UC Davis

UC Davis Previously Published Works

Title

State averaged CASSCF in AMOEBA polarizable water model for simulating nonadiabatic molecular dynamics with nonequilibrium solvation effects

Permalink

Journal

ISSN

Author

Song, Chenchen

Publication Date

DOI

Copyright Information

Peer reviewed

State averaged CASSCF in AMOEBA polarizable water model for simulating nonadiabatic molecular dynamics with nonequilibrium solvation effects

Chenchen Song[1, a)]

Department of Chemistry, University of California Davis, Davis, CA 95616, USA.

This paper presents state-averaged complete active space self-consistent field (SA-CASSCF) in the atomic multipole optimized energetics for biomolecular applications (AMOEBA) polarizable water model, which enables rigorous simulation of nonadiabatic molecular dynamics with nonequilibrium solvation effects. The molecular orbital and configuration interaction coefficients of the solute wavefunction, and the induced dipoles on solvent atoms, are solved by minimizing the state averaged energy variationally. In particular, by formulating AMOEBA water models and polarizable continuum model (PCM) in a unified way, the algorithms developed for computing SA-CASSCF/PCM energies, analytical gradients, and non-adiabatic couplings in our previous work can be generalized to SA-CASSCF/AMOEBA by properly substituting a specific list of variables. Implementation of the method will be discussed with emphasis on how the calculations of different terms are partitioned between the quantum chemistry and molecular mechanics codes. We will present and discuss results that demonstrate the accuracy and performance of the implementation. Next, we will discuss results that compare three solvent models that work with SA-CASSCF, i.e. PCM, fixed-charge force fields, and the newly implemented AMOEBA. Finally, the new SA-CASSCF/AMOEBA method has been interfaced with the ab initio multiple spawning method to carry out non-adiabatic molecular dynamics simulations. The method is demonstrated by simulating the photodynamics of the model retinal protonated Schiff base molecule (PSB3) in water.

---
a)Electronic mail: ccsong@ucdavis.edu

I. INTRODUCTION

Many important photoreactions in life, such as vision[1] in animals and photoprotection[2] in plants, take place in solvents or in macromolecular systems and are affected by their interactions with the environments. Theoretical methods for simulating photoreactions in complex environments have the potential to improve our understanding about the evolutionary advantages of these photoreactions, and may also lead to the possibility of rationally controlling photodynamics through careful design of the environments. Because solvent environments are simpler than heterogenous macromolecular environments, developing theoretical methods for simulating photoreactions in solvent becomes the natural first step. Despite steady improvements in quantum chemistry software and methods over the past few decades, full quantum mechanical treatment of the solvated system is still impractical at the moment. Fortunately, because the excited wavefunction is generally localized on the chromophore, this suggests a natural multi-scale[3–5] description of the solvated system in which the electronically excited chromophore is treated quantum mechanically and the solvents are treated classically.

To help decide suitable classical solvent models for photoreactions, we need to first consider what solvent effects are known to have notable impacts on photoreactions from experimental studies. The electrostatic interaction is one of the fundamental types of solute-solvent interactions. For chromophores with strong intramolecular charge transfer characteristics, such as peridinin[6] and merocyanine[7], their excited state lifetimes have been observed to be sensitive to the polarity of the solvents. In particular, photoreactions in a polar solvent (e.g. water) are subject to nonequilibrium solvent polarization effects.[8] These nonequilibrium effects arise from the fact that the solvent electronic response to the electronic excitation of the chromophore is much faster than the corresponding rearrangements of solvent nuclei. This temporal evolution of the solvent polarization has been a major challenge in simulating photoreactions in solvents. In addition to electrostatic solvent effects, for photoreactions that involve large conformational changes, such as the photoisomerization reactions of diphenyl butadiene[9], stilbene[10,11] and sorivudine[12], the photoreaction rates are observed to be sensitive to the viscosity of the solvents. Other solute-solvent interactions that have been studied either experimentally or theoretically include the solvent steric or caging effects[13], the Pauli

2

repulsion interactions,[14] and the dispersion interactions.[15]

The commonly used classical solvent models can be divided into two categories: implicit models that approximate the solvent as a continuum[16], and explicit models that keep atomistic details of the solvent molecules.[17] In terms of their abilities of describing electrostatic interactions, the classical models can be further classified into polarizable and non-polarizable models. Most implicit models, such as the Onsager reaction field[18] and the polarizable continuum model (PCM)[19–21], are polarizable and can capture the difference in solvent polarization during vertical excitations. However, capturing the temporal evolution of solvent polarization into implicit models is much more difficult. Although there have been efforts along these lines using real-time formulation,[22,23] the corresponding analytical nuclear gradients are difficult to formulate, thus limiting their applicability for studying photoreactions. An important class of explicit solvent models are represented by molecular mechanics (MM) force field models. Traditional non-polarizable force fields, such as the TIP3P and TIP4P water models,[24–26] describe the electrostatic interactions using a set of fixed atomic partial charges, which are normally fitted to reproduce the ground state thermodynamic properties of the liquid solvent. Since the atomic partial charges are fixed and determined independently from the solute, such models are inherently incapable of describing any polarization changes in solvents in response to the solute excitations. In contrast, polarizable force fields[27,28] describe electronic polarization explicitly by introducing induced variables on each atom. Common choices for modeling the induced variables are the fluctuating charge model,[29,30] the Drude oscillator model[31,32] or the induced dipole model.[33,34] A major advantage of the polarizable explicit solvent models is that when combined with molecular dynamics simulations, they provide a natural way to simulate the temporal separation of the solvent response, where the induced variables model the fast response of the solvent electrons, and the dynamical changes of nuclear configurations model the slow response of the solvent nuclei.[35] In addition to electrostatic effects, another advantage of explicit models over implicit models for studying photoreactions is their ability to describe viscosity and steric effects when combined with molecular dynamics (MD).[10,36] In terms of other types of solute-solvent interactions such as the Pauli-exclusion effects and dispersion, these effects have been included as additional terms both in implicit models (e.g. the SMx models[37]) or explicit models (e.g. Lennard-Jones interaction[38]).

From the discussions in the previous paragraph, the explicit polarizable solvent models clearly stand out in their potential capability of capturing a wide range of important solvent effects important for photoreactions. Among the available explicit polarizable solvent models, the atomic multipole optimized energetics for biomolecular applications (AMOEBA)[39] force field is one of the most sophisticated models. It was originally developed for water[40,41] and has been shown to accurately reproduce the phase diagrams of water with systematically improved parameters.[42] Later, AMOEBA force field for organic compounds[43] and biological macromolecules[44] have also been developed and become widely used. The mutual polarization in the AMOEBA force field is described by an atomic induced dipole model,[34] in which the induced dipoles react to the permanent electrostatic moments as well as one another. This model is known to be good at reproducing the anisotropic molecular polarization response,[45] which is important for describing the solute-solvent interactions. When studying the electric fields exerted on a vibrational chromophore created by the solvents, Fried et al.[46] found that compared to fixed-charge and continuum models, AMOEBA is the only model that can provide consistent descriptions of nonpolar, polar, and hydrogen bonding solvent environments. These studies suggest that AMOEBA is suitable for modeling environment effects in photoreactions.

Ground state quantum mechanics/molecular mechanics (QM/MM) embedding approaches that combine density functional theory (DFT)[47,48] with AMOEBA force field have been developed first for solvent environments,[49,50] and later generalized to macromolecular environments that require proper treatments of the covalent QM/MM boundary.[51,52] Developments of excited state methods embedded in AMOEBA are more recent, and there have been continuous efforts towards combining the linear response time-dependent density functional theory (LR-TDDFT)[53,54] with AMOEBA.[55] The developments of the corresponding analytical nuclear gradients have enabled the simulation of adiabatic molecular dynamics on the excited states.[56] Recently, such approach has also been applied to simulating non-adiabatic molecular dynamics using trajectory surface hopping.[57] One advantage of LR-TDDFT is that its ground and excited state calculations are decoupled, making it simple to account for the different solvent polarization in response to each solute electronic state. However, the same decoupling leads to an incorrect topology of the potential energy surfaces around conical intersections,[58] which are critical for photoreactions as the main nonradiative tran-

sition channels. Therefore, multi-reference methods are generally more desirable for the purpose of searching conical intersections and simulating nonadiabatic photodynamics, and one of the most widely used multi-reference methods is the state-averaged complete active space self-consistent field theory (SA-CASSCF).[59–61] Although SA-CASSCF embedded in AMOEBA force fields hasn't been reported to our knowledge, there has been progress on polarizable embedding of SA-CASSCF using induced point dipoles,[62] the Drude oscillator model,[63] or the direct reaction field model.[64].

In this work, we will take the initial step of developing SA-CASSCF embedded in AMOEBA polarizable force field for nonadiabatic MD simulations, where we start with AMOEBA force field for small solvent molecules (e.g. water, chloroform). Recently, we developed SA-CASSCF in PCM[65] based on a self-consistent formulation, where the molecular orbital (MO) coefficients, configuration interaction (CI) coefficients and the induced surface charges are determined by minimizing the state-averaged energy variationally. The dynamic weight scheme proposed by Hagras et al.[63] is used, which smoothly interpolates between state-specific descriptions near the Franck-Condon region to account for different polarization corresponding to each state, and a state-averaged scheme near the conical intersection to ensure the correct topology. Because in the state-averaged formalism, the energy of each state is no longer variational with respect to the wavefunction coefficients and the induced charges, the corresponding analytical gradients and non-adiabatic couplings were developed based on the Lagrangian formulation.[66,67] One advantage of the self-consistent formulation is that the induced surface charges are treated in the same footing as MO and CI coefficients, making the corresponding algorithm a natural extension to the well-established gas phase algorithm. As pointed out in Nottoli et al.,[68] PCM and AMOEBA force fields can be unified in a general polarizable framework. As a result, this work will show how the ideas from the previous SA-CASSCF/PCM work can be generalized to SA-CASSCF/AMOEBA. By further interfacing with ab initio multiple spawning,[69,70] it will enable simulating photoreactions with nonequilibrium solvent effects in polar solvents, which cannot be described properly using PCM with a fixed dielectric constant.

This paper is structured as follows. In Section II A, we will start by reviewing the similarities between PCM and AMOEBA. Based on these similarities, in Section II B and II C, we will discuss what operations from SA-CASSCF/PCM can be carried over to SA-

CASSCF/AMOEBA, and what operations need to be implemented differently. We will then discuss the implementation of the method in Section II D, with the emphasis on how different terms are partitioned between QM codes and MM codes. Next, we will discuss results to help analyze the new method. Section III A discusses the accuracy and performance of the implementation. In Section III B, we will compare SA-CASSCF/AMOEBA with other solvent models and discuss how they differ in the electrostatic potential thus affect the excitation energies of the solvated chromophore. In Section III C, we will demonstrate nonadiabatic MD simulation using SA-CASSCF/AMOEBA, and we will discuss how the simulations in AMOEBA water models compare to that in the gas phase, and what improvements are required for future works.

Below we summarize the list of symbols and notations used throughout this paper:

- Molecular orbital indices: p, q, r, s

- QM atom indices: N, M

- PCM induced surface charge indices: n, m

- MM atom indices: A, B

- Cartesian indices from the set of $\{x, y, z\}$: i, j, k, l

- Arbitrary electronic states: R, S. Electronic states included in the state averaging: I, J. Target electronic state for gradients and non-adiabatic coupling calculations: Θ, Π

- For matrices and tensors, they are represented either by writing the indices explicitly (e.g. P_{pq}) or in bold without indices (e.g. $\mathbf{P}$).

- Parenthesis in superscripts or subscripts are used to denote quantities that are conceptually related but have different values, e.g. $E^{(\mathrm{PCM})}$ versus $E^{(\mathrm{AMOEBA})}$.

6

II. METHOD

A. Comparison of PCM and AMOEBA

Both PCM and AMOEBA are polarizable models, and their connections have been studied in previous research.[68] Here we briefly summarize some of the main points to facilitate later discussions. Because this section focuses on the solvent polarization, the solute polarization and non-polarizable interactions (e.g. van der Waals and valence interactions) will be omitted from the discussion without loss of generality.

We start by considering a solute in the PCM environment. The polarization of PCM is described by the induced surface charges q_n. The corresponding total energy can be written as

$$E^{(\text{PCM})} = E^{(\text{solute})} + \sum_n U_n^{(\text{solute})} q_n + \frac{1}{2} \sum_{nm} q_n A_{nm} q_m \tag{1}$$

$E^{(\text{solute})}$ represents the electrostatic energy of the solute, and $U^{(\text{solute})}$ is the electric potential generated by the solute. Their specific forms depend on the level of theory used for describing the solute. The last term represents the self-interaction among the surface charges. The induced surface charges q_n are determined variationally such that

$$\frac{\partial E^{(\text{PCM})}}{\partial q_n} = U_n^{(\text{solute})} + \sum_m A_{nm} q_m = 0 \tag{2}$$

which is equivalent to the following linear equations

$$\sum_m A_{nm} q_m = -U_n^{(\text{solute})} \tag{3}$$

and can be solved iteratively.

When the solvent polarization is described by the AMOEBA model, the polarization is represented by the induced dipoles d_{Ai}. In addition to induced dipoles, the solvent atoms also come with permanent multipoles (including partial charges, dipoles and quadrupoles). To clearly separate the induced dipoles from other electrostatic quantities, we will use the term "source" to refer to the solute plus the permanent multipoles of the solvent. In most situations, AMOEBA is not presented as a variational minimization. The induced dipoles

are determined by polarizability α_A and the local electric field

$$d_{Ai} = \alpha_A \left(\mathcal{E}_{Ai}^{\text{(source-D)}} + \sum_{A \neq B, j} T_{Ai,Bj}^{\text{(Thole)}} d_{Bj} \right) \tag{4}$$

$\vec{\mathcal{E}}^{\text{(source-D)}}$ represents the electric field generated by the source, and $\mathbf{T}^{\text{(Thole)}}$ is the Thole-damped electrostatic tensor for dipole–dipole interactions.[34] Once the induced dipoles d_{Ai} are solved from Eq.4 , the total energy can then be computed as follows,

$$E^{\text{(AMOEBA)}} = E^{\text{(source)}} - \frac{1}{2} \sum_{Ai} d_{Ai} \mathcal{E}_{Ai}^{\text{(source-P)}} \tag{5}$$

The specific forms of $E^{\text{(source)}}$ and $\vec{\mathcal{E}}^{\text{(source)}}$ in the above equations depend on the level of theory used for describing the solute. Note that there are two types of electric field generated by the source: $\vec{\mathcal{E}}^{\text{(source-D)}}$ in Eq.4 that is used to solve the induced dipoles, and $\vec{\mathcal{E}}^{\text{(source-P)}}$ in Eq.5 that is used to compute the energy. These two electric fields are not identical when intramolecular polarization needs to be included, because AMOEBA was designed to use different intramolecular scaling rules when determining the induced dipoles versus computing the energy.[43] On the other hand, AMOEBA implementations in both Tinker[71] and OpenMM[72] software provide the option of neglecting the intramolecular polarization for two atoms in the same molecule separated by up to 4 covalent bonds, leading to the two electric fields being identical for sufficiently small solvent molecules such as water and chloroform.

