

Cite this: DOI: 00.0000/xxxxxxxxxx

Combining Multilevel Hartree Fock and Multilevel Coupled Cluster with Molecular Mechanics: a Study of Electronic Excitations in Solutions[†]Linda Goletto,^{a‡} Tommaso Giovannini,^{a‡} Sarai D. Folkestad^a and Henrik Koch^{b*}

Received Date

Accepted Date

DOI: 00.0000/xxxxxxxxxx

We investigate the coupling of different quantum-embedding approaches with a third molecular-mechanics layer, which can be either polarizable or non-polarizable. In particular, such a coupling is discussed for the multilevel families of methods, in which the system is divided into an active and an inactive orbital spaces. The computational cost of the resulting three layers approaches is reduced by treating the long-range interactions at the classical level. The developed methods are tested by the calculation of excitation energies of molecular systems in aqueous solution, for which an atomistic description of the environment is crucial to correctly reproduce the specific solute-solvent interactions, such as hydrogen bonding. In particular, we present the results obtained for three different moieties—acrolein, pyridine and para-nitroaniline—showing that an almost perfect agreement with experimental data can be achieved when the relevant physico-chemical interactions are included in the modeling of the condensed phase.

1 Introduction

The study of excited-state properties of systems has become one of the most important topics in computational chemistry. The theoretical investigation of such properties for molecules in the gas-phase can be performed by using different approaches, ranging from time dependent density functional theory (TD-DFT) to the most accurate equation-of-motion coupled cluster (EOM-CC) or coupled cluster response theory (CCRT).^{1–5} When a moiety is dissolved in a solution, a reliable description of the solute-solvent interactions is crucial to reproduce the experimental findings.^{6–12} In fact, although the measured observable is due to the solute, it can be dramatically modified by the external environment, in particular when the solute is embedded in a strongly interacting solvent, such as water.^{13,14}

Condensed systems are usually treated by resorting to the so-called focused approaches, where the attention is centered on the solute, which retains the final property, and the solvent is instead treated at a lower level of theory.^{13,15–20} The most widespread models have been developed in the context of quantum mechanical (QM)/classical approaches, in which the solute is described at the QM level, whereas the environment is treated by means

of classical mechanics, either at the continuum^{18,20–23} or atomistic level.^{24,25} The latter approaches overcome the well-known limits of the continuum description by retaining the atomistic description of the solvent, which is treated by means of Molecular Mechanics (MM) force field. Thus, the resulting QM/MM methods allow for the treatment of specific solute-solvent interactions, such as hydrogen bonding. In most QM/MM approaches, the QM and MM parts interact at the electrostatic level, including also the mutual polarization between the two regions in the polarizable QM/MM approaches.^{26–34} Therefore, the purely quantum forces, such as Pauli repulsion and dispersion, are usually neglected, although they may play a significant role.^{35–40} Moreover, the accuracy of QM/MM methods strongly depends on the quality of the parametrization of the classical force field. In order to overcome such limitations, quantum embedding approaches can be used.^{41–55} The latter are usually based on partitioning the density of the investigated system into two parts: an active, which leads the property, and an inactive part, which perturbs the active moiety.

Among the several quantum-embedding methods, a variety of multilevel (ML) approaches based on Hartree Fock (HF) and coupled cluster (CC) theories have been developed in the last years.^{56–64} These MLHF and MLCC methods rely on a partitioning of the orbitals into active and inactive orbital sets. In MLHF, the active occupied molecular orbitals (MOs) are obtained by means of a Cholesky decomposition of some initial density of the system.^{65,66} Active virtual orbitals can be obtained from orthonormalized projected atomic orbitals (PAOs).^{67,68} The active

^a Department of Chemistry, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

^b Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy; E-mail: henrik.koch@sns.it

[†] Electronic Supplementary Information (ESI) available: data related to Figs. 3–7. See DOI: 10.1039/cXCP00000x/

[‡] L.G. and T.G. contributed equally to this work.

and inactive MOs can also be localized in their pre-defined spatial regions using the MLHF-AB method, which allows for a further reduction in the computational cost.⁶¹ The MLHF(-AB) wave function can be used as a reference state for following CC calculations (CC-in-MLHF).

In MLCC, the higher order excitation operators included in the cluster operator are restricted to a set of active MOs.⁵⁷ This active orbital space can, for instance, be selected as in MLHF—using Cholesky orbitals and PAOs—or by using correlated natural transition orbitals (CNTOs).⁶⁹ In the CC-in-HF family of methods,^{64,66,70} the cluster operator is fully restricted to the active orbital space; the active MOs are selected using Cholesky orbitals and PAOs. The reference state is represented by the full HF wave function, and therefore the CC-in-HF method is more accurate than a CC calculation using an MLHF reference (CC-in-MLHF), whose accuracy is intrinsically determined by the quality of the decomposed density matrix.

In this work, we propose different three-layer approaches which couple the aforementioned multilevel methods — i.e. CC-in-MLHF, CC-in-HF and MLCC — with a MM region, in order to effectively take into account the long range interactions dominated by the electrostatic contribution. The accuracy of these embedding schemes is assessed to the reproduction of the vacuo-to-water solvatochromic shifts.^{71,72} In particular, the water molecules closest to the molecules are included in the QM region, whereas the remaining waters are treated by means of the polarizable fluctuating charge (FQ) force field,^{13,14,73} which has been proved to be particularly suitable for describing the water medium.¹³ Within the proposed schemes, the electrostatic, polarization, Pauli repulsion and dispersion interactions between the solute and the closest water molecules are taken into account, together with the long-range electrostatic and polarization contributions.

The manuscript is organized as follows. In the next section, we outline the computational methods. In particular, we focus on CC2-in-HF, CC2-in-MLHF(-AB) and on MLCC2, and we present their coupling with a third MM layer, which can either be non-polarizable or polarizable (QM/FQ). We then discuss the computational protocol applied in the calculations. The developed methods are applied to the prediction of the vacuo-to-water solvatochromic shifts of acrolein (ACRO), para-nitroaniline (PNA) and pyridine (PY). A summary and the future perspectives of the approaches conclude the manuscript.

2 Theoretical Models

In this section, we briefly review the different theoretical methods that are used to simulate the excitation energies of molecular systems in solution. In particular, we present the theory for non-polarizable QM/MM and polarizable QM/FQ, MLHF, MLHF-AB, CC2, MLCC2 and CC2-in-HF. Also, the coupling between the QM methods and the MM region is discussed.

2.1 Non-polarizable QM/MM and polarizable QM/FQ

In QM/MM calculations the studied system is partitioned into two parts, which are treated at the QM and MM level, respectively.

The energy of the system can be written as:

$$E = E_{\text{QM}} + E_{\text{MM}} + E_{\text{QM/MM}}^{\text{int}}, \quad (1)$$

where, E_{QM} , E_{MM} and $E_{\text{QM/MM}}^{\text{int}}$ are the QM, MM and their interaction energies, respectively. The different QM/MM approaches developed in the literature mainly differ in the description of the interaction energy $E_{\text{QM/MM}}^{\text{int}}$. In most QM/MM models, $E_{\text{QM/MM}}^{\text{int}}$ reads:

$$E_{\text{QM/MM}}^{\text{int}} = E_{\text{QM/MM}}^{\text{ele}} + E_{\text{QM/MM}}^{\text{pol}}, \quad (2)$$

where $E_{\text{QM/MM}}^{\text{ele}}$ and $E_{\text{QM/MM}}^{\text{pol}}$ are the electrostatic and polarization energy terms, respectively. Therefore, the QM-MM Pauli repulsion and dispersion interactions are usually neglected. In electrostatic QM/MM embeddings, only the electrostatic contribution in Eq. 2 is retained. In particular, each MM atom is assigned a fixed charge, which polarizes the QM density. In polarizable QM/MM approaches, both $E_{\text{QM/MM}}^{\text{ele}}$ and $E_{\text{QM/MM}}^{\text{pol}}$ are retained. In this work, we exploit the QM/FQ approach because it is particularly tailored to treat aqueous solutions.^{13,14} In QM/FQ, each atom of the MM portion is assigned a charge (q) which can vary in agreement with the electronegativity equalization principle (EEP),⁷⁴ i.e. a charge flow occurs when two atoms have different chemical potential.^{73,75–78} The FQ force field is defined in terms of two atomic parameters, namely the electronegativity (χ) and the chemical hardness (η).

In both non-polarizable QM/MM and polarizable QM/FQ, the QM/MM interaction energy is given by:

$$E_{\text{QM/MM}}^{\text{int}} = \sum_i q_i V_i(\mathbf{D}), \quad (3)$$

where, $V_i(\mathbf{D})$ is the electric potential generated by the QM density, \mathbf{D} , on the i -th charge (q_i). Notice that, in the case of QM/TIP3P, the values of the charges q_i are fixed, whereas in QM/FQ they are obtained by minimizing the following energy expression:^{75,79}

$$\mathcal{E}[\mathbf{D}, \mathbf{q}, \lambda] = \text{tr} \mathbf{h} \mathbf{D} + \frac{1}{2} \text{tr} \mathbf{D} \mathbf{G}(\mathbf{D}) + \frac{1}{2} \mathbf{q}_\lambda^\dagger \mathbf{M} \mathbf{q}_\lambda + \mathbf{q}_\lambda^\dagger \mathbf{C}_Q + \mathbf{q}_\lambda^\dagger \mathbf{V}(\mathbf{D}). \quad (4)$$

where \mathbf{h} and \mathbf{G} are the one- and two-electron integral matrices, respectively. The term \mathbf{q}_λ is a vector, collecting the Lagrangian multipliers ensuring charge conservation and the electric variables associated with the FQ force field, i.e. charges \mathbf{q} . The \mathbf{M} matrix is the interaction kernel between the charges, containing also the two Lagrangian blocks.^{80,81} The term $\mathbf{q}_\lambda^\dagger \mathbf{C}_Q$ accounts for the interaction between permanent moments, i.e. atomic electronegativities and total charge constraints. The last term $\mathbf{q}_\lambda^\dagger \mathbf{V}(\mathbf{D})$ represents the interaction between the charges and the QM electric potential.

