Quantum Mechanics/Extremely Localized Molecular Orbital Embedding Strategy for Excited States: Coupling to Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster

Giovanni Macetti, Alessandro Genoni

Quantum Mechanics / Extremely Localized Molecular Orbital Embedding Strategy for Excited-States: Coupling to Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster

Giovanni Macetti[(1)], Alessandro Genoni[(1)*]

(1) Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France.

* Correspondence to:

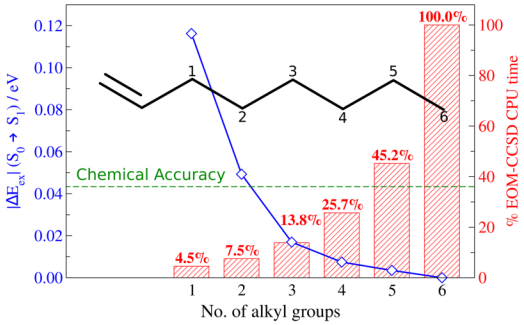
Alessandro Genoni, Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France. E-mail: alessandro.genoni@univ-lorraine.fr; Phone: +33 (0)3 72 74 91 70; Fax: +33 (0)3 72 74 91 87.

ABSTRACT

The QM/ELMO (quantum mechanics / extremely localized molecular orbital) method is a recently developed embedding technique in which the most important region of the system under exam is treated at fully quantum mechanical level, while the rest is described by means of transferred and frozen extremely localized molecular orbitals. In this paper, we propose the first application of the QM/ELMO approach to the investigation of excited-states and, in particular, we present the coupling of the QM/ELMO philosophy with Time-Dependent Density Functional Theory (TDDFT) and Equation-of-Motion Coupled Cluster with single and double substitutions (EOM-CCSD). The proposed TDDFT/ELMO and EOM-CCSD/ELMO strategies underwent a series of preliminary tests that were already considered for the validation of other embedding methods for excited-states. The obtained results showed that the novel techniques allow the accurate description of localized excitations in large systems by only including a relatively small number of atoms in the region treated at fully quantum chemical level. Furthermore, for TDDFT/ELMO, it was also observed that: i) the method enables to avoid the presence of artificial low-lying charge-transfer states that may affect traditional TDDFT calculations, even using functionals that do not take into account long-range corrections; ii) the novel approach can be also successfully exploited to investigate local electronic transitions in quite large systems (e.g., reduced model of the Green Fluorescent Protein) and the accuracy of the results can be improved by including a sufficient number of chemically crucial fragments/residues in the quantum mechanical region. Finally, concerning EOM-CCSD/ELMO, it was also seen that, despite the quite crude approximation of an embedding potential given by frozen extremely localized molecular orbitals, the new strategy is able to satisfactorily account for the effects of the environment. This work

paves the way to further extensions of the QM/ELMO philosophy for the study of local excitations in extended systems, suggesting the coupling of the QM/ELMO approach with other quantum chemical methods for excited-states, from the simplest Δ SCF techniques to the most advanced and computationally expensive multi-references methods.

TOC GRAPHICS



KEYWORDS: excited-states, embedding techniques, QM/QM' strategies, extremely localized molecular orbitals (ELMOs), Time-Dependent Density Functional Theory, Equation-of-Motion Coupled Cluster

1. INTRODUCTION

The accurate description of excited-states is of fundamental importance for a full comprehension of photoinduced processes both in chemical and biological systems. To accomplish this challenging task, quantum chemistry currently offers a large variety of methods and, among them, prominent places are occupied by Time-Dependent Density Functional Theory (TDDF) [1-3] and Equation-of-Motion Coupled Cluster (EOM-CC) [4-6]. The former is certainly the most widely used approach due to its advantageous trade-off between accuracy and computational cost. In fact, notwithstanding its known intrinsic limitations, the strategy is characterized by an M^3/M^4 scaling (with M as the number of basis functions used in the calculation), which allows the application of the technique to the study of excited-states of medium/large systems. The latter, and particularly Equation-of-Motion Coupled Cluster with single and double substitutions (EOM-CCSD), can be considered as the current gold standard method for excited-states,[7,8] although its unfavorable computational scaling prevents a direct application to systems larger than about fifty atoms.

In order to extend the range of applicability of both standard TDDFT and EOM-CCSD, different strategies have been introduced over the years. For instance, we can mention the linear-scaling approaches based on the localized nature of atomic and molecular orbitals,[9-21] the techniques exploiting efficient reductions of the virtual orbitals space,[22,23] or the couplings with fragmentation methods.[24,25] However, to really apply TDDFT and EOM-CCSD to very large systems, they have to be interfaced with embedding and/or multi-scale strategies. In this context, the simplest and most popular options still consist in the coupling with molecular mechanics in QM/MM[26-29] (quantum mechanics / molecular mechanics) computations or with

implicit solvent models, such as PCM (polarizable continuum model).[30-32] Finally, more interestingly and more importantly for the work that will be discussed in the present paper, it is worth noting that TDDFT and EOM-CCSD have been also interfaced with fully quantum mechanical embedding methods, such as the density matrix[33-36] and density functional embedding approaches[37-50].

In particular, TDDFT has been coupled with the Embedded Mean Field Theory (EMFT) [35], as suggested by Manby and Miller, who proposed the so-called Linear-Response Time-Dependent EMFT (LR-TD-EMFT) [51], or by Parkhill and coworkers, who actually exploited the block-orthogonalized version of EMFT and extended it to Real-Time Time-Dependent Density Functional Theory (RT-TDDFT)[52]. A larger number of developments have been actually proposed in the case of density functional embedding strategies. The first examples consisted in implementing TDDFT in the framework of the frozen density embedding (FDE) approach[38,39], thus giving rise to the so-called subsystem-Time-Dependent Density Functional Theory methods.[53-58] The projection-based embedding (PBE) technique[44,45] has also been taken into account. The first attempt in this direction was the one by Chulhai and Jensen,[59] followed by the works conducted by the Neugebauer group[60,61] and recently by Goodpaster and collaborators[62], who exploited their absolutely localized variant of PBE[48,49] to further reduce the TDDFT computational cost.

In analogous way, EOM-CCSD has been also recently interfaced with the projection-based embedding approach. The first work along this line is the one conducted by Bennie and coworkers,[63] who have practically extended the original version of PBE to EOM-CCSD and showed that their new embedding method for excited-states is capable of providing results in excellent agreement with those obtained from the corresponding full EOM-CCSD computations. Furthermore, as done for TDDFT,

Goodpaster and collaborators have shown that the absolutely localized version of the projection-based embedding strategy[48,49] can be exploited also to significantly reduce the computational cost of EOM-CCSD without impacting the accuracy of the results.[62]

In line with the strategies just mentioned above, in this paper we propose the first extensions of the recently developed multiscale embedding method QM/ELMO[64,65] (quantum mechanics / extremely localized molecular orbital) to quantum chemical strategies for excited-states. QM/ELMO is a technique that describes the most chemically relevant region of the system under exam by means of usual quantum chemistry approaches (e.g., Hartree-Fock (HF), Density Functional Theory (DFT), Møller-Plesset perturbation theory, Coupled Cluster, etc.) and the remaining part through frozen extremely localized molecular orbitals[66-68] (ELMOs) that are exported from suitable databanks[69-71] or tailor-made model molecules. In fact, being orbitals strictly localized on small molecular subunits (e.g., atoms, bonds or functional groups), ELMOs are reliably transferable from a molecule to another[69-76], provided that the chemical environments of the fragments on which they are localized are approximately the same in the two molecules. For this reason, libraries of ELMOs covering all the possible elementary subunits of the twenty natural amino acids have been already constructed.[71] They allowed not only instantaneous reconstructions of approximate wave functions and electron densities of large systems, but also fast and accurate refinements of polypeptide and protein crystal structures[77] in the framework of quantum crystallography[78-83] (for interested readers, more details about ELMOs theory, transferability and databanks are given in the Supporting Information).

Preliminary tests on the QM/ELMO approach have shown that transferred and frozen ELMOs provide completely reasonable and accurate embeddings for traditional

ground state quantum mechanical calculations on chemically active regions of extended systems. In fact, considering only a small number of atoms in the QM subunit, the QM/ELMO computations generally give results that differ by less than 1 kcal/mol from the corresponding fully QM ones.[65] In light of this, we have decided to start coupling the QM/ELMO approach with TDDFT and EOM-CCSD, thus giving rise to two new multi-scale embedding strategies for the treatment of localized excitations in large molecules/systems: the TDDFT/ELMO and EOM-CCSD/ELMO methods. Concerning the former, we wanted to prove that the new approach is able i) to describe local excited-states in large systems also when the partition between the QM and ELMO subunits occurs across covalent bonds, ii) to eliminate the typical spurious low-lying charge-transfer states of TDDFT, iii) to provide accurate results when sufficiently large parts of biological systems are considered as chemically active regions in the calculations. Pertaining to the latter, other than proving that the novel strategy is able to accurately and cheaply describe localized excitations in relatively large systems, the goal was also of assessing current capabilities and limitations of the developed technique in taking into account the effects of the environment.

The paper is organized as follows. In section 2, we will present the basic theoretical features of the QM/ELMO philosophy and we will discuss how the environment effects are or can be possibly taken into account in the EOM-CCSD/ELMO approach. In section 3, we will describe the validation tests that were carried out to evaluate the performances of the new techniques. The obtained results will be afterwards shown and discussed in sections 4 and 5, which will be followed by some general conclusions in section 6.

2. THEORY

2.1 The QM/ELMO SCF algorithm. In this section, we will review the philosophy of the QM/ELMO strategy, which is also at the basis of the TDDFT/ELMO and EOM-CCSD/ELMO variants discussed in the present paper.

As already mentioned in the Introduction, the QM/ELMO approach initially consists in subdividing the system under investigation into two regions: i) the QM subunit, which is the most chemically relevant part of the examined system and is treated with a traditional quantum chemistry method; ii) the ELMO subsystem, which represents the environment of the QM region and is described by means of transferred and frozen extremely localized molecular orbitals.

After the transfer of the necessary ELMOs to the ELMO subunit, the QM/ELMO procedure consists i) in the preliminary orthogonalization of the transferred extremely localized molecular orbitals and of the basis functions of the QM region, and ii) in the actual QM/ELMO self-consistent field (SCF) cycle.

The preliminary orthogonalization entails the following three steps:

1. Löwdin orthonormalization of the transferred ELMOs;
2. orthogonalization of the QM basis functions against the Löwdin orthonormalized ELMOs;
3. canonical orthogonalization of the QM basis functions resulting from the previous step.

The procedure above, which is practically negligible in terms of computational cost (see Supporting Information for more details), can be summarized through the following transformation:

$$\boldsymbol{\chi}' = \boldsymbol{\chi}\,\mathbf{B} \quad (1)$$

with $\boldsymbol{\chi} = [|\chi_1\rangle, |\chi_2\rangle, \dots, |\chi_M\rangle]$ as the starting $1 \times M$ array of the M non-orthogonal basis functions for the whole system (which includes the QM and ELMO regions), $\boldsymbol{\chi}' = [|\chi'_1\rangle, |\chi'_2\rangle, \dots, |\chi'_{M_{QM}}\rangle]$ as the final $1 \times M_{QM}$ array of the M_{QM} orthonormal basis functions for the QM subunit only (with $M_{QM} \ll M$), and **B** as an $M \times M_{QM}$ transformation matrix, which is crucial in the QM/ELMO SCF algorithm that will be described below (interested readers can find more details on the preliminary orthogonalization steps in the Supporting Information of this work or in the original papers of the QM/ELMO strategy[64,65]).