When intramolecular polarization is neglected, we have $\mathcal{E}_{Ai}^{\text{(source-D)}} = \mathcal{E}_{Ai}^{\text{(source-P)}} = \mathcal{E}_{Ai}^{\text{(source)}}$. It is then possible to formulate AMOEBA as a variational problem, as pointed out in previous works.[49] When this condition is satisfied, Eq.4 can be rearranged as the following linear system

$$\sum_{Bj} O_{Ai,Bj} d_{Bj} = \mathcal{E}_{Ai}^{\text{(source)}} \tag{6}$$

$$O_{Ai,Bj} = \frac{1}{\alpha_A} \delta_{AB} \delta_{ij} - T_{Ai,Bj}^{\text{(Thole)}} \tag{7}$$

Using Eq.6, we can then rewrite the total energy in Eq.5 as

$$E^{(\text{AMOEBA})} = E^{(\text{source})} - \sum_{Ai} \mathcal{E}_{Ai}^{(\text{source})} d_{Ai} + \frac{1}{2} \sum_{Ai,Bj} d_{Ai} O_{Ai,Bj} d_{Bj} \tag{8}$$

One can then show that the induced dipoles solved from Eq.6 satisfies the variational condition

$$\frac{\partial E^{(\text{AMOEBA})}}{\partial d_{Ai}} = -\mathcal{E}_{Ai}^{(\text{source})} + \sum_{Bj} O_{Ai,Bj} d_{Bj} = 0 \tag{9}$$

The methods in this work are restricted to the special case of neglecting intramolecular polarization as a first step, and inclusion of these effects will be developed in future works as discussed in the conclusion section.

The similarities between PCM and AMOEBA with the condition of $\mathcal{E}_{Ai}^{(\text{source-D})} = \mathcal{E}_{Ai}^{(\text{source-P})}$ are readily apparent when comparing Eq.1 with Eq.8, and comparing Eq.3 with Eq.6. For simplicity, we will only say "AMOEBA" in all the following discussions, but readers should keep in mind that the condition of $\mathcal{E}_{Ai}^{(\text{source-D})} = \mathcal{E}_{Ai}^{(\text{source-P})}$ is assumed throughout the paper.

B. Energy

These similarities discussed in Section II A suggest the possibility of generalizing most of the ideas developed previously for SA-CASSCF in PCM[65] to AMOEBA solvent models for small molecules, most importantly in water. In particular, our previous work used an atomic orbital formulation, which only relies on a few basic operations that are accelerated on graphical processing units (GPUs). Therefore, as we discuss the generalization to AMOEBA, we will identify what basic operations are needed correspondingly.

We start with a brief review of SA-CASSCF/PCM developed in our previous work. The solute is described by the nuclear charges $\mathbf{Z}$ and the electronic densities (including the one-electron density matrix $\mathbf{P}$, and the two-electron density matrix $\mathbf{\Pi}$). Based on Eq.1, the state-averaged energy in PCM is defined as

$$E^{(\text{SA-PCM})} = E^{(\text{SA-solute})} + \sum_{n} U_{n}^{(\text{SA-solute})} q_n + \frac{1}{2} \sum_{nm} q_n A_{nm} q_m \tag{10}$$

In Eq.10, $E^{(\text{SA-solute})}$ is the state-averaged energy from the solute and can be written as

$$E^{(\text{SA-solute})} = \sum_{pq} P_{pq}^{(\text{SA})} K_{pq} + \sum_{pqrs} \Pi_{pqrs}^{(\text{SA})}(pq|rs) + \sum_{pq} P_{pq}^{(\text{SA})} V_{pq}^{(0)}[\mathbf{Z}] + \epsilon_{\text{es}}[\mathbf{Z}] \qquad (11)$$

In the first term of Eq.11, $\mathbf{K}$ is the kinetic integral, and $\mathbf{P}^{(\text{SA})}$ is the state-averaged one-electron density matrix

$$P^{\text{SA}} = \sum_{I} w_I P_I \qquad (12)$$

The dynamic weight scheme[63] can be used such that the weights are adjusted based on the energy gaps from the chosen state of interests (i.e. the "reference" state):

$$w_I = \tilde{\omega}_I(E) = \frac{f(E_I - E_{\text{Ref}})}{\sum_K f(E_K - E_{\text{Ref}})} \qquad (13)$$

The function f can be either a cubic spline function or a secant hyperbolic function. The weight of a state is close to zero if its energy difference from the reference state exceeds a pre-selected bandwidth, and has equal weight as the reference state if the energy difference decreases to zero. The state averaged two-electron density matrix $\Pi^{(\text{SA})}$ in the second term of Eq.11 is defined in a similar way. The third term in Eq.11 represents the electron-nuclear attraction, where the superscript "(0)" denotes interaction with point charges

$$V_{pq}^{(0)}[\mathbf{Z}] = \sum_{N} v_{pq,N}^{(0)} Z_N = -\sum_{N} \int \phi_p(\mathbf{r})\phi_q(\mathbf{r}) \frac{1}{|\mathbf{r}-\mathbf{R}_N|} d\mathbf{r} Z_N \qquad (14)$$

The last term $\epsilon_{\text{es}}[\mathbf{Z}]$ in Eq.11 represents the electrostatic interaction among nuclear charges $\mathbf{Z}$. The state-averaged electric potential $U^{(\text{SA-solute})}$ in Eq.10 contains contributions from the electronic density $\mathbf{P}^{(\text{SA})}$ and nuclear charge $\mathbf{Z}$ respectively as

$$U_n^{(\text{SA-solute})} = \sum_{pq} P_{pq}^{(\text{SA})} \beta_{pq,n} + \sum_{N} Z_N B_{N,n} \qquad (15)$$

The definitions of $\beta_{pq,n}$ and $B_{N,n}$ depend on which PCM model is being used. In our previous and this work, we use the conductor-like polarizable continuum model with switching Gaussian smooth discretization,[73,74] which defines $\beta_{pq,n}$ and $B_{N,n}$ based on damped Coulomb interactions.

10

We now consider SA-CASSCF/AMOEBA. In addition to the solute nuclei and electrons, the source also contains the solvent permanent multipoles including charges $\mathbf{C}$, dipoles $\mathbf{M}$, and quadrupoles $\mathbf{Q}$. The QM nuclear charges ($\mathbf{Z}$) and the MM permanent multipoles ($\mathbf{C}$, $\mathbf{M}$ and $\mathbf{Q}$) will be combined and referred to as "fixed multipoles". Based on Eq.8, the state-averaged energy in AMOEBA is

$$E^{(\text{SA-AMOEBA})} = E^{(\text{SA-source})} - \sum_{Ai} \mathcal{E}_{Ai}^{(\text{SA-source})} d_{Ai} + \frac{1}{2} \sum_{Ai,Bj} d_{Ai} O_{Ai,Bj} d_{Bj} \tag{16}$$

$E^{(\text{SA-source})}$ is the state averaged source energy, and is defined as

$$\begin{aligned} E^{(\text{SA-source})} &= \sum_{pq} P_{pq}^{(\text{SA})} K_{pq} + \sum_{pqrs} \Pi_{pqrs}^{(\text{SA})} (pq|rs) \\ &+ \sum_{pq} P_{pq}^{\text{SA}} V_{pq}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}] + \epsilon_{\text{es}}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}] + \epsilon_{\text{vdW}} + \epsilon_{\text{valence}} \end{aligned} \tag{17}$$

The first two terms in Eq.17 are identical with Eq.11 from PCM. The third term in Eq.17 represents the electrostatic interactions between the electrons and the fixed point multipoles, and we assume they interact through the full Coulomb interaction

$$V_{pq}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}] = V_{pq}^{(0)}[\mathbf{Z}] + V_{pq}^{(0)}[\mathbf{C}] + V_{pq}^{(1)}[\mathbf{M}] + V_{pq}^{(2)}[\mathbf{Q}] \tag{18}$$

$V_{pq}^{(0)}$ is defined in Eq.14. $V_{pq}^{(1)}$ denotes the interaction with the point dipoles:

$$V_{pq}^{(1)}[\mathbf{M}] = \sum_{Ai} v_{pq,Ai}^{(1)} \cdot M_{Ai} = -\sum_{Ai} \int \phi_p(\mathbf{r})\phi_q(\mathbf{r}) \nabla_{R_{Ai}} \frac{1}{|\mathbf{r}-\mathbf{R}_A|} d\mathbf{r} \cdot M_{Ai} \tag{19}$$

and $V_{pq}^{(2)}$ denotes the interaction with the point quadrupoles:

$$V_{pq}^{(2)}[\mathbf{Q}] = \sum_{Aij} v_{pq,Aij}^{(2)} \cdot Q_{Aij} = -\sum_{Aij} \int \phi_p(\mathbf{r})\phi_q(\mathbf{r}) \nabla_{R_{Ai}} \nabla_{R_{Aj}} \frac{1}{|\mathbf{r}-\mathbf{R}_A|} d\mathbf{r} \cdot Q_{Aij} \tag{20}$$

The fourth term ϵ_{es} in Eq.17 represents the electrostatic interaction among all the fixed point multipoles. The valence interaction energy $\epsilon_{\text{valence}}$ and van der Waals interaction energy ϵ_{vdW} from the force field have also been included. The state averaged electric field $\vec{\mathcal{E}}^{(\text{SA-source})}$ in

Eq.16 contains the electric field generated by the electrons and the fixed multipoles as:

$$\mathcal{E}_{Ai}^{(\text{SA-source})} = \mathcal{E}_{Ai}^{(\text{elec})}[\mathbf{P}^{(\text{SA})}] + \mathcal{E}_{Ai}^{(\text{fix})}[\mathbf{Z}, \mathbf{C}, \mathbf{D}, \mathbf{Q}] \tag{21}$$

In this work, because the QM region and the MM region are not covalently bonded, we assume the electrons interact with the induced dipoles through the full Coulomb interaction as used in Mao et.al.[49] Therefore, using the integral $v_{pq,Ai}^{(1)}$ defined in Eq.19, the electric field from the electrons can be computed as

$$-\mathcal{E}_{Ai}^{(\text{elec})}[\mathbf{P}^{(\text{SA})}] = \sum_{pq} P_{pq}^{(\text{SA})} v_{pq,Ai}^{(1)} \tag{22}$$

By comparing Eq.17 and Eq.15 from SA-CASSCF/PCM with Eq.17 and Eq.21 from SA-CASSCF/AMOEBA, it is apparent that there is a one-to-one correspondence between every term in SA-CASSCF/PCM with SA-CASSCF/AMOEBA, as summarized in Table I. By introducing p_μ as the abstract induced variables, where $p_{\mu=n} = q_n$ for PCM, and $p_{\mu=3A+i} = d_{Ai}$ for AMOEBA, we can unify Eq.10 from PCM and Eq.16 from AMOEBA into the following equation, representing the state-averaged energy in a polarizable environment:

$$
\begin{aligned}
E^{(\text{SA-Polarize})} = &\sum_{pq} P_{pq}^{(\text{SA})} \left(K_{pq} + V_{pq}^{(\text{fix})} \right) + \sum_{pqrs} \Pi_{pqrs}^{(\text{SA})}(pq|rs) \\
&+ \sum_\mu \left(W_\mu^{(\text{elec})} + W_\mu^{(\text{fix})} \right) p_\mu + \tfrac{1}{2} \sum_{\mu\nu} p_\mu \Xi_{\mu\nu} p_\nu + \epsilon_{\text{non-polarize}}
\end{aligned}
\tag{23}
$$

The definition for each symbol in Eq.23 is summarized in Table I. Given Eq.23, the molecular orbital (MO) coefficients, configuration interaction (CI) coefficients and the induced variables can then be solved self-consistently such that the state-averaged energy is variationally minimized:

$$\frac{\partial E^{(\text{SA-Polarize})}}{\partial \kappa_{rs}} = 0, \frac{\partial E^{(\text{SA-Polarize})}}{\partial \theta_{RS}} = 0, \frac{\partial E^{(\text{SA-Polarize})}}{\partial p_\mu} = 0 \tag{24}$$

where κ_{rs} and θ_{RS} represent the rotations within the MO and CI space respectively. In the previous work on SA-CASSCF/PCM, we developed a first-order convergence algorithm, which goes back and forth between diagonalizing the Hamiltonian to update the CI coefficients, and using Newton-Raphson to update the MO coefficients and the induced variables

12

simultaneously. This algorithm requires two quantities: one quantity is the gradient with respect to the orbital rotations

$$g_{rs}^{(O;\text{SA-Polarize})} = \frac{\partial E^{(\text{SA-Polarize})}}{\partial \kappa_{rs}} = \frac{\partial}{\partial \kappa_{rs}} \left(\sum_{pq} P_{pq}^{(\text{SA})} \eta_{pq} + \sum_{pqrs} \Pi_{pqrs}^{(\text{SA})}(pq|rs) \right) \tag{25}$$

where the one electron integral η_{pq} in Eq.25 is the sum of three components:

$$\eta_{pq} = K_{pq} + V_{pq}^{(\text{fix})} + V_{pq}^{(\text{induce})} \tag{26}$$

The other quantity is the gradient with respect to the induced variables

$$g_{\mu}^{(Q;\text{SA-Polarize})} = \frac{\partial E^{(\text{SA-Polarize})}}{\partial p_{\mu}} = W_{\mu}^{(\text{elec})} + W_{\mu}^{(\text{fix})} + \sum_{\nu} \Xi_{\mu\nu} p_{\nu} \tag{27}$$

By substituting the variables in Eq.26 and Eq.27 based on Table I, the same convergence algorithm can be used for both SA-CASSCF/PCM as well as SA-CASSCF/AMOEBA. This algorithm may also be applied to SA-CASSCF in other polarizable models as long as the state-averaged energy can be written in the form of Eq.23, and the symbols in Table I represent the basic operations that need to be defined and implemented for each new polarizable model.

Once the converged wavefunctions and induced variables are obtained, the state specific energy can then be computed similar to Eq.23 by replacing the state averaged density matrices ($\mathbf{P^{(SA)}}$ and $\mathbf{\Pi^{(SA)}}$) with state-specific density matrices ($\mathbf{P^{\Theta}}$ and $\mathbf{\Pi^{\Theta}}$).

C. Analytical gradients and non-adiabatic couplings

In Section II B, we have shown how the energies of SA-CASSCF in PCM and AMOEBA can be computed using similar algorithms. It is natural to expect their analytical gradients and non-adiabatic couplings are also closely related.

