

Rarefied Basis Sets for the Calculation of Optical Tensors. 1. The Importance of Gradients on Hydrogen Atoms for the Raman Scattering Tensor

G rard Zuber[†] and Werner Hug^{*}

Department of Chemistry, University of Fribourg, Ch. Du Mus e 9, CH-1700 Fribourg, Switzerland

Received: December 4, 2003

A systematic investigation has been undertaken on the basis set requirements for the calculation of optical tensors. In this first of a series of two papers we present a comparison of the performance of various basis sets for the calculation of the Raman and Raman optical activity (ROA) scattering tensor. We show that it is possible to obtain results in excellent qualitative and in reasonable quantitative agreement with those from complex sets by highly rarefied sets provided the vibrational problem is solved in a separate calculation. It is shown that, for hydrogen-containing molecules, a basic requirement for the effectiveness of rarefied basis sets is the proper description of gradients of electronic tensors on these atoms and that this can only be achieved by the inclusion of moderately diffuse p-type functions on these atoms. In addition, the correct rendering of the Raman and ROA tensor for main atom–hydrogen and main atom–deuterium stretching vibrations requires a layer of diffuse functions with valence angular momentum quantum numbers on all atoms. A split valence shell description matters similarly for most other molecular vibrations. The proper description of the core electrons, on the other hand, is of utter unimportance. Questions which arise for rarefied basis sets, such as balance and gauge origin dependence of the optical activity tensor, will be addressed in a second paper.

Introduction

Light scattering has become an important tool in the investigation of molecular structure. Raman spectroscopy¹ and Raman optical activity (ROA) spectroscopy,² in particular, with many of their long-standing experimental problems solved by solid state multichannel techniques^{3,4} and offset elimination schemes,^{5,6} have joined the gamut of the methods routinely applied to the determination of the primary, secondary, and tertiary structure of molecules. Progress in their computation has also been impressive thanks to the steady advance of ab initio methods. Present computational schemes can yield qualitatively correct and quantitatively acceptable results for a property as complex and delicate as Raman optical activity (ROA) provided large enough basis sets are used.^{7–9}

The decisive molecular quantities required in all such calculations are the gradients of electronic molecular property tensors with respect to nuclear displacements. Their analytical calculation is possible for plain Raman scattering but, at present, not for ROA. The calculation of complete ROA spectra for other than the smallest of chiral molecules, with basis sets known to yield reliable tensor gradients, has therefore up to now been out of question.

Our goal was to extend the scope of applicability of the computation of Raman and ROA spectra to chiral molecules of actual practical interest. With this in mind we have carried out a systematic study of the minimal basis set requirements for the determination of those electronic tensors, and of their gradients, which are needed in the calculation of Rayleigh, Rayleigh optical activity, Raman, and Raman optical activity scattering.² The approach has been pragmatic, with the guiding idea not being the highest possible numerical precision but qualitative reliability. The essential compromise, the adoption of which we

found unavoidable, is that the molecular equilibrium geometry and, for Raman scattering, the vibrational frequencies and coordinates have to be determined by an independent calculation.

In the present work we focus on the role which gradients of electronic tensors, with respect to the coordinates of hydrogen nuclei, play in the Raman and ROA scattering of chiral organic molecules and explore the possibility of rendering small basis sets suitable for such calculations by the sole improvement of the description of the electron distribution on hydrogen atoms. Scant attention only is therefore paid to other expectation values. In a subsequent publication we show that a comparatively modest increase in the sophistication of the sets renders likewise possible the reliable computation of other electronic molecular optical properties.

Theoretical Expressions

The differential Rayleigh or Raman scattering cross section $d\sigma$ for an isotropic sample can be written in the form^{10,11}

$$d\sigma = K(c_1 a^2 + c_2 \beta^2) d\Omega \quad (1)$$

For Rayleigh and Raman optical activity one has

$$\Delta d\sigma = \frac{4K}{c} (c_3 a G' + c_4 \beta_G^2 + c_5 \beta_A^2) d\Omega \quad (2)$$

Ω is the solid angle, and the c_i 's are numerical coefficients which depend on the chosen polarization scheme and on the scattering geometry (for actual values, see Calculated Spectra). K is a constant which depends on the frequency ω_0 of the exciting light and, for Raman scattering, also on the frequency ω_p of the scattered light, where p is the vibrational mode for which a transition between different states occurs:

$$K_p = \frac{1}{90} \left(\frac{\mu_0}{4\pi} \right)^2 \omega_p^3 \omega_0 \quad (3)$$

^{*} Author to whom correspondence should be addressed. E-mail: Werner.Hug@unifr.ch.

[†] E-mail: Gerard.Zuber@unifr.ch.

In the Rayleigh case the invariants occurring in (1) and (2) are combinations of elements of real dynamic property tensors and have the form²

$$a^2 = \frac{1}{9} \sum_{\mu, \nu} \alpha_{\mu\mu} \alpha_{\nu\nu} \quad (4a)$$

$$\beta^2 = \frac{1}{2} \sum_{\mu, \nu} [3\alpha_{\mu\nu} \alpha_{\mu\nu} - \alpha_{\mu\mu} \alpha_{\nu\nu}] \quad (4b)$$

$$aG' = \frac{1}{9} \sum_{\mu, \nu} \alpha_{\mu\mu} G'_{\nu\nu} \quad (4c)$$

$$\beta_G^2 = \frac{1}{2} \sum_{\mu, \nu} [3\alpha_{\mu\nu} G'_{\mu\nu} - \alpha_{\mu\mu} G'_{\nu\nu}] \quad (4d)$$

$$\beta_A^2 = \frac{\omega_0}{2} \sum_{\mu, \nu} \alpha_{\mu\nu} \mathcal{L}_{\mu\nu} \quad (4e)$$

a^2 is the isotropic and β^2 the anisotropic invariant of the electric-dipole–electric-dipole polarizability tensor. aG' and β_G^2 are the isotropic and anisotropic invariants due to cross products of this tensor with the electric-dipole–magnetic-dipole polarizability tensor, and β_A^2 is the anisotropic invariant due to its cross product with the tensor $\mathcal{L}_{\mu\nu}$, obtained by contracting the electric-dipole–electric-quadrupole polarizability tensor with the anti-symmetric unit tensor of Levi–Civita.

In the Raman case transition tensors replace the property tensors. Within the Placzek polarizability theory, which we assume in all formulas here, and for a fundamental transition, the various invariants can be written in the general form^{11,12}

$$I_p = \frac{\hbar}{400\pi c \Delta \tilde{\nu}_p} \sum_{\alpha, \beta} \mathbf{L}_{\alpha, p}^x \cdot \mathbf{V}_{\alpha, \beta} \cdot \mathbf{L}_{\beta, p}^x \quad (5)$$

c is the speed of light, \hbar the reduced Planck constant, and $\Delta \tilde{\nu}_p$ is the wavenumber shift in the Raman spectrum (in cm^{-1}). $\mathbf{L}_{\alpha, p}^x$, the Cartesian displacement vector of nucleus α with mass m_α in normal mode p , is normalized so that

$$\sum_{\alpha, i} m_\alpha (\mathbf{L}_{\alpha, i, p}^x)^2 = 1 \quad (6)$$

The elements of the dyadics $\mathbf{V}_{\alpha\beta}$ have, for the various invariants, the form

$$V(a^2)_{ai, \beta j} = \frac{1}{9} \sum_{\mu, \nu} \left(\frac{\partial \alpha_{\mu\mu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial \alpha_{\nu\nu}^e}{\partial x_j^\beta} \right)_0 \quad (7a)$$

$$V(\beta^2)_{ai, \beta j} = \frac{1}{2} \sum_{\mu, \nu} \left[3 \left(\frac{\partial \alpha_{\mu\nu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial \alpha_{\mu\nu}^e}{\partial x_j^\beta} \right)_0 - \left(\frac{\partial \alpha_{\mu\mu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial \alpha_{\nu\nu}^e}{\partial x_j^\beta} \right)_0 \right] \quad (7b)$$

$$V(aG')_{ai, \beta j} = \frac{1}{9} \sum_{\mu, \nu} \left(\frac{\partial \alpha_{\mu\mu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial G'_{\nu\nu}}{\partial x_j^\beta} \right)_0 \quad (7c)$$

$$V(\beta_G^2)_{ai, \beta j} = \frac{1}{2} \sum_{\mu, \nu} \left[3 \left(\frac{\partial \alpha_{\mu\nu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial G'_{\mu\nu}}{\partial x_j^\beta} \right)_0 - \left(\frac{\partial \alpha_{\mu\mu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial G'_{\nu\nu}}{\partial x_j^\beta} \right)_0 \right] \quad (7d)$$

$$V(\beta_A^2)_{ai, \beta j} = \frac{\omega_0}{2} \sum_{\mu, \nu} \left(\frac{\partial \alpha_{\mu\nu}^e}{\partial x_i^\alpha} \right)_0 \left(\frac{\partial \mathcal{L}_{\mu\nu}}{\partial x_j^\beta} \right)_0 \quad (7e)$$

The index 0 indicates that the derivatives are to be taken at the nuclear equilibrium position.

