

On the use of property-oriented basis sets for the simulation of vibrational chiroptical spectroscopies

Brendan M. Shumberger, Ethan H. Fink, Rollin A. King & T. Daniel Crawford

To cite this article: Brendan M. Shumberger, Ethan H. Fink, Rollin A. King & T. Daniel Crawford (23 Dec 2023): On the use of property-oriented basis sets for the simulation of vibrational chiroptical spectroscopies, Molecular Physics, DOI: [10.1080/00268976.2023.2293232](https://doi.org/10.1080/00268976.2023.2293232)

To link to this article: <https://doi.org/10.1080/00268976.2023.2293232>



View supplementary material [↗](#)



Published online: 23 Dec 2023.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

On the use of property-oriented basis sets for the simulation of vibrational chiroptical spectroscopies

Brendan M. Shumberger^a, Ethan H. Fink^a, Rollin A. King^b and T. Daniel Crawford^a

^aDepartment of Chemistry, Virginia Tech, Blacksburg, Virginia, U.S.A.; ^bDepartment of Chemistry, Bethel University, St. Paul, Minnesota U.S.A.

ABSTRACT

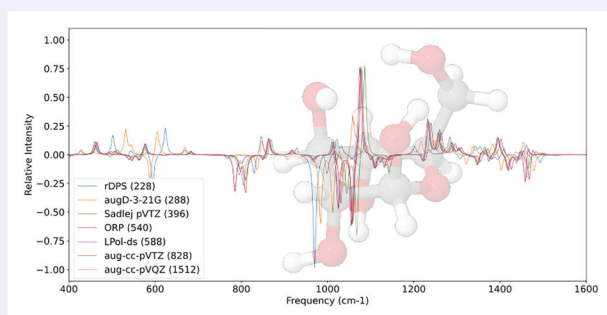
We computed vibrational circular dichroism (VCD) and Raman optical activity (ROA) spectra for a test set of six chiral compounds using two standard density-functionals and an array of basis sets. We analysed the performance of property-oriented basis sets using a quadruple-zeta basis as a reference against four key metrics. We find little qualitative difference between the spectra produced by the larger basis sets (ORP, LPolX, aug-cc-pVTZ, and aug-cc-pVQZ), though their quantitative metrics exhibit wide variations. The smaller basis sets (rDPS, augD-3-21G, augT3-3-21G, Sadlej-pVTZ, and aug-cc-pVDZ) performed better for VCD rotatory strengths than for the corresponding ROA circular intensity differences (CIDs). However, this trend diminishes as the basis-set size is increased, lending validity to the conclusion that more robust property-oriented basis sets are required for ROA spectral generation than that of VCD. We observed improved performance in the mid-infrared region compared to the high-frequency regime, as well as overestimation of VCD rotatory strengths in the latter region as compared to the reference. We conclude that the ORP and LPol-ds basis sets are the most efficient and effective choices of basis set for the prediction of VCD and ROA spectra, as they provide both highly accurate results at reduced computational expense.

ARTICLE HISTORY

Received 9 November 2023
Accepted 5 December 2023

KEYWORDS

Density-functional theory; chirality; vibrational circular dichroism; Raman optical activity



1. Introduction

Over the last several decades, two complementary infrared characterisation techniques, vibrational circular dichroism (VCD) and Raman optical activity (ROA), have been developed for the determination of absolute configuration and conformational analysis [1]. VCD and ROA are defined as the differential absorption and differential scattering, respectively, of left- and right-circularly polarised light [2]. Due to their flexible experimental requirements as compared to electronic circular dichroism (ECD), X-ray crystallography, or nuclear magnetic resonance (NMR) spectroscopy,

VCD and ROA both can be applied to a wide array of molecules in a variety of conditions [1,3], though the two methods have seen application in somewhat distinct areas of chemistry. Whereas, VCD has primarily been used in the determination of absolute configuration for small organic molecules, ROA has seen greater application towards the conformational analysis of large water-soluble biological molecules [4]. A recent publication by Bogaerts et. al., however, suggested that utilising both VCD and ROA spectroscopies together allows stronger, unambiguous assignments of diastereomers [5].

CONTACT T. Daniel Crawford  crawdad@vt.edu

 Supplemental data for this article can be accessed online at <http://dx.doi.org/10.1080/00268976.2023.2293232>.

Unfortunately, the relationships between molecular structure and molecular properties are often highly complex and frequently require assistance from theoretical computations for robust interpretation. Simulation of chiroptical spectra is particularly challenging, however, as it requires the computation of the response of the electronic wave function with respect to external electric and magnetic fields: through dynamic molecular property tensors in the case of ROA or through static wave function derivatives for VCD (*vide infra*). The determination of such responses can be computationally expensive and the properties themselves tend to be highly sensitive to environmental effects, often requiring the inclusion of dynamic solvent models and conformational averaging [6,7].

A key component in the quantum chemical computation of molecular properties, including ROA and VCD spectra, is the choice of basis set. Traditional Gaussian-type basis sets come in a variety of flavours but one commonality between them is that their orbital exponents are often optimised with respect to atomic energies and with contraction schemes designed to provide additional computational efficiency [8–12]. The quality of the basis set is pivotal to the robustness of simulations of electronic and vibrational chiroptical spectra, but balancing accuracy with computational cost is necessary for practical applications.

A number of efforts to elaborate on the basis-set dependence of VCD and ROA spectra have been reported in the literature. In 2004, Zuber and Hug [13] carried out a set of careful computations of the effect of using relatively simple basis sets (which they referred to as ‘rarefied’) to obtain the required circular intensity differences of ROA spectra. They found that they could obtain reasonable quantitative agreement as compared to larger basis sets, provided the vibrational normal modes were obtained separately using a more complete set, as well. In addition, they found that, for a test set of small chiral molecules containing hydrogen atoms, it was necessary to include diffuse *p*-type functions on those atoms, forming what they referred to as ‘rDP’ and ‘rDPS’ functions, depending on the initial basis set. This work was followed by a 2005 study by Reiher, Ligeois, and Ruud [14] that examined the performance of the Zuber and Hug minimal basis sets, as well as Dunning’s correlation-consistent basis sets, and Sadlej’s polarised triple-zeta basis set (described below). Using a set of five representative chiral compounds ranging from (*S*)-methyloxirane up to (*M*)- σ -[4]-helicene, they found agreement regarding Zuber and Hug’s observation that their rarefied basis sets yielded a good balance between accuracy and cost for ROA intensities, provided a more complete basis set was used to obtain the force field. In 2011, Cheeseman

and Frisch [15] carried out a larger study involving 11 molecular test cases, and concurred with the findings of Reiher *et al.* In addition, they found that they could modify Dunning’s augmented correlation-consistent double-zeta (aug-cc-pVDZ) basis set by removing a set of diffuse *d*-type functions to improve the computational costs of the basis set with little impact on the accuracy of the resulting ROA spectra.

Until recently, fewer studies have been reported analysing basis set effects for VCD spectra. The earliest effort by Jalkanen, Stephens, Amos, and Handy [16] focussed on the basis-set dependence of the atomic axial tensors and atomic polar tensors underlying the VCD rotatory strengths. Using the NHDT isotopomer of ammonia as a chiral test case and a series of 12 relatively small basis sets, they determined that polarisation functions are essential for accurate calculations. Much more recently, Scholten, Engelage, and Merten [17] expanded on the literature surrounding density functionals and basis sets involved in VCD spectral prediction by exploring tosylates and sulfinates. In their study, they systematically varied the number of polarisation functions in the Pople 6-311++G(2d,p) basis set. They found that the stretching modes of S=O are red shifted when the basis set does not include higher order polarisation functions and stressed that the minimum Pople style basis that should be used for these types of molecules is the 6-311G(3df,2dp) set. In 2022, Eikås, Beerepoot, and Ruud [7] included an analysis of basis-set effects as part of their larger effort to develop a systematic protocol for simulating VCD spectra of cyclic oligopeptides. They focussed on the standard series of split-valence basis sets by Pople and co-workers including polarisation and diffuse functions. Using the 6-311++G** basis set as the reference, they found that the effect of polarisation functions was more significant than diffuse functions, and heavy-atom polarisation functions were more important than for hydrogen atoms. However, they also observed that shifts in the vibrational frequencies between basis sets led to poorer overlap between the reference spectra and those computed using smaller basis sets. Earlier this year, Groß *et al.* [18] compared the performance of different density functionals, basis sets, and solvation models for VCD spectra including the Pople, Karlsruhe, and Dunning series of basis sets for a test set of six molecules. They found that the largest Dunning basis set they tested, aug-cc-pVTZ, performed best overall though they considered this basis too large for application to larger molecules.

In addition, to the above studies, which focussed primarily on the use of conventional or slightly modified conventional basis sets, a number of researchers have designed basis sets specifically to target response and

related properties. The aim of such property-oriented basis sets is two-fold: to increase the accuracy of the computed property and to reduce the number of basis functions required to obtain it [19]. One of the earlier attempts at developing such a bespoke basis set was reported by Sadlej [20], who used the basis set polarisation method to develop a medium-sized basis with the goal of predicting static electric dipole polarisabilities and related properties. Using the atomic gaussian-type orbital (GTO) basis set of van Duijneveldt as the source, each shell was augmented by one diffuse GTO with orbital exponents determined by assuming an even-tempered sequence of the two most diffuse orbitals in that shell. Expansion coefficients were determined through self-consistent field Hartree-Fock (SCF HF) calculations and first-order polarisation functions were derived by applying the basis set polarisation method to the outermost shell. These functions were appended to the augmented set and a contraction scheme was determined based on the computed values of the polarisability anisotropy. As noted by Sadlej, this basis set, referred to as the Sadlej-pVTZ basis set, did not involve any explicit, property-oriented, optimisation of orbital exponents. It did, however, provide a systematic method for introducing polarisation functions. This new basis set provided comparable results to much larger basis sets at both Hartree-Fock and correlated levels of theory.

In 2010, Baranowska and Sadlej developed the LPolX basis set using a similar method [21]. Again using the largest optimised GTO basis of van Duijneveldt as the source set, each shell was augmented with one diffuse GTO and the exponents determined by geometric progression. Expansion coefficients were computed through restricted open-shell Hartree-Fock (ROHF) calculations and first-order polarisation functions were derived. Following this, second-order polarisation functions were derived and their contraction coefficients were determined using ROHF calculations on the uncontracted set including first-order polarisation functions which resulted in the LPol-fl set. Reducing the size of these basis sets resulted in the LPol-fs, LPol-dl, and LPol-ds sets where ‘f’ and ‘d’ refer to the highest angular momentum type of the basis set and ‘l’ and ‘s’ refer to the size of the chosen contraction scheme (‘large’ or ‘small’). Using a test of four small molecules, they found that the LPolX basis sets provided a high level of saturation for describing the polarisation effects caused by external electric fields, yielding accurate dipole moments, polarisabilities, and first hyperpolarizabilities. In addition, they speculated that the LPolX basis sets would also be useful for mixed electric-field/magnetic-field response properties such as optical rotations and circular dichroism spectra.

Following this work, Baranowska-Łączkowska and Łączkowski developed the optical rotation prediction

(ORP) basis set, which, as the name indicates, targeted computation of specific rotation in chiral molecules [22]. Using the uncontracted VTZ basis of Ahlrichs and coworkers [23], they augmented each shell with one diffuse function which, again, had orbital exponents determined through geometric progression. Next, three uncontracted first-order polarisation functions with initial guesses at their orbital exponents were appended to the set. The orbital exponents were then optimised by minimising the errors in atomic polarisabilities using finite-field ROHF calculations. Finally, a contraction scheme was determined using ROHF calculations of the initial set with the polarisation functions remaining uncontracted. They found excellent agreement with rotations obtained using the larger aug-cc-pVTZ basis set. Three years later, Baranowska-Łączkowska [24] introduced a smaller version of the ORP basis set, R-ORP, which was similarly constructed but includes fewer polarisation functions on each atom.

In 2020, Aharon and Caricato [25] introduced two highly compact basis sets also targeting optical rotations, constructed by combining the standard 3-21G Pople-type basis set with the diffuse functions from the Dunning aug-cc-pVDZ or aug-cc-pVTZ basis sets, but then reoptimizing the latter to minimise statistical errors in long-wavelength specific rotations as compared to the full aug-cc-pVXZ basis set. Using a training set of 21 chiral compounds to build the new basis sets – named augD-3-21G and augT3-3-21G – they carried out a series of computations on a control set consisting of another 30 compounds. They found that their new basis sets yielded mean unsigned errors in sodium D-line specific rotations of approximately 4% at considerably less cost than the larger ORP basis set.

The current work is intended to be a component of a larger effort to develop a robust computational protocol for simulating the VCD and ROA spectra of glycans in solution. Such a goal requires exploration of numerous relevant variables, including choice of quantum chemical model (e.g. density functional, level of electron correlation), implicit vs. explicit solvent model, molecular dynamics forcefield (in the case of explicit solvation models), conformational/configurational sampling, and, of course, basis set.

The central question we seek to address is whether the recently-developed basis sets targeting *electronic* response properties such as dipole polarisabilities and optical rotation provide similar accuracy at reduced computational cost for *vibrational* response properties. Given that these seemingly disparate properties depend on the same sundry response tensors and wavefunction derivatives, a compelling argument can be made that these new basis sets will perform admirably beyond their original

purpose. Thus, we have carried out a series of benchmark VCD and ROA computations across a range of general and property-optimised basis sets to measure their performance against the effectively complete reference quadruple-zeta (aug-cc-pVQZ) basis set. We make use of several quantitative measures for this purpose, all oriented towards identifying the basis set(s) that provide practical balance between accuracy and computational efficiency.

2. Theoretical background

Although VCD and ROA are both vibrational chiroptical spectroscopies, their theoretical formulations are distinct. Theoretical simulation of VCD requires computation of the rotatory strength, which is the dot product of the electric-dipole and magnetic-dipole vibrational transition moments, *viz.*,

$$R_{\nu\nu'}^n = \text{Im}\{\langle \Psi_{n\nu} | \vec{\mu} | \Psi_{n\nu'} \rangle \cdot \langle \Psi_{n\nu'} | \vec{m} | \Psi_{n\nu} \rangle\}, \quad (1)$$

where $\vec{\mu}$ and $\vec{m}$ are the electric- and magnetic-dipole operators, respectively, n denotes the electronic state, and ν denotes the vibrational state. The principal challenge in computing $R_{\nu\nu'}^n$ is that the electronic contribution to the magnetic-dipole transition moment unphysically vanishes within the Born-Oppenheimer approximation. As shown by Stephens nearly 40 years ago, [26] this problem may be overcome by first-order corrections of the adiabatic wave functions via Taylor expansions in the nuclear positions and the external magnetic field, leading to the expression,

$$\langle \Psi_{n\nu} | m_{\beta}^{\text{elec}} | \Psi_{n\nu'} \rangle = -2\hbar\omega_{\nu\nu'} \sum_{\alpha} \left\langle \left(\frac{\partial \psi_n}{\partial R_{\alpha}} \right)_0 \middle| \left(\frac{\partial \psi_n}{\partial B_{\beta}} \right)_0 \right\rangle \langle \chi_{n\nu} | (R_{\alpha} - R_{\alpha}^0) | \chi_{n\nu'} \rangle, \quad (2)$$

where the sum runs over all nuclear coordinates, ψ_n is the electronic wave function, $\chi_{n\nu}$ is the vibrational wave function, B_{β} is a particular Cartesian direction of the magnetic field, and the subscript/superscript 0 denotes the reference geometry or field-free state. The overlap between wave function derivatives, known as the electronic contribution to the atomic axial tensor, is the key challenge of formulating and computing VCD rotatory strengths.

