Fully Polarizable Multiconfigurational Self-Consistent Field/Fluctuating Charge Approach

Chiara Sepali,[†] Linda Goletto,[†] Piero Lafiosca,[†] Matteo Rinaldi,[†] Tommaso Giovannini,[‡] and Chiara Cappelli[*,†,¶]

†*Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

‡*Department of Physics, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133, Rome, Italy*

¶*IMT School for Advanced Studies Lucca, Piazza San Francesco 19, I-55100 Lucca, Italy*

E-mail: chiara.cappelli@sns.it

Abstract

A multiscale model based on the coupling of the multiconfigurational self-consistent field (MCSCF) method and the classical atomistic polarizable Fluctuating Charges (FQ) force field is presented. The resulting MCSCF/FQ approach is validated by exploiting the CASSCF scheme through application to compute vertical excitation energies of formaldehyde and para-nitroaniline in aqueous solution. The procedure is integrated with molecular dynamics simulations to capture the solute's conformational changes and the dynamic aspects of solvation. Comparative analysis with alternative solvent models, gas-phase calculations, and experimental data provides insights into the model's accuracy in reproducing solute-solvent molecular interactions and spectral signals.

Introduction

Multiconfiguration self-consistent field (MCSCF) techniques offer a useful and widely used alternative to single reference quantum mechanical (QM) methods, that often fail when applied to highly correlated chemical systems[1,2] such as carotenoids[3] transition metal complexes,[4] and especially photochemical processes.[5,6]

In MCSCF, wavefunctions are expressed as the superposition of multiple electronic configurations; molecular orbitals (MOs) and configuration interaction (CI) coefficients of electronic configurations are variationally optimized, and MOs are categorized into inactive, active, and virtual spaces.[7–11] Molecular states can be optimized using either a state-specific (SS) approach or a state-average (SA) scheme, in which a weighted average of the energy of all selected states is minimized over a set of common orbitals.[12–14] Currently, the most widespread approach for performing MCSCF calculations consists of resorting to the complete active space self-consistent field (CASSCF) approach, where the wavefunction expansion includes all potential configuration state functions within a given set of active orbitals.[8,11] The scalability of this method has advanced considerably, with the largest calculations performed using an active space of 22 electrons in 22 active orbitals.[15] Also, its accuracy is excellent, especially after incorporating dynamic correlation effects, for instance by using second-order perturbation theory based on a CASSCF reference state, i.e. the CASPT2 method. All this makes the approach a gold standard for accurate ab initio calculations of excited-state energies, properties,[16–22] and especially non-adiabatic excited state dynamics.[20,21,23]

The MCSCF methods, when applied to molecular systems in the gas phase, are highly developed and thoroughly validated. However, most excited-state photophysical and photochemical processes occur in the condensed phase, which is why methods have been developed to couple an accurate MCSCF description with (atomistic or continuum) classical or quantum embedding methods, in a multiscale fashion.[24–30]

Multiscale hybrid QM/Classical models constitute an efficient way to study systems in the condensed phase.[26,28,31–33] In these approaches, the target is accurately treated at the QM level, while the environment is described with classical physics. This allows for taking into account the effects of the environment on the target's properties in a computationally efficient way. The MCSCF method has been integrated with the Polarizable Continuum Model (PCM) in an SS framework,[34–38] and extended to an SA formulation to study conical intersections and photoreaction energy pathways of solvated systems.[39] A bunch of polarizable

atomistic models such as Polarizable Embedding (PE), MMPol, and AMOEBA[40–45] have been coupled with MCSCF wave functions using a variety of strategies, including SS or SA approaches, or hybrid methods in which the solvent polarization is converged to the solute SA density and an SS correction is implemented to refine energy values.[44]

In this work, we propose for the first time the coupling of an MCSCF wavefunction (within a CASSCF formalism) and the polarizable Fluctuating Charge (FQ) force field,[46] in a QM/MM fashion.[47–50] FQ is a polarizable classical force field, where polarization effects are modeled by assigning to each atom belonging to the classical portion (i.e. to the solvent) a charge (q) that is not fixed, like in standard classical force field, but can vary according to the electronegativity equalization principle (EEP).[51] In this way, a charge flow occurs when two atoms have different chemical potentials.[46] When coupled to DFT or TD-DFT Hamiltonians, FQ has been demonstrated to yield excellent results in the simulation of various spectral properties of aqueous and non-aqueous solutions[48–50,52,53]

The paper is organized as follows. First, the theoretical foundations of MCSCF and FQ are recalled, and then the MCSCF/FQ coupling is discussed. After a detailed description of the computational protocol, the method is validated by simulating vertical excitation energies of formaldehyde and para-nitroaniline in aqueous solution. Formaldehyde has been selected for its compact size and is used to set the computational protocol, by examining variations in the basis set, starting orbitals, active space selection, and solvent modeling (from continuum to atomistic methods). Para-nitroaniline is a prototypical push-pull system with a π-conjugated link between electron-donor and electron-acceptor groups. Its optical properties, which are significantly influenced by solvation, make it an ideal system for showing the performance of the newly developed method.[38,54–56] Vertical excitation energies of PNA in aqueous solution are discussed by highlighting the role of dynamical solvent fluctuations and finally compared with experimental data. Conclusion and future perspectives end the paper.

1 Theory

This section presents the coupling of MCSCF with the FQ polarizable force field. Initially, the main features of MCSCF (section 1.1) and FQ (section 1.2) are summarized to provide a background for the description

of their coupling, which is presented in section 1.3. The integration of FQ with MCSCF is challenging, due to the intrinsic complexity of multireference approaches. In this work, MCSCF (CASSCF formulation) is integrated with FQ by exploiting a State-specific (SS) approach, where each solute's state is optimized separately and interacts with the polarizable solvent. This strategy appears reasonable for calculating excited states' energies and properties. However, a state-average (SA) scheme should be preferred to study phenomena involving different excited states, such as conical intersections, photochemistry, and excited state dynamics.[43] The extension of MCSCF/FQ to an SA formalism is not a trivial extension of the SS formulation, therefore it will be treated in future works.

1.1 The MCSCF approach

The MCSCF wavefunction is expressed as a linear combination of Slater determinants $|\phi_m\rangle$ built over a common set of molecular orbitals (MOs), which in turn are expressed as a linear combination of atomic orbitals (AOs).[57] By following the notation of Roos,[57] the state of interest $|\Psi_0\rangle$ and the excited states $|\Psi_K\rangle$ assume the following form:

$$|\Psi_0\rangle = \sum_m C_{m0} |\phi_m\rangle , \quad |\Psi_K\rangle = \sum_m C_{mK} |\phi_m\rangle \tag{1}$$

where C_{m0} and C_{mK} indicate CI coefficients. In MCSCF, $|\Psi_0\rangle$ is variationally optimized with respect to C_{m0} and MOs coefficients. To this end, the MCSCF wavefunction $|\Psi_0\rangle$ is conveniently expressed as follows:[57]

$$|\Psi_0\rangle = \text{exp}(\hat{T})\text{exp}(\hat{S}) |0\rangle \tag{2}$$

where $|0\rangle$ is a reference state, while exp($\hat{T}$) and exp($\hat{S}$) are unitary transformations among MOs and states, respectively. $\hat{T}$ and $\hat{S}$ are anti-Hermitian operators which read:

$$\hat{T} = \sum_{pq} T_{pq} \hat{E}_{pq}; \quad \hat{S} = \sum_{K\neq 0} S_{K0}(|\Psi_K\rangle \langle\Psi_0| - |\Psi_0\rangle \langle\Psi_K|) \tag{3}$$

where $\hat{E}_{pq}$ is the singlet excitation operator for generic p, q MOs , while T_{pq} and S_{K0} are variational parameters.

The energy of $|\Psi_0\rangle$ can thus be written as:

$$E_{\mathrm{MCSCF}} = \langle 0| \exp(-\hat{S})\exp(-\hat{T})\hat{H}\exp(\hat{T})\exp(\hat{S}) |0\rangle \tag{4}$$

where $\hat{H}$ indicates the Hamiltonian, which reads (in second quantization):

$$\hat{H} = \sum_{pq} h_{pq}\hat{E}_{pq} + \frac{1}{2}\sum_{pqrs} g_{pqrs}(\hat{E}_{pq}\hat{E}_{rs} - \delta_{rq}\hat{E}_{ps}) \tag{5}$$

h_{pq} and g_{pqrs} are one- and two-electron integrals in the MO basis, respectively. Equation (6) can be re-written as a function of $\mathbf{D}$ and $\mathbf{P}$, i.e. the one-particle and two-particle density matrices associated with $|\Psi_0\rangle$:[58]

$$E_{\mathrm{MCSCF}}(\mathbf{D},\mathbf{P}) = \sum_{pq} D_{pq}h_{pq} + \sum_{pqrs} P_{pqrs}g_{pqrs} \tag{6}$$

By applying the Baker-Campbell-Hausdorff (BCH) expansion, the electronic MCSCF gradients and Hessians are obtained by differentiating the energy with respect to the wavefunction parameters. The stationarity of the energy functional is guaranteed when gradients with respect to CI coefficients ($g_K^{(c)}$) and MOs coefficients ($g_{rs}^{(o)}$) satisfy the two following conditions:

$$g_K^{(c)} = 2\langle\Psi_0|\hat{H}|\Psi_K\rangle = 0 \tag{7}$$

$$g_{rs}^{(o)} = \langle\Psi_0|[\hat{H},\hat{E}_{rs}^{-}]|\Psi_0\rangle = 2\langle\Psi_0|\hat{H}\hat{E}_{rs}^{-}|\Psi_0\rangle = 2\langle\Psi_0|\hat{H}|rs\rangle = 0 \tag{8}$$

where $\hat{E}_{rs}^{-} = \hat{E}_{rs} - \hat{E}_{sr}$ and $|rs\rangle = \hat{E}_{rs}^{-}|\Psi_0\rangle$ are the so-called Brillouin states.[59] Equation (7) is equivalent to the secular problem $(\mathbf{H} - E\mathbf{1})\mathbf{C} = \mathbf{0}$ and eq. (8) is the Brillouin-Levy-Berthier (BLB) theorem, also known as the *Extended Brillouin Theorem*.[57,60,61]

The optimization of MCSCF wavefunctions is generally carried out using first-order or second-order methods (see Refs. 1,62 and references therein) depending on whether the MOs and CI coefficients are optimized alternatively or simultaneously, respectively. In this work, we resort to a first-order approach implemented in OpenMolcas,[63,64] in particular by adopting the Super-CI (SCI) approach[8,9,65,66] based on the BLB theorem.[60,61] SCI can be adopted in combination with the most common method for MCSCF

calculations, i.e., the complete active space self-consistent field (CASSCF) approach, where the wavefunction expansion includes all potential configuration state functions within a given set of active orbitals.[8,11] CASSCF is the method employed in the present study.

