

QM/ELMO: A Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals

Giovanni Macetti, Erna K Wieduwilt, Alessandro Genoni

▶ **To cite this version:**

Giovanni Macetti, Erna K Wieduwilt, Alessandro Genoni. QM/ELMO: A Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals. Journal of Physical Chemistry A, 2021, 10.1021/acs.jpca.0c11450 . hal-03161232

HAL Id: hal-03161232

https://hal.univ-lorraine.fr/hal-03161232

Submitted on 5 Mar 2021

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

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

This document is the Accepted Manuscript version of a Published Work that appeared in final form in the Journal of Physical Chemistry A, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://pubs.acs.org/doi/10.1021/acs.jpca.0c11450

QM/ELMO: a Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals

Giovanni Macetti[(1)], Erna K. Wieduwilt[(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 development of computationally advantageous methods for the study of large systems is a long-standing research topic in theoretical chemistry. Among these techniques, a prominent place is certainly occupied by the multi-scale embedding strategies, from the well-known QM/MM (quantum mechanics / molecular mechanics) methods to the latest and promising fully quantum mechanical approaches. In this Feature Article, we will briefly review the recently proposed QM/ELMO (quantum mechanics / extremely localized molecular orbital) scheme, namely a new multiscale embedding strategy in which the most chemically relevant region of the investigated system is treated at fully quantum chemical level, while the remaining part (namely, the environment) is described by means of transferred extremely localized molecular orbitals that remain frozen throughout the computation. Other than highlighting the theoretical bases, here we will also review the main results obtained through all the currently available variants of the novel method. In particular, we will show how the QM/ELMO embedding scheme has been successfully exploited to perform both ground and excited state calculations, reproducing the results of corresponding fully quantum mechanical computations but with a much lower computational cost. A first application to crystallography will be also discussed and we will describe how the QM/ELMO approach has been recently coupled with the Hirshfeld atom refinement technique to accurately determine the positions of hydrogen atoms from X-ray diffraction data. Given the reliability and quality of the obtained results, future applications of the current versions of the QM/ELMO embedding strategy to different types of chemical problems are to be expected in the near future. Moreover, further algorithmic improvements and methodological developments are also envisaged, such as the development of a polarizable QM/ELMO scheme accounting for the effects of the QM

region on the ELMO subsystem, or the use of the new embedding approach in the context of quantum crystallography to perform unprecedented accurate refinements of macromolecular crystal structures.

KEYWORDS: Embedding methods, Extremely Localized Molecular Orbitals (ELMOs), QM/ELMO approach, ground state calculations, excited states calculations, Hirshfeld atom refinement (HAR), X-ray structural refinement, hydrogen atom positions.

1. INTRODUCTION

Nowadays, one of the still open and most tantalizing goals of theoretical chemistry is represented by the accurate description of large systems through fully quantum mechanical methods. A proof of this fact is the relentless work of several research groups that took on this challenging task over the years, steadily proposing successful and promising methods with a favorable trade-off between chemical accuracy and computational cost (see Table 1 for a schematic overview).[1,2]

In this context, a prominent place is certainly occupied by the so-called fragmentation techniques. They consist in i) partitioning the system of interest into smaller and computationally treatable fragments, ii) carrying out the computations on these subunits and, finally, iii) properly combining the obtained results to reconstruct the overall wave function or electron density of the large system under exam. Noteworthy examples belonging to this group are the "Divide & Conquer" strategy,[3-7] the Molecular Tailoring Approach,[8,9] the elongation method,[10-17] and various techniques based on databanks of fuzzy fragment electron densities or density matrices[18-24]. Other notable fragmentation techniques are the "fragment interaction methods" in which the total energy/electron density of the examined system is computed as sum of the energies/electron densities of the single subunits plus corrections due to the interactions in increasingly larger groups of fragments (namely, dimers, trimers or tetramers of subunits). Among this class of strategies, we can mention the molecular fractionation with conjugated caps (MFCC) approach[25-27] with its more recent energy-corrected version[28] (ECMFCC), the very popular and well-established fragment molecular orbital (FMO) technique[29-32] and the kernel energy method[33-35] (KEM).

Other than the fragmentation strategies briefly outlined above, another important group of techniques for the treatment of large systems at quantum mechanical level is certainly

represented by the multiscale embedding approaches, namely methods in which the chemically active region of the system under exam is described at (very) high level of theory, while the remaining part is treated through lower-level techniques. In this framework, we can include the quantum mechanics/molecular mechanics methods,[36-39] whose importance has also been recently acknowledged by the 2013 Nobel Prize in Chemistry.[40-42] Along this line, Morokuma and coworkers afterwards developed the ONIOM approach,[43,44] in which the system of interest is partitioned into multiple layers that can be also potentially described at different quantum mechanical levels, thus leading to the first type of QM/QM' method.

Table 1. Schematic overview of quantum chemistry methods for the treatment of large systems. Each method is associated with the corresponding bibliographical references in the paper.

Type of Techniques	Subgroups	Representative Examples
Fragmentation Methods	Divide & Conquer-like approaches	Divide & Conquer,[3-7] Molecular Tailoring Approach,[8,9] elongation method[10-17]
	Databanks of election densities, density matrices, and molecular orbitals	MEDLA,[18] ADMA,[19,20] Transferable Atom Equivalent method,[21] ELMO libraries[101-103]
	Fragment Interaction Methods	MFCC,[25-27] EMFCC,[28] FMO,[29-32] KEM[33-35]
Multiscale Embedding Methods	QM / Classical techniques	QM/MM,[36-42] ONIOM[43,44] (QM:MM)
	Fully QM strategies	ONIOM[43,44] (QM:QM), density matrix embedding methods[45-52] (e.g., DMET, DET, EMFT, Bootstrap Embedding), density functional embedding techniques[53-95] (e.g., FDET, PbE, OCBSE), **QM/ELMO**[96,97,109,123]

Although the QM/MM and ONIOM strategies already represent usual and well-performing tools of computational chemistry when dealing with large molecular systems, more sophisticated approaches are those based on fully quantum mechanical treatments of the environment. These techniques have known a rapid and successful development in the last twenty years and can be schematically labelled as i) density matrix[45-52] or ii) density functional embedding strategies[53-82]. The former subgroup comprises methods where suitable quantum baths are used to describe the environment and embed the fragments of interest without the need of introducing link atoms or boundary regions. Among them it is worth citing the density matrix embedding theory (DMET) initially proposed by Knizia and Chan,[45,46] the density embedding theory (DET) later introduced as a simplification of DMET by Scuseria and coworkers,[47,48] the embedded mean-field theory (EMFT) developed by Manby and Miller,[49] and, more recently, the bootstrap embedding approach devised by the Van Voorhis group.[50-52]

Much more numerous techniques have been proposed in the framework of the density functional embedding methods.[53-82] Just to explicitly mention two prominent examples, we have the pioneering frozen density embedding theory[54-56] (FDET) and the more recent projection-based embedding (PbE) technique[69-79]. FDET strongly relies on the Hohenberg & Kohn theorem and allows the optimization of the electron density/wave function of the active subunit for the investigated system under the constraint given by the electron distribution of the environment region, which, if desired, can be in turn optimized through a sort of freeze-and-thaw procedure. PbE is a very promising method recently introduced jointly by the Manby and Miller research groups. Two are the main features of this technique: i) the use of localized (Kohn-Sham) molecular orbitals (MOs) that are preliminarily assigned to the subunits (typically, active region and environment) of the system under exam; ii) the adoption of a level-shifting operator

that keeps the MOs of one subsystem orthogonal to those corresponding to the other one. The latter aspect enforces the Pauli exclusion principle between the electrons belonging to different subunits and, consequently, circumvents the problem of evaluating nonadditive kinetic potential (NAKP) contributions, which usually affects the DFT embedding strategies. The projection-based approach has been successfully exploited to perform very accurate wavefunction-in-DFT (e.g., Coupled Cluster-in-DFT or MP2-in-DFT) computations characterized by reduced computational costs compared to those of corresponding correlated calculations on the full systems. Furthermore, it has been afterwards coupled with molecular mechanics, allowing the successful modeling of enzyme reactions.[77,78]

Before concluding this brief overview, it is also worth mentioning that fully quantum mechanical embedding strategies have been exploited not only for ground state studies, but also for the investigation of excited states. In particular, they have been used to extend the applicability range of Time-Dependent Density Functional Theory (TDDFT) and, above all, of Equation-of-Motion Coupled Cluster (EOM-CC). In the former case, both density matrix and density functional embedding strategies have been considered, particularly the above-mentioned EMFT,[83,84] FDET[85-90] and PbE approaches[91-94]. Concerning EOM-CC, the projection-based embedding method has been recently exploited, both in its original version[95] and in the absolutely localized variant introduced by Goodpaster and coworkers[94].

An alternative embedding technique has been recently proposed by our group: the quantum mechanics / extremely localized molecular orbital (QM/ELMO) approach,[96,97] a method in which the most important region for the molecule/system under investigation is treated at fully quantum chemical level, while the remaining part (i.e., the environment) is described by means of frozen extremely localized molecular

orbitals[98-100] (ELMOs) transferred from recently constructed libraries[101-103] or suitable tailor-made model molecules. In fact, ELMOs are molecular orbitals strictly localized on small molecular fragments (e.g., atoms, bonds and functional groups) and, for this reason, easily transferable from one molecule to another, provided that the environments of the subunit on which the transferred orbital is localized are approximately equivalent in the two systems.[101-108]

The current version of the QM/ELMO technique presents common features both to the projection-based embedding approach[69-79] and to the well-known QM/MM strategy[36-42]. In fact, similarly to the PbE method (and particularly to its absolutely localized variant[75,76,94]), frozen (extremely) localized molecular orbitals are used to describe the environment region and to embed/polarize the active subsystem. The difference is that, while in PbE these orbitals result from a traditional localization procedure that follows a standard quantum mechanical calculation (usually at DFT level) on the full system under exam, in the novel QM/ELMO method ELMOs are transferred from external databanks that can be seen as the quantum mechanical analogs of classical non-polarizable force fields. Moreover, again in analogy with the PbE technique, also in the QM/ELMO approach it is not necessary to compute NAKP terms, although, as we will see in Section 2, this is not obtained by introducing a level-shifting operator, but by performing some preliminary orthogonalizations. For the sake of completeness, the QM/ELMO strategy also presents a certain degree of similarity with the elongation method,[10-17] a linear scaling approach that allows calculations of the electronic structure of macromolecules by performing series of computations on increasingly larger subsystems of the target molecule with a considerable number of localized molecular orbitals that are kept frozen.

The QM/ELMO technique has been originally developed in the framework of the Hartree-Fock strategy[96] and later extended to Density Functional Theory and correlated post-Hartree-Fock (post-HF) methods[97] (e.g., Møller-Plesset perturbation theory and Coupled Cluster). The performed validation tests have shown that the new approach is able to provide results that agree with the corresponding standard quantum chemical ones within the threshold of chemical accuracy (1.0 kcal/mol), even when only a small number of atoms is included in the fully quantum mechanical region (see Subsection 3.1). This also leads to significant reductions in terms of computational cost, especially when post-HF/ELMO calculations are performed (see Subsection 3.3). Afterwards, in analogy with other embedding approaches, the QM/ELMO strategy has been coupled with Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster to carry out excited state computations on extended systems (see Subsection 3.2).[109] Finally, the new embedding scheme has been also interfaced with the Hirshfeld atom refinement (HAR) technique[110-115] of quantum crystallography[116-122] for the accurate determination of hydrogen atom positions in crystal structures from X-ray diffraction data (see Subsection 3.4).[123]

In the rest of the paper, after an overview on the theoretical bases of the QM/ELMO approach (Section 2), we will illustrate the main results obtained through our new embedding scheme in the different facets briefly mentioned above (Section 3). When possible, throughout the text we will try to highlight analogies and differences with the other quantum mechanical embedding strategies. In the last section, we will draw our final conclusions and we will outline some possible future directions of our research in this field.