ROA, on the other hand, requires geometric derivatives of three dynamic linear response tensors, namely, the electric dipole polarisability, [27]

$$\alpha_{\alpha\beta}(\omega) = \frac{2}{\hbar} \sum_{n \neq 0} \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2} \text{Re}\{\langle \Psi_0 | \mu_{\alpha} | \Psi_n \rangle \langle \Psi_n | \mu_{\beta} | \Psi_0 \rangle\}, \quad (3)$$

the electric-dipole/magnetic-dipole polarisability,

$$G'_{\alpha\beta}(\omega) = -\frac{2}{\hbar} \sum_{n \neq 0} \frac{\omega}{\omega_{n0}^2 - \omega^2} \text{Im}\{\langle \Psi_0 | \mu_{\alpha} | \Psi_n \rangle \langle \Psi_n | m_{\beta} | \Psi_0 \rangle\}, \quad (4)$$

and the electric-dipole/electric-quadrupole polarisability

$$A_{\alpha\beta\gamma}(\omega) = \frac{2}{\hbar} \sum_{n \neq 0} \frac{\omega_{n0}}{\omega_{n0}^2 - \omega^2} \text{Re}\{\langle \Psi_0 | \mu_{\alpha} | \Psi_n \rangle \langle \Psi_n | \Theta_{\beta\gamma} | \Psi_0 \rangle\}. \quad (5)$$

In Equations (3)–(5), the subscripts α , β , and γ refer to Cartesian directions, ω is the frequency of incident radiation, ω_{n0} is the excitation energy between the ground and n th excited electronic states, and Θ is the electric-quadrupole operator.

From the above, we see the key differences between VCD and ROA: the former requires an accurate description of the derivative of the electronic wave function with respect to the external magnetic field, while the latter involves derivatives of electric- and magnetic-field-dependent response tensors. (It is interesting to note that VCD can also be formulated in terms of dynamic response theory, namely as the frequency derivative of the linear response function involving the electronic part of the geometrical gradient operator and the magnetic-dipole operator. [28]) Thus, we wish to methodically test the performance of both conventional and property-optimised basis sets for these methods.

3. Computational details

Our test set includes six relatively small molecules (Figure 1), selected because of their representative structures and bonding patterns: (*P*)-hydrogen peroxide, (*P*)-2,3-pentadiene, (*R*)-fluorooxirane, (*S*)-methyloxirane, (*1S,4S*)-norbornenone, and β -D-glucose. For each molecule, we carried out VCD and ROA computations using optimised geometries and vibrational force fields computed with the given basis set being tested. While some previous studies, such as those reported by Zuber and Hug [13] and by Reiher and co-workers [14], have used small basis sets to obtain the VCD and/or ROA intensities, but larger basis sets for the geometries and vibrational force fields, this is not the conventional workflow in modern quantum chemistry packages. Given that the long-term purpose of the present work is to develop protocols for general use, we have elected to maintain consistent choice of basis set for all components of the workflow. The selection of basis sets

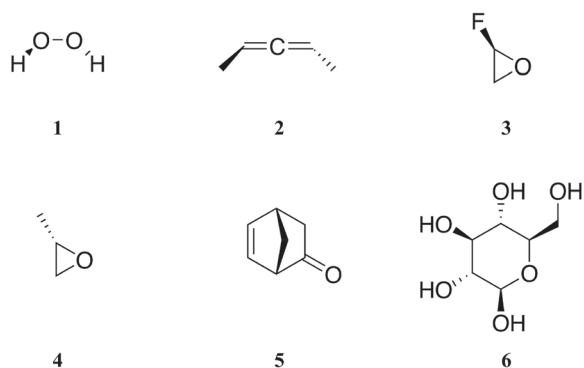


Figure 1. The molecular test set for testing conventional and property-optimised basis sets.

includes (1) the augmented correlation-consistent (aug-cc-pVXZ) sets of Dunning and co-workers [10] with double-, triple-, and quadruple-zeta quality (X=D, T, and Q, respectively), (2) the rDPS basis set of Zuber and Hug [13], (3) the Sadlej-pVTZ basis [20], (4) the LPolX sets of Baranowska and Sadlej [21], (5) the ORP sets of Baranowska-Łaczkowska and Łaczkowski [22] and R-ORP set of Baranowska-Łaczkowska [24], and (6) the compact augD-3-21G and augT3-3-21G sets of Aharon and Caricato [25]. As available, all basis sets were obtained from the Basis Set Exchange of the Molecular Sciences Software Institute [29]. The number of basis functions for each molecule and basis set are listed in Table 1. For each molecule and basis-set combination, the corresponding structures and vibrational spectra were computed using the B3LYP [30–32] and CAM-B3LYP [33] density functionals. All computations were carried out using the Gaussian-09 quantum chemistry package [34]. Workflow control, spectrum generation, and subsequent analysis were carried out using the SoAPy package developed in our research group [35]. The spectrum simulations employ a Lorentzian lineshape for both VCD and ROA, using a full-width at half-maximum (FWHM) of 10^{-3} eV (8.06573 cm^{-1}). Geometry optimizations used tight convergence parameters,

including maximum and root-mean-squared forces of $1.5 \times 10^{-5} E_h/a_0$. For glucose, only the lowest-energy conformer was considered, i.e. the computed spectra are not conformationally averaged. An external field wavelength of 532 nm was used with the in-phase dual circular polarisation (DCPI) at 180° for the ROA spectral simulations.

We have used four primary metrics to analyse the performance of these basis sets. The first of these is the singly normalised overlap (SNO) function,

$$S_s = \frac{\int f_s(\tilde{\nu}) f_r(\tilde{\nu}) d\tilde{\nu}}{\int f_r(\tilde{\nu})^2 d\tilde{\nu}}, \quad (6)$$

which includes integration over the entire (or selected) spectral region(s) between the sample spectrum, f_s , and the reference spectrum, f_r , normalised with respect to only the reference function. The values produced by this function can range from $-\infty$ to ∞ , though for well-behaved functions, such as those examined here, we observe limits typically between $+1$ and -1 , where the former indicates that the spectra are identical and the latter would occur for the same spectrum but for the opposite enantiomer. Artificially inflated values can arise when the sample spectrum has a significantly greater area than that of the reference. This circumstance can be detected, however, by our second and third metrics which are the doubly normalised overlap (DNO) function and the integrated difference function, respectively. The DNO function is,

$$S_d = \frac{\int f_s(\tilde{\nu}) f_r(\tilde{\nu}) d\tilde{\nu}}{\sqrt{\int f_s(\tilde{\nu})^2 d\tilde{\nu} \int f_r(\tilde{\nu})^2 d\tilde{\nu}}}, \quad (7)$$

which is similar in design to the SNO, but, as the name indicates, it includes normalisation factors for both the sample and reference spectra, which formally limits bounds from $+1$ to -1 . Thus, the DNO avoids the artificial inflation that can potentially occur for the SNO. However, it neglects differences in contributions of

Table 1. Number of basis functions for each molecule and basis set.

Basis Set	(P)-hydrogen peroxide	(P)-2,3-pentadiene	(R)-fluorooxirane	(S)-methyloxirane	(1S,4S)-norbornenone	β -D-glucose
rDPS	38	113	70	88	152	228
augD-3-21G	48	138	90	108	192	288
augT3-3-21G	48	138	90	108	192	288
R-ORP	64	187	119	146	256	384
aug-cc-pVDZ	64	187	119	146	256	384
Sadlej-pVTZ	66	192	123	150	264	396
ORP	90	264	167	206	360	540
LPol-ds	98	284	183	222	392	588
LPol-dl	124	361	231	282	496	744
LPol-fs	146	434	269	338	584	876
LPol-fl	196	586	360	456	784	1176
aug-cc-pVTZ	138	414	253	322	552	828
aug-cc-pVQZ	252	768	458	596	1008	1512

the peak intensities to the overlap between the two functions. For example, if a sample spectrum had a magnitude twice that of the reference, but was otherwise identical, the DNO function would still produce an overlap of 1. Thus, we choose to include both the SNO and DNO for our analysis.

We refer to the third metric as the integrated difference function,

$$D = \frac{\int f_r(\tilde{\nu})^2 d\tilde{\nu} - \int f_s(\tilde{\nu})^2 d\tilde{\nu}}{\int f_r(\tilde{\nu})^2 d\tilde{\nu}}, \quad (8)$$

which has the advantage that it is invariant to frequency shifts and is primarily a measure of the error in intensity differences between two spectra. This allows resolution of spectral differences into those from vibrational frequency shifts versus those from differences in absorption/scattering intensities. The integrated difference function ranges from +1 to $-\infty$ where negative values indicate that the squared area of the sample is larger than that of the reference. Values closest to zero indicate that the differences between spectra are minimal while values further away, either negative or positive indicate greater differences between the two functions. For this reason, we refer to the results of the integrated difference function as ‘errors’ and our analysis primarily involves their absolute values. We choose to retain the sign in our reported results to offer the opportunity to identify potentially inflated singly normalised overlap values.

Our final metric for basis set evaluations are ‘events,’ which include discrepancies in the sign between the sample and the reference, normal mode reordering, or a combination of the two. This metric has the advantage that it accumulates discrepancies that are not necessarily visible in the spectra for further comparison. Events are determined by comparing the sign of the sample intensity with that of the reference intensity for a given normal mode index.

Ideally, we would define values for each of these metrics indicating a minimal level of acceptable performance, such as whether it was sufficient for assigning a molecule’s absolute stereochemical configuration. However, such metrics must ultimately be determined based on comparison to experimental data, whereas the goal of the present work is to measure the efficacy of each basis set relative to more complete sets. In many ways this is a more stringent requirement than configuration assignment, and smaller basis sets may offer significant advantages for experimental comparisons. Furthermore, it should be noted that we have chosen not to include the parent IR and Raman vibrational spectra, which are often analysed alongside their chiroptical counterparts. This is especially important when comparing to experimental

data as the parent spectra are critical for determining frequency scaling factors that are commonly applied. These scaling factors are dependent on the choice of density functional, basis set, and functional groups associated with the molecule of interest [17]. Given that our study involves only comparisons to computed spectra in an effort to identify the most computationally effective and efficient protocol for VCD and ROA, inclusion of the parent IR and Raman spectra would only serve to provide redundant information in our basis set evaluation, especially given the greater sensitivity of VCD and ROA spectra than their chiral versions to the level of theory and choice of basis set. In addition, while conformational dynamics and solvent effects are certainly critical for comparison to experiment, we have not considered them in the present work in order to focus specifically on basis set performance.

4. Results

The general trends observed between the B3LYP and CAM-B3LYP functionals are sufficiently similar that we choose to focus on the former unless otherwise stated. However, all data, including spectra and metrics for all molecules, basis sets, and density-functionals, are provided in the Supplementary Information. The results will be presented in the following format for each molecule. First the overlaps, integrated differences, and number of events will be presented in groupings based on the performance of each basis set for VCD and ROA. Then a brief synopsis of the best performing basis sets will be given, including contributions from basis set size and the three metrics used for evaluation of the basis sets. Finally, the events for each basis set will be described in more detail including the types of normal modes and their respective effects on the final spectrum.

4.1. (*P*)-Hydrogen peroxide

We choose to begin with the smallest of our molecular test cases because its small number of vibrational modes make it simultaneously the simplest to analyse in detail, but also the most challenging for our four metrics. In addition, hydrogen peroxide is also of interest because it is one of two compounds in our test set that exhibits axial, rather than point chirality. The computed VCD and ROA spectra for (*P*)-hydrogen peroxide are given in Figures 2 and 3 where we have grouped the basis sets into ‘small’ (rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, and Sadlej-pVTZ) and ‘large’ (ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, and aug-cc-pVTZ), respectively, for easier comparison and analysis. The overlaps,

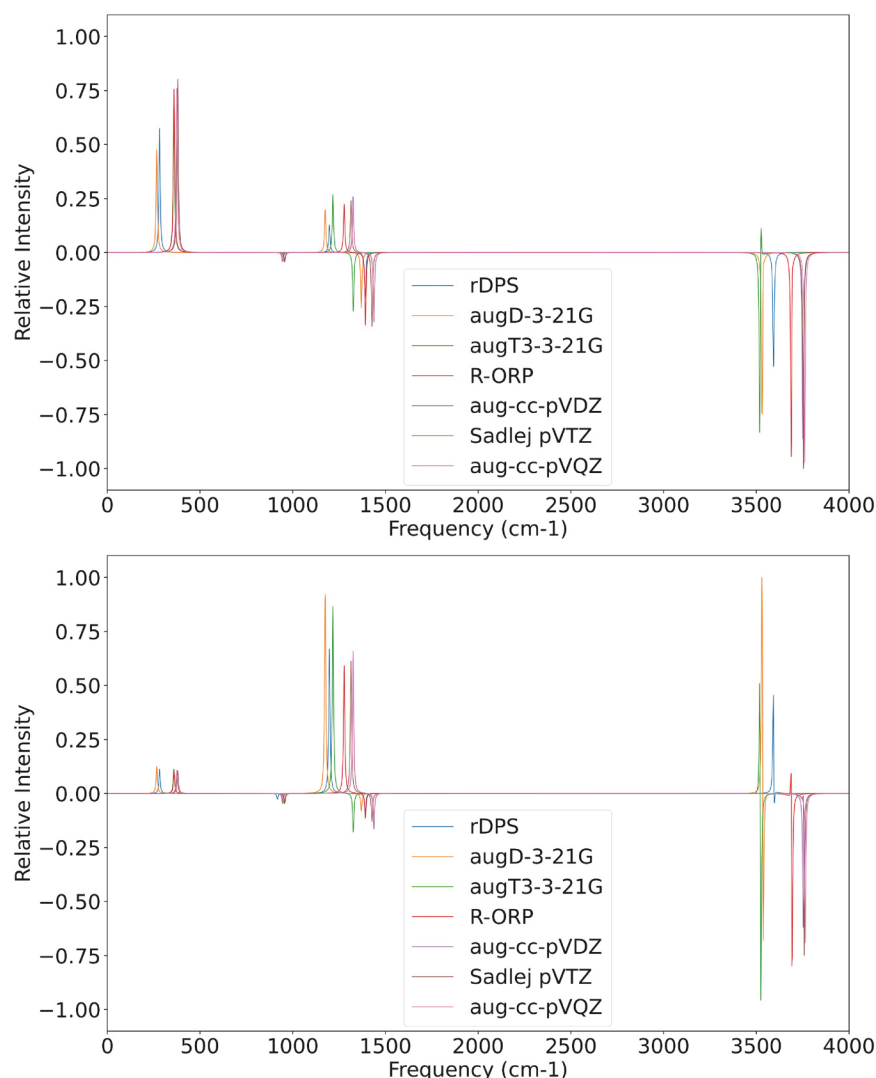


Figure 2. VCD (top) and ROA (bottom) spectra of (*P*)-hydrogen peroxide computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

integrated differences, and number of events for (*P*)-hydrogen peroxide are presented in Table 2.