In SCI, the CI coefficients are calculated by solving the secular problem (see eq. (7)) for a given set of MOs. The improved SCI wavefunction is then set up as a superposition of $|\Psi_0\rangle$ and all the Brillouin states corresponding to non-redundant orbital rotations as follows:[65]

$$|\mathrm{SCI}\rangle = |\Psi_0\rangle + \sum_{p>q} T_{pq}\hat{E}^{-}_{pq}|\Psi_0\rangle = |\Psi_0\rangle + \sum_{p>q} T_{pq}|pq\rangle \tag{9}$$

Therefore, the T_{pq} coefficients are determined by solving the generalized eigenvalue equation:[65]

$$\mathbf{HT} = E_{\mathrm{SCI}}\overline{\mathbf{S}}\mathbf{T} \tag{10}$$

where $H_{rs,pq} = \langle rs|\hat{H}|pq\rangle$ is the Hamiltonian matrix and $\overline{\mathbf{S}}$ is the overlap matrix in the basis of $|\Psi_0\rangle$ and Brillouin states.[59] Because calculating the matrix elements between the Brillouin states $\langle rs|\hat{H}|pq\rangle$ is extremely expensive, alternative schemes have been proposed and amply tested in the literature.[8,65,66]

1.2 The Fluctuating Charges (FQ) force field

The FQ force field[46,67] endows each atom with a charge q which dynamically adapts to the external potential. The total Lagrangian functional of the FQ model is obtained from a second-order Taylor expansion of the energy with respect to charges.[46] For a system composed of various molecules, it is generally written as follows:[46]

$$E_{\mathrm{FQ}} = \sum_{i\alpha} \mathrm{q}_{i\alpha}\chi_{i\alpha} + \frac{1}{2}\sum_{i\alpha}\sum_{j\beta} \mathrm{q}_{i\alpha}\mathrm{T}^{\mathrm{qq}}_{i\alpha,j\beta}\mathrm{q}_{j\beta} + \sum_{\alpha}\left[\lambda_\alpha \sum_i \mathrm{q}_{i\alpha} - \mathrm{Q}_\alpha\right] \tag{11}$$

where (i,j) and (α,β) indices run over FQ atoms and molecules, respectively. $\chi_{i\alpha}$ indicates the atomic electronegativity, while $\mathrm{T}^{\mathrm{qq}}_{i\alpha,j\beta}$ is the charge-charge interaction kernel. To prevent the so-called "polarization catastrophe",[68] the Ohno kernel is adopted.[69] Its diagonal elements $T^{qq}_{i\alpha,i\alpha}$ are expressed in terms of the atomic chemical hardness $\eta_{i\alpha}$. A set of Lagrangian multipliers λ_α are introduced in eq. (11) to constrain the total charge of each molecule to Q_α. This is done to avoid artifacts arising from unphysical charge transfer

between different molecules.[70] As a result of its formulation, FQ relies on two atomic parameters, $\chi_{i\alpha}$ and $\eta_{i\alpha}$, which can be rigorously defined in the framework of Conceptual Density Functional Theory.[71,72]

Charge exchange between atoms is driven by the Electronegativity Equalization Principle,[51] which is satisfied by imposing the stationarity conditions of the Lagrangian functional with respect to atomic charges and multipliers. This translates into solving the following linear system:[47,48]

$$\begin{pmatrix} \mathbf{T}^{\mathrm{qq}} & \mathbf{1}_\lambda \\ \mathbf{1}_\lambda^\dagger & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\chi} \\ \mathbf{Q}_\alpha \end{pmatrix} \tag{12}$$

where $\mathbf{1}_\lambda$ are rectangular blocks collecting Lagrangian multipliers.

1.3 The MCSCF/FQ approach

MCSCF and FQ force field are coupled in a QM/MM fashion. The starting point is the definition of the energy of the total system:

$$E_{tot} = E_{\mathrm{MCSCF}} + E_{\mathrm{FQ}} + E^{\mathrm{int}}_{\mathrm{MCSCF/FQ}} \tag{13}$$

where E_{MCSCF} is defined in eq. (6), and E_{FQ} in eq. (11).

The interaction energy $E^{\mathrm{int}}_{\mathrm{MCSCF/FQ}}$ is expressed in terms of the electrostatic interaction between the FQ charges and the QM solute. It can be written as:[47,48]

$$E^{\mathrm{int}}_{\mathrm{MCSCF/FQ}} = \sum_{i\alpha} q_{i\alpha} V_{i\alpha}(\mathbf{D}) \tag{14}$$

where $\mathbf{D}$ is the QM one-particle density matrix, and $V_{i\alpha}(\mathbf{D})$ is the molecular QM electrostatic potential acting on the charge $q_{i\alpha}$, i.e.:

$$V_{i\alpha}(\mathbf{D}) = \sum_{N}^{\mathrm{nuclei}} \frac{Z_N}{|\mathbf{r}_{i\alpha} - \mathbf{R}_N|} - \sum_{pq} D_{pq} V^{\mathrm{FQ}}_{pq,i\alpha} \tag{15}$$

In eq. (15), the first term is the nuclear potential, generated by the nucleus N with charge Z_N located at

position $\mathbf{R}_N$, whereas the second term is the electronic potential, defined as:

$$V^{\mathrm{FQ}}_{pq,i\alpha} = \left\langle \phi_p \left| \frac{1}{|\mathbf{r} - \mathbf{r}_{i\alpha}|} \right| \phi_q \right\rangle \tag{16}$$

where $\mathbf{r}_{i\alpha}$ is the position of the i-th FQ belonging to molecule α.

Within a SS approach, the MCSCF wavefunction is optimized with reference to a specific state ℓ, which interacts with the polarizable environment. Thus, the total MCSCF/FQ energy functional defined in eq. (13) becomes:

$$\begin{aligned} E^{\ell}(\mathbf{D}^{\ell}, \mathbf{P}^{\ell}, \mathbf{q}, \boldsymbol{\lambda}) =& E_{\mathrm{MCSCF}}(\mathbf{D}^{\ell}, \mathbf{P}^{\ell}) \\ &+ \sum_{i\alpha} q_{i\alpha}(\mathbf{D}^{\ell}) \chi_{i\alpha} + \frac{1}{2} \sum_{i\alpha, j\beta} q_{i\alpha}(\mathbf{D}^{\ell}) T^{\mathrm{qq}}_{i\alpha, j\beta} q_{j\beta}(\mathbf{D}^{\ell}) + \sum_{\alpha} \lambda_{\alpha} \left[\sum_{i} q_{i\alpha}(\mathbf{D}^{\ell}) - Q_{\alpha} \right] \\ &+ \sum_{i\alpha} q_{i\alpha}(\mathbf{D}^{\ell}) V_{i\alpha}(\mathbf{D}^{\ell}) \end{aligned} \tag{17}$$

where $\mathbf{D}^{\ell}$ and $\mathbf{P}^{\ell}$ represent the one-particle and two-particle density matrices of the state ℓ. FQ charges are obtained by minimizing the total functional in eq. (17) with respect to FQ charges and Lagrangian multipliers. Therefore, the linear system in eq. (12) is modified by accounting for the QM electronic potential as an additional polarization source, i.e.:

$$\begin{pmatrix} \mathbf{T}^{\mathrm{qq}} & \mathbf{1}_{\lambda} \\ \mathbf{1}^{\dagger}_{\lambda} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{q}(\mathbf{D}^{\ell}) \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\boldsymbol{\chi} \\ \mathbf{Q}_{\alpha} \end{pmatrix} + \begin{pmatrix} -\mathbf{V}(\mathbf{D}^{\ell}) \\ \mathbf{0} \end{pmatrix} \tag{18}$$

To account for mutual solute-solvent polarization, the MCSCF Hamiltonian in eq. (5) is also perturbed by the presence of the FQ charges as follows:

$$\hat{H} = \sum_{pq} \left[h_{pq} + [\mathbf{q}(\mathbf{D}^{\ell})]^{\dagger} \mathbf{V}^{\mathrm{FQ}}_{pq} \right] \hat{E}_{pq} + \frac{1}{2} \sum_{pqrs} g_{pqrs} \left(\hat{E}_{pq} \hat{E}_{rs} - \delta_{rq} \hat{E}_{ps} \right) \tag{19}$$

where $\mathbf{V}^{\mathrm{FQ}}_{pq}$ is the one-electron operator defined in eq. (15). In addition, gradients with respect to CI and MOs coefficients defined in eqs. (7) and (8), respectively, are modified to account for the additional FQ

contributions, i.e.:

$$g^{(c)}_{\mathrm{tot},K} = g^{(c)}_{K} + 2\,\langle\Psi_0|\sum_{pq}[\mathbf{q}(\mathbf{D}^{\ell})]^{\dagger}\mathbf{V}^{\mathrm{FQ}}_{pq}\hat{E}_{pq}|\Psi_K\rangle \tag{20}$$

$$g^{(o)}_{\mathrm{tot},rs} = g^{(o)}_{rs} + 2\,\langle\Psi_0|\sum_{pq}[\mathbf{q}(\mathbf{D}^{\ell})]^{\dagger}\mathbf{V}^{\mathrm{FQ}}_{pq}\hat{E}_{pq}|rs\rangle \tag{21}$$

These gradient expressions are used during the optimization of the MCSCF wavefunction. In particular, $g^{(c)}_{\mathrm{tot},K}$ is used to obtain the CI coefficients from the secular equation (see eq. (7)), whereas $g^{(o)}_{\mathrm{tot},pq}$ enters the SCI eigenvalue equation for the determination of the MO coefficients (see eq. (10)).

To summarize, an SS MCSCF/FQ calculation implies:

1. computing initial $\mathbf{T}^{(0)}$ and $\mathbf{S}^{(0)}$ values;
2. calculating the starting density matrices $\mathbf{D}^{\ell,(0)}$ and $\mathbf{P}^{\ell,(0)}$;
3. computing the starting FQ charges $\mathbf{q}^{(0)}(\mathbf{D}^{\ell})$ from eq. (18);
4. for $k = 1, 2, \ldots$ until convergence:
   (a) $\mathbf{S}^{(k)}$ is computed from eq. (7) with the inclusion of FQ contributions in eq. (20);
   (b) $\mathbf{T}^{(k)}$ are calculated from eq. (10) with the inclusion of FQ contributions in eq. (21);
   (c) FQ charges $\mathbf{q}^{(k)}(\mathbf{D}^{\ell})$ are updated from eq. (18);
   (d) the SS MCSCF/FQ energy is finally computed by exploiting eq. (17).

The computational procedure may be generalized to SA MCSCF schemes, by taking inspiration from what has been proposed for alternative polarizable embedding methods.[43,45] Because such a generalization is not straightforward, this work limits the treatment to SS CASSCF wavefunctions. Extension to alternative MCSCF schemes will be presented in a future communication.

2 Computational Details

The quality of the approach is investigated by computing vertical excitation energies of formaldehyde (FORM) and para-nitroaniline (PNA) solvated in aqueous solution. A multi-step protocol, adapted from a protocol

that was specifically designed for modeling spectral signals of molecules in solution[48] at the QM/MM level is exploited. The protocol consists of the following steps:[48,73]

1. *Definition of the system*: The QM and MM regions are defined according to the physicochemical features of the system.[48] In this case, the solutes (FORM and PNA) are treated at the CASSCF level, while the polarizable FQ force field is used to represent the aqueous solvent.

2. *Conformational sampling*: In line with previous studies of some of the present authors,[48] the solute-solvent phase-space is explored by using classical non-polarizable molecular dynamics (MD) simulations to obtain a representative conformational sampling of the solvated system. This strategy has been proven to be particularly effective when combined with QM/FQ to describe various kinds of spectral signals.[48,74] Specifically, a 5 ns MD simulation of aqueous FORM (NVT) using the GROMACS package[75] is performed using the general AMBER force field (GAFF) to treat intramolecular FORM (fixed at the equilibrium geometry) and FORM-water intermolecular interactions, while TIP3P is exploited for water molecules [see section S1 in the Supporting Information (SI) for further details]. In the case of PNA, two NVT MD simulations are considered: i) the PNA geometry is kept fixed in its minimum energy, as optimized at the CAM-B3LYP/aug-cc-pVDZ/PCM level of theory (MD_{rigid}) or ii) PNA can move freely (MD_{free}). The two MDs are taken from Refs. 52 (2.5 ns production run) and 56 (5 ns production run), respectively. In both cases, intra- and intermolecular interactions are described by the GAFF force field, while TIP3P is used for water molecules.

3. *Extraction of structures*: A set of uncorrelated snapshots is extracted from the production phase of each MD run. 200 snapshots of both FORM and PNA are extracted from their respective MD simulations, with PNA snapshots derived from both MD_{rigid} and MD_{free}. For each snapshot, a spherical droplet is cut. The droplet is centered on the solute with a radius large enough to account for relevant solute-solvent interactions. Specifically, the radius is 15 Å for FORM and 18 Å for PNA. Representative snapshots of FORM and PNA are depicted in fig. 1.

