Development and Application of Hybrid Methods

Sijin Ren

Submitted to the graduate degree program in Department of Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee members

Marco Caricato, Chairperson

James D. Blakemore

Krzysztof Kuczera

Kyle Camarda

Ward H. Thompson

Date defended: May 29, 2020

The Dissertation Committee for Sijin Ren certifies
that this is the approved version of the following dissertation :

Development and Application of Hybrid Methods

Marco Caricato, Chairperson

Date approved: June 5, 2020

Abstract

This thesis focuses on the development and application of hybrid methods for modeling excited state properties of large systems. The rapid growth of technology and computer processing power allows the wide application of accurate quantum mechanical (QM) methods to different areas of scientific research. However, more accurate computational methods are usually associated with higher computational cost that limits their application to small to medium sized systems. In many cases, modeling the environment at QM level is not possible, and yet completely neglecting the effect of the environment may not be a reasonable assumption. One popular approach for reducing computational cost is to use hybrid methods, which combine two or multiple computational methods to treat a single system.

An area where major development is still necessary for hybrid methods is modeling excited states. For a realistic description of these phenomena, a combination of two different QM methods (QM/QM) may even be required. The challenging part of such a hybrid approach is to properly describe the mutual polarization among layers. Currently, in the field, there is no accepted standard computational protocol for a multi-layered embedding scheme that provides reliable modeling of excited state properties in complex environments. In order to design a multi-layered, fully mutually-polarizable embedding (MPE) scheme, implementing and thoroughly benchmarking the methods that act as its components is crucial. Our research aims to benchmark and implement several key methods that serve as components of an MPE, and to further expand the area of hybrid methods development for accurately modeling excited state properties.

One focus of our research is the development and application of hybrid methods that use coupled cluster with single and double excitations (CCSD) and its excited state version, equation-of-motion CCSD (EOM-CCSD), as high level. Methods based on CC theory provide consistently accurate results in gas phase. However, due to their steep computational scaling, works that focus

on CCSD as part of hybrid schemes are limited, and even less has been done for excited state CCSD compared to density functional theory (DFT)/time dependent (TD)-DFT.

We first benchmark EOM-CCSD combined with the polarizable continuum model (PCM) of solvation, for computing electronic excitation energies of solvated molecules. We address the PCM shortcoming of neglecting specific interactions, such as H bonding, through micro-solvated clusters inside the PCM cavity. We then combine EOM-CCSD with a classical polarizable molecular mechanics force field (MMPol) based on the induced dipole model to describe such specific interactions more efficiently.

Our work also extends to QM/QM hybrid methods. We consider embedding for the ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) hybrid method, and investigate the importance of including a polarizable embedding based on the induced dipole model. In addition, to simulate large regions of the UV/vis spectrum, we develop a multi-state extrapolation scheme based on the ONIOM extrapolation formula. This method overcomes the state-matching difficulty among sub-calculations in standard ONIOM, and is able to efficiently extrapolate several bands of the spectra at once.

In summary, our research expands the field of hybrid methods development for excited states, by proposing and testing several approaches for the simulation of large chromophores in complex environments.

Acknowledgements

I begin by expressing appreciation to my advisor, Prof. Marco Caricato, for being extremely supportive throughout my Ph.D. study and research. He has mentored and trained me not only in doing research, but also in improving all essential skills required to become a scientist, which include writing, presenting, and coding. These skills are the most valuable assets to my future career growth. I feel very lucky to have an advisor who care about what I learned and improved from doing each project, writing each paper, and giving each presentation. I am grateful for having such a enjoyable learning and research experience in Caricato group.

I would like to thank my other committee members: Prof. James Blakemore, Prof. Krzysztof Kuczera, Prof. Kyle Camarda, and Prof. Ward Thompson, for generously offering their time, support, and guidance throughout the whole process of preparing and reviewing the thesis.

I would also like to thank my group, both current members: Amy Jystad, Ty Balduf, Kaihua Zhang, Isaac Moore, Katherine Vander Laan, and Ryan Lampton, and past members: Dr. Alessandro Biancardi, Dr. Matt Barclay, Dr. Tal Aharon, Robert Adams, Hunter LeBlanc, Clara Stanton, Delaney Lynam, Analila Valenzuela, Jeremy Barnes, Joseph Harms, and Claudiu Caraiani, for being very supportive and creating such a great research environment in the group.

Next, I would like to thank my family for their continued understanding, love, and encouragement. I am thankful for my husband Xin, who has been a great partner, and accompanied me through every step on this journey. I am very grateful for having my son Rui in my life. The love that he expressed with his "baby language" gives me motivation and energy to finish works that are otherwise impossible. I would like to express appreciation to my parents for their support and understanding. They have been a great role model and taught me to take care of my family and be a good mom.

Finally, I would like to thank my undergraduate research advisor Prof. Wafaa Fawzy, who

started me on the journey of computational chemistry, when I was in my freshman year. She was very patient and supportive, and created a memorable research experience that motivate me to further pursue a Ph.D. degree in this area.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Accurate computational methods based on quantum mechanics (QM) have been developed and widely applied in different areas of scientific research. Thanks to the rapid growth of technology and computer processing power, QM calculations are not limited to theoretical studies, but have been used as a powerful tool to support, explain, and predict experiments. However, better performing computational methods are usually associated with higher computational cost that limits their application to small to medium-sized systems. Selecting only strictly the part of the system that is important to the process of interest for the modeling may be a solution to effectively minimize the computational cost, but the results of such calculations can be biased if the rest of the system also has a significant contribution to the process. For example, to study the properties of a molecule in solution, one can perform a gas phase calculation to reduce the computational cost, but this may not give an accurate enough description of the system as the interaction of the solute with the solvent molecules may considerably alter the property of interest. In order to perform QM calculations to effectively model the properties of large systems (such as biological molecules, solvated molecules, and large chromophores with complex interactions with their surroundings), a strategy that provides a good description while maintaining a reasonable computational cost is needed.

A popular approach for reducing the computational cost when performing calculations on large systems is to use hybrid approaches, which combine two or multiple computational methods for treating a single system. This is first partitioned in layers: the core region, such as a reaction/interaction center, which is treated with a more accurate and expensive computational method, and outer layers, which are treated with less accurate and expensive methods to provide a meaningful

description of the environmental effect. This approach provides a more efficient strategy for using computing resources, and its impact is such that, in 2013, Martin Karplus, Micheal Levitt, and Arieh Warshel were awarded the Nobel Prize for their development.[1] Over the past decades, different hybrid approaches have been developed and benchmarked that allow the wide application of QM methods to accurately modeling complex systems in various areas of scientific research.[2–10]

Typically, in a hybrid approach, the core region is described at QM level, whereas the rest of the system can be treated classically or with a QM method that has a lower computational cost than the one used for treating the core region, as shown in Figure 1.1. The choice of methods for treating the different regions depends on their respective size, the desired accuracy and computing time, as well as the amount of available computational resources.

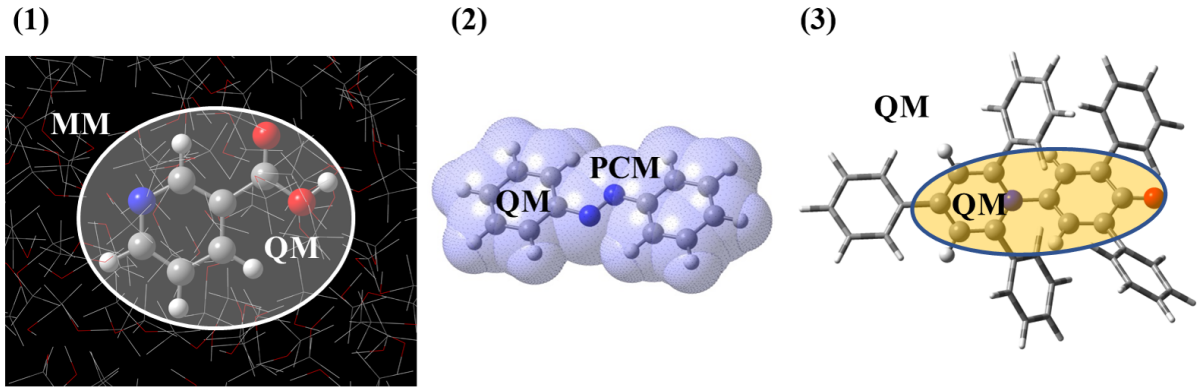

Figure 1.1: Different hybrid methods that combine (1) quantum mechanical (QM) and molecular mechanical (MM) methods, (2) QM and polarizable continuum model (PCM), (3) two QM methods.

A classical environment can be described explicitly or implicitly. In explicit models, the QM core is surrounded by a fixed[1,11–14] or polarizable molecular mechanics (MM) force field.[15–27] In implicit models, the effect of the environment is modeled by a continuum medium. One of the most famous examples for treating the environment implicitly is the polarizable continuum model (PCM) of solvation,[28–30] where the solvent molecules are replaced by a continuum dielectric medium. Both the implicit and explicit descriptions of the environment have pros and cons. Implicit models are unable to describe inhomogeneous environments and the direct interactions with the core region, such as the hydrogen bonding between solute and solvent molecules. Ex-

plicit models provide a more realistic description of the environment, and are able in principle to reproduce all sorts of environments. However, explicit models are more expensive due to the additional effort to account for conformational averaging, which in implicit models is included in the macroscopic dielectric constant of the medium.

While a classical treatment is an efficient way to include some major effects of the environment into QM calculations, it lacks the ability to describe QM effects, which can be important in certain cases. In such situations, a second layer can be treated with another QM method that has a lower computational cost than the one used for describing the core region. There are different approaches being developed to treat the interactions between two QM regions, which can be categorized into additive and subtractive. In the additive approaches, the interaction energy (E_{int}) between the high and low QM levels is computed directly, where in subtractive approaches, E_{int} is obtained indirectly with an extrapolation formula. Typically, low level QM methods are based on density functional theory (DFT) or semiempirical wave function (WF) theory.

A popular family of additive methods is DFT-in-DFT and WF-in-DFT.[31–41] In these methods, the core region (embedded subsystem) described at DFT level or with a correlated WF method, such as coupled-cluster theory, is embedded in the potential created based on the charge density of the environment computed with a low-level DFT method. Although, WF-in-DFT is more accurate than DFT-in-DFT for single-point calculations, it is too expensive for treating large systems and processes that involve multiple calculations, such as exploring reaction pathways. DFT-in-DFT is more efficient, thus it found wider applications.[42–46] In this family of embedding methods, one can employ the frozen-density-embedding (FDE) approximation, where the density of the environment is determined once and kept frozen. Alternatively, both densities can be optimized simultaneously with a freeze-and-thaw procedure, until mutual polarization is achieved. One of the advantages of DFT-in-DFT embedding is that it avoids the approximations associated with link atoms for system partitioning, which exist in other popular hybrid methods. The major limitations of these DFT embedding methods come from issues caused by the density overlap between the embedded subsystem and the environment. Another issue is the non-additive nature of the kinetic

potential (NAKP).[47,48] There have been many attempts to obtain improved approximations of the NAKP,[49–54] but most have limited applications to weakly interacting subsystems. Other popular additive methods are based on fragmentation, such as the fragment molecular orbital (FMO) method.[55–68] These methods separate the entire system into fragments, and then each fragment is treated separately within the embedding of the other fragments. These methods are powerful as they are able to treat very large systems at QM level, but they are only applicable to systems that can be clearly separated into fragments, such as proteins.

One of the most popular subtractive methods is ONIOM (our own n-layered integrated molecular orbital and molecular mechanics),[69–74] developed by Morokuma and co-workers, which has applications in both QM/MM and QM/QM studies. In ONIOM, the core region is usually called the "model", and dangling bonds that result from the partitioning of the system are capped with H link atoms. For 2-layered ONIOM, the energy is obtained as the combination of the energies of three independent subcalculations: 1) a calculation on the entire system (called "real" system) at the low level 2) a calculation on the model at the high level, and 3) a calculation on the model at the low level. The fact that ONIOM requires a low level calculation on the entire system somewhat limits its application to systems of certain sizes for QM/QM hybrid.[75–77] However, the energy is obtained directly within the ONIOM extrapolation without additional modification of quantum chemistry codes. The model calculations can in principle be performed in the presence of a polarizing embedding for a more accurate description of the model region.[78–81]

Both additive and subtractive methods have advantages and disadvantages. For instance, subtractive methods have an intrinsic mechanism to soften the effect of a poor choice of low-level method through the extrapolation formula. On the other hand, the additive methods based on fragmentation usually do not require calculations on the entire system. Thus, they are more easily applicable to larger systems, but they require specific implementations in QM software packages. Therefore, despite all this effort, the development of hybrid methods is still an active field of research. In fact, selecting the proper hybrid approach for a specific application is not trivial.

An area of major development for hybrid methods is the study of excited states. Accurately

modeling the excitation of systems that have complex interactions with their surrounding is challenging, but essential to correctly model those processes. For example, in dye-sensitized solar cells,[82,83] accurately modeling the complex interaction between the chromophore and its surrounding is necessary to obtain qualitatively correct trends of energy absorption and charge transport. Thus, a combination of two different QM methods (QM/QM) is needed to accurately model the chromophore. In such a hybrid approach, the challenging part is to properly describe the mutual polarization among layers. Currently, there is no accepted standard computational protocol for a multi-layered embedding scheme that provides reliable modeling of the excited state properties in complex environments. In order to design a multi-layered fully mutually-polarizable-embedding (MPE) scheme, implementing and thoroughly benchmarking the methods that act as components of the MPE is crucial. The purpose of our research is to implement and benchmark the methods that serve as components of a general MPE, and to further expand the area of hybrid methods development for accurately modeling excited state properties.

In a hybrid system, the choice of QM method used for modeling the core region plays a very important role. Accurate DFT functionals as high level are a popular choice, which has been well benchmarked for excited state calculations in both gas[84,85] and solution phases.[86–90] These functionals have been combined with many classical embedding schemes. For instance, the Mennucci group combined various polarizable embedding schemes with DFT/time dependent (TD)-DFT, and used it in many applications.[2–10] However, the performance of DFT highly depends on the choice of functional, and it is very system-dependent. Methods based on coupled cluster (CC) theory, such as CC with single and double excitations (CCSD) and its excited state version, i.e., equation-of-montion CCSD (EOM-CCSD, see section 2.3.1 in Chapter 2),[91–99] provide consistently accurate results in gas phase.[85,100,101] Although EOM-CCSD is one of the most accurate methods for computing one-electron excitation energies, it has a steep computational scaling: O(N^6), where N is the size of the basis set. Thus, works that focus on CCSD as part of a hybrid approach are limited, and even less has been done for excited state CCSD compared to DFT/TD-DFT.[95–97,100,102] Therefore, part of our work focuses on the development and application of hybrid methods with

CCSD/EOM-CCSD as high level.

In the Chapter 3, we benchmark EOM-CCSD combined with PCM for the calculation of electronic excitation energies of solvated molecules. The goal is to test the reliability of EOM-CCSD-PCM for the evaluation and prediction of excitation energies of solvated molecules compared to TDDFT-PCM. We also investigate a particularly important source of error: the lack of H-bonding interactions in PCM, by adding explicit solvent molecules to form micro-solvated clusters inside the PCM cavity. We define an energy shift, Δ_{HB}, from bare PCM to microsolvation + PCM at DFT level, and show that Δ_{HB} is independent of the functional used, contrary to the absolute value of the excitation energy. Hence, we suggest an efficient protocol where the EOM-CCSD-PCM transition energy is corrected by Δ_{HB} (DFT), which consistently improves the agreement with the experimental measurements. The EOM-CCSD-PCM benchmark confirms that modeling solvation with an implicit model neglects important specific interactions between solvent and solute. Treating explicit solvent molecules with CC is too expensive even for micro-clusters. Although we address this by forming micro-solvated clusters inside the PCM cavity at DFT level, proper sampling is still required for deciding where to place the solvent molecules.

Therefore, in Chapter 4, we combine CCSD with a classical polarizable molecular mechanics force field (MMPol) based on the induced dipole model.[5,19,22,27,103–105] We present the theory and implementation for combining CCSD/EOM-CCSD with MMPol within both the state specific (SS)[95,106–108] and linear response (LR)[108–110] formalisms for the interaction of the QM and MM regions. We also consider an approximate expression of the correlation energy, originally developed for CCSD with implicit solvation models, where the interaction term is linear in the coupled cluster density. This approximation allows us to include the explicit contribution of the environment to the CC equations without increasing the computational effort. We perform tests on a set of microsolvated systems, where the CCSD/MMPol method is compared to full CCSD calculations, and demonstrate the reliability of this computational protocol for all interaction schemes (error < 2%). We also show that it is important to include induced dipoles on all atomic centers of the classical region, and that too diffuse functions in the basis set may be problematic due to a too

strong interaction with the environment. In this work, the induced dipoles included in the MMPol force field are computed by using pre-determined parameters. However, it would be desirable to have a strategy to determine these parameters on-the-fly for a generic system.

In Chapter 5, we introduce an embedding with fixed point charges and induced dipoles for the ONIOM (QM:QM) hybrid method. We devise a strategy for obtaining atomic polarizabilities on-the-fly, which are the key parameter for the induced dipole embedding model. This strategy allows us to treat any generic system. We compare excitation energies from ONIOM extrapolations with different embedding models against the corresponding full QM calculation (i.e., the entire molecule calculated at the high QM level). These preliminary tests show that this on-the-fly polarizable embedding is better than no or fixed embedding when the substituent groups around the model system are polar.

These methods provide excitation energies and transition properties for individual states. However, in many cases, we are interested in an entire region of the absorption spectrum, rather than a single excited state. Unfortunately, it is hard to obtain the full spectrum with standard ONIOM, because proper state-matching between the subcalculations is challenging. In Chapter 6, we present a method to perform multi-state extrapolation of UV-vis spectra, based on the ONIOM extrapolation formula. This method overcomes the state-matching difficulty between sub-calculations, and is able to efficiently extrapolate several bands of the spectrum at once. This is accomplished by extrapolating the parameters that characterize the band position and shape in the spectrum of each subcalculation, e.g., band/shoulder position, width, and height, with a formula similar to that of ONIOM. We perform tests on seven chromophoric molecules that show the efficacy and robustness of this methodology in reproducing the spectrum computed for the entire molecule at a high level of theory.

In summary, our research expand the field of hybrid methods development for excited states, by developing and implementing various hybrid approaches that serve as components of a general multi-layered MPE for simulating large systems in complex environments.

Chapter 2

Theory

In this chapter, we review the theory that underlies the methods developed and utilized in the later chapters of this thesis, in order to provide a theoretical foundation for these methods. This chapter is constructed as follows: Section 2.1 gives an overview of hybrid methods; Section 2.2 introduces the basic theory of the polarizable continuum model (PCM); Section 2.3 reviews coupled cluster theory for ground and excited states, in gas phase and in combination with PCM.

2.1 Hybrid Methods

A hybrid approach in computational chemistry combines multiple computational methods in order to achieve the best balance between computational cost and accuracy. For example, in a two-layered hybrid scheme in Figure 2.1, the system core is selected to be treated with a more accurate (high) method, while the rest of the system (environment) is treated with a less expensive (low) method. The choice of core region depends on the specific problem at hand, and it is where the process of interest occurs, such as a reaction/interaction center. An accurate quantum mechanical (QM) method is usually chosen as the high level to treat the core region, while the choice of low level depends on the size of the system and the desired accuracy. The most common hybrid approaches combine a QM method with molecular mechanical (MM) methods (QM/MM).[1,11–14] or a continuum model, such as PCM (QM/PCM).[28,29,95,99,103,106–109,111,112] Combinations can also be made between a QM method and another QM method with lower computational cost.[31–44] In a

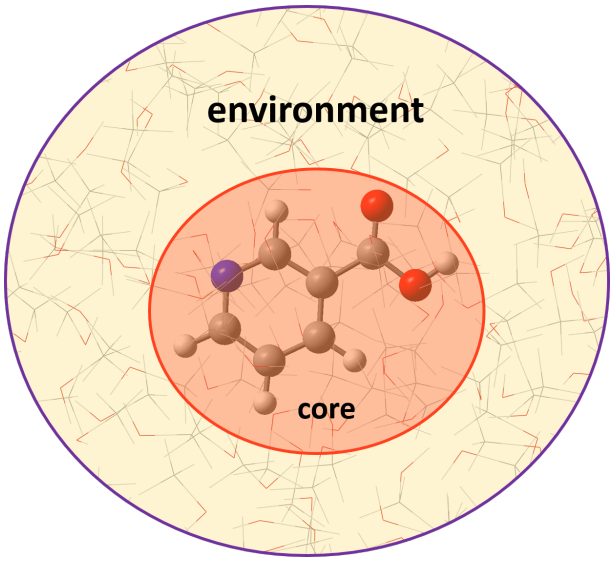

Figure 2.1: Illustration for the partitioning of a system for a hybrid scheme.

hybrid scheme, the energy of the system can be expressed as:[3,113,114]

$$E = E(core) + E(env) + E(int) \tag{2.1}$$

where $E(core)$ is the energy of the core region, $E(env)$ is the energy of the environment, and $E(int)$ is the interaction energy between the core region and the environment. There are two main categories of hybrid methods: additive and subtractive, and they differ in how they compute the interaction energy. In an additive approach, the interaction energy is calculated explicitly. In a subtractive method, such as ONIOM (Our own N-layered Integrated molecular Orbital molecular Mechanics),[69–77] the interaction energy of the system is expressed as:

$$E(int) = E(real, low) - [E(model, low) + E(rest, low)] \tag{2.2}$$

where $E(real, low)$, $E(model, low)$, and $E(rest, low)$ are the energies of the entire system, the core region, and the environment calculated at the same low level, respectively. Inserting Eq. 2.2 into

Eq. 2.1, we obtain the extrapolated energy expression for ONIOM:

$$E(ONIOM) = E(model, high) + E(real, low) - E(model, low) \tag{2.3}$$

Therefore, in a subtractive method, $E(int)$ is never actually calculated. In Eq. 2.2 and 2.3, we use a notation that is specific for ONIOM: comparing to the general energy expression in Eq. 2.1, $E(env) \equiv E(rest, low)$ and $E(core) \equiv E(model, high)$. The expression in Eq. 2.3 can be interpreted as the energy of the entire system calculated with a low level of method ($E(real, low)$) being corrected by the energy difference of the model between low and high methods.

In an additive approach, the accuracy highly depends on the low level method used for obtaining the interaction energy. Additionally, specific implementations of the hybrid scheme are required. In a subtractive approach, if no electronic embedding schemes are used, the subcalculations are independent from each other and the method requires no specific implementation. This permits maximum flexibility in the choice of the high and low levels of theory. In our research, we use both approaches: additive for combining coupled cluster (CC) with PCM and a classical polarizable molecular mechanics force field (MMPol); subtractive for ONIOM with electronic embedding (EE), and for the multi-state extrapolation of UV/vis spectra.

2.2 Polarizable Continuum Model

Including environmental effects is important when modeling processes that occur in solution, which are typically introduced with classical models. One of the most popular solvation models (belonging to the additive family of hybrid methods) is the polarizable continuum model (PCM). In PCM (Figure 2.2),[30] the solute molecule is placed in a cavity surrounded by a polarizable dielectric, and the bulk polarization is represented by an apparent surface charge on the cavity. The

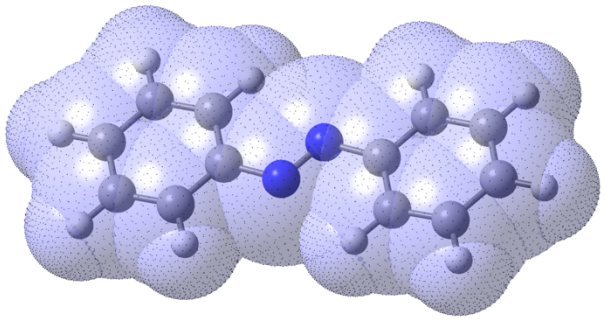

Figure 2.2: An example of a solute molecule (ball and stick) placed in a PCM cavity (light blue).

model starts from the Poisson equations:

$$-\nabla^2 V(\vec{r}) = 4\pi\rho_M(\vec{r}) \quad \text{within C} \tag{2.4}$$

$$-\varepsilon\nabla^2 V(\vec{r}) = 0 \quad \text{outside C} \tag{2.5}$$

where C represent the PCM cavity, ε is a dielectric constant for the solvent medium, and V is the electrostatic potential on point $\vec{r}$:

$$V(\vec{r}) = V_M(\vec{r}) + V_R(\vec{r}) \tag{2.6}$$

$V_M(\vec{r})$ is the electrostatic potential generated by the charge distribution ρ_M inside the cavity, and $V_R(\vec{r})$ is the reaction potential generated by the polarization of the dielectric medium. At the cavity surface Γ, the potential V satisfies jump conditions expressed as:[30]

$$[V] = V_{in} - V_{out} = 0 \quad \text{on } \Gamma \tag{2.7}$$

$$[\partial V] = \left(\frac{\partial V}{\partial \vec{n}}\right)_{in} - \varepsilon\left(\frac{\partial V}{\partial \vec{n}}\right)_{out} = 0 \quad \text{on } \Gamma \tag{2.8}$$

where $\vec{n}$ is a set of outward-pointing vectors perpendicular to the cavity surface. The first jump condition in Eq. 2.7 expresses the continuity of the potential across the surface, whereas the second jump condition in Eq. 2.8 expresses the discontinuity of the gradient of the potential with respect to the $\vec{n}$ vectors. The dielectric constant inside the cavity is 1, i.e., the dielectric constant of the vacuum, while the dielectric constant outside equals to ε (a finite value > 1). From the jump

conditions in Eq. 2.7-2.8, an apparent surface charge (ASC) $\sigma(\vec{s})$ spread on the cavity surface is defined, where $\vec{s}$ is a position on the cavity surface. The ASC defines the reaction field potential V_R in Eq. 2.6 over the whole space as:

$$V_R(\vec{r}) = \int_\Gamma \frac{\sigma(\vec{s})}{|\vec{r}-\vec{s}|} d\vec{s} \tag{2.9}$$

The cavity surface is approximated in terms of a set of finite elements that are small enough to consider a constant $\sigma(\vec{s}_k)$ within each element, and a set of charges q_k is defined in terms of the local value of $\sigma(\vec{s}_k)$ on each of these finite elements times the corresponding area A_k. With this new definition, the integral in Eq. 2.9 is transformed into a finite sum:

$$V_R(\vec{r}) = \sum_k \frac{\sigma(\vec{s}_k)A_k}{|\vec{r}-\vec{s}_k|} = \sum_k \frac{q_k}{|\vec{r}-\vec{s}_k|} \tag{2.10}$$

There are different ACS variants of PCM that differ in their definition of surface charges and potentials.[30] However, most ACS models can be recast in a matrix form for the calculation of the PCM charges:

$$\mathbf{q} = -\mathbf{K}\mathbf{f} \tag{2.11}$$

where $\mathbf{K}$ is a $T \times T$ square matrix that collects the cavity geometrical factors and the dielectric constant of the medium, with T equal to the number of finite elements. $\mathbf{q}$ and $\mathbf{f}$ are vectors that contain the charges and the potential, respectively. For different ACS variants of PCM, the definitions of $\mathbf{K}$ and $\mathbf{f}$ are different, and we use the Integral Equation Formalism (IEF)[115–117] in our research.

The advantages of PCM include its adaptability to virtually any solute and solvent, and its efficient implementation with QM methods allows its wide application for electronic structure calculations in solution.

2.3 Coupled Cluster Theory

Coupled-cluster (CC) theory collects a highly accurate family of QM methods that can be used as the high-level in a hybrid scheme. This section contains an overview of CC theory and its combination with the PCM solvation model for ground and excited states.[95,99,106–112,118]

2.3.1 CC Electronic Ground State

CC theory provides a systematic approach to evaluate approximate solutions of the time-independent Schrödinger equation:[94]

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \tag{2.12}$$

where $\hat{H}$ is the Hamiltonian of the system, Ψ is the exact wave function, and E is the exact energy of the ground state (GS). The wave function ansatz in coupled cluster theory is given by an exponential wave operator acting on a reference Slater determinant $|\Phi^0\rangle$:[119–122]

$$\begin{aligned} |\Psi_{CC}\rangle &= e^{\hat{T}}\left|\Phi^0\right\rangle, \\ \hat{T} &= \hat{T}_1 + \hat{T}_2 + \cdots + \hat{T}_n \end{aligned} \tag{2.13}$$

where $\hat{T}$ is the linear combination of all possible $\hat{T}_n$ excitation operators for a system with n electrons.[123,124] Each $\hat{T}_n$ generates n-fold excitations of the same order:

$$\hat{T}_1 = \sum_{\text{singles}} t_1\hat{\tau}_1 = \sum_i \sum_a t_i^a a^+ i \tag{2.14}$$

$$\hat{T}_2 = \sum_{\text{doubles}} t_2\hat{\tau}_2 = \sum_{i,j} \sum_{a,b} t_{ij}^{ab} a^+ i b^+ j \tag{2.15}$$

$$\vdots \tag{2.16}$$

$$\hat{T}_n = \sum_{\text{n order excitations}} t_n\hat{\tau}_n = \sum_{i_1<i_2\cdots<i_n} \sum_{a_1<a_2\cdots<a_n} t_{i_1 i_2 \cdots i_n}^{a_1 a_2 \cdots a_n} a_1^+ i_1 a_2^+ i_2 \cdots a_n^+ i_m \tag{2.17}$$

More explicitly, $\hat{T}_1$ is the operator that collects all single excitations, $\hat{T}_2$ is the operator that collects all double excitations and so forth. Including different orders of excitations in the operator $\hat{T}$ results in different truncated CC methods, for example, CCSD includes single and double excitations ($\hat{T} = \hat{T}_1 + \hat{T}_2$), whereas CCSDT includes single, double, and triple excitations ($\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$). $\hat{\tau}_n$ is an excitation operator that creates a n-order excited determinant $|\Phi^n\rangle$ when acting on the reference Slater determinant $|\Phi^0\rangle$: $\hat{\tau}_n|\Phi^0\rangle = |\Phi^n\rangle$, and it takes the form of $a^+i, a^+ib^+j, \cdots$, etc. For a sequence of creation/annihilation operators in second quantization, $i, j, k, \cdots (a, b, c, \cdots)$ refer to occupied (virtual) molecular orbitals in the Fermi-vacuum Φ^0.

The expression for the CC ground state energy is obtained by inserting the $|\Psi_{CC}\rangle$ ansatz into the Schrödinger equation in Eq. 2.12, multiplying on the left by $e^{-\hat{T}}$, then projecting onto the reference wave function $\langle\Phi^0|$:

$$\left\langle \Phi^0 \middle| e^{-\hat{T}} \hat{H} e^{\hat{T}} \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| \bar{H} \middle| \Phi^0 \right\rangle = E_0 \tag{2.18}$$

where the overbar indicates the similarity transformation of the Hamiltonian: $\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$. The similarity transformed Hamiltonian in CC theory is non-Hermitian, thus it has different left- and right-hand eigenvectors that result in more complicated equations for excited states. However, the non-Hermitian property of $\hat{H}$ guarantees the size extensivity of the CC ground state energy at every level of truncation of $\hat{T}$. Additionally, using the Baker–Campbell–Hausdorff (BCH) expansion:

$$\bar{H} = \hat{H} + \left[\hat{H}, \hat{T}\right] + \frac{1}{2!}\left[\left[\hat{H}, \hat{T}\right], \hat{T}\right] + \frac{1}{3!}\left[\left[\left[\hat{H}, \hat{T}\right], \hat{T}\right], \hat{T}\right] + \frac{1}{4!}\left[\left[\left[\left[\hat{H}, \hat{T}\right], \hat{T}\right], \hat{T}\right], \hat{T}\right] \tag{2.19}$$

This expansion of $\bar{H}$ ends exactly after four nested commutators, based on Wick's theorem.[124]

It is convenient to express the energy using Lagrange multipliers,[94] for combining CC theory with a polarizable environment later. The Lagrangian for ground state CC in gas phase is expressed as:

$$\mathcal{L}_0 = \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \bar{H} \middle| \Phi^0 \right\rangle \tag{2.20}$$

where $\hat{\Lambda}$ is a linear combination of de-excitation operators:[123,124]

$$\hat{\Lambda} = \sum_n \lambda_n \hat{\tau}_n^\dagger \tag{2.21}$$

where $\hat{\tau}_n^\dagger$ is the Hermitian conjugate of $\hat{\tau}$ in Eq. 2.14, and λ_n is the corresponding amplitude. It is useful to separate the reference contribution to E^0 from the rest. By introducing the normal product form of an operator:[123,124]

$$\hat{X}_N = \hat{X} - \left\langle \Phi^0 \middle| \hat{X} \middle| \Phi^0 \right\rangle = \hat{X} - X^0 \tag{2.22}$$

Eq. 2.20 can be rewritten as:

$$\mathcal{L}_0 = E^0 + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \bar{H}_N \middle| \Phi^0 \right\rangle \tag{2.23}$$

where E^0 is the reference energy. Minimizing $\mathcal{L}_0$ with respect to λ_n amplitudes, we obtain:

$$\frac{\partial \mathcal{L}_0}{\partial \lambda_n} = \left\langle \Phi^n \middle| \bar{H}_N \middle| \Phi^0 \right\rangle = 0 \tag{2.24}$$

which is the equations of t_n amplitudes. At convergence, Eq. 2.20 is reduced to:

$$\mathcal{L}_0 = \mathcal{L}^0 + \left\langle \Phi^0 \middle| \bar{H}_N \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| \bar{H} \middle| \Phi^0 \right\rangle = E_0 \tag{2.25}$$

i.e., the CC ground state energy.

