JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

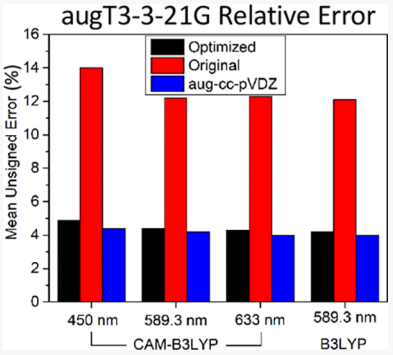
Compact Basis Sets for Optical Rotation Calculations

Tal Aharon and Marco Caricato*

Cite This: J. Chem. Theory Comput. 2020, 16, 4408−4415 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In this work, we present two compact basis sets optimized for the calculation of specific rotation: augD-3-21G and augT3-3-21G. They are obtained by combining the standard 3-21G basis set with the diffuse functions of aug-cc-pVDZ and aug-cc-pVTZ, respectively, followed by a reoptimization of the exponents of the diffuse functions. The exponent optimization is based on minimization of the root-mean-square relative error (RMSE) of the specific rotation computed at 589.3 nm (the sodium D line, $[\alpha]_D$) with CAM-B3LYP compared with the corresponding calculations using the full correlation-consistent basis sets. The training set comprises 21 chiral molecules with $|[\alpha]_D| > 50$ deg dm^{-1} (g/mL)$^{-1}$. For augT3-3-21G, the functions with the highest angular momentum are neglected, so that augD-3-21G and augT3-3-21G are of the same size. The exponents are optimized for four common elements in chiral organic molecules (H, C, N, and O), while the original exponents are maintained for other elements. Tests are conducted on the training set with CAM-B3LYP at 450 and 633 nm and with B3LYP at 589.3 nm; furthermore, a similar comparison is performed on a control set containing 30 more chiral molecules. A comparison with the optical rotatory prediction (ORP) basis set is also presented. The results show that the new compact basis sets are able to reproduce the calculations with the full Dunning basis sets remarkably well, and definitely better than before reoptimization of the exponents, with relative mean unsigned errors of around 4%. More significantly, augT3-3-21G is either of similar quality or better than aug-cc-pVDZ in reproducing the values obtained with aug-cc-pVTZ, even though augT3-3-21G is smaller than aug-cc-pVDZ. The larger ORP basis set outperforms both augT3-3-21G and aug-cc-pVDZ, but it requires a considerably larger computational effort. In summary, augT3-3-21G provides results that are in very good agreement with those obtained using aug-cc-pVTZ, but approximately 20 times faster, and it may be used for quick and reliable calculations of specific rotation of large chiral molecules.



1. INTRODUCTION

Optical activity has been a topic of continuous interest because of the fundamental role that chiral molecules play in pharmaceutical research. Because enantiomers of chiral molecules interact differently with a chiral environment, the correct determination of the absolute configuration (AC) of a molecule is of paramount importance in synthetic and biological fields. Because of the direct link between the structure of the molecule and its optical response, chiroptical spectroscopy and electronic structure calculations have become essential to determine the AC of chiral compounds.[1] Great progress has been made in developing quantum-chemical methods for performing calculations of chiroptical properties in both density functional theory (DFT) and coupled cluster (CC) methods.[2−17] Despite these advancements, chiroptical calculations on large systems are still prohibitively expensive, due to the need for large basis sets to obtain accurate results.[18]

There have been only a few attempts to address the issue of computational cost related to the size of the basis set. Crawford et al. implemented a localized orbital scheme in combination with the local correlation idea of Pulay and Saebø[19,20] to neglect CC amplitudes that do not significantly contribute to the calculation of the property.[21−23] They tested this method with three different localization schemes and found that it works well for pseudolinear systems like 1-fluoroalkanes. However, when the dimensionality of the system increases (e.g., with cage structures), the truncation threshold needs to be tightened to maintain the same level of accuracy, to the detriment of computational efficiency. We have also proposed an approach to select the most important molecular orbitals for the calculation of optical rotation with DFT methods.[24] The selection is based on $\tilde{S}_k$ analysis of the optical rotation in terms of one-electron orbital transitions,[25−27] but applied to the guess density of the linear response equations rather than to the converged response density. Our preliminary results showed very promising trends in reducing the computational cost.[24]

A different approach is to develop basis sets designed for specific properties. To the best of our knowledge, the only example of a basis set designed for chiroptical properties is the

Received: May 4, 2020
Published: June 11, 2020

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

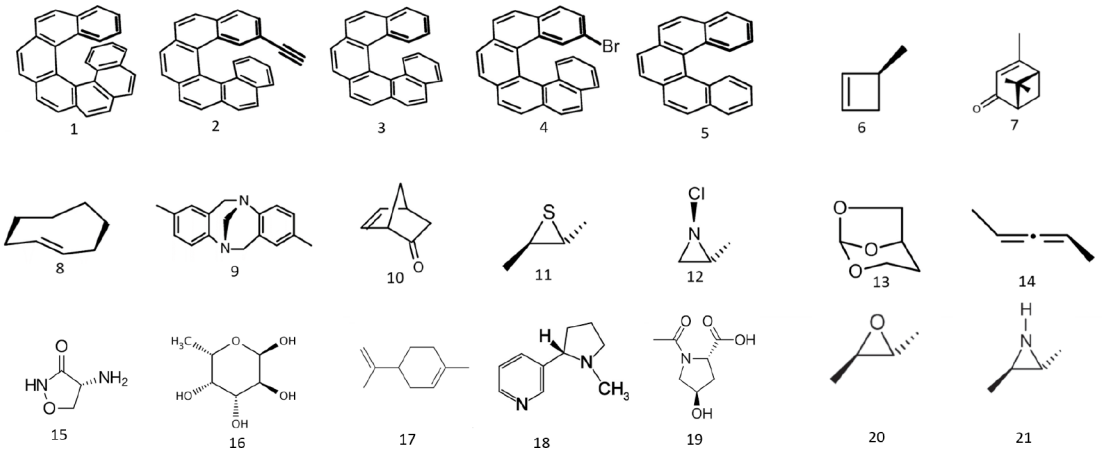


Figure 1. Structures of molecules in the training set.

optical rotatory prediction (ORP) basis set of Baranowska-Łączkowska and Łączkowski.[28,29] ORP was tested on a variety of chiral molecules, including particularly difficult flexible biological molecules,[29] and it outperformed the standard aug-cc-pVDZ basis set in terms of errors relative to aug-cc-pVTZ. More recently, Howard et al. tested the ORP basis set in combination with coupled cluster calculations and also found results comparable to those with aug-cc-pVTZ.[30] However, the size of ORP is between those of standard aug-cc-pVDZ and aug-cc-pVTZ, and it is not defined for all elements.

