
α	-6.4	2.8	-0.5	0.70	0.7
G^{COM}	-3.1	-1.4	-0.1	0.28	0.9
A	-19.0	4.0	1.75	0.30	0.4
HF/eq. (5)					
α	-10.3	4.1	-0.9	1.06	1.1
G^{COM}	-5.5	-2.2	-2.5	0.70	3.7
A	-29.4	8.3	1.2	0.77	4.1
HF/CIS in eq. (1)					
α	-8.1	5.2	-0.5	1.03	0.7
G^{COM}	-4.8	-3.9	-0.8	0.58	1.6
A	-22.0	8.4	1.3	0.78	3.8

6-31G** basis set was used; symbols as in Table I.

of computational time because of the four index integral transformation, the main advantage of the SOS scheme would be lost. Because accuracy of the tensors \mathbf{A} and \mathbf{G}' is restricted primarily by the basis set size for large systems, use of eq. (4) with the B3LYP functional can be considered as a most reasonable model balancing the accuracy and computational cost. This is in agreement with results published elsewhere,^{16, 17, 21, 22, 25, 32} indicating that this and similar hybrid functionals is more suitable for the SOS calculations that pure (without HF exchange) functionals, although the volume of available DFT recipes prevents definite conclusions. However, qualitative differences among various functionals are rather small in this context and thus the ROA/SOS simulation is further documented with B3LYP only.

α -PINENE

To investigate the behavior of the proposed scheme on a bigger system, ROA spectra of α -pinene were simulated with the SOS and CP methods, both in the 6-31G** basis set. Geometry and harmonic force field was obtained at the BPW91/TZ2P, the same as in ref. 33. In Figure 1, simulated spectra

are plotted together with computational times and experiment from ref. 8. Calculated ROA ICP/180 intensities for the experimentally accessible region of frequencies are detailed in Table IV.

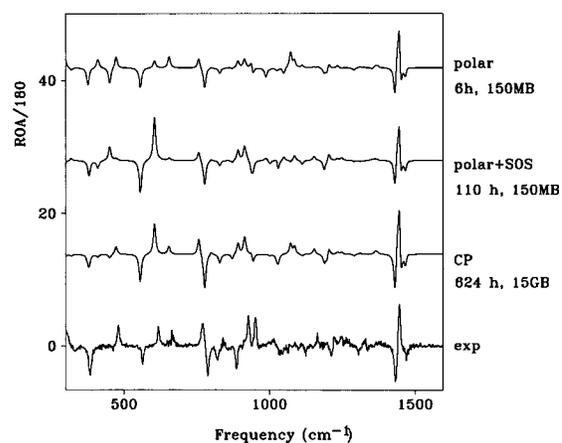


FIGURE 1. ROA backscattering spectra of (1S)-(-)- α -pinene simulated by the polar, SOS, and CP procedures with experimental spectrum from ref. 8. Time and disk space needed for the computations are listed on the right-hand side.

TABLE IV. Calculated ROA Intensities for (S)-(-)- α -Pinene in the mid-IR Region.

ω (cm ⁻¹)	Polar Model	Polar + SOS	CP
1663	-9	-34	-12
1477	11	2	11
1468	-53	-62	-71
1455	-21	-19	-32
1451	-221	-195	-257
1448	338	307	413
1440	84	77	100
1437	0	4	8
1432	-183	-174	-255
1373	11	3	9
1365	10	9	18
1354	-7	5	1
1326	4	-3	0
1312	6	-4	6
1292	-11	-5	-8
1250	-2	13	10
1235	11	8	11
1206	24	34	37
1199	-25	-9	-32
1190	-21	-44	-39
1169	5	2	0
1155	5	19	25
1114	10	-17	-8
1087	27	21	32
1074	61	-7	43
1051	-22	15	2
1031	-2	-40	-30
1026	-9	18	-15
1003	-2	-12	-5
989	-33	10	-1
944	-28	-30	-32
939	25	-29	12
923	-1	6	-2
916	29	47	60
893	22	33	35
873	-3	-8	-16
830	-15	-15	-25
800	3	5	10
777	-53	-65	-95
758	25	24	45
655	24	5	17
605	14	83	60
556	-33	-55	-48
473	16	3	11
451	-20	18	-5
410	10	-8	-3
cc	0.94	0.97	1
a	0.67	0.73	1
δ	22	17	0

Based on a CP/HF/6-31G** and SOS/B3LYP/6-31G** calculations.