4. *QM/MM calculations*: For each spherically-shaped snapshot, vertical excitation energies are calculated at the CASSCF/FQ level using two parametrizations (FQ^a, taken from Ref. 46 and FQ^b, taken from Ref. 76). To highlight the effect of solute-solvent polarization, additional calculations employing

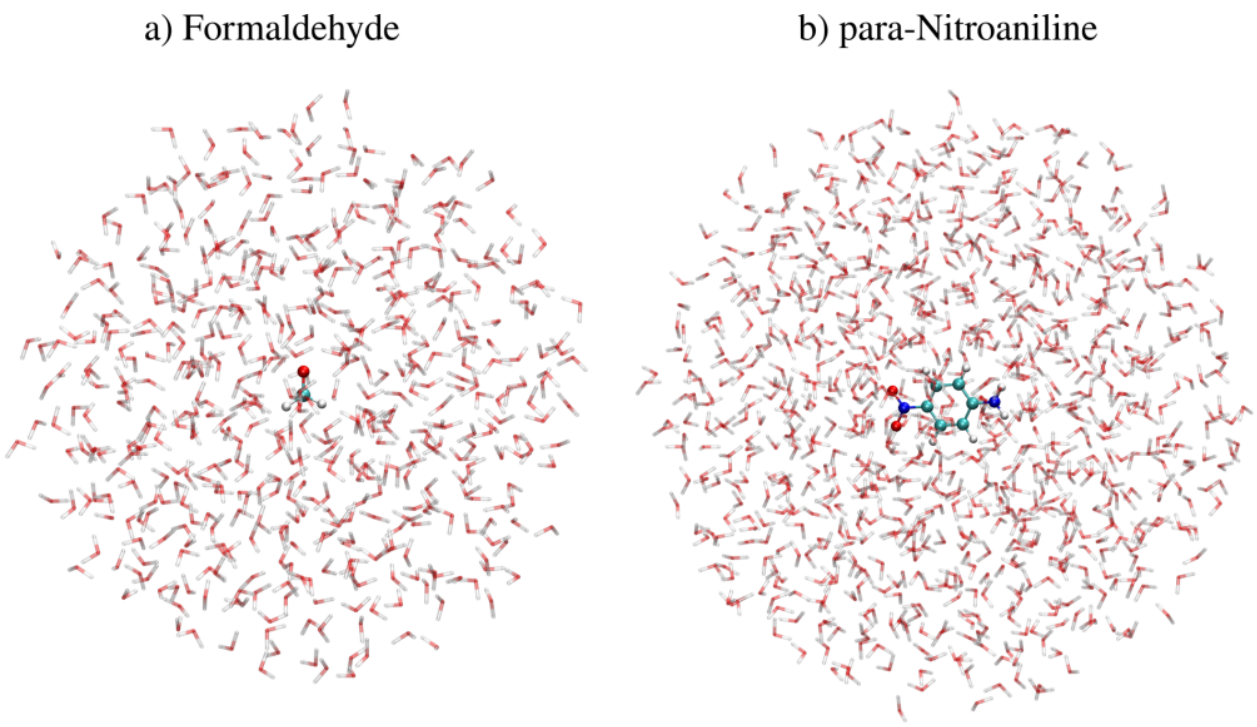


Figure 1: Representative snapshots of formaldehyde (a) and para-nitroaniline (b) in aqueous solution.

the non-polarizable ElectroStatic Potential Fitted (ESPF)[77] approach are performed, using TIP3P charges.[78]

CASSCF computed values depend on several parameters, such as the choice of starting orbitals, the active space, and the basis set.[63,64,79–82] In this paper, the following procedure is exploited:

- Calculation of starting orbitals [at the HF[57]/(classical) or CASSCF/(classical) level] using a small basis set (ANO-RCC-MB).[83] This procedure permits to easily determine the MOs defining in active space;
- Selection of the MOs and definition of the active space;
- Projection of the ANO-RCC-MB basis set into the selected, larger, basis set.;[63,64]
- CASSCF-SS(/classical) calculations on the ground state (GS) and excited state (ES). The vertical excitation energy is computed as the difference between CASSCF-ES and CASSCF-GS energies. QM/MM calculations are performed on a subset of extracted snapshots to ensure that active space and electronic transitions are consistent.

It is important to highlight that a single-state CASSCF/FQ method is employed for calculating the

GS and ES, i.e., a separate CASSCF calculation is performed for both the GS and the ES. FQ charges are equilibrated to the state being optimized, thus describing the polarization of the solvent to each electronic state. Consequently, the GS and ES Hamiltonians differ because they incorporate a different FQ contribution, resulting in non-orthogonal wave functions. This methodology has already been proposed and assessed in the context of MCSCF/Classical models and can be effectively applied to study low-lying excited states.[34,35,40] However, single-state MCSCF/Classical approaches may encounter convergence issues, such as root flipping,[44] and the electronic states obtained from different Hamiltonians can lead to ill-defined transition matrix elements.[84] Several alternative methods have been proposed in the literature to address these limitations.[44,85–87]

For analyzing solvent effects and specific solute-solvent interactions, vertical excitation energies are also computed in the gas phase and by treating the solvent at the implicit PCM level.[26] In both cases, FORM and PNA geometries are optimized with Gaussian16,[88] by using the B2PLYP double hybrid functional combined with the maug-cc-pVTZ-d(H) basis set.[76,89] CASSCF/PCM calculations are performed within the non-equilibrium regime.[38] All CASSCF(/classical) calculations are performed by exploiting a locally modified version of OpenMolcas,[63,64] where CASSCF/FQ has been implemented.

5. *Extraction of final spectra*: For each system, the CASSCF/MM spectrum is obtained by averaging absorption spectra of all snapshots, after convolution with a Gaussian function with a full width at half maximum (FWHM) of 0.3 eV.

3 Results and Discussion

This section starts with a preliminary validation step, which is performed by computing vertical excitation energies of a single structure of FORM in aqueous solution, in particular by analyzing the differences arising from the selection of the basis set, starting orbitals, active spaces, and solvent models. Then, vertical excitation energies of FORM and PNA in aqueous solution are discussed, by highlighting the role of solvent dynamical fluctuations and comparing computed and experimental data.

3.1 Model Validation

To evaluate the quality of CASSCF/FQ, the $n - \pi^*$ excitation energy of FORM in aqueous solution is selected. A single snapshot, randomly extracted from the MD trajectory, is considered. As described in Section 2, vertical excitation energies are computed as the difference between ES and GS SS-CASSCF calculations. Starting orbitals are determined by performing QM(/classical) calculations at the HF level, at the CASSCF-SS level by optimizing the GS (CASSCF-GS), and at the CASSCF-SA level by selecting the first two states [CASSCF-SA(2)]. In the particular case of CASSCF-SA(2)/classical calculations with PCM or FQ, the classical charges are equilibrated with the GS. Both a reduced (6,5) or the full valence (12,10) active spaces are tested, in line with previous studies.[90–93] Four basis sets (6-31G*, cc-pVDZ, aug-cc-pVDZ, and aug-cc-pVTZ) are selected, which allow evaluating the effect of polarization and diffuse functions on calculated excitation energies. Finally, various descriptions of the solvent environment (implicit PCM, non-polarizable ESPF, FQ with two different parametrizations - FQ^a and FQ^b) are compared. Gas-phase values obtained at the same level (basis set, active space, and starting orbitals) are taken as a reference.

Figure 2 graphically shows computed values (raw data are reported in tables S1 to S3 in the SI). By first focusing the attention on the effect of the selection of specific starting orbitals, it comes out that differences are generally negligible (0.01 eV on average). The largest variations occur for CASSCF/FQ^a(12,10)/aug-cc-pVTZ calculations. In this case, vertical excitation energies of 4.64 eV (HF), and 4.70 eV [CASSCF-GS, and CASSCF-SA(2)] are obtained. This suggests that the choice of initial orbitals is not critical for aqueous FORM.

We now focus on how the choice of the active space affects numerical results. The (6,5) space includes the σ, π, π^*, and σ^* orbitals of the carbonyl group and the n_y of the oxygen (see fig. 3). The (12,10) active space includes all MOs except for the 1s of Carbon and Oxygen atoms. As expected, diverse excitation energies are obtained with the two active spaces, independently of the other parameters (see fig. 2). The largest differences (0.17 eV) are reported for the FQ^b parametrization combined with the 6-31G* basis set. Generally, the larger active space (12,10) yields smaller excitation energies. This is because, in the full valence active space, the ES is more stabilized than the GS, thus resulting in a lower excitation energy.

Substantial differences also emerge by changing the basis set. In fact, by exploiting the (6,5) active space,

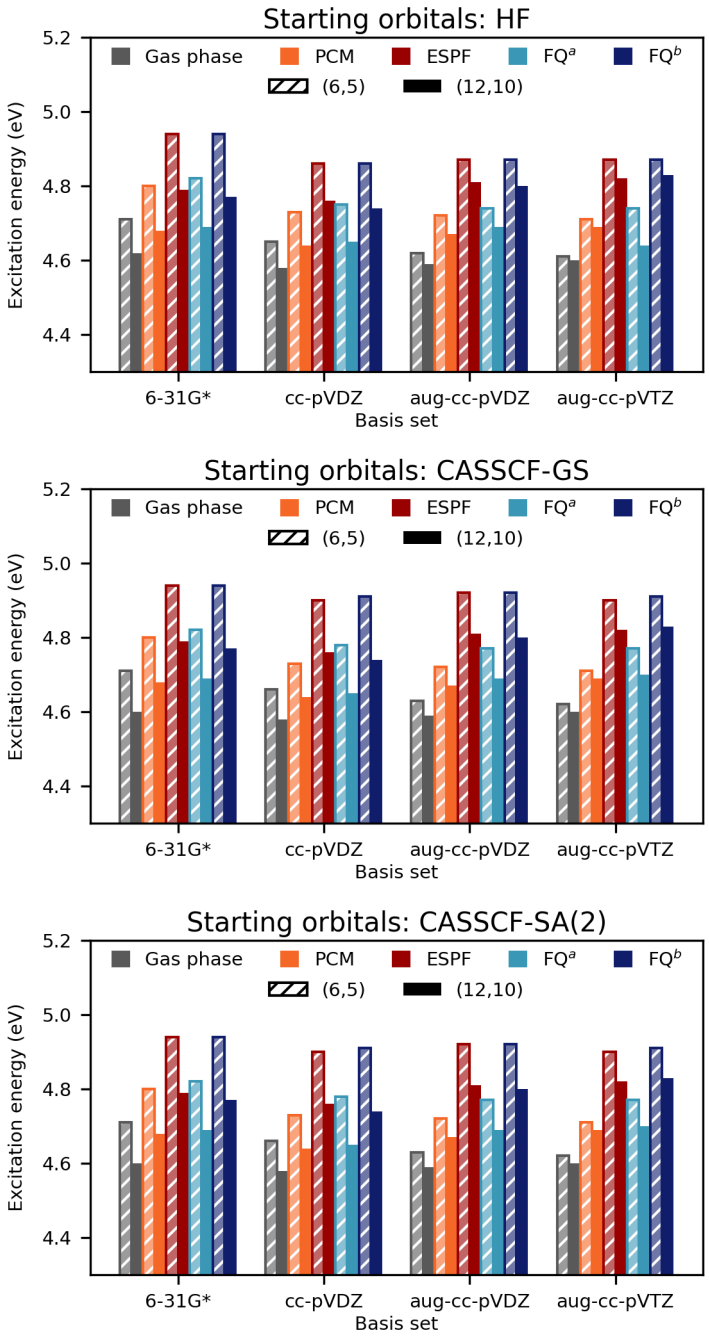


Figure 2: Computed FORM $n - \pi^*$ vertical excitation energies employing HF (top panel), CASSCF-GS (middle panel), and CASSCF-SA(2) (bottom panel) starting orbitals. Selected basis set, active space, and solvation models (gas phase, PCM, ESPF, FQ^a, and FQ^b) are compared. The hatched and filled bars correspond to the (6,5) and (12,10) active spaces, respectively.

a decrease in excitation energies (0.05 eV on average) is observed by moving from the 6-31G* to cc-pVDZ basis set; values further decrease moving to the aug-cc-pVDZ basis (less than 0.01 eV on average) and to the largest aug-cc-pVTZ basis set (0.01 eV on average, see tables S1 to S3 in the SI). When the active space

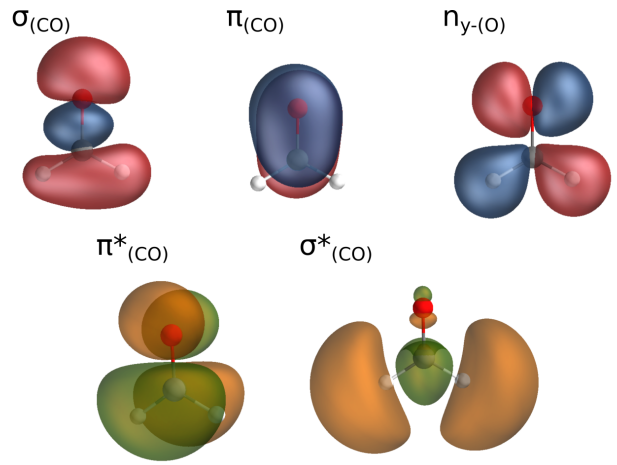


Figure 3: HF/FQ^b selected natural orbitals of aqueous FORM included in the (6,5) active space for subsequent CASSCF calculations. The image illustrates σ, π, π^*, and σ^* orbitals of the carbonyl group, and the n_y orbital of oxygen. FORM lies in the yz plane, with the carbonyl group aligned along the z axis.

is extended to (12,10), the above trends are maintained with the only exception of the aug-cc-pVTZ basis set, which yields higher excitation energies than aug-cc-pVDZ (0.01 eV on average). In general, increasing the size of the basis set stabilizes both the GS and the ES. However, depending on the basis set, such a stabilization can be larger for the GS or the ES, thus resulting in higher or lower excitation energy values, respectively. Note that the above comments yield for both gas-phase and solution and all tested solvation models.

