Introducing QMC/MMpol: Quantum Monte Carlo in polarizable force fields for excited states

Riccardo Guareschi,[†] Habiburrahman Zulfikri,[†] Csaba Daday,[†] Franca Maria Floris,[‡] Claudio Amovilli,[*,‡] Benedetta Mennucci,[‡] and Claudia Filippi[*,†]

MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, and Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Giuseppe Moruzzi 13, 56124 Pisa, Italy

E-mail: claudio.amovilli@unipi.it; c.filippi@utwente.nl

Abstract

We present for the first time a quantum mechanics/molecular mechanics scheme which combines quantum Monte Carlo with the reaction field of classical polarizable dipoles (QMC/MMpol). In our approach, the optimal dipoles are self-consistently generated at the variational Monte Carlo level and then used to include environmental effects in diffusion Monte Carlo. We investigate the performance of this hybrid model in describing the vertical excitation energies of prototypical small molecules solvated in water, namely, methylenecyclopropene and *s*-trans acrolein. Two polarization regimes are explored where either the dipoles are optimized with respect to the ground-state solute density (polGS) or different sets of dipoles are separately brought to equilibrium with the states involved in the electronic transition (polSS). By comparing with reference supermolecular calculations where both solute and solvent are treated quantum mechanically, we find that the inclusion of the response of the environment to the

---

[*]To whom correspondence should be addressed
[†]MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands
[‡]Dipartimento di Chimica e Chimica Industriale, Università di Pisa, Via Giuseppe Moruzzi 13, 56124 Pisa, Italy

excitation of the solute leads to superior results than the use of a frozen environment (point charges or polGS), in particular, when the solute-solvent coupling is dominated by electrostatic effects which are well recovered in the polSS condition. QMC/MMpol represents therefore a robust scheme to treat important environmental effects beyond static point charges, combining the accuracy of QMC with the simplicity of a classical approach.

1 Introduction

The effective treatment of complex processes of chemical and biological relevance imposes practical choices in the construction of the theoretical model. The system is divided into an active region and the surrounding environment, and the two parts are treated at a different level of accuracy. A liquid solution is a typical application of such an approach: One or more active molecules (the solute) are embedded in an external potential generated by the solvent. Within this framework, the challenge is to combine the accuracy of the model with the increasing computational requirements due to the high number of interactions inside the different regions and between them. The most common computational approaches resort to the use of quantum mechanics (QM) to treat the active region through orbital- or density-based theories, and classical molecular mechanics (MM) for the environment. Originally proposed by Warshel and Levitt[1] in 1976, such combination of a QM description with a classical one has gained increasing popularity through the years[2–7] and represents now a standard approach to describe molecular systems of high complexity.

Within quantum-in-classical methods, different formulations can be adopted to describe the classical component either preserving its discrete molecular structure[8,9] or through a coarser treatment as a continuum medium.[10] If we focus on the more realistic discrete methods, the simplest way to simulate the environmental effects is to assign partial atomic charges to the classical atoms, which create a perturbation on the active subsystem through a static external electric field. While its most common use is for ground-state problems, this approach has also been widely employed to compute excitation properties of (bio)molecules in complex environments with a varying degree of success. Several recent studies of the absorption properties of small organic molecules in solutions

as well as photo-active proteins[11–25] have in fact provided substantial evidence that the choice of a static point-charges embedding scheme may qualitatively reproduce some experimental findings but appears inadequate if one strives for higher accuracy.

To cope with these limitations and simulate a more realistic interaction between the environment and the photo-active site, a more accurate embedding scheme is required which takes into account mutual polarization effects both in the ground state and in response to the excitation of the embedded quantum region (as long as a partitioning scheme with fixed number of electrons is applicable to the problem of interest). This can be achieved by describing the classical atoms not only with partial atomic charges but also with induced dipoles determined by the atomic polarizabilities and the total electric field of the system with contributions from the embedded electronic density, the static electric field of the point charges and the other induced dipoles in the environment. In this scheme, the QM subsystem is polarized by an MM region which can in turn respond to the presence of the embedded molecule in its ground or excited state.[26–29] Such a polarizable embedding (QM/MMpol) scheme has been combined with coupled cluster theory (CC),[23,26,30,31] the complete-active-space self-consistent field (CASSCF) method,[32] and time-dependent density functional theory (TDDFT)[27,28,33,34] and has provided a very effective description of excitation energies beyond static multipole models,[23,34–37] allowing the systematic study of different polarization effects induced in the ground state and in response to the excitation of the solute.[15,25]

Here, we combine for the first time quantum Monte Carlo (QMC) methods with the reaction field of polarizable dipoles and investigate the performance of QMC/MMpol for the computation of the excitation energies of small solvated molecules, namely, methylenecyclopropene and acrolein in water. The use of QMC to compute electronic excitations has already been extensively validated in the gas phase[21,38–41] and also employed in combination with standard MM methods[17,21,25] and DFT embedding.[42] To explore the robustness of the QMC/MMpol scheme, we compare the excitation energies obtained in the presence of static and polarizable environments with the results of supermolecular calculations, where both solute and solvent are treated at the QM level. Alongside QMC, we also analyse the response to the different embedding conditions of other quantum me-

chanical methods for the treatment of the solute, namely, time-dependent density functional theory and the complete-active space second-order perturbation theory (CASPT2).

The paper is organized as follows. In Sec. 2, we discuss the theoretical background and the methodological aspects of the computation of the QMC/MMpol excitations and, in Sec. 3, give the computational details. In Sec. 4, we present the complete analysis of our results for methylenecyclopropene and acrolein, and conclude in Sec. 5.

2   Methods

In the QM/MMpol approach presented here, the solute and the solvent are treated at different levels of the theory, namely the solute is described quantum mechanically while the solvent is represented by explicit molecules treated with a classical model which consists of permanent atomic partial charges and atomic polarizabilities. The polarizabilities are used to determine the induced dipoles on the atoms of the solvent molecules.

Thus, the Schrödinger equation for the solute wave function becomes

$$(\hat{\mathscr{H}}_{\text{vac}} + \hat{\mathscr{V}})\Psi = E\Psi \tag{1}$$

where the vacuum Hamiltonian describes the isolated solute and the second term the interaction between the solute and the classical solvent. Since the MM model has both static and polarization contributions, we have

$$\hat{\mathscr{V}} = \sum_{\text{m}}^{N_{\text{ch}}} q_{\text{m}} \hat{V}_{\text{solute}}(\mathbf{r}_{\text{m}}) - \sum_{\text{a}}^{N_{\text{dip}}} \boldsymbol{\mu}_{\text{a}} \cdot \hat{\mathbf{E}}_{\text{solute}}(\mathbf{r}_{\text{a}}), \tag{2}$$

where $\hat{V}_{\text{solute}}(\mathbf{r}_{\text{m}})$ and $\hat{\mathbf{E}}_{\text{solute}}(\mathbf{r}_{\text{a}})$ are the electrostatic potential and electric field operators generated by the electrons and the nuclei at the coordinate $\mathbf{r}_{\text{m}}$ of the point charge q_{m} and $\mathbf{r}_{\text{a}}$ of the dipole $\boldsymbol{\mu}_{\text{a}}$, respectively.