Linear fit $y = ax$ and correlation coefficient cc for all of the 72 normal modes.

In agreement with previous observations,^{8,34} the polar model reproduces the experimental pattern remarkably well. Also, on average, about 67% of the CP intensities consist of the polar contribution; see the coefficient a in Table IV. Statistically, the agreement can be improved by addition of the local SOS contribution up to 73%, with better linear correlation and smaller error. Ultimately, the viability of the SOS procedure is even more convincing if simulated spectra are compared to experiment as in Figure 1. The SOS correction improves occasional failures of the polar model namely in the lower frequency region, and provides intensities comparable with the CP calculation. The lower frequency modes (400–655 cm⁻¹) consist namely of skeletal deformation of the bicyclic rings coupled with out of plane bending of the double-bond system. The polar model is not adequate for these modes probably because of π -electron conjugation, partially present also in the four-member ring, enables intense electronic flow under the vibrations.

As indicated in the figure, the same computer space of about 150 MB was required for the polar and SOS model for computation of the polarizability (α) derivatives. About 15 GB of disk space was needed for the Dalton calculation accompanied by a sixfold increase in the CPU time. Given the inaccuracies stemming from the harmonic approximation and experimental noise, the SOS scheme thus appears sufficient and more convenient for interpretation of ROA spectrum of this system.

PEPTIDE SECONDARY STRUCTURES

As mentioned in the introduction, dependence of ROA signal on protein conformation has been extensively studied experimentally. Lately, VCD spectra of α -helical, 3_{10} -helical, coil, and β -sheet conformations were simulated faithfully with smaller peptide fragments.^{35,36} Here, a similar procedure as in ref. 35 can be pursued for ROA, because the method can be applied for fragments of the same size. According to the reference, molecular force fields of model trimer (with respect to the number of amide groups) CH₃—[CONH—CH(CH₃)₂—CONH—CH₃] were calculated at the BPW91/6-31G** level for each of the four conformations. Then CP polarizability (α) derivatives were obtained by Gaussian using the HF/6-311++G** approximation. Finally, the local SOS contributions to the G' and A tensors were computed at the Becke-3LYP/6-31++G** level by the same procedure as for α -pinene. For the last computation with 387 basis functions about 500 h of CPU time and

73 MB of disk space were needed. Obviously, the numerical differentiation can be run on an arbitrary number of processors or computers, up to the total number of geometries. Using the Cartesian tensor transfer technique,^{13,35,36} calculated harmonic force fields and the tensor derivatives were transferred onto corresponding conformers of an octamer $\text{CH}_3\text{—}[\text{CONH—CH}(\text{CH}_3)]_7\text{—CONH—CH}_3$ approximating a longer peptide chain segment with seven alanine residues.

Simulated ROA spectra of the octamer in the four standard peptide conformations are plotted in Figure 2. Captured region corresponds approximately to the range of experimentally accessible frequencies. For example, ROA spectra of β -sheet-rich proteins α -chymotrypsin and β -lactoglobulin were reported³⁷ in a range of about 850–1750 cm^{-1} , spectra of filamentous bacteriophages containing predominantly α -helix² in a range of 550–1850 cm^{-1} .

