



HAL
open science

Coupling molecular density functional theory with converged selected configuration interaction methods to study excited states in aqueous solution

Maxime Labat, Emmanuel Giner, Guillaume Jeanmairat

► To cite this version:

Maxime Labat, Emmanuel Giner, Guillaume Jeanmairat. Coupling molecular density functional theory with converged selected configuration interaction methods to study excited states in aqueous solution. *The Journal of Chemical Physics*, 2024, 161 (1), pp.014113. 10.1063/5.0213426 . hal-04760563

HAL Id: hal-04760563

<https://hal.science/hal-04760563v1>

Submitted on 30 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Coupling Molecular Density Functional Theory with Converged Selected Configuration Interaction Methods to Study Excited states in Aqueous Solution

Maxime Labat,¹ Emmanuel Giner*,^{2, a)} and Guillaume Jeanmairet*^{1, 3, b)}

¹⁾ Sorbonne Université, CNRS, Physico-Chimie des électrolytes et Nanosystèmes Interfaciaux, PHENIX, F-75005 Paris, France

²⁾ Sorbonne Université, CNRS, Laboratoire de Chimie Théorique, Sorbonne Université, F-75005 Paris, France

³⁾ Réseau sur le Stockage électrochimique de l'énergie (RS2E), FR CNRS 3459, 80039 Amiens Cedex, France

This paper presents the first implementation of a coupling between advanced wave function theories and molecular density functional theory (MDFT). This method enables the modeling of solvent effect into quantum mechanical (QM) calculations by incorporating an electrostatic potential generated by solvent charges into the electronic Hamiltonian. Solvent charges are deduced from the spatially and angularly dependent solvent particle density. Such density is obtained through the minimization of the functional associated to the molecular mechanics (MM) Hamiltonian describing the interaction between the fluid particles. The introduced QM/MDFT framework belongs to QM/MM family of methods but its originality lies in the use of MDFT as the MM solver, offering two main advantages. Firstly, its functional formulation makes it competitive with respect to sampling-based molecular mechanics. Secondly, it preserves a molecular-level description lost in macroscopic continuum approaches. Excited states properties of water and formaldehyde molecules solvated into water have been computed at the selected configuration interaction (SCI) level. Excitation energies and dipole moment have been compared with experimental data and previous theoretical work. A key finding is that using the Hartree-Fock method to describe the solute allows for predicting the solvent charge around the ground-state with sufficient precision for the subsequent SCI calculations of excited-states. This significantly reduces the computational cost of the described procedure, paving the way for the study of more complex molecules.

I. 1. INTRODUCTION

Most of the chemical properties, including reactivity and spectroscopy, are governed by electronic properties. This is the reason why electronic structure calculations play a central role in theoretical and computational chemistry. Energy differences should be computed with a precision of typically 0.05 eV (approximately 1 kcal.mol⁻¹) for reactivity and in the range 10⁻⁴ to 10⁻¹ eV depending on the targeted spectroscopic property. To be able to compare with actual experimental results, one needs to account for the environment in electronic structure calculations, as it can significantly influence the properties of the studied molecules. The most ubiquitous effect to consider is likely the influence of the solvent, given that the majority of chemical processes occur in solution rather than in the gas phase. In organic chemistry, the solvent has essentially a two-fold role. It acts as a medium that brings the reactant species together, but it also subtly influences the reaction path. This aspect has been acknowledged and exploited by chemists since the early 30s¹. In spectroscopy, the solvent impacts the excitation energies of the analyte modifying its spectroscopic signature, the most spectacular example being solvatochromism².

Regarding reactivity, the solvent can modify the reaction profile by stabilizing or destabilizing the reactants and products. Moreover, it can also be directly involved in the reaction mechanism, for instance by forming chemical bonds with the reactants. This is also true in spectroscopy where the solvent can actively contribute to the excitation process (*i.e.* its molecular orbitals are explicitly involved in the electronic excited and/or ground states) or play a passive role when the excitation process is located on the solute.

The most straightforward way to account for solvent effect is to explicitly include solvent molecules into the QM

calculations. However, this considerably increases the computational cost which scales as a polynomial of the number of electrons. Moreover, neglecting temperature in solution is a much cruder assumption than in vacuum (gas phase). Indeed, a solution is a liquid whose constituting molecules are in constant motion. This implies that one should no longer perform a unique (QM) calculation for a given set of coordinates but rather sample the solute and solvent configurations until a statistically converged dataset is obtained. This is the strategy adopted in *ab-initio* molecular dynamics³ (AIMD) where the forces acting between molecules are computed at the QM level and the particles are subsequently moved by discretization of Newton's laws. This is a powerful simulation technique but its numerical cost leads to several limitations. First, since it needs to solve QM equations multiple times, only the most efficient methods are appropriate. In practice it is essentially limited to the use of electronic density functional theory (eDFT) to compute the forces. Second, the length and timescales accessible to AIMD are rather limited (typically up to a few ns and a few nm). Finally, high throughput calculations are not an option since each simulation requires 10⁴⁻⁵ CPU hours.

When the solvent does not react with the solute, an usual strategy is to resort to classical force-field molecular dynamics (MD) to describe the solvent while keeping a QM description of the solute. This hybrid QM/MM⁴ strategy reduces tremendously the computational cost. An even further simplification is to avoid the statistical sampling of the solvent degrees of freedom by modelling it as a continuum. This allows to account for part of solvation effect while maintaining the computational cost close to *in vacuo* calculations. It is thus a popular strategy but it suffers from several limitations. Continuum solvation models (CSMs) have several parameters, which are usually fitted to reproduce a set of reference calculations. Consequently the *ab-initio* nature of the approach is lost in the process, as well as the molecular nature of the solvent.

A third strategy is to conserve the QM/MM partitioning of the system but to resort to liquid state theories

a) Electronic mail: emmanuel.giner@lct.jussieu.fr

b) Electronic mail: guillaume.jeanmairet@sorbonne-universite.fr

(LST) to describe the classical part. There exist several LST that aim at solving the Ornstein-Zernike (OZ) equation that is the fundamental equation describing liquids at the atomic/molecular scale. In the OZ equation, the liquid is described by a continuous field related to its spatially dependant density $n(\mathbf{r})$. Key quantities entering OZ equation are the direct and total correlation function c and h that account for the correlation within the liquid. These functions depend on the interaction potential between the liquid particles and the considered physical conditions, such as the pressure or temperature. It is worth insisting on the fact that such approaches truly conserves a molecular description of the fluid particles while relying on the rather simple continuous solvent density $n(r)$. This represents a compromise between the CSM, which fails to account for the molecular nature of the liquid, and the MD-based strategy, which relies on the tedious sampling of solvent configurations. The appealing properties of LST has motivated several attempts to utilize them for incorporating solvent effects in QM calculations. The RISM approach and its 3D-RISM counterpart are for instance implemented in several QM packages⁵⁻⁷. The idea behind 3D-RISM is to replace the molecular direct correlation function c by a sum of simpler 1D site-site correlation functions. This is an **unphysical** approximation that makes the theory inconsistent with statistical physics principles but it nevertheless drastically simplifies the resolution of the molecular OZ equation. Another LST allowing to predict the in-homogeneous density of the liquid is classical density functional theory (cDFT)^{8,9}. In short, cDFT allows to compute the grand-potential by minimizing a functional of the liquid density. This minimum is reached for the equilibrium density. Similarly to its electronic counterpart, cDFT is theoretically well grounded but approximations in the choice of the functional describing the interactions between particles, here solvent molecules, are required to make its utilisation practical. Several attempts to use cDFT as a solver for QM/MM calculations have been proposed. In the joint DFT framework and its associated software JDFTx^{10,11}, model functionnals whose parameters are adjusted to account for properties of real fluid are proposed. In the so-called reaction density functional theory (RxDFT)¹², solute electronic structure is computed with electronic DFT. Atomic partial charges are fitted to best reproduce the molecular electrostatic potential. These atomic charges, along with **Lennard-Jones** (LJ) sites used to model dispersion and repulsion, are subsequently employed in a cDFT calculation to estimate the solvation free energy. A more elaborate strategy is to self-consistently minimize the electronic and the solvent functional to account for the mutual impact between the solute and the solvent. Even keeping the coupling Hamiltonian at the electrostatic level, it is possible to capture a large part of the mutual polarization between the solute and the solvent. This strategy has been employed to study the reactivity of an SN₂ nucleophilic substitution between chloride and chloromethane in water¹³. In this original work, eDFT was chosen to describe the QM solute.