Afterwards, the QM/ELMO SCF algorithm starts with the construction of the Fock matrix **F** in the original basis $\boldsymbol{\chi}$, whose elements are given by the following expression:

$$
\begin{aligned}
F_{\mu\nu} = {} & \langle \chi_\mu | \hat{h}^{core} | \chi_\nu \rangle \\
& + \sum_{\lambda,\sigma=1}^{M} P_{\lambda\sigma}^{QM} \left[(\chi_\mu \chi_\nu | \chi_\sigma \chi_\lambda) - \frac{1}{2} x \, (\chi_\mu \chi_\lambda | \chi_\sigma \chi_\nu) \right] \\
& + \sum_{\lambda,\sigma \in ELMO} P_{\lambda\sigma}^{ELMO} \left[(\chi_\mu \chi_\nu | \chi_\sigma \chi_\lambda) - \frac{1}{2} x \, (\chi_\mu \chi_\lambda | \chi_\sigma \chi_\nu) \right] \\
& + \langle \chi_\mu | \hat{v}^{XC} [\mathbf{P}^{\mathrm{QM}} + \mathbf{P}^{ELMO}] | \chi_\nu \rangle = \\
= {} & h_{\mu\nu} + F_{\mu\nu}^{QM} + F_{\mu\nu}^{ELMO} + v_{\mu\nu}^{XC} \qquad (2)
\end{aligned}
$$

where $\hat{h}^{core}$ is the usual one-electron Hamiltonian operator, $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$ are the QM and ELMO one-particle density matrices in the starting basis-set $\boldsymbol{\chi}$, respectively, $\langle \chi_\mu | \hat{v}^{XC} [\mathbf{P}^{QM} + \mathbf{P}^{ELMO}] | \chi_\nu \rangle$ is a generic element of the exchange-correlation potential matrix (which disappears in the simple Hartree-Fock case), and x is the fraction of exact exchange used in the calculation (and which is equal to 1 when HF/ELMO computations are performed). In equation (2) it is worth pointing out that the only two terms that are updated during the SCF cycle are $F_{\mu\nu}^{QM}$ and $v_{\mu\nu}^{XC}$, while $h_{\mu\nu}$

and, above all, $F_{\mu\nu}^{ELMO}$ remain constant throughout all the iterations. However, despite the previous observation, it is also fair to note that, at present, the matrix **F** is still evaluated over the whole basis-set of the investigated system (supermolecular basis-set), which is the current algorithmic limitation of the QM/ELMO strategies. This has a larger impact on QM/ELMO methods for ground and excited-states where the Fock matrix evaluation has a non-negligible weight in the calculations (e.g., in TDDFT/ELMO calculations), while it is less important in QM/ELMO computations where the QM region is treated through a more highly correlated approach and the post-SCF procedure is predominant (e.g., in EOM-CCSD/ELMO calculations). To overcome this drawback and thus really improve the scalability of techniques as the TDDFT/ELMO approach, we have already planned to introduce a suitable criterion to truncate the basis functions space over which to initially evaluate the Fock matrix, for instance by following one the procedures proposed by Manby, Miller and coworkers to solve a completely similar problem in their PBE approach.[45]

At the second step, the Fock matrix is transformed to the basis of the orthonormal functions $\boldsymbol{\chi}'$ by means of the following relation:

$$\mathbf{F}' = \mathbf{B}^{\dagger}\mathbf{F}\,\mathbf{B} \quad (3)$$

where $\mathbf{B}^{\dagger}$ is the transpose of matrix **B** discussed for equation (1). The obtained $M_{QM} \times M_{QM}$ matrix $\mathbf{F}'$ is then diagonalized:

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\,\mathbf{E}' \quad (4)$$

with $\mathbf{C}'$ as the matrix containing the coefficients of the (occupied and virtual) molecular orbitals of the QM region in the basis $\boldsymbol{\chi}'$. If the QM subsystem is described at DFT level, the above-mentioned molecular orbitals correspond to Kohn-Sham orbitals.

Afterwards the obtained molecular orbitals are transformed back to the original basis $\boldsymbol{\chi}$ always exploiting the transformation matrix $\mathbf{B}$:

$$\mathbf{C} = \mathbf{B}\,\mathbf{C}' \quad (5)$$

Therefore, we obtain the new matrix $\mathbf{C}$ of the molecular orbitals coefficients, which are used to determine the one-particle density matrix for the QM region (which appears in the second term of the right-hand side of equation (2)):

$$P_{\lambda\sigma}^{QM} = 2\sum_{i=1}^{N} C_{\sigma i}^{*}\, C_{\lambda i} \quad (6)$$

For the sake of completeness, the one-particle density matrix of the ELMO region in the basis $\boldsymbol{\chi}$ (which appears in the third term of the right-hand side of equation (2)) is given by the following expression:

$$P_{\lambda\sigma}^{ELMO} = 2\sum_{i=1}^{N_{ELMO}} C_{\sigma i}^{\perp *}\, C_{\lambda i}^{\perp} \quad (7)$$

where $\{C_{\mu i}^{\perp}\}$ are the coefficients of the Löwdin orthonormalized extremely localized molecular orbitals of the ELMO subsystem. As already stressed above, this quantity is calculated only once before the SCF cycle starts and remains constant throughout all the iterations.

After the computation of the QM density matrix according to equation (6), convergence is inspected. If it is reached, the cycle halts, otherwise a new iteration starts and the one-particle density matrix for the QM region is used to update the Fock matrix $\mathbf{F}$ in the original non-orthogonal basis $\boldsymbol{\chi}$ according to equation (2). The occupied and virtual molecular orbitals resulting from the above-described cycle can be afterwards exploited for subsequent post-HF/ELMO[65], TDDFT/ELMO or EOM-CCSD/ELMO computations.

It is worth noting that, since $M_{QM} \ll M$ (we remind that M is the number of basis functions for the whole system), the QM/ELMO methods are characterized by a significant reduction of the computational cost, especially if one deals with very large systems and if post-HF techniques are used to treat the QM region. In fact, assuming to work with a *2N*-electron closed-shell QM subsystem, the diagonalization of Fock matrix $\mathbf{F}'$ (see equation (4)) in the above-discussed SCF cycle provides N doubly occupied molecular orbitals (with N generally much lower than $N + N_{ELMO}$) and $M_{QM} - N$ virtual molecular orbitals (with $M_{QM} - N$ always much lower than $M - N$). Since the computational cost of correlated calculations for ground and excited-states depends on the number of occupied molecular orbitals and, even more importantly, on the number of the virtual ones (e.g., CCSD(T) and EOM-CCSD calculations scale as o^3v^4 and o^2v^4, respectively, with o and v as the number of occupied and virtual molecular orbitals), it is clear that the reduced dimensions of matrix $\mathbf{F}'$ entail important savings in terms of CPU time, as already shown by means of preliminary validation tests on the QM/ELMO strategy.[65] This feature of the QM/ELMO approach is common to other fully quantum mechanical embedding methods, such as the above-mentioned absolutely localized PBE strategy introduced by the Goodpaster group[48,49] or other recent techniques proposed by Hammes-Schiffer *et al.*[46,47] and by Claudino and Mayhall[50]. Moreover, as one should expect, this characteristic is also very favorable in order to significantly reduce the computational cost of EOM-CCSD calculations, as we will show below and as it was also stressed for the absolutely localized PBE strategy for excited-states[62].

The QM/ELMO SCF algorithm, the post-HF/ELMO strategies and all the QM/ELMO techniques for excited-states presented in this work have been implemented by

modifying the original subroutines of the corresponding fully quantum mechanical methods in the quantum chemistry package *Gaussian09*[84].

2.2 Environment effects. In this subsection we will briefly discuss how the effects of the environment are or can be possibly taken into account in the EOM-CCSD/ELMO calculations of excited-states.

First of all, following Wen *et al*.,[62] the environment effects can be distinguished in ground state polarization and polarization response effects. In our case, the ground state polarization effects are simply given by the embedding potential provided by transferred extremely localized molecular orbitals. Compared to the projection-based embedding approaches for excited-states,[62] our treatment of the ground state polarization is only approximate and not flexible. In fact, while in the PBE-based techniques the embedding potentials for the excited-states are generally constructed from the converged DFT ground state electron densities of the whole system or of its subunits, in our EOM-CCSD/ELMO strategy the embedding potential is represented by transferred ELMOs that remain frozen throughout the preliminary HF/ELMO calculation and that, therefore, are not influenced by the actual ground state electron density of the QM region (it is worth reminding that HF/ELMO or DFT/ELMO calculations are always preliminary steps to carry out QM/ELMO calculations for excited-states). The advantage of our technique is that the transfer of extremely localized molecular orbitals to the environment region is practically instantaneous and it is not necessary to perform a preliminary HF or DFT calculation on the whole molecule under exam. In the future, a possible way to increase the flexibility of the ELMO-based methods might consist in developing polarizable QM/ELMO strategies in which transferred virtual extremely localized molecular orbitals might be used to relax the electronic structure / electron density of the environment (i.e., the ELMO

region) as a response to the influence of the ground state electron distribution of the chemically active subsystem.

Concerning the polarization response of the environment, in analogy with the absolutely localized version of the PBE approaches for excited-states,[62] also the EOM-CCSD/ELMO technique does not intrinsically take into account this effect. There are three possibilities to introduce it: i) including more atoms or molecular orbitals in the active QM region; ii) exploiting a state-averaged approach, as proposed by the Carter[85] and Goodpaster[62] groups; iii) adding a suitable TDDFT correction. As it will be shown below, the first and the third options are those that were tested in the present work.

The first possibility is the most straightforward, but potentially also the most expensive from the computational point of view. In our case, this option consists in properly selecting a set of originally frozen ELMOs, removing them from the ELMO region and treating the associated electrons in the QM subsystem explicitly. This is also the strategy that has been adopted when the conventional projection-based embedding technique[44,45] (not its absolutely localized variant[48,49]) was coupled with wavefunction-based methods for excited-states. In that case, polarization response has been quite easily taken into account by including a certain number of relevant occupied localized molecular orbitals (and consequently the corresponding electrons) of the environment region in the QM subsystem.[63]

The second possibility would consist in developing a self-consistent strategy to relax the ELMO electron density by taking into account the ground state and excited-states electron distributions of the QM region. As indicated above for the ground state polarization, this task might be accomplished in the future by developing a technique that exploits transferred virtual ELMOs.

Finally, following Wen *et al.*,[62] the third option is equivalent to write the excitation energy as follows:

$$\tilde{\omega}_{\mathrm{EOM-CCSD/ELMO}} = \omega_{\mathrm{EOM-CCSD/ELMO}} + \omega_{\mathrm{TDDFT}} - \omega_{\mathrm{TDDFT/ELMO}} \quad (8)$$

where $\tilde{\omega}_{\mathrm{EOM-CCSD/ELMO}}$ is the embedded EOM-CCSD/ELMO excitation energy after the TDDFT correction, $\omega_{\mathrm{EOM-CCSD/ELMO}}$ is the embedded EOM-CCSD/ELMO excitation energy before the TDDFT correction (including only the approximate ground state polarization), ω_{TDDFT} is the traditional TDDFT excitation energy on the full system, and $\omega_{\mathrm{TDDFT/ELMO}}$ is the embedded TDDFT excitation energy obtained through the TDDFT/ELMO strategy. Of course, from the computational perspective, this option is convenient only if the cost of the full TDDFT computation is small compared to the one associated with the corresponding full EOM-CCSD calculation.

An investigation on the capabilities of the new EOM-CCSD/ELMO method in accounting for the environment effects will be described and discussed in subsections 3.2.3 and 5.2.

3. COMPUTATIONAL DETAILS

In order to start testing their general capabilities, the new TDDFT/ELMO and EOM-CCSD/ELMO methods underwent a series of validation tests that were previously conceived and carried out to evaluate the performances of other embedding TDDFT and EOM-CCSD approaches, particularly the Linear-Response Time-Dependent Embedded Mean Field Theory[51] and the TDDFT and EOM-CCSD techniques coupled with the absolutely localized version of the projection-based embedding strategy[62].

Except for particular cases (e.g., calculations of ELMOs; see subsection 3.3), all the computations that will be described below were performed using the *Gaussian09*

package,[84] both in its standard version and in a modified variant where the TDDFT/ELMO and EOM-CCSD/ELMO methods have been implemented.

3.1 TDDFT/ELMO validation tests. In this subsection, we will describe the computational details of the test calculations that were performed to assess the performances of the TDDFT/ELMO approach. In particular, we will describe the tests carried out on long-chain hydrocarbons (subsection 3.1.1), solvated acrylamide (subsection 3.1.2) and a model of the Green Fluorescent Protein (GFP; subsection 3.1.3).

3.1.1 Calculations on long-chain hydrocarbons. We considered two long-chain hydrocarbons with different terminal functional groups: 1-decanal and 1-nonylbenzene (see top panel of Figure S3 in the Supporting Information). The calculations performed on these molecules allowed us to evaluate the performances of the proposed method when the frontier between the QM and ELMO regions is represented by a covalent bond.

The geometries of the molecules under exam were initially optimized at B3LYP level with basis-set cc-pVDZ. The obtained geometries were afterwards used to perform regular TDDFT calculations with two different functionals (CAM-B3LYP and B3LYP) and two different basis-sets (aug-cc-pVDZ and cc-pVDZ). The results of the full TDDFT computations carried out with the combinations of the above-mentioned functionals and basis-sets were used as reference values for the performed TDDFT/ELMO computations. For 1-decanal we considered the excitation energies and the oscillator strengths corresponding to the $n \rightarrow \pi^*$ excited-state. For 1-nonylbenzene we took into account the excitation energies and the oscillator strengths associated with two low-lying excited-states dominated by $\pi \rightarrow \pi^*$ transitions,

namely, the $^1B_{2u}$-like excited-state (mainly characterized by the HOMO → LUMO and HOMO-1 → LUMO+1 excitations) and the $^1B_{1u}$-like excited-state (mainly characterized by the HOMO → LUMO+1 and HOMO-1→LUMO excitations).

Concerning the TDDFT/ELMO calculations, for each combination of functional and basis-set, we gradually increased the size of the QM region in order to assess the effects of the embedding provided by the ELMOs in the description of the localized excitations. In particular, for both 1-decanal and 1-nonylbenzene, in the QM region we considered from one to eight CH_2 alkyl moieties of the hydrocarbon chain along with the terminal functional group (aldehyde group and aromatic ring for 1-decanal and 1-nonylbenzene, respectively).