The FQ charges are obtained by a minimization procedure of the functional in Eq. 4, which can be recasted in the following set of linear equations:

$$\mathbf{M} \mathbf{q}_\lambda = -\mathbf{C}_Q - \mathbf{V}(\mathbf{D}). \quad (5)$$

The QM Fock matrix in both QM/MM approaches is:

$$F_{\mu\nu} = h_{\mu\nu} + G_{\mu\nu}(\mathbf{D}) + \sum_i q_i V_{i,\mu\nu}(\mathbf{D}). \quad (6)$$

Notice that in non-polarizable QM/MM, the QM/MM contribution to the Fock matrix is fixed because the charges are constant, thus it can be computed once at the beginning of the Self Consistent Field (SCF) procedure. In polarizable QM/FQ, the charges depend on the QM density, thus the QM/MM contribution to the Fock matrix has to be computed at each SCF step. In this way, the mutual polarization between the QM and FQ parts is taken into account.

2.2 Multilevel HF

Multilevel Hartree-Fock (MLHF)^{59,60} is a quantum-embedding approach, in which the studied system is partitioned into an active and an inactive (environment) region. The density of the system (\mathbf{D}) is expressed as the sum of an active (\mathbf{D}^A) and an inactive (\mathbf{D}^B) densities:

$$\mathbf{D} = \mathbf{D}^A + \mathbf{D}^B. \quad (7)$$

An initial set of active occupied orbitals, and consequently \mathbf{D}^A , is obtained through a partial Cholesky decomposition^{66,82} of \mathbf{D} , where the pivoting elements of the decomposition are restricted to correspond to AOs centered on a set of active atoms. An initial set of active virtual orbitals is obtained from orthonormalized projected atomic orbitals (PAOs)^{67,68} centered on the atoms in the active region of the system. The MLHF energy of the whole system can be then written as:⁵⁹

$$E = \text{Tr} \mathbf{h} \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) + \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^B) + \text{Tr} \mathbf{h} \mathbf{D}^B + \frac{1}{2} \text{Tr} \mathbf{D}^B \mathbf{G}(\mathbf{D}^B) + h_{\text{nuc}}, \quad (8)$$

where h_{nuc} is the nuclear repulsion. In MLHF, the active density is optimized through rotations among the active orbitals, while the inactive density is fixed. The inactive density takes part in the optimization procedure entering the *effective* Fock matrix expression through its two-electron integral term. The Fock matrix in the atomic orbital (AO) basis ($\{\chi_\mu\}$) reads:

$$F_{\mu\nu} = h_{\mu\nu} + G_{\mu\nu}(\mathbf{D}^A) + G_{\mu\nu}(\mathbf{D}^B). \quad (9)$$

Note that in MLHF both the electrostatic and the Pauli repulsion interactions between the active and inactive fragments are retained.

2.3 Multilevel Hartree-Fock AB

The MLHF energy defined in Eq. 8 allows for the minimization of the active energy in the field generated by the inactive density. However, the different energy terms cannot be assigned to the individual fragments A and B, because they are coupled in the one-electron term $\mathbf{h} = \mathbf{T} + \mathbf{V}^A + \mathbf{V}^B$, where \mathbf{T} is the kinetic operator and \mathbf{V}^A and \mathbf{V}^B are the electron-nuclear attraction operators for

the two parts. Using the definition of \mathbf{h} , Eq. 8 can be written as:⁶¹

$$E = \underbrace{\text{Tr} \mathbf{h}^A \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A)}_{E_A} + h_{\text{nuc}}^A + \underbrace{\text{Tr} \mathbf{h}^B \mathbf{D}^B + \frac{1}{2} \text{Tr} \mathbf{D}^B \mathbf{G}(\mathbf{D}^B)}_{E_B} + h_{\text{nuc}}^B + \underbrace{\text{Tr} \mathbf{V}^B \mathbf{D}^A + \text{Tr} \mathbf{V}^A \mathbf{D}^B + \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^B) + h_{\text{nuc}}^{AB}}_{E_{AB}}, \quad (10)$$

where h_{nuc}^A , h_{nuc}^B and h_{nuc}^{AB} are nuclear repulsion terms, and $\mathbf{h}^X = \mathbf{T} + \mathbf{V}^X$, with $\{X = A, B\}$. Using the Eq. 10, the active and inactive MOs can be localized in their user-specified regions. This can be done by exploiting the energy-based localization procedure recently developed by us.⁶¹ In this approach (called MLHF-AB), the sum of E_A and E_B is minimized in the space spanned by the occupied orbitals of both active and inactive parts, thus the total energy defined in Eq. 8 remains constant during the procedure. The sum of the active and inactive energies reads:

$$E_A + E_B = \text{Tr} \mathbf{h}^A \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) + \text{Tr} \mathbf{h}^B \mathbf{D} - \text{Tr} \mathbf{h}^B \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D} \mathbf{G}(\mathbf{D}) + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) - \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}) = \text{Tr}(\mathbf{V}^A - \mathbf{V}^B) \mathbf{D}^A + \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) - \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}) + \text{Tr} \mathbf{h}^B \mathbf{D} + \frac{1}{2} \text{Tr} \mathbf{D} \mathbf{G}(\mathbf{D}), \quad (11)$$

where the last two terms depend on the total density \mathbf{D} , and are therefore constant energy terms. The Fock matrix can be written as:

$$F_{\mu\nu} = V_{\mu\nu}^A - V_{\mu\nu}^B + G_{\mu\nu}(\mathbf{D}^A) - G_{\mu\nu}(\mathbf{D}^B) = V_{\mu\nu}^A - V_{\mu\nu}^B + 2G_{\mu\nu}(\mathbf{D}^A) - G_{\mu\nu}(\mathbf{D}). \quad (12)$$

Note that minimizing the sum of A and B parts in MLHF-AB (see Eq. 11) is equivalent to maximizing the interaction energy E_{AB} , i.e. maximizing the repulsion between the two fragments.

2.4 CC2, MLCC2, and CC2-in-HF

The coupled cluster wave function is given by

$$|\text{CC}\rangle = e^T |\text{R}\rangle, \quad T = \sum_{\mu} t_{\mu} \tau_{\mu}, \quad (13)$$

where T is the cluster operator, defined in terms of the cluster amplitudes, t_{μ} , and excitation operators, τ_{μ} , and where $|\text{R}\rangle$ is a reference determinant. The amplitudes are obtained by solving the projected coupled cluster equations,

$$\Omega_{\mu} = \langle \mu | e^{-T} H e^T | \text{R} \rangle = 0, \quad (14)$$

where $|\mu\rangle = \tau_\mu|R\rangle$. The coupled cluster ground state energy is given by:

$$E_{CC} = \langle R|e^{-T}He^T|R\rangle, \quad (15)$$

and excitation energies are calculated by determining the eigenvalues of the Jacobian matrix,

$$A_{\mu\nu} = \frac{\partial \Omega_\mu}{\partial t_\nu}. \quad (16)$$

The CC2⁸³ model is obtained by restricting T to only include single and double excitations with respect to the reference,

$$T = T_1 + T_2, \quad (17)$$

and treating the double excitation operator T_2 perturbatively. The projected coupled cluster equations become

$$\Omega_{\mu_1} = \langle \mu_1|\tilde{H} + [\tilde{H}, T_2]|R\rangle \quad (18)$$

$$\Omega_{\mu_2} = \langle \mu_2|\tilde{H} + [F, T_2]|R\rangle, \quad (19)$$

where \tilde{H} is the T_1 -transformed Hamiltonian:

$$\tilde{H} = e^{-T_1}He^{T_1}. \quad (20)$$

The multilevel CC2 (MLCC2)⁶² method belongs to the class of multilevel coupled cluster (MLCC) approaches, introduced by Myhre *et al.*^{56,57} These approaches aim to retain the accurate description of intensive properties that can be obtained with coupled cluster theory, while reducing the computational cost through the restriction of the higher order excitation operators to a subset of the molecular orbitals. In MLCC2, the T_2 operator is restricted to excite within an active orbital space, whereas the T_1 operator is unrestricted. If one also restricts T_1 to the active orbital space, one obtains the CC2-in-HF model. The determination of a sensible active orbital space is a prerequisite for the MLCC2 and CC2-in-HF methods. Active occupied orbitals can, for instance, be constructed through a partial Cholesky decomposition of the Hartree-Fock density, and active virtual orbitals can be obtained as orthonormalized PAOs, as in MLHF. Alternatively, for MLCC2, correlated natural transition orbitals (CNTOs) can be employed.^{62,69} With CNTOs, no active region of the system must be selected *a priori*, since the partition relies on the excitation vectors of a lower level method—i.e., CCS in a MLCC2 calculation.

2.5 Coupling MLHF and MLCC2 with an MM layer

In this section, we present and discuss the coupling between MLHF and MLCC2 with a third layer treated at the MM level. For non-polarizable MLHF/MM, this can be done by rewriting the total energy in Eq. 8 including QM/MM interaction energy terms:

$$E = \text{Tr} \mathbf{h} \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) + \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^B) + \text{Tr} \mathbf{h} \mathbf{D}^B + \frac{1}{2} \text{Tr} \mathbf{D}^B \mathbf{G}(\mathbf{D}^B) + h_{nuc} + \mathbf{q} \mathbf{V}(\mathbf{D}), \quad (21)$$

where $\mathbf{V}(\mathbf{D})$ is the electric potential acting on each MM charge

due to the total density matrix \mathbf{D} . The non-polarizable MLHF/MM Fock matrix is thus modified as:

$$F_{\mu\nu} = h_{\mu\nu} + G_{\mu\nu}(\mathbf{D}^A) + G_{\mu\nu}(\mathbf{D}^B) + \mathbf{q} \mathbf{V}_{\mu\nu}, \quad (22)$$

which corresponds to Eq. 9, with the additional $\mathbf{q} \mathbf{V}_{\mu\nu}$ term that is independent of the active density matrix. Therefore, such a contribution can be calculated once at the beginning of the SCF procedure.

