Towards an Improved Method for the Prediction of Linear Response Properties of Small Organic Molecules

Ruhee Lancelot D'Cunha

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

Daniel Crawford, Chair

Mark Embree

Nicholas Mayhall

Diego Troya

August 6, 2021

Blacksburg, Virginia

Keywords: molecular properties, coupled cluster theory, reduced scaling

Towards an Improved Method for the Prediction of Linear Response Properties of Small Organic Molecules

Ruhee Lancelot D'Cunha

(ABSTRACT)

Quantum chemical methods to predict experimental chiroptical properties by solving the time-dependent Schrödinger equation are useful in the assignment of absolute configurations. Chiroptical properties, being very sensitive to the electronic structure of the system, require highly-accurate methods on the one hand and on the other, need to be able to be computed with limited computational resources.

The calculation of the optical rotation in the solution phase is complicated by solvent effects. In order to capture those solvent effects, we present a study that uses conformational averaging and time-dependent density functional theory calculations that incorporate solvent molecules explicitly in the quantum mechanical region. While considering several controllable parameters along which the system's optical rotation varies, we find that the sampling of the dynamical trajectory and the density functional chosen have the largest impact on the value of the rotation.

In order to eliminate the arbitrariness of the choice of density functional, we would prefer to use coupled cluster theory, a robust and systematically improvable method. However, the high-order polynomial scaling of coupled cluster theory makes it intractable for numerous large calculations, including the conformational averaging required for optical rotation calculations in solution. We therefore attempt to reduce the scaling of a linear response coupled cluster singles and doubles (LR-CCSD) calculation *via* a perturbed pair natural orbital (PNO++) local correlation approach which uses an orbital space created using a perturbed density matrix. We find that by creating a "combined PNO++" space, incorporating a set of orbitals from the unperturbed pair natural orbital (PNO) space into the PNO++ space, we can obtain well-behaved convergence behavior for both CCSD correlation energies and linear response properties, including dynamic polarizabilities and optical rotations, for the small systems considered.

The PNO++ and combined PNO++ methods require aggressive truncation to keep the computational cost low, due to an expensive two-electron integral transformation at the beginning of the calculation. We apply the methods to larger systems than previously studied and refine them for more aggressive truncation by exploring an alternative form of the perturbed density and a perturbation-including weak pair approximation.

This work received support from the National Science Foundation (Grant CHE-1900420) as well as Advanced Research Computing at Virginia Tech for computational resources and technical support.

Towards an Improved Method for the Prediction of Linear Response Properties of Small Organic Molecules

Ruhee Lancelot D'Cunha

(GENERAL AUDIENCE ABSTRACT)

Theoretical chemistry attempts to provide connections between the structure of molecules and their observable properties. One such family of observables are chiroptical properties, or the effect of the medium on the light which passes through it. These properties include the scattering, absorption and change in polarization of light. Light being classically an electromagnetic field, chiroptical properties can be derived by treating molecules quantum mechanically and the light classically. The prediction of chiroptical properties on computers using the principles of quantum mechanics is still a growing field, being very sensitive to the method used, and requiring considerations of factors such as conformations and anharmonic corrections.

Matching experimental properties is an important step in the creation of a reliable method of predicting properties of systems in order to provide more information than can be obtained through experimental observation. This work begins by addressing the problem of matching experimentally obtained quantities. Our results show that current time-intensive methods still fall short in the matching of experimental data. Thus, we then move on to approximating a more robust but computationally expensive method in order to be able to use a more accurate method on a larger scale than is currently possible. On obtaining positive results for small test systems, we test the new method on larger systems, and explore possible improvements to its accuracy and efficiency.

Dedication

For Jax

Acknowledgments

I would like to thank the following people for helping me get to and through grad school:

- My parents and brother for their unconditional love and support.
- Ben for being my partner on this journey and for reading every word I've ever written in a scientific capacity.
- Archita, Zeba, Sravya, and Rewa for being my eternal cheerleaders.
- My chem junta: Hersh, Siddarth, Tahir, Yash, Lakhotia, JD, Abhiram, Santanu, and Harshit for all the lukkha and friendship.
- Vibin for being a 10-year point of reference, a great chemistry teacher, and above all a really good friend.
- Shannon for always being up for a work party and her invaluable help with editing multiple portions of this thesis.
- Crawford group members: Coleman, Roberto, Alex, Monika, AJ, Ashutosh, Kirk, Zhe,

Susana, Brendan, and Jose for always being willing to talk science, and for making the lab a fun workplace.

- Friends in the theory groups: Nakul, Karl, Kalyani, and Marjory, for helping keep me on track this whole time.

- Harley, Taylor, and Ashutosh (again) for putting up with a summer undergrad with a lot of patience and introducing me to scientific programming.

- Prof. Pradeepkumar PI, for introducing me to scientific research and computational chemistry.

- Members of my committee, Dr. Embree, Dr. Mayhall, and Dr. Troya for useful help and feedback throughout my grad school life.

- Last, but not least, Dr. Crawford, who inspired me to do theoretical chemistry as well as a PhD, and helped me become a better scientist.

Attribution

Chapters 3, 4 and 5 have the coauthor T. Daniel Crawford. Calculations, figures and a majority of the text for these chapters was contributed by myself. Dr. Crawford's contribution was serving as a mentor for the projects and editing all three chapters.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Biological homochirality means that human beings are affected deeply by the chirality of things that they consume. This can mean anything from L-sugars not being metabolized in the body,[1] to R-salbutamol being a more effective bronchodilator,[2] to sad cases of birth defects following the racemization of thalidomide in the body.[3] Optical activity is defined as the differential absorption or transmission of left- and right-circularly polarized light, and is displayed by chiral molecules, or molecules that exist as non-superimposable mirror images.[4] They can have one or more chiral centers, but in terms of point group symmetry, must not have an improper axis of rotation, that is, an inversion center, a reflection plane or a rotation-reflection axis. The effects of chirality can be seen very easily by observing the rotation of the plane of plane-polarized light; however, connecting the measured angle of rotation and the three-dimensional absolute configuration of the molecule is not an easy task.

Experimentally, the absolute configuration (or the spatial arrangement of atoms around stereocenters) of molecules that can be crystallized into high-quality crystals can be determined by X-ray crystallography. Practically, it is much easier to measure chiroptical properties such as the optical rotation or circular dichroism spectra. A theoretical connection between the molecular geometry and the measured spectra would assist the determination of absolute configuration of molecules, especially in neat liquid or solution state, which is practically easier to measure the spectra for. Thus, theoretical methods to predict spectra can bridge the gap and help provide experimental chemists an important tool in structure elucidation.

For the energies of weakly-correlated systems, coupled cluster singles and doubles with a perturbed triples correction (CCSD(T)) is considered the gold standard of quantum chemistry;[5] however, response properties, or properties related to the response of a molecule to an external field, such as optical properties, are not as easily predicted to a desired level of accuracy. Being sensitive to small changes in the electronic density, the computation of properties requires more computational effort and careful selection of methods. Unfortunately, there is no reliable black-box method currently for the computation of properties to take on a truly predictive role.

Experiments routinely obtain the optical rotation of organic molecules in solution, and use this information during synthesis in the form of enantiomeric excess. Matching the experimentally-seen solvent effect on the rotation of chiral organic solutes is a long-standing problem that has not yet been solved. Explicitly solvated methods, including the effect of the external field on the electronic density of both the solute and the solvent, seem to be meth-

ods to capture specific interactions missing from continuum solvent methods.[6,7] However, the necessary depth of the solvation shell to which one needs to go has not been explored for a variety of polar and nonpolar solvents. This, as well as other questions surrounding the methodology of using a combined classical molecular dynamics and density functional theory approach has been considered in Chapter 3.

This is the two-fold challenge facing the field of chiroptical property prediction: One, to obtain properties of sufficient accuracy and be able to match experiment so that the automatic backwards interpolation is not only possible but reliable; Two, to be able to do those computations in a reasonable timeframe, on computers accessible to most scientists. In the case of solvated systems, as well as large chiral molecules, the coupled cluster approach, though attractive for its systematic improvability, is too computationally expensive for the system sizes we consider. It is for this reason that several reduced-scaling approaches have been considered. However, it was seen in Ref 8 that local correlation approaches that work well for the correlation energy do not necessarily work well for chiroptical properties. Thus, an alternate approach is needed. One such approach is explored in Chapters 4 and 5.

This work begins by describing and deriving the equations needed in order to compute coupled cluster-level response properties, specifically dynamic polarizabilities and specific rotations, in Chapter 2. Chapter 2 also includes a description of local correlation approaches, focusing on the local pair natural orbital (LPNO) method. A combined molecular dynamics (MD) and time-dependent density functional theory (TD-DFT) study is presented in Chapter 3, in order to explore the extent to which explicit solvent molecules are required in order

to achieve convergence in the rotation value. In Chapter 4, a new orbital space for local correlation, perturbed pair natural orbitals (PNO++), created using perturbation-aware densities, is tested for accuracy for small model systems, and a combined PNO/PNO++ space is created in order to maintain accuracy in both the correlation energy and the linear response properties. Chapter 5 presents data refining the PNO++ method, including some alternate forms of the density, larger molecules whose properties are computed using a more efficient implementation of the simulation code, and the effect of weak pairs on the accuracy of the method. Finally, a summary and some conclusions are presented in Chapter 6.

Chapter 2

Theoretical Background

2.1 Electronic Structure

The ultimate goal of electronic structure theory is to describe the motion of electrons by solving the time-dependent Schrödinger equation,[9]

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(\mathbf{r},t;\mathbf{R})\right\rangle = \hat{H}\left|\Psi(\mathbf{r},t;\mathbf{R})\right\rangle, \tag{2.1}$$

for the wave function $|\Psi(\mathbf{r},t;\mathbf{R})\rangle$ that fully describes the system and can be used to determine properties of the system associated with a given Hermitian operator.[10,11] $|\Psi(\mathbf{r},t;\mathbf{R})\rangle$ depends on the positions of the electrons $\mathbf{r}$ as well as the time t, and, under the Born-Oppenheimer approximation,[12] implicitly depends on the nuclear positions $\mathbf{R}$. Finding approximate solutions to Eq (2.1) under the Born-Oppenheimer approximation is one of the most important problems in the field of quantum chemistry. Many solutions begin via a

separation of variables and the assumption of a time-independent Hamiltonian leading to the time-independent Schrödinger equation, Eq (2.2), in which the time-independent part of the wave function ψ now depends on the positions of the electrons and implicitly on nuclear positions:

$$E\,|\psi(\mathbf{r};\mathbf{R})\rangle = \hat{H}\,|\psi(\mathbf{r};\mathbf{R})\rangle \tag{2.2}$$

where E is the energy of the system. Many-electron wave functions are usually written as a product of single-electron wave functions, also known as orbitals $\chi(\mathbf{x})$, containing both spatial ($\phi(\mathbf{r})$) and spin ($\alpha(\omega)$) components $\chi(\mathbf{x}) = \phi(\mathbf{r})\alpha(\omega)$.[13] Here, ω is the spin coordinate corresponding to α or spin-up and β or spin-down electrons. A popular way of defining a many-electron wave function in terms of orbitals is the Slater determinant:[14]

$$\psi(\mathbf{x}_1, ...\mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \dots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \dots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \dots & \chi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.3}$$

The inherent properties of a determinant make it suitable for this purpose, as the Pauli antisymmetry principle[15] and fermionic antisymmetry is built into the determinantal form. The interchange of any two rows or columns merely changes the sign, thus satisfying the antisymmetry requirement, and any two rows or columns (and thus orbitals) being equal makes the determinant zero. Beginning with such a simple determinantal wave function in terms of single-electron orbitals, we may variationally optimize the orbitals, under the constraint of keeping those orbitals orthonormal, in order to obtain the Hartree-Fock wave

function and energy.[13]

The Hartree-Fock wave function is missing a key piece – dynamical correlation – or the correlation between the movement of two electrons in the same molecular system, above and beyond the correlation between two electrons of parallel spins. And thus, "correlated" methods build upon Hartree-Fock methods in order to obtain the most accurate picture of electronic motion available to us under the constraints of computing resources.

2.2 Coupled Cluster Theory

Coupled cluster (CC) theory is a post-Hartree-Fock method that uses an exponential *Ansatz*, thus obtaining certain useful properties of the truncated wave function, such as size consistency and size extensivity.[16–19] The CC wave function is usually given in terms of the exponential of a cluster operator $\hat{T}$ acting on a reference wave function:

$$|\Psi_{\mathrm{CC}}\rangle = e^{\hat{T}} |0\rangle \tag{2.4}$$

where the reference wave function $|0\rangle$ is usually assumed to be the Hartree-Fock wave function as given in Eq (2.3). The cluster operator is an excitation operator which helps create the multi-determinantal wave function $|\Psi_{\mathrm{CC}}\rangle$ and is of the form:

$$\hat{T} = \sum_{i=1}^{N} \hat{T}_i = \sum_{\mu} t_\mu \hat{\tau}_\mu \tag{2.5}$$

where the sum μ runs over all single, double and so on excitations from an occupied orbital to a virtual orbital. At each excitation level μ, the operator is broken down into an excitation

operator $\hat{\tau}_\mu$ and a scalar amplitude t_μ. The summation runs up to an N-electron excitation, where N is the number of electrons in the system. However, higher excitation-level terms can quickly become restrictive in terms of the computational cost of both memory and storage. Thus in its usual form, the coupled cluster wave function is truncated to a given level of excitation. A truncation to the single and double excitations alone, for example, gives us $\hat{T} = \hat{T}_1 + \hat{T}_2$, the coupled cluster singles and doubles (CCSD) cluster operator, and the CCSD wave function:

$$|\Psi_{\mathrm{CCSD}}\rangle = e^{\hat{T}_1 + \hat{T}_2} |0\rangle . \tag{2.6}$$

Given the form of the untruncated CC wave function in Eq (2.4), we write the time-independent Schrödinger equation from Section 2.1 as:

$$\hat{H} |\Psi_{\mathrm{CC}}\rangle = \hat{H} e^{\hat{T}} |0\rangle = E_{CC} e^{\hat{T}} |0\rangle \tag{2.7}$$

and we obtain the CC energy by projecting on the left with the reference wave function times an inverse exponential $\langle 0| e^{-\hat{T}}$ such that the right hand side reduces to the CC energy:

$$\langle 0| e^{-\hat{T}} H e^{\hat{T}} |0\rangle = E_{\mathrm{CC}}. \tag{2.8}$$

The expression $e^{-\hat{T}} H e^{\hat{T}}$ can be evaluated once the cluster amplitudes are known, and is known as the similarity-transformed Hamiltonian. The eigenvalue spectrum of the similarity-transformed Hamiltonian is identical to the Hamiltonian $\hat{H}$, and thus the same eigenvalues may be found. Importantly, this projection leads to the natural truncation of the CC energy

expression via the Baker-Campbell-Hausdorff expansion:[20]

$$e^{-\hat{T}}He^{\hat{T}} = \hat{H} + \Big[\hat{H},\hat{T}\Big] + \frac{1}{2!}\Big[\Big[\hat{H},\hat{T}\Big],\hat{T}\Big] + \frac{1}{3!}\Big[\Big[\Big[\hat{H},\hat{T}\Big],\hat{T}\Big],\hat{T}\Big] + \frac{1}{4!}\Big[\Big[\Big[\Big[\hat{H},\hat{T}\Big],\hat{T}\Big],\hat{T}\Big],\hat{T}\Big] \tag{2.9}$$

which truncates at the fourth term, since the Hamiltonian is at most a two-electron operator and the cluster operators commute with one another.

The coupled cluster amplitudes can be obtained by projecting on the left with excited determinants, multiplied by the same exponential as above:

$$\langle\mu|e^{-\hat{T}}He^{\hat{T}}|0\rangle = 0. \tag{2.10}$$

Here, μ once again refers to all possible excitation levels, with each set of amplitude equations giving us a set of t_μ amplitudes. These equations may be solved iteratively to obtain cluster amplitudes, which are then plugged into Eq (2.8) to give us the correlation energy. The scaling of a truncated coupled cluster method can be seen as the order of the largest contraction over amplitudes, so CCSD is of order iterative N^6, while CCSDT is of order iterative N^8.

Size consistency is defined as the correctly additive property of the energy of two separately calculated non-interacting subsystem fragments. This is a computed property which is satisfied by truncated coupled cluster theory.[21] Size extensivity, on the other hand, refers to the correct scaling of the energy or other extensive properties of a system with an increase in system size. This property can in theory be proven for the energy obtained from the truncated CC wave function, for example for the CCSD energy, by noting that the coupled cluster amplitude equations do not depend on the coupled cluster energy, and that the

Hamiltonian must be connected to all T amplitudes to the right for the terms in the CC energy and amplitude equations to be non-zero.[22]

CC theory is widely used for accurate and systematically-improvable calculations of ground and excited state energies as well as a wide variety of molecular properties.[23,24] While obtaining the energy from the above equations is fairly straightforward, some extra steps are needed in order to compute properties of interest. Most commonly, excited states and related properties may be computed using the equation-of-motion EOM-CC method,[25,26] and properties relating to the interaction of light and matter are computed using CC response theory,[27,28] described in more detail below.

2.3 Molecular Chiroptical Properties

Chiroptical properties relate to the interaction of light and matter; more specifically, they arise due to either differential absorption, transmission or scattering of circularly polarized light in a medium. To predict chiroptical properties, we examine the oscillation of the molecular multipole moments induced by an external electromagnetic field. This oscillation creates a new electromagnetic field which interferes with the transmitted light and leads to a change in phase or magnitude of the light wave. An in-depth derivation can be found in Ref 4, and we follow this derivation closely.

To begin, we may use Maxwell's equations[29] in an infinite homogeneous medium with no free charges and zero conductivity to obtain the wave equation for light in terms of the electric

field associated with it, **E**:

$$\nabla^2\mathbf{E} = \mu\mu_0\epsilon\epsilon_0\frac{\partial^2\mathbf{E}}{\partial t^2} \tag{2.11}$$

with μ and ϵ being the permeability and permittivity respectively of the medium and μ_0 and ϵ_0 being the same for free space. Assuming plane-polarized light propagating in the z direction, we obtain the following as a solution of the wave equation for a monochromatic light wave with frequency ω, written as a complex expression with amplitude $\tilde{\mathbf{E}}_0$ and photon momentum $\hbar\boldsymbol{\kappa}$,

$$\tilde{\mathbf{E}} = \tilde{\mathbf{E}}_0 e^{i(\boldsymbol{\kappa}.\mathbf{r}-\omega t)}. \tag{2.12}$$

Here the wave vector $\boldsymbol{\kappa}$ is defined as $\frac{\omega}{c}\mathbf{n}$, with c being the speed of light in vacuum and $\mathbf{n}$ being a propagation vector with magnitude $|\mathbf{n}|$ as the refractive index of the medium. Since the refractive indices of a chiral medium are different for left- and right-circularly polarized light, we can break this plane-polarized light down into its left- and right-circularly polarized components as follows:

$$\mathbf{E}_{R/L} = \frac{1}{\sqrt{2}}E_0(\hat{\mathbf{i}} + e^{\mp(i\pi/2)}\hat{\mathbf{j}})e^{i(\boldsymbol{\kappa}.\mathbf{r}-\omega t)} \tag{2.13}$$

with the vectors $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ being the unit vectors in the x and y directions. Having obtained, after the use of the above assumptions, the equation of plane-polarized monochromatic light, we use the form of this equation to derive expressions for chiroptical properties through quantum mechanics.

We may write the time-dependent Schrödinger equation (TDSE) from Eq (2.1) with an interaction Hamiltonian $\hat{V}$:

$$i\hbar\frac{\partial}{\partial t}\psi = (\hat{H}_0 + \hat{V})\psi \tag{2.14}$$

where the Hamiltonian is expressed as the sum of a time-independent part, $\hat{H}_0$, and time-dependent part, which can be expressed in terms of multipole moments, fields E_α, B_α and field derivatives $(E_{\alpha\beta})_0 = \Delta_\alpha E_\beta$ at the coordinate origin (indicated by the subscript 0) as:[30]

$$\begin{aligned} V = \quad & q(\phi)_0 - \mu_\alpha(E_\alpha)_0 - \frac{1}{3}\Theta_{\alpha\beta}(E_{\alpha\beta})_0 \\ & - m_\alpha(B_\alpha)_0 - \frac{1}{2}\chi^{(d)}_{\alpha\beta}(B_\alpha)_0(B_\beta)_0 + \ldots \end{aligned} \tag{2.15}$$

where ϕ is the scalar potential, μ_α is the electric dipole moment, $\Theta_{\alpha\beta}$ is the electric quadrupole moment, m_α is the magnetic dipole moment and $\chi^{(d)}_{\alpha\beta}$ is the diamagnetic susceptibility. Here and afterwards, $\alpha, \beta, \ldots$ indicate Cartesian coordinates and Einstein summation is implied.

This form of the interaction Hamiltonian is not unique, but this choice is useful since molecular properties can be related to molecular property tensors, which serve as a bridge between the induced oscillations of the multipole moments of molecules and the electromagnetic field (i.e. light) that induces those oscillations. The tensors can be obtained as the coefficients of the radiation field components in the expansion of the expectation values of the multipole moments.

We may use perturbation theory to express the perturbed wave function $\psi' = \sum_n \psi'_n$ that solves the form of the TDSE in Eq (2.14) in terms of the unperturbed eigenfunctions of the time-independent part of the Hamiltonian H_0, ψ_n. The eigenfunctions $\psi_n = \psi_n^{(0)} e^{-i\omega_n t}$ have associated eigenvalues of $\hbar\omega_n$. Writing a single perturbed wave function ψ'_n as a linear combination of unperturbed wave functions ψ_j with as-yet undetermined coefficients $\tilde{a}_{jn_\beta}, \tilde{b}_{jn_\beta}$,

of the complex field components $\tilde{E}_\beta, \tilde{B}_\beta$, etc.:

$$\begin{aligned}\psi'_n = \psi_n^{(0)} e^{-i\omega_n t} + \sum_{j\neq n} [\tilde{a}_{jn_\beta}(\tilde{E}_\beta)_0 + \tilde{b}_{jn_\beta}(\tilde{E}^*_\beta)_0 + \tilde{c}_{jn_\beta}(\tilde{B}_\beta)_0 \\ + \tilde{d}_{jn_\beta}(\tilde{B}^*_\beta)_0 + \ldots]\psi_j^{(0)} e^{-i\omega_n t}.\end{aligned} \tag{2.16}$$

Substituting the chosen form of V (Eq (2.15)) and the above-obtained perturbed wave function ψ'_n into Eq (2.14), we obtain:

$$\begin{aligned}-\hbar \sum_{j\neq n} \Big[&(\omega_{jn} - \omega)\tilde{a}_{jn_\beta}(\tilde{E}_\beta)_0 + (\omega_{jn} + \omega)\tilde{b}_{jn_\beta}(\tilde{E_\beta}^*)_0 \\ &+(\omega_{jn} - \omega)\tilde{a}_{jn_\beta}(\tilde{E}_\beta)_0 + (\omega_{jn} + \omega)\tilde{b}_{jn_\beta}(\tilde{E_\beta}^*)_0 + \ldots \Big] \psi_j^{(0)} e^{-i\omega_n t} \\ &= -\frac{1}{2}\left(\mu_\beta[(\tilde{E}_\beta)_0 + (\tilde{E}^*_\beta)_0] + m_\beta[(\tilde{B}_\beta)_0 + (\tilde{B}^*_\beta)_0] + \ldots\right)\psi_n^{(0)} e^{-i\omega_n t}\end{aligned} \tag{2.17}$$

where $\omega_{jn} = \omega_j - \omega_n$. Multiplying both sides by $\psi_j^{(0)*}$ and integrating over all space, we can equate coefficients to obtain the coefficients $\tilde{a}_{jn_\beta}, \tilde{b}_{jn_\beta}$, etc. as:

$$\begin{aligned}\tilde{a}_{jn_\beta} &= \frac{1}{2\hbar}\frac{\langle j|\mu_\beta|n\rangle}{(\omega_{jn} - \omega)}, \\ \tilde{b}_{jn_\beta} &= \frac{1}{2\hbar}\frac{\langle j|\mu_\beta|n\rangle}{(\omega_{jn} + \omega)},\end{aligned} \tag{2.18}$$

with the coefficients containing the magnetic dipole as well as higher-order multipole operators being obtained in a similar way.

Expanding the expectation value of, for example, the dipole moment in terms of electromag-

netic field components, we obtain:

$$\begin{aligned}
\mu_\alpha &= \langle \psi'_n | \mu_\alpha | \psi'_n \rangle \\
&= \langle \psi_n | \mu_\alpha | \psi_n \rangle + \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega_{jn}}{\omega_{jn}^2 - \omega^2} \mathrm{Re}(\langle n|\mu_\alpha|j\rangle \ \langle j|\mu_\beta|n\rangle)(E_\beta)_0 \\
&- \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \mathrm{Im}(\langle n|\mu_\alpha|j\rangle \ \langle j|\mu_\beta|n\rangle)(\dot{E}_\beta)_0 \\
&+ \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega_{jn}}{\omega_{jn}^2 - \omega^2} \mathrm{Re}(\langle n|\mu_\alpha|j\rangle \ \langle j|m_\beta|n\rangle)(B_\beta)_0 \\
&- \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \mathrm{Im}(\langle n|\mu_\alpha|j\rangle \ \langle j|m_\beta|n\rangle)(\dot{B}_\beta)_0 + \dots
\end{aligned} \tag{2.19}$$

or, alternatively, in terms of the dynamic molecular property tensors,

$$\mu_\alpha = \alpha_{\alpha\beta}(E_\beta)_0 + \frac{1}{\omega}\alpha'_{\alpha\beta}(\dot{E}_\beta)_0 + G_{\alpha\beta}(B_\beta)_0 + \frac{1}{\omega}G'(\dot{B}_\beta)_0 + \dots \tag{2.20}$$

The dynamic molecular property tensors, being coefficients of the radiation field components, can then be obtained as sum-over-states expressions. We are specifically concerned with the coefficient of $(E_\beta)_0$, or the component of the dynamic polarizability tensor $\alpha_{\alpha\beta}$ and the coefficient of $(\dot{B}_\beta)_0$, or the component of the Rosenfeld tensor $G'_{\alpha\beta}$, as obtained in Eq (2.21). The dynamic polarizability tensor contributes to light scattering and refraction, while the G' tensor is the major contribution to natural optical rotation and circular dichroism in isotropic media.

$$\begin{aligned}
\alpha_{\alpha\beta} &= \frac{2}{\hbar} \sum_{j \neq n} \frac{\omega_{jn}}{\omega_{jn}^2 - \omega^2} \mathrm{Re}(\langle n|\mu_\alpha|j\rangle \ \langle j|\mu_\beta|n\rangle) \\
G'_{\alpha\beta} &= -\frac{2}{\hbar} \sum_{j \neq n} \frac{\omega}{\omega_{jn}^2 - \omega^2} \mathrm{Im}(\langle n|\mu_\alpha|j\rangle \ \langle j|m_\beta|n\rangle)
\end{aligned} \tag{2.21}$$

2.3.1 Optical Rotation

The natural optical rotation arises from the differential transmission of left and right circularly polarized light. This translates to separate refractive indices n_L and n_R and the optical rotation $\Delta\theta$ can be written in terms of the difference between them, factored by the path length l and the frequency of the monochromatic light ω:

$$\Delta\theta = \frac{\omega l}{2c}(n_L - n_R). \tag{2.22}$$

Using the complex form of a plane wave of monochromatic light, Eq (2.12), and the bulk multipole polarizations in terms of the multipole moments of the individual molecules, we can obtain the Rosenfeld equation which describes the natural optical rotation in radians in an isotropic medium:[31]

$$\Delta\theta \simeq -\frac{1}{3}\omega\mu_0 l N \mathrm{Tr}[G'(\omega)] \tag{2.23}$$

in terms of the mixed electric-dipole magnetic-dipole tensor G' derived as a sum over excited states earlier. The trace over G' and the factor of $\frac{1}{3}$ comes from the isotropic averaging over 3 dimensions, and N is the number density of molecules. A further step to normalize the rotation with respect to path length and concentration gives us the following expression for the specific rotation:

$$[\alpha]_\omega = \frac{V\Delta\theta}{ml} = \frac{1}{3}\frac{\omega\hbar^2 N_A}{c^2 m_e^2 M}\mathrm{Tr}[G'(\omega)] \tag{2.24}$$

where N_A is Avogadro's number, m_e is the mass of an electron in kg and M is the molecular mass in amu. The spectrum obtained by plotting the wavelength-dependence of the optical rotation is an optical rotatory dispersion spectrum, and is shown in Fig 2.1. The Drude

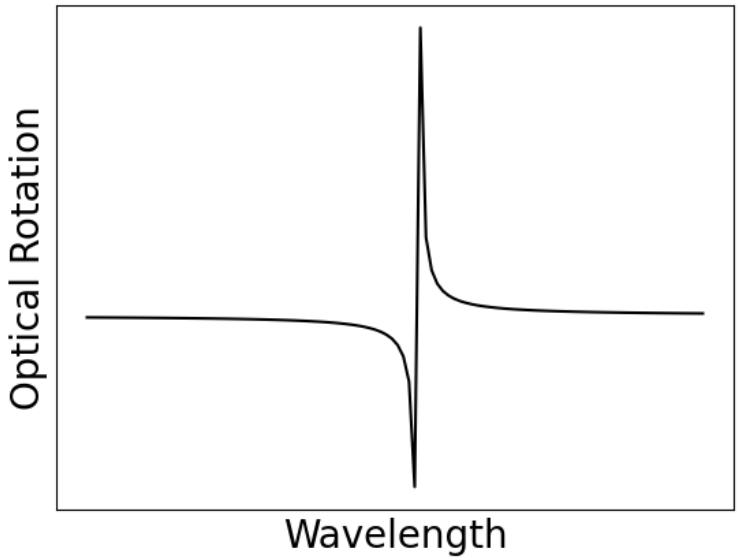


Figure 2.1: An example ORD spectrum, with a point of inflection at the excitation wavelength.

equation:[32]

$$\Delta\theta = \frac{C_k}{\lambda^2 - \lambda_k^2} \tag{2.25}$$

is often used for fitting data at specific wavelengths in order to produce an ORD spectrum.

The sum-over-states expression in Eq (2.21) is impractical because it requires the calculation of matrix elements of the dipole moments for each excited state, and the number of states required to converge the series is in practice very large.[33] Instead, response theory is used to obtain a linear response function, which can be related to the Rosenfeld tensor in the following way:[34]

$$G'_{\alpha\beta}(\omega) = \mathrm{Im}\langle\langle \mu_\alpha; m_\beta \rangle\rangle_\omega \tag{2.26}$$

2.4 Response Theory

Response theory is an application of perturbation theory to calculate the response of a molecule's electronic density to a time-dependent external electromagnetic field.[33,35] This application means that in order for the approximation to hold, the external field must be weak in comparison to the internal atomic fields. This is true for a wide range of spectroscopies, and thus response can be used for the calculation of many different time-dependent properties. Response theory also has the advantage of implicitly including all electronically excited states without needing to obtain each individual state and is thus a more practical alternative to the sum-over-states approach. Vibrational contributions, important in a large number of spectroscopies, can also be calculated more easily in a response framework. Linear response

theory, or the first order contribution in the perturbational expansion, is used to compute time-dependent properties such as dynamic polarizabilities, excitation energies, oscillator strengths, and electronic circular dichroism (ECD) spectra, as well as optical rotations and rotatory strengths.[36–40]

We write the Hamiltonian as a sum of a time-independent piece and the perturbation $\hat{V}^t$, which can be related to a frequency-dependent perturbation $\hat{V}^\omega$ via Fourier transform:

$$\hat{H}(t) = \hat{H}_0 + \hat{V}^t \tag{2.27}$$

$$\hat{V}^t = \int_{-\infty}^{\infty} d\omega \hat{V}^\omega e^{-i\omega t} \tag{2.28}$$

$$\hat{V}^\omega = \sum_\omega \hat{\Omega} F^\omega \tag{2.29}$$

where ω is the frequency of the external electromagnetic field, and the amplitude F^ω of the operator $\hat{\Omega}$ satisfies $(F^\omega)^\dagger = F^{-\omega}$.

Expanding the wave function $|\Psi(t)\rangle$ in orders of the frequency-dependent perturbation gives us a path to expressing the expectation value of an operator of interest, $\hat{A}$, in orders of the perturbation as well.

$$\begin{aligned} |\Psi(t)\rangle =& |\Psi_0\rangle + \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t} |\Psi(\omega)\rangle \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\omega_1\, d\omega_2\, e^{-i(\omega_1+\omega_2)t} |\Psi(\omega_1,\omega_2)\rangle + \ldots \end{aligned} \tag{2.30}$$

The time-dependent expectation value of the operator $\hat{A}$ can then be expressed as:

$$\begin{aligned} \langle\Psi(t)|\hat{A}|\Psi(t)\rangle =& \langle\Psi_0|\hat{A}|\Psi_0\rangle + \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t} \langle\langle \hat{A}; \hat{V}^\omega \rangle\rangle \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\omega_1\, d\omega_2\, e^{-i(\omega_1+\omega_2)t} \langle\langle \hat{A}; \hat{V}^{\omega_1}, \hat{V}^{\omega_2} \rangle\rangle + \ldots \end{aligned} \tag{2.31}$$

where the functions $\langle\langle \hat{A}; \hat{V}^{\omega} \rangle\rangle$ and $\langle\langle \hat{A}; \hat{V}^{\omega_1}, \hat{V}^{\omega_2} \rangle\rangle$ are response functions that represent the effect of the frequency-dependent perturbation on the operator $\hat{A}$.

2.4.1 Quasi-energy formalism

In order to obtain expressions for response functions, we may make use of a formalism that expresses properties as derivatives of a quasi-energy, which is useful since it can be applied via a Lagrangian to nonvariational methods such as coupled cluster theory.[35,41] We will see this application in Section 2.4.2 to derive coupled cluster response functions. The response functions can then be obtained as terms in the perturbational expansion of the quasi-energy.

To begin, we write a general time-dependent wave function $\Psi(t)$ as a product of a phase-isolated part $\bar{\psi}(t)$ and a phase factor:

$$\Psi(t) = e^{-i\phi(t)}\bar{\psi}(t) \tag{2.32}$$

where the phase factor contains $\phi(t)$, a real function. Substituting into Eq (2.1) and projecting on the left with the phase-isolated wave function $\bar{\psi}$, we can obtain an expression for the quasi-energy $Q(t)$:

$$Q(t) = \langle\bar{\psi}|\left(\hat{H} - i\frac{\partial}{\partial t}\right)|\bar{\psi}\rangle\,. \tag{2.33}$$

We can show that for a variation in the phase-isolated wave function $\delta\bar{\psi}$:

$$\delta Q(t) + i\frac{\partial}{\partial t}\langle\bar{\psi}|\delta\bar{\psi}\rangle = 0 \tag{2.34}$$

which forms the basis of a time-dependent variational principle. We choose a variation

$\delta\bar{\psi} = \frac{d\bar{\psi}(t)}{dF^{\omega}}$ and differentiate the quasi-energy in Eq (2.33) to obtain:

$$\frac{dQ(t)}{dF^{\omega}} = \langle\bar{\psi}|\frac{\partial\hat{H}}{\partial F^{\omega}}|\bar{\psi}\rangle + \delta Q(t) \tag{2.35}$$

$$= \langle\bar{\psi}|\frac{\partial\hat{H}}{\partial F^{\omega}}|\bar{\psi}\rangle - i\frac{\partial}{\partial t}\left\langle\bar{\psi}\middle|\frac{d\bar{\psi}}{dF^{\omega}}\right\rangle \tag{2.36}$$

which is the time-dependent Hellmann-Feynman theorem.[42,43]

If we time-average the the time-dependent variational principle from Eq (2.34) and the time-dependent Hellmann-Feynman theorem, by integrating over the period T of the phase-isolated wave function, we can reduce them to the simpler forms:

$$\delta Q_T = 0 \tag{2.37}$$

and

$$\frac{dQ_T}{dF^{\omega}} = \left\{ \langle\bar{\psi}|\frac{\partial}{\partial F^{\omega}}H|\bar{\psi}\rangle \right\}_T \tag{2.38}$$

with the time-averaged quasi-energy Q_T defined as:

$$Q_T = \frac{1}{T}\int_t^{t+T} Q(t')\,dt'. \tag{2.39}$$

The time-averaged quasi-energy is still time-dependent, but we can view this quasi-energy Q_T as a function of the amplitude of the external field F^{ω}. Using the time-dependent Hellmann-Feynman theorem, we can express response functions for general frequency-dependent operators V^{ω} in terms of derivatives of the quasi-energy, for example, the linear and quadratic response functions as below.

$$\langle\langle V^{\omega_1}; V^{\omega_2}\rangle\rangle = 2\frac{d^2Q_T}{dF^{\omega_1}\,dF^{\omega_2}}\bigg|_{F^{\omega}=0} \tag{2.40}$$

$$\langle\langle V^{\omega_1}; V^{\omega_2}, V^{\omega_3}\rangle\rangle = 3!\frac{d^3Q_T}{dF^{\omega_1}\,dF^{\omega_2}\,dF^{\omega_3}}\bigg|_{F^{\omega}=0} \tag{2.41}$$

For a general reference state, equating the derivative of the quasi-energy to the perturbational expansion of the quasi-energy in orders of the external field, we can obtain for instance, the linear response function for our operator $\hat{A}$:[41]

$$\langle\langle \hat{A}; \hat{V}^{\omega} \rangle\rangle = \hat{P}_{A,V^{\omega}} \sum_{n} \frac{\langle \Psi_0 | \hat{A} | \Psi_n \rangle \ \langle \Psi_n | \hat{V}^{\omega} | \Psi_0 \rangle}{\omega_{n0} - \omega} \tag{2.42}$$

where $\hat{P}_{A,V^{\omega}}$ is a permutation operator that exchanges the operators $\hat{A}$ and $\hat{V}^{\omega}$.

2.4.2 Coupled Cluster Linear Response

A time-dependent parametrization of coupled cluster theory expresses the time-dependent left and right CC wave functions as:

$$|\Psi_{CC}\rangle = e^{i\phi(t)} e^{\hat{T}(t)} |0\rangle \tag{2.43}$$

$$\langle \Lambda_{CC}| = \langle 0| \left(1 + \hat{\Lambda}(t)\right) e^{-\hat{T}(t)} e^{-i\phi(t)} \tag{2.44}$$

where the time-dependent cluster operators $\hat{T}(t)$ and $\hat{\Lambda}(t)$ are expressed in terms of time-dependent amplitudes and excitation/de-excitation operators:

$$\hat{T}(t) = \sum_{\mu} t_{\mu}(t) \hat{\tau}_{\mu} \tag{2.45}$$

$$\hat{\Lambda}(t) = \sum_{\mu} \lambda_{\mu}(t) \hat{\tau}_{\mu}^{\dagger} \tag{2.46}$$

We can obtain the quasi-energy, as defined in Eq (2.33) in this parametrization as:

$$Q_{CC}(t) = \langle 0| \left(\hat{H} - i \frac{\partial}{\partial t} \right) e^{\hat{T}(t)} |0\rangle = \langle 0| \hat{H} e^{\hat{T}(t)} |0\rangle \tag{2.47}$$

Projecting on the left with the excited determinants that would usually give us the CC amplitude equations, we obtain a set of constraints:

$$\langle\mu|e^{-\hat{T}}\left(\hat{H}-i\frac{\partial}{\partial t}\right)e^{\hat{T}(t)}|0\rangle=0 \tag{2.48}$$

We can incorporate these constraints by forming a quasi-energy Lagrangian as:

$$\mathcal{L}=Q(t)+\sum_{\mu}\lambda_{\mu}(t)\,\langle\mu|e^{-\hat{T}}\left(\hat{H}-i\frac{\partial}{\partial t}\right)e^{\hat{T}(t)}|0\rangle \tag{2.49}$$

and we can obtain response functions from the derivatives of the time-averaged Lagrangian $\mathcal{L}_T$. We then expand the cluster operators in orders of the perturbation:

$$\hat{T}=\hat{T}^{(0)}+\hat{T}^{(1)}+\hat{T}^{(2)}+\ldots \tag{2.50}$$

$$\hat{\Lambda}=\hat{\Lambda}^{(0)}+\hat{\Lambda}^{(1)}+\hat{\Lambda}^{(2)}+\ldots \tag{2.51}$$

and expanding the Lagrangian also in orders of the perturbation, we may obtain the Lagrangian to nth order in the perturbation, which is proportional to the nth order derivative of the quasi-energy Lagrangian, and thus the corresponding response functions. A clear derivation of these individual pieces can be found in Ref 35. The linear response function is then derived in terms of the zeroth and first order amplitudes and the similarity-transformed perturbation $\bar{A}^{(0)}=e^{-\hat{T}^{(0)}}\hat{A}e^{\hat{T}^{(0)}}$ as:

$$\langle\langle\hat{A};\hat{V}^{\omega}\rangle\rangle=\frac{1}{2}\hat{P}_{A,V^{\omega}}\left\langle 0\left|\left(1+\hat{\Lambda}^{(0)}\right)\left[\bar{A}^{(0)},\hat{T}^{(1)}\right]+\hat{\Lambda}^{(1)}\bar{A}^{(0)}\right|0\right\rangle \tag{2.52}$$

which can also be rewritten assuming the $\hat{T}$ and $\hat{\Lambda}$ operators are zeroth order and defining $\hat{X}=\hat{T}^{(1)}$ and $\hat{Y}=\hat{\Lambda}^{(1)}$ as:

$$\langle\langle\hat{A};\hat{V}^{\omega}\rangle\rangle=\frac{1}{2}\hat{P}_{A,V^{\omega}}\left\langle 0\left|\left(1+\hat{\Lambda}\right)\left[\bar{A},\hat{X}^{V^{\omega}}\right]+\hat{Y}^{V^{\omega}}\bar{A}\right|0\right\rangle. \tag{2.53}$$

The factor of $\frac{1}{2}$ and the permutation operator in the expression for the CC linear response function comes from the requirement that the exact time-dependent expectation value of $\hat{A}$ must be real, $\hat{A}$ being a Hermitian operator, while the expectation value of $\hat{A}$ with the CC wave function is not necessarily real.[44] Equation (2.53) is used to obtain the linear response function and thus the dynamic polarizability and optical rotation tensors using coupled cluster linear response theory.

2.5 Local Correlation

Local correlation techniques are a systematic approach to reducing the scaling of high-accuracy methods such as coupled cluster theory. Localization for electron correlation was first suggested by Sinanoglu[16] and Nesbet[45], and the basic principle remains unchanged. The first step in the reduction of the scaling involves the direct localization of the occupied orbital space, in order to introduce the largest amount of sparsity in the wave function. This is followed by a transformation of the virtual orbital space with the same purpose, and a truncation criterion is then used to maintain accuracy in the energy or properties sought, while speeding up the electronic structure computation.

Projected atomic orbitals (PAOs),[46,47] orbital-specific virtuals (OSVs),[48] as well as fragmentation schemes such as the cluster-in-molecule approach (CIM),[49,50] the divide-and-conquer method[51] and the divide-expand-consolidate (DEC) method[52] have been explored for the approximation of ground state energies. The pair natural orbital (PNO) approach,[53] based

on virtual spaces defined using pairs of localized occupied orbitals, has had great success in reducing the scaling of coupled cluster energy calculations, with an extension of the approach, the domain-based local pair natural orbital with coupled cluster singles and doubles, including perturbative triples (DLPNO-CCSD(T)) method,[54] having been reduced asymptotically to linear scaling in practice on large molecular systems.

While local correlation methods have been applied to excitation energies and some excited state properties,[55–62] response properties, being more sensitive to changes in the electron density, have not yet been approximated in the same successful manner as ground state energies.

2.5.1 Localization of Occupied Orbitals

Occupied orbitals have been localized historically in order to obtain an orbital picture, to help visualize transitions, and to serve as a starting point for local correlation methods. This last application is of interest since a localized occupied basis inherently contains the fewest interactions between occupied orbitals, and thus the magnitude of the integrals containing these localized orbitals should drop off faster with increasing distance between the atoms on which those basis functions are centered.[47] Thus, most local correlation methods begin with a direct localization of occupied orbitals.

Since single Slater determinant wave functions, such as the Hartree-Fock reference wave function, are invariant to unitary transformations within the space spanned by the molecular

orbitals $|\phi\rangle$, the occupied MOs can be localized by looking for unitary transformations U that maximize a specific criterion. The criterion is chosen to be directly correlated with the extent of localization of the orbitals. The resulting localized molecular orbitals (LMOs):

$$|\phi_i'\rangle = \sum_{j=1}^{N} U_{ji} |\phi_j\rangle \,; \quad UU^\dagger = I \tag{2.54}$$

are sparse in terms of contributions to the energy and the cost of the calculation can be reduced by truncating orbitals that have negligible contributions to the energy. Direct localization methods involve finding the extrema of functionals designed specifically to create maximum localization. For example, the Boys-Foster functional[63] minimizes the sum of squares of the difference between the orbital radii, $r_1 - r_2$. The Edmiston-Ruedenberg functional[64] maximizes the self-repulsion energy $|r_1 - r_2|^{-1}$, while the Pipek-Mezey functional[65] minimizes the gross Mulliken atomic population, and thus is not a simple distance criterion. While the Foster-Boys functional is popular and efficient, it has the issue of treating multiple bonds as τ bonds. The Edmiston-Ruedenberg functional is much more expensive, scaling as N^5, which is an issue in local correlation calculations where the localization of occupied orbitals is just the first step in a reduced-scaling scheme. The Pipek-Mezey functional treats σ and π bonds correctly, and is relatively fast, scaling as N^4, and is thus a safe choice when not dealing with genuinely delocalized systems.

Direct localization schemes such as the ones above are not practical for the localization of the virtual space.[47] However, the size of the virtual space usually causes the storage and computational bottleneck for post-Hartree-Fock methods and thus it is important to truncate the space, and therefore to attempt to localize it.

2.5.2 Pair Natural Orbitals

The PNO approach, introduced as pseudonatural orbitals by Edmiston and Krauss,[66] used in local correlation approaches by Ahlrichs and Driessler[53] and Meyer[67] and revived by Neese et al. as LPNOs,[68,69] is based on the creation of a basis of "natural spin orbitals" as defined by Löwdin.[70] In the examination of the properties of density matrices, it was proved that the natural orbital basis created by diagonalization of the first-order reduced density matrix is the basis in which the configuration interaction (CI) energy expansion gives the quickest convergence. Diagonalizing a pair density matrix to obtain an orbital space was seen by Edmiston and Krauss as creating a pseudonatural orbital basis, since the obtained space was an approximation to the true natural orbital basis. Obtaining each set of pair natural orbitals at the higher level of theory would require obtaining the density at that higher level and instead, alternate solutions have been presented. Ahlrichs and Driessler suggested an iterative method that converged to the exact density. Meyer started from the CID equations and neglected terms to obtain a PNO basis.