2. THEORETICAL BASES OF THE QM/ELMO TECHNIQUE

In this section, we will describe the theoretical fundamentals of the QM/ELMO embedding scheme. In particular, we will focus on the QM/ELMO self-consistent field (SCF) algorithm, which is also the preliminary step to obtain occupied and virtual molecular orbitals that are used in post-Hartree-Fock/ELMO ground state calculations[97] and in TDDFT/ELMO and EOM-CCSD/ELMO computations for excited states[109].

The QM/ELMO method starts with the partitioning of the system under exam into the QM and ELMO regions (see Figure 1). Although there is not a general rule for this subdivision, the QM subsystem usually corresponds only to the chemically important part of the system, while the ELMO subunit represents the chemical environment. Moreover, it is worth bearing in mind that the ELMO approach is not generally suitable for the description of conjugated regions of a molecule and, therefore, if they are present, they should be included in the QM part.

Extremely localized molecular orbitals are then transferred to the ELMO subsystem from the constructed libraries[103] or from tailor-made model molecules by exploiting the rotation/transfer strategy proposed by Philipp and Friesner[101,124] in the context of QM/MM approaches that use strictly localized bond orbitals (SLBOs) to describe the frontier between the QM and MM subunits (for more details about theory, transfer and libraries of ELMOs, we refer the readers to the papers on the construction of the ELMO databanks[101-103]). We want to point out that, unlike the parent Local Self-Consistent Field (LSCF) method,[125,126] in this case the transferred ELMOs are formally expanded only on the basis functions of the ELMO region and use only a limited number of atomic orbitals belonging to the QM subsystem, namely those centered on the frontier atoms. Although minimal, this difference was crucial to avoid linear dependency, which, on the contrary, is present in the LSCF approach and must be properly taken into account.



Figure 1. Schematic representation of the QM and ELMO regions in the QM/ELMO embedding scheme, with Y_1 and Y_2 indicating the frontier atoms between the QM and ELMO subsystems. Reprinted with permission from reference 96. Copyright 2019 American Chemical Society.

Before starting the actual QM/ELMO SCF algorithm, a preliminary orthogonalization procedure on the basis functions of the QM region and on the exported ELMOs is carried out. It consists in i) Löwdin orthonormalizing the transferred extremely localized molecular orbitals, ii) projecting out the orthonormalized ELMOs from the basis functions of the QM region, and iii) canonically orthogonalizing the QM basis functions obtained at point (ii). These three steps can be summarized through the following transformation:

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

where $\boldsymbol{\chi}$ is the starting $1 \times M$ array $[|\chi_1\rangle, |\chi_2\rangle, \dots, |\chi_M\rangle]$ of the M non-orthogonal basis functions for the full system, $\boldsymbol{\chi}'$ is the final $1 \times M_{QM}$ array $[|\chi'_1\rangle, |\chi'_2\rangle, \dots, |\chi'_{M_{QM}}\rangle]$ of the M_{QM} final orthonormal basis functions for the QM subunit (with M_{QM} much lower than M), and $\mathbf{B}$ is a global $M \times M_{QM}$ transformation matrix that has a central role in the QM/ELMO SCF algorithm. The main goal of this preliminary procedure is to guarantee the orthogonality between the molecular orbitals describing the active region and those associated with the environment subsystem, thus avoiding the necessity of evaluating non-additive kinetic potential terms. In this regard, the strategy exploited in our QM/ELMO embedding technique is different from the one based on the level-shifting parameter that is used in the projection-based embedding approach,[69-79] but it is more similar to the orthogonality constrained basis-set expansion (OCBSE)

procedure adopted by Hammes-Schiffer and coworkers in their DFT embedding method.[80,81]

After the preliminary orthogonalizations, the real SCF procedure begins and it can be schematized through the following steps.

1. Determination of the $M \times M$ Fock matrix $\mathbf{F}$ in the original and non-orthogonal basis $\boldsymbol{\chi}$.
2. Transformation of matrix $\mathbf{F}$ to the $M_{QM} \times M_{QM}$ Fock matrix $\mathbf{F'}$ for the QM subsystem in the orthogonal basis $\boldsymbol{\chi'}$ by exploiting the transformation $\mathbf{F'} = \mathbf{B}^{\dagger}\mathbf{F}\,\mathbf{B}$, where $\mathbf{B}^{\dagger}$ corresponds to the transpose of the transformation matrix $\mathbf{B}$ seen in equation (1).
3. Diagonalization of matrix $\mathbf{F'}$ to obtain the coefficients of the (occupied and virtual) molecular orbitals of the QM region: $\mathbf{F'C'} = \mathbf{C'}\,\mathbf{E'}$.
4. Transformation of the obtained molecular orbitals in the original and non-orthogonal basis $\boldsymbol{\chi}$ through the relation $\mathbf{C} = \mathbf{B}\,\mathbf{C'}$.
5. Determination of the QM one-particle density matrix: $\mathbf{P}^{QM} = \mathbf{C}\mathbf{C}^{\dagger}$.
6. Inspection of convergence on energy and density matrix. If convergence is reached, the SCF cycle ends, otherwise it restarts from point 1 where the one-particle density matrix $\mathbf{P}^{QM}$ at point 5 is used to update the Fock matrix $\mathbf{F}$.

Two points of the above SCF-cycle have to be discussed in detail: i) the expression of matrix $\mathbf{F}$ in the original basis (step 1) and ii) the reduction of the dimensionality of the Fock matrix from $M \times M$ to $M_{QM} \times M_{QM}$ (step 2).

Concerning the first aspect, the general expression of $\mathbf{F}$ in the starting basis-set of M basis functions is the following one:

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

where the usual core one-electron Hamiltonian operator is indicated as $\hat{h}^{core}$, the QM and ELMO one-electron reduced density matrices in the original basis-set as $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$, respectively, the two-electron repulsion integrals as $(\chi_\alpha \, \chi_\beta | \, \chi_\gamma \, \chi_\delta)$, the fraction of exact exchange as x, and the (μ, ν) element of the exchange-correlation potential matrix (depending on both $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$) as $\langle \chi_\mu | \hat{v}^{XC} [\mathbf{P}^{QM} + \mathbf{P}^{ELMO}] | \chi_\nu \rangle$. In case of a Hartree-Fock/ELMO computation, x becomes equal to 1 and the exchange-correlation contribution is not present.

If now we analyze equation (2) in detail, we can notice that only two terms vary and have to be updated throughout the iterations. They are the contributions corresponding to the QM subsystem ($F_{\mu\nu}^{QM}$) and the exchange-correlation potential ($v_{\mu\nu}^{XC}$). On the contrary, the one-electron term $h_{\mu\nu}$, and the contribution due to the ELMO subunit ($F_{\mu\nu}^{ELMO}$) never change and can be favorably evaluated only once before starting the SCF cycle, thus entailing a significant reduction in terms of computational cost. Nevertheless, it is also important to point out that, at the moment, the Fock matrix **F** is always evaluated in the full/supermolecular basis-set and this clearly represents the current rate-limiting step of the procedure. For this reason, following the example of some strategies already adopted in the context of the PbE approach,[70,72] we are currently trying to introduce suitable criteria with the final goal of limiting the number of atomic

orbitals over which to initially compute the matrix **F**. This will lead to further reductions of the CPU times associated with the QM/ELMO SCF calculations.

The second aspect of the algorithm that deserves a particular comment is the transformation of the $M \times M$ Fock matrix **F** in the supermolecular basis to the $M_{QM} \times M_{QM}$ Fock matrix $\mathbf{F}'$ in the orthogonal basis-set of the QM region (step 2), which clearly leads to a reduction in the dimensionality of the problems to solve and to an important lowering of the computational cost. First of all, since $M_{QM} \ll M$, the diagonalization of matrix $\mathbf{F}'$ for the QM subsystem (step 3) is already more computationally advantageous than the diagonalization of the full matrix **F**. However, the most favorable consequence derives from the fact that a greatly reduced number of virtual molecular orbitals is obtained from the diagonalization: $M_{QM} - N$ instead of $M - N$ virtual orbitals, with N as the number of occupied molecular orbitals for a *2N*-electron closed-shell QM subsystem. Considering that the computational effort associated with correlated calculations strongly depends on the number of involved virtual orbitals, the above-mentioned reduction automatically leads to a significant decrease of the cost corresponding to correlated QM/ELMO calculations for ground and excited states (particularly for Coupled Cluster / ELMO and Equation-of-Motion Coupled Cluster / ELMO calculations; see Subsection 3.3). This feature does not depend on the size of the initial Fock matrix **F** and it is common to other recent fully quantum mechanical embedding techniques, such as the absolutely localized version of the PbE approach developed by Chulhai and Goodpaster[75,76,94] or the methods introduced by the Hammes-Schiffer group[80,81] and by Claudino and Mayhall[82]. On the contrary, the original version of the projection-based embedding technique[69-74] has the disadvantage that the reduction in the number of virtual orbitals can be achieved only through the introduction of a criterion to truncate the number of basis functions over

which the Fock matrix is initially evaluated,[70,72] leading to a slightly higher computational cost. On the other hand, this can also be seen as a non-negligible advantage for the original version of the PbE method because the fact of having a larger number of virtual orbitals delocalized all over the system under exam allows the inclusion of excitations from the active subsystem to the environment.

The QM/ELMO SCF algorithm described in this subsection and the different couplings to the post-HF, TDDFT and EOM-CCSD techniques have been implemented by properly modifying the relevant subroutines of the corresponding standard quantum mechanical methods in the quantum chemistry suite of programs *Gaussian09*.[127]

3. APPLICATIONS OF THE QM/ELMO METHOD

As mentioned above, the QM/ELMO method has been already proposed in different facets. In the next subsections, for each of these variants, we will show representative results that will highlight the potentialities of the novel quantum mechanical embedding approach. In particular, in Subsection 3.1 we will discuss the use of the QM/ELMO technique for ground state calculations, also considering the application to a quite large system as a protein-ligand complex. In Subsection 3.2 we will focus on the capabilities of the TDDFT/ELMO and EOM-CCSD/ELMO techniques when they are exploited for the treatment of localized electronic transitions in relatively large systems, while in Subsection 3.3 we will comment on the significantly reduced computational cost associated with the QM/ELMO strategy when it is used in conjunction with the Coupled Cluster and Equation-of-Motion Coupled Cluster techniques. Finally, in Subsection 3.4 we will show the recent coupling of the QM/ELMO approach (at DFT level) with the Hirshfeld atom refinement method of quantum crystallography.

3.1 QM/ELMO for ground state calculations. As a first example, we consider the application of the QM/ELMO approach to the nucleophilic substitution S_N2 reaction between 1-bromodecane and the chloride anion.[97] In this case, exploiting the geometries along the intrinsic reaction coordinate determined at B3LYP/cc-pVDZ level, we performed different single point QM/ELMO calculations adopting different levels of theory for the QM region (i.e., restricted Hartree-Fock (RHF), DFT-B3LYP, MP2 and CCSD(T)) and using two different basis-sets (namely, cc-pVDZ and mixed aug-cc-pVDZ/cc-pVDZ). For all the performed QM/ELMO computations, we gradually increased the size of the QM subsystem by including from two to eight alkyl groups along with the chlorine and bromine atoms. The outcomes of the QM/ELMO calculations were then compared to those obtained through the corresponding fully quantum mechanical computations on the same geometries. Here we will show the results obtained by adopting the cc-pVDZ basis-set.