2.3.2 Coupled Cluster Theory for Electronic Excited States

There are two approaches for extending CC theory to excited states: linear response (LR) and equation of motion (EOM).[91–94,125] These two approaches start from different points, but they arrive to the same eigenvalue equation. The CC LR function expresses excited states observables as the response of the ground state to external field, and the excitation energies are obtained as the poles of the LR function. On the other hand, EOM-CC begins with the ansatz that expresses the

wave function of the excited state K as:

$$|\Psi_K\rangle = \hat{R}_K |\Psi_{CC}\rangle \tag{2.26}$$

where $\hat{R}_K$ is an excitation operator that satisfies the "equation-of-motion":[126,127]

$$\left[\hat{H}, \hat{R}_K\right] = \omega_K \hat{R}_K \tag{2.27}$$

The EOM ansatz is then inserted into the Schrödinger equation in Eq. 2.12, leading to an eigenvalue equation for excited states. Since $\bar{H}$ is non-Hermitian, its left and right eigenvectors are different. For the right-hand diagonalization, the transition ω_K is found by solving:

$$\left\langle \Phi^n \middle| \left[\bar{H}, \hat{R}_K\right] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^n \middle| \hat{R}_K \middle| \Phi^0 \right\rangle \tag{2.28}$$

From the left-hand side, we have:

$$\left\langle \Phi^0 \middle| \hat{L}_K [\bar{H}, \hat{\tau}_n] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^0 \middle| \hat{L}_K \middle| \Phi^n \right\rangle \tag{2.29}$$

where $\hat{R}_K$ and $\hat{L}_K$ are the right- and left- hand eigenvectors of the similarity transformed Hamiltonian $\bar{H}$, respectively:

$$\hat{R}_K = r_0^K + \sum_n r_n^K \hat{\tau}_n \tag{2.30}$$

$$\hat{L}_K = l_0^K + \sum_n l_n^K \hat{\tau}_n^\dagger \tag{2.31}$$

which satisfy the biorthonormality condition:

$$\left\langle \Phi^0 \middle| \hat{L}\hat{R} \middle| \Phi^0 \right\rangle = 1 \tag{2.32}$$

We can introduce an energy functional for the excited state K using Lagrangian multipliers,[94] again for making a connection with the CC-PCM equations, which will be introduced next. The

Lagrangian for an excited state K is expressed as:

$$\mathcal{L}_K = E^0 + \left\langle 0 \middle| \hat{L}_K \left[\bar{H}_N, \hat{R}_K \right] \middle| 0 \right\rangle + \left\langle \Phi^0 \middle| (1 + \Lambda_K) \bar{H}_N \middle| \Phi^0 \right\rangle + \omega_K \left(1 - \left\langle \Phi_0 \middle| \hat{L}_K \hat{R}_K \middle| \Phi_0 \right\rangle \right) \tag{2.33}$$

Here we used the normal-product form of an operator (Eq. 2.22) to separate the reference wave function contribution to $\mathcal{L}_K$ from the correlation part, where E^0 is the ground state reference energy. Equations for the $\hat{T}_K$, $\hat{\Lambda}_K$, and $\hat{R}_K$ ($\hat{L}_K$) for the ground and excited amplitudes can be obtained by minimizing $\mathcal{L}_K$ in Eq. 2.33 with respect to their amplitude parameters. At convergence, Eq. 2.33 reduces to:

$$\mathcal{L}_K = E_K = E^0 + \left\langle \Phi^0 \middle| \bar{H}_N \middle| \Phi^0 \right\rangle + \omega_K \tag{2.34}$$

2.3.3 CC-PCM Electronic Ground State

In order to perform CC calculations in solution, one can combine CC with a continuum model, like PCM. The CC-PCM energy functional can be conveniently expressed as a Lagrangian, as for gas phase CC:[95,106–110,118]

$$\mathcal{G}_0^{PCM} = \mathcal{G}^0 + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \frac{1}{2} \boldsymbol{Q}_N \cdot \boldsymbol{V}_N \tag{2.35}$$

where $\mathcal{G}^0$ is the reference free energy:

$$\mathcal{G}^0 = \left\langle \Phi^0 \middle| \hat{H} \middle| \Phi^0 \right\rangle + \frac{1}{2} \boldsymbol{Q}^0 \cdot \boldsymbol{V}^0 \tag{2.36}$$

$\mathcal{G}^0$ is a free energy because it includes the energy for polarizing the dielectric. $\bar{H}_N^{PCM}$ is the QM Hamiltonian including the PCM operator with the reference reaction field $\boldsymbol{Q}^0$:

$$\hat{H}_N^{PCM} = \hat{H}_N + \hat{V}_N \cdot \boldsymbol{Q}^0 \tag{2.37}$$

where $\boldsymbol{Q}_N/\boldsymbol{V}_N$ are the surface charge/electrostatic potential induced by the correlation density:

$$\boldsymbol{Q}_N = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) e^{-\hat{T}} \hat{\boldsymbol{Q}}_N e^{\hat{T}} \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{\boldsymbol{Q}}_N \middle| \Phi^0 \right\rangle \tag{2.38}$$

$$\boldsymbol{V}_N = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) e^{-\hat{T}} \hat{\boldsymbol{V}}_N e^{\hat{T}} \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle \tag{2.39}$$

The expression in Eq. 2.35 is quadratic in the CCSD density through the last term, and it is usually referred to as PTED scheme (from perturbation theory energy and density).[128,129]

The free energy for the ground state is evaluated by minimizing the Lagrangian in Eq. 2.35 with respect to the $\hat{T}$ and $\hat{\Lambda}$ amplitudes. The partial derivative of the Lagrangian with respect to λ_n gives the $\hat{T}$ equations:

$$\frac{\partial \mathcal{G}_0^{PCM}}{\partial \lambda_n} = \left\langle \Phi^n \middle| \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \boldsymbol{Q}_N \cdot \left\langle \Phi^n \middle| \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle = 0 \tag{2.40}$$

This is similar to the gas phase $\hat{T}$ equations in Eq. 2.24 with an additional PCM term. The partial derivative of the Lagrangian with respect to t_n gives the $\hat{\Lambda}$ equations:

$$\frac{\partial \mathcal{G}_0^{PCM}}{\partial t_n} = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \left[\bar{H}_N^{PCM}, \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle + \boldsymbol{Q}_N \cdot \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \left[\bar{\boldsymbol{V}}_N, \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle = 0 \tag{2.41}$$

At convergence Eq. 2.35 reduces to:

$$\mathcal{G}_0^{PCM} = \mathcal{G}^0 + \Delta E_0 - \frac{1}{2} \boldsymbol{Q}_N \cdot \boldsymbol{V}_N \tag{2.42}$$

where

$$\Delta E_0 = \left\langle \Phi^0 \middle| \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \boldsymbol{Q}_N \cdot \left\langle \Phi^0 \middle| \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle \tag{2.43}$$

In gas phase, it is not necessary to evaluate the $\hat{\Lambda}$ amplitudes. However, since Eq. 2.35 is quadratic in $\hat{T}$ and $\hat{\Lambda}$ due to the PCM term, contrary to the gas phase case, Eqs. 2.40-2.41 are coupled and must be solved simultaneously. Thus, there is a steep increase in the computational cost

for solving these equations iteratively. To reduce this computational cost, approximate free energy functionals for CC-PCM can be used instead of the complete PTED version of the Lagrangian.[95] The goal is to decouple the $\hat{T}$ and $\hat{\Lambda}$ equations by eliminating the terms that cause the coupling. In order to do this, $\boldsymbol{Q}_N$ in Eq. 2.35 is split into two contributions:

$$\boldsymbol{Q}_N = \boldsymbol{Q}_N^T + \boldsymbol{Q}_N^\Lambda \tag{2.44}$$

defined as:

$$\begin{aligned} \boldsymbol{Q}_N^T &= \left\langle \Phi^0 \middle| \bar{\boldsymbol{Q}}_N \middle| \Phi^0 \right\rangle \\ \boldsymbol{Q}_N^\Lambda &= \left\langle \Phi^0 \middle| \hat{\Lambda} \bar{\boldsymbol{Q}}_N \middle| \Phi^0 \right\rangle \end{aligned} \tag{2.45}$$

If the same separation is performed for $\boldsymbol{V}_N$, the last term in Eq. 2.35 can be written as:

$$\frac{1}{2}\boldsymbol{Q}_N \cdot \boldsymbol{V}_N = \frac{1}{2}\boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^T + \boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^\Lambda + \frac{1}{2}\boldsymbol{Q}_N^\Lambda \cdot \boldsymbol{V}_N^\Lambda \tag{2.46}$$

where in the second term on the right hand side we use the symmetry of the PCM kernel ($\frac{1}{2}\boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^\Lambda = \frac{1}{2}\boldsymbol{Q}_N^\Lambda \cdot \boldsymbol{V}_N^T$). By eliminating the last term in Eq. 2.46, which is responsible for coupling the $\hat{T}$ and $\hat{\Lambda}$ equations, one obtain the approximate energy functional for the PTES scheme (where the S stands for singles):

$$\mathcal{G}_0^{S,PCM} = \mathcal{G}^0 + \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \frac{1}{2}\boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^T + \boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^\Lambda \tag{2.47}$$

The $\hat{T}$ equations become:

$$\frac{\partial \mathcal{G}_0^{S,PCM}}{\partial \lambda_n} = \left\langle \Phi^n \middle| \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \boldsymbol{Q}_N^T \cdot \left\langle \Phi^n \middle| \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle = 0 \tag{2.48}$$

The $\hat{\Lambda}$ equations become:

$$\begin{aligned}\frac{\partial \mathcal{G}_0^{S,PCM}}{\partial t_n} &= \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \left[\bar{H}_N^{PCM}, \tau_n \right] \middle| \Phi^0 \right\rangle + \boldsymbol{Q}_N^T \cdot \left\langle \Phi^0 \middle| \hat{\Lambda} \left[\bar{\boldsymbol{V}}_N, \tau_n \right] \middle| \Phi^0 \right\rangle \\ &+ \boldsymbol{Q}_N \cdot \left\langle \Phi^0 \middle| \bar{\boldsymbol{V}}_N \middle| \Phi^n \right\rangle = 0\end{aligned} \tag{2.49}$$

At convergence, Eq. 2.47 is reduced to:

$$\mathcal{G}_0^{S,PCM} = \mathcal{G}^0 + \left\langle \Phi^0 \middle| \bar{H}_N^{PCM} \middle| \Phi^0 \right\rangle + \frac{1}{2} \boldsymbol{Q}_N^T \cdot \boldsymbol{V}_N^T \tag{2.50}$$

Since in the PTES scheme the Lagrangian is linear in $\hat{\Lambda}$, as in gas phase, the equations for the $\hat{T}$ and $\hat{\Lambda}$ are decoupled, and can be solved seperately.

In terms of computational effort, the cost for evaluating all PCM terms scales as $O(N^5)$, where N is the basis set size, because that is the cost to evaluate the correlation density. Therefore, the PCM terms do not directly increase the cost of CC equations, which scale as $O(N^6)$ for CCSD. The main increase in computational cost comes from the coupling of the amplitude equations for the PTED scheme, contrary to the gas phase CC equations. However, by using the approximate PTES scheme, CC-PCM is equivalent in cost to CC in gas phase.

2.3.4 CC-PCM Electronic Excited States

As in gas phase, there are two strategies for including solvent effects in excited state calculations: the state specific (SS) (equivalent to EOM)[95,106–108] and linear response (LR)[108–110] formalisms. Although the two approaches provide the same result for the gas phase energy, they differ in solution. The origin of the difference comes from the definition of the solvent response, and can be summarized by saying that in the SS formalism the solvent response depends on the excited state density, while in the LR formalism it depends on the transition density. More detailed discussion of the difference between the two approaches can be found in Refs. 130,131.

2.3.4.1 State Specific formalism for CC-PCM

In the SS formalism,[95,106–108] the excited state Lagrangian can be written in a form equivalent to that of the ground state:

$$\begin{aligned}\mathcal{G}_K^{PCM} = \mathcal{G}^0 + \left\langle\Phi^0\middle|(1+\hat{\Lambda}_K)\bar{H}_N^{PCM}\middle|\Phi^0\right\rangle + \left\langle\Phi^0\middle|\hat{L}_K\left[\bar{H}_N^{PCM},\hat{R}_K\right]\middle|\Phi^0\right\rangle \\ + \omega_K\left(1-\left\langle\Phi^0\middle|\hat{L}_K\hat{R}_K\middle|\Phi^0\right\rangle\right) + \frac{1}{2}\boldsymbol{Q}_{KN}\cdot\boldsymbol{V}_{KN}\end{aligned} \tag{2.51}$$

where $\hat{L}_K$ and $\hat{R}_K$ are the left- and right-hand eigenvectors of the EOM-CCSD Hamiltonian for the K-th state, and ω_K is the eigenvalue associated with $\hat{L}_K$ and $\hat{R}_K$. The free energy for the K-th excited state can be computed by minimizing the Lagrangian in Eq. 2.51 with respect to all sets of amplitudes, as in gas phase. At convergence, Eq. 2.51 reduces to:

$$\mathcal{G}_K^{PCM} = \mathcal{G}^0 + \Delta E_K + \omega_K - \frac{1}{2}\boldsymbol{Q}_{KN}\cdot\boldsymbol{V}_{KN} \tag{2.52}$$

where:

$$\Delta E_K = \left\langle\Phi^0\middle|\bar{H}_N^{PCM}\middle|\Phi^0\right\rangle + \boldsymbol{Q}_{KN}\cdot\left\langle\Phi^0\middle|\bar{\boldsymbol{V}}_N\middle|\Phi^0\right\rangle \tag{2.53}$$

For excited states, the equations for the ground state amplitudes ($\hat{T}_K$ and $\hat{\Lambda}_K$ equations) and the equations for the excited state amplitudes ($\hat{R}_K$ and $\hat{L}_K$ equations) are all coupled by the solvent reaction field. Additionally, the $\hat{T}_K$ amplitudes are different from those of the ground state, because they are obtained in the presence of excited state charges $\boldsymbol{Q}_{KN}$. Therefore, the transition energy from the ground to the K-th excited state must be evaluated through two separate calculations, as the difference between $\mathcal{G}_K$ in Eq. 2.52 and $\mathcal{G}_0$ in Eq. 2.42.

The PTED formalism described above is computationally expensive because all the amplitude equations must be solved simultaneously. To reduce the computational cost, as for ground state, approximate schemes can be used to decouple these amplitude equations.[95] An excited state PTES scheme can be defined also for excited states. At convergence, $\mathcal{G}_K^{S,PCM}$ in Eq. 2.51 can be reduced

to:

$$\mathcal{G}_K^{S,PCM} = \mathcal{G}_0^{S,PCM} + \omega_K - \frac{1}{2}\boldsymbol{Q}_N^K \cdot \boldsymbol{V}_N^K \tag{2.54}$$

where $\mathcal{G}_0^{S,PCM}$ is the ground state PTES energy. It is important to note that in the PTES scheme the ground state equations for the $\hat{T}_K$ and $\hat{\Lambda}_K$ amplitudes are decoupled from the excited state equations, as in gas phase, but the $\hat{R}_K$ and $\hat{L}_K$ equations are still coupled. Nevertheless, since the ground and excited state amplitude equations are decoupled, excitation energies can be computed in a single calculation, contrary to the PTED scheme.

2.3.4.2 Linear Response Formalism for CC-PCM

In the LR formalism,[108–110] as in gas phase, the excitation energies are obtained as the poles of the LR function, and are calculated from the right or left-hand diagonalization of the Jacobian matrix, $\left\langle \Phi^m \middle| \left[(\bar{H}_N + \boldsymbol{Q}_N \cdot \bar{\boldsymbol{V}}_N), \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle$. The complete derivation of the CC/PCM LR function can be in found in Ref. 108–110. For the right-hand diagonalization, the transition energy ω_K is found by solving:

$$\left\langle \Phi^n \middle| \left[(\bar{H}_N + \boldsymbol{Q}_N \cdot \bar{\boldsymbol{V}}_N), \hat{R}_K \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^n \middle| \boldsymbol{Q}_N^{R_K} \cdot \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^n \middle| \hat{R}_K \middle| \Phi^0 \right\rangle \tag{2.55}$$

From the left-hand side, we have:

$$\left\langle \Phi^0 \middle| \hat{L}_K \left[(\bar{H}_N + \boldsymbol{Q}_N \cdot \bar{\boldsymbol{V}}_N), \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \left[\boldsymbol{Q}_N^{L_K} \cdot \bar{\boldsymbol{V}}_N, \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^0 \middle| \hat{L}_K \middle| \Phi^n \right\rangle \tag{2.56}$$

The $\boldsymbol{Q}_N^{R_K}$ and $\boldsymbol{Q}_N^{L_K}$ charges (CC has different right-hand and left-hand transition densities) can be interpreted as transition charges, as they express the solvent response to the transition density of

the solute:

$$\boldsymbol{Q}_N^{R_K} = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \left[\bar{\boldsymbol{Q}}_N, \hat{R}_K \right] \middle| \Phi^0 \right\rangle = \boldsymbol{Q}_N^{TR_K} + \boldsymbol{Q}_N^{\Lambda R_K} \tag{2.57}$$

$$\boldsymbol{Q}_N^{L_K} = \left\langle \Phi^0 \middle| \hat{L}_K \bar{\boldsymbol{Q}}_N \middle| \Phi^0 \right\rangle \tag{2.58}$$

These charges depend on the transition density of the solute, and are different from the charges in the SS formalism, which depends on the excited state density. LR transition properties were also derived and implemented for CCSD/PCM.[93,94,109,111]

The PTES approximation can be also introduced for the LR formalism, where the right and left-hand diagonalization equations become:[95,108]

$$\left\langle \Phi^n \middle| \left[(\bar{H}_N^Q + \boldsymbol{Q}_N^T \cdot \bar{\boldsymbol{V}}_N), \hat{R}_K \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^n \middle| \boldsymbol{Q}_N^{TR_K} \cdot \bar{\boldsymbol{V}}_N \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^n \middle| \hat{R}_K \middle| \Phi^0 \right\rangle \tag{2.59}$$

$$\left\langle \Phi^0 \middle| \hat{L}_K \left[(\bar{H}_N^Q + \boldsymbol{Q}_N^T \cdot \bar{\boldsymbol{V}}_N), \tau_n \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| \left[\boldsymbol{Q}_N^{TL_K} \cdot \bar{\boldsymbol{V}}_N, \tau_n \right] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^0 \middle| \hat{L}_K \middle| \Phi^n \right\rangle \tag{2.60}$$

The advantage of the PTES over the PTED scheme for the LR formalism is due to the fact that in the former the ground state $\hat{T}$ and $\hat{\Lambda}$ equations are decoupled. Thus, LR-CCSD/PCM with the LR-PTES scheme is virtually equal in cost to the corresponding gas phase method.

Chapter 3

An EOM-CCSD-PCM Benchmark for Electronic Excitation Energies of Solvated Molecules

(This work is taken from: Sijin Ren, Joseph Harms, and Marco Caricato, *J. Chem. Theory Comput.* **2017**, *13*, 117-124.[132] Supporting information is available online.)

Abstract

In this work, we benchmark the equation of motion coupled cluster with single and double excitations (EOM-CCSD) method combined with the polarizable continuum model (PCM) for the calculation of electronic excitation energies of solvated molecules. EOM-CCSD is one of the most accurate methods for computing one-electron excitation energies, an accounting for the solvent effect on this property is a key challenge. PCM is one of the most widely employed solvation models due to its adaptability to virtually any solute, and its efficient implementation with density functional theory methods (DFT). Our goal in this work is to evaluate the reliability of EOM-CCSD-PCM, especially compared to time-dependent DFT-PCM (TDDFT-PCM). Comparisons between calculated and experimental excitation energies show that EOM-CCSD-PCM consistently overestimates experiment by 0.4-0.5 eV, which is larger than the expected EOM-CCSD error in vacuo. We attribute this decrease in accuracy to the approximated solvation model. Thus, we investigate a particularly important source of error: the lack of H-bonding interactions in PCM. We show that this issue can be addressed by computing an energy shift, Δ_{HB}, from bare-PCM to micro-solvation + PCM at DFT level. Our results show that such shift is independent of the functional used, contrary to the absolute value of the excitation energy. Hence, we suggest an efficient protocol where the EOM-CCSD-PCM transition energy is corrected by Δ_{HB}(DFT), which consistently improves the agreement with the experimental measurements.

3.1 Introduction

The simulation of electronic spectra of molecules in solution is fundamental for basic research since it provides invaluable information about, for instance, the presence of products or intermediates in a chemical reaction. It also allows us to probe the interactions between a chromophore and its surrounding, and in this respect it is the first step in the design of new and efficient dyes in solar energy research[82,83]. Despite tremendous progress in the evaluation of electronic excited states of isolated molecules, including the effect of the environment is still a big challenge. This is due to the large number of molecular species that surround the chromophore of interest, which makes it nearly impossible to treat the entire system quantum mechanically (QM). For solvents, the common approach is to resort to classical models that may or may not include polarization terms. These models can be classified in two large families: explicit and implicit models, each with its advantages and disadvantages. Explicit models maintain an atomistic and thus more realistic representation of the solvent, and they can describe direct solute-solvent interactions such as hydrogen bonding (H-bonding). However, these models require a thorough conformational averaging in order to achieve a statistically appropriate description of a specific molecular property, which implies the repetition of the QM calculation on the solute. Therefore, explicit models are computationally intensive. Additionally, these models can include explicit polarization terms for the classical region (through fluctuating charges or induced atomic dipoles)[133] that further increase the computational cost since mutual polarization between solute and solvent must be achieved for each conformation included in the averaging. On the other hand, implicit solvation models replace the atomistic structure of the solvent with a continuum, polarizable medium characterized by some macroscopic properties (e.g. the dielectric constant). These models are particularly efficient since the averaging is included implicitly in the macroscopic solvent property, and mutual polarization with the solute is achieved with only a moderate increase in the computational effort compared to that required for an isolated molecule. The limitation of implicit models is that dynamical effects for both solute and solvent are neglected, as well as direct interactions. The range of applicability of both strategies must be obtained through a comparison with available experimental data.

In this work, we test one particular QM method for the solute: the equation of motion coupled cluster with single and double excitations (EOM-CCSD)[91,92], with a particular implicit solvation model: the polarizable continuum model (PCM).[30] EOM-CCSD is one of the most accurate methods for computing one-electron excitation energies (which are the most relevant in solar energy research, for instance), but it has a steep computational scaling: $O(N^6)$, where N is the size of the basis set. PCM is one of the most widely employed solvation models due to its adaptability to virtually any solute and solvent, and its efficient implementation with density functional theory methods (DFT). In fact, modern DFT-PCM implementations lead to an increase in the cost of the calculation by less than 20% compared to the corresponding isolated molecule. However, CCSD-PCM calculations are considerably more expensive than DFT-PCM because the solvent terms couple the equations for the CC T and Λ amplitudes,[107,112,118,134] so that the cost of the calculations in solution is 2-3 times larger than in gas phase. We presented the first implementation of CCSD and PCM for ground and excited state properties, and we developed a series of approximations to limit the computational cost increase.[95,96,98,99,102,118,134–136] What is now missing is a systematic benchmarking of EOM-CCSD-PCM so that the reliability of this approach can be assessed.

Before diving into the details of the present study, it is important to note that the interfacing of CC methods with solvation models, especially explicit ones, is an active area of research. Important contributions come from Kongsted and coworkers, who used a polarizable solvation model with induced dipole and higher multipole terms.[137–140] Recent work with the same approach was presented by the Klopper group, which extended this model to two-component CC.[141,142] Other important contributions come from the Slipchenko group, where CC methods are interfaced with the effective fragment potential (EFP) method.[143–148] We also proposed an implementation of EOM-CCSD with a fluctuating charge (FQ) explicit solvation model.[97] As discussed above, the drawback of all of these approaches is that they require the repetition of the EOM-CCSD calculation many times to achieve statically meaningful conformational averages. For continuum models, the first work with CC methods was presented by Christiansen and Mikkelsen[112,112] with a simplified multipole model. In recent years, Cammi has presented various developments related to

PCM.[106,107,109]. Ehara and coworkers, in collaboration with Cammi, also interfaced PCM with the symmetry adapted cluster-configuration interaction (SAC-CI) method,[149,150] which shares many features with EOM-CCSD.

The goal of this work is to test how reliable EOM-CCSD-PCM calculations are for the evaluation and prediction of excitation energies of solvated molecules. We accomplish this by comparing calculated and measured excitation energies, and by discussing the possible sources of disagreement between the two. These sources include the lack of vibronic effects, and the proper description of H-bonding. Both can be significant, but the former requires an intense computational effort at the CC level. On the other hand, the lack of H-bonding interactions in continuum models can be addressed with a simple and computationally inexpensive approach described below. A main outcome of this work is that the average error of EOM-CCSD- PCM is about 0.2 eV larger than for isolated molecules,[84,85] which is not surprising considering the solvation model employed. However, the error is consistent in its sign (i.e., overestimation of experiment), which is an essential feature for predictive studies. Additionally, such error can be considerably reduced by computing a H-bond shift via DFT microsolvation calculations. The key advantages of computing such a shift with DFT are that only one EOM-CCSD calculation is required and that the shift is largely independent of the choice of the functional.

The chapter is structured as follows. Section 3.2 provides some theoretical background and reports the details of the calculations. Section 3.3 contains all of the calculations outcomes and analysis. Finally, Section 3.4 summarizes the findings of the study and reports concluding remarks.

3.2 Theory and Computational Details

The calculations were performed using standard linear response (LR) techniques[151–156] for the solute (EOM-CCSD and LR-CCSD provide the same excitation energies[91–99]). We used the LR approach also for PCM,[157] where the solvent response depends on the solute transition density. Since we consider vertical transitions, all calculations were performed within the non-equilibrium solvation regime,[30] which assumes that only the solvent electrons move fast enough to respond to

changes in the solute electron density (the so-called "dynamic" response) while the solvent nuclei are fixed in the equilibrium position with the solute electronic ground state (i.e., the "inertial" response).

We consider five popular approximate functionals together with EOM-CCSD. Three functionals are global hybrids: B3LYP,[158] PBE0,[159,160] and M06;[161] while two are range-separated hybrids: CAM-B3LYP[162] and LC-ωPBE.[163–165] This list is certainly far from complete, but our goal is to test EOM-CCSD, and the DFT results are used for comparison and for the micro-solvation calculations that were too large for EOM-CCSD. We refer to previous work of Jacquemin, Adamo, and coworkers for a thorough benchmarking of DFT methods.[86–90] We note that some of the DFT results differ from those in Refs. 86–90 because we used an alternative definition of the PCM cavity radii, but the effect is minor.

We tested five double- and triple-ζ (i.e., number of basis functions for each valence atomic orbital) basis sets: TZVP, aug-cc-pVDZ, 6-31++G**, 6-311++G**, 6-311(3+,3+)G**, which include polarization and diffuse functions. While DFT is not very sensitive to the choice of basis set, EOM-CCSD is, and these sets represent a good compromise between computational cost and accuracy. TZVP (a split valence triple zeta basis set without diffuse functions developed by Ahlrichs and coworkers[166]) was used by Kánnár and Szalay when they compared EOM-CCSD and CC3 results in gas phase,[84] while 6-311(3+,3+)G** (with three sets of diffuse functions) was used by Caricato *et al.* when they tested DFT and EOM-CCSD against experimental gas phase data.[85,100,101] The correlation-consistent series is often a preferred choice with CC methods, but we could not go beyond the double-ζ set (aug-cc-pVDZ[167]) due to the size of the molecules considered. Finally, the remaining Pople sets were chosen for their reduced size compared to 6-311(3+,3+)G**. As discussed below, the choice of basis set in solution is not necessarily the same as in gas phase calculations since we need to take into account the escaped charge error, i.e., the error due to the fraction of solute electron density that extends beyond the PCM cavity.[30]

For PCM, we used the symmetric version of the integral equation formalism (IEF-PCM)[168] approach with non-electrostatic contributions from the solvation model based on density (SMD).[169]

These non-electrostatic terms do not affect the excitation energy, but enter the ground state geometry optimization. We also used SMD radii for all calculations. For the excitation energy calculations, we included added spheres in the cavity to avoid unphysical solvent pockets, especially in the micro-solvation calculations.

All ground state geometries were optimized using B3LYP[158] with the cc-pVTZ basis set. The same geometries were then used for the excitation calculations with all other levels of theory in order to avoid geometry effects on the results. The geometrical changes across methods should be relatively small since the test molecules are relatively rigid. All calculations were performed with a Development Version of the GAUSSIAN suite of programs.[170]

3.3 Results and Discussion

We chose a test set of 16 molecules divided in four groups according to the principal structural moiety. The molecules are shown in Figure 3.1, and the four groups are: nitroso, AB, NQ, and U (for uracil). We chose medium size molecules that were affordable with EOM-CCSD with all basis sets, and for which experimental data in polar solvents is available.[171–175] The effect of non-polar solvents on electronic energies is either small or due to non-electrostatic interactions, and in either case using PCM is ineffective. Most of the solvents in this work are protic and can form hydrogen bonds, which are also not well reproduced with PCM. However, one can partially account for H-bonding using a simple micro-solvation + continuum approach, where the H-bond sites are saturated with few solvent molecules and the entire cluster is embedded in the continuum dielectric.

The comparison with experiment is reported in terms of signed average errors stacked on top of each other to form a cumulative error bar. The averaging is done in terms of the total number of molecules (we consider one state per molecule). The total average error is:

$$\Delta\bar{\omega} = \sum_{i=1}^{N} |\Delta\bar{\omega}^i| = \sum_{i=1}^{N} \frac{|\omega^i_{cal} - \omega^i_{exp}|}{N} \tag{3.1}$$

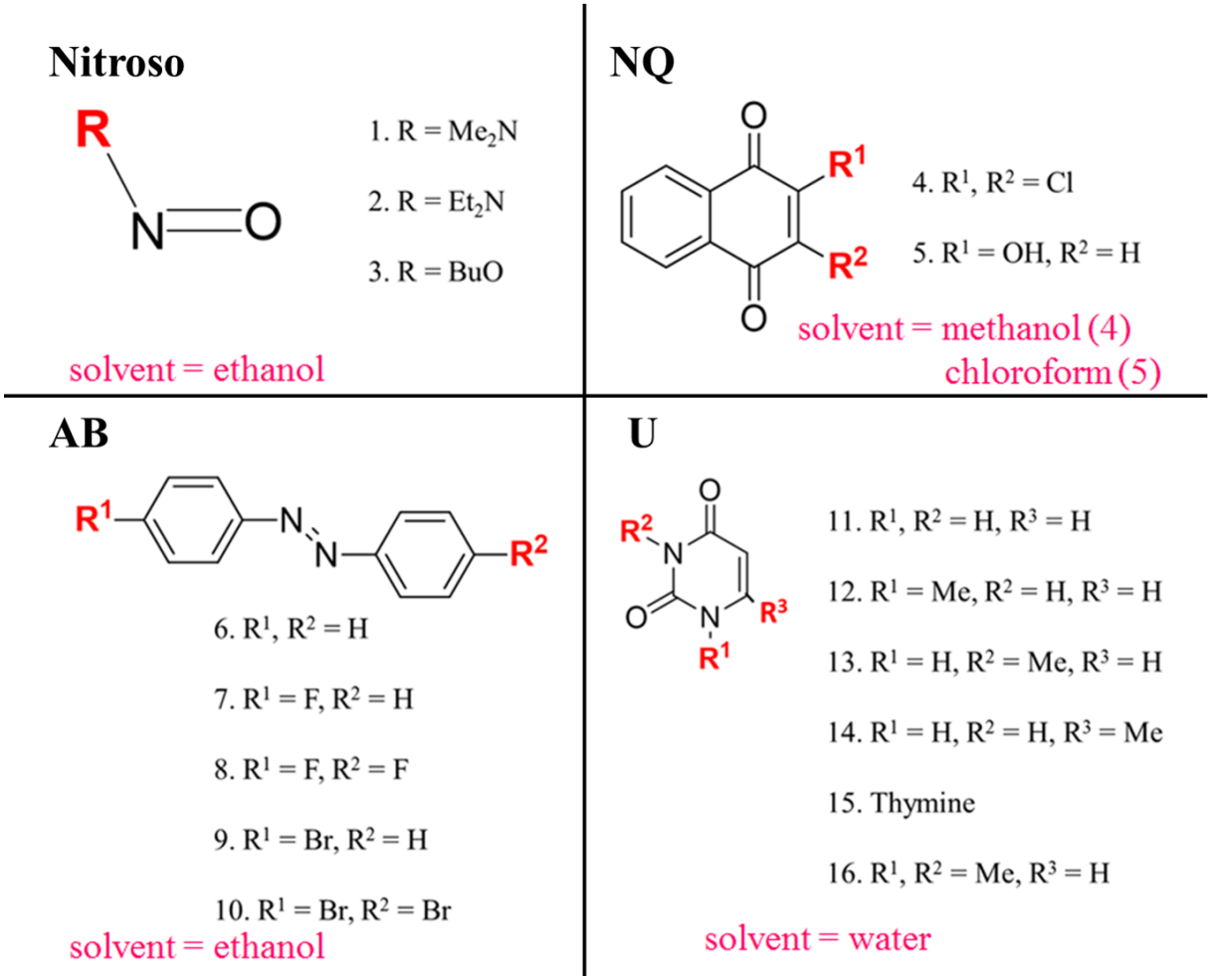

Figure 3.1: Test set used for the benchmarking. This includes 16 molecules divided in 4 groups: Nitroso, NQ, AB, and U, and the solvents in which measurements were performed.

where ω^i_{cal} and ω^i_{exp} are the calculated and experimental excitation energies for compound i, respectively, and N is the total number of molecules. The experimental and calculated excitation energies, and the optimized geometries are reported in the supporting material (SI).

The error bars for all of the methods and basis sets are reported in Figure 3.2. The trend across basis sets is pretty similar for all methods, with EOM-CCSD showing the largest sensitivity whereas the functionals are rather insensitive, as for isolated molecules. For EOM-CCSD, the sets that provide the largest errors are TZVP and 6-311(3+,3+)G** with $\Delta\bar{\omega} > 0.5$ eV. The reason is that TZVP does not contain diffuse functions that are important for excitation energies, while 6-311(3+,3+)G** contains too many diffuse functions that lead to the escaped charge problem. The

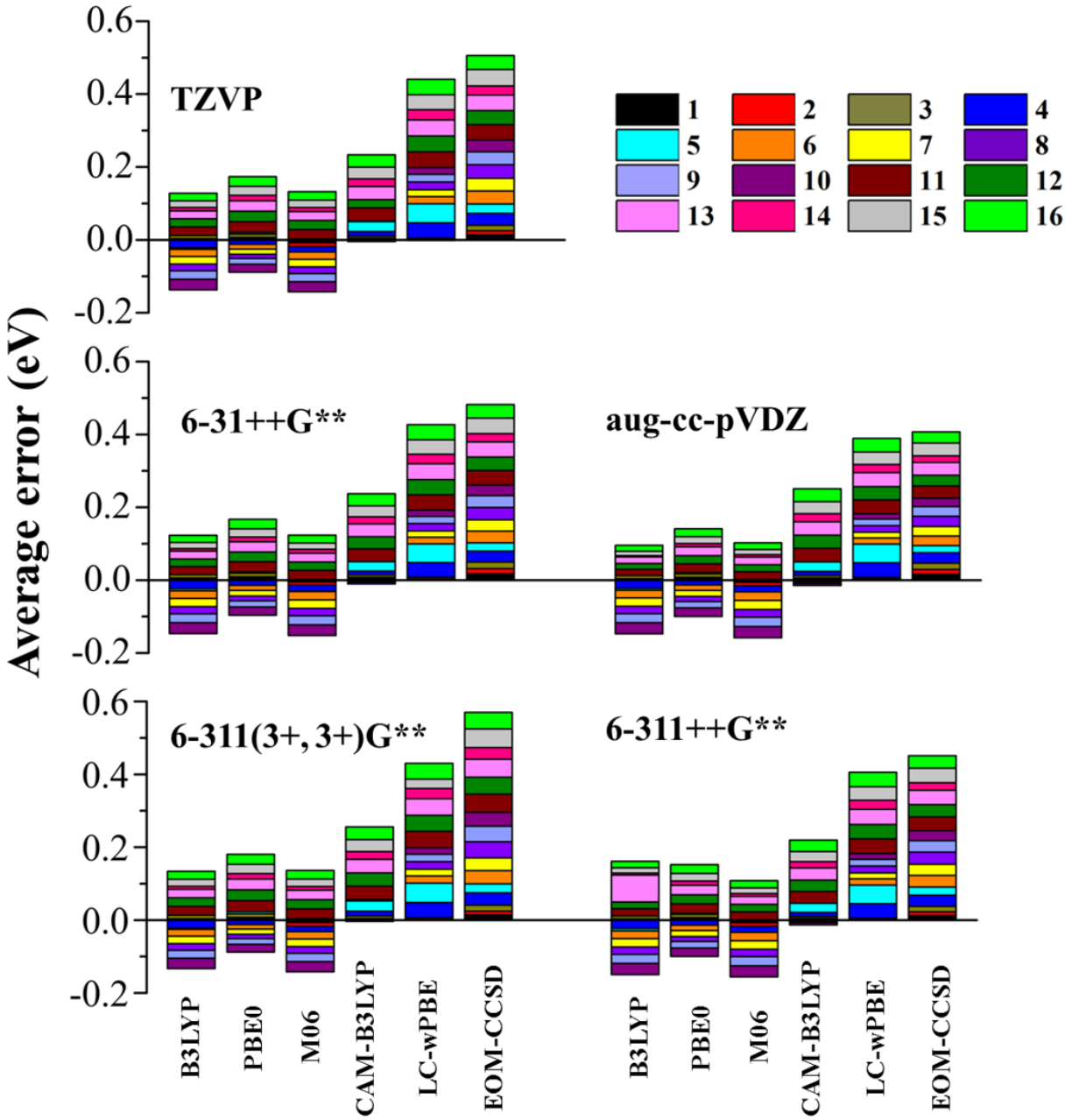

Figure 3.2: Stacked average errors of the excitation energies (eV) calculated with all methods and basis sets. The molecule enumeration follows the order in Figure 3.1.

best results are obtained with 6-311++G** and aug-cc-pVDZ ($\Delta\bar{\omega} \simeq 0.4$ eV), while 6-31++G** is slightly worse because double-ζ Pople basis sets are rather small. It would be interesting to compare EOM-CCSD results obtained with aug-cc-pVTZ and -pVQZ sets, but these are too large for our computational resources. However, it is likely that some of the diffuse functions in these

large sets may contribute to the escaped charge problem as seen in 6-311(3+,3+)G*, and should be pruned.