Neese and co-workers, in their paper on a new local PNO-CCSD method,[69] use doubles amplitudes obtained from second-order Møller Plesset perturbation theory to approximate coupled cluster-level amplitudes and combines the approximation with a weak pair approximation based on a lower-level pair correlation energy cutoff. The approximated amplitudes are used to create an approximated density, which is diagonalized to obtain the LPNO basis.

$$\mathbf{D}^{ij}\mathbf{Q}^{ij} = \mathbf{Q}^{ij}\mathbf{n}^{ij} \tag{2.55}$$

The pair density in terms of PNOs χ_k is written as:

$$\mathbf{D}^{ij} = \sum_a n_a^{ij} \chi_a^{ij*} \chi_a^{ij} \tag{2.56}$$

and for small values of the occupation number n_a^{ij}, the a-th term in the expansion can be neglected, as the contribution to the density (and hence to the energy) is small.

Additional considerations must be thought of for a production-level reduced-scaling approach, including the fact that PNOs are no longer orthonormal between pairs, thus introducing an overlap matrix within the canonical coupled cluster equations. Since there is now a distinct set of PNOs for every pair of occupied orbitals, the PNO space for each pair needs to be truncated aggressively to obtain an efficient method. This truncation of the space may lead to errors and can be accounted for by a lower-level correction, formed by computing the difference between pair energies of the full space and the truncated PNO space. A discussion of the balance of performance and accuracy of the LPNO method can be found in Ref 71.

In this section, we have reviewed the application of local correlation schemes to the reduction of the scaling of calculations of ground- and excited-state energies, as well as response properties. We have paid special attention to the method of pair natural orbitals, having been successfully used in the approximation of large-scale electronic ground state energy calculations.

Chapter 3

Modeling Complex Solvent Effects on the Optical Rotation of Chiral Molecules: A Combined Molecular Dynamics and Density Functional Theory Study

3.1 Introduction

The accurate prediction of almost any property of a molecule in solution can be a daunting task for electronic structure theory, particularly for optical properties of chiral compounds, which can be essential for the determination of absolute stereochemical configurations.[4] While the quantum chemical protocols for determining many molecular properties are now well established, chiroptical properties in the condensed phase such as optical rotation (OR), circular dichroism (CD), vibrational Raman optical activity (ROA), and circularly polarized luminescence (CPL), present unique challenges even for state-of-the-art methods.[72–78] Whereas continuum or other implicit solvent models are sufficient for simpler properties such as reaction energetics, for which the solvent acts as a relatively small perturbation, chiroptical properties are inherently non-addivitive and thus intrinsic to the combined solute/solvent system, at least within an incompletely defined "chiral imprint" region.[79,80] Indeed, it is well known that the choice of solvent can lead even to a change in the *sign* of the chiroptical response in some cases.[81] Thus, a vital question stands as to what extent dynamic, explicit solvation approaches must be employed to obtain reliable, accurate predictions of such complex properties.

Among implicit solvation approaches, the polarizable continuum model (PCM)[82,83] is by far the most successful and widely used. In 2002, Mennucci and co-workers[84] combined density-

functional theory (DFT) with the PCM to predict the specific rotations of methyloxirane in a range of solvents. While this approach has yielded much fruit for a variety of spectroscopic properties, it was unable to reproduce the sodium D-line ordering of the solvents examined by Kumata and co-workers in 1970[81] and more recently by Vaccaro and co-workers,[85] — errors which they attributed to non-electrostatic effects.

In 2005, Marchesan and co-workers[86] reported a DFT/PCM analysis of the sodium D-line OR of the conformationally flexible compound paraconic acid in methanol and found that they could reproduce the experimental absolute configurational assignment. While the Boltzmann-averaged rotations depended substantially on the choice of DFT vs. second-order perturbation theory (MP2) optimized geometries, the primary impact of the PCM model was on the relative energies (and thus the Boltzmann populations) of the individual conformers rather than the specific rotation of any individual structure. Kongsted and Ruud reported[87] DFT/PCM-based calculations of the OR dispersion (ORD) of methyloxirane in several solvents, including zero-point vibrational corrections, which have been found to be significant in some cases. Although the correct sign was obtained for methyloxirane in water (albeit at a factor of three too large at short wavelengths), the incorrect sign of the OR as compared to experiment was found for both cyclohexane and acetonitrile solvents.

Interestingly, four years later Egidi and co-workers[88] carried out DFT/PCM calculations of the OR of methyloxirane including vibrational corrections and reported quite the opposite results of Kongsted and Ruud, namely good agreement with experimental specific rotations for cyclohexane and acetonitrile, but poorer comparison for water. They attributed these

discrepancies in part to differences in the choice of PCM cavity, as well as local field effects and the formulation of the vibrational corrections. Egidi and co-workers also observed that solvent and vibrational effects are not additive, *i.e.*, vibrational corrections must be determined in the presence of solvent for a physically realistic description of the chiroptical response. Most recently, Aharon and co-workers[89] reported OR calculations for a series of rigid organic compounds using the PCM approach in conjunction with both DFT and the coupled cluster (CC) method, as well as zero-point vibrational corrections. They observed that, while CC response calculations (using the modified velocity gauge) generally provided better comparison to experiment, the PCM does not overall yield correct solvent shifts relative to vacuum or to other solvents.

Our own work in the realm of implicit solvent models has focused on the continuum dielectric model of Fattebert and Gygi[90] that is similar in spirit to the PCM but utilizes a definition of the dielectric permittivity that is a continuous function of the molecular electron density. For methyloxirane and other small, rigid organic molecules, we found[91] that, similar to the PCM, this smooth dielectric potential in concert with CC linear response theory generally fails to predict the correct sign and magnitude of the OR solvent shifts, in spite of its success with other properties such as excitation energies.[92]

As these examples make clear, and as noted by Mennucci and co-workers in their 2011 review,[72] the use of continuum-based models of solvation for the prediction of chiroptical properties ultimately fails to capture the key solute-solvent interactions associated with the chiral imprinting region surrounding the solute. Thus, extension of these approaches to more

expensive explicit or combined implicit/explicit approaches is necessary.

One strategy to the latter is the frozen density embedding (FDE) approach originally explored by Wesolowski and Warshel[93,94] in which the approximate electron density of the complete system is partitioned into solute and solvent components. The latter is then transformed to a one-electron embedding potential that can subsequently be incorporated into simulations of the complete solute-solvent system. In our own work in this domain, we examined a "wave-function-in-DFT" embedding approach that utilized less expensive DFT calculations of the full system to obtain the external solvent potential. However, we observed that the lack of response by the (frozen) solvent density to the external field resulted in too small an overall chiroptical response as compared to calculations that described the solvent-solute system explicitly. Niemeyer *et al.*[95] recently used subsystem DFT to investigate the failure of the uncoupled frozen density embedding method to describe chiroptical properties and concluded that missing solute-solvent couplings were the source of the discrepancy. They introduced the coupling of the environment to the perturbation, obtaining predictions of polarizability and optical rotation parameters comparable to explicitly-solvated systems computed using TD-DFT, for microsolvated systems of (*S*)-methyloxirane and (*P*)-dimethylallene in water, but using a more efficient DFT-in-DFT embedding approach. The authors also emphasized the importance of the choice of basis set in order to reproduce explicit solvent results.

In one of the most extensive dynamical simulations of the OR of methyloxirane in water, Lipparini and co-workers[96] utilized molecular dynamics simulations of the solute in a

box of water molecules, followed by a combined quantum mechanics/molecular mechanics (QM/MM) calculation on a set of snapshots extracted from the resulting trajectory. For each snapshot, the they chose DFT as the QM component, whereas the MM component was taken to be a polarizable force field described by a set of fluctating charges (FQ), along with PCM to describe bulk solvation effects and a vibrational correction obtained using the solute's equilibrium geometry. Using both B3LYP and CAM-B3LYP functionals, they obtained a good comparison to experimental measurements in water, though they reported that the lack of dominant solute-solvent configurations required a large number of snapshots (2000). They further emphasized the importance of vibrational corrections in cases such as methyloxirane where the electronic contribution to the OR is small, a conclusion similar to that observed for the gas-phase property.[87,97,98] On the other hand, a similar QM/FQ approach recently reported by Egidi and co-workers[99] on the OR of the conformationally flexible compound glycidyl methyl ether in water was unable to yield acceptable comparison with experiment, and that the simulations were highly sensitive to the details of the QM/MM computational protocols.

In 2006 and 2007, Beratan and co-workers[79,80] carried out sophisticated dynamical simulations of the OR of methyloxirane in both water and benzene. For snapshots extracted from classical dynamics trajectories, they retained a limited number of explicit solvent molecules closest to the solute using radial cutoffs, and subjected these to DFT calulations of the chiroptical response in the presence of a continuum solvent field using the COnductor-like Screening Model (COSMO) developed by Klamt and co-workers.[100] The selected radial cut-

offs retained up to nine water molecules and up to ten benzene molecules in the snapshots. For water, the explicit solvation calculations reproduced the experimental measurements with quantitative accuracy, whereas for benzene, the theoretical results were roughly a factor of two too small, but with the correct sign. In addition, for both solvents the authors carried out calculations of the OR of the solvent alone with the solute removed — *i.e.*, the "imprint" of the solute — and found that pure water makes a minor contribution to the total chiroptical response of methyloxirane, but the benzene contribution is nearly the total OR of the solute-solvent system. In 2008 and 2009, Kundrat and Autschbach[7,101] explored the effect of explicit solvent in the form of point charge water molecules on chiroptical properties of chiral amino acids, following a similar conformational sampling methodology to Beratan and coworkers. Their results emphasized the necessity of conformational sampling for the accurate modeling of chiroptical properties of solvated molecules, since even a single explicit water molecule had a large impact at close range on the properties of interest. The 2008 paper also reinforced, with timing data for the first excitation CD for 128 snapshots with up to 8 water molecules, the high computational cost of an averaged result when solvent molecules are treated explicitly. The comparison to experimental molar rotations proved once again the sensitivity of this property, especially for systems with small values of the rotation.

These pioneering studies have revealed many of the challenges associated with the *ab initio* calculation of chiroptical properties such as ORD. Clearly, some accounting of the explicit contribution of the solvent must be included, as well as sampling of the dynamical contri-

butions of the solute's configurational space. Furthermore, these studies, as well as our own recent work focused on the many-body expansion,[102] demonstrate that we cannot correctly view such properties as arising from a mere *perturbation* of the solute by the solvent, but instead as originating from the combined solute-solvent system extending to some as yet undetermined "imprinted" regime.

The focus of the present work is to extend the previous studies of explicit solvation to additional solutes and solvents, and to compare emerging computational protocols for explicit and QM/MM simulations. To that end, we will simulate the ORD of methyloxirane and methylthiirane — which, in spite of their apparent similarity in chemical composition, exhibit substantially different chiroptical response behavior — in water as well as the ORD of methyloxirane in methanol and carbon tetrachloride. In addition, we will consider the convergence of the dynamical ORD with respect to the number of snapshots taken over a given trajectory, as well as the impact of the choice of basis set and density functional on the resulting dispersion curves. In the same spirit as the earlier studies by Beratan and co-workers, we will also examine the extent of the chiral imprint on the solvent by the solute.

3.2 Computational Details

We used a combination of classical molecular dynamics (MD) and time-dependent DFT (TD-DFT) to compute the specific rotations of two chiral compounds: (*S*)-methyloxirane in water, methanol and CCl_4, and (*R*)-methylthiirane in water. Each MD simluation included

the solute molecule surrounded by a 5 nm box of solvent molecules (containing 5182 water molecules, 4354 methanol molecules, and 796 CCl_4 molecules, respectively) using an all-atom OPLS/AA forcefield[103] for the solutes, methanol, and CCl_4, and the SPC/E model for water.[104] Each 5 ns trajectory with a 2 fs time step was carried out in the NVT ensemble, with the solute and solvent coupled separately to a temperature bath at 300 K using a modified Berendsen thermostat and a coupling time of 0.1 ps. We recorded snapshots at every 2 ps along each trajectory, and extracted sets of 50, 100, and 250 coordinate sets evenly-spaced along the dynamical simulation. Finally, in order to make the subsequent quantum chemical calculations manageable, we reduced the solvent shell in each snapshot to include only up to the 12 solvent molecules nearest to the center-of-mass of the solute. Autocorrelation times were computed using the dipole moment of the solute plus the nearest 12 solvent molecules. The dipole autocorrelation function at time t was computed as a function of the dipole μ_i for all particles i in the selected subsystem:

$$C_\mu(t) = \langle \mu_i(t)\dot{\mu}_i(0)\rangle \tag{3.1}$$

with the autocorrelation time τ given as the integral

$$\tau_\mu = \int_0^\infty C_\mu(t)dt. \tag{3.2}$$

The autocorrelation function for each system computed over all saved snapshots of the trajectory can be found in the Supporting Information. Autocorrelation times were found to be 5 ps, 3 ps and 6 ps for the systems of (*S*)-methyloxirane in water, methanol and CCl_4 respectively. Autocorrelation times for the system of (*R*)-methylthiirane in water were

found to be 13 ps. Thus the snapshots sampled 20 ps apart were considered to be sufficiently uncorrelated.

For each snapshot, we computed specific rotations at four typical wavelengths of plane polarized light — 633 nm, 589 nm (the sodium D-line), 436 nm, and 355 nm — using TD-DFT with both B3LYP[105–107] and the long-range corrected CAM-B3LYP[108] functionals. While the former has been widely utilized in the calculation of chiroptical properties, the latter has recently been demonstrated to provide improved comparison with experiment for gas-phase compounds.[109,110] Gauge-including atomic orbitals (GIAOs) were used to ensure coordinate origin-independence.[111–114] To retain clear comparison between explicit and implicit solvent systems, as well as between explicit solvent systems with variable numbers of solvent molecules, we used only the mass of the solute in the calculation of the specific rotation. (This is in contrast to previous work by Beratan and co-workers, who used the mass of the explicit solute-solvent system.[79,80]) Bulk solvent effects beyond those described using explicit solvent molecules were incorporated into each snapshot using PCM.[82,83] Total specific rotations were computed both as simple averages and Boltzmann-like averages based on computed correlation energies over a given number of snapshots. Since the system energies were narrowly distributed, the Boltzmann-like weighting did not have a large impact on the final value of the averaged OR and simple averaging was used in the sections below. We tested the impact of the choice of basis set using both the correlation-consistent, diffuse function-augmented aug-cc-pVDZ[115,116] basis set, and the optical rotation prediction (ORP)[117,118] basis set in the TD-DFT calculations. The ORP basis set contains 13 basis

functions for hydrogen and helium and 32 for second-row elements, while the aug-cc-pVDZ basis set contains 9 and 23 functions, respectively. The ORP basis set is optimized with respect to atomic polarizabilities, and is thus designed for optical rotation calculations. All MD simulations were carried out using Gromacs 4.0.5[119] and all quantum chemical calculations using Gaussian09.[120]

3.3 Results and Discussion

3.3.1 Sampling

Figure 3.1 provides plots of the specific rotation of (*R*)-methylthiirane in a cage of 12 water molecules calculated using up to 250 evenly-spaced snapshots over the full molecular dynamics trajectory, as well as the cumulative running average of the specific rotation for four common wavelengths of polarized light. As shown in Fig. 3.1(a), the variations in the specific rotation with structure — both the configuration of the solvent and the internal vibrations of the solute — are large, ranging from nearly -800 to $+800$ deg dm^{-1} (g/mL)$^{-1}$. This is consistent with previous studies of conformational dependence[121] and solvent impact[96] on optical rotations. The magnitude of the variation is larger in the case of (*R*)-methylthiirane than for (*S*)-methyloxirane by roughly a factor of seven due to the presence of lower-lying diffuse (Rydberg-type) electronic states in the former.[122] However, as shown in Fig. 3.1(b), the running average converges reasonably quickly, particularly for the longer wavelengths,

suggesting that 250 snapshots are enough to sample the trajectory well. We observe similar trends for (*S*)-methyloxirane in water, methanol, and carbon tetrachloride (see Supporting Information).

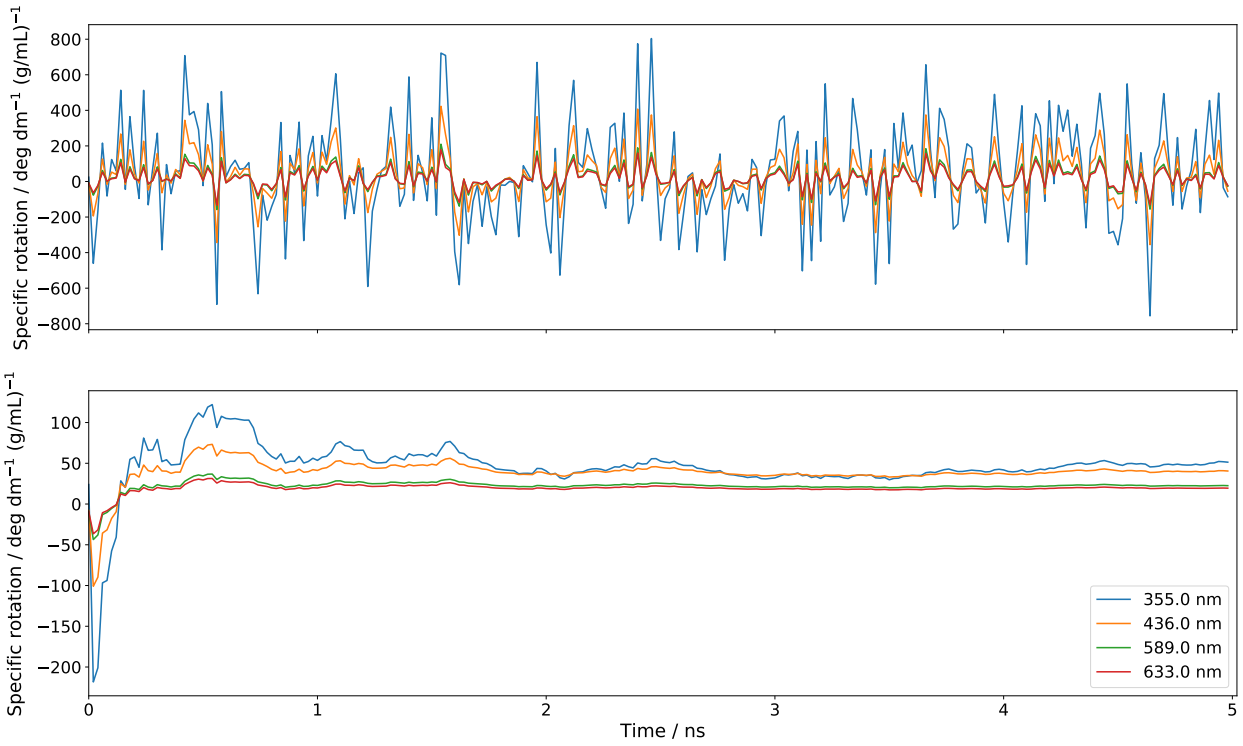


Figure 3.1: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

Figure 3.2 depicts the convergence of the average specific rotation with respect to the number of explicit water molecules included in the calculation for (*R*)-methylthiirane, whereas Figures 3.3-3.5 depict the corresponding values for (*S*)-methyloxirane in water, methanol,

and CCl_4, respectively. For both compounds and long wavelengths of polarized light, the specific rotation in water converges between eight and ten solvent molecules for 250 trajectory snapshots, and is nearly unchanged even for as few as 50 snapshots. The convergence behavior of the averaged rotation is poorer for shorter wavelengths, though for 250 snapshots convergence appears finally to be reached between ten and 12 solvent molecules. Comparing our results with the work of Kundrat and Autschbach,[101] in which the first solvation shell around the solute ended at around 16 waters, this is in agreement with our number of 10-12 solvent molecules for the converged systems. We note that, while the number of explicit solvent molecules required is consistent with previous work by Beratan and co-workers,[79,80] this is a significantly smaller number of snapshots than reported in those previous studies, in which 1000-2000 structures were used for (*S*)-methyloxirane in both water and benzene. However, Figure S3 of the Supporting Information to their work on (*S*)-methyloxirane in water[79] indicates approximate convergence of the average to within 1 deg dm^{-1} $(g/mL)^{-1}$ in approximately 200 snapshots, with subsequent sampling serving to reduce the error-bar estimate.

Table 3.1 shows the 3-point standard deviation for three consecutive points on the curves averaged over 50 and 250 snapshots in Figure 3.2. The standard deviation values decrease with increasing numbers of solvent molecules. This remains true for all wavelengths across systems, with the smallest values being < 6 deg dm^{-1} $(g/mL)^{-1}$, usually between 8-10 solvent molecules when averaged over 250 snapshots. The small deviations indicate better sampling when 250 snapshots are taken into account. The largest deviations are seen for the

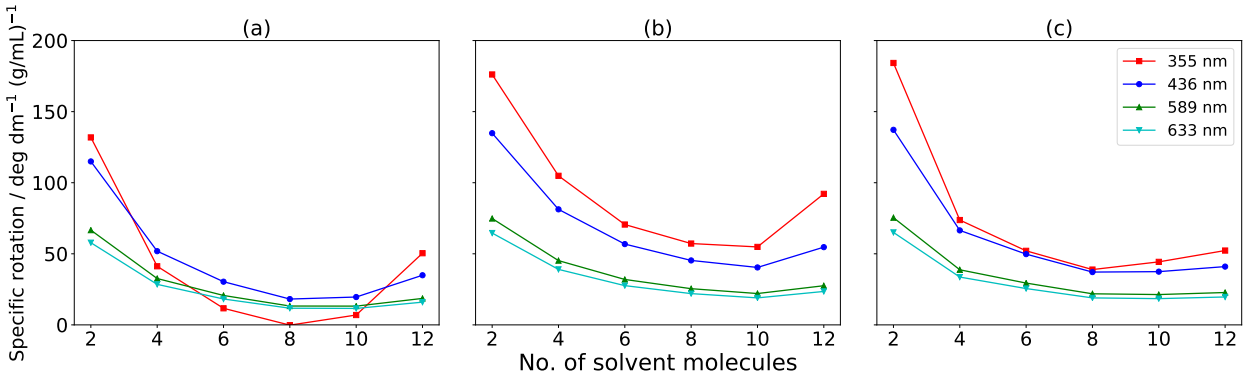


Figure 3.2: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of (*R*)-methylthiirane in water for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained for (a) 50, (b) 100, and (c) 250 snapshots along the MD trajectory.

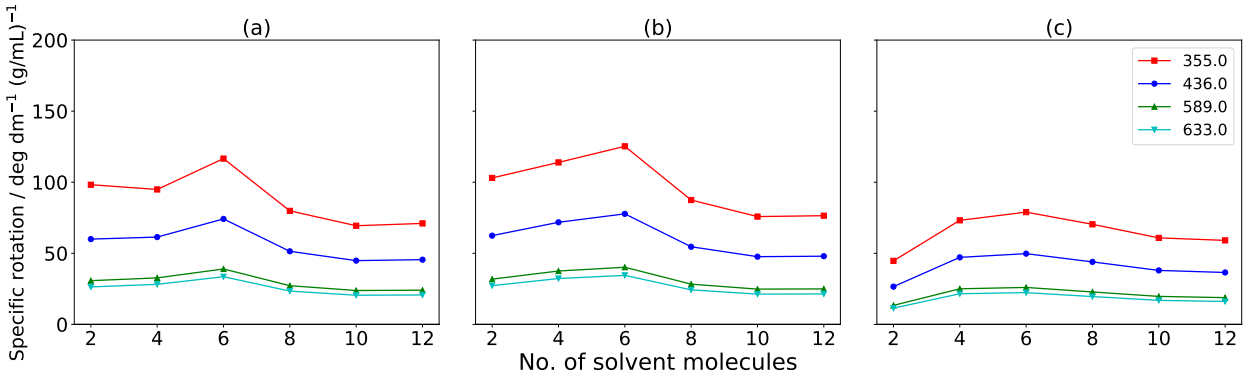


Figure 3.3: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of (*S*)-methyloxirane in water for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained for (a) 50, (b) 100, and (c) 250 snapshots along the MD trajectory.

2-4-6 solvent molecules, often significantly different from the deviations for larger numbers of solvent molecules.

As shown in Fig. 3.4, for (*S*)-methyloxirane in methanol, convergence is overall somewhat

Table 3.1: 3-point standard deviation (in deg dm^{-1} $(g/mL)^{-1}$) for the OR of (*R*)-methylthiirane in water, averaged over 50 and 250 trajectory snapshots, at the CAM-B3LYP/aug-cc-pVDZ level of theory. Standard deviations were computed for i number of solvent molecules using the averaged optical rotation for the range $(i-2, i+2)$ of solvent molecules.

Solvent	50 snapshots				250 snapshots			
molecules	355	436	589	633	355	436	589	633
4	51.14	35.90	19.47	16.77	57.87	37.90	19.85	17.03
6	17.41	13.92	7.97	6.91	14.36	12.01	6.91	5.99
8	4.87	5.45	3.52	3.09	5.41	5.88	3.69	3.23
10	22.36	7.58	2.54	2.04	5.50	1.75	0.59	0.48

slower even for the longer wavelengths, but a final value is still reached between ten and 12 explicit solvent molecules. However, for (*S*)-methyloxirane in CCl_4, Fig. 3.5 shows that the observed behavior is significantly different from the other solvents, with smaller numbers of explicit solvent molecules and smaller numbers of snapshots yielding negative rotations. With 250 snapshots and ten CCl_4 molecules, the rotation turns positive, but does not yet appear to be converged. (Larger numbers of solvent molecules were beyond the capabilities of our current computational resources.) This behavior can be traced, in part, to the presence of longer wavelength Rydberg-type excited states for (*S*)-methyloxirane in CCl_4 than in CH_3OH or H_2O. For example, CCl_4 exhibits a low-lying $4p$ transition that is roughly 0.5 eV lower in energy than the lowest $3s$ state in methanol (at the CAM-B3LYP/aug-cc-pVDZ//B3LYP/6-

31G* level of theory). Such solvent-centered states yield larger variations in (and thus greater sensitivity of) the ORD to environmental effects, and they affect the short-wavelength OR more strongly than their long-wavelength counterparts due to the contribution of the excitation energies to the denominator of the Rosenfeld tensor.[4,123]

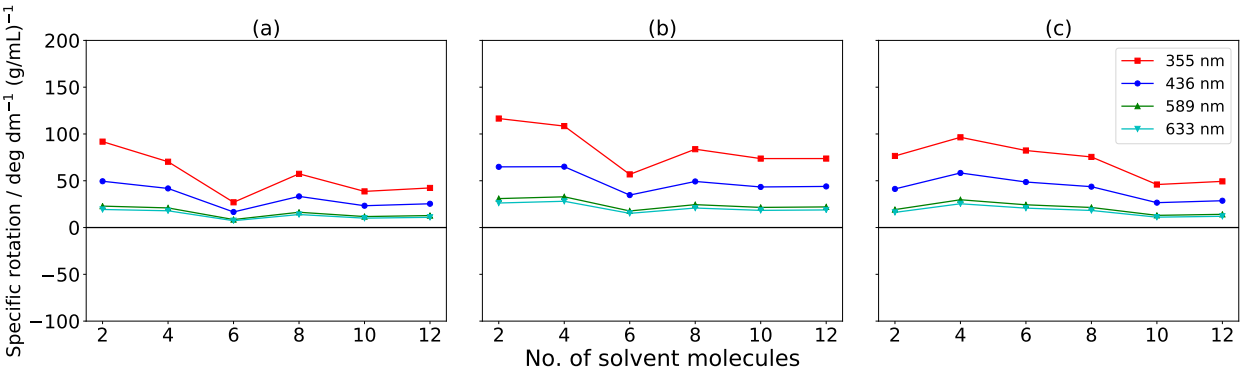


Figure 3.4: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of (*S*)-methyloxirane in methanol for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained for (a) 50, (b) 100, and (c) 250 snapshots along the MD trajectory.

3.3.2 Choice of Density Functional

To examine the effect of the choice of density functional on the convergence behavior with respect to number of explicit solvent molecules, the hybrid B3LYP and the long range-corrected CAM-B3LYP functionals were chosen. Figure 3.6 shows specific rotations computed for the aug-cc-pVDZ basis set for (*R*)-methylthiirane in water. A large oscillation is seen in the sign of specific rotation values computed using B3LYP at the smallest wavelength considered,

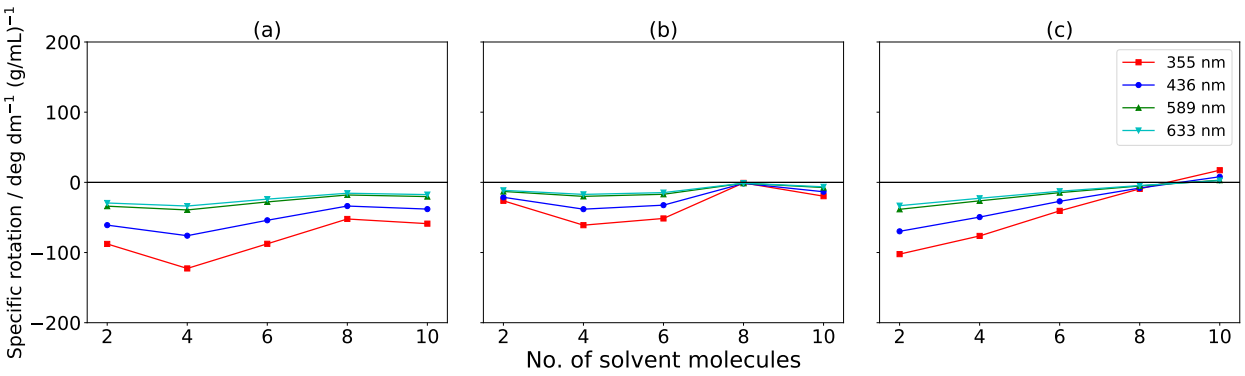


Figure 3.5: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of (*S*)-methyloxirane in CCl_4 for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained for (a) 50, (b) 100, and (c) 250 snapshots along the MD trajectory.

355 nm, as more solvent molecules are added to each snapshot. This observation is consistent across the two basis sets used. In comparison, specific rotation values computed using B3LYP at the longer wavelengths of 589 nm and 633 nm show little oscillation for both basis sets. On using the long range-corrected hybrid functional CAM-B3LYP for calculations on the same system and geometries, there are no oscillations in the sign of rotation values computed at 355 nm, and improved convergence at higher wavelengths is seen on the addition of more explicit solvent molecules.

The oscillations can be quantitatively seen through the slope data in Table 3.2, generated as the slope of the line between data points in Figure 3.6. For the B3LYP functional at 355 nm, the slope varies from −112.08 to +2302.42 and then to −973.60 as the number of explicit solvent molecules is increased from 6 to 8 to 10, while the same set of calculations with the CAM-B3LYP functional shows a non-oscillating, though still not converged, trend

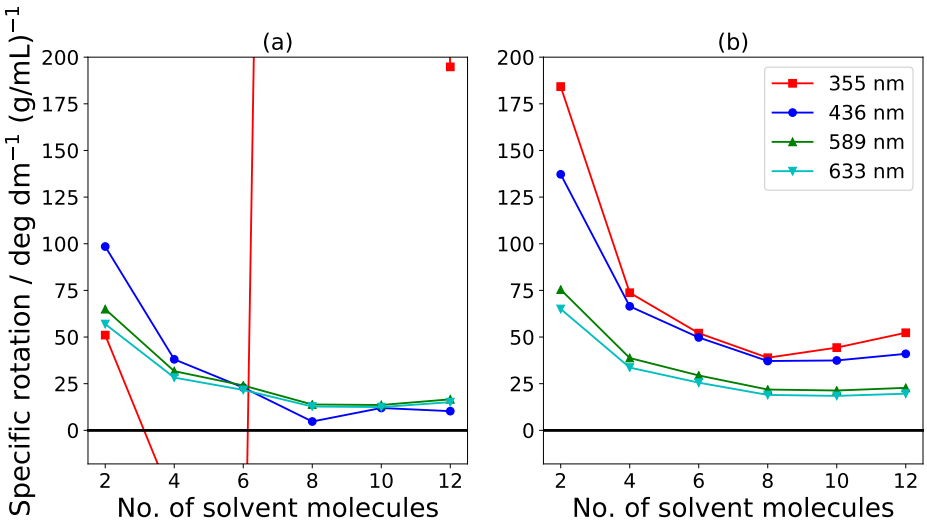


Figure 3.6: Average specific rotations of (*R*)-methylthiirane in water for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained calculated with the (a) B3LYP and (b) CAM-B3LYP functionals along the MD trajectory.

to a final slope value of 7.96 from 10 to 12 solvent molecules. The longer wavelengths with the CAM-B3LYP functional show a slope of less than a degree for the 8-10 and 10-12 values, indicating convergence at 8-10 water molecules.

The averages computed using B3LYP at an external-field wavelength of 355 nm were skewed by relatively small sets of snapshots, each of which had specific rotations of magnitude $>$ 10000 deg dm^{-1} (g/mL)$^{-1}$. TD-DFT excitation energy calculations for the lowest few excited states of snapshots with abnormally large B3LYP optical rotation values using the B3LYP functional exhibit first excitation wavelengths of between 355-356 nm, with the corresponding CAM-B3LYP excitation wavelength being between 262-277 nm for the same system. The B3LYP functional is known to underestimate the excitation energy and overestimate the

Table 3.2: Slope data in deg/[dm g/mL] per added molecule for the OR of the system of *(R)*-methylthiirane in water, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the B3LYP and CAM-B3LYP DFT functionals and the aug-cc-pVDZ basis set.

Solvent	B3LYP				CAM-B3LYP			
molecules	355	436	589	633	355	436	589	633
4	−90.70	−60.46	−33.22	−28.68	−110.49	−70.75	−36.65	−31.40
6	−112.08	−15.12	−7.66	−6.64	−21.65	−16.69	−9.35	−8.09
8	2302.42	−18.20	−10.19	−8.88	−13.18	−12.63	−7.55	−6.57
10	−973.60	7.24	−0.27	−0.30	5.42	0.30	−0.53	−0.52
12	−982.27	−1.69	3.04	2.66	7.96	3.55	1.43	1.18

excitation wavelength,[124] thus the TD-DFT response approaches a singularity at 355 nm, modelling an electronic transition at a longer wavelength than the corresponding CAM-B3LYP response calculation. Using the CAM-B3LYP functional shifts the transition to a shorter wavelength, thus the convergence at 355 nm with respect to the number of solvent molecules is better-behaved.

While we have mainly considered the system of (R)-methylthiirane in water, it is to highlight the marked difference in behavior seen while using the B3LYP functional versus the CAM-B3LYP functional. The system of (S)-methyloxirane in water does not show a large deviation based on the functional used. Based on the well-behaved convergence of the OR using the

CAM-B3LYP functional for the specific case of (*R*)-methylthiirane in water, we have used the CAM-B3LYP functional to compute rotations for the larger solvents of methanol and CCl_4 in our comparison across systems solvating (*S*)-methyloxirane.

3.3.3 Choice of Basis Set

Two basis sets were chosen to probe the effect of the choice of basis set on the averaged optical rotation, the augmented correlation-consistent aug-cc-pVDZ basis set,[115,116] and the larger ORP basis set.[117,118] On average across systems, the number of basis functions used in calculations with the ORP basis set were 1.5× the number used in calculations with the aug-cc-pVDZ basis. The ORP basis set is expected to perform better for the property of specific rotation, thus the trade-off between computational accuracy and expense is considered.

Standard deviation data for (*R*)-methylthiirane and (*S*)-methyloxirane in water, as seen in Tables 3.3 and 3.4, show very similar values and trends for both basis sets. For the (*R*)-methylthiirane system, calculations with the aug-cc-pVDZ basis set have deviations between 17-58 deg dm^{-1} $(gm/mL)^{-1}$ across wavelengths for the smallest number of molecules, while calculations with the ORP basis set have deviations spanning 14-46 deg dm^{-1} $(gm/mL)^{-1}$ across wavelengths. Similar trends are observed as the number of solvent molecules included in the calculation are increased, with the ORP basis set showing smaller standard deviation values, indicating better convergence with the number of solvent molecules added.

Figure 3.7 plots the specific rotation as a function of the number of solvent molecules added

Table 3.3: 3-point standard deviation (in deg dm^{-1} $(g/mL)^{-1}$) for the OR of the system of (*R*)-methylthiirane in water, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the CAM-B3LYP functional. Standard deviations were computed for i number of solvent molecules using the averaged optical rotation for the range $(i-2, i+2)$ of solvent molecules.

Solvent	aug-cc-pVDZ				ORP			
molecules	355	436	589	633	355	436	589	633
4	57.87	37.90	19.85	17.03	46.06	30.99	16.42	14.10
6	14.36	12.01	6.91	5.99	9.55	9.55	5.76	5.01
8	5.41	5.88	3.69	3.23	4.97	3.37	2.37	2.10
10	5.50	1.75	0.59	0.48	6.28	1.54	0.37	0.30

for both basis sets for the system of (*S*)-methyloxirane in methanol. We see an improvement in the convergence, as indicated by the change in slope between values, particularly between 10 and 12 solvent molecules, for each wavelength considered and conclude that the ORP basis set constitutes an improvement over the aug-cc-pVDZ basis. The slight improvement seen here has a corresponding additional computational cost of a 50% increase in basis functions. Even so, in order to obtain the most accurate values for comparison to experiment, we chose to use the ORP basis in further calculations on the system of (*S*)-methyloxirane in CCl_4.

Table 3.4: 3-point standard deviation (in deg dm^{-1} $(g/mL)^{-1}$) for the OR of the system of (*S*)-methyloxirane in water, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the CAM-B3LYP functional. Standard deviations were computed for *i* number of solvent molecules using the averaged optical rotation for the range (i−2, i+2) of solvent molecules.

Solvent	aug-cc-pVDZ				ORP			
molecules	355	436	589	633	355	436	589	633
4	14.98	10.38	5.74	4.97	9.93	7.10	3.99	3.46
6	3.57	2.37	1.33	1.16	4.36	2.94	1.65	1.43
8	7.40	4.82	2.57	2.22	7.51	4.94	2.66	2.29
10	4.96	3.22	1.71	1.48	4.37	2.89	1.56	1.35

3.3.4 Use of Implicit Solvent

To capture longer-range solvent interactions than the relatively few explicit solvent molecules composing our system allowed, we used an implicit solvent model, in this case the polarizable continuum model (PCM),[125] whose constant dielectric continuum began outside of the explicit solvent shell. This would, hypothetically, add in bulk solvent effects beyond the explicit solvent molecules and thus improve the solvated specific rotation value.

Comparison of Figures 3.8(a) and (b) shows the effect of the addition of the PCM implicit solvent model to the calculation for (*R*)-methylthiirane in water using the CAM-B3LYP

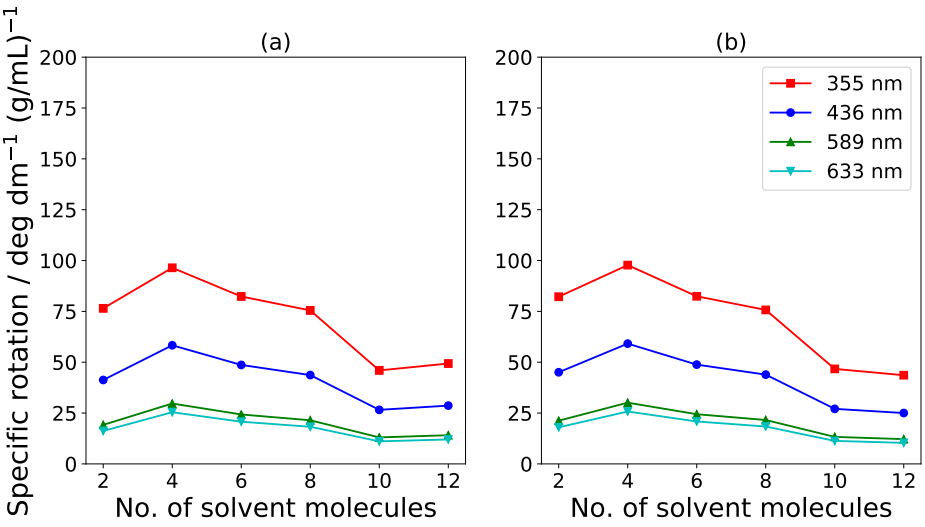


Figure 3.7: Average CAM-B3LYP specific rotations of (*S*)-methyloxirane in methanol for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained using the (a) aug-cc-pVDZ and (b) ORP basis sets over 250 snapshots along the MD trajectory.

functional. The CAM-B3LYP calculation shows slightly better convergence without the addition of implicit solvent at 10-12 water molecules, as compared to the calculation with implicit solvent. On the other hand, the B3LYP calculation shows a large change in trend from the oscillation seen at an external-field wavelength of 355 nm in Figure 3.9(a) to a fairly well-behaved graph in Figure 3.9(b). Examining the excitation energies for the snapshots with abnormally large specific rotations, we see a shift in the computed first excitation wavelength with the addition of the implicit solvent model, such that it no longer coincides with the smallest wavelength at which the specific rotation is computed, 355 nm. That the addition of the implicit solvent shifts the wavelength towards shorter wavelengths does not

necessarily indicate a correction, since the computed CAM-B3LYP excitation energies show a similar shift towards shorter wavelengths.

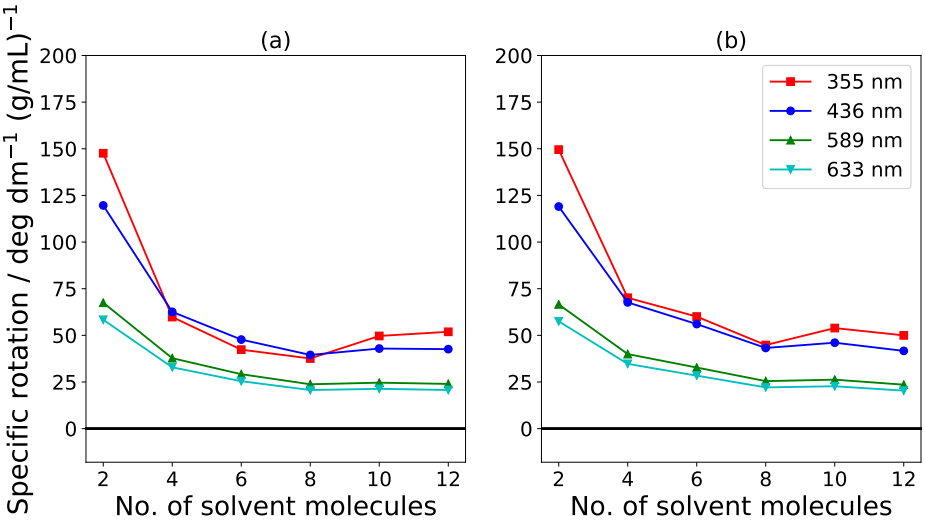


Figure 3.8: Average specific rotations of (*R*)-methylthiirane in water for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained calculated (a) without and (b) with the inclusion of the PCM implicit solvent model and the CAM-B3LYP functional, and the ORP basis set along the MD trajectory.

Tables 3.5 - 3.7 detail the effect of adding in implicit solvent around the explicit solvent shell for our (*S*)-methyloxirane systems in water, methanol, and carbon tetrachloride, respectively. As the comparison between standard deviations in Table 3.6 indicates, the methyloxirane in methanol system shows slightly better convergence with the use of implicit solvent. However, the values of averaged optical rotation at each number of solvent molecules vary only by 7.63 deg dm^{-1} (g/mL)$^{-1}$ at maximum, thus the effect of adding the implicit solvent is quantitatively small.

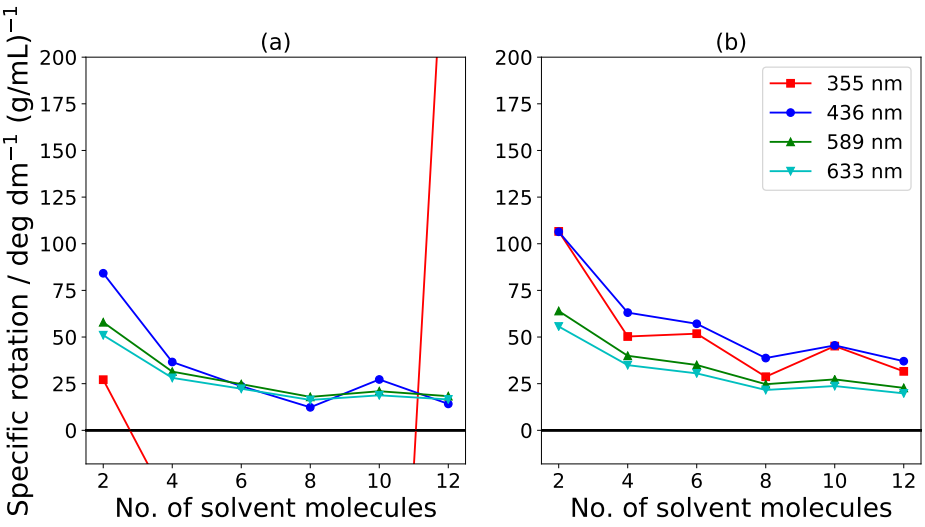


Figure 3.9: Average specific rotations of (*R*)-methylthiirane in water for several wavelengths of light as a function of the number of solvent molecules closest to the solute retained calculated (a) without and (b) with the inclusion of the PCM implicit solvent model and the B3LYP functional, and the ORP basis set along the MD trajectory.

For (*S*)-methyloxirane in water and methanol, the addition of an implicit solvent model does not significantly change the convergence of the specific rotation value with respect to the number of solvent molecules. For the same solute in CCl_4, we see a more significant shift on the addition of the implicit solvent towards larger standard deviation values and indicating that the rotation values do not converge even more strongly. These results are in line with results obtained by Schieschke *et al.*,[126] on the small effect of adding a continuum solvent model on top of frozen density embedding on solvent shifts of the vertical excitation energy.

Table 3.5: Standard deviation data for the OR of the system of (*S*)-methyloxirane in water, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the CAM-B3LYP functional.

Solvent	Without PCM				With PCM			
molecules	355	436	589	633	355	436	589	633
4	9.93	7.10	3.99	3.46	2.60	2.88	1.95	1.73
6	4.36	2.94	1.65	1.43	7.34	4.86	2.64	2.28
8	7.51	4.94	2.66	2.29	9.04	5.70	2.98	2.56
10	4.37	2.89	1.56	1.35	5.72	3.60	1.88	1.61

3.3.5 Comparison with Experiment

The simulated ORD spectra of methyloxirane in water, methanol and carbon tetrachloride were compared to experimental solvated data in order to judge the effectiveness of this mixed QM/MM methodology, as outlined in Section 2. In light of results from Sections 3.1-3.4, the average was taken over 251 snapshots, with the largest explicit solvent shell used, 12 molecules for water and methanol and 10 for carbon tetrachloride, and the specific rotation calculated using TD-DFT with the CAM-B3LYP functional and the ORP basis set. Fitted ORD spectra computed using data points at 4 wavelengths for (*S*)-methyloxirane (Figure 3.10) show a significant error when compared to experimental ORD curves, obtained from

Table 3.6: Standard deviation data for the OR of the system of (*S*)-methyloxirane in methanol, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the CAM-B3LYP functional.