3.1.2 Computations on solvated acrylamide. One of the well-known shortcomings of Time-Dependent Density Functional Theory is that spurious low-lying charge-transfer states are sometimes obtained when local and semi-local exchange-correlation functionals are employed. The most common way to partially circumvent this problem consists in exploiting long-range corrected functionals (e.g., CAM-B3LYP, LRC-ωPBEh, LRC-ωPBE and ωB97XD). Another possible solution to the drawback is represented by the use of embedding TDDFT methods. In fact, since in these techniques the electron density of each subsystem is expanded only on basis functions of a local basis-set, low-lying charge-transfer excitations between far subunits that interact weekly are excluded *a priori*. Goodpaster and coworkers recently proved this for their PBE-TDDFT approach for excited-states,[62] which is particularly suitable to solve the above-mentioned problem due the absolute localization of the electron density on each subsystem. For the sake of completeness, it is worth noting that, as side effect, these strategies unfortunately lead to also neglect global and real charge-transfer excited-states.

The TDDFT/ELMO approach proposed in this work also provides electron densities distinctly localized on the regions in which the investigated system is preliminarily subdivided. For this reason, we decided to prove whether the new embedding TDDFT strategy is also capable of avoiding spurious low-lying charge-transfer states. To accomplish this task, we submitted our method to a validation test similar to the one proposed by Goodpaster *et al.* for their PBE-TDDFT technique.[62]

We considered acrylamide solvated by water molecules. We initially performed a Molecular Dynamics simulation (see Supporting Information for details) and, from the production phase, we extracted a frame characterized by the presence of two hydrogen bond contacts between the carbonyl group of acrylamide and the solvent molecules (see Figure S4 in the Supporting Information). Considering this frame, we carried out traditional TDDFT calculations at B3LYP/aug-cc-pVDZ and CAM-B3LYP/aug-cc-pVDZ levels by increasing the number of water molecules surrounding acrylamide from 0 to 25. Corresponding TDDFT/ELMO computations were afterwards performed exploiting functional B3LYP and the aug-cc-pVDZ basis-set. The QM region always consisted only of acrylamide, while the ELMO subsystem was gradually expanded from 5 to 25 water molecules. For the sake of precision, for both the traditional TDDFT and TDFFT/ELMO calculations, we added 5 water molecules at each step according to their distance from the barycenter of the solute. From all the computations we extracted and compared the excitation energies associated with the dark $n \rightarrow \pi^*$ and the bright $\pi \rightarrow \pi^*$ excited-states. It is worth noting that, in the gas-phase, the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions correspond to the first and second excited-states, respectively. Therefore, for each TDDFT and TDDT/ELMO calculation including water molecules, we also took note of the number of states that were actually computed before obtaining the bright $\pi \rightarrow \pi^*$ transition.

This is a useful indication to evaluate the capability of the considered methods in avoiding artificial low-lying charge-transfer states. Furthermore, for all the performed TDDFT and TDDFT/ELMO calculations, we also computed and analyzed the natural transition orbitals (NTOs) associated with the different excited-states, which also helped us to identify the desired $\pi \rightarrow \pi^*$ transition and to evaluate the extent of its delocalization in the different cases.

3.1.3 Application to the Green Fluorescent Protein. To conclude our validations of the TDDFT/ELMO approach, we finally decided to test the technique on a 161-atom model for the A-form of the Green Fluorescent Protein (GFP), which was previously extracted from a crystal structure (PDB code: 1EMB) by Kaila *et al*.[86] and which was also used by Goodpaster and coworkers[62] to validate their PBE-based TDDFT method for excited-states. This model comprises the chromophore molecule (p-hydroxybenzylidene-imidazolinone, pHBDI), the side-chains (cut at the carbon Cβ) of nine residues of GFP (Thr62, Gln69, Gln94, Arg96, His148, Val150, Thr203, Ser205 and Glu222), other than four water molecules observed from the crystallographic experiment (see Figure S5 in the Supporting Information). Using this 161-atom model, we performed benchmark TDDFT calculations with functional B3LYP and basis-sets 6-311G(d,p) and cc-pVDZ. From these computations, we have particularly considered the brightest low-lying excited-state corresponding to excitation energies of 3.14 eV and 3.15 eV for basis-sets 6-311G(d,p) and cc-pVDZ, respectively, which are in good agreement with the experimental values of 3.12-3.14 eV.[87-89]

Afterwards, following Goodpaster and coworkers,[62] we performed embedded TDDFT/ELMO computations on the 161-atom model of the Green Fluorescent Protein by gradually increasing the chemically active region (i.e., the chromophore pHBDI) in two different ways: i) sequentially including the H_2O, Ser205, Arg96,

Glu222 and $3H_2O$ subunits; ii) sequentially adding the moieties of case (i), but in reverse order. Moreover, we also carried out traditional TDDFT calculations without embedding, only on the QM subsystems defined for the TDDFT/ELMO computations; in other words, we performed traditional TDDFT calculations without embedding on further reduced GFP models. Also all of these computations were performed at B3LYP/6-311G(d,p) and B3LYP/cc-pVDZ levels and the excitations of interest were identified by inspecting the oscillator strengths and the obtained NTOs.

3.2 EOM-CCSD/ELMO validation tests. This subsection is dedicated to the computational details for the validation tests on the EOM-CCSD/ELMO approach. We will particularly describe the test calculations that were performed to assess the performances of the new approach when QM and ELMO regions are separated by a covalent and non-covalent frontier (subsections 3.2.1 and 3.2.2, respectively). Finally, in subsection 3.2.3, we will describe the test computations that were carried out to evaluate the capabilities of the EOM-CCSD/ELMO technique in accounting for the environment effects. If not differently specified, in this work, all the EOM-CCSD and EOM-CCSD/ELMO computations were performed with basis-set aug-cc-pVDZ.

3.2.1 EOM-CCSD/ELMO calculations with a covalent frontier. As first standard validation test, we decided to assess the convergence of EOM-CCSD/ELMO calculations as a function of the QM region size when the frontier between the QM and the ELMO subsystems coincides with a covalent bond. Also in this case, to accomplish this task, we considered two molecules characterized by a relatively long hydrocarbon chain: 1-octene and octanoic acid (see low panel of Figure S3 in the Supporting Information), whose geometries were preliminarily optimized at B3LYP/cc-pVDZ level. The geometries were subsequently exploited to carry out full

EOM-CCSD computations. For each molecule, the first three transitions were taken into account and the excitation energies and the oscillator strengths resulting from the full EOM-CCSD calculations were afterwards used as benchmark values.

Pertaining to the EOM-CCSD/ELMO computations, we gradually increased the size of the QM region by including the alkyl groups of the hydrocarbon chain together with the terminal functional group of the considered molecules (i.e., carboxylic group for octanoic acid and carbon-carbon double bond for 1-octene). Also for these calculations we took into account excitation energies and oscillator strengths associated with the first three excitations.

3.2.2 EOM-CCSD/ELMO calculations with a non-covalent frontier. As second step, we decided to investigate the convergence of embedded EOM-CCSD/ELMO computations in cases of non-covalent frontiers between the QM and ELMO regions. To this purpose we investigated the case of four different molecules solvated by water: formaldehyde, acetaldehyde, acrolein and acrylamide.

For each of the above-mentioned systems, we initially carried out a preliminary Molecular Dynamics (MD) simulation (see Supporting Information for more details), from which we extracted a frame where the solute establishes two plausible hydrogen bond contacts with the surrounding water molecules. Considering that frame, we kept only the solute and the six closest solvent molecules (see Figure S6 in the Supporting Information). The resulting systems were then used to perform benchmark full EOM-CCSD calculations. In analogy with the test calculations performed on their methods by Bennie *et al.*[63] and by Goodpaster and coworkers[62], also in our case we focused only on the $n \rightarrow \pi^*$ transition (first excited-state).

Concerning the EOM-CCSD/ELMO computations, we gradually enlarged the QM region by including the surrounding solvent subunits: in the cheapest EOM-

CCSD/ELMO calculations, only the solute was included in the QM subsystem; we afterwards performed EOM-CCSD/ELMO computations by adding to the QM region only the two water molecules establishing hydrogen bond contacts with the solute; finally, we sequentially included the remaining water molecules by considering their distance from the oxygen atom of the carbonyl group of the solute (formaldehyde, acetaldehyde, acrolein or acrylamide) taken into account. The excitation energies and the oscillator strengths for the $n \rightarrow \pi^*$ excitation resulting from the EOM-CCSD/ELMO calculations were then compared to the corresponding benchmark EOM-CCSD values.

3.2.3 Environment effects. In order to assess the capabilities of the new embedding method in accounting for the effects of the environment, we considered two of the molecules taken into account in the previous subsection: formaldehyde and acrylamide. For both of them, from the same MD simulation-frame used for the test calculations described above, we extracted the geometry of the system consisting of the solute and of the two water molecules involved in hydrogen bond contacts. We then performed the following calculations: i) traditional EOM-CCSD computations on the full system, which account for ground state polarization and polarization response; ii) traditional EOM-CCSD calculation on the solute molecule only, which does not account for any environment effect; iii) EOM-CCSD/ELMO computation on the full system with only the solute molecule in the QM region, which should include only an approximate ground state polarization; iv) EOM-CCSD/ELMO calculation as the previous one, but with the additional inclusion of properly selected ELMOs/electrons in the QM region (see subsection 5.2 for more details) to partially account for the polarization response; v) TDDFT and TDDFT/ELMO (with only the solute molecule in the QM region) computations on the full system to determine the TDDFT

corrections for the polarization response according to equation (8). Again, we considered only the excitation energies obtained for the $n \rightarrow \pi^*$ transition.

Finally, to conclude the investigation on the effects due to the environment, for both formaldehyde and acrylamide, we monitored the variation of the $n \rightarrow \pi^*$ excitation energy as a function of the number of surrounding water molecules when the following type calculations are carried out: i) EOM-CCSD(0)/ELMO, namely an ELMO-embedded EOM-CCSD computation without water molecules in the QM region; ii) EOM-CCSD(2)/ELMO, namely an ELMO-embedded EOM-CCSD computation with two water molecules in the QM subsystem; iii) traditional TDDFT (CAM-B3LYP functional); iv) traditional Time-Dependent Hartree-Fock (TDHF). All the calculations were performed with basis-set aug-cc-pVDZ and the global number of water molecules was gradually increased from 2 to 30 taking into account their distance from the oxygen atom of the carbonyl group of the solute and always using the MD simulation-frames considered for the benchmark tests described in subsection 3.2.2.

3.3 Calculation of ELMOs. The extremely localized molecular orbitals employed in the TDDFT/ELMO and EOM-CCSD/ELMO computations described in the previous subsections were previously obtained from the available ELMO libraries[71] or by means of calculations performed on suitable model systems through a modified version of the GAMESS-UK[90] package that implements the Stoll equations[66] (see Supporting Information for more details about the ELMO theory).

For the TDDFT/ELMO and EOM-CCSD/ELMO computations on 1-decanal, 1-nonylbenzene, octanoic acid and 1-octene, the ELMOs describing the alkyl groups were calculated on the butane molecule using a geometry optimized at B3LYP/cc-pVDZ level. The extremely localized molecular orbitals used in the TDDFT/ELMO

calculations on 1-decanal and 1-nonylbenzene were determined with basis-sets aug-cc-pVDZ and cc-pVDZ, while the ones employed in the EOM-CCSD/ELMO computations on octanoic acid and 1-octene were calculated only with the aug-cc-pVDZ set of basis functions.

For all the validation tests on the solvated molecules, the only necessary extremely molecular orbitals were those localized on the surrounding water molecules and were computed with basis-set aug-cc-pVDZ on a geometry always optimized at B3LYP/cc-pVDZ level.

Finally, for the TDDFT/ELMO calculations on the 161-atom model of the Green Fluorescent Protein, the ELMOs describing residues and water molecules belonging to the ELMO regions were those stored in the recently constructed ELMO databanks (6-311G(d,p) and cc-pVDZ basis-sets).[71]

The transfers of the pre-calculated ELMOs to the geometries of the investigated systems were carried out by exploiting the *ELMOdb* program[71], namely the software that is associated with the ELMO libraries and that implements the strategy originally devised by Philipp and Friesner[91] for the rotation of strictly localized bond orbitals (see Supporting Information for more details).

4. TDDFT/ELMO: RESULTS AND DISCUSSION

In this section we will show and discuss the results of the validation tests carried out to evaluate the performances of the TDDFT/ELMO approach. At first we will focus on the test calculations on 1-decanal and 1-nonylbenzene (Section 4.1) to assess the convergence of the new computational strategy when the frontier between the QM and ELMO regions corresponds to a covalent bond. We will afterwards analyze the results of the computations that mainly aimed at evaluating the capability of the new

technique in avoiding low-lying charge-transfer states (Section 4.2). Finally, we will comment on the application of the TDDFT/ELMO method on the 161-atom model of the Green Fluorescent Protein (Section 4.3).