The solvent contribution to the energy is given by

$$
\begin{aligned}
U_{\text{solvent}} \quad = \quad & \frac{1}{2} \sum_{n}^{N_{\text{ch}}} \sum_{m \neq n}^{N_{\text{ch}}} \frac{q_m q_n}{|\mathbf{r}_m - \mathbf{r}_n|} - \sum_{a}^{N_{\text{dip}}} \boldsymbol{\mu}_a \cdot \mathbf{E}_q(\mathbf{r}_a) \\
+ \quad & \frac{1}{2} \sum_{a}^{N_{\text{dip}}} \sum_{b \neq a}^{N_{\text{dip}}} \boldsymbol{\mu}_a \cdot \mathbf{T}_{ab} \cdot \boldsymbol{\mu}_b + \frac{1}{2} \sum_{a}^{N_{\text{dip}}} \frac{\mu_a^2}{\alpha_a}.
\end{aligned}
\tag{3}
$$

where α_a is the isotropic polarizability at site a, $\mathbf{E}_q(\mathbf{r}_a)$ the static electric field generated by the point charges of the solvent at the sites of the dipoles,

$$
\mathbf{E}_q(\mathbf{r}_a) = \sum_{b \neq a}^{N_{\text{ch}}} f_1(r_{ab}) \frac{q_m (\mathbf{r}_a - \mathbf{r}_b)}{|\mathbf{r}_a - \mathbf{r}_b|^3},
\tag{4}
$$

and the dipole field tensor $\mathbf{T}_{ab}$ is defined as

$$
(\mathbf{T}_{ab})_{ij} = \frac{f_1(r_{ab})}{r_{ab}^3} \left[\delta_{ij} - 3 f_2(r_{ab}) \frac{(\mathbf{r}_{ab})_i (\mathbf{r}_{ab})_j}{r_{ab}^2} \right],
\tag{5}
$$

with $\mathbf{r}_{ab} = \mathbf{r}_a - \mathbf{r}_b$ and $r_{ab} = |\mathbf{r}_{ab}|$. The first three terms in the solvent energy are the charge-charge, dipole-charge, and dipole-dipole interaction energies while the last term is the self-energy cost to induce the dipoles. In the last two equations, we have introduced two screening functions following the Thole linear model[43–45] to prevent divergences in the energy when two MM sites are too close. The two functions are defined as

$$
\begin{aligned}
f_1(r_{ab}) \quad &= \quad 4 \left(\frac{r_{ab}}{c} \right)^3 - 3 f_2(r_{ab}) \\
f_2(r_{ab}) \quad &= \quad \left(\frac{r_{ab}}{c} \right)^4,
\end{aligned}
\tag{6}
$$

for $r_{ab}/c \leq 1$ with $c = k(\alpha_a \alpha_b)^{1/6}$, and are identical to one otherwise. **The dimensionless parameter k is optimized together with the atomic polarizabilities in Ref.[46] to reproduce the experimental molecular polarizabilities of an extensive set of compounds**. In this work, we consider water as solvent and interactions between charges and dipoles belonging to the same

water molecule are excluded.

For a given solute wave function, Ψ, the energy in the presence of this polarizable solvent model is given by

$$E[\Psi, \boldsymbol{\mu}] = \langle \Psi | \hat{\mathscr{H}}_{\text{vac}} + \hat{\mathscr{V}} | \Psi \rangle + U_{\text{solvent}}, \tag{7}$$

where the formalism so far has not assumed that the solute and the induced dipoles are in equilibrium, so wave function and dipoles are treated as independent entries in the energy expression. For a given wave function, the induced dipoles in equilibrium with the solute can then be obtained as the dipoles that minimize the energy $E[\Psi, \boldsymbol{\mu}]$. This minimization leads to the expected relation between the induced dipoles and the total electric field:

$$\boldsymbol{\mu}_{\text{a}} = \alpha_{\text{a}} [\mathbf{E}_{\text{static}}(\mathbf{r}_{\text{a}}) - \sum_{\text{b} \neq \text{a}}^{\text{N}_{\text{dip}}} \mathbf{T}_{\text{ab}} \cdot \boldsymbol{\mu}_{\text{b}}], \tag{8}$$

where the static field due to the solute charge density and the solvent permanent charge distribution is given by

$$\mathbf{E}_{\text{static}}(\mathbf{r}_{\text{a}}) = \langle \Psi | \hat{\mathbf{E}}_{\text{solute}}(\mathbf{r}_{\text{a}}) | \Psi \rangle + \mathbf{E}_q(\mathbf{r}_{\text{a}}), \tag{9}$$

with

$$\langle \Psi | \hat{\mathbf{E}}_{\text{solute}}(\mathbf{r}_{\text{a}}) | \Psi \rangle = \int \rho_{\text{solute}}(\mathbf{r}) \frac{(\mathbf{r}_{\text{a}} - \mathbf{r})}{|\mathbf{r} - \mathbf{r}_{\text{a}}|^3} d\mathbf{r}, \tag{10}$$

and ρ_{solute} the solute nuclear and electronic charge density.[47] This set of equations defines the optimal dipoles and can be rewritten as

$$\mathbf{A}\boldsymbol{\mu} = \mathbf{E}_{\text{static}} \tag{11}$$

where $\boldsymbol{\mu}$ and $\mathbf{E}_{\text{static}}$ are vectors of dimension 3N_{dip}, which contain all the dipoles and the static electric field at the corresponding positions, respectively, while the $3\text{N}_{\text{dip}} \times 3\text{N}_{\text{dip}}$ matrix $\mathbf{A}$ only

depends on the values of α_a and the positions of the solvent molecules.

Importantly, since the dipoles $\boldsymbol{\mu}$ and the wave function Ψ are mutually dependent (Eqs. 8 and 9), they must be brought to self consistency, a procedure which results in a different sets of induced dipoles for the ground and the excited states. In other words, the Schrödinger equation (Eq. 1) coupling the solute to classical polarizable dipoles is non-linear being characterized by a potential $\hat{\mathscr{V}}$ which depends on the wave function quadratically through the induced dipoles (Eq. 8). Solving this equation leads to a set of non-orthogonal wave functions in equilibrium with their corresponding dipoles. When wave function and dipoles are in equilibrium, the energy of the solute and the solvent (Eq. 7) becomes

$$
\begin{aligned}
E[\Psi, \boldsymbol{\mu}(\Psi)] \;=\; & \langle\Psi|\,\mathscr{H}_{\text{vac}} + \sum_{\text{m}}^{N_{\text{ch}}} q_{\text{m}} \hat{V}_{\text{solute}}(\mathbf{r}_{\text{m}})\,|\Psi\rangle \\
& - \frac{1}{2}\sum_{\text{a}}^{N_{\text{dip}}} \boldsymbol{\mu}_{\text{a}}(\Psi) \cdot [\langle\Psi|\,\hat{\mathbf{E}}_{\text{solute}}(\mathbf{r}_{\text{a}})\,|\Psi\rangle + \mathbf{E}_q(\mathbf{r}_{\text{a}})] \\
& + \frac{1}{2}\sum_{\text{n}}^{N_{\text{ch}}}\sum_{\text{m}\neq\text{n}}^{N_{\text{ch}}} \frac{q_{\text{m}}q_{\text{n}}}{|\mathbf{r}_{\text{m}}-\mathbf{r}_{\text{n}}|},
\end{aligned}
\tag{12}
$$

where the dipoles $\boldsymbol{\mu}(\Psi)$ are determined through equation (Eq. 8). It is important to stress that this simplified expression for the energy does not hold when the dipoles and the electric field of the solute are not in equilibrium. This occurs for instance when one computes an excitation energy in the presence of dipoles which are in equilibrium with one of the two states of interest: One solves the Schrödinger equation for state I to determine Ψ_I and $\boldsymbol{\mu}(\Psi_I)$ and computes the other state J in the potential $\mathscr{V}$ of the fixed dipoles $\boldsymbol{\mu}(\Psi_I)$ to obtain the excitation energy as $\Delta E = E[\Psi_J, \boldsymbol{\mu}(\Psi_I)] - E[\Psi_I, \boldsymbol{\mu}(\Psi_I)]$, where the solvent contribution U_{solvent} cancels out in this energy difference since the dipoles are the same for both states. Another non-equilibrium situation is encountered when we optimize a wave function in the presence of dipoles which are optimal for the initial starting wave function. Then, we must employ the general expressions for the coupling of the solute wave function to a set of non-equilibrium induced dipoles (Eqs. 2, 3, and 7).

In this work, we focus on the computation of the excitation energies of a solvated molecule

in such a classical polarizable model using QMC to describe the solute, and consider two different polarization regimes. In the first case, the induced dipoles are in equilibrium with the solute ground-state density and used in the computation of the ground and excited states (polGS). In the second case, state-specific dipoles are generated with respect to the ground- and the excited-state static electric field separately (polSS). To this aim, two parallel routes (I and II) are followed to obtain the polGS and polSS excitation energies according to the steps outlined below, where we denote with the subscripts 0 and 1 the ground and the excited state and with the superscripts "g" and "e" whether the wave functions are computed in the potential of the induced dipoles $\boldsymbol{\mu}^g$ and $\boldsymbol{\mu}^e$ polarized to the ground- and the excited-state solute density, respectively. Starting from iteration $i = 1$, we have

Ia) Determine $\boldsymbol{\mu}^g(i-1)$ in equilibrium with $\Psi_0^g(i-1)$.