Computational Approach

The method used for the calculation of the optical tensors is the Hartree–Fock linear response theory¹³ as implemented in the DALTON¹⁴ program. There are several reasons for this choice. For one, DALTON appears to be at present the only program which provides a gauge origin invariant implementation of dynamic molecular property tensors. As we will show in the second paper of this series, gauge origin independence of calculated magnetic properties is of utmost importance in the calibration of our rarefied basis sets. For another, even though we rely on density-functional theory (DFT) for the geometry and the determination of vibrational modes, and even though DFT can yield better results than Hartree–Fock theory also for electronic tensors, we judged it unsuitable, in the absence of a detailed investigation of the influence of the choice of the density functional on computed electronic tensors, as the actual calibration method. As a semiempirical method, DFT will unavoidably leave the imprint of the particular density functional used on the optimized set, which is undesirable.

Wherever possible we have judged the quality of the rarefied basis sets by comparing computed electronic tensors with those obtained by more comprehensive reference sets. This avoids the influence of approximations which unavoidably enter the comparison of theoretical with experimental data. The calculations for which results were compared were always done for identical molecular geometries and vibrational modes, obtained by DFT with the B3LYP hybrid functional¹⁵ and a 6-311++G** basis set.¹⁶ The Hessian matrix was calculated with GAUSSIAN¹⁷ and transformed to the DALTON format with an interface routine distributed with DALTON.¹⁸ This procedure guarantees that observed variations are due to a basis set's ability to account for the electronic tensor part, rather than to variations in geometry.

Reference Set and Comparison of Results

An important question concerns the choice of the reference basis set. While it might appear desirable to choose a reference set close to the basis set limit this is, at present, not possible for ROA calculations with DALTON. The best choice therefore is a comparatively modest set but with a known ability to correctly reproduce the sign and semiquantitative size of measured ROA and Raman data. The two moderately sized double- ζ sets which do well in this respect, according to a comparison with the currently best experimental data,⁶ are aug-cc-pVDZ¹⁹ and the basis set of Sadlej.²⁰ Sadlej, in particular, has recently been shown to yield Raman data of a quality close to aug-cc-pVTZ.²¹ The apparently better performance of the somewhat heavier Sadlej, as compared to aug-cc-pVDZ, has decided the issue of the choice of a reference set in its favor.

Another key question is how to judge the suitability for the calculation of Raman and ROA data of a trial relative to the reference basis sets. An obvious way appears to be the comparison of the sums of deviations for all molecular vibrations, either squared or taken absolute, from the values of the reference set; another is the comparison of the average percentage deviation from the values of the reference set, with deviations taken positive. We found both measures not well suited to our purpose. In the former it is relatively small errors in invariants with large values, typically for vibrations of the CH-stretch type, which determine the outcome, while in the latter absolutely small, but in relation to often negligibly small

values of invariants nevertheless large, errors tend to turn the scale, particularly in ROA.

As a first step, to avoid domination by MH- and MD-stretch vibrations we have separated them from the other vibrations of a molecule, with M standing for a main atom. For the two groups of vibrations we have then calculated the expression

$$\delta(I) = \frac{\sum_p |I_p^{\text{trial}} - I_p^{\text{ref}}|}{\sum_p |I_p^{\text{ref}}|} \quad (8)$$

I_p^{ref} and I_p^{trial} stand for the values invariants take for vibration p , calculated with the reference and the trial set, respectively. $\delta(I)$ is a dimensionless quantity representing the normalized average absolute deviation the invariant has for the vibrations of the group. This makes it possible to compare mean values of $\delta(I)$ for different invariants and for different groups of vibrations.

Importance of Basis Functions on Hydrogen Atoms. In our work on the visualization of the generation of Raman and ROA scattering intensities in vibrating molecules,^{11,12} it was found that, for organic molecules which contain hydrogen atoms, the proper description of the diffuse part of the electron distribution in the vicinity of the hydrogen nuclei is the single most important ingredient for the qualitatively correct rendering of Raman and ROA spectra. In all such molecules, it is the hydrogen nuclei which move most, not just in those vibrations traditionally associated with hydrogen motion but in skeletal modes as well, notably because hydrogen atoms have a low mass and, as monovalent atoms, tend to be located at the outskirts of molecules. They are therefore responsible for the generation of much of the Raman and ROA intensity observed in almost all vibrations.

The proper description of gradients of optical tensors with respect to the displacement of the hydrogen nuclei requires the correct rendering of the anisotropy of the diffuse part of the electron distribution in their vicinity. There is no way to achieve this without basis functions with angular momentum quantum numbers higher than zero on hydrogens. In small basis sets such functions are invariably lacking. Medium-sized basis sets may include p-type functions on hydrogens, the so-called polarization functions, but their exponents are optimized according to an energy criterion which makes the functions compact and ineffective in improving the description of the diffuse part of the electron distribution. More diffuse p-type functions on hydrogen atoms are normally only found in rather elaborate basis sets. In general, adding such functions is, moreover, done in conjunction with augmenting the basis set also with diffuse functions with valence angular momentum quantum numbers, which for hydrogen means at least one additional diffuse s-type function. The model case for this way to proceed is the passage from Dunning's cc-pVDZ set²² to the aug-cc-pVDZ set.¹⁹ In other basis sets, such as the well-known Pople sets, even in their largest standard implementations, diffuse p-type functions on hydrogens never occur.²³ As can be seen from the results given below, this is one of their decisive shortcomings with respect to the computation of the Raman scattering tensor of hydrogen-containing molecules.

The procedure of first adding polarization type functions to a valence basis set and, in further steps only, diffuse functions with higher angular momentum quantum numbers leads to far too heavy sets for our purposes. As we are not interested in improving bonding energies or molecular geometries, we can

concentrate on optimizing a single set of p-type functions on hydrogen atoms in order to achieve the best possible description of gradients of optical tensors on them. As shown in the following sections, this leads to highly efficient basis sets for the calculation of Raman and ROA spectra of organic molecules.

Structure of Suitable Sets. An important consideration in a scheme where functions are added to, or removed from, an existing set is the structure of the starting set. We found starting sets of the split valence type to be the best choice if diffuse p-type functions on hydrogen atoms only are added. For sets without a split valence description acceptable results could only be obtained if diffuse functions with valence angular momentum quantum numbers were also added onto all atoms. The best results by far, however, were obtained by adding moderately diffuse p-type functions on hydrogen atoms as well as diffuse functions with valence angular momentum quantum numbers on all atoms to a split valence type set. These observations point to the requirement that the charge distribution be able to properly "breathe" upon nuclear motion. The detailed description of the core electrons, on the other hand, turned out to be unimportant.

Figure 1 illustrates this for (*R*)-(+)-methyloxirane by comparing the values of Raman and ROA tensor invariants calculated with four trial basis sets with those obtained by the basis set of Sadlej.²⁰ The trial sets rDP:3-21G and rDP:STO-3G were obtained by augmenting the split valence 3-21G set²⁴ and the minimal, nonsplit valence STO-3G set²⁵ with diffuse p-type functions on hydrogen atoms and the trial sets rDPS:3-21G and rDPS:STO-3G by adding diffuse p-type functions on hydrogen atoms to the 3-21++G²⁶ and the STO-3++G set, the latter related to STO-3G in the same way as 3-21++G is to 3-21G. DP stands for diffuse polarization function augmented, and DPS for diffuse polarization function and shell augmented. The letter r, for "reduced", is added to distinguish these sets from larger ones to be presented in further work. The two sets obtained from 3-21G, namely, rDPS:3-21G and rDP:3-21G, are in the following also simply designated as rDPS and rDP, respectively. The cc-pVDZ and the aug-cc-pVDZ set were included in Figure 1 for the sake of comparison.

A remarkable aspect of Figure 1 is the almost quantitative agreement between the values calculated by the basis set of Sadlej and the aug-cc-pVDZ set, though the detailed structure of the two sets is quite different. An exception is aG' for CH-stretch vibrations. The cc-pVDZ set on the other hand performs badly, particularly for vibrations other than CH-stretch. The two invariants β_G^2 and β_A^2 which determine ROA backscattering are particularly poorly rendered by it.