To analyze how computed values are affected by the selection of the solvent model, the $n-\pi^*$ transition is considered and gas phase results are taken as a reference. The inclusion of solvent effects blueshift excitation energies as compared to gas phase values. By focusing on solvatochromic shifts, i.e. the difference between the absorption energy in solution (E^{solv}) and gas-phase (E^{vac}) (see tables S1 to S3 in the SI) it comes out that QM/PCM gives the smallest values. This is probably due to the lack of accounting for hydrogen bonding.[24,48] In fact, all atomistic solvent descriptions yield more pronounced solvatochromic shifts. Excitation energies computed at the QM/ESPF level result in solvatochromic shifts that are almost twice those calculated at the QM/PCM level. Large differences arise when the two FQ parametrizations are employed. Indeed, QM/FQ^a calculations produce systematically lower excitation energies than QM/FQ^b, resulting in a smaller solvatochromic shift. This can be explained by the fact that FQ^a has been parametrized to reproduce the properties of bulk water,[46] whereas FQ^b is designed to reproduce solute-solvent electrostatic and polarization interactions in aqueous solutions,[76] leading to a more pronounced solvent effect. The largest QM/FQ^b shift

(0.29 eV) is obtained with the reduced active space (6,5) in combination with the aug-cc-pVDZ and aug-cc-pVTZ basis sets and CASSCF and CASSCF-SA(2) starting orbitals (see tables S2 and S3 in the SI). Note that excitation energies computed at the QM/ESPF and QM/FQ^b level are very similar for the selected snapshot, but this trend is not confirmed in the subsequent calculations (*vide infra*). Generally, among the various parameters considered in the above analysis, the choice of the solvent model and its parametrization are those that impact computed excitation energies more largely.

3.2 Formaldehyde in Aqueous Solution

In this section, CASSCF/FQ is applied to the simulation of the dark $n - \pi^*$ excitation of FORM in aqueous solution. The solute-solvent phase space is sampled by resorting to classical MD in which the solute is kept frozen to the optimized GS geometry (see section 2).

200 spherically-shaped snapshots are extracted from the MD simulation, then absorption energies are computed on each snapshot at the non-polarizable CASSCF/ESPF and polarizable CASSCF/FQ^a and CASSCF/FQ^b levels. All CASSCF calculations are performed by employing the aug-cc-pVDZ basis set and HF starting orbitals, in line with the analysis reported above for the single structure of aqueous FORM (see section 3.1). The active spaces (6,5) and (12,10) are tested.

In fig. 4, computed CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b excitation energies extracted from the 200 snapshots are reported as "stick spectra"; both active spaces are considered. Each vertical stick represents the excitation energy of a given snapshot with its associated oscillator strength. Absorption energies and oscillator strengths undergo large variations as a function of the snapshot, related to the dynamic fluctuations of water molecules surrounding the FORM molecule. Remarkably, this way of proceeding results in a solvent-induced inhomogeneous broadening of spectral bands to be naturally taken into account; since it results from solvent fluctuations, the spreading of excitation energies depends on the solvation model, and it is less accentuated for ESPF and FQ^a as compared to FQ^b. In particular, the lowest and largest variations of absorption energies are reported for CASSCF(12,10)/FQ^a (0.34 eV) and CASSCF(6,5)/FQ^b (0.74 eV), thus highlighting a diverse description of solute-solvent interactions provided by the various atomistic approaches.

Before moving to comment on spectra, it is critical to ensure that the absorption signal is at convergence with respect to the number of snapshots extracted from MD simulations and exploited to compute final

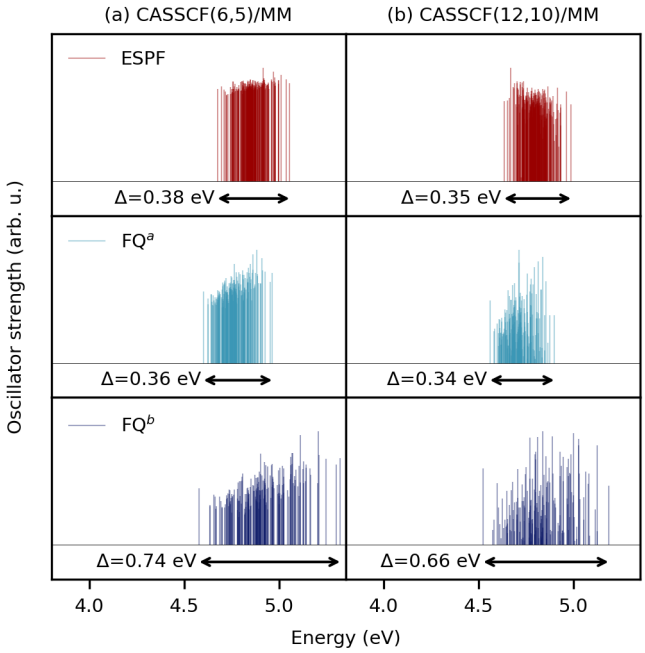


Figure 4: Computed transitions (stick spectra) of FORM in aqueous solution employing CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b with active spaces (a) (6,5) and (b) (12,10). The spread in the excitation energy among all snapshots (Δ) for each level of theory is also reported.

spectra. Averaged spectra which are obtained by using an increasing number of snapshots are shown in fig. S2 in the SI. As a measure of the statistical significance of the averaged excitation energy, table S5 in the SI reports average excitation energies and the 95% confidence interval. The largest error (at 95% confidence) is reported for CASSCF/FQ^b (0.003 eV) for both active spaces.

Figure 5 shows averaged absorption spectra at each CASSCF/MM level of theory for both active spaces. Absorption energies computed in the gas phase and using the implicit PCM approach at the same level of theory are also shown. Corresponding excitation energies (ε) and associated solvatochromic shifts (δ) for each level of theory are reported in table 1.

Moving from the gas phase (gray line) to the aqueous solution, the $n - \pi^*$ transition is blue-shifted. The shift varies as a function of the solvent description and the selected active space. As noted in the previous section (see section 3.1), by enlarging the active space from (6,5) to (12,10) the excitation energy of aqueous FORM becomes smaller for all solvent models. This is due to a larger stabilization of the ES with respect to the GS, as also observed in the gas phase (see bottom panel of fig. 5). In particular, the excitation energy is reduced by 0.10 eV for CASSCF/PCM, 0.07 eV for CASSCF/ESPF, 0.04 eV for CASSCF/FQ^a, and 0.07 eV for CASSCF/FQ^b when the active space is enlarged from (6,5) to (12,10) (see table 1).

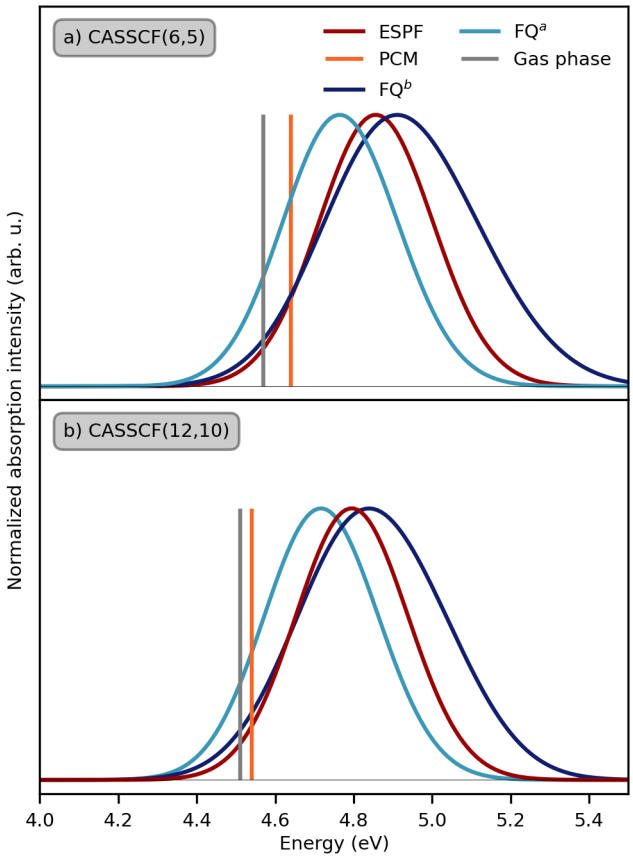


Figure 5: FORM normalized averaged absorption spectra obtained from snapshots extracted from the MD simulation. (a) (8,7) and (b) (12,10) active spaces. Computed excitation energies in gas-phase and aqueous solution as obtained with the implicit PCM description are indicated.

Table 1: Excitation energies (ε) and vacuo-to-water solvatochromic shifts ($\delta = \omega_{\text{gas}} - \omega_{\text{solv}}$) of rigid FORM in aqueous solution as computed at CASSCF/aug-cc-pVDZ level of theory with the active spaces (6,5) and (12,10) and different solvent models. For the sake of comparison, experimental data in the gas phase (see Refs. 94–96 and references therein) and in aqueous solution[97] are reported.

Model	ε (eV)		δ (eV)	
	(6,5)	(12,10)	(6,5)	(12,10)
Gas phase	4.57	4.51	-	-
PCM	4.64	4.54	-0.07	-0.03
ESPF	4.86	4.79	-0.29	-0.28
FQ^a	4.76	4.72	-0.19	-0.21
FQ^b	4.91	4.84	-0.34	-0.33
Exp. (gas phase)	3.79,[94] 3.86,[95] 4.00[96]		-	
Exp. (water)	4.28[97]		-	

The experimental spectrum of highly-concentrated FORM ($\sim$ 10 M) in aqueous solution has been measured by Bercovici *et al,*[97] who have extracted a value for the excitation energy of 4.28 eV. As pointed out by several authors,[97–100] the extraction of accurate excitation energy values for the $n - \pi^*$ transition of

aqueous FORM is especially difficult, due to the formation of oligomers and methylene glycol in aqueous solution. Therefore, in the following only a qualitative comparison between computed and experimental values is proposed. Furthermore, given the data reported in table 1, the experimental solvatochromic shift of FORM in aqueous solution is expected to be between -0.49 and -0.28 eV. However, the definition of the experimental value is challenging, given the relatively small value and the large variability of experimental data found in the literature (see Ref. 96 and references therein). Several authors[98,101–103] have proposed to take as a reference the solvatochromic shift of acetone in water, which yields a solvatochromic shift of about -0.21 eV.[104,105]

From an inspection of table 1, the variation of the active space has a limited effect on the solvatochromic shift. The largest variation is reported for CASSCF/PCM (0.04 eV), which is also the model yielding the smallest solvatochromic shift. All atomistic approaches generally provide larger solvatochromic shifts, which increase by moving from CASSCF/FQ^a to CASSCF/ESPF. The largest shift is obtained for CASSCF/FQ^b [-0.34 eV and -0.33 eV for (6,5) and (12,10) respectively]. By taking the solvatochromic shift of acetone as a reference, CASSCF/FQ^a in combination with the (12,10) active space yields the best agreement for both active spaces. Note that the solvatochromic shifts obtained with our calculations are in line with previous calculations reported in the literature (see Ref. 106 and references therein).

The diverse picture provided by the various solvent models can be explained by considering that, while ESPF and FQ^a force fields are parameterized for reproducing bulk liquid water, FQ^b is parameterized to reproduce electrostatic and polarization interaction energies of molecular systems in an aqueous environment. Consequently, CASSCF/FQ^b predicts a larger solvent effect, resulting in an overall overestimation of the solvatochromic shift (error: 0.13/0.12 eV), while CASSCF/FQ^a provides the best agreement with the experimental value (error: 0.02/0.00 eV). However, it is worth remarking that CASSCF/FQ coupling completely neglects solute-solvent non-electrostatic interactions. In particular, we have recently shown that solute-solvent Pauli repulsion is particularly relevant for consistently modeling vacuo-to-water solvatochromic shifts.[53] Specifically, Pauli repulsion effects are expected to confine the QM density, overall decreasing the absolute value of solvatochromic shift,[53,107] and possibly improving the agreement of CASSCF/FQ^b calculation with experimental data. On the contrary, the excellent agreement of CASSCF/FQ^a with the reference value of acetone may be due to error cancellation associated with the parametrization strategy of FQ^a,[46] as

noted in previous works.[56,108]

From the inspection of fig. 5 and table 1, it can be noticed that all simulations yield a larger value of the transition energy as compared to the experimental value, independently of the choice of the active space and solvent model. This can be associated with the lack of dynamic electron correlation; in fact, also the computed gas phase transition energy overestimates the experimental value.[94–96] The effect of dynamic correlation can be estimated by exploiting the CASPT2 approach,[16] in combination with the same computational protocol described in section 2. Gas phase CASPT2 calculations in combination with the active spaces (6,5) and (12,10) yield an excitation energy of 3.88 eV and 3.90 eV, i.e., the energy contribution due to dynamic correlation is 0.69 eV and 0.61 eV, respectively. These values are in good agreement with experimental data (see table 1) and with calculated transition energies reported by Lupieri *et al.* of 3.82–3.89 eV (TDDFT) and 3.84–3.86 eV (CASPT2).[106] In the case of aqueous FORM, CASPT2/PCM computed energies with active spaces (6,5) and (12,10) blueshift to 4.00 eV and 3.99 eV, with a dynamic correlation contribution of 0.65 eV and 0.55 eV, respectively.