Ib) Optimize $\Psi_0^g(i)$ and $\Psi_1^g(i)$ in the presence of the ground-state dipoles $\boldsymbol{\mu}^g(i-1)$.

Ic) Return to Ia.

At self-consistency, we compute the polGS excitation energy as

$$\Delta E_{\text{polGS}} = E\left[\Psi_1^g, \boldsymbol{\mu}^g\right] - E\left[\Psi_0^g, \boldsymbol{\mu}^g\right]. \tag{13}$$

For the polSS calculation,

IIa) Determine $\boldsymbol{\mu}^e(i-1)$ in equilibrium with $\Psi_1^e(i-1)$.

IIb) Optimize $\Psi_0^e(i)$ and $\Psi_1^e(i)$ in the presence of the excited-state dipoles $\boldsymbol{\mu}^e(i-1)$.

IIc) Return to IIa.

The polSS excitation energy is then obtained as

$$\Delta E_{\text{polSS}} = E\left[\Psi_1^e, \boldsymbol{\mu}^e\right] - E\left[\Psi_0^g, \boldsymbol{\mu}^g\right]. \tag{14}$$

The starting wave functions $\Psi_0(0)$ and $\Psi_1(0)$ are obtained in the presence of the static classical point charges and, at each iteration, the wave functions are optimized in a state-average (SA) fashion since the systems considered here have no symmetry. We note that, while in the polGS excitation energy the solvent energy $U_{solvent}$ cancels out as already mentioned above, this is not the case in the polSS energy difference since two different sets of dipoles are employed.

To describe the quantum solute, we use both the variational (VMC) and diffusion Monte Carlo (DMC) methods. We employ many-body molecular wave functions of the Jastrow-Slater form,

$$\Psi = \mathscr{J} \sum_{i=1}^{N_{CSF}} c_i C_i, \tag{15}$$

where c_i are the expansion coefficients of the configuration state functions (CSFs), C_i, and $\mathscr{J}$ is the Jastrow correlation factor which has an explicit dependence on the inter-particle distances. In VMC, the energy of the electronic states is computed according to Eq. 7 where the expectation values are calculated via Monte Carlo sampling of the square of the wavefunction, Ψ^2. For the given wave function, the average electric field due to the solute sampled at the dipole sites is then used to compute the induced dipoles (Eq. 8). The Jastrow-Slater wave functions are fully optimized by energy minimization in VMC[48] in a state-average fashion,[38] using $\sum_I |\Psi_I|^2$ as square of the guiding wave function.

After we have followed schemes I and II above to obtain the self-consistent wave functions and dipoles in VMC, we perform DMC calculations to compute the polGS and polSS excitations energies, where we employ Ψ_0^g and Ψ_1^g as trial wave functions in the potential generated by $\boldsymbol{\mu}^g$, and Ψ_1^e in combination with $\boldsymbol{\mu}^e$. For the molecules considered here, the VMC and DMC excitation energies are converged after two wave function-dipole optimization cycles as reported for the methylenecyclopropene molecule in polarizable water solvent in Table 1.

Table 1: Convergence of the polGS and polSS excitation energies (eV) obtained in VMC and DMC with the number of wave function-dipole optimization cycles for the $\pi \rightarrow \pi^*$ state of methylenecyclopropene in water. The statistical error on the QMC values is given in brackets.

Method	iter. no.	E_{exc} (eV)	
		VMC	DMC
QM/polGS	1	5.138(3)	5.120(7)
	2	5.134(3)	5.119(7)
	3	5.128(3)	5.110(7)
QM/polSS	1	5.010(3)	4.999(7)
	2	5.012(3)	5.006(7)
	3	5.015(3)	5.009(7)

3   Computational details

The QMC calculations are carried out with the program package CHAMP.[49] We employ scalar-relativistic energy-consistent Hartree-Fock pseudopotentials[50,51] and obtain the starting determinantal component of the Jastrow-Slater wave function in complete-active-space self-consistent field (CASSCF) calculations performed with the program GAMESS(US).[52,53] For acrolein, we truncate the CAS expansion expressed on the state-average natural orbitals with an appropriate threshold of 0.03 on the coefficients of the CSFs (the total CASSCF weight of all states is greater than 98.5%), and retain the union set of surviving CSFs for the states of interest. We use instead the complete CAS space for methylenecyclopropene (MCP). We employ a two-body Jastrow factor ($\mathcal{J}_{2-\text{body}}$) to account for electron-nuclear and electron-electron correlations and use different Jastrow factors to describe different atom types.[54] The effect of using a three-body Jastrow factor ($\mathcal{J}_{3-\text{body}}$) including electron-electron-nucleus correlation is also tested. The QMC wave functions are fully optimized within VMC using the linear method[48] and its extension to state-average calculations,[38] and used to compute excitation energies at both the VMC and the DMC level. We treat the pseudopotentials beyond the locality approximation[55] and use an imaginary time step of 0.075 a.u. in the DMC supermolecular calculations and of 0.05 a.u. in all other DMC runs. In order to prevent divergences of the electrostatic interactions at the coalescence points, we introduce a small cutoff distance of 0.04 a.u. and, when an electron is closer than this distance to the site of a point charge or a dipole, we set the interaction equal to the value it takes at the cutoff. For historical rea-

10

sons, the QMC calculations with point charges only employ charges which are screened according to the function $v_q(r) = q[1 - \exp(-Br^2)]/r$.

We use the MOLCAS 7.4 suite of programs[56] for the CASPT2[57,58] calculations and always employ the recommended zero-order Hamiltonian[59] with the IPEA shift set to 0.25 a.u. We use an additional constant imaginary shift[60] of 0.1 a.u. and, unless otherwise noted, we report the excitation energies computed at the single-state level. For direct comparison with the QMC results, we employ pseudopotentials[50,51] also in the MOLCAS calculations. The CASPT2/MMpol results are computed in a two-step procedure recently presented,[32] where the MM induced dipoles and charges from an all-electron CASSCF/MMpol calculation with the Gaussian code[61] are later used in MOLCAS as a static external potential to obtain the CASPT2 results. Such CASSCF/MMpol calculations are performed using a locally modified version of Gaussian09, revision A.02, and are obtained using the state-average procedure for the ground and excited states and adapting the MM polarization either to the ground state (polGS), or to the ground and excited states in two separate calculations (polSS). All-electron CASPT2 calculations are reported in the SI, where we freeze as many σ orbitals as there are heavy atoms.

We perform the TDDFT calculations using a locally modified version of Gaussian09, revision A.02[61] and the CAM-B3LYP[62] exchange-correlation functional. The range-separated hybrid functional CAM-B3LYP is chosen to limit possible spurious charge-tranfer effects in the supermolecular calculations.