There are clear differences in the behavior of the basis sets not only for different invariants but also for CH-stretch and other vibrations. For CH-stretch vibrations, higher level results can clearly more easily be reproduced with simplified basis sets than is the case for other vibrations. This does not mean that the computation of Raman spectra for CH-stretch vibrations is easier in general as anharmonicity, which is not considered here, can play a more important role for them. For CH-stretch vibrations rDPS:STO-3G outperforms the 3-21G-based rDP set by a good margin, while for other vibrations it is the other way around. As both sets contain the same number of contractions one might conclude that, in addition to moderately diffuse p-type functions on hydrogen atoms, truly diffuse functions are mostly important for CH-stretch and possibly other pure stretch vibrations, while a split valence description is necessary for other vibrations.

The clear winner in Figure 1 is the rDPS basis set derived from 3-21G. It performs well for all vibrations, in agreement with the above reasoning, combining a split valence description, moderately diffuse p-type functions on hydrogen atoms, and

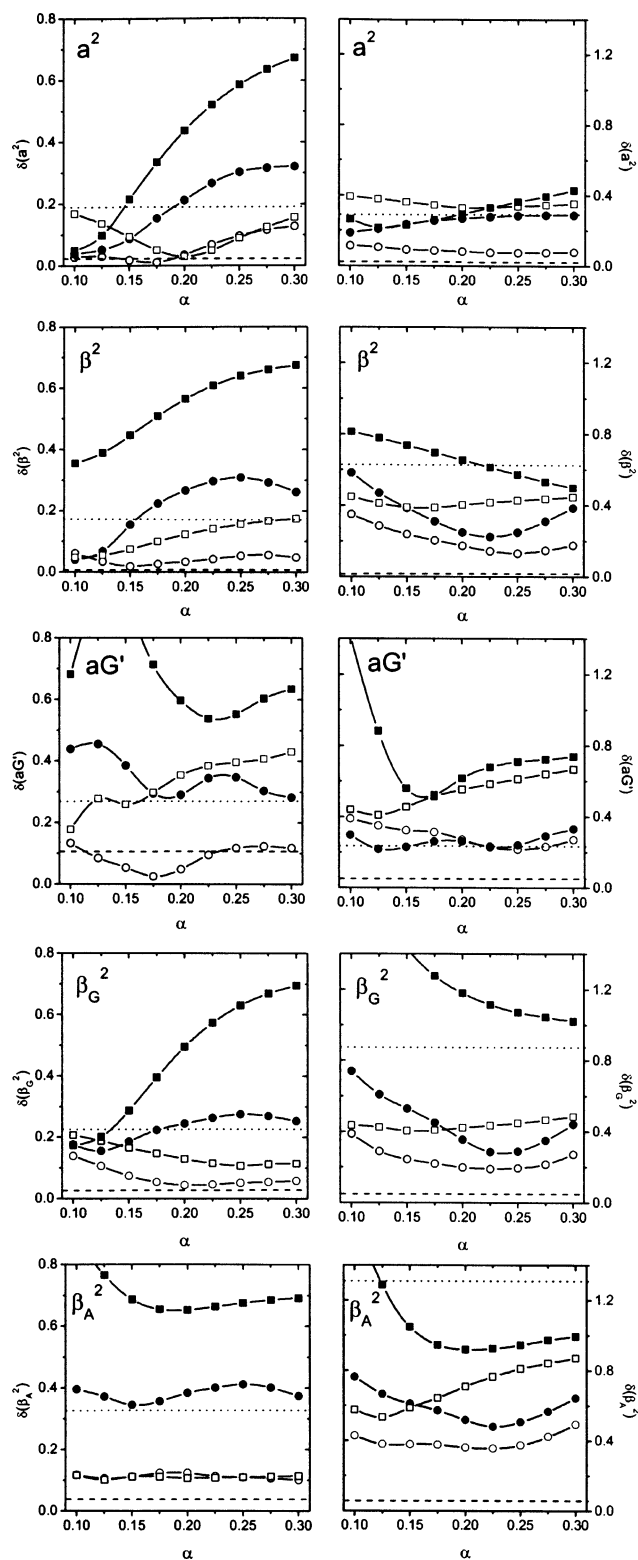


Figure 1. Normalized average absolute difference according to eq 8 between values calculated with trial sets and with the set of Sadlej for Raman and ROA invariants, as a function of the value of the exponent α of the diffuse p-type functions on hydrogen atoms. The CH-stretch vibrations of (*R*)-(+)-methyloxirane are shown on the left, the remainder of the molecules' vibrations on the right. \circ : rDPS:3-21G. \bullet : rDP:3-21G. \square : rDPS:STO-3G. \blacksquare : rDP:STO-3G. Dotted lines: cc-pVDZ. Dashed lines: aug-cc-pVDZ. The geometry and the vibrational coordinates were calculated at the B3LYP/6-311++G** level of theory and are identical for all electronic tensor calculations. Electronic tensors have been calculated for a wavelength of the incident light of 532 nm, and a step length of 0.001 bohr has been used in their numerical differentiation.

truly diffuse functions with valence orbital angular momentum quantum numbers on all atoms. From Figure 1 one also sees that a compromise value of the exponent α of the diffuse p-type functions on hydrogen atoms exists for rDPS, at least for the molecule methyloxirane, somewhere between 0.15 and 0.25, which simultaneously yields good results for all invariants and all vibrations. For the simpler sets, and in particular for rDP:STO-3G, the simplest one, this is less obvious. rDP:STO-3G was therefore dropped altogether from further consideration, and we have in the following concentrated on optimizing the value of α for the 3-21G-based rDPS basis set, with rDP retained for the sake of comparison.

For rDPS, no systematic attempt was made to optimize the diffuse s- and p-type functions on main atoms and the diffuse s-type functions on hydrogen atoms. Pilot calculations showed that little can be gained by this. Their values therefore correspond to those of the 3-21++G set where they have been optimized for the calculation of negative ions.²⁶ We note that, while not identical, the diffuse layer of functions with valence angular momentum quantum numbers of 3-21++G is actually quite similar to that found in aug-cc-pVDZ with its known ability to account for electronic molecular property tensors. The diffuse s- and p-type functions on carbon atoms, for example, have the exponents 0.0469 and 0.0404 in aug-cc-pVDZ and 0.0438 in 3-21++G, while for the diffuse s-type functions on hydrogens one has 0.0297 and 0.0360, respectively, with somewhat more differing values for other atoms.

Choice of the Exponent α . The molecules used in determining an optimum value of the exponent α are shown in Chart 1. A problem is the size of chiral molecules, not for the calculations with the trial sets but for those which need to be done with the reference set of Sadlej because DALTON is at present limited to 255 contractions if integrals are written to disk. A larger number of contractions can, in principle, be used if the integrals are calculated directly, but we found that this leads to computational times beyond reasonable limits. It is for this reason that a number of molecules chiral by isotope substitution only, and also achiral molecules for which only the Raman invariants a^2 and β^2 can be calculated, have been included in the set.

The dependence of $\delta(I)$, eq 8, on the exponent α , averaged for all invariants and all molecules, provides little guidance for the choice of α because it is rather weak in the range represented in Figure 1, with only a shallow minimum for $\alpha = 0.16$ for MH- and MD-stretch vibrations, and $\alpha = 0.26$ for other vibrations.

Linear correlation analysis for individual invariants provides more help for choosing α because it can give insight into whether values calculated with two different methods differ in a systematic or random way. A deviation in size, if it is similar for different invariants, is of little concern in Raman spectroscopy where absolute values for scattering cross sections are rarely ever measured. A compromise value of the exponent α which minimizes the difference in the slope of different invariants, rather than aiming at a slope of 1 by all means, and which minimizes the random scatter in a linear correlation plot can therefore be preferable to one which simply reduces the sum of deviations.