Because the MO coefficients, CI coefficients and induced variables are determined by variationally minimizing the state-averaged energy, the state-specific energies are no longer variational with respect to these variables. Therefore, the Lagrangian formulation[66,67] is required for evaluating the corresponding analytical gradients. Based on the state-averaged

13

energy in Eq.23, the variational conditions in Eq.24 and the definition of dynamic weights in Eq.13, the Lagrangian corresponding to the target state Θ is

$$L_\Theta^{(\text{SA-Polarize})} = E_\Theta^{(\text{Polarize})} + \sum_{rs} \bar{\kappa}_{rs}^\Theta \frac{\partial E^{(\text{SA-Polarize})}}{\partial \kappa_{rs}} + \sum_{RS} \bar{\theta}_{RS}^\Theta \frac{\partial E^{(\text{SA-Polarize})}}{\partial \theta_{RS}}$$
$$+ \sum_\mu \bar{p}_\mu^\Theta \frac{\partial E^{(\text{SA-Polarize})}}{\partial p_\mu} + \sum_I \bar{w}_I^\Theta \tau_I \tag{28}$$

Eq.28 introduces four sets of Lagrange multipliers. The first two sets of Lagrange multipliers $\bar{\boldsymbol{\kappa}}^\Theta$ and $\bar{\boldsymbol{\theta}}^\Theta$ impose constraints on the orbital rotations and CI rotations respectively. The third set of Lagrange multipliers $\bar{\boldsymbol{p}}^\Theta$ imposes constraints on the induced variables, corresponding to $\bar{p}_{\mu=n}^\Theta = \bar{q}_n$ for PCM and $\bar{p}_{\mu=3A+i}^\Theta = \bar{d}_{Ai}^\Theta$ for AMOEBA. The last set of Lagrange multipliers $\bar{\mathbf{w}}^\Theta$ constrains the values of dynamic weights, where $\tau_I = w_I - \tilde{\omega}_I(E)$ from Eq.13. The Lagrange multipliers are solved by imposing the following invariant conditions on the Lagrangian,

$$\frac{\partial L_\Theta^{(\text{SA-Polarize})}}{\partial \kappa_{rs}} = 0, \frac{\partial L_\Theta^{(\text{SA-Polarize})}}{\partial \theta_{RS}} = 0, \frac{\partial L_\Theta^{(\text{SA-Polarize})}}{\partial p_\mu} = 0, \frac{\partial L_\Theta^{(\text{SA-Polarize})}}{\partial w_I} = 0 \tag{29}$$

This leads to the coupled-perturbed equation that can be divided into four by four blocks, where each block corresponds to one set of the Lagrange multipliers

$$\begin{pmatrix} \mathbb{H}^{(OO)} & \mathbb{H}^{(OC)} & \mathbb{W}^{(OW)} & \mathbb{H}^{(OQ)} \\ \mathbb{H}^{(CO)} & \mathbb{H}^{(CC)} & \mathbf{0} & \mathbb{H}^{(CQ)} \\ \mathbb{H}^{(WO)} & \mathbf{0} & \mathbf{I} & \mathbb{H}^{(WQ)} \\ \mathbb{H}^{(QO)} & \mathbb{H}^{(QC)} & \mathbb{W}^{(QW)} & \mathbb{H}^{(QQ)} \end{pmatrix} \begin{pmatrix} \bar{\boldsymbol{\kappa}}^\Theta \\ \bar{\boldsymbol{\theta}}^\Theta \\ \bar{\mathbf{p}}^\Theta \\ \bar{\mathbf{w}}^\Theta \end{pmatrix} = \begin{pmatrix} -\mathbf{g}^{(O;\text{Polarize}),\Theta} \\ \mathbf{0} \\ \mathbf{0} \\ -\mathbf{g}^{(Q;\text{Polarize}),\Theta} \end{pmatrix} \tag{30}$$

In our previous works on SA-CASSCF/PCM,[65] we have derived the expressions for each of the terms in Eq.30. Moreover, we showed that the coupled-perturbed equation can be solved iteratively using operations that are readily available from the energy calculations. Therefore, by making the substitutions of variables following Table I, the Lagrange multipliers for SA-CASSCF/AMOEBA can be solved using the same algorithm as SA-CASSCF/AMOEBA .

Once the Lagrange multipliers are solved, they are used to construct the effective density matrices $\boldsymbol{\gamma}^\Theta$, $\boldsymbol{\Gamma}^\Theta$ and $\tilde{\mathbf{X}}^\Theta$ as derived in the previous work on SA-CASSCF/PCM. The analytical nuclear gradients for state energy then require evaluating the nuclear gradients for

each type of interactions existing in the system

$$\frac{dE_\Theta^{(\text{Polarize})}}{d\xi} = \sum_{pq} \gamma_{pq}^\Theta \frac{\partial K_{pq}}{\partial \xi} + \sum_{pqrs} \Gamma_{pqrs}^\Theta \frac{\partial (pq|rs)}{\partial \xi} - \sum_{pq} \tilde{X}_{pq}^\Theta \frac{\partial S_{pq}}{\partial \xi} + V^{(\text{fix}),\xi}[\boldsymbol{\gamma}^\Theta]$$
$$+ W^{(\text{elec}),\xi}[\boldsymbol{\gamma}^\Theta, \mathbf{p}] + W^{(\text{elec}),\xi}\left[\mathbf{P}^{(\text{SA})}, \bar{\mathbf{p}}^\Theta\right] + W^{(\text{fix}),\xi}[\mathbf{p} + \bar{\mathbf{p}}^\Theta] + \Xi^\xi\left[\mathbf{p}, \tfrac{1}{2}\mathbf{p} + \bar{\mathbf{p}}^\Theta\right] + \frac{d\epsilon_{\text{non-polarize}}}{d\xi} \tag{31}$$

The first three terms in Eq.31 correspond to nuclear gradients from kinetic integrals, two-electron repulsion integrals (ERI) and overlap integrals respectively. The remaining terms in Eq.31 are defined in Table I based on whether PCM or AMOEBA is used. Below, we will discuss the definitions of these terms in AMOEBA.

In the case of AMOEBA, $\frac{d}{d\xi}\epsilon_{\text{non-polarize}}$ is the nuclear gradient contributions from all non-polarizable interactions, including the valence interactions, the van der Waals interactions, and the electrostatic interactions among the fixed multipoles. $\Xi^\xi[\mathbf{d}^{(1)}, \mathbf{d}^{(2)}]$ represents the nuclear gradients from the mutual interactions between two sets of induced dipoles with values $\mathbf{d}^{(1)}$ and $\mathbf{d}^{(2)}$, i.e.

$$\Xi^\xi[\mathbf{d}^{(1)}, \mathbf{d}^{(2)}] = \sum_{AiBj} d_{Ai}^{(1)} \frac{\partial O_{Ai,Bj}}{\partial \xi} d_{Bj}^{(2)} \tag{32}$$

$W^{(\text{fix}),\xi}[\mathbf{d}]$ represents the nuclear gradients from the interaction between fixed multipoles and a set of induced dipoles $\mathbf{d}$

$$W^{(\text{fix}),\xi}[\mathbf{d}] = -\sum_{Ai} \frac{\partial \mathcal{E}_{Ai}^{(\text{fix})}[\mathbf{Z}, \mathbf{C}, \mathbf{M}, \mathbf{Q}]}{\partial \xi} d_{Ai} \tag{33}$$

The above three terms $W^{(\text{fix}),\xi}$, Ξ^ξ and $\frac{d}{d\xi}\epsilon_{\text{non-polarize}}$ all arise from purely classical interactions, and the corresponding gradients will be computed by the molecular dynamics codes with details described in the Appendix.

The other two terms $V^{(\text{fix}),\xi}$ and $W^{(\text{elec}),\xi}$ both involve interactions with the electrons. Given a one-electron density matrix $\mathbf{P}$ and a set of induced dipoles $\mathbf{d}$, $W^{(\text{elec}),\xi}[\mathbf{P},\mathbf{d}]$ computes the nuclear gradients from their electrostatic interactions

$$W^{(\text{elec}),\xi}[\mathbf{P},\mathbf{d}] = \sum_{pq,Ai} P_{pq} \frac{\partial v_{pq,Ai}^{(1)}}{\partial \xi} d_{Ai} \tag{34}$$

where the integral $v_{pq,Ai}^{(1)}$ is defined in Eq.19. Similarly, $V^{(\text{fix}),\xi}[\mathbf{P}]$ represents the gradients

15

from the electrostatic interaction between the electrons and the fixed multipoles. Based on Eq.18, it can be further decomposed into four contributions corresponding to four groups of fixed multipoles

$$V^{(\text{fix}),\xi}[\mathbf{P}] = V^{(0),\xi}[\mathbf{P},\mathbf{Z}] + V^{(0),\xi}[\mathbf{P},\mathbf{C}] + V^{(1),\xi}[\mathbf{P},\mathbf{M}] + V^{(2),\xi}[\mathbf{P},\mathbf{Q}] \tag{35}$$

$V^{(0),\xi}$ computes the gradients from interactions between electrons and point charges, and are readily available in most electronic structure software

$$V^{(0),\xi}[\mathbf{P},\mathbf{Z}] + V^{(0),\xi}[\mathbf{P},\mathbf{C}] = \sum_{pqN} P_{pq} \frac{\partial v_{pq,N}^{(0)}}{\partial \xi} Z_N + \sum_{pqA} P_{pq} \frac{\partial v_{pq,A}^{(0)}}{\partial \xi} C_A \tag{36}$$

The gradients from interactions between electrons with permanent point dipoles $V^{(1),\xi}$ or permanent point quadrupoles $V^{(2),\xi}$ are more complicated. Because the permanent dipoles and quadrupoles are attached to local frames of MM molecules, there are additional contributions to the gradients due to the rotations of the local frames. There are two ways of treating this. One way is to follow the original implementation of AMOEBA, which first computes the torque exerted on the point dipoles and quadrupoles, and then converts the torques to gradients on the atoms which the local frames are attached to.[40] The second way is to explicitly take derivatives of the local frame rotations.[75] In this work, we follow the first method and use the torques as intermediates. As a result, the gradients from the electron-point dipole interactions can be decomposed as

$$V^{(1),\xi}[\mathbf{P},\mathbf{M}] = V_{\text{no-rotate}}^{(1),\xi}[\mathbf{P},\mathbf{M}] + \text{TorqueToGrad}\left(\tau^{(1)}[\mathbf{P},\mathbf{M}]\right) \tag{37}$$

$$V_{\text{no-rotate}}^{(1),\xi}[\mathbf{P},\mathbf{M}] = \sum_{pq,Ai} P_{pq} \frac{\partial v_{pq,Ai}^{(1)}}{\partial \xi} M_{Ai} \tag{38}$$

The torque exerted on a point dipole $\vec{M}$ depends on the local electric field $\vec{\mathcal{E}}$ as $\vec{\tau} = \vec{M} \times \vec{\mathcal{E}}$. By using Levi-Civita symbol, it can also be written as $\tau_i = \sum_{jk} \epsilon_{ijk} M_j \mathcal{E}_k$. Thus the torque in Eq.37 is computed as

$$\tau_{Ai}^{(1)} = -\sum_{jk} \epsilon_{ijk} M_j \left(\sum_{pq} P_{pq} v_{pq,Ak}^{(1)} \right) \tag{39}$$

16

Similarly, the gradients from the electron-point quadrupole interactions can also be decomposed into two parts:

$$V^{(2),\xi}[\mathbf{P},\mathbf{Q}] = V^{(2),\xi}_{\text{no-rotate}}[\mathbf{P},\mathbf{Q}] + \text{TorqueToGrad}\big(\tau^{(2),\xi}[\mathbf{P},\mathbf{Q}]\big) \tag{40}$$

$$V^{(2),\xi}_{\text{no-rotate}}[\mathbf{P},\mathbf{Q}] = \sum_{pq,Aij} P_{pq}\frac{\partial v^{(2)}_{pq,Aij}}{\partial \xi}Q_{Aij} \tag{41}$$

The torque exerted on a point quadrupole $\overleftrightarrow{Q}$ depends on the local electric field gradient $\overleftrightarrow{\mathcal{G}}$, and using the Levi-Civita symbol it can be written as $\vec{\tau} = \sum_{jk}\epsilon_{ijk}\left(\sum_l Q_{jl}\mathcal{G}_{lk}\right)$. Thus, the torque in Eq.40 can be computed as

$$\tau^{(2)}_{Ai}[\mathbf{P},\mathbf{Q}] = -\sum_{jk}\epsilon_{ijk}\left(\sum_l Q_{Ajl}\left(\sum_{pq}P_{pq}v^{(2)}_{pq,Alk}\right)\right) \tag{42}$$

The torques are then converted to nuclear gradients on the MM atoms by the molecular dynamics codes, as discussed in Section II D.

The evaluations of the non-adiabatic couplings are very similar to the analytical nuclear gradients. When solving the non-adiabatic couplings between state Θ and Π, the Lagrangian is defined by replacing $E_\Theta^{(\text{Polarize})}$ in Eq.28 with $\Omega_{\Theta\Pi}^{(\text{Polarize})}$

$$\Omega_{\Theta\Pi}^{(\text{Polarize})} = \sum_{pq}T_{pq}^{\Theta\Pi}\eta_{pq} + \sum_{pqrs}\Pi_{pqrs}^{\Theta\Pi}(pq|rs) \tag{43}$$

where $\mathbf{T}^{\Theta\Pi}$ and $\mathbf{\Pi}^{\Theta\Pi}$ are transition density matrices between the two states, and η_{pq} is defined in Eq.26. The corresponding Lagrange multipliers $\bar{\boldsymbol{\kappa}}^{\Theta\Pi}$, $\bar{\boldsymbol{\theta}}^{\Theta\Pi}$, $\bar{\mathbf{p}}^{\Theta\Pi}$, $\bar{\mathbf{w}}^{\Theta\Pi}$ can be solved from the coupled perturbed equation in Eq.30. By using the Lagrange multipliers to construct the effective density matrices $\boldsymbol{\gamma}^{\Theta\Pi}$, $\boldsymbol{\Gamma}^{\Theta\Pi}$ and $\tilde{\mathbf{X}}^{\Theta\Pi}$, the non-adiabatic couplings $\chi_\xi^{\Theta\Pi}$ can then be computed as

$$\chi_\xi^{\Theta\Pi} = \delta_\xi^{\Theta\Pi} - \frac{1}{\Delta^{\Theta\Pi}}\left\{\sum_{pq}\gamma_{pq}^{\Theta\Pi}\frac{\partial K_{pq}}{\partial \xi} + \sum_{pqrs}\Gamma_{pqrs}^{\Theta\Pi}\frac{\partial(pq|rs)}{\partial \xi} - \sum_{pq}\tilde{X}_{pq}^{\Theta\Pi}\frac{\partial S_{pq}}{\partial \xi}\right.$$
$$\left. +V^{(\text{fix}),\xi}\left[\boldsymbol{\gamma}^{\Theta\Pi}\right] + W^{(\text{elec}),\xi}\left[\boldsymbol{\gamma}^{\Theta\Pi},\mathbf{p}\right] + W^{(\text{elec}),\xi}\left[\mathbf{P}^{(\text{SA})},\bar{\mathbf{p}}^{\Theta\Pi}\right] + W^{(\text{fix}),\xi}\left[\bar{\mathbf{p}}^{\Theta\Pi}\right] + \Xi^\xi\left[\mathbf{p},\bar{\mathbf{p}}^{\Theta\Pi}\right]\right\} \tag{44}$$

In Eq.44, $\Delta^{\Theta\Pi} = E_\Theta^{(\text{Polarize})} - E_\Pi^{(\text{Polarize})}$ is the energy gap, and $\delta_\xi^{\Theta\Pi}$ is computed from the

derivative of overlap integrals

$$\delta_\xi^{\Theta\Pi} = \frac{1}{2} \sum_{rs} \left(T_{rs}^{\Theta\Pi} - T_{sr}^{\Theta\Pi} \right) \cdot \left(\left\langle \phi_r | \frac{d}{d\xi} \phi_s \right\rangle - \left\langle \frac{d}{d\xi} \phi_r | \phi_s \right\rangle \right) \tag{45}$$

The other operations appearing in Eq.44 are identical to the ones in analytical gradients.

Similar to the discussions on energy in Section II B, the above discussions on gradients and non-adiabatic couplings may also be applied to SA-CASSCF in other polarizable models that can be written in the form of Eq.23. The symbols in Table I summarize the quantities and operations that need to be implemented and substituted for each polarizable model respectively.

D. Implementation

Section II B and section II C show that by making the substitutions of variables following Table I, the algorithms developed for SA-CASSCF/PCM can be generalized to SA-CASSCF/AMOEBA. However, compared to PCM that can be implemented completely within the quantum chemistry codes, SA-CASSCF/AMOEBA requires interfacing the QM codes with the MM codes, thus introducing more complications.