Turning now to the theoretical description of spectroscopic properties in solution, the latter has focussed an increasing attention over the last thirty years. As for ground state or free energy calculations, there are mainly two approaches to describe the solvent: either the solvent is considered as a continuous medium¹⁴⁻²⁵, or explicitly treated as a molecular object either at the QM or MM level²⁶⁻²⁸.

Whether the solvent is treated at the continuum or molecular level, its effect on the spectrum of the Hamilto-

nian is obtained by adding an effective one-body potential accounting for the QM/MM interaction. Nevertheless, the main difference between the continuum and the explicit description from the point of view of the spectroscopy comes from the way the excited states properties are computed. In the case of the CSM, once the converged charge density of the dielectric is obtained, a single QM calculation is performed and correspondingly all the excited states properties. When the solvent is described using classical MD, once the solvent trajectory have been generated either using a quantum or a classical description of the solute, two strategies can be employed to compute the excited states. A first one is to determine the statistically averaged structure of the solvent and to use it as an the external field acting on the solute in a single QM calculation allowing to compute the excited states. Another one is to compute the excited states properties as a statistical average over a set of distinct QM calculations performed along the classical trajectory of the whole QM/MM system. The latter is of course more computationally demanding than the former since it requires to perform $10^2 - 10^3$ QM calculations.

Again, as in the case of the ground state, there is a trade off between the accuracy of a molecular description and the appealing computational cost of a continuum description.

In the present work, we propose to use the hybrid continuum/molecular description of the solvent offered by molecular density functional theory (MDFT) to compute excited state properties. Our approach consists in a self-consistent ground state MDFT calculation, which allows to obtain both the ground state and solvent equilibrium charge distribution. A single excited states calculation in the presence of the electrostatic field of the equilibrium solvent distribution is subsequently performed. **Note that our MDFT-based methodology is similar to MD based calculations which use an average representation of the solvent to perform a single QM calculation of the excitation. Indeed both methods assume that the (statistical) average of an observable over a set of solvent configurations can be well approximated by a single calculation on the averaged solvent configuration. However, this is not a trivial approximation since the electrons respond instantaneously to movement of the nuclei. This approximation has been thoroughly tested by González-Espinosa *et al.*²⁹ for the acetone in water system. They have shown that, for the two observables considered here — namely, the electronic excitation and dipole moment — using an averaged environment effectively represents the fluctuating solvent environment.**

We tested the proposed methodology by studying the excited states properties of water and formaldehyde in aqueous solution using near full-configuration interaction level of theory, namely converged selected configuration interaction (SCI)³⁰⁻³⁶, for the calculation of excited states,

We compare with state-of-the-art approaches using a coupled cluster (CC) level of theory for the QM part and a molecular²⁷ or continuum description^{25,37} of the solvent. For water in water, the results obtained with our QM/MDFT approach are in better agreement with the reference values than the continuum approaches^{25,37}. A reasonable accuracy is also found for the formaldehyde in aqueous solution. We also investigate the impact of the level of theory used to describe the QM part in the MDFT calculation. To rationalize predicted excitation energies we also computed the dipole moment of both ground and excited state, in vacuum and solution. Contrary to previous claims³⁸, we found no clear evidence of a correlation

between the intensity of the blue shifts and the dipole moments.

II. THEORY

A. MDFT as an effective point charges model for QM/MM calculations

As the primary goal of this paper is to demonstrate that MDFT is an appropriate method to account for solvent effects in QM calculations, we will expose the formalism from the QM perspective, *i.e.*, as a self-consistent classical electrostatic embedding of the usual QM problem.

In the present approach, the QM solute, treated within the Born-Oppenheimer approximation, interacts with the solvent described as a volumetric charge density ρ_c . Therefore, the additional QM/MM potential $\hat{V}_C[\rho_c](\mathbf{r})$ is simply the Coulomb potential produced by ρ_c , which can be explicitly written as

$$\hat{V}_C[\rho_c](\mathbf{r}) = \int \frac{\rho_c(\mathbf{r}')}{|\mathbf{r}' - \mathbf{r}|} d\mathbf{r}', \quad (1)$$

using the standard atomic units, $1/4\pi\epsilon_0 = 1$. Note that, as opposed to a continuous description, there is no relative permittivity involved in the expression of the Coulomb potential but only the vacuum permittivity ϵ_0 . In the present work, both the classical solvent charge density ρ_c and the QM counterpart are obtained through the self-consistent resolution of the QM and MDFT equations, allowing to take into account the explicit coupling between the QM and MM part. Before describing in more details the MDFT equations in Sec. II B, we would like to emphasize that ρ_c depends mainly on three quantities: the QM density, the choice of the classical functional describing the solvent and the description of the QM/MM interaction *i.e.* the coupling Hamiltonian.

Although the classical density ρ_c is in principle a continuous object, we represent it on a three-dimensional grid and therefore \hat{V}_C is obtained as a sum of point charges contribution

$$\hat{V}_C[\rho_c](\mathbf{r}) = \sum_{m=1}^{N_g} \frac{\rho_c(\mathbf{r}_m)\Delta V}{|\mathbf{r}_m - \mathbf{r}|}, \quad (2)$$

where \mathbf{r}_m is the coordinate of the grid point m , ΔV the volume of the grid voxel, and N_g the number of grid points.

Given a classical charge density ρ_c , the total effective QM Hamiltonian $\hat{H}_{\text{tot}}[\rho_c]$ is therefore the sum of the usual Born-Oppenheimer Hamiltonian and of the $\hat{V}_C[\rho_c]$ potential

$$\hat{H}_{\text{tot}}[\rho_c] = V_{\text{NN}} + \hat{T} + \hat{V}_{\text{Ne}} + \hat{W}_{\text{ee}} + \hat{V}_C[\rho_c] \quad (3)$$

where

$$V_{\text{NN}} = \sum_{A>B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}, \quad (4)$$

$$\hat{T} = -\frac{1}{2} \sum_{i=1}^{N_e} \Delta_{\mathbf{r}_i}, \quad (5)$$

$$\hat{V}_{\text{Ne}} = - \sum_{i=1}^{N_e} \sum_A \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}, \quad (6)$$

$$\hat{W}_{\text{ee}} = \sum_{i>j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (7)$$

with Z_A and \mathbf{R}_A being the nuclear charge and coordinates, respectively, and \mathbf{r}_i the electronic coordinates. Developed on a AO basis set $\mathcal{B} = \{\chi_i(\mathbf{r})\}$, the classical Coulomb potential is then simply given as a sum of usual one-electron Coulomb integrals weighted by the classical charge density ρ_c

$$\langle \chi_j | \hat{V}_C[\rho_c] | \chi_i \rangle = \sum_{m=1}^{N_g} \Delta V \rho_c(\mathbf{r}_m) \langle \chi_j | \frac{1}{|\mathbf{r}_m - \mathbf{r}|} | \chi_i \rangle. \quad (8)$$

For a given classical density of charge ρ_c , we can then solve the Schrödinger equation as usually done in any quantum chemistry calculations

$$\hat{H}_{\text{tot}}[\rho_c] |\Psi_i[\rho_c]\rangle = E_i[\rho_c] |\Psi_i[\rho_c]\rangle, \quad (9)$$

where the solutions $|\Psi_i[\rho_c]\rangle$ and $E_i[\rho_c]$ depend on the solvent charge density ρ_c . Therefore one can obtain any usual properties with a computational cost similar to vacuum calculation while explicitly taking into account the electrostatic effect of the solvent.