Solvent	Without PCM				With PCM			
molecules	355	436	589	633	355	436	589	633
4	7.27	5.95	3.66	3.21	6.05	3.97	2.47	3.02
6	9.22	6.35	3.54	3.07	7.98	5.45	3.02	2.62
8	15.49	9.31	4.73	4.04	13.62	8.24	4.20	3.60
10	14.46	8.46	4.20	3.58	13.33	7.99	4.00	3.42

Ref. 127. All spectra were fitted using an equation of the form:

$$[\alpha]_\lambda = \frac{A}{B^2 - \lambda^2} + C \tag{3.3}$$

where A, B and C are constant parameters determined by non-linear least-squares fitting, and λ is the external-field wavelength corresponding to the specific rotation value $[\alpha]_\lambda$.

The averaged value of the solute's specific rotation is, in all cases, large relative to the solvated system rotation, and positive in sign. The addition of the explicit solvent molecules shifts the curve closer to the experimental curve for all three solvents tested. The fitted curve for water-solvated (*S*)-methyloxirane in Figure 3.10(a) has the correct sign and trend of rotation at the given wavelengths; however, the values are off from the experimental values by between 14-45 deg dm^{-1} (g/mL)$^{-1}$. For the curve in Figure 3.10(b) for (*S*)-meth-

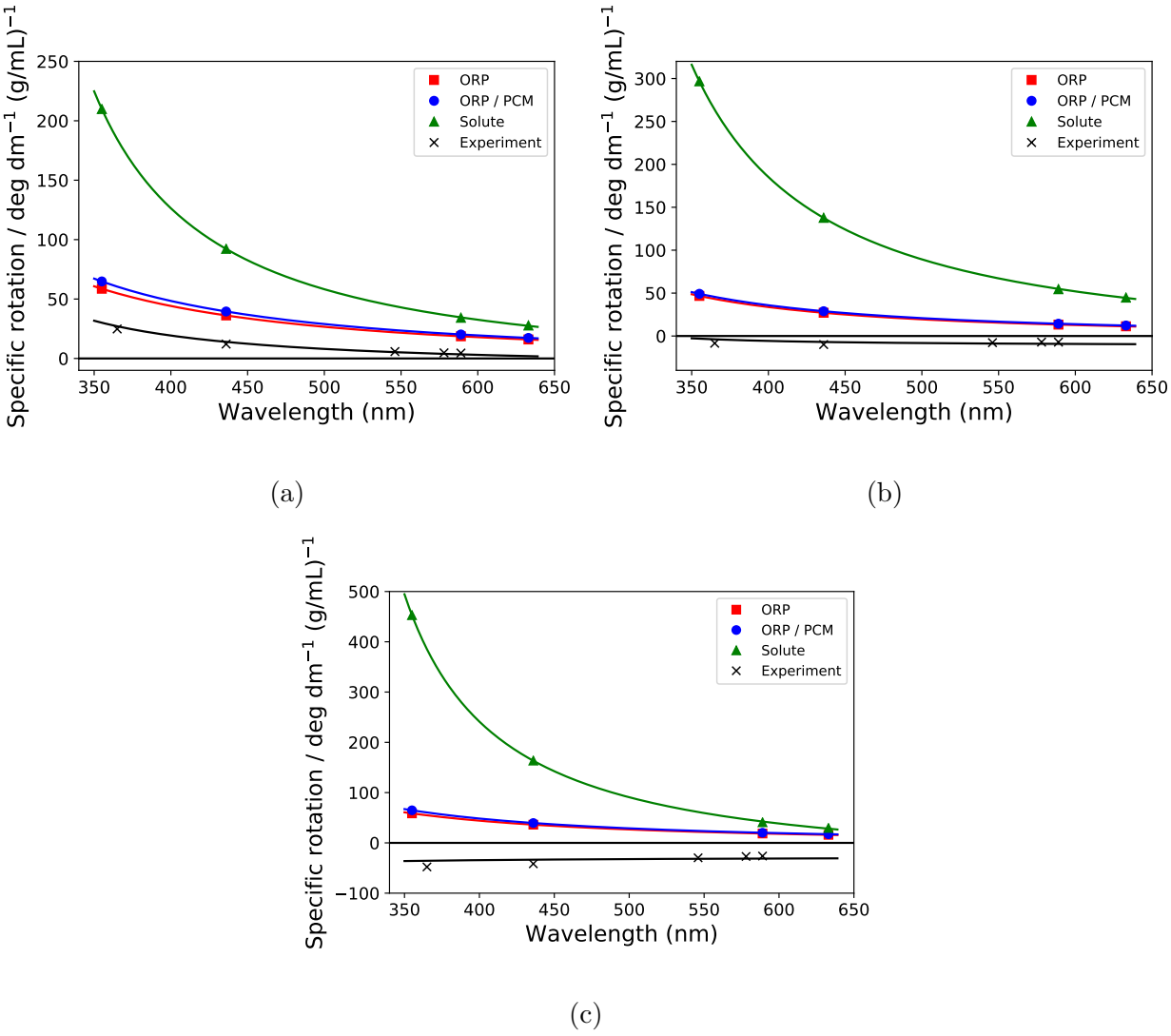


Figure 3.10: ORD spectra of *(S)*-methyloxirane in (a) water, (b) methanol and (c) CCl_4 over 250 evenly-spaced snapshots of a 5 ns trajectory. Optical rotation calculations were carried out with 12 explicit solvent molecules for the H_2O and methanol systems, and 10 explicit solvent molecules for the CCl_4 system using the CAM-B3LYP functional.

Table 3.7: Standard deviation data for the OR of the system of (*S*)-methyloxirane in carbon tetrachloride, averaged over 250 trajectory snapshots. OR values were computed with TD-DFT using the CAM-B3LYP functional.

Solvent	Without PCM				With PCM			
molecules	355	436	589	633	355	436	589	633
4	25.26	17.45	9.69	8.40	11.71	9.49	5.75	5.04
6	27.42	16.89	8.67	7.43	27.92	17.20	8.83	7.57
8	23.68	14.30	7.24	6.19	32.06	19.50	9.93	8.50

yloxirane in methanol, the signs of the optical rotation computed at each wavelength are incorrect, although the trend is correctly captured. However, for the system of (*S*)-methyloxirane in CCl_4 (Fig. 3.10(c)), both the sign and the trend of the rotation do not match the experimental curve. The inclusion of implicit solvent around the explicit solvent shell shifts the computed curves away from the experimental values by a small amount, not exceeding 6 deg dm^{-1} $(g/mL)^{-1}$.

Table 3.8 shows the errors with respect to experiment at the wavelengths of 436 and 589 nm for the systems of (*S*)-methyloxirane in different solvents. Beyond 6 solvent molecules, the error monotonically decreases with each additional molecule for the systems in water and methanol. For both solvents, the difference between the errors with respect to experiment for 10 and 12 solvent molecules are approximately 1-2 deg dm^{-1} $(g/mL)^{-1}$ for both wavelengths. This convergent behavior suggests that the addition of explicit solvent molecules would not

Table 3.8: Errors with respect to experiment in deg dm^{-1} $(g/mL)^{-1}$ for the OR of the system of *(S)*-methyloxirane in three solvents, averaged over 250 trajectory snapshots. OR values were computed with the CAM-B3LYP functional and the ORP basis set.

Solvent	H_2O		CH_3OH		CCl_4	
molecules	436	589	436	589	436	589
2	21.93	12.81	54.94	28.35	−48.77	−26.40
4	35.55	20.76	69.03	37.21	−28.49	−14.35
6	38.10	21.70	58.71	31.54	−6.05	−2.66
8	30.98	17.83	53.79	28.7	12.83	6.86
10	26.08	15.23	36.96	20.40	31.79	16.47
12	24.09	14.09	35.67	19.66	–	–

improve the result of the calculation, indicating a deficiency in the level of theory used or inadequate sampling of the MD simulation. The system in CCl_4 begins with a large negative error, that becomes more positive with larger numbers of added solvent molecules, until we obtain a large positive error, which when compared to the small negative experimental values, gives us specific rotation values that are incorrect in sign. This, together with the fact that we do not see convergence at 8-10 solvent molecules, is evidence that larger numbers of solvent molecules are necessary to obtain the right sign for the CCl_4 system. Given the error shown by the other systems on convergence with respect to the number of solvent molecules, it is expected that the converged CCl_4 system will still have significant error related to the problems in QM method or MD sampling mentioned above.

3.3.6 Solvent-Only Specific Rotations

Similar to calculations done by Mukhopadhyay *et al.*,[6,128] a solvent-only shell was created by removing the solute atoms from the snapshot geometries, and the above variations in calculations were run on this shell for all four solvated systems. The purpose was to find out if the specific rotation of the solvent shell alone (the "imprint") formed a large contribution to the total optical rotation of the system. Figure 3.11 plots specific rotation values as a function of the number of solvent molecules making up the shell. The behavior of the rotation varies dramatically with the solvent, with the values of the rotation for the solvent-only system remaining close to 0 for (*S*)-methyloxirane in water, while it fluctuates from small to fairly large values with the addition of explicit solvent molecules for (*S*)-methyl-

oxirane in both CCl_4 and methanol. Unlike the converging results obtained for the system optical rotation for (*S*)-methyloxirane in water and methanol, the rotation values obtained for the solvent shells alone do not converge with respect to the number of solvent molecules. Interestingly, the sign change seen in calculations on the full system of (*S*)-methyloxirane in CCl_4 is mirrored by the solvent-only system, with the specific rotation going from negative to positive on moving from 6 to 8 explicit solvent molecules.

Figure 3.12 includes ORD curves of the solvent-only as well as the solute and solvent systems, plotted together with experimental curves of methyloxirane in solution. We compare the converged 12 solvent molecule systems to their solvent-only shells to examine the contribution of the solvent-only curves to the system optical rotation. Although the solvent-only curve for the system in water more closely approximates experimental values at each wavelength than the solute+solvent curve, it does not follow the trend as well. Similarly, for the methanol system, the curve remains positive and relatively flat, in contrast with the negative experimental values. The experimental curve for the methanol system curves upwards on moving towards smaller wavelengths due to an experimental value of $+-$ at 302 nm, not shown in Fig. 3.12(b). In both the water and methanol systems, the values at all wavelengths are very small, especially when seen as a percentage of the system optical rotation value. This is in agreement with findings by Mukhopadhyay *et al.* for the water system, that the contribution of the solvent-only shell is a fraction of the total system optical rotation. However, the larger solvent of methanol also shows small values of the OR for the solvent shell alone, and while the largest system in CCl_4 examined shows a large positive value of

the OR relative to the calculated system OR, it is not converged at 10 solvent molecules. Thus, in contrast to their finding that the rotation of the solvent shell of benzene formed a large contribution to the system OR, our calculations find that the solvent shell OR is small relative to the calculated system rotation.

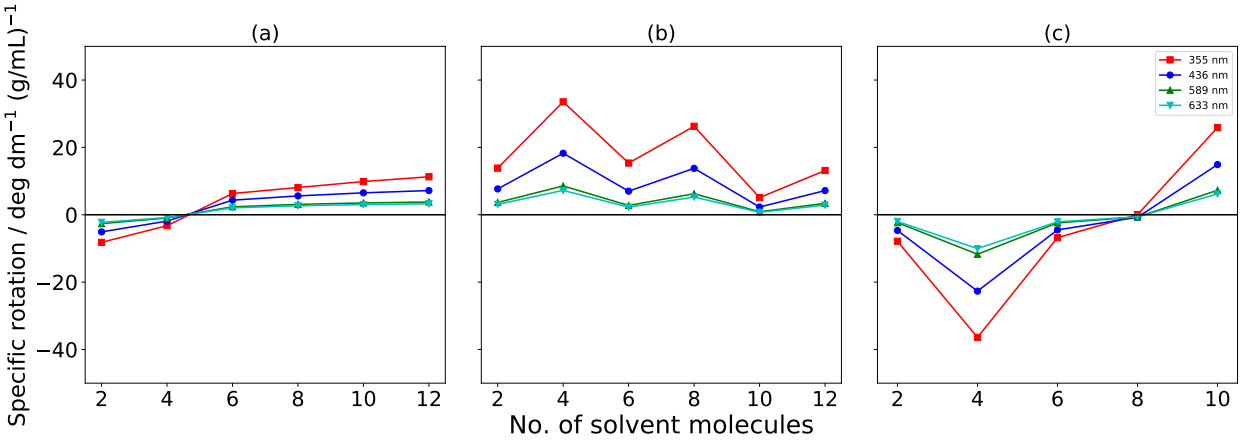


Figure 3.11: Specific rotation [α] of *(S)*-methyloxirane in (a) water (b) methanol and (c) CCl_4 over 250 evenly-spaced snapshots of a 5 ns trajectory. Optical rotation calculations were carried out using the CAM-B3LYP functional and the ORP basis set.

3.4 Conclusions

In this work, we have explored a computational protocol for the specific rotations of chiral compounds in a solvent environment. In particular, we have used a combination of molecular dynamics and time-dependent density functional theory to simulate the multi-wavelength optical rotation of (*S*)-methyloxirane in water, methanol, and carbon tetrachloride, and of (*R*)-methylthiirane in water. In spite of dramatic variation of the optical rotation obtained

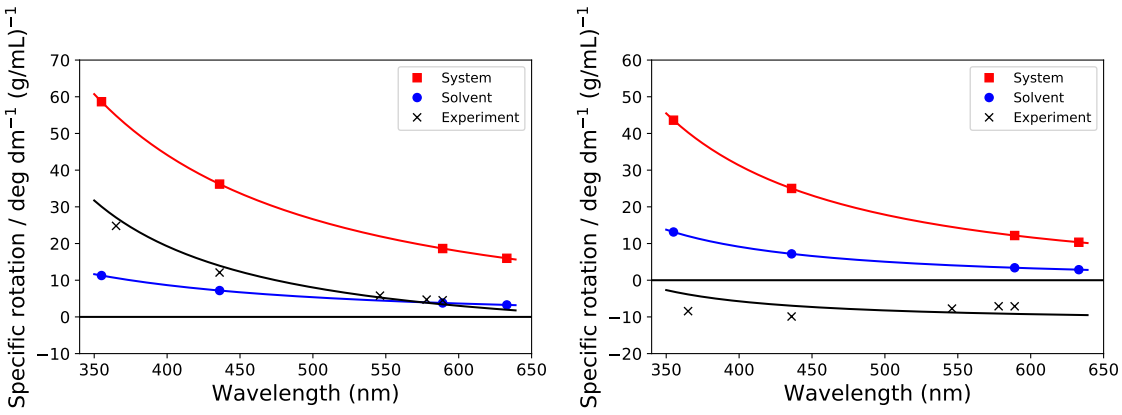


Figure 3.12: ORD spectra of *(S)*-methyloxirane in (a) water and (b) methanol over 250 evenly-spaced snapshots of a 5 ns trajectory. Optical rotation calculations were carried out with 10 explicit solvent molecules using the CAM-B3LYP functional and the ORP basis set.

at each step of the MD simulation, we find that a set of 250 snapshots taken from a 5 ns MD trajectory and a cutoff of 10-12 explicit solvent molecules yields sufficient convergence of the OR of each test system, with the exception of (*S*)-methyloxirane in CCl_4. This can be explained by the presence of Rydberg-type excited states occurring at longer wavelengths for (*S*)-methyloxirane in CCl_4 than in the other solvents, leading to greater sensitivity of the OR to environmental effects.

Other parameters potentially affecting the converged rotations include the choice of density functional, the basis set, and the use of implicit solvent models. The B3LYP and CAM-B3LYP functionals differed significantly in their prediction of OR due to their concomitant variation in the prediction of the excitation wavelength of the lowest lying excited states,

particularly for the case of (*R*)-methylthiirane in water. We found that B3LYP significantly underestimated the excitation energies, leading to large-magnitude OR values and poor convergence of the snapshot averaged. CAM-B3LYP, on the other hand, shifted these states to higher energies and away from the wavelength of the external plane-polarized fields, yielding much better behaved convergence of the average OR. The ORP basis set provided only a small improvement in the predicted OR for each test case as compared to the aug-cc-pVDZ basis, in spite of the fact that it has been optimized for such response properties and contains approximately 50% more basis functions. Finally, the addition of an implicit solvent outside of the explicit solvent shell via the polarizable continuum model had little effect on the averaged OR values.

In spite of the extensive computational effort invested in the above protocol and the careful convergence of the methodology with respect to each degree of freedom, the comparison with experimental solution-phase data remains inadequate. For (*S*)-methyloxirane in water, the comparison to experimental measurements was reasonably good, with the CAM-B3LYP/ORP averages providing the correct sign and shape of the dispersion curve across all wavelengths, though overestimating the quantitative results by approximately a factor of two. For (*S*)-methyloxirane in less polar CH_3OH and non-polar CCl_4 solvents, however, we were unable to reproduce the correct sign of the experimental OR, though the shift relative to the isolated solvent was correct. Using the largest solvent shell, CAM-B3LYP and the ORP basis set shifted the gas-phase ORD spectrum closer to the experimental solvated value for all systems.

There are several possible reasons for the incomplete comparison with experiment, first and foremost of which is the use of TD-DFT as the primary quantum chemical model. While DFT is convenient due to its relatively low computational cost, previous demonstrations of its occasional failure for sensitive properties such as OR lead to reduced confidence in its predictions, though dispersion-corrected functionals may yield improvements. Wavefunction-based methods such as coupled cluster theory may ultimately be a better choice for the QM model, but their higher cost precludes their use at present. It is also possible that the choice of MD force field may play a role, particularly given the partial success of the water simulations, which used the SPC/E water model vs. the methanol and CCl_4 simulations, which used OPLS/AA. The slightly poorer convergence of the averaging with the number of solvent molecules for CCl_4 also suggests that a more extensive solvent shell is needed for larger solvents.

Chapter 4

PNO++: Perturbed Pair Natural Orbitals for Coupled Cluster Linear Response Theory

4.1 Introduction

For molecules and materials subjected to weak electromagnetic fields, key molecular properties may be defined via time-dependent perturbation theory – response theory – by orders of their component multipole operators.[35,43,129] Beyond "zeroth order" properties such as dipole or magnetic moments, linear, quadratic, and higher-order dynamic response functions provide computational routes to frequency-dependent polarizabilities/magnetizabilities, hyperpolarizabilities, and other mixed electric-/magnetic-field properties.[37,130–132] The ultimate accuracy of such computations depends critically on the choice of quantum chemical model.

Among the simplest such models is time-dependent Hartree-Fock (TDHF) (also known as the random-phase approximation (RPA)),[133,134] whose modest computational cost is offset by its lack of treatment of dynamic electron correlation effects, leading to, for example, its overestimation of electronic excitation energies. More sophisticated methods that include at least some approximation of electron correlation can provide significantly improved predictions of molecular response properties, including time-dependent density-functional theory (TD-DFT),[135,136] the second-order polarization propagator (SOPPA),[137,138] the algebraic diagrammatic construction (ADC),[139–141] and more. One of the most robust quantum chemical models is coupled cluster (CC) theory,[16,17,19,142] which relies on an exponential expansion of the multideterminant wave function not only to recover electron correlation effects systematically, but also to ensure the correct (intensive or extensive) scaling of response properties with respect to the size of the molecular system.

The principal disadvantage of coupled cluster theory lies in its high-order polynomial scaling with system size, due to the combination of delocalized canonical molecular orbitals and electron correlation effects which are local in nature.[47] Efforts to reduce the cost of coupled cluster calculations have resulted in the development of several reduced-scaling methods based on orbital localization,[46,48,68,143,144] tensor decompositions,[145–150] and fragmentation schemes,[51,151–153] among others. The aim of orbital localization schemes in particular is to utilize the intrinsic sparsity in correlated wave functions in order to reduce the scaling of ground- and excited-state property calculations. Local correlation methods such as projected atomic orbitals (PAOs) introduced by Pulay and Saebo[46,47] and extended to coupled cluster theory by Hampel and Werner,[143] pair natural orbitals (PNOs) first pioneered by Meyer,[67] Edmiston and Krauss,[66,154] as well as Ahlrichs and coworkers,[53] and resurrected by Neese and coworkers,[68,69] and orbital-specific virtuals (OSVs) created more recently by Yang *et al.*[48] introduce controllable sparsity into the wave function via orbital localization and have been successfully applied to the computation of properties related to ground-state energies without significant loss of accuracy. Though a number of local correlation schemes have been introduced for properties such as excitation energies,[55,56,58,62,155–157] due to the complexity of calculations of higher-order response properties, the application of localization methods to such properties has been limited so far.[158–163]

In a first application of the PNO and OSV approaches to CCSD linear response properties, McAlexander and Crawford compared the performance of three local correlation approaches (PAO, PNO, and OSV) in the calculation of response properties,[8] and found that the system

showed reduced sparsity in the wave function when perturbed by an external field. This resulted in increased sensitivity of the response properties to truncation of the wave function, particularly the mixed electric dipole-magnetic dipole property of specific rotation. Kumar and Crawford further explored the effects of truncation of a virtual natural orbital (NO) space on linear response properties,[164] and concluded that the virtual NO space did not preserve accuracy in those properties following aggressive truncation. It was seen that the truncation of NOs contributing little to electron correlation effects corresponded to the truncation of those that were most diffuse, and therefore essential for the description of field-response properties. A natural alternative is to build an orbital space optimized for those properties sought. Transition-specific virtual spaces have been built by utilizing correlated natural transition orbitals, first introduced by Høyvik, Myhre, and Koch for use with the multilevel coupled cluster scheme,[165] and later used by Baudin and Kristensen in combination with their local framework to reduce the scaling of excitation energy calculations.[57,59] Effective natural transition orbitals have also been constructed by Höfener and Klopper by utilizing both ground and excited state densities.[60] Mester, Nagy and Kállay have combined MP2 and CIS(D) densities to produce state-specific natural orbitals and natural auxiliary functions for the computation of excitation energies.[166,167]

In this work, we examine the PNO++ approach, which incorporates the external field perturbation into the density and thus the orbital space, and uses this perturbed basis throughout the calculation. We create a union of the PNO and PNO++ bases, called the "combined PNO++" basis, in order to accurately describe both the ground and excited state wave

functions. We further explore the PNO++ and combined PNO++ approaches, examining the sparsity introduced by the use of the PNO++ basis, followed by analysis of the truncation error for both correlation energies and second-order response properties, i.e. dipole polarizabilties and specific rotations. We also examine the effect of an MP2-level correction for response properties, analogous to the MP2-level external space correction commonly used in PNO-related methods.

4.2 Theory and Computational Details

4.2.1 Coupled Cluster Linear Response

Coupled cluster response theory is an accurate way to simulate the response of a molecular system to an external electromagnetic field, with the time-dependent field treated as a perturbation.[27,28,36,131,132] The coupled cluster linear response function is the first-order term in the perturbative expansion of the time-independent operator $\hat{A}$ in response to the external field $\hat{B}$,[35] and can be expressed in terms of perturbed wavefunction operators, dependent on the frequency ω of the external field:

$$\langle\langle \hat{A}; \hat{B} \rangle\rangle_\omega = \frac{1}{2}\hat{P}(\hat{A}(-\omega), \hat{B}(+\omega))\Big[\langle 0|(1+\hat{\Lambda})\Big[\bar{A}, \hat{X}^B_\omega\Big]|0\rangle + \langle 0|\hat{Y}^B_\omega \bar{A}|0\rangle\Big] \tag{4.1}$$

where $\hat{P}$ is a permutation operator that ensures that the response function is real, $\bar{A}$ is the similarity-transformed operator, and $\hat{\Lambda}$ are the left-hand coupled cluster amplitudes. The frequency-dependent perturbed operators $\hat{X}^B_\omega$ and $\hat{Y}^B_\omega$ are obtained using sets of perturbed

wavefunction equations for each of the Cartesian directions:

$$\langle\mu|(\bar{H}-\omega I)\hat{X}^B_\omega|0\rangle = -\langle\mu|\bar{B}|0\rangle \tag{4.2}$$

$$\langle\mu|\hat{Y}^B_\omega(\bar{H}+\omega I)|0\rangle = -\langle 0|(1+\Lambda)\big[\bar{B},\tau\big]|\mu\rangle - \langle 0|(1+\Lambda)\big[\big[\bar{H},\tau\big],\tau\big]\hat{X}^B_\omega|\mu\rangle\,. \tag{4.3}$$

Second-order properties, such as dynamic polarizabilities and optical rotations, are then obtained by computing molecular property tensors as the appropriate linear response functions. The polarizability tensor $\boldsymbol{\alpha}(\omega)$ is computed using the electric dipole moment operator $\hat{\mu}$ and the Rosenfeld tensor $\boldsymbol{\beta}(\omega)$ is computed using both the electric and magnetic dipole moment ($\hat{m}$) operators:[31,123,129]

$$\boldsymbol{\alpha}(\omega) = -\langle\langle\hat{\mu};\hat{\mu}\rangle\rangle_\omega \tag{4.4}$$

$$\boldsymbol{\beta}(\omega) = \mathrm{Im}\langle\langle\hat{\mu};\hat{m}\rangle\rangle_\omega. \tag{4.5}$$

Coupled cluster linear response involves the solving of first the coupled cluster amplitude and lambda equations, followed by the solving of twelve sets of perturbed amplitude equations for each wavelength of interest. The main drawback of the coupled cluster method is the high-degree polynomial scaling with the system size. Coupled cluster singles and doubles (CCSD), used in this work, scales as $\mathcal{O}(N^6)$, where N is a measure of the system size. This necessitates some form of reduced-scaling method to make coupled cluster viable for larger molecules.

4.2.2 Local Pair Natural Orbitals

The pair natural orbital (PNO) approach, resurrected by Neese et al. as local pair natural orbitals (LPNO),[68,69] is a reduced-scaling approach that relies upon the creation of a more compact virtual-orbital space using the pair density. In the LPNO-CCSD approach, the pair density $\mathbf{D}^{ij}$ for a given pair of localized occupied orbitals ij, is defined for second-order Møller–Plesset perturbation theory (MP2)[168,169] amplitudes $\mathbf{T}^{ij}$ as:

$$\mathbf{D}^{ij} = \frac{2}{1+\delta_{ij}}\Big(\mathbf{T}^{ij}\widetilde{\mathbf{T}}^{ij\dagger} + \mathbf{T}^{ij\dagger}\widetilde{\mathbf{T}}^{ij}\Big) \tag{4.6}$$

with

$$\widetilde{\mathbf{T}}^{ij} = 2\mathbf{T}^{ij} - \mathbf{T}^{ij\dagger}. \tag{4.7}$$

The PNO basis $\mathbf{Q}^{ij}$ is obtained by diagonalizing this density:

$$\mathbf{D}^{ij}\mathbf{Q}^{ij} = \mathbf{Q}^{ij}\mathbf{n}^{ij}. \tag{4.8}$$

A truncation is carried out by comparing the set of occupation numbers $\{\mathbf{n}^{ij}\}$ to a predetermined threshold T_{cutPNO} and discarding any virtual orbitals with occupation numbers below the threshold. A further truncation is applied by discarding weakly-interacting occupied pairs using a criterion based on the semicanonical MP2 pair correlation energy for each pair, defined as:

$$\epsilon_{ij} = \sum_{ijab} t_{ij}^{ab}\langle ij\|ab\rangle \tag{4.9}$$

with localized occupied pairs ij and canonical virtual orbitals ab. The occupied pairs ij are sorted into weak and strong pairs based on if ϵ_{ij} is lesser or greater than another fixed

threshold $T_{cutPairs}$ and weak pairs are then neglected in the calculation. The occupied space is localized using one of several mathematical criteria, which form the basis of the Boys-Foster,[63] Pipek-Mezey,[65] Edmiston-Ruedenberg,[64] Magnasco-Perico,[170] and von Niessen[171] localization schemes. This localization of the occupied space is important in the truncation of weak pairs, based on the logic that two-electron integrals $\langle ij||ab\rangle$ remain small if, given that the occupied orbitals i and j are spatially distant, the overlap between the occupied-virtual pairs ia and jb is small.

The PNO basis $\mathbf{Q}^{ij}$ is much more compact in terms of contributions to the correlation energy, and thus can be truncated to *e.g.* 1/10th of the total virtual space while preserving 99% of the correlation energy (*vide infra*). This aggressive truncation is an essential part of the reduced scaling of the method, since a distinct set of virtual orbitals is created for every pair of occupied orbitals, and thus a large number of integrals must be constructed and stored, one set for each occupied pair. The virtual orbitals for a given pair are orthonormal within a set, however, the sets of virtual orbitals are non-orthogonal between pairs, and an overlap matrix $\mathbf{S}^{ij,kl}$ between the pairs ij and kl must be constructed and inserted into the regular coupled cluster equations.

Since its reappearance, the LPNO method has been extended to include open-shell cases[172] as well as explicit correlation,[173] parallelization,[174] and other optimizations[175,176]. The PNO method has also been applied to the calculation of excitation energies, ionization potentials and electron affinities.[61,155–157,177–179] An application of pair natural orbitals to linear response theory by Frank and Hattig[62] led to state-specific PNOs being used to obtain excitation

energies as the eigenvalues of the Jacobian. The optimized PNO method has also been extended to obtain analytic energy derivatives[180] leading to first-order molecular properties. McAlexander and Crawford[8] published the first application of the PNO method to second-order properties, namely dynamic polarizabilities and specific rotations, however, the three reduced-scaling methods compared in that work showed oscillations in the value of the CCSD-level optical rotation with the truncation of the virtual space, and hence could not directly be used in the same way to compute these more sensitive properties.

4.2.3 Perturbation-aware Densities

We see from the form of the coupled cluster linear response function (Eq. (4.1)) that the final value of the response property sought is dependent on not only the ground state of the system as defined in the coupled cluster ansatz but also on the external perturbation via the perturbed wavefunction amplitudes. Thus, in order to find a more effective compact basis in which to compute molecular properties, we modify the definition of the PNO basis to include the effect of the perturbation on the system.

As seen from previous work done in our group,[181] the PNO++ pair density is formulated in an analogous fashion to the PNO pair density (Eq. (4.6)), with the first-order perturbed amplitudes $\mathbf{X}_B^{ij}$, corresponding to the perturbed wavefunction operators $\hat{X}_\omega^B$ seen above, replacing the ground state amplitudes $\mathbf{T}^{ij}$:

$$\mathbf{D}^{ij}(B,\omega) = \frac{2}{1+\delta_{ij}}\Big(\mathbf{X}_B^{ij}\widetilde{\mathbf{X}}_B^{ij\dagger} + \mathbf{X}_B^{ij\dagger}\widetilde{\mathbf{X}}_B^{ij}\Big) \tag{4.10}$$

with $\widetilde{\mathbf{X}}_B^{ij}$ defined similarly to Eq. (4.7) above.

In order to create a density from MP2-level calculations, first-order perturbed guess amplitudes must be created using a perturbation $\bar{B}$ and Hamiltonian matrix elements $\bar{H}_{pp}$, both similarity-transformed:

$$X_{ab}^{ij} = \frac{\bar{B}}{\bar{H}_{aa} + \bar{H}_{bb} - \bar{H}_{ii} - \bar{H}_{jj} + \omega}. \tag{4.11}$$

This is the usual form of the guess amplitudes in the coupled cluster linear response equations above (Eq. (4.2)). The Hamiltonian matrix elements are given in Eq. (4.12) in terms of Fock matrix elements f_{pq}, MP2 amplitudes t_{ab}^{ij} and antisymmetrized two-electron integrals $\langle ij \| ab \rangle$.

$$\bar{H}_{ii} = f_{ii} + \frac{1}{2} t_{ef}^{in} \langle in \| ef \rangle \tag{4.12a}$$

$$\bar{H}_{aa} = f_{aa} - \frac{1}{2} t_{fa}^{mn} \langle mn || fa \rangle \tag{4.12b}$$

Similarity-transformed Hamiltonian matrix elements are found by applying the Baker-Campbell-Hausdorff expansion to the second-quantized Hamiltonian[182] and setting the terms in the final expression which contain t_a^i to zero. Once the perturbation-aware pair density $\mathbf{D}^{ij}(B, \omega)$ is created, it can be used to create the PNO++ basis in an identical way to the PNO basis, and thus a similar approach is used to reduce the scaling of the coupled cluster linear response equations.

4.2.4 MP2-level correction

The MP2-level perturbative correction added to the LPNO-CCSD correlation energy, as described in Ref. 68, constitutes a substantial improvement to the value of the correlation energy at larger truncations. The correction is defined as the difference between the MP2-level correlation energy of the full, untruncated space and the correlation energy of the truncated space:

$$\Delta_{MP2} = E_{MP2}^{full} - E_{MP2}^{PNO}. \tag{4.13}$$

This correction is then added to the LPNO-CCSD correlation energy to obtain corrected energy values. In order to improve the PNO++-CCSD property values at large truncations in a similar way, an MP2 correction to the PNO++ polarizability and specific rotation values is defined as:

$$\Delta\alpha_{MP2} = \alpha_{MP2}^{full} - \alpha_{MP2}^{PNO++} \tag{4.14}$$

$$\Delta\beta_{MP2} = \beta_{MP2}^{full} - \beta_{MP2}^{PNO++}. \tag{4.15}$$

However, due to the absence of an MP2-level linear response analogue to the MP2 correlation energy, the MP2-level properties are defined as the linear response function containing MP2-level amplitudes, as well as perturbed X_ω and Y_ω amplitudes derived from these by using Eq. (4.12).

The linear response function contains a frequency-dependence that is incorporated via the dependence of the perturbed X_ω and Y_ω amplitudes on the frequency of the external field. When computing the MP2-level correction in the modified velocity gauge, we select the

zero-frequency correction, since the difference in the frequency-dependent and zero-frequency corrections is minimal.

4.2.5 Computational Details

In order to check the accuracy of our method without a full production-level implementation involving the creation and storage of PNO-basis integrals and overlap matrices, a local filter approach was used, as used previously for PAO calculations,[143] and described in Ref. 8. To begin, the occupied space was localized using the Pipek-Mezey localization criterion,[65] due to its relative efficiency and accuracy. These orbitals were then used to run a canonical CCSD linear response calculation, with a simulation code applied at each iteration. The simulation code transforms the residuals $\mathbf{R}^{ij}$ obtained at each iteration of the coupled cluster amplitude equations to first, the PNO basis, obtained as $\mathbf{Q}^{ij}$ in Eq. (4.8), and then to the semi-canonical basis in which the virtual block of the Fock matrix is diagonal:

$$\bar{\mathbf{R}}^{ij} = \mathbf{L}^{ij\dagger}\mathbf{Q}^{ij\dagger}\mathbf{R}^{ij}\mathbf{Q}^{ij}\mathbf{L}^{ij} \tag{4.16}$$

where $\mathbf{L}^{ij}$ is the result of the diagonalization of the virtual block of the Fock matrix in the PNO basis, $\mathbf{F}^{ij}$:

$$\mathbf{F}^{ij}\mathbf{L}^{ij} = \mathbf{L}^{ij}\bar{\epsilon}^{ij}. \tag{4.17}$$

The denominator consisting of diagonal Fock matrix elements f_{ii} and semicanonical orbital energies $\bar{\epsilon}_a$ is then applied to obtain the increment in the local basis, which is then converted

back into the canonical MO basis.

$$\bar{\Delta}_{ab}^{ij} = \frac{\bar{R}_{ab}^{ij}}{f_{ii} + f_{jj} - \bar{\epsilon}_a - \bar{\epsilon}_b} \tag{4.18}$$

$$\mathbf{\Delta}^{ij} = \mathbf{Q}^{ij}\mathbf{L}^{ij}\overline{\mathbf{\Delta}}^{ij}\mathbf{L}^{ij\dagger}\mathbf{Q}^{ij\dagger} \tag{4.19}$$

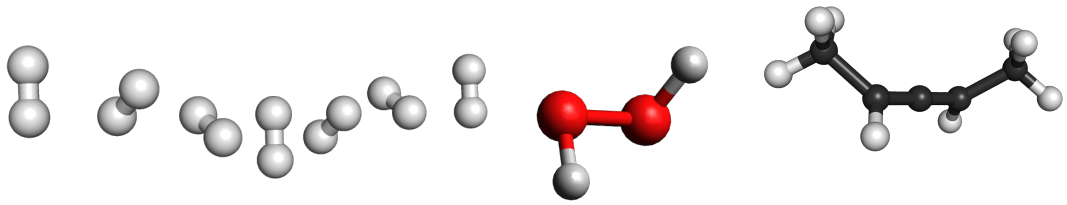
Thus, the final increment, containing no nonlocal contributions, is added to the previous iteration's coupled cluster amplitudes and the loss in accuracy in the converged energy values can be obtained with minimal alteration to the existing coupled cluster response code. A similar procedure is applied to the $\hat{\Lambda}$ and the perturbed wavefunction equations, in order to obtain the loss in accuracy in the converged property values.

The PNO and PNO++ pair densities were created using the forms from Eqs. (4.6) and (4.10) above. For the PNO++ density, since the perturbed wavefunction depends on the direction of the external field relative to the molecule, a separate density was created for each of the Cartesian directions. The densities were then averaged to obtain a single density for natural orbitals. This density was then diagonalized to obtain the PNO++ basis, which was then truncated using a threshold $T_{cutPNO++}$. To create a combined space, the PNO and PNO++ spaces were both created by individually diagonalizing the PNO and PNO++ pair density matrices, followed by truncation using two separate thresholds, T_{cutPNO} and $T_{cutPNO++}$. The spaces were then concatenated as in Eq. (4.20) and a QR decomposition was used to orthogonalize the space that remained.

$$\mathbf{Q}_{combined}^{ij} = [\mathbf{Q}_{PNO}^{ij} \quad \mathbf{Q}_{PNO++}^{ij}] \tag{4.20}$$

A few test systems were used in order to clearly see the convergence of the method for both

optimally localized as well as general molecular systems. $(H_2)_n$ systems, with $n = 4, 5, 6, 7$ were used as a best-case scenario for the MO localization, as well as having large values of optical rotation in the arrangement chosen [Fig. 4.1]. The $(H_2)_n$ systems were also organized to lie along the Y-axis, in order to maximize the directionality of the perturbed wavefunction. As more general small-molecule systems, H_2O_2 and (*P*)-1,3-dimethylallene (DMA) were used. The aug-cc-pVDZ basis set was used for the correlation energy, dynamic polarizability and specific rotation calculations.[183] Specific rotations and dynamic polarizabilities were computed at the sodium D-line wavelength, 589 nm.



(a) (b) (c)

Figure 4.1: Systems used in this work. (a) an $(H_2)_7$ helix (b) H_2O_2 and (c) (*P*)-1,3-dimethylallene (DMA).

(a)

(b)

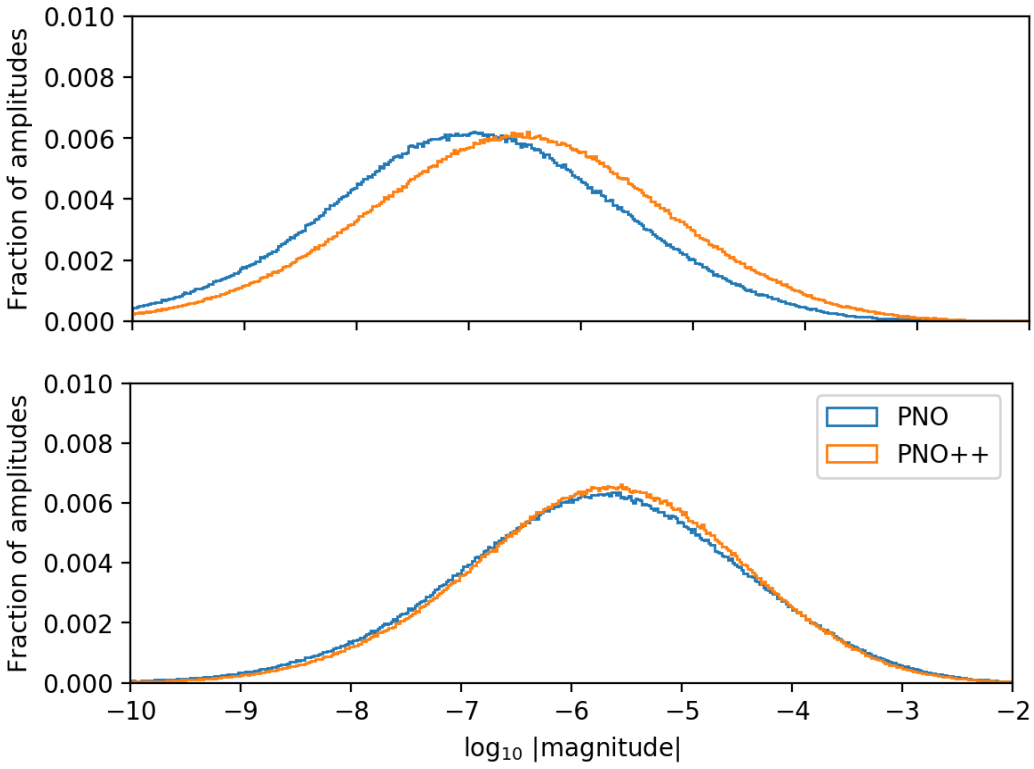


Figure 4.2: Fractions of (a) unperturbed and (b) perturbed CCSD wave function amplitudes for the PNO space (blue) and the PNO++ space (orange) as a function of their magnitude for 1-fluoro-heptane using the 6-31G basis set.

4.3 Results and Discussion

4.3.1 Amplitude Sparsity

Figure 4.2 shows the distribution of the converged CCSD T_2 amplitudes for a linear system, 1-fluoro-heptane, as well as the converged CCSD perturbed wavefunction amplitudes X_2^μ that contribute to the linear response function. The distribution was created by localizing the occupied space using the Pipek-Mezey localization scheme, followed by localization of the virtual space into the PNO and PNO++ spaces. These spaces were not truncated in order to analyze the sparsity introduced by each method. We may use the magnitude distribution of wave function amplitudes as a measure of the sparsity introduced by the localization of the virtual space. Such sparsity in the wave function should, in principle, allow a more aggressive truncation threshold without a commensurate loss in accuracy. However, we note that our system of truncation using occupation numbers does not necessarily imply removal of amplitudes of small magnitudes, but instead, amplitudes of any magnitude may be removed. Amplitudes with values below 10^{-10}, the coupled cluster convergence threshold, are not shown, in order to focus on the most significant values.

The sparsity introduced by the PNO approach was contrasted by McAlexander and Crawford[8] with the projected atomic orbital[46,47,143] (PAO) and orbital specific virtual[48] (OSV) approaches and it was concluded that the PNO distribution contained the most sparsity, especially for the unperturbed amplitudes. Fig. 4.2a shows that the PNO++ method conserves much of this sparsity, showing an unperturbed peak within an order of magnitude of

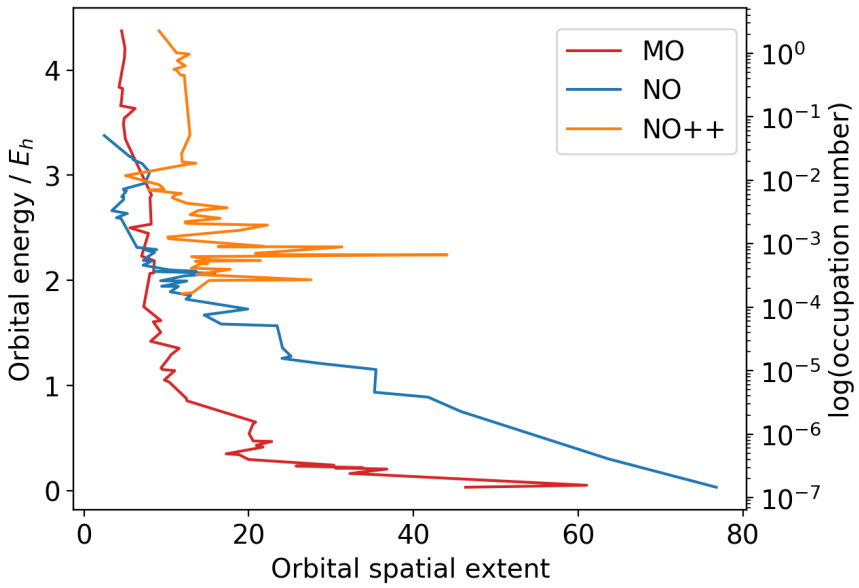


Figure 4.3: Orbital energies and occupation numbers plotted against orbital spatial extents for the canonical MO, PNO and PNO++ spaces for the H_2O_2 molecule using the 6-31G basis set.

the peak for the PNO method. The perturbed amplitudes that contribute to the response equation have a right-shifted distribution with larger magnitudes compared to the unperturbed amplitudes, as seen in Fig. 4.2b. The PNO and PNO++ spaces have very similar distributions of perturbed amplitudes, and thus could be expected to show similar truncation errors; however, as seen later, the PNO++ method allows us to truncate the space more aggressively than the PNO method.

4.3.2 Orbital extent

Accurate calculations of response properties require more diffuse basis functions[183], thus in order to obtain an idea of the relative diffuseness of the orbital spaces computed using the natural orbital (NO) and perturbed natural orbital (NO++) methods, we look at the spatial extent of the virtual orbitals, defined as $\langle r^2 \rangle$, in Figure 4.3. Plotted against them are virtual orbital energies in the case of canonical molecular orbitals (MOs) and occupation numbers on the log scale in the case of natural orbitals and perturbed natural orbitals. The natural orbitals and perturbed natural orbitals are created by diagonalizing MP2-level densities given by:

$$D_{ab} = \sum_{ijc} 2t_{bc}^{ij}\big(2t_{ac}^{ij} - t_{ca}^{ij}\big) \tag{4.21}$$

and

$$D_{ab}(B,\omega) = \sum_{ijc} 2x_{bc}^{ij}\big(2x_{ac}^{ij} - x_{ca}^{ij}\big) + \sum_{i} x_i^a x_i^b \tag{4.22}$$

$$x_{ab}^{ij} = \frac{\bar{B}}{\bar{H}_{aa} + \bar{H}_{bb} - \bar{H}_{ii} - \bar{H}_{jj} + \omega} \tag{4.23}$$

$$x_{ab}^{ij} = \frac{\bar{B}}{\bar{H}_{aa} - \bar{H}_{ii} + \omega}. \tag{4.24}$$

The similarity-transformed $\bar{H}$ and $\bar{B}$ elements are formed as described in Section 4.2.3 above.