The remarkable result in Figure 3.2 is the worse performance of EOM-CCSD compared to all DFT methods. Such poor performance can be only attributed to the approximations introduced in the solvation model (assuming that the experimental values are accurate), since there is no reason to believe that EOM-CCSD would represent the solute electronic density any worse than in gas phase. The approximations introduced in this comparison are several: we are comparing calculated vertical excitation energies with experimental maxima of absorption, the calculations neglect direct solute-solvent interactions like hydrogen bonds, and they lack non-electrostatic interactions. As discussed above, the last term should not represent a considerable source of error for polar solvents. To account for the first term, we should include dynamical solvation effects, which is typically done by computing and averaging the excitation energies of many explicit solute and solvent configurations obtained from a molecular dynamics (MD) simulation. However, using such an approach with EOM-CCSD is particularly costly because it requires a considerable number of CC calculations in order to obtain a proper averaging, even when a classical model is used for the surrounding solvent molecules. The effect of hydrogen bonding, on the other hand, can be investigated with the micro-solvation model introduced above.

Before we describe the micro-solvation results, it is instructive to compare the EOM-CCSD error trends with those of the functionals. EOM-CCSD always overestimates experiment, consistently with gas phase[84] and previous LR PCM calculations[102]. This is an advantage since one always knows the direction of the error when making predictions on new compounds. Only LC-ωPBE shows the same trend, and to a large extent also CAM-B3LYP (except for $Me_2N{-}N{=}O$). These are the two range-separated functionals, which display a better long range behavior. LC-ωPBE provides a performance that is only slightly better than that of EOM-CCSD, and less basis set dependent. CAM-B3LYP provides a better performance, with $\Delta\bar{\omega} \simeq 0.25$ eV. The three global functionals provide an overall good performance in terms of total average error: $0.2 \leq \Delta\bar{\omega} \leq 0.3$ eV, but experiment may be over- or underestimated, thus making the use of these methods for pre-

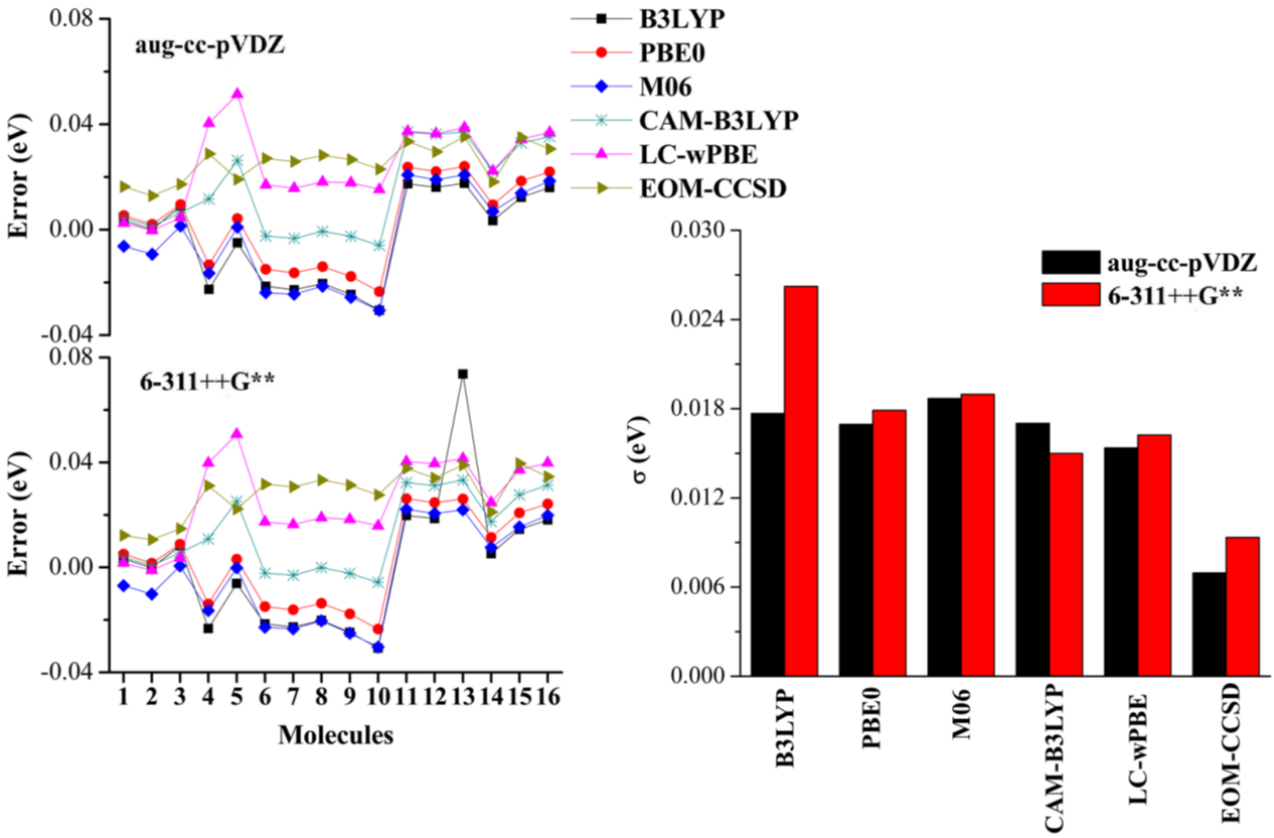

Figure 3.3: On the left: Errors (eV) for all molecules and all methods with the two best basis sets, aug-cc-pVDZ and 6-311++G**. On the right: the corresponding error standard deviations.

dictive studies inadvisable. Another important advantage of EOM-CCSD compared to DFT is the smaller error spread. This is shown in Figure 3.3, which reports the signed, unscaled error for all of the methods with the two best basis sets, aug-cc-pVDZ and 6-311++G**, and the corresponding standard deviation. The figure shows how for EOM-CCSD the error standard deviation is 2-3 times smaller than that for any functional.

In order to study the effect of hydrogen bonding, we first consider the 2,3−Cl NQ molecule, whose experimental data was measured in methanol. We built micro-solvated clusters with up to four solvent molecules forming H-bonds with the carbonyl groups. All of the micro-solvated clusters are also embedded in PCM. We define a H-bond shift Δ_{HB} with respect to the PCM-only calculations:

$$\Delta_{HB} = \omega_{HB} - \omega_{PCM} \tag{3.2}$$

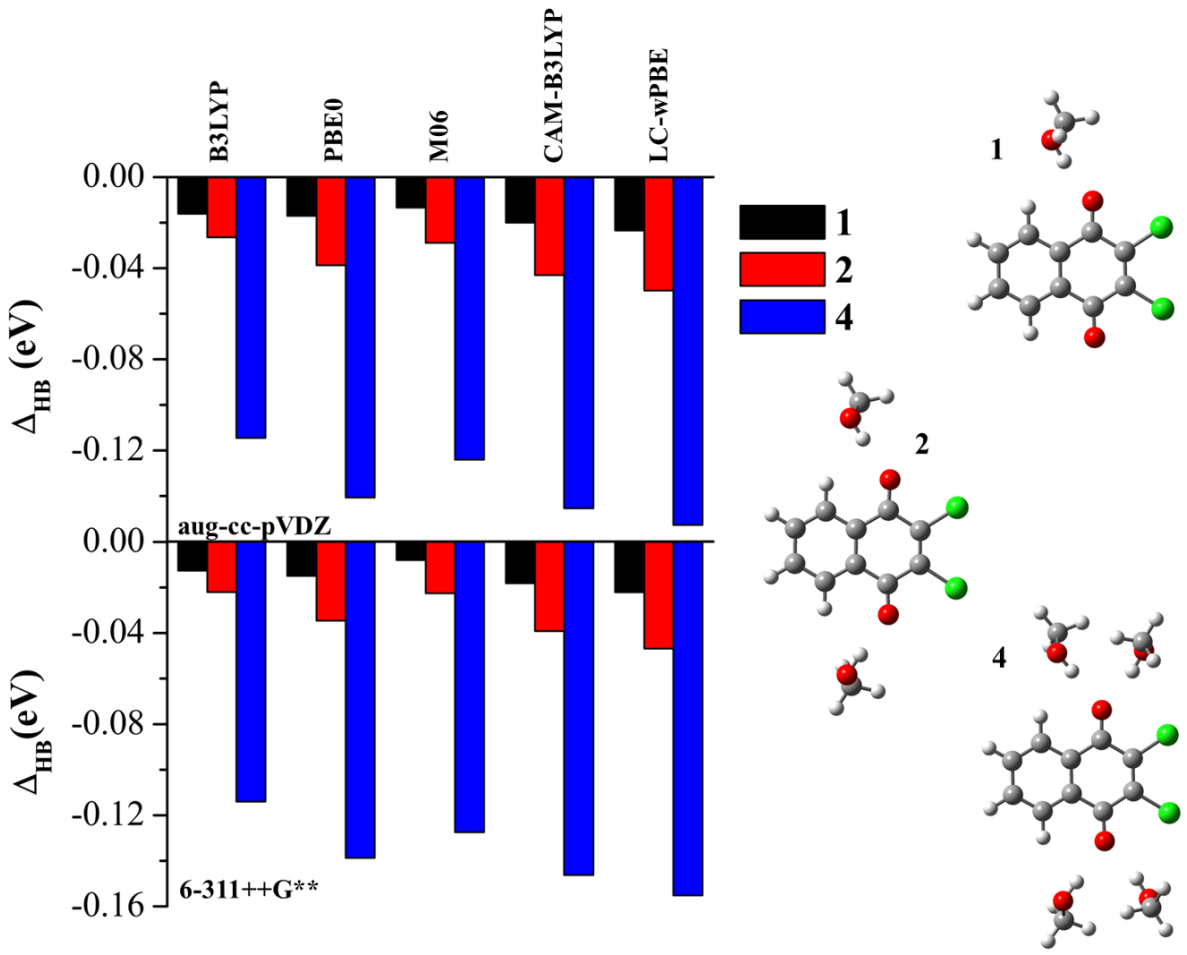

Figure 3.4: Solvent shifts for the excitation energy (eV) of 2, 3-Cl NQ after adding 1, 2, and 4 explicit methanol molecules. The shifts are computed as in Eq. 3.2.

where ω represents the excitation energy, the subscript *HB* refers to the explicit solvent + PCM system, and the subscript *PCM* refers to the PCM-only system. The results for this solute with all of the functionals are shown in Figure 3.4 for the two basis sets that work best with EOM-CCSD: aug-cc-pVDZ and 6-311++G**, which is in agreement with a recent work by Laurent and coworkers.[176] The figure also reports the optimized geometries for the clusters. This data shows a decrease in the excitation energy when H-bonds are treated explicitly. The red shift is particularly pronounced when all of the H-bond sites are saturated: the average across all functionals is 0.14 eV. An important result in this figure is that all of the functionals provide a rather consistent value of the shift, with deviations of at most 0.02 eV from the average.

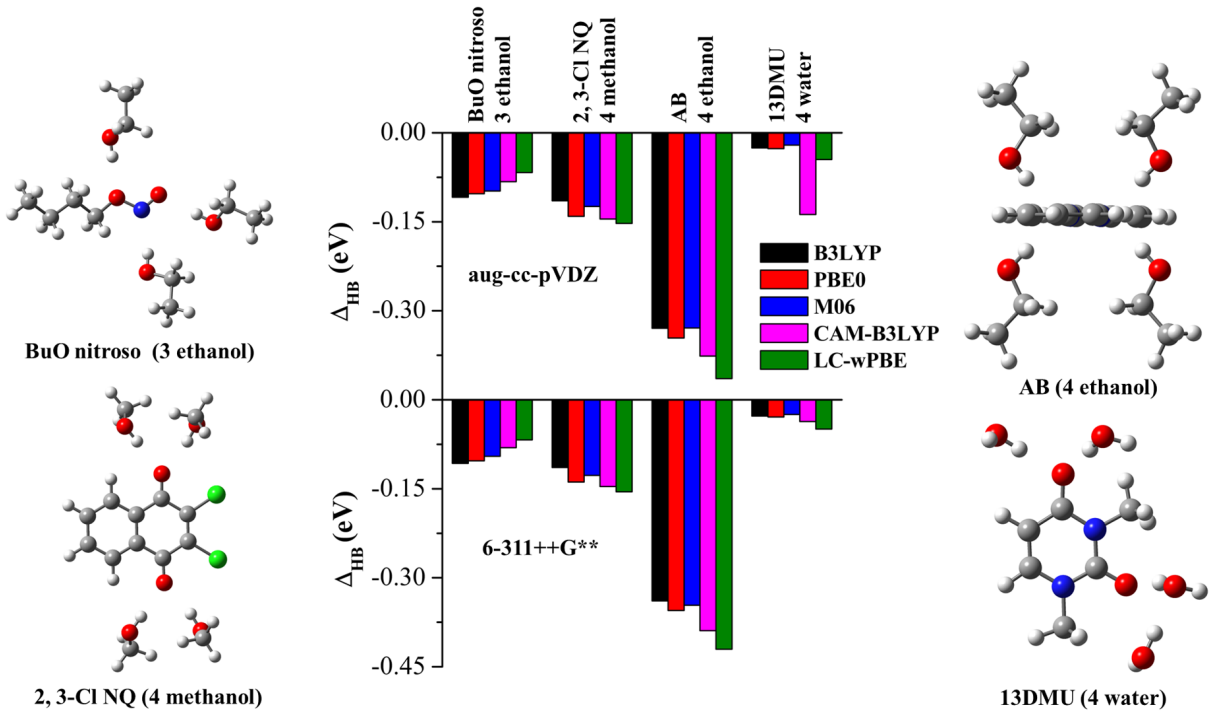

Figure 3.5: Solvent shifts for the excitation energy (eV) of one molecule for each group in Figure 3.1 after saturation of the H-bond sites with explicit solvent molecules. The shifts are computed as in Eq. 3.2.

Therefore, we selected one member from each of the four molecular groups and computed the shifts Δ_{HB} by saturating the H-bond sites (using the same two basis sets). The results, shown in Figure 3.5 together with the optimized micro-solvated structures, confirm the trends of the first molecule: Δ_{HB} is always negative and its value is fairly consistent across functionals (with the exception of the 13DMU molecule with CAM-B3LYP and the aug-cc-pVDZ basis set). The shift can be small as for 13DMU, 0.04 eV on average, medium as for BuO and NQ, 0.1 eV on average, and rather large as for AB, 0.4 eV on average. If we apply the average shift Δ_{HB} to the EOM-CCSD-PCM results, we obtain a considerable improvement of the agreement with experiment. This is shown in Figure 3.6, which reports the error with and without Δ_{HB}. The corrected EOM-CCSD-PCM results get closer to what is expected from gas phase EOM-CCSD calculations. When the shift is small, as for 13DMU, it is an indication that dynamical effects (which originate from not averaging over many solvent and solute configurations) are important, and an explicit

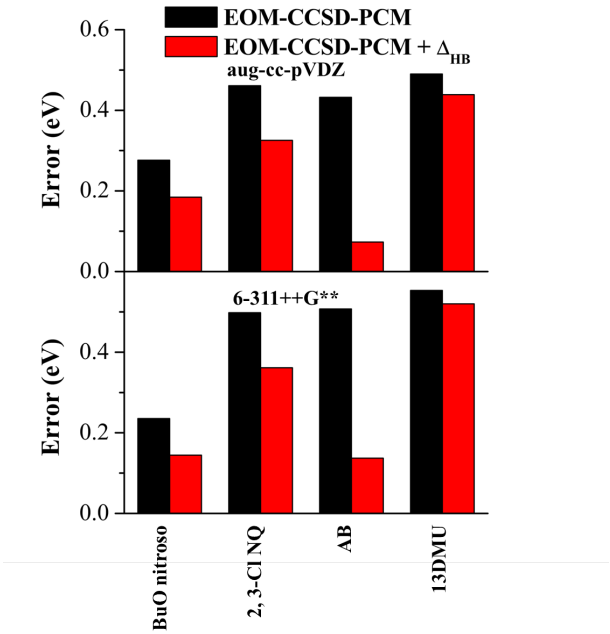

Figure 3.6: Errors in excitation energy (eV) for EOM-CCSD-PCM with and without Δ_{HB} for the four molecules considered in Figure 3.5.

MD simulation may be necessary. It is noteworthy that adding a negative shift would worsen the performance of those functionals that already underestimate experiment with a continuum-only approach, thus providing more evidence that their performance relies on cancellation of errors. However, one must be careful that improving the description of the solvent may in fact worsen the DFT agreement with experiment.

3.4 Conclusions

In this work, we compare electronic excitation energies of solvated molecules computed at the EOM-CCSD-PCM level with various basis sets and experimental results for a series of 16 molecules. All of the solvents are polar so that electrostatic contributions play a central role. We also include DFT results from a subset of popular global and range-separated hybrid functionals. These results confirm some trends that are common to EOM-CCSD calculations in gas phase: a considerable basis set dependence compared to DFT, and a consistent overestimation of experiment. The latter

point is a positive feature, as one is always aware of the error direction when making predictions. The error is due to the intrinsic limitations of this approach: incompleteness of the electron correlation description and lack of vibronic effects. The basis set issue is more delicate than in gas phase, as one needs to take into account the escaped charge problem so that basis sets with too many diffuse functions perform as badly as those without diffuse functions. The overall error of EOM-CCSD in solution is, not surprisingly, larger than in gas phase by about 0.2 eV on overage. This is due to the additional approximations introduced by the solvation model.

To demonstrate the last point, we investigate the effect of hydrogen bonding on the excitation energy with a micro-solvation + PCM approach, where the H-bond sites are saturated with explicit solvent molecules and the entire cluster is embedded in the continuum dielectric. These calculations, performed entirely at DFT level, uniformly show a decrease of the excitation energy, i.e. in the direction that would increase the agreement of the CC calculations with experiment. Furthermore, this preliminary data suggests that the EOM-CCSD performance in solution can be improved through a computationally inexpensive route when protic solvents are involved, i.e. performing a DFT geometry optimization and subsequent LR calculation at the same level of theory for a micro-solvated + continuum solvation. Computing Δ_{HB} as in Eq. 3.2, and adding it to the PCM-only EOM-CCSD excitation energy provides a more reliable estimate of the excitation. It is remarkable that Δ_{HB} is rather independent of the nature of the functional, contrary to the absolute value of the excitation energy itself. This observation greatly reduces the burden of the proper choice of functional, which is one of the greatest drawbacks of modern DFT. This simple protocol has the advantage that only one expensive EOM-CCSD calculation is required, thus minimizing the computational effort. Obviously, H-bonds are not the only factor that can influence the excitation energy, as shown by the 13DMU case, and in those situations a proper averaging over many MD snapshots may be necessary, although we hypothesize that even in that case a shift from a continuum-only picture can be computed employing only DFT. We are working towards verifying this conjecture. In summary, EOM-CCSD-PCM + Δ_{HB}(DFT) can provide reliable predictions of excitation energies of solvated systems, with the added bonus that, contrary to DFT, one is confi-

dent that the simulations can be improved by improving the solvation model.

3.5 Acknowledgment

The authors thank the support from the General Research Fund and startup fund of the University of Kansas.

3.6 Supporting information available

See the supporting information for the values of the excitation energies of all systems, and their optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

Chapter 4

Coupled Cluster Theory with Induced-Dipole Polarizable Embedding for Ground and Excited States

(This work is taken from: Sijin Ren, Filippo Lipparini, Benedetta Mennucci, and Marco Caricato, *J. Chem. Theory Comput.* **2019**, *15*, 4485-4496.[177] Supporting information is available online.)

Abstract

In this work, we present the theory and implementation of the coupled cluster single and double excitations (CCSD) method combined with a classical polarizable molecular mechanics force field (MMPol) based on the induced dipole model. The method is developed to compute electronic excitation energies within the state specific (SS) and linear response (LR) formalisms for the interaction of the quantum mechanical and classical regions. Furthermore, we consider an approximate expression of the correlation energy, originally developed for CCSD with implicit solvation models, where the interaction term is linear in the coupled cluster density. This approximation allows to include the explicit contribution of the environment to the CC equations without increasing the computational effort. The test calculations on microsolvated systems, where the CCSD/MMPol method is compared to full CCSD calculations, demonstrates the reliability of this computational protocol for all interaction schemes (errors $< 2\%$). We also show that it is important to include induced dipoles on all atom centers of the classical region, and that too diffuse functions in the basis set may be problematic due to too strong interaction with the environment.

4.1 Introduction

Hybrid methods allow one to describe large systems in complex environments by treating the core region at an accurate level of theory and the rest with a less computationally intensive method. Typically, the core region is treated at a quantum mechanical (QM) level and the environment is treated classically using two common strategies: 1) a continuum dielectric such as the polarizable continuum model (PCM)[28,29] for bulk solvation, or 2) atomistic molecular mechanics force fields (MM).[1,11–14] Most approaches use an effective Hamiltonian (H_{eff}) to define the electronic properties of the QM region embedded in the classical environment. In the most basic form, H_{eff} includes a term that describes the coupling of the QM region with the environment through a Coulomb interaction with the embedding field, often represented in MM methods by fixed point charges. Implicit and explicit embedding models have advantages and disadvantages: the former automatically take into account mutual polarization with the environment and are more computationally efficient because the conformational averaging of the medium is implicit in the dielectric constant, but they cannot describe inhomogeneous environments or direct interactions with the QM region (such as H bonding); explicit models are more realistic and can reproduce all sorts of environments, but mutual polarization can be computationally costly and averaging over multiple configurations is required to obtain statistically meaningful results. Thus, the proper choice of embedding depends on the problem at hand, but it is desirable to have both in one's computational toolbox.

Classical polarizable force fields (MMPol) can be designed in multiple ways,[15–27] and this work focuses on the induced dipoles scheme.[2,4,5,104,105,133,178–183] In this QM/MMPol approach, part or all classical centers are endowed with a point dipole, whose magnitude and direction is determined by the electric field generated by the QM and MM regions (including the other induced dipoles), thus providing mutual polarization with the system core. Some of us presented an efficient implementation of this approach for electronic ground and excited states with time-dependent density functional theory (TD-DFT).[2–10]

The scope of this work is to extend this implementation to coupled cluster methods,[123,124] in particular the version that includes single and double excitations (CCSD) and the equation of

motion/linear response (EOM/LR) variants for excited states.[91,92] The combination of CC with induced multipole embedding was first presented by Kongsted and coworkers in the context of linear response (LR) theory.[5,19,22,27,103–105] However, in this contribution we go beyond the LR formalism to include state specific (SS) solvation, where the mutual polarization depends on the excited state density rather than just on the ground state density response. More importantly, the similarity of the functional form of the CCSD/MMPol and CCSD/PCM equations,[95,99,103,106–109,111,112] discussed in section 4.2, allows us to employ the approximate schemes that some of us have developed for continuum models.[95,99,108,110,118] In fact, one of the complications of CC/polarizable embedding theory is the considerable increase of the cost of the calculation compared to the gas phase theory (i.e., the prefactor of the scaling increases by a factor between 2 and 5). The approximations we developed for CCSD/PCM allow us to obtain accurate results for CCSD/MMPol at a computational cost virtually identical to that of gas phase CCSD.[95,118] We test these schemes on a series of microsolvated systems, where we can compare QM/MMPol against full QM excitation energies to determine the hybrid protocol with the best compromise between accuracy and computational cost.

The chapter is organized as follows. The theory is presented in section 4.2, the computational details of the test calculations are described in section 4.3, the results are discussed in section 4.4, and concluding remarks are reported in section 4.5.

4.2 Theory

In this section, we present the theory for combining a polarizable classical molecular mechanics force field (MMPol)[2,4,5,104,105,133,178–183] with the coupled cluster singles and doubles (CCSD) method and its excited state versions, i.e., the equation of motion (EOM)[92] and linear response (LR)[91] approaches. In this version of MMPol, the environment polarization is created by including induced dipoles on specific classical sites, although very similar equations apply to other types of polarizable force fields (e.g., fluctuating charges and higher induced multipoles). We report a summary of the main MMPol equations for completeness, but a more detailed treatment can be

found in Refs. 95,103,104,108,110. In the following, the "gas phase" expression is loosely defined as the absence of MMPol embedding. In a system that is partitioned into a core region described at a quantum mechanical (QM) level while the rest is described at MM level, the energy can be expressed as:[3,113,114]

$$\begin{aligned}\varepsilon(P,\mu) &= \varepsilon^{QM}(P) + \varepsilon^{MM}(\vec{\mu}) + \varepsilon^{Coup}(P,\vec{\mu}) \\ &= \varepsilon^{QM}(P) + \varepsilon^{Env}(P,\vec{\mu})\end{aligned} \tag{4.1}$$

where P and $\vec{\mu}$ indicate the electronic density and induced dipoles, respectively, $\varepsilon^{QM}(P)$ is the energy of the QM region, $\varepsilon^{MM}(\vec{\mu})$ is the energy of the MM region, and $\varepsilon^{Coup}(P,\vec{\mu})$ is the interaction energy between the QM and MM regions. The last two terms in Eq. 4.1 represent the environment energy, $\varepsilon^{Env}(P,\vec{\mu})$. The energy also depends on the QM nuclear charges and the fixed contributions in the MM force field, e.g. point charges, but since these do not vary during the solution of the CCSD/MMPol equations, we do not mention them explicitly. The environment energy can be further expanded into three parts:

$$\varepsilon^{Env}(P,\vec{\mu}) = \varepsilon^{FF} + \varepsilon^{QM/MM}(P) + \varepsilon^{Pol}(P,\vec{\mu}) \tag{4.2}$$

where ε^{FF} is a constant term that corresponds to the energy of the static force field, $\varepsilon^{QM/MM}(P)$ is the interaction energy between the QM density and the fixed MM field, and $\varepsilon^{Pol}(P,\vec{\mu})$ is the polarization energy that depends on the induced dipoles:

$$\varepsilon^{Pol}(P,\vec{\mu}) = -\vec{\boldsymbol{\mu}} \cdot \left[\vec{\boldsymbol{E}}^{MM} + \vec{\boldsymbol{E}}^{QM}(P)\right] + \frac{1}{2}\vec{\boldsymbol{\mu}} \cdot \boldsymbol{T}\vec{\boldsymbol{\mu}} \tag{4.3}$$

which includes the interaction of the induced dipoles with the electric field generated by the QM and MM regions on the dipole sites, and the induced dipole/induced dipole interactions. The bold

font collectively indicates all polarizable sites. The interaction matrix $\boldsymbol{T}$ takes the form:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{\alpha}_1^{-1} & \mathcal{T}_{12} & \cdots & \mathcal{T}_{1N_{pol}} \\ \mathcal{T}_{21} & \boldsymbol{\alpha}_2^{-1} & \cdots & \mathcal{T}_{2N_{pol}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{T}_{N_{pol}1} & \mathcal{T}_{N_{pol}2} & \cdots & \boldsymbol{\alpha}_{N_{pol}}^{-1} \end{bmatrix} \tag{4.4}$$

where $\mathcal{T}_{ij}$ is the classical damped interaction tensor between sites i and j, N_{pol} is the number of polarizable sites, and $\boldsymbol{\alpha}_i$ is the polarizability tensor of site i, which is taken as the isotropic value and is a parameter of the force field. The polarization energy is computed by minimizing Eq. 4.3 with respect to the induced dipoles:

$$\begin{aligned} \frac{\partial \varepsilon^{Pol}}{\partial \vec{\boldsymbol{\mu}}} &= \boldsymbol{T}\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{E}}^{MM} - \vec{\boldsymbol{E}}^{QM}(P) = 0 \\ &\Longrightarrow \boldsymbol{T}\vec{\boldsymbol{\mu}} = \vec{\boldsymbol{E}}^{MM} + \vec{\boldsymbol{E}}^{QM}(P) \end{aligned} \tag{4.5}$$

Combining Eq. 4.5 and Eq. 4.3, we get:

$$\varepsilon^{Pol} = -\frac{1}{2}\vec{\boldsymbol{\mu}} \cdot \left(\vec{\boldsymbol{E}}^{MM} + \vec{\boldsymbol{E}}^{QM}(P) \right) = -\frac{1}{2}\vec{\boldsymbol{\mu}} \cdot \vec{\boldsymbol{E}} \tag{4.6}$$

4.2.1 CCSD/MMPol Electronic Ground State

The expression for ε^{Pol} in Eq. 4.6 takes a form similar to the PCM energy,[28,29] with the surface charges replaced by the induced dipoles and electrostatic potential replaced by the electrostatic field. Because of this analogy, the CCSD/MMPol equations are basically identical to CCSD/PCM equations, and the same approximate schemes can be employed.[95,99,103,106–112,118] Thus, we only report the key equations, and refer the reader to Refs. 95,106–110,118 for a complete derivation.

The ground state Lagrangian with MMPol embedding can be expressed as the sum of a term

that includes the gas phase QM and the fixed MM parts, and a polarization term:

$$\mathcal{G}_0^{\mu} = \mathcal{G}_0^{fix} - \frac{1}{2}\vec{\boldsymbol{\mu}}_0 \cdot \vec{\boldsymbol{E}}_0 \tag{4.7}$$

This expression is quadratic in the CCSD density through the last term, and it is usually referred to as PTED scheme (from perturbation theory energy and density). The ground state induced dipoles and electric field, $\vec{\boldsymbol{\mu}}_0$ and $\vec{\boldsymbol{E}}_0$, are:

$$\vec{\boldsymbol{\mu}}_0 = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) e^{-\hat{T}} \hat{\boldsymbol{\mu}} e^{\hat{T}} \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{\boldsymbol{\mu}} \middle| \Phi^0 \right\rangle \tag{4.8}$$

$$\vec{\boldsymbol{E}}_0 = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) e^{-\hat{T}} \hat{\boldsymbol{E}} e^{\hat{T}} \middle| \Phi^0 \right\rangle = \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{\boldsymbol{E}} \middle| \Phi^0 \right\rangle \tag{4.9}$$

where $\hat{T}$ and and $\hat{\Lambda}$ are the CC excitation and de-excitation operators, respectively, and the overbar indicates similarity transformation.[123,124] By separating the reference wave function contribution to $\mathcal{G}_0^{\mu}$ from the correlation part through the normal-product form of an operator:[123,124]

$$\hat{X}_N = \hat{X} - \left\langle \Phi^0 \middle| \hat{X} \middle| \Phi^0 \right\rangle = \hat{X} - X^0 \tag{4.10}$$

Eq. 4.7 can be rewritten as:

$$\mathcal{G}_0^{\mu} = \mathcal{G}^0 + \left\langle \Phi^0 \middle| (1+\hat{\Lambda}) \bar{H}_N^{\mu} \middle| \Phi^0 \right\rangle - \frac{1}{2}\vec{\boldsymbol{\mu}}_N \cdot \vec{\boldsymbol{E}}_N \tag{4.11}$$

where $\mathcal{G}^0$ is the reference free energy, $\bar{H}_N^{\mu}$ is the QM Hamiltonian including the fixed embedding and the dipole embedding induced by the reference density, and $\vec{\boldsymbol{\mu}}_N$ are the dipoles induced by the correlation density ($\vec{\boldsymbol{\mu}}_N = \vec{\boldsymbol{\mu}}_0 - \vec{\boldsymbol{\mu}}^0$).

The free energy for the ground state is evaluated by minimizing the Lagrangian in Eq. 4.11 with respect to the $\hat{T}$ and $\hat{\Lambda}$ amplitudes, whose explicit expressions are reported in Eqs. S1-S2 of the Supporting Information (SI). Since Eq. 4.11 is quadratic in $\hat{T}$ and $\hat{\Lambda}$ due to the induced dipole term, and contrary to the gas phase case, Eqs. S1-S2 are coupled and must be solved simultaneously. At

convergence Eq. 4.11 reduces to:

$$\mathcal{G}_0^{\mu} = \mathcal{G}^0 + \Delta\varepsilon_0 + \frac{1}{2}\vec{\boldsymbol{\mu}}_N \cdot \vec{\boldsymbol{E}}_N \tag{4.12}$$

where

$$\Delta\varepsilon_0 = \left\langle \Phi^0 \middle| \bar{H}_N^{\mu} \middle| \Phi^0 \right\rangle - \vec{\boldsymbol{\mu}}_N \cdot \left\langle \Phi^0 \middle| \bar{\boldsymbol{E}}_N \middle| \Phi^0 \right\rangle \tag{4.13}$$

4.2.2 State Specific Formalism for Electronic Excited States

There are two strategies for including the polarizable embedding effect in excited state calculations, called state specific (SS) and linear response (LR) formalisms. The difference between the two is discussed in detail in Refs. 130,131, and it can be summarized by saying that in the SS formalism the environment response depends on the excited state density, while in the LR formalism it depends on the transition density. We implemented both approaches, and we report the main equations in the following.

In the SS formalism, the excited state Lagrangian can be written in a form equivalent to that of the ground state, where the polarizable quantities depend on the excited state density:

$$\vec{\boldsymbol{\mu}}_{KN} = \left\langle \Phi^0 \middle| \hat{L}_K \left[\bar{\boldsymbol{\mu}}_N, \hat{R}_K \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}_K) \bar{\boldsymbol{\mu}}_N \middle| \Phi^0 \right\rangle \tag{4.14}$$

$$\vec{\boldsymbol{E}}_{KN} = \left\langle \Phi^0 \middle| \hat{L}_K \left[\bar{\boldsymbol{E}}_N, \hat{R}_K \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}_K) \bar{\boldsymbol{E}}_N \middle| \Phi^0 \right\rangle \tag{4.15}$$

where $\hat{L}_K$ and $\hat{R}_K$ are the left and right-hand eigenvectors of the EOM-CCSD Hamiltonian for the K-th state. The excited state Lagrangian is:

$$\begin{aligned} \mathcal{G}_K^{\mu} = \mathcal{G}^0 &+ \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}_K) \bar{H}_N^{\mu} \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| \hat{L}_K \left[\bar{H}_N^{\mu}, \hat{R}_K \right] \middle| \Phi^0 \right\rangle \\ &+ \omega_K \left(1 - \left\langle \Phi^0 \middle| \hat{L}_K \hat{R}_K \middle| \Phi^0 \right\rangle \right) - \frac{1}{2} \vec{\boldsymbol{\mu}}_{KN} \cdot \vec{\boldsymbol{E}}_{KN} \end{aligned} \tag{4.16}$$

where ω_K is the eigenvalue associated with $\hat{L}_K$ and $\hat{R}_K$. The free energy for the K-th excited state

can be computed by minimizing the Lagrangian in Eq. 4.16 with respect to all sets of amplitudes, which are all coupled through the environment term that is quadratic in the excited state density. The explicit expressions for the amplitude equations are reported in Eqs. S3-S7 of the SI. At convergence, Eq. 4.16 reduces to:

$$\mathcal{G}_K^{\mu} = \mathcal{G}^0 + \Delta\varepsilon_K + \omega_K + \frac{1}{2}\vec{\boldsymbol{\mu}}_{KN} \cdot \vec{\boldsymbol{E}}_{KN} \tag{4.17}$$

where:

$$\Delta\varepsilon_K = \left\langle \Phi^0 \middle| \bar{H}_N^{\mu} \middle| \Phi^0 \right\rangle - \vec{\boldsymbol{\mu}}_{KN} \cdot \left\langle \Phi^0 \middle| \bar{\boldsymbol{E}}_N \middle| \Phi^0 \right\rangle \tag{4.18}$$

Since all amplitude equations are coupled in the excited state Lagrangian, the $\hat{T}$ amplitudes are different from those of the ground state. Therefore, the transition energy from the ground to the K-th excited state must be evaluated through two separate calculations as the difference between $\mathcal{G}_K$ in Eq. 4.17 and $\mathcal{G}_0$ in Eq. 4.12.