In this paper, we shall be mostly concerned with electronic properties of both the ground and excited states, such as excitation energies

$$\omega_i[\rho_c] = E_i[\rho_c] - E_0[\rho_c], \quad (10)$$

and the dipole moments

$$\boldsymbol{\mu}_i[\rho_c] = \boldsymbol{\mu}_{\text{nucl}} - \int n_i(\mathbf{r}) \mathbf{r} d\mathbf{r}, \quad (11)$$

where

$$\boldsymbol{\mu}_{\text{nucl}} = \sum_A Z_A \mathbf{R}_A, \quad (12)$$

and $n_i(\mathbf{r})$ is the one-electron density of the state $\Psi_i[\rho_c]$, which depends implicitly on the classical charge density. The electronic wave-function depends explicitly on the solvent charge density as evidenced in Eq. (9). It will be shown in section II B that the solvent charge density is also influenced by the electronic density n . The primary objective of this paper is to introduce a self-consistent QM/MDFT resolution to predict a reasonable solvent charge density and to compute the electronic structure properties in its presence.

B. Summary of Molecular Density Functional Theory

Molecular density functional theory^{13,39,40} is a particular flavour of cDFT that has been designed to tackle solvation problems. We shall start by recalling the main ansatz of cDFT^{8,9}. We consider a system made of classical particles, *i.e.* whose interactions are described by a classical force-field, which is perturbed by the presence of an external potential V_{ext} . Following Mermin and Evans pioneering work, it is possible to show that in the grand-canonical ensemble, at finite temperature T , there exists a unique functional $\Omega[\rho]$ of the particle density, ρ . This functional

is equal to the grand potential Ω at its minimum, which is reached for the equilibrium solvent density ρ_{eq} . In the framework of MDFT, the external potential is due to the presence of the solute and the particle density describes the solvent. We will consider solvent constituted of rigid molecules such as the knowledge of the position of their center of mass \mathbf{r} and orientation $\mathbf{\Omega}$ is sufficient to fully describe molecule coordinates. Consequently, the solvent density $\rho(\mathbf{r}, \mathbf{\Omega})$ have a spatial and angular dependancy. We introduce the functional F , defined as the difference between the grand potential functional in the presence of solute $\Omega[\rho]$ and the grand potential Ω_0 of the non perturbed fluid, *i.e* with the homogeneous density ρ_0 ,

$$F[\rho] = \Omega[\rho] - \Omega_0. \quad (13)$$

The minimization of this functional directly yields the solvation free energy (SFE) and the equilibrium solvent density. It is noteworthy to emphasize the challenge associated with computing SFE using methods based on statistical sampling of the phase space, such as MD. This difficulty arises because the free energy cannot be expressed as an ensemble average but instead necessitates knowledge of the grand-partition function⁴¹.

The functional F is usually written as

$$F[\rho] = F_{\text{id}}[\rho] + F_{\text{exc}}[\rho] + F_{\text{ext}}[\rho]. \quad (14)$$

In this equation, the first term F_{id} represents the ideal term, corresponding to the entropic contribution of the non-interacting fluid^{9,42}. The second term, F_{exc} , is due to solvent-solvent interactions. Similarly to the exchange-correlation term in eDFT, this is the problematic term. Note that, unlike eDFT, the excess functional is not unique, as it depends on the force-field model chosen to describe the interaction between particles. While an analytical expression exists, it requires approximations to be used in practice. In this work, we use the so-called HNC excess functional F_{exc} ⁴³,

$$F_{\text{exc}}[\rho] = -\frac{k_{\text{B}}T}{2} \iint \Delta\rho(\mathbf{r})c(|\mathbf{r}-\mathbf{r}'|, \mathbf{\Omega}, \mathbf{\Omega}')\Delta\rho(\mathbf{r}')d\mathbf{r}d\mathbf{r}'d\mathbf{\Omega}d\mathbf{\Omega}' \quad (15)$$

where k_{B} is the Boltzmann constant and $\Delta\rho = \rho - \rho_0$. The remaining term, F_{ext} , represents contribution due to the external perturbation V_{ext} of the solute acting on the solvent. It can be formally expressed as

$$\begin{aligned} F_{\text{ext}}[\rho] &= \iint \rho(\mathbf{r}, \mathbf{\Omega})V_{\text{ext}}(\mathbf{r}, \mathbf{\Omega})d\mathbf{r}d\mathbf{\Omega} \\ &= F_{\text{ext,C}}[n, \rho] + F_{\text{ext,RD}}[\rho, n] \end{aligned} \quad (16)$$

which can be split into the electrostatic contribution $F_{\text{ext,C}}$ and the repulsion-dispersion contribution $F_{\text{ext,RD}}$. The electrostatic interaction between the quantum solute and the classical solvent, which depends on the electronic density n is

$$F_{\text{ext,C}}[n, \rho] = \int \rho_c(\mathbf{r})V_{\text{QM}}[n](\mathbf{r})d\mathbf{r}. \quad (17)$$

where $V_{\text{QM}}[n]$ is the electrostatic potential generated by the nuclei and the solute electronic density. The solvent charge density can be computed from the solvent density with

$$\rho_c(\mathbf{r}) = \iint \rho(\mathbf{r}', \mathbf{\Omega})\gamma(\mathbf{r}-\mathbf{r}', \mathbf{\Omega})d\mathbf{r}'d\mathbf{\Omega}, \quad (18)$$

where $\gamma(\mathbf{r}, \mathbf{\Omega})$ is the charge density of a water molecule taken at the origin with orientation $\mathbf{\Omega}$. In addition to

the electrostatic term of Eq. (17), one should also include short-range repulsion and dispersion interactions between the solute and the solvent. Using a popular strategy in QM/MM literature⁴⁴, this is done by adding LJ sites on the nuclei of the solute. The repulsion-dispersion part of the functional then takes the following form

$$F_{\text{ext,RD}}[\rho, n] \approx F_{\text{ext,LJ}}[\rho] = \iint \rho(\mathbf{r}, \mathbf{\omega})V_{\text{LJ}}(\mathbf{r}, \mathbf{\omega})d\mathbf{r}d\mathbf{\Omega}, \quad (19)$$

with

$$\begin{aligned} V_{\text{LJ}}(\mathbf{r}, \mathbf{\Omega}) &= \sum_{i \in \text{solute}} \sum_{j \in \text{solvent}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{|\mathbf{r} + \mathbf{r}_{j\mathbf{\Omega}} - \mathbf{r}_i|} \right)^{12} \right. \\ &\quad \left. - \left(\frac{\sigma_{ij}}{|\mathbf{r} + \mathbf{r}_{j\mathbf{\Omega}} - \mathbf{r}_i|} \right)^6 \right]. \end{aligned} \quad (20)$$

where ϵ_{ij} and σ_{ij} are the mixed LJ parameters. In Eq. (20), \mathbf{r}_i is the position of the i^{th} site of the solute and $\mathbf{r}_{j\mathbf{\Omega}}$ denotes the position with respect to the center of mass of the j^{th} site of a solvent molecule located in \mathbf{r} and having an orientation $\mathbf{\Omega}$. There is no prescribed method to choose the LJ parameters but it impacts the predicted solvation free energies and structure¹³. As a result, the proposed methodology is no longer truly ab initio. An alternative could be to include repulsion and dispersion interactions through the use of an electron-solvent pseudopotential^{45,46}. The QM and MM part are thus only coupled at the electrostatic level through Eqs. (2) and (17), which should in principle give the same result if no approximations were made in their numerical implementation. In the next subsection, we detail how the joint resolution of Eqs. (9) and (14) are performed to self-consistently compute the electronic density and the solvent charge density.