Overall, we see that the canonical MOs follow a trend of the orbitals with smaller orbital energies having the largest spatial extent. Comparing the NO space to this, we see that natural orbitals follow a similar trend, with the orbitals with the smallest occupation numbers being the most diffuse. In the usual truncation scheme, this corresponds to the most diffuse

orbitals being the first ones truncated. This can explain the requirement of a large number of orbitals kept in order to accurately compute response properties. On the other hand, while the perturbed natural orbitals have larger magnitudes of occupation numbers, their extent lies between 5-25, with just one outlier. We see that these orbitals continue to be at the same diffuseness even at the larger occupation numbers that are kept as the space is truncated. This points to a more even distribution of the orbital extents with occupation numbers, and can explain the method's improved accuracy as seen in Sections 4.3.2 and 4.3.3 below.

4.3.3 PNO++ results

Correlation energies

CCSD correlation energies were computed for all systems using the simulation code described in Section 4.2.5. In Figure 4.4, errors in correlation energy values are plotted against the T_2 ratio, defined as the ratio of the number of wave function amplitudes present in the truncated space to the total number of wave function amplitudes in the full, untruncated space. The errors in correlation energy were computed using: $\Delta E = E_{PNO/PNO++} - E_{CCSD}$. The T_2 ratio is used as a measure of the cost of a production-level calculation, since contractions over doubles amplitudes and two-electron integrals are the most expensive portions of a CCSD linear response calculation. Although a truncation threshold was used to remove orbitals from the calculation, an analysis using the T_2 ratio removes any ambiguity caused by the differing sparsity of the PNO and PNO++ spaces and provides a clearer picture of the

computational savings obtained. The relationship between the T_2 ratio and the truncation threshold used for each method is shown in the supporting information.

Figure 4.4 shows truncation errors for the smallest and largest $(H_2)_n$ systems considered, as well as H_2O_2 and DMA. The PNO++ method has a more sharply-rising error with truncation as compared to the PNO method, which is seen most clearly for the larger molecules of H_2O_2 and DMA. The PNO density is created using T_2 amplitudes, while the PNO++ density is created using X_2^B amplitudes, and thus is not optimized for sparsity in the T_2 amplitudes. While the PNO++ method is optimized for response properties, it is desirable for the method to also describe the ground state wave function in an accurate way. One method of recapturing accuracy in the correlation energy involves including both the PNO and the PNO++ spaces in the calculation, in a combined PNO++ method as described in Section 4.2.3 above.

Figure 4.4 also reports a comparison of the errors in correlation energies for the PNO, PNO++ and combined PNO++ methods as a function of the T_2 ratio. The combined PNO++ method has two distinct truncation thresholds, T_{cutPNO} for the PNO space and $T_{cutPNO++}$ for the PNO++ space, and the variation of both these thresholds affects the accuracy of the method. As seen later, the variation of the perturbed threshold makes a much larger difference than the same variation of the unperturbed threshold, and thus we have selected and fixed a reasonable threshold for the PNO space (T_{cutPNO}) of 10^{-6}, while varying the perturbed threshold only. This combined method introduces only a slight error at the same T_2 ratio or the same fraction of the space as compared to the original PNO

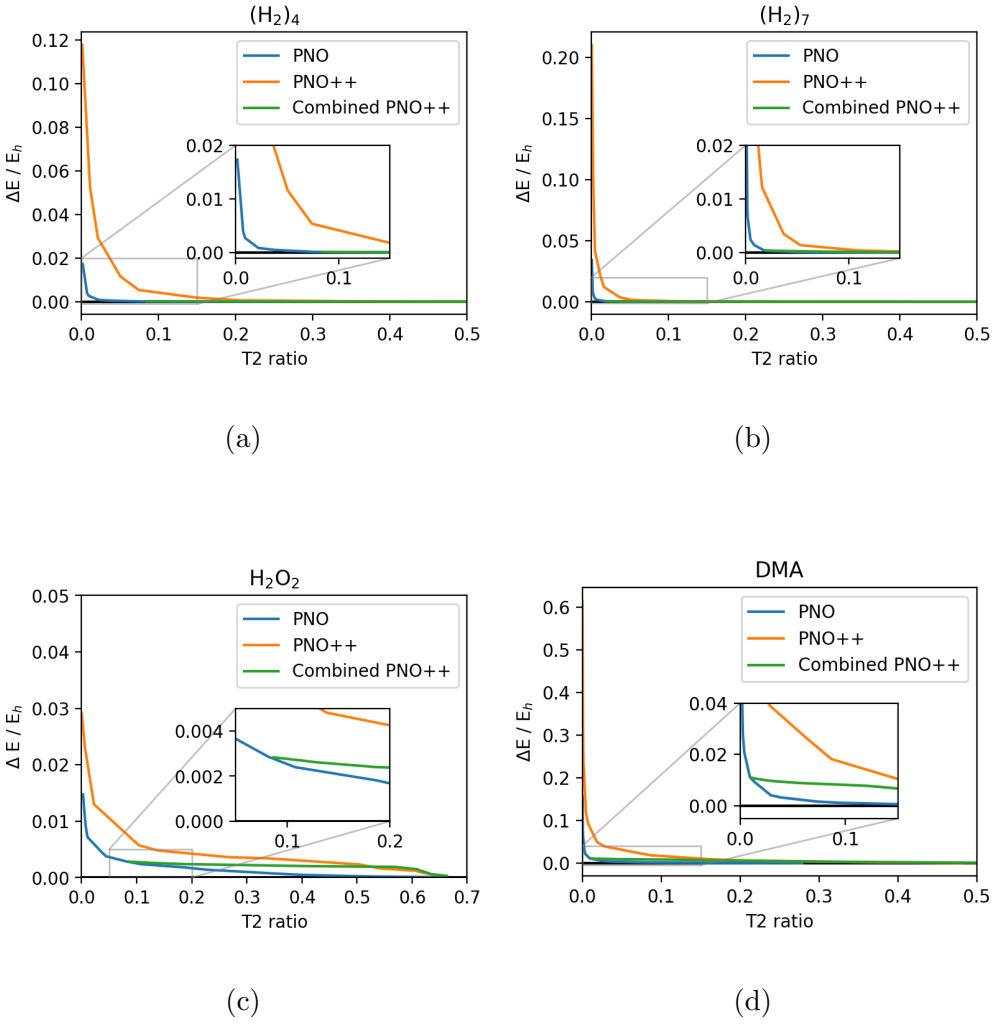


Figure 4.4: Truncation errors in CCSD correlation energy in Hartree for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set.

method. We see larger errors for the combined method for DMA, the largest molecule we examine; however, we do not see a steep increase in the error at large truncations, in contrast to the PNO++ method. While the combined method appears to be simply creating the same space as the PNO approach, the results for the dipole polarizabilities and specific rotations in Sections 4.3.2 and 4.3.3 indicate that this is not the case. The combined space created is optimized for both correlation energies as well as properties.

Dynamic polarizabilities

Figure 4.5 reports dynamic dipole polarizabilities computed at 589 nm using the PNO, PNO++ and combined PNO++ approaches as a function of the T_2 ratio. The PNO++ method shows a significant improvement over the PNO method, obtaining 99% of the CCSD reference polarizability value for our smallest system, $(H_2)_4$ at a low truncation threshold with a T_2 ratio of 0.074. A similar value requires a T_2 ratio above 0.60 when using PNOs. As the size of the $(H_2)_n$ systems increases, we see that the PNO++ method requires smaller fractions of the space to be kept to obtain 99% of the reference value, with the $(H_2)_7$ system requiring only 3.68% of the T_2 amplitudes needed in a conventional CCSD calculation. The data for the H_2O_2 and DMA systems also shows the PNO++ method achieving 99% accuracy at T_2 ratios of 0.61 and 0.087 respectively, while the PNO method required a T_2 ratio of 0.70 for the H_2O_2 system to achieve similar levels of accuracy and did not obtain 99% of the CCSD value at the smallest threshold used.

For all systems and methods considered, monotonic convergence towards the polarizability

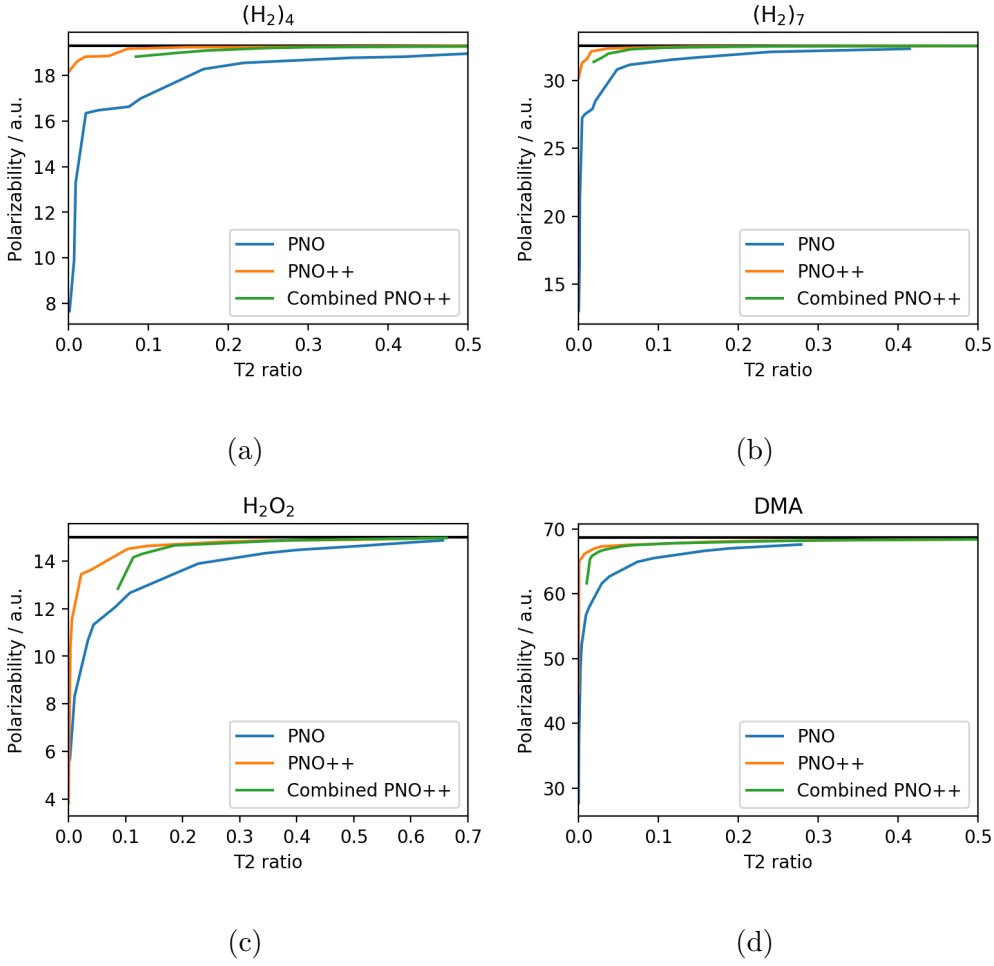


Figure 4.5: CCSD linear response dynamic polarizabilities at 589 nm in a.u. for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set.

value was observed; albeit much more slowly for the PNO method. The combined PNO++ method, in this case, introduces an error relative to the PNO++ method alone, but the largest truncation error is still within 14% of the reference, and is traded off by the improvement seen in the CCSD correlation energy in Section 4.3.1. The T_{cutPNO} thresholds present different T_2 ratios at the same threshold for the different methods, since the densities are distinct and thus different numbers of amplitudes are kept. The combined PNO++ curves begin at a larger T_2 ratio due to the T_{cutPNO} threshold being set to 10^{-6}, with the $T_{cutPNO++}$ varying.

The dynamic polarizability data suggests that the PNO++ density, created using the X_2^B amplitudes, is more efficient than the PNO density at creating a compact space with which to approximate external field-dependent properties. Combining the spaces and orthogonalizing them, as in the combined PNO++ approach, captures most of the efficiency for properties seen in the PNO++ method while also describing the ground state wave function with higher accuracy.

Specific rotations

Figure 4.6 shows specific rotations using the modified-velocity gauge representation of the electric dipole moment operator, computed at 589 nm using the PNO and PNO++ methods. For this and the following graphs, the electric dipole moment operator was chosen as the perturbation in the creation of the perturbed density. The use of the magnetic dipole moment operator in the perturbed density was considered; however, convergence behavior suffered

(see SI for more details). For the $(H_2)_n$ systems, the behavior of the specific rotations is similar to that of the polarizabilities, in that the PNO++ method converges to within 5% of the reference value at a T_2 ratio of 0.68 for $(H_2)_4$ and 0.26 for $(H_2)_7$, while the PNO method does not seem to converge even at the fairly large T_2 ratios examined here. The DMA system also follows a similar trend, with the PNO value unconverged at a T_2 ratio of 0.48, the largest T_2 ratio considered, while the PNO++ value converges to within 5% at a T_2 ratio of 0.16. H_2O_2 exhibits the slowest convergence, with the PNO rotation value showing slow convergence while the PNO++ method oscillates about the reference. However, the PNO++ method does remain within ± 15% of the CCSD value beginning at the fairly low T_2 ratio of 0.34.

Examining Figure 4.6, we see that similar to the polarizability data, the combined PNO++ approach introduces an error relative to the PNO++ approach and begins at a higher T_2 ratio due to the orbitals kept from the PNO space at a T_{cutPNO} threshold of 10^{-6}. Unlike the polarizability, contributions to the specific rotation are not always positive, but can also be negative; this means that the truncation error shows itself as an oscillation instead of approaching the reference value from the same side. This is most clearly seen in Figure 4.6(d), the DMA system. The combined PNO++ method continues to show convergence towards the reference value below a T_2 ratio of 0.5, in contrast to the PNO method for all systems.

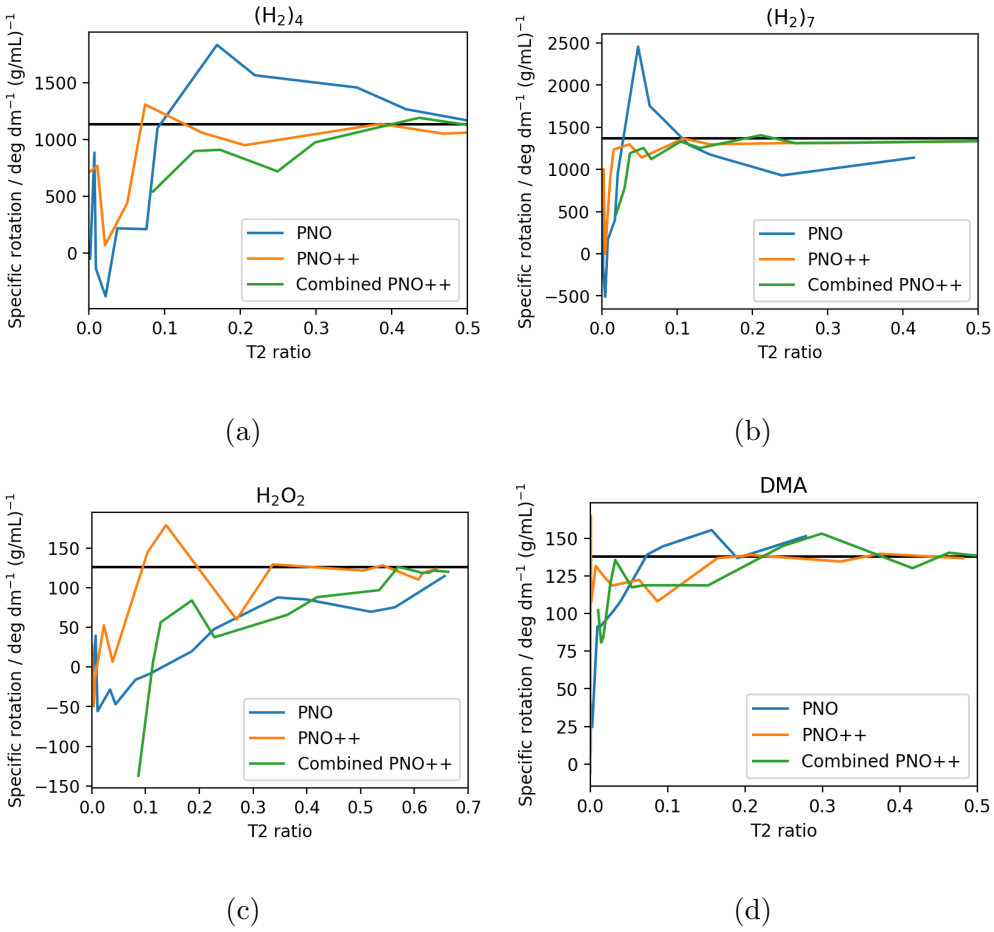


Figure 4.6: CCSD specific rotations in deg dm^{-1} $(gm/mL)^{-1}$ at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set.

4.3.4 The Combined PNO++ Method

To explore the variation in accuracy of the combined PNO++ method with truncation of the PNO and PNO++ spaces we plot the energies and specific rotations keeping one threshold, either T_{cutPNO} or $T_{cutPNO++}$, constant at 10^{-6} while varying the other in Figures 4.7 and 4.8. We compare the results for correlation energies and specific rotations, selecting the best result from the previous section as a standard of comparison for each property. Thus, we compare CCSD-level correlation energies using the PNO method, and dynamic polarizabilities and specific rotations using the PNO++ method.

A natural outcome of combining the spaces is that while fixing the orbitals obtained from one space, the convergence behavior follows that of the other space. For example, when keeping a fixed number of orbitals from the PNO space, the combined PNO++ follows the PNO++ method as we add orbitals from the PNO++ space. By comparing the leftmost point in the energy and rotation graphs we can see that fixing the PNO++ threshold to 10^{-6} versus fixing the PNO threshold to 10^{-6} results in a larger space to begin with, due to larger numbers of orbitals being kept at the same threshold; however, this difference is effectively avoided by plotting against T_2 ratios, and the convergence behavior can be fairly compared.

When comparing the CCSD correlation energies in Fig. 4.7, we see that while neither combined method has the quick convergence of the PNO method, fixing the T_{cutPNO} threshold brings us closer to the reference value than fixing the $T_{cutPNO++}$ threshold at a given truncation. The addition of the PNO++ orbitals does not improve the correlation energy as

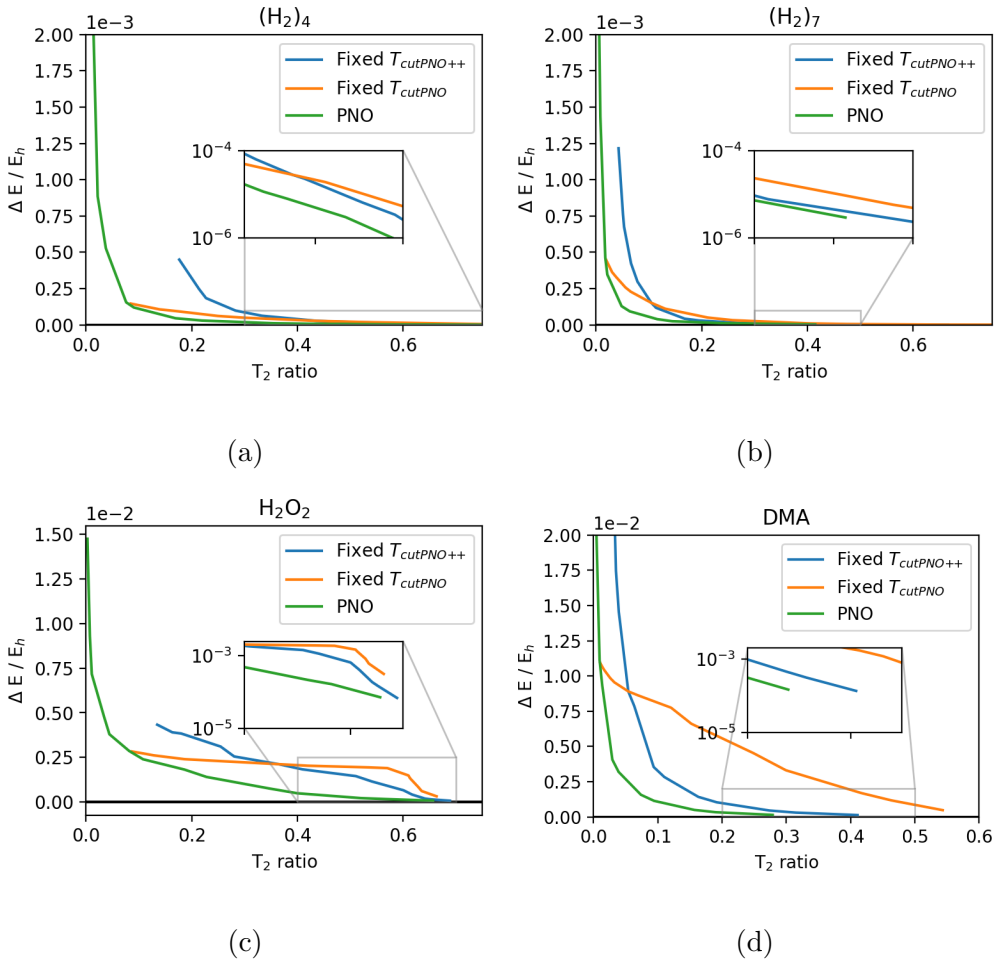


Figure 4.7: Truncation errors in CCSD correlation energy in Hartree for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set for the PNO and PNO++ methods. Inset axes show errors on a logarithmic scale.

efficiently as the addition of the PNO orbitals, causing a crossover point in all systems. Though the error in the energy computed using a fixed T_{cutPNO} is not seen to converge for H_2O_2 and DMA (Fig. 4.7 (c) and (d)) at the largest T_2 ratios considered, the energy will converge to the full CCSD value as either of the thresholds T_{cutPNO} or $T_{cutPNO++}$ are tightened.

Fig. 4.8 shows the specific rotations computed using the same sets of spaces as Fig. 4.7, but compared here to the PNO++ data, which was seen in the previous section to converge to a value close to the reference relatively quickly for all systems. Once again, the combined methods show slower convergence; however, the combined method with the fixed number of unperturbed orbitals shows convergence behavior to within ± 14% of the reference value at a T_2 ratio of 0.4 and higher for all systems. Thus, the convergence behavior of the method with the fixed T_{cutPNO} follows the PNO++ behavior, while the method with the fixed $T_{cutPNO++}$ follows the non-converging PNO behavior seen in Section 4.3.3. Thus, in order to achieve the well-behaved convergence of specific rotation seen with the PNO++ method, it is preferred to fix T_{cutPNO}, in effect adding a fixed number of PNO orbitals to the PNO++ space.

4.3.5 MP2-level corrections

MP2-level corrections were computed for the PNO++ and combined PNO++ methods using the method described in Section 4.2.4. The goal of the MP2-level corrections was to recapture some of the accuracy lost due to the truncation of the PNO++ space, by adding in corrections

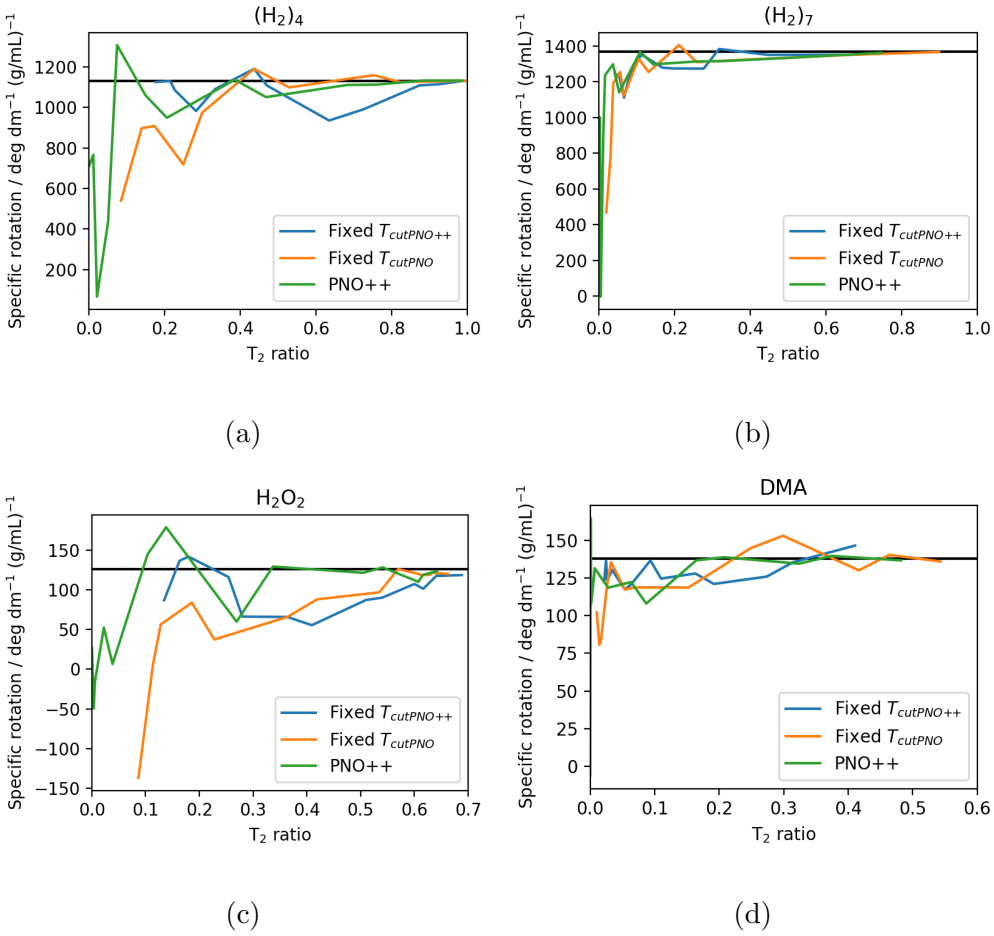


Figure 4.8: CCSD specific rotations in deg dm^{-1} $(gm/mL)^{-1}$ at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set.

computed as in equations (4.14) and (4.15).

Figure 4.9 looks at the effect of corrections on the PNO++ and the combined PNO++ methods for the $(H_2)_7$ and DMA systems. As expected, the CCSD correlation energies are improved by the addition of MP2-level corrections; however, there is some tendency to overcorrect at very small T_2 ratios (Fig. 4.9 (a) and (b)). The polarizability and rotation data in Fig. 4.9 (c), (d), (e) and (f) does not show a substantial improvement in accuracy at a given T_2 ratio on the addition of an MP2-level correction.

Adding corrections computed in this way at the MP2 level does not improve the description of the system's response to an external electromagnetic field. Since one normally computes coupled cluster linear response properties using converged coupled cluster amplitudes, and the corrections are computed using MP2-level amplitudes, it could be argued that simply adding the external field in the form of a perturbation to MP2-level amplitudes does not take into account the sensitive nature of response properties.

4.4 Conclusions

In this work, we have examined a reduced-scaling approach optimized for response properties, computing truncation errors in CCSD correlation energies, dynamic polarizabilities and specific rotations for a few test systems. This approach, dubbed PNO++, conserves the sparsity introduced into the wave function by the PNO approach, while also having a relatively diffuse orbital space, required for the accurate calculation of higher-order response

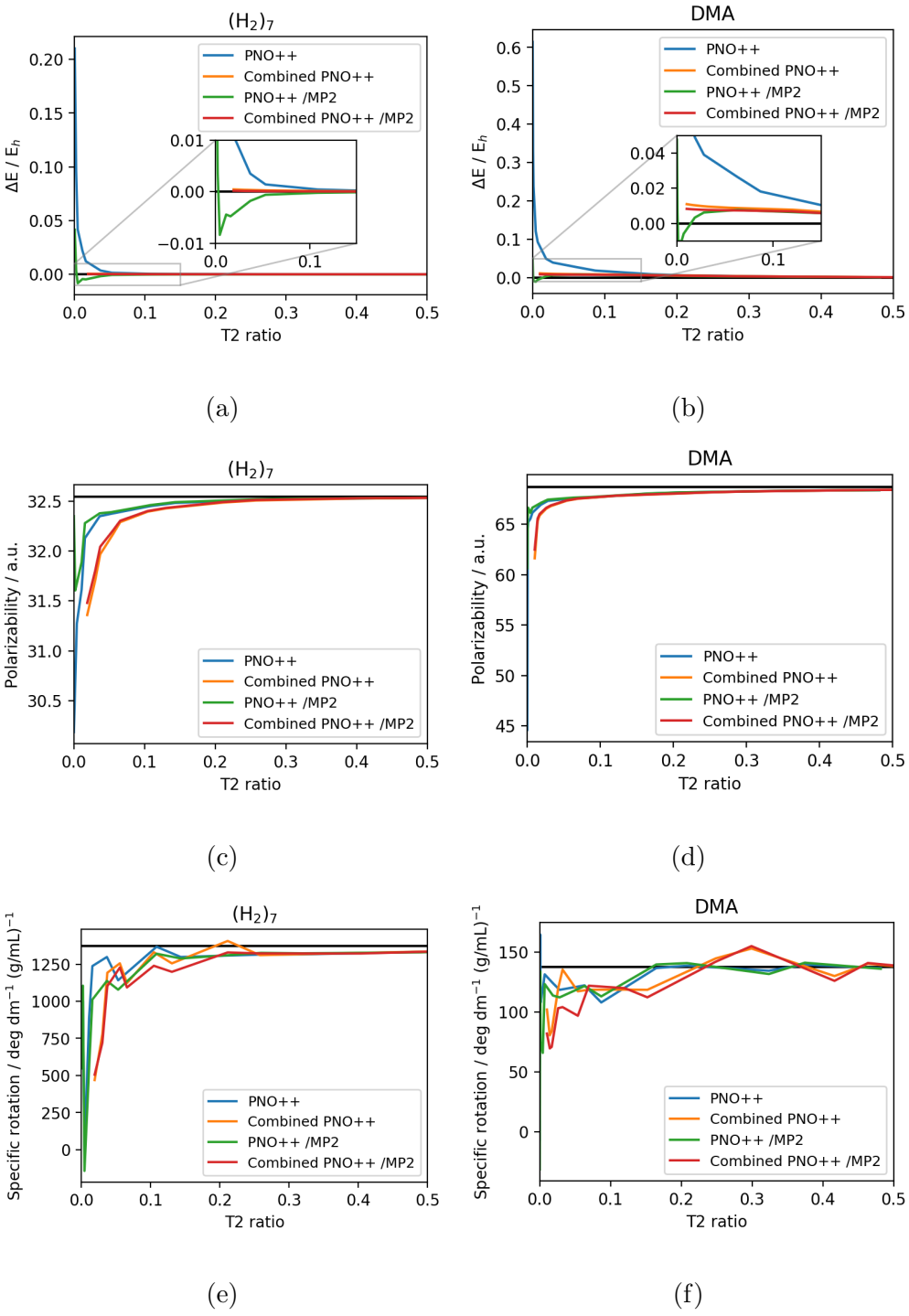


Figure 4.9: CCSD correlation energies, dynamic polarizabilties and specific rotations at 589 nm for $(H_2)_7$ ((a), (c) and (e) respectively) and 1,3-dimethylallene (DMA) ((b), (d) and (f) respectively), computed using the aug-cc-pVDZ basis set.

properties. The PNO++ method is not, however, optimized for correlation energies and thus shows large truncation errors in the CCSD correlation energy at small fractions of the space. We find that incorporating a fixed number of orbitals from the PNO space into the PNO++ method recovers accuracy in the CCSD correlation energy. This "combined PNO++" method allows for the truncation of a large portion of the space, up to a T_2 ratio of 0.5, without sacrificing accuracy in both the response property as well as the CCSD correlation energy. MP2-level corrections, while improving the value of the correlation energy at highly-truncated system sizes, do not improve the quality of the more sensitive dynamic polarizability and optical rotation at similarly-truncated sizes. A production-level implementation which includes screening for weak pairs as well as method-specific optimizations will be required in order to estimate the true computational savings obtained through truncation of the PNO++ space created using this form of perturbation-aware density for the computation of linear response properties.

Chapter 5

Improvements to the PNO++ Method

5.1 Introduction

Coupled cluster theory has become one of the most popular methods used for electronic structure calculations, due to its accuracy and systematic improvability.[19,24] Its accuracy can be seen through its folding in of higher level excitations in an exponential *Ansatz*, and its improvability, with full coupled cluster at N-level excitations providing an exact nonrelativistic solution to the time-independent electronic Schrödinger equation within a finite basis set. However, coupled cluster theory suffers from the disadvantage of steep scaling of the computational cost with system size. Coupled cluster truncated at single

and double excitations (CCSD) scales as N^6, while including triples excitations (CCSDT) increases the scaling to N^8, where N is a measure of the size of the system.[184]

This scaling problem has been tackled by the development of methods including fragmentation approaches,[51,151–153] tensor decompositions,[145–150] and localization schemes.[46,48,68,143,144] The main motivation for localization methods is that the dynamic electron correlation which coupled cluster aims to capture is localized, and thus a localized basis would better represent the sparsity present in the wave function than the usual delocalized molecular orbital (MO) basis. Localization and local correlation schemes such as projected atomic orbitals (PAOs),[46,47,143] pair natural orbitals (PNOs),[53,66–69,154] and orbital specific virtuals (OSVs)[48] have been successfully applied to coupled cluster ground state energies, and several inventive methods have been developed in order to apply them to excited state energies.[58,62,155–157,185] However, their application to response properties has been limited.[158–161]

Russ and Crawford investigated an alternative domain-selection scheme for PAOs that incorporated the perturbation into the domain definition.[162,186] McAlexander and Crawford,[8] in a first application of PNOs to coupled cluster linear response properties, found that due to the reduced sparsity of a field-perturbed wave function, the sensitivity of response properties to truncation of the localized basis was higher. Kumar and Crawford,[164] exploring the effect of truncation of a natural orbital space on linear response properties, saw an inverse correlation between the diffuseness of the orbitals and their occupation number, thus the truncation of low occupation number orbitals led to the truncation of more diffuse orbitals, which are important for response properties. Consideration of the transitions that must be described

well led to the creation of transition-specific natural orbitals by Høyvik, Myhre and Koch,[165] and their use by Baudin and Kristensen in combination with the local framework to reduce the scaling of excitation energy calculations.[57,59] Höfener and Klopper formed effective natural transition orbitals using ground and excited state densities,[60] while Mester, Nagy, and Kállay combined MP2 and CIS(D) densities to produce state-specific natural orbitals for excitation energy calculations.[166,167]

In our group's most recent work, the incorporation of the external perturbation into the orbital space for the PNO method, compact for ground state energies, led to the creation and exploration of the PNO++ approach for coupled cluster response properties.[181,187] The PNO++ method described in Ref 187 has shown encouraging results for small, localizable systems. The combined PNO++ method, including a fixed number of PNOs in the final orbital space, has smaller localization errors in the correlation energy than the PNO++ method as well as smaller localization errors in the response properties studied than the PNO method. However, in order for either method to be implemented at production-level, validation on larger systems than available to a Python-based code is still required.

The sensitive nature of response properties means that a larger proportion of the orbital space needs to be kept to maintain accuracy in the value of the property. But the nature of the PNO method means that aggressive truncation is required since the initial integral transformation into the PNO basis becomes expensive. Thus, a production-level code requires careful consideration as to an optimal method, including the ability to truncate aggressively for larger, more delocalized systems. The simulation code on the other hand is only

marginally slower than a regular canonical coupled cluster code, since it only incorporates a single extra transformation at each iteration. A simulation code implemented in an optimized coupled cluster linear response code, such as the one present in PSI4,[188] is a good testing environment for further approximations as well as improvements to the method for specific properties, as described below.

In this work, we apply the PNO++ and combined PNO++ methods as implemented as a simulation code in PSI4 to larger organic molecules in order to validate the methods and compare them to the performance of the PNO method for the same systems. We also explore refinements to the PNO++ method, including a new formulation of the PNO++ density specifically for optical rotation calculations, and the application of energy- and perturbation-based weak pair approximations to reduce the computational expense of the methods.

5.2 Theory and Computational Details

5.2.1 Larger Benchmark Calculations

Large molecule benchmark calculations were run using an implementation of a local correlation simulation code in a development version of the PSI4 program package.[188] In the code, nonlocal contributions to the residual were filtered out at each iteration in the coupled cluster amplitude, left-hand amplitude and perturbed wave function equations. For further details of the simulation refer to Ref 187. The combined PNO++ method had its T_{cutPNO}

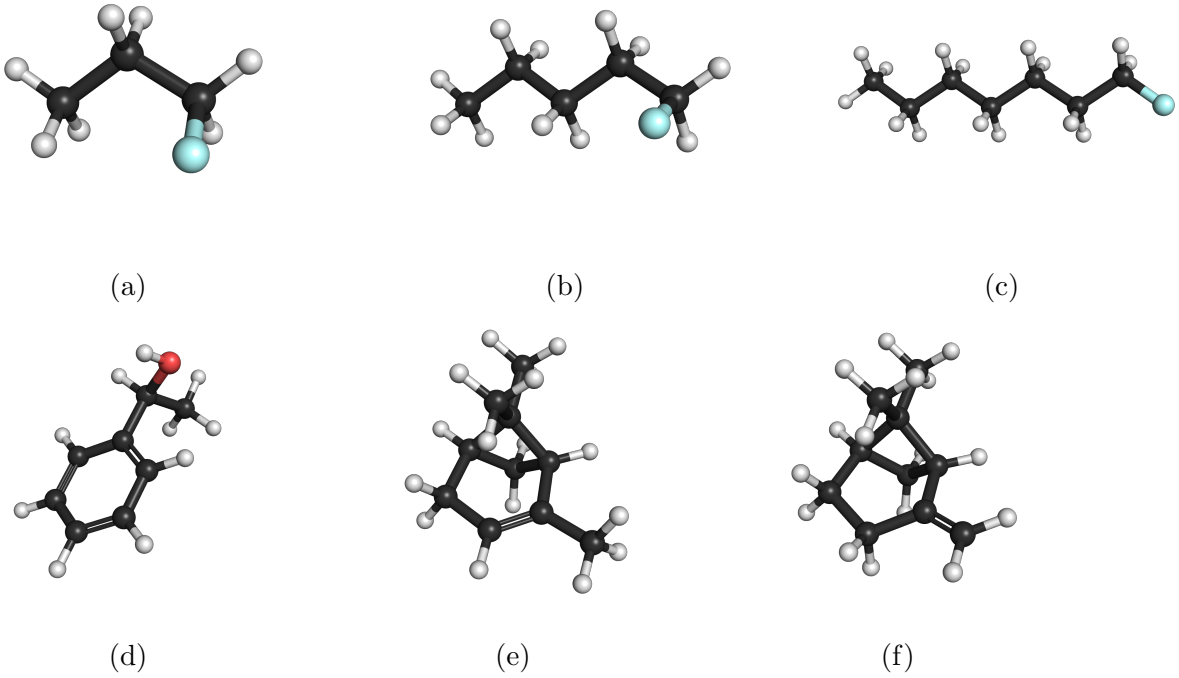
threshold fixed to a reasonable value of 10^{-6}. All molecular geometries were optimized at the B3LYP[105–107]/aug-cc-pVDZ[115,116] level using PSI4's optking module. The molecules studied in this work are given in Fig 5.1. They include a set of fluoroalkane chains, a phenyl-substituted alcohol, and two bicyclic alkenes. Dipole polarizabilities and specific rotations were computed using CCSD linear response[131,132] at 589 nm, with and without the local simulation applied. The specific rotation was computed using both the length and the velocity gauge representation of the dipole moment operator, with a shift at 0 nm giving us the modified velocity gauge value of the specific rotation.[44,189] All calculations used the augmented correlation-consistent double zeta basis set of Dunning and coworkers, aug-cc-pVDZ.[183]

5.2.2 Product-based densities

Pair densities in the PNO and PNO++ approaches are formed using MP2-level amplitudes in the form of the MP2 reduced density matrix. Spaces of natural orbitals are then formed by diagonalization of the pair densities, with the PNO++ density including some measure of the external perturbation in order to improve the compactness of the space for response properties. However, the PNO++ density, written as:

$$\mathbf{D}^{ij}(B,\omega) = \frac{2}{1+\delta_{ij}}\left(\mathbf{X}_B^{ij}\widetilde{\mathbf{X}}_B^{ij\dagger} + \mathbf{X}_B^{ij\dagger}\widetilde{\mathbf{X}}_B^{ij}\right) \tag{5.1}$$

only contains dependence on a single perturbation B, which can be either $\hat{\mu}$ or $\hat{m}$. The optical rotation tensor $\beta(\omega)$, on the other hand, contains both the electric dipole moment



(a) (b) (c)

(d) (e) (f)

Figure 5.1: Systems used in this work. (a) (*M*)-1-fluoropropane, (b) (*M*)-1-fluoropentane, (c) (*M*)-1-fluoroheptane, (d) (*S*)-1-phenylethanol, (e) (1*R*,5*R*)-α-pinene, and (f) (1*R*,5*R*)-β-pinene. All optimized geometries can be found in the Supporting Information.

operator $\hat{\mu}$ and the magnetic dipole moment operator $\hat{m}$:

$$\beta(\omega) = \mathrm{Im}\langle\langle\hat{\mu};\hat{m}\rangle\rangle_{\omega}. \tag{5.2}$$

The linear response function can further be formulated in terms of the first-order perturbed density as:

$$\langle\langle\hat{\mu};\hat{m}\rangle\rangle_{\omega} = \frac{1}{2}\hat{P}(\hat{\mu},\hat{m})\left[\sum_{pq}\hat{\mu}_{pq}\left[D_{pq}^{\hat{m}^{\omega}}\right]^{(1)}\right] \tag{5.3}$$

where the permutation operator $\hat{P}$ ensures that the linear response function contains both $[D_{pq}^{\hat{\mu}^{\omega}}]^{(1)}$ and $[D_{pq}^{\hat{m}^{\omega}}]^{(1)}$. Thus, a reasonable form of the pair density used for creating the space

could include an element-wise product of the perturbed pair densities containing $\hat{\mu}$ and $\hat{m}$.

$$\mathbf{D}^{ij} = \mathbf{D}^{ij}(\hat{\mu}, \omega) \circ \mathbf{D}^{ij}(\hat{m}, \omega) \tag{5.4}$$

Orbitals created using this product density should in principle contain more information about the response of the system in the case of the mixed response function in Eq (5.2) above, leading to a more compact space for the specific rotation.

5.2.3 Weak Pairs

The locality of a localized occupied basis can be exploited by treating at a lower level or neglecting certain sets of PNOs. Those PNOs which correspond to the pairs of occupied orbitals ij whose approximated contributions to the quantity of interest are smaller than a predetermined cutoff can be neglected. These contributions are used as a measure of the importance of those pairs, and neglected pairs are known as weak pairs. For the PNO method, this criterion is of the form:

$$|\epsilon_{ij}| < T_{cutPairs} \tag{5.5}$$

where ϵ_{ij} is defined as the pair correlation energy and is computed at the MP2 level for a given ij pair. The method works by neglecting or otherwise treating with a less expensive method the pairs whose integrals

$$\int \phi_i^*(r_1)\phi_a^*(r_2)\frac{1}{r_{12}}\phi_j(r_1)\phi_b(r_2)dr_1dr_2 \tag{5.6}$$

are negligible, and thus their contribution to the energy is also negligible. Here, the pairs ij are occupied, with the pairs ab being virtual orbitals. This is true when the orbitals are

far enough away spatially to have minimal overlap. This works well for localized occupied orbitals on large molecules, and thus the number of pairs and therefore the total number of PNOs can be reduced to a much more computationally efficient number.

Previous studies neglecting weak pairs using the energy criterion saw that adding a second truncation threshold in addition to the virtual-space truncation led to larger errors in response properties at similar fractions of the space kept. A similar perturbation-including criterion can be formulated as:

$$|\bar{\mu}_{ij}| < T_{cutPairs} \tag{5.7}$$

where

$$\bar{\mu}_{ij} = \sum_{ab} \bar{\mu}_{ij}^{ab} \tag{5.8}$$

with $\bar{\mu}_{ij}^{ab}$ being the similarity-transformed perturbation using MP2-level doubles amplitudes, given below:

$$\bar{\mu}_{ij} = e^{-T_2} \mu e^{T_2}. \tag{5.9}$$

In this work, we neglect contributions from the weak pairs, determined either using the MP2 pair correlation energy or the MP2-level similarity transformed perturbation, in order to quantify the loss in accuracy at various truncation thresholds. The goal is to combine this perturbation-including weak pair criterion with the PNO++ method in order to truncate the space even further than in previous work and attempt to maintain accuracy in the correlation energy as well as response properties.

5.3 Results and Discussion

5.3.1 Larger Benchmark Calculations

Fluoroalkane Chains

Figures 5.2-5.4 show the effect of truncation of the PNO, PNO++ and combined spaces on the errors in correlation energy, dipole polarizabilities and specific rotations for the systems of 1-fluoropropane, -pentane and -heptane. The systems are linear and can provide information about the effect of an increasing linear system size on the truncation errors in energy and properties. The errors in the correlation energy are plotted as a function of the T_2 ratio, which is the ratio of the number of truncated T_2 amplitudes to the full untruncated number, which can be used as a measure of the computational savings obtained by the corresponding truncation of the space.

Across systems, we see that the truncation errors in correlation energy are larger for the PNO++ method, with the PNO method achieving values within chemical accuracy of the CCSD reference at T_2 ratios of 0.11, 0.04 and 0.02 respectively, while the PNO++ requires T_2 ratios of more than 0.36, 0.22 and 0.19 for the same result. The combined method, with a set threshold for unperturbed PNOs to be included in the space, regains a significant portion of the accuracy at small truncations.

Figure 5.3 shows the convergence in the values of the dipole polarizabilities with respect to the T_2 ratio, with the error for the PNO method being larger than that for the PNO++

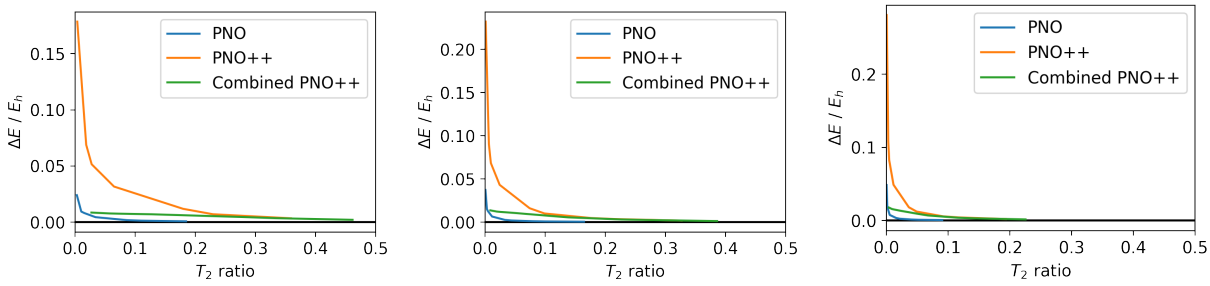
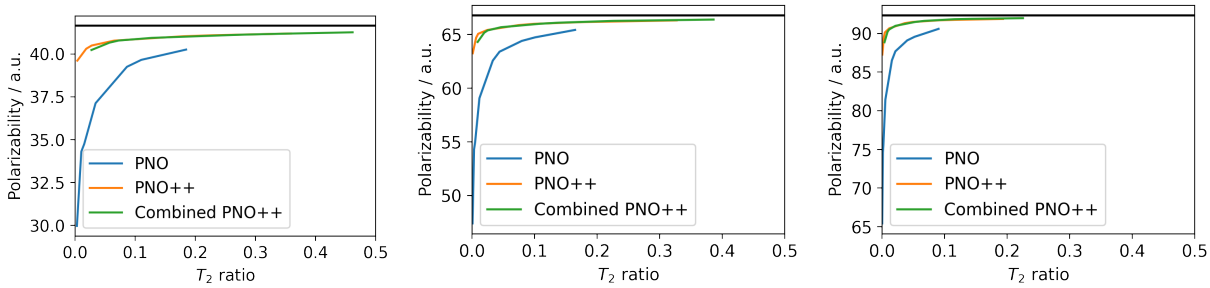


Figure 5.2: Truncation errors in CCSD correlation energy in Hartree for (a) (M)-1-fluoropropane, (b) (M)-1-fluoropentane and (c) (M)-1-fluoroheptane systems, computed using the aug-cc-pVDZ basis set.

and combined methods at each truncation. The PNO++ and combined methods show a slowly-tapering error, reaching within 5% of the reference value at a T_2 ratio of 0.03 but requiring a larger amount of the space for complete convergence to the reference value.