4.1 Convergence of the TDDFT/ELMO calculations. Let us initially consider the results obtained for the $n \rightarrow \pi^*$ excitation energy of 1-decanal (see Figure 1A). Regardless of the chosen exchange-correlation functional (B3LYP or CAM-B3LYP) or the adopted basis-set (cc-pVDZ and aug-cc-pVDZ), the TDDFT/ELMO values always agree with the corresponding TDDFT ones within the limit of chemical accuracy (0.043 eV), also when only one alkyl group is included in the QM subunit. Moreover, except for the CAM-B3LYP/cc-pVDZ case, which is characterized by a monotonically decreasing trend, the absolute discrepancies between the TDDFT/ELMO and TDDFT excitation energies generally present a peak (always lower than 0.043 eV) when two CH_2 moieties are included in the chemically active region and then gradually reduce as more and more atoms are treated at fully quantum mechanical level. It is worth noting that, when only three alkyl groups belong to the QM subsystem, all the computed TDDFT/ELMO excitation energies differ from the corresponding benchmark TDDFT values by less than 0.015 eV. We also observe that the discrepancies obtained through the CAM-B3LYP functional are generally lower than those resulting from the computations performed with the B3LYP functional. Finally, pertaining to the oscillator strength of the $n \rightarrow \pi^*$ excited-state (see Figure 1B), we see that the TDDFT/ELMO results clearly converge towards the TDDFT reference values for all the functional/basis-set combinations taken into account. In particular, it is easy to notice that, in all the cases, a *plateau* is practically achieved when four alkyl groups are included in the quantum mechanical region for the TDDFT/ELMO computations.

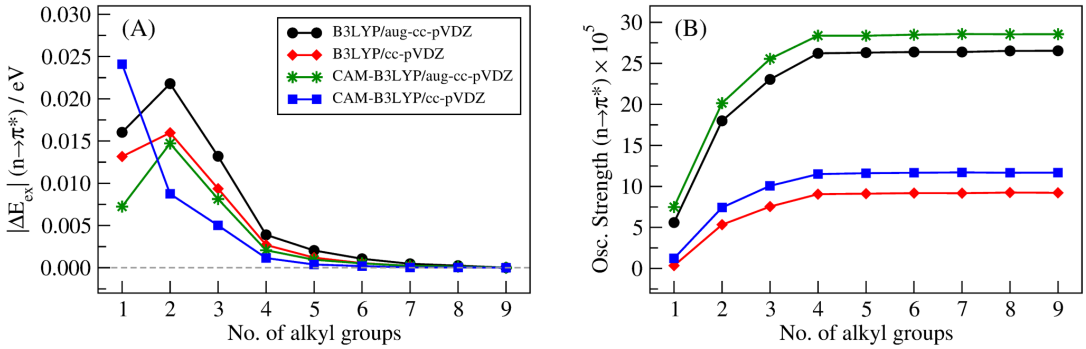


Figure 1. Results of the full TDDFT and TDDFT/ELMO calculations on 1-decanal: (A) absolute discrepancies between the $n \rightarrow \pi^*$ excitation energies obtained at TDDFT and TDDFT/ELMO levels; (B) $n \rightarrow \pi^*$ oscillator strengths. For the oscillator strengths, the reference TDDFT values are those obtained for nine alkyl groups in the QM region.

Quite similar results were also obtained for the $^1B_{2u}$- and $^1B_{1u}$-like excited-states of 1-nonylbenzene (see Figure S7 in the Supporting Information), with the only difference that, in these cases, it was almost always necessary to include at least two alkyl groups in the chemically active region in order to obtain discrepancies between the TDDFT/ELMO and TDDFT excitation energies lower than 0.043 eV. Convergence was also observed for the oscillator strengths, although it is sometimes slower than the one noticed for the excitation energies.

4.2 Elimination of spurious low-lying charge-transfer states. In Figure 2A, we have reported the $n \rightarrow \pi^*$ excitation energies obtained at B3LYP, CAM-B3LYP and B3LYP/ELMO levels for acrylamide as a function of the number of surrounding solvent water molecules (values are also given in Table S1 of the Supporting Information). We notice that, except for the case of 25 water molecules, the B3LYP and B3LYP/ELMO calculations provided very similar results that always differ by less than the chemical accuracy limit of 0.043 eV. Traditional CAM-B3LYP calculations gave higher values for the excitation energies, which is due to the larger contribution of Hartree-Fock exchange at long distance. Also for the excitation

energies for the bright $\pi \rightarrow \pi^*$ transition (see Figure 2B and Table S2 in the Supporting Information), the traditional TDDFT and TDDFT/ELMO computations with functional B3LYP gave excitation energies that are in very good agreement. In fact, with the only exception of the case for 15 water molecules, the absolute differences between the B3LYP and B3LYP/ELMO values are always lower than 1 kcal/mol. The TDDFT calculations with functional CAM-B3LYP always predicted higher excitation energies, which is consistent with what was also observed for the dark $n \rightarrow \pi^*$ transition. Since in the ELMO-embedded B3LYP computations only the solute was included in the QM region, the results obtained for the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ excitation energies further confirm the capability of the proposed TDDFT/ELMO approach in reproducing the results of traditional TDDFT computations by only treating a limited number of atoms at a fully quantum chemical level. For the sake of completeness, it is also worth noting that, for all the levels of theory taken into account, the $n \rightarrow \pi^*$ and the $\pi \rightarrow \pi^*$ excitation energies practically reach a *plateau* value when at least ten water molecules are taken into account. This is due to the fact that the two water molecules establishing hydrogen-bond contacts with the carbonyl group of acrylamide are included in the calculations only after considering those ten surrounding solvent molecules (see also Figure S4 in the Supporting Information; we remind that, in this case, the water molecules are gradually included in the calculations considering their increasing distance from the barycenter of the solute).

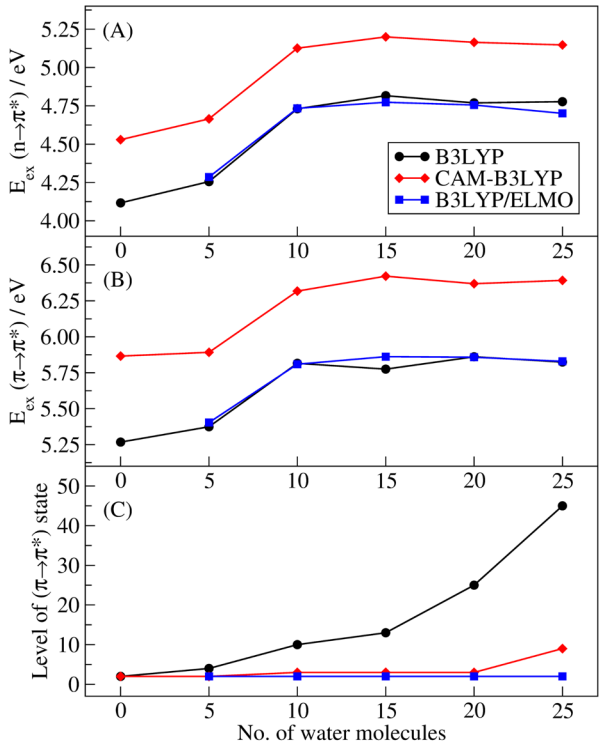


Figure 2. Results of full TDDFT (B3LYP and CAM-B3LYP functionals) and TDDFT/ELMO (B3LYP functional) calculations on solvated acrylamide as a function of the number of surrounding water molecules: (A) $n \rightarrow \pi^*$ excitation energies; (B) $\pi \rightarrow \pi^*$ excitation energies; (C) level of the $\pi \rightarrow \pi^*$ excited-states.

Let us now focus on the number of excited-states that we had to compute in order to obtain the bright $\pi \rightarrow \pi^*$ transition (see Figure 2C). As already anticipated in the section dedicated to the computational details, when traditional TDDFT calculations are carried out on the isolated acrylamide molecule (gas-phase computation), the bright $\pi \rightarrow \pi^*$ transition corresponds to the second excited-state, regardless of the exchange-correlation functional that is used. Nevertheless, as the number of surrounding water molecules increases, we can observe that the presence of spurious low-lying charge-transfer states between acrylamide and the solvent molecules entails that more and more intermediate transitions are obtained through traditional TDDFT calculations before identifying the desired $\pi \rightarrow \pi^*$ excited-state. This problem particularly affects the traditional TDDFT computations with functional B3LYP, for

which, for instance, the $\pi \rightarrow \pi^*$ transition corresponds to the 45-th excited-state when twenty-five water molecules surround the solute. The situation improves when functional CAM-B3LYP is adopted, with the $\pi \rightarrow \pi^*$ transition occurring at the 9-th excited-state when twenty-five water molecules are taken into account. On the contrary, the problem is completely avoided when the new TDDFT/ELMO approach is applied, although we used the B3LYP functional, which is not a long-range corrected one. In fact, inspecting Figure 2C, it is possible to see that, irrespective of the number of water molecules, the $\pi \rightarrow \pi^*$ transition resulting from the embedded B3LYP/ELMO computations always corresponds to the second excited-state. This clearly proves the capability of the novel TDDFT/ELMO approach in eliminating spurious low-lying charge-transfer states of Time-Dependent Density Functional Theory. A further evidence of this feature is also given in Figure S8 of the Supporting Information, where, for the B3LYP and B3LYP/ELMO calculations, we reported the number of excited-states with excitation energy lower than 6.0 eV. We can notice that, for the full TDDFT computations, the number of these excited-states drastically increases as a function of the surrounding water molecules, while it remains quite low for the TDDFT/ELMO calculations.

To better characterize the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$ transitions discussed above, we also analyzed the associated natural transition orbitals (see Figures S9-S11 in the Supporting Information). From this analysis, it emerges that the $n \rightarrow \pi^*$ excited-state is quite localized in all the considered cases, independently of the chosen functional or the use of the ELMO embedding. On the contrary, for the $\pi \rightarrow \pi^*$ transition, the NTOs show quite large delocalizations for the traditional TDDFT calculations with the B3LYP functional, a smaller delocalized character when functional CAM-B3LYP is used, and a completely localized nature for the TDDFT/ELMO computations at

B3LYP level. This is consistent with the previous observation that the novel embedding TDDFT/ELMO approach allows the elimination of artificial low-lying charge-transfer states when a solute is surrounded by a large number of solvent molecules.

As already mentioned, the test calculations on solvated acrylamide described in this subsection are analogous to those performed by Wen *et al*. on solvated acrolein to assess the capabilities of their absolutely localized PBE-TDDFT approach.[62] Except for the $\pi \rightarrow \pi^*$ excitation energies, which in the projection-based embedding computations were closer to the CAM-B3LYP values, the two different embedding strategies (i.e., TDDFT/ELMO and PBE-TDDFT) provided completely analogous and comparable results, especially for the elimination of the low-lying charge-transfer states of standard Time-Dependent Density Functional Theory.

4.3 Test calculations on the Green Fluorescent Protein. In Figure 3, we reported the absolute discrepancies of the excitation energies resulting from TDDFT/ELMO and reduced non-embedded TDDFT calculations on models of GFP (see subsection 3.1.3) with basis-set 6-311G(d,p). As already mentioned, the reference values are the excitation energies obtained by means of full TDDFT/6-311G(d,p) calculations on the considered 161-atom model. Corresponding results obtained with the cc-pVDZ set of basis functions are given in Figure S12 of the Supporting Information.

We can easily observe that, for the TDDFT computations without embedding, the deviations from the benchmark values are not necessarily lower than the chemical accuracy limit as we increase the number of residues explicitly included in the calculation. For example, when the residues are gradually added in the direct order (see again subsection 3.1.3), the absolute discrepancy drops to 0.003 eV when the Arg96 subunit is included, but it afterwards increases up to 0.050 eV when the

Glu222 residue and three water molecules are taken into account (see Figure 3A). More importantly, if we consider the reverse order (see Figure 3B), the TDDFT calculations without embedding provide excitation energies that always differ from the reference value by more than 0.043 eV. On the contrary, for the embedding TDDFT/ELMO computations, we can notice that if a sufficient number of fragments is included in the QM region, the deviations from the benchmark value stay always below the chemical accuracy threshold, regardless of the chosen order with which the subunits are included in the chemically active subsystem. These different trends are certainly ascribable to the fact that the TDDFT/ELMO calculations take into account the influence of the environment, despite it is only approximately described through transferred and frozen extremely localized molecular orbitals. Analogous results were obtained by adopting the cc-pVDZ basis-set in the computations (see Figure S12 in the Supporting Information).