As already mentioned, solvatochromic shifts of FORM are relatively small. The lack of dynamic correlation may therefore significantly impact the final value. Moving from the gas phase to PCM, the solvatochromic shift calculated at the CASPT2 level in combination with active spaces (6,5) and (12,10) become -0.12 eV and -0.09 eV, respectively, which are almost double the corresponding CASSCF values (see table 1) even if the absolute value variation is below 0.06 eV.

3.3 Para-Nitroaniline in Aqueous Solution

In this section, CASSCF/FQ is applied to the simulation of the bright $\pi - \pi^*$ excitation of PNA in aqueous solution. The solute-solvent phase space is sampled by resorting to classical MDs, from which a set of uncorrelated snapshots is extracted (see section 2). First, the solute is kept frozen to the optimized GS geometry (MD_{rigid}, section 3.3.1). Then, PNA is allowed to freely move (MD_{free}, section 3.3.2). In this way, effects arising from the solvent dynamics around the solute are isolated from those following the reorganization of both PNA geometry and the solvent around it.

200 spherically-shaped snapshots are extracted from MD_{rigid} or MD_{free}, then absorption energies are computed on each snapshot at the non-polarizable CASSCF/ESPF and polarizable CASSCF/FQ^a and

CASSCF/FQ^b levels. All CASSCF calculations are performed by employing the aug-cc-pVDZ basis set and HF starting orbitals, in line with the analysis reported above for FORM (see section 3.1). In the case of CASSCF(12,10)/PCM calculation, the starting orbitals are obtained from CASSCF-GS(8,7)/PCM, to ensure convergence to the same electronic states for both active spaces.

Two active spaces, i.e. (8,7) and (12,10), are tested for structures extracted from MD_{rigid}, whereas only the (8,7) active space is applied to structures extracted from MD_{free}. The (12,10) active space is selected based on Ref. 109, while the (8,7) is chosen to investigate how the choice of the active space affects the final results. MOs belonging to the two active spaces and their classification are graphically reported in fig. 6; the $\pi - \pi^*$ transition involves orbitals 36 and 37. The (12,10) active space consists of five occupied π orbitals (29, 31, 34, 35, 36), one occupied n orbital (33), and four lower-energy virtual π^* orbitals (37, 38, 39, 40). In contrast, the (8,7) active space includes four occupied π orbitals (31, 34, 35, 36) and three lower-energy virtual π^* orbitals (37, 38, 39). The π^*-type orbital 40 is excluded due to its low occupation number, as indicated by the natural orbital analysis (see table S4 in the SI).

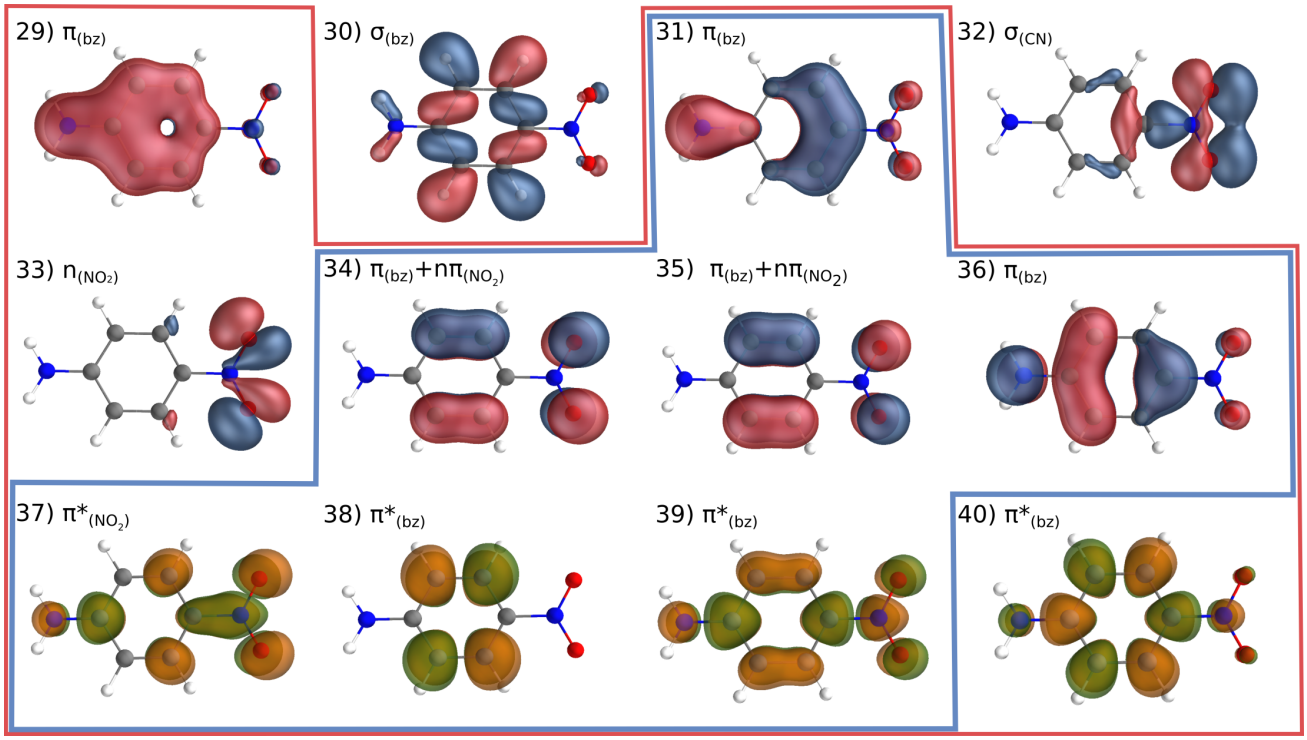


Figure 6: HF/FQ^b selected PNA natural orbitals adopted for subsequent CASSCF calculations. The image represents π and π^* orbitals associated with the benzene ring (bz) and the nitro group, σ orbitals of the benzene ring and CN group and n orbital of oxygen atoms of the nitro group (the labeling follows Ref. 110). The blue box collects orbitals defining the (8,7) active space, while the red box defines the (12,10) active space.

3.3.1 PNA/water structures extracted from MD_{rigid}

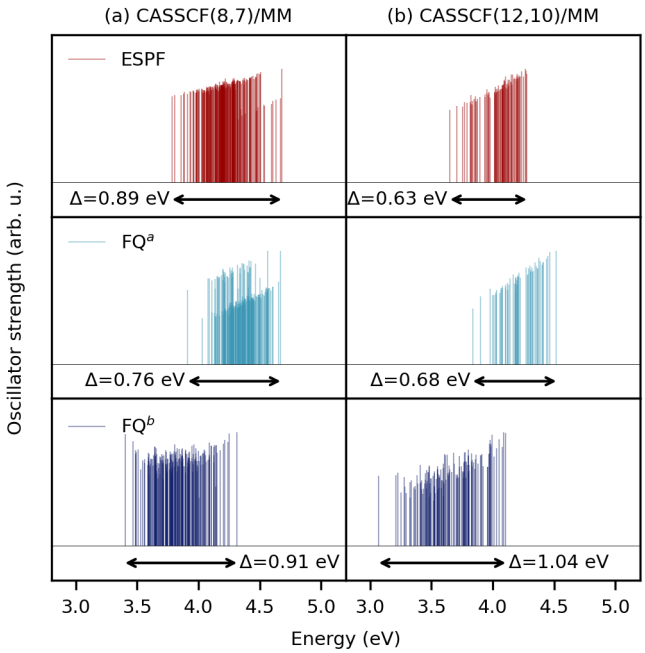


Figure 7: Computed transitions (stick spectra) of PNA in aqueous solution (structure are taken from MD_{rigid}) employing CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b with active spaces (a) (8,7) and (b) (12,10). The spread in the excitation energy among all snapshots (Δ) for each level of theory is also reported.

Stick spectra as computed at the CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b levels for both active spaces are reported in fig. 7. Similar to aqueous FORM (see section 3.2), the variation of both absorption energies and oscillator strengths is linked to dynamic fluctuations of water molecules around the PNA molecule. The spreading of excitation energies is less accentuated for ESPF and FQ^a as compared with FQ^b. CASSCF(12,10)/ESPF yields the smallest variation of absorption energies ($\Delta = 0.63$ eV) while the largest variation is observed for CASSCF(12,10)/FQ^b calculations ($\Delta = 1.04$ eV).

The convergence analysis of convoluted spectra with respect to the number of snapshots is reported in fig. S3 and table S6 in the SI. The largest error (at 95% confidence) is reported for CASSCF(12,10)/FQ^b (0.016 eV), while CASSCF(12,10)/FQ^a requires the smallest number of snapshots (57) to reach convergence (error: 0.01 eV).

Figure 8 shows averaged absorption spectra at each CASSCF/MM level of theory for both active spaces. Absorption energies computed in the gas phase and using the implicit PCM approach at the same level of theory are also shown, along with the experimental profile reproduced from Ref. 111.

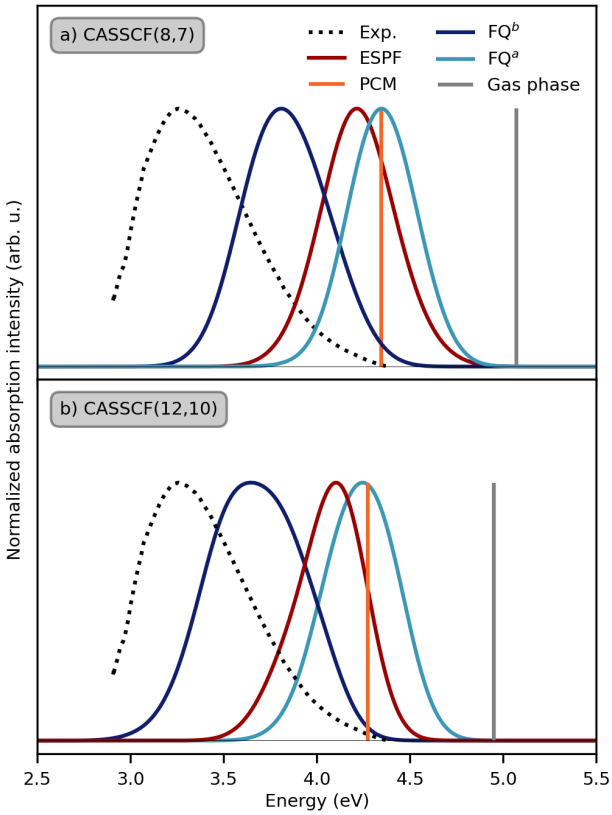


Figure 8: PNA normalized averaged absorption spectra obtained from snapshots extracted from MD_{rigid}. (a) (8,7) and (b) (12,10) active spaces. Computed excitation energies in gas-phase and aqueous solution as obtained with the implicit PCM description are indicated. Experimental spectra taken from Ref.[111] are reported for the sake of comparison.

Moving from the gas phase (gray line) to the aqueous solution, the $\pi - \pi^*$ transition undergoes a red-shift due to the increased polarizability of the excited state and the resulting greater stabilization induced by the solvent.[56,112] The shift varies as a function of the solvent description and the active space. Corresponding excitation energies (ε) and associated solvatochromic shifts (δ) for each level of theory are reported in table 2.

It is also worth highlighting that the (12,10) active space gives smaller excitation energies than (8,7). This is due to a larger stabilization of the ES with respect to the GS, as also observed in the gas phase (see bottom panel of fig. 8). As a result, the excitation energy is reduced by 0.08 eV for CASSCF/PCM, 0.11 eV for CASSCF/ESPF, 0.10 eV for CASSCF/FQ^a, and 0.16 eV for CASSCF/FQ^b when the active space is enlarged from (8,7) to (12,10) (see table 2). Consistently, absorption energies computed using the largest active space (12,10) are remarkably shifted towards the experimental value, with CASSCF/FQ^b yielding the best reproduction of the experimental absorption band.