In the QM/MMpol calculations, the water molecules are described by point charges derived from a standard restrained electrostatic potential (RESP) fit using the electrostatic potential computed at the MP2/aug-cc-pVTZ level on the TIP3P water geometry ($q_O = -0.726$, $q_H = 0.363$ a.u.)[63] and by isotropic atomic polarizabilities taken from the Amber pol12 force field (AL model in Ref.[46,64]). The parameter k (Eq. 6) is set to 2.5874 (AL model in Ref.[46]). We also use the non-polarizable TIP3P model ($q_O = -0.834$ and $q_H = 0.417$ a.u.).[65] The screening parameters in the QMC calculations with point charges are $B_O = 0.52548$ and $B_H = 2.73465$ a.u.
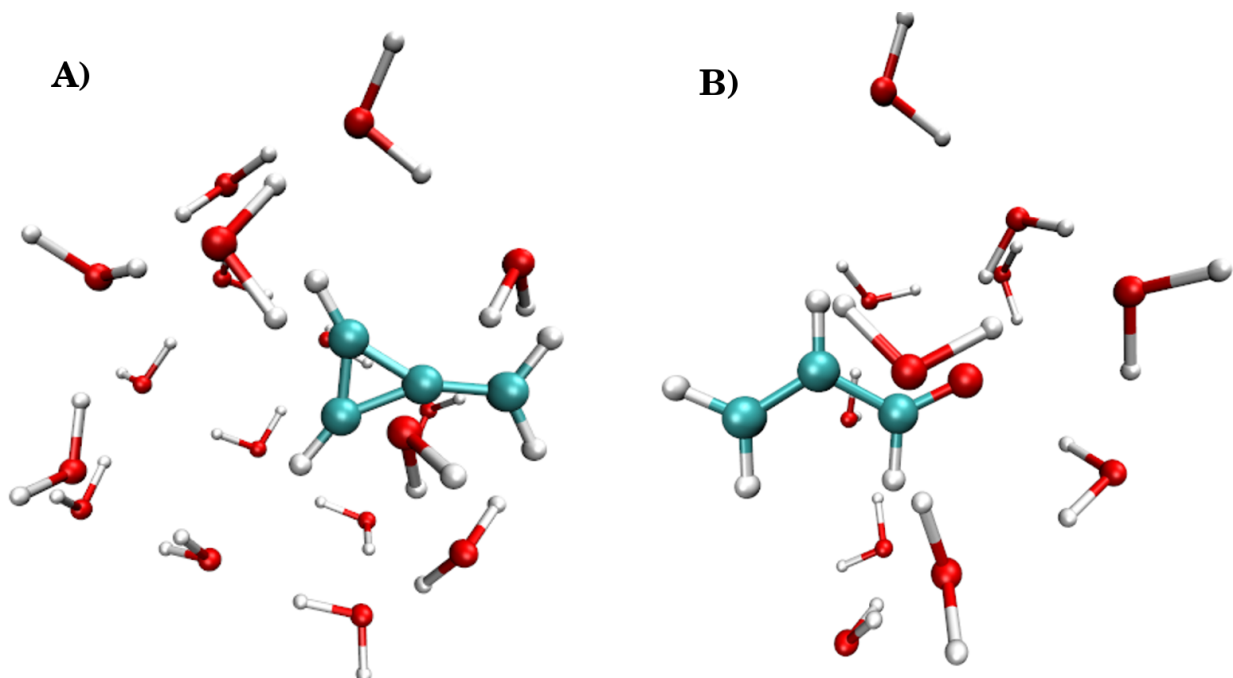
Figure 1: Cluster models of solvated (A) methylenecyclopropene and (B) *s*-trans acrolein in water.

3.1 Cluster models

In this study, we use small water clusters of solvated methylenecyclopropene (MCP) and *s*-trans acrolein. This allows us to perform supermolecular reference calculations of the excitation energies which can be used to assess the quality of the results obtained with different embedding approaches.

We generate the cluster models by solvating MCP and acrolein in approximately one shell of water molecules (17 and 19 molecules, respectively) and optimize the geometries with Gaussian09 in DFT with the BLYP exchange-correlation functional[66,67] and the cc-pVDZ basis set. For acrolein, due to the high number of tests performed to ascertain the supermolecular value of the excitation energy, we reduce the size of the cluster and retain the 11 water molecules closer to the carbonyl oxygen of acrolein. The final cluster models used in this study are shown in Figure 1.

3.2 Choice of basis set and Jastrow factor in QMC

In the QMC calculations, we employ the Gaussian basis sets[50,68] specifically constructed for our pseudopotentials. In Table 2, we assess the basis-set convergence of the QMC excitation energies of MCP isolated and in TIP3P water using the cc-pVTZ′, aug-cc-pVDZ′, and aug-cc-pVTZ′ basis sets, where the prime indicates that cc-pVDZ is employed for hydrogen. We find that the use of augmentation significantly accelerates the convergence of the VMC and DMC excitation energies and that the aug-cc-pVDZ′ basis set yields results at the DMC level that are only 0.02–0.03 eV higher than those obtained with the larger aug-cc-pVTZ′. In Table 2, we also investigate the effect of including three-body terms in the Jastrow factor and find that it generally leads to higher excitation energies, the shift being larger in the isolated system **and in combination with the aug-cc-pVDZ′ basis set. The error with respect to the triple-ζ result is about 0.05 eV at the VMC level and the basis-set convergence of the excitation energy therefore appears to be slower in the presence of a three-body Jastrow factor.** Consequently, all QMC results presented below are obtained with the aug-cc-pVDZ′ basis set and the two-body Jastrow factor, which offer a good compromise between accuracy and computational cost.

Table 2: Basis-set convergence of the VMC and DMC excitation energies (eV) of the $\pi \to \pi^*$ state of MCP isolated and in TIP3P water. Either a two- or a three-body Jastrow factor is employed. The statistical error on the QMC values is given in brackets.

	pVTZ′	aug-cc-pVDZ′		aug-cc-pVTZ′	
	$\mathcal{J}_{2-\text{body}}$	$\mathcal{J}_{2-\text{body}}$	$\mathcal{J}_{3-\text{body}}$	$\mathcal{J}_{2-\text{body}}$	$\mathcal{J}_{3-\text{body}}$
VMC/isolated	4.265(1)	4.179(1)	4.225(1)	4.161(1)	4.170(1)
DMC/isolated	4.229(1)	4.159(1)	4.185(1)	4.141(1)	4.151(1)
VMC/TIP3P	5.010(0)	4.889(1)	4.908(1)	4.857(0)	4.862(1)
DMC/TIP3P	4.963(1)	4.862(1)	4.867(1)	4.831(1)	4.829(1)

In the QMC supermolecular calculations, we employ the aug-cc-pVDZ′ basis set for the solute and the closest water molecules and describe the remaining solvent molecules with the cheaper cc-pVDZ on the oxygens and cc-VDZ basis set (without p functions) on the hydrogens. In the MCP and acrolein clusters, we identify eight and five close water molecules, respectively, so that the use of different basis sets for the rest of the solvent affects the CASSCF excitation energies to

less than 0.01 eV. In the wave function optimization, we localize the orbitals and do not to optimize those of the far water molecules to limit the computational cost.

3.3   Basis-set convergence in CASPT2 and TDDFT

We test the basis set convergence of the CASPT2 and TDDFT excitation energies for MCP using the Dunning's correlation consistent cc-pVXZ and aug-cc-pVXZ series up to quintuple and quadruple ζ, respectively.[69–72] As shown in Figure S1 for MCP isolated and in water, the inclusion of augmentation is very important (more so than in the QMC calculations) to accelerate the basis set convergence, and the excitation energies computed with the aug-cc-pVDZ basis are compatible with the aug-cc-pVQZ and cc-pV5Z values within 0.02-0.03 eV at both level of theory.

Therefore, we select the aug-cc-pVDZ as default basis set for the computation of the TDDFT excitation energies. For CASPT2, we compare the excitation energies computed with pseudopotentials in combination with the aug-cc-pVDZ′ basis set described above to the all-electron values obtained with the Dunning's aug-cc-pVDZ basis. As reported in Table S4, the resulting excitation energies are compatible within 0.04 eV for MCP and the n $\rightarrow$ π^* excitation of acrolein, while the difference is of the order of 0.1 eV for the $\pi \rightarrow \pi^*$ excitation energy of acrolein (see Table S4). Surprisingly, in all three cases, the CASSCF values computed with and without pseudopotential agree to better than 0.02 eV. To allow a direct comparison with the QMC results, we report below the CASPT2 values obtained with pseudopotentials and the aug-cc-pVDZ′ basis set.

4   Results

We investigate here how the MMpol description of the environment affects the excitation energies computed with QMC for methylenecyclopropene and s-trans acrolein solvated in water. We treat the environment using two different MMpol schemes: In the polGS one, the environment is polarized only to the ground state and frozen in the computation of the excitation energy while, in the polSS approach, we include the response of the solvent to the excitation of the solute, equilibrating

the dipoles with the ground- and excited-state density separately for each state. As elaborated in the Method Section, the resulting state-specific dipoles are then employed to compute the polSS excitation energy and to account in this manner for differential polarization effects. The polGS and polSS treatments of the environment capture the purely electrostatic interaction of the solute with the solvent, so possible discrepancies with the reference can be attributed not only to the approximate nature of the model but also to the incomplete description of solute-solvent coupling.