In Figure 2, we reported the absolute discrepancies $|\Delta\Delta E_{TS-R}|$ and $|\Delta\Delta E_{TS-P}|$ as a function of the number of alkyl groups included in the quantum mechanical region for the QM/ELMO calculations. $|\Delta\Delta E_{TS-R}|$ and $|\Delta\Delta E_{TS-P}|$ are the absolute deviations of the energy variations $\Delta E_{TS-R} = E_{TS} - E_{reactants}$ and $\Delta E_{TS-P} = E_{TS} - E_{products}$ (where *TS* stands for transition state) calculated at QM/ELMO level from the values of the same quantities computed through the corresponding fully QM methods. We can notice that the results obtained through the novel QM/ELMO embedding technique always differ from the fully quantum mechanical ones by less than 1.0 kcal/mol, independently of the size of the QM subsystem and of the chosen level of theory. Furthermore, as one should expect, the absolute discrepancies tend to progressively decrease as the number of alkyl groups in the chemically active region increases, although the trend is more oscillating when the ΔE_{TS-P} values are taken into account.

Analogous trends were observed also for QM/ELMO calculations performed with the mixed sets of basis functions (aug-cc-pVDZ/cc-pVDZ) for the active and environment regions.[97]

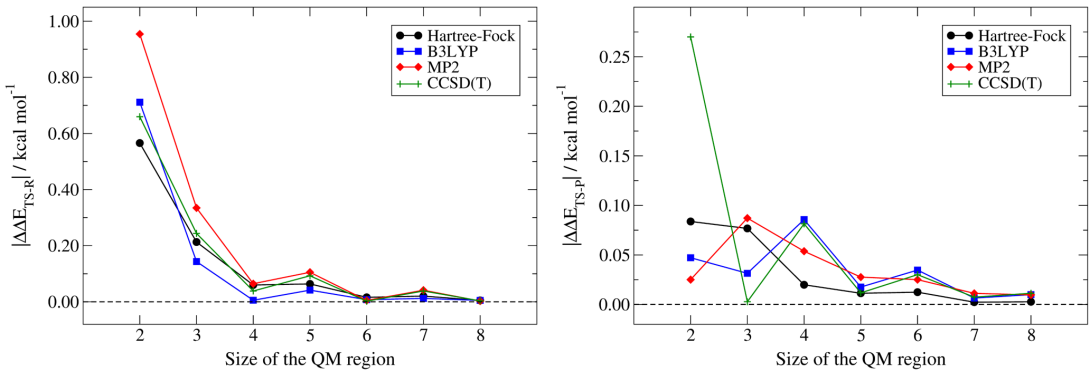


Figure 2. Absolute discrepancies of the energy variations $\Delta E_{TS-R} = E_{Transition\ State} - E_{reactants}$ (left panel) and $\Delta E_{TS-P} = E_{Transition\ State} - E_{products}$ (right panel) for the S_N2 reaction between 1-bromodecane and chloride anion computed at QM/ELMO levels (QM = HF, B3LYP, MP2 and CCSD(T)) from those obtained through the corresponding fully quantum mechanical computations. The variation of the absolute discrepancies as a function of the QM region size (i.e., number of alkyl groups) is explicitly shown. Reprinted in part and adapted with permission from reference 97. Copyright 2020 American Chemical Society.

To better assess the performances of the QM/ELMO embedding approach, the S_N2 reaction profiles obtained at RHF/ELMO, B3LYP/ELMO, MP2/ELMO and CCSD(T)/ELMO levels with basis-set cc-pVDZ and increasingly larger quantum mechanical regions were also compared point-by-point to those resulting from the corresponding fully quantum mechanical computations. For the sake of brevity, in Figure 3, we depicted only the results obtained at CCSD(T) level. We can observe that only for the smallest QM subunit there are few points along the reaction coordinate for which the deviation from the fully quantum mechanical profile is in absolute value greater than 1.0 kcal/mol. Already from the second smallest quantum mechanical region, all the discrepancies are well below the chemical accuracy limit (which is

1.0 kcal/mol or, equivalently, 0.043 eV), with the situation that systematically improves as the size of the active region increases.

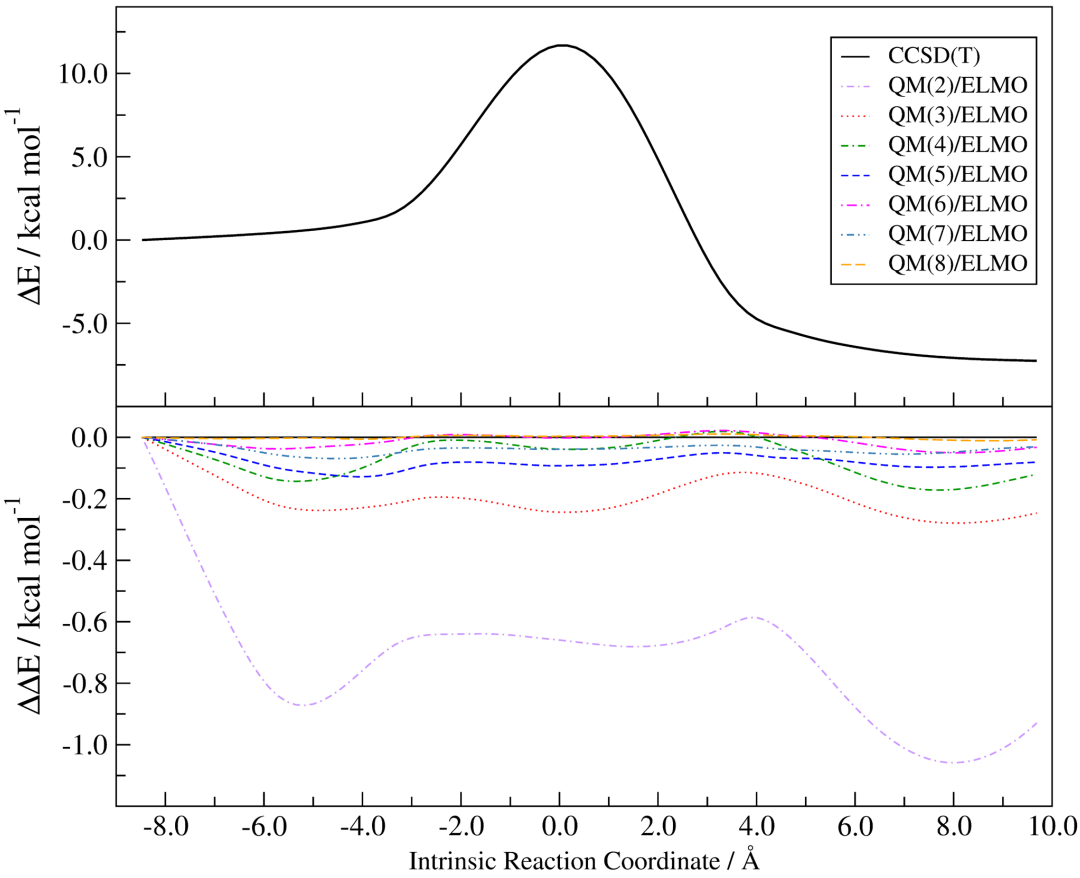


Figure 3. Energy profile for the S_N2 reaction between 1-bromodecane and chloride anion obtained at CCSD(T) level (top panel) and deviations from it when CCSD(T)/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves were obtained with the cc-pVDZ basis-set. Adapted with permission from reference 97. Copyright 2020 American Chemical Society.

Completely similar results were also observed for the other levels of theory. They are briefly summarized in Table 2, where we showed the maximum and average absolute discrepancies observed for all the reaction profiles obtained by means of QM/ELMO calculations. While for the RHF/ELMO and B3LYP/ELMO cases the maximum absolute deviations are always lower than the chemical accuracy threshold, for the MP2/ELMO and CCSD(T)/ELMO computations the maximum error is greater than 1.0 kcal/mol only for the smallest QM subsystem and it afterwards rapidly decreases as

the QM region becomes larger. Concerning the average absolute deviations, the values are always within the chemical accuracy limit, regardless of the level of theory and the dimension of the quantum mechanical subunit. From this example, it is therefore clear that, including only a small number of atoms in the chemically active region, the QM/ELMO embedding approach provides results that are in close agreement with those resulting from more expensive fully quantum chemical computations.

Table 2. Maximum and average absolute deviations between corresponding QM/ELMO and fully QM reaction energy profiles for the S_N2 reaction between 1-bromodecane and chloride anion (cc-pVDZ basis-set). Data reprinted with permission from reference 97. Copyright 2020 American Chemical Society.[(a)]

QM region size	Maximum Absolute Deviations				Average Absolute Deviations			
	HF	B3LYP	MP2	CCSD(T)	HF	B3LYP	MP2	CCSD(T)
QM(2)	0.585	0.826	1.130	1.058	0.468	0.647	0.813	0.695
QM(3)	0.213	0.241	0.335	0.279	0.146	0.148	0.228	0.195
QM(4)	0.130	0.154	0.179	0.171	0.058	0.051	0.071	0.061
QM(5)	0.009	0.115	0.138	0.128	0.054	0.055	0.084	0.080
QM(6)	0.046	0.048	0.051	0.050	0.019	0.019	0.018	0.016
QM(7)	0.043	0.051	0.073	0.068	0.021	0.021	0.040	0.039
QM(8)	0.015	0.013	0.012	0.011	0.006	0.006	0.006	0.005

[(a)] The acronym QM(*N*) indicates that *N* alkyl groups were included in the quantum mechanical region for the QM/ELMO calculations.

As another very important application of the QM/ELMO method to ground state problems, now we will show how the novel embedding technique was also profitably used to investigate large biomolecular systems. In particular, we will discuss the results obtained for the complex between the CFTR associated ligand PDZ domain and the polypeptide iCAL36 (see Figure 4A, showing also the extended network of hydrogen bonds in the protein-ligand complex).[96] In this case, we carried out RHF/ELMO

computations exploiting two different QM regions: i) region QM1, which comprises all the residues of the polypeptide plus those residues of the PDZ domain that are directly involved in hydrogen-bonds or $\pi - \pi$ interactions with iCAL36; ii) region QM2, which consists of region QM1 and also of those residues of the PDZ domain that indirectly participate in hydrophobic interactions with the ligand.