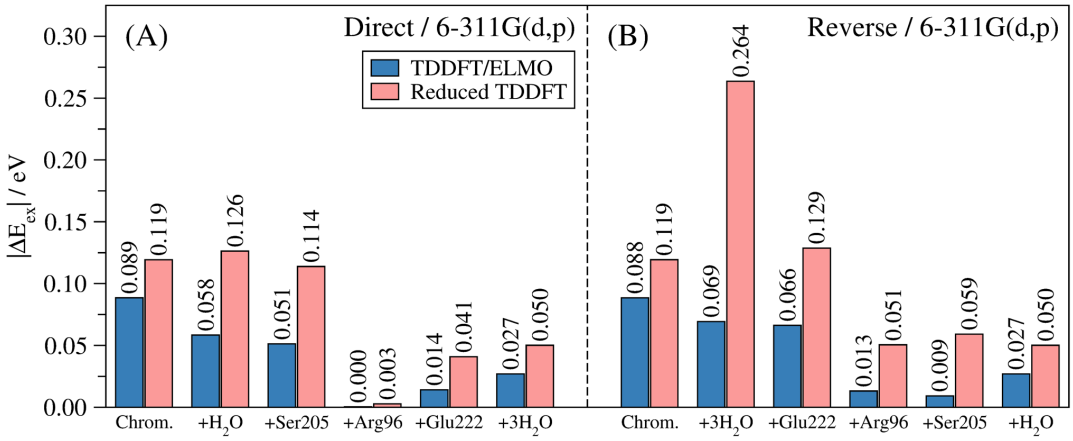


Figure 3. Absolute deviations from the full TDDFT excitation energy associated with the brightest low-lying excited-state of the 161-atom model of GFP, as resulting from TDDFT/ELMO calculations and reduced TDDFT computations (without embedding): (A) direct pathway; (B) reverse order. All the values obtained with functional B3LYP and basis-set 6-311G(d,p).

The TDDFT/ELMO calculations performed on the 161-atom model of the Green Fluorescent Protein also allowed us to roughly determine the relative importance of the different residues/moieties to the global excitation energy. In fact, in Figures 4A (which reports the results obtained with basis-set 6-311G(d,p)) we can observe that the TDDFT/ELMO calculations are quite stable, independently of the order with which the fragments are included in the QM region. For example, the inclusion of the single water molecule in the QM subsystem according to the direct order (first subunit added to the chromophore) entails a reduction of the excitation energy of 0.030 eV. Similarly, when the same water molecule is included in the QM subsystem following the reverse order (last subunit added to the chromophore), we have a lowering of the excitation energy equal to 0.036 eV. Similar results were practically obtained for all the other subunits. On the contrary, from Figures 4B, we can easily notice that the TDDFT/6-311G(d,p) calculations without embedding do not show the same stability. For instance, always considering the inclusion of the single water molecule, the excitation energy increases by 0.007 eV when the direct order is taken into account, while it decreases by 0.109 eV when the reverse pathway is followed. Similar discrepancies are also observed for the other subunits taken into account. Again, analogous results were obtained with the cc-pVDZ basis-set (see Figure S13 in the Supporting Information).

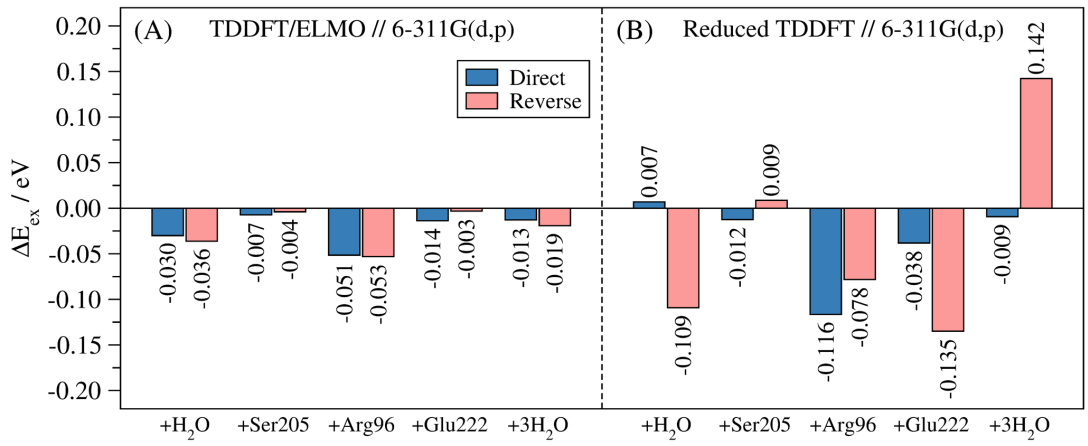


Figure 4. Contributions of the residues/moieties to the global excitation energy associated with the brightest low-lying excited-state of the 161-atom model of GFP: (A) TDDFT/ELMO calculations (direct and reverse order); (B) reduced TDDFT computations without embedding (direct and reverse pathway). All the calculations performed with functional B3LYP and basis-set 6-311G(d,p).

Furthermore, also for these test calculations, it is important to point out that all the above-discussed results are very similar to those obtained through the application of the absolutely localized version of the projection-based embedding approach extended to TDDFT. In fact, by carrying out computations on the same systems used for our validation tests, Goodpaster and coworkers had previously observed that also their PBE-TDDFT technique is able to provide more accurate results as the size of the high-level region is increased and to consistently determine the approximate contribution of each residue/moiety to the global excitation energy of the investigated system, thus showing also in that case the importance of a suitable embedding to correctly describe the effects of the environment on the chemically active region.[62]

5. EOM-CCSD/ELMO: RESULTS AND DISCUSSION

5.1 Convergence of the EOM-CCSD/ELMO calculations. In this subsection, we will analyze the convergence of the EOM-CCSD/ELMO calculations as a function of the size of the quantum mechanical region: at first, we will focus on the QM/ELMO

computations with a frontier occurring at a covalent bond between the QM and the ELMO subsystems; afterwards we will consider the results of the QM/ELMO calculations with a non-covalent boundary between the two subunits. Finally, the cost of the performed EOM-CCSD/ELMO computations will be discussed.

First of all, let us consider the results for the first three excited-states of 1-octene and let us analyze the obtained excitation energies (see Figure 5A). We can notice that, for all the three electronic transitions, the EOM-CCSD/ELMO excitation energies clearly converge towards the fully EOM-CCSD ones, with the $|\Delta E_{ex}|$ discrepancies that start being lower than 0.043 eV (chemical accuracy threshold) when at least three alkyl groups are included in the QM region. This result can be interpreted considering the natural transition orbitals (NTOs) analysis associated with the full EOM-CCSD calculation on 1-octene (see Figure S14 in the Supporting Information), from which we can evince that the extent of localization for the first three excited-states is practically equivalent.

Pertaining to the oscillator strengths of 1-octene (see Figure 5B), a gradual convergence of the EOM-CCSD/ELMO values towards the reference fully quantum mechanical ones is also observed. For this quantity, convergence is faster for the first excited-state, while for the $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$ transitions a *plateau* is reached when four alkyl moieties are included in the QM subsystem, although it is also worth noting that the oscillator strengths obtained for the $S_0 \rightarrow S_2$ excitation are one order of magnitude smaller than those obtained for the $S_0 \rightarrow S_3$ transition (see details in the caption of Figure 5 and also Tables S3 and S4 in the Supporting Information for the actual values of excitation energies and oscillator strengths obtained for 1-octene).

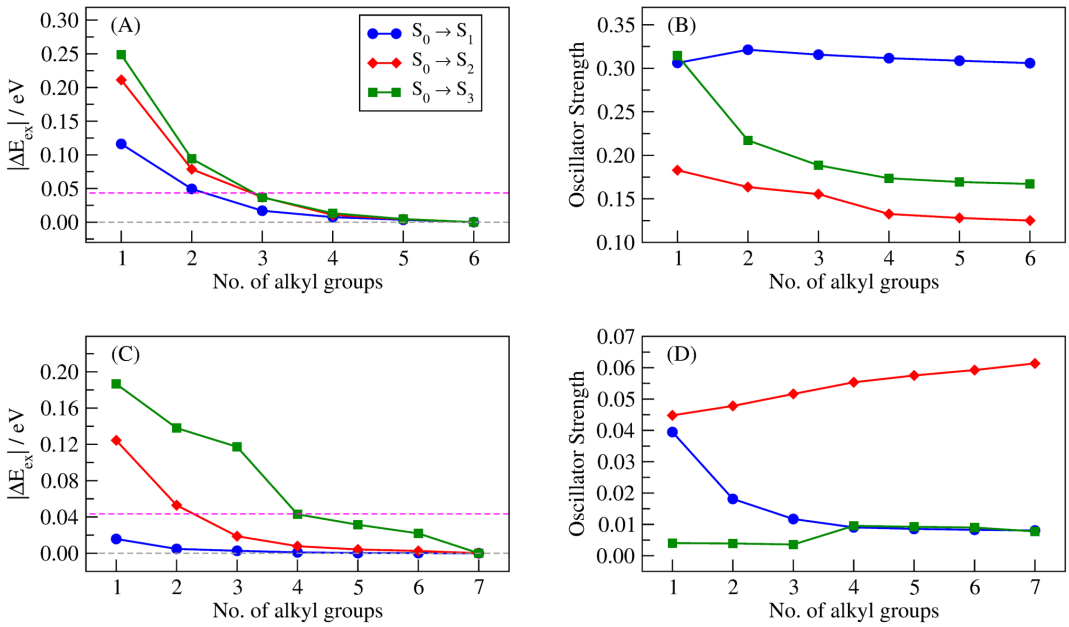


Figure 5. Results of the full EOM-CCSD and EOM-CCSD/ELMO calculations on 1-octene and octanoic acid for their first three excited-states: (A) absolute discrepancies between the EOM-CCSD and EOM-CCSD/ELMO excitation energies of 1-octene; (B) oscillator strengths of 1-octene (the $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ values are multiplied by 10; for all the transitions, the benchmark EOM-CCSD values are those obtained for six alkyl groups in the QM region); (C) absolute discrepancies between the EOM-CCSD and EOM-CCSD/ELMO excitation energies of octanoic acid; (D) oscillator strengths of octanoic acid (the $S_0 \rightarrow S_1$ values are multiplied by 100; for all the transitions, the benchmark EOM-CCSD values are those obtained for seven alkyl groups in the QM subsystem). The magenta-dashed lines in (A) and (C) indicate the chemical accuracy threshold.

Now, let us analyze the results obtained for octanoic acid. Concerning the excitation energies (see Figure 5C), the EOM-CCSD/ELMO values converge to the fully quantum mechanical results as the size of the QM region increases, regardless of the considered excited-state. Nevertheless, while for the first transition $S_0 \rightarrow S_1$ the discrepancy with respect to the EOM-CCSD value is already well below 0.043 eV when only one alkyl group is included in the quantum mechanical subsystem (discrepancy of 0.016 eV), in the other two cases the convergence towards chemical accuracy is slower. In particular, for transitions $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$, the $|\Delta E_{ex}|$

discrepancy starts being lower than or equal to the chemical accuracy threshold when at least 3 and 4 alkyl groups are considered in the QM subunit, respectively. This can be explained again considering the extent of localization for the different excitations. In fact, from the inspection of the NTOs associated with the full EOM-CCSD calculation for the first three excited-states of octanoic acid (see Figure S15 in the Supporting Information) it can be easily noticed that transition $S_0 \rightarrow S_1$ is well localized, while excitations $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$ are gradually more delocalized over the examined molecule. Convergence is also observed for the oscillator strengths. From Figure 5D, we can see that, for the first ($S_0 \rightarrow S_1$) and third ($S_0 \rightarrow S_3$) excitations, the EOM-CCSD/ELMO values practically reach a *plateau* when four alkyl groups are treated at quantum mechanical level, with the differences that, for transition $S_0 \rightarrow S_1$, the trend is monotonically decreasing and the obtained oscillator strengths are two order of magnitude lower than those observed for transition $S_0 \rightarrow S_3$ (see details in the caption of Figure 5). Concerning the second ($S_0 \rightarrow S_2$) excited-state, the values of the oscillator strength resulting from the EOM-CCSD/ELMO computations gradually approach the EOM-CCSD one with a monotonically increasing trend as the size of the QM region becomes larger, although in this case a clear *plateau* is not reached (actual values of excitation energies and oscillator strengths for octanoic acid are also respectively given in Tables S3 and S4 of the Supporting Information). The previous observations can be also seen as a clear indication that the QM/ELMO embedding methods for excited-states proposed in this work (and particularly the EOM-CCSD/ELMO strategy discussed in this section) are completely suitable for the description of local electronic transitions, while they are generally less appropriate to treat delocalized excited-states that require larger fully QM regions.