The general design principle in this work is to let the MM codes store and compute terms that are purely classical. For example, we let the MM codes handle the solute nuclear charges $\mathbf{Z}$ even though they reside within the QM region. The purpose of such design is to avoid duplicate implementations of classical interactions that are already well-optimized in the MM codes, such as the electrostatic interactions between point nuclear charges and point dipoles or quadrupoles. Another purpose is to keep any force field specific information (such as local frames) within the MM codes, such that the amount of information to be passed between the QM and MM codes is minimized.

Because Table I has summarized all the differences between SA-CASSCF/PCM and SA-CASSCF/AMOEBA, it provides guidance about what new operations need to be implemented when AMOEBA is used. There are 7 operations in Table I involving electrons, and these are implemented in the quantum chemistry codes. All of these operations involve evaluating or differentiating integrals following the McMurchie-Davidson algorithm,[76] and are

computed on the graphical processing units (GPUs) using codes generated by the automated code engine.[77] These include:

- Computing the electric field generated by the electrons (Eq. 22);

- Computing the Fock matrix from a set of point dipoles (Eq.19) or point quadrupoles (Eq.20);

- Computing the gradients and torques from electrons interacting with point dipoles (Eq.38 and Eq.39) or point quadrupoles (Eq.41 and Eq.42).

The remaining terms in Table I other than the 7 operations listed above are all handled by the MM codes. In this work, we use the AMOEBA plugin[39] in the OpenMM software. Some of the information to be exchanged between QM codes and MM codes (e.g. getting the coordinates of the system, resetting the permanent multipole values of the QM atoms) can already be performed using the existing OpenMM interfaces (APIs). For information that are not directly available using the current interfaces, we added a few new APIs to the OpenMM-AMOEBA plugins and the codes are publicly available on GitHub (see descriptions in SI-Section 4). These new APIs include:

- Get all the permanent multipoles $(\mathbf{C}, \mathbf{M}, \mathbf{Q})$ in the lab frame. This information is needed by multiple terms, such as Eq.18 and 34.

- Get the electric field $\vec{\mathcal{E}}^{(\text{fix})}$ generated by the fixed multipoles. (used in Eq.21 and Eq.27).

- Get the mutual interaction matrix $O_{Ai,Bj}$ in Eq.7.

- Given an array representing torques exerted on the MM atoms, convert the torques to gradients. (used in Eq.37 and Eq.40).

- Given two sets of arrays $\boldsymbol{\mu}^{(\text{D})}$ and $\boldsymbol{\mu}^{(\text{P})}$ both of the same dimension as induced dipoles, set the induced dipoles in OpenMM according to $\boldsymbol{\mu}^{(\text{D})}$ and $\boldsymbol{\mu}^{(\text{P})}$, and compute the corresponding energy $\tilde{E}^{(\text{ES})}[\boldsymbol{\mu}^{(\text{D})}, \boldsymbol{\mu}^{(\text{P})}]$ or gradients $\tilde{G}^{(\text{ES}),\xi}[\boldsymbol{\mu}^{(\text{D})}, \boldsymbol{\mu}^{(\text{P})}]$.

In the Appendix, we describe how through different combinations of $\boldsymbol{\mu}^{(\text{D})}$ and $\boldsymbol{\mu}^{(\text{P})}$, $\tilde{E}^{(\text{ES})}[\boldsymbol{\mu}^{(\text{D})}, \boldsymbol{\mu}^{(\text{P})}]$ and $\tilde{G}^{(\text{ES}),\xi}[\boldsymbol{\mu}^{(\text{D})}, \boldsymbol{\mu}^{(\text{P})}]$ can be used to evaluate the pure classical contributions, including $\epsilon_{\text{non-polarize}}$, $\frac{d\epsilon_{\text{non-polarize}}}{d\xi}$, $\Xi^{\xi}[\mathbf{d}^{(1)}, \mathbf{d}^{(2)}]$ (Eq.32), and $W^{(\text{fix}),\xi}[\mathbf{d}]$ (Eq.33). All

of the new APIs are simply wrappers to let the QM codes access intermediate variables or internal functions in OpenMM that were previously not publicly accessible. With minimal modifications to the OpenMM codes, this strategy helps to ensure that the AMOEBA force field is handled accurately.

The overall algorithms for computing energy, nuclear gradients and nonadiabatic couplings for SA-CASSCF/AMOEBA are summarized in Table II. The Table highlights which operations are carried out inside the QM codes, and which operations are performed in the MM codes by calling the OpenMM APIs.

III. RESULTS AND DISCUSSION

The method described in this work is implemented in the Terachem[78,79] quantum chemistry package (currently supporting atom-centered *spd* Gaussian basis sets) and is interfaced to a modified version of the OpenMM[72] software version 7.4 (using the "Reference" platform) for computing classical terms related to the AMOEBA force field. Geometry optimizations and minimal energy conical intersection (MECI) searches are calculated with the geomeTRIC[80] software. Nonadiabatic molecular dynamics using ab initio multiple spawning (AIMS) is performed with the FMS software.[70] All calculations are performed on computing nodes with NVIDIA GeForce GTX 1080Ti GPUs and Intel Xeon E5-2637 CPUs. Figures and tables in the supporting information (SI) will be referred to as SI-Figures and SI-Tables.

Unlike calculations in PCM where the solvation cavity is automatically generated based on the solute geometry, setting up explicit-solvent QM/MM calculations requires more careful treatments. In SI-Section 2, we provide a detailed description of the procedure we used to generate the configurations of the solvated system throughout this work. Given a solute structure and a choice of solvent with existing AMOEBA parameters, the procedure described in SI-Section 2 can generally be divided into three parts: (1) Parametrize the AMOEBA force field for the solute following established workflows using Tinker.[81] The force field parameters for the solute are mainly used in classical molecular dynamics simulations to sample solvent configurations in presence of the solute. (2) Create a periodic solvent box with the solute at the center. Apply restraint to the positions of the solute atoms, and equilibrate the solvent configurations by first running minimization, then running molecular

dynamics for 200 ps in the NPT ensemble at 298K temperature and 1 atm pressure. These simulations are carried out using Tinker[71]. (3) Take samples from the equilibrated trajectory, and convert the AMOEBA force field parameters and coordinates to OpenMM format (with periodic boundary conditions removed). Save the OpenMM State and System objects to XML files, which serve as the input files to SA-CASSCF/AMOEBA calculations. To make sure we have set up the simulations in OpenMM correctly, each energy component computed by OpenMM is compared with Tinker to ensure that they are in good agreement, i.e. energy from the electrostatic component computed by OpenMM or by Tinker should differ at most on the order of 1.0e-4 kcal/mol, and all other energy components should differ on the order of 1.0e-9 kcal/mol.

A. Accuracy and performance

We start with testing the accuracy of the new method by comparing the analytical nuclear gradients with numerical ones, and the results are presented in SI-Section 1. The calculations are performed with SA(2)-CASSCF(4,4)/AMOEBA/6-31G* on the model retinal protonated Schiff base molecule (PSB3, $C_4H_9N^+$) solvated in more than 400 water molecules. Energy calculations are deemed converged when two conditions are met: (1) changes in state averaged energy from the previous iteration is smaller than 1.0e-8 Hartree, and (2) the derivative elements with respect to each orbital rotation (i.e. Eq.25) and each induced variable (i.e. Eq.27) are all below 1.0e-8 Hartree. In analytical gradient calculations, the convergence threshold for solving the coupled-perturbed equation is set to 1.0e-7 Hartree. The numerical gradients are evaluated for all the atoms in the PSB3 molecule (described quantum mechanically) as well as atoms in the two water molecules (described with AMOEBA) closest to PSB3. The numerical gradients are computed with 3-point central difference using a displacement of 0.0005 Angstrom. In SI-Table S1 and S2, we compute nuclear gradients for ground and excited state respectively with polarization from the induced dipoles turned on, and the root mean square in the differences between analytical and numerical results are both within 4e-7 Hartree/Bohr. For reference, in SI-Table S3 and S4, we have also carried out similar calculations at the same geometry but without polarization, i.e. electrostatic interactions with the MM region only coming from fixed multipoles, and the root mean square in the

differences between analytical and numerical results are both within 4.5e-7 Hartree/Bohr. These results suggest that the analytical gradients are implemented correctly. Polarizations are enabled in all the calculations below.

In contrast to PCM where the number of induced surface charges is proportional to the size of the solute, the number of induced dipoles in AMOEBA is proportional to the number of solvent molecules, and this will affect both the excitation energies and the computational cost. To test the effects from the size of the MM region, the test molecule we use is a recently reported dipolar merocyanine dye (merocyanine 4-(dicyanomethylene)-2-tert-butyl-6-[3-(3-butyl-benzothiazol-2-ylidene)1-propenyl]-4H-pyran,[7] abbreviated as "DCBT", see the chemical structure in Figure 2. The molecule has been shown to exhibit large positive solvatochromism both experimentally as well as in our previous SA-CASSCF/PCM studies. In Figure 1, we solvate DCBT in increasing numbers of water molecules, and show how the first excitation energies and the computational costs change accordingly.

Figure 1 has indicated the sizes of the MM region both in terms of the number of water molecules, as well as the radius of the solvent sphere surrounding the solute. The changes in the first excitation energies are shown in Figure 1(a), and each data point is computed from 10 different solvent configurations sampled from molecular dynamics simulations in equilibrium. Figure 1(a) shows that the averaged excitation energies fluctuate between 3.5eV to 3.7eV when the solvent sphere radius is less than 10 Angstrom. When the solvent sphere radius exceeds 10 Angstrom, the averaged excitation energies are more consistent and fall within 3.5 eV to 3.53 eV. Based on this observation, in all calculations presented later in this work, the radius of the solvent sphere is at least 20 Angstrom. One reason why a large radius of MM region is needed for the excitation energies to converge is due to the edge effects. Because SA-CASSCF/AMOEBA calculations are performed on an isolated cluster without periodic boundary condition, the edge of the solvent sphere is in direct contact with vacuum. As a result, a large solvent sphere is required to make the environment experienced by the solute resemble a condensed phase environment. This artificial edge effect is also discussed in Section III C, where we further show how it may affect the nonadiabatic molecular dynamics and discuss possible ways to address it.

Figure 1(b) shows the corresponding computational time with increasing sizes of the MM region. The computational time for gas phase (stars) and PCM (circles) calculations have also

22

been included for comparison. The computational cost of SA-CASSCF/PCM only increases by a small amount from the gas phase calculations, which are consistent with results from our previous work.[65] When AMOEBA is used, the computational cost has strong dependency on the number of solvent molecules as expected. For this particular chromophore, it is observed that the computational cost of SA-CASSCF/AMOEBA stays close to the cost of PCM when the MM region contains fewer than 2000 atoms, but quickly increases as the MM region includes more atoms. This behavior is likely due to that the quantum mechanical operations dominate the computational cost when the MM region size is relatively small, but AMOEBA related operations start to gradually take over as the MM region size increases.

Close examination on the timing data shows that this increasing of computational cost is mainly correlated with the increasing cost of solving the induced dipoles, i.e. the linear system in the form of Eq.6. In particular, the solver to this linear system needs to be called repeatedly during the calculations, i.e. at each iteration during the energy calculations, as well as at each iteration when solving the coupled-perturbed equation in gradient or NAC calculations. As a result, the performance of the codes can be improved significantly if the solver can be made more efficient. In the current implementation, the matrix $O_{Ai,Bj}$ in Eq.6 is explicitly computed by the Reference platform of OpenMM and then stored in the memory of the QM codes, and the preconditioned conjugate gradient solver (originally implemented for PCM) is used to solve the linear system. One promising way to improve the solver is to use the GPU-accelerated platform (i.e. "CUDA") from OpenMM that solves the induced dipoles using a direct inversion in the iterative subspace (DIIS) algorithm. Numerical tests have shown that the solver from the CUDA platform is notably faster than our current solver. However, we were unable to use the CUDA platform in this work because it led to the competition for GPU resources between Terachem and OpenMM. We will explore possible ways to partition the GPU between Terachem and OpenMM in future works.

B. Electrostatic potential and excitation energies

In this section, we will investigate how the first excitation energies are affected by the types of solvent models. We will compare three different solvent models, including PCM, AMOEBA, and fixed-charge force field. We will continue to use the DCBT molecule as the

test system, since its first excitation energy is sensitive to the solvent environment.

Because the electrostatic interaction plays a fundamental role in both PCM and AMOEBA, we start by comparing the electrostatic potential (ESP) generated by the solvents when the molecule is surrounded by water. In the SA-CASSCF/PCM calculations, the dielectric constant is set to $\epsilon = 80.1$, which assumes the solvent to be in equilibrium with the solute. The AMOEBA14[82] force field parameters are used in the SA-CASSCF/AMOEBA calculations, TIP3P-FB[24] force field parameters are used in the SA-CASSCF/fixed charge force field calculations, and both models have parameters systematically improved using ForceBalance[83]. The same solvent configuration is used in the AMOEBA and fixed-charge force field calculations, and the radius of the water sphere is 24 Angstrom. As introduced by the experimental paper that synthesized DCBT, this molecule consists of a donor and an acceptor group that are covalently connected by a polymethine chain. From the chemical structure of DCBT in Figure 2, the acceptor group lies on the left hand side and the donor group lies on the right hand side. When the molecule is in its ground electronic state, in order to stabilize the solute-solvent electrostatic interaction, we expect the ESP from the solvent to be more positive on the acceptor side (left) and more negative on the acceptor side (right). As shown in the first column of Figure 2, the ESP generated by the three solvent models are all consistent with this qualitative analysis when the solute is in its ground state. However, it is also observed that notable differences occur in a few local regions when comparing PCM and the explicit solvent models (AMOEBA or fixed-charge force fields) . For example, the ESP generated by the two explicit solvent models are both locally positive around the sulfur atom, which is not observed in the PCM. By examining the corresponding structure, this positive contribution to the ESP is associated with a hydrogen bond donated by a nearby water molecule. This difference highlights the inefficiency of PCM for describing solvation effects arising from local interactions such as hydrogen bonding.

We now consider how the ESP from the solvent changes when the molecule is electronically excited. To help understand how the electronic density changes upon excitation, Figure 2(a) shows the attachment-detachment density of the first excited state of the molecule. It is observed that the detachment density is mostly on the side of the donor group, whereas the attachment density is mostly on the side of the acceptor group. This suggests that the electronic density will move from the donor to the acceptor group upon excitation. To

compensate for this changes in electronic distributions, we expect the ESP generated by the solvent changes to be even more negative on the the donor side and even more positive on the acceptor side. The changes in the solvent ESP are shown in the second column Figure 2(b), and the three solvent models are observed to behave differently. The fixed-charge model completely misses the polarization, thus is not able to describe any changes in the solvent ESP in response to the solute excitation. The two polarizable models (i.e. PCM and AMOEBA) both predict changes that are consistent with the qualitative analysis, i.e. negative changes on the donor side and positive changes on the acceptor side. However, the two polarizable models differ notably in the magnitudes of changes. Because the fast-slow separation of dielectric constant in PCM is currently not supported in SA-CASSCF due to the state-averaged formulation, the PCM calculation is carried out with $\epsilon = 80.1$, which assumes both the electrons and nuclei of the solvent to be in immediate equilibrium with the excited molecule. This is expected to overestimate the polarization effect from the solvent, since it neglects the time delay in the slow response of nuclear relaxation as discussed in the introduction. This fast-slow separation is captured in the AMOEBA calculations, as the changes in the induced dipoles reflect the fast response from the electrons, while the nuclear polarization effects remain unchanged at the same nuclear configuration. As a result, the changes in the solvent ESP from AMOEBA that only come from electrons is much weaker than PCM that has both electronic and nuclear contributions.