In case of QM/FQ the total MLHF/FQ energy instead reads:

$$E = \text{Tr} \mathbf{h} \mathbf{D}^A + \frac{1}{2} \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^A) + \text{Tr} \mathbf{D}^A \mathbf{G}(\mathbf{D}^B) + \text{Tr} \mathbf{h} \mathbf{D}^B + \frac{1}{2} \text{Tr} \mathbf{D}^B \mathbf{G}(\mathbf{D}^B) + h_{nuc} + \frac{1}{2} \mathbf{q}_\lambda^\dagger \mathbf{M} \mathbf{q}_\lambda + \mathbf{q}_\lambda^\dagger \mathbf{C}_Q + \mathbf{q}_\lambda^\dagger \mathbf{V}(\mathbf{D}). \quad (23)$$

In order to calculate the FQ charges, Eq. 5 can be used, by setting $\mathbf{D} = \mathbf{D}^A + \mathbf{D}^B$, where \mathbf{D}^B is constant during the SCF procedure. The QM/FQ Fock matrix reduces to Eq. 22, however the charges vary during the SCF procedure, thus the last term is calculated at each step. In this way, mutual polarization between the MLHF active part and FQ charges is taken into account. The coupling between MLCC2 and a third MM layer is instead trivial because the MM contributions are introduced at the SCF level only. Therefore, if the reference state is calculated at the HF/MM level, the coupling between MLCC2 and the MM layer is automatically included.

3 Computational Details

The CC2-in-MLHF/MM, CC2-in-MLHF-AB/MM, CC2-in-HF/MM and MLCC2/MM methods are implemented in a development version of the electronic structure program *eT*.⁷⁰ The MLHF calculations are performed by first diagonalizing the initial Fock matrix constructed by using a superposition of molecular densities.^{84–86} Then, the obtained density is partitioned into A and B densities, using Cholesky decomposition (threshold: 10^{-2}) and PAOs for active occupied and virtual orbitals, respectively.^{59,66–68,82} Finally, the MLHF energy in Eq. 8 is minimized in the active MO space. The MLHF-AB/MM calculations are performed using the procedure developed in Ref. 61. In particular, the MLHF and occupied-occupied rotations minimizing Eq. 10 are iterated until convergence. In the MLHF-AB calculations, the number of Cholesky vectors in the decomposition corresponds to the number of occupied orbitals n_o for the active atoms, i.e. the largest n_o pivots corresponding to AOs centered on the active atoms are selected.⁶¹ For both MLHF and MLHF-AB, the excitation energies are computed by means of a CC2 calculation (CC2-in-MLHF(-AB)) performed in the same orbital space defined as MLHF(-AB) active.

The MLCC2 calculations are instead performed in the framework of the CNTOs approximation. These orbitals are tailored for the excited state calculation performed, specifying the corresponding state numbers. The number of occupied CNTOs is requested as well, while the number of virtual is automatically defined as $n_v = n_o \frac{N_v}{N_o}$, where the lowercase letters refer to the active virtual and occupied sets and the uppercase letters to full space.⁶² The double excitation operator indices are then restricted to the

active orbitals.

In addition to the aforementioned methods we also perform CC2-in-HF calculations, where only a subspace of the HF space is treated at CC2 level. A partial Cholesky decomposition (threshold: 10^{-2}) is used to define the occupied orbitals, while the virtual orbitals are separated through orthogonalized PAOs.

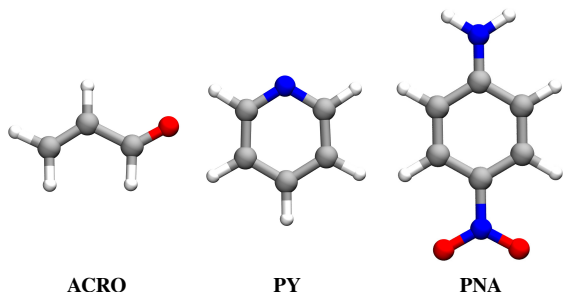


Fig. 1 Molecular structures of ACRO (left), PY (middle) and PNA (right).

All the discussed methods are applied to the calculations of excitation energies of acrolein (ACRO), pyridine (PY) and par-nitroaniline (PNA) in aqueous solution. Such molecules are chosen because their vacuo-to-water solvatochromism has been amply studied both theoretically and experimentally.^{65,87–110} In order to take into account the specific and dynamical aspects of solvation, we follow the computational protocol reported in Ref. 13, which is designed to accurately describe the particular case of aqueous solutions. More specifically, for each molecule a classical molecular dynamics (MD) simulation in aqueous solution was performed as reported in Refs. 34,37, where the MD analysis is also discussed. We therefore refer the interested reader to both references for more details on the MD runs. A set of 50 uncorrelated snapshots is extracted from the MDs and a solute-centered sphere is cut. The selected number of snapshots has been chosen to ensure convergence of the results (see Figs. S1-S3 given in ESI). The radius of the sphere is chosen to retain all solute-solvent interactions, and it is set to 15 Å (containing approximately 400 water molecules).^{34,37}

On the obtained droplets, we calculate CC2/aug-cc-pVDZ excitation energies of the solute moieties using the frozen core approximation and the different methods exposed above. In particular, the QM regions are defined by selecting the solute and all the water molecules that are placed at a distance lower than 3.5 Å with respect to each solute atom. In this way, the first solvation sphere is included.^{34,37} All the remaining water molecules are described using the non-polarizable TIP3P force field¹¹¹ (QM/TIP3P) or the polarizable FQ force field¹³ (QM/FQ), with the parametrization proposed in Ref. 36.

The default thresholds of the e^T program⁷⁰ are used, with the exception of the residual threshold in the CC excited state equations and of the Cholesky solver decomposition threshold,¹¹² both set to 10^{-4} . The excitation energy of the state s given a set of N_{snap} snapshots is calculated as:¹¹³

$$\langle \omega^s \rangle_{MD} = \frac{\sum_i^{N_{snap}} f_i^s \omega_i^s}{\sum_i^{N_{snap}} f_i^s}, \quad (24)$$

where, f_i^s and ω_i^s are the oscillator strength and the excitation energy of the i -th snapshot, respectively. The error bars at the 68% confidence are computed as the standard error (e_s), i.e. $e_s = \frac{\sigma}{\sqrt{n}}$, where σ is the standard deviation and n is the number of frames.

The current implementation of MLCC2 in e^T is limited to the calculation of excitation energies. Therefore, for all MLCC2 calculations, we use the CC2-in-HF oscillator strengths in Eq. 24.

4 Numerical Results

The selected molecules are characterized by $n \rightarrow \pi^*$ (PY), $\pi \rightarrow \pi^*$ (PNA) and both $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ (ACRO) electronic transitions. In this section, we first report a benchmark of the different theoretical methods on a single randomly selected snapshot of ACRO, as extracted from the MD simulation. Then, the methods that are found to provide a good compromise between the accuracy and the computational cost are applied to the full set of extracted snapshots of ACRO, PY and PNA. The chosen approaches are compared to the experimental results taken from the literature.^{106,113}

4.1 Benchmark

The performances of all the different combinations of the developed theoretical approaches are tested for a randomly selected snapshot of ACRO in aqueous solution. Acrolein is selected because its small size allows for a comparison with high level calculations.

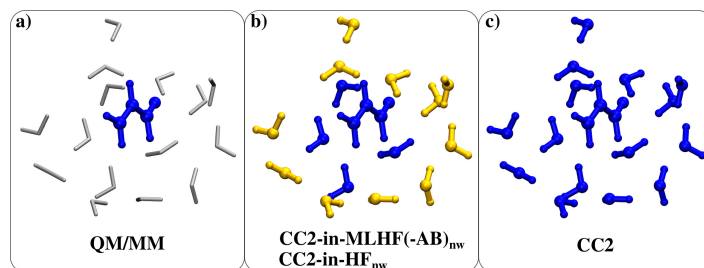


Fig. 2 Graphical depiction of the reduced snapshot of ACRO in aqueous solution as partitioned by (a) QM/MM approaches; (b) CC2-in-MLHF_{nw}, CC2-in-MLHF-AB_{nw} and CC2-in-HF_{nw} (the active part is represented in blue, the inactive one in yellow); (c) CC2.

First, we study the performances of the different methods on a reduced snapshot obtained by retaining all the water molecules that are closer than 3.5 Å from each acrolein atom, and by removing the remaining water molecules. This results in a snapshot with ACRO and 15 water molecules. The obtained system is described at different levels of theory:

- QM/MM: the ACRO moiety is treated at the CC2 level, whereas the 15 water molecules are described either at the TIP3P (i.e. electrostatic embedding) or at the FQ level (see Fig. 2 a).
- CC2-in-MLHF(-AB): only the ACRO moiety is treated as the active fragment at the MLHF level, whereas the 15 water molecules are inactive (see Fig. 2 b).
- CC2-in-MLHF(-AB)_{nw}: the ACRO moiety and the water

molecules closest to the ACRO center of mass are included in the active MLHF part. We include up to 5 water molecules (see Fig. 2 b). The n in the acronym specifies the number of water molecules included in the active part, while the remaining water molecules are inactive.

- MLCC2_{nCNTOS}: we select a number n of active occupied CNTOs equal to 25, 30, 35. In this way, we select the same number of occupied orbitals as present in ACRO and two, three or four water molecules.
- CC2-in-HF_{nw}: the reduced snapshot is treated with HF, and only ACRO and n water molecules are active in the CC2 calculation (see Fig. 2 b).
- CC2: the reduced snapshot is fully described at the CC2/aug-cc-pVDZ level of theory. This is the reference value for all the used methods (see Fig. 2 c).

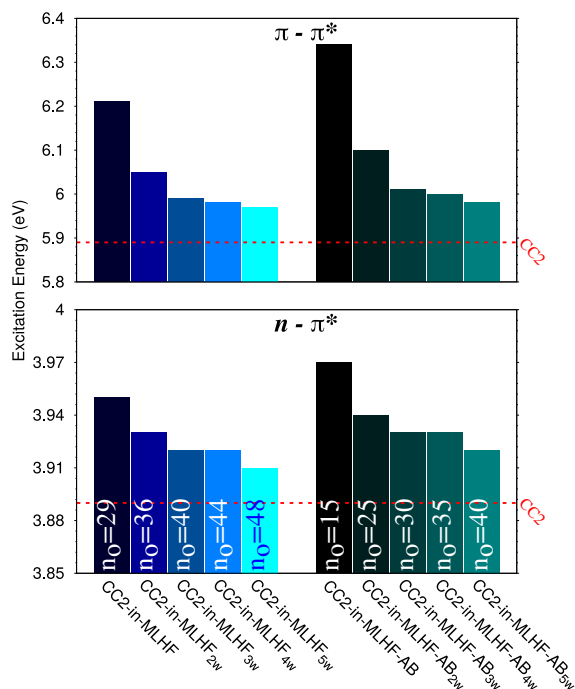


Fig. 3 The $n - \pi^*$ and $\pi - \pi^*$ excitation energies of the reduced snapshot of ACRO in aqueous solution computed at the CC2/aug-cc-pVDZ level, using the MLHF (left) and MLHF-AB (right) reference states.