C. QM/MM Coupling

The first iteration of our framework begins at the quantum level by computing the energy of the solute in vacuum. The electrostatic potential $V_{\text{QM}}[n]$, which depends on the ground-state electronic density is subsequently employed in the calculation of the external functional of Eq. (16) in the MDFT calculation. After optimization of the classical functional, the solvent charge density computed as in Eq. (18) is introduced as a set of point charges in a new QM calculation as in Eq. (2). This allow to obtain the ground-state energy of the solute in solution. This process is iterated until convergence is achieved, that is when the solute energy variation is below $5 \cdot 10^{-4}$ a.u and the classical free energy variation is below 10^{-4} kJ · mol⁻¹ in this paper.

The introduced QM/MDFT properly accounts for a large part of the mutual polarization of the electronic density and the molecular density of the solvent. Moreover, solvation structure is predicted in great details since the equilibrium density $\rho(\mathbf{r}, \mathbf{\Omega})$ measures the average number of molecules at specific locations but also quantifies the preferred orientations. Nevertheless, as this paper exclusively focuses on a non-polarizable water model, it overlooks the electronic polarizability of the solvent molecules. This is a strong limitation since the electronic density of the solvent relaxes almost instantaneously upon a modification of the electric field.

3. COMPUTATIONAL DETAILS

This article focuses on the aqueous solvation of two polar solutes: water and formaldehyde. Except for equation of motion coupled cluster restricted to single and double substitution (EOM-CCSD) calculations which have been performed with PySCF⁴⁷, all QM calculations are performed using the Quantum Package³⁶. The electrostatic potential created by the solute is evaluated in a 25^3 \AA^3 cubic box with 100 grid nodes in each direction, and serves as an external potential in the following MDFT calculation.

Solvent densities are obtained using our homemade Fortran-written MDFT program. Minimizations are performed using the same spatial grid as the one used to compute the QM electrostatic potential, and the three Euler angles are described with 196 discrete orientations. Water solvent molecules are described using the non-polarizable SPC/E model⁴⁸. The $c^{(2)}$ direct correlation function, entering the excess functional of Eq. (15) has been kindly provided by Luc Belloni. He computed it for an homogeneous density $n_0 = 0.0332891 \text{ \AA}^{-3}$ using a mixture of Monte-Carlo simulation data and integral equation theory⁴⁹. As described earlier in the text, dispersion and repulsion forces between the solute and the solvent molecules are modelled by adding LJ sites onto the solute nuclei. We use the same LJ parameters as the QM/MD we compare to^{25,50-53}. When comparing to CSM results which do not need LJ sites, we take our parameters from the second generation generalized amber force field (gaff2)⁵⁴.

The proposed QM/MDFT procedure can be used to predict the equilibrium solvent density from which the solvent charge density can be deduced from Eq. (18). As mentioned earlier, the choice of the QM method used to obtain the ground state electronic density impacts the predicted solvent charge density. To assess the impact of this choice, three QM methods were considered: Hartree-Fock (HF), Configuration Interaction Single Double (CISD) and SCI³⁶. Nevertheless, the excitation energies and dipole moment are always computed using SCI, regardless of the QM/MDFT method employed, in order to avoid any approximations on these subtle properties.

III. APPLICATIONS

A. Water in Water

We start by the investigation of the three lowest singlet excited states of water, both in vacuum and in solution. This system has been studied previously using various techniques. In particular, two research groups used the equation of motion coupled cluster restricted to single and double substitution (EOM-CCSD) to describe the solute^{25,50-53}. However, these studies relied on two different approaches to account for the solvent effects. The group of Christiansen used a MD approach in which they first run a fully classical simulation to generate a set of relevant solvent configurations that are subsequently use as an external potential in an EOM-CCSD calculation. In practice, they assigned partial point charges to the MM nuclei which are introduced into the one-electron part of the vacuum Hamiltonian. They still resort to LJ sites on the quantum nuclei to account for Pauli repulsion and dispersion. Then, they either perform a single QM calculation on a configuration representative of the average over the

State	EOM-CCSD						SCI			
	AVDZ		AVTZ		d-AVTZ		AVDZ		AVTZ	
1B_1	7.38	7.46	7.54	7.61	7.53	7.62	7.47	7.54	7.56	7.64
1A_2	9.16	9.22	9.31	9.37	9.30	9.37	9.26	9.33	9.35	9.42
1A_1	9.82	9.87	9.93	9.96	9.84	9.88	9.92	9.94	9.96	9.99

Table I. Excitation energies (eV) for water in vacuum computed at the EOM-CCSD and SCI level of theory using the geometry of Howard *et. al.* (left) and the geometry of Kongsted *et. al.*⁵² using various Dunning basis sets.

trajectory⁵⁰⁻⁵². Another strategy is to perform a series of several hundreds of QM calculation with a set of solvent configuration selected along the MM trajectory⁵³. The first strategy is an equilibrium approach and closely resembles the MDFT-based methodology proposed here. The average solvent configuration employed is nothing more than some sort of projection of the equilibrium density onto the most representative set of coordinates. In contrast, the second approach encompasses some **informations about the fluctuations** of the solvent. This, however, comes at the cost of a drastic increase in computational needs.

On the other hand, Howard *et. al.*²⁵ used an implicit solvent model based on a inhomogeneous permittivity⁵⁵⁻⁵⁷ which depends locally on the electronic density. The dielectric constant transition smoothly between $\epsilon = 1$ for high density region, i.e. close to solute, towards the permittivity of the homogeneous fluid far from it.

1. Benchmark of water in vacuum

Our primary goal is to be able to compare our QM/MDFT results with existing data. However, different studies often use different geometries and basis sets. Therefore, in order to allow for a fair comparison, we start by discussing how the geometry, the basis set and the level of theory influence the predicted excitation energies in vacuum.

We start by assessing the impact of the solute coordinates by comparing the predictions using the geometries reported by Kongsted *et. al.*⁵⁰ and Howard *et. al.*²⁵. In Table I we report the excitation energies computed both at the EOM-CCSD and SCI level of theory, in the Dunning aug-cc-pVXZ basis sets. In Kongsted’s geometry, the OH bond length is about 0.01 \AA larger and the HOH angle is wider by roughly 1° with respect to Howard’s. From Table I we can notice that the geometry plays a substantial role because, despite their similarities, the excitation energies computed using Howard’s geometry are between 0.04 eV to 0.08 eV lower than the values obtained using Kongsted’s. This is a significant difference since our target accuracy is about 0.04 eV (*i.e.* 1 kcal/mol).

Interestingly, the changes in excitation energy are very similar for both geometries when changing the basis set. We will thus focus on Howard’s geometry to discuss the influence of the basis set, as similar conclusions can be drawn with Kongsted’s geometry. Within the EOM-CCSD framework, the changes in excitation energy between the aug-cc-pVDZ and aug-cc-pVTZ basis sets are 0.16 eV, 0.15 eV and 0.11 eV for the first, second and third excited state, respectively, while using the SCI level of theory they are slightly smaller (*i.e.* 0.09 eV, 0.09 eV, 0.04 eV, respectively). Nevertheless, using the EOM-CCSD method, the change in excitation energies between the aug-cc-pVTZ and d-aug-cc-pVTZ basis sets is much smaller (*i.e.* 0.01 eV, 0.01 eV

and 0.09 eV, for the first, second and third excited states, respectively), and we can expect similar trends at the SCI level of theory. We can conclude that an aug-cc-pVTZ basis set is sufficient for the description of the first and second excited state, while the aug-cc-pVDZ basis set is sufficient for the third excited state as error cancellations lead to an excitation energy close to the d-aug-cc-pVTZ description.