The variance in intensities of simulated spectra in the figure suggests that the ROA technique is, in principle, able to monitor detailed molecular conformation. In the experiment, scattering on amino acid side chains, hydration, and aggregation obscures the protein backbone signal. Nevertheless several useful conclusions can be drawn based on a comparison of simulated spectra and experimental data from refs. 2 and 38. Calculated frequencies of amide I vibration (C=O stretch, at about 1720 cm^{-1}) are higher than experimental (1600–1680 cm^{-1}). This is, unfortunately, a common feature in *ab initio* simulations of peptides, originating probably in a combination of Hamiltonian and basis set limitations, neglecting anharmonic interactions and solvents effects.^{34–36,38} However, this vibration is virtually uncoupled to other modes and easily

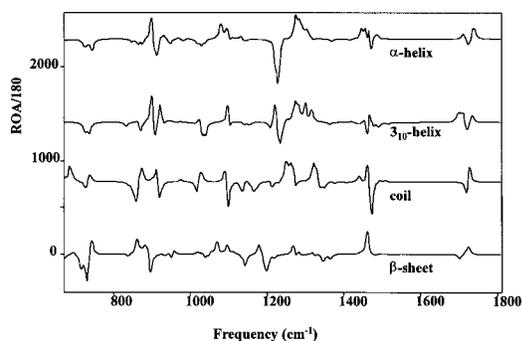


FIGURE 2. Simulated ICP backscattering ROA spectra of $\text{L-CH}_3\text{—}[\text{CONH—CH}(\text{CH}_3)]_7\text{—CONH—CH}_3$ for the four regular peptide conformations. Arbitrary scale on y-axis used.

identifiable in most of infrared spectra. A $+++$ pattern predicted for the α -helix can be seen experimentally, although the relative intensities of these three bands vary up to a $-+$ pattern. The $+++$ pattern is conserved for its closest structure, 3_{10} -helix. 3_{10} -helical ROA spectra, however, have not yet been identified experimentally. Neither was the coil conformation, the structure of which is thought to coincide with polyproline II geometry. A $-+$ pattern is predicted for amide I of the coil and β -sheet. For the latter, this is consistent with the observations for β -sheet containing proteins. Thus, it can be concluded that while the simulated spectral shapes correspond to the experiment for the α -helix and β -sheet, the shapes are similar for all of the four conformations, and thus the amide I region is not convenient for conformational monitoring.

A greater conformational specificity can be observed for the amide II vibrations consisting of C—N stretching and N—H bending vibrations of the amide groups, calculated at about 1470 cm^{-1} , observed around 1455 cm^{-1} . The predominantly negative signal for the α -helix as well as the positive peak for the β -sheet can be clearly identified experimentally. The signal is somewhat obscured by the CH scissoring motion on the methyl groups.

The next region of high ROA intensity is associated with the amide III (C—N stretching and N—H bending) vibrations coupled with the α C—H bending motion. Perhaps the negative band calculated at 1347 cm^{-1} for the β -sheet is related to the negative lobe observed at 1363 cm^{-1} for α -chymotrypsin and β -lactoglobulin, while the low intensity predicted for the α -helix is consistent with the observed absence of signal in this region.

The dominant negative (1228) and positive (at 1277 cm^{-1}) bands can be seen experimentally (at 1244 and 1308 cm^{-1} , respectively) and correspond to delocalized skeletal modes caused by C—N stretching coupled with C—H bending. The $-+$ pattern was found experimentally also for the β -sheets, although a large positive signal in this region was not reproduced. A solitary positive peak calculated at 1180 for the β -sheet can be observed at about 1174 cm^{-1} .

The shape of the ROA signal from the C—C, C—N stretching skeletal modes (1000–1150 cm^{-1}) as well as that from the lower frequency vibrations also tightly depends on the conformation, but cannot be reliably compared to experiment, where it is probably overlapped with contribution from peptide side chains and other groups.