An alternative strategy for the construction of compact basis sets for optical activity was proposed by Wiberg et al.[31] Based on the hypothesis that chiroptical properties are based on the interaction of the external field with the periphery of the wave function, the authors suggested combining the 3-21G and STO-3G basis sets with the diffuse functions from aug-cc-pVDZ or aug-cc-pVTZ. Even without any exponent reoptimization, this work showed errors in specific rotation ($[\alpha]_\lambda$) on the order of 14% relative to the calculations with the full Dunning basis sets. In this work, we aim to further improve this remarkable result by optimizing the exponents of the s, p, and d diffuse functions from aug-cc-pVDZ and aug-cc-pVTZ in combination with the 3-21G basis set. We call these compact basis sets augD-3-21G and augT3-3-21G, respectively. The optimization is based on reproducing reference data on specific rotation at the sodium D line (589.3 nm, $[\alpha]_D$) computed with the full Dunning basis sets and the CAM-B3LYP functional, which was shown to be a reliable functional for this property,[18] for a training set of 21 chiral molecules. We then test the results at two other wavelengths (450 and 633 nm), with a different functional (B3LYP), and for a separate control set of 30 chiral molecules. We show that these compact basis sets are indeed able to reproduce the $[\alpha]_\lambda$ values computed with the corresponding full basis sets. Furthermore, we show that augT3-3-21G is competitive with the larger aug-cc-pVDZ and ORP basis sets in reproducing the aug-cc-pVTZ results and may be used for fast and reliable calculations of specific rotation.

This work is organized as follows. In section 2, we discuss the procedure for the optimization and report the resulting exponents. Section 3 contains the results of the calculations with our basis sets and the comparison with ORP, and section 4 contains a discussion of these tests and concluding remarks.

2. METHODS

The specific rotation (in deg dm^{-1} $(g/mL)^{-1}$) of a molecule in an isotropic medium is calculated as[32]

$$[\alpha]_\omega = \frac{(72.0 \times 10^6)\hbar^2 N_A \omega}{c^2 m_e{}^2 M} \times \left[\frac{1}{3}\mathrm{Tr}(\mathbf{G}'(\omega))\right] \quad (1)$$

where ω is the frequency of incident light in atomic units, M is the molecular mass in amu, m_e is the mass of the electron in kg, c is the speed of light in m/s, N_A is Avogadro's number, $\hbar$ is Planck's constant divided by 2π, and $\mathbf{G}'$ is the Rosenfeld mixed electric dipole−magnetic dipole polarizability tensor in atomic units, computed with standard linear response techniques.[33−36] Since it is customary to report the external field in terms of wavelength rather than frequency, we shall use the symbol $[\alpha]_\lambda$ rather than $[\alpha]_\omega$ in the following.

The two compact basis sets were built by augmenting the 3-21G basis set with the diffuse functions of aug-cc-pVDZ or aug-cc-pVTZ. These basis sets are named augD-3-21G and augT3-3-21G, respectively, where T3 indicates that only three sets of diffuse functions are included for the heavier elements (i.e., the f diffuse functions were not included, and for H the d diffuse functions were also not included). Wiberg et al. showed that valence basis functions are not very important for accuracy in the evaluation of $[\alpha]_\lambda$.[31] Therefore, we left the 3-21G functions unchanged and optimized the exponents of the diffuse functions for hydrogen, carbon, oxygen, and nitrogen, as they are very common elements in chiral organic molecules. Some of the molecules we used for training and testing contained other elements: F, Cl, Br, and S. Since these elements are not common, we did not optimize their exponents but instead retained those from the parent basis sets. The optimization is based on the minimization of the root-mean-square relative error (RMSE) in the calculation of $[\alpha]_D$ with the compact basis set using the $[\alpha]_D$ values computed with the corresponding correlation-consistent basis set as the reference. The training set includes 21 organic molecules with $|[\alpha]_D| > 50$ deg dm^{-1} $(g/mL)^{-1}$, which are shown in Figure 1. These molecules come from the OR45 set of Srebro et al.,[18] except for compounds **15**−**19**.

The minimization is based on the Newton−Raphson procedure. As analytical derivatives of $[\alpha]_D$ with respect to the exponents of the diffuse functions are complicated, we employed finite difference methods to compute the first and

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

Table 1. Exponents for augD-3-21G before and after Optimization

function	H	C	O	N
		Optimized		
s	1.823999×10^{-2}	1.769533×10^{-1}	7.702713×10^{-2}	3.172590×10^{-2}
p	1.019984×10^{-1}	1.924718×10^{-2}	4.945624×10^{-2}	9.601104×10^{-2}
d		1.984507×10^{-1}	1.988384×10^{-1}	4.998359×10^{-1}
		Original		
s	2.526000×10^{-2}	4.402000×10^{-2}	7.376000×10^{-2}	5.760000×10^{-2}
p	1.020000×10^{-1}	3.569000×10^{-2}	5.974000×10^{-2}	4.910000×10^{-2}
d		1.000000×10^{-1}	2.140000×10^{-1}	1.510000×10^{-1}