Eventually, we can compare the results obtained with different levels of theory, *i.e.* EOM-CCSD and SCI, keeping in mind that our converged SCI results converge towards the Full Configuration Interaction (FCI) and therefore are the reference within a given basis set. From the results of Table I we can notice that, in the aug-cc-pVDZ basis set, the EOM-CCSD method systematically underestimates the excitation energies by typically the same amount for both geometries, which ranges between 0.07 eV to 0.11 eV. Interestingly, when using the aug-cc-pVTZ basis set, the bias of EOM-CCSD excitation energies with respect to SCI is significantly lowered, ranging from 0.02 eV to 0.05 eV.

To conclude this part, it appears clearly that the geometry, the QM method and the basis set can all together have a significant influence on the predicted excitation energy. For instance, the first excitation energy differs by about 0.3 eV between the EOM-CCSD in the aug-cc-pVDZ basis set using Howard’s geometry and the SCI in the aug-cc-pVTZ basis set using Kongsted’s geometry. Regarding both the choice of the QM method and the basis set, it is of course driven by computational resources considerations. While in the case of a small system such as the water molecule we could obtain near FCI calculations, it becomes soon impossible when increasing the system size, even with efficient SCI algorithms. In the case of the water molecule, we can safely estimate that the EOM-CCSD in the aug-cc-pVTZ is within 0.05 eV to FCI, and that the use of larger basis set only impacts the third excited state by about 0.1 eV. This allows to estimate the first source of error between our framework and the literature. Finally, the choice of the geometry may have the least straightforward effect to rationalize, as even a slight change in geometry can impact by about 0.08 eV the predicted excitation energies. This can become problematic since while it is rather straightforward to define a geometry in vacuum - for instance by geometry optimization -, this not the case in solution. Indeed, the geometry of the solute is most likely fluctuating due to thermal agitation. It is crucial to keep this in mind when comparing theoretical predictions to experimental data.

2. Water in water

We now turn to the study of water, described at the QM level, solvated into liquid water, described at the MDFT level using the methodology described above. An important point to remember is that the solvent is not allowed to relax during the excitation process. Consequently, even in the computation of the excited states, the solvent charge density is the one in equilibrium with the ground-state electronic density. Thus, regardless of the choice of the QM method used to compute the excitation energies (e.g., SCI in our case), the solvatochromism ultimately depends on the classical solvent charge density obtained through the self-consistent QM/MDFT procedure.

Since the QM/MDFT procedure involves multiple QM and MDFT serial calculations, it is worth investigating how the predicted excitation energies are impacted by the level of theory used to describe the ground state electronic den-

sity in this process.

To address this question, three wave function methods of increasing level of complexity were employed to predict the solvent structure with the QM/MDFT procedure: HF, CISD and SCI. Each QM method produces a different solvent density, which are labeled ρ_0 , ρ_2 and ρ_∞ , respectively, according to the highest level of excitation considered in the method.

The excitation energies computed with SCI in the aug-cc-pVDZ basis set using the equilibrium solvent densities are reported in Table II .

As appears in Table II, accounting for the solvent environment results in a notable increase in the excitation energies. This is true for both the geometries and for all three QM level of description tested to describe the ground state density during the QM/MDFT calculation. This observation is relatively easy to rationalize: since the solvent density is optimized for the ground-state electronic density, it tends to stabilize the ground-state energy more than the excited states, leading to a blue-shift in the excitation energies.

More importantly, the excitation energies exhibit a weak dependency on the chosen solvent density, with the observed deviation being at most 0.04 eV (less than 1 kcal/mol). The latter is very encouraging because it means that, at least for this system, performing the QM/MDFT optimization at the simple HF level is sufficient to determine the solvent equilibrium density, which can subsequently serve in more elaborate quantum methods. We can also remark that the excitation energies computed using ρ_∞ are lower than the ones computed using ρ_2 , which are themselves lower than the ones computed using ρ_0 . This can be rationalized by noticing that the permanent dipole of the water molecule, computed in vacuum and reported in Table III, follows a similar trend: $\mu_{\text{HF}} > \mu_{\text{CISD}} > \mu_{\text{SCI}}$.

Therefore, a more polar solute tends to induce larger polarization in the surrounding solvent, which itself leads to stronger stabilization of the ground state compared to the excited states. This justifies the observed trend where both the dipole moment and the excitation energies follow the same pattern.

We also report in Table IV the ground state and excited states dipole moments of the water molecule in solution, computed at the SCI level. As observed in Table IV, the dipole moment of the ground state increases in solution with respect to vacuum, as expected due to the polarisation of the QM solute by the polar solvent. The strength of the ground state dipole moment of the water molecule follows a trend similar to the one observed for the excitation energies: $\mu_{\text{GS}}[\rho_0] > \mu_{\text{GS}}[\rho_2] > \mu_{\text{GS}}[\rho_\infty]$. This is again a consequence of the fact that the solvent density predicted with HF, ρ_0 , is more polarized than the densities predicted with the correlated methods ρ_2 and ρ_∞ .

Regarding now the dipole moments of the excited states, one can first notice that in vacuum they have an opposite sign with respect to the ground state. The latter indicates a flipping of the charge distribution during excitation. In the presence of the solvent, one can notice that, with respect to the vacuum calculations, the dipole moments of the excited states are much lower in absolute value: they tend to align with the ground state dipole. This can be explained since the solvent density is optimised for the ground state charge distribution, it therefore favors dipole moments aligned with the one of the ground state. The excited states have then to adapt to this surrounding electrostatic potential and tend to flip.

The tendency of the solvent to favor dipole moments

State	QM/MDFT			Vacuum
	ρ_∞	ρ_2	ρ_0	
1B_1	8.04/8.12	8.05/8.12	8.08/8.15	7.47/7.54
1A_2	9.94/10.00	9.95/10.00	9.98/10.03	9.26/9.33
1A_1	10.42/10.46	10.43/10.46	10.46/10.49	9.92/9.94

Table II. H₂O, aug-cc-pVDZ: excitation energies computed with SCI in vacuum and in solution using the solvent charges obtained with our QM/MDFT approach. The solvent density obtained using HF, CISD and SCI to describe the ground state electronic density in the QM/MDFT procedure are denoted as ρ_0 , ρ_2 and ρ_∞ respectively. Calculations are reported with the geometry of Howard *et al.* (left) and that of Kongsted *et al.* (right).

QM Method	SCI	CISD	HF
μ	0.732	0.745	0.796

Table III. Dipole moments μ_i (u.a) of the ground state of the water molecule, computed in vacuum using HF, CISD & SCI with the aug-cc-pVDZ basis set. Calculations were made for Howard’s geometry.

State	Vacuum	Solvent		
		ρ_∞	ρ_2	ρ_0
GS	0.732	0.880	0.882	0.891
1B_1	-0.548	-0.028	-0.020	0.021
1A_2	-0.397	-0.083	-0.078	-0.054
1A_1	-0.434	0.074	0.082	0.120

Table IV. Dipole moments (u.a) of the ground state and three first excited state of the water molecule in vacuum and water. All calculations were performed at the SCI level. Calculations in solution were performed with the solvent charge density generated using different QM methods: HF, CISD & SCI. Calculations were made for Howard’s geometry using the aug-cc-pVDZ basis set.

aligned along the O-H direction is illustrated in a pictorial way in Figure 1 where the charge density of the solvent surrounding the solute is displayed in the parallel and the bisector plane of the molecule. The first solvation shell is characterised by a positive charge surrounding the oxygen atom while a negative charge surrounds the hydrogen centers. The charged regions in the immediate vicinity of the solvent are surrounded by regions of opposite charges, which is typical of a polar solvent. The second and third solvation shells are also captured, as evidenced by the oscillating behavior of the charge distribution. Overall, MDFT is able to predict the non-trivial three-dimensional structure of the solvent, as shown by the two slices displayed in Figure 1. In particular, the tetrahedral shape of the charge distribution, resulting from hydrogen bonding, is well rendered. This is a clear advantage of MDFT as compared to simpler CSM, which do not capture the molecular nature of the solvent but instead describe it as a continuum.