Given the narrow margin for error because of the limited number of vibrational modes, it is unsurprising that most of the smallest basis sets under scrutiny – rDPS, augD-3-21G, augT3-3-21G, and R-ORP – exhibit wide variations in their representations of the VCD and ROA spectra of (*P*)-hydrogen peroxide, both in terms of the positions of the vibrational transitions and the corresponding rotatory strengths/scattering intensities. While only the very small rDPS basis exhibits an event (sign discrepancy relative to the reference), the error, which occurs in the O–H stretching region above 3500 cm^{-1} is visibly obscured in the VCD spectrum because of another vibrational mode lying ca. 1 cm^{-1} away with the same sign. The pair is slightly more separated with the augT3-3-21G basis set and so the opposite sign rotatory strengths are visible in the spectrum. The pair of O–H

stretching vibrations are more distinct in the ROA spectrum for these small basis sets, but that, too, is at odds with the aug-cc-pVQZ spectrum, which exhibits only one peak due to our chosen FWHM. The SNO and DNO values for these basis sets fall into the range of 0.00–0.10 for VCD and $-0.13 - 0.02$ for ROA, which is exceedingly low relative to the reference aug-cc-pVQZ results. However, the integrated difference formula, which is expected to emphasise differences in intensities and deemphasize vibrational frequency shifts, reveals that the rDPS, augD-3-21G, and augT3-3-21G basis sets still exhibit substantial errors, but the discrepancies for the R-ORP basis are at least somewhat due to errors in the harmonic forcefield.

The aug-cc-pVDZ basis set performs similarly for both VCD and ROA spectra with SNO/DNO values of ca. 0.65–0.66 and small integrated difference values. The Sadlej-pVTZ basis is somewhat poorer with

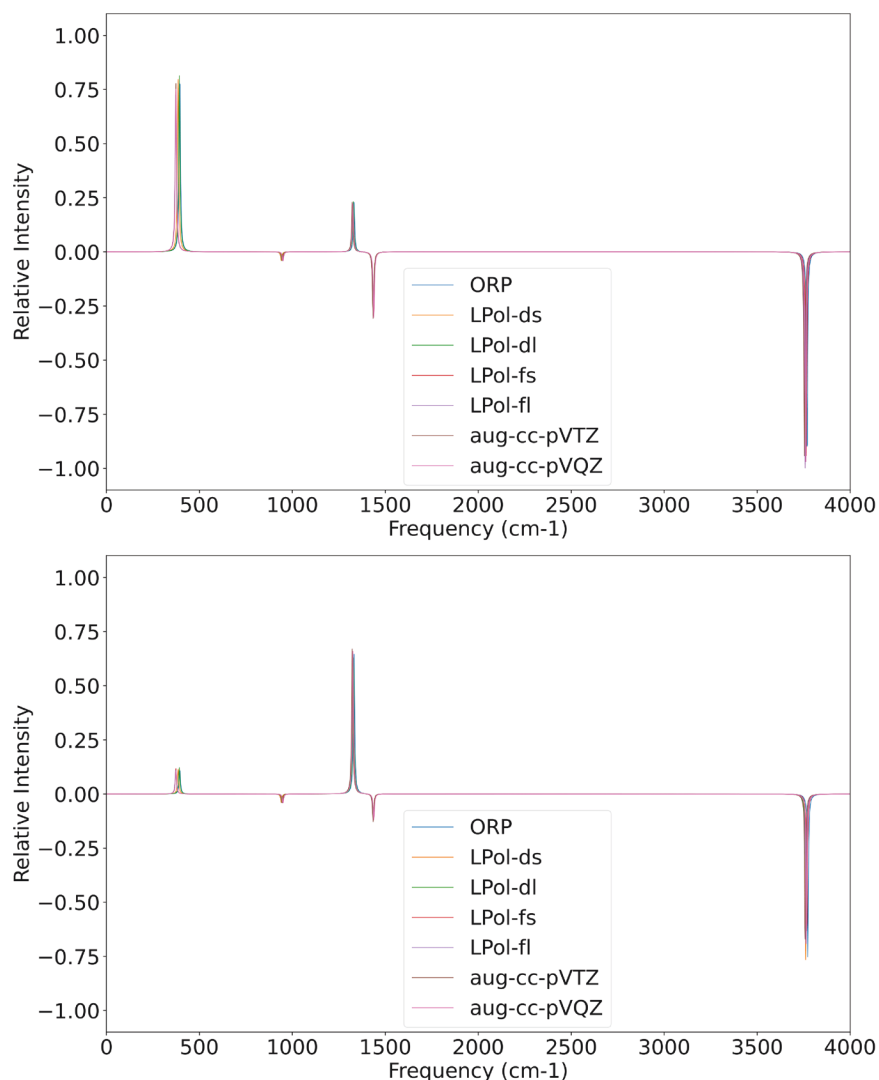


Figure 3. VCD (top) and ROA (bottom) spectra of (*P*)-hydrogen peroxide computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 2. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of (*P*)-hydrogen peroxide using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	0.0035	0.0053	0.5612	1	0.0010	0.0011	0.2459	0
augD-3-21G	0.0015	0.0019	0.4279	0	0.0015	0.0010	−1.4351	0
augT3-3-21G	0.0355	0.0400	0.2127	0	−0.1331	−0.0935	−1.0256	0
R-ORP	0.0956	0.0976	0.0410	0	0.0205	0.0197	−0.0837	0
aug-cc-pVDZ	0.6558	0.6602	0.0132	0	0.6580	0.6494	−0.0265	0
Sadlej-pVTZ	0.5839	0.5554	−0.1053	0	0.4053	0.4109	0.0268	0
ORP	0.3789	0.3774	−0.0080	0	0.5502	0.5359	−0.0538	0
LPol-ds	0.6006	0.5966	−0.0135	0	0.8754	0.8429	−0.0786	0
LPol-dl	0.6143	0.6131	−0.0039	0	0.8520	0.8575	0.0127	0
LPol-fs	0.8984	0.9169	0.0400	0	0.8718	0.9012	0.0642	0
LPol-fl	0.8628	0.8667	0.0089	0	0.8544	0.8680	0.0310	0
aug-cc-pVTZ	0.7187	0.7245	0.0160	0	0.6526	0.6390	−0.0429	0
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

SNO/DNO values of 0.58/0.56 for VCD and 0.41/0.41 for ROA. Somewhat surprisingly, the ORP basis set, which is the next largest in the list, has a lower than expected SNO/DNO values of 0.38/0.38 for the VCD

spectrum of (*P*)-hydrogen peroxide, though the overlaps increase somewhat to 0.55/0.54 for the ROA spectrum. However, similarly to the smaller basis sets, its integrated difference formula remains relatively small at

−0.008 for VCD and −0.05 for ROA, which suggests that the principal source of the discrepancy is the vibrational frequency shifts rather than the rotatory strengths or CIDs.

The LPol-ds and LPol-dl basis sets perform worse for the VCD spectrum of (*P*)-hydrogen peroxide than the aug-cc-pVDZ basis (SNO/DNO values of 0.60/0.60 for LPol-ds, 0.61/0.61 for LPol-dl even though they are considerably larger. However, the performance improves considerably for the ROA spectrum increasing to 0.88/0.84 for LPol-ds and 0.85/0.86 for LPol-dl. Among the larger basis sets, LPol-fs and LPol-fl yield very good results for both VCD and ROA spectra while the aug-cc-pVTZ basis performs somewhat more poorly than expected, given its size. Unfortunately, the aug-cc-pVTZ basis set yields somewhat worse results with SNO/DNO values of 0.71/0.72 for VCD and 0.65/0.64 for ROA. However, in spite of the variations in overlaps and integrated

differences, it should be noted that, for all of the larger basis sets – from ORP through aug-cc-pVTZ – the visual differences in both VCD and ROA spectra are relatively small.

4.2. (*P*)-2,3-Pentadiene

We include (*P*)-2,3-pentadiene in our test set because of its axial chirality, as well as its adjacent C=C bonds. The computed VCD and ROA spectra for (*P*)-2,3-pentadiene are given in Figures 4 and 5 using the same basis-set grouping as the previous section. In addition, the overlaps, integrated differences, and number of events for (*P*)-2,3-pentadiene are given in Table 3.

The overlaps observed for the VCD and ROA spectra draw a clear distinction between the small basis sets – rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, and Sadlej pVTZ – and their larger counter

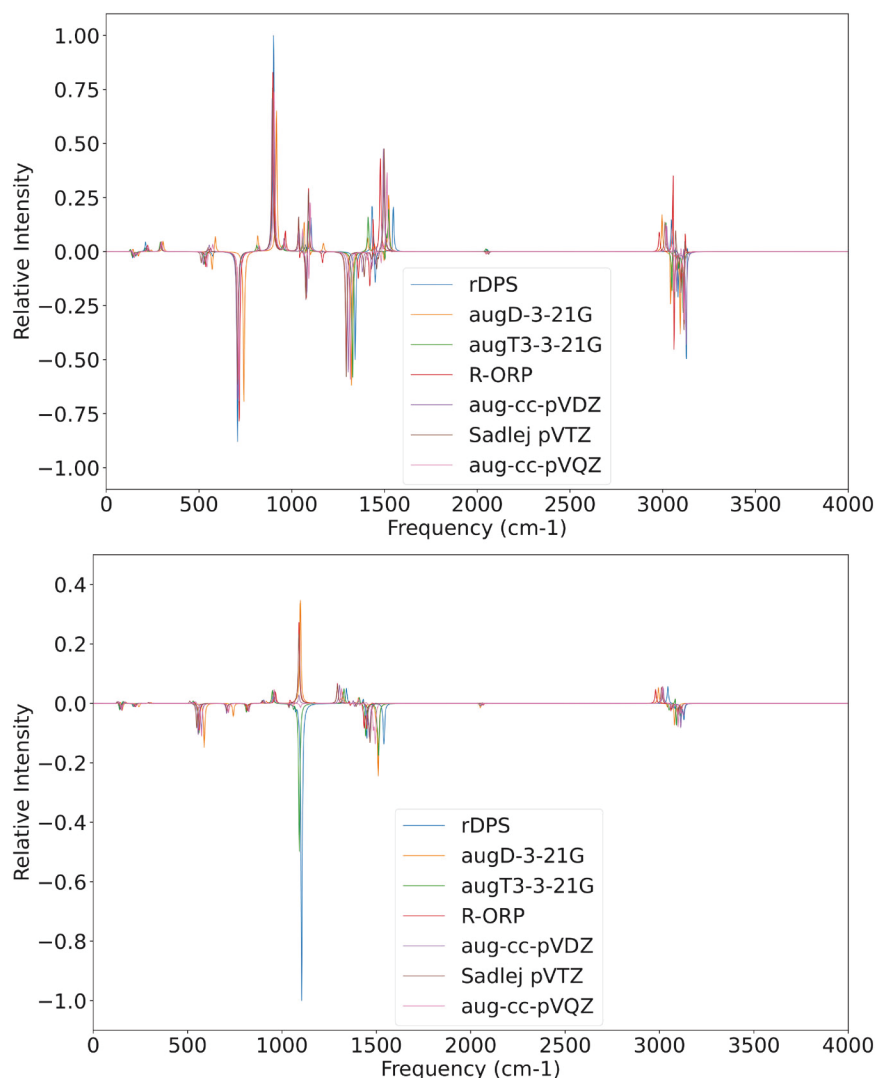


Figure 4. VCD (top) and ROA (bottom) spectra of (*P*)-2,3-pentadiene computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

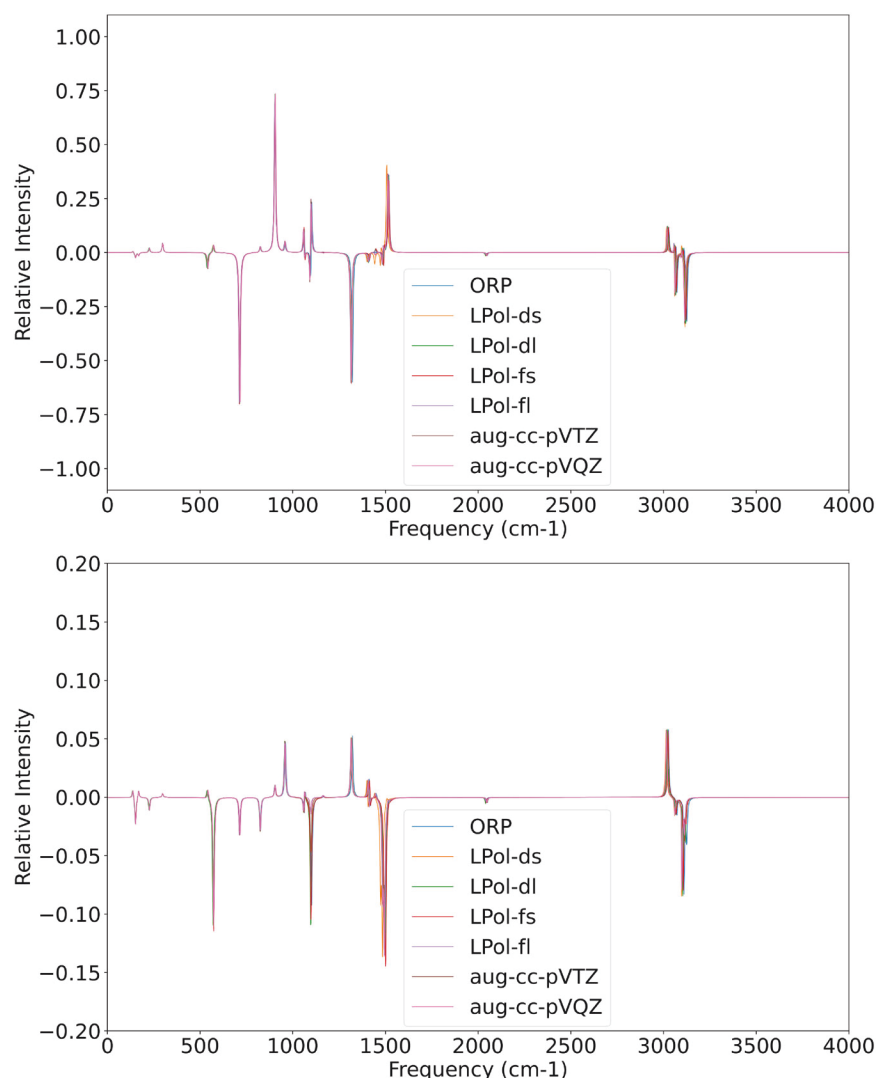


Figure 5. VCD (top) and ROA (bottom) spectra of (*P*)-2,3-pentadiene computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 3. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of (*P*)-2,3-pentadiene using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	0.6077	0.5176	−0.3782	7	0.4634	0.1080	−17.4018	4
augD-3-21G	0.2670	0.2774	0.0736	4	0.2902	0.1441	−3.0564	3
augT3-3-21G	0.4047	0.4294	0.1116	6	0.3826	0.1619	−4.5829	4
R-ORP	0.6264	0.5372	−0.3595	3	0.1241	0.0888	−0.9557	3
aug-cc-pVDZ	0.3478	0.3327	−0.0932	2	0.2460	0.2561	0.0771	1
Sadlej-pVTZ	0.3824	0.3773	−0.0273	5	0.2043	0.1538	−0.7654	7
ORP	0.7749	0.7776	0.0071	3	0.8081	0.7474	−0.1690	0
LPol-ds	0.9614	0.9496	−0.0249	2	0.8006	0.7775	−0.0605	0
LPol-dl	0.9268	0.9298	0.0064	5	0.8752	0.7872	−0.2360	0
LPol-fs	0.8681	0.8773	0.0209	5	0.8360	0.7447	−0.2602	0
LPol-fl	0.9270	0.9322	0.0112	3	0.9685	0.9722	0.0076	0
aug-cc-pVTZ	0.9904	0.9876	−0.0056	2	0.9979	0.9815	−0.0338	0
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