The differences in the solvent electrostatic potentials discussed in the previous paragraph will have a direct impact on the excitation energies, and the results are shown in Figure 3. In addition to water, we have also calculated the excitation energies in three other organic solvents, i.e. chloroform ($\epsilon = 4.81$), tetrahydrofuran ($\epsilon = 7.58$) and acetonitrile ($\epsilon = 37.5$). All the solvent molecules are small enough such that the intramolecular polarization can be turned off. The AMOEBA force field parameters for chloroform are taken from Mu et al.[84] and the AMOEBA parameters for tetrahydrofuran or acetonitrile are taken from Fried et al.[46] The fixed charge force field parameters are obtained from the general Amber force field (GAFF).[85] From the experimental measurements, the absorption spectrum of the DCBT molecule exhibits a red shift with increasing dielectric constant of the solvent. In Figure 3, all the three solvent models are able to predict this trend qualitatively. However, there are notable differences when comparing the excitation energies between the models. Qualitatively,

the solvent's polarization response to the solute excitation is expected to further stabilize the excited state thus reduce the excitation energy. Because PCM overestimates solvent polarization in response to the solute excitation, we expect it to over-stabilize the excited state, leading to the excitation energy being the smallest among the three solvent models. In contrast, because the fixed-charge model predicts no changes in the solvent polarization, it provides no additional stabilization of the excited state, leading to the excitation energy being the largest among the three solvent models. The AMOEBA model captures changes in the solvent polarization only from the electrons, and the excitation energy is expected to lie between PCM and fixed-charge model. From Figure 3, it is observed that the excitation energies in chloroform, tetrahydrofuran and acetonitrile are all consistent with the above qualitative analysis. However, when fixed-charge models are used, the excitation energy in the least polar solvent (i.e. chloroform) is unexpectedly higher than gas phase. Another un-expected behavior occurs in water, where although the excitation energy in PCM is still the lowest, the differences between AMOEBA and fixed-charge model is only within 0.01eV with AMOEBA being slightly higher on average. We found that similar unexpected observations were reported in Fried et al.[46] who study the solvent effects on the bond vibration frequency of acetophenone. First, the authors found that only polarizable models can account for the condensed-phase shift in the bond vibrational frequency when changing from gas phase to non-polar solvents, which is consistent with what we observed in chloroform. In addition, when comparing the electric field exerted on the carbonyl group from different solvents, the authors also observe that the differences between AMOEBA and fixed-charge model in wa-ter is much smaller compared to the other solvents studied, ranging from nonpolar organic solvents like hexane to polar organic solvents like dimethyl sulfoxide (DMSO). The authors attribute this abnormal behavior of water to hydrogen bonding, i.e. the "extra" electric field exerted by water compared to a polar organic solvent (e.g. DMSO) is mostly due to the existence of a few water molecules in close contact and form hydrogen-bonding with the solute. Since the both AMOEBA and fixed-charge model can describe hydrogen bonding, the two models exhibit a much smaller difference when describing solute-solvent electrostatic interactions in water compared to other solvents without hydrogen bonds. Based on these discussions, we think the hydrogen-bonding may also explain the result in Figure 3 (d), where the first excitation energies exhibit much smaller differences between AMOEBA and

26

fixed-charge model in water compared to other solvents.

C.   Nonadiabatic molecular dynamics

The main advantage of using a multi-reference electronic structure method is that it is capable of simulating nonadiabatic molecular dynamics with correct topology around conical intersections. When combined with AMOEBA polarizable force field for describing the solvents, it leads to the possibility of simulating the nonequilibrium solvation effects on photoreactions. To carry out nonadiabatic ab initio molecular dynamics simulations, we interface the SA-CASSCF/AMOEBA method developed in this work with the ab initio multiple spawning methods[69,70] in the FMS software. This is achieved by using the existing Terachem-FMS interfaces that already support SA-CASSCF calculations in gas phase, with small modifications in Terachem to make sure the positions and velocities of the AMOEBA solvents are sent and received properly. In this section, we perform some preliminary test calculations and discuss the issues suggested by the test results that need to be addressed in future works.

The test molecule used in this section is PSB3, which is a simplified model molecule for the retinal protonated Schiff base in vision. Ruckenbauer et al.[86] studied the effects of the solvent polarity over the photodynamics of PSB3 using multireference configuration-interaction with fixed charge solvent model, and observed that the excited state lifetimes of PSB3 in water is very close to that in the gas phase. In Figure 4, we present the results corresponding to a single FMS trajectory for PSB3 in the gas phase (using SA-CASSCF), and a single FMS trajectory for PSB3 solvated in water sphere with radius around 20 Angstrom (using SA-CASSCF/AMOEBA). The initial positions for the solvated system is randomly sampled from the equilibrated molecular dynamics simulations, and the initial velocities are randomly sampled from the Maxwell-Boltzmann distribution at 300K. As a result, the initial conditions used in this work are generated from purely classical statistics, which needs to be improved in the future as discussed at the end of this section. The initial positions and velocities of the solvated PSB3 molecule are then used as the initial conditions for launching the dynamics in the gas phase.

Figure 4 (a) and (b) show how the population changes along the trajectory as a function of

time in the gas phase and water respectively, and the molecule is fully relaxed to the ground state at around 300 fs in both simulations. In addition, we observed that in both trajectories, the main photoreaction coordinate corresponds to rotation around the central double bond. As shown in Figure 4 (e) and (f), the transfer of population to the ground state is strongly correlated with the changes of the dihedral angle formed by the four atoms highlighted in the inset structure in Figure 4. The inset structure illustrates the geometry when the molecule starts to transfer population to the ground state for the first time, and most transfers occur when the dihedral angle highlighted changes from flat to near perpendicular. For the two trajectories shown, the first occurrence of population transfer in the gas phase is at around 100 fs, whereas in AMOEBA it first occurs at around 150 fs. This time difference might be due to that in order for PSB3 to rotate around the central bond in solvent, the surrounding water molecules need to rearrange accordingly, although more substantial samplings of trajectories are required to determine whether the presence of explicit water will lead to notable differences in the excited state lifetime. In SI-Figure S1, we have also optimized the minimal energy conical intersection (MECI) in water using SA-CASSCF/AMOEBA, and the corresponding dihedral angle is -95 degrees. All of these observations are qualitatively consistent with previous theoretical studies, which suggests that the non-adiabatic dynamics simulations using SA-CASSCF/AMOEBA is set up correctly.

Despite the similarities discussed above, one notable difference between simulations in the gas phase and in AMOEBA is found when comparing Figure 4 (c) and (d), which show the time-dependent changes of the potential energies along the two trajectories. When following PSB3 on the first excited state, the potential energy in the gas phase changes less by 2eV during the whole trajectory. In contrast, for the trajectory in AMOEBA, the potential energy decreases sharply by more than 10eV during the first 50 fs. To help understand the source of this rapid decrease of potential energy, we analyze the time-dependent changes in the kinetic energy. We divide the solvated system into three components: the solute (PSB3), the inner shell of solvents (i.e. water molecules within 16 Angstrom from the solute), and the outer shell of solvents (i.e. water molecules beyond 16 Angstrom up to 20 Angstrom). The time-dependent changes of the kinetic energy corresponding to the three components are shown in SI-Figure S2. As shown in Figure SI-Figure S2(b), the kinetic energy of the solute and the inner solvent shell don't have notable changes in the first

50 fs. In contrast, the kinetic energy of the outer solvent shell increases rapidly in the first 50 fs, which strongly correlates with the rapid decrease of the potential energy. This observation suggests that the initial decrease of potential energy in AMOEBA is likely due to an artificial edge effect. When we generate samples of equilibrated solvent configurations using molecular dynamics simulations, periodic boundary conditions are enabled, thus all the water molecules are experiencing true condensed phase environment. However, the SA-CASSCF/AMOEBA method currently only supports non-periodic systems. As a result, at the start of the nonadiabatic MD simulations, the water molecules at the edge are suddenly exposed to vacuum, resulting in a large force compared to when they were in the condensed phase. Based on these considerations, the initial rapid decrease of potential energy is non-physical, and should be remedied in future works as discussed below.

The above simulation results demonstrate the potential of using SA-CASSCF/AMOEBA for performing non-adiabatic molecular dynamics simulations, but also point out a few limitations that need to addressed before more rigorous simulations can be carried out. One issue is the artificial edge effect pointed out in the last paragraph. There are two possible approaches to address this issue that we would like to try in future works. One possible approach is to restrain the outer solvent layer in the nonadiabatic MD simulations, preventing its kinetic energy from increasing too quickly. Although restraints are well established in classical MD simulations, it is less clear how restraints should be applied when the nuclei are treated quantum-mechanically as wavepackets as in the AIMS method. Another possible approach is to mimic the condensed phase environment by surrounding the AMOEBA explicit water molecules with a PCM cavity. Such multi-layer polarizable embedding approach has recently been reported in the context of TDDFT,[87] and would require describing the mutual polarization between the electrons, the induced dipoles from the AMOEBA atoms, and the induced surface charges in PCM.

A second limitation found in the above test calculations is related to the sampling of initial conditions. The current way of generating the initial positions and velocities for the solvated system is purely classical. However, because AIMS is a type of quantum-dynamics method, the proper way of sampling initial conditions should include quantum statistical mechanical effects. For small molecules in the gas phase, the initial conditions are often generated from the Wigner distribution,[88] which requires the Hessian of the system and can be computed by

finite difference of analytical nuclear gradients. For a system with N_{atom} atoms, computing the Hessian this way requires $3N_{\text{atom}}$ of nuclear gradient calculations, which is impractical for the solvated system with tens of thousands of atoms. Previous studies have proposed a combination of Wigner sampling for the QM region and classical sampling for the MM region,[89] which may potentially be a good solution to work around this issue.

A third limitation present in the test calculations is that the dynamic-weight scheme for SA-CASSCF/AMOEBA is not supported in the current Terachem-FMS interface, despite that the corresponding analytical gradients and non-adiabatic couplings are all available for single-point calculations. The advantage of using a dynamic weighting scheme for dynamics simulations has been discussed in the introduction section. However, the dynamic weight formula in Eq.13 requires specifying a reference state, which needs to be automatically adjusted based on which electronic state the molecule is currently propagating on. This requires adding new interface to the FMS codes to send the corresponding state information to the SA-CASSCF/AMOEBA codes, which we plan to develop in future works.

IV. CONCLUSIONS

In this work, we have developed SA-CASSCF in AMOEBA polarizable force fields for water as well as small solvent molecules, which enables the possibility of simulating nona-diabatic molecular dynamics with non-equilibrium solvent effects. In particular, we implemented a unified method for SA-CASSCF in polarizable model that supports both PCM and AMOEBA water model. The main advantage for the unified framework is that it may also be applied to polarizable models in addition to the two models discussed in this work. As long as the energy can be written in the general form of Eq.23, the corresponding energies, analytical gradients and non-adiabatic couplings can be computed using the existing algorithms by properly substituting the variables listed in Table I .

The preliminary test results presented in this work have suggested several things to improve in future works. In Section III A, we have observed that the computational cost is dominated by AMOEBA-related operations when the number of solvent molecules is large, and we have discussed the possibility of using the CUDA platform in OpenMM to help accelerate these operations. At the end of Section III C, we have pointed out three limitations

in the current nonadiabatic MD simulations using SA-CASSCF/AMOEBA. These include the edge effects leading to unphysical rapid decrease of potential energies at the beginning of the dynamics, missing of quantum effects when sampling initial conditions of the solvated system, and lack of supports for dynamic-weighted scheme in the nonadiabatic MD simulations. In Section III C, we have also discussed possible ways to address these limitations, so that rigorous nonadiabatic MD simulations using SA-CASSCF/AMOEBA can be routinely applied to the studies of photoreactions in solvents in the future.

By combining SA-CASSCF with AMOEBA models, it opens the door to simulating photoreactions in bio-macromolecular environments such as proteins. The study of solvent environment in this work serves as the initial step towards this goal, and we anticipate that there are two questions to be addressed for incorporating protein environment. One is the inclusion of intramolecular polarization. As discussed in Eq.II A, accounting for intramolecular polarization leads to two different electric fields used for separate purposes,[43] which makes the current self-consistent formulation insufficient. Fortunately, such difference can be naturally incorporated with the Lagrangian formulation, as previously shown in the context of TDDFT/AMOEBA.[56] Since the Lagrangian formulation has already been used in this work, including intramolecular polarization into SA-CASSCF/AMOEBA should primarily involve identifying the additional terms to be included into the coupled-perturbed equation. The other question for simulating protein environment is the treatment of the covalent QM/MM boundary. Compared to fixed charge force field models, the additional challenge from using polarizable models is how to avoid the polarization catastrophe. Methods specialized for describing QM/MM boundary with polarizable models have been developed before,[51] and we can potentially include these ideas in our future implementations.

SUPPLEMENTARY MATERIAL

See supplementary material for additional test results, descriptions of the detailed procedure used for setting up the QM/MM systems, descriptions of new OpenMM APIs added for this work, as well as sample input and output files from this work.

AUTHOR DECLARATION

The author has no conflicts to disclose.

DATA AVAILABILITY

For each system studied in this work, one sample calculation that includes input and output files is provided in the supplementary information. Additional data is available from the corresponding author upon reasonable request.

Appendix A

Here we describe how the the purely classical contributions to the energy (i.e. $\epsilon_{\text{non-polarize}}$) and gradients (i.e. the last three terms in Eq.31, and the last two terms in Eq.44) are evaluated using the newly added OpenMM APIs. Because the valence and van der Waals contributions can be evaluated with the standard OpenMM APIs, here we will focus on the electrostatic contributions.

In Section II A, we have mentioned that in the more general situations for AMOEBA, there are two types of electric field generated from the fixed multipoles, i.e. $\vec{\mathcal{E}}^{(\text{fix-D})}$ and $\vec{\mathcal{E}}^{(\text{fix-P})}$. Correspondingly, we can define two sets of induced dipoles:

$$\mu_{Ai}^{(\text{D})} = \alpha_A \mathcal{E}_{Ai}^{(\text{fix-D})}, \mu_{Ai}^{(\text{P})} = \alpha_A \mathcal{E}_{Ai}^{(\text{fix-P})} \tag{A1}$$

Furthermore, there are two different "polarization types" in AMOEBA, both supported by OpenMM. The one presented in section II A corresponds to the "mutual" polarization type. Its energy is evaluated as Eq.5, and the gradients are derived in the appendix of Ren et al.[43]

32

as

$$G_{\text{Mutual}}^{(\text{es}),\xi} = \frac{d\epsilon_{\text{es-fix}}}{d\xi} - \frac{1}{2} \sum_{Ai} \left(\frac{\partial \mathcal{E}_{Ai}^{(\text{fix-D})}}{\partial \xi} \mu_{Ai}^{(\text{P})} + \mu_{Ai}^{(\text{D})} \frac{\partial \mathcal{E}_{Ai}^{(\text{fix-P})}}{\partial \xi} \right) + \frac{1}{2} \sum_{AiBj} \mu_{Ai}^{(\text{P})} \frac{\partial T_{Ai,Bj}^{(\text{Thole})}}{\partial \xi} \mu_{Bj}^{(\text{D})} \quad \text{(A2)}$$

$\epsilon_{\text{es-fix}}$ represents the electrostatic interactions among the fixed multipoles, i.e. $\epsilon_{\text{es}}[\mathbf{Z}, \mathbf{C}, \mathbf{M}, \mathbf{Q}]$ in Eq.17. The second polarization type is called "direct", which can be considered as an approximation to Eq.5 by neglecting the mutual interaction and approximating the energy as

$$E_{\text{Direct}}^{(\text{es})} = \epsilon_{\text{es-fix}} - \frac{1}{2} \sum_{Ai} \mu_{Ai}^{(\text{D})} \mathcal{E}_{Ai}^{(\text{fix-P})} \quad \text{(A3)}$$

The corresponding analytical gradients thus become

$$G_{\text{Direct}}^{(\text{es}),\xi} = \frac{d\epsilon_{\text{es-fix}}}{d\xi} - \frac{1}{2} \sum_{Ai} \left(\frac{\partial \mathcal{E}_{Ai}^{(\text{fix-D})}}{\partial \xi} \mu_{Ai}^{(\text{P})} + \mu_{Ai}^{(\text{D})} \frac{\partial \mathcal{E}_{Ai}^{(\text{fix-P})}}{\partial \xi} \right) \quad \text{(A4)}$$

The various types of classical electrostatic contributions to energies or gradients can then be evaluated by properly choosing different combinations of "mutual" or "direct" polarization types, and choosing the values of $\boldsymbol{\mu}^{(\text{P})}$ and $\boldsymbol{\mu}^{(\text{D})}$.