In Figure 2 the values obtained with $\alpha = 0.2$ for rDPS are plotted against those of the basis set of Sadlej. For MH- and MD-stretch vibrations linear correlation is generally excellent. A few substantial deviations occur for aG' and some lesser ones for a^2 . For other vibrations the picture is less favorable, with a number of values found far off the 45° line. They stem from those compounds in Chart 1 which contain aromatic bonds, CC

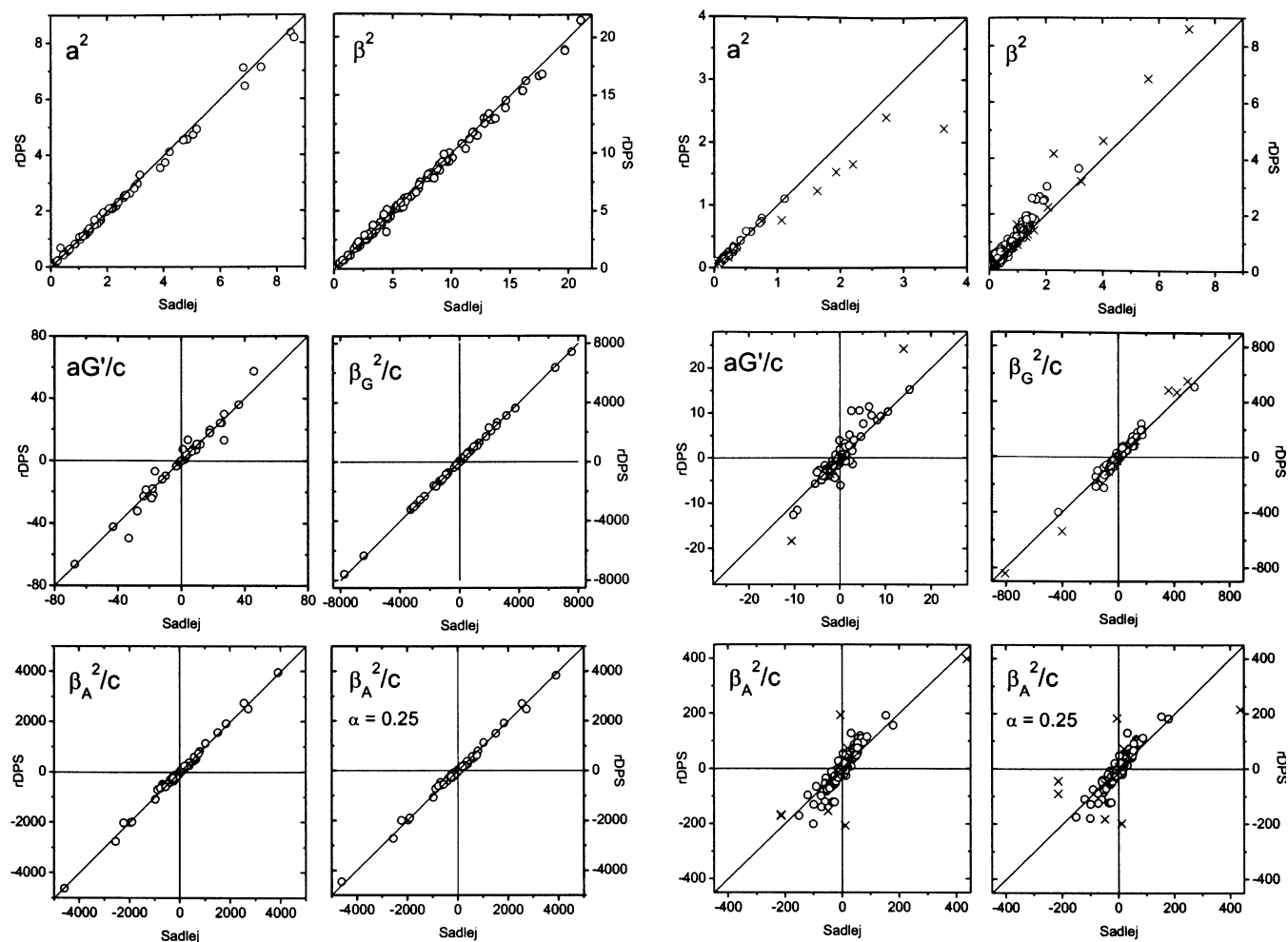


Figure 2. Correlation of values of Raman and ROA invariants for individual MH- and MD-stretch vibrations (left graphs) and for other vibrations (right graphs) of the molecules of Chart 1 calculated with rDPS ($\alpha = 0.2$, except where indicated otherwise) and the set of Sadlej. The straight line is not a fit but rather represents the locus for perfect correlation, if the values from rDPS and those from the set of Sadlej were coinciding. The units of the axes are $\text{\AA}^4/\text{amu}$ for the Raman tensors and $10^{-6} \text{\AA}^4/\text{amu}$ for the ROA tensors, with 1 amu equal to $1/12$ the mass of ^{12}C .

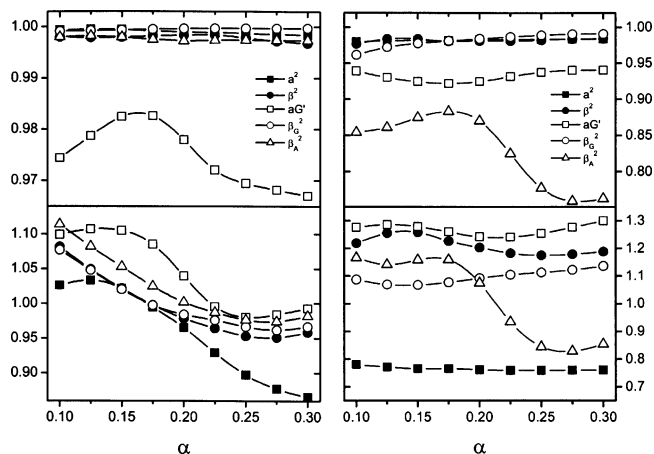


Figure 3. Correlation coefficient r (top) and slope (bottom) for MH- and MD-stretch vibrations (left) and for other vibrations (right), as a function of the exponent α of the diffuse p-type functions on hydrogen atoms, of a linear correlation analysis of values of rDPS against those of the set of Sadlej, for Raman and ROA invariants of the molecules of Chart 1.

double bonds, or CC triple bonds. Values due to these compounds are marked with a cross instead of a circle in Figure 2.

The correlation coefficient r and the slope obtained from a least-squares fit of a straight line to rDPS values plotted against

those of Sadlej are shown in Figure 3 as a function of α . For the MH- and MD-stretch vibrations correlation is high for all values of α . Even for aG' , where it is lowest, r remains close to 0.98. The variation of the slope is more pronounced and extends from 1.1 downward to 0.85. Figure 3 suggests a compromise value of 0.2 for α . The slope of all invariants is then close to 1.

For other vibrations, as seen from Figure 3, the picture is more varied. The slope is generally too high except for a^2 where it is too low everywhere and for β_A^2 where it varies from too high for $\alpha < 0.2$ to too low for $\alpha > 0.225$. Correlation is also far less satisfactory than for MH- and MD-stretch vibrations. For β_A^2 , in particular, it degrades to unacceptably low levels for $\alpha > 0.2$. This is due to 1,3-dideuterioallene, the only chiral molecule of Chart 1 with multiple CC bonds.

The slope of β_A^2 in Figure 3 yields a further argument for the choice of α in the vicinity of 0.2 because it is then similar to that of β_G^2 for this value. This matters because the balance of the two invariants determines the sign of ROA in backscattering and in depolarized right-angle scattering.

From the foregoing discussion it is obvious that the correlation coefficient r and the slope of β_A^2 are decisively influenced by the inclusion of 1,3-dideuterioallene in the correlation analysis. While 1,3-dideuterioallene, with two cumulated double bonds and no hydrogens on the central carbon atom, might be con-