To understand the interplay between these environmental effects, the limitation of the embedding model, and the choice of quantum method for the solute, we employ TDDFT and CASPT2 in addition to QMC to compute the polGS and polSS excitation energies. Furthermore, we obtain the TDDFT excitation energies also in the linear response regime (polLR) of MMpol, where the environment does relax upon excitation of the solute but the induced dipoles respond to the transition density instead of the density difference between the ground and the excited state. The polLR model has been shown to miss the electrostatic response of the environment described in the polSS approximation but to capture the solvent polarization oscillating at the frequency of the solute excitation.[73] Both terms should however be present to describe the full environmental response and the analysis of the difference in the polSS and polLR responses at the TDDFT level will also help us to unravel the performance of the MMpol model in combination with QMC and other wave function approaches. **Finally, we also compute the excitation energies of the solute at all levels of theory in a non-polarizable solvent of only point charges. We consider two different set of charges, namely, the point charges of the TIP3P water model and those (denoted as "nopol") from the polarizable force field, which we also employ in the MMpol calculations but use in this case without the corresponding polarizabilities.**

We focus here on the vertical excitation energies of the $\pi \rightarrow \pi^*$ state of MCP and the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ states of acrolein. From absorption experiments in different solvents, we infer that water solvent induces a red-shift in the $\pi \rightarrow \pi^*$ excitation energy of acrolein with respect to the isolated value and a blue-shift for the other two states.[74,75] Since we only consider one geometrical configuration for each system, we expect a qualitatively similar behavior of the excitation energies

15

Table 3: Excitation energies (eV) of the $\pi \to \pi^*$ state of MCP isolated and solvated in a water cluster. The statistical error on the QMC values is given in brackets.

Method	isolated	TIP3P	nopol	polGS	polSS	polLR	super
VMC	4.179(1)	4.889(1)	4.800(3)	5.134(3)	5.012(3)	–	5.07(2)
DMC	4.159(1)	4.862(1)	4.773(2)	5.119(7)	5.006(7)	–	5.04(2)
CASPT2	4.133	4.902	4.801	5.045	4.928	–	4.978
TDDFT	4.224	5.017	4.916	5.190	5.077	5.173	5.187

computed on our cluster models but a detailed comparison with absorption experiments in solution is outside the scope of this paper.

4.1 Methylenecyclopropene

MCP is a cross-conjugated molecule with interesting electrostatic features in the ground and first $\pi \to \pi^*$ excited states. Although without polar bonds with heteroatoms, MCP displays a relatively large dipole moment in the ground state (with an experimental value of about -1.9 Debye[74,76]) being characterized by an accumulation of π-electron density on the exocyclic double bond. The $\pi \to \pi^*$ excited state exhibits strong charge transfer from this double bond to the molecular ring, and the direction of the molecular dipole moment is reversed in going from the ground to the ex- cited state. Consequently, a polar solvent in equilibrium with the ground-state solute is expected to destabilize the $\pi \to \pi^*$ excited state and to induce a blue-shift of the vertical excitation with respect to the gas-phase value.[77] Given the dramatic change in electronic structure upon excitation, we an- ticipate a strong response of the excitation energy to changes in the description of the environment. In particular, accounting for differential polarization effects via state-specific embedding should lead to a significant improvement on the use of a frozen (non-polarizable or polGS) solvent.

Our findings on the MCP-water model described above are summarized in Table 3 and Figure 2. In all calculations, the $\pi \to \pi^*$ excited state is always the second root and, at the correlated level, we employ a reference minimal CAS expansion correlating four electrons in two bonding and two anti-bonding π orbitals. In line with the physical picture given above, we find that all methods predict a large blue-shift in the excitation energy (about 0.7–0.8 eV) from isolated MCP to MCP

solvated in TIP3P water. The use of different point charges extracted from the polarizable force field (nopol) results in a rather similar trend. Polarizing the dipoles to the ground state of the solute (polGS) induces an additional significant blue-shift since a frozen environment optimal in the ground state further destabilizes the excited state. The shift with respect to nopol is somewhat larger in QMC and of the order of 0.35 eV. As expected, accounting for differential polarization effects via polSS reverses the trend, lowering the excitation energies with respect to the polGS values. Given the greatly different nature of the two states, the polSS correction to polGS is significant (about 0.1 eV) in all methods and somewhat overshoots the supermolecular values. Nevertheless, the use of state-specific embedding is crucial to reach an agreement as good as 0.05 eV with the supermolecular excitation energies at the correlated level.
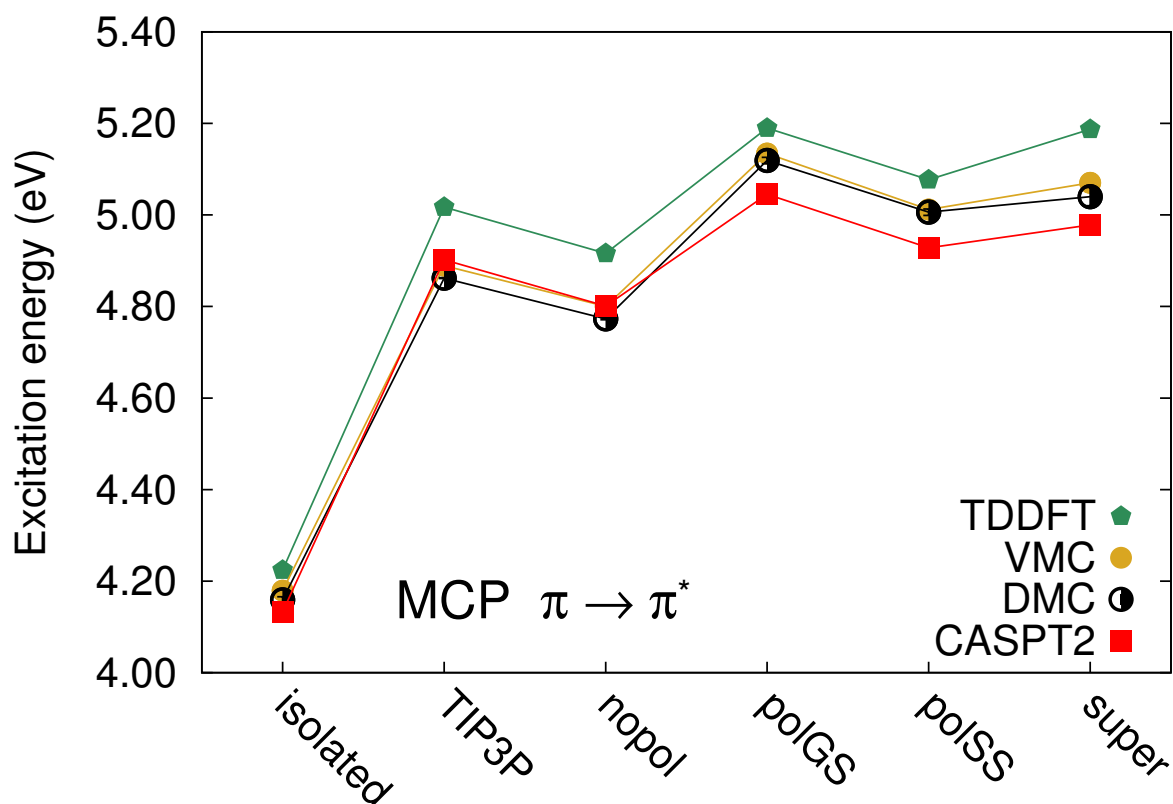

Figure 2: Excitation energies of the $\pi \rightarrow \pi^*$ state of MCP with different quantum methods and embedding schemes.

We find that QMC yields excitation energies in very good agreement with the CASPT2 val-

ues in non-polarizable embedding while the two quantum methods predict a different shift from static point charges to polGS, and maintain hereinafter a similar behavior at the polSS and supermolecular levels. For TDDFT, the excitation energies computed with the CAM-B3LYP exchange-correlation functional respond less strongly to polGS than the QMC ones, with a shift from nopol somewhat closer to the CASPT2 value. Finally, we observe that the inclusion of the solvent relaxation in linear response (polLR) yields an excitation energy only slightly lower than the polGS value. The smaller effect with respect to a state-specific electrostatic treatment is to be expected since the response of the dipoles in the polLR model is proportional to the transition dipole moment which in the $\pi \rightarrow \pi^*$ transition of MCP is rather small.

4.2 *s*-trans acrolein

As in the case of MCP, the n $\rightarrow \pi^*$ excited state of acrolein has charge-transfer character with the electronic density moving from the oxygen to the carbon skeleton upon excitation. This flux of charge is responsible for the reduction of the excited-state molecular dipole moment with respect to the ground state. Therefore, the vertical n $\rightarrow \pi^*$ excitation energy blue-shifts in a polar environment with respect to the isolated molecule since the excited state is destabilized with respect to the ground state.[78] Again, we expect a significant difference between the excitation energies computed within a polarizable and a frozen environment and, in particular, the relaxation of the excited-state dipoles in the polSS embedding scheme should have a strong effect on the excitation. Since the n $\rightarrow \pi^*$ transition is symmetry forbidden in the gas phase and retains a small transition dipole moment also in water, we do not expect polLR to considerably affect the excitation energies with respect to the polGS treatment.