The PTED formalism described in section 4.2.2 is computationally expensive because all the amplitude equations must be solved simultaneously. To reduce the computational cost, similar to EOM-CCSD/PCM,[95] we introduce an approximate free energy functional for CCSD/MMPol, called PTES scheme, where S stands for singles. The goal is to decouple the ground state from the excited state equations, similar to gas phase EOM-CCSD. We split $\vec{\boldsymbol{\mu}}_{KN}$ in Eq. 4.16 into three contributions:

$$\vec{\boldsymbol{\mu}}_{KN} = \vec{\boldsymbol{\mu}}_N^K + \vec{\boldsymbol{\mu}}_N^T + \vec{\boldsymbol{\mu}}_N^{\Lambda} \tag{4.19}$$

defined as:

$$\begin{aligned}\vec{\boldsymbol{\mu}}_N^K &= \left\langle \Phi^0 \middle| \hat{L}_K \left[\bar{\boldsymbol{\mu}}_N, \hat{R}_K \right] \middle| \Phi^0 \right\rangle \\ \vec{\boldsymbol{\mu}}_N^T &= \left\langle \Phi^0 \middle| \bar{\boldsymbol{\mu}}_N \middle| \Phi^0 \right\rangle \\ \vec{\boldsymbol{\mu}}_N^\Lambda &= \left\langle \Phi^0 \middle| \hat{\Lambda}_K \bar{\boldsymbol{\mu}}_N \middle| \Phi^0 \right\rangle\end{aligned} \tag{4.20}$$

If the same separation is performed for $\vec{\boldsymbol{E}}_{KN}$, the last term in Eq. 4.16 can be written as:

$$\begin{aligned}-\frac{1}{2}\vec{\boldsymbol{\mu}}_{NK} \cdot \vec{\boldsymbol{E}}_{NK} = &-\frac{1}{2}\vec{\boldsymbol{\mu}}_N^K \cdot \vec{\boldsymbol{E}}_N^K - \vec{\boldsymbol{\mu}}_N^K \cdot \vec{\boldsymbol{E}}_N^T - \frac{1}{2}\vec{\boldsymbol{\mu}}_N^T \cdot \vec{\boldsymbol{E}}_N^T \\ &- \vec{\boldsymbol{\mu}}_N^T \cdot \vec{\boldsymbol{E}}_N^\Lambda - \frac{1}{2}\vec{\boldsymbol{\mu}}_N^\Lambda \cdot \vec{\boldsymbol{E}}_N^\Lambda - \vec{\boldsymbol{\mu}}_N^\Lambda \cdot \vec{\boldsymbol{E}}_N^K\end{aligned} \tag{4.21}$$

The last two terms in Eq. 4.21 are responsible for coupling the ground and excited state equations, and they are neglected in the PTES scheme, so that the approximate Lagrangian is:

$$\begin{aligned}\mathcal{G}_K^{S,\mu} = \mathcal{G}^0 &+ \left\langle \Phi^0 \middle| \hat{L}_K \left[\bar{H}_N^\mu, \hat{R}_K \right] \middle| \Phi^0 \right\rangle + \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}_K) \bar{H}_N^\mu \middle| \Phi^0 \right\rangle \\ &+ \omega_K \left(1 - \left\langle \Phi^0 \middle| \hat{L}_K \hat{R}_K \middle| \Phi^0 \right\rangle \right) - \frac{1}{2}\vec{\boldsymbol{\mu}}_N^{TK} \cdot \vec{\boldsymbol{E}}_N^{TK} - \vec{\boldsymbol{\mu}}_N^T \cdot \vec{\boldsymbol{E}}_N^\Lambda\end{aligned} \tag{4.22}$$

where $\vec{\boldsymbol{\mu}}_N^{TK} = \vec{\boldsymbol{\mu}}_N^T + \vec{\boldsymbol{\mu}}_N^K$ and $E_N^{TK} = \vec{\boldsymbol{E}}_N^T + \vec{\boldsymbol{E}}_N^K$. The partial derivative equation for the $\hat{T}$ amplitudes in the PTES scheme can be found in Eqs. S8 of the SI. The $\hat{R}_K$ and $\hat{L}_K$ equations are similar to those in the PTED with $\vec{\boldsymbol{\mu}}_{KN}$ being replaced with $\vec{\boldsymbol{\mu}}_N^{TK}$. At convergence, $\mathcal{G}_K^{S,\mu}$ in Eq. 4.22 can be reduced to:

$$\mathcal{G}_K^{S,\mu} = \mathcal{G}_0^S + \omega_K + \frac{1}{2}\vec{\boldsymbol{\mu}}_N^K \cdot \vec{\boldsymbol{E}}_N^K \tag{4.23}$$

where $\mathcal{G}_0^S$ is the ground state PTES energy:

$$\mathcal{G}_0^S = \mathcal{G}^0 + \left\langle \Phi^0 \middle| \bar{H}_N^\mu \middle| \Phi^0 \right\rangle - \frac{1}{2}\vec{\boldsymbol{\mu}}_N^T \cdot \vec{\boldsymbol{E}}_N^T \tag{4.24}$$

It is important to note that in the ground state, the equations for the $\hat{T}$ and $\hat{\Lambda}$ amplitudes are also

decoupled in the PTES scheme, so that the computational cost is virtually equivalent to that of a gas phase calculation. For excited states with the SS formalism, however, the $\hat{R}_K$ and $\hat{L}_K$ equations are still coupled because of the first term on the right-hand side of Eq. 4.21. Nevertheless, since the ground and excited state amplitude equations are decoupled, excitation energies can be computed in a single calculation.

4.2.3 Linear Response Formalism for Electronic Excited States

In the LR formalism, excitation energies are obtained as the poles of the LR function, which are calculated from the right or left-hand diagonalization of the Jacobian matrix. In gas phase, LR-CCSD and EOM-CCSD provide the same excitation energies but different transition properties.[93,94,125] In the presence of a polarizable environment, LR-CCSD and EOM-CCSD differ also in the excitation energy. The derivation of the CC/MMPol LR function follows that of the CC/PCM method, and more details can be in found in Refs. 98,99,103,107,108,110–112. For the right-hand diagonalization, the transition energy ω_K is found by solving:

$$\left\langle \Phi^n \middle| \left[(\bar{H}_N^{\mu} - \vec{\boldsymbol{\mu}}_N \cdot \bar{\boldsymbol{E}}_N), \hat{R}_K \right] \middle| \Phi^0 \right\rangle - \left\langle \Phi^n \middle| \vec{\boldsymbol{\mu}}_N^{R_K} \cdot \bar{\boldsymbol{E}}_N \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^n \middle| \hat{R}_K \middle| \Phi^0 \right\rangle \tag{4.25}$$

where $\hat{R}_K$ is the right-hand eigenvector and Φ^n is an excited Slater determinant. From the left-hand side, we have:

$$\left\langle \Phi^0 \middle| \hat{L}_K \left[(\bar{H}_N^{\mu} - \vec{\boldsymbol{\mu}}_N \cdot \bar{\boldsymbol{E}}_N), \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle - \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \left[\vec{\boldsymbol{\mu}}_N^{L_K} \cdot \bar{\boldsymbol{E}}_N, \hat{\tau}_n \right] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^0 \middle| \hat{L}_K \middle| \Phi^n \right\rangle \tag{4.26}$$

where $\hat{L}_K$ is the left-hand eigenvector and $\hat{\tau}_n$ is an elementary excitation operator: $\hat{\tau}_n |\Phi^0\rangle = |\Phi^n\rangle$. The $\vec{\boldsymbol{\mu}}_N^{R_K}$ and $\vec{\boldsymbol{\mu}}_N^{L_K}$ dipoles should be interpreted as induced transition dipoles:

$$\vec{\boldsymbol{\mu}}_N^{R_K} = \left\langle \Phi^0 \middle| (1 + \hat{\Lambda}) \left[\bar{\boldsymbol{\mu}}_N, \hat{R}_K \right] \middle| \Phi^0 \right\rangle = \vec{\boldsymbol{\mu}}_N^{TR_K} + \vec{\boldsymbol{\mu}}_N^{\Lambda R_K} \tag{4.27}$$

$$\vec{\boldsymbol{\mu}}_N^{L_K} = \left\langle \Phi^0 \middle| \hat{L}_K \bar{\boldsymbol{\mu}}_N \middle| \Phi^0 \right\rangle \tag{4.28}$$

Although the results in this work concentrate on transition energies, we also derived and implemented LR transition properties, following the equations for CCSD/PCM. 98,99,103,107,108,110–112

The PTES approximation can be also introduced for the LR formalism, where the right and left-hand diagonalization equations become:

$$\left\langle \Phi^n \middle| \left[(\bar{H}_N^{\mu} - \vec{\boldsymbol{\mu}}_N^T \cdot \bar{\boldsymbol{E}}_N), \hat{R}_K \right] \middle| \Phi^0 \right\rangle - \left\langle \Phi^n \middle| \vec{\boldsymbol{\mu}}_N^{TR_K} \cdot \bar{\boldsymbol{E}}_N \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^n \middle| \hat{R}_K \middle| \Phi^0 \right\rangle \tag{4.29}$$

$$\left\langle \Phi^0 \middle| \hat{L}_K \left[(\bar{H}_N^{\mu} - \vec{\boldsymbol{\mu}}_N^T \cdot \bar{\boldsymbol{E}}_N), \tau_n \right] \middle| \Phi^0 \right\rangle - \left\langle \Phi^0 \middle| \left[\vec{\boldsymbol{\mu}}_N^{L_K} \cdot \bar{\boldsymbol{E}}_N, \tau_n \right] \middle| \Phi^0 \right\rangle = \omega_K \left\langle \Phi^0 \middle| \hat{L}_K \middle| \Phi^n \right\rangle \tag{4.30}$$

The advantage of the PTES over the PTED scheme for the LR formalism is due to the fact that in the former the ground state $\hat{T}$ and $\hat{\Lambda}$ equations are decoupled. Thus, LR-CCSD/MMPol with the LR-PTES scheme is virtually equal in cost to the corresponding gas phase method.

In terms of overall computational effort, the evaluation of all induced dipole terms for ground and excited state equations (both SS and LR) scale as $O(N^5)$, where N is the basis set size, because that is the cost to evaluate the correlation ground, excited state, or transition densities. The cost of including the induced dipole terms in the CCSD equations is also no larger than $O(N^5)$, and these terms can be usually folded into contractions already necessary for the gas phase equations. Thus, these terms do not directly increase the cost of the CCSD equations, which scale as $O(N^6)$. The cost increase comes from the coupling of the amplitude equations, which is not a feature of the gas phase method. According to the discussion above, the four excited state schemes can be ranked in terms of computational cost as: SS-PTED > SS-PTES > LR-PTED > LR-PTES = gas phase.

4.3 Computational Details

All ground state geometries were optimized using the B3LYP hybrid functional[158] with the aug-cc-pVDZ basis set.[167] These structures were then used for the excitation energy calculations with all levels of theory in order to avoid geometrical effects on the results. The excitation energies

are computed with the time-dependent (TD) B3LYP,[151–153] the configuration interaction singles (CIS),[184] and the EOM/LR-CCSD methods[91,92] (in the following, we will use only the EOM-CCSD acronym for simplicity, as the use of the EOM or LR approaches should be clear from the context). The QM/MMPol calculations treat the solute at QM level and the solvent molecules with the Amber force field[185] plus the induced dipoles described in section 4.2. The theory and implementation for the evaluation of excitation energies with TD-B3LYP/MMPol and CIS/MMPol are described in Refs. 2,3,5. The results in the next sections are discussed in terms of the performance of the QM/MMPol approach compared to the corresponding full QM calculation rather than with experiment. This means that the comparison is not, for instance, between the EOM-CCSD and the CIS results, but rather full-EOM-CCSD vs. EOM-CCSD/MMPol and full-CIS vs. CIS/MMPol. For CCSD, we test the four computational schemes described in the previous section: SS-PTED, SS-PTES, LR-PTED, and LR-PTES. We utilize the same five basis sets from our previous benchmark paper for EOM-CCSD/PCM:[132] TZVP, aug-cc-pVDZ, 6-31++G**, 6-311++G**, 6-311(3+,3+)G**, plus cc-PVDZ. These basis sets include a variety of polarization and diffuse functions, and they can give us information on how well the classical model can reproduce solvent polarization. We consider two solvents, water and chloroform, where the induced dipoles are located on each atomic center (MMPol1). For water, we also consider a version where there is only one induced dipole for the whole molecule, located at the oxygen atom (MMPol2). The polarizability parameters for the induced dipoles are obtained from Refs. 178,180,182,183. All calculations were performed with a development version of the GAUSSIAN suite of programs.[170]

4.4 Results and Discussion

We use three test molecules for the benchmarking of the QM/MMPol methods: formaldehyde, azobenzene, and acrolein. These compounds are surrounded by 2-6 water or 2 chloroform molecules, so that we can investigate the MMPol performance with an increasing number of polarizable sites and with different solvents. Since the full-EOM-CCSD calculations are computationally expensive, we arranged the solvent in a symmetrical fashion around the solute. The structures of the

clusters and their point group symmetry are reported in Figure 4.1.

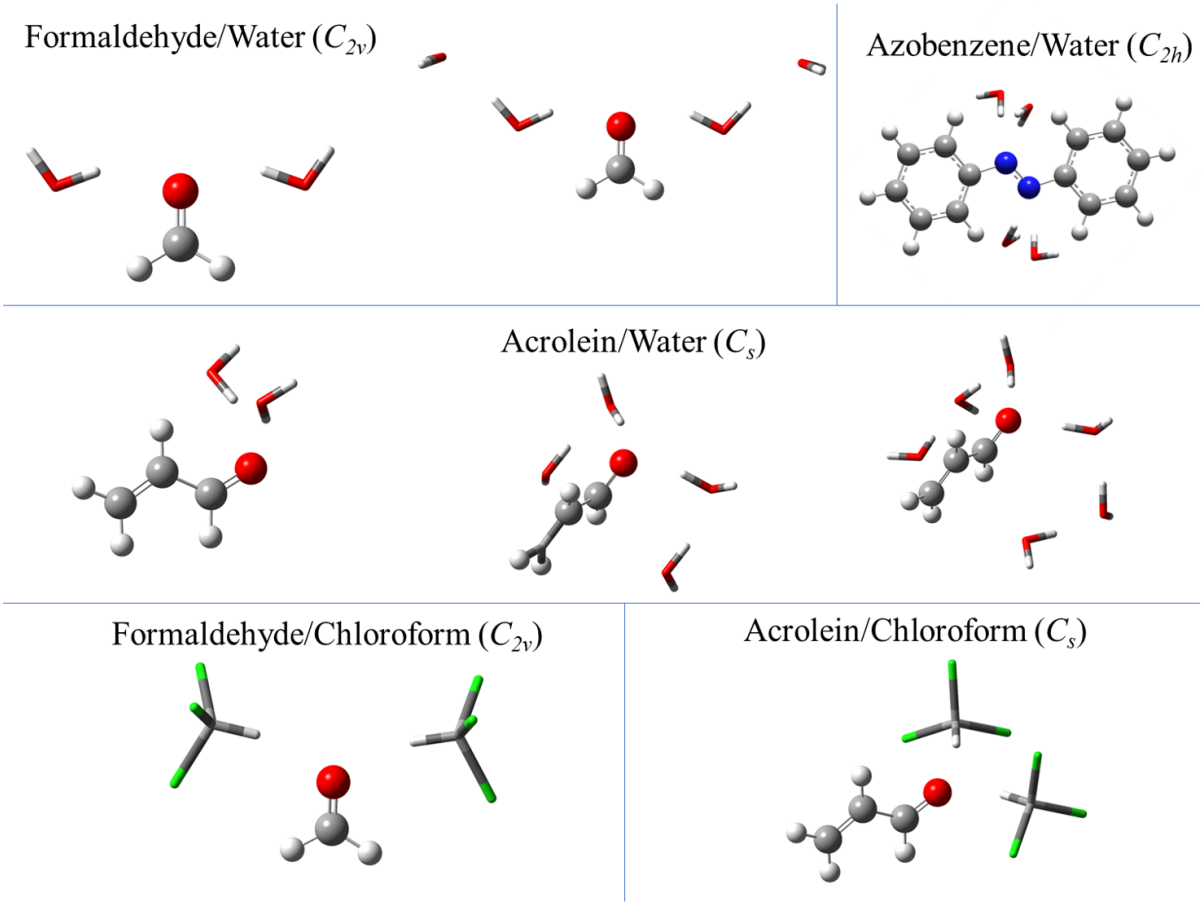

Figure 4.1: Test systems: formaldehyde, azobenzene, and acrolein microsolvated with water and chloroform molecules. The solute molecules are in "ball and stick" form, and the solvent molecules are in "tube" form. The point group symmetry of each system (solute + solvent molecules) is shown in parentheses.

In the following, the results are shown in a series of bar plots where we compare the relative solvatochromic shift computed with the full-QM methods and the corresponding QM/MMPol methods (panel a), and the relative full-QM shift with the relative errors for the QM/MMPol methods (panel b). Note that the relative solvatochromic shift is defined slightly differently for the two sets of bar plots so that the denominator is the same across plots in the same panel (the expressions are reported explicitly in the figures). The first set of plots carries visual information about the ability of the force field to reproduce the solvatochromic shift, while the second set of plots shows whether the relative error for the hybrid approach is indeed smaller than the solvatochromic shift

itself. These plots also include the data computed with only the non-polarizable part of the classical field (QM/Q), which provides an idea on the importance of the induced dipole contributions. The CCSD/MMPol results reported in this series of plots are computed with the LR-PTED method. Another series of bar plots reports the relative error for the four EOM-CCSD/MMPol schemes compared to the full-EOM-CCSD calculations. All transition energies are reported in Table S1-S9 of the SI.

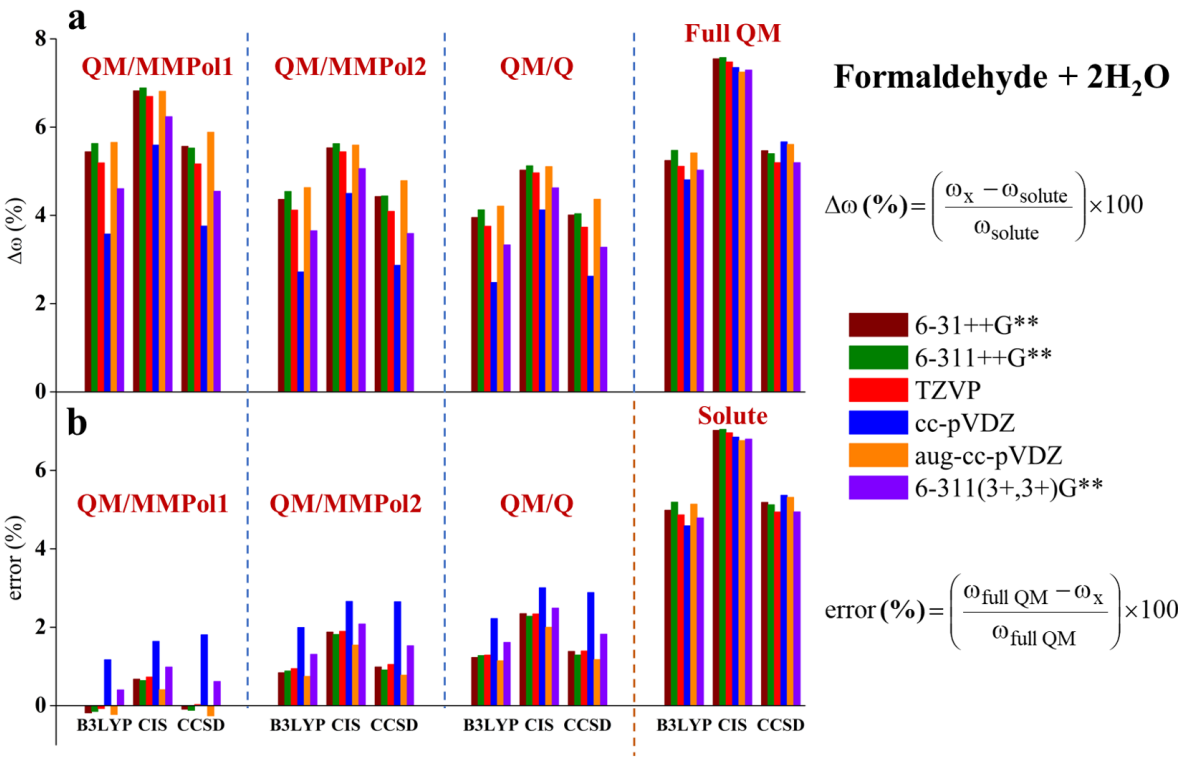

Figure 4.2: Relative solvatochromic shift (a) and relative error (b) for the excitation energy of formaldehyde + $2H_2O$ for all methods and basis sets. The subscript X in the formulas on the side refers to the red label in the plots.

The comparisons of the methods for formaldehyde with two water molecules are shown in Figures 4.2-4.3. We consider the lowest excitation, which has $n \rightarrow \pi^*$ character and is dipole-forbidden for perfectly symmetrical cases. The excitation energies and oscillator strengths for all method combinations are reported in Table S2 of the SI. For reference here, the EOM-CCSD/aug-cc-pVDZ transition energy for the isolated formaldehyde is 4.1 eV and $f = 0.00$. From a cursory

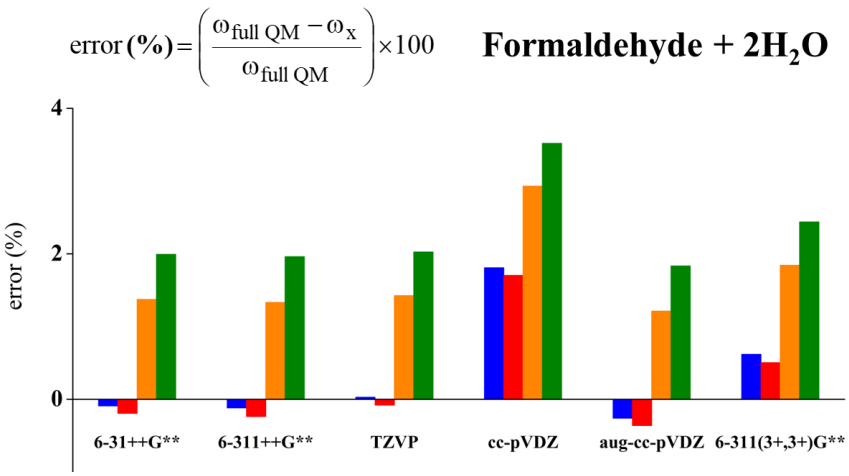

Figure 4.3: Relative errors for the excitation energy of formaldehyde + $2H_2O$ using the four EOM-CCSD/MMPol schemes for all basis sets.

inspection of the relative shifts and relative errors in Figure 4.2, it is evident that the MMPol1 method is able to reproduce the full-QM results rather well for all QM methods. Using only the non-polarizable part of the force field already provides good results, but the inclusion of the induced dipoles on each atomic site reduces the error to < 1% for all methods and basis sets except for the smallest one (cc-pVDZ). The MMPol2 option improves the performance compared to QM/Q, but the improvement is somewhat small considering that the difference in computational cost between MMPol1 and MMPol2 is negligible. The largest errors are with CIS, not surprisingly since this method is the least flexible in terms of wave function polarization. The MMPol approach works equally well when combined with DFT and CC methods. The best overall results are obtained with the basis sets that include a reasonable amount of diffuse functions. The lack of diffuse functions (cc-pVDZ), or too many diffuse functions (6-311(3+,3+)G**) lead to poor performance. This trend is consistent with our previous findings for the EOM-CCSD/PCM method,[132] and it can be explained by realizing that diffuse functions are necessary for an accurate representation of excited states, but too many diffuse functions create issues due to the close contact of the QM density and the classical region. For this case, the TZVP basis set performs rather well, probably because the triple-ζ functions provide enough flexibility to the electron density. The four EOM-CCSD/MMPol1 solvation schemes are compared in Figure 4.3. The relative errors are be-

low 2% with all schemes and basis sets except the worst two choices described above (cc-pVDZ and 6-311(3+,3+)G**), and the LR scheme provides the best performance with errors < 0.5%. The difference between PTED and PTES is very small for the LR formalism, and smaller than the relative error for the SS formalism (the PTES/PTED difference is expected to be larger for the latter formalism because it takes into account the full polarization of the excited state).

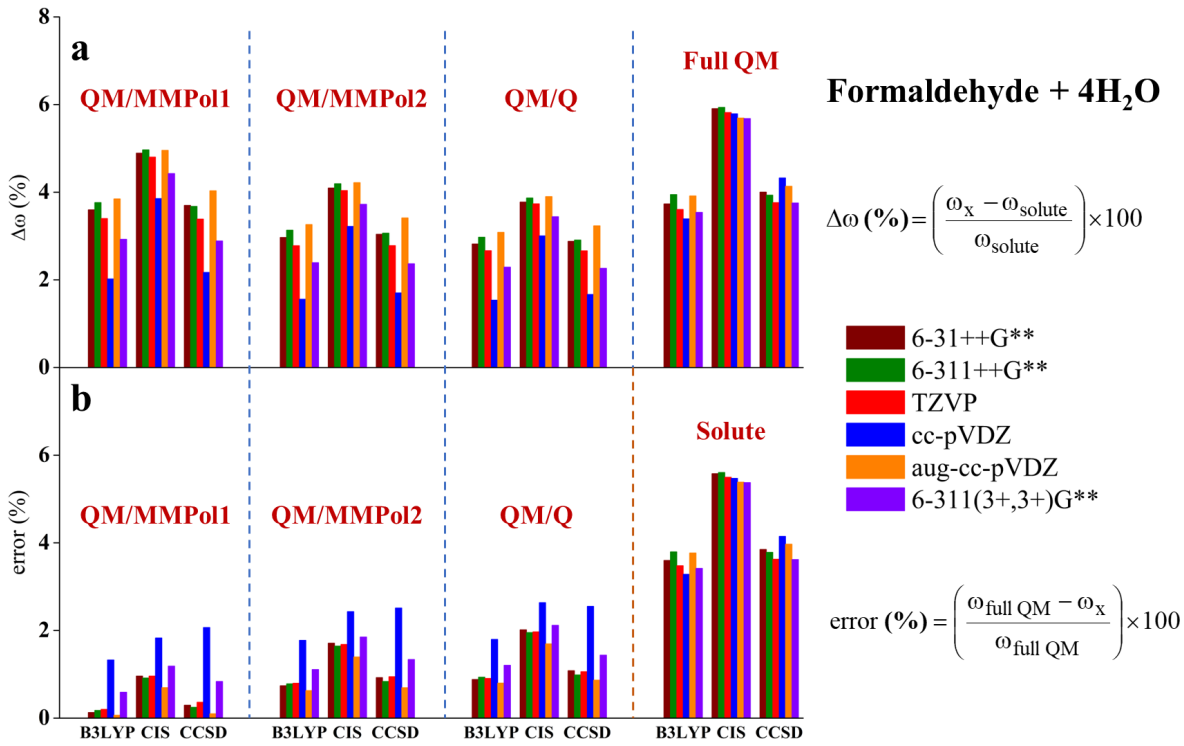

Figure 4.4: Relative solvatochromic shift (a) and relative error (b) for the excitation energy of formaldehyde + $4H_2O$ for all methods and basis sets. The subscript X in the formulas on the side refers to the red label in the plots.

The results for formaldehyde with four water molecules are shown in Figures 4.4-4.5. These results refer to the same dark transition as in the previous example. The solvatochromic shift (4%-6%) is smaller than with two solvent molecules because the second set of solvent units is positioned away from the solute, thus reducing the interaction with the closer solvent units. The QM/MMPol calculations reproduce this trend with all QM levels of theory and with all classical embedding models. MMPol1 has again the smallest errors, see Figure 4.4. The relative error is basically the same as that with two solvent molecules. All basis sets provide good results except for cc-pVDZ

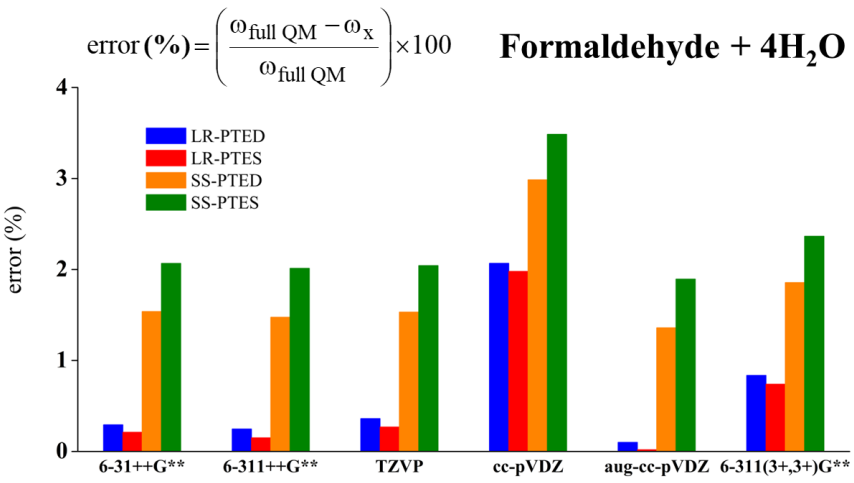

Figure 4.5: Relative errors for the excitation energy of formaldehyde + $4H_2O$ using the four EOM-CCSD/MMPol schemes for all basis sets.

and 6-311(3+,3+)G** as in the previous case, and the best performance is obtained with aug-cc-pVDZ. The difference between the various EOM-CCSD/MMPol computational schemes, shown in Figure 4.5, is rather similar to the case with two water molecules: the LR formalism provides errors < 1% with all basis sets except cc-pVDZ, and the difference between PTED and PTES is negligible; the errors with the SS formalism are larger but still < 2% with the PTED scheme, and around 2% with the PTES scheme.

The microsolvated azobenzene cluster has 4 water molecules H-bonded to the N centers, see Figure 4.1. We selected the lowest $\pi \rightarrow \pi^*$ (bright) excitation for this compound; the EOM-CCSD/aug-cc-pVDZ calculation for the isolated molecule predicts $\omega = 4.4$ eV and $f = 0.76$, while the rest are reported in Table S4 of the SI. The data on solvatochromic shifts and relative errors are reported in Figures 4.6-4.7. The red-shift is well reproduced by all methods, with negligible errors with MMPol1 for B3LYP and EOM-CCSD and with MMPol2 for CIS (the latter trend is probably due to cancellation of errors, given the lack of flexibility of both the QM method and the MMPol2 scheme). The relative errors of MMPol1 compared to those from the QM/Q calculations emphasize the importance of polarization in the force field. Contrary to formaldehyde, for azobenzene the SS-PTED scheme has the best performance, as shown in Figure 4.7. However, the errors are so small with most basis sets (< 0.5%) that the difference between the SS and LR formalisms is

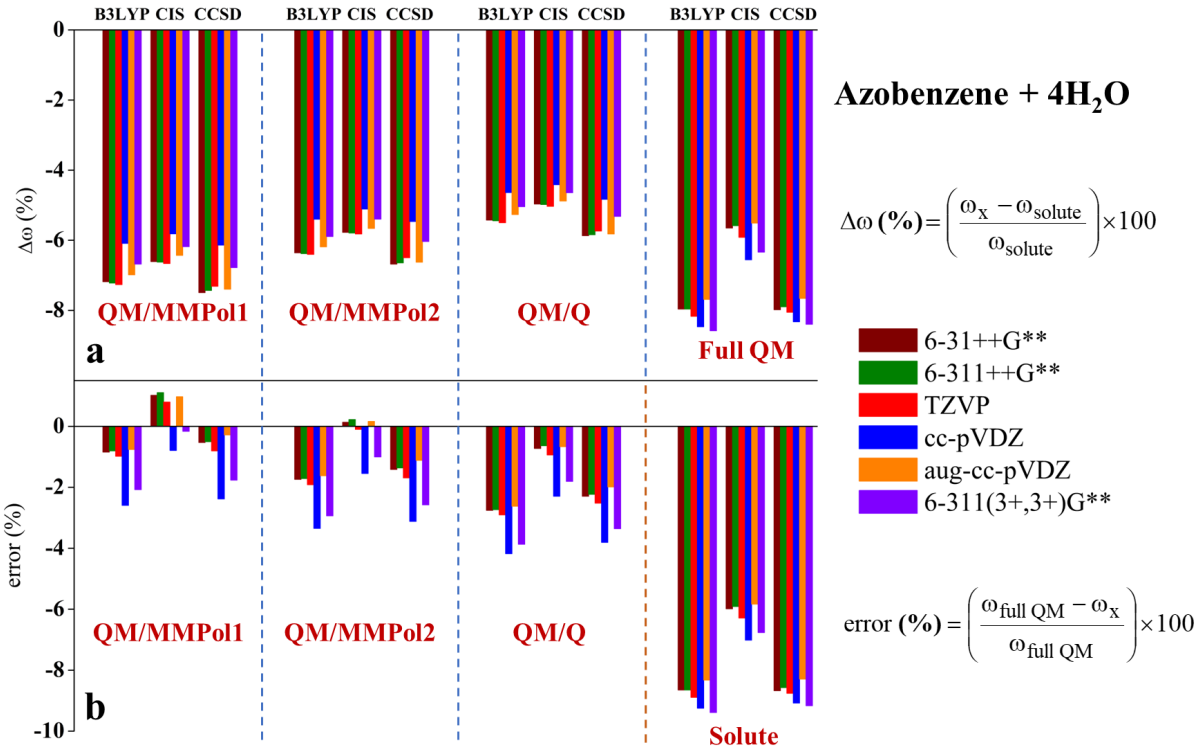

Figure 4.6: Relative solvatochromic shift (a) and relative error (b) for the excitation energy of azobenzene + $4H_2O$ for all methods and basis sets. The subscript X in the formulas on the side refers to the red label in the plots.

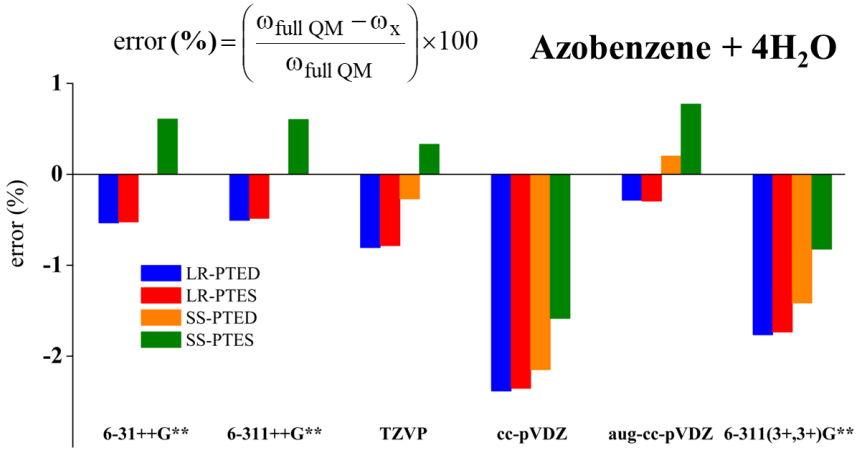

Figure 4.7: Relative errors for the excitation energy of azobenzene + $4H_2O$ using the four EOM-CCSD/MMPol schemes for all basis sets.

effectively negligible.