From Eq.A3, it is obvious that $\epsilon_{\text{es-fix}}$ in $\epsilon_{\text{non-polarize}}$ can be obtained if $\boldsymbol{\mu}^{(\text{D})}$ is equal to zero

$$\epsilon_{\text{es-fix}} = \tilde{E}_{\text{Direct}}^{(\text{es})}[\boldsymbol{\mu}^{(\text{D})} = 0] \quad \text{(A5)}$$

Similarly, from Eq.A4, the gradients of $\epsilon_{\text{es-fix}}$ can be evaluated by passing zero to both $\boldsymbol{\mu}^{(\text{P})}$ and $\boldsymbol{\mu}^{(\text{D})}$

$$\frac{d\epsilon_{\text{es-fix}}}{d\xi} = \tilde{G}_{\text{Direct}}^{(\text{es}),\xi} \left[\boldsymbol{\mu}^{(\text{D})} = 0, \boldsymbol{\mu}^{(\text{P})} = 0 \right] \quad \text{(A6)}$$

We now consider $W^{(\text{fix}),\xi}[\mathbf{d}]$ defined in Eq.33. Recall that this work assumes $\mathcal{E}_{Ai}^{(\text{fix-D})} = \mathcal{E}_{Ai}^{(\text{fix-P})}$. By comparing Eq.33 and Eq.A4, it's easy to see that

$$W^{(\text{fix}),\xi}[\mathbf{d}] = \tilde{G}_{\text{Direct}}^{(\text{es}),\xi} \left[\boldsymbol{\mu}^{(\text{D})} = \mathbf{d}, \boldsymbol{\mu}^{(\text{P})} = \mathbf{d} \right] - \frac{d\epsilon_{\text{es-fix}}}{d\xi} \quad \text{(A7)}$$

where $\frac{d\epsilon_{\text{es-fix}}}{d\xi}$ is computed using Eq.A6. Finally, for $\Xi^{\xi}[\mathbf{d}^{(1)}, \mathbf{d}^{(2)}]$ defined in Eq.32, by com-

paring the equation with Eq.A2, it can be evaluated if all the fixed multipoles in Eq.A2 are set to zero,

$$\Xi^{\xi}[\mathbf{d}^{(1)}, \mathbf{d}^{(2)}] = \tilde{G}_{\text{Mutual}}^{(\text{es}),\xi}\left[\boldsymbol{\mu}^{(D)} = \mathbf{d}^{(1)}, \boldsymbol{\mu}^{(P)} = \mathbf{d}^{(2)}, \{\mathbf{Z}, \mathbf{C}, \mathbf{M}, \mathbf{Q}\} = 0\right] \tag{A8}$$

where the values of the fixed multipoles can be modified using the standard OpenMM APIs.

TABLE I. Comparison of SA-CASSCF/PCM with SA-CASSCF/AMOEBA. The "Symbol" column summarizes the symbols used in the unified formulation, i.e. eq.(23), eq.(31) and eq.(44). The "PCM" and "AMOEBA" columns summarize the corresponding definitions in each model. Notations in the "PCM" column are consistent with Ref.[65]. For the "AMOEBA" column, related equation numbers have been provided for each term.

	Quantity	Symbol	PCM	AMOEBA	eq.
Variable	Electron density matrix	P_{pq}	Same	Same	
	Fixed multipoles	-	Solute nuclear charges $\mathbf{Z}$	Solute nuclear charges $\mathbf{Z}$ Solvent permanent charges $\mathbf{C}$ dipoles $\mathbf{M}$, quadrupoles $\mathbf{Q}$	
	Induced variables	p_μ	Induced charges $p_{\mu=n}=q_n$	Induced dipoles $p_{\mu=3A+i}=d_{Ai}$	
Energy	Fock from fixed	$V_{pq}^{(\mathrm{fix})}$	$V_{pq}^{(0)}[\mathbf{Z}]$	$V_{pq}^{(0)}[\mathbf{Z}]+V_{pq}^{(0)}[\mathbf{C}]$ $+V_{pq}^{(1)}[\mathbf{M}]+V_{pq}^{(2)}[\mathbf{Q}]$	(18)-(20)
	Fock from induced	$V_{pq}^{(\mathrm{induce})}$	$\sum_n \beta_{pq,n}q_n$	$\sum_{Ai} v_{pq,Ai}^{(1)} d_{Ai}$	(19)
	Induction from electrons	$W_\mu^{(\mathrm{elec})}$	$\sum_{pq} P_{pq}\beta_{pq,n}$	$\sum_{pq} P_{pq}v_{pq,Ai}^{(1)}$	(22)
	Induction from fixed	$W_\mu^{(\mathrm{fix})}$	$\sum_N Z_N B_{N,n}$	$-\mathcal{E}_{Ai}^{(\mathrm{fix})}[\mathbf{Z},\mathbf{C},\mathbf{D},\mathbf{Q}]$	(21)
	Mutual induction	$\Xi_{\mu\nu}$	A_{nm}	$O_{Ai,Bj}$	(7)
	Non-polarizable	$\epsilon_{\mathrm{non\text{-}polarize}}$	$\epsilon_{\mathrm{es}}[\mathbf{Z}]$	$\epsilon_{\mathrm{es}}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}]+\epsilon_{\mathrm{valence}}+\epsilon_{\mathrm{vdW}}$	(A5)
Gradient	Electrons - fixed	$V^{(\mathrm{fix}),\xi}[\mathbf{P}]$	$V^{(0),\xi}[\mathbf{P},\mathbf{Z}]$	$V^{(0),\xi}[\mathbf{P},\mathbf{Z}]+V^{(0),\xi}[\mathbf{P},\mathbf{C}]$ $+V^{(1),\xi}[\mathbf{P},\mathbf{M}]+V^{(2),\xi}[\mathbf{P},\mathbf{Q}]$	(35)-(42)
	Electrons - induced	$W^{(\mathrm{elec}),\xi}[\mathbf{P},\mathbf{p}]$	$\sum_{pq,n} P_{pq}\frac{\partial\beta_{pq,n}}{\partial\xi}q_n$	$\sum_{pq,Ai} P_{pq}\frac{\partial v_{pq,Ai}^{(1)}}{\partial\xi}d_{Ai}$	(34)
	Fixed - induced	$W^{(\mathrm{fix}),\xi}[\mathbf{p}]$	$\sum_{Nn} Z_N\frac{\partial B_{N,n}}{\partial\xi}q_n$	$-\sum_{Ai}\frac{\partial\mathcal{E}_{Ai}^{(\mathrm{fix})}[\mathbf{Z},\mathbf{C},\mathbf{D},\mathbf{Q}]}{\partial\xi}d_{Ai}$	(33),(A7)
	Induced - induced	$\Xi^\xi[\mathbf{p},\mathbf{p}']$	$\sum_{nm} q_n\frac{\partial A_{nm}}{\partial\xi}q_m'$	$\sum_{AiBj} d_{Ai}\frac{\partial O_{Ai,Bj}}{\partial\xi}d_{Bj}'$	(32),(A8)
	Non-polarizable	$\frac{d\epsilon_{\mathrm{non\text{-}polarize}}}{d\xi}$	$\frac{d\epsilon_{\mathrm{es}}[\mathbf{Z}]}{d\xi}$	$\frac{d\epsilon_{\mathrm{es}}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}]}{d\xi}+\frac{d\epsilon_{\mathrm{valence}}}{d\xi}+\frac{d\epsilon_{\mathrm{vdW}}}{d\xi}$	(A6)

TABLE II. Algorithms of computing SA-CASSCF/AMOEBA energies, analytical nuclear gradients and non-adiabatic couplings. Definitions for symbols appearing in the table can be found in Table I. "CAS/PCM" refers to Ref.[65]. The arrow $\leftarrow$ represents the corresponding calculations are carried out in the MM codes, and the results are passed to and stored in the QM codes.

| QM codes	Call MM API
Initialization	
	1 Construct pure MM system from XML files
2 Read list of QM atoms	
	3 Delete valence forces that involve QM atoms
	4 Exclude vdW interactions between QM atoms
	5 Set permanent charges of QM atoms to nuclear charges $\mathbf{Z}$
	6 Set permanent dipoles and quadrupoles of QM atoms to zero
7 Save fixed point multipoles: $[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}]$	$\leftarrow$ Get permanent multipoles in lab frame
8 Compute one-electron kinetic integral: $h_{pq} = K_{pq}$	
9 Add Fock matrix from fixed: $h_{pq}+=V_{pq}^{(\text{fix})}[\mathbf{Z},\mathbf{C},\mathbf{M},\mathbf{Q}]$	
10 Save electric fields exerted on MM atoms: $\mathcal{E}_{Ai}^{(\text{fix})}$	$\leftarrow$ Get electric fields from fixed multipoles at all atom positions
11 Save mutual induction matrix between MM atoms: $O_{Ai,Bj}$	$\leftarrow$ Get mutual induction matrix between all atoms
Energy: minimize eq.(23)	
12 Solve MO, CI, dynamic weights, and MM induced dipoles $\mathbf{d}$	
Follow Steps 4-17 of Algorithm Table I in CAS/PCM	
Substitute variables according to Table I in this work	
Analytical nuclear gradients of state Θ: eq.(31)	
13 Solve Lagrange multipliers $\bar{\boldsymbol{\kappa}}^{\Theta}, \bar{\boldsymbol{\theta}}^{\Theta}, \bar{\mathbf{d}}^{\Theta}$	
Follow Steps 5-21 of Algorithm Table II in CAS/PCM	
Substitute variables according to Table I in this work	
14 Build effective density matrices $\boldsymbol{\gamma}^{\Theta}, \boldsymbol{\Gamma}^{\Theta}, \tilde{\mathbf{X}}^{\Theta}$	
Follow Steps 22-24 of Table II in CAS/PCM	
15 Add gradients from kinetic, overlap, and ERI	
16 Add gradients from electron-induced: $W^{(\text{elec}),\xi}[\boldsymbol{\gamma}^{\Theta},\mathbf{d}]$,	
17 Add gradients from electron-induced: $W^{(\text{elec}),\xi}\left[\mathbf{P}^{\text{SA}},\bar{\mathbf{d}}^{\Theta}\right]$	
18 Add partial gradients from electron-fixed: $V_{\text{no-rotate}}^{(\text{fix}),\xi}[\boldsymbol{\gamma}^{\Theta}]$	
19 Compute torques from electron-fixed: $\tau^{(\text{fix})}[\boldsymbol{\gamma}^{\Theta}]$	
	20 Convert torques from Step 19 and add to gradients
	21 Add gradients from fixed-induced: $W^{(\text{fix}),\xi}[\mathbf{d}+\bar{\mathbf{d}}^{\Theta}]$
	22 Add gradients from induced-induced: $\Xi^{\xi}\left[\mathbf{d},\frac{1}{2}\mathbf{d}+\bar{\mathbf{d}}^{\Theta}\right]$
	23 Add gradients from non-polarizable: $\frac{d\epsilon_{\text{non-polarize}}}{d\xi}$
Nonadiabatic couplings between states Θ and Π: eq.(44)	
24 Solve Lagrange multipliers $\bar{\boldsymbol{\kappa}}^{\Theta\Pi}, \bar{\boldsymbol{\theta}}^{\Theta\Pi}, \bar{\mathbf{d}}^{\Theta\Pi}$	
25 Build effective density matrices $\boldsymbol{\gamma}^{\Theta\Pi}, \boldsymbol{\Gamma}^{\Theta\Pi}, \tilde{\mathbf{X}}^{\Theta\Pi}$	
Follow Steps 5-7 of Table III in CAS/PCM	
26 Save gradients from kinetic, overlap, and ERI to G^{ξ}	
27 Add gradients from electron-induced: $W^{(\text{elec}),\xi}\left[\boldsymbol{\gamma}^{\Theta\Pi},\mathbf{d}\right]$ to G^{ξ}	
28 Add gradients from electron-induced: $W^{(\text{elec}),\xi}\left[\mathbf{P}^{\text{SA}},\bar{\mathbf{d}}^{\Theta\Pi}\right]$ to G^{ξ}	
29 Add partial gradients electron-fixed: $V_{\text{no-rotate}}^{(\text{fix}),\xi}[\boldsymbol{\gamma}^{\Theta\Pi}]$ to G^{ξ}	
30 Compute torques from electron-fixed: $\tau^{(\text{fix})}[\boldsymbol{\gamma}^{\Theta\Pi}]$	
	31 Convert torques from Step 30 and add to gradients G^{ξ}
	32 Add gradients from fixed-induced $W^{(\text{fix}),\xi}[\bar{\mathbf{d}}^{\Theta\Pi}]$ to G^{ξ}
	33 Add gradients from induced-induced $\Xi^{\xi}\left[\mathbf{d},\bar{\mathbf{d}}^{\Theta\Pi}\right]$ to G^{ξ}
34 Compute NAC: $\chi_{\xi}^{\Theta\Pi} = \delta_{\xi}^{\Theta\Pi} - \frac{1}{\Delta^{\Theta\Pi}}G^{\xi}$	
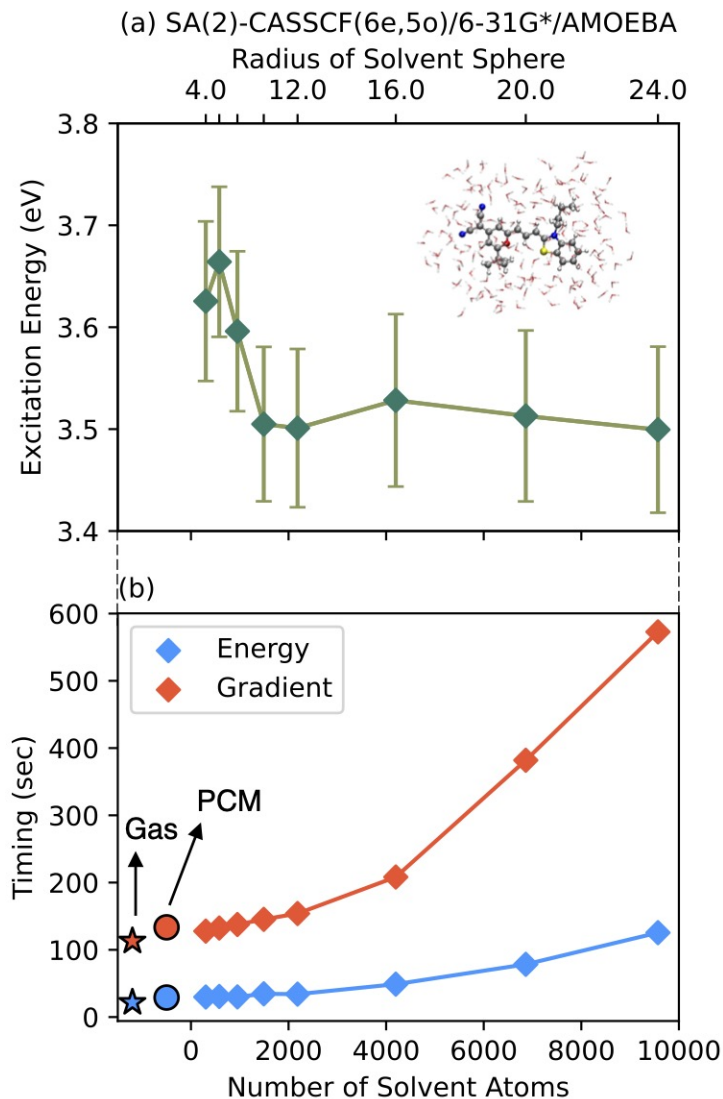
FIG. 1. Excitation energies and computational time of SA-CASSCF/AMOEBA with respect to increasing sizes of the MM solvent environment. The inset structure illustrates the chromophore DCBT solvated in water molecules. (a) shows changes in the first excitation energies. Each data point is computed from 10 different geometries sampled from the equilibrated AMOEBA molecular dynamics simulations, and the error bar indicates one standard deviation. (b) shows changes in computational time for both energy (blue) and gradient (red) calculations. The timing of gas phase calculations (star) and PCM ($\epsilon = 80.1$) calculations (circle) have also been included for comparison. (a) and (b) share identical x-axis, which labels the sizes of the MM region both in terms of the number of MM atoms, as well as the radius of the solvent sphere.