Table 2. Exponents for augT3-3-21G before and after Optimization

function	H	C	O	N
		Optimized		
s	2.341077×10^{-2}	1.964813×10^{-1}	3.568807×10^{-2}	3.078583×10^{-2}
p	1.016736×10^{-1}	4.837758×10^{-2}	7.727372×10^{-2}	1.544036×10^{-1}
d		1.720940×10^{-1}	2.533549×10^{-1}	5.596826×10^{-1}
		Original		
s	2.974000×10^{-2}	4.690000×10^{-2}	7.896000×10^{-2}	6.124000×10^{-2}
p	1.410000×10^{-1}	4.041000×10^{-2}	6.856000×10^{-2}	5.611000×10^{-2}
d		1.510000×10^{-1}	3.320000×10^{-1}	2.300000×10^{-1}

second derivatives of $[\alpha]_D$ (f' and f''in the following equations):

$$f'(x) = \lim_{\delta \to 0} \frac{f(x+\delta) - f(x-\delta)}{2\delta}$$
$$f''(x) = \lim_{\delta \to 0} \frac{f(x+\delta) + f(x-\delta) - 2f(x)}{\delta^2} \quad (2)$$

Because initial tests showed that the simultaneous optimization of all basis functions was unstable, we opted for an iterative procedure in which the functions for each element were optimized separately while the others were kept fixed at the current value. We started with H, followed by C, O, and N, and then the procedure was repeated for a new pass. Each element was considered converged when the change in the RMSE between two successive Newton−Raphson steps was less than 5%. After five passes across all of the elements, the RMSE stopped decreasing, and we performed a full optimization of all 11 exponents.

In Table 1, we report the original and optimized exponents for augD-3-21G, and a typical input for GAUSSIAN using these exponents can be found in Table S1 of the Supporting Information (SI). The s functions become more diffuse for hydrogen and nitrogen, slightly less diffuse for oxygen, and considerably less diffuse for carbon. All of the p functions except those for nitrogen become more diffuse. Carbon and nitrogen d functions become less diffuse with optimization. All of the hydrogen and most of the oxygen functions become more diffuse. The original and optimized exponents for augT3-3-21G are reported in Table 2, and a typical input for GAUSSIAN can be found in Table S2 of the SI. All of the s functions except that for carbon become more diffuse. All of the p functions except those for hydrogen become less diffuse. All of the d functions become less diffuse with optimization, except in the case of oxygen. The changes in the exponents after optimization are similar for the two basis sets, with a few exceptions: the carbon p functions become more diffuse during the augD-3-21G optimization and less diffuse in the augT3-3-21G optimization; the oxygen s functions become less diffuse during the augD-3-21G optimization and more diffuse in the augT3-3-21G optimization, while the opposite happens for the p functions.

In Table 3, we compare the numbers of basis functions for each element for different basis sets, including ORP. To get an idea of the relative cost of a calculation with each basis set, we note that although DFT formally scales as $O(N^4)$, practical implementations scale as $O(N^3)$ with respect to basis set size. Thus, we can estimate the speedup of the calculation in passing from the full correlation-consistent basis set to the corresponding augX-3-21G one (X = D, T3) with the computational ratio r_c:

Table 3. Numbers of Contracted Basis Functions for Each Element for Different Basis Sets

element	3-21G	aug(D/T3)-3-21G	aug-cc-pVDZ	ORP	aug-cc-pVTZ
H	2	6	9	13	23
C	9	19	23	32	46
O	9	19	23	32	46
N	9	19	23	32	46

$$r_c = \left(\frac{N^{\text{target}}}{N^{\text{opt}}} \right)^3 \quad (3)$$

where N^{target} is the number of basis functions of the full Dunning basis set and N^{opt} is the number of basis functions in our optimized basis set. We compare estimated average r_c values for the entire set to the actual average speedup (i.e., based on wall-clock timings) for the five largest molecules in the training set (**1**−**5** in Figure 1), as shorter calculation times on small molecules are more prone to variations due to computational noise.

All of the $[\alpha]_\lambda$ calculations were performed with the GAUSSIAN 09 suite of programs[37] on the Comet machines of the Extreme Science and Engineering Discovery Environment (XSEDE) supercomputing cluster. Newton−Raphson steps were performed using a locally modified version of the GAUSSIAN *gauopt* utility program. Geometries were opti-

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

mized using the CAM-B3LYP/aug-cc-pVDZ model chemistry.[38,39] The optimized geometries of the molecules in the training set can be found in the Supporting Information of ref 24 and in Tables S3−S7 of the SI of this work. Calculations of specific rotation were performed at the sodium D line (589.3 nm), 450 nm, and 633 nm with CAM-B3LYP[40−42] and at the sodium D line with B3LYP using aug-cc-pVDZ, aug-cc-pVTZ, ORP, and the optimized and unoptimized versions of the augX-3-21G basis sets. The values of $[\alpha]_\lambda$ for the training set can be found in Tables S12−S14, S18−S21, and S28 of the SI.

3. TEST CALCULATIONS

We performed a series of tests to determine the transferability of the augD-3-21G and augT3-3-21G basis sets. These tests included calculations on the molecules in the training set with CAM-B3LYP at two other wavelengths, 450 and 633 nm, and with B3LYP at 589.3 nm. In order to have a more comprehensive assessment, we also chose a second set of molecules, labeled the control set (see Figure 2). Most of these

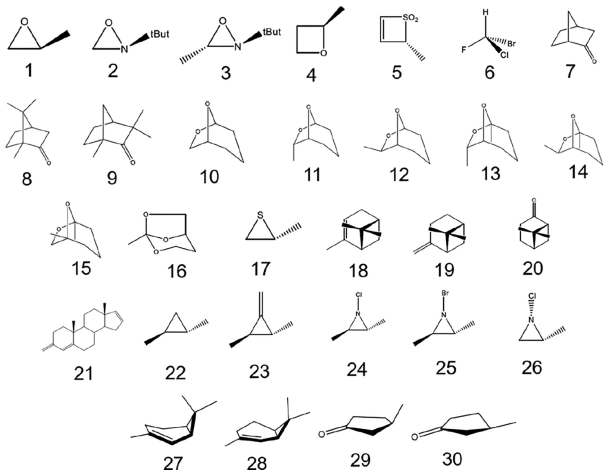


Figure 2. Structures of molecules in the control set.

molecules have $|[\alpha]_D| < 100$ deg dm^{-1} (g/mL)$^{-1}$, providing a more stringent test for the new basis sets. These compounds were taken from the OR45 set,[18] except for **27**−**30**. The optimized geometries of the molecules in the control set can be found in the Supporting Information of ref 24 and in Tables S8−S11 of the SI of this work. The values of $[\alpha]_\lambda$ for the control set are reported in Tables S15−S17, S22−S25, and S29 of the SI.