parts with the small basis sets having SNO/DNO values ranging between 0.27/0.28–0.63/0.54 for VCD and 0.12/0.09–0.46/0.26 for ROA. Additionally, this is where the first significant inflation of the SNO values arise

in the ROA spectrum which is attributed to an asymmetric methyl stretching motion at ca. 1100 cm^{−1}. The extent of this inflation is most notable in the rDPS, augD-3-21G, and augT3-3-21G sets which returned absolute

values computed with the integrated difference function upwards of 17.40.

The remaining, larger basis sets performed considerably better with VCD SNO/DNO values ranging between 0.77/0.78–0.99/0.99 and ROA SNO/DNO values between 0.80/0.74–1.00/0.98. Most of the values produced by the integrated difference function for VCD showed only minor discrepancies and inflations though, of note, was a marked increase in the values for ROA which were an order of magnitude greater or more in some cases. Further investigations revealed that this was a result of a poor prediction in the CID value for the same troublesome peak as that for the smaller basis sets ca. 1100 cm^{-1} , though the sign was correctly predicted with these larger sets. Given the performance and size of the ORP and LPol-ds basis sets, further evaluation of these promising bases is required.

In the VCD spectra, the ORP basis exhibits events at 1167.31 cm^{-1} , 3124.98 cm^{-1} , and 3125.00 cm^{-1} , while the LPol-ds basis produces events at 3114.27 cm^{-1} and 3114.37 cm^{-1} . The frequency at 1167.31 cm^{-1} corresponds to a symmetric allene (C=C) stretch, while the higher-frequency modes that are not clearly distinguishable in the spectrum are C–H stretches, and so are less relevant to experimental comparisons. Nevertheless, the discrepancy of the high-frequency modes occurs because of reordering of the symmetric and antisymmetric stretching modes as compared to the aug-cc-pVQZ basis set. Interestingly, none of these discrepancies occur for the corresponding ROA spectra for either the ORP or LPol-ds basis sets, which means that their VCD spectra have inverted signs because of the mode reordering, but they give the correct sign of the rotatory strength for the modes themselves. Meanwhile, we do not observe similar events in the corresponding ROA spectra even though the same modes are re-ordered; hence, the ORP and LPol-ds appear to give the correct ordering of the signs because they produce incorrect signs for the modes themselves. In addition, in comparing the ORP and LPol-ds basis sets, it is noteworthy that the ORP basis has a better absolute error for the integrated difference function for its VCD spectrum while the opposite is true for the ROA spectra.

4.3. (R)-Fluorooxirane

The next test case, (R)-fluorooxirane, is the only compound studied here containing fluorine and thus increases the diversity of the test set for analysing the various basis sets [36]. (Several of the sets under consideration here have not yet been defined beyond fluorine; hence the limitation to first- and second-row elements in our benchmark.) The computed VCD and

ROA spectra for (R)-fluorooxirane are given in Figures 6 and 7 using the same basis-set grouping as in the previous section. The overlaps, integrated differences, and number of events obtained for (R)-fluorooxirane are provided in Table 4.

As with previous molecules in our test set, the small bases – rDPS, augD-3-21G, augT3-3-21G, R-ORP, and aug-cc-pVDZ – performed rather poorly with VCD and ROA SNO/DNO values falling within the ranges $-0.02/-0.02-0.15/0.18$ and $-0.42/-0.29-0.06/0.05$, respectively. These large negative overlap values were attributed to frequency shifts of the smaller bases compared to the reference in regions of the spectrum with a large density of normal modes of alternating signs. Surprisingly, the LPol-ds and LPol-fs did not perform as well as was the case for previous molecules in our test set. The SNO/DNO values observed for the LPol-ds and LPol-fs were 0.59/0.62 and 0.52/0.62 for VCD and 0.60/0.59 and 0.34/0.35 for ROA, respectively. In addition, the LPol-fs basis set had a relatively large value from the integrated difference function for the VCD spectrum in contrast to the other larger basis sets.

The remaining bases performed well for VCD and ROA with SNO/DNO values between 0.69/0.73–0.97/0.97 and 0.75/0.73–0.92/0.93, respectively. This grouping included the Sadlej-pVTZ basis which outperformed the ORP and LPol-fl sets in the VCD spectrum. In the ROA spectrum, however, the SNO/DNO of the Sadlej-pVTZ basis were significantly poorer. Interestingly, within this group, only the LPol-fl basis had a larger than expected value of 0.20 produced by the integrated difference function. Unsurprisingly, aug-cc-pVTZ performed very well, with metrics indicating that it produced results nearly identical to those of aug-cc-pVQZ.

Of the larger/better performing bases, only the Sadlej-pVTZ and LPol-fl basis sets exhibited events. For both VCD and ROA, the Sadlej-pVTZ basis exhibited only one event. In the VCD spectra, the event occurred at 1154 cm^{-1} while for ROA the event was at 3203 cm^{-1} . These normal modes corresponded to an asymmetric hydrogen bending motion in the direction parallel to the plane of the ring and an asymmetric stretching motion involving the same hydrogen atoms, respectively. The LPol-fl basis exhibited events at 515 cm^{-1} and 1154 cm^{-1} and at 1398 cm^{-1} , for VCD and ROA respectively. These vibrational modes involved a bending motion of the fluorine and oxygen atoms relative to the chiral carbon, the same asymmetric hydrogen rocking motion, and a new asymmetric hydrogen bending perpendicular to the ring plane. All of these events, for both basis sets, were apparent in the spectra.

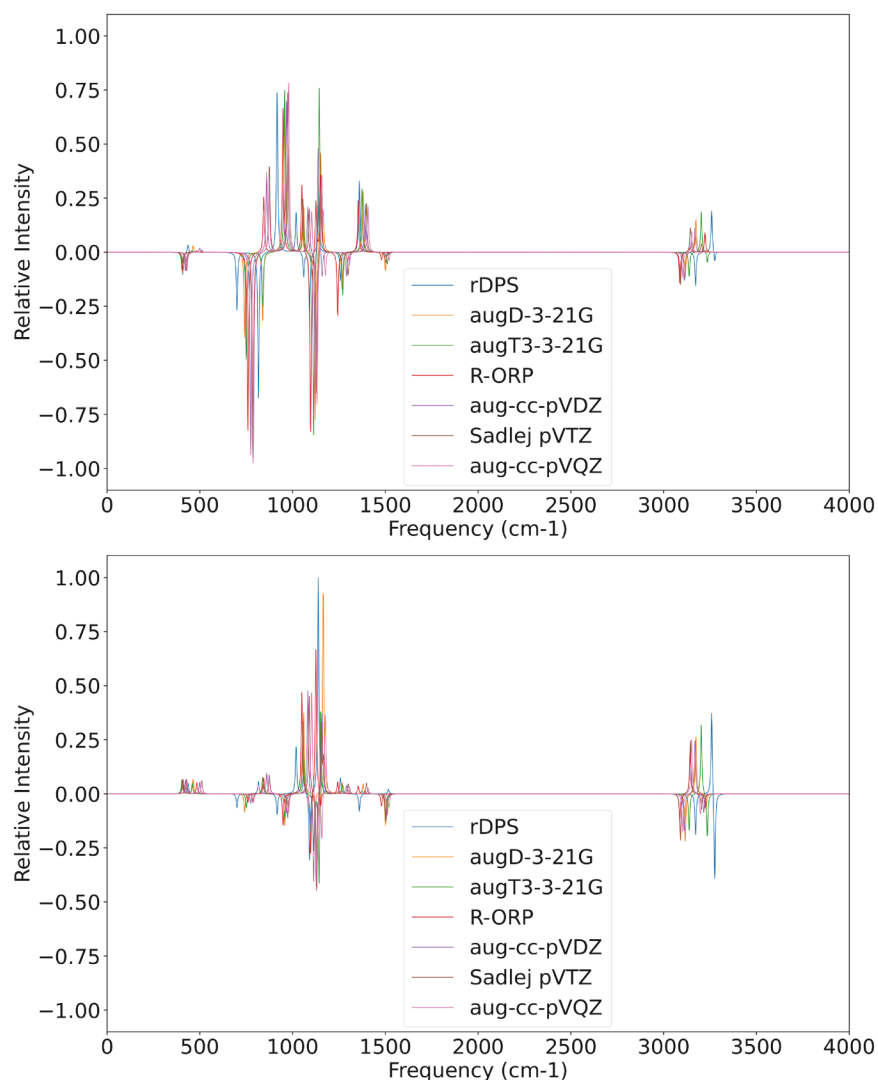


Figure 6. VCD (top) and ROA (bottom) spectra of (*R*)-fluorooxirane computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

4.4. (*S*)-Methyloxirane

(*S*)-Methyloxirane is similar in structure to (*R*)-fluorooxirane, but its methyl torsional vibrations provide additional low-frequency modes for comparison. The computed VCD and ROA spectra for (*S*)-methyloxirane are given in Figures 8 and 9 using the same basis-set grouping as in the earlier sections. The overlaps, integrated differences, and number of events obtained for (*S*)-methyloxirane are provided in Table 5.

Paralleling trends from our previous test molecules, the smaller basis sets including rDPS, augD-3-21G, augT3-3-21G, rORP, aug-cc-pVDZ, and Sadlej-pVTZ yielded poor SNO/DNO overlaps ranging from 0.09/0.10–0.21/0.21 for VCD and from –0.02/–0.03–0.47/0.39 for ROA. The exceptions to these ranges were the rDPS ROA spectrum which had an SNO/DNO of 1.13/0.35 (with an integrated difference –9.20 marking a

perfect example of the type of inflation the SNO function can experience) and the Sadlej-pVTZ VCD spectrum which had an SNO/DNO of 0.54/0.63. On average, these bases exhibited three or more events for both spectroscopies.

A drastic improvement was noted for the ORP, LPol-ds, LPol-dl, and LPol-fs sets which yielded SNO/DNO values between 0.69/0.68–0.78/0.80 and 0.62/0.61–0.77/0.76 for VCD and ROA, respectively. Of import, was that most of these basis sets had absolute errors from the integrated difference function below 0.10, which is rather reasonable. Additionally, these bases averaged one event for VCD and two events for ROA.

The remaining sets – LPol-fl and aug-cc-pVTZ – performed at nearly the aug-cc-pVQZ level with SNO/DNO greater than 0.95/0.94 for VCD and 0.97/0.92 for ROA. Interestingly, the VCD spectra yielded absolute errors for LPol-fl and aug-cc-pVTZ of 0.01 for VCD with no events.

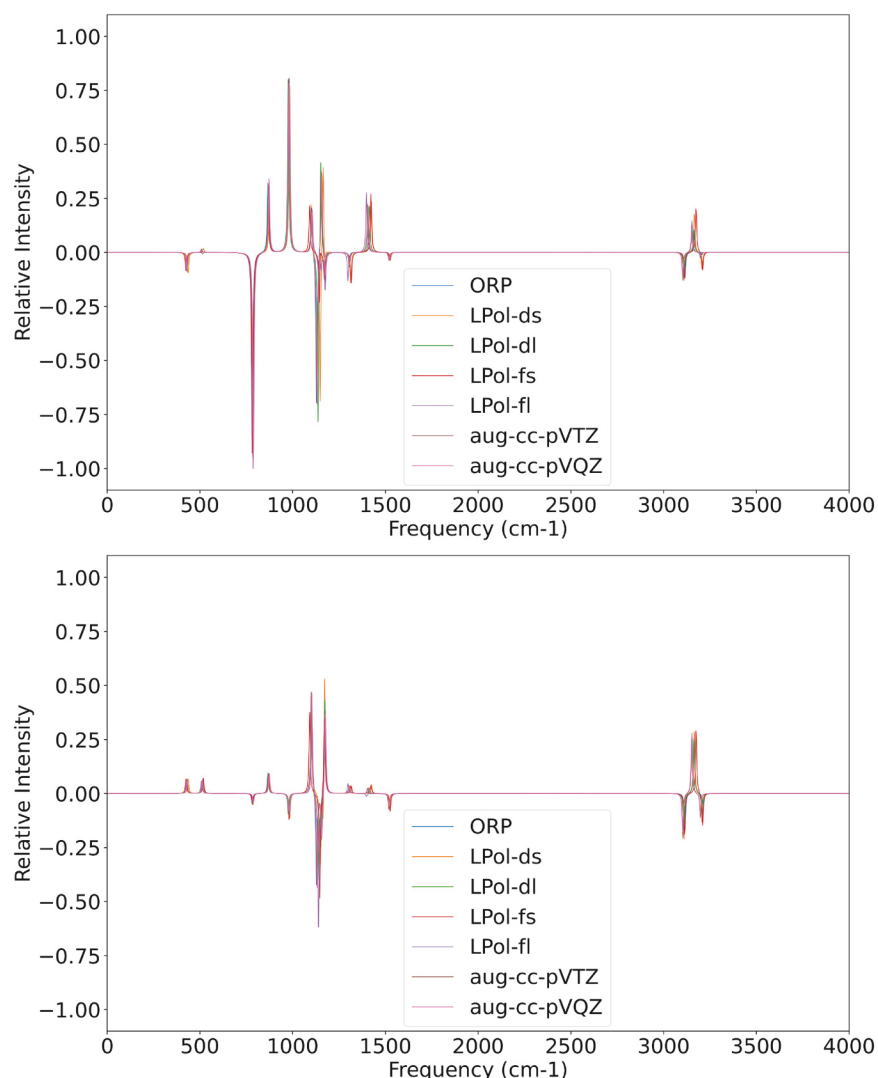


Figure 7. VCD (top) and ROA (bottom) spectra of (*R*)-fluorooxirane computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 4. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of (*R*)-fluorooxirane using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	−0.0092	−0.0120	0.4135	3	−0.4166	−0.2900	−1.0643	3
augD-3-21G	0.1515	0.1761	0.2603	3	0.0625	0.0477	−0.7208	0
augT3-3-21G	0.0104	0.0107	0.0643	1	−0.0679	−0.0723	0.1189	0
R-ORP	−0.0165	−0.0167	0.0290	2	−0.2864	−0.2687	−0.1362	2
aug-cc-pVDZ	0.1009	0.1014	0.0098	1	−0.0061	−0.0063	0.0532	0
Sadlej-pVTZ	0.7452	0.7756	0.0767	1	0.3019	0.3073	0.0346	1
ORP	0.7052	0.7291	0.0645	0	0.7970	0.8157	0.0454	0
LPol-ds	0.5872	0.6176	0.0961	0	0.5983	0.5925	−0.0196	0
LPol-dl	0.7708	0.7769	0.0157	0	0.7658	0.7525	−0.0357	0
LPol-fs	0.5156	0.6184	0.3048	0	0.3368	0.3489	0.0682	0
LPol-fl	0.6903	0.7723	0.2012	2	0.7454	0.7289	−0.0457	1
aug-cc-pVTZ	0.9687	0.9707	0.0039	0	0.9175	0.9258	0.0179	0
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

In contrast, the ROA spectra for this metric increased to 0.11 and 0.09, respectively with one event each.