Solvatochromic shifts (δ) are reported in table 2. By varying the active space, the largest variation of

solvatochromic shift is reported for CASSCF/FQ^b and CASSCF/PCM (0.04 eV), and QM/PCM calculations yield the smallest solvatochromic shift for both active spaces. On the contrary, shifts computed with atomistic approaches are larger than QM/PCM, and $\delta(\text{FQ}^a) < \delta(\text{ESPF}) < \delta(\text{FQ}^b)$ similarly to what already commented in section 3.2. The polarizable CASSCF/FQ^b method gives the largest solvatochromic shift [1.26 eV and 1.30 eV for (8,7) and (12,10) respectively], almost doubling CASSCF/FQ^a. CASSCF/PCM, CASSCF/ESPF, and CASSCF/FQ^a underestimate the solvatochromic shift compared to the experimental value (1.0 eV),[111] while CASSCF/FQ^b overestimates it for both active spaces (see table 2). The solvatochromic shift computed

Table 2: Excitation energies (ε) and vacuo-to-water solvatochromic shifts ($\delta = \omega_{\text{gas}} - \omega_{\text{solv}}$) of PNA in aqueous solution (structures are extracted from MD$_{\text{rigid}}$) as computed at CASSCF/aug-cc-pVDZ level of theory with the active spaces (8,7) and (12,10) and different solvent models. Experimental data taken from Ref. 111 are also reported.

Model	ε (eV)		δ (eV)	
	(8,7)	(12,10)	(8,7)	(12,10)
Gas phase	5.07	4.95	-	-
PCM	4.35	4.27	0.72	0.68
ESPF	4.21	4.10	0.86	0.85
FQ^a	4.35	4.25	0.72	0.70
FQ^b	3.81	3.65	1.26	1.30
Exp. (water)[111]	3.23		1.00	

at the CASSCF/FQ^b level overestimates the experimental value (error: 0.26/0.30 eV), while CASSCF/ESPF yields the best agreement with the experimental value (error: 0.15/0.16 eV). As noted in the case of aqueous FORM, the introduction of solute-solvent Pauli repulsion effects may improve the agreement of CASSCF/FQ^b calculation with experimental data. In parallel, the good agreement of CASSCF/ESPF with experiments might be ascribed to a favorable error cancellation (in fact, it lacks both polarization and Pauli repulsion effects).

From the inspection of fig. 8, it comes out clearly that all embedding methods overestimate the transition energy as compared to experiments, independently of the choice of the active space. This is in agreement with previous data reported in the literature and is again associated with the lack of dynamic electron correlation.[92] Similarly to the case of aqueous FORM in section 3.2, dynamic correlation effects can be estimated by performing CASPT2 calculations.[16] Specifically, CASPT2(8,7) simulations in the gas phase and in combination with PCM and ESPF solvent models yield excitation energies of 4.31 eV, 3.60 eV, and 3.40 eV, with dynamic correlation contributions of 0.76 eV, 0.75 eV, and 0.81 eV, respectively (absorption

spectra are reported in fig. S5 in the SI). The solvatochromic shift values calculated at the CASPT2(8,7) level in combination with PCM and ESPF solvent models are 0.71 eV and 0.91 eV, which are in good agreement with the CASSCF values reported in table 2. Differently from the case of aqueous FORM, solvatochromic shifts for aqueous PNA are less sensitive to dynamic correlation effects in terms of relative magnitude.

3.3.2 PNA/water structures extracted from MD_{free}

To achieve a more accurate description, the absorption spectrum of aqueous PNA is simulated by allowing both solute and solvent molecules to freely move (MD_{free}). As described in Ref. 56, during the MD PNA oscillates predominantly around its planar configuration, thus indicating that the benzene ring is substantially rigid. The dihedral distribution functions of the three main dihedral angles, as shown in Fig. 5 of Ref. 56, reveal that the nitro group exhibits rotational freedom, while the NH_2 group is confined to oscillations around its equilibrium geometry.

As already mentioned, the (8,7) active space is exploited. This choice is guided by two practical aspects. First, as shown in section 3.3.1, the solvatochromic shift slightly depends on the choice of the active space, therefore we can speed up the computation by reducing the active space to the minimum required to study the $\pi - \pi^*$ transition. Second, given the conformational flexibility of PNA, MOs undergo significant mixing and changes across different conformations, making the consistent selection of the (12,10) active space challenging (see fig. S6 in the SI for an example of mixed orbitals in a random snapshot).

In fig. 9, stick and averaged spectra, average excitation energies (ε), and solvatochromic shifts (δ) for each CASSCF(8,7)/MM level of theory are reported. Notably, PNA geometrical rearrangement significantly impacts the variability of the absorption energy among sampled conformations, almost doubling the spreading of excitation energy (Δ) as compared to MD_{rigid} (see fig. 7). To investigate the origin of CASSCF/MM inhomogeneous broadening observed in fig. 9, CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b excitation energies are analyzed as functions of the three PNA main dihedral angles (see fig. S7 in the SI). The variability in excitation energies is quite similar across the three CASSCF/MM approaches, showing no correlation with dihedral angles. Although the solute's flexibility directly affects the calculated property, the different broadening between CASSCF/ESPF, CASSCF/FQ^a, and CASSCF/FQ^b calculations are mainly due to the physical description inherited in the specific MM model, as previously noted for rigid PNA, where the

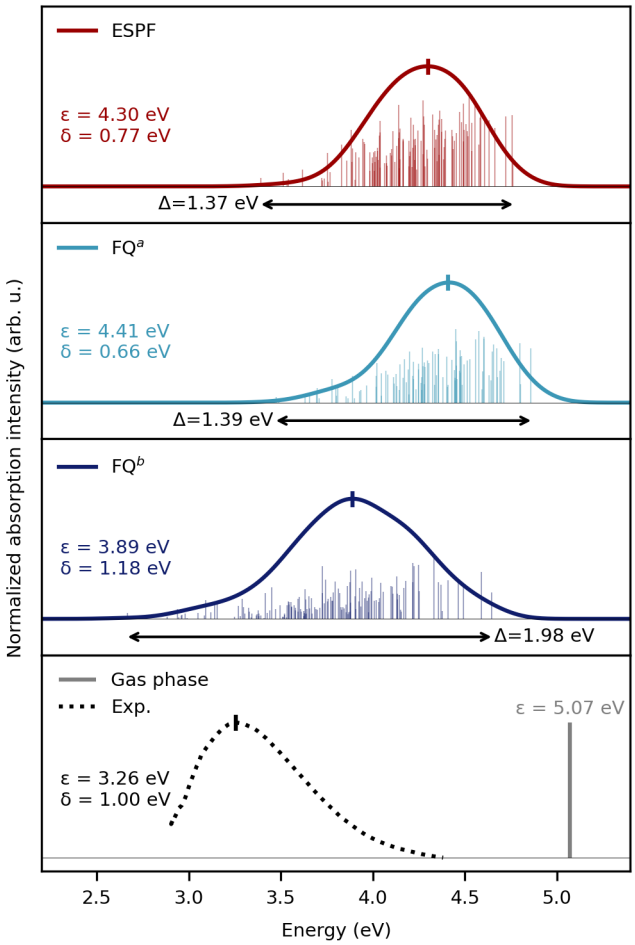


Figure 9: PNA normalized stick and absorption spectra in aqueous solution obtained with different solvent models and by using the (8,7) active space. Structures are extracted from MD_{free}. The intensity of each stick is multiplied by a factor of 5 for ease of visualization. Maximum excitation energy (ε) and solvatochromic shift (δ) in eV are reported in each panel. Gas-phase values and experimental spectra[111] are shown in the lowest panel.

flexibility of the solute is not taken into account.

The convergence of spectra as increasing the number of snapshots is analyzed in fig. S4 and table S7 in the SI. Although the convergence is slower compared to MD_{rigid}, the number of selected snapshots is sufficient to converge the final excitation energy with a maximum error of 0.077 eV in the case of CASSCF/FQ^b calculations.

Similar to MD_{rigid}, final absorption profiles are obtained by convoluting each stick spectrum with a Gaussian lineshape with a FWHM of 0.3 eV; the results are reported in fig. 9. Clearly, the broadening of absorption peaks increases, as it was previously observed for the stick distributions. The relative positions of FQ^b, FQ^a, and ESPF absorption maxima are consistent with MD_{rigid} (see fig. 8). Furthermore, excitation

energies in fig. 9 exhibit only a slight blueshift with respect to table 2,: 0.09 eV for ESPF, 0.06 eV for FQ^a and 0.08 eV for FQ^b.

Given the small increment in excitation energies, solvatochromic shifts are reduced with respect to MD_{rigid}, as it is shown by comparing fig. 9 and table 2. As a consequence, solvatochromic shifts obtained using CASSCF/ESPF and CASSCF/FQ^a are strongly underestimated, whereas CASSCF(8,7)/FQ^b is in the best agreement with the experimental value, yielding a value of 1.18 eV. As discussed in the previous sections, the inclusion of non-electrostatic interactions and dynamic correlation may bring the CASSCF/FQ^b solvatochromic shift closer to the experimental value.

4 Summary and future perspectives

A fully atomistic multiscale model based on the coupling between an MCSCF Hamiltonian and the polarizable FQ force field has been developed within a state-specific approach. Fluctuations in calculated values, arising from the structural rearrangement of solute and solvent, have been recovered by averaging computed values on hundreds of structures generated from a conformational sampling performed by running MD simulations (by either keeping the solute frozen or allowing it to freely move). The computational protocol has required meticulous attention to ensure the consistency of the active space across the set, and also a careful inspection of the consistency of the modeled transition.

The approach and the associated computational protocol have been validated by computing the $n - \pi^*$ transition of aqueous formaldehyde and the $\pi - \pi^*$ transition of aqueous para-nitroaniline. As expected, computed excitation energies and vacuo-to-solvent solvatochromic shifts vary with different choices of active space. However, even larger modifications are observed as a function of the selected solvation model (continuum vs explicit, non-polarizable, or polarizable). In the case of PNA, a good agreement between computed and experimental solvatochromic shifts is obtained by using the CASSCF(8,7)/FQ^b method and allowing the solute and solvent to rearrange during the MD run employed for the conformational sampling.

The reported method has some limitations, which will be overcome in future works. First, dynamic correlation can be recovered, for instance by extending the approach to the second-order perturbation theory with a CASSCF reference wavefunction, as in the CASPT2 model.[16]

Additionally, the state-specific approach that has been exploited here may result unphysical when dealing with multiple states simultaneously, e.g. when the model is coupled to methods to describe conical intersections and more generally non-adiabatic dynamics. In these cases, a state-average formulation appears more physically consistent.[43]

From the purely numerical point of view, the accuracy of the method could be further enhanced by developing on-purpose FQ parametrizations based on CASSCF reference data. In fact, FQ^a and FQ^b parametrization, which have been tested in the numerical segment of this study, have been developed and tested for the calculation of spectral properties at DFT level.

Moreover, similarly to any other classical electrostatic force field, FQ lacks the description of non-electrostatic interactions, which could play a significant role, especially in combination with very accurate wavefunctions. A possible way to account for such interactions would be to extend to MCSCF approaches the method that has been developed recently by some of the present authors,[113] and which formulates Pauli repulsion and dispersion in terms of the QM density. As an alternative, an intermediate layer between the CASSCF and FQ regions could be exploited, thus recovering a fully quantum interaction between the CASSCF region and the first solvation shells, and resorting to FQ to describe long-range electrostatic interactions, in line of previous works of some of us.[107,114]

Acknowledgement

We gratefully acknowledge the Center for High-Performance Computing (CHPC) at SNS for providing the computational infrastructure. CC acknowledges funding from MUR-FARE Ricerca in Italia: Framework per l'attrazione ed il rafforzamento delle eccellenze per la Ricerca in Italia - III edizione. Prot. R20YTA2BKZ.

Supporting Information Available

Computational details and analysis of MD simulations of FORM in aqueous solution. Raw excitation energies of FORM in aqueous solution by varying the starting orbitals, active space, basis set and solvent description. Occupation numbers of PNA natural orbitals. Convergence analysis for CASSCF/MM calculations on FORM and PNA snapshots extracted from MD_{rigid} and MD_{free}. Computed CASPT2(8,7) absorption spectra

in vacuo and by exploiting PCM and ESPF solvent models for PNA. Natural orbitals of a random PNA snapshot extracted from MD_{free}. Correlation plot of excitation energies and PNA dihedral angles for CASSCF(8,7)/MM calculations.