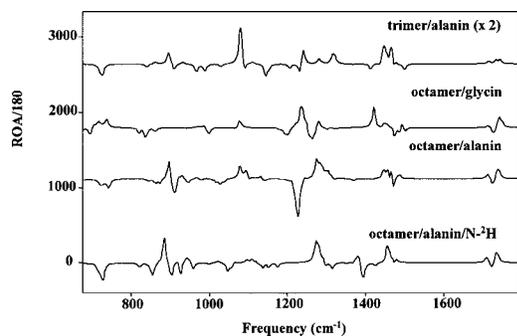


FIGURE 3. Simulated ICP backscattering ROA spectra of L- α -helical peptides. From top to bottom: $\text{CH}_3\text{—}[\text{CONH—CH}(\text{CH}_3)]_2\text{—CONH—CH}_3$ (intensity multiplied twice), $\text{CH}_3\text{—}[\text{CONH—CH}_2]_7\text{—CONH—CH}_3$, and $\text{CH}_3\text{—}[\text{CONH—CH}(\text{CH}_3)]_7\text{—CONH—CH}_3$, and $\text{CH}_3\text{—}[\text{COND—CH}(\text{CH}_3)]_7\text{—COND—CH}_3$. Arbitrary scale on y-axis used.

In Figure 3, ROA spectra of four α -helical fragments are simulated. The similarity of sign patterns between alanine trimer and octamer, with the exception of the amide I region, is in agreement with the generally accepted perception of vibrational optical activity as a short-distance structural probe. Although the role of hydrogen bonds could not be elucidated with the present model, the results suggest their limited influence on spectral intensities. On the other hand, unlike for VCD, contribution of side chains cannot be neglected for ROA, c.f. the glycine and alanine octamer spectra in the figure. However, the peptide skeletal modes (amide I, amide II, C—C, C—N stretch) can still be identified in experimental spectra because of signal additivity and a partial cancellation of “randomly” oriented amino acid residues, especially in globular proteins.

Protein deuteration can serve as a useful tool for spectra interpretation. As is apparent from the lowest spectrum in Figure 2, amide I is almost unaffected by the hydrogen exchange, while the negative signal of amide II (1473 cm^{-1}) disappears, part of the intensity around 1300 cm^{-1} is lost, the negative at 1027 (experimentally at about 1090) cm^{-1} and positive at 1079 (exp 1128 cm^{-1}) bands move upwards to 1047 (exp 1100) and 1097 (exp $1130\text{--}1140$), respectively, all of which was observed in ref. 2. Obviously, many intensity features could not be simulated, and clearly more experimental and theoretical models are needed to elucidate the link between ROA intensities and peptide structure.

Conclusions

The sum-over-state calculation of polarizabilities combined with the numerical differentiation provided a way of a convenient simulation of the Raman optical activity for bigger systems. The benchmark calculations on formamide revealed a similar basis set dependence and reasonable convergence of the CP and SOS methods, sufficient for reliable ROA simulations. An inclusion of the diffuse functions is essential for obtaining accurate quadrupole polarizability and optical activity tensors. Simulated ROA spectra of α -pinene are in an excellent agreement with experiments with inaccuracies stemming rather from frequency than intensity calculations. Theoretical spectra of peptide fragments documented a significant conformational dependence of ROA and many spectral features could be directly compared to available measurements on native proteins.