We first comment on the results obtained using MLHF or MLHF-AB to describe the ground state reference. The main difference between the two approaches is that the number of occupied active orbitals in MLHF is determined by the Cholesky threshold, which is set to 10^{-2} , according to previous studies.^{57,64} In MLHF-AB, the number of occupied orbitals centered on the active atoms is instead selected to be localized as described in section 2.3, resulting in a different orbital set.⁶¹ MLHF and MLHF-AB are applied to the small snapshot represented in Fig. 2b, treating ACRO and n water molecules as active, whereas the remaining water molecules are inactive. The CC2-in-MLHF and CC2-in-MLHF-AB $n - \pi^*$ and $\pi - \pi^*$ excitation energies are reported in Fig. 3. The two approaches mostly differ for the results reported when only

the solute is treated as active, with CC2-in-MLHF-AB predicting an higher excitation energy for both $n - \pi^*$ (0.02 eV) and $\pi - \pi^*$ (0.13 eV) transitions. This is due to the fact that the number of occupied orbitals selected in MLHF is almost twice the number of those used in MLHF-AB (29 vs. 15). To better understand the differences between the CC2-in-MLHF and CC2-in-MLHF-AB methods, in Fig. 4 we report the computed ground state active and inactive densities. As it can be seen, with the MLHF reference the active density is spread also into the inactive region, whereas with MLHF-AB it is well localized in the ACRO moiety only. This is the origin of the observed differences in the excitation energies reported in Fig. 3.

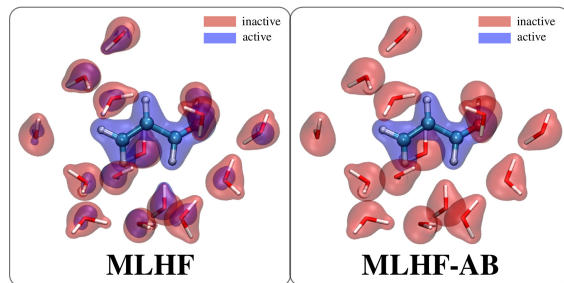


Fig. 4 MLHF and MLHF-AB active (blue) and inactive (red) HF densities of the reduced snapshot with ACRO being the required active part.

The discrepancy between the two methods rapidly decreases as a larger number of water molecules is included in the active region. In particular, CC2-in-MLHF-AB_{5w} excitation energies only differ by 0.01 eV with respect to the results obtained by means of the CC2-in-MLHF_{5w} approach (see also Table S1 given in the Electronic Supplementary Information - ESI). However, if the same active atoms are requested, the number of occupied in MLHF-AB are almost 10 less than MLHF. This yields a reduction of the computational cost which ranges from 50% to 33% going from zero to five water molecules included in the active part.

We now move on to comment on the results obtained by using all the other approaches, which are reported in Fig. 5 and in Tab. S1 in the ESI. We recall that the full space CC2 result is used as reference (see the red line in Fig. 5). Both QM/TIP3P and QM/FQ approaches are not able to correctly reproduce the reference value for either $n - \pi^*$ or $\pi - \pi^*$ excitations. These results are in line with what has already been reported in the literature. In fact, Pauli repulsion effects, which are neglected in the considered methods, play a crucial role in the correct description of ACRO transitions.^{37,106} Such a confinement effect is taken into account by all the quantum-embedding approaches used in this work, i.e. CC2-in-MLHF(-AB), CC2-in-HF and MLCC2. By enlarging the active part in the CC2-in-MLHF(-AB), CC2-in-HF and MLCC2 methods, the excitation energies of both studied transitions approach the reference CC2 values. In particular, for CC2-in-MLHF(-AB) and CC2-in-HF the best agreement is clearly achieved by including 5 water molecules in the active region; however, a good compromise between computational cost and accuracy is also reached by including 3 to 4 water molecules. The MLCC2 approach, finally, gives the best agreement with the reference data, and the results are almost constant by considering 25, 30 or 35 active

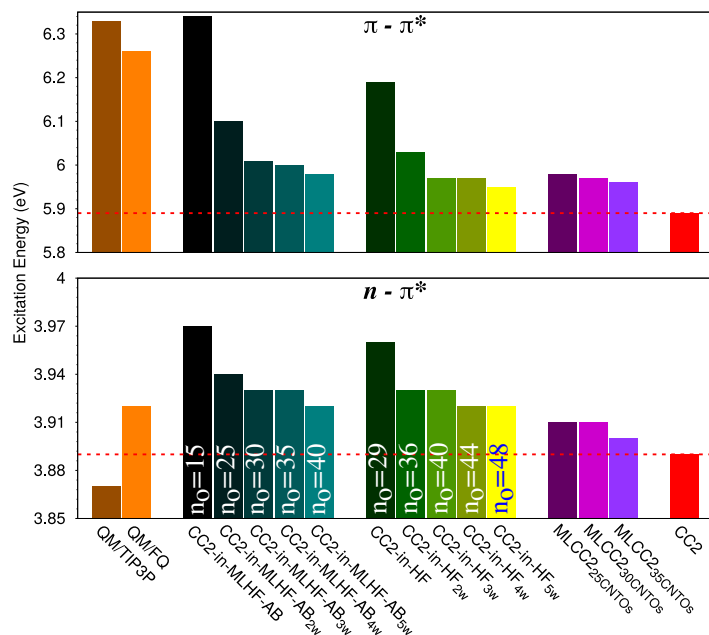


Fig. 5 The $n - \pi^*$ and $\pi - \pi^*$ excitation energies of the reduced snapshot of ACRO in aqueous solution (see Fig. 2) computed at CC2/aug-cc-pVDZ level of theory, using different approaches to restrict the orbital space.

CNTOs. Moreover, whereas in all the other multilevel methods used, the n active water molecules are selected on a geometrical criterion, in MLCC2 the active-inactive partitioning is performed based on the CCS excitation vectors.

All the investigated methods are also applied to the full snapshot, i.e. the previously discarded water molecules are included in the calculations, and treated in all cases by means of the non-polarizable TIP3P¹¹¹ or the polarizable FQ force field. This allows for the study of the performances of the three-layer approaches (CC2-in-MLHF(-AB)/MM, CC2-in-HF/MM, MLCC2/MM) we have developed. In particular, the long-range interactions are retained at the electrostatic level (with polarization effects in case of QM/FQ), thus neglecting Pauli repulsion. This is justified by the fact that Pauli repulsion is a short-range interaction. For the full snapshot, the reference approach is represented by the CC2-in-HF_{15w} results. In the latter method, the whole snapshot is treated at the HF level, whereas the small subsystem constituted by ACRO and 15 water molecules is CC2 active. A schematic picture of the partitioning used in the different methods is shown in Fig. 6.

The $n - \pi^*$ and $\pi - \pi^*$ excitation energies obtained using QM/TIP3P, QM/FQ, CC2-in-MLHF-AB/FQ, CC2-in-HF/FQ, MLCC2 and CC2-in-HF_{15w} are reported in Fig. 7 and in Tab. S2 in the ESI. Note that we also perform a comparison between CC2-in-MLHF-AB/FQ and CC2-in-MLHF/FQ, as previously presented in Fig. 3 (see Fig. S1 given in ESI). From the comparison between the reference values reported in Fig. 7 and 5, we notice that the inclusion of the previously discarded water molecules leads to an opposite shift for the two transitions. In particular, the CC2-in-HF_{15w} $n - \pi^*$ excitation energy increases of almost 0.1 eV with respect to the CC2 value of the reduced structure; for the bright $\pi - \pi^*$ excitation, the absorption energy is redshifted of about 0.1

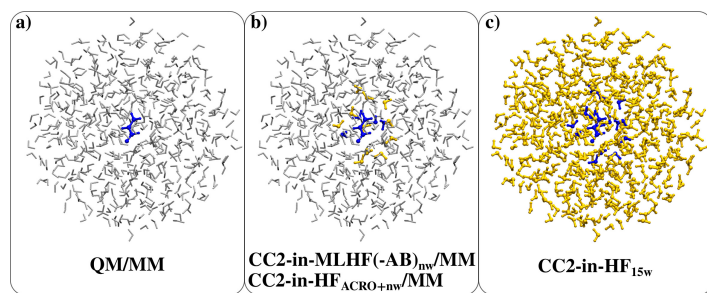


Fig. 6 Graphical depiction of the full snapshot of ACRO in aqueous solution as partitioned by (a) QM/MM approaches; (b) CC2-in-MLHF_{nw}/MM, CC2-in-MLHF-AB_{nw}/MM and CC2-in-HF_{nw}/MM, (the active part is represented in blue, the inactive one in yellow, the MM portion in silver licorice); (c) CC2-in-HF_{15w} (CC2 active in blue, HF active: blue + yellow).

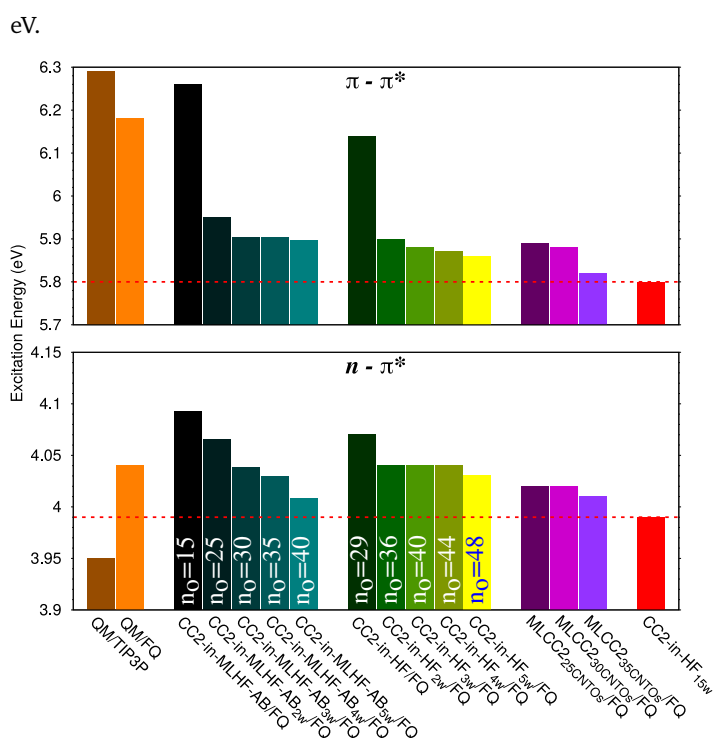


Fig. 7 The $n - \pi^*$ and $\pi - \pi^*$ excitation energies of the full snapshot of ACRO in aqueous solution (see Fig. 6) computed at CC2/aug-cc-pVDZ level of theory, using different approaches to describe the reference state.