For acrolein, we consider another bright $\pi \rightarrow \pi^*$ transition ($\omega = 6.8$ eV with $f = 0.29$ at EOM-CCSD/aug-cc-pVDZ level for the isolated molecule). This compound is surrounded by two, four,

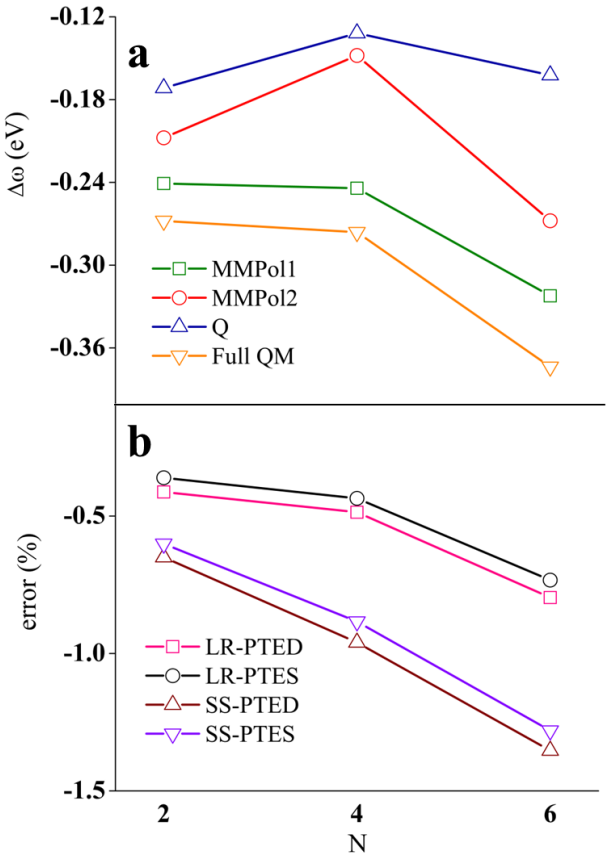

Figure 4.8: Panel a: solvatochromic shifts in the excitation energies of acrolein with two to six water molecules (N) computed at EOM-CCSD/aug-cc-pVDZ level. Panel b: corresponding relative errors for the four EOM-CCSD/MMPol1 schemes. For the gas phase solute $\omega = 6.8$ eV and $f = 0.29$.

and six water molecules as shown in Figure 4.1. The usual bar graphs for this test case are reported in Figures S1-S6 in the SI. The trends are very similar to those for the previous compounds: MMPol1 provides the best agreement with the full-QM calculations, and aug-cc-pVDZ performs best among the various basis sets. Thus, in Figure 4.8 we report the actual solvatochromic shifts with EOM-CCSD/aug-cc-pVDZ, and the relative errors for the four CCSD/MMPol schemes as a function of the number of water molecules. All choices of force field reproduce the correct sign of the shift, but only MMPol1 reproduces the full-QM results both qualitatively and quantitatively. MMPol2 produces the wrong trend for $N = 4$, while the non-polarizable option fails to reproduce the trend for $N = 4,\ 6$. All four coupling schemes show small errors (Figure 4.8b), with the best performance obtained with the LR formalism and small differences between the PTES and PTED schemes.

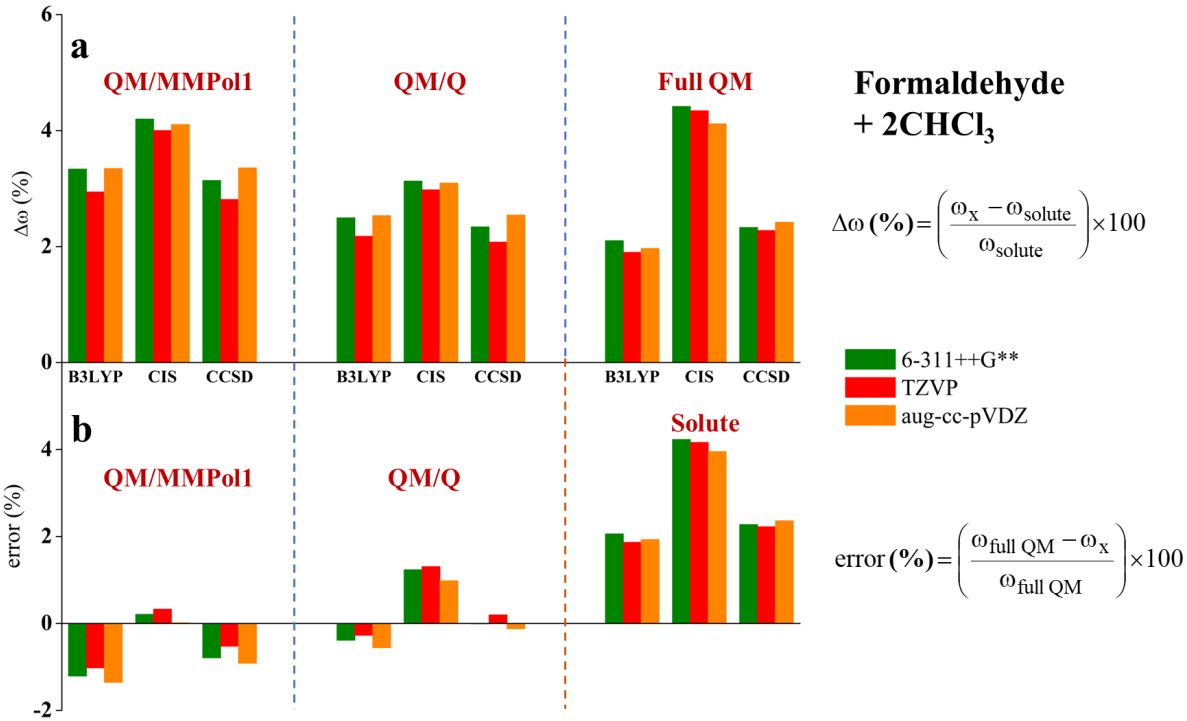

Figure 4.9: Relative solvatochromic shift (a) and relative error (b) for the excitation energy of formaldehyde + $2CHCl_3$ for all methods and basis sets. The subscript X in the formulas on the side refers to the red label in the plots.

To test the performance of the polarizable force field with a different solvent, we performed

calculations for formaldehyde and acrolein surrounded by chloroform, as shown in Figure 4.1. Given the size of $CHCl_3$, we limited the number of solvent molecules to two, and we considered only the 6-311++G**, TZVP, and aug-cc-pVDZ basis sets. The formaldehyde plots are in Figures 4.9-4.10. The solvatochromic shift for the full QM calculation is smaller than with water, since chloroform is a less polar solvent. The QM/Q calculations perform slightly better than QM/MMPol1, as shown in Figure 4.9b. However, the difference between the MMPol1 and Q results is significant, as shown in panel a of the same figure, indicating that polarization effects on the solvent are non-negligible. Thus, the better performance of the non-polarizable force field is likely fortuitous. Furthermore, the QM/MMPol1 errors are $< 1\%$, thus overall negligible. The comparison between the CC schemes shows a slightly better performance of the SS formalism, but the largest errors with LR are still $< 1\%$.

The results for acrolein with chloroform are shown in Figures 4.11-4.12. Also in this case, the overall solvatochromic shift is smaller than with water, compare with Figures S1-S6 in the SI, but the agreement with the full QM calculations is considerably better with the polarizable force field, see Figure 4.11. The best performance is obtained with aug-cc-pVDZ, and the plot in Figure 4.12 indicates that the LR formalism provides better results than SS, although the difference between all schemes is again small.

4.5 Conclusions

In this work, we present an implementation of CCSD with a polarizable force field based on the induced dipole model for ground and electronic excited states. We test different formalisms and approximations for the correlation part of the embedding, based on schemes that some of us developed for CCSD with implicit solvation models. We utilize microsolvated molecules to compare the QM/MMPol simulations with the corresponding full QM calculations (including lower level methods such as TD-B3LYP and CIS) employing six different basis sets, two MMPol models, and two solvents (water and chloroform). The results show that this flavor of hybrid QM/MMPol is able to reproduce the relative shift in transition energies due to the environment with good accuracy

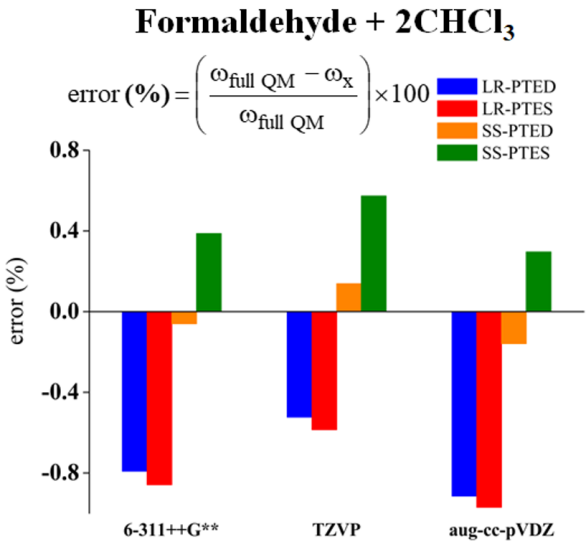

Figure 4.10: Relative errors for the excitation energy of formaldehyde + $2CHCl_3$ using the four EOM-CCSD/MMPol schemes for all basis sets.

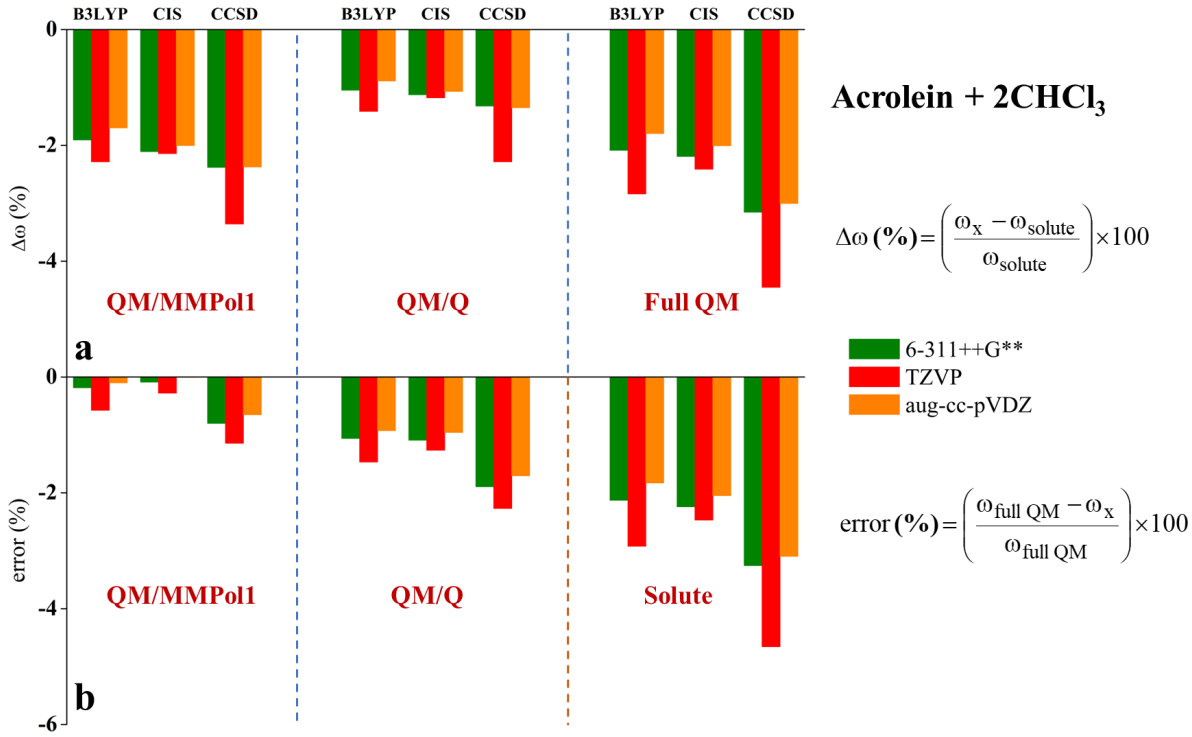

Figure 4.11: Relative solvatochromic shift (a) and relative error (b) for the excitation energy of acrolein + $2CHCl_3$ for all methods and basis sets. The subscript X in the formulas on the side refers to the red label in the plots.

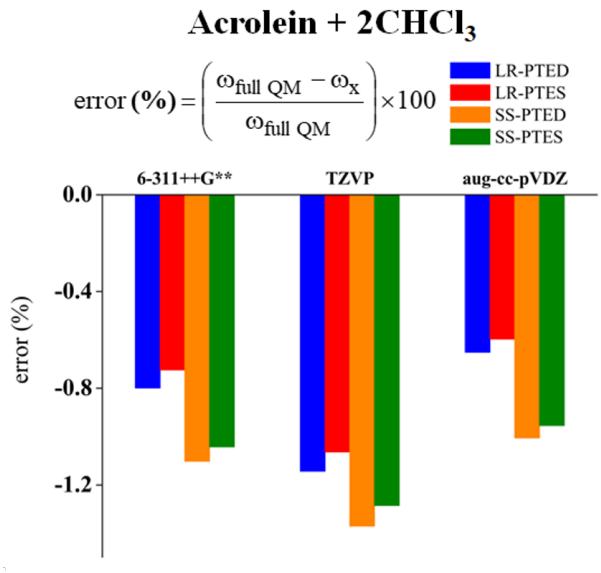

Figure 4.12: Relative errors for the excitation energy of acrolein + $2CHCl_3$ using the four EOM-CCSD/MMPol schemes for all basis sets.

for all levels of theory and most basis sets. In fact, a comparison with non-polarizable embedding shows the importance of including mutual polarization between the QM region and the environment. The simulations show similar performance with multiple basis sets, except for those with no diffuse functions (cc-pVDZ) that are important for describing excited states, and too many diffuse functions (6-311(3+,3+)G**) where the QM density is too close to the classical region. With the best basis set, aug-cc-pVDZ, the errors are consistently below 2% with any combination of level of theory and embedding scheme. The QM/MMPol approach works equally well with both solvents tested, water and chloroform, as long as induced dipoles are located on each atomic center. The four CCSD/MMPol embedding schemes (LR-PTES, LR-PTED, SS-PTES, and LR-PTED) perform similarly well for all cases, with the LR or SS formalisms being the best in different situations. Given the different computational cost and comparable performance, the best compromise is obtained with the LR-PTES scheme, which is equivalent in cost to a gas phase calculation.

The tests presented in this work show that CCSD/MMPol with the LR-PTES scheme is a promising tool to study chromophores in complex, non-homogeneous environments. Future studies will apply the method to more realistic systems for a comparison with experimental results.

4.6 Supporting Information

The Supporting Information includes the equations of the partial derivatives of the Lagrangians for (EOM-)CCSD/MMPol, the excitation energies and oscillator strengths for all systems studied, the bar plots for acrolein with 2-6 water molecules, and the optimized geometries for all test systems.

Chapter 5

Induced-Dipole Polarizable Embedding for the ONIOM(QM:QM) Hybrid Method

Abstract

In this chapter, we develop a new polarizable embedding for ONIOM (our own n-layered integrated molecular orbital and molecular mechanics), and apply it to the calculation of vertical excitation energies. In ONIOM without electronic embedding (EE), commonly called mechanical embedding (ME), the model subcalcualtions do not experience any polarization effects from the rest of the system. Since this effect may be important, EE is introduced in two ways: 1) fixed point-charges embedding (QE), 2) polarizable embedding (PQE) based on fixed charges and induced dipoles. The charges are computed with a fitting method previously developed in our group specifically to address overpolarization issues at the boundary. The PQE is based on the same idea of classical polarizable force field (MMPol), but the atomic polarizabilities (the central parameter) are computed on-the-fly. The ONIOM excitation energies with and without embedding are compared against the target values, (i.e., the excitation energies obtained from the calculation with the higher level of quantum mechanical method on the entire system). The method is tested on several related molecules: betaine-30 (B30), which has been used to define a polarity scale for solvents due to the sensitivity of its absorption spectrum to the environment, and the functionalized analogs with polar substituent groups. Preliminary results show that the PQE performs better than QE and ME, especially for the cases with polar groups, and it represent a promising approach for performing reliable and efficient calculations of excitation energies with the ONIOM (QM:QM) hybrid method.

5.1 Introduction

Accurate modeling of excited state properties of large systems is important in different areas of research. For example, in energy science, these simulations are used to understand reactions occurring in photovoltaic devices based on organic chromophores.[82,83] Other examples include excitations occurring in biological systems, light-sensitive drug molecules, and photocatalysts. Hybrid methods allow efficient modeling of large systems by combining multiple computational methods of different cost and accuracy to study systems that are otherwise too large for standard quantum mechanical (QM) simulations. When applying a hybrid method, the system is partitioned in a core region where the process occurs, (e.g. a reaction center), which is treated with a more accurate QM method, and the rest of the system (environment), which is treated with a less accurate but more computationally efficient method. This allows us to effectively utilize limited computing resource to model process with high accuracy, while still take into account the effects from the rest of the system.

One of the most popular treatment for the environment is with classical physics methods.[1,11–14] A classical method allows one to model the effects of a large section of the environment with a relatively modest additional computational cost. However, in cases where a classical treatment of the environment is not sufficient for accurately describing complex interactions with the core region, the outer layers must be treated at QM level. A typical example is a large chromophore with active substituent groups.

Describing the interactions between QM regions is important but challenging. There are various approaches being developed in the field, and they can be categorized into additive and subtractive. In an additive approach, the interaction energy (E_{int}) between the high and low QM levels is computed directly, whereas in a subtractive approach, E_{int} is obtained indirectly with an extrapolation formula. Typical choices of low level QM methods are based on density functional theory (DFT) or semiempirical wave function (WF) theory. One of the most successful examples of additive QM/QM hybrid method is DFT-in-DFT,[31–44] where a subsystem is selected and embedded in the density of the rest of the system. The most popular subtractive method is ONIOM (our

own n-layered integrated molecular orbital and molecular mechanics),[69–74] which can be used as a QM/molecular mechanics (MM) or QM/QM method. One of the distinct advantages of ONIOM is its intrinsic error canceling mechanism, which reduces the burden for a perfect choice of low level method.

In ONIOM, the entire structure is usually called the "real" system, and the core region is called the "model" system. A 2-layered ONIOM requires three subcalcualtions: 1) a low-level calculation on the real system (rl) 2) a high-level calculation on the model (mh), and 3) a low-level calculation on the model (ml). These are combined to obtain the total energy with an extrapolation formula. The same extrapolation method also applies to excitation energies. If no electronic embedding (EE) is used, the ONIOM subcalculations are independent, and the model subcalculations do not feel the existence of the rest of the system. To recover this effect, the model calculations can be performed with the EE obtained from the real system subcalculation. The Raghavachari group has made great contributions to ONIOM-EE methods for ground state problems.[74,78–81] One of the basic form of EE is based on fixed point charges (QE) placed at the positions of the atoms in the real system surrounding the model region. One possible issue with QE is the unphysical overpolarization of the model system electronic density at the boundary region, caused by the capping atoms being too close to the embedding charges. To address this issue, our group developed a method for creating embedding charges specifically fitted to minimize such overpolarization.[186]

However, in QE, the embedding charges are not polarizable, thus, the environment is unable to reach mutual polarization with the core region. A typical approach in classical mechanics to create polarizable force field is to use induced dipoles, which adjust to the presence of the QM electron density. These classical polarizable force fields (MMPol) have been combined with various levels of theory.[5,19,22,27,104,105] In this work, we adapt this approach to ONIOM(QM:QM)-EE methods by computing the required parameters (i.e., atomic polarizabilities) on-the-fly. We test this new approach for excitation energies against standard ONIOM with no embedding (called mechanical embedding, ME), and with ONIOM-QE. Preliminary results show that the PQE approach performs better than QE and ME, especially for the cases with polar substituent groups around the model

region.

The rest of the chapter is organized as follows. The theory is presented in section 5.2, the computational details of the test calculations are described in section 5.3, the results are discussed in section 5.4, and conclusions are reported in section 5.5.

5.2 Theory

In this section, we discuss a new approach to include polarizable embedding in ONIOM(QM:QM). For completeness, we first review the ONIOM method. In a 2-layered case, the entire system (real) is partitioned as shown in Figure 5.1.[69–74] The core region (called "model" system) is capped

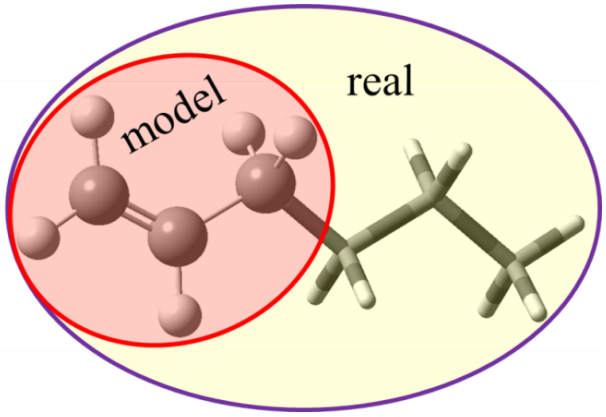

Figure 5.1: Illustration of the ONIOM partitioning of a molecule.

with link atoms, and is treated at a high level of theory. The ONIOM energy is defined with an extrapolation formula that combines the energies from separate subcalculations:

$$E_{ONIOM} = E^{low}_{real} + E^{high}_{model} - E^{low}_{model} \tag{5.1}$$

where *high* and *low* represent the levels of theory, and *real* and *model* refer to the partition shown in Figure 5.1. For excited states, the ONIOM transition energy is obtained with a similar extrapolation:

$$\omega_{ONIOM} = \omega^{low}_{real} + \omega^{high}_{model} - \omega^{low}_{model} \tag{5.2}$$

In standard ONIOM-ME calculations, the model subcalculations do not include polarization effect from the rest of the system, thus, the individual subcalculations are independent from each other. In contrast, in ONIOM-EE, the model system feels the effect of the environment through the embedding. In this work, we consider a fixed embedding through a set of point charges (QE), and a polarizable embedding that also includes induced dipoles (PQE).

In the QE model, fixed point charges are placed at the positions of the atoms in the environment. However, these charges may create unphysical overpolarization of the electronic density at the boundary region. To address this issue, our group previously developed a method for creating an improved charge distribution to minimize this overpolarization, by fitting the charges on the electrostatic potential (ESP) of the entire structure in the presence of the link atoms. A more detailed discussion of this approach can be found in Ref. 186.

The QE is a fixed embedding computed with the ground state density, and it cannot respond to changes induced by an electronic excitation. In order to add this flexibility to the embedding field, we borrow an idea used in classical polarizable force field (MMPol).[5,19,22,27,104,105] We endow the embedding sites (where the fixed charges are located) with a set of induced dipoles. Here, we review the key equations for this approach for completeness, and a detailed derivation of the QM/MMPol method can be found in Refs. 5,19,22,27,104,105. The embedding dipoles are computed through a minimization of the polarization energy with respect to the induced dipoles:

$$\begin{aligned} \frac{\partial E^{Pol}}{\partial \vec{\boldsymbol{\mu}}} &= \boldsymbol{T}\vec{\boldsymbol{\mu}} - \vec{\boldsymbol{E}}^{MM} - \vec{\boldsymbol{E}}^{QM}(P) = 0 \\ &\Longrightarrow \boldsymbol{T}\vec{\boldsymbol{\mu}} = \vec{\boldsymbol{E}}^{MM} + \vec{\boldsymbol{E}}^{QM}(P) \end{aligned} \tag{5.3}$$

where $\boldsymbol{T}$ is the interaction matrix, which takes the form:

$$\boldsymbol{T} = \begin{bmatrix} \boldsymbol{\alpha}_1^{-1} & \mathcal{T}_{12} & \cdots & \mathcal{T}_{1N_{pol}} \\ \mathcal{T}_{21} & \boldsymbol{\alpha}_2^{-1} & \cdots & \mathcal{T}_{2N_{pol}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{T}_{N_{pol}1} & \mathcal{T}_{N_{pol}2} & \cdots & \boldsymbol{\alpha}_{N_{pol}}^{-1} \end{bmatrix} \tag{5.4}$$

$\mathcal{T}_{ij}$ is the classical damped interaction tensor between sites i and j, which are the positions of the atoms in the environment, N_{pol} is the number of polarizable sites, and $\boldsymbol{\alpha}_i$ is the polarizability tensor of site i, which is replaced by the isotropic value for simplicity:

$$\alpha_i = \frac{1}{3}[\alpha_{xx}(i) + \alpha_{yy}(i) + \alpha_{zz}(i)] \tag{5.5}$$

For an environment that can be subdivided in individual fragments, such as a solvent or a protein, we could use the same set of pre-calculated atomic polarizabilities as in a classical force field. However, when we have a generic chromophore with novel substituent groups, no parameters are generally available. For these cases, we need to calculate the atomic polarizability parameters on-the-fly.

We propose a strategy to obtain atomic polarizability parameters based on a decomposition of the polarizability of the entire molecule (i.e., the real system), computed at the low level of theory. Such an approach is completely general, and it allows us to obtain these parameters with minimum computational effort. The molecular polarizability is computed with standard linear response techniques,[187] as a contraction between the dipole matrix ($\mu^i_{\lambda\sigma}$) and the density matrix ($P^i_{\lambda\sigma}$). For instance, for xx element we have:

$$\alpha_{xx} = \sum_{\lambda\sigma} \langle \lambda | \hat{\mu}_x | \sigma \rangle P^x_{\lambda\sigma} = \sum_{\lambda\sigma} \mu^x_{\lambda\sigma} P^x_{\lambda\sigma} \tag{5.6}$$

where λ and σ represent atomic orbitals (AOs), and the summation runs over all AOs in the molecule. Therefore, $\mu^i_{\lambda\sigma}$ and $P^i_{\lambda\sigma}$ are square matrices that can be split according to the atoms where the AO functions are located. Using a water molecule as an example, the $\mu^x_{\lambda\sigma}$ and $P^x_{\lambda\sigma}$ matrices in the x direction take the form shown in Figure 5.2, where the elements in the diagonal blocks represent the contributions from the AOs of each atom, and in the off-diagonal blocks λ

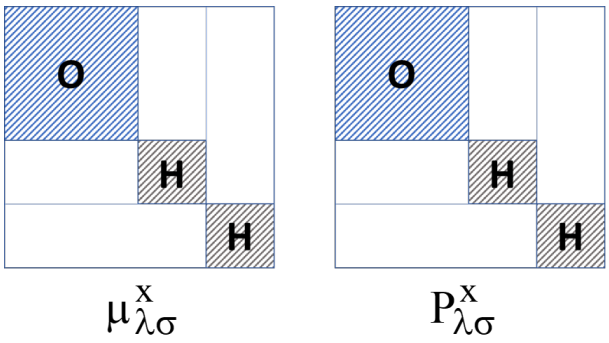

Figure 5.2: A pictorial representation of the $\mu^x_{\lambda\sigma}$ and $P^x_{\lambda\sigma}$ matrices for a water molecule.

and σ belong to different atoms. The atomic polarizability for an atom A is computed as:

$$\alpha_{xx}(A) = \sum_{\substack{\lambda\sigma \\ \in A}} \langle\lambda|\hat{\mu}_x|\sigma\rangle P^x_{\lambda\sigma} + \frac{1}{2}\sum_{\substack{\lambda\in A \\ \sigma\in B}} \langle\lambda|\hat{\mu}_x|\sigma\rangle P^x_{\lambda\sigma} \tag{5.7}$$

The first term is the contraction for the diagonal block corresponding to atom A, see Figure 5.2. The second term corresponds to the contraction of the off-diagonal blocks, where we assign half of the contribution to atom A and the other half to atom B, to avoid double-counting. Using this method, the sum of all atomic polarizabilities equals the molecular polarizability $\left(\alpha_{xx} = \sum_A \alpha_{xx}(A)\right)$. For the y and z directions, $\alpha_{yy}(A)$ and $\alpha_{zz}(A)$ are obtained in the same way.

Once the atomic polarizabilities are computed, we can perform a ONIOM-PQE calculation by making use of the Gaussian[170] program by the Mennucci group. Although there are no MM force field parameters for a generic chromofore, the energy expressions for ONIOM in Eq. 5.1 and 5.2 lead to a complete cancellation of the fixed MM force field contributions in the two model subcalculations.

5.3 Computational Details

All geometries were optimized using the B3LYP hybrid functional[151–153,158] with the aug-cc-pVDZ basis set.[167] For the excited state calculations, we used the CAM-B3LYP functional with the 6-311++G** basis set as the high level. The entire system treated with this method provides the target values used as references. For the low level, we used the configuration interaction sin-

gles (CIS)[184] method or the B3LYP functional with the 3-21G basis set. Therefore, we considered two sets of method combinations for ONIOM: CAM-B3LYP/6-311++G**:CIS/3-21G and CAM-B3LYP/6-311++G**:B3LYP/3-21G. The method in Ref. 186 is used to evaluate the QE scheme. For the PQE embedding, we perform QM/MMPol calculations using the Amber[185] force field with QE charges. As explained in the previous section, the choice of force field is irrelevant, given that it is contributions to the model subcalculations cancel out. The only requirement is that the chosen force field is defined for all atoms. The atomic polarizability parameters for the PQE embedding were obtained as discussed in section 5.2, by performing a molecular polarizability calculation on the entire molecule at the low level, and separating the diagonal elements in atomic contributions as in Eq. 5.7. All calculations were performed with a development version of the GAUSSIAN suite of programs.[170]

5.4 Results and Discussion

We consider four test molecules, shown in Figure 5.3: betaine-30 (B30) and three functionalized analogs with polar substituent groups. The model is chosen as the two central rings (in "ball and stick" form in the figure). We compare ONIOM-ME and ONIOM-EE, the latter considersing various EE models: 1) fixed point charge embedding (QE) 2) polarizable embedding (PQE), which includes induced dipoles, and for testing purposes, we also consider 3) polarizable embedding (PE), where fixed point charges are set to zero. For all test cases, the excitation energies are compared with the target calculations (i.e., the entire molecule calculated with the high level: CAM-B3LYP/6-311++G**).

The results for **1** (B30) are shown in Table 5.1. For B3LYP as low level, QE performs the best at reproducing the target excitation energy with a positive error of 0.04 eV, whereas ME performs the worst with a negative error of -0.29 eV. PQE gives an error of 0.18 eV, which is 0.14 eV larger than that of QE but 0.11 eV smaller in magnitude than that of ME. To test explicitly the effect of the induced dipoles, we set the fixed point charges to zero (PE). It is shown that PE has a same error sign as ME, but it is 0.07 eV smaller. This indicates that the effect of the induced dipoles is much

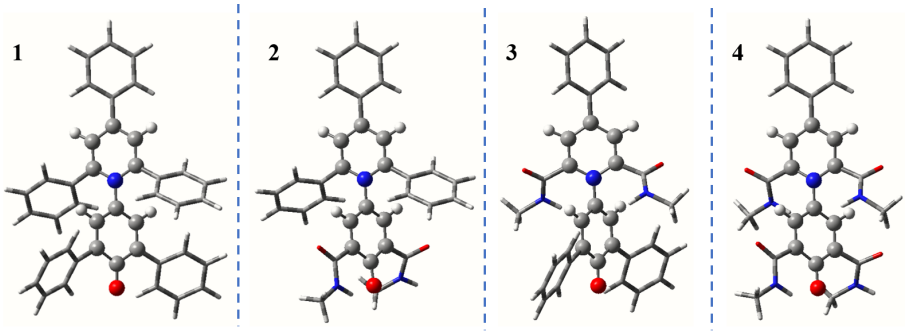

Figure 5.3: stuctures of all test cases

smaller than that of the point charges. For CIS as low level, PE and ME give the smallest error, opposite in sign compared to both QE and PQE, and PE works best. For B3LYP, the ML excitation energies are affected only slightly by the electronic embedding (changes < 0.05 eV), while for CIS there is a significant shift (1 eV) in excitation energies going from ME to PQE. This is consistent also in the following cases. The MH excitation energies increase with both the B3LYP and CIS embeddings with a 0.5 eV difference in excitation energies between ME and PQE.

	PQE	QE	PE	ME
B3LYP				
RL	1.48			
MH	2.35	2.22	1.92	1.88
ML	1.93	1.93	1.89	1.93
ext	1.90	1.77	1.51	1.43
error	0.18	0.04	-0.22	-0.29
CIS				
RL	2.69			
MH	2.46	2.35	1.91	1.88
ML	3.75	3.60	2.82	2.72
ext	1.40	1.44	1.78	1.85
error	-0.33	-0.29	0.05	0.12

Table 5.1: Excitation energies (eV) of B30 (**1**) from the ONIOM subcalculations: real low (RL), model high (MH), and model low (ML), with different embedding models: polarizable point charge embedding (PQE), point charge embedding (QE), polarizable embedding (PE), and mechanical embedding (ME). "ext" represents the ONIOM extrapolated excitation energies, and "error" is calculated with respect to the target calculation (1.73 eV). Top: B3LYP as low level, Bottom: CIS as low level.

To investigate the effect of polar substituent groups, we replace the rings around the model

with amide functional groups. For **2**, we replace the bottom two rings. The results for **2** are shown in Table 5.2. For both B3LYP and CIS as low level, ME performs the worst. Especially for B3LYP, the ME error is 0.77 eV, while for PQE it is only 0.11 eV. The errors for B3LYP are all negative, and decrease in magnitude with a trend: ME>PE>QE>PQE. The difference in error is 0.26 eV between QE and PQE; 0.52 eV between PE and PQE. These comparisons indicate again that point charges have a greater impact than the induced dipoles, and the combination of these embedding schemes reduces the error considerably in this case. For CIS, QE performs the best, and the error difference between PQE and QE is small (0.04 eV). Both QE and PQE reduce the error considerably compared to ME (0.3 eV).

	PQE	QE	PE	ME
B3LYP				
RL	1.37			
MH	2.31	2.03	1.72	1.63
ML	1.61	1.60	1.55	1.60
ext	2.07	1.80	1.54	1.40
error	-0.11	-0.37	-0.63	-0.77
CIS				
RL	3.56			
MH	2.42	2.21	1.68	1.63
ML	3.85	3.60	2.79	2.64
ext	2.12	2.18	2.44	2.55
error	-0.05	0.01	0.27	0.38

Table 5.2: Excitation energies (eV) of a functionalized B30 analog (**2**) from the ONIOM subcalculations: real low (RL), model high (MH), and model low (ML), with different embedding models: polarizable point charge embedding (PQE), point charge embedding (QE), polarizable embedding (PE), and mechanical embedding (ME). "ext" represents the ONIOM extrapolated excitation energies, and "error" is calculated with respect to the target calculation (1.73 eV). Top: B3LYP as low level, Bottom: CIS as low level.

For **3**, we replace the top two rings to investigate the effect of polar groups at different positions. The results for **3** are shown in Table 5.3. For B3LYP, the same trend of decreasing errors (ME>PE>QE>PQE) is seen, consistently with **2**, and PQE reduces the error of ME by 0.25 eV. In this case, QE and PE perform similarly, with a 0.03 eV difference in error. For CIS as low level, both QE and PE reduce the error of ME, whereas PQE increases the error. However, all errors are rather small in this case.

	PQE	QE	PE	ME
B3LYP				
RL	1.61			
MH	2.22	2.08	2.07	2.01
ML	2.06	2.08	2.04	2.10
ext	1.77	1.60	1.63	1.52
error	-0.09	-0.26	-0.23	-0.34
CIS				
RL	2.69			
MH	2.09	2.08	2.05	2.01
ML	3.07	3.00	2.94	2.74
ext	1.72	1.77	1.80	1.96
error	-0.15	-0.09	-0.06	0.10

Table 5.3: Excitation energies (eV) of a functionalized B30 analog (**3**) from the ONIOM subcalculations: real low (RL), model high (MH), and model low (ML), with different embedding models: polarizable point charge embedding (PQE), point charge embedding (QE), polarizable embedding (PE), and mechanical embedding (ME). "ext" represents the ONIOM extrapolated excitation energies, and "error" is calculated with respect to the target calculation (1.73 eV). Top: B3LYP as low level, Bottom: CIS as low level.