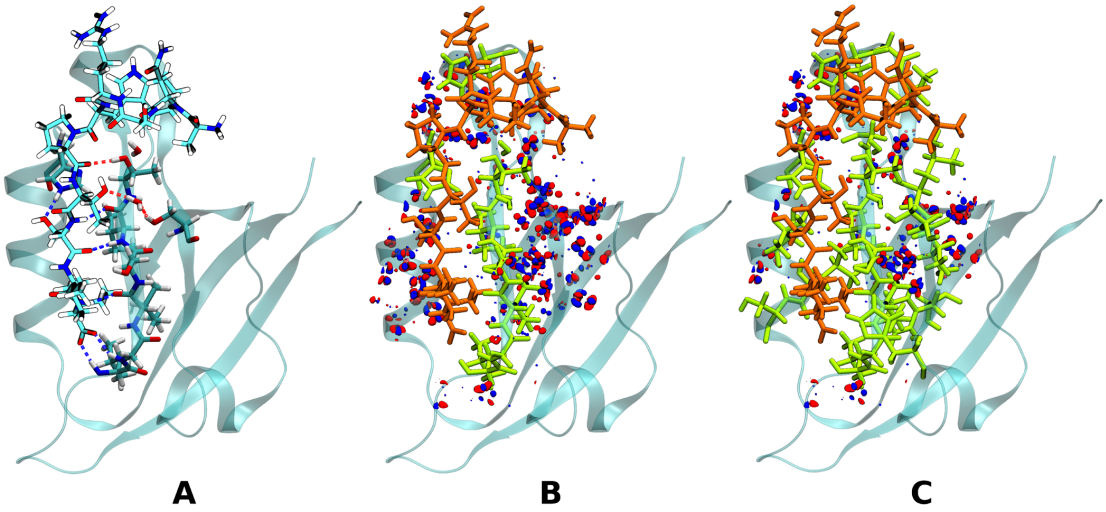


Figure 4. Three-dimensional structures of the PDZ:iCAL36 complex, with ELMO regions in "new cartoon representation" and QM regions in "licorice representation": (A) highlights of the main hydrogen-bond interactions between the ligand and the PDZ domain (blue dashed lines: N-H···O interactions; red dashed lines: O-H···O interactions); (B) residual density $\rho_{RHF/ELMO} - \rho_{RHF}$ when the quantum mechanical region QM1 is adopted; (C) residual density $\rho_{RHF/ELMO} - \rho_{RHF}$ when the quantum mechanical region QM2 is adopted. For Figures (B) and (C), the isovalues are set to 0.01 e/bohr^3, with positive and negative isosurfaces in blue and red, respectively; the ligand molecule is colored in orange, while the protein residues included in the QM regions are colored in lime green. Adapted with permission from reference 96. Copyright 2019 American Chemical Society.

In Table 3, we showed the comparison of the protein-ligand interaction energies as obtained (with and without counterpoise (CP) correction to account for the basis-set superposition error (BSSE)) through full RHF, RHF/ELMO and ONIOM calculations (see the caption of Table 2 for the details about the used basis-sets and force fields). For the sake of completeness, it is worth specifying that, in this case, the high- and low-

level regions for the ONIOM calculations respectively corresponded to the QM and ELMO subsystems adopted in the QM/ELMO computations. Notwithstanding the quite large number of intermolecular contacts between the PDZ domain and the iCAL36 polypeptide (see again Figure 4A), the QM/ELMO approach provided interaction energies that are very close to the Hartree-Fock ones, with deviations that never exceed 5 kcal/mol for both the two adopted quantum mechanical regions and regardless of the application of the counterpoise correction. Pertaining to the comparison with the results of the ONIOM calculations, we can observe that the QM/ELMO technique clearly outperforms the ONIOM method of QM:MM type. On the other hand, the situation is less clear if we consider the results of the QM:QM'-like ONIOM computations. In fact, the results oscillate depending on the size of the high-level region and on the basis-set used to treat the low-level subunit in the ONIOM calculations. Nevertheless, it is worth pointing out that the QM/ELMO results are fully in line with the QM:QM' ONIOM ones, thus confirming the full reliability of the novel quantum mechanical embedding approach when it is applied to large systems.

Finally, as one should expect, Table 3 also indicates that the results of the QM/ELMO computations improve when a larger QM region is adopted. Other than from numerical values, this can be also evinced in the plots that graphically display the differences between the RHF and RHF/ELMO electron distributions (see Figures 4B and 4C), where the region characterized by the presence of residuals becomes clearly larger when the smaller quantum mechanical subunit is exploited.

Table 3. Interaction energies between PDZ domain and iCAL36 polypeptide calculated at RHF/cc-pVDZ level ($E_{int,RHF/cc-pVDZ}$, with and without CP correction for BSSE) and deviations from them when different embedding methods are used ($\Delta E_{int,X} = E_{int,X} - E_{int,RHF/cc-pVDZ}$, where X indicates the method). Adapted with permission from reference 96. Copyright 2019 American Chemical Society. [(a)]

	QM region 1		QM region 2	
	No CP correction	CP correction	No CP correction	CP correction
$E_{int,RHF/cc-pVDZ}$	-134.84	-99.65	-134.84	-99.65
$\Delta E_{int,RHF/ELMO}$ [(b)]	4.23	2.52	-1.20	-1.39
$\Delta E_{int,\mathrm{ONIOM(RHF/cc-pVDZ:RHF/6-31G)}}$	-0.30	//	-1.60	//
$\Delta E_{int,\mathrm{ONIOM(RHF/cc-pVDZ:RHF/3-21G)}}$	-6.86	//	0.45	//
$\Delta E_{int,\mathrm{ONIOM(RHF/cc-pVDZ:UFF,emb)}}$	-8.07	//	-50.17	//
$\Delta E_{int,\mathrm{ONIOM(RHF/cc-pVDZ:UFF)}}$	-31.37	//	6.44	//

[(a)] All the energies are expressed in kcal/mol; [(b)] For the RHF/ELMO calculation, the cc-pVDZ basis-set was used for both QM and ELMO subsystems.

3.2 QM/ELMO for excited states. As already discussed in the Introduction, the QM/ELMO embedding scheme was also exploited to extend the range of applicability of two popular methods for the treatment of excited states: TDDFT and EOM-CCSD.[109] For the sake of clarity and precision, we point out that in the current versions of the TDDFT/ELMO and EOM-CCSD/ELMO strategies, the transferred and frozen ELMOs are ground state molecular orbitals that approximately account only for ground state polarization, but not for polarization response. At the moment, the latter can be included by only extending the QM region. However, another possibility that will be investigated in the future could consist in developing a self-consistent strategy to relax the ELMO electron density by taking into account the ground and excited state electron distributions of the QM subsystem.

To start assessing capabilities and performances of the developed TDDFT/ELMO and EOM-CCSD/ELMO techniques, we initially performed some validation tests on long-

chain hydrocarbons. This also enabled us to evaluate the new strategies in case of covalent boundaries between the QM and ELMO subunits. As representative examples, here we will show the results obtained through the TDDFT/ELMO approach on 1-nonylbenzene for its $^1B_{2u}$-like excited state (see Figure 5) and those obtained by means of the EOM-CCSD/ELMO method on octanoic acid for its first three electronic transitions (see Figure 6).[109]

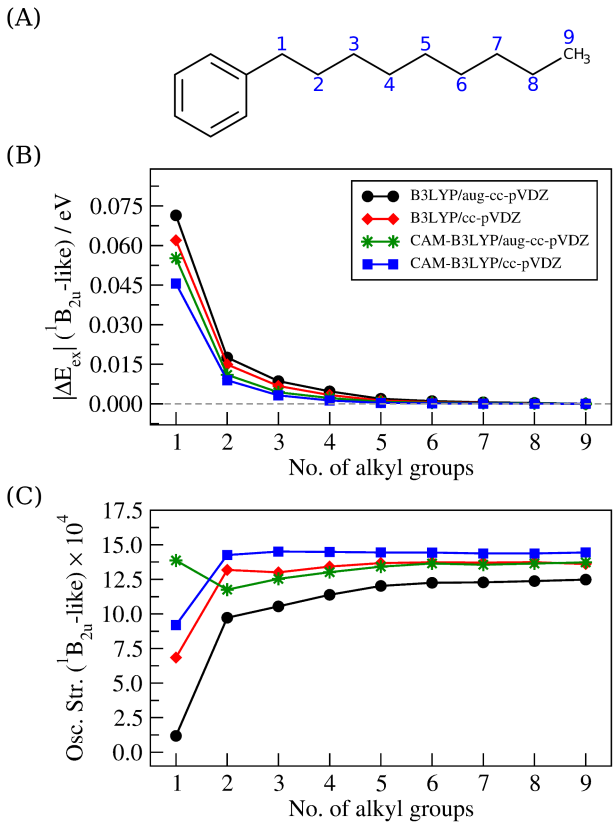


Figure 5. Results of the full TDDFT and TDDFT/ELMO calculations on 1-nonylbenzene ($^1B_{2u}$-like excited state): (A) schematic representation of 1-nonylbenzene, with the numbers indicating the labels of the alkyl groups progressively included in the QM subsystem; (B) absolute discrepancies between the excitation energies obtained at full TDDFT and TDDFT/ELMO levels; (C) oscillator strengths, with the reference TDDFT values corresponding to those obtained for nine alkyl groups in the QM region.

In Figure 5, we can observe that, for different combinations of exchange-correlation functionals and basis-sets, the TDDFT/ELMO calculations are able to excellently

reproduce the corresponding fully TDDFT results, with only a small number of alkyl groups of the hydrocarbon chain included in the quantum mechanical subsystem. In fact, concerning the excitation energies (see Figure 5B), we can notice that when only two CH_2 groups are included in the QM region along with the terminal aromatic ring, the deviations between the TDDFT/ELMO and TDDFT results start being lower than the chemical accuracy limit (0.043 eV), with the largest one (0.018 eV) observed for the TDDFT/ELMO computation at B3LYP/aug-cc-pVDZ level. The results clearly improve as larger portions of the chain are considered in the active subsystem. A similar trend can be also seen for the computed oscillator strengths (see Figure 5C), with a clear convergence of the TDDFT/ELMO results towards the TDDFT benchmark values as the size of the QM subsystem increases. In particular, we can observe that, in the different cases, a plateau is practically always reached when one includes from two to five alkyl groups in the chemically active region.

Now, let us consider the validation of the EOM-CCSD/ELMO approach, with the example for the first three excited states of octanoic acid (see Figure 6). Pertaining to the excitation energies (see Figure 6B), we can see that, also in this case, we have a clear convergence of the QM/ELMO results towards the reference EOM-CCSD values. However, it is also clear that, while for the first excited state $S_0 \rightarrow S_1$ the difference between the EOM-CCSD/ELMO and full EOM-CCSD excitation energies is already within the chemical accuracy limit when only one alkyl group is included in the QM subunit, for transitions $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$ the $|\Delta E_{ex}|$ deviation becomes lower than 0.043 eV only when three and four CH_2 groups are considered in the active subsystem, respectively. This is ascribable to the higher degree of localization associated with the first electronic transition compared to those of excitations $S_0 \rightarrow S_2$ and $S_0 \rightarrow S_3$, as it was also highlighted by the natural transition orbitals (NTOs) related to the full EOM-CCSD

calculation.[109] For the oscillator strengths, we also observed convergence (see Figure 6C). In particular, a clear plateau is reached for the first and third excited states when four alkyl moieties are included in the QM region, while for transition $S_0 \rightarrow S_2$ the EOM-CCSD/ELMO values monotonically increase and are closer and closer to the EOM-CCSD benchmark as the size of the QM region becomes larger.