Specific rotations computed for the fluoroalkane systems, which are optically active due to a twist in the methyl group containing the fluorine away from the molecular mirror plane, can be seen in Figure 5.4. The data shows better convergence for the PNO++ and combined PNO++ methods as compared to the PNO method, at T_2 ratios below 0.2 for the smaller 1-fluoropropane and 1-fluoropentane systems, and below 0.1 for the larger 1-fluoroheptane system. The large fluctuations in specific rotation values seen with PNOs become smaller fluctuations with the PNO++ and combined PNO++ methods, an observation that aligns with the results of the previous study. However, rotations computed at T_2 ratios below 0.5 are slow to converge to the reference value for both the PNO++ and the combined PNO++



(a) (b) (c)

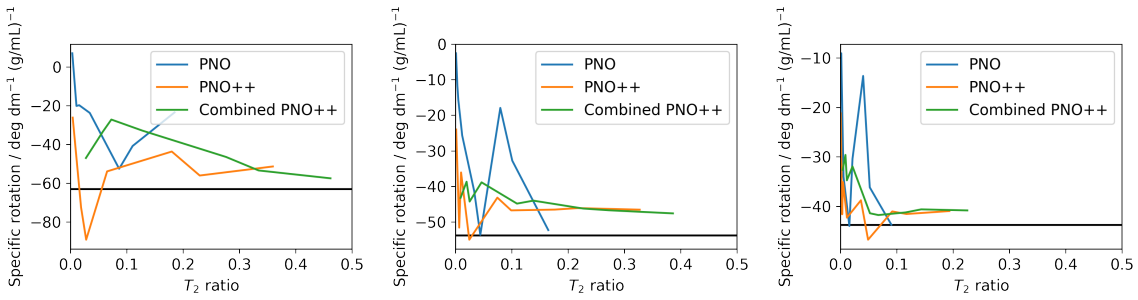
Figure 5.3: CCSD linear response dynamic polarizabilities at 589 nm in a.u. for (a) (*M*)-1-fluoropropae, (b) (*M*)-1-fluoropentane and (c) (*M*)-1-fluoroheptane systems, computed using the aug-cc-pVDZ basis set.

methods. Rotations computed using the length-gauge representation of the electric dipole moment operator show quicker convergence to the reference value, results which are specific to this set of molecules and are not reproduced for any other system studied (See Supporting Information).

While convergence of the linear response properties to the reference CCSD value is not seen at T_2 ratios below 0.5, the results indicate that the PNO++ and combined PNO++ methods still perform better in terms of reducing the cost of property calculations than the PNO method for this set of linear systems.

α- and β-pinene

Figures 5.5 and 5.6 show the effect of truncation with the PNO, PNO++ and combined PNO++ methods on correlation energies and linear response properties computed for the



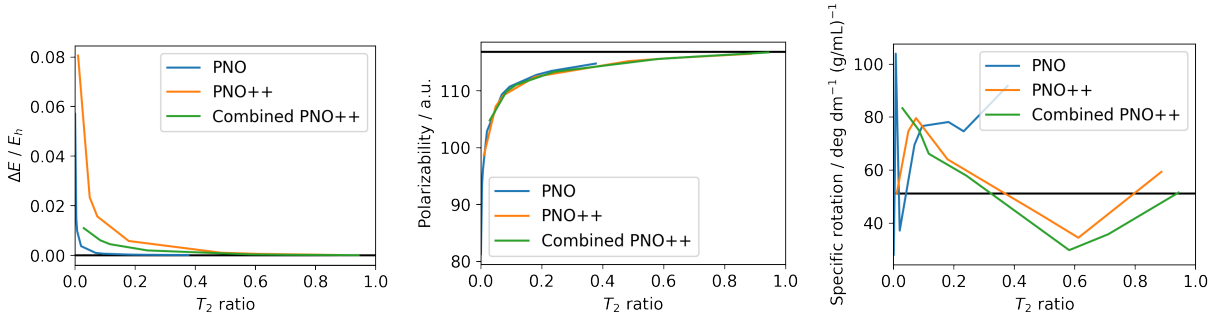
(a) (b) (c)

Figure 5.4: CCSD specific rotations in deg dm^{-1} (g/mL)$^{-1}$ at 589 nm (a) (*M*)-1-fluoropropane, (b) (*M*)-1-fluoropentane and (c) (*M*)-1-fluoroheptane systems, computed using the aug-cc-pVDZ basis set.

systems of (1*R*,5*R*)-α-pinene and (1*R*,5*R*)-β-pinene. Errors in correlation energies (Fig 5.5(a) and 5.6(a)) follow the same trend as seen for the fluoroalkane systems earlier, with the PNO method having much smaller errors in comparison with the PNO++ at large truncations, and the combined method recovering accuracy at the same truncations.

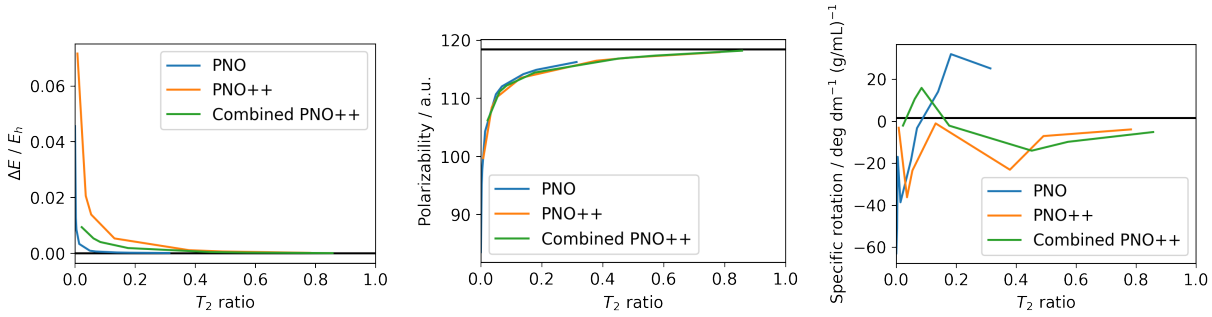
Dipole polarizabilities for both pinene systems calculated using the PNO, PNO++ and combined methods have similar errors at similar T_2 ratios. This is in contrast to the results seen with the fluoroalkanes and indicates that all three spaces show similar levels of compactness for the polarizability for these bicyclic molecules.

Figures 5.5(c) and 5.6(c) contain specific rotations as a function of the T_2 ratio. The pinene systems differ only by the placement of the double bond, but their specific rotations calculated using CCSD linear response differ significantly, with the modified velocity gauge value for α-pinene being 51.19 deg dm^{-1} (g/mL)$^{-1}$, and the same for β-pinene being 1.72 deg



(a) (b) (c)

Figure 5.5: (a) Truncation errors in CCSD correlation energy in Hartree, (b) CCSD dynamic polarizabilities at 589 nm in a.u., and (c) CCSD specific rotations in deg dm^{-1} (g/mL)$^{-1}$ at 589 nm for (1*R*,5*R*)-α-pinene, computed using the aug-cc-pVDZ basis set.



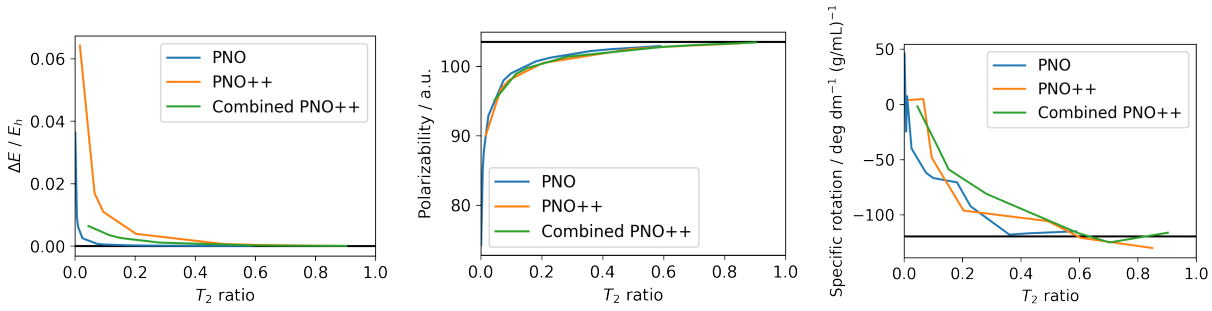
(a) (b) (c)

Figure 5.6: (a) Truncation errors in CCSD correlation energy in Hartree, (b) CCSD dynamic polarizabilities at 589 nm in a.u., and (c) CCSD specific rotations in deg dm^{-1} (g/mL)$^{-1}$ at 589 nm for (1*R*,5*R*)-β-pinene, computed using the aug-cc-pVDZ basis set.

dm^{-1} $(g/mL)^{-1}$. We see that for α-pinene, all three methods show oscillations around the reference value, with the error being fairly large up to the T_2 ratio of 1.0, indicating that even small truncations of the space introduce a large error (of up to 33% at a T_2 ratio of 0.6) in the specific rotation value. The small value of the specific rotation of β-pinene introduces additional problems, as a negative truncation error at small T_2 ratios means that the sign of the specific rotation for the PNO and PNO++ methods is incorrect at all T_2 ratios considered. We see this in Figure 5.6(c). For both pinene systems, while the error using the PNO method is larger than that using the PNO++ method at several T_2 ratios, neither the PNO++ nor the combined method can be used to reliably maintain accuracy in the specific rotation value.

(*S*)-1-phenylethanol

Figure 5.7 plots the truncation errors in correlation energy as well as the dynamic polarizabilities and specific rotations for all three methods for the system of (S)-1-phenylethanol, a challenging system for reduced scaling methods exploiting spatial locality due to the presence of the aromatic ring substituent. The PNO method, along with showing lower errors in correlation energy relative to the PNO++ and combined methods, also shows similar errors for the polarizability (within 2 a.u.) to the PNO++ and combined methods. The convergence behavior of the specific rotation is better for the PNO method for this system than the PNO++ method, with a point at a T_2 ratio of 0.36 falling within 3 deg dm^{-1} $(g/mL)^{-1}$ of the reference value. In comparison, the PNO++ and combined methods require T_2 ratios



(a) (b) (c)

Figure 5.7: (a) Truncation errors in CCSD correlation energy in Hartree, (b) CCSD dynamic polarizabilities at 589 nm in a.u, and (c) CCSD specific rotations in deg dm^{-1} (g/mL)$^{-1}$ at 589 nm for (*S*)-1-phenylethanol, computed using the aug-cc-pVDZ basis set.

of between 0.6-0.7 to achieve similarly close values. The system presents a challenge to the local correlation methods tested here, with very large portions of the space required in order to minimize the error in specific rotation value.

5.3.2 Product Densities

Correlation Energies

Figure 5.8 contains data contrasting the PNO++ method explored in Ref 187 with the method using a product density described in Section 5.2.2, for four small test systems: two hydrogen molecule helices, H_2O_2 and 1,3-dimethylallene (DMA). At the same T_2 ratio, the truncation error in the correlation energy is higher for the product-based density than for the regular PNO++ density using only the electric dipole moment operator. The product-based

density requires T_2 ratios greater than 0.5 in order to maintain reasonable accuracy in the correlation energy. Since the PNO++ method itself shows higher truncation errors than the PNO method for correlation energies at small T_2 ratios, this indicates that the product-based density is not a good method for obtaining correlation energies at truncated system sizes.

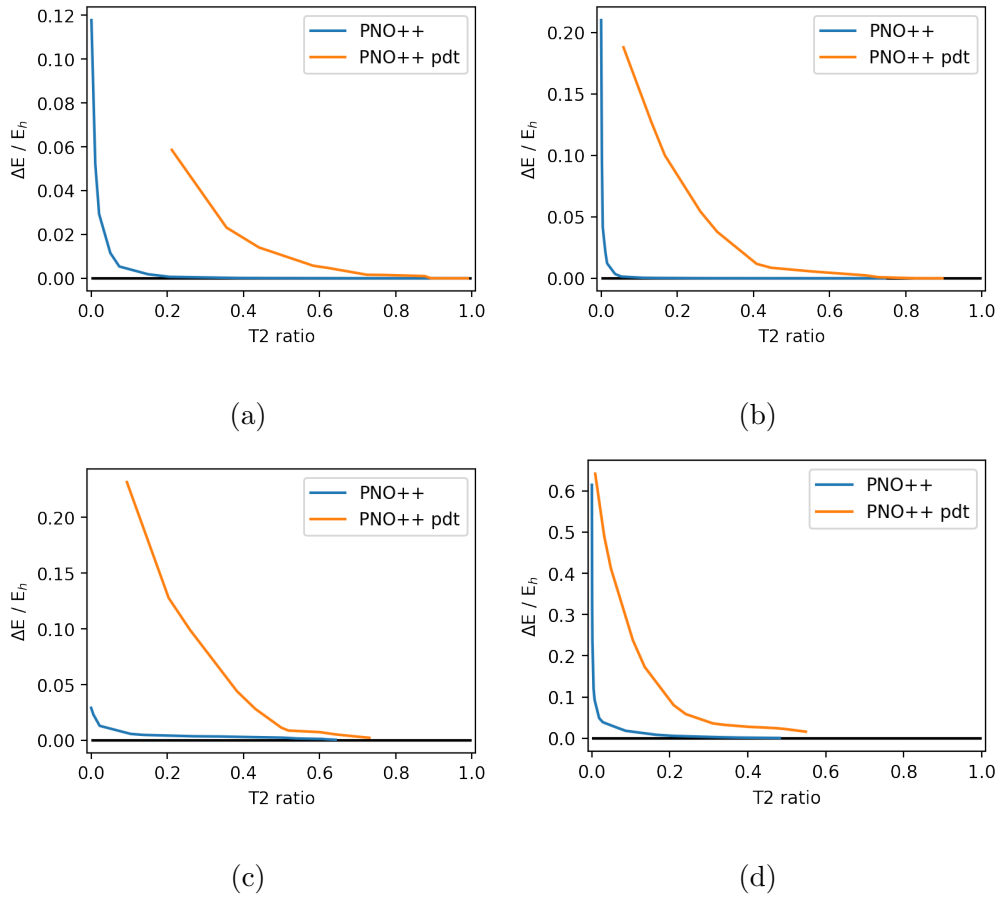


Figure 5.8: Truncation errors in CCSD correlation energy in Hartree for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set for the PNO++ methods with the regular perturbed (PNO++, blue) and product (PNO++ pdt, orange) densities.

Optical Rotation

Figure 5.9 shows optical rotations computed using the PNO++ method with regular and product densities. While convergence is seen for both methods, it can be seen from the figures that the product-based density offers no improvement to the convergence behavior shown by the PNO++ method for the four test systems. At T_2 ratios below 0.25, we see a large magnitude of errors in the rotation value for the product density space, with the specific rotation for DMA at very large truncation being the wrong sign. Thus the space created by the product density is not more optimized for the mixed property of rotation than the density containing only a single external perturbation operator.

One possible reason for the worse convergence behavior of the product density versus the density using only the electric dipole moment operator is the worse convergence behavior of the PNO++ method using only the magnetic dipole moment operator to create the perturbed density (See SI for Ref 187).

5.3.3 Weak Pairs

Table 5.1 contains T_2 ratios and errors in correlation energies for the pair energy criterion ϵ as well as the perturbation-including criterion $\bar{\mu}$ at a given weak pair threshold $T_{cutPairs}$ for the 1-fluoropropane, -pentane, and -heptane systems. As the system size increases, we see the T_2 ratio becoming uniformly smaller at a given threshold for the ϵ criterion indicating that fewer pairs have contributions to the energy larger in magnitude than the chosen cutoff.

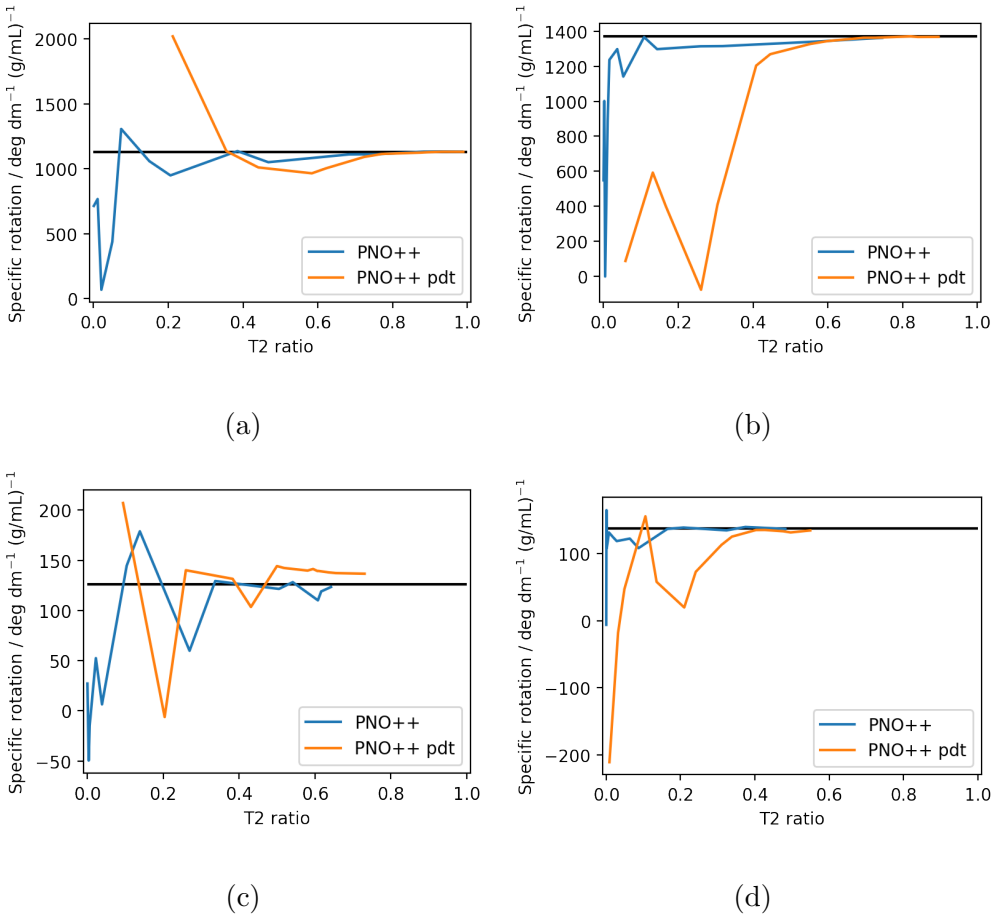


Figure 5.9: CCSD specific rotations in deg dm^{-1} (g/mL)$^{-1}$ at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_7$, (c) H_2O_2 and (d) DMA systems, computed using the aug-cc-pVDZ basis set for the PNO++ methods with the regular perturbed (PNO++, blue) and product (PNO++ pdt, orange) densities.

This is also true for the $\bar{\mu}$ criterion. This aligns with the discussion in Refs 186 and 163 that on increasing molecular sizes the computational efficiency of local correlation methods improves.

For both of the considered criteria, errors in the correlation energy require T_2 ratios above 0.83 in order to remain below chemical accuracy or 1.6 mE_h, thus keeping a significant portion of the space. However, the errors in correlation energy seen using the $\bar{\mu}$ criterion are larger at a given T_2 ratio than the errors using the ϵ criterion. As an example, for the very similar T_2 ratios of 0.805 and 0.795 for 1-fluoropentane, the error in correlation energy was less than 1 mE_h for the pair energy criterion, while it was above 13 mE_h for the perturbation-including criterion. The pair energy criterion thus provides a better way to identify weak pairs specifically for maintaining accuracy in the correlation energy.

Table 5.2 presents polarizabilities and specific rotations computed using the modified velocity gauge as a function of the threshold for both criteria for the 1-fluoropropane, -pentane, and -heptane systems. For all three fluoroalkanes and all values of the threshold the polarizability value remains within 1 a.u. of the reference, even at the smallest T_2 ratios. Thus the polarizability is not strongly affected by the neglect of weak pairs using either criterion. The specific rotation, on the other hand, is affected by the truncation and is seen to require a T_2 ratio between 0.5-0.7 for the ϵ criterion and 0.4-0.8 for the $\bar{\mu}$ criterion to remain within 10% of the reference CCSD value.

Table 5.3 presents data on the errors in correlation energy with the neglect of weak pairs using various thresholds for (*S*)-1-phenylethanol. This system represents a difficult case for

Table 5.1: T_2 ratios and errors in correlation energy (mE_h) computed at the CCSD level for the three 1-fluoroalkane systems as a function of the $T_{cutPairs}$ threshold, using the aug-cc-pVDZ basis set.

Molecule	Threshold	ϵ_{ij}		$\bar{\mu}_{ij}$	
		T_2 ratio	ΔE	T_2 ratio	ΔE
1-fluoropropane	10^{-3}	0.211	51.1838	0.457	136.677
	10^{-4}	0.612	2.75454	0.827	2.64050
	10^{-5}	0.709	1.57325	0.958	0.22850
	10^{-6}	0.882	0.11981	1.000	0.00000
1-fluoropentane	10^{-3}	0.146	83.3302	0.413	70.5225
	10^{-4}	0.456	6.20241	0.795	13.6891
	10^{-5}	0.568	3.20850	0.955	0.05368
	10^{-6}	0.805	0.15424	0.987	0.00912
1-fluoroheptane	10^{-3}	0.111	112.815	0.328	87.2679
	10^{-4}	0.361	9.77172	0.722	16.5131
	10^{-5}	0.469	4.85166	0.935	2.22753
	10^{-6}	0.726	0.18945	0.991	0.01013

Table 5.2: T_2 ratios, dynamic polarizabilities (a.u.), and specific rotations (deg dm^{-1} (g/mL)$^{-1}$) at 589 nm computed at the CCSD level for the three 1-fluoroalkane systems as a function of the $T_{cutPairs}$ threshold, using the aug-cc-pVDZ basis set. α_{Ref} 1-fluoropropane: 41.65, 1-fluoropentane: 66.77, 1-fluoroheptane: 92.30. MVG_{Ref} 1-fluoropropane: -63.03, 1-fluoropentane: -53.72, 1-fluoroheptane: -43.75.

Molecule	Threshold	ϵ_{ij}			$\bar{\mu}_{ij}$		
		T_2 ratio	α	MVG	T_2 ratio	α	MVG
1-fluoropropane	10^{-3}	0.211	42.02	-83.06	0.457	42.69	-58.10
	10^{-4}	0.612	41.69	-77.05	0.827	41.67	-60.30
	10^{-5}	0.709	41.66	-64.56	0.958	41.66	-57.66
	10^{-6}	0.882	41.65	-63.12	1.000	41.65	-62.89
1-fluoropentane	10^{-3}	0.146	67.26	-77.99	0.413	67.61	-71.85
	10^{-4}	0.456	66.64	-65.14	0.795	66.95	-57.39
	10^{-5}	0.568	66.77	-57.89	0.955	66.77	-54.78
	10^{-6}	0.805	66.77	-54.96	0.987	66.77	-54.11
1-fluoroheptane	10^{-3}	0.111	92.73	-62.17	0.328	92.99	-50.76
	10^{-4}	0.361	91.89	-53.42	0.722	92.47	-43.19
	10^{-5}	0.469	92.21	-47.57	0.935	92.30	-42.51
	10^{-6}	0.726	92.30	-44.82	0.991	92.30	-44.21

the application of local correlation methods to properties. The T_2 ratio data shows that the ϵ criterion keeps less of the space for a given threshold than the $\bar{\mu}$ criterion. The errors in correlation energy are similar at similar T_2 ratios, being 1.57 mE_h at a T_2 ratio of 0.93, and between 0.4-11 mE_h between the T_2 ratios of 0.92-0.98 for the ϵ and $\bar{\mu}$ criterion respectively.

Table 5.3: T_2 ratios and errors in correlation energy (mE_h) computed at the CCSD level for the (S)-1-phenylethanol system as a function of the $T_{cutPairs}$ threshold, using the aug-cc-pVDZ basis set.

Molecule	Threshold	ϵ_{ij}		$\bar{\mu}_{ij}$	
		T_2 ratio	ΔE	T_2 ratio	ΔE
(S)-1-phenyl-ethanol	10^{-3}	0.281	85.5406	0.917	11.9979
	10^{-4}	0.722	4.99202	0.983	0.38369
	10^{-5}	0.934	1.57325	0.993	0.02812
	10^{-6}	1.000	0.00000	1.000	0.00000

Finally, trends in polarizabilities and specific rotations are shown for the (S)-1-phenylethanol system in Table 5.4. Once again, the change in the polarizability value is minimal even at very small T_2 ratios, remaining within 1.5 a.u. of the reference. Using both weak pair criteria, a T_2 ratio of at least 0.7 is required in order to remain within 10% of the reference rotation value. A comparison of the points with T_2 ratios above 0.9 suggests that the ϵ criterion loses less accuracy with truncation for this system when compared to the perturbation-including criterion. These results are consistent with the results from Section 5.3.1, where the PNO method had lower errors in the specific rotation at a given T_2 ratio than the PNO++. Thus (S)-1-phenylethanol continues to present a challenge for perturbation-including truncation

schemes.

Table 5.4: T_2 ratios, dynamic polarizabilities (a.u.), and specific rotations (deg dm^{-1} (g/mL)$^{-1}$) at 589 nm computed at the CCSD level for the (*S*)-1-phenylethanol system as a function of the $T_{cutPairs}$ threshold, using the aug-cc-pVDZ basis set. α_{Ref}: 103.47. MVG_{Ref}: -119.32

Molecule	Threshold	ϵ_{ij}			$\bar{\mu}_{ij}$		
		T_2 ratio	α	MVG	T_2 ratio	α	MVG
(*S*)-1-phenyl-ethanol	10^{-3}	0.281	104.21	-159.15	0.917	103.57	-109.97
	10^{-4}	0.722	103.30	-119.94	0.983	103.48	-97.661
	10^{-5}	0.934	103.43	-120.94	0.993	103.47	-119.23
	10^{-6}	1.000	103.47	-119.31	1.000	103.47	-119.31

5.4 Conclusions

In this work, we have compared the performance of the PNO++ and combined PNO++ methods to the PNO method for larger molecules than considered in previous work using an implementation of the simulation code in PSI4. We have also tested a new product-based density in the creation of the PNO++ space, as well as the accuracy limits of both an energy- and perturbation-based weak pair threshold.

For the series of 1-fluoroalkane systems tested, we see better convergence behavior and lower truncation errors for the PNO++ and combined PNO++ methods than the PNO for the linear response properties tested. While the truncation errors in correlation energy

were large for the PNO++ method, we recover accuracy by incorporating a number of pair natural orbitals in the combined PNO++ space. α- and β-pinene were more difficult systems than the linear alkane chains for all three methods tested, requiring large amounts of the space to be kept in order to obtain accuracy in the dynamic polarizability, and showing large errors in the specific rotation value even at small truncations. A surprising result is the very similar convergence behavior of the polarizability for all three methods, with all three having approximately the same error at a given truncation. Finally, truncation errors in both the correlation energy and the specific rotation are smaller for the PNO method than the PNO++ or combined methods for the system of (S)-1-phenylethanol.

The product density, in theory, should result in a more compact space for specific rotations, having combined two perturbed densities with an element-wise multiplication. However, the results suggest that the density using a single electric field perturbation is more effective than the product density and that further testing is required in order to understand the impact of using a mixed density in order to create the PNO++ space.

The weak pair criteria underscored several observations from the larger benchmark calculations with the PNO, PNO++, and combined PNO++ methods. T_2 ratios are larger at a given threshold for the $\bar{\mu}$ than for the ϵ criterion, suggesting that this criterion naturally contains larger magnitude contributions than the pair energy criterion. The ϵ criterion provides a better way to identify weak pairs specifically for maintaining accuracy in the correlation energy, while the polarizability is not affected by aggressive truncation, and thus either criterion could be used to truncate the space. For the specific rotation, the ability to truncate

at a lower threshold for a given criterion is system-dependent, with the linear fluoroalkane chains requiring slightly less of the space with the $\bar{\mu}$ than with the ϵ criterion and the (*S*)-1-phenylethanol system requiring more of the space with the $\bar{\mu}$ criterion. This mirrors observations of the truncation error with the PNO method versus the PNO++ method for the same systems.

Through this study we have studied the effect of truncation on CCSD correlation energies, dynamic polarizabilities, and specific rotations using three local correlation methods, the PNO, PNO++ and combined PNO++ methods. While the results for the linear alkane chains are encouraging in their better convergence behavior, it is clear that the PNO++ method is not a black-box method for the local correlation approximation of linear response properties. Reasons for the greatly reduced accuracy of the method for some molecules still need further elucidation.

Chapter 6

Conclusions

The main aim of this work is to explore and improve on current methods of predicting molecular chiroptical properties. To this end, we began in Chapter 1 by providing context to the problem of chiroptical property-prediction in quantum chemistry. Chapter 2 focused on deriving the equations needed to compute chiroptical properties using electronic structure theory, specifically dynamic polarizabilities and specific rotations using coupled cluster linear response theory.

In Chapter 3, we have explored a methodology to attempt to match the experimental specific rotations of small organic molecules in solution. Using increasing sets of explicitly-treated solvent molecules and averaging over a classical molecular dynamics trajectory, we see convergence with respect to the number of solvent molecules for all but one system studied, and significant variation in the specific rotation with the orientation of solvent molecules.

The impact of the density functional used is most clearly seen in the underestimation by the B3LYP functional[105,106] of excitation energies for the system of (*R*)-methylthiirane in water, resulting in large-magnitude optical rotation values which skew the averages over the trajectory.

Having found that averages of time-dependent density functional theory calculations over a classical molecular dynamics trajectory fail to match experimental results, and are, in fact, heavily dependent on the density functional used, we turn to coupled cluster theory as a more robust and improvable method. However, coupled-cluster level linear response properties are computationally expensive enough to discourage attempts to average over hundreds of calculations, and thus we then considered methods of approximation in order to reduce the computational cost of a single calculation and to bring sets of high-accuracy calculations within our reach.

Chapter 4 presented a discussion of a local correlation reduced-scaling method which includes the perturbation in the density used to create virtual orbital spaces in a similar way to the established pair natural orbital (PNO)[53,67,68] method. The PNO++ method is relatively compact for linear response properties of the small test systems studied, but shows large truncation errors in the correlation energy relative to the PNO method. Incorporating a number of orbitals from the PNO space into the PNO++ method recovers accuracy in the correlation energy, thus creating a "combined PNO++" method that can be used for both correlation energies as well as linear response properties.

The PNO++ and combined PNO++ methods, while converging quickly for the small, local-

ized systems studied in Chapter 4, require further benchmarking for larger organic molecules than in the previous study. Chapter 5 contained a study testing the performance of the PNO++ and combined PNO++ methods on larger organic molecules, brought within reach via an implementation in the quantum chemistry code PSI4. It also presents energy and specific rotation data comparing the original PNO++ space to one created using an element-wise product of perturbed densities, finding that the mixed density so created is not more compact for the rotation than the original PNO++ space. In the case of larger molecules, a weak pair screening criterion has proven to be effective in order to remove those terms from the calculation whose approximated contributions are small in magnitude. Our results suggest that a weak pair approximation does well for the reduced scaling of dynamic polarizabilities, but specific rotations require a tight threshold for the neglect of weak pairs in order to maintain accuracy.

There are several avenues for further study of improvements to the PNO++ method. Having considered the effect of the occupation number and weak pair thresholds individually on the truncation error for the PNO++ method, pairing the PNO++ and combined PNO++ methods with the weak pair criteria studied could present a more complete picture of the computational savings that could be obtained. Further exploration of different forms of the perturbation-including density is also warranted from the results in Chapter 5. Finally, we observed in Chapter 4 that MP2-level corrections do not improve on the response properties as they do on the correlation energies, however, one could formulate lower-scaling corrections that would still include information required for accurate properties.

A key goal would be an optimized production-level black-box method that would retain accuracy while allowing us to calculate properties within a reasonable timeframe. An additional future path would be the combination of that optimized method with averaging over a trajectory to obtain accurate predictions of chiroptical properties of small organic molecules in solution.

Bibliography

[1] Danchin, A.; Sekowska, A. The logic of metabolism and its fuzzy consequences. *Environ. Microbiol.* **2014**, *16*, 19–28.

[2] Slattery, D.; Wong, S. W.; Colin, A. A. Levalbuterol hydrochloride. *Pediatr. Pulmonol.* **2002**, *33*, 151–157.

[3] D'Amato, R. J.; Loughnan, M. S.; Flynn, E.; Folkman, J. Thalidomide is an inhibitor of angiogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 4082–4085.

[4] Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press: Camridge, U.K., 2004.

[5] Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479–483.

[6] Mukhopadhyay, P.; Zuber, G.; Goldsmith, M.-R.; Wipf, P.; Beratan, D. N. Solvent

Effect on Optical Rotation: A Case Study of Methyloxirane in Water. *ChemPhysChem* **2006**, *7*, 2483–2486.

[7] Kundrat, M. D.; Autschbach, J. Modeling of the chiroptical response of chiral amino acids in solution using explicit solvation and molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 1051–1060.

[8] McAlexander, H. R.; Crawford, T. D. A Comparison of Three Approaches to the Reduced-Scaling Coupled Cluster Treatment of Non-Resonant Molecular Response Properties. *J. Chem. Theory Comput.* **2016**, *12*, 209–222.

[9] Schrödinger, E. *Collected Papers on Wave Mechanics*, 3rd ed.; Chelsea Publishing Company: New York, 1982.

[10] Atkins, P. W.; Friedman, R. S. *Molecular Quantum Mechanics*, 5th ed.; Oxford University Press, 2011.

[11] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1996.

[12] Born, M.; Oppenheimer, R. Zur Quantentheorie der Molekeln. *Ann. Phys.* **1927**, *389*, 457–484.

[13] Roothaan, C. C. J. New developments in molecular orbital theory. *Rev. Mod. Phys.* **1951**, *23*, 69–89.

[14] Slater, J. C. The theory of complex spectra. *Phys. Rev.* **1929**, *34*, 1293–1322.

[15] Pauli, W. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Zeitschrift für Phys.* **1925**, *31*, 765–783.

[16] Sinanoglu, O. *Adv. Chem. Phys.*; 1964; Vol. 6; pp 315–412.

[17] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods. *J. Chem. Phys.* **1966**, *45*, 4256–4266.

[18] Čižek, J.; Paldus, J. Correlation problems in atomic and molecular systems III. Rederivation of the coupled-pair many-electron theory using the traditional quantum chemical methods. *Int. J. Quantum Chem.* **1971**, *5*, 359–379.

[19] Crawford, T. D.; Schaefer, H. F. In *Rev. Comput. Chem.*; Kenny B. Lipkowitz, D. B. B., Ed.; John Wiley and Sons, Inc.: New York, 2000; Chapter 2, pp 33–136.

[20] Achilles, R.; Bonfiglioli, A. The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin. *Arch. Hist. Exact Sci.* **2012**, *66*, 295–358.

[21] Nooijen, M.; Shamasundar, K. R.; Mukherjee, D. Reflections on size-extensivity, size-consistency and generalized extensivity in many-body theory. *Mol. Phys.* **2005**, *103*, 2277–2298.

[22] Bartlett, R. J. Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359–401.

[23] Cársky, P.; Paldus, J.; Pittner, J. *Recent Progress in Coupled Cluster Methods: Theory and Applications*; 2010.

[24] Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291–352.

[25] Geertsen, J.; Rittby, M.; Bartlett, R. J. The equation-of-motion coupled-cluster method: Excitation energies of Be and CO. *Chem. Phys. Lett.* **1989**, *164*, 57–62.

[26] Stanton, J. F.; Bartlett, R. J. The equation of motion coupled-cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state properties. *J. Chem. Phys.* **1993**, *981*.

[27] Monkhorst, H. J. Calculation of properties with the coupled-cluster method. *Int. J. Quantum Chem.* **1977**, *12*, 421–432.

[28] Dalgaard, E.; Monkhorst, H. J. Some aspects of the time-dependent coupled-cluster approach to dynamic response functions. *Phys. Rev. A* **1983**, *28*, 1217–1222.

[29] Maxwell, J. C. *A Treatise on Electricity and Magnetism Vol. I*; The Clarendon Press: Oxford, 1873.

[30] Barron, L. D.; Gray, C. G. The multipole interaction Hamiltonian for time dependent fields. *J. Phys. A Gen. Phys.* **1973**, *6*, 59–61.

[31] Rosenfeld, L. Quantenmechanische Theorie der naturlichen optischen Aktivitaet von Fluessigkeiten und Gasen. *Z. Phys.* **1929**, *52*, 161–174.

[32] Drude, P. *The Theory of Optics*; Longmans, Green and Co.: London, 1902.

[33] Norman, P. A perspective on nonresonant and resonant electronic response theory for time-dependent molecular properties. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20519.

[34] Norman, P.; Ruud, K.; Helgaker, T. Density-functional theory calculations of optical rotatory dispersion in the nonresonant and resonant frequency regions. *J. Chem. Phys.* **2004**, *120*, 5027–5035.

[35] Helgaker, T.; Coriani, S.; Jørgensen, P.; Kristensen, K.; Olsen, J.; Ruud, K. Recent advances in wave function-based methods of molecular-property calculations. *Chem. Rev.* **2012**, *112*, 543–631.

[36] Mukherjee, D.; Mukherjee, P. A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism. *Chem. Phys.* **1979**, *39*, 325–335.

[37] Olsen, J.; Jørgensen, P. Linear and nonlinear response functions for an exact state and for an MCSCF state. *J. Chem. Phys.* **1984**, *82*, 3235–3264.

[38] Weiss, H.; Ahlrichs, R.; Häser, M. A direct algorithm for self-consistent-field linear response theory and application to C60: Excitation energies, oscillator strengths, and frequency-dependent polarizabilities. *J. Chem. Phys.* **1993**, *99*, 1262–1270.

[39] Autschbach, J.; Ziegler, T. Calculating molecular electric and magnetic properties

from time-dependent density functional response theory. *J. Chem. Phys.* **2002**, *116*, 891–896.

[40] Bast, R.; Ekstro, U.; Gao, B.; Helgaker, T.; Ruud, K.; Thorvaldsen, A. J. The ab initio calculation of molecular electric, magnetic and geometric properties. *Phys. Chem. Chem. Phys. Phys. Chem. Chem. Phys* **2011**, *13*, 2627–2651.

[41] Christiansen, O.; Jørgensen, P.; Hättig, C. Response functions from Fourier component variational perturbation theory applied to a time-averaged quasienergy. *Int. J. Quantum Chem.* **1998**, *68*, 1–52.

[42] Hayes, E. F.; Parr, R. G. Time-Dependent Hellmann-Feynman Theorems. *J. Chem. Phys.* **1965**, *43*, 1831–1832.

[43] Langhoff, P. W.; Epstein, S. T.; Karplus, M. Aspects of time-dependent perturbation theory. *Rev. Mod. Phys.* **1972**, *44*, 602–644.

[44] Pedersen, T. B.; Koch, H. Coupled cluster response functions revisited. *J. Chem. Phys.* **1997**, *106*, 8059–8072.

[45] Nesbet, R. K. *Adv. Chem. Phys.*; 1965; Vol. IX; pp 321–363.

[46] Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151–154.

[47] Saebø, S.; Pulay, P. Local Treatment Of Electron Correlation*. *Annu. Rev. Phys. Chem* **1993**, *44*, 213–36.

[48] Yang, J.; Chan, G. K.-L.; Manby, F. R.; Schütz, M.; Werner, H.-J. The orbital-specific-virtual local coupled cluster singles and doubles method. *J. Chem. Phys.* **2012**, *136*, 144105.

[49] Li, W.; Piecuch, P. Improved Design of Orbital Domains within the Cluster-in-Molecule Local Correlation Framework: Single-Environment Cluster-in-Molecule Ansatz and its Application to the Coupled-Cluster Approach with Singles and Doubles. **1952**,

[50] Ni, Z.; Li, W.; Li, S. Fully optimized implementation of the cluster-in-molecule local correlation approach for electron correlation calculations of large systems. *J. Comput. Chem.* **2019**, *40*, 1130–1140.

[51] Li, W.; Li, S. Divide-and-conquer local correlation approach to the correlation energy of large molecules. *J. Chem. Phys.* **2004**, *121*, 6649–6657.

[52] Ziółkowski, M.; Jansík, B.; Kjrgaard, T.; Jørgensen, P. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys.* **2010**,

[53] Ahlrichs, R.; Driessler, F. Direct determination of pair natural orbitals. *Theor. Chim. Acta* **1975**, *36*, 275–287.

[54] Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*.

[55] Korona, T.; Werner, H.-J. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.* **2003**, *118*, 3006–3019.

[56] Kats, D.; Korona, T.; Schütz, M. Local CC2 electronic excitation energies for large molecules with density fitting. *J. Chem. Phys.* **2006**, *125*.

[57] Baudin, P.; Kristensen, K. LoFEx - A local framework for calculating excitation energies: Illustrations using RI-CC2 linear response theory. *J. Chem. Phys.* **2016**, *144*.

[58] Myhre, R. H.; Koch, H. The multilevel CC3 coupled cluster model. *J. Chem. Phys.* **2016**, *145*.

[59] Baudin, P.; Kristensen, K. Correlated natural transition orbital framework for low-scaling excitation energy calculations (CorNFLEx). *J. Chem. Phys.* **2017**, *146*.

[60] Höfener, S.; Klopper, W. Natural transition orbitals for the calculation of correlation and excitation energies. *Chem. Phys. Lett.* **2017**, *679*, 52–59.

[61] Dutta, A. K.; Nooijen, M.; Neese, F.; Izsák, R. Exploring the Accuracy of a Low Scaling Similarity Transformed Equation of Motion Method for Vertical Excitation Energies. *J. Chem. Theory Comput.* **2018**, *14*, 72–91.

[62] Frank, M. S.; Hättig, C. A pair natural orbital based implementation of CCSD excitation energies within the framework of linear response theory. *J. Chem. Phys.* **2018**, *148*.

[63] Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.

[64] Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–465.

[65] Pipek, J.; Mezey, P. G. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

[66] Edmiston, C.; Krauss, M. Pseudonatural orbitals as a basis for the superposition of configurations. I. He2+. *J. Chem. Phys.* **1966**, *45*, 1833–1839.

[67] Meyer, W. PNO–CI Studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane. *J. Chem. Phys.* **1973**, *58*, 1017–1035.

[68] Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *J. Chem. Phys.* **2009**, *130*, 114108–134101.

[69] Neese, F.; Hansen, A.; Liakos, D. G. Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis. *J. Chem. Phys.* **2009**, *131*, 64103–114108.

[70] Löwdin, P. O. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Phys. Rev.* **1955**, *97*, 1474–1489.

[71] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M.; Neese, F. Exploring the accuracy limits of local pair natural orbital coupled-cluster theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525–1539.

[72] Mennucci, B.; Cappelli, C.; Cammi, R.; Tomasi, J. Modeling Solvent Effects on Chiroptical Properties. *Chirality* **2011**, *23*, 717–729.

[73] Cappelli, C.; Corni, S.; Mennucci, B.; Cammi, R.; Tomasi, J. Vibrational Circular Dichroism within the Polarizable Continuum Model: A Theoretical Evidence of Conformation Effects and Hydrogen Bonding for (S)-(-)-3-butyn-2-ol in CCl_4 Solution. *J. Phys. Chem. A* **2002**, *106*, 12331–12339.

[74] Pecul, M.; Ruud, K. The Ab Initio Calculation of Optical Rotation and Electronic Circular Dichroism. *Adv. Quantum Chem.* **2005**, *50*, 185–212.

[75] Pecul, M.; Marchesan, D.; Ruud, K.; Coriani, S. Polarizable continuum model study of solvent effects on electronic circular dichroism parameters. *J. Chem. Phys.* **2005**, *122*, 024106.

[76] Pecul, M.; Lamparska, E.; Cappelli, C.; Frediani, L.; Ruud, K. Solvent Effects on Raman Optical Activity Spectra Calculated Using the Polarizable Continuum Model. *J. Phys. Chem. A* **2006**, *110*, 2807–2815.

[77] Crawford, T. D. In *Comprehensive Chiroptical Spectroscopy*; Berova, N., Nakanishi, K., Woody, R. W., Polavarapu, P., Eds.; Wiley: Hoboken, New Jersey, 2012; Vol. 1; Chapter 23, pp 675–697.

[78] Crawford, T. D. In *Frontiers of Quantum Chemistry*; Wójcik et al., M. J., Ed.; Springer Nature: Singapore, 2018; Chapter 3.

[79] Mukhopadhyay, P.; Zuber, G.; Goldsmith, M. R.; Wipf, P.; Beratan, D. N. Solvent effect on optical rotation: A case study of methyloxirane in water. *Chem. Phys. Chem.* **2006**, *7*, 2483–2486.

[80] Mukhopadhyay, P.; Zuber, G.; Wipf, P.; Beratan, D. N. Contribution of a solute's chiral solvent imprint to optical rotation. *Angew. Chem. Int. Ed.* **2007**, *46*, 6450–6452.

[81] Kumata, Y.; Furukawa, J.; Fueno, T. The Effect of Solvents on the Optical Rotation of Propylene Oxide. *Bull. Chem. Soc. Japan* **1970**, *43*, 3920–3921.

[82] Tomasi, J.; Persico, M. Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent. *Chem. Rev.* **1994**, *94*, 2027–2094.

[83] Tomasi, J.; Cammi, R.; Mennucci, B.; Cappelli, C.; Corni, S. Molecular properties in solution described with a continuum solvation model. *Phys. Chem. Chem. Phys.* **2002**, *4*, 5697–5712.

[84] Mennucci, B.; Tomasi, J.; Cammi, R.; Cheeseman, J. R.; Frisch, M. J.; Devlin, F. J.; Gabriel, S.; Stephens, P. J. Polarizable continuum model (PCM) calculations of solvent effects on optical rotation of chiral molecules. *J. Phys. Chem. A* **2002**, *106*, 6102–6113.