We now compare the excitation energies predicted with the proposed SCI/MDFT framework to the results obtained by Howard *et al.* using EOM-CCSD/PCM and by Kongsted *et al.* using the CCSD-MM. We report in Table V the excitation energies obtained with SCI/MDFT in the aug-cc-pVDZ and aug-cc-pVTZ basis sets for water in an environment described by the solvent density ρ_2 , *i.e.* a density generated by a solute description at the CISD level. We performed computations with both the geometry of Howard *et al.* and Kongsted *et al.*

A first observation is that our predicted excitation energies are consistently higher than the one obtained using

the CSM, by a 0.2 to 0.4 eV. However, the agreement with the explicit modelling of the solvent is much better. This could be related to the similar level of description in MD and MDFT. The water model employed in MDFT being non-polarizable, we start by comparing to the non-polar CCSD-MM description. We can notice that we underestimate the first and third excitation energy by 0.02 eV and 0.15 eV, respectively while overestimating the second excited state by 0.08 eV. Overall, the agreement is quite satisfactory considering the numerous differences between the two calculations. For instance, the QM methods differ, which might induce *per se* a difference of about 0.05 eV, as mentioned in Sec. III A 1. Also, the classical water force field are different, and, finally the excess functional of Eq. (15) is approximated and therefore the solvent density is not exact.

Interestingly, the overall agreement with Kongsted’s pre-

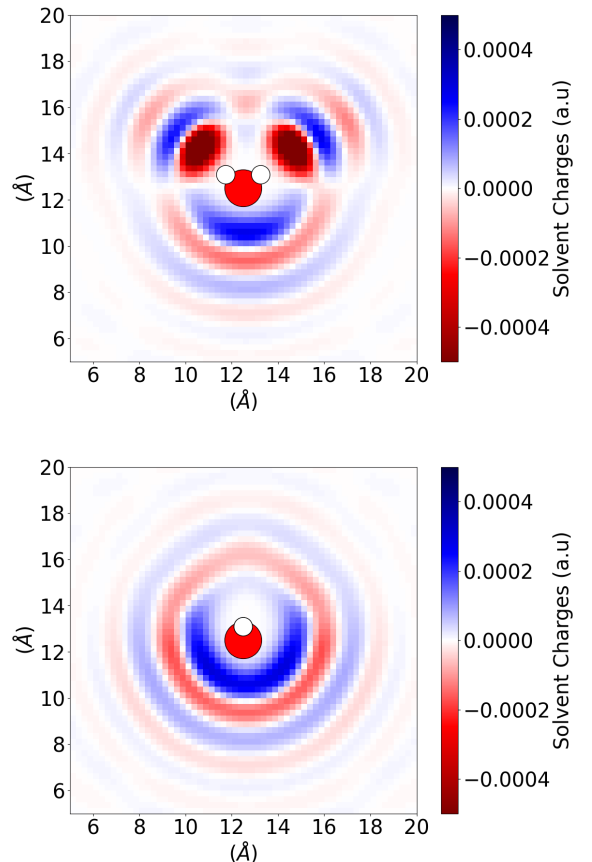


Figure 1. Solvent charges (u.a) in the plane of the molecule (upper) and in the bisector plane (lower). Calculations were made for Howard’s geometry using the aug-cc-pVDZ basis set. CISD has been used to describe the electronic density during the QM/MDFT procedure.

ditions including polarization effect is better, which can likely be attributed to error cancellations.

To conclude this part, we have demonstrated that the proposed QM/MDFT methodology enables the computation of excitation energies and dipole moments. These computations are in good agreement with state-of-the-art computations, in the case of the water molecule. Moreover, the main trends of "blue" solvatochromism can be relatively well rationalized with simple considerations based on electrostatic and dipole moments. From a computational perspective, an important finding is that performing the self-consistent QM/MDFT optimization with a low-level QM method, such as CISD or even HF, does not significantly alter the predicted excitation energies.

States	AVDZ		AVTZ			experiment ⁵⁸
	ρ_2^a	PCM	ρ_2^a	PCM	MD(pol/nopol) ^b	
¹ B ₁	8.05/8.12	7.82	8.15/8.22	7.99	8.18/8.24	8.2
¹ A ₂	9.94/10.00	9.56	10.03/10.10	9.72	9.97/10.02	XX
¹ A ₁	10.43/10.46	10.16	10.48/10.51	10.27	10.56/10.66	9.9

Table V. Excitation energies (u.a) for the water molecule in spce water solvent in aug-cc-pVDZ and aug-cc-pVTZ basis sets compared with results from the literature.

^a: results (left) are obtained with the geometry of Howard while results (right) are obtained with the geometry of Kongsted.

^b: results from Kongsted *et. al.* using an d-AVTZ basis set with a polarizable or non polarizable force-field.

B. Formaldehyde in Water

We now turn to the study of excitation properties of formaldehyde, a small organic molecule, in aqueous solution. This choice was partly guided by the fact that it was also studied by Kongsted *et. al.*²⁷ and Howard *et. al.*²⁵ using the same methodology as the one applied to the study of water. Here again, Kongsted and Howard's geometry slightly differs, but in both cases the molecules are planar with a C₂ symmetry axis along the CO bond. Notably, Kongsted *et. al.* only studied the first excited state of formaldehyde, which corresponds to the $n \rightarrow \pi^*$ transition while Howard *et. al.* computed the 8 first excited states. Our study focused on the four first excited states. Similarly to the study of water, several factors may impact the predicted excited states, such as the choice of the geometry and the LJ parameters used to model dispersion and repulsion, the choice of the solvent model and the level of theory employed to describe the quantum solute. For the sake of compactness, we will limit ourselves to the geometry of Howard *et. al.*. Regarding the choice basis set, Loos *et. al.*³⁵ have shown that going from aug-cc-pVDZ to aug-cc-pVTZ change the excitation energies of the four first excited states of formaldehyde by at most 0.12 eV when using SCI. Since these differences in energy are comparable to the effect of geometry modification, there is no interest to go beyond aug-cc-pVDZ in our calculations. The excitation energies computed in vacuum are reported in Table VI, where they are compared to those obtained by Howard *et. al.* and Kongsted *et. al.*. Comparing the EOM-CCSD results of Howard *et. al.* with our SCI results, we observe a maximum deviation of 0.04 eV for the third excited state, 0.01 eV for the fourth excited state, and below 0.01 eV for the first two excited states. These observations indicate that the EOM-CCSD approach does not introduce bias from the electronic structure for these states. Additionally, when comparing the EOM-CCSD re-

sults of Howard *et. al.* with those of Kongsted *et. al.*, we observe that the impact of changing the geometry on the first excitation energy is only 0.03 eV, suggesting a weak effect.