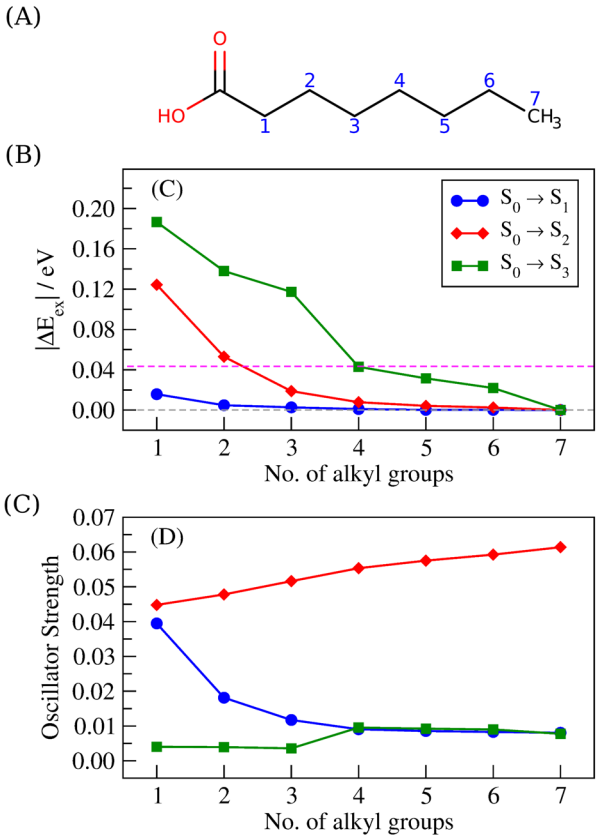


Figure 6. Results of the full EOM-CCSD and EOM-CCSD/ELMO calculations with basis-set aug-cc-pVDZ on octanoic acid (first three excited states): (A) schematic representation of octanoic acid, with the numbers indicating the labels of the alkyl groups progressively included in the QM subsystem; (B) absolute discrepancies between the excitation energies obtained at full EOM-CCSD and EOM-CCSD/ELMO levels, with the magenta-dashed line indicating the chemical accuracy threshold; (C) oscillator strengths, with the $S_0 \rightarrow S_1$ values multiplied by 100 and with the reference EOM-CCSD values corresponding to those obtained for seven alkyl groups in the QM region. Reprinted in part and adapted with permission from reference 109. Copyright 2020 American Chemical Society.

The previous validation tests showed that, also for excited states, the QM/ELMO embedding scheme is able to provide results that are in close agreement (i.e., within the chemical accuracy limit) with the fully quantum mechanical ones when only a small number of atoms are included in the chemically active subsystem. However, it also emerged that the QM/ELMO approach is particularly useful for the description of local excited states, while it could be less suitable to describe highly delocalized electronic transitions.

The performances of the developed TDDFT/ELMO method were also tested on a relatively large system, namely a 161-atom model for the A-form of the Green Fluorescent Protein (GFP; see Figure 7).[109] In particular, we computed the excitation energies associated with the brightest low-lying excited state. At a first step we carried out a series of TDDFT/ELMO calculations with a larger and larger QM region, initially consisting only of the chromophore (p-hydroxybenzylidene-imidazolinone) and then gradually including surrounding residues/moieties (see Figure 8). For comparison, we also performed a series of analogous standard TDDFT computations without embedding on further reduced models of GFP consisting only of the QM regions considered for the above-mentioned TDDFT/ELMO calculations. By including an increasing number of crucial residues in the quantum mechanical region other than the chromophore, the new ELMO-based TDDFT embedding approach provided excitation energies that are in close agreement with those resulting from corresponding benchmark TDDFT computations, where the chromophore and all the residues/moieties of the 161-atom model were treated quantum mechanically (see blue bars in Figures 8A and 8B). On the contrary, this does not always occur if one performs standard TDDFT calculations on the increasingly larger truncated systems that completely neglect the environment of the remaining and surrounding residues/moieties (see red bars in

Figures 8A and 8B). Moreover, this does not depend on the order of inclusion of the residues in the active subsystem (compare Figures 8A and 8B). Therefore, it indicates that, although approximate, the embedding description given by frozen extremely localized molecular orbitals is somehow crucial to obtain more accurate results.

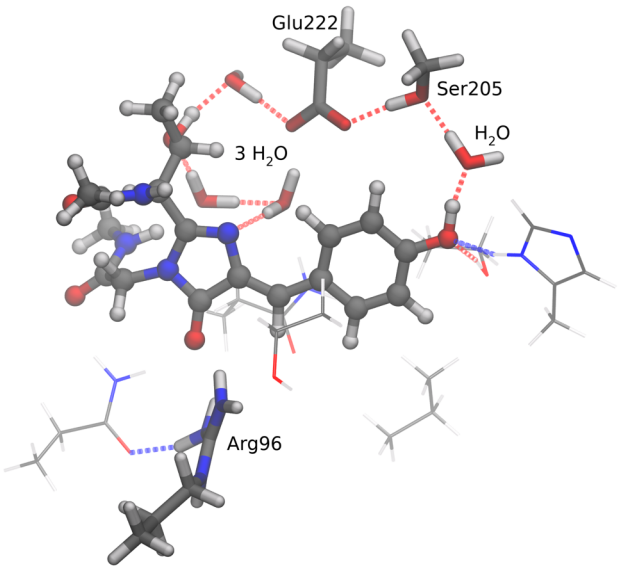


Figure 7. 161-atom model for the A-form of the Green Fluorescent Protein (GFP): chromophore in ball-and-stick representation; residues/moieties gradually included in the QM subsystem depicted in licorice representation; all the remaining subunits shown in wireframe. Hydrogen, carbon, nitrogen and oxygen atoms are depicted in light grey, dark grey, blue and red, respectively.

Very interestingly, the performed test calculations also showed that the TDDFT/ELMO approach is able to consistently quantify the contribution of each residue/subunit to the global excitation energy of the system under exam. This was again observed by analyzing the excitation energies resulting from TDDFT/ELMO calculations with larger and larger quantum mechanical regions. It was indeed noted that the variations of the excitation energies corresponding to the different residues/subunits are practically independent on the order with which these residues/subunits are included in the QM subsystem (compare corresponding blue bars in Figures 8C and 8D). Quite the opposite, this does not happen for the standard TDDFT calculations without embedding

on truncated systems, for which the excitation energy variations also markedly change according to the order of inclusion of the different residues/subunits in the calculations (compare corresponding red bars in Figures 8C and 8D).

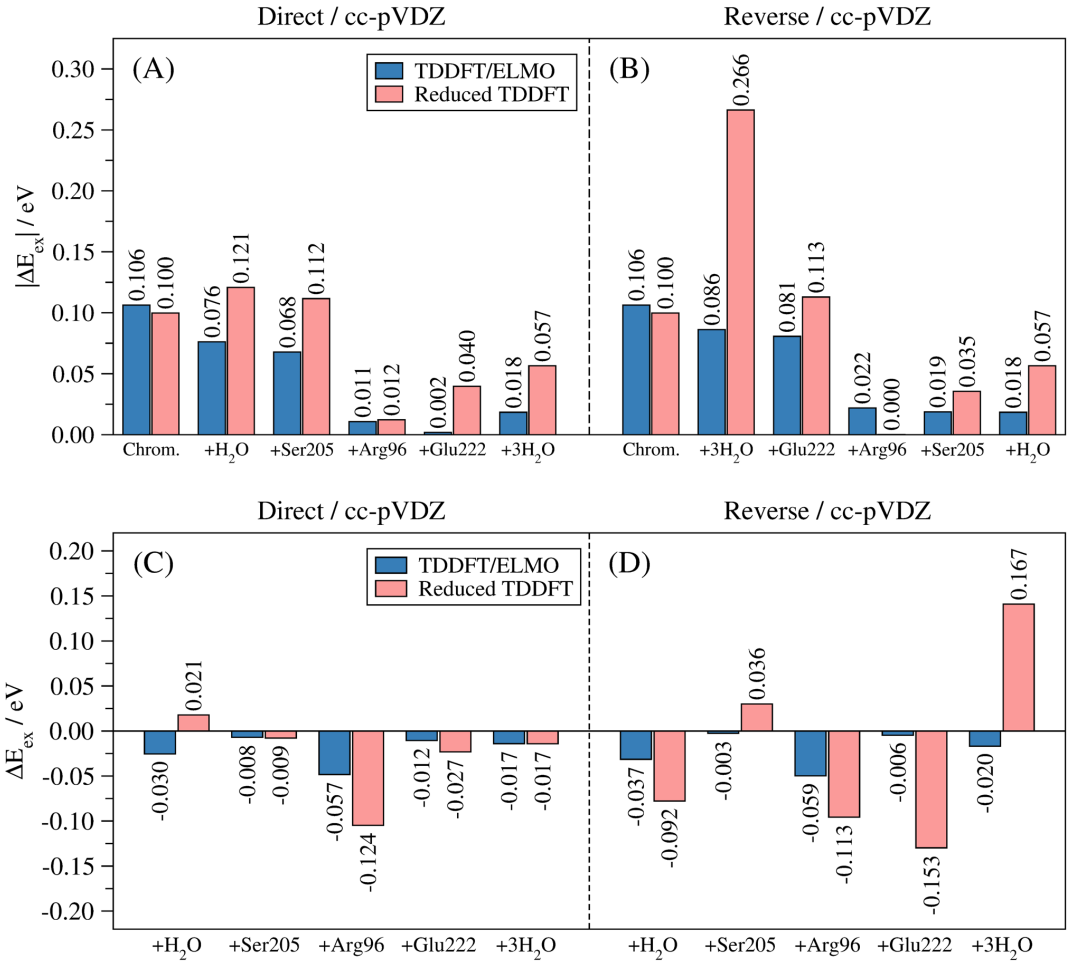


Figure 8. Validation of the TDDFT/ELMO approach on the 161-atom model of GFP: (A and B) absolute deviations from the full TDDFT excitation energies associated with the brightest low-lying excited state, as resulting from TDDFT/ELMO calculations and reduced TDDFT computations without embedding (A: direct order of including the residues/moieties; B: reverse order of including the residues/moieties); (C and D) contributions of the residues/moieties to the global excitation energy associated with the brightest low-lying excited state, as resulting from TDDFT/ELMO calculations and reduced TDDFT computations without embedding (C: direct order of including the residues/moieties; D: reverse order of including the residues/moieties). All the reported values were obtained through calculations with the cc-pVDZ basis-set.

Finally, we close this subsection showing how the effects of the environment are also properly taken into account in EOM-CCSD/ELMO calculations.[109] For this purpose, in Figure 9, we reported the values of the $n \rightarrow \pi^*$ excitation energies for formaldehyde and acrylamide as the number of surrounding solvent water molecules increases from 2 to 30. These excitation energies were obtained by performing calculations at different levels of theory with basis-set aug-cc-pVDZ: i) EOM-CCSD(0)/ELMO, namely an EOM-CCSD/ELMO computation that does not include any solvent molecule in the QM region; ii) EOM-CCSD(2)/ELMO, namely a EOM-CCSD/ELMO computation that includes two solvent molecules in the QM subsystem; iii) standard TDDFT with functional CAM-B3LYP; iv) standard Time-Dependent Hartree-Fock (TDHF). We can see that the trends obtained through the EOM-CCSD/ELMO techniques are completely analogous to those resulting from the other fully QM methods, which confirms that the effects of the external environment on the local electronic excitation are completely captured by our new ELMO-based embedding strategy. Finally, the insets of Figure 9 also show that, when the number of surrounding water molecules is small enough to allow benchmark EOM-CCSD computations, the EOM-CCSD/ELMO calculations optimally reproduce the corresponding reference values with discrepancies that are always lower than 0.043 eV. This confirms the reliability of the new EOM-CCSD/ELMO approach, which can indeed reduce the computational cost of the Equation-of-Motion Coupled Cluster method (see also Section 3.3) without significantly affecting the accuracy of the obtained results.