As mentioned in the previous section, the $[\alpha]_\lambda$ values computed with the full aug-cc-pVDZ basis set are used as the reference for the augD-3-21G basis set, and those computed with aug-cc-pVTZ are used as the reference for the augT3-3-21G basis set. The performance of the optimized exponents in the reduced basis set is compared with that obtained using the original values of the exponents. Results for the training set are reported in terms of relative (%) errors, whereas control set values are presented as absolute errors (calculated as $[\alpha]_\lambda^{opt} - [\alpha]_\lambda^{ref}$) because the magnitude of $[\alpha]_\lambda$ is generally small. Tables collect the mean signed error (MSE), mean unsigned error (MUE), maximum error (Max), and signed-error standard deviation (σ). We also evaluate the relative computational cost of the reduced basis sets relative to the full Dunning ones using the theoretical estimate of the speedup r_c given in eq 3 and wall-clock relative timings for molecules **1**−**5** of the training set. The latter calculations were performed on the same machine with the same number of processors and memory.

It is important to discuss the errors for the reduced basis sets in the context of the expected errors for the full basis sets. The latter are known for aug-cc-pVDZ thanks to Srebro et al.[18] and Stephens et al.[43] Srebro et al. benchmarked optically active compounds with B3LYP/aug-cc-pVDZ and CAM-B3LYP/aug-cc-pVDZ against experimental results using a test set of molecules with $[\alpha]_D$ values ranging from −400 to +500 deg dm^{-1} (g/mL)$^{-1}$.[18] They found average deviations of 25−30 deg dm^{-1} (g/mL)$^{-1}$ with CAM-B3LYP and 20−25 deg dm^{-1} (g/mL)$^{-1}$ for B3LYP, also in agreement with previous findings by Stephens et al.[43]

3.1. augD-3-21G. A statistical analysis of the results for the augD-3-21G basis set with the training set is presented in Table 4. The results obtained with CAM-B3LYP at λ = 589.3 nm, i.e., the optimization set (set in bold font in the table), show a reduction of the error by about half with respect to those obtained with the original exponents for all statistical metrics. For instance, MUE and σ are reduced from 6.8% to 3.7% and from 8.7% to 4.4%, respectively. More interestingly, the control calculations at the other wavelengths and with B3LYP follow the same trends, with MUE and σ also decreasing by factors of about 2, from 7.8% and 10.1% to 4.5% and 5.4% at 450 nm, from 6.7% and 8.6% to 3.7% and 4.3% at 633 nm, and from 6.7% and 8.7% to 3.8% and 4.5% when using B3LYP at 589.3 nm, respectively. All of the calculations on the training set with the reduced basis sets, both before and after optimization, produce $[\alpha]_\lambda$ with the same sign as with aug-cc-pVDZ.

Table 4. MSE, MUE, Max, and σ of the Relative Errors (in %) for the Training Set with the augD-3-21g Basis Set[a]

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
		Optimized		
MSE	−1.9	**−1.0**	−0.8	−1.9
MUE	4.5	**3.7**	3.7	3.8
Max	11.1	**7.3**	7.2	8.7
σ	5.4	**4.4**	4.3	4.5
		Original		
MSE	−1.3	−0.7	−0.6	−1.2
MUE	7.8	6.8	6.7	6.7
Max	23.6	18.0	17.4	19.4
σ	10.1	8.7	8.6	8.7

[a]The data set in bold font refer to the calculations used in the exponent optimization.

The statistical data for the absolute errors using the control set are reported in Table 5. The results with CAM-B3LYP at 589.3 nm do not present a large change in the error after the optimization, as both the MUE and σ change by less than 1 deg dm^{-1} (g/mL)$^{-1}$. This is the case because the results obtained with the original exponents are already extremely good for this series of molecules. The calculations at the other two wavelengths and with B3LYP provide similar trends, with the results obtained with the optimized exponents being only slightly worse (generally less than 1 deg dm^{-1} (g/mL)$^{-1}$) relative to those with the original exponents. Only one calculation using the control set across all sets of calculations produced an incorrect sign of $[\alpha]_\lambda$ relative to the aug-cc-pVDZ results. This is for molecule **20** at 450 nm, which has $[\alpha]_{450}$ =

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

Table 5. MSE, MUE, Max, and σ of the Absolute Errors (in deg dm^{-1} (g/mL)$^{-1}$) for the Control Set with the augD-3-21g Basis Set

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
		Optimized		
MSE	1.1	1.2	0.3	5.5
MUE	10.5	4.8	4.5	10.8
Max	28.1	17.2	13.2	58.9
σ	13.6	6.6	6.2	18.4
		Original		
MSE	1.7	0.4	1.0	6.1
MUE	9.7	5.5	3.7	9.3
Max	32.4	15.4	15.1	53.4
σ	12.7	7.2	5.7	15.1

−1.9 deg dm^{-1} (g/mL)$^{-1}$ with the full aug-cc-pVDZ basis set. The augD-3-21G results are $[\alpha]_{450}$ = 2.1 deg dm^{-1} (g/mL)$^{-1}$ with the original exponents and 11 deg dm^{-1} (g/mL)$^{-1}$ with the optimized exponents. Both values are well within the average error expected for this level of theory (see section 3). At 450 nm, the Max error with the optimized basis set in Table 5 is due to molecule **22**, which has $[\alpha]_{450}$ = 98.1 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVDZ, while with the reduced basis set $[\alpha]_{450}$ = 87.1 before the optimization and 70.0 deg dm^{-1} (g/mL)$^{-1}$ after optimization. The Max value with the original exponents is due to molecule **3**, which at this wavelength has $[\alpha]_{450}$ = −153.6 deg dm^{-1} (g/mL)$^{-1}$ with the full basis set and errors of 32.4 deg dm^{-1} (g/mL)$^{-1}$ with the original exponents and 22.1 deg dm^{-1} (g/mL)$^{-1}$ with the optimized exponents of the reduced basis set. The two large Max errors for B3LYP in Table 4 are also due to two different molecules, molecule **30** for the original exponents ($[\alpha]_D$ = −274.6, −328.0, and −329.6 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVDZ and augD-3-21g with the original and optimized exponents, respectively), and molecule **29** for the optimized exponents ($[\alpha]_D$ = 235.9 with aug-cc-pVDZ and errors of 53.4 and 55.0 deg dm^{-1} (g/mL)$^{-1}$ for augD-3-21g with the original and the optimized exponents, respectively). It should be noted that although the Max errors are due to the molecules with larger values of the specific rotation, their relative errors are similar to those reported in Table 4 for molecules in the training set with similar $[\alpha]_\lambda$ magnitude.