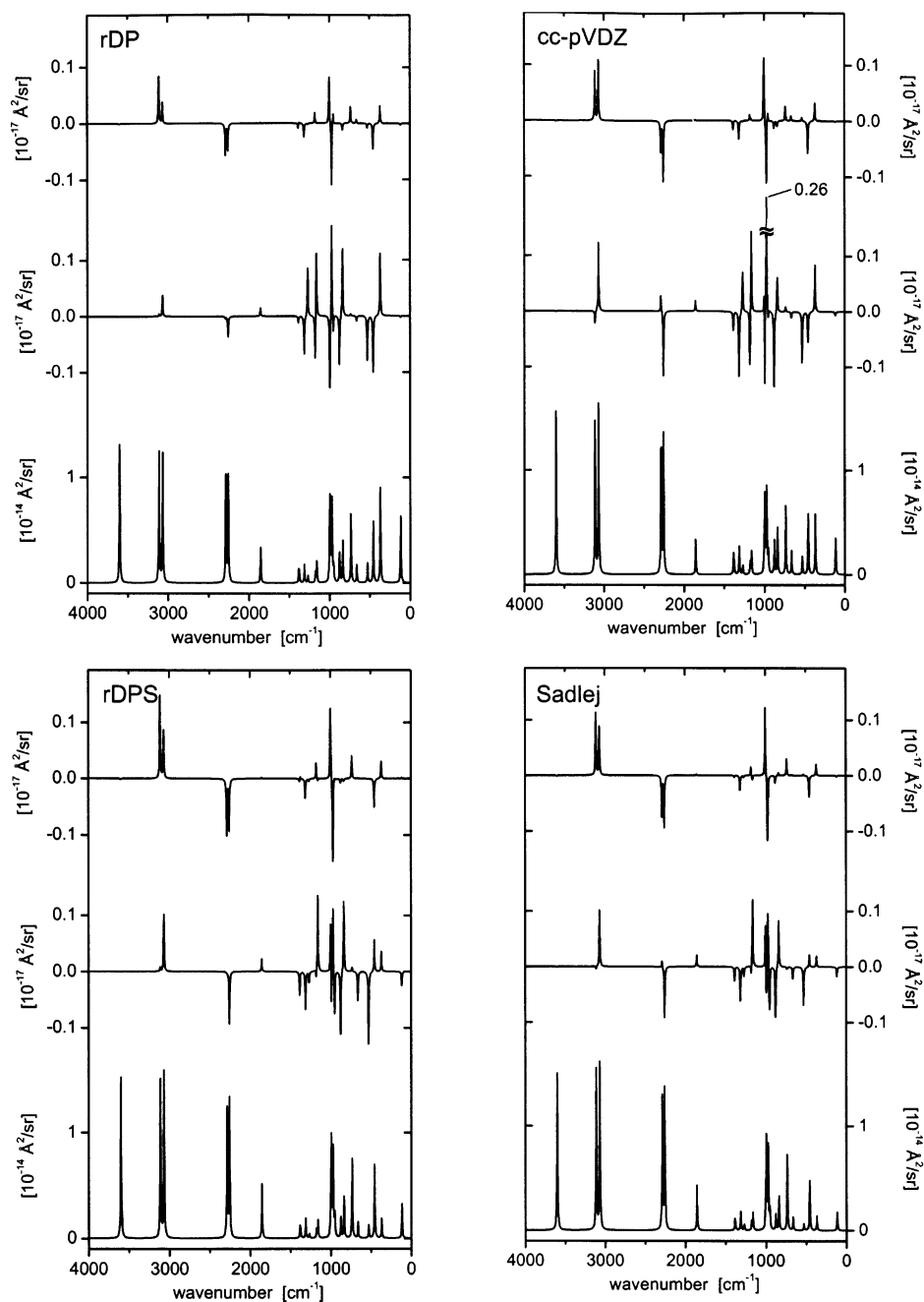


Figure 4. Computed spectra of (3*S*,4*R*)-dideuterio-azetidin-2-one. In each frame the top spectrum represents ROA in forward scattering, the middle spectrum ROA in backward scattering, and the bottom spectrum Raman scattering for both geometries. The spectra were obtained from theoretical data by assuming a 10 cm⁻¹ fwhm Lorentzian band shape (see text for details). A value of 0.2 was used for the exponent α of the diffuse p-type functions on hydrogen atoms in the calculations with rDPS and rDP.

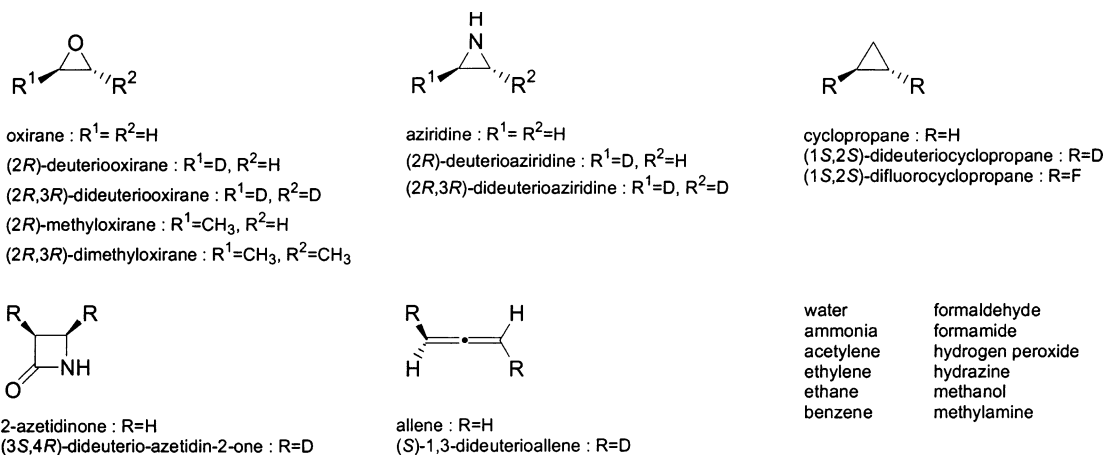
sidered an extreme case, the similarly diverging behavior also observed for a^2 and β^2 for certain vibrations of benzene, ethylene, acetylene, and allene points to a basic weakness of the rDPS basis set in properly describing gradients of electronic tensors of unsaturated hydrocarbons. The vibrations for which the rendering of the Raman invariants suffers most are the ring breathing mode of benzene and the CC multiple bond stretching vibrations of the other compounds, while for the ROA invariants the same role is played by the C=C=C bending and =CHD deformation modes of 1,3-dideuterioallene.

Removing unsaturated hydrocarbons from the analysis improves correlation and slopes and also suggests an optimal value of the exponent α somewhat higher than 0.2, namely, between 0.225 and 0.25, but without that such a higher value leads to a decisive improvement of the calculated data of the compounds

retained in the analysis. In view of the advantage lower values entail for unsaturated compounds, we therefore have settled for a compromise value $\alpha = 0.2$.

The composition and exponents of the rDPS set are shown in Table 1.

Calculated Spectra. From a practical point of view the usefulness of a basis set rests with its ability to provide, via the comparison of measured and calculated data, information on a molecule's structure, on its absolute configuration, and on its conformations. This ability can only partly be judged from knowing how well individual invariants are reproduced because it is combinations of invariants which determine actual spectra. In ROA, in particular, not only can the linear combination of invariants relevant to a particular experimental configuration show the wrong sign even if all signs of individual ROA

CHART 1: List of Molecules Used for the Calibration of the Exponent α of the Diffuse p-Type Functions on Hydrogen Atoms**TABLE 1: Values of the Exponents and Contraction Coefficients of the 3-21G-Based rDPS Set for Hydrogen^a**

CGTO	α_i	c_i
	s Subset	
1	5.447 178	0.156 285
	0.824 547	0.904 691
2	0.183 192	1.000 000
3	0.036000	1.000 000
	p Subset	
1	0.200 000	1.000 000

^a The rDPS set is obtained by removing the diffuse s- and p-type functions on main atoms and the diffuse s-type functions on hydrogen atoms.

TABLE 2: Composition of Some Basis Sets and Comparison of Relative CPU Times for the Electronic Tensor Part of the 3,4-Dideuterioazetidin-2-one with HF Linear Response Theory and the DALTON Program^a

basis set	primitives	contractions	relative CPU time
rDP:STO-3G	105	45	1
rDP	105	70	2.3
rDPS:STO-3G	130	70	2.9
rDPS	130	95	5.8
cc-pVDZ	165	95	12
aug-cc-pVDZ	230	160	70
Sadlej	330	165	130

^a All calculations were done on LINUX stations equipped with Athlon 1800+ processors.

invariants are calculated correctly, but cancellation between close-lying vibrations with opposite signs can also strongly affect the appearance of calculated spectra.

A preferred experimental arrangement for analytical purposes in Raman and ROA is backscattering.^{27,28} The coefficients in eqs 1 and 2 for this geometry and the scattered circular polarization (SCP) configuration are¹¹ $c_1 = 90$, $c_2 = 14$, $c_3 = 0$, $c_4 = 12$, and $c_5 = 4$. An interesting ROA configuration is also forward scattering,²⁹ with $c_3 = 90$, $c_4 = 2$, and $c_5 = -2$, with c_1 and c_2 unchanged.

Figure 4 and Figure 6 show calculated spectra for both scattering geometries, for two chiral molecules with electronic structures which pose different demands on basis sets. The spectra were simulated by Lorentzian band shapes with a 10 cm^{-1} full width at half-maximum height (fwhm), with the height of individual Lorentzians, before superposition, corresponding to the differential scattering cross sections as calculated with eqs 1 and 2. This is admittedly a crude way to simulate Raman spectra, where isotropic and anisotropic components have differing and often variable bandwidths, but it is sufficient for

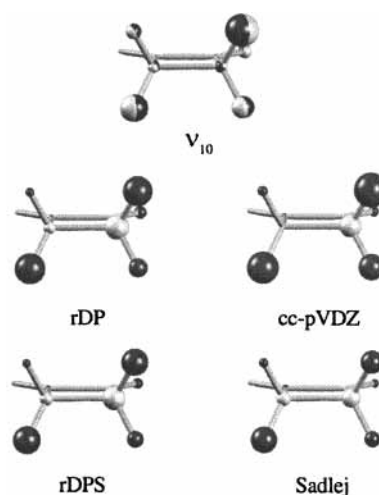


Figure 5. The atomic contribution pattern (ACP) of the backscattering ROA of vibration 10 of (3*S,4R*)-dideuterioazetidin-2-one. The shape of vibration 10, which is calculated to appear at 971 cm^{-1} at the B3LYP/6-311++G** level of theory, is the same for all ACPs. Dark gray is used for positive and light gray for negative contributions.

the present purpose which is less the comparison of calculated and measured data than the comparison of theoretical spectra obtained with different basis sets.