DW-SA(2)-CASSCF(6e, 5o)/6-31G*

(a) Attach (Blue)–Detach(Red) density of first excitation

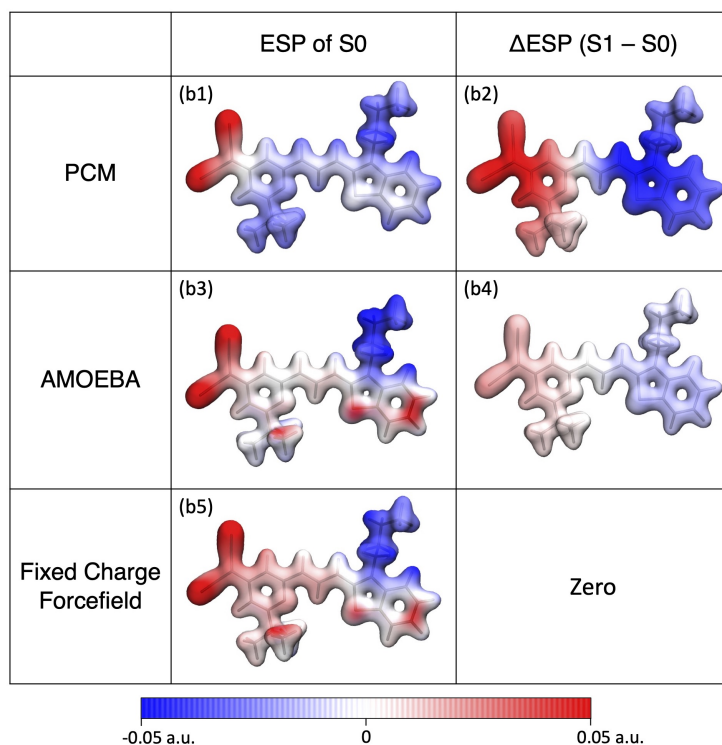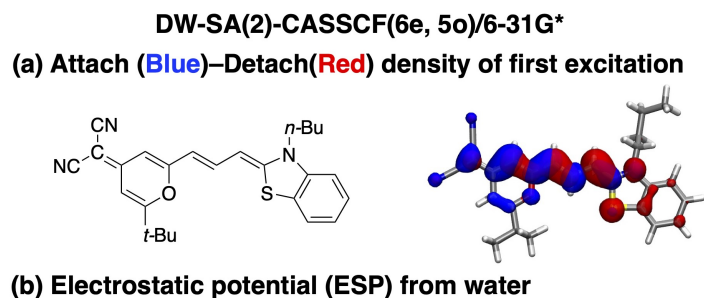**(b) Electrostatic potential (ESP) from water**

FIG. 2. (a) Attachment density (blue) and detachment density (red) for the first excitation of DCBT chromophore in the gas phase using SA-CASSCF. (b) Comparison of electrostatic potential (ESP) generated by the solvent in DW-SA-CASSCF calculations using three different water models, including PCM ($\epsilon = 80.1$), AMOEBA, and fixed-charge force field. The ESP is plotted on the total electronic density isosurface of the chromophore with isovalue 0.05. The first column "ESP-S0" represents the total ESP generated by the solvent when the chromophore is in its ground state. The second column "Δ ESP(S1-S0)" represents changes in the solvent ESP when the chromophore is excited from S0 to S1. The bandwidth used in the dynamic weight formula is 3 eV. The QM/MM geometry is randomly sampled from the equilibrated AMOEBA molecular dynamics simulations. The radius of the solvent sphere in QM/MM is around 24 angstrom.
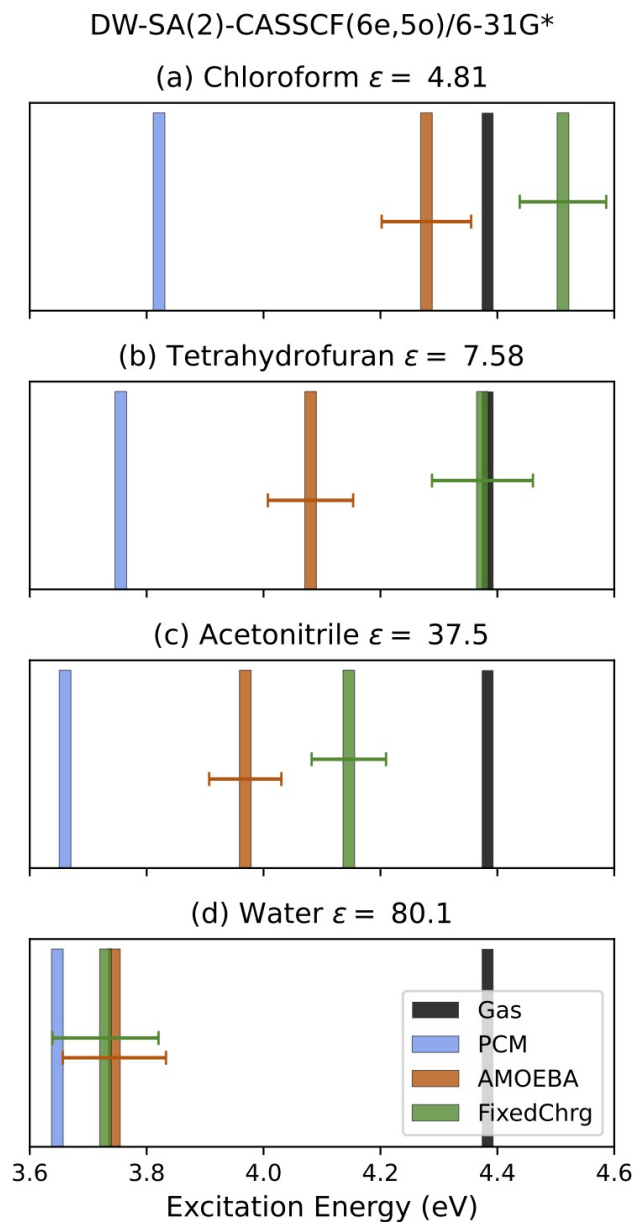
FIG. 3. First excitation energies of DCBT computed with DW-SA-CASSCF in the gas phase and in 4 different solvents. The figure compares three solvent models, including PCM (blue), AMOEBA (orange) and fixed charge force field (green). Gas phase and PCM are computed using the optimized geometry in gas phase. Excitation energies in QM/MM calculations (AMOEBA or fixed charge force field) are computed from 10 structures sampled from equilibriated MD simulations. The error bar represents one standard deviation. The radius of the solvent spheres in QM/MM calculations is around 24 angstrom. Intramolecular polarizations in solvent molecules have been neglected in all SA-CASSCF/AMOEBA calculations.
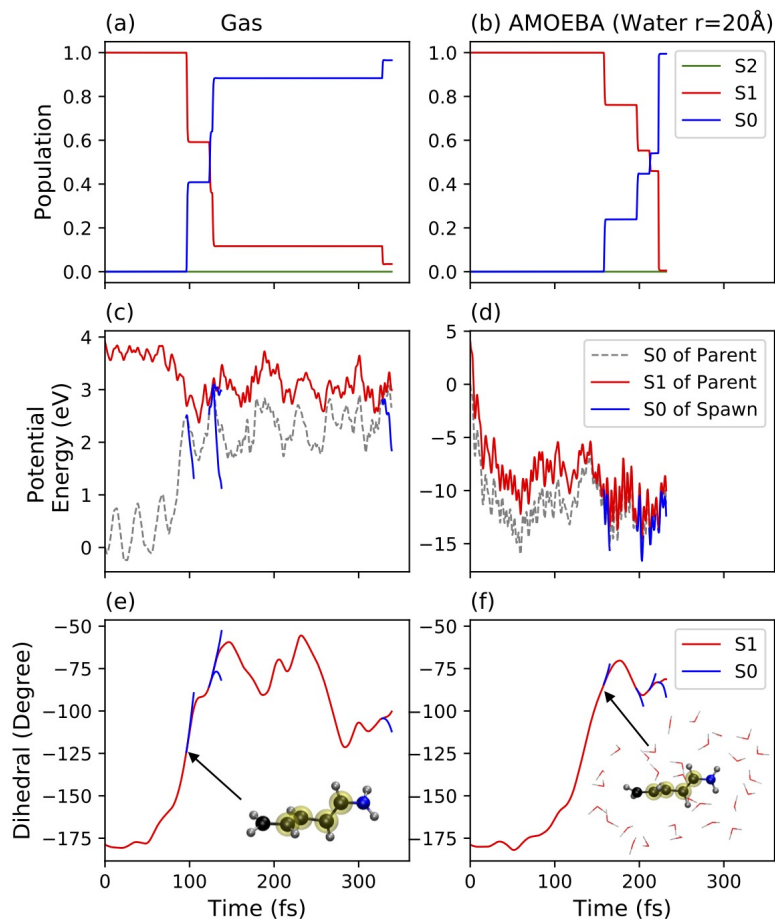
FIG. 4. Nonadiabatic molecular dynamics of PSB3 using SA-CASSCF in gas phase (left column) and in AMOEBA water model (right column, solvent contains around 5000 atoms). One AIMS trajectory is simulated in each environment. (a) and (b) show changes in the population on each electronic state as a function of time along the trajectory. (c) and (d) show changes in the potential energies as a function of time, where "parent" refers to the initial wavepacket, and "spawns" are created when internal conversions take place. (e) and (f) show changes in the highlighted dihedral angle as a function of time. The inset structure illustrates the geometry where the first spawned wavepacket is created, and only the first solvation shell is shown.

REFERENCES

[1] A. Wand, I. Gdor, J. Zhu, M. Sheves, and S. Ruhman, "Shedding New Light on Retinal Protein Photochemistry," Annual Review of Physical Chemistry **64**, 437–458 (2013).

[2] L. A. Baker, M. D. Horbury, S. E. Greenough, F. Allais, P. S. Walsh, S. Habershon, and V. G. Stavros, "Ultrafast Photoprotecting Sunscreens in Natural Plants," The Journal of Physical Chemistry Letters **7**, 56–61 (2016).

[3] M. Orozco and F. J. Luque, "Theoretical Methods for the Description of the Solvent Effect in Biomolecular Systems," Chemical Reviews **100**, 4187–4226 (2000).

[4] H. Lin and D. G. Truhlar, "QM/MM: What have we learned, where are we, and where do we go from here?" Theoretical Chemistry Accounts **117**, 185–199 (2007).

[5] J. Gao, "Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry," Accounts of Chemical Research **29**, 298–305 (1996).

[6] H. A. Frank, J. A. Bautista, J. Josue, Z. Pendon, R. G. Hiller, F. P. Sharples, D. Gosztola, and M. R. Wasielewski, "Effect of the Solvent Environment on the Spectroscopic Properties and Dynamics of the Lowest Excited States of Carotenoids," The Journal of Physical Chemistry B **104**, 4569–4577 (2000).

[7] J. Hoche, A. Schulz, L. M. Dietrich, A. Humeniuk, M. Stolte, D. Schmidt, T. Brixner, F. Würthner, and R. Mitric, "The origin of the solvent dependence of fluorescence quantum yields in dipolar merocyanine dyes," Chemical Science **10**, 11013–11022 (2019).

[8] M. Nottoli, L. Cupellini, F. Lipparini, G. Granucci, and B. Mennucci, "Multiscale Models for Light-Driven Processes," Annual Review of Physical Chemistry **72**, 489–513 (2021).

[9] G. R. Fleming, S. H. Courtney, and M. W. Balk, "Activated barrier crossing: Comparison of experiment and theory," Journal of Statistical Physics **42**, 83–104 (1986).

[10] L. Nikowa, D. Schwarzer, J. Troe, and J. Schroeder, "Viscosity and solvent dependence of low-barrier processes: Photoisomerization of cis-stilbene in compressed liquid solvents," The Journal of Chemical Physics **97**, 4827–4835 (1992).

[11] J. Saltiel and J. T. D'Agostino, "Separation of viscosity and temperature effects on the singlet pathway to stilbene photoisomerization," Journal of the American Chemical Society **94**, 6445–6456 (1972).

[12]D. Desai, D. S. Desai, D. Li, R. Janjikhel, B. Lang, and N. Jain, "Effects of Light Intensity, n-Alcohols, Water-Soluble Colorants, and Solution Viscosity on Photoisomerization of Sorivudine," Pharmaceutical Development and Technology **6**, 99–106 (2001).

[13]A. Vong, D. R. Widmer, and B. J. Schwartz, "Nonequilibrium Solvent Effects during Photodissociation in Liquids: Dynamical Energy Surfaces, Caging, and Chemical Identity," The Journal of Physical Chemistry Letters **11**, 9230–9238 (2020).

[14]W. J. Glover, R. E. Larsen, and B. J. Schwartz, "How Does a Solvent Affect Chemical Bonds? Mixed Quantum/Classical Simulations with a Full CI Treatment of the Bonding Electrons," The Journal of Physical Chemistry Letters **1**, 165–169 (2010).

[15]B. Dereka and E. Vauthey, "Solute–Solvent Interactions and Excited-State Symmetry Breaking: Beyond the Dipole–Dipole and the Hydrogen-Bond Interactions," The Journal of Physical Chemistry Letters **8**, 3927–3932 (2017).

[16]J. Tomasi, B. Mennucci, and R. Cammi, "Quantum Mechanical Continuum Solvation Models," Chemical Reviews **105**, 2999–3094 (2005).