References

(1) Schmidt, M. W.; Gordon, M. S. The construction and interpretation of MCSCF wavefunctions. *Annu. Rev. Phys. Chem.* **1998**, *49*, 233–266.

(2) Szalay, P. G.; Muller, T.; Gidofalvi, G.; Lischka, H.; Shepard, R. Multiconfiguration self-consistent field and multireference configuration interaction methods and applications. *Chem. Rev.* **2012**, *112*, 108–181.

(3) Spezia, R.; Zazza, C.; Palma, A.; Amadei, A.; Aschi, M. A DFT study of the low-lying singlet excited states of the all-trans peridinin in vacuo. *J. Phys. Chem. A* **2004**, *108*, 6763–6770.

(4) Neese, F.; Petrenko, T.; Ganyushin, D.; Olbrich, G. Advanced aspects of ab initio theoretical optical spectroscopy of transition metal complexes: Multiplets, spin-orbit coupling and resonance Raman intensities. *Coord. Chem. Rev.* **2007**, *251*, 288–327.

(5) Curchod, B. F. E.; Martínez, T. J. Ab Initio Nonadiabatic Quantum Molecular Dynamics. *Chem. Rev.* **2018**, *118*, 3305–3336.

(6) Crespo-Otero, R.; Barbatti, M. Recent advances and perspectives on nonadiabatic mixed quantum–classical dynamics. *Chem Rev.* **2018**, *118*, 7026–7068.

(7) Grein, F.; Chang, T. Multiconfiguration wavefunctions obtained by application of the generalized Brillouin theorem. *Chem. Phys. Lett.* **1971**, *12*, 44–48.

(8) Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157–173.

(9) Malmqvist, P. Å.; Rendell, A.; Roos, B. O. The restricted active space self-consistent-field method, implemented with a split graph unitary group approach. *J. Phys. Chem.* **1990**, *94*, 5477–5482.

(10) Werner, H.-J.; Knowles, P. J. A second order multiconfiguration SCF procedure with optimum convergence. *J. Chem. Phys.* **1985**, *82*, 5053–5063.

(11) Olsen, J. The CASSCF method: A perspective and commentary. *Int. J. Quantum Chem.* **2011**, *111*, 3267–3272.

(12) Ruedenberg, K.; Cheung, L.; Elbert, S. MCSCF optimization through combined use of natural orbitals and the Brillouin–Levy–Berthier theorem. *Int. J. Quantum Chem.* **1979**, *16*, 1069–1101.

(13) Werner, H.-J.; Meyer, W. A quadratically convergent MCSCF method for the simultaneous optimization of several states. *J. Chem. Phys.* **1981**, *74*, 5794–5801.

(14) Docken, K. K.; Hinze, J. LiH Potential Curves and Wavefunctions for X 1 Σ+, A 1 Σ+, B 1 Π, 3 Σ+, and Π 3. *J. Chem. Phys.* **1972**, *57*, 4928–4936.

(15) Vogiatzis, K. D.; Ma, D.; Olsen, J.; Gagliardi, L.; De Jong, W. A. Pushing configuration-interaction to the limit: Towards massively parallel MCSCF calculations. *J. Chem. Phys.* **2017**, *147*, 184111.

(16) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483–5488.

(17) Andersson, K.; Malmqvist, P.-Å.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218–1226.

(18) Schreiber, M.; Silva-Junior, M. R.; Sauer, S.; Thiel, W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J. Chem. Phys.* **2008**, *128*.

(19) Silva-Junior, M. R.; Schreiber, M.; Sauer, S.; Thiel, W. Benchmarks of electronically excited states: Basis set effects on CASPT2 results. *J. Chem. Phys.* **2010**, *133*.

(20) Lischka, H.; Nachtigallova, D.; Aquino, A. J.; Szalay, P. G.; Plasser, F.; Machado, F. B.; Barbatti, M. Multireference approaches for excited states of molecules. *Chem. Rev.* **2018**, *118*, 7293–7361.

(21) González, L.; Escudero, D.; Serrano-Andrés, L. Progress and challenges in the calculation of electronic excited states. *ChemPhysChem* **2012**, *13*, 28–51.

(22) Barneschi, L.; Kaliakin, D.; Huix-Rotllant, M.; Ferré, N.; Filatov(Gulak), M.; Olivucci, M. Assessment of the Electron Correlation Treatment on the Quantum-Classical Dynamics of Retinal Protonated Schiff Base Models: XMS-CASPT2, RMS-CASPT2, and REKS Methods. *J. Chem. Theory Comput.* **2023**, *19*, 8189–8200.

(23) Mai, S.; Marquetand, P.; González, L. Nonadiabatic dynamics: The SHARC approach. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1370.

(24) Giovannini, T.; Cappelli, C. Continuum vs. atomistic approaches to computational spectroscopy of solvated systems. *Chem. Commun.* **2023**, *59*, 5644–5660.

(25) Mennucci, B.; Corni, S. Multiscale modelling of photoinduced processes in composite systems. *Nature Reviews Chemistry* **2019**, *3*, 315–330.

(26) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999–3094.

(27) Olsen, J. M. H.; Kongsted, J. *Adv. Quantum Chem.*; Elsevier, 2011; Vol. 61; pp 107–143.

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

(29) Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

(30) Marrazzini, G.; Giovannini, T.; Scavino, M.; Egidi, F.; Cappelli, C.; Koch, H. Multilevel density functional theory. *J. Chem. Theory Comput.* **2021**, *17*, 791–803.

(31) Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(32) Lin, H.; Truhlar, D. G. QM/MM: what have we learned, where are we, and where do we go from here? *Theor. Chem. Acc.* **2007**, *117*, 185–199.

(33) Morzan, U. N.; Alonso de Armino, D. J.; Foglia, N. O.; Ramirez, F.; Gonzalez Lebrero, M. C.; Scherlis, D. A.; Estrin, D. A. Spectroscopy in complex environments from QM–MM simulations. *Chem. Rev.* **2018**, *118*, 4071–4113.

(34) Cammi, R.; Frediani, L.; Mennucci, B.; Tomasi, J.; Ruud, K.; Mikkelsen, K. V. A second-order, quadratically convergent multiconfigurational self-consistent field polarizable continuum model for equilibrium and nonequilibrium solvation. *J. Chem. Phys.* **2002**, *117*, 13–26.

(35) Cossi, M.; Barone, V.; Robb, M. A. A direct procedure for the evaluation of solvent effects in MC-SCF calculations. *J. Chem. Phys.* **1999**, *111*, 5295–5302.

(36) Aguilar, M.; Olivares del Valle, F.; Tomasi, J. Nonequilibrium solvation: An ab initio quantum-mechanical method in the continuum cavity model approximation. *J. Chem. Phys.* **1993**, *98*, 7375–7384.

(37) Amovilli, C.; Mennucci, B.; Floris, F. M. MCSCF study of the SN2 Menshutkin reaction in aqueous solution within the polarizable continuum model. *J. Phys. Chem. B* **1998**, *102*, 3023–3028.

(38) Cammi, R.; Frediani, L.; Mennucci, B.; Ruud, K. Multiconfigurational self-consistent field linear response for the polarizable continuum model: Theory and application to ground and excited-state polarizabilities of para-nitroaniline in solution. *J. Chem. Phys.* **2003**, *119*, 5818–5827.

(39) Song, C. State-averaged CASSCF with polarizable continuum model for studying photoreactions in solvents: Energies, analytical nuclear gradients, and non-adiabatic couplings. *J. Chem. Phys.* **2022**, *156*.

(40) Hedegård, E. D.; Jensen, H. J. A.; Kongsted, J. Polarizable embedding based on multiconfigurational methods: Current developments and the road ahead. *Int. J. Quantum Chem.* **2014**, *114*, 1102–1107.

(41) Hedegård, E. D.; List, N. H.; Jensen, H. J. A.; Kongsted, J. The multi-configuration self-consistent field method within a polarizable embedded framework. *J. Chem. Phys.* **2013**, *139*, 044101.

(42) List, N. H.; Jensen, H. J. A.; Kongsted, J.; Hedegård, E. D. A unified framework for the polarizable

embedding and continuum methods within multiconfigurational self-consistent field theory. *Adv. Quantum Chem.* **2013**, *66*, 195–238.

(43) Song, C. State averaged CASSCF in AMOEBA polarizable water model for simulating nonadiabatic molecular dynamics with nonequilibrium solvation effects. *J. Chem. Phys.* **2023**, *158*.

(44) Li, Q.; Mennucci, B.; Robb, M. A.; Blancafort, L.; Curutchet, C. Polarizable QM/MM multiconfiguration self-consistent field approach with state-specific corrections: Environment effects on cytosine absorption spectrum. *J. Chem. Theory Comput.* **2015**, *11*, 1674–1682.

(45) Song, C.; Wang, L.-P. A Polarizable QM/MM Model That Combines the State-Averaged CASSCF and AMOEBA Force Field for Photoreactions in Proteins. *J. Chem. Theory. Comput.* **2024**,

(46) Rick, S. W.; Stuart, S. J.; Berne, B. J. Dynamical fluctuating charge force fields: Application to liquid water. *J. Chem. Phys.* **1994**, *101*, 6141–6156.

(47) Cappelli, C. Integrated QM/polarizable MM/continuum approaches to model chiroptical properties of strongly interacting solute–solvent systems. *Int. J. Quantum Chem.* **2016**, *116*, 1532–1542.

(48) Giovannini, T.; Egidi, F.; Cappelli, C. Molecular spectroscopy of aqueous solutions: a theoretical perspective. *Chem. Soc. Rev.* **2020**, *49*, 5664–5677.

(49) Giovannini, T.; Egidi, F.; Cappelli, C. Theory and algorithms for chiroptical properties and spectroscopies of aqueous systems. *Phys. Chem. Chem. Phys.* **2020**, *22*, 22864–22879.

(50) Gómez, S.; Giovannini, T.; Cappelli, C. Multiple facets of modeling electronic absorption spectra of systems in solution. *ACS Phys. Chem. Au* **2023**, *3*, 1–16.

(51) Sanderson, R. An interpretation of bond lengths and a classification of bonds. *Science* **1951**, *114*, 670–672.

(52) Ambrosetti, M.; Skoko, S.; Giovannini, T.; Cappelli, C. Quantum mechanics/fluctuating charge protocol to compute solvatochromic shifts. *J. Chem. Theory Comput.* **2021**, *17*, 7146–7156.

(53) Giovannini, T.; Ambrosetti, M.; Cappelli, C. Quantum confinement effects on solvatochromic shifts of molecular solutes. *J. Phys. Chem. Lett.* **2019**, *10*, 5823–5829.

(54) Kosenkov, D.; Slipchenko, L. V. Solvent effects on the electronic transitions of p-nitroaniline: A QM/EFP study. *J. Phys. Chem. A* **2011**, *115*, 392–401.

(55) Makwani, D.; Vijaya, R. Solvent effects on the dispersion of linear polarizability and first hyperpolarizability of para-nitroaniline using polarizable continuum model. *J. Nonlinear Opt. Phys. Mater.* **2009**, *18*, 85–97.

(56) Giovannini, T.; Riso, R. R.; Ambrosetti, M.; Puglisi, A.; Cappelli, C. Electronic transitions for a fully polarizable qm/mm approach based on fluctuating charges and fluctuating dipoles: linear and corrected linear response regimes. *J. Chem. Phys.* **2019**, *151*, 174104.

(57) Roos, B. O. *Lecture notes in quantum chemistry: European summer school in quantum chemistry*; Springer, 1992; pp 177–254.

(58) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular electronic structure theory*; John Wiley & Sons, 2013.

(59) Kreplin, D. A.; Knowles, P. J.; Werner, H.-J. MCSCF optimization revisited. II. Combined first-and second-order orbital optimization for large molecules. *J. Chem. Phys.* **2020**, *152*, 074102.

(60) Levy, B.; Berthier, G. Generalized Brillouin theorem for multiconfigurational SCF theories. *Int. J. Quantum Chem.* **1968**, *2*, 307–319.

(61) Levy, B.; Berthier, G. Generalized Brillouin theorem for multiconfigurational SCF theories. *Int. J. Quantum Chem.* **1969**, *3*, 247–247.

(62) Kreplin, D. A.; Knowles, P. J.; Werner, H.-J. Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence. *J. Chem. Phys.* **2019**, *150*.