All spectra were also calculated with aug-cc-pVDZ. The difference between the spectra obtained with aug-cc-pVDZ and the set of Sadlej is minor, however, and only the spectra calculated with the set of Sadlej are shown.

Computational times are compared in Table 2 for 3,4-dideuterioazetidin-2-one. The set of Sadlej is slowest, followed by aug-cc-pVDZ. While the nonaugmented cc-pVDZ set is 10 times faster for this molecule than the set of Sadlej, it is still 4 times slower than rDP. This last set yields Raman and ROA spectra which resemble in many respects those calculated with cc-pVDZ, particularly for vibrations below 2000 cm^{-1} . As is the case for cc-pVDZ, several bands show the wrong sign in backscattering. In contrast to cc-pVDZ, however, no single band in Figure 4 is calculated so completely off the mark with rDP as is the 971 cm^{-1} ROA backscattering band due to vibration 10 with cc-pVDZ. The reason for the complete failure of cc-pVDZ to account for the ROA of this particular vibration is not immediately obvious. The vibrational motion, which is graphically represented in Figure 5, exhibits no particular characteristics which would distinguish it from other vibrations with a similar frequency.

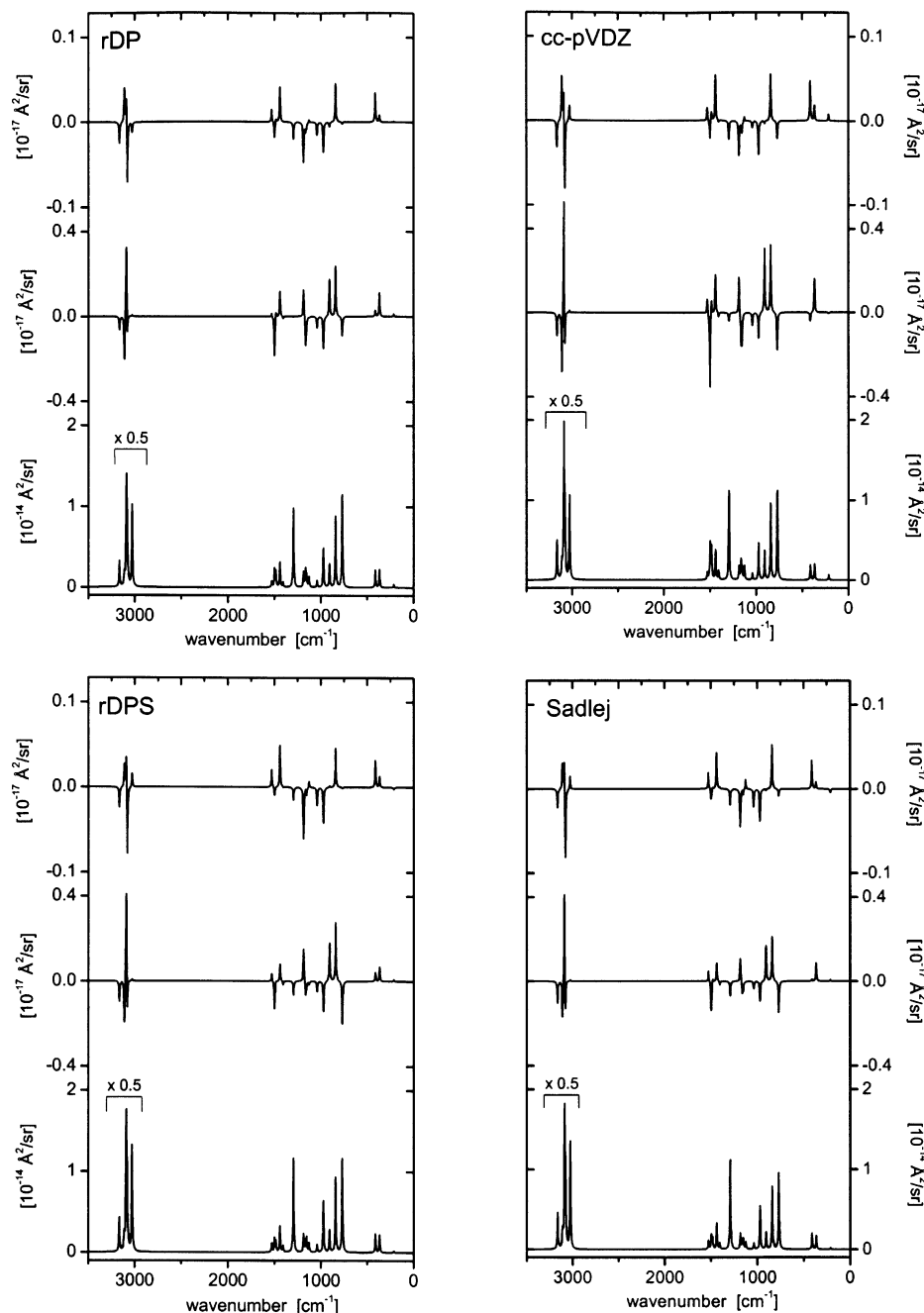


Figure 6. Computed spectra of (*R*)-(+)-methyloxirane. Computational details are as described in Figures 4 and 5. Raman intensities above 2000 cm^{-1} were multiplied by a factor 0.5.

TABLE 3: Calculated Frequencies and $\Delta(\pi)$ -Values Used in Generating the Theoretical Spectra of Figure 7^a

ν	freq [cm^{-1}]	$\Delta(\pi)$ Sadlej	$\Delta(\pi)$ rDPS	$\Delta(\pi/2)$ Sadlej	$\Delta(\pi/2)$ rDPS	ν	freq [cm^{-1}]	$\Delta(\pi)$ Sadlej	$\Delta(\pi)$ rDPS	$\Delta(\pi/2)$ Sadlej	$\Delta(\pi/2)$ rDPS
18	1528	0.46	0.34	0.59	0.68	9	1125	-0.02	-0.13	-0.14	-0.24
17	1497	-0.80	-0.78	-0.39	-0.38	8	1039	-0.45	-0.51	-0.24	-0.16
16	1483	0.16	0.08	0.09	0.03	7	971	-0.17	-0.22	-0.33	-0.36
15	1439	0.27	0.25	0.24	0.23	6	906	0.79	0.65	0.27	0.22
14	1405	-0.23	-0.24	-0.16	-0.21	5	840	0.27	0.30	0.16	0.14
13	1293	-0.06	-0.06	-0.10	-0.10	4	768	-0.16	-0.17	-0.03	-0.04
12	1184	0.57	0.70	1.18	1.50	3	411	0.05	0.18	0.09	0.23
11	1162	-0.50	-0.54	-0.22	-0.16	2	365	0.51	0.31	0.29	0.18
10	1151	-0.34	-0.15	-0.16	-0.08	1	212	0.39	0.45	0.30	0.31

^a Also included are $\Delta(\pi/2)$ -values obtained with RDPS, which refer to depolarized right-angle scattering. Vibrational frequencies and coordinates: DFT with B3LYP functional and 6-311++G** basis set.