For **4**, we replace four rings around the model. The results of **4** is shown in Table 5.4. The same trend in error (ME>PE>QE>PQE) is seen again for B3LYP as low level, and PQE decrease the error of ME by 0.5 eV. The performance of QE and PE are very similar, with a 0.02 eV difference

in error similar to **3**. For CIS as low level, all EE models reduces the error of ME with QE being the best, and PQE close behind.

	PQE	QE	PE	ME
B3LYP				
RL	1.74			
MH	2.34	2.08	1.98	1.90
ML	1.89	1.93	1.85	1.95
ext	2.18	1.88	1.86	1.68
error	-0.19	-0.49	-0.51	-0.69
CIS				
RL	3.57			
MH	2.39	2.15	1.96	1.90
ML	3.67	3.33	2.97	2.71
ext	2.28	2.39	2.56	2.76
error	-0.09	0.02	0.19	0.38

Table 5.4: Excitation energies (eV) of a functionalized B30 analog (**4**) from the ONIOM subcalculations: real low (RL), model high (MH), and model low (ML), with different embedding models: polarizable point charge embedding (PQE), point charge embedding (QE), polarizable embedding (PE), and mechanical embedding (ME). "ext" represents the ONIOM extrapolated excitation energies, and "error" is calculated with respect to the target calculation (1.73 eV). Top: B3LYP as low level, Bottom: CIS as low level.

As we test the embedding on these functionalized B30 analogs by replacing the rings around the model system with polar substituent groups, we see the same trend in errors (ME>PE>QE>PQE) for B3LYP as low level. In Figure 5.4, we plot the errors against the number of polar groups surrounding the model, and the figure shows that as we replace the non-polar rings around the model with polar ones, ME fails at reproducing the results of the target calculations, and the trend in errors (ME>PE>QE>PQE) becomes clearer. For these functionalized cases, there are significant differences in errors (0.2-0.3 eV) between PQE and QE. This indicates that the polarization of

the environment becomes more important when the groups surrounding the model are polar, and including the induced dipoles in the embedding significantly reduces the errors. However, for CIS as low level, the trend that we see in B3LYP does not completely apply. PQ performs better than PQE for all test molecules. Both PE and QE give lower error than ME for the functionalized analogs with polar substituent groups, with QE performs much better than PE.

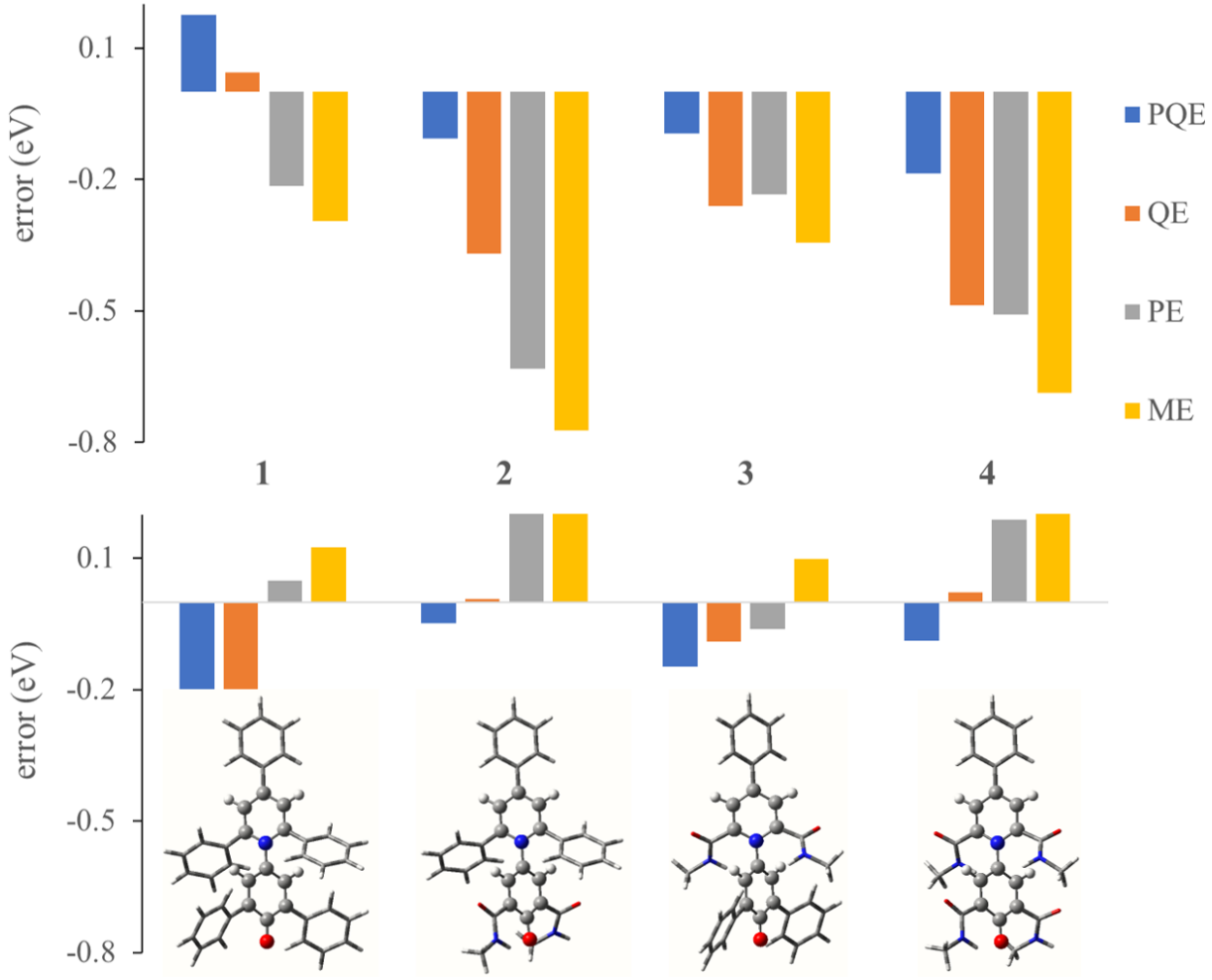

Figure 5.4: Errors (eV) plotted for all test molecules. Different embedding models are considered: polarizable point charge embedding (PQE), point charge embedding (QE), polarizable embedding (PE), mechanical embedding (ME). Top: B3LYP as low level, Bottom: CIS as low level.

5.5 Conclusions

In this work, we present a polarizable embedding scheme based on induced dipoles that can be used within ONIOM(QM:QM) hybrid method. This embedding is used together with the fixed point-charge embedding for ONIOM(QM:QM) previously developed in our group. We devise a strategy for obtaining atomic polarizabilities, which is the key parameter required for the induced-dipole embedding, by decomposing the polarizability of the entire molecule in atomic contributions. We perform tests using B30 and functionalized analogs to compare ONIOM excitation energies with various EE models (PQE, QE, and PE) against standard full QM calculations. These preliminary tests show that this on-the-fly polarizable embedding is better than no or fixed embedding when substituent groups around the model are polar.

Chapter 6

Multi-State Extrapolation of UV/Vis Absorption Spectra with QM/QM Hybrid Methods

(This work is taken from: Sijin Ren and Marco Caricato, *J. Chem. Phys.* **2016**, *18*, 184102.[188] Supporting information is available online.)

Abstract

In this work, we present an approach to simulate absorption spectra from hybrid QM/QM calculations. The goal is to obtain reliable spectra for compounds that are too large to be treated efficiently at a high level of theory. The present approach is based on the extrapolation of the entire absorption spectrum obtained by individual subcalculations. Our program locates the main spectral features in each subcalculation, e.g. band peaks and shoulders, and fits them to Gaussian functions. Each Gaussian is then extrapolated with a formula similar to that of ONIOM (Our own N-layered Integrated molecular Orbital molecular Mechanics). However, information about individual excitations is not necessary so that difficult state-matching across subcalculations is avoided. This multi-state extrapolation thus requires relatively low implementation effort while affording maximum flexibility in the choice of methods to be combined in the hybrid approach. The test calculations show the efficacy and robustness of this methodology in reproducing the spectrum computed for the entire molecule at a high level of theory.

6.1 Introduction

Theoretical simulations of UV/Vis absorption spectra of large chromophores have become an important tool for the design of new compounds in materials and renewable energy science. For instance, it is important to develop compounds that present a large absorption in the visible region where the sun emits most photons so that large amounts of solar energy can be harvested[82,189]. The photochemical characteristics of chromophores and dyes can be tuned by proper choices of substituent groups. Hence, accurate theoretical simulations of such effects largely benefit the design process by screening promising candidates[83]. The most successful tool to perform such calculations is time-dependent density functional theory (TD-DFT), since it provides the best compromise between accuracy and computational effort. However, an important limitation of DFT is that the approximate functionals available perform differently for different compounds, and a sensible choice for each case is often difficult[83,190]. Thus, much effort has been devoted to the benchmarking and calibration of approximate functionals[85,101,191–193]. On the other hand, more reliable methods may simply not be affordable in practical situations.

A possible strategy to overcome this difficulty is to use multi-scale or hybrid techniques. In general, the goal of multi-scale methods is to reduce the computational cost by partitioning the system into regions, where the core is treated at a high level of theory while the rest is treated at a lower and less computationally demanding level. This reduction in cost should be obtained without a significant loss in accuracy compared to the calculation on the entire system at the high level of theory. An example of such partitioning is shown in Figure 6.1, where "model" refers to the core region, and "real" refers to the entire system. Such approaches have been used extensively in biochemical applications where, for instance, the reactive pocket of a protein is described quantum mechanically (QM) while the rest is treated classically. These methods can also combine two QM methods, QM/QM, and they have been used in many applications[194–200]. The use of QM/QM methods has also been extended in recent times to excited state calculations[201–208].

One of the most successful hybrid methods is ONIOM (Our own N-layered Integrated molecular Orbital molecular Mechanics)[69–74], which is briefly described in the next Section. The main

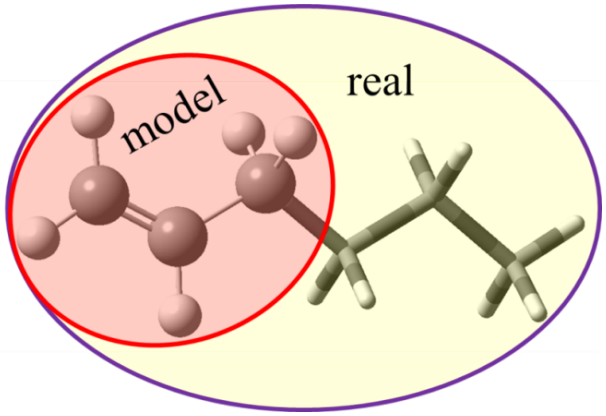

Figure 6.1: A sample system partitioning scheme for two layered ONIOM, in which the entire system (real system) is labeled as real, and the model system is labeled as model.

feature of ONIOM is that the combined energy is expressed as an extrapolation of the energy computed for the real and model systems, see Figure 6.1. Thus, if no electronic embedding schemes are used[78–81], these subcalculations are independent and the method requires no specific implementation. This also permits maximum flexibility in the choice of the high and low levels of theory. We showed that ONIOM can accurately reproduce excitation energies and transition properties computed at a high level of theory at a fraction of the computational cost[75–77]. However, a drawback of this approach is that its success depends on the correct matching of states between the subcalculations. This is relatively easy for low-lying, isolated, and bright transitions, but it becomes a significant issue if one is interested in various regions of the spectrum.

In this work, we propose a proof-of-concept approach to extrapolate multiple states at once or, in other words, to extrapolate the entire spectrum within an energy range. The extrapolation formula is borrowed from ONIOM, but no state-matching is required. Therefore, this approach maintains attractive ONIOM features (e.g., simplicity, favorable cost/accuracy ratio) while avoiding any connection to a particular description of individual excited states. Perfect state-matching would indeed provide a direct interpretation of the extrapolated spectrum in terms of characteristic transition types (e.g. $\pi \rightarrow \pi^*$). Unfortunately, a typical band is composed of a relatively large number of excitations with varying oscillator strength, especially in higher energy regions. This often makes a definite assignment difficult even with standard methods, and it renders the

use of QM/QM hybrid methods like ONIOM virtually impossible. The reason is that states that are close in energy often present various degrees of orbital mixing that are different across levels of theory and model vs. real systems. It is this mixing that prevents a clear state-matching, thus precluding the extrapolation in the higher energy regions. Our method bypasses this limitation by extrapolating spectral features (i.e., bands and shoulders) rather than individual states. However, once the multi-state extrapolation is performed, a qualitative analysis of the types of transitions that form a particular band can always be obtained by examination of the corresponding bands of the subcalculations that were used for the extrapolation. An important application of this multi-state extrapolation is the design of new dyes for solar energy harvesting. As mentioned above, a critical goal in this field is the tuning of the substituents on the main chromophoric moiety for maximum absorption of sunlight. Our method may allow to explore a large number of options by treating the core moiety at a high level of theory and the substituents at a lower level without sacrificing accuracy. These calculations would provide valuable screening information of promising candidates for sunlight harvesting dyes, which could then be synthesized and tested in actual devices. From a computational perspective, our method may permit, for instance, the use of global functionals as low level methods without worrying about the unphysical low-lying excitations with small oscillator strength (typical of these functionals), since these will not contribute to the overall spectrum. The computer program that we developed to perform the extrapolation can be downloaded from our group web site[209], and it can read the necessary information for the extrapolation (i.e., excitation energies and related oscillator strength) from a simple text file. As shown by our test calculations, this proof-of-concept approach is quite robust, and provides spectra that are in very good agreement with the corresponding high level calculations on the entire system.

The chapter is organized as follows. Section 6.2 reviews the ONIOM method and discusses the multi-state extrapolation. Details of the calculations are presented in Section 6.3, while test calculations are described in Section 6.4. Section 6.5 reports an overall discussion of the results and final remarks.

6.2 Theory

In this work, we propose to simulate UV-Vis spectra of large chromophores through a multi-state extrapolation of spectra obtained by combining two QM levels of theory, following a strategy similar to the ONIOM method[69–77]. Hence, let us briefly review ONIOM for two layers, and introduce the notation that we will use throughout the chapter. Following the layer separation in Figure 6.1, open valencies resulting from the severing of covalent bonds are capped with link atoms (usually hydrogens), e.g. X–Y $\longrightarrow$ X–H. The link atom bond length is based on the ratio between the current X–Y length and its typical value. The energy extrapolation is performed combining the energies from separate subcalculations:

$$E_{ext} = E^{low}_{real} + E^{high}_{model} - E^{low}_{model} \tag{6.1}$$

where *high* and *low* refer to the levels of theory, and *model* and *real* to the system partition as in Figure 6.1. If no electronic embedding is introduced, the three subcalculations are truly independent. The same extrapolation can also be used to compute excitation energies:

$$\omega_{ext} = \omega^{low}_{real} + \omega^{high}_{model} - \omega^{low}_{model} \tag{6.2}$$

Finally, transition properties are evaluated by extrapolation of transition moments with formulas equivalent to Eqs. 6.1 and 6.2[77]. In previous studies[75–77], we showed that excitation energies and properties can be accurately evaluated with ONIOM with great savings in computational cost, and that the use of link atoms does not affect the results when the model system is sensibly chosen. Despite these promising results and the simplicity of the extrapolation in Eq. 6.2, a straightforward application of ONIOM to compute excitation energies is difficult. The issue resides in the proper matching of states across subcalculations. Indeed, a clear classification of an excitation is often difficult, and clear-cut cases such as very bright $\pi \rightarrow \pi^*$ excitations with large oscillator strength (f) are not the norm. On the contrary, matching based on oscillator strength, energy ordering

or even orbital contribution can be ambiguous due to considerable differences between levels of theory or between model and real systems.

Thus, here we propose to abandon individual states and attempt to directly extrapolate entire absorption bands. In particular, we concentrate on three key parameters for the definition of an extrapolated band i: position ω_i^{ext}, height ε_i^{ext}, and half-width σ_i^{ext}. These parameters, collectively called $\{P^{ext}\}$, are obtained through an ONIOM-type extrapolation from individual subcalculations as:

$$P^{ext} = P_{real}^{low} + P_{model}^{high} - P_{model}^{low} \tag{6.3}$$

Once the P^{ext}s are computed with Eq. 6.3, a Gaussian envelope is assigned to the extrapolated band as:

$$\varepsilon^{ext}(\omega) = \varepsilon_i^{ext} \exp\left[-\left(\frac{\omega - \omega_i^{ext}}{\sigma_i^{ext}}\right)^2\right] \tag{6.4}$$

The final extrapolated spectrum is then built from the sum of the individual extrapolated bands. In this approach, the spectral bands from each subcalculation may include any number of excited states with different oscillator strengths. Thus, the issue of individual state-matching is transformed into the issue of finding and matching a relatively small number of bands from each subcalculation (more precisely, finding and matching the band parameters $\{P^{sub}\}$). Although defining and matching bands may still be ambiguous in certain cases, as we shall see in Section 6.4, it is still a much simpler and better defined problem than matching a large number of states individually. We note that information about the nature of the states that mostly contribute to a band in the extrapolated spectrum can always be recovered from an analysis of the subcalculations spectra. This qualitative analysis can help in the characterization of the band contributions while still avoiding the cumbersome state-matching in the extrapolation.

The first step is to simulate the spectrum for each subcalculation. This is achieved by assigning a Gaussian function to each excitation, which is proportional to the extinction coefficient through[210]:

$$\varepsilon(\omega) = \frac{D_i \omega_i}{4 * 2.926 \times 10^{-39} \sqrt{\pi}\sigma} exp\left[-\left(\frac{\omega - \omega_i}{\sigma}\right)^2\right] \tag{6.5}$$

where ω_i is the excitation energy in eV, σ is the half-width at ε_{max}/e (we set σ = 0.4 eV), and D_i is the dipole strength (in cgs units), which is related to the oscillator strength (f_i) by:

$$D_i = \frac{f_i}{3.7922 \times 10^{33} \omega_i} \tag{6.6}$$

The final spectrum is obtained by summing all of the Gaussian functions. Note that this choice of fitting is not unique, and others can be selected without altering the procedure for obtaining $\{P^{sub}\}$ described below.

Once the full spectrum for a subcalculation is obtained, our program performs a scan to find each peak, and their position ω_i^{sub} and height ε_i^{sub} are stored. Then, σ_i^{sub} is taken for each peak as the half-width at the height of ε_i^{sub}/e. If two bands are too close to each other, an alternative value of σ is taken at $2\varepsilon_i^{sub}/e$. A particularly delicate point is the treatment of band shoulders. These cannot be well represented by a single Gaussian function for the main band peak, so a different Gaussian is assigned to each shoulder. The detection of shoulders in the subcalculation bands is performed with a second scan after the primary Gaussian parameters for the peaks have been assigned. The shoulder detection scheme is shown in Figure 6.2. The shoulder detection is performed on both the

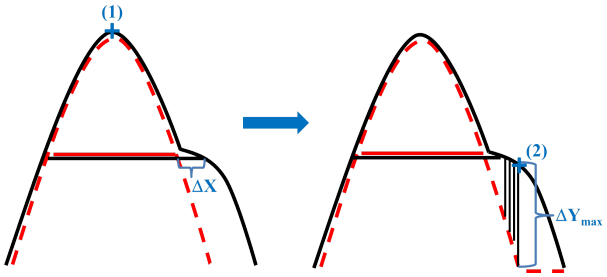

Figure 6.2: A schematic representation of shoulder detection. The dotted red curve represents the fitted Gaussian assigned to the mean peak (1). ΔX is the width difference between the fitted Gaussian and the actual spectrum, and it is used to assign the presence of a shoulder (2). ΔY is the height of the shoulder at (2).

right- and the left-hand sides of each band peak. As shown in Figure 6.2 for a right-hand search, starting from position 1 the difference between the assigned Gaussian and the actual spectrum (ΔX) is monitored. If $\Delta X > \alpha$, where α is a parameter, the presence of a shoulder is detected. At this

point, the height of the shoulder is obtained by scanning the difference in height between the actual spectrum and the Gaussian function assigned to the main peak (ΔY in Figure 6.2). The position for the shoulder Gaussian is chosen at the largest value of ΔY, i.e. point 2 in Figure 6.2. We use $\alpha = 0.1$ in this work. The band width parameter (σ) of the new Gaussian is taken as the width difference between the peak Gaussian and the actual band at shoulder height/e. However, since a shoulder is described by a single Gaussian that is very close to the main peak, the direct sum of the two Gaussian functions (one for the peak and one for the shoulder) would overestimate the height for both peak and shoulder. Therefore, we implemented an iterative optimization of both height and width of the two Gaussians so that their sum matches the original height $(\varepsilon_1, \varepsilon_2)$ and half-width (σ_1, σ_2) parameters.

Once the main features (peak/shoulder) of each subcalculation spectrum are found and a Gaussian function is assigned to each of them (i.e., $\{P^{sub}\}$), the extrapolated Gaussian functions are obtained using Eqs. 6.3 and 6.4. In our program, we can extrapolate a subset of the bands, and a sanity check is implemented so that the number of requested bands is equal or smaller than those found in the subcalculations. The complete implementation steps are outlined in Figure 6.3.

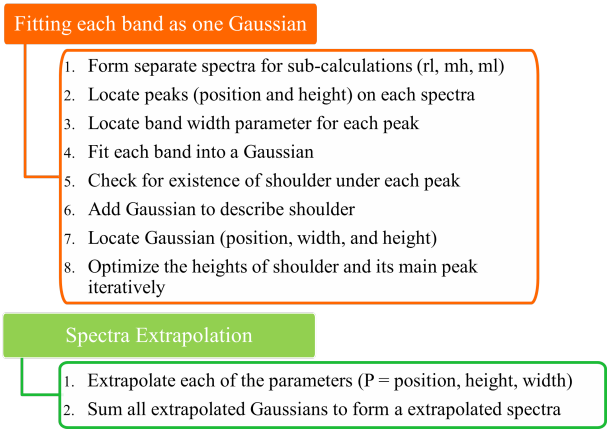

Figure 6.3: Scheme of the steps involved in the spectra extrapolation.

6.3 Computational Details

All calculations were performed with a development version of the GAUSSIAN suite of programs[170]. The geometries were optimized at the CAM-B3LYP[162]/6-311++G** level of theory, and are available in the Supporting Information. The same level is used for the *target* and *model/high* excited state calculations using the usual linear response approach[151–153]. The *target* spectra, i.e. the spectra obtained considering the entire molecule and the target level of theory, are used as reference for the comparison of the extrapolated and subcalculation spectra. We chose CAM-B3LYP because it has shown the ability to provide a balanced description of excitations of different nature (e.g., valence and Rydberg) thanks to its range separation[85,101,191]. In the layer definition, we follow the standard ONIOM approach to use hydrogen link atoms to cap severed covalent bonds. We test two methods as *low* level of theory: CIS[211] and B3LYP[158,212,213] with the 6-31+G* basis set. However, we report only the CIS result in the main text since the B3LYP results are qualitatively the same, and can be found in the Supporting Information.

6.4 Results

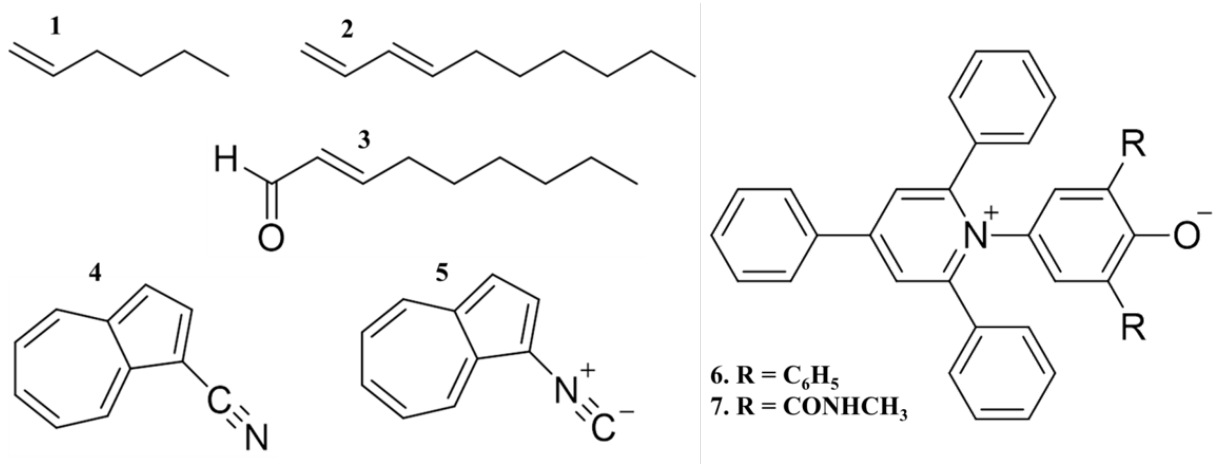

Figure 6.4: Test molecules used in evaluating the method.

We apply our method on six test molecules, shown in Figure 6.4: 1-hexene (**1**), 1,3-decadiene (**2**), 2-nonenal (**3**), 1-cyanoazulene (**4**), 1-isocyanoazulene (**5**), betaine-30 (**6**), and a methylamino-

carbonyl-substituted betaine (**7**). The size of the main chromophoric moiety is different across the test set so that we can evaluate the performance of our method in different conditions. In the following, the model system is shown in a ball-and-stick representation, while the other layer is shown in a tube-frame representation. The figures also show all spectra (i.e. target, subcalculations, and extrapolated) overlaid on each other for a direct comparison of the band positions, heights, and widths. Insets also show the target and subcalculation spectra individually, where each excited state is represented as a stick with length proportional to the oscillator strength. In this way, we are able to visualize the number and the relative importance of the individual excitations that really contribute to each band.

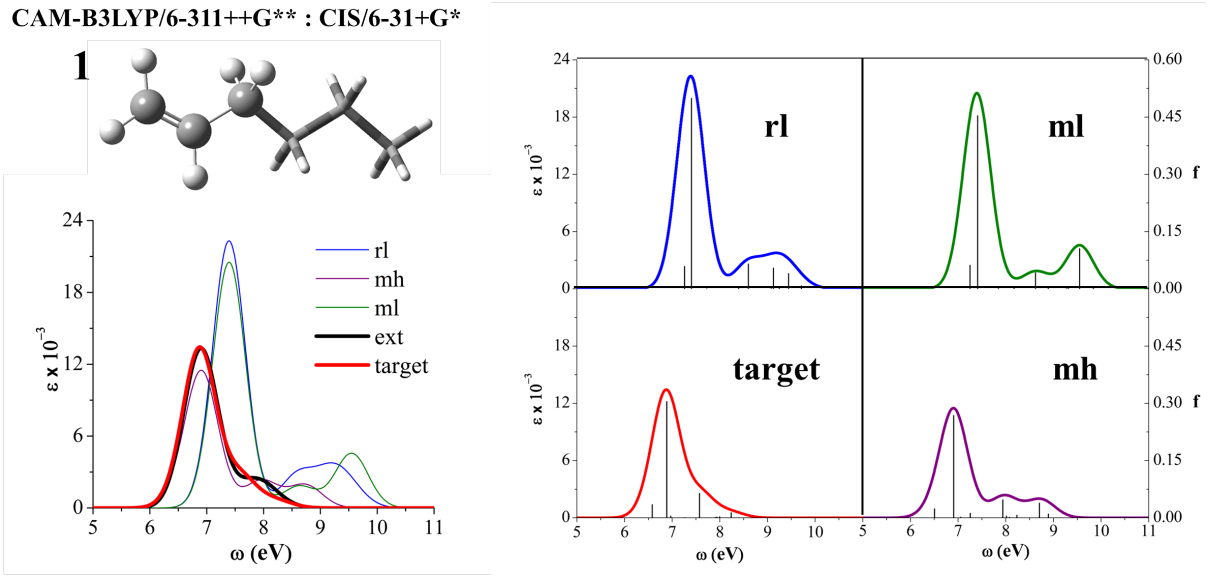

Figure 6.5: Absorption spectra for 1-hexene (**1**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

The results for **1** are shown in Figure 6.5. In this case, the chromophore is the alkene group. We consider ten states for the target and each subcalculation. As shown by the stick spectra in the insets of the figure, only a few states provide significant contribution to the final spectrum. Additionally, a shoulder in the target calculation is present to the right of the main peak (i.e., at higher energy). The same band is found as individual peaks in the subcalculations. The extrapolation is able to

correctly handle these differences so that the extrapolated spectrum is in excellent agreement with the target. Indeed, the position, height, and width of both peak and shoulder of the target spectrum are better reproduced by the extrapolated spectrum than by any of the subcalculations.

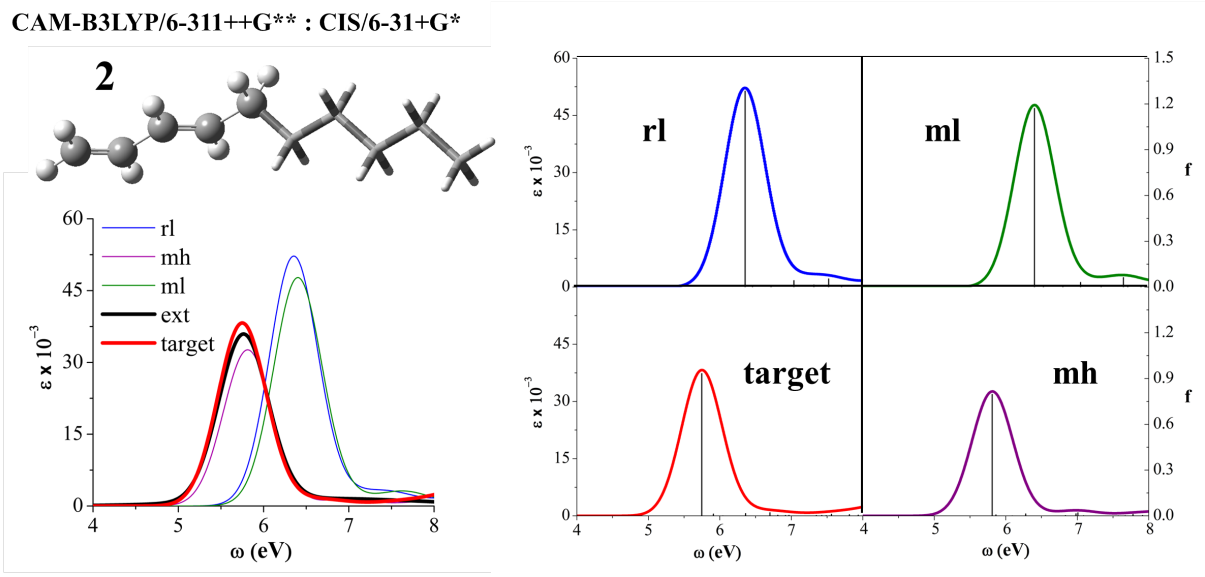

Figure 6.6: Absorption spectra for 1,3-decadiene (**2**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

For **2**, shown in Figure 6.6, two conjugated double bonds form the main chromophore. 100 states are used to build the spectrum. There is only one main peak within this subset of states, which is mainly due to one $\pi \rightarrow \pi^*$ excitation. However, a large number of states with small oscillator strength are also present that form a small tail in the 6.5-7 eV region. The extrapolated spectrum is able to describe both the large peak and the tail with excellent accuracy. Figure 6.6 clearly shows that all subcalculations individually do not reproduce the main peak in position or height. For this case in particular, the advantage of extrapolating bands rather than individual states is particularly evident for the tail part of the spectrum where many states with small oscillator strength contribute.

For **3**, shown in Figure 6.7, we consider the conjugated carbonyl and alkene groups as the chromophore. Considering 50 states produces three spectral features: one intense peak at lower

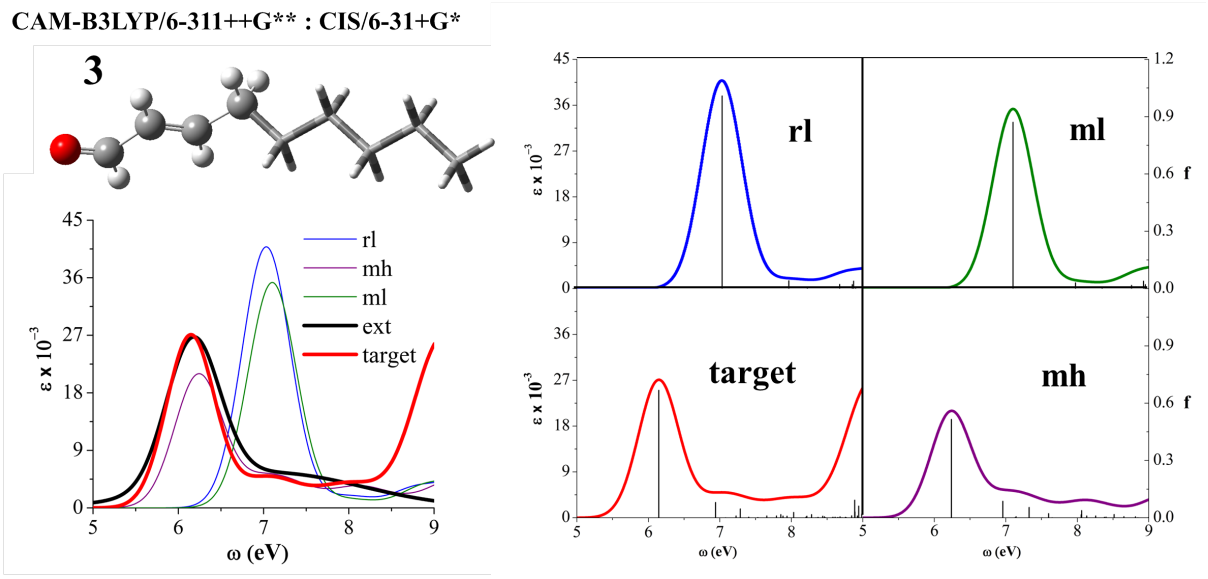

Figure 6.7: Absorption spectra for 2-nonenal (**3**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

energy, and two moderate peaks. The extrapolated spectrum provides the best match for the first, intense band both in terms of position and intensity. The two small bands, on the other hand, are merged together into a tail of reasonable intensity. This is due to the fact that the two shallow peaks strongly overlap, and the two Gaussians assigned to them merge together into one. It is not surprising that close, shallow peaks are problematic for our approach. Nonetheless, we find the simulation of this region acceptable at this stage, especially compared to that of the individual subcalculations, see Figure 6.7. One way to improve the description of these cases may be to extend the iterative optimization of the shoulders described in Section 6.2 to the entire spectrum in a recursive manner.