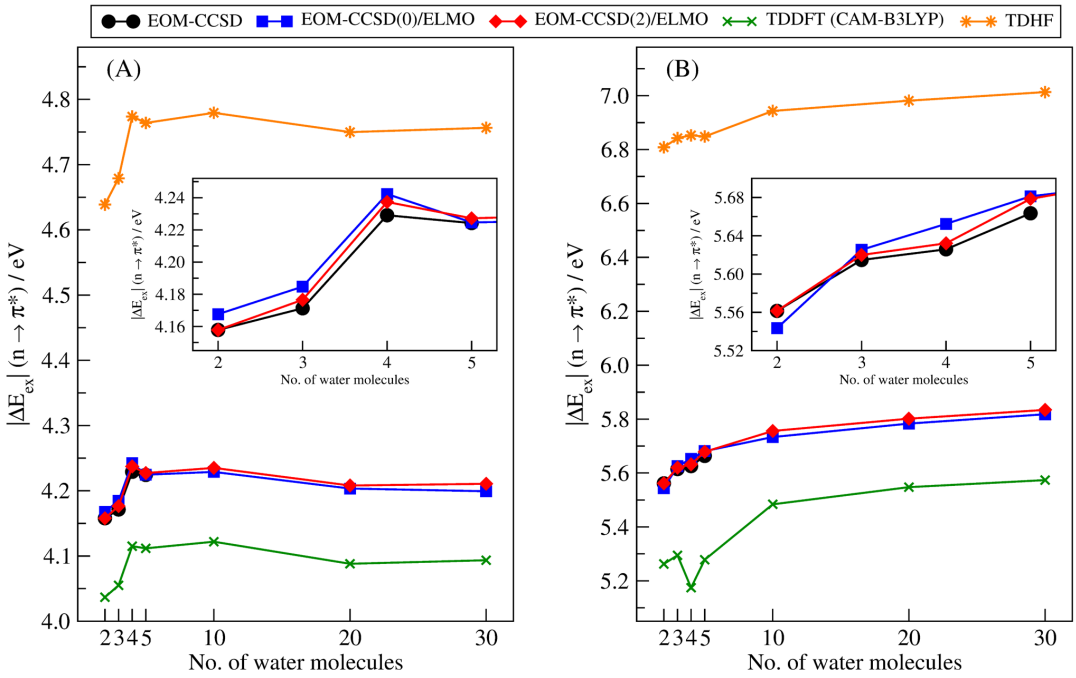


Figure 9. $n \rightarrow \pi^*$ excitation energies obtained at EOM-CCSD(0)/ELMO (no water molecules in the QM region), EOM-CCSD(2)/ELMO (two water molecules in the QM subsystem), full EOM-CCSD (when possible), TDDFT (CAM-B3LYP functional) and TDHF levels for solvated (A) formaldehyde and (B) acrylamide as the number of surrounding water molecules is gradually increased from 2 to 30; the insets highlight the EOM-CCSD(0)/ELMO, EOM-CCSD(2)/ELMO and full EOM-CCSD trends from 2 to 5 water molecules. Reprinted with permission from reference 109. Copyright 2020 American Chemical Society.

3.3 Computational cost. The QM/ELMO embedding scheme significantly reduces the computational cost of the parent quantum mechanical method. This is particularly true when the QM subunit is described through post-HF techniques for ground states (e.g., Coupled Cluster) or by means of correlated strategies for excited states (e.g., Equation-of-Motion Coupled Cluster). As already discussed in the Theory section, the reason is the limited number of virtual orbitals that are obtained from the diagonalization of the reduced Fock matrix $\mathbf{F}'$ (see Section 2).

As an example, in Table 4 we reported the number of occupied and virtual molecular orbitals, and the recorded CPU times for i) the CCSD(T)/ELMO computations performed on the transition-state geometry for the S_N2 reaction between the chloride

anion and 1-bromodecane[97] (see Subsection 3.1), and ii) the EOM-CCSD/ELMO calculations of the first three excited states of octanoic acid[109] (see Subsection 3.2).

Table 4. Number of occupied molecular orbitals (N_{occ}), number of virtual molecular orbitals (N_{virt}) and timings for i) the CCSD(T)/ELMO calculations (cc-pVDZ basis-set) performed on the transition-state geometry for the S_N2 reaction between 1-bromodecane and the chloride anion, and ii) the EOM-CCSD/ELMO calculations (aug-cc-pVDZ basis-set) of the first three excited states of octanoic acid. Some of the data are reprinted with permission from reference 97. Copyright 2020 American Chemical Society.[(a)]

Calculations	CCSD(T)/ELMO for S_N2 reaction				EOM-CCSD/ELMO on octanoic acid			
	N_{occ}	N_{virt}	CPU time[(c)] (s)	%	N_{occ}	N_{virt}	CPU time[(c)] (s)	%
QM(1)/ELMO	15	43	2553.5	0.8	11	112	39021.4	2.1
QM(2)/ELMO	18	64	3962.6	1.3	14	151	62937.9	3.4
QM(3)/ELMO	21	85	6413.0	2.1	17	190	114415.7	6.2
QM(4)/ELMO	24	106	11207.8	3.7	20	229	378907.4	20.7
QM(5)/ELMO	27	127	18194.1	5.9	23	268	582271.5	31.7
QM(6)/ELMO	30	148	33806.8	11.0	26	307	1126575.0	61.4
QM(7)/ELMO	33	169	65108.8	21.2	//	//	//	//
QM(8)/ELMO	36	190	109481.7	35.7	//	//	//	//
Full QM[(b)]	43	236	306446.5	100.0	30	354	1835239.0	100.0

[(a)] The acronym QM(*N*)/ELMO indicates that *N* alkyl groups were included in the QM region for the CCSD(T)/ELMO or EOM-CCSD/ELMO calculation; [(b)] the data for the fully quantum mechanical CCSD(T) and EOM-CCSD computations are also reported; [(c)] the recorded timings were obtained by performing parallel calculations on 16 Intel Xeon Gold 6130 2.1 GHz processors.

In both cases we can indeed notice that the numbers of occupied and virtual molecular orbitals employed in the QM/ELMO computations are much lower than those used in the corresponding benchmark fully quantum mechanical calculations. This is especially true for virtual orbitals and when the size of the QM subunit remains quite small. As mentioned above, this has a direct and important impact on the cost of the different computations. For instance, if we consider only three alkyl groups in the active

subsystems, the CCSD(T)/ELMO calculation on the transition-state of the S_N2 reaction and the EOM-CCSD/ELMO computation on octanoic acid respectively use 85 of 236 and 190 of 354 virtual orbitals, which translates into the fact that the two embedding computations take only 2% and 6% of the CPU time associated with the corresponding fully quantum chemical calculations. The computational cost obviously increases with the size of the QM region, but it is also worthwhile to observe that the largest CPU times associated with the considered CCSD(T)/ELMO and EOM-CCSD/ELMO computations are never larger than 36% and 61% of the respective standard calculations.

3.4 QM/ELMO applied to crystallography. The QM/ELMO embedding method has been recently exploited also for the refinement of crystal structures from X-ray diffraction data, and especially for the accurate determination of hydrogen atom positions. In this regard, it is worth reminding that standard structural refinements from X-ray diffraction data are based on the basic independent atom model (IAM), a very approximate method that considers the electron density of the investigated system as sum of spherically averaged atomic electron distributions and that, for this reason, completely neglects the presence of chemical bonds between atoms. As a consequence of this fact, the element-hydrogen (E-H) bond lengths that result from IAM refinements are generally too short compared to those obtained from neutron diffraction measurements. Therefore, given the intrinsic limitations of the neutron diffraction technique (such as the need of nuclear reactors or spallation sources to perform experiments) and, at the same time, the crucial importance of correctly locating the positions of hydrogen atoms in many fields of chemistry, over the years different strategies have been proposed to improve the IAM results. All of these methods

introduced the aspherical deformation of the atomic electron densities due to chemical bonding and, among them, the Hirshfeld atom refinement is the technique that recently emerged as the most promising.[110-115] This strategy consists in performing a quantum chemistry calculation at each step of the refinement (usually at HF or DFT level, even if post-HF methods have also been recently considered[128,129]) to provide quantum mechanically rigorous electron distributions that are afterwards partitioned into aspherical atomic contributions through the stockholder Hirshfeld partitioning technique[130,131]. The obtained aspherical atomic densities are used to compute thermally averaged structure factor amplitudes $\{|F_{\boldsymbol{h}}^{calc}|\}$ that, along with the experimental ones $\{|F_{\boldsymbol{h}}^{exp}|\}$, give the χ^2 statistical agreement:

$$\chi^2 = \frac{1}{N_r - N_p} \sum_{\boldsymbol{h}} \frac{\left(\eta|F_{\boldsymbol{h}}^{calc}| - |F_{\boldsymbol{h}}^{exp}|\right)^2}{\sigma_{\boldsymbol{h}}^2} \quad (3)$$

where $\boldsymbol{h}$ is the generic triad of Miller indices characterizing each reflection, N_r the number of considered reflections, N_p the number of model parameters, η a scale factor that puts $\{|F_{\boldsymbol{h}}^{calc}|\}$ and $\{|F_{\boldsymbol{h}}^{exp}|\}$ on the same scale, and $\sigma_{\boldsymbol{h}}$ the experimental uncertainty for the generic structure factor amplitude $|F_{\boldsymbol{h}}^{exp}|$. χ^2 is then minimized to obtain η, the atomic positions and the atomic anisotropic displacement parameters (ADPs). The procedure is iterated until convergence in the atomic structural parameters is achieved. Recent investigations have shown that, by exploiting the HAR technique, the positions of the hydrogen atoms can be generally determined with the same precision and accuracy that is attained through neutron diffraction measurements,[113-115] although the same quality is not usually obtained for the corresponding ADPs.[132,133] Moreover, it was also shown that HAR provides better results when the crystal environment is taken into account. In standard Hirshfeld atom refinements, this is accomplished by performing the underlying quantum mechanical calculations with external point

charges placed within a given radius at symmetry-generated positions around the reference crystal unit.

Although very simple, the embedding strategy based on clusters charges works in many situations, but it is still insufficient to achieve neutron accuracy in cases of crystals characterized by very strong intermolecular interactions. To overcome this drawback, we recently proposed to describe the crystal environment in HAR at a fully quantum mechanical level by exploiting the above-described QM/ELMO embedding approach.[123] In particular, the QM/ELMO method was used to perform the underlying quantum chemical calculations, with the chosen reference crystal unit corresponding to the QM region and the crystal environment (namely, the symmetry-generated units within a given radius from the reference unit) described by transferred and frozen ELMOs.

Test refinements were carried out exploiting X-ray diffraction data collected by Madsen and coworkers[134] for the xylitol crystal structure, which is characterized by strong hydrogen bond interactions. The underlying QM/ELMO computations were performed treating the QM subsystem at DFT-B3LYP level with basis-sets cc-pVDZ and cc-pVTZ. A preliminary version of the under-development QM/ELMO/MM method was also employed to start assessing the advantages of using a three-layer approach. All the refinements were carried out by exploiting an in-house *Bash* script that interfaced our modified version of *Gaussian09*[127], where the QM/ELMO and QM/ELMO/MM approaches are implemented, with the quantum crystallographic software *Tonto*[135] (https://github.com/dylan-jayatilaka/tonto), which is able to read the electron densities resulting from the QM/ELMO or QM/ELMO/MM calculations, performs the Hirshfeld partitioning and manages the structural refinement.