The atomic contribution pattern,¹² or ACP, which combines nuclear motion and the gradients of the electronic tensors into a representation of the contributions individual atoms make to

invariants and scattering cross sections, provides more insight into why ROA computed with cc-pVDZ is so far off the mark for this vibration. From Figure 5 one sees that cc-pVDZ

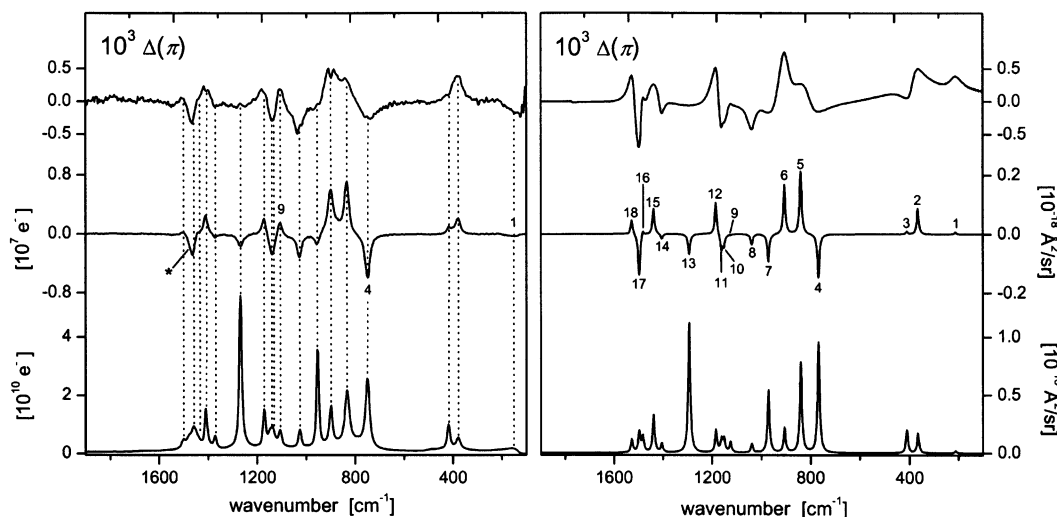


Figure 7. Comparison of the measured (left) and the computed (right) backscattering ROA and Raman spectra of (*R*)-(+)-methyloxirane. The dotted lines in the graph with the experimental spectra correspond to the peak height of the Raman bands. The experimental $\Delta(\pi)$ -curve was obtained by direct division of the ROA through the Raman spectrum, without any baseline correction. Instrumental resolution (fwhm): 7 cm^{-1} . Exciting wavelength: 532 nm. Value for the fwhm used for the computed spectrum: 10 cm^{-1} . Basis set for the electronic tensor calculation: Sadlej. Vibrational coordinates: B3LYP/6-311++G**. The star marks an overtone.

overestimates, in comparison to the set of Sadlej, by far the positive contributions hydrogen and deuterium atoms make. As nuclear motion is calculated with B3LYP/6-311++G** and is therefore the same for all ACPs in Figure 5, this must be due to a misrepresentation of gradients of electronic tensors on hydrogen atoms. The reason is the lack of moderately diffuse p-type functions on these atoms.

For most vibrations the consequences of the misrepresentation of gradients on hydrogen atoms are less obvious than for the 971 cm^{-1} vibration, as opposing contributions tend to lead to spurious signs, rather than to exceptionally large calculated ROA.

The ACP from rDPS is in best agreement with the basis set of Sadlej. While rDPS is over 20 times faster it still yields, as is evident from Figure 4, Raman and ROA spectra of a comparable quality. There are some minor differences, notably in ROA backscattering for the small bands due to the in-phase coupling of the two CH- and of the two CD-stretch motions, which occur at 3112 and 2286 cm^{-1} , respectively. There is no major failure, however, with rDPS for any of the important ROA bands.

The experimental data of the (3*S*,4*R*)-dideuterioazetidin-2-one are not known. If they were, and if it were possible to determine the molecules' absolute configuration by a comparison with the backscattering ROA spectrum calculated with the basis set of Sadlej, then this would clearly also be possible with rDPS. In contrast, one would probably not want to make a firm statement on an unknown absolute configuration based on the backscattering spectrum calculated with either cc-pVDZ or rDP. Surprisingly, though, such a statement would appear justified from a comparison of forward scattering spectra, which only occasionally have been measured up to now.⁸⁵

As illustrated by Figure 6, the computation of the electronic part is a far easier task for (*R*)-(+)-methyloxirane than for (3*S*,4*R*)-dideuterioazetidin-2-one. Again, rDPS comes closest by far to the spectra calculated with the set of Sadlej. cc-pVDZ yields ROA intensities in backscattering which are too high, and it also produces several sign errors for small bands at low-frequency shifts, in backward as well as forward scattering. rDP leads to a sign error in forward scattering for the symmetric CH_3 -stretch vibration calculated at 3027 cm^{-1} . This is not

unexpected in view of the set's limited performance for CH- and CD-stretch vibration. It also misses the negative backscattering ROA of the prominent Raman band calculated at 1293 cm^{-1} . The vibration is due to a combination of tertiary CH bending and ring breathing. Relative to the basis set of Sadlej, all calculated ROA spectra would, however, easily allow the assignment of the absolute configuration of (*R*)-(+)-methyloxirane, if it were not known.

Comparison with Experimental Data. The recent availability of high-quality, offset-free experimental ROA spectra⁶ for (*R*)-(+)-methyloxirane allows a judgment on the usefulness of the computed data. Figure 7 compares the experimental Raman and ROA spectra of this molecule, measured for the condensed phase in SCP backscattering, with the gas-phase spectra computed with the basis set of Sadlej. We include curves of the ratio Δ of the ROA signal divided by the Raman signal because this dimensionless quantity is independent of instrumental light detection efficiency.

Numerical values of Δ extracted from experimental spectra have been preferred in the past,^{7,9,30} notably because of the limited quality of the experimental data, but we feel that representing the measured curve for Δ has advantages.⁸ As the data of Figure 7 demonstrate, extracting numerical values for individual vibrations from peak heights in measured spectra can actually have its pitfalls. The position of the peaks in the Δ -spectrum is often displaced from that found in the parent Raman and ROA spectra, and sometimes two Δ -peaks occur for a single vibrational band. This is due to a difference of the band shape in the measured ROA and Raman spectra and not to noise, which tends to be low in Δ -spectra where Raman intensity is high, and vice versa. The numerical values underlying the computed Δ -spectrum are given in Table 3. Values for right-angle scattering are also indicated so that a comparison with the values cited in refs 9 and 30 is possible.

General qualitative agreement between calculated and measured spectra is clearly satisfactory. There are three bands, however, which are not rendered correctly by the calculation. One concerns an overtone of vibration 4 found at about 1475 cm^{-1} , marked by a star in Figure 7 and discussed in ref 12. It can obviously not be accounted for in the harmonic approximation used here. For two bands the calculated sign of the

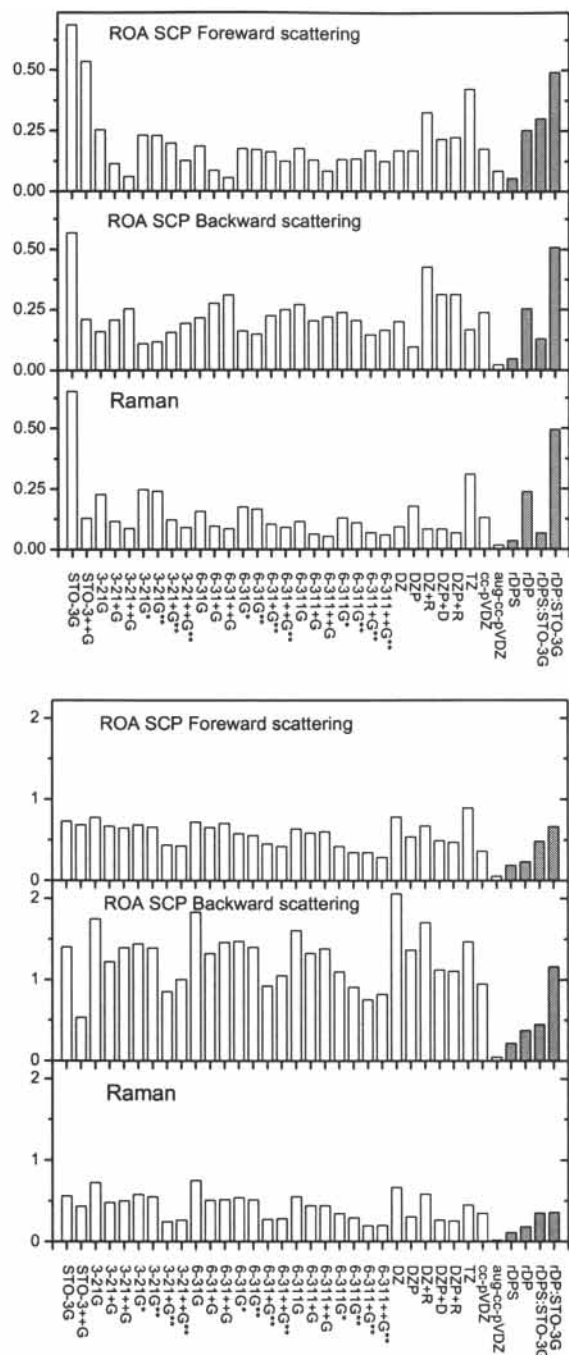


Figure 8. Normalized average absolute deviation according to eq 8 with respect to the set of Sadlej for MH- and MD-stretch vibrations (top) and for other vibrations (bottom) of (*R*)-(+)-methyloxirane. In addition to the new sets rDPS and rDP, the two similarly constructed STO-3G-based sets are included for comparison. The value of the exponent α of the diffuse p-type functions added onto hydrogen atoms was set to 0.2.

underlying vibrational mode does not agree with the experimental one. The positive band at 1107 cm^{-1} is due to vibration 9, calculated at 1125 cm^{-1} , for which the computed ROA is slightly negative. With present knowledge it is not possible to decide if the discrepancy is due to the vibrational coordinates, the electronic tensors, or the fact that we compare condensed phase spectra with theoretical gas-phase data. We did find a strong dependence of the calculated ROA of this vibration on the method used to calculate the vibrational coordinates. The calculated ROA also varies with the basis set used for the electronic tensor calculation. From the extensive tabulated data

presented in ref 9 a strong dependence on the chosen geometry and on the method with which the electronic tensor is calculated is also evident.