We also consider the $\pi \rightarrow \pi^*$ singlet transition, which has rather a different nature than the n $\rightarrow \pi^*$ one, being dipole allowed and associated with an increment of the molecular dipole moment. Therefore, the $\pi \rightarrow \pi^*$ transition is red-shifted in polar environments with respect to the isolated value.[79] In addition to a polSS correction on the polGS excitation energy due to the non-negligible change in molecular dipole moment between the ground and excited states, we expect

that, when computing the excitations within TDDFT, polLR should recover sizable polarization effects considering the bright nature of the excited state and the correspondingly large transition dipole moment (see Table S6).

We describe the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitation energies of acrolein at the correlated level using a CAS(6,10) expansion, which includes two bonding and six anti-bonding π orbitals on the molecular chain, the σ lone pair localized on the carbonyl oxygen and its anti-bonding σ^* orbital. This large active space allows a balanced treatment of the two excited states at the same time in a SA-CASSCF calculation. In particular, the choice of the CAS(6,10) reference active space is crucial to give a proper description of the $\pi \rightarrow \pi^*$ excited state, which shows a slow convergence with the size of the active space (see table S3).

While the $n \rightarrow \pi^*$ state is always the second root of the SA-CASSCF wavefunction, the bright $\pi \rightarrow \pi^*$ state in the isolated system is the fifth root and becomes the fourth state after the perturbation correction is introduced in CASPT2. The same behavior is observed in QMC after the inclusion of dynamic correlation through the Jastrow factor. The presence of a solvent simplifies the treatment of the $\pi \rightarrow \pi^*$ state, which becomes the third root at the SA-CASSCF level, and the perturbation and QMC calculations can then be performed using only three states. In the following, we present the CASPT2 excitation energies computed at the single-state level since, for the isolated system, we observe a strong mixing of the eigenvectors of the three highest roots in the multi-state CASPT2 calculations. The $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitation energies of our solvated cluster are reported in Table 4 and Figure 3.

For the $n \rightarrow \pi^*$ state, all quantum methods respond rather similarly to the presence of a non-polarizable environment, predicting a blue-shift of about 0.5–0.7 eV depending on the set of point charges (TIP3P or nopol) considered. Improving only the ground-state description of the solvent with the polGS scheme consistently induces a further blue-shift of more than 0.2 eV with respect to the nopol values, **thus generally overshooting the value of the supermolecular reference.** The inclusion of differential polarization effects corrects the polGS results, albeit to a lesser extent than in MCP, bringing the excitation energies in closer agreement with the reference. As expected, the

19

Table 4: Excitation energies (eV) of the $n \to \pi^*$ and $\pi \to \pi^*$ states of acrolein isolated and solvated in a water cluster. The statistical error on the QMC values is given in brackets.

	isolated	TIP3P	nopol	polGS	polSS	polLR	super
			$n \to \pi^*$				
VMC	3.637(3)	4.266(3)	4.191(3)	4.435(3)	4.376(3)	–	4.40(2)
DMC	3.621(6)	4.232(5)	4.172(5)	4.434(5)	4.396(5)	–	4.30(2)
CASPT2	3.559	4.213	4.127	4.385	4.297	–	4.224
TDDFT	3.549	4.145	4.068	4.311	4.250	4.304	4.112
			$\pi \to \pi^*$				
VMC	6.944(3)	6.557(3)	6.608(3)	6.548(3)	6.521(3)	–	6.58(2)
DMC	6.771(6)	6.405(5)	6.455(5)	6.392(6)	6.314(6)	–	6.33(2)
CASPT2	6.727	6.431	6.494	6.366	6.337	–	6.221
TDDFT	6.150	6.038	6.057	6.008	5.994	5.913	5.831

polLR correction is negligible for this state. We must note that a TIP3P description appears here to outperform polSS within TDDFT and CASPT2, which is possibly a fortuitous coincidence due to the use of a single, highly asymmetric configuration with a limited number of water molecules to describe the solution. Finally, it is interesting to observe that the supermolecular excitation energies have a considerable spread, with the TDDFT value being much smaller than the CASPT2 and DMC ones. Moreover, DMC significantly corrects VMC red-shifting the excitation energy by about 0.1 eV while, for the embedded results, the differences between the two flavors of QMC is of the order of 0.03 eV or smaller.

For the $\pi \to \pi^*$ state, the spread in the excitation energies is particularly large with the correlated values being systematically higher by 0.4–0.6 eV than the corresponding CAM-B3LYP results. Other signatures of the complexity of this state are that the perturbation correction on the zero-order CASSCF excitation energies is more than 1 eV in the isolated system and remains as large as 0.7 eV for solvated acrolein (see Table S3). The DMC correction on VMC is around 0.2 eV and therefore larger (albeit the same in the isolated and solvated cases) than what observed so far for MCP and the $n \to \pi^*$ state of acrolein. We find that CASPT2 and DMC agree in predicting a very comparable red-shift induced by the point charges as well as further changes in the excitation energy due to the different treatments of polarization. This level of agreement is somewhat surprising since the dipoles included in the zero-order CASPT2 Hamiltonian are induced in a CASSCF
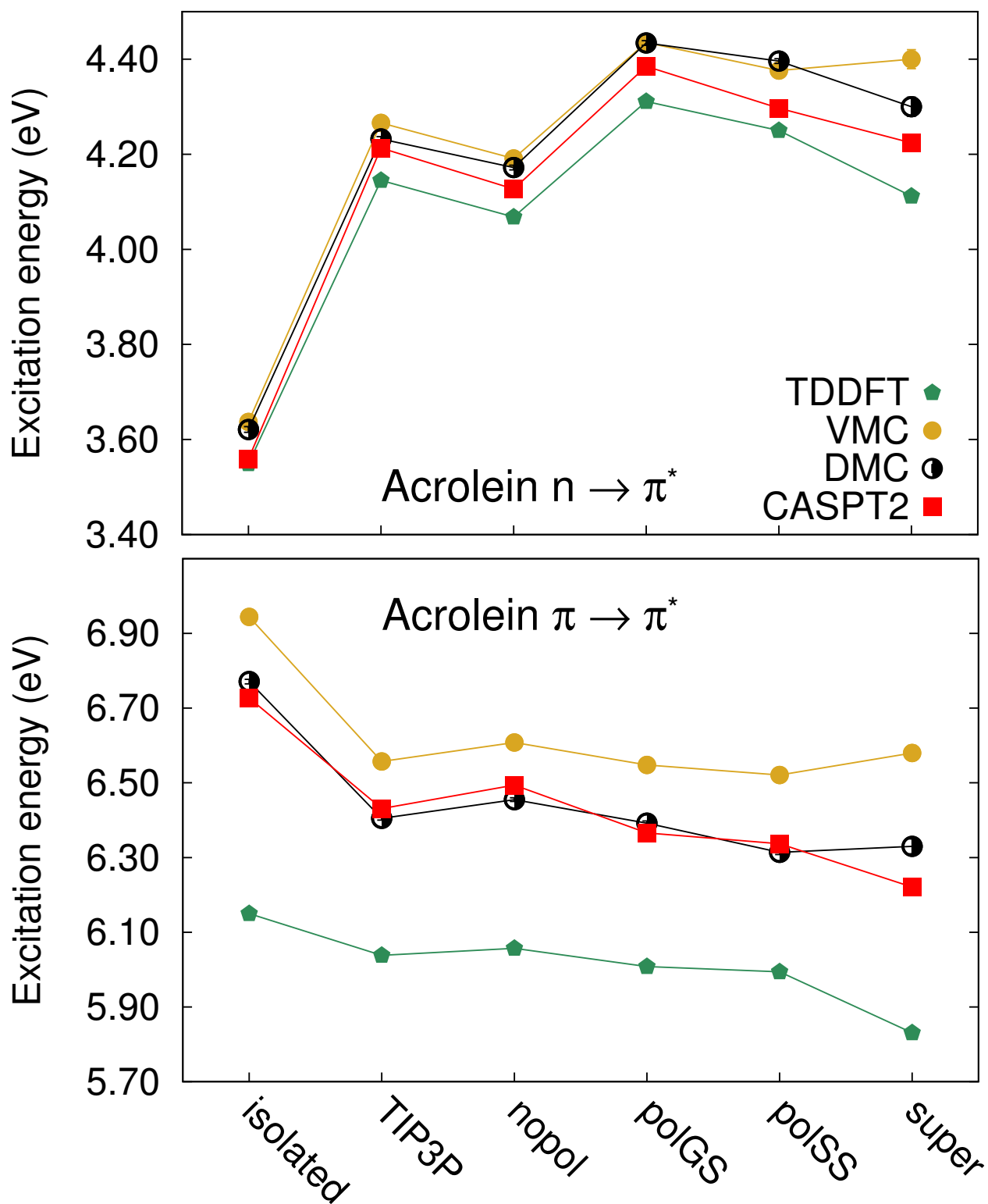
Figure 3: Excitation energies of the n → π^* (top) and π → π^* (bottom) state of *s*-trans acrolein with different QM methods and embedding schemes.