We now focus on the spectroscopic properties of formaldehyde in aqueous solution. The excitation energies computed using the SCI/MDFT procedure with the solvent charge density predicted using HF, CISD and SCI for the aug-cc-pVDZ basis set are compared to that of Howard *et. al.* and Kongsted *et. al.* in Table VII. The choice of the QM method used to predict the solvent charge has little impact on the excitation energies of the two first excited states, as the energies differences are no greater than 0.04 eV. The impact on the higher excited states is more marked, reaching up to 0.18 eV for the fourth excited state. This might be explained by the fact third and fourth excited states correspond to very diffuse Rydberg states. For the lowest excited state, our computed excitation energies are systematically underestimated with respect to EOM-CCSD/CSM and EOM-CCSD/MD which are consistently larger by roughly 0.2 eV. Again, several approximations can be the cause of this underestimation but we believe that the lack of polarization of our solvent model could play the major role. For the three following excited states, our SCI/MDFT predicts a significant increase in the excitation energies with respect to vacuum. In contrast, Howard *et. al.* surprisingly report almost no solvent effect using EOM-CCSD/CSM. Unfortunately, Kongsted *et. al.* did not study these higher excited states, which would have helped to determinate which results between our SCI/MDFT and EOM-CCSD/CSM are correct.

States	QP	EOM-CCSD	
	SCI	Howard	Kongsted
¹ A ₂	4.03	4.03	4.0
¹ B ₂	7.03	7.03	N/A
¹ B ₂	8.02	7.98	N/A
¹ A ₁	8.04	8.05	N/A

Table VI. Excitation energies (a.u.) of formaldehyde in vacuum computed at the SCI level using aug-cc-pVDZ basis set compared to the predictions of Howard *et. al.* and Kongsted *et. al.* using EOM-CCSSD

States	QP/MDFT			Litt.	
	ρ_∞	ρ_2	ρ_0	CCSD/PCM	CCSD/MD*
¹ A ₂	4.14	4.15	4.17	4.32	4.34
¹ B ₂	7.59	7.61	7.63	7.09	-
¹ B ₂	8.52	8.54	8.63	8.03	-
¹ A ₁	8.59	8.61	8.77	8.13	-

Table VII. Excitation energies (a.u.) of the formaldehyde in aqueous solution. All calculations are made with the aug-cc-pVDZ basis set, except the EOM-CCSSD calculation (*) in which the aug-cc-pVTZ basis set were used.

IV. CONCLUSIONS

In this paper, we propose to couple molecular density functional theory to advanced wavefunction methods as a way to account for solvent effects while computing excited states properties. While this approach belongs to the more general family of QM/MM method, its originality

lies in the description of the solvent through its molecular density. This differs from the conventional MD method in which the solvent is depicted by a set of explicit molecules whose configurations need to be sampled. In practice, the equilibrium solvent density can be computed through functional minimization and the associated three dimensional charge density is directly obtained through a convolution product. From the point of view of the quantum solute, the solvent charge density is simply a set of additional point charges, creating the corresponding Coulomb potential that is added in the one body term in the Hamiltonian. We do not consider any Pauli repulsion or dispersion forces coming from the MM part; therefore, the interaction between the quantum solute and the MM part is purely electrostatic. On the other hand, the electronic density modifies the classical Hamiltonian through the electrostatic potential it generates. To avoid the classical solvent to overlap with the quantum solute, Lennard-Jones sites are added on top of the solute nuclei. This classical model of dispersion and repulsion creates the solute’s cavity in the solvent. The sensibility of the QM/MM predictions on the LJ parameters have already been pointed out^{13,59} but we resort here to parameters from standard MD forcefield without further investigation. Another important physical aspect that is missing from our description is the polarizability of the solvent molecules, which refers to how the electron cloud is altered by the presence of the solute and other solvent molecules. Our QM/MM approach is performed self-consistently until both the MM solvation free energy and the QM ground state energy converge. Once the solvent charge in equilibrium with the ground state is obtained, the excited states properties are computed using converged SCI calculations. This ensures that the electronic structure is described with a nearly exact precision. Therefore, the quality of the excitation energies can be directly attributed to the quality of the equilibrium classical charge density and of the QM/MM coupling. From the computational perspective, several QM and MDFT calculations are required to obtain the equilibrium solvent charge density. We would like to highlight that the cost of an MDFT calculation is independent of the complexity of the solute and relatively inexpensive compared to the QM calculation. Therefore, it is desirable to use a QM method that provides reasonable computational efficiency.

We began our investigation by studying the quantum water molecule in the classical water solvent. We studied the impact of the quality of the ground state QM method (*i.e.* HF, CISD an SCI) on the converged classical solvent charge by comparing the corresponding SCI excitation energies. The predicted SCI excitation energies appear to depend very weakly on the method chosen to optimize the solvent charge. This means that using Hartree-Fock is sufficient to converge the solvent charge. While this assertion have to be further checked for more complex solutes, particularly for strongly correlated systems, it is very encouraging to keep the computational cost of the introduced methodology tractable. The excitation energies predicted using SCI in the presence of the solvent charge obtained with QM/MDFT were found in satisfactory agreement with previously reported high quality QM/MM calculations⁵⁰.

We emphasize that unlike the conventional CSM, the MDFT framework provides a detailed picture of the solvent structure surrounding the solute. For instance, it captures the complex structure of water around the water solute, as illustrated by the tetrahedral shape of the solvent charge density in the first solvation shell, depicted in Fig. 1.

We continued our study by computing the excitation en-

ergies of formaldehyde as similar EOM-CCSD/MD calculations have been previously reported. However, the quality of the results is here more reserved. We underestimate the first excitation energy by roughly 0.2 eV, and it is difficult to assess the quality of the prediction for the three following excited states due to the lack of references.

These results suggest that further developments are needed to refine our QM/MDFT model. We believe that there are two main limitations to the presented approach. Firstly, a framework should be developed to account for the polarizability of the solvent molecules. This could be achieved by making the classical functional dependant on an additional field, namely the electronic polarization field. This field would naturally influence the external part of the functional, which account for the solute-solvent interaction but also the excess part, which models the solvent-solvent interaction. Secondly, the Pauli-dispersion interaction should be included in a more ab-initio manner, both in the quantum part and in the classical part. Overall, the MDFT description of the solvent coupled with advanced wavefunction methods to describe the solute appears to be a promising method to study electronic properties in solution, as it results in a simple one-body operator that could be coupled easily to any QM method or packages.

V. ACKNOWLEDGEMENTS

The authors want to thank Luc Belloni for providing the SPC/E direct correlation functions. Financial support for this project was provided by the Institute of Materials Science (iMAT) of the Alliance Sorbonne Université. ML acknowledges iMAT for a PhD grant.

REFERENCES

- 1E. D. Hughes and C. K. Ingold, *J. Chem. Soc.*, 244 (1935).
- 2V. G. Machado and C. Machado, *J. Chem. Ed.* **78**, 649 (2001).
- 3P. L. Silvestrelli and M. Parrinello, *Phys. Rev. Lett.* **82**, 3308 (1999).
- 4H. Lin and D. G. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2007).
- 5S. Ten-no, F. Hirata, and S. Kato, *Chem. Phys. Lett.* **214**, 391 (1993).
- 6K. Naka, H. Sato, A. Morita, F. Hirata, and S. Kato, *Theoretical Chemistry Accounts* **102**, 165 (1999).
- 7D. Casanova, S. Gusarov, A. Kovalenko, and T. Ziegler, *Journal of Chemical Theory and Computation* **3**, 458 (2007).
- 8N. D. Mermin, *Physical Review* **137**, A1441 (1965).
- 9R. Evans, *Advances in Physics* **28**, 143 (1979).
- 10J. Lischner and T. A. Arias, *Physical Review B* **84**, 125143 (2011).
- 11R. Sundararaman, K. Letchworth-Weaver, K. A. Schwarz, D. Gunceler, Y. Ozhables, and T. Arias, *SoftwareX* **6**, 278 (2017).
- 12W. Tang, C. Cai, S. Zhao, and H. Liu, *The Journal of Physical Chemistry C* **122**, 20745 (2018).
- 13G. Jeanmairet, M. Levesque, and D. Borgis, *Journal of Chemical Theory and Computation* (2020).
- 14M. Karelson and M. C. Zerner, *J Am Chem Soc* **112**, 9405 (1990).
- 15M. M. Karelson and M. C. Zerner, *J Phys Chem* **96**, 6949 (1992).
- 16B. Mennucci, R. Cammi, and J. Tomasi, *J Chem Phys* **109**, 2798 (1998).
- 17J. Tomasi and M. Persico, *Chem Rev* **94**, 2027 (1994).
- 18M. Cossi and V. Barone, *J Chem Phys* **112**, 2427 (2000).
- 19P. Bandyopadhyay and M. S. Gordon, *J Chem Phys* **113**, 1104 (2000).
- 20J.-L. Fattebert and F. Gygi, *J Comput Chem* **23**, 662 (2002).
- 21J.-L. Fattebert and F. Gygi, *Int J Quantum Chem* **93**, 139 (2003).
- 22M. Chiba, D. G. Fedorov, and K. Kitaura, *J Comput Chem* **29**, 2667 (2008).
- 23T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa, and S. Sekiguchi, *J Comput Chem* **31**, 447 (2010), 19499534.