The main results of the test calculations on the systems without a covalent frontier between the QM and ELMO regions are summarized in Figure 6A, where we graphically reported the absolute deviations of the EOM-CCSD/ELMO $n \rightarrow \pi^*$ excitation energies from the benchmark EOM-CCSD values as a function of the size of the QM subsystem (see also Tables S5 and S6 in the Supporting Information for the actual values of the excitation energies and of the corresponding oscillator strengths). It is possible to notice that, in practically all the cases, the discrepancy is already below the chemical accuracy limit when only the solute molecule is included in the QM region and all the surrounding water molecules belong to the ELMO subsystem. In particular, for formaldehyde and acrylamide, the initial $|\Delta E_{ex}|$ deviations are already very small (0.010 eV and 0.008 eV, respectively) and indicate the reliability of the description of the environment at the approximate ELMO level. This is even more worthy considering the fact that the corresponding gas-phase calculations provide excitation energies that are quite far from those obtained in presence of the six surrounding water molecules, with absolute discrepancies that amount to 0.285 eV and 0.619 eV for formaldehyde and acrylamide, respectively. In the other two cases (acetaldehyde and acrolein), the excitation energy values resulting from EOM-CCSD/ELMO calculations with all the water molecules in the ELMO region are slightly below 0.043 eV (0.033 eV and 0.040 eV for acetaldehyde and acrolein, respectively). However, increasing the size of the QM subsystem, the situation significantly improves, with discrepancies that drop below 0.007 eV when only the two water molecules involved in hydrogen bonds are treated in a fully quantum mechanical way.

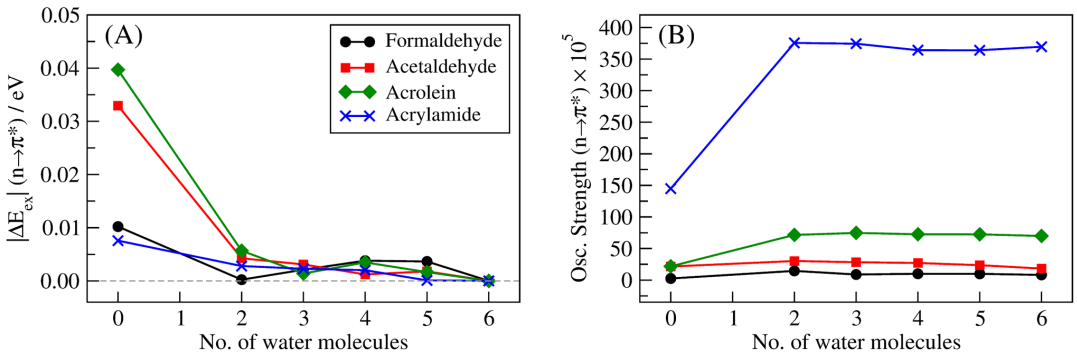


Figure 6. Results of the full EOM-CCSD and EOM-CCSD/ELMO calculations on solvated formaldehyde, acetaldehyde, acrolein, and acrylamide: (A) absolute discrepancies between the EOM-CCSD and EOM-CCSD/ELMO excitation energies associated with the $n \rightarrow \pi^*$ transitions as a function of the QM region size; (B) oscillator strengths corresponding to the $n \rightarrow \pi^*$ transitions as a function of the QM region size (the reference EOM-CCSD values are those obtained for six water molecules in the QM region).

Concerning the obtained oscillator strengths, the EOM-CCSD/ELMO computations provided values that are completely comparable to and, above all, almost always of the same order of magnitude of those obtained through the full EOM-CCSD method (see again Table S6 in the Supporting Information). In the graphs reporting the values of the oscillator strengths as a function of QM region size (see Figure 6B), we can always notice a clear increase when the number of water molecules treated at fully quantum mechanical level rises from zero to two, after which the oscillator strengths converge towards the benchmark EOM-CCSD values. In all the examined cases, a *plateau* value is practically reached when two solvent molecules are included in the QM subsystem.

To conclude this section, we focus on the computational cost of the performed EOM-CCSD/ELMO calculations. For this purpose, in Table 1 we provided the number of (frozen and active) occupied molecular orbitals, the number of active virtual orbitals and the CPU times corresponding to the EOM-CCSD/ELMO and full EOM-CCSD calculations performed on solvated acrylamide. Analyzing the collected data, we can

clearly observe that the number of active occupied molecular orbitals and virtual molecular orbitals used in the EOM-CCSD/ELMO computations are significantly lower than those in the reference EOM-CCSD calculations, especially when the size of the QM region remains quite small. As already mentioned in the Introduction and in the Theory section, this directly affects the computational cost of the EOM-CCSD/ELMO calculations, which are characterized by important reductions in terms of CPU time. For example, in Table 1 we can observe that when only two water molecules are included in the QM region (which allows to reach convergence for both the $n \rightarrow \pi^*$ excitation energy and oscillator strength) the EOM-CCSD/ELMO computation takes only 8.4% of the time taken by the corresponding full EOM-CCSD calculation. Finally, as expected, we can see that the computational cost of the EOM-CCSD/ELMO calculations gradually increases with the size of the QM subunit, although the recorded times never exceed 44.4% of the CPU time reported for the benchmark fully QM computation. Completely analogous trends were observed for the calculations carried out on the other systems taken into account in our validation tests (see Tables S7-S11 in the Supporting Information).

Table 1. Number of (frozen and active) occupied molecular orbitals (N_{occ}), number of virtual molecular orbitals (N_{virt}) and timings associated with the EOM-CCSD/ELMO and EOM-CCSD calculations (aug-cc-pVDZ basis-set) performed on acrylamide surrounded by six water molecules.[a, b]

Calculations	N_{occ}		N_{virt}	CPU time (s)	%
	Frozen	Active			
QM(0)/ELMO	35	14	151	53498.0	2.36
QM(2)/ELMO	27	22	227	189200.3	8.35
QM(3)/ELMO	23	26	265	345155.1	15.22
QM(4)/ELMO	19	30	303	569756.8	25.13
QM(5)/ELMO	15	34	341	1006533.1	44.40
Full QM	11	38	379	2267035.0	100.00

[a] The acronym QM(*N*)/ELMO indicates that *N* water molecules were included in the QM region for the EOM-CCSD/ELMO calculation; [b] the recorded timings were obtained by performing parallel calculations on 16 Intel Xeon Gold 6130 2.1 GHz processors.

5.2. Effects of the environment. The results of the test calculations to assess the capabilities of the EOM-CCSD/ELMO approach in taking into account the effects of the environment are reported in Table 2. Let us initially focus on the validation tests performed on formaldehyde. We can notice that, by including the two hydrogen-bonded water molecules in the ELMO region, the EOM-CCSD/ELMO method is able to satisfactorily take into account the environment effects, practically recovering completely the difference between the $n \rightarrow \pi^*$excitation energy of formaldehyde in the gas phase and the $n \rightarrow \pi^*$excitation energy obtained for the solvated system (i.e., formaldehyde plus two water molecules) at full EOM-CCSD level. In particular, the discrepancy drops from -0.169 eV to 0.010 eV. However, we can observe that the EOM-CCSD/ELMO description slightly overestimates the benchmark value and this might be ascribed to the concurrence of two different factors: i) the approximate description of the ground state polarization through the embedding potential given by

frozen ELMOs; ii) the complete lack of polarization response. As already mentioned in the Computational Details section, to further improve the EOM-CCSD/ELMO description, and particularly to try to partially introduce the polarization response effects, we have adopted and tested two different strategies. One consisted in introducing a TDDFT-based correction, as indicated by equation (8) in the Theory section. The results are shown in Table 2 and indicate that this strategy strongly depends on the adopted functional. In fact, it provided better agreements with the full EOM-CCSD benchmark value when functionals CAM-B3LYP and PBE0 were used (perfect agreement and discrepancy of 0.004 eV, respectively), while a worsening is observed when functional B3LYP was adopted (deviation of -0.030 eV). The other possible approach consisted in including properly chosen ELMOs/electrons in the QM region, as also successfully done by Bennie and coworkers in their coupling of the EOM-CCSD method with the projection-based embedding technique[63]. To evaluate this option, we thus performed an additional EOM-CCSD/ELMO calculation where we included in the QM subsystem eight additional electrons of the surrounding water molecules. In the initial EOM-CCSD/ELMO computation, four of these electrons had been described through ELMOs localized on the O-H bonds involved in the hydrogen bonds with formaldehyde; the other four had been described through ELMOs corresponding to lone-pairs of the oxygen atoms. The results show that the inclusion of these electrons improved the initial EOM-CCSD/ELMO description, practically leading to the EOM-CCSD benchmark value.

Table 2. $n \rightarrow \pi^*$ excitation energies resulting from EOM-CCSD and EOM-CCSD/ELMO (with and without corrections to account for polarization response) calculations on formaldehyde and acrylamide in gas-phase and in presence of two water molecules.[(a)]

Calculation	formaldehyde		acrylamide	
	E_{ex} (eV)	ΔE_{ex} (eV)	E_{ex} (eV)	ΔE_{ex} (eV)
EOM-CCSD (full; with two water molecules)	4.158	//	5.561	//
EOM-CCSD (gas-phase; no water molecules)	3.962	-0.196	5.069	-0.492
EOM-CCSD/ELMO (two water molecules in the ELMO region)	4.168	0.010	5.543	-0.018
EOM-CCSD/ELMO + TDDFT correction (B3LYP)	4.128	-0.030	5.734	0.172
EOM-CCSD/ELMO + TDDFT correction (CAM-B3LYP)	4.158	0.000	5.527	-0.035
EOM-CCSD/ELMO + TDDFT correction (PBE0)	4.162	0.004	5.805	0.243
EOM-CCSD/ELMO + ELMOs selection	4.158	0.000	5.560	-0.001

[(a)] The ΔE_{ex} discrepancies are computed with respect to the $n \rightarrow \pi^*$ excitation energies resulting from the full EOM-CCSD calculations on formaldehyde and acrylamide surrounded by two water molecules.

Analogous test calculations were carried out on acrylamide. Also in this case, the simple EOM-CCSD/ELMO computation with two water molecules treated at ELMO level almost allowed the full recovery of the gap between the $n \rightarrow \pi^*$ excitation energies obtained at full EOM-CCSD level in gas-phase and in presence of the two surrounding water molecules, with the discrepancy that decreased in absolute value from 0.492 eV to 0.018 eV. Unfortunately, in this situation, the TDDFT corrections did not improve the original EOM-CCSD/ELMO description. In fact, regardless of

the chosen functional, the deviation from the benchmark values always increased, in two cases even above the chemical accuracy threshold (0.172 eV and 0.243 eV for the B3LYP and PBE0 functionals, respectively). On the contrary, following the same procedure used for formaldehyde, the inclusion of the proper set of electrons in the QM region enabled to practically recover the reference EOM-CCSD value, with a discrepancy reduced to only -0.001 eV.

To further show the performances of the new embedded EOM-CCSD technique in capturing the effects of the environment, in Figure 7 we also reported how the $n \rightarrow \pi^*$ excitation energies of formaldehyde (see Figure 7A) and acrylamide (see Figure 7B) obtained from EOM-CCSD/ELMO, TDDFT and TDHF calculations vary when the number of solvent water molecules increases from 2 to 30. It is easy to observe that the obtained trends are practically analogous for all the methods considered in our computations, thus confirming the capability of the new EOM-CCSD/ELMO approach in properly capturing the effects of the surrounding molecules on the local electronic transitions. Furthermore, from the inset of Figure 7 we can also notice that, in the cases in which the number of solvent molecules is small enough to easily perform full EOM-CCSD calculations (i.e., from two to five water molecules), both the EOM-CCSD(0)/ELMO and the EOM-CCSD(2)/ELMO computations provided excitation energies that perfectly agree with the reference ones (i.e., always within the threshold of chemical accuracy; see Tables S13 and S14 in the Supporting Information for the actual values of the obtained excitation energies). This proves again that the new embedding EOM-CCSD/ELMO approach can be indeed used to reliably extend the range of applicability of the parent EOM-CCSD technique, without affecting the accuracy of the results, but significantly reducing the computational cost.

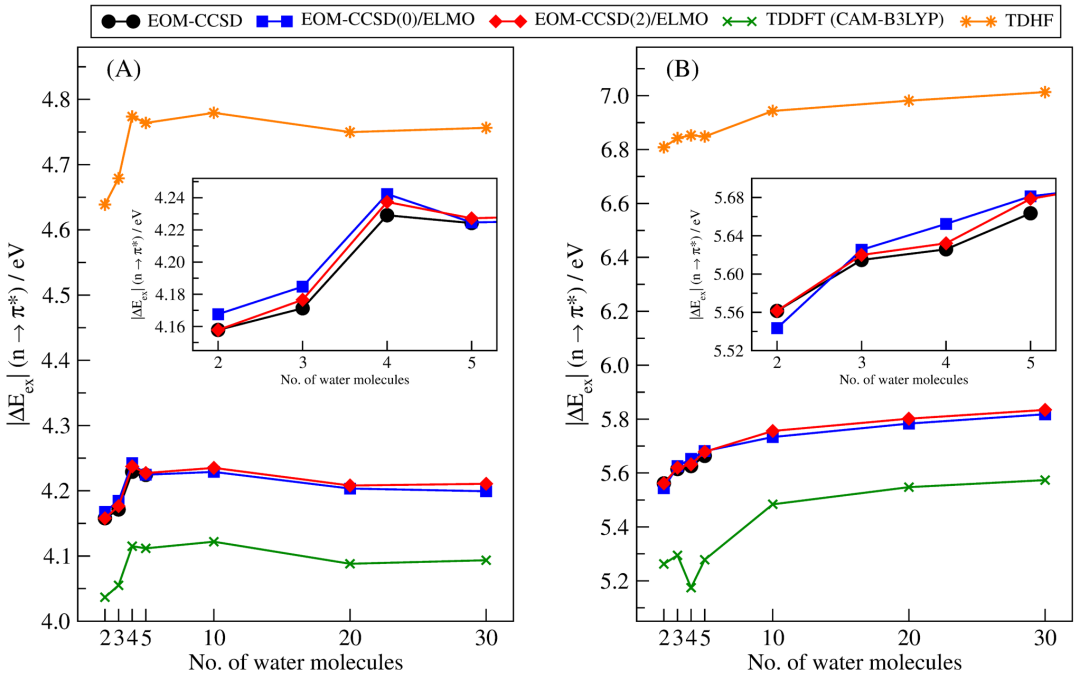


Figure 7. $n \rightarrow \pi^*$ excitation energies obtained at EOM-CCSD(0)/ELMO (no water molecules in the QM region), EOM-CCSD(2)/ELMO (two water molecules in the QM subsystem), full EOM-CCSD (when possible), TDDFT (CAM-B3LYP functional) and TDHF levels for solvated (A) formaldehyde and (B) acrylamide as the number of surrounding water molecules is gradually increased from 2 to 30; the inset highlights the EOM-CCSD(0)/ELMO, EOM-CCSD(2)/ELMO and full EOM-CCSD trends from 2 to 5 water molecules.