The QM/TIP3P approach is not able to reproduce such shifts with respect to the reduced structure. In fact, the $n - \pi^*$ excitation energy increases of almost 0.08 eV, whereas the $\pi - \pi^*$ value remains almost unchanged (-0.04 eV). On the other hand, both excitations are correctly shifted when the polarizable QM/FQ model is used (see Tabs. S2-S3 in the ESI). Therefore, the polarization effects need to be taken into account to correctly reproduce electrostatic long-range effects. For this reason, we will present the results obtained by using the different QM approaches (CC2-in-MLHF, CC2-in-MLHF-AB, CC2-in-HF, MLCC2) coupled with the polarizable FQ approach. The results obtained by coupling the quantum-embedding approaches with the non-polarizable TIP3P are given in Tab. S3 in the ESI.

The results reported for all the methods coupled with the FQ force field confirm the correct description provided by FQ in the prediction of the shifts with respect to the reduced snapshots (see Tab. S2 in the ESI). Moreover, the general trends already discussed with reference to Fig. 5 continue to be valid also in this case. A good agreement with the CC2-in-HF_{15w} reference is obtained when 3 to 4 water molecules are included in the active part of CC2-in-MLHF-AB/FQ and CC2-in-HF/FQ approaches. Finally, the MLCC2 provides the best agreement with the reference data, and the results are almost constant when considering 25, 30 or 35 active occupied CNOs, with the lowest discrepancy shown by the MLCC2_{35CNOs}/FQ.

Based on this benchmark, we have selected six different approaches to be used in the calculation of the excitation energies of ACRO, PY and PNA in aqueous solution. In particular, we will present results obtained by exploiting the non-polarizable QM/TIP3P and different combinations of polarizable QM/FQ methods. In both QM/TIP3P and QM/FQ, only the solvated solutes are treated at the CC2 level. CC2-in-MLHF(-AB)/FQ calculations are performed by including three water molecules in the active part, whereas in CC2-in-HF, four active water molecules are selected. For MLCC2/FQ calculations, a number of active CNOs corresponding to the number of occupied orbitals of the solute and three water molecules is selected. The final results are obtained as an average over a set of representative structures extracted from classical MD simulations.

4.2 ACRO in aqueous solution

The first system we have studied is acrolein (ACRO) in aqueous solution. As discussed above, such a system is characterized by both a dark $n - \pi^*$ and a bright $\pi - \pi^*$ transitions. They present opposite solvatochromic shifts when passing from the gas-phase to the water solution.¹⁰⁶ In particular, the vacuo-to-solvent experimental $n - \pi^*$ solvatochromic shift consists in about -0.25 eV, i.e. the excitation energy increases. On the other hand, the $\pi - \pi^*$ absorption energy decreases when ACRO is dissolved in water, yielding a positive solvatochromism of 0.52 eV.¹⁰⁶ Although ACRO is a relatively small system, the theoretical prediction of the experimental data is particularly challenging, because quantum effects, such as Pauli repulsion, must be taken into account.^{37,106} With the exception of QM/MM methods, all the other quantum-embedding approaches we have benchmarked in the previous section can effectively reproduce the quantum confinement effects due to Pauli repulsion. The calculated excitation energies and the vacuo-to-solvent solvatochromic shifts are reported in Tab. 1, together with their experimental counterparts. The solvatochromic shifts are also graphically illustrated in Fig. 8. The gas-phase excitation energies have been calculated with a geometry optimized at the B2PLYP/cc-pVTZ level. We first comment on the trends of the excitation energies when moving from the gas phase to the aqueous solution. For the $n - \pi^*$ transition, all the investigated methods show a shift to higher energies with respect to the vacuo value; the opposite is the case for the $\pi - \pi^*$ excitation.

Comparing the gas-phase computed values with their experimental counterparts, a discrepancy of almost 0.3 eV is observed.

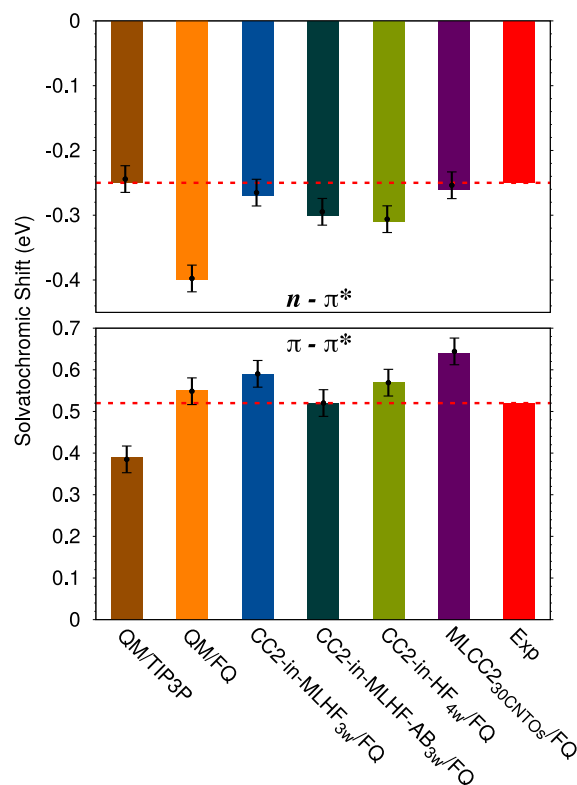


Fig. 8 Computed and experimental $n - \pi^*$ (top) and $\pi - \pi^*$ (bottom) vacuo-to-water solvatochromic shifts of ACRO (values are given in Tab. 1). Computed standard error bars at the 68% confidence interval are also plotted. The horizontal dashed lines follow the experimental data reproduced from Ref. 106.

Method	$n - \pi^*$		$\pi - \pi^*$	
	Exc. Ene. (eV)	Shift (eV)	Exc. Ene. (eV)	Shift (eV)
vacuo	3.86	–	6.78	–
exp (vac)	3.69	–	6.42	–
QM/TIP3P	4.11 \pm 0.02	-0.25 \pm 0.02	6.39 \pm 0.03	0.39 \pm 0.03
QM/FQ	4.26 \pm 0.02	-0.40 \pm 0.02	6.23 \pm 0.03	0.55 \pm 0.03
CC2-in-MLHF _{3w} /FQ	4.13 \pm 0.02	-0.27 \pm 0.02	6.19 \pm 0.03	0.59 \pm 0.03
CC2-in-MLHF-AB _{3w} /FQ	4.16 \pm 0.02	-0.30 \pm 0.02	6.26 \pm 0.03	0.52 \pm 0.03
CC2-in-HF _{4w} /FQ	4.17 \pm 0.02	-0.31 \pm 0.02	6.21 \pm 0.03	0.57 \pm 0.03
MLCC2 _{30CNTOS} /FQ	4.12 \pm 0.02	-0.26 \pm 0.02	6.14 \pm 0.03	0.64 \pm 0.03
exp (wat)	3.94	-0.25	5.90	0.52

Table 1 $n - \pi^*$ and $\pi - \pi^*$ excitation energies and vacuo-to-solvent solvachromic shifts of ACRO in vacuo and in aqueous solution. Theoretical solvatochromic shifts are calculated with respect to the vacuo value. Computed standard errors at the 68% confidence interval are also given. Experimental data are reproduced from Ref. 106.

This is due to the level of theory (CC2/aug-cc-pVDZ) which was used in all the calculations. Although such a difference is quite large, we point out that we are interested in the calculation of the vacuo-to-water solvatochromic shift, which, being a difference, is less affected by absolute errors of the excitation energies.²⁶ The solvatochromic shifts are reported both in Tab. 1 and in Fig. 8. In agreement with the experimental data, the computed values show an opposite vacuo-to-water solvatochromic shift, being positive for $\pi - \pi^*$, and negative for $n - \pi^*$.

The results obtained for QM/TIP3P and QM/FQ show a good agreement only with one of the two studied transitions — $n - \pi^*$ and $\pi - \pi^*$, respectively. This behavior has already been reported at the TD-DFT level in Ref. 37, in which it is also shown that the inclusion of Pauli repulsion, which is neglected in common QM/MM approaches, deteriorates the agreement with the experiment. Therefore, the almost perfect agreement reported for QM/MM methods is primarily due to error cancellation.

Pauli repulsion contributions, which lead to a confinement of the solute wave function, are automatically introduced in all the employed quantum-embedding approaches. The largest errors with respect to the experimental data are reported for the CC2-in-HF_{4w}/FQ and for the MLCC2_{30CNTOS}/FQ approaches, for the $n - \pi^*$ and the $\pi - \pi^*$ transitions, respectively. However, for the lowest dark transition, the CC2-in-HF discrepancy is of about 0.06 eV, and when the computed errors are taken into account, it is reduced to about 0.04 eV. This means that all the methods are suitable to accurately reproduce the experimental vacuo-to-water solvatochromic shift of the $n - \pi^*$ excitation. For the second excitation, the largest error is of about 0.12 \pm 0.03 eV for MLCC2_{30CNTOS}/FQ, whereas a perfect agreement is reported for CC2-in-MLHF-AB_{3w}/FQ. However, we note that the calculated solvatochromic shifts depend on the value of the excitation energy in the gas-phase, which is highly sensible to the geometry optimization level.¹⁰⁶ As an example, if the gas-phase excitation energies are computed on the same snapshots used for the calculations in aqueous solution by removing all water molecules, then the agreement with the experiments is improved in all cases (see Table S4 given in ESI).

As a final comment on ACRO, we note that in Ref. 106, applying a TD-DFT/MM scheme, a good prediction of the experimen-

tal solvatochromic shift is only achieved by including 12 water molecules in the QM region. In this work, we show that an excellent agreement with the experiment can be achieved by only including in the active part three or four water molecules. Therefore, we can state that the confinement effects introduced by the inactive water molecules are crucial to achieve the correct result.