In Table S26 of the SI, we present the numbers of basis functions for each molecule with both aug-cc-pVDZ and augD-3-21G. Using eq 3, we estimate that a calculation with augD-3-21G should be on average 2 times faster than an aug-cc-pVDZ calculation. Calculations on the five largest molecules in our set are 3.5−5 times faster with the smaller basis set (r_c = 4.25 on average).

3.2. augT-3-21G. In this section, we discuss the performance of augT3-3-21G compared with the full aug-cc-pVTZ basis set. We also present the results obtained using the reduced basis set with the original exponents, a variant of the reduced basis set that includes all of the diffuse functions with original exponents (called augT4-3-21G), and the full aug-cc-pVDZ basis set. The last of these is an important test because augT3-3-21G is significantly more compact than aug-cc-pVDZ (in fact, it has the same number of functions per element as augD-3-21G) and can provide considerable computational savings.

The statistical results for the relative error with the training set are reported in Table 6, where again the values set in bold font are from the actual exponent optimization. Not surprisingly, the augT3-3-21G results with CAM-B3LYP at 589.3 nm are considerably better with the optimized exponents than those with the original exponents across the board. More interestingly, the optimized augT3-3-21G basis set outperforms augT4-3-21G as well as the full aug-cc-pVDZ on virtually all metrics. This excellent performance is also replicated at the other two wavelengths with CAM-B3LYP and at 589.3 nm with B3LYP. Remarkably, the calculations with the full aug-cc-pVDZ basis set are only marginally better than those with augT3-3-21G for the MUE and significantly worse for all of the other metrics. All of the calculations with the small basis sets produce $[\alpha]_D$ values of the same sign as those with the aug-cc-pVTZ basis set.

Table 6. MSE, MUE, Max, and σ of the Relative Errors (in %) for the Training Set with the augT3-3-21g Basis Set[a]

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
		Optimized augT3-3-21G		
MSE	−2.2	**−1.6**	−1.4	−1.4
MUE	4.9	**4.4**	4.3	4.2
Max	12.7	**10.1**	8.8	8.3
σ	6.0	**5.3**	5.0	4.9
		Original augT4-3-21G		
MSE	1.0	1.1	1.3	1.3
MUE	5.7	5.4	5.1	5.0
Max	18.6	17.8	15.0	14.5
σ	7.1	6.6	5.9	5.8
		Original augT3-3-21G		
MSE	0.7	1.9	1.7	1.8
MUE	14.0	12.2	12.3	12.1
Max	52.0	48.8	46.2	45.4
σ	19.6	17.7	17.3	17.0
		aug-cc-pVDZ		
MSE	3.3	3.1	3.1	3.0
MUE	4.4	4.2	4.0	4.0
Max	22.4	22.0	20.9	20.7
σ	7.3	6.9	6.6	6.6

[a]The data set in bold font refer to the calculations used in the exponent optimization.

Table 7 shows the statistics for the absolute error of the control set. The optimization clearly improves the performance of the augT3-3-21G basis set compared with the original exponents, with a decrease in the MUE by more than 30%. A performance improvement is also recorded compared with the augT4-3-21G basis set with original exponents across all metrics. However, the most interesting comparison is again with aug-cc-pVDZ. The results in Table 7 show that augT3-3-21G is definitely competitive with the considerably larger aug-cc-pVDZ basis set, with error values that are a few deg dm^{-1} (g/mL)$^{-1}$ larger or smaller than those obtained using aug-cc-pVDZ. The augT3-3-21G basis set provides the wrong sign of $[\alpha]_{450}$ compared with aug-cc-pVTZ for two molecules, **20** and **25**, because the reference results are very close to 0: molecule **20** has $[\alpha]_{450}$ = −3.2, 3.6, −0.7, 15.2, and −1.9 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVTZ, augT4-3-21G, unoptimized augT3-3-21G, optimized augT3-3-21G, and aug-cc-pVDZ, respectively, while molecule **25** has $[\alpha]_{450}$ = −2.2, 38.5, 63.2, 2.4, and

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

Table 7. MSE, MUE, Max, and σ of the Absolute Errors (in deg dm^{-1} (g/mL)$^{-1}$) for the Control Set with the augT3-3-21g Basis Set

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
	Optimized augT3-3-21G			
MSE	1.4	1.0	0.9	1.0
MUE	10.4	5.9	5.0	6.4
Max	27.5	15.1	13.0	17.0
σ	13.8	7.6	6.5	8.3
	Original augT4-3-21G			
MSE	6.4	3.1	2.6	3.3
MUE	12.4	6.4	5.4	6.4
Max	51.2	27.0	23.0	29.0
σ	17.5	8.9	7.6	9.2
	Original augT3-3-21G			
MSE	7.6	3.7	3.1	3.6
MUE	18.4	9.2	7.8	9.6
Max	65.7	35.9	30.8	37.9
σ	26.6	13.6	11.5	14.3
	aug-cc-pVDZ			
MSE	3.2	1.7	1.4	1.4
MUE	5.9	3.1	2.6	8.2
Max	47.5	25.3	21.5	50.3
σ	10.6	5.7	4.8	15.8