For (*S*)-methyloxirane LPol-ds is clearly the preferred basis for VCD and ROA calculations with good

performance at a moderate cost. Though, once again, we focus our discussion of events around the larger basis sets which make for reasonable candidates for use in the development of robust computational protocols for

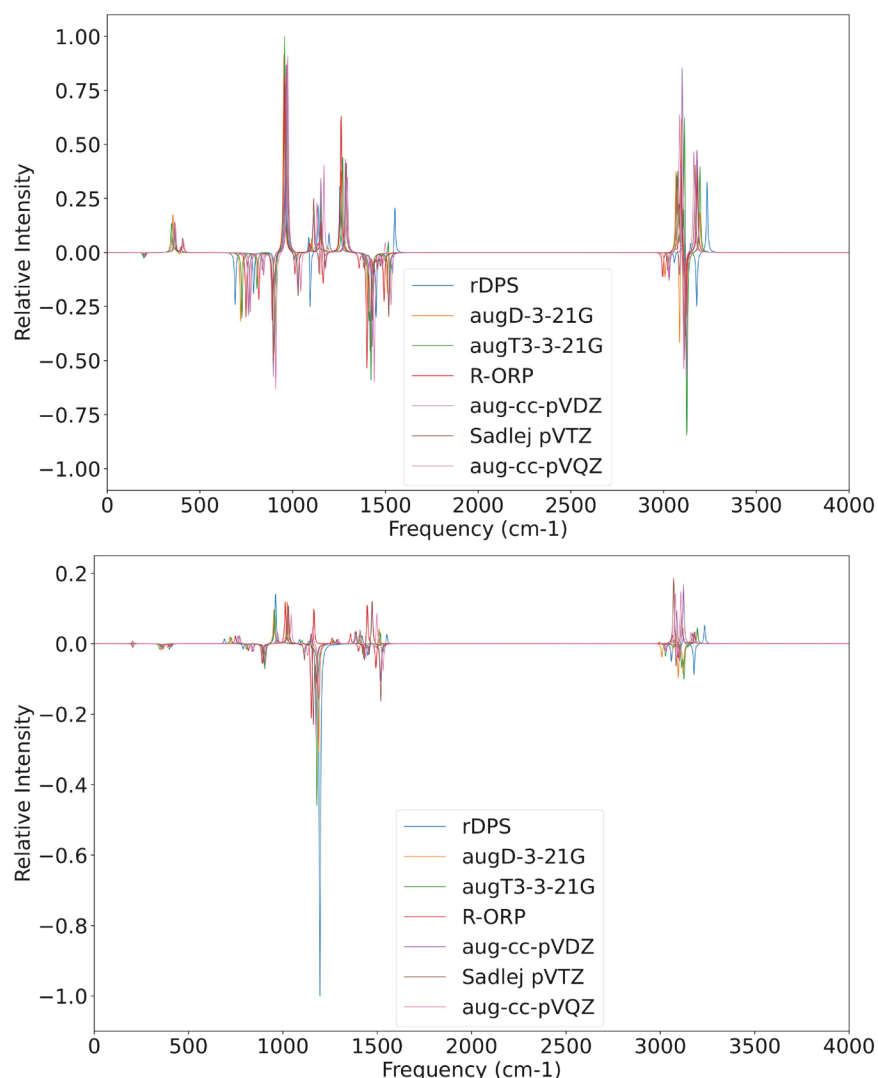


Figure 8. VCD (left) and ROA (right) spectra of (*S*)-methyloxirane computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

simulating glycans in solution. The single event noted for the ORP basis in the VCD spectrum occurs at 3091.52 cm^{-1} and corresponds to a sign change and normal mode reordering with a neighbouring peak at 3094.41 cm^{-1} (though neither is clearly observable in the simulated spectrum). In the ROA spectrum an event arises at 1168.68 cm^{-1} , which may be ascribed to hydrogen bending motions on all three carbon atoms. The resulting sign change is relatively subtle in the spectrum but does allow resolution of an additional peak not seen in the reference aug-cc-pVQZ spectrum. The same phenomenon as in the ORP VCD spectrum appears in the LPol-ds VCD spectrum at 3085.81 cm^{-1} with the neighbouring mode at 3082.23 cm^{-1} . The modes in question are associated with a symmetric C–H stretch on the achiral carbon in the epoxide ring and an asymmetric C–H stretching on the methyl group, respectively. In this case, the reordering does not cause significant changes in the

shape of the VCD spectrum with either basis set (even with our choice of linewidth). The ROA spectrum produced by the LPol-ds spectrum also yields events at these same normal modes, however, the impact is much more easily observed in the ROA spectrum due to the large shift in intensity relative to the aug-cc-pVQZ reference spectrum. Finally, the LPol-ds spectrum also exhibits events located at 1166.88 cm^{-1} and 3087.60 cm^{-1} in the ROA spectrum corresponding to the same events as the ORP ROA and ORP VCD spectra.

4.5. (1*S*,4*S*)-Norbornenone

The next test case, (1*S*,4*S*)-norbornenone, includes important new structural features such as the bicyclic cage and the carbonyl moiety. This chiral compound has been the focus of numerous computational and experimental studies, particularly regarding its electronic CD

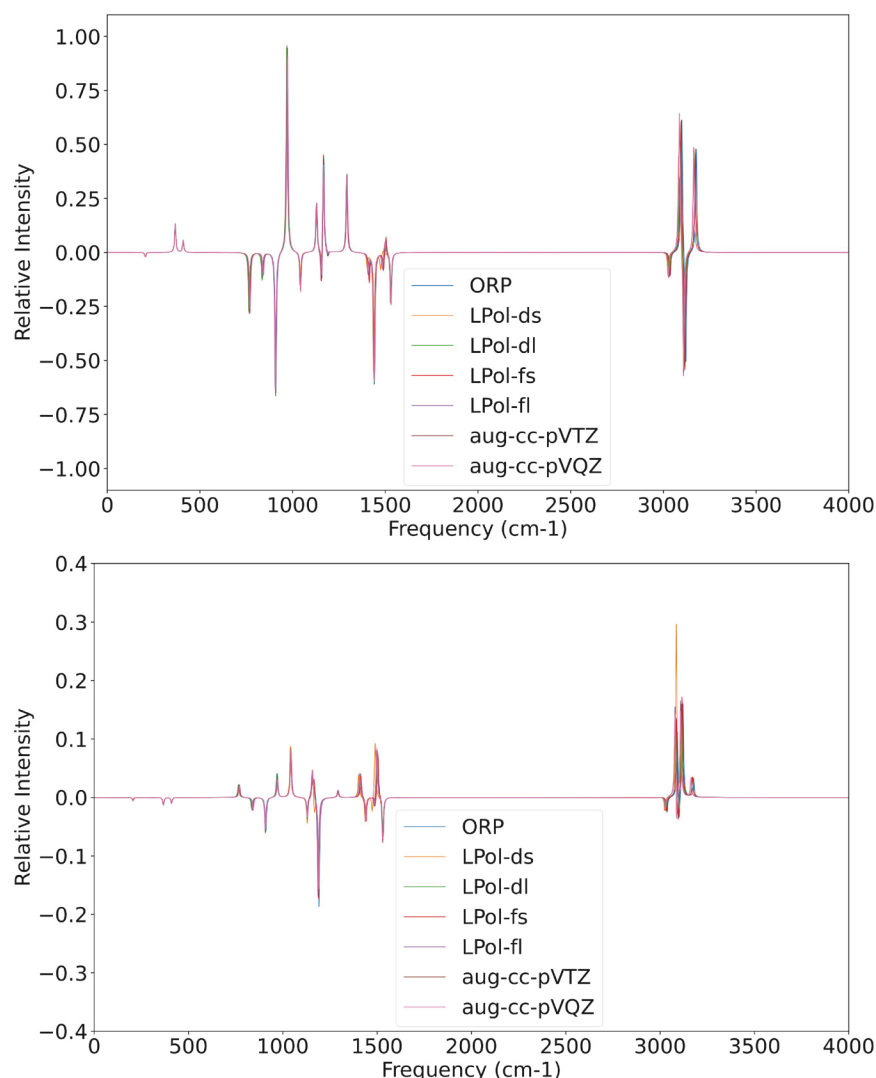


Figure 9. VCD (left) and ROA (right) spectra of (*S*)-methyloxirane computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 5. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of (*S*)-methyloxirane using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	0.1217	0.1691	0.4821	5	1.1259	0.3526	−9.1950	7
augD-3-21G	0.0889	0.0986	0.1876	5	0.4665	0.3881	−0.4447	7
augT3-3-21G	0.1368	0.1314	−0.0835	2	0.2094	0.1322	−1.5072	5
R-ORP	0.1679	0.1918	0.2338	2	−0.0245	−0.0256	0.0778	2
aug-cc-pVDZ	0.2088	0.2080	−0.0079	2	0.1976	0.1568	−0.5878	3
Sadlej-pVTZ	0.5438	0.6287	0.2519	2	0.1984	0.1518	−0.7094	1
ORP	0.6861	0.6765	−0.0285	1	0.6211	0.6058	−0.0509	1
LPol-ds	0.7656	0.8037	0.0924	1	0.7729	0.6291	−0.5095	2
LPol-dl	0.7816	0.7916	0.0250	0	0.7362	0.7666	0.0777	2
LPol-fs	0.7788	0.7719	−0.0177	1	0.6886	0.6525	−0.1137	1
LPol-fl	0.9495	0.9430	−0.0137	0	0.9705	0.9211	−0.1100	1
aug-cc-pVTZ	0.9867	0.9816	−0.0103	0	1.0109	0.9702	−0.0855	1
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

spectrum and its long-wavelength specific rotation. [37–39] The latter has proved particularly challenging to theoretical prediction due to its large magnitude and wide

variation between solution- and gas-phase properties. The computed VCD and ROA spectra for (*1S,4S*)-norbornenone are given in Figures 10 and 11 using

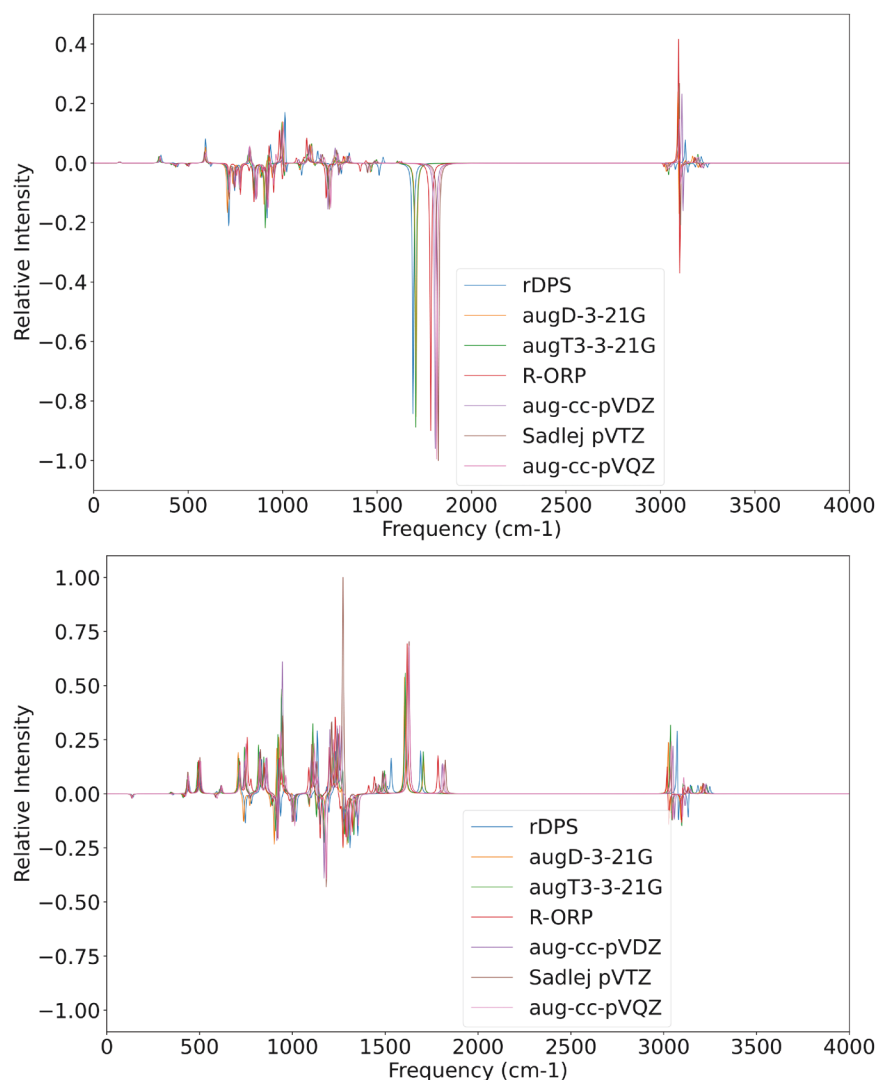


Figure 10. VCD (top) and ROA (bottom) spectra of (1*S*,4*S*)-norbornenone computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

the same basis-set grouping as in the previous section. The overlaps, integrated differences, and number of events obtained for (1*S*,4*S*)-norbornenone are provided in Table 6.