As one can already evince from Figure 10, regardless of the size of the embedding region (4 or 8 Å), the new QM/ELMO-based HAR technique provided significantly improved results, with E-H bond distances that are almost always statistically equivalent to the neutron ones[136], but very different from those resulting from Hirshfeld atom refinements without any embedding or with embeddings given by point charges.[123] Slight improvements were also observed for the hydrogen ADPs, although they are not statistically significant as those seen for the bond lengths.[123]

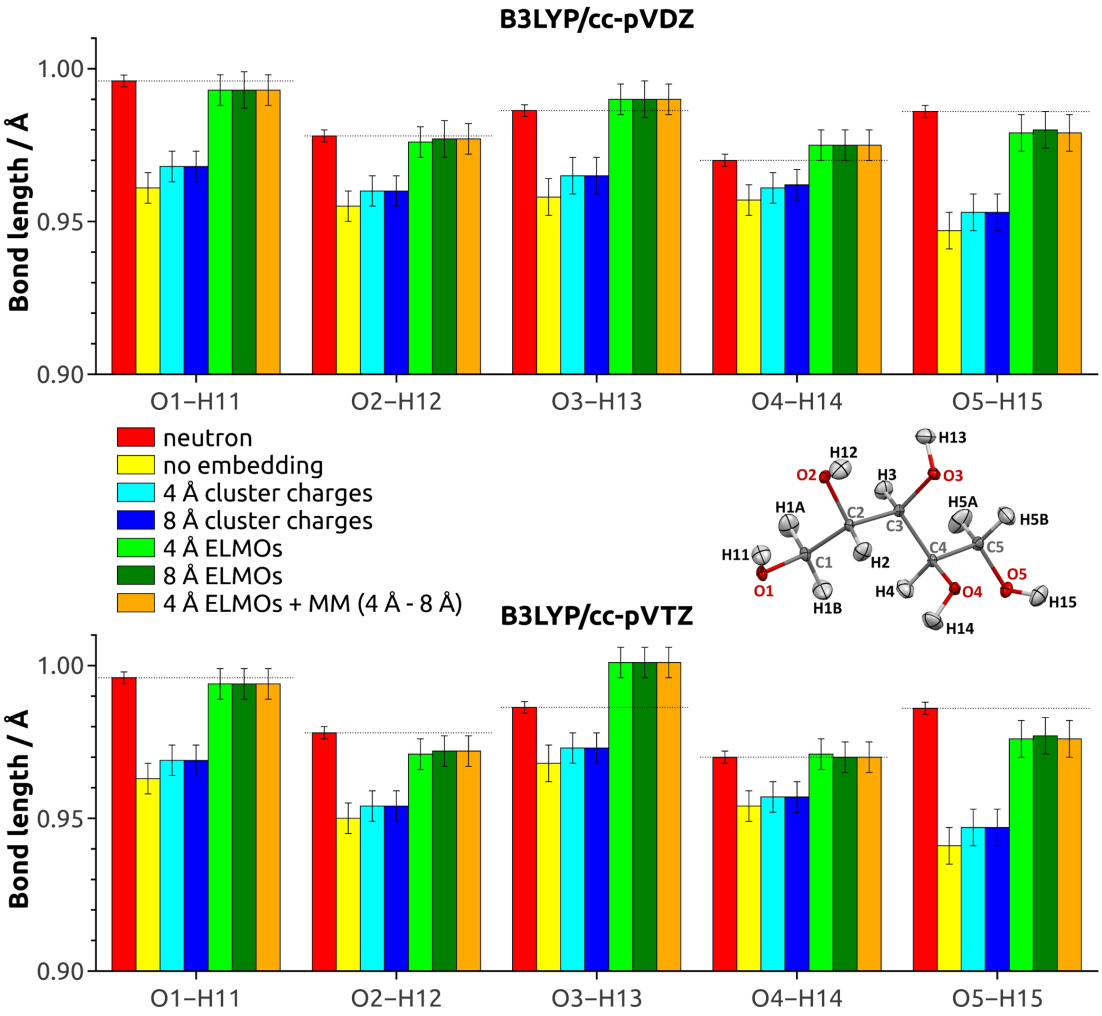


Figure 10. O-H bond lengths in xylitol as resulting from standard HARs without embedding, standard HARs with clusters charges, and HARs based on QM/ELMO(/MM) calculations; the neutron crystal structure of xylitol with specification of the atomic labels is also shown. Reprinted with permission from reference 123. Copyright 2021 American Chemical Society.

The very good results obtained through the new variant of Hirshfeld atom refinement are certainly ascribable to the fact that the ELMO embedding allows a better description of the crystal-field effects and, consequently, the capture of all those electron density variations that are due to well-defined and well-oriented intermolecular interactions. This is not possible if only point charges are used. This study obviously paved the way also to the combination of HAR with other quantum mechanical embedding approaches, with the final goal of obtaining X-ray crystal structures closer and closer to the reference neutron ones.

4. CONCLUSIONS AND PERSPECTIVES

In this paper, we have reviewed the recently developed QM/ELMO embedding scheme, showing its theoretical bases and presenting the main features and capabilities of all its current variants. In the context of ground state calculations, the new embedding strategy has been initially developed in the framework of the Hartree-Fock approach[96] and later extended to DFT and correlated post-HF techniques,[97] such as Møller-Plesset perturbation theory or Coupled Cluster method. Concerning excited state calculations, the QM/ELMO approach has been interfaced to Time-Dependent Functional Theory and Equation-of-Motion Coupled Cluster,[109] namely, two of the most popular techniques for the theoretical investigation of electronic excitations. For all these different versions, we have consistently observed that, thanks to the ELMO description of the environment, the treatment of only a small part of the examined system at a fully quantum mechanical level generally leads to optimally reproduce results that can be usually obtained by means of standard quantum chemical computations, with the additional and advantageous consequence of significantly reducing the computational cost.[97,109]

More specifically, concerning the QM/ELMO approach for ground state calculations, we have also seen that the technique can be already used to start investigating large systems of biological interest, providing reliable results that are completely comparable to those obtained by means of other well-established techniques, such as ONIOM.[96] Regarding the applications to excited states,[109] remarkable results are i) the TDDFT/ELMO capability of approximately evaluating the contributions of the different subunits to the global excitation energy, especially when large systems are studied, and ii) the possibility of extending the applicability range of the EOM-CCSD approach through the novel EOM-CCSD/ELMO strategy with a completely reliable description of the chemical environment and a lower computational cost.

Finally, we have seen that the novel ELMO-based embedding strategy has also been coupled with the promising Hirshfeld atom refinement of quantum crystallography allowing extremely accurate and precise determinations of hydrogen atom positions from X-ray diffraction data also for molecular crystals characterized by very strong intermolecular interactions.[123]

Although the currently available variants of the QM/ELMO embedding scheme already allow reliable investigations of a large number of chemical problems, there is still large room for further algorithmic improvements and other methodological developments.

First of all, it will be certainly necessary to implement the QM/ELMO analytic gradient, which will be fundamental not only to perform geometry optimizations, but also to couple the novel multiscale embedding technique to well-known quantum chemistry-based strategies for the refinement of X-ray crystal structures.[137-142]

An important upgrade of the current version of the QM/ELMO SCF algorithm will also be the adoption of a suitable strategy to truncate the number of basis functions over which constructing the initial Fock matrix. For instance, following the direction already

taken by Miller, Manby and coworkers in the context of the projection-based embedding approach,[70,72] a possible solution could consist in devising a Mulliken population-based criterion that will not affect the accuracy of the final results, but that will further speed up the QM/ELMO calculations at HF and DFT levels.

Another fundamental step forward from the algorithmic point of view will certainly be the introduction of a polarizable environment. In fact, at the moment, the environment effects are taken into account through an embedding potential given by transferred extremely localized molecular orbitals. This description is only approximate and rigid because the exported ELMOs remain unchanged during the QM/ELMO computations and are not polarized by either the ground state or the excited states of the chemically active subunit. On the contrary, in methods such as the PbE approaches, the (localized) molecular orbitals that describe the environment region are obtained through HF or DFT calculations on the whole system under exam and thus intrinsically include the (ground state) influence of the active region; in other words, the molecular orbitals are tailor-made for the system that one wants to study. Nevertheless, the clear advantage of the QM/ELMO technique is that the transfer of ELMOs is practically instantaneous and no preliminary fully QM computations are needed, thus allowing the application of the approach to much larger systems. Therefore, strategies to improve the flexibility of the QM/ELMO technique are necessary and already envisaged. For example, a possibility might consist in the development of polarizable QM/ELMO methods that could take advantage of transferred virtual ELMOs (which are already available in the current ELMO libraries) to relax the electron distribution / wave function of the environment in response to the actual ground state or excited state electron density / wave function of the active subsystem. This is in line with the state-averaged approaches already

proposed by the Carter[61] and Goodpaster[94] groups in the framework of other embedding techniques.

Concerning the methodological development, the most obvious and immediate extension of the technique is represented by the introduction of a third Molecular Mechanics layer to give rise to the QM/ELMO/MM approach. This strategy is currently under development and a preliminary version has been already used to perform some of the ELMO-embedded Hirshfeld atom refinements discussed above. After it will be completely developed and finely tuned, we envisage applications of the novel three-layer method to the investigation of interesting biochemical problems, such as enzyme reactions. Another possible methodological advancement is the interface of the QM/ELMO scheme to other traditional methods of quantum chemistry. Following this direction, we are now working to couple our embedding approach to other techniques for the investigation of excited states, such as the (Initial) Maximum Overlap Method (MOM[143] and IMOM[144]) introduced by Gill and collaborators.

Finally, given its already successful coupling with HAR for small molecules, its capability of properly accounting for the effects of the environment, and its quite straightforward applicability to large systems, we can imagine that the QM/ELMO embedding technique could be also advantageously exploited to perform structural refinements of macromolecules of biological interest. This is consistent with the recent and successful coupling of HAR with the ELMO libraries[103], which gave rise to the HAR-ELMO approach for the fast refinement of macromolecular and organometallic crystal structures.[145] Therefore, another tantalizing possibility offered by the QM/ELMO method is represented by the future development of the HAR-QM/ELMO technique that could provide more and more accurate and unprecedented structural

details for crucial regions of biomolecules, such as active sites or binding pockets of proteins.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMNTS

The French Research Agency (ANR) is gratefully acknowledged for the financial support of the Young Investigator Project *QuMacroRef* (Grant No. ANR-17-CE29-0005-01).

REFERENCES

[1] Gordon, M. S.; Slipchenko, L. V. Introduction: Calculations on Large Systems. *Chem. Rev.* **2015**, *115*, 5605-5606.

[2] Jones, L. O.; Mosquera, M. A; Schatz, G. C.; Ratner, M. A. Embedding Methods for Quantum Chemistry: Applications from Materials to Life Sciences. *J. Am. Chem. Soc.* **2020**, *142*, 3281-3295.

[3] Yang, W. Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **1991**, *66*, 1438-1441.

[4] Yang, W. Direct calculation of electron density in density-functional theory: Implementation for benzene and a tetrapeptide. *Phys. Rev. A* **1991**, *44*, 7823-7826.

[5] Dixon, S. L.; Merz, K. M., Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643-6649.

[6] Dixon, S. L.; Merz, K. M., Jr. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107*, 879-893.

[7] He, X.; Merz, K. M., Jr. Divide and Conquer Hartree-Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405-411.

[8] Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular Tailoring Approach for Simulation of Electrostatic Properties. *J. Phys. Chem.* **1994**, *98*, 9165-9169.

[9] Sahu, N.; Gadre, S. R. Molecular Tailoring Approach: A Route for *ab initio* Treatment of Large Clusters. *Acc. Chem. Res.* **2014**, *47*, 2739-2747.

[10] Imamura, A.; Aoki, Y.; Maekawa, K. A theoretical synthesis of polymers by using uniform localization of molecular orbitals: Proposal of an elongation method. *J. Chem. Phys.* **1991**, *95*, 5419-5431.

[11] Aoki, Y.; Imamura, A. Local density of states of aperiodic polymers using the localized orbitals from ab initio elongation method. *J. Chem. Phys.* **1992**, *97*, 8432-8440.

[12] Aoki, Y.; Suhai, S. An efficient cluster elongation method in density functional theory and its application to poly-hydrogen-bonding molecules. *J. Chem. Phys.* **1994**, *101*, 10808-10823.