−0.6 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVTZ, augT4-3-21G, unoptimized augT3-3-21G, optimized augT3-3-21G, and aug-cc-pVDZ, respectively. At 589.3 nm with B3LYP, the calculation for molecule **23** has the wrong sign compared with the reference for all of the basis sets, including aug-cc-pVDZ. Again, this is due to the small magnitude of $[\alpha]_D$: −3.7, 25.2, 34.1, 10.0, and 11.8 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVTZ, augT4-3-21G, unoptimized augT3-3-21G, optimized augT3-3-21G, and aug-cc-pVDZ, respectively. All of the calculations with CAM-B3LYP using the optimized augT3-3-21G basis set produce the same sign of $[\alpha]_\lambda$ as those using the aug-cc-pVTZ basis set at 589.3 and 633 nm. The largest Max error with augT3-3-21G with optimized exponents is 27.5 deg dm^{-1} (g/mL)$^{-1}$ at 450 nm (see Table 7), which is due to molecule **28**; this molecule has a large $[\alpha]_{450}$ of 403.7 deg dm^{-1} (g/mL)$^{-1}$, and the errors with the other basis sets are 2.9 deg dm^{-1} (g/mL)$^{-1}$ with augT4-3-21G, 17.1 deg dm^{-1} (g/mL)$^{-1}$ with augT3-3-21G (original exponents), and −1.2 deg dm^{-1} (g/mL)$^{-1}$ with aug-cc-pVDZ. However, it should be noted that the Max errors are consistently larger with the other basis sets, including aug-cc-pVDZ.

Table S27 of the SI shows the total numbers of basis functions for each molecule with the aug-cc-pVTZ and augT3-3-21G basis sets. From these values, we estimate an average r_c = 22 with augT3-3-21G compared with aug-cc-pVTZ. This agrees with the actual times of our calculations for the five largest molecules, with r_c = 18−26.

3.3. Comparison with the ORP Basis Set. In this section, we present a direct comparison between augT3-3-21G, ORP, and aug-cc-pVDZ using aug-cc-pVTZ as the reference. The statistical analyses for the relative error on the training set and the absolute error on the control set are reported in Tables 8 and 9, respectively. However, since ORP is not defined for Br and has provided mixed performance for S and Cl,[30] these results do not include molecules **4**, **11**, and **12** for the training set and molecules **5**, **6**, **17**, **24**, **25**, and **26** for the control set.

Table 8. MSE, MUE, Max, and σ of the Relative Errors (in %) for the Training Set with the ORP Basis Set

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
	ORP			
MSE	−0.4	−0.3	−0.3	−0.3
MUE	0.5	0.4	0.4	0.5
Max	2.4	1.8	1.7	2.1
σ	0.7	0.6	0.6	0.8
	augT3-3-21G			
MSE	−1.6	−0.7	−0.6	−0.8
MUE	4.1	3.4	3.4	3.5
Max	12.7	8.8	8.3	10.1
σ	6.1	5.0	4.8	5.2
	aug-cc-pVDZ			
MSE	2.8	2.6	3.0	2.6
MUE	3.3	3.0	4.0	3.0
Max	22.4	20.9	20.7	21.1
σ	6.3	5.7	6.6	5.6

Table 9. MSE, MUE, Max, and σ of the Absolute Errors (in deg dm^{-1} (g/mL)$^{-1}$) for the Control Set with the ORP Basis Set

	CAM-B3LYP			B3LYP
	450 nm	589.3 nm	633 nm	589.3 nm
	ORP			
MSE	0.2	0.0	0.1	0.2
MUE	1.8	1.0	0.8	1.1
Max	9.3	4.9	4.2	4.6
σ	3.5	1.8	1.5	2.0
	augT3-3-21G			
MSE	7.4	3.2	2.7	3.5
MUE	8.7	4.9	4.1	4.9
Max	27.5	15.1	13.0	17.0
σ	12.5	7.2	6.3	7.4
	aug-cc-pVDZ			
MSE	4.1	2.0	1.8	−4.2
MUE	3.8	1.9	1.6	6.6
Max	26.4	14.0	11.9	50.3
σ	5.9	3.2	2.8	16.1

This is not a problem for augT3-3-21G because we can use the original exponents from the parent basis sets. The values of $[\alpha]_\lambda$ computed with the ORP basis set for the training and control sets are reported in Tables S28 and S29 of the SI, respectively.

The results in the tables show that ORP outperforms augT3-3-21G and aug-cc-pVDZ for both sets of molecules. This is not surprising since ORP contains almost as many functions as aug-cc-pVTZ, while augT3-3-21G is less than half the size of the target basis set (see Table 3). Nonetheless, the augT3-3-21G results on the training set are only about 4% different from the target $[\alpha]_\lambda$ on average (MUE in Table 8), with σ = 5−6%; for the control set, the MUE is 5−9 deg dm^{-1} (g/mL)$^{-1}$ with augT3-3-21G against 1−2 deg dm^{-1} (g/mL)$^{-1}$ with ORP, while σ = 6−12 deg dm^{-1} (g/mL)$^{-1}$ against 1.5−3.5 deg dm^{-1} (g/mL)$^{-1}$, respectively. Considering the size of our compact basis set, we consider this performance rather satisfactory.

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415

4. DISCUSSION AND CONCLUSIONS

In this work, we propose two compact basis sets obtained from combining the standard 3-21G basis set with diffuse functions obtained from the aug-cc-pVDZ basis set (augD-3-21G) or the aug-cc-pVTZ basis set (augT3-3-21G). For the latter, we exclude the diffuse functions of highest angular momentum, so that augD-3-21G and augT3-3-21G include the same number of functions per element. The exponents of the diffuse functions for four elements (H, C, O, and N) were optimized to minimize the RMSE difference from the $[\alpha]_D$ values computed with CAM-B3LYP and the corresponding full Dunning basis sets using a training set of 21 organic molecules. The new basis sets were then tested on the same training set but at different wavelengths (450 and 633 nm and with B3LYP at 589.3 nm) and on a control set of different molecules with all of the previous functionals and wavelengths. The results are compared with those obtained with the reduced basis set but using the original exponents and, in the case of augT3-3-21G, also against the full aug-cc-pVDZ and ORP basis sets.