[85] Wilson, S. M.; Wiberg, K. B.; Cheeseman, J. R.; Frisch, M. J.; Vaccaro, P. H. The

nonresonant optical activity of isolated organic molecules. *J. Phys. Chem. A* **2005**, *109*, 11752–11764.

[86] Marchesan, D.; Coriani, S.; Forzato, C.; Nitti, P.; Pitacco, G.; Ruud, K. Optical Rotation Calculation of a Highly Flexible Molecule: The Case of Paraconic Acid. *J. Phys. Chem. A* **2005**, *109*, 1449–1453.

[87] Kongsted, J.; Ruud, K. Solvent effects on zero-point vibrational corrections to optical rotations and nuclear magnetic resonance shielding constants. *Chem. Phys. Lett.* **2008**, *451*, 226–232.

[88] Egidi, F.; Barone, V.; Bloino, J.; Cappelli, C. Toward an Accurate Modeling of Optical Rotation for Solvated Systems: Anharmonic Vibrational Contributions Coupled to the Polarizable Continuum Model. *J. Chem. Theory Comput.* **2012**, *8*.

[89] Aharon, T.; Lemler, P.; Vaccaro, P. H.; Caricato, M. Comparison of measured and predicted specific optical rotation in gas and solution phases: A test for the polarizable continuum model of solvation. *Chirality* **2018**, *30*.

[90] Fattebert, J.-L.; Gygi, F. Density functional theory for efficient ab initio molecular dynamics simulations in solution. *J. Comp. Chem.* **2002**, *23*, 662–666.

[91] Howard, J. C.; Crawford, T. D. Calculating Optical Rotatory Dispersion Spectra in Solution Using a Smooth Dielectric Model. *J. Phys. Chem. A* **2018**, *122*, 8557–8564.

[92] Howard, J. C.; Womack, J.; Dziedzic, J.; Skylaris, C.-K.; Pritchard, B. P.; Craw-

ford, T. D. Electronically Excited States in Solution via a Smooth Dielectric Model Combined with Equation-of-Motion Coupled Cluster Theory. *J. Chem. Theory Comput.* **2017**, *11*, 5572–5581.

[93] Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1997**, *97*, 8050–8053.

[94] Wesolowski, T. A. In *Computational Chemistry: Reviews of Current Trends*; Leszczynski, J., Ed.; World Scientific: Singapore, 2006; Vol. 10; pp 1–82.

[95] Niemeyer, N.; Tölle, J.; Neugebauer, J. Approximate versus Exact Embedding for Chiroptical Properties: Reconsidering Failures in Potential and Response. *J. Chem. Theory Comput.* **2020**, *16*, 3104–3120.

[96] Lipparini, F.; Egidi, F.; Cappelli, C.; Barone, V. The Optical Rotation of Methyloxirane in Aqueous Solution: A Never Ending Story? *J. Chem. Theory Comp.* **2013**, *9*, 1880–1884.

[97] Kongsted, J.; Pedersen, T. B.; Jensen, L.; Hansen, A. E.; Mikkelsen, K. V. Coupled cluster and density functional theory study of the vibrational contribution to the optical rotation of (*S*)-propylene oxide. *J. Am. Chem. Soc.* **2006**, *128*, 976–982.

[98] Pedersen, T. B.; Kongsted, J.; Crawford, T. D.; Ruud, K. On the importance of vibrational contributions to small-angle optical rotation: Fluorooxirane in gas phase and solution. *J. Chem. Phys.* **2009**, *130*, 034310.

[99] Egidi, F.; Giovannini, T.; Del Frate, C.; Lemler, P. M.; Vaccaro, P. H.; Cappelli, C. A combined experimental and theoretical study of optical rotatory dispersion for (*R*)-glycidyl methyl ether in aqueous solution. *Phys. Chem. Chem. Phys.* **2019**, *21*, 3644–3655.

[100] Klamt, A.; Schüürmann, G. COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient. *J. Chem. Soc. Perkin Trans.* **1993**, *2*, 799–805.

[101] Kundrat, M. D.; Autschbach, J. Ab initio and density functional theory modeling of the chiroptical response of glycine and alanine in solution using explicit solvation and molecular dynamics. *J. Chem. Theory Comput.* **2008**, *4*, 1902–1914.

[102] Peyton, B. G.; Crawford, T. D. Basis Set Superposition Errors in the Many-Body Expansion of Molecular Properties. *J. Phys. Chem. A* **2019**,

[103] Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

[104] Berendsen, H. J.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

[105] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[106] 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* **1988**, *37*, 785–789.

[107] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Ab initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Theory. *J. Phys. Chem.* **1994**, *98*, 11623–11627.

[108] 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.

[109] Haghdani, S.; Åstrand, P.-O.; Koch, H. Optical Rotation from Coupled Cluster and Density Functional Theory: The Role of Basis Set Convergence. *J. Chem. Theory Comput.* **2016**, *12*, 535–548.

[110] Haghdani, S.; Hoff, B. H.; Koch, H.; Åstrand, P.-O. Solvent Effects on Optical Rotation: On the Balance between Hydrogen Bonding and Shifts in Dihedral Angles. *J. Phys. Chem. A* **2017**, *121*, 4756–4777.

[111] 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.

[112] London, F., Théorie quantique des courants interatomiques dans les combinaisons aromatiques. *J. Phys. Radium* **1937**, *8*, 397–409.

[113] Ditchfield, R. Self-consistent perturbation theory of diamagnetism. *Molecular Physics* **1974**, *27*, 789–807.

[114] Helgaker, T.; Jørgensen, P. An electronic Hamiltonian for origin independent calculations of magnetic properties. *J. Chem. Phys.* **1991**, *95*, 2595–2801.

[115] Dunning, T. H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

[116] Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

[117] Baranowska-Łączkowska, A.; Łączkowski, K. Z. The ORP basis set designed for optical rotation calculations. *J. Comput. Chem.* **2013**, *34*, 2006–2013.

[118] 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.

[119] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[120] Frisch, M. J. et al. Gaussian 09 Revision E.01. 2009.

[121] Wiberg, K. B.; Vaccaro, P. H.; Cheeseman, J. R. Conformational effects on optical rotation. 3-Substituted 1-butenes. *J. Am. Chem. Soc.* **2003**, *125*, 1888–1896.

[122] Crawford, T. D.; Tam, M. C.; Abrams, M. L. The problematic case of (*S*)-methylthiirane: Electronic Circular Dichroism Spectra and Optical Rotatory Dispersion. *Mol. Phys.* **2007**, *105*, 2607–2617.

[123] Crawford, T. D. *Ab Initio* Calculation of Molecular Chiroptical Properties. *Theor. Chem. Acc.* **2006**, *115*, 227–245.

[124] Peach, M. J. G.; Benfield, P.; Helgaker, T.; Tozer, D. J. Excitation energies in density functional theory: An evaluation and a diagnostic test. *J. Chem. Phys.* **2008**, *128*.

[125] Miertuš, S.; Scrocco, E.; Tomasi, J. Electrostatic interaction of a solute with a continuum. A direct utilizaion of ab initio molecular potentials for the prevision of solvent effects. *Chemical Physics* **1981**, *55*, 117 – 129.

[126] Schieschke, N.; Di Remigio, R.; Frediani, L.; Heuser, J.; Höfener, S. Combining frozen-density embedding with the conductor-like screening model using Lagrangian techniques for response properties. *J. Comput. Chem.* **2017**, *38*, 1693–1703.

[127] Wilson, S. M.; Wiberg, K. B.; Cheeseman, J. R.; Frisch, M. J.; Vaccaro, P. H. Nonresonant Optical Activity of Isolated Organic Molecules. *J. Phys. Chem. A* **2005**, *109*, 11752–11764.

[128] Mukhopadhyay, P.; Zuber, G.; Wipf, P.; Beratan, D. N. Contribution of a Solute's

Chiral Solvent Imprint to Optical Rotation. *Angew. Chemie - Int. Ed.* **2007**, *46*, 6450–6452.

[129] Barron, L. D. *Molecular Light Scattering and Optical Activity*, 2nd ed.; Cambridge University Press: Cambridge, 2004.

[130] Crawford, T. D. Ab initio calculation of molecular chiroptical properties. *Theor. Chem. Acc.* **2006**, *115*, 227–245.

[131] Koch, H.; Jørgensen, P. Coupled-Cluster Response Functions. *J. Chem. Phys.* **1990**, *93*, 3333–3344.

[132] Pedersen, T. B.; Koch, H. Coupled cluster response functions revisited. *J. Chem. Phys.* **1997**, *106*, 8059.

[133] Jørgensen, P. Molecular and Atomic Applications of Time-Dependent Hartree-Fock Theory. *Annu. Rev. Phys. Chem.* **1975**, *26*, 359–380.

[134] Oddershede, J. *Adv. Quantum Chem.*; 1978; Vol. 11; pp 275–352.

[135] Casida, M. E. In *Recent Adv. Density Funct. Methods, Part I*; D. P. Chong,, Ed.; World Scientific, 1995; Chapter 5, pp 155–192.

[136] Casida, M. E. Time-dependent density-functional theory for molecules and molecular solids. *J. Mol. Struct. THEOCHEM* **2009**, *914*, 3–18.

[137] Nielsen, E. S.; Jørgensen, P.; Oddershede, J. Transition moments and dynamic polarizabilities in a second order polarization propagator approach. *J. Chem. Phys.* **1980**, *73*, 6238–6246.

[138] Packer, M. J.; Dalskov, E. K.; Enevoldsen, T.; Jensen, H. J. A.; Oddershede, J. A new implementation of the second-order polarization propagator approximation (SOPPA): The excitation spectra of benzene and naphthalene. *J. Chem. Phys.* **1996**, *105*, 5886–5900.

[139] Schirmer, J. Beyond the random-phase approximation: A new approximation scheme for the polarization propagator. *Phys. Rev. A* **1982**, *26*, 2395–2416.

[140] Trofimov, A. B.; Krivdina, I. L.; Weller, J.; Schirmer, J. Algebraic-diagrammatic construction propagator approach to molecular response properties. *Chem. Phys.* **2006**, *329*, 1–10.

[141] Dreuw, A.; Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 82–95.

[142] Čížek, J. In *Adv. Chem. Phys.*; Lefebvre, R., Moser, C., Eds.; 1969; Vol. XIV; pp 35–89.

[143] Hampel, C.; Werner, H.-J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286–6297.

[144] Schütz, M.; Werner, H.-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.* **2001**, *114*, 661–681.

[145] Kinoshita, T.; Hino, O.; Bartlett, R. J. Singular value decomposition approach for the approximate coupled-cluster method. *J. Chem. Phys.* **2003**, *119*, 7756–7762.

[146] Koch, H.; Sánchez De Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481–9484.

[147] Hohenstein, E. G.; Parrish, R. M.; Martínez, T. J. Tensor hypercontraction density fitting. I. Quartic scaling second- and third-order Møller-Plesset perturbation theory. *J. Chem. Phys.* **2012**, *137*, 1085.

[148] Schutski, R.; Zhao, J.; Henderson, T. M.; Scuseria, G. E. Tensor-structured coupled cluster theory. *J. Chem. Phys.* **2017**, *147*.

[149] Parrish, R. M.; Zhao, Y.; Hohenstein, E. G.; Martínez, T. J. Rank reduced coupled cluster theory. I. Ground state energies and wavefunctions. *J. Chem. Phys.* **2019**, *150*.

[150] Pawłowski, F.; Olsen, J.; Jørgensen, P. Cluster perturbation theory. I. Theoretical foundation for a coupled cluster target state and ground-state energies. *J. Chem. Phys.* **2019**, *150*.

[151] Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632–672.

[152] Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyaev, K.; Shao, Y.; Krylov, A. I. General implementation of the resolution-of-the-identity and Cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks. *J. Chem. Phys.* **2013**, *139*.

[153] Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*.

[154] Edmiston, C.; Krauss, M. Pseudonatural Orbitals as a Basis for the Superposition of Configurations. II. Energy Surface for Linear H 3. *J. Chem. Phys.* **1968**, *49*, 192–205.

[155] Helmich, B.; Hättig, C. Local pair natural orbitals for excited states. *J. Chem. Phys.* **2011**, *135*.

[156] Dutta, A. K.; Neese, F.; Izsák, R. Towards a pair natural orbital coupled cluster method for excited states. *J. Chem. Phys.* **2016**, *145*.

[157] Peng, C.; Clement, M. C.; Valeev, E. F. State-Averaged Pair Natural Orbitals for Excited States: A Route toward Efficient Equation of Motion Coupled-Cluster. *J. Chem. Theory Comput.* **2018**, *14*, 5597–5607.

[158] Gauss, J.; Werner, H.-J. NMR chemical shift calculations within local correlation methods: The GIAO-LMP2 approach. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2083–2090.

[159] Kats, D.; Korona, T.; Schütz, M. Transition strengths and first-order properties of

excited states from local coupled cluster CC2 response theory with density fitting. *J. Chem. Phys.* **2007**, *127*.

[160] Ledermüller, K.; Kats, D.; Schütz, M. Local CC2 response method based on the Laplace transform: Orbital-relaxed first-order properties for excited states. *J. Chem. Phys.* **2013**, *139*.

[161] Schütz, M. Oscillator strengths, first-order properties, and nuclear gradients for local ADC(2). *J. Chem. Phys.* **2015**, *142*.

[162] Russ, N. J.; Crawford, T. D. Local correlation in coupled cluster calculations of molecular response properties. *Chem. Phys. Lett.* **2004**, *400*, 104–111.

[163] 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*, 7830–6.

[164] Kumar, A.; Crawford, T. D. Frozen virtual natural orbitals for coupled-cluster linear-response theory. *J. Phys. Chem. A* **2017**, *121*, 708–716.

[165] Høyvik, I. M.; Myhre, R. H.; Koch, H. Correlated natural transition orbitals for core excitation energies in multilevel coupled cluster models. *J. Chem. Phys.* **2017**, *146*.

[166] Mester, D.; Nagy, P. R.; Kállay, M. Reduced-cost linear-response CC2 method based on natural orbitals and natural auxiliary functions. *J. Chem. Phys.* **2017**, *146*.

[167] Mester, D.; Nagy, P. R.; Kállay, M. Reduced-cost second-order algebraic-diagrammatic construction method for excitation energies and transition moments. *J. Chem. Phys.* **2018**, *148*.

[168] Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618–622.

[169] Bartlett, R. J.; Silver, D. M. Pair-correlation energies in sodium hydride with many-body perturbation theory. *Phys. Rev. A* **1974**, *10*, 1927–1931.

[170] Magnasco, V.; Perico, A. Uniform Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1967**, *47*, 971–981.

[171] von Niessen, W. Density Localization of Atomic and Molecular Orbitals. I. *J. Chem. Phys.* **1972**, *56*, 4290–4297.

[172] Hansen, A.; Liakos, D. G.; Neese, F. Efficient and accurate local single reference correlation methods for high-spin open-shell molecules using pair natural orbitals. *J. Chem. Phys.* **2011**, *135*.

[173] Krause, C.; Werner, H.-J. Comparison of explicitly correlated local coupled-cluster methods with various choices of virtual orbitals. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7591–7604.

[174] Liakos, D. G.; Hansen, A.; Neese, F. Weak molecular interactions studied with parallel

implementations of the local pair natural orbital coupled pair and coupled cluster methods. *J. Chem. Theory Comput.* **2011**, *7*, 76–87.

[175] Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse maps - A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J. Chem. Phys.* **2016**, *144*.

[176] Clement, M. C.; Zhang, J.; Lewis, C. A.; Yang, C.; Valeev, E. F. Optimized Pair Natural Orbitals for the Coupled Cluster Methods. *J. Chem. Theory Comput.* **2018**, *14*, 4581–4589.

[177] Helmich, B.; Hättig, C. A pair natural orbital implementation of the coupled cluster model CC2 for excitation energies. *J. Chem. Phys.* **2013**, *139*.

[178] Helmich, B.; Hättig, C. A pair natural orbital based implementation of ADC(2)-x: Perspectives and challenges for response methods for singly and doubly excited states in large molecules. *Comput. Theor. Chem.* **2014**, *1040-1041*, 35–44.

[179] Dutta, A. K.; Saitow, M.; Riplinger, C.; Neese, F.; Izsák, R. A near-linear scaling equation of motion coupled cluster method for ionized states. *J. Chem. Phys.* **2018**, *148*.

[180] Datta, D.; Kossmann, S.; Neese, F. Analytic energy derivatives for the calculation of the first-order molecular properties using the domain-based local pair-natural orbital coupled-cluster theory. *J. Chem. Phys.* **2016**, *145*.

[181] Crawford, T. D.; Kumar, A.; Bazanté, A. P.; Di Remigio, R. Reduced-scaling coupled cluster response theory: Challenges and opportunities. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9*, 1–25.

[182] Shavitt, I.; Bartlett, R. J. *Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory*; Cambridge Molecular Science; Cambridge University Press, 2009.

[183] 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.

[184] Helgaker, T.; Klopper, W.; Tew, D. P. Quantitative quantum chemistry. *Mol. Phys.* **2008**, *106*, 2107–2143.

[185] Crawford, T.; King, R. A. Locally correlated equation-of-motion coupled cluster theory for the excited states of large molecules. *Chem. Phys. Lett.* **2002**, *366*, 611–622.

[186] 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.

[187] D'Cunha, R.; Crawford, T. D. PNO++: Perturbed Pair Natural Orbitals for Coupled Cluster Linear Response Theory. *J. Chem. Theory Comput.* **2021**, *17*, 290–301.

[188] Smith, D. G. et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *J. Chem. Phys.* **2020**, *152*.

[189] Pedersen, T. B.; Koch, H.; Boman, L.; Sánchez de Merás, A. M. Origin invariant calculation of optical rotation without recourse to London orbitals. *Chem. Phys. Lett.* **2004**, *393*, 319–326.

Appendix A

Solvent Effects: Supporting Information

The Supporting Information contains molecular geometries, additional optical rotation data for 2-12 solvent systems, running averages over the MD trajectory, and additional convergence data with respect to numbers of snapshots as well as numbers of explicit solvent molecules included in the calculation for (R)-methylthiirane in water and (S)-methyloxirane in water, methanol and carbon tetrachloride.

Table A.1: Atomic coordinates of (R)-methylthiirane (Ångstroms)

Atomic Symbol	X	Y	Z
C	-0.887000	-0.990000	0.655000
S	0.737000	-1.683000	0.099000
C	0.000000	0.000000	0.000000
C	-2.150000	-1.466000	-0.025000
H	-2.010000	-1.533000	-1.108000
H	-2.449000	-2.453000	0.342000
H	-2.972000	-0.766000	0.178000
H	-0.952000	-0.919000	1.740000
H	-0.259000	0.329000	-1.004000
H	0.486000	0.757000	0.611000

Table A.2: Atomic coordinates of (S)-methyloxirane (Ångstroms)

Atomic Symbol	X	Y	Z
C	0.000000	0.000000	0.000000
O	0.000000	0.000000	1.400000
C	1.256000	0.000000	0.725000
C	-0.077000	1.184000	-0.834000
H	-1.020000	1.184000	-1.378000
H	-0.019000	2.073000	-0.208000
H	0.751000	1.184000	-1.541000
H	-0.513000	-0.889000	-0.363000
H	1.827000	-0.889000	0.462000
H	1.827000	0.889000	0.462000

A.1 Input Geometries

A.2 Additional data

A.2.1 Autocorrelation functions

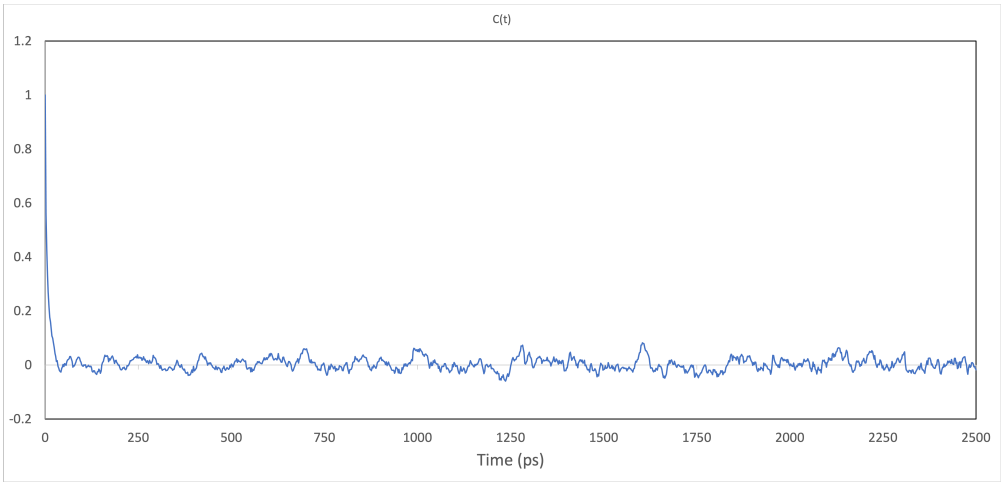


Figure A.1: Dipole autocorrelation function for (*R*)-methylthiirane in a shell of 12 water molecules over 2500 frames of a 5 ns MD trajectory.

A.2.2 Running averages

(*R*)-methylthiirane in water

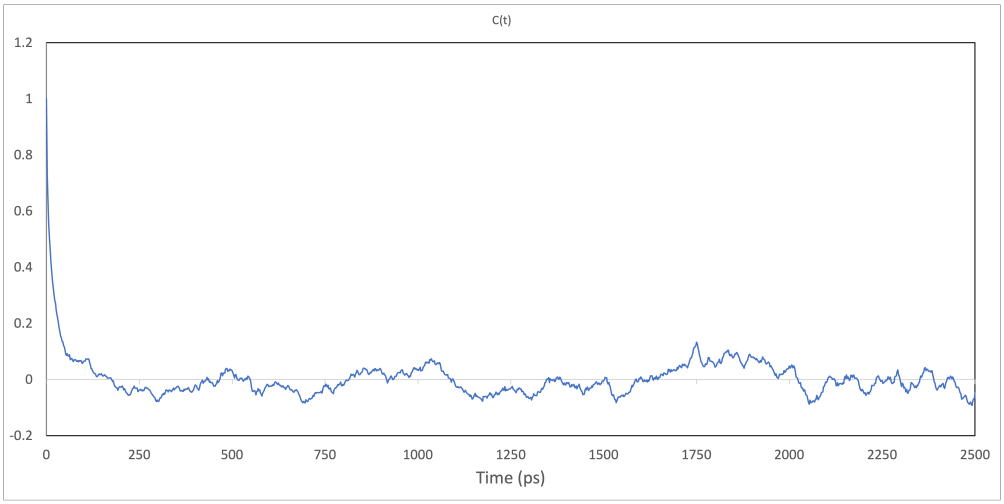


Figure A.2: Dipole autocorrelation function for (*S*)-methyloxirane in a shell of 12 water molecules over 2500 frames of a 5 ns MD trajectory.

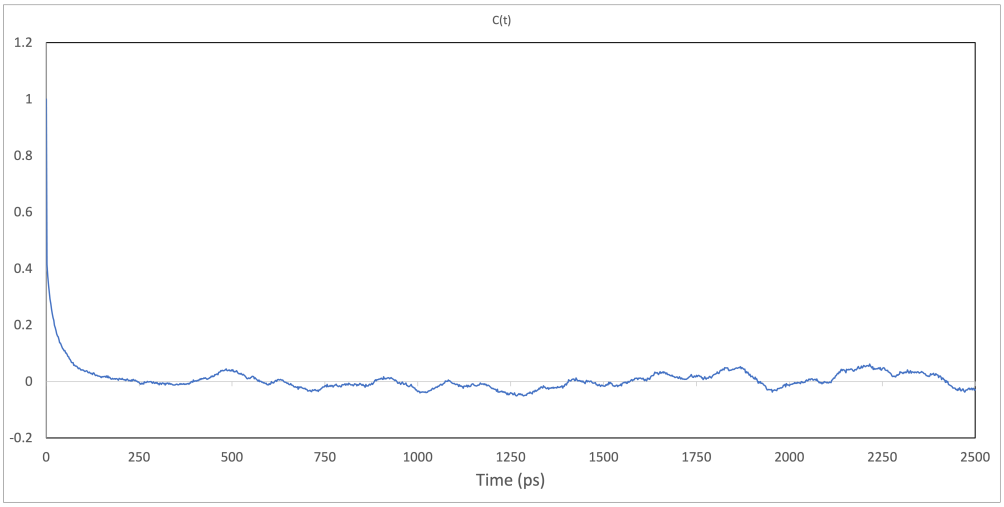


Figure A.3: Dipole autocorrelation function for (*R*)-methylthiirane in a shell of 12 methanol molecules over 2500 frames of a 5 ns MD trajectory.

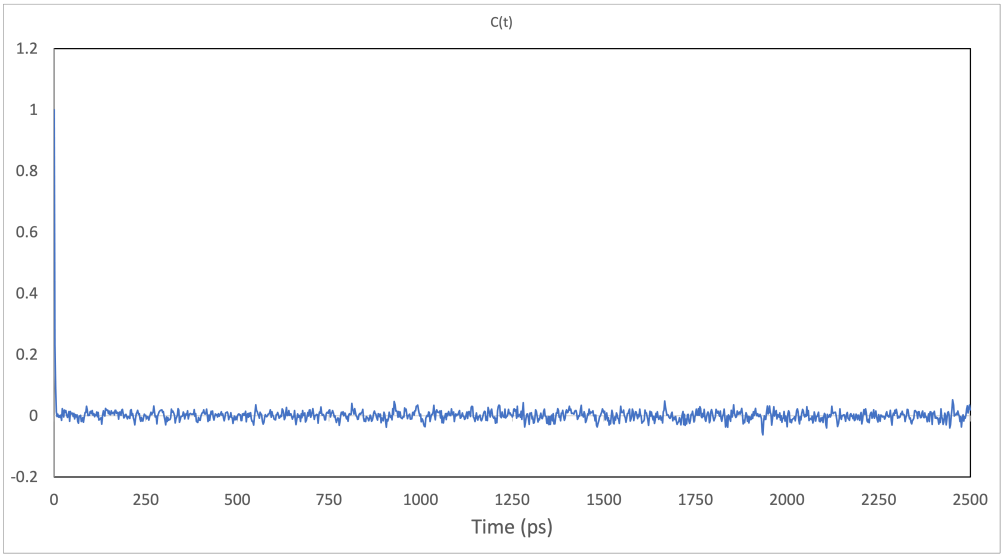


Figure A.4: Dipole autocorrelation function for (*R*)-methylthiirane in a shell of 12 carbon tetrachloride molecules over 2500 frames of a 5 ns MD trajectory.

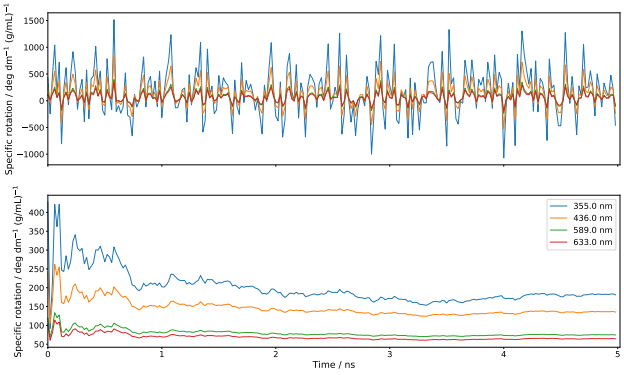


Figure A.5: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

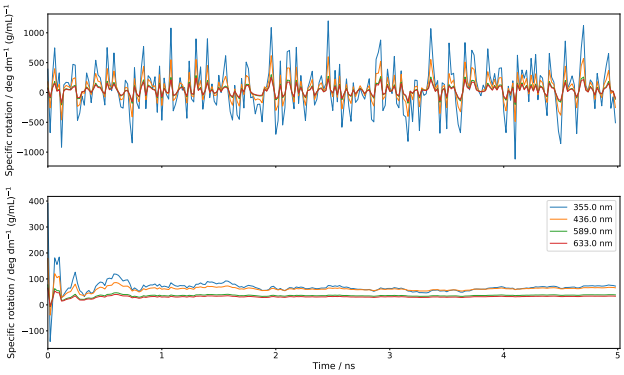


Figure A.6: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

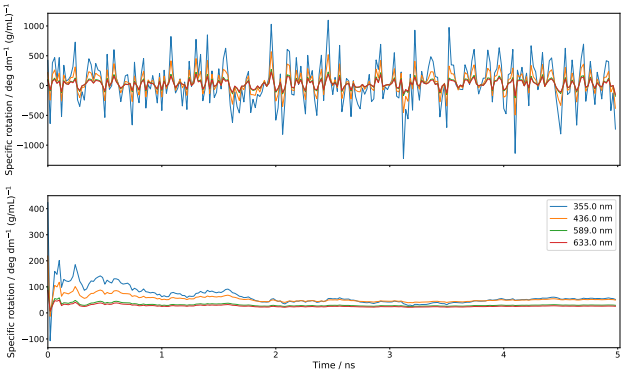


Figure A.7: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

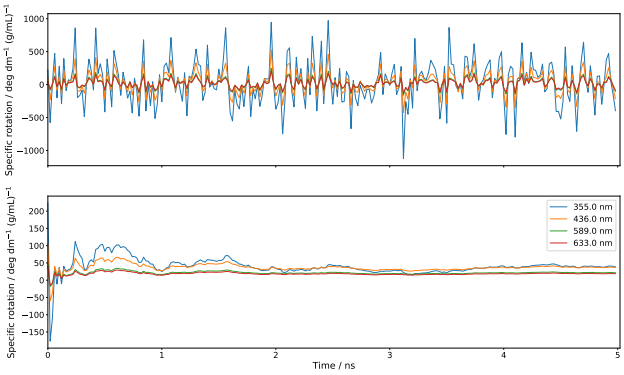


Figure A.8: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

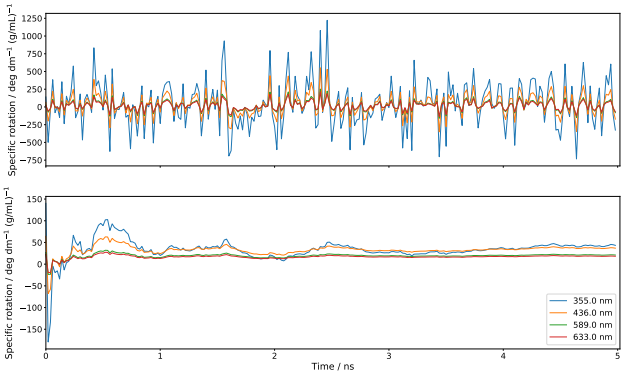


Figure A.9: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

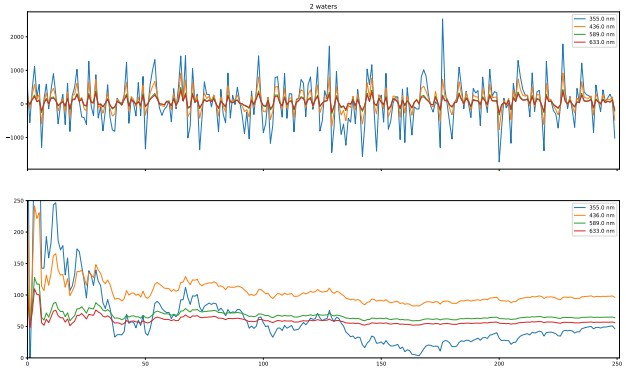


Figure A.10: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

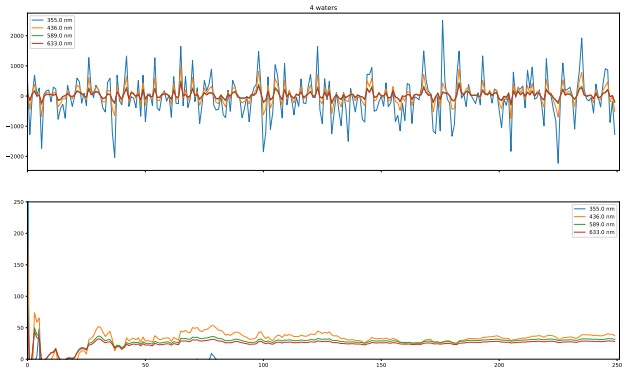


Figure A.11: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

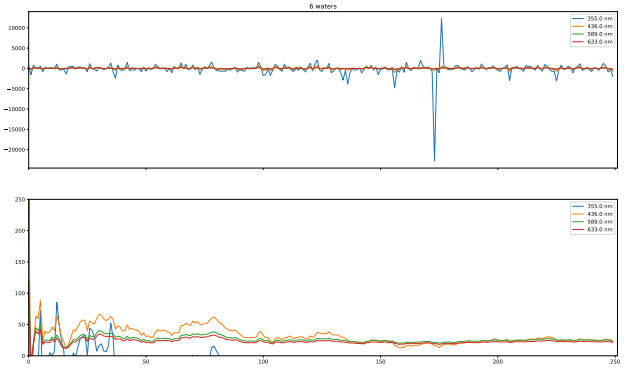


Figure A.12: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

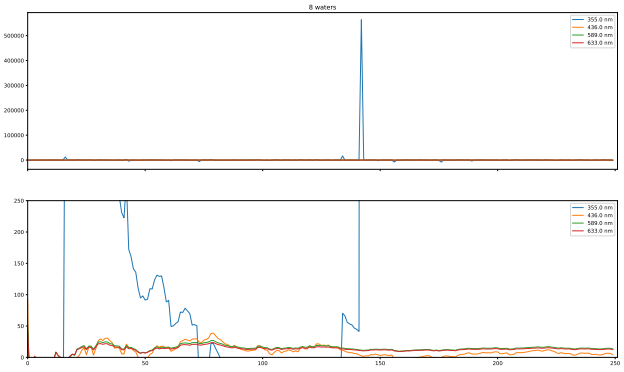


Figure A.13: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

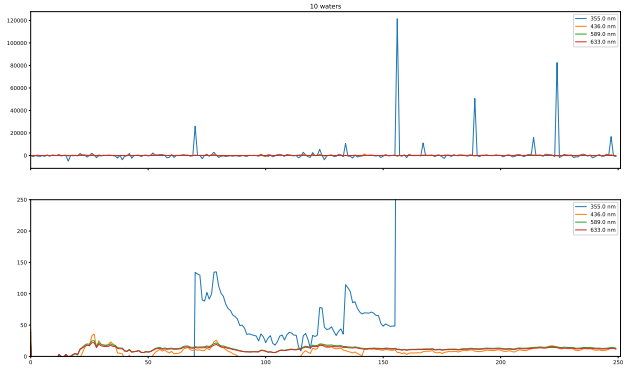


Figure A.14: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

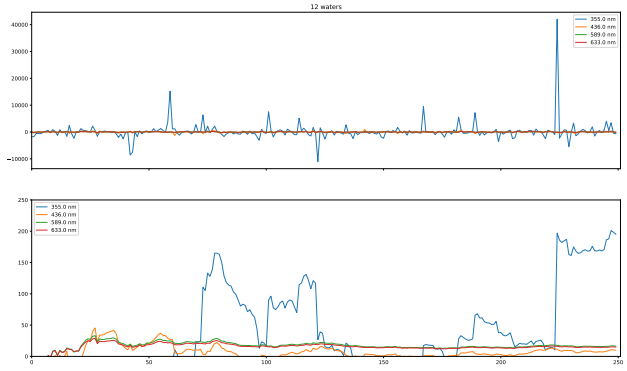


Figure A.15: Specific rotation at the b3lyp/aug-cc-pVDZ level of theory of (*R*)-methylthiirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

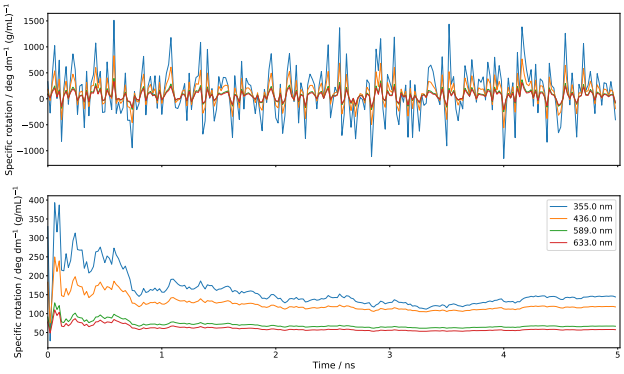


Figure A.16: Specific rotation at the CAM-B3LYP/orp level of theory of (R)-methylthiirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

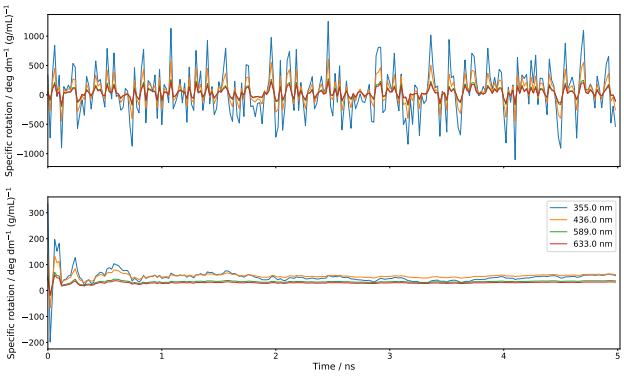


Figure A.17: Specific rotation at the CAM-B3LYP/orp level of theory of (*R*)-methylthiirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

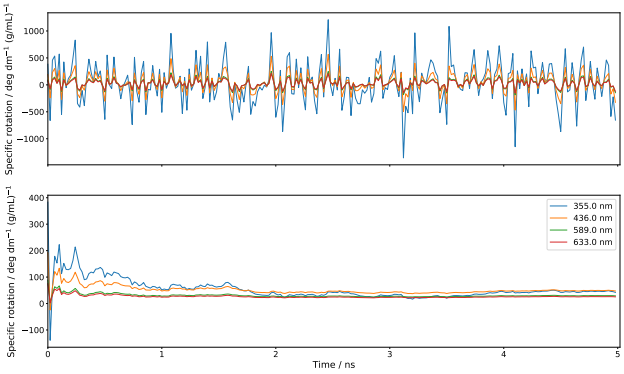


Figure A.18: Specific rotation at the CAM-B3LYP/orp level of theory of (*R*)-methylthiirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

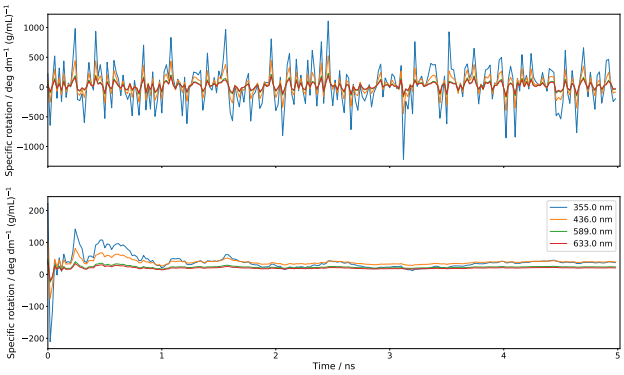


Figure A.19: Specific rotation at the CAM-B3LYP/orp level of theory of (*R*)-methylthiirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

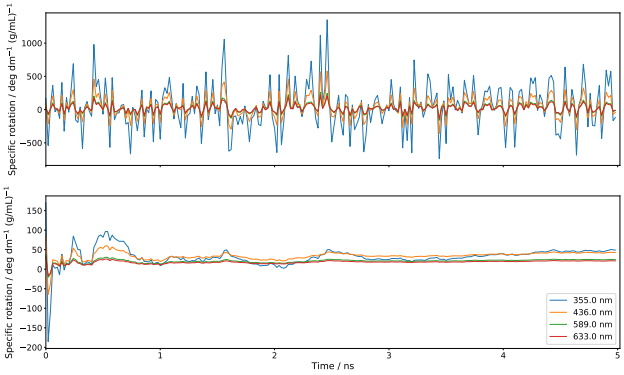


Figure A.20: Specific rotation at the CAM-B3LYP/orp level of theory of (*R*)-methylthiirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

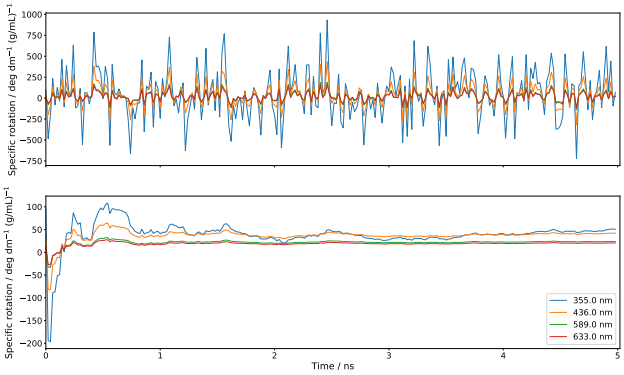


Figure A.21: Specific rotation at the CAM-B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

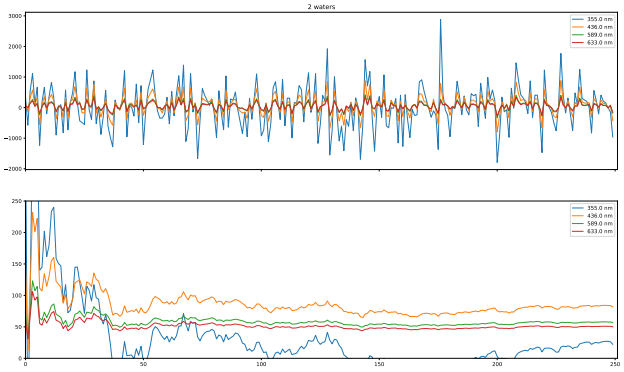


Figure A.22: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

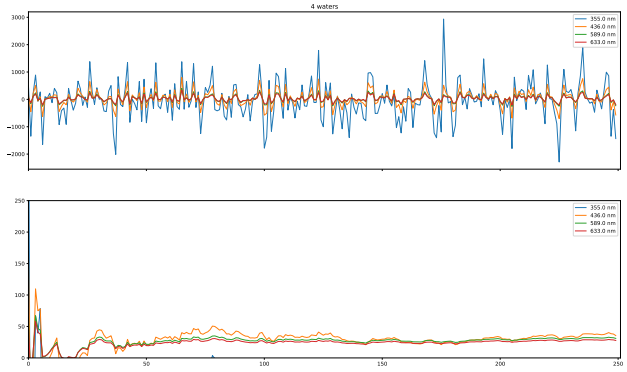


Figure A.23: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

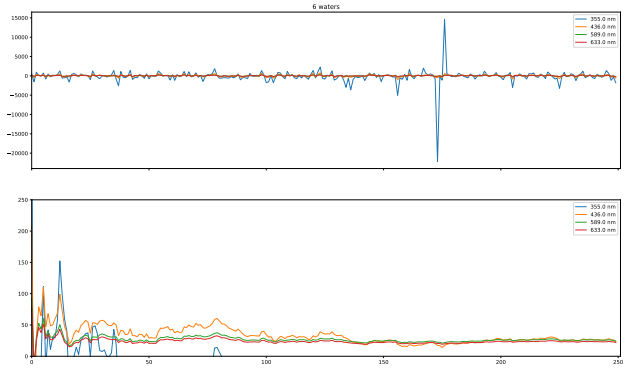


Figure A.24: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

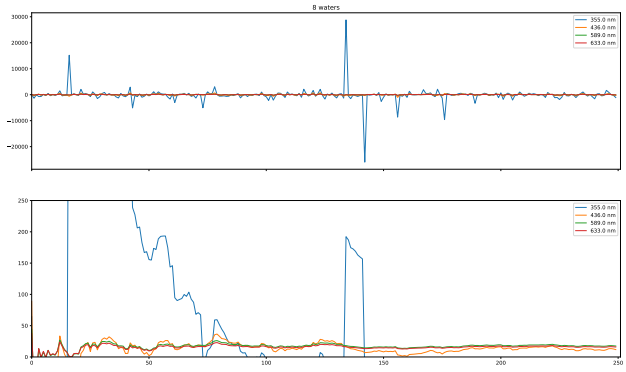


Figure A.25: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

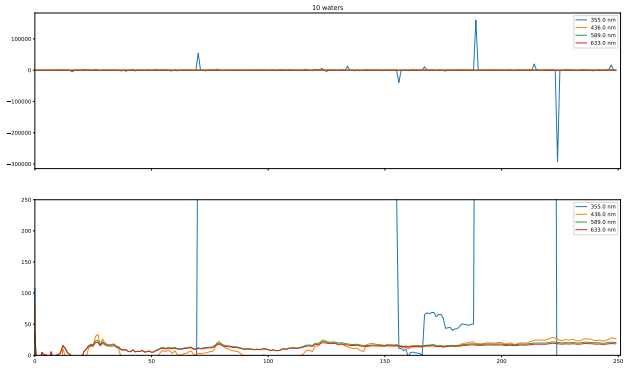


Figure A.26: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

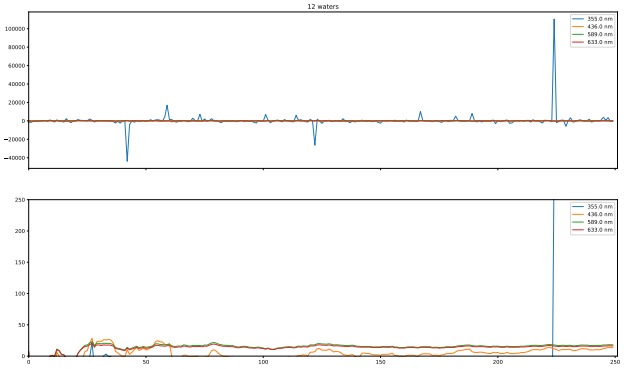


Figure A.27: Specific rotation at the B3LYP/ORP level of theory of (*R*)-methylthiirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