Test cases **4** and **5** are two isomers: 1-cyanoazulene and 1-isocyanoazulene[214]. The results for these molecules are shown in Figure 6.8. They represent an unusual choice of model system since the core layer (azulene) is larger than the second layer (substituent). However, they are interesting because they allow us to study the effect of different substituents on the absorption of the main chromophore. Indeed, azulene is a particularly versatile compound whose photochemical

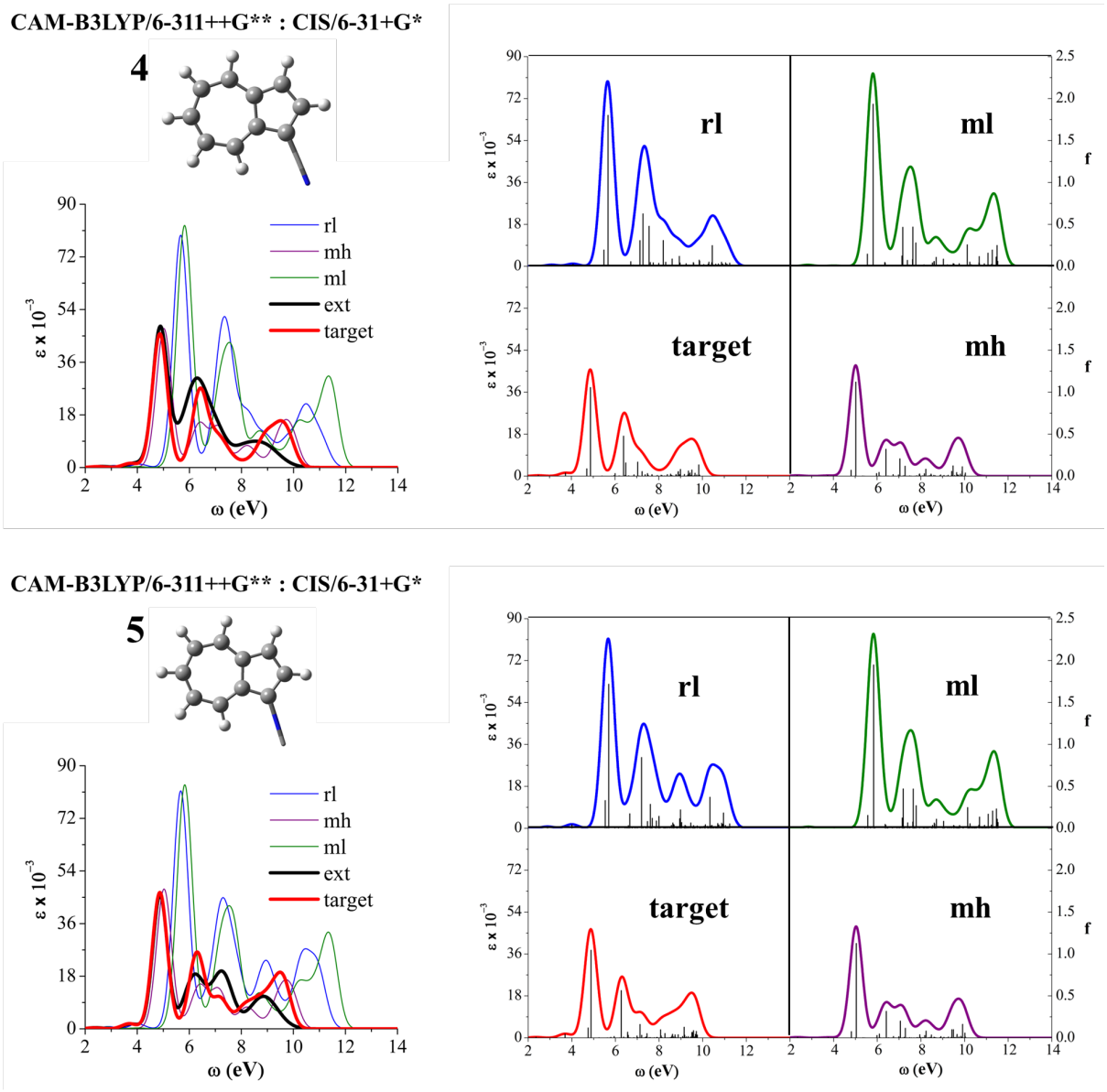

Figure 6.8: Absorption spectra for 1-cyanoazulene (**4**, top) and 1-isocyanoazulene (**5**, bottom). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

characteristics can be finely tuned by substitution in any of the eight possible positions[214]. Here we want to probe how our approach handles relatively small differences due to similar substituents. We considered 100 states for the target and each subcalculation to simulate the spectrum. As shown in the insets in Figure 6.8, there are a handful of states with large oscillator strength, but the band

structure in the 7-8 eV and 9-10 eV regions is due to a large number of states with moderate value of oscillator strength. The main effect of the substituents is to add a small band at low energy (< 4 eV), and to modify the higher energy bands in the 6-10 eV region, whereas the main peak at 5 eV is clearly due to azulene since it is qualitatively similar in both molecules and across subcalculations. From a comparison of the *target* insets of Figure 6.8, the effect of the different substituents in **4** and **5** is evident in the 7-9 eV region. We also point out that the small differences between the model subcalculations in the two molecules are due to the slightly different link atom bond lengths (1.025 and 1.080 Å, respectively) induced by the different substituents. The extrapolated spectrum for the cyanoazulene is in very good agreement with the target for the main peak as well as for the higher energy region. The extrapolated data improves the agreement both in terms of position and intensity of the bands compared to the subcalculations. The agreement for the isocyano isomer is excellent in the energy region below 5 eV, but it is not quite as good in the higher energy region. In the 6-8 eV region, the first peak is slightly underestimated while the second is slightly overestimated. However, this description is overall closer to the target spectrum than any of the subcalculations. Moreover, the extrapolation is able to capture the differences between the spectra of compounds **4** and **5** qualitatively and quantitatively.

Betaine-30 (**6**) is a large conjugated molecule, which has been used to define a polarity scale for solvents due to the sensitivity of its absorption spectrum to the environment[215]. The layer separation and simulated spectra are shown in Figure 6.9 (top). This is a particularly challenging case because there is no clear way to define a layer separation. The model system shown in the figure is directly conjugated, but weaker conjugation with the phenyl substituents may be important for the absorption spectrum. Additionally, the phenyl substituents are chromophores themselves, and they may contribute to the spectrum independently. We considered 100 states for the simulation of the spectrum, which results in a peak below 2 eV, a peak/shoulder around 3.5 eV, a peak around 5 eV, and a peak around 6 eV. The latter two are the result of many states with moderate oscillator strength. The model system calculations are able to reproduce the low energy peak, but the remaining parts of the spectrum are shifted at higher energies compared to the target calculation,

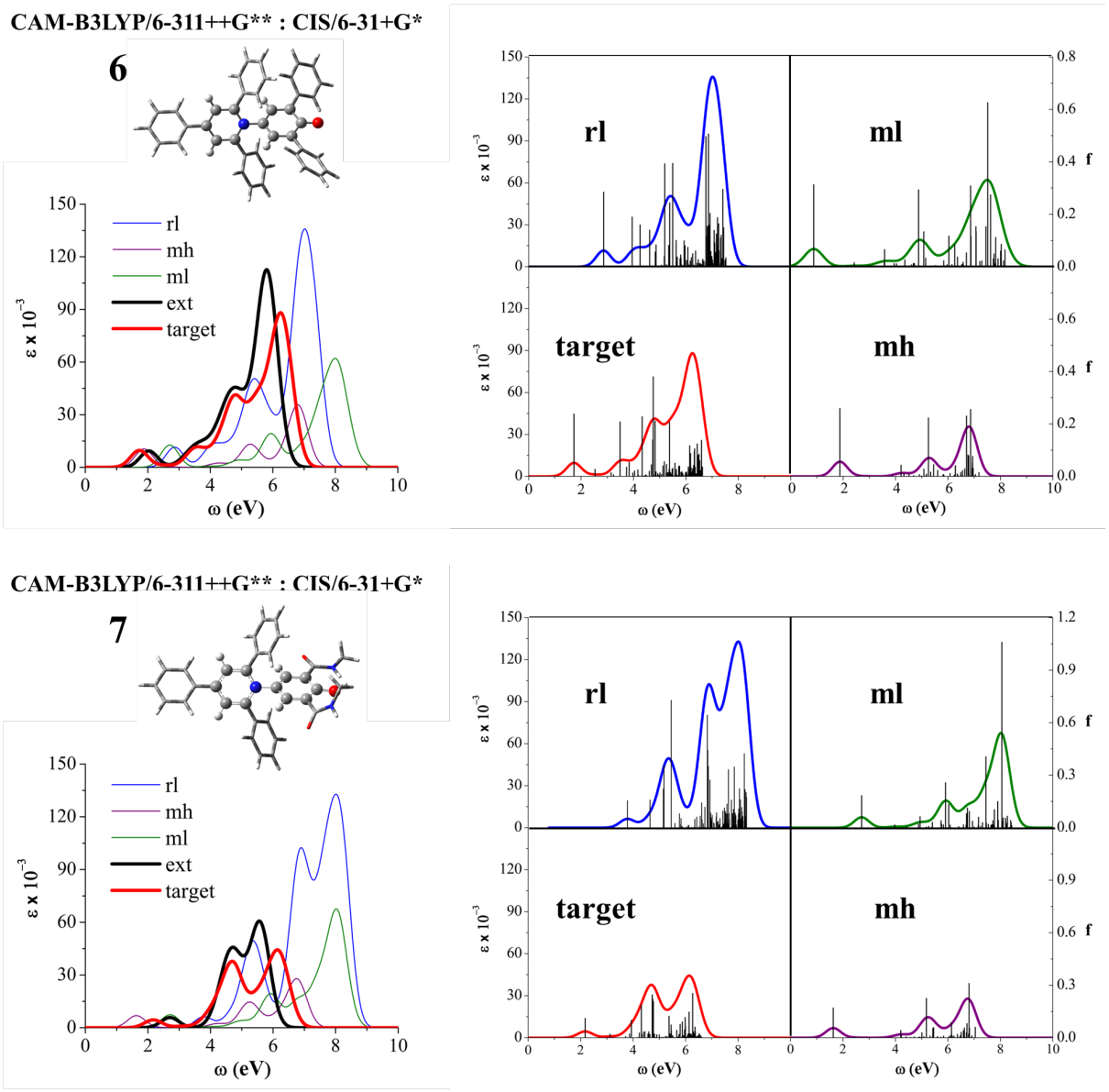

Figure 6.9: Absorption spectra for betaine-30 (**6**). The ball-and-stick representation describes the model system, while the tube-frame representation indicates the rest of the molecule. The insets report the subcalculations and target spectra with the stick representation of individual excitations. rl: real/low, ml: model/low, mh: model/high, ext: extrapolated.

which indicates that the effect of the substituents is significant. The first band is reproduced well by the extrapolation despite the overestimation of the low level subcalculations. The second and third features (at 3.5 and 5 eV, respectively) are in excellent agreement with the target, and show considerable improvement with respect to the subcalculations. The last band is slightly overestimated in intensity and underestimated in position, which is due to the increasing contributions of

the substituents. Nonetheless, the agreement with the target is remarkable considering the performance of the subcalculations. This case shows the robustness of the extrapolation approach even when the core layer definition is pushed towards the limit of acceptability. Compound **7** is a variant of betaine 30 where two of the phenyl rings are substituted with a methylamino-carbonyl group, see Figure 6.9 (bottom). The substitution considerably changes the target spectrum both in terms of position and intensity of the bands. For instance, the shoulder at 3.5 eV disappears, and the last peak is considerably less intense. We used the same model system as for **6** so that the model subcalculations are basically the same as before. The extrapolated spectrum is able to reproduce these modifications, providing a performance that is comparable to that for betaine 30: the first peak is slightly shifted towards higher energy, the second peak is well reproduced, and the last peak is shifted towards lower energy and overestimated in intensity. As before, the extrapolated spectrum is in better qualitative and quantitative agreement with the target one than any of the subcalculations.

6.5 Discussion and Conclusions

In this work, we present an extrapolation approach for the simulation of absorption spectra with hybrid QM/QM methods. The extrapolation is based on the ONIOM formula, but instead of considering individual states we extrapolate entire bands. Thus, we shift the problem of matching individual states, which is easy only for few low-lying bright states, to that of matching spectral features that are fewer in number. One of the key points is to recognize these features. This is easy for peaks but less so for shoulders, and in Section 6.2 we discuss how we propose to handle this issue. The results in Section 6.4, where the method is applied to a number of test cases, show the ability of the band-extrapolation idea to reproduce spectra computed at higher levels of theory on the entire molecule.

The main issue for the success of the method is the proper layer separation. This is not specific to our case, but it is common to all multi-layer methods. In previous work, we suggested a number of guidelines for the proper choice of the model system for excited state calculations[75–77], and

the same guidelines apply here. An issue that is specific to our method is the overlap between bands. Since we assign a Gaussian function to each spectral feature with extrapolated values of position, height, and width, overlap between adjacent Gaussian functions may lead to intensity overestimation in the extrapolated band. This is particularly problematic for shoulders and their relative main peak. To solve, or at least alleviate this problem, we suggest an iterative procedure for the definition of the optimal choice of these parameters within each subcalculation. The final parameters are then extrapolated in the usual way, see Eq. 6.3. This provides excellent results for test molecules **1**, **2**, and **4**. However, band overlap can also happen between peaks, and we have not addressed this issue yet. This results in overestimations of certain bands, e.g. in molecule **3**, although the overall shape of the spectrum is still well described. A promising approach to overcome this problem is to extend the iterative process used for shoulders to the entire spectrum recursively, but we leave this development for future work. Finally, test molecules **6** and **7** indicate that the extrapolation method is robust even when the layer separation is not optimal.

The range of applicability of this method is connected to the choice of model system. Since the latter is meant to be the main chromophoric moiety in the molecule, its excitations are likely to be concentrated in the low energy region of the spectrum. Thus, this is the region where the extrapolation method will work best. At higher energies, excitations from the second layer will contribute significantly to the overall spectrum, and the extrapolation may fail. However, this is not a strong limitation since practical applications in materials and energy research are focused on the low energy region. Therefore, we expect that this extrapolation approach will be useful, for instance, in the design of efficient dyes where the absorption spectrum is tuned by different ligands.

6.6 Acknowledgment

This material is based upon work supported by the National Science Foundation under Award No. EPS-0903806 and matching support from the State of Kansas through the Kansas Board of Regents.

Chapter 7

Concluding Remarks

This thesis focuses on the development and application of hybrid methods to efficiently perform accurate excited state calculations on large systems. In fact, accurate standard QM methods are usually associated with a high computational cost that limits their application to small to medium sized systems. In many cases, describing the effect of the environment surrounding the system of interest at quantum mechanical (QM) level is not possible, while eliminating the environment from the modeling is not a good approximation. A popular approach to reduce computational cost of treating large system is to use hybrid methods, which combine two or multiple computational methods to describe different regions of the system. Our research has pushed the boundaries in the area of hybrid methods in different directions, by making contributions to both additive and subtractive hybrid methods.

EOM-CCSD is one of the most accurate methods for computing one-electron excitation energies in molecules, although it has a steep computational scaling: O(N^6), where N is the size of the basis set. In Chapter 3, we presented an EOM-CCSD-PCM benchmark to evaluate its performance for electronic excitation energies of solvated molecules. Excitation energies computed with EOM-CCSD-PCM and time-dependent density functional theory-PCM (TD-DFT-PCM) for 16 test molecules were compared with experiment. EOM-CCSD-PCM consistently overestimates experimental excitation energies, with an error larger than in the gas phase by about 0.2 eV on average. For TD-DFT-PCM, the size and the sign of the error for different functionals vary among test molecules. These results are consistent among all tested basis sets. The consistent overestimation of experimental results by EOM-CCSD is due to the additional approximations introduced by the solvation model. Testing different basis sets shows that aug-cc-pVDZ and 6-311++G**

perform best, and than it is important to include a proper number of diffuse functions for the solute molecule, as too many diffuse functions that interact with the PCM region can be problematic. One of the important effects neglected by using PCM is the specific interaction between solvent and solute, such as hydrogen bonding (HB). To recover the effect of HB, we considered a micro-solvation + PCM approach, where QM solvent molecules are added explicitly to saturate the HB sites of the solute molecule, and the entire cluster is embedded in the continuum dielectric. A shift (Δ_{HB}) in excitation energy was calculated at DFT level between calculations with and without microsolvation. It was shown that this shift is not functional dependent, and applying the shift calculated at DFT level to the EOM-CCSD excitation energy considerably reduces its error. While introducing explicit solvation with forming micro-solvated clusters is necessary for describing the specific interaction between solvent and solute, treating explicit solvent molecules with CC is too expensive. Also, proper sampling is required for deciding where to place the solvent molecules. Thus, it is necessary to use a classical explicit solvation model to reproduce the effect of specific interactions efficiently.

In Chapter 4, we presented a combination of coupled cluster theory with a polarizable embedding scheme for ground and excited states. We combined CC theory with a classical polarizable force field (MMPol) based on induced dipoles with both the linear response (LR) and state specific (SS) formalisms for the response of the dipoles. We considered an approximate expression of the correlation environment response, originally developed for CC/PCM. This scheme allows us to include the explicit induced dipoles in the CC equations without increasing the computational cost. For testing the implementation, we considered several microsolvated systems, and compared their solvatochromic shifts obtained with classical embeddings and full CC calculations. We compared MMPol embedding with point charge embedding, and showed that the former performs best. We also showed that it is important to include induced dipoles on all of the atom centers of the classical region, and that too many diffuse functions in the basis set may be problematic due to too strong interaction with the environment (as for CC-PCM). We demonstrated the reliability of this computational protocol for the approximate scheme with the LR formalism because of its lower

computational cost. The induced dipoles scheme is part of a classical force field, where parameters are pre-computed. However, it is desirable to have a strategy to obtain such parameters on-the-fly for describing a generic environment.

Thus, in Chapter 5, we presented an induced dipole polarizable embedding scheme that can be used within QM/QM hybrid methods. Our group had already introduced fixed point-charge embedding for QM/QM based on the ONIOM (our own n-layered integrated molecular orbital and molecular mechanics) hybrid method. We now made this embedding polarizable using the induced-dipole model. We devised a strategy for obtaining atomic polarizabilities, which is the key parameter required for the induced-dipole embedding on-the-fly, by partitioning the molecular polarizability in atomic contributions. We used several related test molecules to compare excitation energies with different embedding methods against standard full QM calculations. These preliminary tests showed that this on-the-fly polarizable embedding is better than no or fixed embedding when substituent groups around the core region are polar.

However, we are often interested in an extended region of the absorption spectrum, instead of a single excited state. This is hard to do with standard ONIOM, as state-matching among the required subcalculations is challenging. In Chapter 6, we addressed this problem and presented a multi-state extrapolation method for UV/Vis absorption spectra based on a formula similar to that of ONIOM. Instead of state-by-state extrapolation, the method performs a ONIOM-like extrapolation on entire bands (using the band location and shape). The method was tested on seven chromophoric molecules, and compared with standard calculations of the entire molecule with the high QM method. We obtained extrapolated spectra that closely match the target spectra, and the results demonstrated the efficacy and robustness of this methodology.

In our research, the systems were chosen to be small enough to compare with standard full QM calculations on the entire system, with the purpose of performing benchmarks and testing the implementations. The next step is to perform systematic testing of EOM-CCSD-MMPol against experimental data. This will require molecular dynamic (MD) simulations to sample configurations of the solvent, from which we can extract snapshots to average out the spectra. Another

application is to use the QM/QM hybrid method with polarizable embedding along with the multi-state extrapolation of UV/Vis spectra to simulate the spectra of large chromophores, and compare against experimental spectra. This application would provide essential testing on different embedding schemes in QM/QM hybrid for large systems, and would determine the reliability of this hybrid scheme. These applications would be challenging, because the source of error would not only come from the intrinsic error of the QM methods, but also from the extrapolation procedure. Performing extensive testing on large-sized systems with hybrid approach, and make comparisons with experiment will help to devise a systematic protocol for treating novel systems. Other applications of these methods could include reactions in solution or on solid supports, which are of great interest in material and energy science, and in biological research.

Two-layered hybrid methods may not be sufficient for accurately describing complex interaction with the environment. More complex multi-layered hybrid approaches may be needed to strike the best balance between limited computational resources and the need for a reliable description of the entire system. My research has developed and benchmarked hybrid methods that act as a starting point to design more complex methods. The next step would be to design multi-layered hybrid methods, which may include two layers treated at different QM levels, a layer treated with explicit classical models, and the outmost layer treated with an implicit model. This development is challenging as the mutual polarization between layers may be significant. Thus, special computational mechanisms would need to be designed for the system to reach complete mutual polarization efficiently.

In the future, hybrid methods could be combined with machine learning, which has also been a popular area of development with wide applications in different fields. One could achieve automatic system analysis, partitioning, and method selection based on the specific system under study. The program would start by performing preliminary calculations to analyze the system, then these preliminary information would be processed by a machine learning algorithm to partition the system into layers based on its size, the amount of computational resources, and desired accuracy/computing time. The program would determine the best hybrid approach based on number of

layers, available implementation, and the result of existing benchmarking. Finally, the program would perform the calculations, and output the results. The most challenging part of such project is to have various hybrid approaches benchmarked for a large number of test systems, in order to create an effective training set for the machine learning algorithm. In principle, this development could achieve complete automation for performing calculations with hybrid methods, which would expand their application to even more scientific areas. This would allow researchers who are not expert in theoretical calculations and method development to take advantage of hybrid methods, and perform complex simulations in a black-box manner.

Chapter 8

References

[1] Senn, H. M.; Thiel, W. QM/MM Methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[2] Caprasecca, S.; Jurinovich, S.; Viani, L.; Curutchet, C.; Mennucci, B. Geometry optimization in polarizable QM/MM Models: the induced dipole formulation. *J. Chem. Theory Comput.* **2014**, *10*, 1588–1598.

[3] Loco, D.; Polack, É.; Caprasecca, S.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P.; Mennucci, B. A QM/MM approach using the AMOEBA polarizable embedding: from ground state energies to electronic excitations. *J. Chem. Theory Comput.* **2016**, *12*, 3654–3661.

[4] Caprasecca, S.; Curutchet, C.; Mennucci, B. Toward a unified modeling of environment and bridge-mediated contributions to electronic energy transfer: a fully polarizable QM/MM/PCM approach. *J. Chem. Theory Comput.* **2012**, *8*, 4462–4473.

[5] Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. Electronic energy transfer in condensed phase studied by a polarizable QM/MM model. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

[6] Curutchet, C.; , J.; Muñoz-Losa, A.; Hossein-Nejad, H.; Scholes, G. D.; Mennucci, B. Photosynthetic light-aarvesting is tuned by the heterogeneous polarizable environment of the protein. *J. Am. Chem. Soc.* **2011**, *133*, 3078–3084.

[7] Menger, M. F. S. J.; Caprasecca, S.; Mennucci, B. Excited-state gradients in polarizable

QM/MM Models: An induced dipole formulation. *J. Chem. Theory Comput.* **2017**, *13*, 3778–3786.

[8] Wildman, A.; Donati, G.; Lipparini, F.; Mennucci, B.; Li, X. Nonequilibrium environment dynamics in a frequency-dependent polarizable embedding model. *J. Chem. Theory Comput.* **2019**, *15*, 43–51.

[9] Menger, M. F. S. J.; Plasser, F.; Mennucci, B.; González, L. Surface hopping within an exciton picture. an electrostatic embedding scheme. *J. Chem. Theory Comput.* **2018**, *14*, 6139–6148.

[10] Loco, D.; Lagardère, L.; Caprasecca, S.; Lipparini, F.; Mennucci, B.; Piquemal, J.-P. Hybrid QM/MM molecular dynamics with AMOEBA polarizable embedding. *J. Chem. Theory Comput.* **2017**, *13*, 4025–4033.

[11] Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

[12] Gao, J.; Xia, X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631–635.

[13] Maseras, F.; Morokuma, K. IMOMM: A new integrated ab initio + molecular mechanics geometry optimization scheme of equilibrium structures and transition states. *J. Comput. Chem.* **1995**, *16*, 1170–1179.

[14] Lin, H.; Truhlar, D. QM/MM: Ahat have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

[15] Thompson, M. A.; Schenter, G. K. Excited states of the bacteriochlorophyll b dimer of rhodopseudomonas viridis: a QM/MM Study of the photosynthetic reaction center that includes MM polarization. *J. Phys. Chem.* **1995**, *99*, 6374–6386.

[16] Bryce, R. A.; Buesnel, R.; Hillier, I. H.; Burton, N. A. A solvation model using a hybrid quantum mechanical/molecular mechanical potential with fluctuating solvent charges. *Chem. Phys. Lett.* **1997**, *279*, 367–371.

[17] Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. The effective fragment potential method: A QM-based MM approach to modeling environmental effects in chemistry. *J. Phys. Chem. A* **2001**, *105*, 293–307.

[18] Jensen, L.; van Duijnen, P. T.; Snijders, J. G. A discrete solvent reaction field model for calculating molecular linear response properties in solution. *J. Chem. Phys.* **2003**, *119*, 3800–3809.

[19] Nielsen, C. B.; Christiansen, O.; Mikkelsen, K. V.; Kongsted, J. Density functional self-consistent quantum mechanics/molecular mechanics theory for linear and nonlinear molecular properties: Applications to solvated water and formaldehyde. *J. Chem. Phys.* **2007**, *126*, 154112.

[20] Illingworth, C. J. R.; Parkes, K. E. B.; Snell, C. R.; Ferenczy, G. G.; Reynolds, C. A. Toward a consistent treatment of polarization in model QM/MM calculations. *J. Phys. Chem. A* **2008**, *112*, 12151–12156.

[21] Lipparini, F.; Barone, V. Polarizable force fields and polarizable continuum model: a fluctuating charges/PCM ppproach. 1. theory and implementation. *J. Chem. Theory Comput.* **2011**, *7*, 3711–3724.

[22] Steindal, A. H.; Ruud, K.; Frediani, L.; Aidas, K.; Kongsted, J. Excitation energies in solution: the fully polarizable QM/MM/PCM method. *J. Phys. Chem. B* **2011**, *115*, 3027–3037.

[23] Boulanger, E.; Thiel, W. Solvent boundary potentials for hybrid QM/MM computations using classical drude oscillators: a fully polarizable model. *J. Chem. Theory Comput.* **2012**, *8*, 4527–4538.

[24] 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.

[25] Gao, J.; Truhlar, D. G.; Wang, Y.; Mazack, M. J. M.; Löffler, P.; Provorse, M. R.; Rehak, P. Explicit polarization: a quantum mechanical framework for developing next feneration force fields. *Acc. Chem. Res.* **2014**, *47*, 2837–2845.

[26] Zeng, Q.; Liang, W. Analytic energy gradient of excited electronic state within TDDFT/MMpol framework: benchmark tests and parallel implementation. *J. Chem. Phys.* **2015**, *143*, 134104.

[27] Olsen, J. M. H.; Steinmann, C.; Ruud, K.; Kongsted, J. Polarizable density embedding: a new QM/QM/MM-based computational strategy. *J. Phys. Chem. A* **2015**, *119*, 5344–5355.

[28] Cramer, C. J.; Truhlar, D. G. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161–2200.

[29] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation Models. *Chem. Rev* **2005**, *105*, 2999–3094.

[30] Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3093.

[31] Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.

[32] Libisch, F.; Huang, C.; Carter, E. A. Embedded correlated wavefunction schemes: Theory and applications. *Acc. Chem. Res.* **2014**, *47*, 2768–2775.

[33] Goodpaster, J. D.; Barnes, T. A.; Miller III, T. F. Embedded density functional theory for covalently bonded and strongly interacting subsystems. *J. Chem. Phys.* **2011**, *134*, 164108.

[34] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller III, T. F. Accurate basis set truncation for wavefunction embedding. *J. Chem. Phys.* **2013**, *139*, 024103.

[35] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. Accurate and systematically improvable density functional theory embedding for correlated wavefunctions. *J. Chem. Phys.* **2014**, *140*, 18A507.

[36] Daday, C.; König, C.; Valsson, O.; Neugebauer, J.; Filippi, C. State-specific embedding potentials for excitation-energy calculations. *J. Chem. Theory Comput.* **2013**, *9*, 2355–2367.

[37] Gomes, A. S. P.; Jacob, C. R.; Visscher, L. Calculation of local excitations in large systems by embedding wave-function theory in density-functional theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353–5362.

[38] Höfener, S.; Severo Pereira Gomes, A.; Visscher, L. Molecular properties via a subsystem density functional theory formulation: A common framework for electronic embedding. *J. Chem. Phys.* **2012**, *136*, 044104.

[39] Höfener, S.; Visscher, L. Calculation of electronic excitations using wave-function in wave-function frozen-density embedding. *J. Chem. Phys.* **2012**, *137*, 204120.

[40] Khait, Y. G.; Hoffmann, M. R. Embedding theory for excited states. *J. Chem. Phys.* **2010**, *133*, 044107.

[41] Wesolowski, T. A. Embedding potentials for excited states of embedded species. *J. Chem. Phys.* **2014**, *140*, 18A530.

[42] Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.

[43] Neugebauer, J.; Baerends, E. J. Exploring the ability of frozen-density embedding to model induced circular dichroism. *J. Phys. Chem. A* **2006**, *110*, 8786–8796.

[44] Höfener, S.; Gomes, A. S. P.; Visscher, L. Solvatochromic shifts from coupled-cluster theory embedded in density functional theory. *J. Chem. Phys.* **2013**, *139*, 104106.

[45] Wesołowski, T. A. *Comput. Chem.: Rev. Curr. Trends*; World Scientific, 2006; pp 1–82.

[46] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *J. Chem. Phys.* **2012**, *137*, 224113.

[47] Gotz, A. W.; Beyhan, S. M.; Visscher, L. Performance of kinetic energy functionals for interaction energies in a subsystem formulation of density functional theory. *J. Chem. Theory Comput.* **2009**, *5*, 3161–3174.

[48] Iannuzzi, M.; Kirchner, B.; Hutter, J. Density functional embedding for molecular systems. *Chem. Phys. Lett.* **2006**, *421*, 16–20.

[49] Lee, S. J.; Welborn, M.; Manby, F. R.; Miller III, T. F. Projection-based wavefunction-in-DFT embedding. *Accounts of chemical research* **2019**, *52*, 1359–1368.

[50] Roncero, O.; de Lara-Castells, M.; Villarreal, P.; Flores, F.; Ortega, J.; Paniagua, M.; Aguado, A. An inversion technique for the calculation of embedding potentials. *J. Chem. Phys.* **2008**, *129*, 184104.

[51] Ryabinkin, I. G.; Kananenka, A. A.; Staroverov, V. N. Accurate and efficient approximation to the optimized effective potential for exchange. *Phys. Rev. Lett.* **2013**, *111*, 013001.

[52] Yang, W.; Wu, Q. Direct method for optimized effective potentials in density-functional theory. *Phys. Rev. Lett.* **2002**, *89*, 143002.

[53] Kümmel, S.; Perdew, J. P. Optimized effective potential made simple: Orbital functionals, orbital shifts, and the exact Kohn-Sham exchange potential. *Phys. Rev. B* **2003**, *68*, 035103.

[54] Jensen, D. S.; Wasserman, A. Numerical methods for the inverse problem of density functional theory. *Int. J. Quantum Chem.* **2018**, *118*, e25425.

[55] Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D.; Basch, H.; Cohen, D. An effective fragment method for modeling solvent effects in quantum mechanical calculations. *J. Chem. Phys.* **1996**, *105*, 1968–1986.

[56] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701–706.

[57] Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. An integrated effective fragment—polarizable continuum approach to solvation: Theory and application to glycine. *J. Chem. Phys.* **2002**, *116*, 5023–5032.

[58] Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. Accurate methods for large molecular systems. *J. Phys. Chem. B* **2009**, *113*, 9646–9663.

[59] 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.

[60] Li, S.; Shen, J.; Li, W.; Jiang, Y. An efficient implementation of the "cluster-in-molecule" approach for local electron correlation calculations. *J. Chem. Phys.* **2006**, *125*, 074109.

[61] Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, 114109.

[62] 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 local coupled-cluster approach with singles and doubles. *J. Phys. Chem. A* **2010**, *114*, 8644–8657.

[63] Mayhall, N. J.; Raghavachari, K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336–1343.

[64] Sengupta, A.; Raghavachari, K. Prediction of accurate thermochemistry of medium and large sized radicals using connectivity-based hierarchy (CBH). *J. Chem. Theory Comput.* **2014**, *10*, 4342–4350.

[65] Sengupta, A.; Ramabhadran, R. O.; Raghavachari, K. Accurate and computationally efficient prediction of thermochemical properties of biomolecules using the generalized connectivity-based hierarchy. *J. Phys. Chem. B* **2014**, *118*, 9631–9643.

[66] Gao, J. Toward a molecular orbital derived empirical potential for liquid simulations. *J. Phys. Chem. B* **1997**, *101*, 657–663.

[67] Gao, J. A molecular-orbital derived polarization potential for liquid water. *J. Chem. Phys.* **1998**, *109*, 2346–2354.

[68] Song, L.; Han, J.; Lin, Y.-l.; Xie, W.; Gao, J. Explicit polarization (X-Pol) potential using ab initio molecular orbital theory and density functional theory. *J. Phys. Chem. A* **2009**, *113*, 11656–11664.

[69] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: a Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels – Alder Reactions and $Pt(P(t\text{-}Bu)_3)_2 + H_2$ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357–19363.

[70] Vreven, T.; Morokuma, K. The accurate calculation and prediction of the bond dissociation energies in a series of hydrocarbons using the IMOMO (integrated molecular orbital+molecular orbital) methods. *J. Chem. Phys.* **1999**, *111*.

[71] Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Jr., J. A. M.; Morokuma, K.; Frisch, M. J. Combining quantum mechanics methods with molecular mechanics methods in ONIOM. *J. Chem. Theory Comput.* **2006**, *2*, 815–826.

[72] Morokuma, K.; Wang, Q.; Vreven, T. Performance evaluation of the three-layer ONIOM method: case Study for a zwitterionic peptide. *J. Chem. Theory Comput.* **2006**, *2*, 1317–1324.

[73] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM method and its applications. *Chem. Rev* **2015**, *115*, 5678–5796.

[74] Mayhall, N. J.; Raghavachari, K. Charge transfer across ONIOM QM:QM boundaries: the impact of model system preparation. *J. Chem. Theory Comput.* **2010**, *6*, 3131–3136.

[75] Caricato, M.; Vreven, T.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. Using the ONIOM hybrid method to apply equation of motion CCSD to larger systems: Benchmarking and comparison with time-dependent density functional theory, configuration interaction singles, and time-dependent Hartree-Fock. *J. Chem. Phys.* **2009**, *131*.

[76] Caricato, M.; Vreven, T.; Trucks, G. W.; Frisch, M. J. Link atom bond length effect in ONIOM excited state calculations. *J. Chem. Phys.* **2010**, *133*.

[77] Caricato, M.; Vreven, T.; Trucks, G. W.; Frisch, M. J. Oscillator strengths in ONIOM excited state calculations. *J. Chem. Theory Comput.* **2011**, *7*, 180–187.

[78] Jovan Jose, K. V.; Raghavachari, K. "Electrostatic potential-based method of balancing charge transfer across ONIOM QM:QM boundaries". *J. Chem. Theory Comput.* **2014**, *10*, 4351–4359.

[79] Hratchian, H. P.; Krukau, A. V.; Parandekar, P. V.; Frisch, M. J.; Raghavachari, K. "QM:QM embedding using electronic densities within an ONIOM framework: Energies and analytic gradients". *J. Chem. Phys.* **2011**, *135*, 014105.

[80] Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. "ONIOM-based QM:QM electronic em-

bedding method using Löwdin atomic charges: Energies and analytic gradients". *J. Chem. Phys.* **2010**, *132*, 114107.

[81] Mayhall, N. J.; Raghavachari, K. "Charge transfer across ONIOM QM: QM boundaries: The impact of model system preparation". *J. Chem. Theory Comput.* **2010**, *6*, 3131–3136.

[82] Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-sensitized solar cells. *Chem. Rev.* **2010**, *110*, 6595–6663.

[83] Pastore, M.; Fantacci, S.; De Angelis, F. Modeling excited states and alignment of energy levels in dye-sensitized solar cells: Successes, failures, and challenges. *J. Phys. Chem. C* **2013**, *117*, 3685–3700.

[84] Kánnár, D.; Szalay, P. G. Benchmarking coupled cluster methods on valence singlet excited states. *J. Chem. Theory Comput.* **2014**, *10*, 3757–3765.