Both the augT3-3-21G and augD-3-21G basis set optimizations change the exponents in mostly similar ways. The main differences are for the carbon p functions (more diffuse after the augD-3-21G optimization and less diffuse after the augT3-3-21G optimization) and the oxygen s functions (less diffuse after the augD-3-21G optimization and more diffuse after the augT3-3-21G optimization). All of the s functions become more diffuse except that for carbon (and also slightly for oxygen with augD-3-21G), while the p functions tend to become less diffuse, except those for hydrogen and augD-3-21G oxygen and carbon. All of the d functions tend to become less diffuse in both basis sets, except those for oxygen.

The tests using the training set show that the augD-3-21G basis set with optimized exponents performs better than that with the original exponents, as the MUE and σ are reduced by nearly half at all frequencies and with all functionals. The tests with the control set do not show a particular improvement with the optimization of the exponents because the results with the original values are already rather close to those with aug-cc-pVDZ (MUE and σ for the optimized and original exponents are almost all within 1 deg dm^{-1} (g/mL)$^{-1}$ of each other). The difference between the augD-3-21G and aug-cc-pVDZ results are also on average smaller than the accuracy of the full basis set relative to experiment with both functionals. In fact, the average expected error relative to the experimental values from B3LYP is 20−25 deg dm^{-1} (g/mL)$^{-1}$, and the average expected error from CAM-B3LYP is 25−30 deg dm^{-1} (g/mL)$^{-1}$,[18,43] whereas the average errors with augD-3-21G with respect to aug-cc-pVDZ are 8.6 and 4.9 deg dm^{-1} (g/mL)$^{-1}$, respectively. From the relative sizes of the basis sets, and using the $O(N^3)$ scaling of DFT, calculations with the augD-3-21G basis set should be at least 2 times faster than those with the aug-cc-pVDZ basis set (actual timings with the five largest molecules in the set showed an average speedup of 4 times).

The optimization of the exponents for the augT3-3-21G basis set leads to a large improvement compared with the results with the original exponents for both the training and control sets. The optimized augT3-3-21G basis set performs better than the augT4-3-21G basis set, which includes all of the diffuse functions from aug-cc-pVTZ with the original exponents. More importantly, augT3-3-21G provides results that are comparable to (control set, Table 7) or better than (training set, Table 6) those with the full aug-cc-pVDZ basis set compared with aug-cc-pVTZ at a fraction of the computational cost. The comparison with ORP indicates that the latter is considerably more accurate than both augT3-3-21G and aug-cc-pVDZ. However, ORP is undefined for several elements and is almost as large as aug-cc-pVTZ, thus providing limited computational savings. In contrast, calculations with the augT3-3-21G basis set are estimated to be nearly 22 times faster on average than those with the full aug-cc-pVTZ basis set (which is confirmed by the actual relative timings with the five largest molecules in the set).

In summary, both the augD-3-21G and augT3-3-21G basis sets perform well compared with the full Dunning basis sets, and they allow calculations of specific rotation of the same quality as with larger basis sets at a fraction of the computational work. This is particularly true for augT3-3-21G, which provides results as close to aug-cc-pVTZ as the full aug-cc-pVDZ basis set. These results are quite promising, and we will now test these basis sets with higher-level methods (e.g., CC) and on other chiroptical properties.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00446.

Optimized exponents of the augD-3-21G and augT3-3-21G basis sets in GAUSSIAN input file format, optimized geometries for compounds **15**−**19** of the training set and compounds **27**−**30** of the control set, $[\alpha]_\lambda$ values computed with all methods and basis sets, and numbers of basis functions per molecule with the augD-3-21G, aug-cc-pVDZ, augT3-3-21G, and aug-cc-pVTZ basis sets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marco Caricato − *Department of Chemistry, University of Kansas, Lawrence, Kansas 66045, United States;* orcid.org/0000-0001-7830-0562; Email: mcaricato@ku.edu

Author

Tal Aharon − *Department of Chemistry, University of Kansas, Lawrence, Kansas 66045, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.0c00446

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors are grateful for the support from the National Science Foundation under Grant CHE-1650942 and the Extreme Science and Engineering Discovery Environment (XSEDE) under Project CHE-170088.

■ REFERENCES

(1) Vaccaro, P. H. Optical Rotation and Intrinsic Optical Activity. In *Comprehensive Chiroptical Spectroscopy, Instrumentation, Methodologies, and Theoretical Simulations, Volume 1*; Berova, N., Polavarapu, P. L., Nakanishi, K., Woody, R. W., Eds.; John Wiley & Sons, 2012; Chapter 11, pp 275−323.

(2) Polavarapu, P. L. Ab initio molecular optical rotations and absolute configurations. *Mol. Phys.* **1997**, *91*, 551−554.