- ²⁴J. Dziedzic, H. H. Helal, C.-K. Skylaris, A. A. Mostofi, and M. C. Payne, *Europhysics Letters* **95**, 43001 (2011).
- ²⁵J. C. Howard, J. C. Womack, J. Dziedzic, C.-K. Skylaris, B. P. Pritchard, and T. D. Crawford, *Journal of Chemical Theory and Computation* **13**, 5572 (2017).
- ²⁶M. S. Gordon, M. A. Freitag, P. Bandyopadhyay, J. H. Jensen, V. Kairys, and W. J. Stevens, *J Phys Chem A* **105**, 293 (2001).
- ²⁷J. Kongsted, A. Osted, K. V. Mikkelsen, P.-O. Åstrand, and O. Christiansen, *J Chem Phys* **121**, 8435 (2004).
- ²⁸A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev, and M. S. Gordon, *J Phys Chem Lett* **2**, 2184 (2011).
- ²⁹C. E. Gonzalez-Espinoza, C. A. Rumble, D. Borgis, and T. A. Wesolowski, *Journal of Chemical Theory and Computation* **18**, 1072 (2022).
- ³⁰B. Huron, J. P. Malrieu, and P. Rancurel, *J. Chem. Phys.* **58**, 5745 (1973).
- ³¹S. Evangelisti, J.-P. Daudey, and J.-P. Malrieu, *Chemical Physics* **75**, 91 (1983).
- ³²E. Giner, A. Scemama, and M. Caffarel, *Can. J. Chem.* **91**, 879 (2013).
- ³³A. A. Holmes, N. M. Tubman, and C. J. Umrigar, *J. Chem. Theory Comput.* **12**, 3674 (2016).
- ³⁴J. B. Schriber and F. A. Evangelista, *J Chem Theory Comput* **13**, 5354 (2017).
- ³⁵P.-F. Loos, A. Scemama, A. Blondel, Y. Garniron, M. Caffarel, and D. Jacquemin, *J Chem Theory Comput* **14**, 4360 (2018).
- ³⁶Y. Garniron, K. Gasperich, T. Applencourt, A. Benali, A. Ferté, J. Paquier, B. Pradines, R. Assaraf, P. Reinhardt, J. Toulouse, P. Barbaresco, N. Renon, G. David, J. P. Malrieu, M. Vêril, M. Caffarel, P. F. Loos, E. Giner, and A. Scemama, *J. Chem. Theory Comput.* **15**, 3591 (2019).
- ³⁷M. Caricato, *J Chem Phys* **139** (2013), 10.1063/1.4816482.
- ³⁸M. De Vetta, M. F. S. J. Menger, J. J. Nogueira, and L. Gonzalez, *The Journal of Physical Chemistry B* **122**, 2975 (2018).
- ³⁹G. Jeanmairet, M. Levesque, R. Vuilleumier, and D. Borgis, *The Journal of Physical Chemistry Letters* **4**, 619 (2013).
- ⁴⁰D. Borgis, S. Luukkonen, L. Belloni, and G. Jeanmairet, *The Journal of Physical Chemistry B* **124**, 6885 (2020).
- ⁴¹D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, second edition ed. (Academic Press, 2002).
- ⁴²J.-P. Hansen and I. McDonald, *Theory of Simple Liquids, Third Edition*, 3rd ed. (Academic Press, 2006).
- ⁴³L. Ding, M. Levesque, D. Borgis, and L. Belloni, *The Journal of Chemical Physics* **147**, 094107 (2017).
- ⁴⁴Freindorf Marek and Gao Jiali, *Journal of Computational Chemistry* **17**, 386 (1998).
- ⁴⁵L. Turi, M.-P. Gaigeot, N. Levy, and D. Borgis, *The Journal of Chemical Physics* **114**, 7805 (2001).
- ⁴⁶A. Marefat Khah, P. Reinholdt, J. M. H. Olsen, J. Kongsted, and C. Httig, *Journal of Chemical Theory and Computation* **16**, 1373 (2020).
- ⁴⁷Q. Sun, X. Zhang, S. Banerjee, P. Bao, M. Barbry, N. S. Blunt, N. A. Bogdanov, G. H. Booth, J. Chen, Z.-H. Cui, J. J. Eriksen, Y. Gao, S. Guo, J. Hermann, M. R. Hermes, K. Koh, P. Koval, S. Lehtola, Z. Li, J. Liu, N. Mardirossian, J. D. McClain, M. Motta, B. Mussard, H. Q. Pham, A. Pulkin, W. Purwanto, P. J. Robinson, E. Ronca, E. R. Sayfutyarova, M. Scheurer, H. F. Schurkus, J. E. T. Smith, C. Sun, S.-N. Sun, S. Upadhyay, L. K. Wagner, X. Wang, A. White, J. D. Whitfield, M. J. Williamson, S. Wouters, J. Yang, J. M. Yu, T. Zhu, T. C. Berkelbach, S. Sharma, A. Y. Sokolov, and G. K.-L. Chan, *The Journal of Chemical Physics* **153**, 024109 (2020).
- ⁴⁸H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *The Journal of Physical Chemistry* **91**, 6269 (1987).
- ⁴⁹L. Belloni, *The Journal of Chemical Physics* **147**, 164121 (2017).
- ⁵⁰J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, *Chemical Physics Letters* **364**, 379 (2002).
- ⁵¹J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, *The Journal of Chemical Physics* **118**, 1620 (2003).
- ⁵²J. Kongsted, A. Osted, K. V. Mikkelsen, and O. Christiansen, *J Phys Chem A* **107**, 2578 (2003).
- ⁵³A. Osted, J. Kongsted, K. V. Mikkelsen, P.-O. strand, and O. Christiansen, *The Journal of Chemical Physics* **124**, 124503 (2006).
- ⁵⁴X. He, V. H. Man, W. Yang, T.-S. Lee, and J. Wang, *The Journal of Chemical Physics* **153**, 114502 (2020).
- ⁵⁵J.-L. Fattebert and F. Gygi, *Journal of Computational Chemistry* **23**, 662 (2002).
- ⁵⁶D. A. Scherlis, J.-L. Fattebert, F. Gygi, M. Cococcioni, and N. Marzari, *The Journal of Chemical Physics* **124**, 074103 (2006).
- ⁵⁷O. Andreussi, I. Dabo, and N. Marzari, *The Journal of Chemical Physics* **136**, 064102 (2012).
- ⁵⁸G. D. Kerr, R. N. Hamm, M. W. Williams, R. D. Birkhoff, and L. R. Painter, *Physical Review A* **5**, 2523 (1972).
- ⁵⁹J. A. Rackers and J. W. Ponder, *The Journal of Chemical Physics* **150**, 084104 (2019).