(63) Aquilante, F.; Autschbach, J.; Baiardi, A.; Battaglia, S.; Borin, V. A.; Chibotaru, L. F.; Conti, I.; De Vico, L.; Delcey, M.; Ferré, N., et al. Modern quantum chemistry with [Open] Molcas. *J. Chem. Phys.* **2020**, *152*, 214117.

(64) Li Manni, G.; Fdez. Galvaén, I.; Alavi, A.; Aleotti, F.; Aquilante, F.; Autschbach, J.; Avagliano, D.; Baiardi, A.; Bao, J. J.; Battaglia, S., et al. The OpenMolcas Web: A community-driven approach to advancing computational chemistry. *J. Chem. Theory Comput.* **2023**, *19*, 6933–6991.

(65) Roos, B. O. The complete active space SCF method in a fock-matrix-based super-CI formulation. *Int. J. Quantum Chem.* **1980**, *18*, 175–189.

(66) Siegbahn, P. E.; Almlöf, J.; Heiberg, A.; Roos, B. O. The complete active space SCF (CASSCF) method in a Newton–Raphson formulation with application to the HNO molecule. *J. Chem. Phys.* **1981**, *74*, 2384–2396.

(67) Rick, S. W.; Stuart, S. J.; Bader, J. S.; Berne, B. Fluctuating charge force fields for aqueous solutions. *J. Mol. Liq.* **1995**, *65*, 31–40.

(68) Thole, B. T. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.* **1981**, *59*, 341–350.

(69) Ohno, K. Some remarks on the Pariser-Parr-Pople method. *Theor. Chim. Acta* **1964**, *2*, 219–227.

(70) Chen, J.; Martínez, T. J. In *Advances in the Theory of Atomic and Molecular Systems: Conceptual and Computational Adv. Quantum Chem.*; Piecuch, P., Maruani, J., Delgado-Barrio, G., Wilson, S., Eds.; Springer Netherlands: Dordrecht, 2009; pp 397–415.

(71) Chelli, R.; Procacci, P. A transferable polarizable electrostatic force field for molecular mechanics based on the chemical potential equalization principle. *J. Chem. Phys.* **2002**, *117*, 9175–9189.

(72) Mortier, W. J.; Van Genechten, K.; Gasteiger, J. Electronegativity equalization: application and parametrization. *J. Am. Chem. Soc.* **1985**, *107*, 829–835.

(73) Gómez, S.; Giovannini, T.; Cappelli, C. Multiple facets of modeling electronic absorption spectra of systems in solution. *ACS Phys. Chem. Au* **2023**, *3*, 1–16.

(74) Gómez, S.; Lafiosca, P.; Egidi, F.; Giovannini, T.; Cappelli, C. Uv-resonance Raman spectra of systems in complex environments: A multiscale modeling applied to doxorubicin intercalated into dna. *J. Chem. Inf. Model.* **2023**, *63*, 1208–1217.

(75) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(76) Giovannini, T.; Lafiosca, P.; Chandramouli, B.; Barone, V.; Cappelli, C. Effective yet reliable computation of hyperfine coupling constants in solution by a QM/MM approach: Interplay between electrostatics and non-electrostatic effects. *J. Chem. Phys.* **2019**, *150*, 124102.

(77) Ferré, N.; Ángyán, J. G. Approximate electrostatic interaction operator for QM/MM calculations. *Chem. Phys. Lett.* **2002**, *356*, 331–339.

(78) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.

(79) Bao, J. J.; Truhlar, D. G. Automatic active space selection for calculating electronic excitation energies based on high-spin unrestricted hartree–fock orbitals. *J. Chem. Theory Comput.* **2019**, *15*, 5308–5318.

(80) Tóth, Z.; Pulay, P. Finding symmetry breaking Hartree-Fock solutions: The case of triplet instability. *J. Chem. Phys.* **2016**, *145*, 164102.

(81) Roos, B. O.; Veryazov, V.; Widmark, P.-O. Relativistic atomic natural orbital type basis sets for the alkaline and alkaline-earth atoms applied to the ground-state potentials for the corresponding dimers. *Theor. Chem. Acc.* **2004**, *111*, 345–351.

(82) Roos, B. O.; Lindh, R.; Malmqvist, P.-Å.; Veryazov, V.; Widmark, P.-O. New relativistic ANO basis sets for transition metal atoms. *J. Phys. Chem. A* **2005**, *109*, 6575–6579.

(83) Veryazov, V.; Malmqvist, P. Å.; Roos, B. O. How to select active space for multiconfigurational quantum chemistry? *Int. J. Quantum Chem.* **2011**, *111*, 3329–3338.

(84) DeFusco, A.; Minezawa, N.; Slipchenko, L. V.; Zahariev, F.; Gordon, M. S. Modeling solvent effects on electronic excited states. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2184–2192.

(85) Hagras, M. A.; Glover, W. J. Polarizable embedding for excited-state reactions: dynamically weighted polarizable QM/MM. *Journal of Chemical Theory and Computation* **2018**, *14*, 2137–2144.

(86) Liu, X.; Humeniuk, A.; Glover, W. J. Conical intersections in solution with polarizable embedding: Integral-exact direct reaction field. *Journal of Chemical Theory and Computation* **2022**, *18*, 6826–6839.

(87) Humeniuk, A.; Glover, W. J. Multistate, polarizable QM/MM embedding scheme based on the direct reaction field method: Solvatochromic shifts, analytical gradients and optimizations of conical intersections in solution. *Journal of Chemical Theory and Computation* **2024**, *20*, 2111–2126.

(88) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. Gaussian 16 Revision C.01. 2016; Gaussian Inc. Wallingford CT.

(89) Puzzarini, C.; Barone, V. Diving for accurate structures in the ocean of molecular systems with the help of spectroscopy and quantum chemistry. *Acc. Chem. Res.* **2018**, *51*, 548–556.

(90) Angeli, C.; Borini, S.; Ferrighi, L.; Cimiraglia, R. A CASSCF theoretical study of the vibrational frequencies and structure of formaldehyde, acetaldehyde and acetone valence excited states. *J. Mol. Struct.: THEOCHEM* **2005**, *718*, 55–69.

(91) Angeli, C.; Borini, S.; Ferrighi, L.; Cimiraglia, R. Ab initio n-electron valence state perturbation theory study of the adiabatic transitions in carbonyl molecules: Formaldehyde, acetaldehyde, and acetone. *J. Chem. Phys.* **2005**, *122*, 114304.

(92) Hoyer, C. E.; Ghosh, S.; Truhlar, D. G.; Gagliardi, L. Multiconfiguration pair-density functional theory is as accurate as CASPT2 for electronic excitation. *J. Phys. Chem. Lett.* **2016**, *7*, 586–591.

(93) Merchán, M.; Roos, B. O. A theoretical determination of the electronic spectrum of formaldehyde. *Theor. Chim. Acta* **1995**, *92*, 227–239.

(94) Walzl, K.; Koerting, C.; Kuppermann, A. Electron-impact spectroscopy of acetaldehyde. *J. Chem. Phys.* **1987**, *87*, 3796–3803.

(95) Taylor, S.; Wilden, D. G.; Comer, J. Electron energy-loss spectroscopy of forbidden transitions to valence and Rydberg states of formaldehyde. *Chem. Phys.* **1982**, *70*, 291–298.

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

(97) Bercovici, T.; King, J.; Becker, R. S. Formaldehyde: comprehensive spectral investigation as a function of solvent and temperature. *J. Chem. Phys.* **1972**, *56*, 3956–3963.

(98) Kongsted, J.; Osted, A.; Pedersen, T. B.; Mikkelsen, K. V.; Christiansen, O. The n$\rightarrow \pi^*$ electronic transition in microsolvated formaldehyde. A coupled cluster and combined coupled cluster/molecular mechanics study. *J. Phys. Chem. A* **2004**, *108*, 8624–8632.

(99) Xu, Z.; Matsika, S. Combined Multireference Configuration Interaction/Molecular Dynamics Approach for Calculating Solvatochromic Shifts: Application to the n O$\rightarrow \pi^*$ Electronic Transition of Formaldehyde. *J. Phys. Chem. A* **2006**, *110*, 12035–12043.

(100) Renge, I. Solvent induced transformations of n- π^* absorption in formaldehyde, acetaldehyde, and acetone. *J. Phys. Chem. A* **2015**, *119*, 8599–8610.

(101) Coutinho, K.; Canuto, S. Solvent effects in emission spectroscopy: A Monte Carlo quantum mechanics study of the n$\leftarrow \pi^*$ shift of formaldehyde in water. *J. Chem. Phys.* **2000**, *113*, 9132–9139.

(102) Canuto, S.; Coutinho, K. From hydrogen bond to bulk: Solvation analysis of the n-π^* transition of formaldehyde in water. *Int. J. Quantum Chem.* **2000**, *77*, 192–198.

(103) Kawashima, Y.; Dupuis, M.; Hirao, K. Monte Carlo microsolvation simulations for excited states using a mixed-Hamiltonian model with polarizable and vibrating waters: Applications to the blueshift of the H 2 CO 1 ($\pi^*\leftarrow$ n) excitation. *J. Chem. Phys.* **2002**, *117*, 248–257.

(104) Bayliss, N.; Wills-Johnson, G. Solvent effects on the intensities of the weak ultraviolet spectra of ketones and nitroparaffins—I. *Spectrochim. Acta A* **1968**, *24*, 551–561.

(105) Renge, I. Solvent Dependence of n- π^* Absorption in Acetone. *J. Phys. Chem. A* **2009**, *113*, 10678–10686.

(106) Lupieri, P.; Ippoliti, E.; Altoe, P.; Garavelli, M.; Mwalaba, M.; Carloni, P. Spectroscopic Properties of Formaldehyde in Aqueous Solution: Insights from Car- Parrinello and TDDFT/CASPT2 Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 3403–3409.

(107) Egidi, F.; Angelico, S.; Lafiosca, P.; Giovannini, T.; Cappelli, C. A polarizable three-layer frozen density embedding/molecular mechanics approach. *J. Chem. Phys.* **2021**, *154*, 164107.

(108) Nicoli, L.; Giovannini, T.; Cappelli, C. Assessing the quality of QM/MM approaches to describe vacuo-to-water solvatochromic shifts. *J. Chem. Phys.* **2022**, *157*, 214101.

(109) Frutos-Puerto, S.; Aguilar, M. A.; Fdez. Galvaén, I. Theoretical study of the preferential solvation effect on the solvatochromic shifts of para-nitroaniline. *The Journal of Physical Chemistry B* **2013**, *117*, 2466–2474.

(110) Soto, J.; Algarra, M. Electronic structure of nitrobenzene: A benchmark example of the accuracy of the multi-state CASPT2 theory. *J. Phys. Chem. A* **2021**, *125*, 9431–9437.

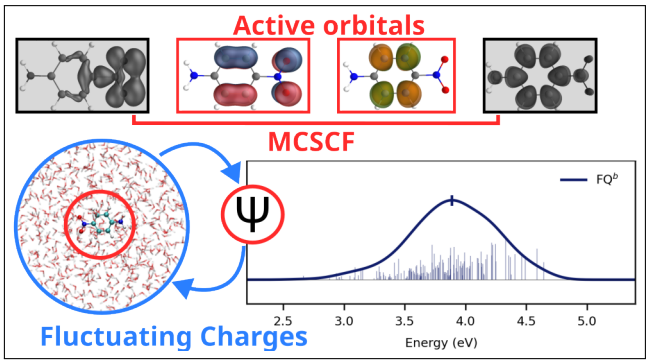
(111) Kovalenko, S.; Schanz, R.; Farztdinov, V.; Hennig, H.; Ernsting, N. Femtosecond relaxation of photoexcited para-nitroaniline: solvation, charge transfer, internal conversion and cooling. *Chem. Phys. Lett.* **2000**, *323*, 312–322.

(112) Persico, M.; Granucci, G. *Photochemistry: A Modern Theoretical Perspective*; Theoretical Chemistry and Computational Modelling; Springer International Publishing, 2018.

(113) Giovannini, T.; Lafiosca, P.; Cappelli, C. A general route to include Pauli repulsion and quantum dispersion effects in QM/MM approaches. *J. Chem. Theory Comput.* **2017**, *13*, 4854–4870.

(114) Lafiosca, P.; Rossi, F.; Egidi, F.; Giovannini, T.; Cappelli, C. Multiscale Frozen Density Embedding/Molecular Mechanics Approach for Simulating Magnetic Response Properties of Solvated Systems. *J. Chem. Theory Comput.* **2023**, *20*, 266–279.

TOC Graphic


Active orbitals
MCSCF
Ψ
FQ^b
2.5
3.0
3.5
4.0
4.5
5.0
Energy (eV)
Fluctuating Charges