The results for (1*S*,4*S*)-norbornenone followed similar trends as the other molecules for the small rDPS, augD-3-21G, augT3-3-21G, and R-ORP bases, with overlaps between 0.03–0.07 for VCD and somewhat larger from 0.16–0.32 for ROA. The integrated difference values for the R-ORP set, however, are smaller than the other basis sets in this group by roughly an order of magnitude, which, again, suggests that at least some of the discrepancy is due to vibrational frequency shifting. On the other hand, the number of events observed for these small basis sets is large, in part, because of the increased number of vibrational modes for this medium-sized molecule. Visual inspection of the spectra is consistent with these numerical data, with

the r-DPS, augD-3-21G, and augT3-3-21G basis sets exhibiting significant frequency shifts for several large peaks and a number of sign errors associated smaller peaks.

As we have observed for other molecules in our test set, the aug-cc-pVDZ and Sadlej-pVTZ basis sets perform markedly better than their smaller counterparts (and better than R-ORP, which is identical in size to aug-cc-pVDZ), with SNO/DNO overlaps of 0.45/0.45 (VCD) and 0.66/0.58 (ROA) for the former and 0.56/0.55 (VCD) and a particularly impressive 0.83/0.62 (ROA) for the latter. Their integrated difference values are also reasonably small, though significantly worse for ROA than VCD, and with a reasonably small number of events. The next largest basis set, ORP, gives particularly good results for (1*S*,4*S*)-norbornenone, with overlaps of 0.93/0.94 for VCD and 0.84/0.82 for ROA and small integrated difference errors.

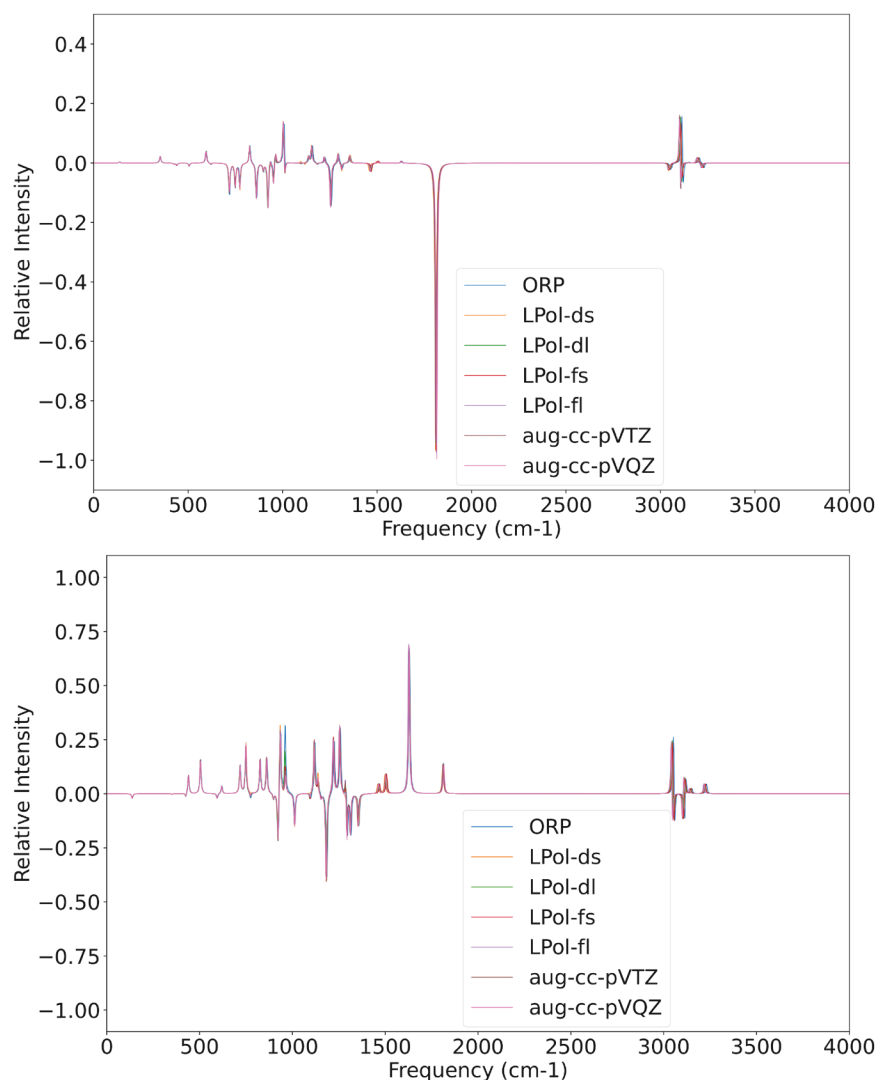


Figure 11. VCD (top) and ROA (bottom) spectra of (1*S*,4*S*)-norbornenone computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, LPol-fs, LPol-fl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 6. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of (1*S*,4*S*)-norbornenone using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	0.0543	0.0617	0.2254	9	0.2078	0.2499	0.3087	7
augD-3-21G	0.0277	0.0310	0.1970	5	0.1624	0.1799	0.1850	8
augT3-3-21G	0.0507	0.0557	0.1720	5	0.3162	0.3065	−0.0644	4
R-ORP	0.0731	0.0736	0.0136	9	0.2832	0.2984	0.0990	5
aug-cc-pVDZ	0.4533	0.4482	−0.0228	3	0.6618	0.5818	−0.2938	2
Sadlej-pVTZ	0.5648	0.5476	−0.0639	5	0.8334	0.6193	−0.8112	1
ORP	0.9331	0.9405	0.0156	1	0.8429	0.8171	−0.0639	2
LPol-ds	0.6390	0.6422	0.0099	0	0.9866	0.9831	−0.0070	0
LPol-dl	0.7358	0.7374	0.0045	1	0.9217	0.9140	−0.0169	1
LPol-fs	0.7559	0.7582	0.0062	1	0.9241	0.9132	−0.0242	1
LPol-fl	0.7635	0.7636	0.0004	2	0.9351	0.9419	0.0143	1
aug-cc-pVTZ	0.9168	0.9173	0.0012	2	0.9752	0.9803	0.0103	0
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

Interestingly, the LPol-X basis sets all perform approximately the same in spite of the variation in sizes from LPol-ds (392 functions) to LPol-fl (784 functions) with

SNO/DNO overlaps ranging from 0.64/0.64–0.76/0.76 for VCD and significantly better for ROA from 0.92/0.91–0.98/0.98 (and only a handful of sign events). Only the

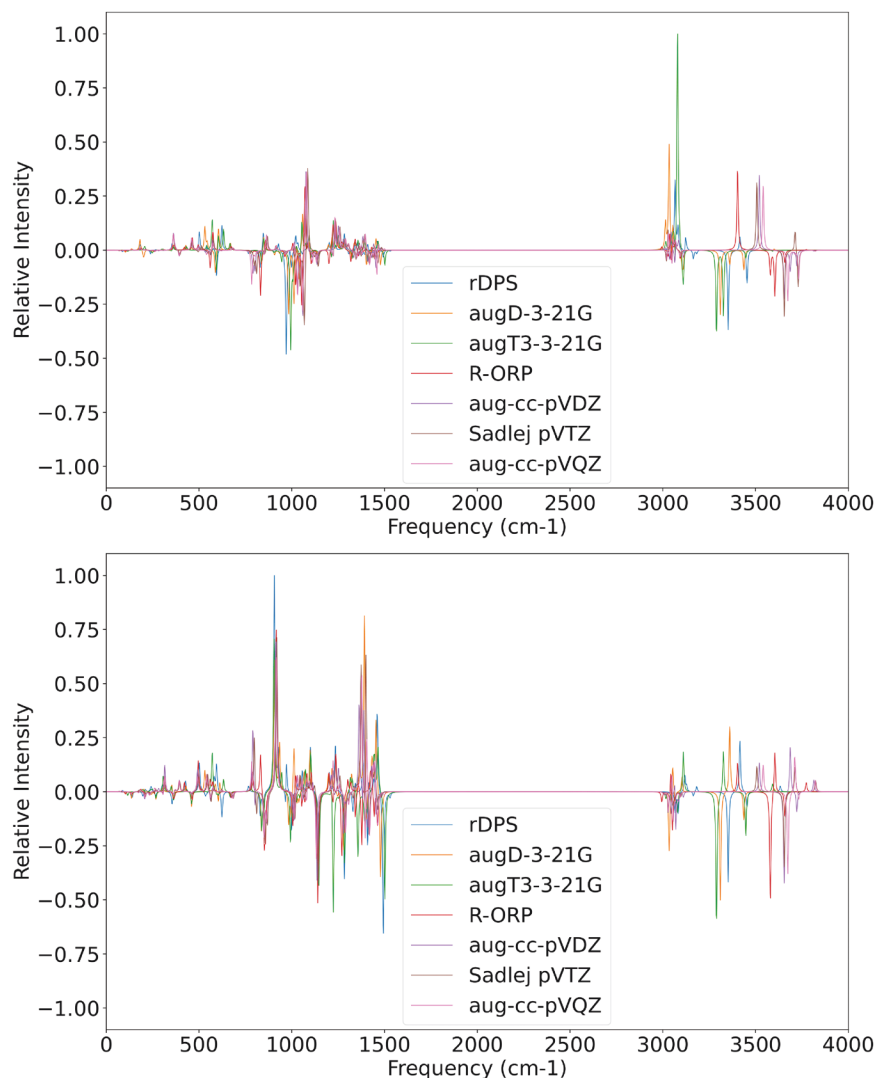


Figure 12. VCD (top) and ROA (bottom) spectra of β -D-glucose computed with B3LYP across the entire spectral region using the rDPS, augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, Sadlej-pVTZ, and aug-cc-pVQZ basis sets.

large aug-cc-pVTZ basis set yields comparably strong performance for both VCD and ROA with all overlaps greater than 0.9 and integrated difference values 0.01 or less. In addition, for all of the basis sets from ORP and larger, visual inspection of the spectra reveals few apparent differences relative to aug-cc-pVQZ. The most significant discrepancy occurs in the ROA spectrum at 961 cm^{-1} for ORP and 959 cm^{-1} for LPol-dl, which is an antisymmetric hydrogen out-of-plane bending motion. For this mode, both basis sets yield significantly larger scattering intensity differences compared to the reference basis set, though the sign of the peak is still correct.

4.6. β -D-Glucose

The last test molecule is β -D-glucose, which we have selected in part because of its less-rigid ring structure

compared to (1*S*,4*S*)-norbornenone. In addition, the longer-term goal of this work is to develop a computational protocol for simulating the vibrational chiroptical spectra of glycans, of which β -D-glucose, as the most abundant monosaccharide, is among the more important building blocks. [40] Although β -D-glucose is conformationally flexible, our inquiry in this work focuses only on basis-set convergence, and thus we have selected only the lowest-energy configuration of the β epimer. The computed VCD and ROA spectra for β -D-glucose are given in Figures 12 and 13 using the same basis-set grouping as in the previous section. The overlaps, integrated differences, and number of events obtained for β -D-glucose are provided in Table 7.

As with the previous test cases, the smaller rDPS, augD-3-21G, augT3-3-21G, and R-ORP basis sets performs relatively poorly, though R-ORP was clearly the best of this group. While the SNO/DNO values for the

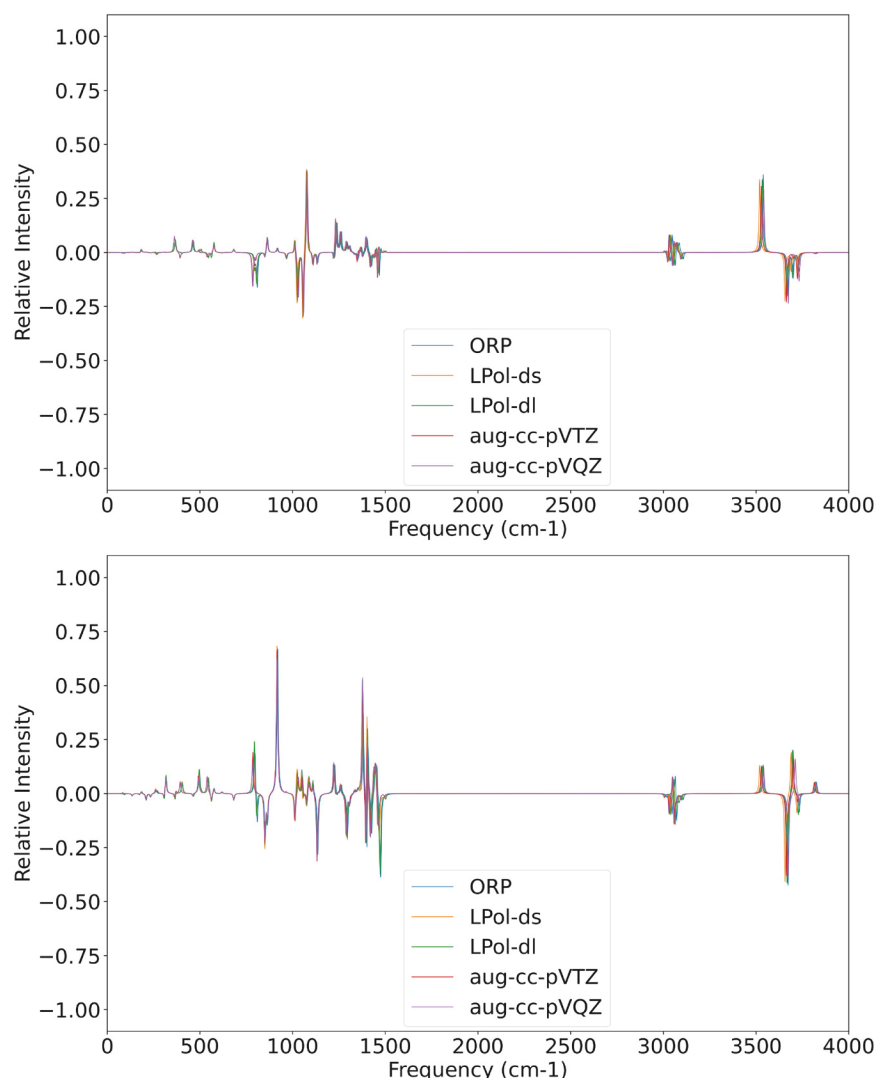


Figure 13. VCD (top) and ROA (bottom) spectra of β -D-glucose computed with B3LYP across the entire spectral region using the ORP, LPol-ds, LPol-dl, aug-cc-pVTZ, and aug-cc-pVQZ basis sets.

Table 7. Overlaps, integrated differences, and number of events for the VCD and ROA spectra of β -D-glucose using B3LYP.