(*S*)-methyloxirane in water

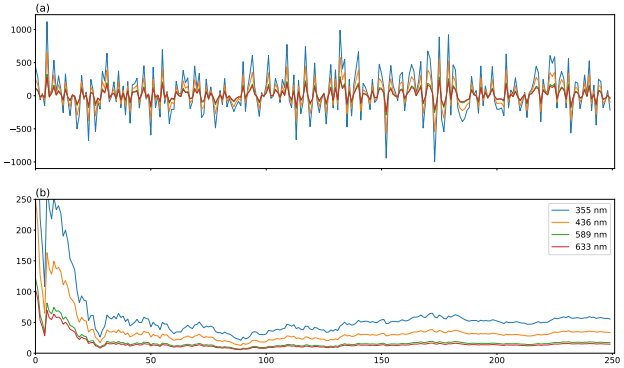


Figure A.28: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

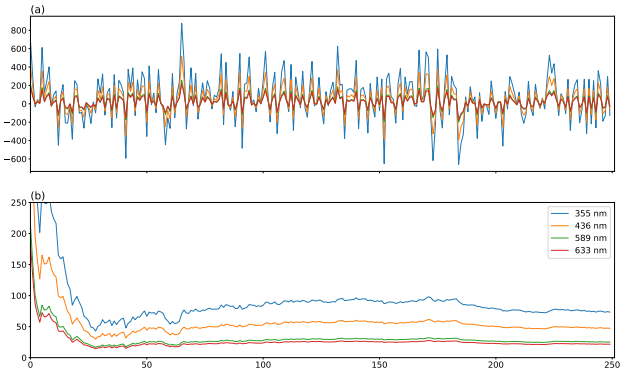


Figure A.29: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

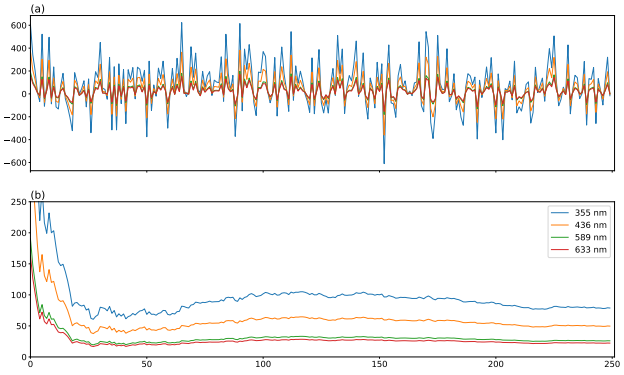


Figure A.30: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

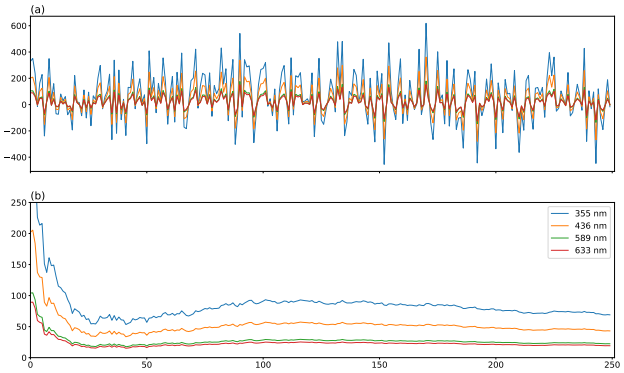


Figure A.31: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

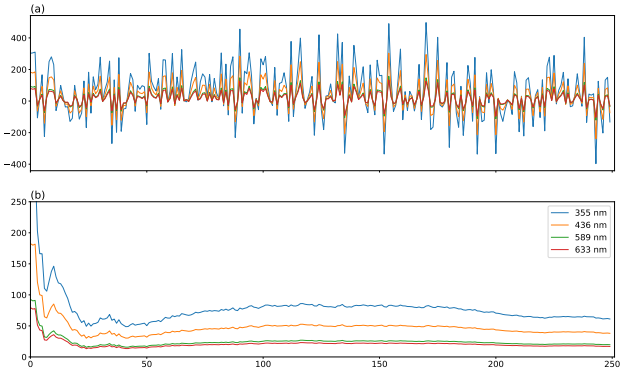


Figure A.32: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

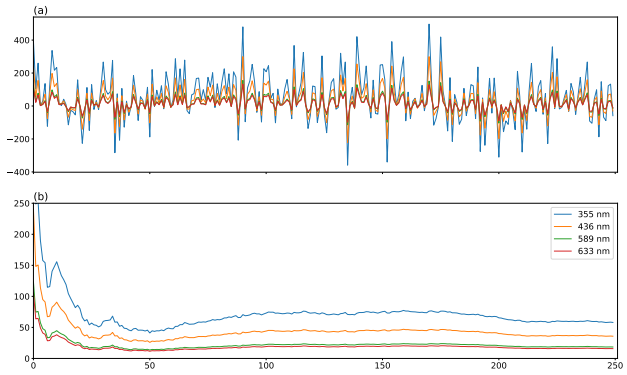


Figure A.33: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

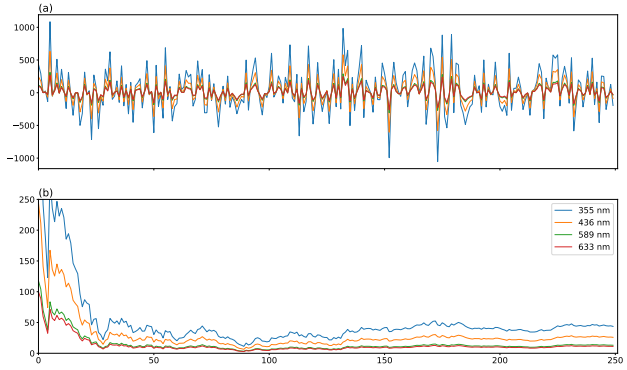


Figure A.34: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 2 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

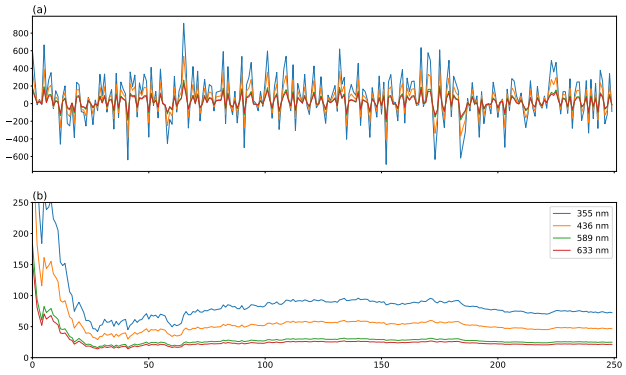


Figure A.35: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 4 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

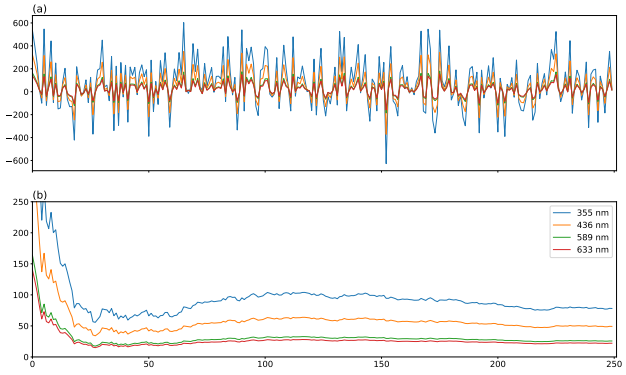


Figure A.36: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 6 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

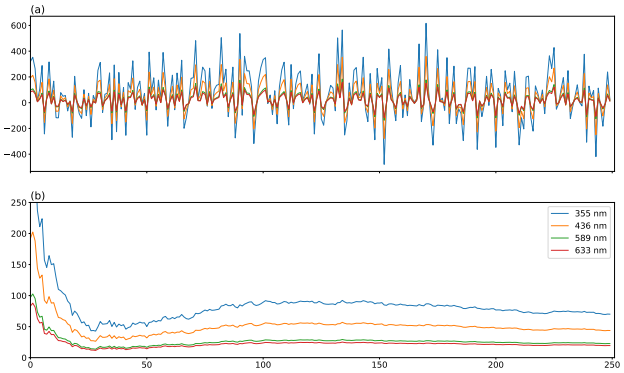


Figure A.37: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 8 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

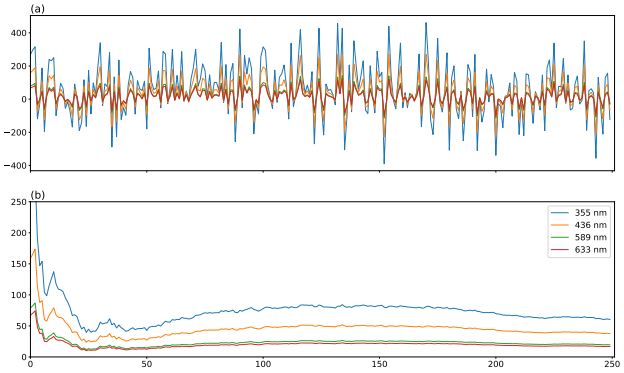


Figure A.38: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 10 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

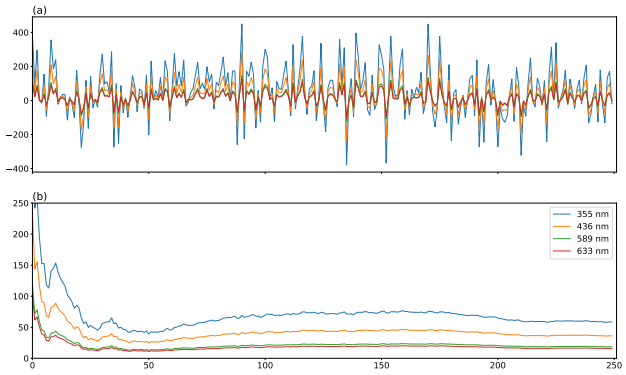


Figure A.39: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 12 water molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

(*S*)-methyloxirane in methanol

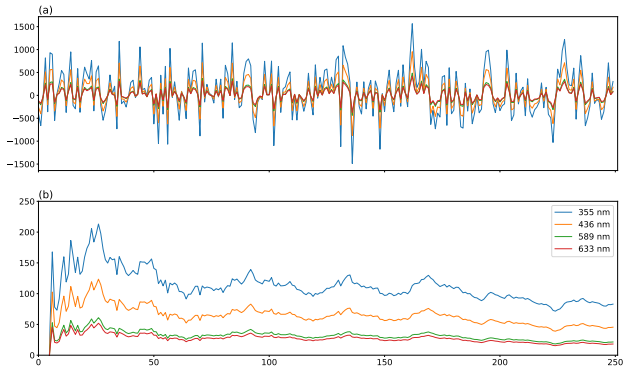


Figure A.40: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 2 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

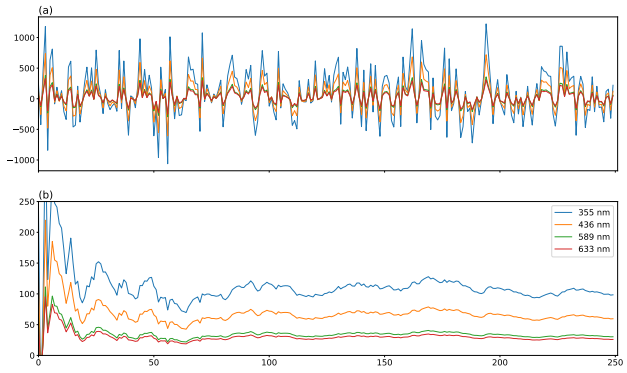


Figure A.41: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 4 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

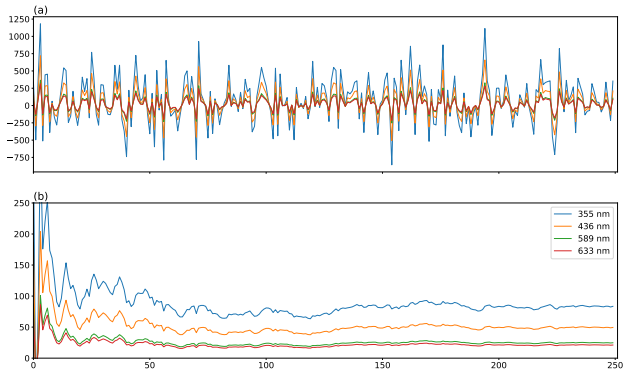


Figure A.42: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 6 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

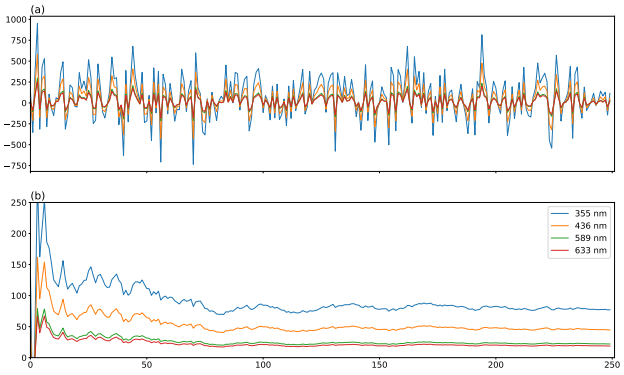


Figure A.43: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 8 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

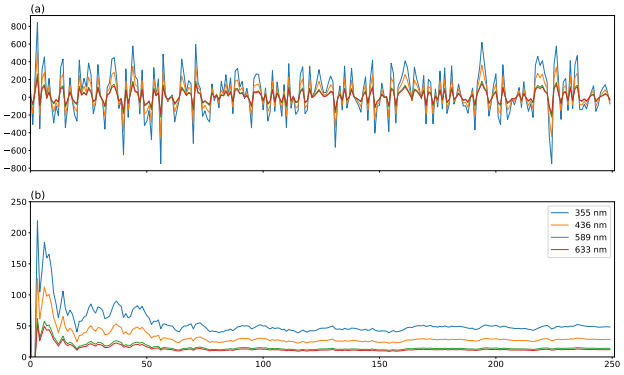


Figure A.44: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 10 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

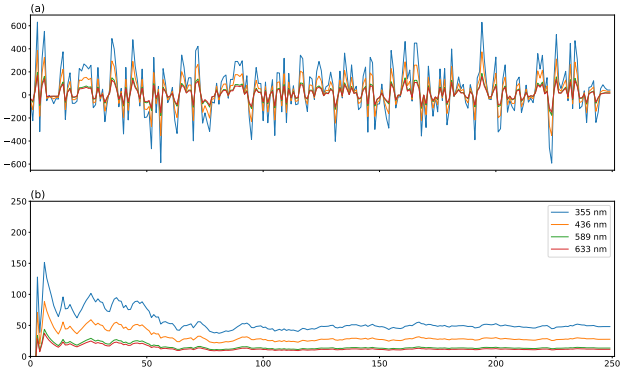


Figure A.45: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 12 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

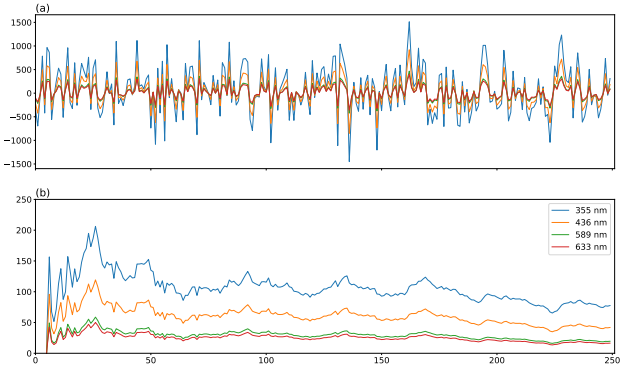


Figure A.46: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 2 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

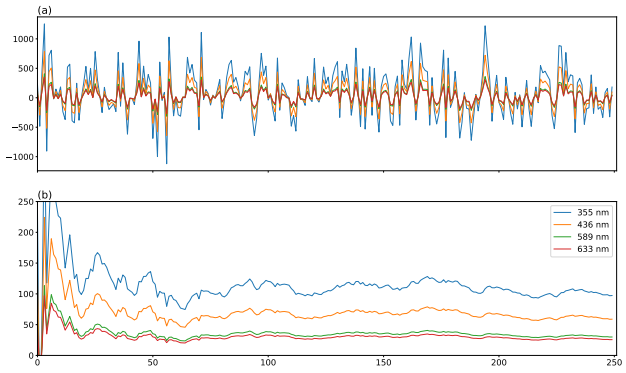


Figure A.47: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 4 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

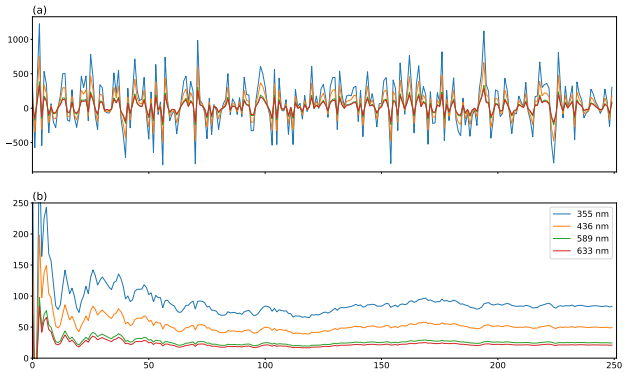


Figure A.48: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 6 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

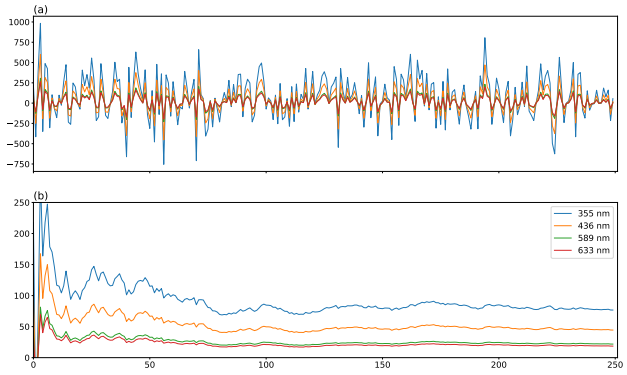


Figure A.49: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 8 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

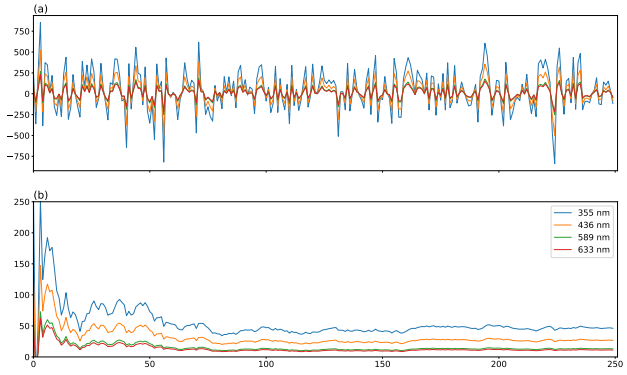


Figure A.50: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 10 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

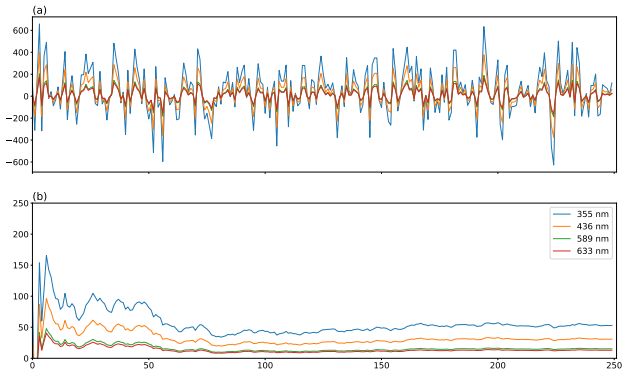


Figure A.51: Specific rotation at the CAM-B3LYP/aug-cc-pVDZ level of theory of (*S*)-methyloxirane in a cage of 12 methanol molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

(*S*)-methyloxirane in carbon tetrachloride

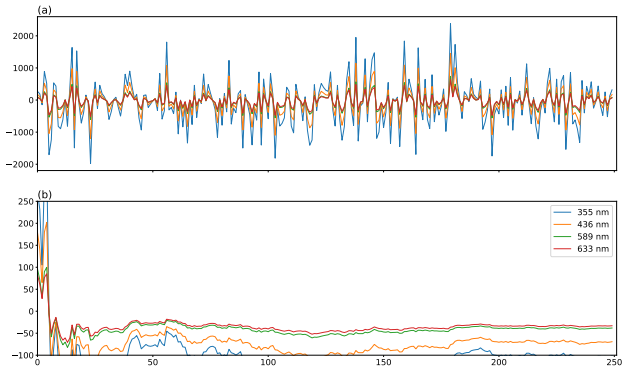


Figure A.52: Specific rotation at the CAM-B3LYP/ORP level of theory of (S)-methyloxirane in a cage of 2 CCl_4 molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

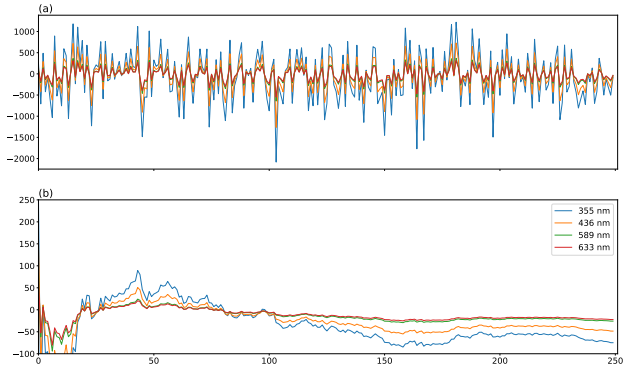


Figure A.53: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 4 CCl_4 molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

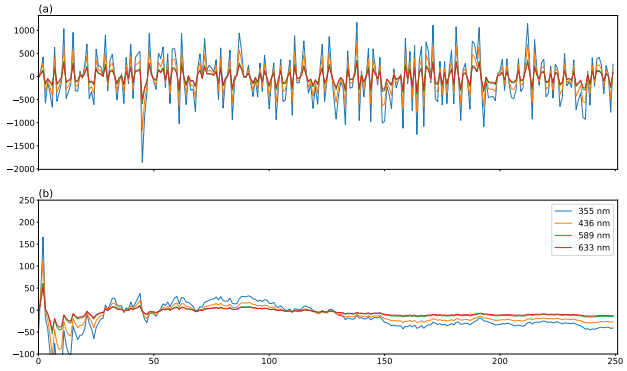


Figure A.54: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 6 CCl_4 molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

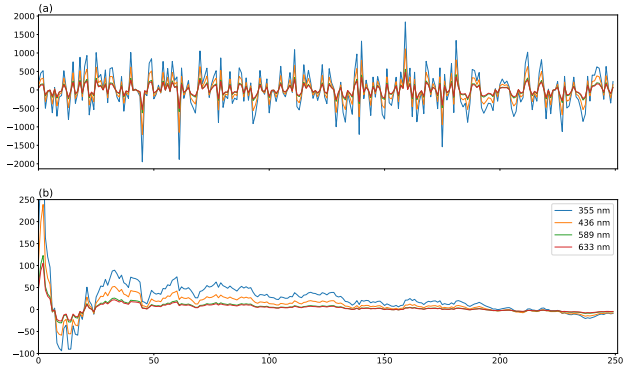


Figure A.55: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 8 CCl_4 molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

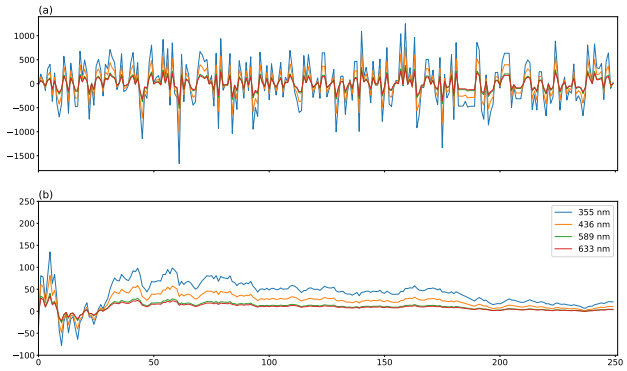


Figure A.56: Specific rotation at the CAM-B3LYP/ORP level of theory of (*S*)-methyloxirane in a cage of 10 CCl_4 molecules sampled over 250 evenly-spaced snapshots along a 5 ns MD trajectory: (a) specific rotation at each snapshot and (b) the cumulative running average.

A.2.3 Convergence with Number of Snapshots

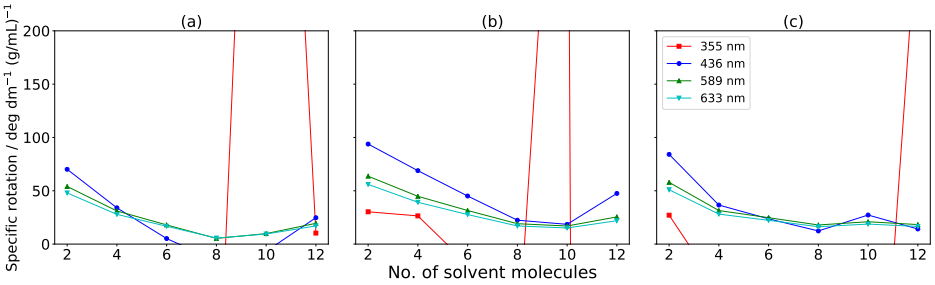


Figure A.57: Average B3LYP/ORP specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

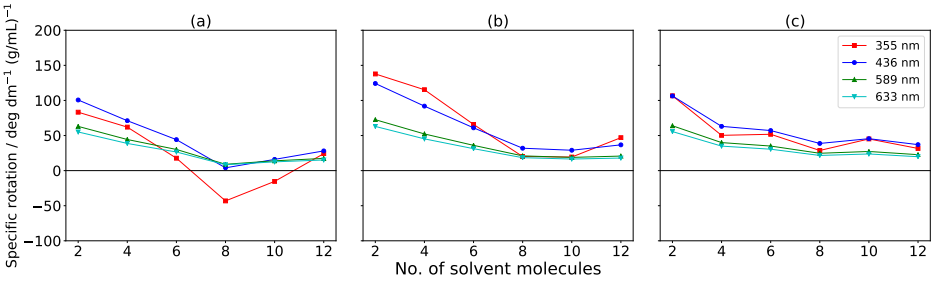


Figure A.58: Average B3LYP/ORP specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

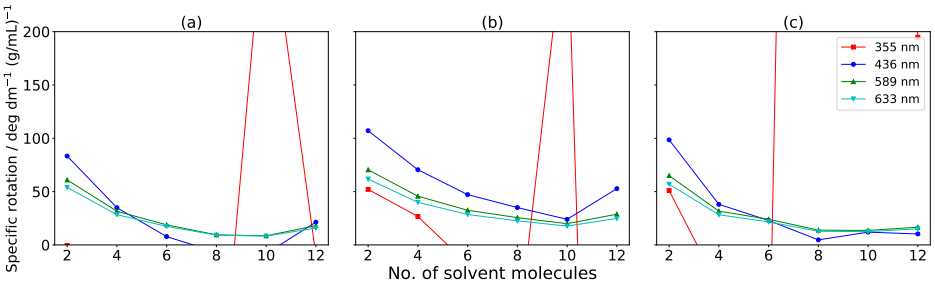


Figure A.59: Average B3LYP/aug-cc-pVDZ specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

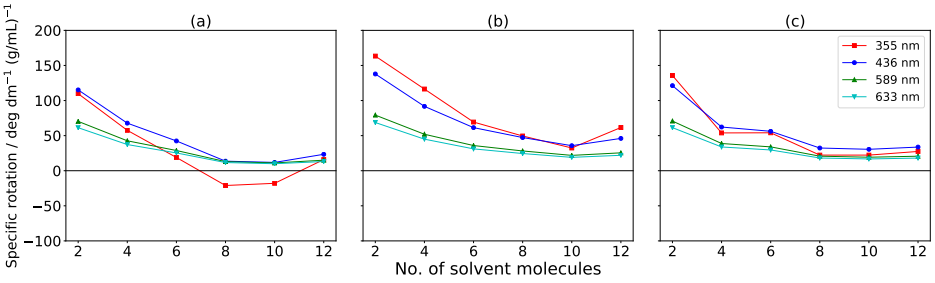


Figure A.60: Average B3LYP/aug-cc-pVDZ specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

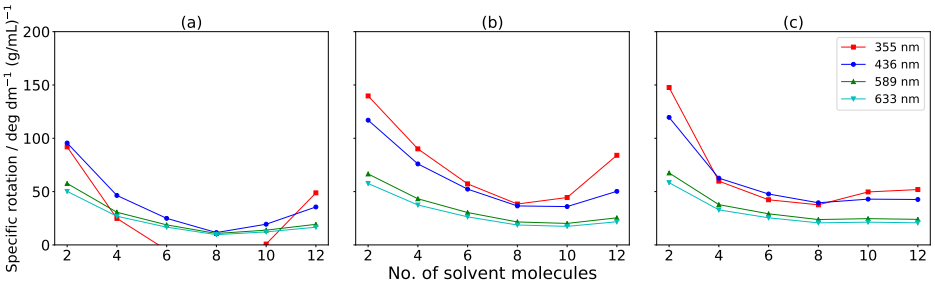


Figure A.61: Average CAM-B3LYP/ORP specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

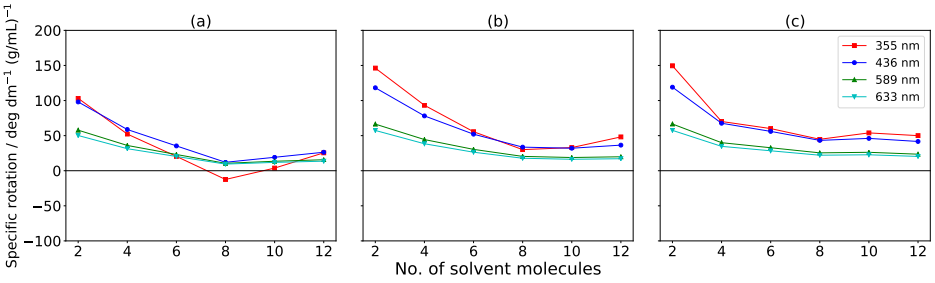


Figure A.62: Average CAM-B3LYP/ORP specific rotations of (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

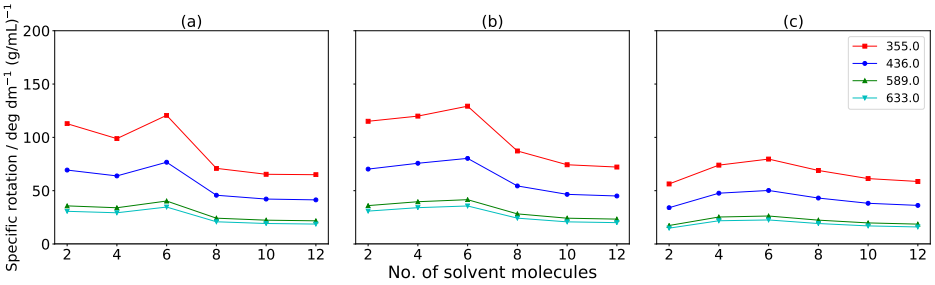


Figure A.63: Average CAM-B3LYP/ORP specific rotations of (*S*)-methyloxirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

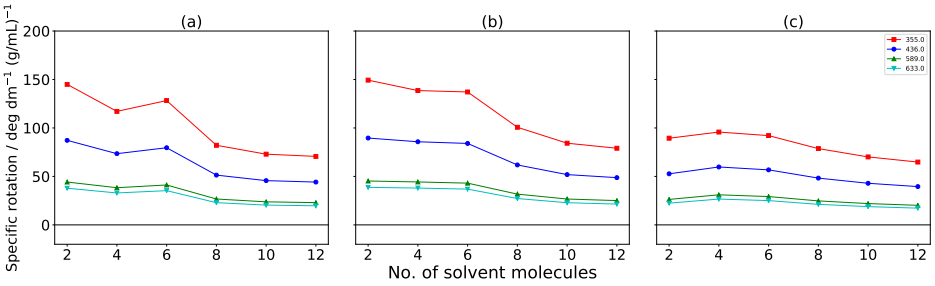


Figure A.64: Average CAM-B3LYP/ORP specific rotations of (*S*)-methyloxirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

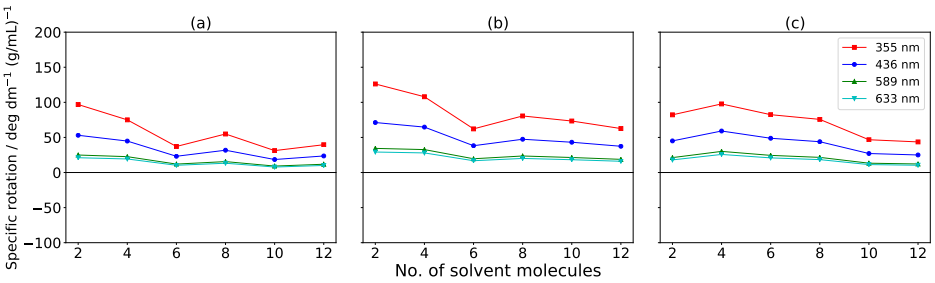


Figure A.65: Average CAM-B3LYP/ORP specific rotations of (*S*)-methyloxirane in methanol for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

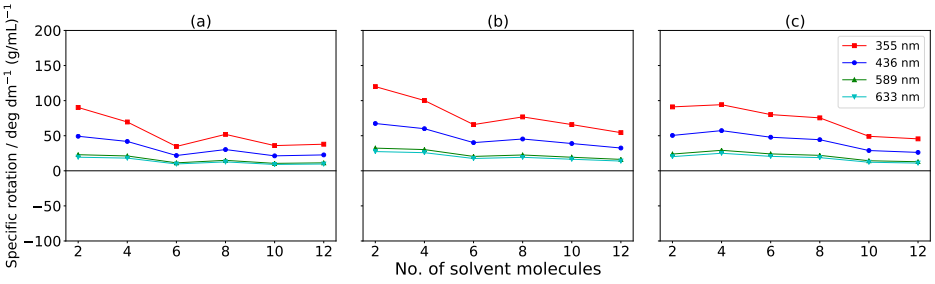


Figure A.66: Average CAM-B3LYP/ORP specific rotations of (*S*)-methyloxirane in methanol for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

A.2.4 Solvent-only plots

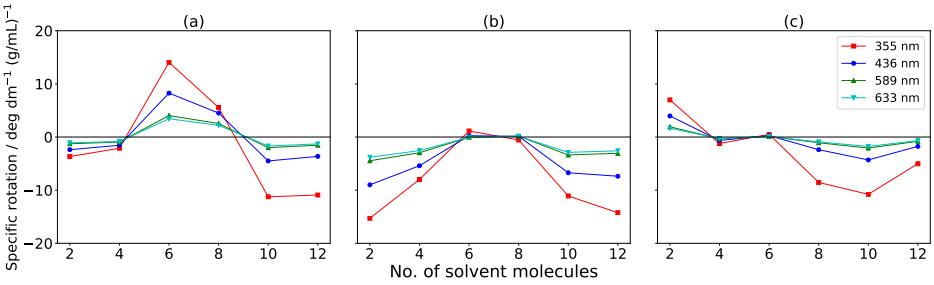


Figure A.67: Average B3LYP/aug-cc-pVDZ specific rotations of solvent-only for (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

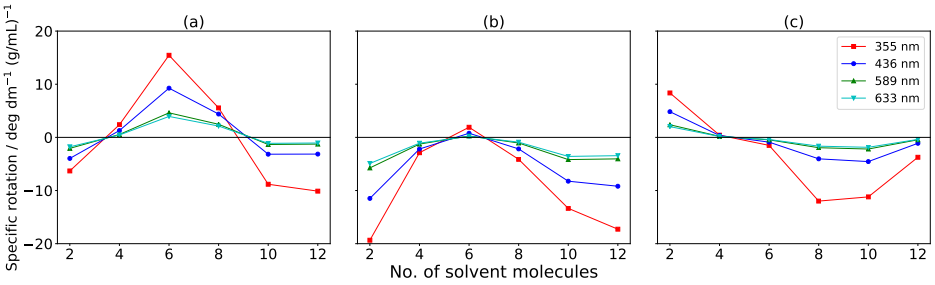


Figure A.68: Average B3LYP/ORP specific rotations of solvent-only for (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

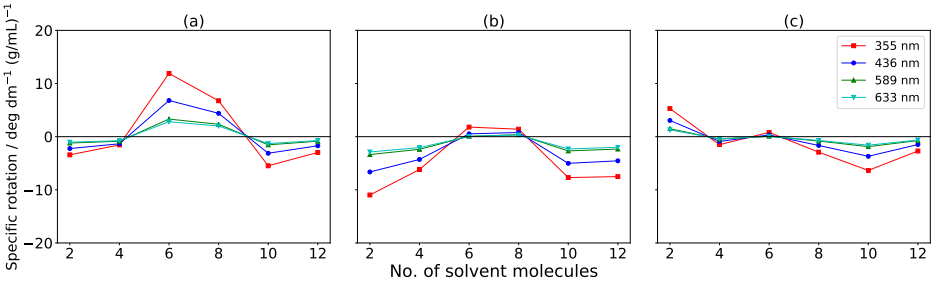


Figure A.69: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of solvent-only for (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

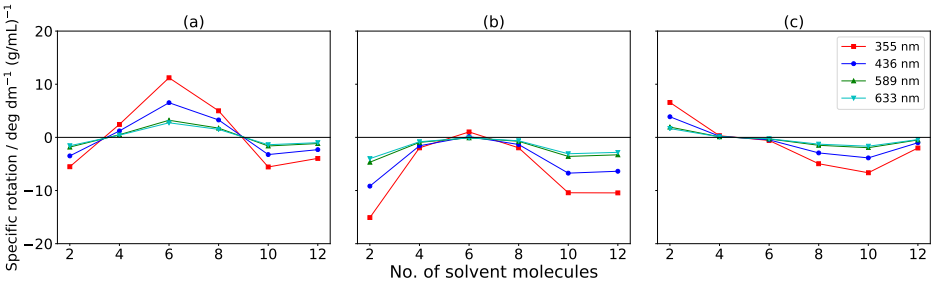


Figure A.70: Average CAM-B3LYP/ORP specific rotations of solvent-only for (*R*)-methylthiirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

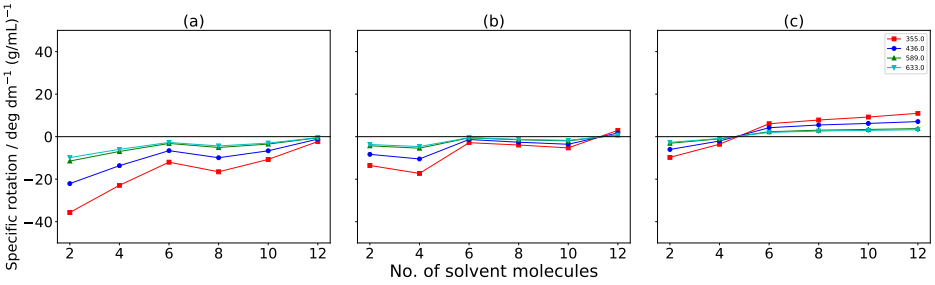


Figure A.71: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of solvent-only for (*S*)-methyloxirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

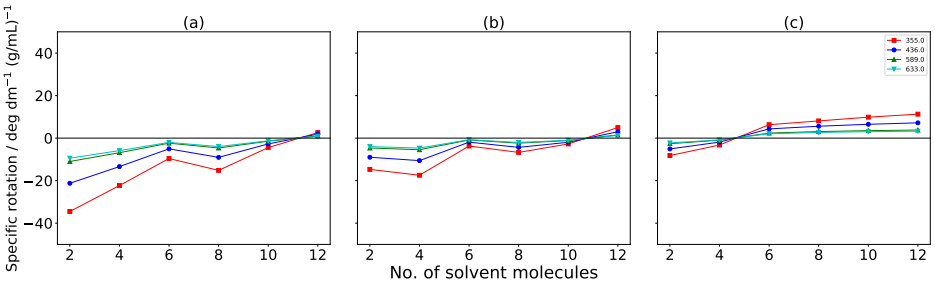


Figure A.72: Average CAM-B3LYP/ORP specific rotations of solvent-only for (*S*)-methyloxirane in water for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

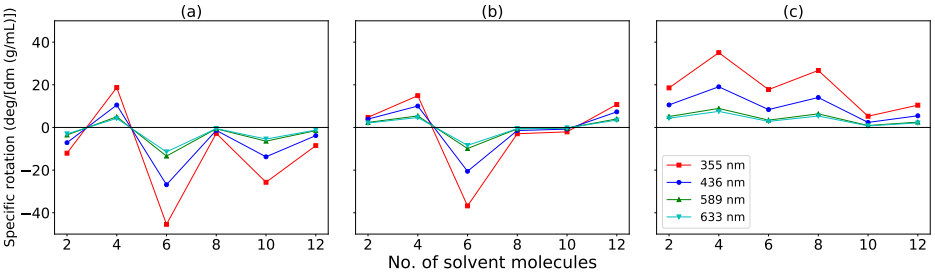


Figure A.73: Average CAM-B3LYP/aug-cc-pVDZ specific rotations of solvent-only for (*S*)-methyloxirane in methanol for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained along the MD trajectory.

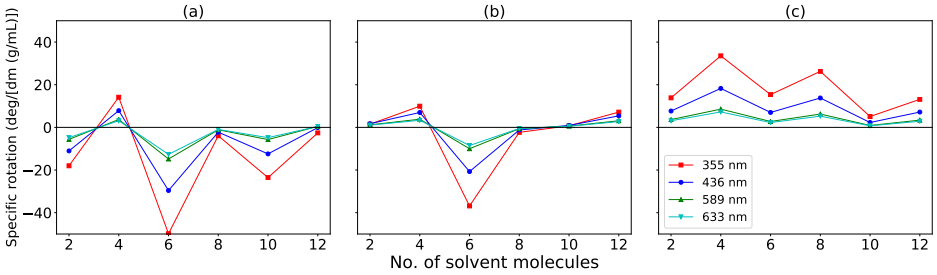


Figure A.74: Average CAM-B3LYP/ORP specific rotations of solvent-only for (*S*)-methyloxirane in methanol for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

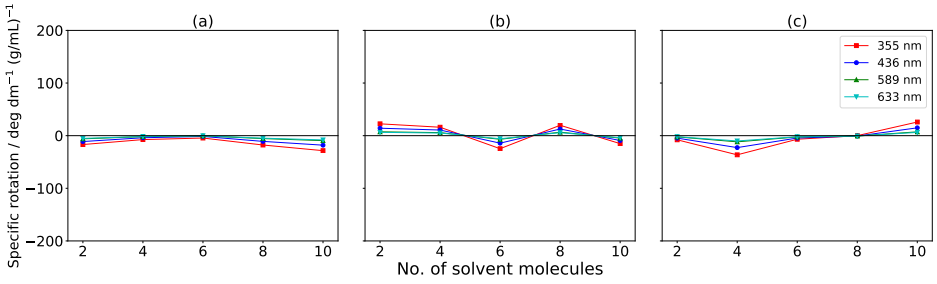


Figure A.75: Average CAM-B3LYP/ORP specific rotations of solvent-only for (*S*)-methyloxirane in CCl_4 for several wavelengths of light averaged over (a) 50 (b) 100 and (c) 250 snapshots as a function of the number of solvent molecules closest to the solute retained using the PCM implicit solvent model along the MD trajectory.

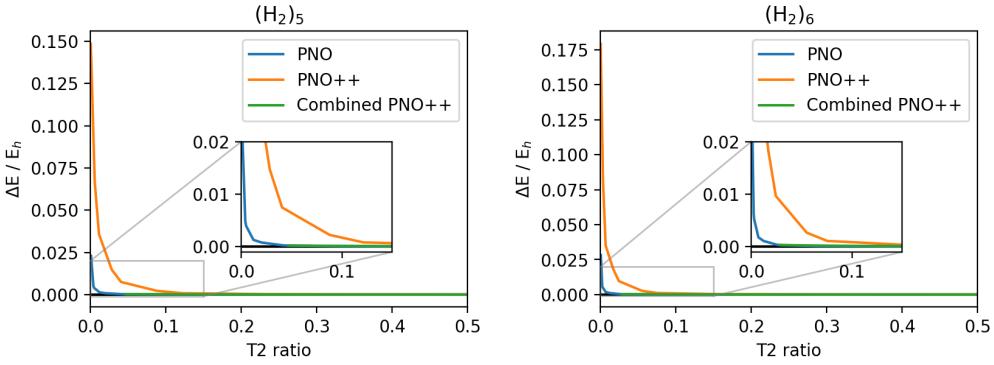
Appendix B

PNO++: Supporting Information

Molecular geometries used, CCSD energy, polarizability and specific rotation data for additional systems, comparisons of CCSD energy, polarizability and specific rotation data for different representations of the electric dipole moment operator, PNO++ data using the magnetic dipole moment operator, T_2 ratios as a function of T_{cut} thresholds.