calculation leading to significantly higher excitation energies, while the dipoles employed in DMC have been optimized accounting for dynamical correlation at the VMC level. In general, all methods confirm that a frozen environment only slightly red-shifts the excitation energies with respect to the point-charge embedding and that differential polarization effects are rather small with an average correction on polGS of about 0.04 eV. While electrostatic polarization does not dominate the solute-solvent coupling in this state, other effects can play a role as it might be revealed in a linear response formulation of polarizable embedding. We find in fact that, at the TDDFT level, polLR corrects the polGS excitation energy by about 0.1 eV, consistently with the bright nature of the $\pi \rightarrow \pi^*$ state. Since both effects should be accounted for in a realistic treatment of the solute-solvent system, one should move beyond a quantum-in-classical description of the $\pi \rightarrow \pi^*$ state of acrolein, ultimately extending the correlated quantum calculation to a larger region (see SI).

Finally, we note that the QMC supermolecular $\pi \rightarrow \pi^*$ excitation energy is higher than the CASPT2 value even though the two methods agree in all other conditions. If we exclude possible shortcomings of CASPT2 due to the large perturbation correction (about 0.7 eV for embedded acrolein and 0.8 eV in the supermolecular calculation) on the zero-order CASSCF excitation energies, the high supermolecular DMC value could indicate that the trial wave function must be extended to include excitations coupling the solute and the solvent molecules. While these transitions are not included in the reference CASSCF wave functions used in both DMC and CASPT2, they are however present in the perturbation correction and their importance seems to be here confirmed by the improved treatment of the solvent-solute coupling observed at the polLR level. This possibility must be the subject of further investigations.

5   Conclusions

In this paper, we combine for the first time quantum Monte Carlo with the self-consistent optimization of polarizable dipoles (QMC/MMpol) and investigate how such a discrete polarizable model affects the vertical excitation energies of prototypical solvated organic molecules, namely,

MCP and acrolein in water. We consider two regimes of polarization where the induced dipoles are optimized with respect to the ground-state density (polGS) and in a state-specific manner (polSS), and assess the quality of the results against a complete quantum (supermolecular) calculation of the same solvated cluster. The induced dipoles optimized with respect to the solute described at the VMC level are then used in a DMC calculation of the excitation energies. We also compare the QMC results obtained in the different embedding schemes against available QM/MMpol implementations of CASPT2 and TDDFT. Within TDDFT we also consider a third way to polarize the environment in a linear response formalism (polLR), where the dipoles equilibrate with the transition density matrix.

We find that our QMC/MMpol approach correctly predicts the response of the excitation energy to the different polarizable environments and, through the inclusion of differential polarization effects, leads in general to an improved description of the excitation with respect to the use of static point charges, in particular when the solute couples mainly electrostatically to the solvent. In MCP and the $n \rightarrow \pi^*$ transition of acrolein, we find that improving the description of the ground-state solvent through polGS does not improve (in fact, worsens) the agreement with the reference excitation energies with respect to the use of a standard QM/MM scheme. Introducing a state-specific description of polarization through polSS corrects the shortcomings of polGS and brings the excitation energies in closer agreement with the supermolecular values. On the other hand, as expected, when such electrostatic effects are small as in the case of the $\pi \rightarrow \pi^*$ state of acrolein, the polarization schemes developed here are not sufficient as indicated by the better performance of polLR at the TDDFT level and one must move to a correlated description of a larger quantum region to capture the quantum coupling in the excited solute-solvent system. Importantly, this study offers a clear demonstration that, through the comparison of different quantum approaches (QMC, CASPT2, and TDDFT) and the influence of the various embedding schemes (with point charges, polGS, polSS, and polLR), it is possible to gain a deep understanding of the nature of the coupling between the solute and the environment and well characterize the electronic transition in the system of interest.

In summary, our QMC/MMpol scheme in the polGS and polSS flavors with dipoles optimized in the presence of dynamical correlation offers a robust treatment of the different electrostatic effects which might affect the excitation energy. Considering the relatively small computational cost of this hybrid approach and the higher accuracy with respect to the standard point-charge embedding, QMC/MMpol represents a particularly useful tool for the study of larger photoactive (bio)systems of chemical interest, where the specific environment-chromophore interactions play a major role and must be modeled within a realistic framework.

6 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Basis-set convergence study of the TDDFT and CASPT2 excitation energies; dependence of the excitation energies of acrolein on the choice of the CAS space; comparison of all-electron and pseudopotential CASSCF and CASPT2 calculations; dipole moments; XYZ coordinates of cluster models.

References

(1) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

(2) Lin, H.; Truhlar, D. *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(3) Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(4) Tu, Y.; Laaksonen, A. *Combining Quantum Mechanics and Molecular Mechanics. Some Recent Progresses in QM/MM Methods*; 2010; Vol. 59, pp 1–15.

(5) Zhang, R.; Lev, B.; Cuervo, J. E.; Noskov, S. Y.; Salahub, D. R. Chapter 10 - A Guide to QM/MM Methodology and Applications. In *Combining Quantum Mechanics and Molecular Mechanics. Some Recent Progresses in QM/MM Methods*; 2010; Vol. 59, pp 353–400.

(6) Wallrapp, F. H.; Guallar, V. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 315–322.

(7) Mennucci, B. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6583–6594.

(8) Gao, J. *Reviews in Computational Chemistry*; 2007; Vol. 7, pp 119–185.

(9) Seabra, G.; Swails, J.; Roitberg, A. *Multi-scale Quantum Models for Biocatalysis*; 2009; Vol. 7, pp 3–20.

(10) Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999–3094.

(11) Wanko, M.; Hoffmann, M.; Strodel, P.; Koslowski, A.; Thiel, W.; Neese, F.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2005**, *109*, 3606–3615.

(12) Wanko, M.; Hoffmann, M.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11462–11467.

(13) Wanko, M.; Hoffmann, M.; Frähmcke, J.; Frauenheim, T.; Elstner, M. *J. Phys. Chem. B* **2008**, *112*, 11468–11478.

(14) Send, R.; Kaila, V. R. I.; Sundholm, D. *J. Chem. Phys.* **2011**, *134*, 214114.

(15) Sneskov, K.; Schwabe, T.; Christiansen, O.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18551–18560.

(16) Schwabe, T.; Olsen, J. M. H.; Sneskov, K.; Kongsted, J.; Christiansen, O. *J. Chem. Theory Comput.* **2011**, *7*, 2209–2217.

(17) Filippi, C.; Buda, F.; Guidoni, L.; Sinicropi, A. *J. Chem. Theory Comput.* **2012**, *8*, 112–124.

(18) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. *J. Chem. Theory Comput.* **2012**, *8*, 5092–5106.

(19) Amat, P.; Nifosi, R. *J. Chem. Theory Comput.* **2013**, *9*, 497–508.

(20) Wolter, T.; Welke, K.; Phatak, P.; Bondar, A.-N.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 12582–12590.

(21) Valsson, O.; Campomanes, P.; Tavernelli, I.; Rothlisberger, U.; Filippi, C. *J. Chem. Theory Comput.* **2013**, *9*, 2441–2454.

(22) Kaila, V. R. I.; Send, R.; Sundholm, D. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491–4495.

(23) Schwabe, T.; Beerepoot, M. T. P.; Olsen, J. M. H.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2015**, *17*, year.

(24) Send, R.; Suomivuori, C.-M.; Kaila, V. R. I.; Sundholm, D. *J. Phys. Chem. B* **2015**, *119*, 2933–2945.