Basis Set	VCD				ROA			
	SNO	DNO	Int. Diff.	Events	SNO	DNO	Int. Diff.	Events
rDPS	0.0626	0.0598	−0.0976	29	0.2646	0.2003	−0.7454	25
augD-3-21G	−0.0014	−0.0013	−0.2147	27	0.2169	0.1692	−0.6437	27
augT3-3-21G	0.0090	0.0051	−2.0544	26	0.1175	0.0922	−0.6244	22
R-ORP	0.2148	0.2231	0.0724	16	0.4037	0.3685	−0.1997	10
aug-cc-pVDZ	0.5827	0.5903	0.0256	4	0.5468	0.5124	−0.1385	4
Sadlej-pVTZ	0.3863	0.3648	−0.1211	9	0.7321	0.6389	−0.3130	6
ORP	0.7551	0.7403	−0.0403	3	0.7225	0.6901	−0.0960	2
LPol-ds	0.4949	0.4883	−0.0271	1	0.7843	0.7762	−0.0210	2
LPol-dl	0.6814	0.6709	−0.0316	1	0.8803	0.8144	−0.1684	2
aug-cc-pVTZ	0.7732	0.7743	0.0028	0	0.8703	0.8865	0.0363	1
aug-cc-pVQZ	1.0000	1.0000	0.0000	0	1.0000	1.0000	0.0000	0

first three basis sets range from -0.001 – 0.06 for VCD and from 0.10 – 0.26 for ROA, the R-ORP basis yields values of $0.21/0.22$ for VCD and $0.40/0.37$ for ROA, as well as smaller values of the integrated difference error. Not surprisingly, the large number of vibrational modes for

β -D-glucose (66) results in a larger number of sign discrepancies, as many as 29 in the VCD spectrum with the rDPS basis set, many of which are clearly discernible in both the VCD and ROA spectra, especially in the higher-frequency domains.

The aug-cc-pVDZ and Sadlej-pVTZ basis sets once again perform similarly to one another, though the latter exhibits a significant increase in SNO/DNO values between VCD (0.39/0.36) vs. ROA (0.73/0.64), while the accuracy of the former is relatively constant at 0.58/0.59 for VCD and 0.55/0.51 for ROA. Integrated difference errors are relatively small for these basis sets, apart from that for the ROA spectrum predicted by Sadlej-pVTZ basis with a value of -0.31 . The most prominent visual difference between the reference aug-cc-pVQZ spectra and that produced by the aug-cc-pVDZ and Sadlej-pVTZ basis sets occurs just near 800 cm^{-1} in the ROA (a hydroxyl torsional motion); the smaller basis sets exhibit a strong, positive Cotton effect, while the larger basis set shows no sign change, but instead a strong positive peak with a small shoulder. Interestingly, the ORP basis set exhibits comparable SNO/DNO values for both VCD and ROA (ranging from 0.69–0.76) and integrated difference values that are significantly smaller than those of the Sadlej-pVTZ basis set, but, like its smaller counterpart, also incorrectly exhibits the positive Cotton peaks near 800 cm^{-1} in its ROA spectrum. However, the ORP basis set also exhibits fewer events/sign errors as compared to the Sadlej basis set, which is also a contributing factor to its strong overlaps and small errors.

We observe a larger performance difference between LPol-ds and LPol-dl for β -D-glucose than for most of the other molecules in this study, particularly for VCD for which the two basis sets give SNO/DNO values of 0.49/0.49 and 0.68/0.67, respectively. The difference is reduced for the ROA spectrum, for which the two basis sets give 0.78/0.77 (LPol-ds) and 0.88/0.81 (LPol-dl), and both exhibit only one or two sign discrepancies and small integrated difference errors. Once again, however, the ROA spectrum with both basis set displays the same erroneous Cotton effect at 800 cm^{-1} as observed

for the smaller sets, though the rest of the spectrum bears strong visual similarity to the aug-cc-pVQZ reference spectrum. This spectral feature is finally reproduced correctly with the aug-cc-pVTZ basis set, which also exhibits strong overlaps, small integrated difference errors, and zero or one sign discrepancies, as compared to the reference. Of note, however, is the approximately 20% deviation of the aug-cc-pVTZ VCD spectrum from that of the reference, which we attribute to frequency shifts in the high frequency region. Indeed, focussing on the $0 - 2000\text{ cm}^{-1}$ region, the SNO/DNO for the aug-cc-pVTZ basis set is very high at 0.9625/0.9599 while the $2000 - 4000\text{ cm}^{-1}$ region produces SNO/DNO values of 0.3968/0.4007. The integrated difference values in these two regions, on the other hand, are -0.0054 and 0.0192 , respectively indicating that the positions of the vibrational modes from these calculations are to blame for discrepancy. (See Tables S90 and S91 in the Supplementary Information.) We were unable to obtain VCD and ROA spectra for β -D-glucose using the LPol-fs and LPol-fl basis sets due to numerical instabilities in the Kohn-Sham self-consistent-field procedure during the geometry optimisation step. Thus, we are unable to make a performance comparison for those basis sets for this test case.

5. Discussion

Table 8 summarises the SNO and DNO values, as well as integrated difference errors, for each type of spectrum, averaged across the set of test molecules for each basis set. From these data, we can reasonably classify the basis sets into four groups based on their relative performance: (1) the small rDPS, augD-3-21G, augT3-3-21G, and R-ORP sets; (2) aug-cc-pVDZ and Sadlej-pVTZ; (3) ORP, LPol-ds, LPol-dl, and LPol-fs; and (4) LPol-fl and aug-cc-pVTZ.

Table 8. Average overlaps and integrated differences for the basis sets across all molecules in the test set for the VCD and ROA spectra computed with B3LYP across the entire spectral region.

Basis Set	VCD			ROA		
	SNO	DNO	Int. Diff.	SNO	DNO	Int. Diff.
rDPS	0.1401	0.1336	0.2011	0.2744	0.1037	-4.6420
augD-3-21G	0.0892	0.0973	0.1553	0.2000	0.1550	-1.0193
augT3-3-21G	0.1079	0.1121	-0.2629	0.1375	0.0878	-1.2809
R-ORP	0.1936	0.1844	0.0051	0.0868	0.0802	-0.1998
aug-cc-pVDZ	0.3916	0.3901	-0.0126	0.3840	0.3584	-0.1527
Sadlej-pVTZ	0.5344	0.5416	0.0018	0.4459	0.3803	-0.4229
ORP	0.7056	0.7069	0.0017	0.7236	0.7020	-0.0647
LPol-ds	0.6748	0.6830	0.0222	0.8030	0.7669	-0.1160
LPol-dl	0.7518	0.7533	0.0027	0.8385	0.8154	-0.0611
LPol-fs	0.7634	0.7885	0.0708	0.7315	0.7121	-0.0531
LPol-fl	0.8386	0.8556	0.0416	0.8948	0.8864	-0.0206
aug-cc-pVTZ	0.8924	0.8927	0.0013	0.9041	0.8972	-0.0163
aug-cc-pVQZ	1.0000	1.0000	0.0000	1.0000	1.0000	0.0000

The basis sets in *Group 1* typically had poorer overlaps and absolute integrated difference values, and they exhibited a much larger number of events/sign discrepancies as compared to the larger basis sets, recognising, of course, that these basis sets were designed principally for efficient computations of optical rotation (augD-3-21G, augT3-3-21-G, and R-ORP) or ROA spectra using harmonic force fields from higher levels of theory (rDPS). (The current study has taken the route of matching the level of theory used for the geometry optimisation and vibrational computation with that used for the rotatory strength/scattering-intensity difference, which is the common protocol in quantum chemical packages.) *Group 2* includes the Sadlej-pVTZ and aug-cc-pVDZ basis sets, with SNO/DNO overlaps roughly a factor 2–2.5 larger than those in *Group 1*. These bases exhibited larger variances in their metrics as compared to the larger basis sets, but significantly smaller than those in *Group 1*. The ORP, LPol-ds, and LPol-dl, and LPol-fs basis sets comprise *Group 3*, which generally performed well across our molecular test set, though clearly not as well as their larger counterparts. Based on both relative overlap values and integrated differences, the ORP basis stands out in this group, providing perhaps the best price/performance ratio. *Group 4* consists of the LPol-fl and aug-cc-pVTZ basis sets, though the latter is clearly superior to the former based on both its better metrics and smaller size. (In addition, it is noteworthy that the LPol-fs in *Group 3* and the LPol-fl basis set in *Group 4* exhibited numerical problems – convergence sensitivity in the Kohn-Sham self-consistent field algorithm – for the (P)-2,3-pentadiene (CAM-B3LYP) and β -D-glucose (both B3LYP and CAM-B3LYP.)

Based on these apparent groupings, Table 9 reports the corresponding metric averages for comparison and analysis. For Groups 1, 3, and 4, we observe slight improvements in the performance of the ROA spectra over that of the VCD. It is possible that this shift is related to the design of many of these basis sets to describe molecular response/field-dependent properties, but larger tests sets will be necessary to assess such a conclusion. In addition, the basis sets in *Group 2* more favourably predict VCD rotatory strengths than the ROA CIDs, but

the data collected here suggest that they may ultimately provide the best balance of efficiency and accuracy for conventional applications for both spectra. Furthermore, given that the average integrated difference values for ROA were all poorer than for VCD across all basis-set groups, it is clear that more stringent requirements must be met for the basis sets to accurately predict ROA spectra than VCD.

The experimentally relevant ‘fingerprint’ region is typically considered to lie between $400 - 1600\text{ cm}^{-1}$, and thus we have carried out additional analyses focussed on this domain. Integrations over this window result in improved overlaps, but the relatively arbitrary choice of cutoff can significantly affect the values produced by the integrated difference function due to frequency shifts between basis sets. Thus, to ensure that relevant transitions just below and above this region are not improperly excluded, we have extended our integration from $0 - 2000\text{ cm}^{-1}$, above which we observe a wide gap of zero intensity until near 3000 cm^{-1} for all computed spectra. The values obtained for the overlaps and integrated difference functions for these ranges are presented in Tables 10 and 11, respectively.

In the fingerprint region, we observe a general improvement in performance for the VCD spectra for all basis sets, based on the SNO/DNO overlaps. The most significant improvement in VCD prediction is that of *Group 2*, while *Group 3* improved only slightly as compared to the full spectral domain. Integrated difference values for VCD do not change dramatically for the fingerprint region vs. the entire domain. For the ROA spectra, the SNO/DNO overlaps improve for all groups for the $0 - 2000\text{ cm}^{-1}$ region, especially for *Group 3*, which increases from 0.77/0.75 to 0.85/0.84. Changes in the integrated difference values for the ROA spectra slightly improve, but not nearly as significantly as the overlaps.

Given that our computations use the same vibrational Hessian for both VCD and ROA spectra for a given basis set, improvements in the SNO/DNO overlap values in conjunction with relatively small changes in the integrated difference errors suggest that the vibrational frequencies in the fingerprint region are more accurate compared to that of the full spectrum. This

Table 9. Average overlaps and integrated difference values for the groups across all molecules for the VCD and ROA spectra computed with B3LYP across the entire spectral region.

Group	VCD			ROA		
	SNO	DNO	Int. Diff.	SNO	DNO	Int. Diff.
Group 1	0.1327	0.1318	0.0246	0.1746	0.1067	−1.7855
Group 2	0.4630	0.4659	−0.0054	0.4150	0.3694	−0.2878
Group 3	0.7222	0.7305	0.0223	0.7760	0.7507	−0.0746
Group 4	0.8680	0.8758	0.0196	0.8998	0.8923	−0.0182
All Groups	0.4985	0.5027	0.0179	0.5274	0.4868	−0.6889

Table 10. Average overlaps and integrated difference values across all molecules for the VCD and ROA spectra computed with B3LYP in the fingerprint region between 0 – 2000 cm^{-1} .

Group	VCD			ROA		
	SNO	DNO	Int. Diff.	SNO	DNO	Int. Diff.
Group 1	0.1484	0.1509	0.1529	0.2372	0.1225	−2.3904
Group 2	0.5287	0.5378	0.0278	0.4305	0.3805	−0.3933
Group 3	0.7422	0.7492	0.0159	0.8529	0.8369	−0.0386
Group 4	0.9082	0.9176	0.0225	0.9089	0.9112	0.0040
All Groups	0.5281	0.5343	0.0660	0.5782	0.5254	−0.8990

Table 11. Average overlaps and integrated difference values across all molecules for the VCD and ROA spectra computed with B3LYP in the high-frequency region between 2000 – 4000 cm^{-1} .

Group	VCD			ROA		
	SNO	DNO	Int. Diff.	SNO	DNO	Int. Diff.
Group 1	0.1068	0.0935	−0.8713	0.0872	0.0862	−0.2449
Group 2	0.4211	0.3468	−0.4973	0.3211	0.3474	0.0732
Group 3	0.4643	0.4604	−0.1171	0.4721	0.4418	−0.1603
Group 4	0.9066	0.8648	−0.0947	0.8920	0.8602	−0.0727
All Groups	0.4038	0.3787	−0.4373	0.3802	0.3695	−0.1355

conclusion is supported by comparison to the overlaps in the higher-frequency C–H stretching regime (Table 11). The SNO/DNO overlaps between 2000 and 4000 cm^{-1} are on average 0.10 – 0.40 smaller than those of the fingerprint region for both VCD and ROA, with Group 3 exhibiting the largest shift. In addition, VCD integrated differences values are significantly worse for all the groups in the high-frequency domain vs. the fingerprint region and slightly improved for ROA, which we interpret to indicate that the absolute VCD intensities are generally overestimated in the C–H stretching region.

The overlap and integrated-difference metrics that we have chosen to make spectral comparisons result in very similar trends for the B3LYP and CAM-B3LYP functionals (see Supporting Information), but a number of interesting discrepancies arising for specific cases are noteworthy. The first major difference between the two functionals appears in the (*P*)-2,3-pentadiene ROA spectrum shown in Figure 14. At approximately 1100 cm^{-1} in the B3LYP spectrum, the augD-3-21G, R-ORP, aug-cc-pVDZ, and Sadlej-pVTZ basis sets incorrectly predict the opposite sign of a rather prominent peak compared to that given by the aug-cc-pVQZ basis. For this same peak the CAM-B3LYP functional produces a much more consistent result, with all basis sets yielding the same sign. This discrepancy between B3LYP and CAM-B3LYP does not appear in the VCD spectrum.

We note another discrepancy around 100 cm^{-1} for the ROA spectrum of β -D-glucose, as shown in Figure S25 (see Supplementary Information) where the rDPS, Sadlej-pVTZ, aug-cc-pVTZ, and aug-cc-pVQZ bases yield positive CIDs while the rest of the basis sets produce negative values in the B3LYP spectrum. In the

CAM-B3LYP computations, on the other hand, all the basis sets give the same sign for this low-frequency torsional motion. As before, this discrepancy between B3LYP and CAM-B3LYP does not appear in the VCD spectrum.