6. CONCLUSIONS

In this paper, we have proposed the coupling of the recently developed multiscale embedding method QM/ELMO with Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster with single and double substitutions. This gave rise to the new TDDFT/ELMO and EOM-CCSD/ELMO techniques, which potentially allow the treatment of local excited-states embedded in a potential given by transferred and frozen extremely localized molecular orbitals. To assess their performances, the new strategies were subjected to a series of validation tests previously devised to evaluate the capabilities of other embedding approaches for excited-states. From the performed calculations, it emerged that, treating only a

limited number of atoms at quantum mechanical level, the two novel embedding techniques enable to reproduce the results of the corresponding full TDDFT and EOM-CCSD calculations within the limit of chemical accuracy, for both covalent and non-covalent boundaries between the QM and ELMO regions. However, it is also important to point out that, while the novel QM/ELMO approaches are completely suitable to describe local excited-states, they are probably less recommended to deal with highly delocalized electronic transitions, for which a larger portion of the investigated system should be treated at fully quantum mechanical level to reach the desired chemical accuracy.

Concerning the TDDFT/ELMO method, it was also proved that, due to the strictly localized treatment of the region involved in the excitation, the novel method intrinsically avoids the presence of spurious low-lying charge-transfer states typical of Time-Dependent Density Functional Theory, even when one uses an exchange-correlation functional that is not long-range corrected. Furthermore, from the application to a reduced model of the Green Fluorescent Protein, it was shown that the new approach can be potentially applied to large systems and that more accurate results can be obtained when a sufficient amount of crucial fragments/residues for the electronic transition under exam is included in the fully quantum mechanical subsystem. Finally, the performed test calculations also revealed that the TDDFT/ELMO strategy can be successfully used to assess the contribution of chemical subunits to the global excitation.

Pertaining to the EOM-CCSD/ELMO technique, the validation tests also showed that, despite the use of only a non-flexible embedding potential given by transferred and frozen extremely localized molecular orbitals, the new method is able to satisfactorily take into account the effects of the environment, whose description can be further

improved through the suitable selection of ELMOs (and, consequently, of the associated electrons) to be included in the quantum mechanical subsystem.

Although the current versions of the TDDFT/ELMO and EOM-CCSD/ELMO approaches already provide quite satisfactory results, there is also room for future improvements. The most important one will consist in overcoming the limitations of the frozen and rigid embedding potential given by the transferred ELMOs. To accomplish this task, in the future we could imagine the transfer and use of virtual extremely localized molecular orbitals (already available in the constructed ELMO libraries) to account for a more reliable ground state polarization and to directly introduce the response polarization of the environment. This would clearly lead to a more flexible technique for the treatment of local excitations in large systems.

Finally, it is also worth stressing that the present work clearly showed that the QM/ELMO approach can be successfully extended also to quantum chemical methods for the investigation of excited-states. Therefore, in light of the obtained results, we envisage to couple the QM/ELMO philosophy both with simple ΔSCF methods (such as, the MOM[92] and IMOM[93] strategies proposed by Gill and coworkers) and with more sophisticated and computationally expensive multi-reference techniques (such as, CASSCF[94,95] or CASPT2[96-98]), for which the inclusion of the above-mentioned polarizable embedding will be crucial to treat the state-specific response.

ASSOCIATED CONTENT

Supporting Information. Brief review about theory, transfer and libraries of extremely localized molecular orbitals (comprising Figure S1 reporting examples of ELMOs and Figure S2 schematically depicting reference frames and atomic triads that are necessary for the ELMOs rotation). Theoretical description of the preliminary

orthogonalization procedure of the QM/ELMO method. Details about the Molecular Dynamics simulations of formaldehyde, acetaldehyde, acrolein and acrylamide in water. Figures S3-S6 depicting the systems considered for the validation tests of the TDDFT/ELMO and EOM-CCSD/ELMO methods. Figure S7 reporting the results of the full TDDFT and TDDFT/ELMO computations on the $^1B_{2u}$- and $^1B_{1u}$-like excited-states of 1-octene. Figure S8 showing the number of excited-states with excitation energy lower than 6.0 eV as obtained from B3LYP and B3LYP/ELMO calculations on solvated acrylamide. Figures S9-S11 showing the NTO analysis associated with full TDDFT and TDDFT/ELMO calculations on acrylamide surrounded by different numbers of water molecules. Figure S12 and S13 showing the results of TDDFT/ELMO and non-embedded TDDFT calculations with the cc-pVDZ basis-set on reduced models of GFP. Figures S14 and S15 showing the NTO analysis corresponding to the full EOM-CCSD calculations (first three excited-states) on 1-octene and octanoic acid, respectively. Tables S1 and S2 respectively reporting the $n \rightarrow \pi^*$ and $\pi \rightarrow \pi^*$excitation energies of acrylamide as a function of the number of surrounding water molecules (TDDFT/ELMO and full TDDFT calculations). Tables S3-S6 and Tables S13-S14 with excitation energies or oscillator strengths resulting from the validation tests carried for the ELMO-CCSD/ELMO method. Tables S7-S11 reporting further computational timings for the performed EOM-CCSD/ELMO and full EOM-CCSD calculations. Table S12 showing the computational cost of the preliminary orthogonalization procedure in TDDFT/ELMO and EOM-CCSD/ELMO calculations. (PDF)

Coordinates of all the systems taken into account in the test computations. (XYZ)

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

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMNTS

The French Research Agency (ANR) is gratefully acknowledged for financial support of this work through the Young Investigator Project *QuMacroRef* (Grant No. ANR-17-CE29-0005-01). The High-Performance Computing Center *EXPLOR* of the University of Lorraine is thanked for providing computing time through the projects 2019CPMXX0966, 2019CPMXX0886 and 2019CPMXX1332. Fabien Pascale is also acknowledged for the set-up and maintenance of our local cluster, which was used to perform most of the calculations reported in this paper.

REFERENCES

[1] Runge, E.; Gross, E. K. U. Density-Functional Theory for Time- Dependent Systems. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.

[2] Petersilka, M.; Gossmann, U. J.; Gross, E. K. U. Excitation Energies from Time-Dependent Density-Functional Theory. *Phys. Rev. Lett.* **1996**, *76*, 1212−1215.

[3] Casida, M. E. In *Recent Advances in Density Functional Methods, Part I*; Dong, D. P., Ed.; Recent Advances in Computational Chemistry; World Scientific: Singapore, 1995; pp 155−192.

[4] Rowe, D. J. Equations-of-Motion Method and the Extended Shell Model. *Rev. Mod. Phys.* **1968**, *40*, 153-166.

[5] Sekino, H.; Bartlett, R. J. A linear response, coupled-cluster theory for excitation energy. *Int. J. Quantum Chem.* **1984**, *26*, 255-565.

[6] Mukherjee, D.; Mukherjee, P. K. A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism. *Chem. Phys.* **1979**, *39*, 325−335.

[7] Stanton, J. F.; Bartlett, R. J. The equation of motion coupled cluster method. A systematic biorthogonal approach to molecular excitation energies, transition probabilities, and excited state proper- ties. *J. Chem. Phys.* **1993**, *98*, 7029-7039.

[8] Krylov, A. I. Equation-of-motion coupled-cluster methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space. *Annu. Rev. Phys. Chem.* **2008**, *59*, 433.

[9] van Gisbergen, S. J.; Fonseca Guerra, C.; Baerends, E. J. Towards Excitation Energies and (Hyper) Polarizability Calculations of Large Molecules. Application of Parallelization and Linear Scaling Techniques to Time-Dependent Density Functional Response Theory. J. Comput. Chem. **2000**, *21*, 1511−1523.

[10] Coriani, S.; Høst, S.; Jansík, B.; Thøgersen, L.; Olsen, J.; Jørgensen, P.; Reine, S.;

Pawłowski, F.; Helgaker, T.; Sałek, P. Linear- Scaling Implementation of Molecular Response Theory in Self-Consistent Field Electronic-Structure Theory. *J. Chem. Phys.* **2007**, *126*, 154108.

[11] Kjærgaard, T.; Jørgensen, P.; Olsen, J.; Coriani, S.; Helgaker, T. Hartree-Fock and Kohn-Sham Time-Dependent Response Theory in A Second-Quantization Atomic-Orbital Formalism Suitable for Linear Scaling. *J. Chem. Phys.* **2008**, *129*, 054106.

[12] Yam, C.; Yokojima, S.; Chen, G. Linear-Scaling Time-Dependent Density-Functional Theory. *Phys. Rev. B* **2003**, *68*, 153105.

[13] Cui, G.; Fang, W.; Yang, W. Reformulating Time-Dependent Density Functional Theory with Non-Orthogonal Localized Molecular Orbitals. *Phys. Chem. Chem. Phys.* **2010**, *12*, 416−421.

[14] Wu, F.; Liu, W.; Zhang, Y.; Li, Z. Linear-Scaling Time- Dependent Density Functional Theory Based on the Idea of "From Fragments to Molecule". *J. Chem. Theory Comput.* **2011**, *7*, 3643−3660.

[15] Liu, J.; Herbert, J. M. An Efficient and Accurate Approximation to Time-Dependent Density Functional Theory for Systems of Weakly Coupled Monomers. *J. Chem. Phys.* **2015**, *143*, 034106.

[16] Liu, J.; Herbert, J. M. Local Excitation Approximations to Time-Dependent Density Functional Theory for Excitation Energies in Solution. *J. Chem. Theory Comput.* **2016**, *12*, 157−166.

[17] Korona, T.; Werner, H.-J. Local treatment of electron excitations in the EOM-CCSD method. *J. Chem. Phys.* **2003**, *118*, 3006-3019.

[18] Baudin, P.; Kristensen, K. LoFEx-A local framework for calculating excitation energies: Illustrations using RI-CC2 linear response theory. *J. Chem. Phys.* **2016**, *144*, 224106.

[19] Höfener, S.; Klopper, W. Natural transition orbitals for the calculation of correlation and excitation energies. *Chem. Phys. Lett.* **2017**, *679*, 52-59.

[20] Dutta, A. K.; Neese, F.; Izsák, R. Towards a pair natural orbital coupled cluster method for excited states. *J. Chem. Phys.* **2016**, *145*, 034102.

[21] Baudin, P.; Bykov, D.; Liakh, D.; Ettenhuber, P.; Kristensen, K. A local framework for calculating coupled cluster singles and doubles excitation energies (LoFEx-CCSD). *Mol. Phys.* **2017**, *115*, 2135-2144.

[22] Kaliman, I. A.; Krylov, A. I. New algorithm for tensor contractions on multi-core CPUs, GPUs, and accelerators enables CCSD and EOM-CCSD calculations with over 1000 basis functions on a single compute node. *J. Comput. Chem.* **2017**, *38*, 842-853.

[23]Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyaev, K.; Shao, Y.; Krylov, A. I. General implementation of the resolution-of-the-identity and Cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: theory and benchmarks. *J. Chem. Phys.* **2013**, *139*, 134105.

[24] Chiba, M.; Fedorov, D. G.; Kitaura, K. Time-Dependent Density Functional Theory Based Upon the Fragment Molecular Orbital Method. *J. Chem. Phys.* **2007**, *127*, 104108.

[25] Haldar, S.; Dutta, A. K. A Multilayer Approach to the Equation of Motion Coupled-Cluster Method for the Electron Affinity. *J. Phys. Chem. A* **2020**, *124*, 3947-3962.

[26] 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.

[27] Field, M. J.; Bash, P. A.; Karplus, M. A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations. *J. Comput Chem.* **1990**, *11*, 700-733.

[28] Gao, J. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29*, 298-305.