(25) Daday, C.; Curutchet, C.; Sinicropi, A.; Mennucci, B.; Filippi, C. *J. Chem. Theory Comput.* **2015**, *11*, 4825–4839.

(26) Thompson, M. A. *J. Phys. Chem.* **1996**, *100*, 14492–14507.

(27) Curutchet, C.; Muñoz-Losa, A.; Monti, S.; Kongsted, J.; Scholes, G. D.; Mennucci, B. *J. Chem. Theory Comput.* **2009**, *5*, 1838–1848.

(28) Olsen, J. M.; Aidas, K.; Kongsted, J. *J. Chem. Theory Comput.* **2010**, *6*, 3721–3734.

(29) Slipchenko, L. V. *J. Phys. Chem. A* **2010**, *114*, 8824–8830.

(30) Sneskov, K.; Schwabe, T.; Kongsted, J.; Christiansen, O. *J. Chem. Phys.* **2011**, *134*, 104108.

(31) Sneskov, K.; Olsen, J. M. H.; Schwabe, T.; Hattig, C.; Christiansen, O.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7567–7576.

(32) Li, Q.; Mennucci, B.; Robb, M. A.; Blancafort, L.; Curutchet, C. *J. Chem. Theory Comput.* **2015**, *11*, 1674–1682.

(33) Jensen, L.; van Duijnen, P. T.; Snijders, J. G. *J. Chem. Phys.* **2003**, *119*, 3800–3809.

(34) Beerepoot, M. T. P.; Steindal, A. H.; Kongsted, J.; Brandsdal, B. O.; Frediani, L.; Ruud, K.; Olsen, J. M. H. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4735–4743.

(35) Steindal, A. H.; Olsen, J. M. H.; Ruud, K.; Frediani, L.; Kongsted, J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 5440–5451.

(36) Beerepoot, M. T.; Steindal, A. H.; Ruud, K.; Olsen, J. M.; Kongsted, J. *Comp. Theor. Chem.* **2014**, *1040-1041*, 304–311.

(37) Pikulska, A.; Steindal, A. H.; Beerepoot, M. T. P.; Pecul, M. *J. Phys. Chem. B* **2015**, *119*, 3377–3386.

(38) Filippi, C.; Zaccheddu, M.; Buda, F. *J. Chem. Theory Comput.* **2009**, *5*, 2074–2087.

(39) Valsson, O.; Filippi, C. *J. Chem. Theory Comput.* **2010**, *6*, 1275–1292.

(40) Send, R.; Valsson, O.; Filippi, C. *J. Chem. Theory Comput.* **2011**, *7*, 444–455.

(41) Valsson, O.; Angeli, C.; Filippi, C. *Phys. Chem. Chem. Phys.* **2012**, *14*, 11015–11020.

(42) Daday, C.; König, C.; Neugebauer, J.; Filippi, C. *ChemPhysChem.* **2014**, *15*, year.

(43) Thole, B. T.; van Duijnen, P. T. *Theor. Chim. Acta* **1980**, *55*, 307–318.

(44) van Duijnen, P. T.; Swart, M. *J. Phys. Chem. A* **1998**, *102*, 2399–2407.

(45) Masia, M.; Probst, M.; Rey, R. *J. Chem. Phys.* **2005**, *123*, .

(46) Wang, J.; Cieplak, P.; Li, J.; Hou, T.; Ray, L.; Yong, D. *J. Chem. Phys. B* **2011**, *8*, 3091–3099.

(47) We note that, in a QMC calculation, the electronic contribution of the solute to the electric field at $\mathbf{r}_a$ is computed by sampling the square of the many-body wave function and collecting the average of $\sum_i (\mathbf{r}_a - \mathbf{r}_i)/|\mathbf{r}_i - \mathbf{r}_a|^3$, where the sum is over the electrons.

(48) Umrigar, C. J.; Toulouse, J.; Filippi, C.; Sorella, S.; Hennig, R. G. *Phys. Rev. Lett.* **2007**, *98*, 110201–110203.

(49) CHAMP is a quantum Monte Carlo program package written by C. J. Umrigar, C. Filippi, S. Moroni, and collaborators.

(50) Burkatzki, M.; Filippi, C.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 234105.

(51) For the hydrogen atom, we use a more accurate BFD pseudopotential and basis set. Dolg, M.; Filippi, C., private communication.

(52) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Jr, J. A. M. *J. Comput. Chem.* **1993**, *14*, 1347–1363.

(53) Gordon, M. S.; Schmidt, M. W. Advances in electronic structure theory: GAMESS a decade later. In *Theory and Applications of Computational Chemistry: the first forty years*; Dykstra, C., Frenking, G., Kim, K., Scuseria, G., Eds.; Elsevier: Amsterdam, 2011; Chapter 41, pp 1167–1190.

(54) Filippi, C.; Umrigar, C. J. *J. Chem. Phys.* **1996**, *105*, 213–226, As Jastrow correlation factor, we use the exponential of the sum of three fifth-order polynomials of the electron-nuclear (e-n), the electron-electron (e-e). The Jastrow factor is adapted to deal with pseudo-atoms, and the scaling factor κ is set to 0.6 a.u.The 2-body Jastrow factor includes five parameters in the e-e terms and four parameters for each atom type in the e-n terms.

(55) Casula, M. *Phys. Rev. B* **2006**, *74*, 161102.

(56) Aquilante, F.; De Vico, L.; Ferré, N.; Ghigo, G.; Malmqvist, P.-å.; Neogrády, P.; Pedersen, T. B.; PitoÅĹák, M.; Reiher, M.; Roos, B. O.; Serrano-Andrés, L.; Urban, M.; Veryazov, V.; Lindh, R. *J. Comput. Chem.* **2010**, *31*, 224–247.

(57) Andersson, K.; Malmqvist, P.-A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(58) Andersson, K.; Malmqvist, P.-A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(59) Ghigo, G.; Roos, B. O.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **2004**, *396*, 142–149.

(60) Forsberg, N.; Malmqvist, P.-Å. *Chem. Phys. Lett.* **1997**, *274*, 196–204.

(61) Frisch, M. J. et al. *Gaussian 09 Revision A.02*, Gaussian Inc. Wallingford CT 2009.

(62) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51–57.

(63) Carles Curutchet, private communication.

(64) Wang, J.; Cieplak, P.; Li, J.; Wang, J.; Cai, Q.; Hsieh, M.; Lei, H.; Luo, R.; Duan, Y. *J. Chem. Phys. B* **2011**, *8*, 3100–3111.

(65) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(66) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(67) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(68) We include diffuse functions only for the heavy atoms and take them from the aug-cc-pVXZ basis sets in the EMSL Basis Set Library (http://bse.pnl.gov). For the double and triple ζ basis sets, we omit the p and f diffuse functions, respectively.

(69) Dunning Jr, T. H. *J. Chem. Phys.* **1989**, *90*, 1007–1023.

(70) Peterson, K. A.; Woon, D. E.; Dunning Jr, T. H. *J. Chem. Phys.* **1994**, *100*, 7410–7415.

(71) Wilson, A.; van Mourik, T.; Dunning Jr, T. H. *J. Mol. Struct. (Theochem)* **1997**, *388*, 339–349.

(72) Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057–1065.

(73) Corni, S.; Cammi, R.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2005**, *123*, 134512–134510.

(74) Staley, S. W.; Norden, T. D. *J. Am. Chem. Soc.* **1984**, *106*, 3699–3700.

(75) Moskvin, A.; Yablonskii, O.; Bondar, L. *Theor. Exp. Chem.* **1966**, *2*, 469–472.

(76) Norden, T. D.; Staley, S. W.; Taylor, W. H.; Harmony, M. D. *J. Am. Chem. Soc.* **1986**, *108*, 7912–7918.

(77) Merchán, M.; González-Luque, R.; Roos, B. O. *Theor. Chim. Acta* **1996**, *94*, 143–154.

(78) Losa, A. M.; Galván, I. F.; Aguilar, M. A.; ; Martín, M. E. *J. Phys. Chem. B* **2007**, *111*, 9864–9870.

(79) do Monte, S. A.; Müller, T.; Dallos, M.; Lischka, H.; Diedenhofen, M.; Klamt, A. *Theor. Chem. Acc.* **2004**, *111*, 78–89.