Table B.1: Atomic coordinates of $(H_2)_4$ (Ångstroms)

Atomic Symbol	X	Y	Z
H	0.000000	0.000000	0.000000
H	0.750000	0.000000	0.000000
H	0.000000	1.500000	0.000000
H	0.375000	1.500000	-0.649520
H	0.000000	3.000000	0.000000
H	-0.375000	3.000000	-0.649520
H	0.000000	4.500000	-0.000000
H	-0.750000	4.500000	-0.000000



(a) (b)

Figure B.1: Errors in CCSD correlation energies for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

Table B.2: Atomic coordinates of $(H_2)_5$ (Ångstroms)

Atomic Symbol	X	Y	Z
H	0.000000	0.000000	0.000000
H	0.750000	0.000000	0.000000
H	0.000000	1.500000	0.000000
H	0.375000	1.500000	-0.649520
H	0.000000	3.000000	0.000000
H	-0.375000	3.000000	-0.649520
H	0.000000	4.500000	-0.000000
H	-0.750000	4.500000	-0.000000
H	0.000000	6.000000	-0.000000
H	-0.375000	6.000000	0.649520

Table B.3: Atomic coordinates of $(H_2)_6$ (Ångstroms)

Atomic Symbol	X	Y	Z
H	0.000000	0.000000	0.000000
H	0.750000	0.000000	0.000000
H	0.000000	1.500000	0.000000
H	0.375000	1.500000	-0.649520
H	0.000000	3.000000	0.000000
H	-0.375000	3.000000	-0.649520
H	0.000000	4.500000	-0.000000
H	-0.750000	4.500000	-0.000000
H	0.000000	6.000000	-0.000000
H	-0.375000	6.000000	0.649520
H	0.000000	7.500000	-0.000000
H	0.375000	7.500000	0.649520

Table B.4: Atomic coordinates of $(H_2)_7$ (Ångstroms)

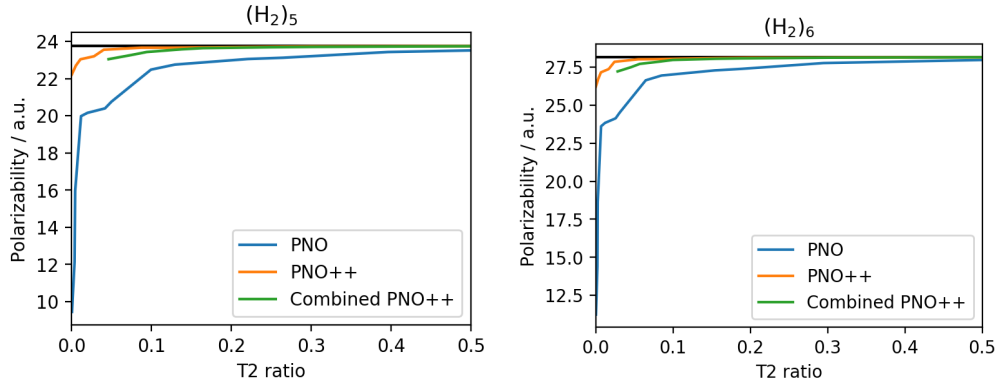
Atomic Symbol	X	Y	Z
H	0.000000	0.000000	0.000000
H	0.750000	0.000000	0.000000
H	0.000000	1.500000	0.000000
H	0.375000	1.500000	-0.649520
H	0.000000	3.000000	0.000000
H	-0.375000	3.000000	-0.649520
H	0.000000	4.500000	-0.000000
H	-0.750000	4.500000	-0.000000
H	0.000000	6.000000	-0.000000
H	-0.375000	6.000000	0.649520
H	0.000000	7.500000	-0.000000
H	0.375000	7.500000	0.649520
H	-0.000000	9.000000	-0.000000
H	0.750000	9.000000	0.000000

Table B.5: Atomic coordinates of 1,3-dimethylallene (Ångstroms)

Atomic Symbol	X	Y	Z
C	-0.000002	-0.414018	-0.000082
C	1.246074	-0.412429	-0.389815
C	-1.245909	-0.411914	0.390189
H	-1.537390	-1.089758	1.192281
C	2.335357	0.460029	0.185131
H	1.537956	-1.091281	-1.190907
C	-2.335503	0.459710	-0.185425
H	1.951298	1.101018	0.979882
H	2.774272	1.096860	-0.590312
H	3.146275	-0.150721	0.596180
H	-3.146688	-0.151602	-0.595103
H	-1.951888	1.099517	-0.981343
H	-2.773938	1.097702	0.589337

Table B.6: Atomic coordinates of H_2O_2 (Ångstroms)

Atomic Symbol	X	Y	Z
O	-0.182400	-0.692195	-0.031109
O	0.182400	0.692195	-0.031109
H	0.533952	-1.077444	0.493728
H	-0.533952	1.077444	0.493728

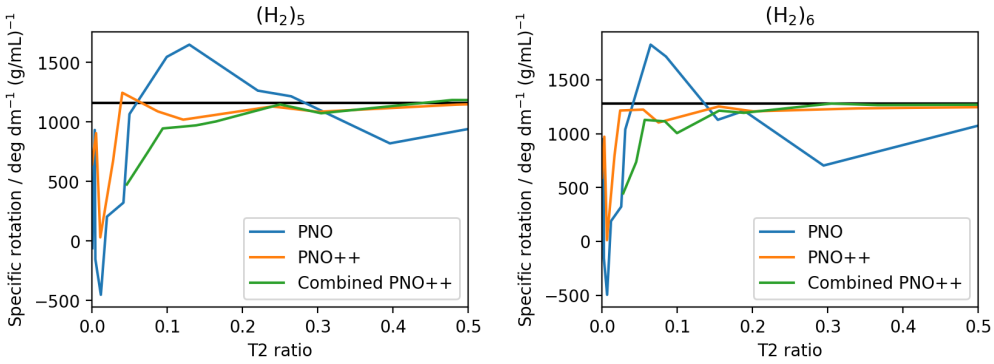


(a) (b)

Figure B.2: CCSD dynamic polarizabilities at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

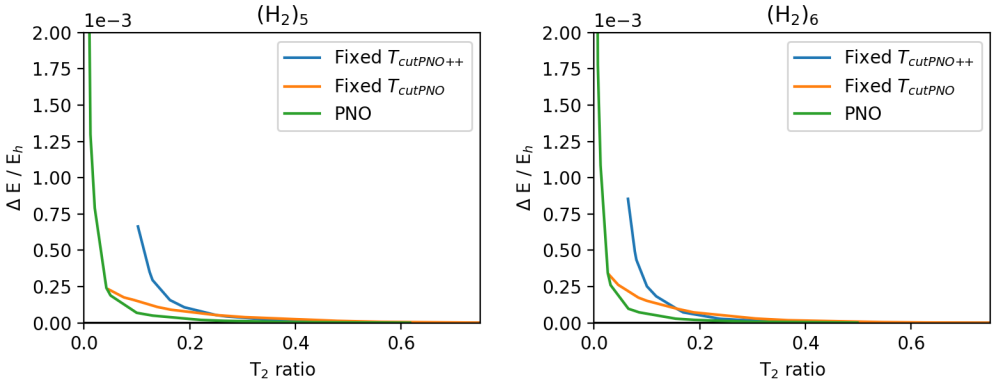
Table B.7: Atomic coordinates of 1-fluoro-heptane (Ångstroms)

Atomic Symbol	X	Y	Z
C	70.008200	10.041100	0.000100
C	68.692300	9.251900	-0.000100
C	68.891800	7.731300	0.000100
C	69.800600	11.560900	-0.000100
C	67.576600	6.941300	-0.000100
C	71.109600	12.333200	0.000100
C	67.782300	5.423200	0.000100
F	70.852500	13.702900	-0.000100
H	70.603100	9.752000	0.879200
H	70.603500	9.751800	-0.878600
H	68.097100	9.540800	-0.878800
H	68.096800	9.541000	0.878300
H	69.487600	7.442400	0.878900
H	69.488000	7.442200	-0.878400
H	69.219200	11.862900	-0.881100
H	69.218800	11.863100	0.880600
H	66.981500	7.229700	-0.878300
H	66.981100	7.229900	0.877800
H	71.708900	12.099700	0.890300
H	71.709300	12.099500	-0.889700



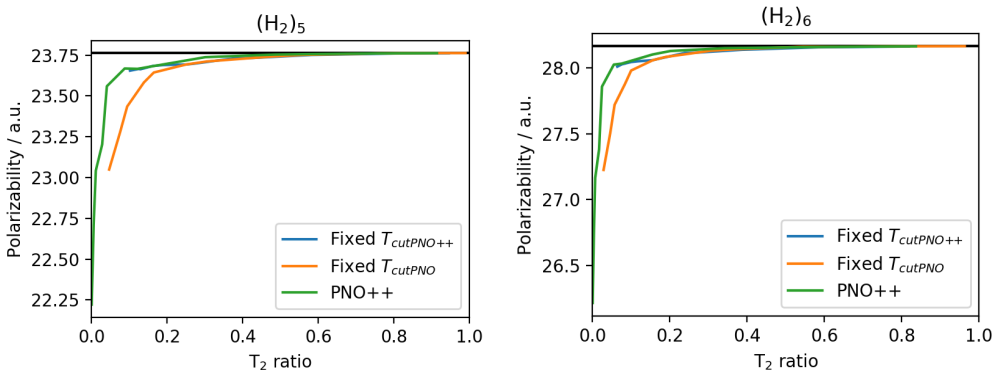
(a) (b)

Figure B.3: CCSD specific rotations at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.



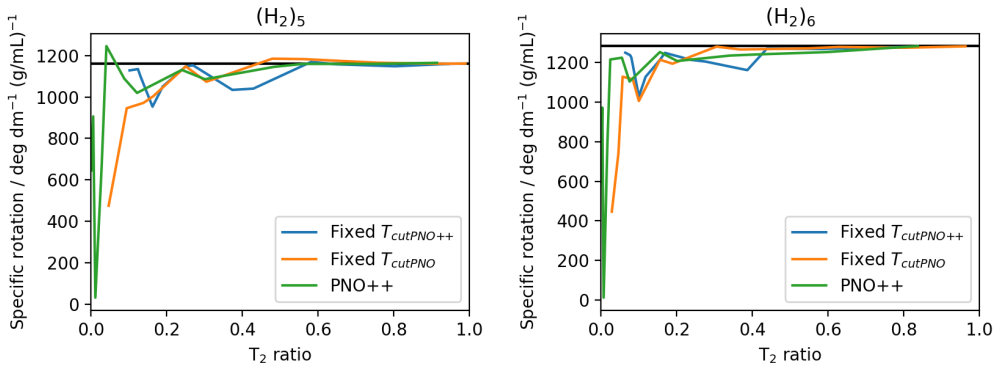
(a) (b)

Figure B.4: CCSD correlation energies for the combined PNO++ methods at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.



(a) (b)

Figure B.5: CCSD correlation energies for the combined PNO++ methods at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.



(a) (b)

Figure B.6: CCSD specific rotations for the combined PNO++ methods at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

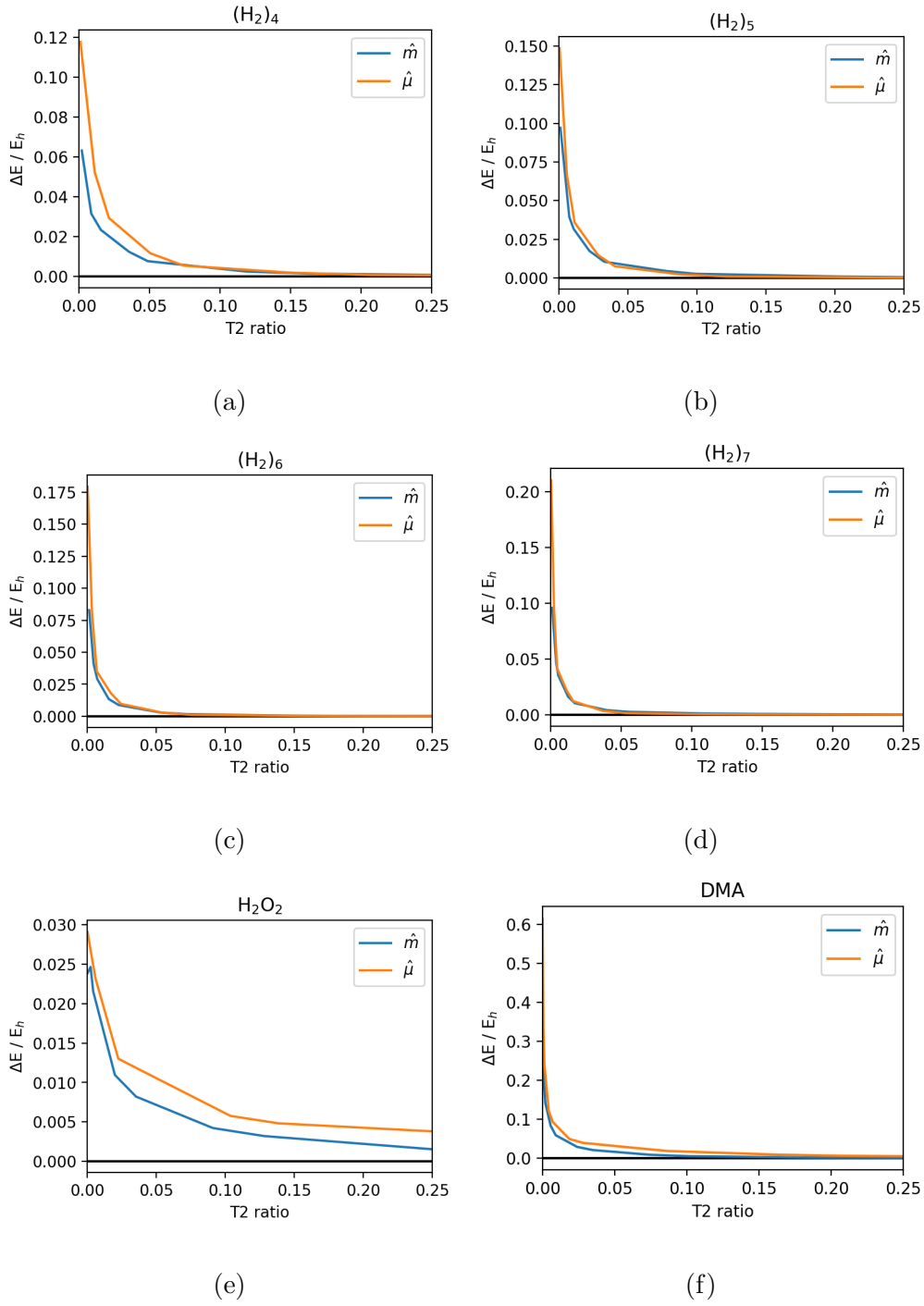


Figure B.7: CCSD correlation energies using the PNO++ method for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, (d) $(H_2)_7$, (e) H_2O_2, and (f) DMA systems, computed using the aug-cc-pVDZ basis set.

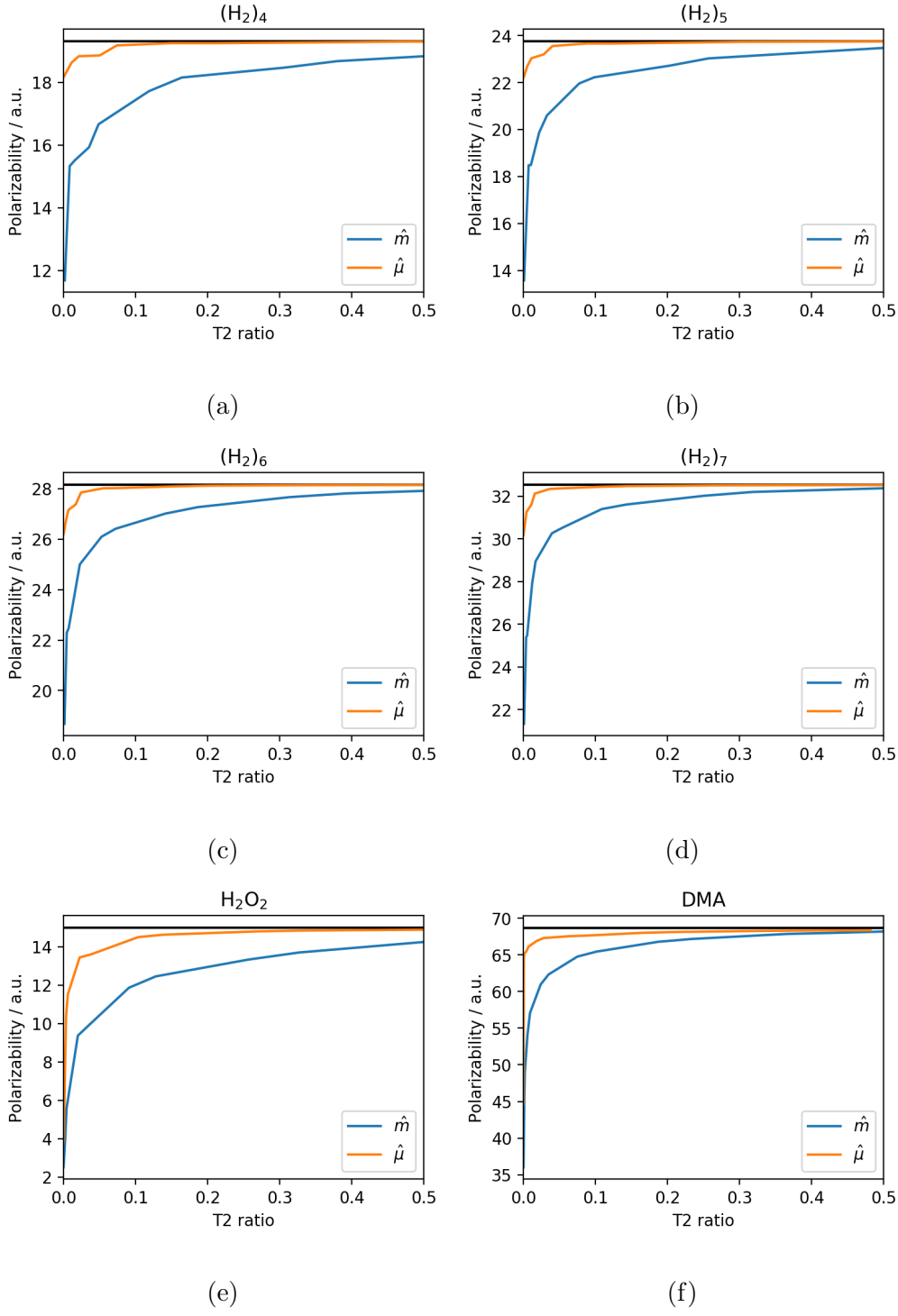


Figure B.8: CCSD dynamic polarizabilities using the PNO++ method at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, (d) $(H_2)_7$, (e) H_2O_2, and (f) DMA systems, computed using the aug-cc-pVDZ basis set.

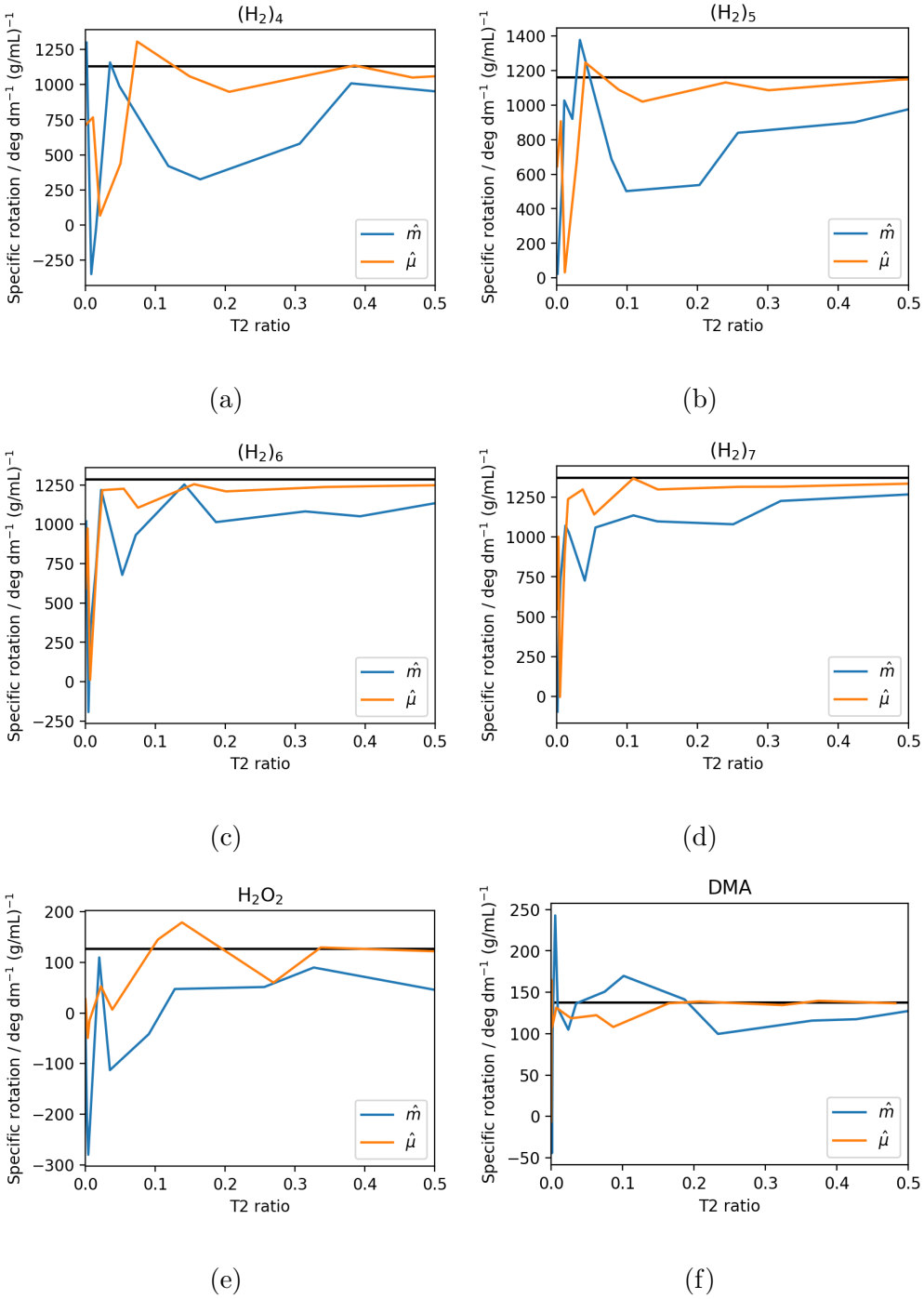


Figure B.9: CCSD specific rotations using the PNO++ method at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, (d) $(H_2)_7$, (e) H_2O_2, and (f) DMA systems, computed using the aug-cc-pVDZ basis set.

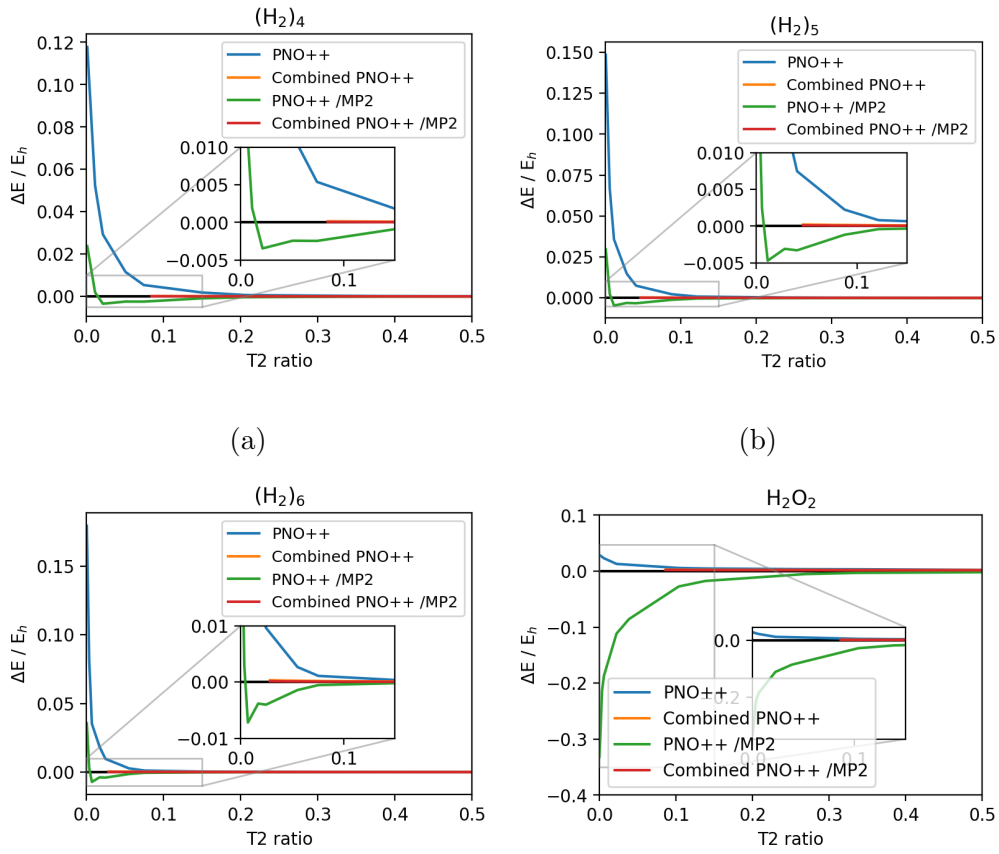


Figure B.10: CCSD specific rotations with and without MP2-level corrections at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

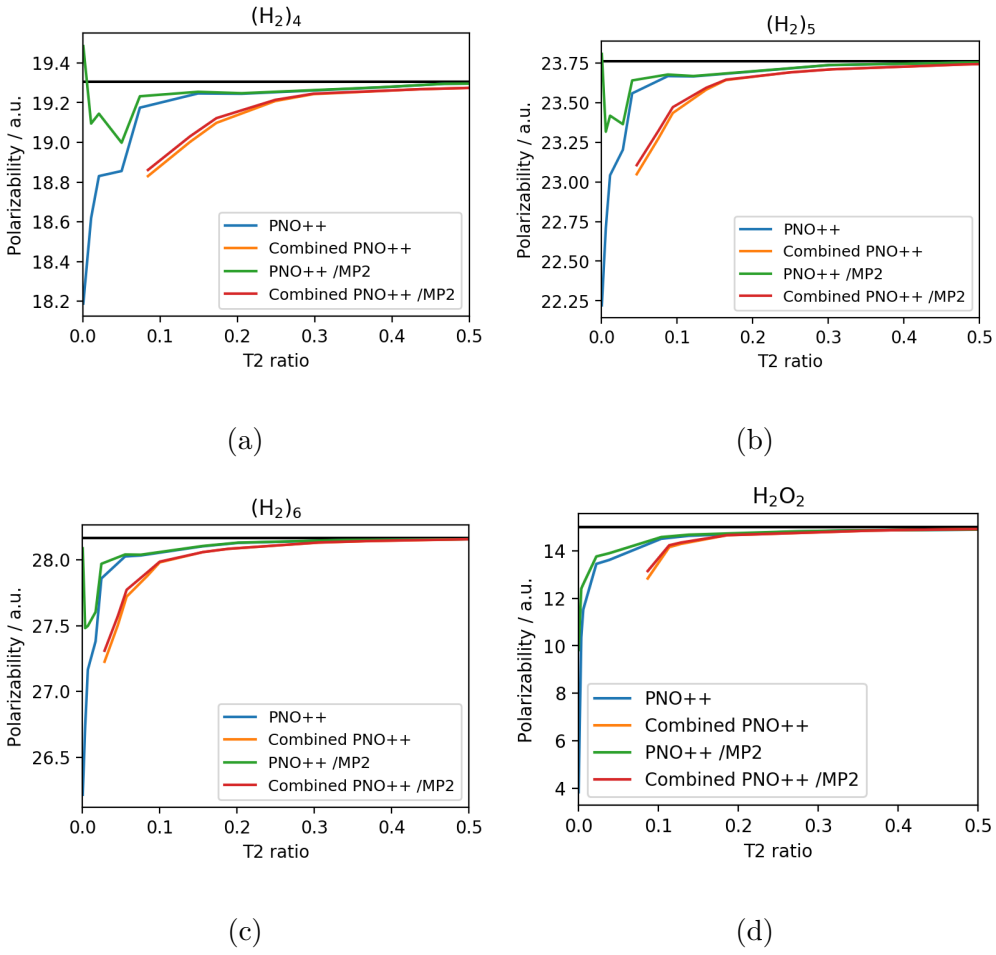


Figure B.11: CCSD specific rotations with and without MP2-level corrections at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

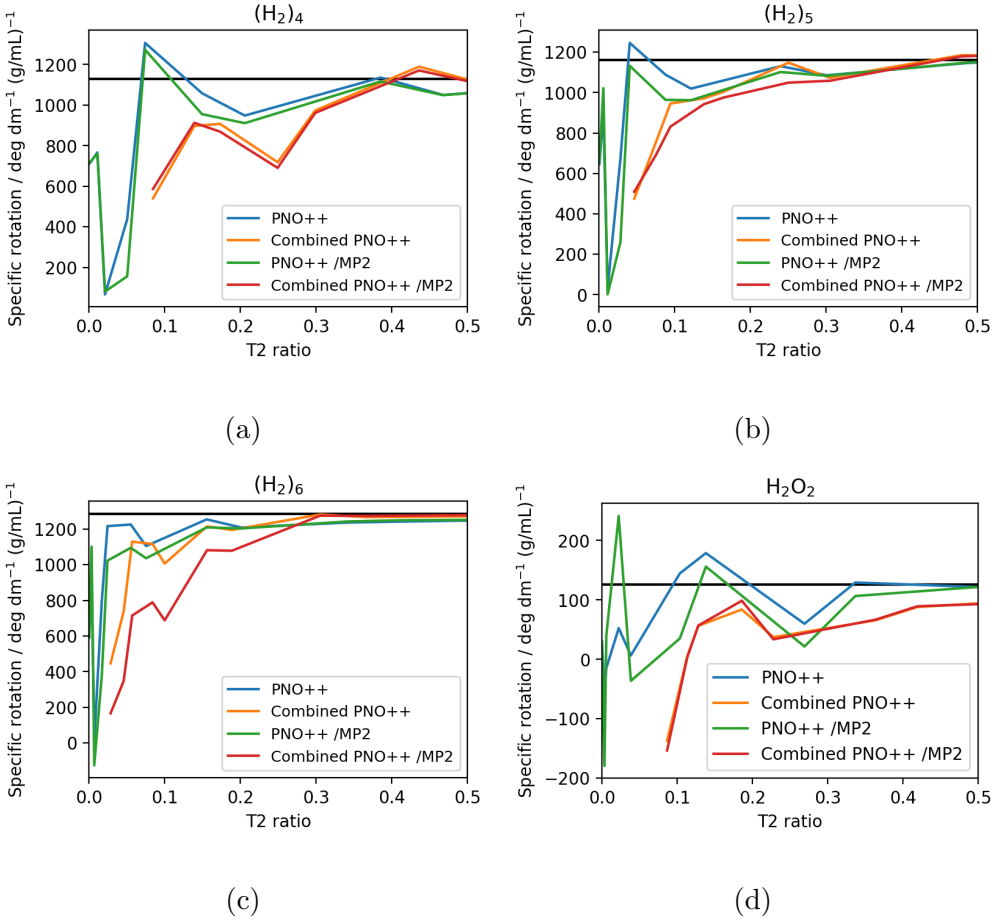


Figure B.12: CCSD specific rotations with and without MP2-level corrections at 589 nm for (a) $(H_2)_5$ and (b) $(H_2)_6$ systems, computed using the aug-cc-pVDZ basis set.

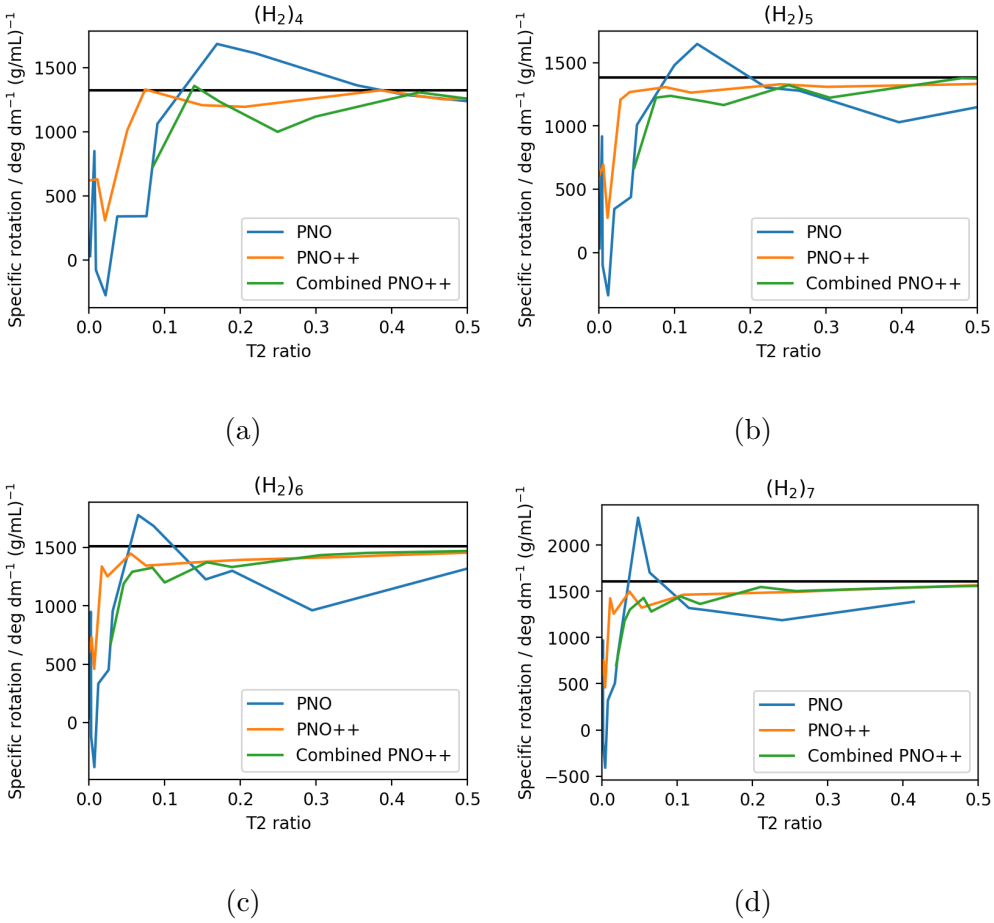
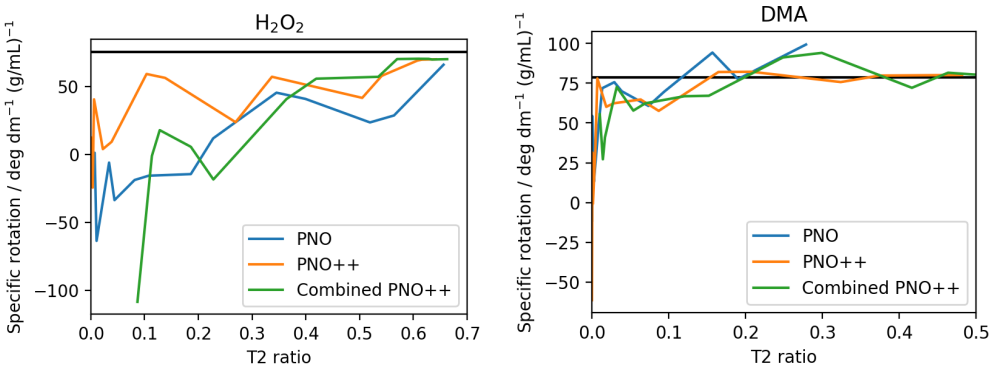


Figure B.13: CCSD specific rotations using the length gauge representation of the dipole moment operator at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, and (d) $(H_2)_7$ systems, computed using the aug-cc-pVDZ basis set.



(a) (b)

Figure B.14: CCSD specific rotations using the length gauge representation of the dipole moment operator at 589 nm for (a) H_2O_2 and (b) DMA systems, computed using the aug-cc-pVDZ basis set.

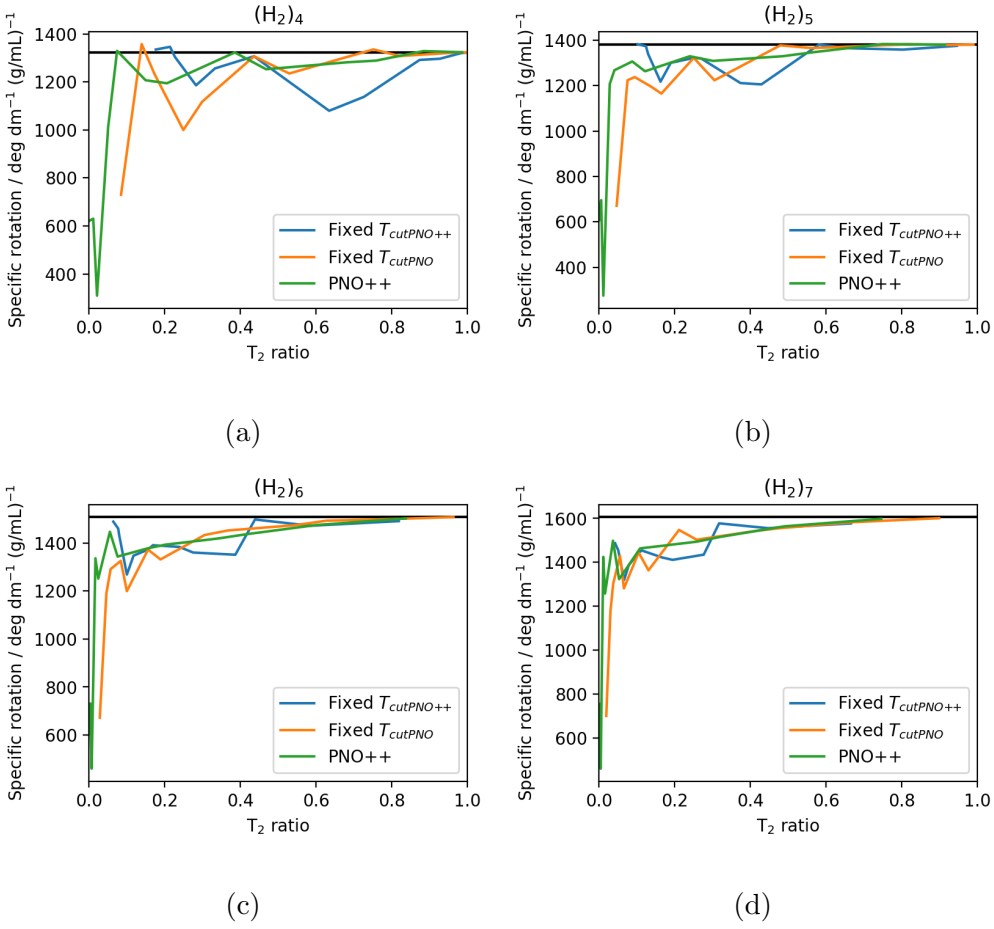
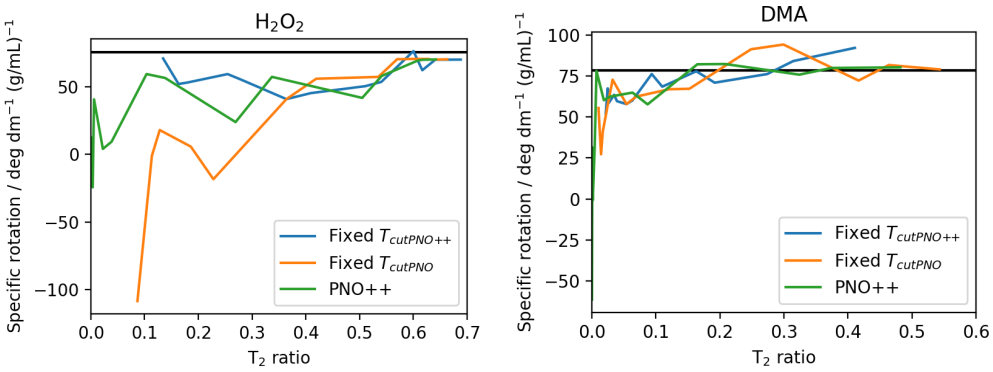


Figure B.15: CCSD specific rotations for the combined PNO++ methods using the length gauge representation of the dipole moment operator at 589 nm for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, and (d) $(H_2)_7$ systems, computed using the aug-cc-pVDZ basis set.



(a) (b)

Figure B.16: CCSD specific rotations for the combined PNO++ methods using the length gauge representation of the dipole moment operator at 589 nm for (a) H_2O_2 and (b) DMA systems, computed using the aug-cc-pVDZ basis set.

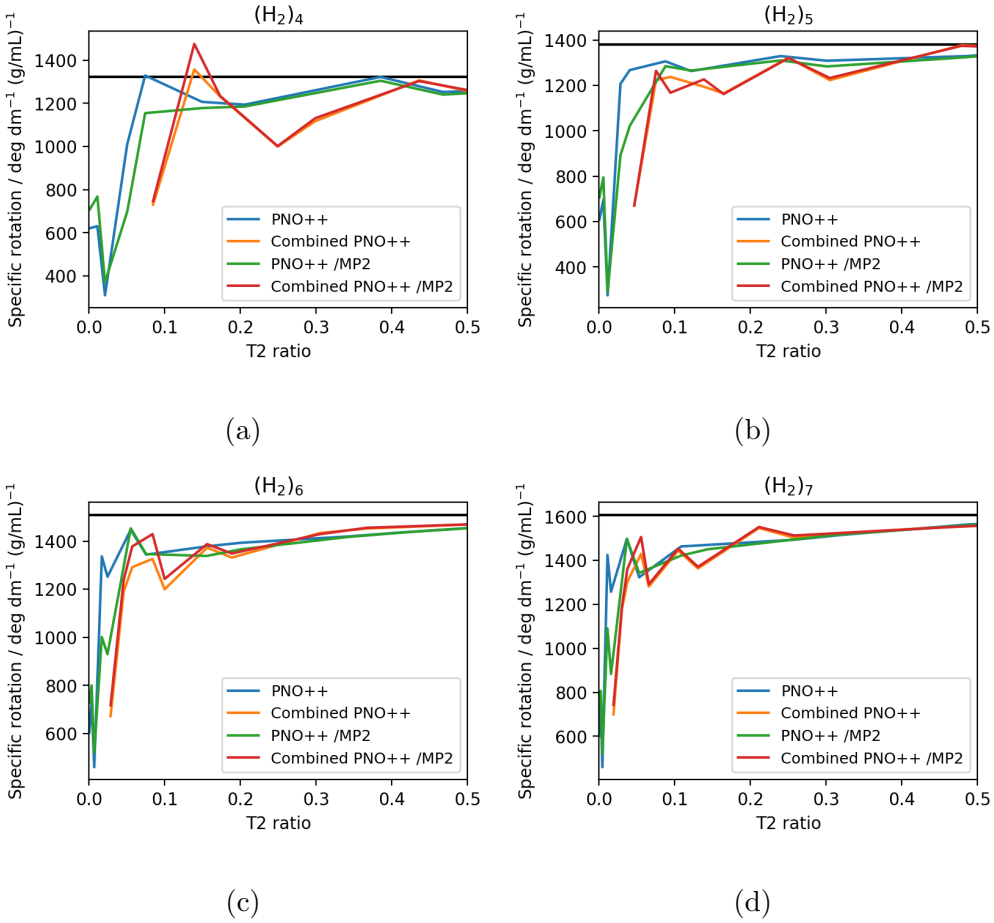
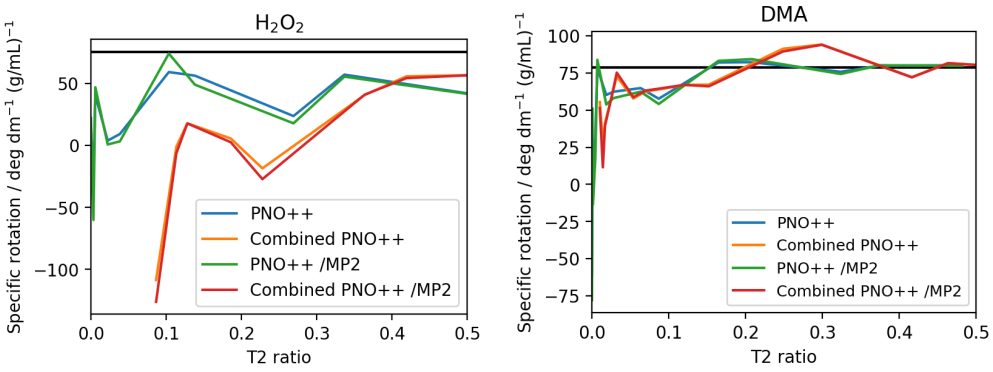


Figure B.17: CCSD specific rotations with and without MP2-level corrections using the length gauge representation of the dipole moment operator at 589 nm for ((a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, and (d) $(H_2)_7$ systems, computed using the aug-cc-pVDZ basis set.



(a) (b)

Figure B.18: CCSD specific rotations with and without MP2-level corrections using the length gauge representation of the dipole moment operator at 589 nm for (a) H_2O_2 and (b) DMA systems, computed using the aug-cc-pVDZ basis set.

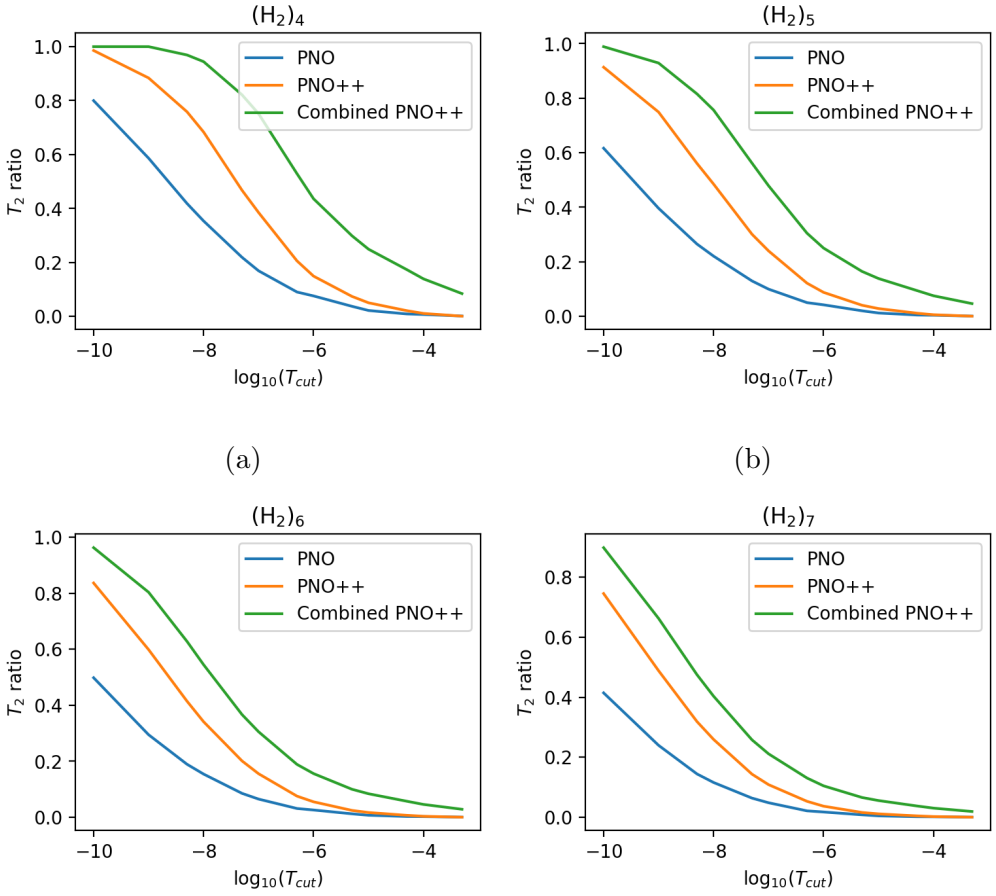
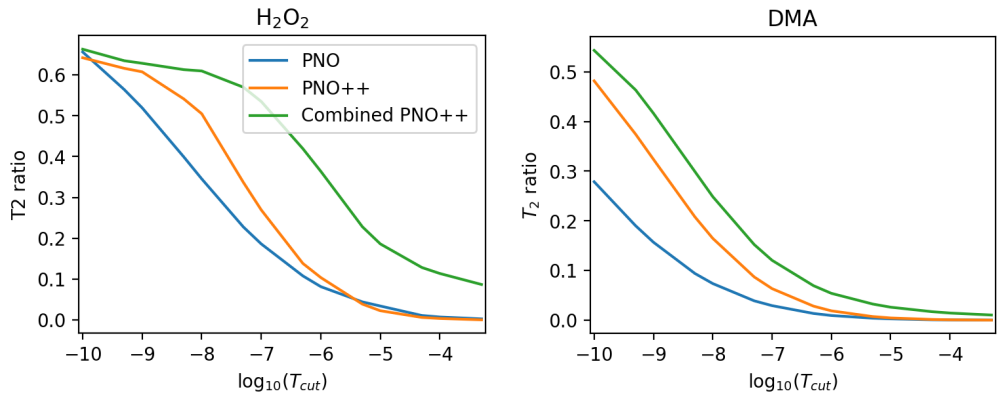


Figure B.19: T_2 ratios plotted as a function of T_{cut} thresholds for the PNO, PNO++ and combined PNO++ methods for (a) $(H_2)_4$, (b) $(H_2)_5$, (c) $(H_2)_6$, and (d) $(H_2)_7$ systems.



(a) (b)

Figure B.20: T_2 ratios plotted as a function of T_{cut} thresholds for the PNO, PNO++ and combined PNO++ methods for (a) H_2O_2 and (b) DMA systems.