[29] Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198-1229.

[30] Caricato, M. Absorption and Emission Spectra of Solvated Molecules with the EOM-CCSD-PCM Method. *J. Chem. Theory Comput.* **2012**, *8*, 4494−4502.

[31] Caricato, M. Exploring Potential Energy Surfaces of Electronic Excited States in Solution with the EOM-CCSD-PCM Method. *J. Chem. Theory Comput.* **2012**, *8*, 5081−5091.

[32] Caricato, M.; Lipparini, F.; Scalmani, G.; Cappelli, C.; Barone, V. Vertical Electronic Excitations in Solution with the EOM-CCSD Method Combined with a Polarizable Explicit/Implicit Solvent Model. *J. Chem. Theory Comput.* **2013**, *9*, 3035−3042.

[33] Knizia, G.; Chan, G. K.-L. Density matrix embedding: A strong coupling quantum embedding theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428−1432.

[34] Bukik, I. W.; Scuseria, G. E.; Dukelsky, J. Density matrix embedding from broken symmetry lattice mean fields. *Phys. Rev. B* **2014**, *89*, 035140.

[35] Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller, T. F., III. Embedded mean-field theory. *J. Chem. Theory Comput.* **2015**, *11*, 568−580.

[36] Ye, H,-Z.; Van Voorhis, T. Atom-Based Bootstrap Embedding For Molecules. *J. Phys. Chem. Lett.* **2019**, *10*, 6368-6374.

[37] Cortona, P. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B* **1991**, *44*, 8454−8458.

[38] Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab-initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050−8053.

[39] Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891-5928.

[40] Henderson, T. M. Embedding wave function theory in density functional theory. *J. Chem. Phys.* **2006**, *125*, 014105.

[41] Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.

[42] Genova, A.; Ceresoli, D.; Pavanello, M. Periodic subsystem density-functional theory. *J. Chem. Phys.* **2014**, *141*, 174101.

[43] Goodpaster, J. D.; Ananth, N.; Manby, F. R.; Miller, T. F., III. Exact nonadditive kinetic potentials for embedded density functional theory. *J. Chem. Phys.* **2010**, *133*, 084103.

[44] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, T. F., III. A simple, exact density-functional theory embedding scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564−2568.

[45] Lee, S. J. R.; Welborn, M.; Manby, F. R.; Miller, T. F., III. Projection-Based Wavefunction-in-DFT Embedding. *Acc. Chem. Res.* **2019**, *52*, 1359-1368.

[46] Culpitt, T.; Brorsen, K. R.; Pak, M. V.; Hammes-Schiffer, S. Multicomponent density functional theory embedding formulation. *J. Chem. Phys.* **2016**, *145*, 044106.

[47] Culpitt, T.; Brorsen, K. R.; Hammes-Schiffer, S. Density functional theory embedding with the orthogonality constrained basis set expansion procedure. *J. Chem. Phys.* **2017**, *146*, 211101.

[48] Chulhai, D. V.; Goodpaster, J. D. Improved Accuracy and Efficiency in Quantum Embedding though Absolute Localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503-1508.

[49] Chulhai, D. V.; Goodpaster, J. D. Projection-Based Correlated Wave Function in Density Functional Theory Embedding for Periodic Systems. *J. Chem. Theory Comput.* **2018**, *14*, 1928−1942.

[50] Claudino, D.; Mayhall, N. J. Automatic Partition of Orbital Spaces Based on Singular Value Decomposition in the Context of Embedding Theories. *J. Chem. Theory Comput.* **2019**, *15*, 1053-1064.

[51] Ding, F.; Tsuchiya, T.; Manby, F. R.; Miller, T. F. Linear-Response Time-Dependent Embedded Mean-Field Theory. *J. Chem. Theory Comput.* **2017**, *13*, 4216−4227.

[52] Koh, K. J.; Nguyen-Beck, T. S.; Parkhill, J. Accelerating Realtime TDDFT with Block-Orthogonalized Manby-Miller Embedding Theory, *J. Chem. Theory Comput.* **2017**, *13*, 4173-4178.

[53] Casida, M. E.; Wesołowski, T. A. Generalization of the Kohn- Sham equations with constrained electron density formalism and its time-dependent response theory formulation. *Int. J. Quantum Chem.* **2004**, *96*, 577-588.

[54] Neugebauer, J.; Louwerse, M. J.; Baerends, E. J.; Wesolowski, T. A. The Merits of the Frozen-Density Embedding Scheme to Model Solvatochromic Shifts. *J. Chem. Phys.* **2005**, *122*, 094115.

[55] Neugebauer, J. Couplings between electronic transitions in a subsystem formulation of time-dependent density functional theory. *J. Chem. Phys.* **2007**, *126*, 134116.

[56] Gomes, A. S. P.; Jacob, C. R.; Visscher, L. Calculation of Local Excitations in Large Systems By Embedding Wave-Function Theory in Density-Functional Theory. *Phys. Chem. Chem. Phys.* **2008**, *10*, 5353-5362.

[57] Höfener, S.; Gomes, A. S. P.; Visscher, L. Molecular Properties via a Subsystem Density Functional Theory Formulation: A Common Framework for Electronic Embedding. *J. Chem. Phys.* **2012**, *136*, 044104.

[58] Humbert-Droz, M.; Zhou, X.; Shedge, S. V.; Wesołowski, T. A. How to choose the frozen density in Frozen-Density Embedding Theory-based numerical simulations of local excitations? *Theor. Chem. Acc.* **2013**, *133*, 1405.

[59] Chulhai, D. V.; Jensen, L. External orthogonality in subsystem time-dependent density functional theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 21032-21039.

[60] Tölle, J.; Böckers, M.; Neugebauer, J. Exact subsystem time- dependent density-functional theory. *J. Chem. Phys.* **2019**, *150*, 181101.

[61] Scholz, L.; Tölle, J.; Neugebauer, J. Analysis of environment response effects on excitation energies within subsystem-based time-dependent density-functional theory. *Int. J. Quantum Chem.* **2020**, e26213, DOI: 10.1002/qua.26213.

[62] Wen, X.; Graham, D. S.; Chulhai, D. V.; Goodpaster, J. D. Absolutely Localized Projection-based Embedding for Excited-States. *J. Chem. Theory. Comput.* **2020**, *16*, 385-398.

[63] Bennie, S. J.; Curchod, B. F. E.; Manby, F. R.; Glowacki, D. R. Pushing the Limits of EOM-CCSD with Projector-Based Embedding for Excitation Energies. *J. Phys. Chem. Lett.* **2017**, *8*, 5559-5565.

[64] Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules. *J. Phys. Chem. A* **2019**, *123*, 9420-9428.

[65] Macetti, G.; Wieduwilt, E. K.; Assfeld, X.; Genoni. A. Localized Molecular Orbital-Based Embedding Scheme for Correlated Methods. *J. Chem. Theory Comput.* **2020**, *16*, 3578-3596.

[66] Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169−178.

[67] Fornili, A.; Sironi, M.; Raimondi, M. Determination of Extremely Localized Molecular Orbitals and Their Application to Quantum Mechanics/Molecular Mechanics Methods and to

the Study of Intramolecular Hydrogen Bonding. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 157-172.

[68] Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685-698.

[69] Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability. *J. Chem. Theory. Comput.* **2016**, *12*, 1052-1067.

[70] Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 2. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory. Comput.* **2016**, *12*, 1068-1081.

[71] Meyer, B.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. 3. Construction and Preliminary Assessment of the New Databanks. *J. Phys. Chem. A* **2018**, *122*, 8965-8981.

[72] Genoni, A; Sironi, M. A Novel Approach to Relax Extremely Localized Molecular Orbitals: the Extremely Localized Molecular Orbital-Valence Bond Method. *Theor. Chem. Acc.* **2004**, *112*, 254-262.

[73] Genoni, A.; Fornili, A.; Sironi, M. Optimal Virtual Orbitals to Relax Wave Functions Built Up with Transferred Extremely Localized Molecular Orbitals. *J. Comput. Chem.* **2005**, *26*, 827-835.

[74] Genoni, A.; Ghitti, M.; Pieraccini, S.; Sironi, M. A novel extremely localized molecular orbitals based technique for the one-electron density matrix computation. *Chem. Phys. Lett.* **2005**, *415*, 256-260.

[75] Genoni, A.; Merz. K. M., Jr.; Sironi, M. A Hylleras functional based perturbative technique to relax extremely localized molecular orbitals. *J. Chem. Phys.* **2008**, *129*, 054101.

[76] Sironi, M.; Ghitti, M.; Genoni, A.; Saladino, G.; Pieraccini, S. DENPOL: A new program to determine electron densities of polypeptides using extremely localized molecular orbitals. *J. Mol. Struct. (THEOCHEM)* **2009**, *898*, 8-16.

[77] Malaspina, L. A.; Wieduwilt, E. K.; Bergmann, J.; Kleemiss, F.; Meyer, B.; Ruiz-López, M.-F.; Pal, R.; Hupf, E.; Beckmann, J.; Piltz, R. O.; Edwards, A. J.; Grabowsky, S.; Genoni, A. Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy. *J. Phys. Chem. Lett.* **2019**, *10*, 6973-6982.

[78] Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159-4176.

[79] Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M.; Jayatilaka, D.; Macchi, P.; Madsen, A. Ø.; Massa, L. J.; Matta, C. F.; Merz, K. M., Jr.; Nakashima, P. N. H.; Ott, H.; Ryde. U.; Schwarz, K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. Eur. J.* **2018**, *24*, 10881-10905.

[80] Massa, L.; Matta, C. F. Quantum Crystallography: A perspective. *J. Comput. Chem.* **2018**, *39*, 1021-1028.

[81] Tsirelson, V. Early days of quantum crystallography: A personal account. *J. Comput. Chem.* **2018**, *39*, 1029-1037.

[82] Genoni, A.; Macchi, P. Quantum Crystallography in the Last Decade: Developments and Outlooks. *Crystals* **2020**, *10*, 473.

[83] Grabowsky, S.; Genoni, A.; Thomas, S. P.; Jayatilaka, D. The Advent of Quantum Crystallography: Form and Structure Factors from Quantum Mechanics for Advanced Structure Refinement and Wavefunction Fitting. In *Structure and Bonding. 21st Century Challenges in Chemical Crystallography 2 - Structural Correlations and Data Interpretation*; Mingos, D. M. P., Rathby, P., Eds.; Springer: Berlin, Heidelberg; DOI: 10.1007/430_2020_62.

[84] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc., Wallingford, CT, USA, 2009.

[85] Kluner, T.; Govind, N.; Wang, Y. A.; Carter, E. A. Periodic density functional embedding theory for complete active space self-consistent field and configuration interaction calculations: Ground and excited states. *J. Chem. Phys.* **2002**, *116*, 42-54.

[86] Kaila, V. R. I.; Send, R.; Sundholm, D. Electrostatic spectral tuning mechanism of the green fluorescent protein. *Phys. Chem. Chem. Phys.* **2013**, *15*, 4491-4495.

[87] Brejc, K.; Sixma, T. K.; Kitts, P. A.; Kain, S. R.; Tsien, R. Y.; Ormö, M.; Remington, S. J. Structural basis for dual excitation and photoisomerization of the Aequorea victoria green fluorescent protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2306−2311.

[88] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V. V.; Jovin, T. M.; Völker, S. Three photoconvertible forms of green fluorescent protein identified by spectral hole-burning. *Nat. Struct. Biol.* **1999**, *6*, 557-560.

[89] Creemers, T. M. H.; Lock, A. J.; Subramaniam, V.; Jovin, T. M.; Völker, S. Photophysics and optical switching in green fluorescent protein mutants. *Proc. Natl. Acad. Sci. U.S.A.*

2000, *97*, 2974−2978.

[90] Guest, M. F.; Bush, I. J.; van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. The GAMESS-UK Electronic Structure Package: Algorithms, Developments and Applications. *Mol. Phys.* **2005**, *103*, 719−747.

[91] Philipp, D. M.; Friesner, R. A. Mixed Ab Initio QM/MM Modeling Using Frozen Orbitals and Tests with Alanine Dipeptide and Tetrapeptide. *J. Comput. Chem.* **1999**, *20*, 1468-1494.

[92] Gilbert, A. T. B.; Besley, N. A.; Gill, P. M. W. Self-Consistent Field Calculations of Excited States Using the Maximum Overlap Method (MOM). *J. Phys. Chem. A* **2008**, *112*, 13164-13171.

[93] Barca, G. M. J.; Gilbert, A. T. B.; Gill, P. M. W. Simple Models for Difficult Electronic Excitations. *J. Chem. Theory Comput.* **2018**, *14*, 1501-1509.

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

[95] 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.

[96] 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.

[97] Andersson, K.; Malmqvist, P. A.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218−1226.

[98] Roos, B. O.; Andersson, K.; Fülscher, M. P.; Malmqvist, P. A.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M. Multiconfigurational perturbation theory: applications in electronic spectroscopy. *Adv. Chem. Phys.* **2007**, *93*, 219-331.