4.3 PY in aqueous solution

Method	$n - \pi^*$	
	Exc. Ene. (eV)	Shift (eV)
vacuo	5.01	–
exp (vac)	4.59 ^a , 4.63 ^b , 4.74 ^c	–
QM/TIP3P	5.42 \pm 0.02	-0.41 \pm 0.02
QM/FQ	5.43 \pm 0.02	-0.42 \pm 0.02
CC2-in-MLHF _{3w} /FQ	5.42 \pm 0.02	-0.41 \pm 0.02
CC2-in-MLHF-AB _{3w} /FQ	5.41 \pm 0.02	-0.40 \pm 0.02
CC2-in-HF _{4w} /FQ	5.39 \pm 0.02	-0.38 \pm 0.02
MLCC2 _{36CNTOS} /FQ	5.37 \pm 0.02	-0.36 \pm 0.02
exp (wat)	4.94	-0.31 \pm 0.04

Table 2 $n - \pi^*$ excitation energies and vacuo-to-solvent solvatochromic shifts of PY in vacuo and in aqueous solution. Computed standard errors at the 68% confidence interval are also given. The experimental excitation energy of PY in aqueous solution is taken from Ref. 113.

^a Reproduced from Ref. 101

^b Reproduced from Ref. 108

^c Reproduced from Ref. 100

In the case of pyridine (PY), we focus on the dark $n - \pi^*$ transition, for which a vacuo-to-water solvatochromic shift of about 0.3-0.4 eV is reported in the literature.¹¹³ We calculate the gas-phase excitation energy at the geometry optimized using B3LYP/aug-cc-pVDZ level. We obtain an absorption energy of 5.01 eV at CC2/aug-cc-pVDZ level, which is shifted from the experimental data of about 0.3-0.4 eV (see Tab. 2), similarly to the case of ACRO. However, we note that different experimental gas-phase data are reported in the literature, ranging from 4.59 to 4.74 eV.^{100,101,108} Among them, the 4.63 eV value is taken as reference for the calculation of the vacuo-to-water solvatochromic shift,¹¹³ and we assign an uncertainty of about 0.04 eV.

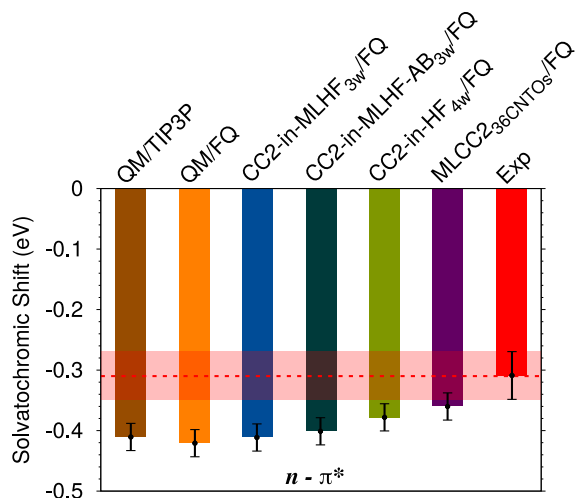


Fig. 9 Computed and experimental $n - \pi^*$ vacuo-to-water solvatochromic shifts of PY (Values are given in Tab. 2). Computed and experimental standard error bars are also plotted. The horizontal dashed line and the red box follow the experimental value and its error bar.

The absorption energies of PY in aqueous solution are computed by means of QM/TIP3P, polarizable QM/FQ, CC2-in-MLHF(-AB)/FQ, CC2-in-HF/FQ and MLCC2/FQ approaches. In particular, three and four water molecules are included in the active part of CC2-in-MLHF(-AB) and CC2-in-HF calculations, respectively. In MLCC2 we considered 36 active CNTOs, which correspond to PY and three water molecules. The computed absorption energies and the vacuo-to-water solvatochromic shifts are reported in Tab. 2 and in Fig. 9, along with their experimental counterparts. The experimental excitation energy in aqueous solution is taken from Ref. 113.

The excitation energies computed using the different approaches range from 5.37 eV with MLCC2_{36CnTOs}/FQ to 5.43 eV with QM/FQ, and are shifted of about 0.4 eV with respect to the experimental value. The corresponding solvatochromic shift ranges from -0.36 eV to -0.42 eV. Also for the reproduction of solvatochromism, the QM/FQ method yields the worst agreement with the experiment, whereas the best agreement is shown for MLCC2_{36CnTOs}/FQ. This confirms the results already commented for ACRO $n - \pi^*$ excitation, which presents the same trends.

From the investigation of Fig. 9, we note that only MLCC2_{36CnTOs}/FQ accurately reproduce the experimental value (within its uncertainty range). However, all the other quantum-embedding/FQ methods present an absolute error ranging from 0.01 (CC2-in-HF_{4w}/FQ) to 0.04 eV (CC2-in-MLHF_{3w}/FQ), thus resulting in an almost perfect agreement with the experimental reference.

4.4 PNA in aqueous solution

Finally, we investigate the vacuo-to-water solvatochromic shift of the bright $\pi - \pi^*$ transition of para-nitroaniline (PNA). The solvatochromism of the PNA moiety has been theoretically studied in previous works.^{34,88,93,109,113} From the experimental point of view, a solvatochromic shift of about 0.99 eV has been measured.^{99,103,113} As reported in Ref. 34, similarly to the other

Method	$\pi - \pi^*$	
	Exc. En. (eV)	Shift (eV)
vacuo	4.38	–
exp (vac)	4.25	–
QM/TIP3P	3.77 ± 0.01	0.61 ± 0.01
QM/FQ	3.43 ± 0.01	0.95 ± 0.01
CC2-in-MLHF _{3w} /FQ	3.61 ± 0.01	0.77 ± 0.01
CC2-in-MLHF-AB _{3w} /FQ	3.66 ± 0.01	0.72 ± 0.01
CC2-in-HF _{4w} /FQ	3.54 ± 0.01	0.84 ± 0.01
MLCC2 _{51CnTOs} /FQ	3.47 ± 0.01	0.93 ± 0.01
exp (wat)	3.26	0.99

Table 3 $\pi - \pi^*$ excitation energies and vacuo-to-solvent solvachromic shifts of PNA in vacuo and in aqueous solution. Computed standard errors at the 68% confidence interval are also given. Vacuo and water experimental values are reproduced from Ref. 99 and Ref. 103, respectively.

molecules, the gas phase excitation energy is computed at the CC2/aug-cc-pVDZ level using the CAM-B3LYP/aug-cc-pVDZ optimized geometry. Both vacuo and aqueous computed absorption energies are reported in Tab. 3, together with their experimental counterparts. The vacuo-to-water solvatochromic shifts are also given, and graphically depicted in Fig. 10. The MLCC2 calculations are performed by including 51 CNTOs in the active part, which correspond to PNA and three water molecules.

In this case, the computed and experimental gas phase values are in a good agreement, with a discrepancy of 0.13 eV. The vacuo-to-water solvatochromic shifts computed using the different methods employed to model solvation range from 0.61 eV with QM/TIP3P to 0.95 with QM/FQ. All the results are in line with what has already been commented for ACRO $\pi - \pi^*$ transition. In fact, the excitation energy in aqueous solution decreases with respect to the gas phase, resulting in a positive solvatochromism. Also, the two QM/MM methods predict the worst (QM/TIP3P) and the best (QM/FQ) agreement with the experiment. However, similarly to ACRO, such a result is due to error cancellation, in particular the neglect of any solute-solvent Pauli repulsion.

We now move on to comment on the results obtained using the investigated quantum-embedding approaches. Among them, the largest discrepancy with the experiment is reported for both CC2-in-MLHF/FQ and CC2-in-MLHF-AB/FQ. A slight improvement is shown by CC2-in-HF/FQ, for which the computed error is about 0.16 eV. We recall that for CC2-in-HF the reference state is represented by a HF calculation on the full QM system, whereas in CC2-in-MLHF(-AB) the reference state is obtained by partitioning the QM system into the active and inactive parts. As a consequence, for CC2-in-HF the mutual polarization between the active and the inactive components is fully taken into account, whereas in CC2-in-MLHF(-AB) it is only included at the first step of the computational protocol, i.e. diagonalizing the starting Fock matrix. We can therefore ascribe the better performance of CC2-in-HF to the importance of polarization terms in the reference state. Finally, among the studied quantum-embedding approaches, the most accurate result is observed for MLCC2_{51CnTOs}/FQ, for which the error with respect to the experiment is of about 0.06 eV. We

note that the better performance of MLCC2 can be related to the inclusion of polarization and dispersion terms at the CC level.

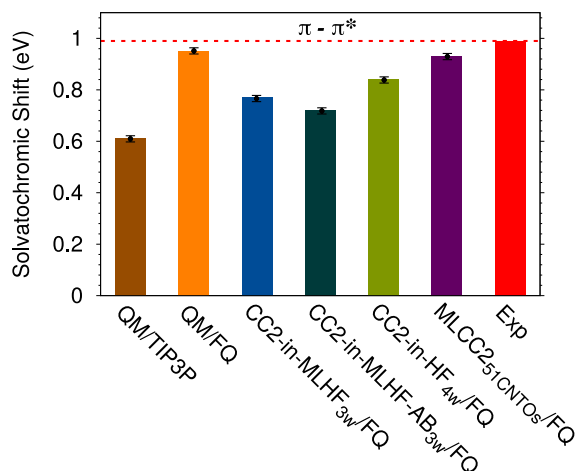


Fig. 10 Computed and experimental $\pi - \pi^*$ vacuo-to-water solvatochromic shifts of PNA (Values are given in Tab. 3). Computed standard error bars are also plotted. The horizontal dashed line follows the experimental value.

5 Summary and Conclusions

In this work, we have presented the coupling of different quantum-embedding approaches with a third MM layer, which can be either polarizable or non-polarizable. In particular, such coupling is discussed for the multilevel families of methods, such as CC2-in-MLHF, CC2-in-MLHF-AB, CC2-in-HF and MLCC2, which have been recently developed in our group. The resulting three layers approaches are particularly useful, because the computational cost of the quantum-embedding methods can be reduced by treating the long-range interactions at the MM level.