References

- Bell, A. F.; Hecht, L.; Barron, L. D. *J Am Chem Soc* 1998, 120, 5820.
- Blanch, E. W.; Bell, A. F.; Hecht, L.; Day, L. A.; Barron, L. D. *J Mol Biol* 1999, 290, 1.
- Hanzliková, J.; Praus, P.; Baumruk, V. *J Mol Struct* 1999, 481, 431.
- Buckingham, A. D. *Faraday Discuss* 1994, 99, 1.
- Nafie, L. A.; Qu, X. H.; Lee, E. N.; Yu, G. S.; Freedman, T. B. *Mikrochim Acta* 1997, 14, 807.
- Tam, C. N.; Bouř, P.; Keiderling, T. A. *J Am Chem Soc* 1997, 119, 7061.
- Nafie, L. A. *Vibrational Spectra and Structure*; Durig, J. R., Ed.; Elsevier: Amsterdam, 1981, p. 153.
- Bouř, P.; Baumruk, V.; Hanzliková, J. *Coll Czech Chem Commun* 1997, 9, 1384.
- Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, 1982.
- Helgaker, T.; Ruud, K.; Bak, K. L.; Jørgensen, P.; Olsen, J. *Faraday Disc* 1994, 99, 165.
- Ditchfield, R. *Mol Phys* 1974, 27, 789.
- Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. *Chem Phys Lett* 1996, 252, 211.
- Bouř, P.; Sopková, J.; Bednářová, L.; Maloň, P.; Keiderling, T. A. *J Comput Chem* 1997, 18, 646.
- Helgaker, T.; Jensen, H. J. A.; Jørgensen, P.; Olsen, J.; Ruud, K.; Ågren, H.; Andersen, T.; Bak, K. L.; Bakken, V.; Christiansen, O.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Heiberg, H.; Hettner, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Saue, T.; Taylor, P. R.; Vahtras, O. *Dalton*, Release 1.0, 1997.
- Amos, R. D., Cadpac 5, Darresbury Laboratory: Darresbury, 1990.

16. Bouř, P. *Chem Phys Lett* 1997, 265, 65.
17. Bouř, P. *Chem Phys Lett* 1998, 288, 363.
18. Barron, L. D.; Hecht, L. *Circular Dichroism Principles and Applications*; Nakanishi, K.; Berova, N.; Woody, R. W., Eds.; VCH Publishers: New York, 1994, p. 179.
19. Nafie, L. A.; Che, D. *Modern Nonlinear Optics, Part 3*; Evans, M.; Kielich, S., Eds.; John Wiley & Sons: New York, 1994, p. 105.
20. Colwell, S. M. *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley & Sons: Chichester, 1998, p. 805.
21. Bouř, P. *J Comput Chem* 2000, 21, 8.
22. Bouř, P. *J Phys Chem A* 1999, 103, 5099.
23. Filippi, C.; Umrigar, C. J.; Gonze, X. *J Chem Phys* 1997, 107, 9994.
24. Stephens, P. J. *J Phys Chem* 1987, 91, 1712.
25. Bouř, P.; Buděšínský, M. *Chem Phys* 1999, 110, 2836.
26. Malkin, V. G.; Malkina, O. L.; Eriksson, L. A.; Salahub, D. R. *Modern Density Functional Theory*; Seminario, J. M.; Politzer, P., Eds.; Elsevier: Amsterdam, 1995, p. 273.
27. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzales, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98, Revisions A.3 and A.7*; Gaussian, Inc.: Pittsburgh, PA, 1998.
28. Long, D. A. *Raman Spectroscopy*; Mc Graw-Hill: New York, 1997.
29. Becke, A. D. *J Chem Phys* 1993, 98, 5648.
30. Lee, C.; Yang, W.; Parr, R. G. *Phys Rev B* 1988, 37, 785.
31. Perdew, J. P.; Wang, Y. *Phys Rev B* 1992, 45, 13244.
32. Bouř, P.; McCann, J.; Wieser, H. *J Chem Phys* 1988, 108, 8782.
33. Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. *J Phys Chem A* 1997, 101, 9912.
34. Bouř, P.; Kapitán, J.; Baumruk, V. *J Phys Chem A*, submitted.
35. Bouř, P.; Kubelka, J.; Keiderling, T. A. *Biopolymers* 2000, 53, 380.
36. Silva, R. A. G. D.; Kubelka, J.; Bouř, P.; Decatur, S. M.; Keiderling, T. A. *Proc Natl Acad Sci USA* 2000, 97, 8318.
37. Wen, Z. Q.; Hecht, L.; Barron, L. D. *Protein Sci* 1994, 3, 435.
38. Tam, C. N.; Bouř, P.; Eckert, J.; Trouw, F. R. *J Phys Chem* 1997, 101, 5877.