[17]R. M. Levy and E. Gallicchio, "Computer simulations with explicit solvent: Recent progress in the thermodynamic decomposition of free energies and in modeling electrostatic effects," Annual Review of Physical Chemistry **49**, 531–567 (1998).

[18]L. Onsager, "Electric Moments of Molecules in Liquids," Journal of the American Chemical Society **58**, 1486–1493 (1936).

[19]A. Klamt and G. Schüürmann, "COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," Journal of the Chemical Society, Perkin Transactions 2 , 799–805 (1993).

[20]J. M. Herbert, "Dielectric continuum methods for quantum chemistry," WIREs Computational Molecular Science **11**, e1519 (2021).

[21]B. Mennucci, "Polarizable continuum model," WIREs Computational Molecular Science **2**, 386–404 (2012).

[22]S. Corni, S. Pipolo, and R. Cammi, "Equation of Motion for the Solvent Polarization Apparent Charges in the Polarizable Continuum Model: Application to Real-Time TDDFT," The Journal of Physical Chemistry A **119**, 5405–5416 (2015).

[23]F. Ding, D. B. Lingerfelt, B. Mennucci, and X. Li, "Time-dependent non-equilibrium dielectric response in QM/continuum approaches," The Journal of Chemical Physics **142**,

034120 (2015).

[24] L.-P. Wang, T. J. Martinez, and V. S. Pande, "Building Force Fields: An Automatic, Systematic, and Reproducible Approach," The Journal of Physical Chemistry Letters **5**, 1885–1891 (2014).

[25] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura, and T. Head-Gordon, "Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew," The Journal of Chemical Physics **120**, 9665–9678 (2004).

[26] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," The Journal of Chemical Physics **79**, 926–935 (1983).

[27] T. A. Halgren and W. Damm, "Polarizable force fields," Current Opinion in Structural Biology **11**, 236–242 (2001).

[28] A. Warshel, M. Kato, and A. V. Pisliakov, "Polarizable Force Fields: History, Test Cases, and Prospects," Journal of Chemical Theory and Computation **3**, 2034–2045 (2007).

[29] S. Patel and C. L. Brooks, "Fluctuating charge force fields: Recent developments and applications from small molecules to macromolecular biological systems," Molecular Simulation **32**, 231–249 (2006).

[30] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: Application to liquid water," The Journal of Chemical Physics **101**, 6141–6156 (1994).

[31] G. Lamoureux and B. Roux, "Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm," The Journal of Chemical Physics **119**, 3025–3039 (2003).

[32] J. A. Lemkul, J. Huang, B. Roux, and A. D. J. MacKerell, "An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications," Chemical Reviews **116**, 4983–5013 (2016).

[33] J. Applequist, J. R. Carl, and K.-K. Fung, "Atom dipole interaction model for molecular polarizability. Application to polyatomic molecules and determination of atom polarizabilities," Journal of the American Chemical Society **94**, 2952–2960 (1972).

[34] B. T. Thole, "Molecular polarizabilities calculated with a modified dipole interaction," Chemical Physics **59**, 341–350 (1981).

[35] M. Bondanza, M. Nottoli, L. Cupellini, F. Lipparini, and B. Mennucci, "Polarizable embed-

ding QM/MM: The future gold standard for complex (bio)systems?" Physical Chemistry Chemical Physics **22**, 14433–14448 (2020).

[36]B. Hess, "Determining the shear viscosity of model liquids from molecular dynamics simulations," The Journal of Chemical Physics **116**, 209–217 (2002).

[37]C. J. Cramer and D. G. Truhlar, "A Universal Approach to Solvation Modeling," Accounts of Chemical Research **41**, 760–768 (2008).

[38]T. A. Halgren, "The representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters," Journal of the American Chemical Society **114**, 7827–7843 (1992).

[39]J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. J. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, and T. Head-Gordon, "Current Status of the AMOEBA Polarizable Force Field," The Journal of Physical Chemistry B **114**, 2549–2564 (2010).

[40]P. Ren and J. W. Ponder, "Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation," The Journal of Physical Chemistry B **107**, 5933–5947 (2003).

[41]P. Ren and J. W. Ponder, "Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations," Journal of Computational Chemistry **23**, 1497–1506 (2002).

[42]L.-P. Wang, T. Head-Gordon, J. W. Ponder, P. Ren, J. D. Chodera, P. K. Eastman, T. J. Martinez, and V. S. Pande, "Systematic Improvement of a Classical Molecular Model of Water," The Journal of Physical Chemistry B **117**, 9956–9972 (2013).

[43]P. Ren, C. Wu, and J. W. Ponder, "Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules," Journal of Chemical Theory and Computation **7**, 3143–3161 (2011).

[44]Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder, and P. Ren, "Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins," Journal of Chemical Theory and Computation **9**, 4046–4063 (2013).

[45]C. Liu, R. Qi, Q. Wang, J.-P. Piquemal, and P. Ren, "Capturing Many-Body Interactions with Classical Dipole Induction Models," Journal of Chemical Theory and Computation **13**, 2751–2761 (2017).

[46]S. D. Fried, L.-P. Wang, S. G. Boxer, P. Ren, and V. S. Pande, "Calculations of the Electric

Fields in Liquid Solutions," The Journal of Physical Chemistry B **117**, 16236–16248 (2013).

[47]W. Kohn, A. D. Becke, and R. G. Parr, "Density Functional Theory of Electronic Structure," The Journal of Physical Chemistry **100**, 12974–12980 (1996).

[48]R. G. Parr, "Density Functional Theory of Atoms and Molecules," in *Horizons of Quantum Chemistry*, Académie Internationale Des Sciences Moléculaires Quantiques / International Academy of Quantum Molecular Science, edited by K. Fukui and B. Pullman (Springer Netherlands, Dordrecht, 1980) pp. 5–15.

[49]Y. Mao, Y. Shao, J. Dziedzic, C.-K. Skylaris, T. Head-Gordon, and M. Head-Gordon, "Performance of the AMOEBA Water Model in the Vicinity of QM Solutes: A Diagnosis Using Energy Decomposition Analysis," Journal of Chemical Theory and Computation **13**, 1963–1979 (2017).

[50]D. Loco, L. Lagardère, S. Caprasecca, F. Lipparini, B. Mennucci, and J.-P. Piquemal, "Hybrid QM/MM Molecular Dynamics with AMOEBA Polarizable Embedding," Journal of Chemical Theory and Computation **13**, 4025–4033 (2017).

[51]E. G. Kratz, A. R. Walker, L. Lagardère, F. Lipparini, J.-P. Piquemal, and G. A. Cisneros, "LICHEM: A QM/MM Program for Simulations with Multipolar and Polarizable Force Fields," Journal of computational chemistry **37**, 1019–1029 (2016).

[52]D. Loco, L. Lagardère, G. A. Cisneros, G. Scalmani, M. Frisch, F. Lipparini, B. Mennucci, and J.-P. Piquemal, "Towards large scale hybrid QM/MM dynamics of complex systems with advanced point dipole polarizable embeddings," Chemical Science **10**, 7200–7211 (2019).

[53]M. Casida and M. Huix-Rotllant, "Progress in Time-Dependent Density-Functional Theory," Annual Review of Physical Chemistry **63**, 287–323 (2012).

[54]E. Runge and E. K. U. Gross, "Density-Functional Theory for Time-Dependent Systems," Physical Review Letters **52**, 997–1000 (1984).

[55]D. Loco, É. Polack, S. Caprasecca, L. Lagardère, F. Lipparini, J.-P. Piquemal, and B. Mennucci, "A QM/MM Approach Using the AMOEBA Polarizable Embedding: From Ground State Energies to Electronic Excitations," Journal of Chemical Theory and Computation **12**, 3654–3661 (2016).

[56]M. Nottoli, B. Mennucci, and F. Lipparini, "Excited state Born–Oppenheimer molecular dynamics through coupling between time dependent DFT and AMOEBA," Physical

Chemistry Chemical Physics **22**, 19532–19541 (2020).

[57]M. Bondanza, B. Demoulin, F. Lipparini, M. Barbatti, and B. Mennucci, "Trajectory Surface Hopping for a Polarizable Embedding QM/MM Formulation," The Journal of Physical Chemistry A **126**, 6780–6789 (2022).

[58]B. G. Levine, C. Ko, J. Quenneville, and T. J. MartÍnez, "Conical intersections and double excitations in time-dependent density functional theory," Molecular Physics **104**, 1039–1051 (2006).

[59]B. O. Roos, P. R. Taylor, and P. E. M. Sigbahn, "A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach," Chemical Physics **48**, 157–173 (1980).

[60]B. O. Roos, "The complete active space SCF method in a fock-matrix-based super-CI formulation," International Journal of Quantum Chemistry **18**, 175–189 (1980).

[61]M. W. Schmidt and M. S. Gordon, "The Construction and Interpretation of Mcscf Wavefunctions," Annual Review of Physical Chemistry **49**, 233–266 (1998).

[62]Q. Li, B. Mennucci, M. A. Robb, L. Blancafort, and C. Curutchet, "Polarizable QM/MM Multiconfiguration Self-Consistent Field Approach with State-Specific Corrections: Environment Effects on Cytosine Absorption Spectrum," Journal of Chemical Theory and Computation **11**, 1674–1682 (2015).

[63]M. A. Hagras and W. J. Glover, "Polarizable Embedding for Excited-State Reactions: Dynamically Weighted Polarizable QM/MM," Journal of Chemical Theory and Computation **14**, 2137–2144 (2018).

[64]X. Liu, A. Humeniuk, and W. J. Glover, "Conical Intersections in Solution with Polarizable Embedding: Integral-Exact Direct Reaction Field," Journal of Chemical Theory and Computation (2022), 10.1021/acs.jctc.2c00662.

[65]C. Song, "State-averaged CASSCF with polarizable continuum model for studying photoreactions in solvents: Energies, analytical nuclear gradients, and non-adiabatic couplings," The Journal of Chemical Physics **156**, 104102 (2022).

[66]J. STÅLRING, A. BERNHARDSSON, and R. LINDH, "Analytical gradients of a state average MCSCF state and a state average diagnostic," Molecular Physics **99**, 103–114 (2001).

[67]Y. Osamura, Y. Yamaguchi, and H. F. Schaefer, "Generalization of analytic configuration

interaction (CI) gradient techniques for potential energy hypersurfaces, including a solution to the coupled perturbed Hartree–Fock equations for multiconfiguration SCF molecular wave functions," The Journal of Chemical Physics **77**, 383–390 (1982).

[68] M. Nottoli and F. Lipparini, "General formulation of polarizable embedding models and of their coupling," The Journal of Chemical Physics **153**, 224108 (2020).

[69] M. Ben-Nun, J. Quenneville, and T. J. Martínez, "Ab Initio Multiple Spawning: Photochemistry from First Principles Quantum Molecular Dynamics," The Journal of Physical Chemistry A **104**, 5161–5175 (2000).

[70] B. G. Levine, J. D. Coe, A. M. Virshup, and T. J. Martínez, "Implementation of ab initio multiple spawning in the Molpro quantum chemistry package," Chemical Physics Ultrafast Photoinduced Processes in Polyatomic Molecules, **347**, 3–16 (2008).

[71] J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren, and J. W. Ponder, "Tinker 8: Software Tools for Molecular Design," Journal of Chemical Theory and Computation **14**, 5273–5289 (2018).

[72] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," PLOS Computational Biology **13**, e1005659 (2017).

[73] F. Liu, N. Luehr, H. J. Kulik, and T. J. Martínez, "Quantum Chemistry for Solvated Molecules on Graphical Processing Units Using Polarizable Continuum Models," Journal of Chemical Theory and Computation **11**, 3131–3144 (2015).

[74] A. W. Lange and J. M. Herbert, "A smooth, nonsingular, and faithful discretization scheme for polarizable continuum models: The switching/Gaussian approach," The Journal of Chemical Physics **133**, 244111 (2010).

[75] F. Lipparini, L. Lagardère, B. Stamm, E. Cancès, M. Schnieders, P. Ren, Y. Maday, and J.-P. Piquemal, "Scalable Evaluation of Polarization Energy and Associated Forces in Polarizable Molecular Dynamics: I. Toward Massively Parallel Direct Space Computations," Journal of Chemical Theory and Computation **10**, 1638–1651 (2014).

[76] L. E. McMurchie and E. R. Davidson, "One- and two-electron integrals over cartesian gaussian functions," Journal of Computational Physics **26**, 218–231 (1978).

[77] C. Song, L.-P. Wang, and T. J. Martínez, "Automated Code Engine for Graphical Process-

ing Units: Application to the Effective Core Potential Integrals and Gradients," Journal of Chemical Theory and Computation **12**, 92–106 (2016).

[78]S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, S. I. L. Kokkila-Schumacher, N. Luehr, J. W. Snyder, C. Song, A. V. Titov, I. S. Ufimtsev, and T. J. Martínez, "TeraChem: Accelerating electronic structure and ab initio molecular dynamics with graphical processing units," The Journal of Chemical Physics **152**, 224110 (2020).

[79]S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. L. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder Jr., C. Song, A. V. Titov, I. S. Ufimtsev, L.-P. Wang, and T. J. Martínez, "TeraChem: A graphical processing unit-accelerated electronic structure package for large-scale ab initio molecular dynamics," WIREs Computational Molecular Science **11**, e1494 (2021).

[80]L.-P. Wang and C. Song, "Geometry optimization made simple with translation and rotation coordinates," The Journal of Chemical Physics **144**, 214108 (2016).

[81]R. T. Bradshaw and J. W. Essex, "Evaluating Parametrization Protocols for Hydration Free Energy Calculations with the AMOEBA Polarizable Force Field," Journal of Chemical Theory and Computation **12**, 3871–3883 (2016).

[82]M. L. Laury, L.-P. Wang, V. S. Pande, T. Head-Gordon, and J. W. Ponder, "Revised Parameters for the AMOEBA Polarizable Atomic Multipole Water Model," The Journal of Physical Chemistry B **119**, 9423–9437 (2015).

[83]L.-P. Wang, J. Chen, and T. Van Voorhis, "Systematic Parametrization of Polarizable Force Fields from Quantum Chemistry Data," Journal of Chemical Theory and Computation **9**, 452–460 (2013).

[84]X. Mu, Q. Wang, L.-P. Wang, S. D. Fried, J.-P. Piquemal, K. N. Dalby, and P. Ren, "Modeling Organochlorine Compounds and the σ-Hole Effect Using a Polarizable Multipole Force Field," The Journal of Physical Chemistry B **118**, 6456–6465 (2014).

[85]J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," Journal of Computational Chemistry **25**, 1157–1174 (2004).

[86]M. Ruckenbauer, M. Barbatti, T. Müller, and H. Lischka, "Nonadiabatic Photodynamics of a Retinal Model in Polar and Nonpolar Environment," The Journal of Physical Chemistry A **117**, 2790–2799 (2013).

[87] M. Nottoli, R. Nifosì, B. Mennucci, and F. Lipparini, "Energy, Structures, and Response Properties with a Fully Coupled QM/AMOEBA/ddCOSMO Implementation," Journal of Chemical Theory and Computation **17**, 5661–5672 (2021).

[88] J. Liu, "Two more approaches for generating trajectory-based dynamics which conserves the canonical distribution in the phase space formulation of quantum mechanics," The Journal of Chemical Physics **134**, 194110 (2011).

[89] D. Avagliano, M. Bonfanti, M. Garavelli, and L. González, "QM/MM Nonadiabatic Dynamics: The SHARC/COBRAMM Approach," Journal of Chemical Theory and Computation **17**, 4639–4647 (2021).