We have applied the developed methods to the calculation of the solvatochromic shifts of molecular systems in aqueous solution, for which an atomistic description of the environment is crucial to correctly describe the specific solute-solvent interactions, such as hydrogen bonding. The absorption energies have been calculated at the CC2 level of theory, which guarantees a good compromise between the computational cost and the accuracy. In particular, we have selected three moieties, ACRO, PY and PNA, for which experimental vacuo-to-solvent solvatochromic shifts have been reported in the literature.

Firstly, we have discussed a benchmark of the methods for a randomly selected snapshot of ACRO, for which we have shown that the third MM layer can accurately take into account the shift related to the long-range interactions. As a consequence, the quantum region can be limited to the first solvation shell. We have shown this to be sufficient to take into account the most relevant short-range interactions, such as Pauli repulsion. Among the different investigated approaches, we show that the MLCC2/FQ method can provide an accurate prediction of both $n - \pi^*$ and $\pi - \pi^*$ excitations, which are the most diffuse transitions in organic molecules. This can be achieved by only including a number of CNTOs corresponding to the solute and three water molecules. A good agreement with the experiment is achieved by including

three or four waters in all the other approaches, although the accuracy of the results strongly depends on the considered system. We finally remark that the three-layer approaches were applied on a set of uncorrelated snapshots extracted from a classical MD simulation in order to take into account the dynamical effects of the solvation phenomenon.¹³ However, the results are indeed dependent on the quality of the classical force field used in the MD sampling, which can be refined by resorting to *ab-initio* MD or QM/MM MD simulations.¹¹⁴ Also, we note that in our modelling vibronic effects are not taken into account, but they can play a relevant role when comparing computed data with experiments.^{115,116}

To conclude, we point out that the polarization in polarizable QM/MM calculations is only included in the ground state reference. However, it can also be introduced in the following CC calculations.^{117–123} Such an extension, together with the inclusion of QM/MM Pauli repulsion³⁵ at the CC level, will be the topic of future communications. Finally, we note that the three layer approach that we have developed can also be applied to any QM/classical approach,¹²⁴ as for instance the polarizable continuum model (PCM),^{20,125} due to its theoretical similarity with the QM/FQ approach.¹²⁶

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge funding from the Marie Skłodowska-Curie European Training Network “COSINE - Computational Spectroscopy In Natural sciences and Engineering”, Grant Agreement No. 765739, and the Research Council of Norway through FRINATEK projects 263110 and 275506.

Notes and references

- 1 J. F. Stanton and R. J. Bartlett, *The Journal of chemical physics*, 1993, **98**, 7029–7039.
- 2 T. Helgaker, P. Jorgensen and J. Olsen, *Molecular electronic-structure theory*, John Wiley & Sons, 2014.
- 3 H. Koch, H. J. Jensen, P. Jorgensen and T. Helgaker, *J. Chem. Phys.*, 1990, **93**, 3345–3350.
- 4 H. Koch and P. Jorgensen, *The Journal of chemical physics*, 1990, **93**, 3333–3344.
- 5 T. B. Pedersen and H. Koch, *The Journal of chemical physics*, 1997, **106**, 8059–8072.
- 6 O. Cannelli, T. Giovannini, A. Baiardi, B. Carlotti, F. Elisei and C. Cappelli, *Phys. Chem. Chem. Phys.*, 2017, **19**, 32544–32555.
- 7 B. Carlotti, A. Cesaretti, O. Cannelli, T. Giovannini, C. Cappelli, C. Bonaccorso, C. G. Fortuna, F. Elisei and A. Spalletti, *J. Phys. Chem. C*, 2018, **122**, 2285–2296.
- 8 B. Mennucci, R. Cammi and J. Tomasi, *J. Chem. Phys.*, 1998, **109**, 2798–2807.
- 9 D. Jacquemin, B. Mennucci and C. Adamo, *Phys. Chem. Chem. Phys.*, 2011, **13**, 16987–16998.

- 10 S. Corni, R. Cammi, B. Mennucci and J. Tomasi, *J. Chem. Phys.*, 2005, **123**, 134512.
- 11 L. Cupellini, M. Corbella, B. Mennucci and C. Curutchet, *WIREs Comput. Mol. Sci.*, 2019, **9**, e1392.
- 12 B. Mennucci and S. Corni, *Nat. Rev. Chem.*, 2019, **3**, 315–330.
- 13 T. Giovannini, F. Egidi and C. Cappelli, *Chem. Soc. Rev.*, 2020, **49**, 5664–5677.
- 14 T. Giovannini, F. Egidi and C. Cappelli, *Phys. Chem. Chem. Phys.*, 2020, **22**, 22864–22879.
- 15 A. Warshel and M. Levitt, *J. Mol. Biol.*, 1976, **103**, 227–249.
- 16 A. Warshel and M. Karplus, *J. Am. Chem. Soc.*, 1972, **94**, 5612–5625.
- 17 S. Miertuš, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
- 18 J. Tomasi and M. Persico, *Chem. Rev.*, 1994, **94**, 2027–2094.
- 19 M. Orozco and F. J. Luque, *Chem. Rev.*, 2000, **100**, 4187–4226.
- 20 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.
- 21 B. Mennucci, *J. Phys. Chem. Lett.*, 2010, **1**, 1666–1674.
- 22 B. Mennucci, *Phys. Chem. Chem. Phys.*, 2013, **15**, 6583–6594.
- 23 F. Lipparini and B. Mennucci, *J. Chem. Phys.*, 2016, **144**, 160901.
- 24 H. M. Senn and W. Thiel, *Angew. Chem. Int. Ed.*, 2009, **48**, 1198–1229.
- 25 H. Lin and D. G. Truhlar, *Theor. Chem. Acc.*, 2007, **117**, 185–199.
- 26 D. Loco, É. Polack, S. Caprasecca, L. Lagardere, F. Lipparini, J.-P. Piquemal and B. Mennucci, *J. Chem. Theory Comput.*, 2016, **12**, 3654–3661.
- 27 C. Curutchet, A. Muñoz-Losa, S. Monti, J. Kongsted, G. D. Scholes and B. Mennucci, *J. Chem. Theory Comput.*, 2009, **5**, 1838–1848.
- 28 J. M. H. Olsen and J. Kongsted, *Adv. Quantum Chem.*, 2011, **61**, 107–143.
- 29 A. H. Steindal, K. Ruud, L. Frediani, K. Aidas and J. Kongsted, *J. Phys. Chem. B*, 2011, **115**, 3027–3037.
- 30 E. Boulanger and W. Thiel, *J. Chem. Theory Comput.*, 2012, **8**, 4527–4538.
- 31 S. Jurinovich, C. Curutchet and B. Mennucci, *ChemPhysChem*, 2014, **15**, 3194–3204.
- 32 T. Giovannini, A. Puglisi, M. Ambrosetti and C. Cappelli, *J. Chem. Theory Comput.*, 2019, **15**, 2233–2245.
- 33 T. Giovannini, L. Grazioli, M. Ambrosetti and C. Cappelli, *J. Chem. Theory Comput.*, 2019, **15**, 5495–5507.
- 34 T. Giovannini, R. R. Riso, M. Ambrosetti, A. Puglisi and C. Cappelli, *J. Chem. Phys.*, 2019, **151**, 174104.
- 35 T. Giovannini, P. Lafiosca and C. Cappelli, *J. Chem. Theory Comput.*, 2017, **13**, 4854–4870.
- 36 T. Giovannini, P. Lafiosca, B. Chandramouli, V. Barone and C. Cappelli, *J. Chem. Phys.*, 2019, **150**, 124102.
- 37 T. Giovannini, M. Ambrosetti and C. Cappelli, *J. Phys. Chem. Lett.*, 2019, **10**, 5823–5829.
- 38 H. Gokcan, E. G. Kratz, T. A. Darden, J.-P. Piquemal and G. A. Cisneros, *J. Phys. Chem. Lett.*, 2018, **9**, 3062–3067.
- 39 P. Arora, L. V. Slipchenko, S. P. Webb, A. DeFusco and M. S. Gordon, *J. Phys. Chem. A*, 2010, **114**, 6742–6750.
- 40 C. Curutchet, L. Cupellini, J. Kongsted, S. Corni, L. Frediani, A. H. Steindal, C. A. Guido, G. Scalmani and B. Mennucci, *J. Chem. Theory Comput.*, 2018, **14**, 1671–1681.
- 41 M. S. Gordon, Q. A. Smith, P. Xu and L. V. Slipchenko, *Annu. Rev. Phys. Chem.*, 2013, **64**, 553–578.
- 42 M. S. Gordon, L. Slipchenko, H. Li and J. H. Jensen, *Annu. Rep. Comput. Chem.*, 2007, **3**, 177–193.
- 43 Q. Sun and G. K.-L. Chan, *Acc. Chem. Res.*, 2016, **49**, 2705–2712.
- 44 G. Knizia and G. K.-L. Chan, *J. Chem. Theory Comput.*, 2013, **9**, 1428–1432.
- 45 D. V. Chulhai and J. D. Goodpaster, *J. Chem. Theory Comput.*, 2018, **14**, 1928–1942.
- 46 D. V. Chulhai and J. D. Goodpaster, *J. Chem. Theory Comput.*, 2017, **13**, 1503–1508.
- 47 X. Wen, D. S. Graham, D. V. Chulhai and J. D. Goodpaster, *J. Chem. Theory Comput.*, 2020, **16**, 385–398.
- 48 F. Ding, F. R. Manby and T. F. Miller III, *J. Chem. Theory Comput.*, 2017, **13**, 1605–1615.
- 49 J. D. Goodpaster, T. A. Barnes, F. R. Manby and T. F. Miller III, *J. Chem. Phys.*, 2012, **137**, 224113.
- 50 J. D. Goodpaster, T. A. Barnes, F. R. Manby and T. F. Miller III, *J. Chem. Phys.*, 2014, **140**, 18A507.
- 51 F. R. Manby, M. Stella, J. D. Goodpaster and T. F. Miller III, *J. Chem. Theory Comput.*, 2012, **8**, 2564–2568.
- 52 J. D. Goodpaster, N. Ananth, F. R. Manby and T. F. Miller III, *J. Chem. Phys.*, 2010, **133**, 084103.
- 53 K. Zhang, S. Ren and M. Caricato, *J. Chem. Theory Comput.*, 2020, **16**, 4361–4372.
- 54 P. Ramos, M. Papadakis and M. Pavanello, *J. Phys. Chem. B*, 2015, **119**, 7541–7557.
- 55 M. Pavanello and J. Neugebauer, *J. Chem. Phys.*, 2011, **135**, 234103.
- 56 R. H. Myhre, A. M. Sánchez de Merás and H. Koch, *Molecular Physics*, 2013, **111**, 1109–1118.
- 57 R. H. Myhre, A. M. Sánchez de Merás and H. Koch, *J. Chem. Phys.*, 2014, **141**, 224105.
- 58 R. H. Myhre and H. Koch, *J. Chem. Phys.*, 2016, **145**, 044111.
- 59 S. Sæther, T. Kjærgaard, H. Koch and I.-M. Høyvik, *J. Chem. Theory Comput.*, 2017, **13**, 5282–5290.
- 60 I.-M. Høyvik, *Mol. Phys.*, 2020, **118**, 1626929.
- 61 T. Giovannini and H. Koch, *J. Chem. Theory Comput.*, 2021, **17**, 139–150.
- 62 S. D. Folkestad and H. Koch, *J. Chem. Theory Comput.*, 2019, **16**, 179–189.
- 63 S. D. Folkestad and H. Koch, *J. Chem. Theory Comput.*, 2020, **16**, 6869–6879.