In (1*S*,4*S*)-norbornenone, we note two discrepancies in the ROA spectrum between B3LYP and CAM-B3LYP, as shown in Figures S26 and S27 (Supplementary Information). The first occurs near 350 cm^{-1} , in which augD-3-21G, augT3-3-21G, R-ORP, aug-cc-pVDZ, and Sadlej-pVTZ all give the incorrect sign of the scattering intensity difference in the B3LYP spectrum but, in the CAM-B3LYP spectrum, all basis sets give a consistent sign. The second discrepancy lies around 725 cm^{-1} where a consistent sign is observed among the basis sets, but inverted between the functionals. Once again, neither of these are distinguishable in the VCD spectrum.

6. Conclusion

The objective of this work was to determine if basis sets designed for electric-field-dependent response properties are also effective for simulating vibrational response properties. To that end, we selected a broad range of ten property-oriented basis sets, as well as three standard augmented correlation-consistent basis sets, to simulate the VCD and ROA spectra of six test molecules: (*P*)-hydrogen peroxide, (*P*)-2,3-pentadiene, (*R*)-fluorooxirane, (*S*)-methyloxirane, (1*S*,4*S*)-norbornenone, and β -D-glucose. Our chosen metrics include overlaps between the spectrum obtained with each test basis set (in conjunction with standard B3LYP and CAM-B3LYP functionals), as well as an integrated difference (error), and the number of sign-discrepancies ('events')

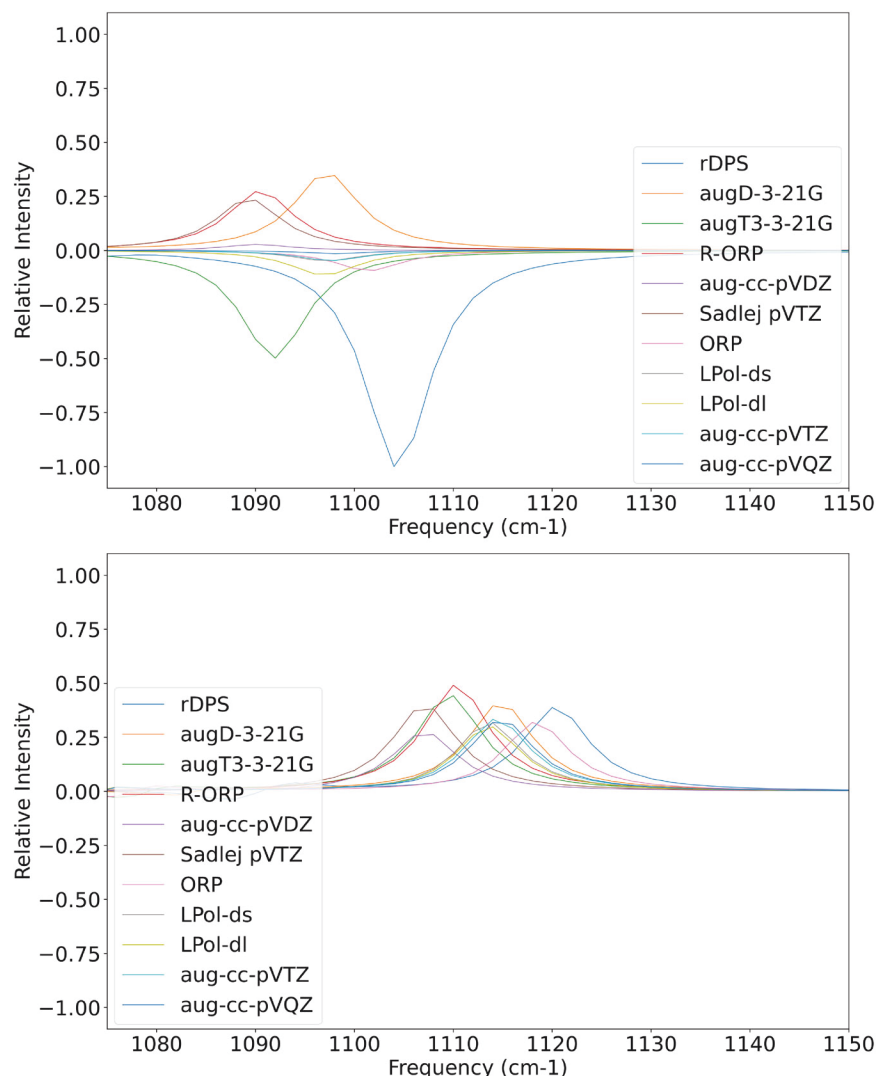


Figure 14. Discrepancy between B3LYP (top) and CAM-B3LYP (bottom) ROA spectra of (*P*)-2,3-pentadiene near 1100 cm^{-1} .

for each vibrational mode, all determined relative to the spectra obtained with the large aug-cc-pVQZ basis set. Based on these metrics – as well as visual comparison of the various spectra – we were able to organise the basis sets into four groups of increasing accuracy.

The smallest group, comprised of the rDPS, augD-3-21G, augT3-3-21G, and R-ORP basis sets, generally yielded poor VCD and ROA spectra compared to the QZ reference basis set and thus is not recommended for robust results. The second group (aug-cc-pVDZ and Sadlej-pVTZ) performed reasonably well in most cases and may ultimately yield the best price/performance ratio among all the sets considered here. For the larger basis sets in groups three (ORP, LPol-ds, LPol-dl, and LPol-fs) and four (LPol-fl and aug-cc-pVTZ), however, the ROA and VCD spectra were reproduced well compared to the much larger aug-cc-pVQZ basis.

In addition, we found that the smaller basis sets performed better in the prediction of VCD rotatory

strengths as compared to ROA CIDs. This feature diminished as basis set size increase with the performance being nearly indistinguishable between groups three and four. Additionally, vibrational frequencies were better predicted in the low-frequency (‘fingerprint’) region, $0 - 2000\text{ cm}^{-1}$, of the spectrum resulting in improved overlaps between the spectra in this region as compared to the high-frequency region, $2000 - 4000\text{ cm}^{-1}$. Interestingly, we also observed overestimation of VCD rotatory strengths by most of the basis sets for the high-frequency region, but not for the ROA CIDs.

The performance of the basis sets revealed promising results for the two-fold aim of property-oriented basis sets. The Sadlej-pVTZ basis set performs at approximately the same level as the slightly smaller aug-cc-pVDZ set, though, it is known that for some molecules more robust basis sets are required for accurate property prediction. The LPol-dl basis typically yields better VCD and ROA spectra compared to the ORP and LPol-ds sets

making it a good alternative when higher accuracy is required, though its usage does come at greater computational expense (it is ca. 40% larger than the ORP basis set).

On average, the LPol-fs and LPol-fl basis set provide better performance than their smaller counterparts, but given their large size (LPol-fl is approximately 2.2 times larger than ORP) and minor performance improvements (as well as producing numerical instabilities in the Kohn-Sham SCF procedure for (*P*)-2,3-pentadiene and β -D-glucose), we conclude that they are not ideal candidates for usage in robust computational protocols for simulating the VCD and ROA spectra. On the other hand, the performance of the ORP and LPol-ds basis sets is particularly noteworthy: on average, they provided highly accurate results at a significantly reduced size (aug-cc-pVQZ is nearly three times larger than the ORP basis set). We therefore recommend these basis sets for standard computational protocols for the prediction of ROA and VCD spectra. However, if the size of the molecular system precludes their use, the aug-cc-pVDZ or Sadlej-pVTZ basis sets may provide sufficient accuracy.

Acknowledgments

The authors are grateful to Advanced Research Computing at Virginia Tech for providing computational resources and technical support that have contributed to the results reported within the paper. The authors would like to dedicate this work to the memory of Dr. Timothy J. Lee, a collaborator, mentor, and dear friend of TDC.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the U.S. National Science Foundation (grant DMR-1933525), and EHF was supported by an undergraduate research fellowship (grant DMR-2244483).

References

- [1] L.A. Nafie, *Chirality* **32** (5), 667–692 (2020). doi:10.1002/chir.v32.5
- [2] L. Barron, *Molecular Light Scattering and Optical Activity*, 2nd ed. (Cambridge University Press, New York, 2004).
- [3] P.L. Polavarapu, *Molecules* **21** (8), 1–16 (2016). doi:10.3390/molecules21081056
- [4] R. Aerts, J. Bogaerts, W. Herrebout and C. Johannessen, *Phys. Chem. Chem. Phys.* **24** (16), 9619–9625 (2022). doi:10.1039/D2CP00746K
- [5] J. Bogaerts, F. Desmet, R. Aerts, P. Bultinck, W. Herrebout and C. Johannessen, *Phys. Chem. Chem. Phys.* **22** (32), 18014–18024 (2020). doi:10.1039/D0CP03257C
- [6] P.J. Stephens and F.J. Devlin, *Chirality* **12**, 172–179 (2000). doi:10.1002/(ISSN)1520-636X
- [7] K.D.R. Eikås, M.T.P. Beerepoot and K. Ruud, *J. Phys. Chem. A* **126** (32), 5458–5471 (2022). doi:10.1021/acs.jpca.2c02953
- [8] W.J. Hehre, R.F. Stewart and J.A. Pople, *J. Chem. Phys.* **51** (6), 2657–2664 (1969). doi:10.1063/1.1672392
- [9] T.H. Dunning and P.J. Hay, in *Methods of Electronic Structure Theory*, edited by H. F. Schaefer, *Modern Theoretical Chemistry*, Vol. 3, Chap. 1 (Plenum, New York, 1977), pp. 1–27.
- [10] T.H. Dunning, *J. Chem. Phys.* **90** (2), 1007–1023 (1989). doi:10.1063/1.456153
- [11] R.A. Kendall, T.H. Dunning and R.J. Harrison, *J. Chem. Phys.* **96** (9), 6796–6806 (1992). doi:10.1063/1.462569
- [12] J. Almlöf and P.R. Taylor, *J. Chem. Phys.* **86** (7), 4070–4077 (1987). doi:10.1063/1.451917
- [13] G. Zuber and W. Hug, *J. Phys. Chem. A* **108**, 2108–2118 (2004). doi:10.1021/jp031284n
- [14] M. Reiher, V. Liégeois and K. Ruud, *J. Phys. Chem. A* **109**, 7567–7574 (2005). doi:10.1021/jp052123h
- [15] J.R. Cheeseman and M.J. Frisch, *J. Chem. Theory Comput.* **7** (10), 3027–3460 (2011). doi:10.1021/ct200507e
- [16] K. Jalkanen, P. Stephens, R. Amos and N. Handy, *Chem. Phys. Lett.* **142** (3), 153–158 (1987). doi:10.1016/0009-2614(87)80913-2
- [17] K. Scholten, E. Engelage and C. Merten, *Phys. Chem. Chem. Phys.* **22**, 27979–27986 (2020). doi:10.1039/D0CP05420H
- [18] J. Groß, J. Kühnborn, S. Pusch, C. Weber, L. Andernach, G. Renzer, P. Eckhardt, J. Brauer and T. Opatz, *Chirality* **35**, 753–765 (2023). doi:10.1002/chir.v35.10
- [19] F. Jensen, *Introduction to Computational Chemistry*, 3rd ed. (John Wiley & Sons, Ltd., Chichester, 2017).
- [20] A.J. Sadlej, *Collect. Czech. Chem. Commun.* **53** (9), 1995–2016 (1988). doi:10.1135/cccc19881995
- [21] A. Baranowska and A.J. Sadlej, *J. Comput. Chem.* **31**, 552–560 (2010). doi:10.1002/jcc.v31:3
- [22] A. Baranowska-Łączkowska and K.Z. Łączkowska, *J. Comput. Chem.* **34** (23), 2006–2013 (2013). doi:10.1002/jcc.v34.23
- [23] A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571–2577 (1992). doi:10.1063/1.463096
- [24] A. Baranowska-Łączkowska, K.Z. Łączkowski, C. Henriksen, B. Fernandez, M. Kozak and S. Zielinska, *RSC Adv.* **6** (24), 19897–19902 (2016). doi:10.1039/C5RA20186A
- [25] T. Aharon and M. Caricato, *J. Chem. Theory Comput.* **16** (7), 4408–4415 (2020). doi:10.1021/acs.jctc.0c00446
- [26] P.J. Stephens, *J. Am. Chem. Soc.* **89**, 748–752 (1985).
- [27] L.D. Barron, L. Hecht, I.H. McColl and E.W. Blanch, *Mol. Phys.* **102** (8), 731–744 (2004). doi:10.1080/00268970410001704399
- [28] S. Coriani, A.J. Thorvaldsen, K. Kristensen and P. Jørgensen, *Phys. Chem. Chem. Phys.* **13**, 4224–4229 (2011). doi:10.1039/c0cp02230f
- [29] B.P. Pritchard, D. Altarawy, B. Didier, T.D. Gibson and T.L. Windus, *J. Chem. Inf. Model.* **59** (11), 4814–4820 (2019). doi:10.1021/acs.jcim.9b00725
- [30] A.D. Becke, *J. Chem. Phys.* **98**, 5648–5652 (1993). doi:10.1063/1.464913

- [31] C. Lee, W. Yang and R.G. Parr, *Phys. Rev. B.* **37**, 785–789 (1988). doi:10.1103/PhysRevB.37.785
- [32] P.J. Stephens, F.J. Devlin, C.F. Chabalowski and M.J. Frisch, *J. Phys. Chem.* **98** (45), 11623–11627 (1994). doi:10.1021/j100096a001
- [33] T. Yanai, D.P. Tew and N.C. Handy, *Chem. Phys. Lett.* **393**, 51–57 (2004). doi:10.1016/j.cplett.2004.06.011
- [34] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.J.A. Montgomery, J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski and D.J. Fox, Gaussian 09 Revision A.1.
- [35] <https://github.com/CrawfordGroup/SoAPy>.
- [36] N. Kreienborg and C. Merten, *Phys. Chem. Chem. Phys.* **21**, 3506–3511 (2019). doi:10.1039/C8CP02395F
- [37] K. Ruud, P.J. Stephens, F.J. Devlin, P.R. Taylor, J.R. Cheeseman and M.J. Frisch, *Chem. Phys. Lett.* **373**, 606–614 (2003). doi:10.1016/S0009-2614(03)00667-5
- [38] P. Lahiri, K.B. Wiberg, P.H. Vaccaro, M. Caricato and T.D. Crawford, *Angew. Chem. Int. Ed. Engl.* **53**, 1386–1389 (2014). doi:10.1002/anie.v53.5
- [39] M. Caricato, P.H. Vaccaro, T.D. Crawford, K.B. Wiberg and P. Lahiri, *J. Phys. Chem. A* **118**, 4863–4871 (2014). doi:10.1021/jp504345g
- [40] B.O. Fraser-Reid, K. Tatsuta and J. Thiem, editors, *Glycoscience I*, Vol. 1, (Springer, Berlin, 2001).