(3) Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Stephens, P. J. Hartree-Fock and density functional theory ab initio calculation of optical rotation using GIAOs: basis set dependence. *J. Phys. Chem. A* **2000**, *104*, 1039−1046.
(4) Grimme, S. Calculation of frequency dependent optical rotation using density functional response theory. *Chem. Phys. Lett.* **2001**, *339*, 380−388.
(5) Polavarapu, P. L. Optical rotation: recent advances in determining the absolute configuration. *Chirality* **2002**, *14*, 768−781.
(6) Autschbach, J.; Patchkovskii, S.; Ziegler, T.; van Gisbergen, S. J. a.; Jan Baerends, E. Chiroptical properties from time-dependent density functional theory. II. Optical rotations of small to medium sized organic molecules. *J. Chem. Phys.* **2002**, *117*, 581−592.
(7) Ruud, K.; Helgaker, T. Optical rotation studied by density-functional and coupled-cluster methods. *Chem. Phys. Lett.* **2002**, *352*, 533−539.
(8) Stephens, P. J.; McCann, D. M.; Cheeseman, J. R.; Frisch, M. J. Determination of absolute configurations of chiral molecules using ab initio time-dependent density functional theory calculations of optical rotation: How reliable are absolute configurations obtained for molecules with small rotations? *Chirality* **2005**, *17*, S52−S64.
(9) Grimme, S.; Bahlmann, A.; Haufe, G. Ab initio calculations for the optical rotations of conformationally flexible molecules: A case study on six-, seven-, and eight-membered fluorinated cycloalkanol esters. *Chirality* **2002**, *14*, 793−797.
(10) Ruud, K.; Stephens, P. J.; Devlin, F. J.; Taylor, P. R.; Cheeseman, J. R.; Frisch, M. J. Coupled-cluster calculations of optical rotation. *Chem. Phys. Lett.* **2003**, *373*, 606−614.
(11) Tam, M. C.; Russ, N. J.; Crawford, T. D. Coupled cluster calculations of optical rotatory dispersion of (*S*)-methyloxirane. *J. Chem. Phys.* **2004**, *121*, 3550−3557.
(12) Crawford, T. D. Ab initio calculation of molecular chiroptical properties. *Theor. Chem. Acc.* **2006**, *115*, 227−245.
(13) Crawford, T. D.; Stephens, P. J. Comparison of time-dependent density-functional theory and coupled cluster theory for the calculation of the optical rotations of chiral molecules. *J. Phys. Chem. A* **2008**, *112*, 1339−1345.
(14) Pedersen, T. B.; Koch, H.; Boman, L.; Sánchez De Merás, A. M. J. Origin invariant calculation of optical rotation without recourse to London orbitals. *Chem. Phys. Lett.* **2004**, *393*, 319−326.
(15) Crawford, T. D.; Owens, L. S.; Tam, M. C.; Schreiner, P. R.; Koch, H. Ab initio calculation of optical rotation in (*P*)-(+)-[4]-triangulane. *J. Am. Chem. Soc.* **2005**, *127*, 1368−1369.
(16) Krykunov, M.; Autschbach, J. Calculation of optical rotation with time-periodic magnetic-field-dependent basis functions in approximate time-dependent density-functional theory. *J. Chem. Phys.* **2005**, *123*, 114103.
(17) Autschbach, J. Computing chiroptical properties with first-principles theoretical methods: background and illustrative examples. *Chirality* **2009**, *21*, E116−E152.
(18) Srebro, M.; Govind, N.; de Jong, W. A.; Autschbach, J. Optical Rotation Calculated with Time-Dependent Density Functional Theory: The OR45 Benchmark. *J. Phys. Chem. A* **2011**, *115*, 10930−10949.
(19) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151−154.
(20) Saebø, S.; Pulay, P. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(21) Russ, N. J.; Crawford, T. D.; Tschumper, G. S. Real versus artifactual symmetry-breaking effects in Hartree-Fock, density-functional, and coupled-cluster methods. *J. Chem. Phys.* **2004**, *120*, 7298−7306.
(22) Russ, N. J.; Crawford, T. D. Local correlation domains for coupled cluster theory: optical rotation and magnetic-field perturbations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3345−3352.
(23) McAlexander, H. R.; Mach, T. J.; Crawford, T. D. Localized optimized orbitals, coupled cluster theory, and chiroptical response properties. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7830−7836.
(24) Aharon, T.; Caricato, M. A molecular orbital selection approach for fast calculations of specific rotation with density functional theory. *Chirality* **2020**, *32*, 243−253.
(25) Caricato, M. Orbital analysis of molecular optical activity based on configuration rotatory strength. *J. Chem. Theory Comput.* **2015**, *11*, 1349−1353.
(26) Caricato, M. Conformational effects on specific rotation: A theoretical study based on the $\tilde{S}_k$ method. *J. Phys. Chem. A* **2015**, *119*, 8303−8310.
(27) Aharon, T.; Caricato, M. Configuration Space Analysis of the Specific Rotation of Helicenes. *J. Phys. Chem. A* **2019**, *123*, 4406−4418.
(28) Baranowska-Łączkowska, A.; Łączkowski, K. Z. The ORP basis set designed for optical rotation calculations. *J. Comput. Chem.* **2013**, *34*, 2006−2013.
(29) Baranowska-Łączkowska, A.; Łączkowski, K. Z.; Henriksen, C.; Fernández, B.; Kozak, M.; Zielińska, S. New basis set for the prediction of the specific rotation in flexible biological molecules. *RSC Adv.* **2016**, *6*, 19897−19902.
(30) Howard, J. C.; Sowndarya S. V., S.; Ansari, I. M.; Mach, T. J.; Baranowska-Łączkowska, A.; Crawford, T. D. Performance of Property-Optimized Basis Sets for Optical Rotation with Coupled Cluster Theory. *J. Phys. Chem. A* **2018**, *122*, 5962−5969.
(31) Wiberg, K. B.; Caricato, M.; Wang, Y.-g.; Vaccaro, P. H. Towards the accurate and efficient calculation of optical rotatory dispersion using augmented minimal basis sets. *Chirality* **2013**, *25*, 606−616.
(32) Crawford, T. D.; Tam, M. C.; Abrams, M. L. The current state of ab initio calculations of optical rotation and electronic circular dichroism spectra. *J. Phys. Chem. A* **2007**, *111*, 12057−68.
(33) Rosenfeld, L. Quantenmechanische Theorie der natürlichen optischen Aktivität von Flüssigkeiten und Gasen. *Eur. Phys. J. A* **1929**, *52*, 161−174.
(34) Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333−3344.
(35) Pedersen, T. B.; Koch, H. Coupled cluster response functions revisited. *J. Chem. Phys.* **1997**, *106*, 8059−8072.
(36) Olsen, J.; Jørgensen, P. Linear and nonlinear response functions for an exact state and for an MCSCF state. *J. Chem. Phys.* **1985**, *82*, 3235−3264.
(37) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian Development Version*, 2009.
(38) Yanai, T.; Tew, D. P.; Handy, N. C. A new hybrid exchange-correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **2004**, *393*, 51−57.
(39) Woon, D. E.; Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties. *J. Chem. Phys.* **1994**, *100*, 2975−2988.
(40) Becke, A. D. A new mixing of Hartree-Fock and local density functional theories. *J. Chem. Phys.* **1993**, *98*, 1372−1377.
(41) Becke, A. D. Density functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(42) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(43) Stephens, P. J.; Devlin, F. J.; Cheeseman, J. R.; Frisch, M. J. Calculation of optical rotation using density functional theory. *J. Phys. Chem. A* **2001**, *105*, 5356−5371.

https://dx.doi.org/10.1021/acs.jctc.0c00446
J. Chem. Theory Comput. 2020, 16, 4408−4415