[85] Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. Electronic transition energies: a study of the performance of a large range of single reference density functional and wave function methods on valence and Rydberg states compared to experiment. *J. Chem. Theory Comput.* **2010**, *6*, 370–383.

[86] Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT benchmark: singlet-excited states of organic molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420–2435.

[87] Jacquemin, D.; Perpète, E. A.; Vydrov, O. A.; Scuseria, G. E.; Adamo, C. Assessment of long-range corrected functionals performance for n$\rightarrow \pi^*$ transitions in organic dyes. *J. Chem. Phys.* **2007**, *127*.

[88] Jacquemin, D.; Perpète, E. A.; Scuseria, G. E.; Ciofini, I.; Adamo, C. TD-DFT performance for the visible absorption spectra of organic dyes: conventional versus long-range hybrids. *J. Chem. Theory Comput.* **2008**, *4*, 123–135.

[89] Jacquemin, D.; Planchat, A.; Adamo, C.; Mennucci, B. TD-DFT assessment of functionals for optical 0–0 transitions in solvated dyes. *J. Chem. Theory Comput.* **2012**, *8*, 2359–2372.

[90] Guido, C. A.; Jacquemin, D.; Adamo, C.; Mennucci, B. Electronic excitations in solution: The interplay between state specific approaches and a time-dependent density functional theory description. *J. Chem. Theory Comput.* **2015**, *11*, 5782–5790.

[91] Koch, H.; Jorgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333–3344.

[92] 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**, *98*, 7029–7039.

[93] Koch, H.; Kobayashi, R.; de Merás, A. S.; Jorgensen, P. Calculation of size-intensive transition moments from the coupled cluster singles and doubles linear response function. *J. Chem. Phys.* **1994**, *100*, 4393–4400.

[94] Kállay, M.; Gauss, J. Calculation of excited-state properties using general coupled-cluster and configuration-interaction models. *J. Chem. Phys.* **2004**, *121*, 9257–9269.

[95] Caricato, M. Exploring potential energy surfaces of electronic excited states in solution with the EOM-CCSD-PCM method. *J. Chem. Theory Comput.* **2012**, *8*, 5081–5091.

[96] Caricato, M. Absorption and emission spectra of solvated molecules with the EOM–CCSD–PCM method. *J. Chem. Theory Comput.* **2012**, *8*, 4494–4502.

[97] Caricato, M.; Lipparini, F.; Scalmani, G.; Cappelli, C.; Barone, V. Vertical electronic excitations in solution with the EOM-CCSD method combined with a polarizable explicit/implicit solvent model. *J. Chem. Theory Comput.* **2013**, *9*, 3035–3042.

[98] Caricato, M. A comparison between state-specific and linear-response formalisms for the

calculation of vertical electronic transition energy in solution with the CCSD-PCM method. *J. Chem. Phys.* **2013**, *139*, 044116.

[99] Caricato, M. Implementation of the CCSD-PCM linear response function for frequency dependent properties in solution: Application to polarizability and specific rotation. *J. Chem. Phys.* **2013**, *139*, 114103.

[100] Wiberg, K. B.; de Oliveira, A. E.; Trucks, G. W. A Comparison of the electronic transition energies for ethene, isobutene, formaldehyde, and acetone calculated using RPA, TDDFT, and EOM-CCSD. Effect of basis sets. *J. Phys. Chem. A* **2002**, *106*, 4192–4199.

[101] Caricato, M.; Trucks, G. W.; Frisch, M. J.; Wiberg, K. B. Oscillator strength: how does TDDFT compare to EOM-CCSD? *J. Chem. Theory Comput.* **2011**, *7*, 456–466.

[102] Caricato, M. In *Photochemistry*; Fasani, E., Albini, A., Eds.; The Royal Society of Chemistry, 2015; Vol. 42; pp 197–214.

[103] Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. The polarizable embedding coupled cluster method. *J. Chem. Phys.* **2011**, *134*, 104108.

[104] Olsen, J. M.; Aidas, K.; Kongsted, J. Excited states in solution through polarizable embedding. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

[105] List, N. H.; Olsen, J. M. H.; Kongsted, J. Excited states in large molecular systems through polarizable embedding. *Phys. Chem. Chem. Phys.* **2016**, *18*, 20234–20250.

[106] Cammi, R. Coupled-cluster theories for the polarizable continuum model. II. Analytical gradients for excited states of molecular solutes by the equation of motion coupled-cluster method. *Int. J. Quantum Chem.* **2010**, *110*, 3040–3052.

[107] Cammi, R. Quantum cluster theory for the polarizable continuum model. I. The CCSD level with analytical first and second derivatives. *J. Chem. Phys.* **2009**, *131*, 164104.

[108] Caricato, M. Coupled cluster theory with the polarizable continuum model of solvation. *Int. J. Quantum Chem.* **2019**, *119*, e25710.

[109] Cammi, R. Coupled-cluster theory for the polarizable continuum model. III. A response theory for molecules in solution. *Int. J. Quantum Chem.* **2012**, *112*, 2547–2560.

[110] Caricato, M. Linear response coupled cluster theory with the polarizable continuum model within the singles approximation for the solvent response. *J. Chem. Phys.* **2018**, *148*, 134113.

[111] Christiansen, O.; Mikkelsen, K. V. Coupled cluster response theory for solvated molecules in equilibrium and nonequilibrium solvation. *J. Chem. Phys.* **1999**, *110*, 8348–8360.

[112] Christiansen, O.; Mikkelsen, K. V. A coupled-cluster solvent reaction field method. *J. Chem. Phys.* **1999**, *110*, 1365–1375.

[113] Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341 – 350.

[114] Lipparini, F.; Lagardère, L.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable evaluation of polarization energy and associated forces in polarizable molecular dynamics: I. toward massively parallel direct space computations. *J. Chem. Theory Comput.* **2014**, *10*, 1638–1651.

[115] Cances, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041.

[116] Mennucci, B.; Cances, E.; Tomasi, J. Evaluation of solvent effects in isotropic and anisotropic dielectrics and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation, and numerical applications. *J. Phys. Chem. B* **1997**, *101*, 10506–10517.

[117] Cancès, E.; Mennucci, B. New applications of integral equations methods for solvation continuum models: ionic solutions and liquid crystals. *J. Math. Chem.* **1998**, *23*, 309–326.

[118] Caricato, M. CCSD-PCM: improving upon the reference reaction field approximation at no cost. *J. Chem. Phys.* **2011**, *135*, 074113.

[119] Oliphant, N.; Adamowicz, L. Multireference coupled-cluster method using a single-reference formalism. *J. Chem. Phys.* **1991**, *94*, 1229–1235.

[120] Piecuch, P.; Oliphant, N.; Adamowicz, L. A state-selective multireference coupled-cluster theory employing the single-reference formalism. *J. Chem. Phys.* **1993**, *99*, 1875–1900.

[121] Kállay, M.; Szalay, P. G.; Surján, P. R. A general state-selective multireference coupled-cluster algorithm. *J. Chem. Phys.* **2002**, *117*, 980–990.

[122] Hackbusch, W. *Integral Equations*; Springer, 1995; pp 266–317.

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

[124] Shavitt, I.; Bartlett, R. J. *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory*; Cambridge Molecular Science; Cambridge University Press, 2009.

[125] Caricato, M.; Trucks, G. W.; Frisch, M. J. On the difference between the transition properties calculated with linear response- and equation of motion-CCSD approaches. *J. Chem. Phys.* **2009**, *131*, 174104.

[126] Simons, J.; Smith, W. D. Theory of electron affinities of small molecules. *J. Chem. Phys.* **1973**, *58*, 4899–4907.

[127] Bobrowicz, F.; Goddard III, W. Methods of electronic structure theory. *Modern theoretical chemistry* **1977**, *3*, 79–127.

[128] del Valle, F.; Tomasi, J. Electron correlation and solvation effects. I. basic formulation and preliminary attempt to include the electron correlation in the quantum mechanical polarizable continuum model so as to study solvation phenomena. *Chem. Phys.* **1991**, *150*, 139 – 150.

[129] Aguilar, M.; del Valle, F.; Tomasi, J. Electron correlation and solvation effects. II. The description of the vibrational properties of a water molecule in a dielectric given by the application of the polarizable continuum model with inclusion of correlation effects. *Chem. Phys.* **1991**, *150*, 151 – 161.

[130] Cammi, R.; Corni, S.; Mennucci, B.; Tomasi, J. Electronic excitation energies of molecules in solution: State specific and linear response methods for nonequilibrium continuum solvation models. *J. Chem. Phys.* **2005**, *122*, 104513.

[131] Corni, S.; Cammi, R.; Mennucci, B.; Tomasi, J. Electronic excitation energies of molecules in solution within continuum solvation models: Investigating the discrepancy between state-specific and linear-response methods. *J. Chem. Phys.* **2005**, *123*, 134512.

[132] Ren, S.; Harms, J.; Caricato, M. An EOM-CCSD-PCM benchmark for electronic excitation energies of solvated molecules. *J. Chem. Theory Comput.* **2017**, *13*, 117–124.

[133] Cieplak, P.; Dupradeau, F.-Y.; Duan, Y.; Wang, J. Polarization effects in molecular mechanical force fields. *J. Phys. Condens. Matter* **2009**, *21*, 333102.

[134] Caricato, M.; Scalmani, G.; Trucks, G. W.; Frisch, M. J. Coupled cluster calculations in solution with the polarizable continuum model of solvation. *J. Phys. Chem. Lett.* **2010**, *1*, 2369–2373.

[135] Caricato, M.; Scalmani, G.; Frisch, M. J. Brueckner doubles coupled cluster method with the polarizable continuum model of solvation. *J. Chem. Phys.* **2011**, *134*, 244113.

[136] Caricato, M. A corrected-linear response formalism for the calculation of electronic excitation energies of solvated molecules with the CCSD-PCM method. *Comp. Theor. Chem.* **2014**, *1040-1041*, 99–105.

[137] Aidas, K.; Kongsted, J.; Osted, A.; Mikkelsen, K. V.; Christiansen, O. Coupled cluster calculation of the n $\rightarrow \pi^*$ electronic transition of acetone in aqueous solution. *J. Phys. Chem. A* **2005**, *109*, 8001–8010.

[138] Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. The polarizable embedding coupled cluster method. *J. Chem. Phys.* **2011**, *134*, 104108.

[139] Eriksen, J. J.; Sauer, S. P. A.; Mikkelsen, K. V.; Jensen, H. J. A.; Kongsted, J. On the importance of excited state dynamic response electron correlation in polarizable embedding methods. *J. Comput. Chem.* **2012**, *33*, 2012–2022.

[140] Schwabe, T.; Sneskov, K.; Haugaard Olsen, J. M.; Kongsted, J.; Christiansen, O.; Hättig, C. PERI-CC2: a polarizable embedded RI-CC2 method. *J. Chem. Theory Comput.* **2012**, *8*, 3274–3283.

[141] Krause, K.; Klopper, W. Communication: A simplified coupled-cluster Lagrangian for polarizable embedding. *J. Chem. Phys.* **2016**, *144*, 041101.

[142] Krause, K.; Bauer, M.; Klopper, W. Approaching phosphorescence lifetimes in solution: the two-component polarizable-embedding approximate coupled-cluster method. *J. Chem. Theory Comput.* **2016**, *12*, 2853–2860.

[143] Ghosh, D.; Isayev, O.; Slipchenko, L. V.; Krylov, A. I. Effect of Solvation on the Vertical Ionization Energy of Thymine: From Microhydration to Bulk. *J. Phys. Chem. A* **2011**, *115*, 6028–6038.

[144] DeFusco, A.; Minezawa, N.; Slipchenko, L. V.; Zahariev, F.; Gordon, M. S. Modeling solvent effects on electronic excited states. *J. Phys. Chem. Lett.* **2011**, *2*, 2184–2192.

[145] Smith, Q. A.; Gordon, M. S.; Slipchenko, L. V. Effective fragment potential study of the interaction of DNA bases. *J. Phys. Chem. A* **2011**, *115*, 11269–11276.

[146] 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.

[147] Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. V. In *Annual Review of Physical Chemistry*; Johnson, M. A., Martinez, T. J., Eds.; 2013; Vol. 64; pp 553–578.

[148] Hoffman, G. J.; Gurunathan, P. K.; Francisco, J. S.; Slipchenko, L. V. Excited states of OH-(H_2O)(n) clusters for n=1-4: An ab initio study. *J. Chem. Phys.* **2014**, *141*, 104315.

[149] Cammi, R.; Fukuda, R.; Ehara, M.; Nakatsuji, H. Symmetry-adapted cluster and symmetry-adapted cluster-configuration interaction method in the polarizable continuum model: theory of the solvent effect on the electronic excitation of molecules in solution. *J. Chem. Phys.* **2010**, *133*, 024104.

[150] Fukuda, R.; Ehara, M.; Nakatsuji, H.; Cammi, R. Nonequilibrium solvation for vertical photoemission and photoabsorption processes using the symmetry-adapted cluster–configuration interaction method in the polarizable continuum model. *J. Chem. Phys.* **2011**, *134*, 104109.

[151] Bauernschmitt, R.; Ahlrichs, R. Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory. *Chem. Phys. Lett.* **1996**, *256*, 454–464.

[152] Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. Molecular excitation energies to high-lying bound states from time-dependent density-functional response theory: Characterization and correction of the time-dependent local density approximation ionization threshold. *J. Chem. Phys.* **1998**, *108*, 4439–4449.

[153] Stratmann, E. R.; Scuseria, G. E.; Frisch, M. J. An efficient implementation of time-dependent density-functional theory for the calculation of excitation energies of large molecules. *J. Chem. Phys.* **1998**, *109*, 8218–8224.

[154] Caillie, C. V.; Amos, R. D. Geometric derivatives of excitation energies using SCF and DFT. *Chem. Phys. Lett.* **1999**, *308*, 249–255.

[155] Caillie, C. V.; Amos, R. D. Geometric derivatives of density functional theory excitation energies using gradient-corrected functionals. *Chem. Phys. Lett.* **2000**, *317*, 159–164.

[156] Furche, F.; Ahlrichs, R. Adiabatic time-dependent density functional methods for excited state properties. *J. Chem. Phys.* **2002**, *117*, 7433–7447.

[157] Scalmani, G.; Frisch, M. J.; Mennucci, B.; Tomasi, J.; Cammi, R.; Barone, V. Geometries and properties of excited states in the gas phase and in solution: Theory and application of a time-dependent density functional theory polarizable continuum model. *J. Chem. Phys.* **2006**, *124*, 094107.

[158] Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[159] Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

[160] Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.* **1999**, *110*, 6158–6170.

[161] Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

[162] 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.

[163] Vydrov, O. A.; Scuseria, G. E. Assessment of a long-range corrected hybrid functional. *J. Chem. Phys.* **2006**, *125*.

[164] Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. Importance of short-range versus long-range Hartree-Fock exchange for the performance of hybrid density functionals. *J. Chem. Phys.* **2006**, *125*.

[165] Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. Tests of functionals for systems with fractional electron number. *J. Chem. Phys.* **2007**, *126*.

[166] Schäfer, A.; Horn, H.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets for atoms Li to Kr. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

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

[168] Lipparini, F.; Scalmani, G.; Mennucci, B.; Cances, E.; Caricato, M.; Frisch, M. J. A variational formulation of the polarizable continuum model. *J. Chem. Phys.* **2010**, *133*, 014106.

[169] Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

[170] Frisch, M. J. et al. *Gaussian Development Version, Revision H.09+, Gaussian, Inc., Wallingford CT* **2010**,

[171] Singh, I.; Ogata, R.; Moore, R.; Chang, C.; Scheuer, P. Electronic spectra of substituted naphthoquinones. *Tetrahedron* **1968**, *24*, 6053 – 6073.

[172] Grasselli, J. G. *CRC Atlas of spectral data and physical constants for organic compounds*; Chemical Rubber Company, 1973.

[173] Gore, P. H.; Wheeler, O. H. Absorption spectra of aromatic azo and related compounds. III. substituted azobenzenes. *J. Org. Chem.* **1961**, *26*, 3295–3298.

[174] Gustavsson, T.; Bányász, A.; Lazzarott, E.; Markovitsi, D.; Scalmani, G.; Frisch, M. J.; Barone, V.; ; Improta, R. Singlet excited-state behavior of uracil and thymine in aqueous solution: a combined experimental and computational study of 11 uracil derivatives. *J. Am. Chem. Soc.* **2006**, *128*, 607–619.

[175] Haszeldine, R. N.; Jander, J. Studies in spectroscopy. Part VI. Ultra-violet and infra-red spectra of nitrosamines, nitrites, and related compounds. *J. Chem. Soc.* **1954**, 691–695.

[176] Laurent, A. D.; Blondel, A.; Jacquemin, D. Choosing an atomic basis set for TD-DFT, SOPPA, ADC (2), CIS (D), CC2 and EOM-CCSD calculations of low-lying excited states of organic dyes. *Theoretical Chemistry Accounts* **2015**, *134*, 1–11.

[177] Ren, S.; Lipparini, F.; Mennucci, B.; Caricato, M. Coupled cluster theory with induced dipole polarizable embedding for ground and excited states. *J. Chem. Theory Comput.* **2019**, *15*, 4485–4496.

[178] Elking, D.; Darden, T.; Woods, R. J. Gaussian induced dipole polarization model. *J. Comput. Chem.* **2007**, *28*, 1261–1274.

[179] Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. Solvation effects on electronic transitions: exploring the performance of advanced solvent potentials in polarizable embedding calculations. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

[180] Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Luo, R.; Duan, Y. Development of polarizable models for molecular mechanical calculations I: parameterization of atomic polarizability. *J. Phys. Chem. B* **2011**, *115*, 3091–3099.

[181] Thompson, M. A. QM/MMpol: a consistent model for solute/solvent polarization. applica-

tion to the aqueous solvation and spectroscopy of formaldehyde, acetaldehyde, and acetone. *J. Phys. Chem.* **1996**, *100*, 14492–14507.

[182] Thole, B. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341–350.

[183] van Duijnen, P. T.; Swart, M. Molecular and atomic polarizabilities: Thole's model revisited. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

[184] Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **1992**, *96*, 135–149.

[185] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[186] Biancardi, A.; Barnes, J.; Caricato, M. Point charge embedding for ONIOM excited states calculations. *J. Chem. Phys.* **2016**, *145*, 224109.

[187] Miller, K. J. Calculation of the molecular polarizability tensor. *J. Am. Chem. Soc.* **1990**, *112*, 8543–8551.

[188] Ren, S.; Caricato, M. Multi-state extrapolation of UV/Vis absorption spectra with QM/QM hybrid methods. *J. Chem. Phys.* **2016**, *144*, 184102.

[189] Hardin, B. E.; Snaith, H. J.; McGehee, M. D. "The renaissance of dye-sensitized solar cells". *Nat. Photonics* **2012**, *6*, 162–169.

[190] Dev, P.; Agrawal, S.; English, N. J. "Determining the appropriate exchange-correlation functional for time-dependent density functional theory studies of charge-transfer excitations in organic dyes". *J. Chem. Phys.* **2012**, *136*, 224301.

[191] Jacquemin, D.; Planchat, A.; Adamo, C.; Mennucci, B. "A TD-DFT assessment of functionals for optical 0-0 transitions in solvated dyes". *J. Chem. Theory Comput.* **2012**, *8*, 2359–2372.

[192] Jacquemin, D.; Duchemin, I.; Blase, X. 0–0 energies using hybrid schemes: benchmarks of TD-DFT, CIS(D), ADC(2), CC2, and BSE/GW formalisms for 80 real-life compounds. *J. Chem. Theory Comput.* **2015**, *11*, 5340–5359.

[193] Jacquemin, D.; Adamo, C. In *Density-functional methods for excited states*; Ferré, N., Filatov, M., Huix-Rotllant, M., Eds.; Springer International Publishing: Cham, 2016; Chapter Computational Molecular Electronic Spectroscopy with TD-DFT, pp 347–375.

[194] Huang, C.; Pavone, M.; Carter, E. A. "Quantum mechanical embedding theory based on a unique embedding potential". *J. Chem. Phys.* **2011**, *134*, 154110.

[195] Huang, C.; Carter, E. A. "Potential-functional embedding theory for molecules and materials.". *J. Chem. Phys.* **2011**, *135*, 194104.

[196] Libisch, F.; Huang, C.; Carter, E. A. "Embedded correlated wavefunction schemes: theory and applications". *Acc. Chem. Res.* **2014**, *47*, 2768–2775.

[197] Goodpaster, J. D.; Barnes, T. A.; Miller, T. F. "Embedded density functional theory for covalently bonded and strongly interacting subsystems". *J. Chem. Phys.* **2011**, *134*, 164108.

[198] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller III, T. F. "Accurate and systematically improvable density functional theory embedding for correlated wavefunctions". *J. Chem. Phys.* **2014**, *140*, 18A507.

[199] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller III, T. F. "Accurate basis set truncation for wavefunction embedding". *J. Chem. Phys.* **2013**, *139*, 024103.

[200] Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller III, T. F. "Embedded mean-field theory". *J. Chem. Theory Comput.* **2015**, *11*, 568–580.

[201] Daday, C.; König, C.; Valsson, O.; Neugebauer, J.; Filippi, C. "State-specific embedding potentials for excitation-energy calculations". *J. Chem. Theory Comput.* **2013**, *9*, 2355–2367.

[202] Neugebauer, J.; Baerends, E. J. "Exploring the ability of frozen-density embedding to model induced circular dichroism". *J. Phys. Chem. A* **2006**, *110*, 8786–8796.

[203] Gomes, A. S. P.; Jacob, C. R. "Quantum-chemical embedding methods for treating local electronic excitations in complex chemical systems". *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2012**, *108*, 222–277.

[204] Höfener, S.; Gomes, A. S. P.; Visscher, L. "Solvatochromic shifts from coupled-cluster theory embedded in density functional theory". *J. Chem. Phys.* **2013**, *139*, 104106.

[205] Höfener, S.; Severo Pereira Gomes, A.; Visscher, L. "Molecular properties via a subsystem density functional theory formulation: A common framework for electronic embedding". *J. Chem. Phys.* **2012**, *136*, 044104.

[206] Höfener, S.; Visscher, L. "Calculation of electronic excitations using wave-function in wave-function frozen-density embedding". *J. Chem. Phys.* **2012**, *137*, 204120.

[207] Khait, Y. G.; Hoffmann, M. R. "Embedding theory for excited states". *J. Chem. Phys.* **2010**, *133*, 044107.

[208] Wesolowski, T. A. "Embedding potentials for excited states of embedded species". *J. Chem. Phys.* **2014**, *140*, 18A530.

[209] `https://caricatogroup.ku.edu/software`, accessed 02/19/2016.

[210] Harada, N.; Chen, S.-M. L.; Nakanishi, K. Quantitative definition of exciton chirality and the distant effect in the exciton chirality method. *J. Am. Chem. Soc.* **1975**, *97*, 5345–5352.

[211] Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **1992**, *96*, 135–149.

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

[213] Lee, C. T.; Yang, W. T.; 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.

[214] Robinson, R. E.; Holovics, T. C.; Deplazes, S. F.; Powell, D. R.; Lushington, G. H.; Thompson, W. H.; Barybin, M. V. Five possible isocyanoazulenes and electron-rich complexes thereof: a quantitative organometallic approach for probing electronic inhomogeneity of the azulenic framework. *organometallics* **2005**, *24*, 2386–2397.

[215] Reichardt, C.; Welton, T. *Solvents and solvent effects in organic chemistry*, 4th ed.; Wiley-VCH, 2011.

Chapter 9

Appendix for Induced-Dipole Polarizable Embedding for the ONIOM(QM:QM) Hybrid Method

Table 9.1: Optimized geometry (Å) for **1**

Atoms	x	y	z
C	2.61016	3.56223	-0.85390
C	4.27851	2.17901	-1.87962
C	2.58256	-2.81140	2.41396
C	4.24449	-2.83720	0.68902
C	3.16748	-3.92186	3.00550
C	4.82779	-3.94832	1.27897
C	4.87375	3.31003	-2.41790
C	3.20791	4.69322	-1.39117
C	4.29368	-4.49992	2.43681
C	4.34239	4.57188	-2.18160
C	-6.06987	1.13323	-0.21361
C	-6.04304	-1.24802	0.11216
C	-7.42885	-1.26315	0.11071
C	-7.45565	1.11580	-0.22161
C	-0.09860	-1.77762	-2.36314
C	-1.27676	-3.43028	-1.06719
C	-0.62992	-4.44034	-1.76602
C	0.27523	-4.12039	-2.76569
C	0.53331	-2.78754	-3.06605
C	-0.16542	1.83923	2.30210
C	-1.34371	3.42884	0.93045
C	-0.71648	4.47182	1.59924
C	0.44882	2.88135	2.97312
C	0.18100	4.19948	2.61971

Atoms	x	y	z
C	-8.13882	-0.08185	-0.05781
H	4.81606	5.45281	-2.59901
H	2.79406	5.67278	-1.17910
H	1.74545	3.67339	-0.20885
H	4.71163	1.20326	-2.04319
H	5.76670	3.20469	-3.02355
H	4.68016	-2.39790	-0.19599
H	5.71403	-4.38421	0.83187
H	4.75838	-5.36362	2.89794
H	2.75029	-4.32681	3.92066
H	1.72194	-2.35128	2.88668
H	-2.03001	3.64227	0.11910
H	-0.92316	5.49552	1.31237
H	0.68042	5.01053	3.13516
H	1.15445	2.66319	3.76472
H	0.07580	0.81623	2.55832
H	-1.97107	-3.68143	-0.27361
H	-0.82687	-5.47620	-1.51909
H	0.78783	-4.90667	-3.30600
H	1.24326	-2.53241	-3.84269
H	0.13287	-0.74322	-2.58001
H	-5.54618	2.06959	-0.36420
H	-9.22181	-0.09456	-0.06147
H	-7.95610	-2.19925	0.24766
H	-8.00379	2.03927	-0.36226
H	-5.49836	-2.17178	0.26614
C	1.04452	-1.07405	0.57011
C	2.41772	-1.11121	0.59153
C	3.19984	-0.03497	-0.02434
C	2.43180	1.09493	-0.55424
C	1.05873	1.07042	-0.54382
C	0.35790	-0.00113	0.00498
N	-1.07795	-0.00611	-0.02375
C	-1.77655	1.02030	0.55638
C	-1.74847	-1.03886	-0.62394
O	4.43878	-0.06894	-0.07721
C	3.12280	2.28399	-1.09991

Atoms	x	y	z
C	3.09756	-2.25572	1.23880
C	-3.15564	1.02041	0.52147
C	-3.12867	-1.06662	-0.60465
C	-3.87109	-0.03246	-0.04296
C	-5.34627	-0.04905	-0.04835
C	-1.06589	2.10837	1.27113
C	-1.00895	-2.09522	-1.35482
H	0.48357	-1.90342	0.98636
H	0.50955	1.89146	-0.99073
H	-3.62025	-1.88140	-1.11687
H	-3.66927	1.82962	1.02078

Table 9.2: Optimized geometry (Å) for **2**

Atoms	x	y	z
C	5.71851	1.30013	0.56075
C	5.83829	-0.88960	-0.42409
C	7.22232	-0.82498	-0.39751
C	7.10265	1.35919	0.59561
C	-0.01935	-2.51603	1.66163
C	1.17588	-3.56030	-0.15084
C	0.60123	-4.78055	0.17223
C	-0.27395	-4.87157	1.24304
C	-0.57646	-3.73987	1.98978
C	-0.13632	2.38805	-1.80843
C	0.81260	3.53179	0.08651
C	0.14710	4.69260	-0.28072
C	-0.78496	3.55324	-2.17979
C	-0.64945	4.70549	-1.41502
C	7.85767	0.29819	0.11444
C	-5.46398	1.83590	3.20114
C	-5.00912	-0.13513	-2.74591
H	1.42218	3.51996	0.98240
H	0.24529	5.58284	0.32783
H	-1.17115	5.61048	-1.70130
H	-1.40644	3.55915	-3.06662
H	-0.26121	1.49011	-2.39917
H	1.84684	-3.48831	-0.99901

Atoms	x	y	z
H	0.82796	-5.65582	-0.42348
H	-0.73008	-5.82195	1.49128
H	-1.26222	-3.80605	2.82505
H	-0.27867	-1.63667	2.23605
H	5.13862	2.12476	0.95726
H	8.93943	0.34641	0.13870
H	7.80630	-1.65217	-0.78180
H	7.59294	2.23406	1.00445
H	5.35222	-1.76239	-0.84301
H	-4.66544	-2.11758	-3.41216
H	-5.32647	2.91854	3.14437
H	-6.52195	1.60463	3.08128
H	-5.13957	1.51579	4.19469
H	-5.15543	0.54925	1.46883
H	-4.32066	0.64705	-2.42930
H	-5.34900	0.09994	-3.75692
H	-5.85153	-0.13573	-2.05487
O	-2.72773	1.97546	2.85288
O	-2.71900	-2.90464	-2.31985
N	-4.72055	1.16092	2.16440
N	-4.32813	-1.41364	-2.77309
C	-3.38562	1.28580	2.07299
C	-3.24947	-1.82682	-2.06448
C	5.06987	0.17359	0.05262
C	0.66740	2.37215	-0.66934
C	0.86081	-2.42019	0.58415
C	-1.29133	-0.85783	-0.99833
C	-2.66552	-0.92801	-1.01925
C	-3.46872	-0.28465	0.00211
C	-1.36068	0.61399	0.90592
C	-0.63828	-0.08046	-0.05170
N	0.81271	-0.01969	-0.03865
C	1.44300	1.16453	-0.29221
C	1.53546	-1.14388	0.24393
O	-4.71319	-0.44463	0.05762
C	2.82091	1.23442	-0.24020
C	2.91607	-1.08571	0.24937

Atoms	x	y	z
C	3.59475	0.10693	0.02038
H	-0.73715	-1.43100	-1.73349
H	-0.87645	1.22744	1.65840
H	3.45392	-1.98594	0.50945
H	3.28536	2.18102	-0.47637
C	-2.73234	0.52620	0.95846

Table 9.3: Optimized geometry (Å) for **3**

Atoms	x	y	z
C	2.51669	3.73022	-0.26330
C	4.30795	2.55926	-1.37592
C	2.60063	-3.27467	1.73650
C	4.37496	-2.83740	0.16396
C	3.21732	-4.46604	2.11255
C	4.98913	-4.03089	0.54053
C	4.89315	3.78231	-1.69979
C	3.10479	4.95099	-0.58748
C	4.41642	-4.85330	1.51229
C	4.29755	4.98381	-1.31185
C	-6.10290	1.17026	0.08210
C	-6.09837	-1.23260	-0.19333
C	-7.49078	-1.23284	-0.19903
C	-7.49528	1.16716	0.07169
C	-8.19458	-0.03370	-0.06766
C	0.48469	2.20071	3.51722
C	0.55383	-2.27989	-3.52190
H	4.76189	5.93420	-1.56220
H	2.63548	5.87636	-0.26280
H	1.60306	3.72608	0.32603
H	4.78672	1.63333	-1.66396
H	5.82512	3.79442	-2.25963
H	4.83643	-2.19675	-0.57505
H	5.92627	-4.31823	0.07007
H	4.90340	-5.77957	1.80609
H	2.76605	-5.08542	2.88378
H	1.68433	-2.97301	2.23773
H	-5.57265	2.11460	0.16308

Atoms	x	y	z
H	-9.28083	-0.03502	-0.07398
H	-8.02692	-2.17223	-0.29930
H	-8.03483	2.10527	0.16542
H	-5.56481	-2.17551	-0.26882
H	0.26848	0.48711	2.28083
H	0.35128	-0.57975	-2.26222
H	1.31796	2.74789	3.05927
H	-0.23255	2.91989	3.91443
H	0.86906	1.57869	4.32891
H	1.34581	-2.87221	-3.04749
H	-0.17631	-2.95941	-3.96305
H	0.99262	-1.65436	-4.30275
O	-1.60993	2.98031	1.81917
O	-1.63709	-3.01069	-1.93055
N	-0.19288	1.34758	2.54868
N	-0.12352	-1.42395	-2.55625
C	3.09716	2.50800	-0.65801
C	3.15603	-2.43983	0.74728
C	-5.38110	-0.03029	-0.05138
C	-1.14098	1.85067	1.72105
C	-1.12584	-1.90509	-1.78269
C	1.06567	-1.18875	0.30515
C	2.44845	-1.20480	0.32805
C	3.21134	0.00833	-0.04640
C	2.42214	1.22672	-0.33709
C	1.04033	1.17631	-0.31817
C	0.35207	-0.01475	-0.01056
N	-1.07698	-0.02753	-0.02597
C	-1.78503	0.89394	0.72187
C	-1.77346	-0.95022	-0.78318
O	4.45539	0.00618	-0.10269
C	-3.16567	0.93537	0.66118
C	-3.15549	-0.99050	-0.73803
C	-3.90666	-0.02863	-0.04264
H	0.51239	-2.09669	0.53411
H	0.46575	2.05715	-0.59480
H	-3.63123	-1.75502	-1.33938

Atoms	x	y	z
H	-3.65054	1.70015	1.25507

Table 9.4: Optimized geometry (Å) for **4**

Atoms	x	y	z
C	5.74287	-1.31193	-0.33858
C	5.86075	1.06330	0.11084
C	7.25027	1.00525	0.04258
C	7.13224	-1.36479	-0.41381
C	7.89027	-0.20762	-0.22186
C	-5.33648	-2.98202	-2.52950
C	-5.08221	1.05036	2.54218
C	-0.88248	-3.16494	2.72663
C	-0.91217	3.56329	-2.07864
H	-0.60432	1.47609	-1.84667
H	-0.22783	4.40789	-2.16463
H	-1.40717	3.39366	-3.03746
H	5.16381	-2.21359	-0.51490
H	8.97422	-0.25077	-0.27798
H	7.83393	1.90727	0.20186
H	7.62288	-2.30971	-0.62840
H	5.37723	2.00757	0.34311
H	-4.85865	3.14888	2.28669
H	-5.13949	-3.98948	-2.14497
H	-6.41078	-2.78622	-2.48662
H	-5.01089	-2.96044	-3.57672
H	-5.13278	-1.20321	-1.28714
H	-4.34646	0.24413	2.58900
H	-5.44350	1.24213	3.55818
H	-5.90767	0.71769	1.90680
H	-0.73881	-1.18218	1.96344
H	-1.55970	-3.70156	2.05142
H	-0.12237	-3.86141	3.08364
H	-1.45049	-2.77551	3.57440
H	-1.65775	3.77298	-1.30191
O	-2.59144	-2.88501	-2.10367
O	-2.94863	3.49547	0.89578
O	1.37985	-3.35621	1.10671

Atoms	x	y	z
O	1.41821	3.48190	-0.48651
N	-4.64848	-1.98017	-1.74225
N	-4.45159	2.25661	2.03553
N	-0.22826	-2.05124	2.05060
N	-0.14023	2.36819	-1.73179
C	-3.30413	-2.01150	-1.59495
C	-3.39837	2.38054	1.18435
C	5.08599	-0.09571	-0.07722
C	0.82312	-2.27048	1.22790
C	0.87973	2.44162	-0.85106
C	-1.36430	1.10287	0.69815
C	-2.74317	1.14300	0.63291
C	-3.50466	0.15148	-0.12663
C	-1.33384	-0.92415	-0.64966
C	-0.64887	0.06542	0.07364
N	0.79430	0.05262	0.10199
C	1.47112	-1.06604	0.53867
C	1.51025	1.13886	-0.34953
O	-4.75603	0.23812	-0.23865
C	2.84986	-1.13147	0.43408
C	2.89431	1.11318	-0.35258
C	3.61369	-0.03886	-0.00355
H	-0.84090	1.90027	1.22020
H	-0.80786	-1.72487	-1.16514
H	3.39168	2.01106	-0.69655
H	3.31078	-2.05082	0.77157
C	-2.71092	-0.90569	-0.75920