- 64 S. D. Folkestad, E. F. Kjønsstad, L. Goletto and H. Koch, *J. Chem. Theory Comput.*, 2021, DOI: 10.1021/acs.jctc.0c00590.
- 65 F. Aquilante, V. Barone and B. O. Roos, *J. Chem. Phys.*, 2003, **119**, 12323–12334.
- 66 A. M. Sánchez de Merás, H. Koch, I. G. Cuesta and L. Boman, *J. Chem. Phys.*, 2010, **132**, 204105.
- 67 P. Pulay, *J. Chem. Phys.*, 1983, **78**, 5043–5051.
- 68 S. Saebo and P. Pulay, *Annu. Rev. Phys. Chem.*, 1993, **44**, 213–236.
- 69 I.-M. Høyvik, R. H. Myhre and H. Koch, *J. Chem. Phys.*, 2017, **146**, 144109.
- 70 S. D. Folkestad, E. F. Kjønsstad, R. H. Myhre, J. H. Andersen, A. Balbi, S. Coriani, T. Giovannini, L. Goletto, T. S. Haugland, A. Hutcheson *et al.*, *J. Chem. Phys.*, 2020, **152**, 184103.
- 71 C. Reichardt, *Chem. Soc. Rev.*, 1992, **21**, 147–153.
- 72 C. Reichardt, *Chem. Rev.*, 1994, **94**, 2319–2358.
- 73 C. Cappelli, *Int. J. Quantum Chem.*, 2016, **116**, 1532–1542.
- 74 R. Sanderson, *Science*, 1951, **114**, 670–672.
- 75 F. Lipparini, C. Cappelli and V. Barone, *J. Chem. Theory Comput.*, 2012, **8**, 4153–4165.
- 76 S. W. Rick, S. J. Stuart and B. J. Berne, *J. Chem. Phys.*, 1994, **101**, 6141–6156.
- 77 S. W. Rick, S. J. Stuart, J. S. Bader and B. Berne, *J. Mol. Liq.*, 1995, **65**, 31–40.
- 78 S. W. Rick and B. J. Berne, *J. Am. Chem. Soc.*, 1996, **118**, 672–679.
- 79 F. Lipparini, C. Cappelli, G. Scalmani, N. De Mitri and V. Barone, *J. Chem. Theory Comput.*, 2012, **8**, 4270–4278.
- 80 F. Lipparini and V. Barone, *J. Chem. Theory Comput.*, 2011, **7**, 3711–3724.
- 81 T. Giovannini, M. Ambrosetti and C. Cappelli, *Theor. Chem. Acc.*, 2018, **137**, 74.
- 82 F. Aquilante, T. Bondo Pedersen, A. Sánchez de Merás and H. Koch, *J. Chem. Phys.*, 2006, **125**, 174101.
- 83 O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.
- 84 J. Neugebauer, M. J. Louwerse, E. J. Baerends and T. A. Wesolowski, *J. Chem. Phys.*, 2005, **122**, 094115.
- 85 X. He and K. M. Merz Jr, *J. Chem. Theory Comput.*, 2010, **6**, 405–411.
- 86 G. Marrazzini, T. Giovannini, M. Scavino, F. Egidi, C. Cappelli and H. Koch, *arXiv preprint arXiv:2009.05333*, 2020.
- 87 S. Sok, S. Y. Willow, F. Zahariev and M. S. Gordon, *J. Phys. Chem. A*, 2011, **115**, 9801–9809.
- 88 D. Kosenkov and L. V. Slipchenko, *J. Phys. Chem. A*, 2010, **115**, 392–401.
- 89 J. J. Eriksen, S. P. Sauer, K. V. Mikkelsen, O. Christiansen, H. J. A. Jensen and J. Kongsted, *Mol. Phys.*, 2013, **111**, 1235–1248.
- 90 K. Sneskov, T. Schwabe, O. Christiansen and J. Kongsted, *Phys. Chem. Chem. Phys.*, 2011, **13**, 18551–18560.
- 91 J. M. Olsen, K. Aidas and J. Kongsted, *J. Chem. Theory Comput.*, 2010, **6**, 3721–3734.
- 92 J. J. Eriksen, S. P. Sauer, K. V. Mikkelsen, H. J. A. Jensen and J. Kongsted, *J. Comput. Chem.*, 2012, **33**, 2012–2022.
- 93 S. Frutos-Puerto, M. A. Aguilar and I. Fdez. Galv'an, *J. Phys. Chem. B*, 2013, **117**, 2466–2474.
- 94 A. DeFusco, N. Minezawa, L. V. Slipchenko, F. Zahariev and M. S. Gordon, *J. Phys. Chem. Lett.*, 2011, **2**, 2184–2192.
- 95 B. Mennucci, *J. Am. Chem. Soc.*, 2002, **124**, 1506–1515.
- 96 M. Pagliai, G. Mancini, I. Carnimeo, N. De Mitri and V. Barone, *J. Comput. Chem.*, 2017, **38**, 319–335.
- 97 M. Cossi and V. Barone, *J. Chem. Phys.*, 2001, **115**, 4708–4717.
- 98 S. Mason, *J. Chem. Soc.*, 1959, 1240–1246.
- 99 S. Millefiori, G. Favini, A. Millefiori and D. Grasso, *Spectrochim. Acta A*, 1977, **33**, 21–27.
- 100 Z.-L. Cai and J. R. Reimers, *J. Phys. Chem. A*, 2000, **104**, 8389–8408.
- 101 M. Schreiber, M. R. Silva-Junior, S. P. Sauer and W. Thiel, *J. Chem. Phys.*, 2008, **128**, 134110.
- 102 L. Prabhunirashi and S. Kunte, *Spectrochim. Acta A*, 1986, **42**, 435–441.
- 103 S. Kovalenko, R. Schanz, V. Farztdinov, H. Hennig and N. Ernstring, *Chem. Phys. Lett.*, 2000, **323**, 312–322.
- 104 K. J. de Almeida, K. Coutinho, W. B. De Almeida, W. R. Rocha and S. Canuto, *Phys. Chem. Chem. Phys.*, 2001, **3**, 1583–1587.
- 105 A. Moskvina, O. Yablonskii and L. Bondar, *Theor. Exp. Chem.*, 1966, **2**, 469–472.
- 106 K. Aidas, A. Møgelhøj, E. J. Nilsson, M. S. Johnson, K. V. Mikkelsen, O. Christiansen, P. Sönderhjelm and J. Kongsted, *J. Chem. Phys.*, 2008, **128**, 194503.
- 107 G. Brancato, N. Rega and V. Barone, *J. Chem. Phys.*, 2006, **125**, 164515.
- 108 A. V. Marenich, C. J. Cramer, D. G. Truhlar, C. A. Guido, B. Mennucci, G. Scalmani and M. J. Frisch, *Chem. Sci.*, 2011, **2**, 2143–2161.
- 109 I. Duchemin, C. A. Guido, D. Jacquemin and X. Blase, *Chem. Sci.*, 2018, **9**, 4430–4443.
- 110 C. Bistafa, L. Modesto-Costa and S. Canuto, *Theor. Chem. Acc.*, 2016, **135**, 129.
- 111 P. Mark and L. Nilsson, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.
- 112 S. D. Folkestad, E. F. Kjønsstad and H. Koch, *J. Chem. Phys.*, 2019, **150**, 194112.
- 113 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2015, **119**, 958–967.
- 114 E. R. Kjellgren, J. M. Haugaard Olsen and J. Kongsted, *J. Chem. Theory Comput.*, 2018, **14**, 4309–4319.
- 115 F. Santoro and D. Jacquemin, *WIREs Comput. Mol. Sci.*, 2016, **6**, 460–486.
- 116 A. Segalina, J. Cerezo, G. Prampolini, F. Santoro and M. Pastore, *J. Chem. Theory Comput.*, 2020, **16**, 7061–7077.
- 117 M. Caricato, *Comput. Theor. Chem.*, 2014, **1040**, 99–105.

- 118 M. Caricato, G. Scalmani, G. W. Trucks and M. J. Frisch, *J. Phys. Chem. Lett.*, 2010, **1**, 2369–2373.
- 119 M. Caricato, G. W. Trucks, M. J. Frisch and K. B. Wiberg, *J. Chem. Theory Comput.*, 2010, **6**, 370–383.
- 120 M. Caricato, *J. Chem. Phys.*, 2013, **139**, 114103.
- 121 M. Caricato, *Int. J. Quantum Chem.*, 2018, e25674.
- 122 R. Cammi, *Int. J. Quantum Chem.*, 2012, **112**, 2547–2560.
- 123 M. Caricato, F. Lipparini, G. Scalmani, C. Cappelli and V. Barone, *J. Chem. Theory Comput.*, 2013, **9**, 3035–3042.
- 124 M. Nottoli and F. Lipparini, *J. Chem. Phys.*, 2020, **153**, 224108.
- 125 B. Mennucci, *WIREs Comput. Mol. Sci.*, 2012, **2**, 386–404.
- 126 R. Di Remigio, T. Giovannini, M. Ambrosetti, C. Cappelli and L. Frediani, *J. Chem. Theory Comput.*, 2019, **15**, 4056–4068.