The third discrepancy concerns the methyl torsional mode calculated with a positive sign at 212 cm^{-1} and observed as an extended negative band at somewhat lower wavenumbers. The band shape indicates that anharmonicity and the Boltzmann population of vibrational states, and possibly also interactions in the condensed phase, cannot be neglected in the interpretation of this ROA signal. This precludes a meaningful comparison with our computed data.

Concluding Discussion

From the data presented in this work it is clear that Raman and ROA spectra in excellent qualitative and reasonable quantitative agreement with those from heavier basis sets can be calculated with much simpler sets, provided the vibrational problem is solved by a separate calculation. Such rarefied sets for electronic tensor calculations need to be carefully chosen, though, with an important aspect being the proper description of gradients on hydrogen atoms. Figure 8 summarizes these findings by a comparison of the average absolute deviation, relative to the basis set of Sadlej, of Raman and ROA intensities, calculated with a large number of standard basis sets and with the new sets with optimized exponents.

Some of the sets included in Figure 8 are expected to yield marginal results, such as STO-3G. Unexpectedly, however, the largest average error, for all compared sets, is due to a split valence type set, namely the DZ set of Dunning. Its average deviation for ROA backscattering is more than twice the ROA itself, calculated with the set of Sadlej! Adding polarization or Rydberg type functions to DZ leads to only a moderate reduction in error. Likewise, going from DZ to TZ provides little encouragement. Considering such facts, the quality of the results which one obtains with a set as simple as rDPS appears amazing.

None of the Pople style sets in Figure 8 yield uniformly good results. There is marginal improvement only in going from 3-21G to the far more elaborate description of the core and valence electrons which 6-311G provides. Adding diffuse s- and p-type functions onto main atoms (the first + sign in Pople's notation) in general improves results, but adding s-type functions also onto hydrogen atoms (++) does so only occasionally. d-type polarization functions on main atoms (*) lead to a slight improvement in many cases, and additional p-type polarization functions on hydrogen atoms (**) have a similar effect. The improvements are not commensurate in any way, however, with the increase in computational expense which they incur, and even 6-311++G** is hardly a set one will recommend for electronic tensor calculations.

As aug-cc-pVDZ shows, it is only diffuse functions with higher than valence angular momentum quantum numbers which decisively improve computed results. Except for aug-cc-pVDZ, the rDPS set with optimized exponents does best by far. This impressive performance is one of the clearest indications that it is moderately diffuse p-type functions on hydrogen atoms which matter most. The relatively large errors found for rDP:STO-3G, on the other hand, also show that simply adding diffuse p-type functions onto hydrogen atoms is not by itself a cure-all. The marginal STO-3G set certainly improves by this procedure but without becoming a set well suited to the calculation of electronic tensors.

From the point of view of assigning absolute configurations, the predictive power of rDPS appears on a par with that of the set of Sadlej or of aug-cc-pVDZ. Except for truly small

molecules, moreover, the precision of the vibrational calculation will often be the limiting factor, rather than that of the electronic tensor calculation. It thus appears that the precision obtainable with rDPS is about as far as one reasonably wants to go, for analytical purposes, in the computation of the electronic tensor part of organic molecules.

Acknowledgment. The authors are grateful to Professor K. Ruud, University of Tromsø, for his help in getting early versions of the DALTON program running in Switzerland. The work was supported by the Swiss National Foundation under Grant Nos. 2000-56905.99 and 2000-66679.01.

References and Notes

- (1) (a) Placzek, G. Rayleigh-Streuung und Raman-Effekt. In *Handbuch der Radiologie*; Marx, E., Ed.; Akademische Verlagsgesellschaft: Leipzig, 1934; Vol. VI, p 205. (b) Long, D. A. *The Raman Effect*; John Wiley and Sons Ltd.: Chichester, U.K., 2002.
- (2) Barron, L. D. *Molecular Light Scattering and Optical Activity*; Cambridge University Press: Cambridge, U.K., 1982.
- (3) (a) Hug, W.; Surbeck, J. *J. Raman Spectrosc.* **1982**, *13*, 38. (b) Deffontaine, A.; Bridoux, M.; Delhay, M.; Da Silva, E.; Hug, W. *Rev. Phys. Appl.* **1984**, *19*, 415.
- (4) Hug, W.; Hangartner, G. *J. Raman Spectrosc.* **1999**, *30*, 841.
- (5) Hug, W. *Appl. Spectrosc.* **1981**, *35*, 115.
- (6) Hug, W. *Appl. Spectrosc.* **2003**, *57*, 1.
- (7) Helgaker, T.; Ruud, K.; Bak, K. L.; Joergensen, P.; Olson, J. *Faraday Discuss.* **1994**, *99*, 165.
- (8) Hug, W.; Zuber, G.; de Meijere, A.; Khlebnikov, A. F.; Hansen, H.-J. *Helv. Chim. Acta* **2001**, *84*, 1.
- (9) Ruud, K.; Helgaker, T.; Bour, P. *J. Phys. Chem. A* **2002**, *106*, 7448.
- (10) Hecht, L.; Nafie, L. A. *Mol. Phys.* **1991**, *72*, 441.
- (11) Hug, W. Raman Optical Activity Spectroscopy. In *Handbook of Vibrational Spectroscopy*; Chalmers, J. M., Griffiths, P. R., Eds.; John Wiley and Sons Ltd.: Chichester, U.K., 2002; Vol. 1, pp 745–758.
- (12) Hug, W. *Chem. Phys.* **2001**, *264*, 53.
- (13) (a) Oddershede, J.; Jorgensen, P.; Yeager, D. L. *Comput. Phys. Rep.* **1984**, *2*, 33. (b) Jorgensen, P.; Jensen, H. J. Aa.; Olsen, J. *J. Chem. Phys.* **1988**, *89*, 3654.
- (14) Helgaker, T.; Jensen, H. J. Aa.; Joergensen, P.; Olsen, J.; Ruud, K.; Aagren, H.; Bak, K. L.; Bakken, V.; Christiansen, O.; Dahle, P.; Dalskov, E. K.; Enevoldsen, T.; Fernandez, B.; Heiberg, H.; Hiettema, H.; Jonsson, D.; Kirpekar, S.; Kobayashi, R.; Koch, H.; Mikkelsen, K. V.; Norman, P.; Packer, M. J.; Ruden, T. A.; Saue, T.; Sauer, S. P. A.; Sylvester-Hvid, K. O.; Taylor, P. R.; Vahtras, O. *DALTON, An Electronic Structure Program*, release 1.1, 2000.
- (15) (a) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (16) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650.
- (17) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; 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.; Gonzalez, 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*; Gaussian, Inc.: Pittsburgh, PA, 1998.
- (18) Hangartner, G. *FChk2Hes*. Unpublished work, 1997.
- (19) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
- (20) Sadlej, A. J. *Collect. Czech. Chem. Commun.* **1988**, *53*, 1995.
- (21) Halls, M. D.; Schlegel, H. B. *J. Chem. Phys.* **1999**, *111*, 8819.
- (22) Dunning, T. H., Jr. *J. Chem. Phys.* **1989**, *90*, 1007.
- (23) Davidson, E. R.; Feller, D. *Chem. Rev.* **1986**, *86*, 681.
- (24) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939.
- (25) Hehre, W. J.; Stewart, R. F.; Pople, J. A. *J. Chem. Phys.* **1969**, *51*, 2657.
- (26) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294.
- (27) Hug, W. Instrumental and Theoretical Advances in Raman Optical Activity. In *Raman Spectroscopy, Linear and Nonlinear*; Lascomb, J., Huang, P., Eds.; Wiley-Heyden: Chichester, U.K., 1982; pp 3–12.
- (28) Hecht, L.; Barron, L. D.; Hug, W. *Chem. Phys. Lett.* **1989**, *158*, 341.
- (29) Barron, L. D.; Hecht, L.; Gargaro, A. R.; Hug, W. *J. Raman Spectrosc.* **1990**, *21*, 375.
- (30) Bose, P. K.; Polavarapu, P. L.; Barron, L. D.; Hecht, L. *J. Phys. Chem.* **1990**, *94*, 1734.