[13] Imamura, A.; Aoki, Y.; Nishimoto, K.; Kurihara, Y.; Nagao, A. Calculations of the electronic structure of various aperiodic polymers by an elongation method. *Int. J. Quantum Chem.* **1994**, *52*, 309-319.

[14] Mitani, M.; Aoki, Y. A novel molecular orbital method for the calculations of polymer systems with local aperiodic part: The combination of the elongation method with the supercell method. *J. Chem Phys.* **1994**, *100*, 2346-2358.

[15] Kurihara, Y.; Aoki, Y.; Imamura, A. Calculations of the excitation energies of all-trans and 11,12s-dicis retinals using localized molecular orbitals obtained by the elongation method. *J. Chem. Phys.* **1997**, *107*, 3569-3575.

[16] Kurihara, Y.; Aoki, Y.; Imamura, A. Calculations of phase transition of polydiacetylenes using localized molecular orbitals by elongation method. *J. Chem. Phys.* **1998**, *108*, 10303-10308.

[17] Aoki, Y.; Gu, F. L. An elongation method for large systems toward bio-systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7640-7668.

[18] Walker, P. D.; Mezey, P. G. Ab Initio Quality Electron Densities for Proteins: A MEDLA Approach. *J. Am. Chem. Soc.* **1994**, *116*, 12022-12032.

[19] Exner, T. E.; Mezey, P. G. Ab initio-quality electrostatic potentials for proteins: An application of the ADMA approach. *J. Phys. Chem. A* **2002**, *106*, 11791-11800.

[20] Exner, T. E.; Mezey, P. G. Ab initio quality properties for macromolecules using the ADMA approach. *J. Comput. Chem.* **2003**, *24*, 1980-1986.

[21] Breneman, C. M.; Thompson, T. R.; Rhem, M.; Dung, M. Electron density modeling of large

systems using the transferable atom equivalent method. *Comput. Chem.* **1995**, *19*, 161-179.

[22] Chang, C.; Bader, R. F. W. Theoretical construction of a polypeptide. *J. Phys. Chem.* **1992**, *96*, 1654-1662.

[23] Bader, R. F. W.; Martín, F. J. Interdeterminancy of basin and surface properties of an open system. *Can. J. Chem.* **1998**, *76*, 284-29.

[24] Matta, C. F. Theoretical reconstruction of the electron density of large molecules from fragments determined as proper open quantum systems: the properties of the oripavine PEO, enkephalins, and morphine. *J. Phys. Chem. A* **2001**, *105*, 11088-11101.

[25] Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. *J. Chem. Phys.* **2003**, *119*, 3599-3605.

[26] Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett.* **2004**, *394*, 293-297.

[27] He, X.; Zhang, J. Z. H. The generalized molecular fractionation with conjugate caps/molecular mechanics method for direct calculation of protein energy. *J. Chem. Phys.* **2006**, 124, 184703.

[28] Li, S.; Li, W.; Fang, T. An efficient fragment-based approach for predicting the ground-state energies and structures of large molecules. *J. Am. Chem. Soc.* **2005**, *127*, 7251-7226.

[29] Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem. Phys. Lett.* **1999**, *313*, 701-706.

[30] Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment

molecular orbital method: application to polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614-618.

[31] Fedorov, D. G.; Kitaura, K. Theoretical development of the fragment molecular orbital (FMO) method. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; Chapter 1, pp 3-38.

[32] Fedorov, D. G.; Kitaura K. Theoretical Background of the Fragment Molecular Orbital (FMO) Method and Its Implementation in GAMESS. In *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Fedorov, D. G., Kitaura, K., Eds.; CRC Press - Taylor & Francis Group: Boca Raton, FL, 2009; Chapter 2, pp 5-36.

[33] Huang, L.; Massa, L.; Karle, J. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808-817.

[34] Huang, L.; Massa, L.; Karle, J. Kernel energy method applied to vesicular stomatitis virus nucleoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1731-1736.

[35] Huang, L.; Bohorquez, H.; Matta, C. F.; Massa, L. The Kernel Energy Method: Application to Graphene and Extended Aromatics. *Int. J. Quantum Chem.* **2011**, *111*, 4150-4157.

[36] 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.

[37] 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.

[38] Gao, J. Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers, Inc.: Weinheim, Germany, 1996; Vol. 7, pp 119-186.

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

[40] Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H_2 to Biomolecules (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 9992-10005.

[41] Levitt, M. Birth and Future of Multiscale Modeling for Macromolecular Systems (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 10006-10018.

[42] Warshel, A. Multiscale Modeling of Biological Functions: From Enzymes to Molecular Machines (Nobel Lecture). *Angew. Chem., Int. Ed.* **2014**, *53*, 10020-10031.

[43] Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO+MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt($P(t\text{-}Bu)_3)_2$ + H_2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

[44] Chung, L. W.; Sameera, W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F. et al. The ONIOM Method and Its Application. *Chem. Rev.* **2015**, *115*, 5678-5796.

[45] Knizia, G.; Chan, G. K.-L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett*. **2012**, *109*, 186404.

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

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

[48] Bulik, I. W.; Chen, W.; Scuseria, G. E. Electron correlation in solids via density embedding theory. *J. Chem. Phys.* **2014**, *141*, 035140.

[49] 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.

[50] Welborn, M.; Tsuchimochi, T.; Van Voorhis, T. Bootstrap embedding: An internally consistent fragment-based method. *J. Chem. Phys.* **2016**, *145*, 074102.

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

[52] Ye, H,-Z.; Ricke, N. D.; Tran, H. K.; Van Voorhis, T. Bootstrap Embedding for Molecules. *J. Chem/. Theory Comput.* **2019**, *15*, 4497-4506.

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

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

[55] Wesolowski, T. A. Embedding a Multideterminantal Wave Function in an Orbital-Free Environment. *Phys. Rev. A* **2008**, *77*, 012504.

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

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

[58] Iannuzzi, M.; Kirchner, B.; Hutter, J. Density functional embedding for molecular systems. *Chem. Phys. Lett.* **2006**, *421*, 16−20.

[59] Jacob, C. R.; Neugebauer, J.; Visscher, L. Software news and update: A flexible implementation of frozen-density embedding for use in multilevel simulations. *J. Comput. Chem.* **2008**, *29*, 1011−1018.

[60] Govind, N.; Wang, Y. A.; da Silva, A. J. R.; Carter, E. A. Accurate ab initio energetics of extended systems via explicit correlation embedded in a density functional environment. *Chem. Phys. Lett.* **1998**, *295*, 129−134.

[61] 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 sates. *J. Chem. Phys.* **2002**, *116*, 42-54.

[62] Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.

[63] Elliott, P.; Cohen, M. H.; Wasserman, A.; Burke, K. Density functional partition theory with fractional occupations. *J. Chem. Theory Comput.* **2009**, *5*, 827−833.

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

[65] Mi, W.; Pavanello, M. Nonlocal Subsystem Density Functional Theory. *J. Phys. Chem. Lett.* **2020**, *11*, 272-279.

[66] 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.

[67] Goodpaster, J. D.; Barnes, T. A.; Miller, T. F., III. Embedded density functional theory for covalently bonded and strongly interacting subsystems. *J. Chem. Phys.* **2011**, *134*, 164108.

[68] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III. Density functional theory embedding for correlated wavefunctions: Improved methods for open-shell systems and transition metal complexes. *J. Chem. Phys.* **2012**, *137*, 224113.

[69] 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.

[70] Barnes, T. A.; Goodpaster, J. D.; Manby, F. R.; Miller, T. F., III. Accurate basis-set truncation for wavefunction embedding. *J. Chem. Phys.* **2013**, *139*, 024103.

[71] Goodpaster, J. D.; Barnes, T. A.; Manby, F. R.; Miller, T. F., III. Accurate and systematically improvable density functional theory embedding for correlated wave functions. *J. Chem. Phys.* **2014**, *140*, 18A507.

[72] Bennie, S. J.; Stella, M.; Miller, T. F., III; Manby, F. R. Accelerating wavefunction in density-functional-theory embedding by truncating the active basis set. *J. Chem. Phys.* **2015**, *143*, 024105.

[73] Pennifold, R. C. R.; Bennie, S. J.; Miller, T. F., III; Manby, F. R. Correcting density-driven errors in projection-based embedding. *J. Chem. Phys.* **2017**, *146*, 084113.

[74] Welborn, M.; Manby, F. R.; Miller, T. F., III. Even-handed subsystem selection in projection-based embedding. *J. Chem. Phys.* **2018**, *149*, 144101.

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

[76] 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.

[77] Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J. A Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase. *J. Chem. Theory Comput.* **2016**, *12*, 2689-2697.

[78] Ranaghan, K. E.; Shchepanovska, D.; Bennie, S. J.; Lawan, N.; Macrae, S. J.; Zurek, J.; Manby, F. R.; Mulholland, A. J. Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solutions. *J. Chem. Inf. Model.* **2019**, *59*, 2063-2078.

[79] 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.

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

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

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

[88] 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.

[89] 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.

[90] 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.

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

[92] Tölle, J.; Böckers, M.; Neugebauer, J. Exact subsystem time-dependent density-functional

theory. *J. Chem. Phys.* **2019**, *150*, 181101.

[93] 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**, *120*, e26213.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

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

[99] 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.

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

[101] 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.

[102] 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.

[103] 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.

[104] 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.

[105] 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.

[106] 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.

[107] 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.

[108] 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.

[109] Macetti, G.; Genoni, A. Quantum Mechanics/Extremely Localized Molecular Orbital Embedding Strategy for Excited States: Coupling to Time-Dependent Density Functional Theory and Equation-of-Motion Coupled Cluster. *J. Chem. Theory Comput.* **2020**, *16*, 7490-7506.

[110] Jayatilaka, D.; Dittrich, B. X-ray structure refinement using aspherical atomic density functions obtained from quantum mechanical calculations. *Acta Crystallogr., Sect. A* **2008**, *64*, 383-393.

[111] Capelli, S.C.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld Atom Refinement. *IUCrJ* **2014**, *1*, 361-379.

[112] Woińska, M.; Jayatilaka, D.; Spackman, M. A.; Edwards, A. J.; Dominiak, P. M.; Woźniak, K.; Nishibori, E.; Sugimoto, K.; Grabowsky, S. Hirshfeld atom refinement for modeling strong hydrogen bonds. *Acta Crystallogr., Sect. A* **2014**, *70*, 483-498.

[113] Woińska, M.; Grabowsky, S.; Dominiak, P. M.; Woźniak, K.; Jayatilaka, D. Hydrogen atoms can be located accurately and precisely by x-ray crystallography. *Sci. Adv.* **2016**, *2*, e1600192.

[114] Fugel, M.; Jayatilaka, D.; Hupf, E.; Overgaard, J.; Hathwar, V. R.; Macchi, P.; Turner, M. J.; Howard, J. A. K.; Dolomanov, O. V.; Puschmann, H. et al. Probing the accuracy and precision of Hirshfeld atom refinement with HARt interfaced with Olex2. *IUCrJ* **2018**, *5*, 32-44.

[115] Kleemiss, F.; Dolomanov, O. V.; Bodensteiner, M.; Peyerimhoff, N.; Midgley, L.; Borhis, L. J.; Genoni, A.; Malaspina, L. A.; Jayatilaka, D.; Spencer, J. L. et al. Accurate Crystal Structures and Chemical Properties from NoSpherA2. *Chem. Sci.* **2021**, DOI:10.139/D0SC05526C.

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

[117] Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M. et al. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. Eur. J.* **2018**, *24*, 10881-10905.

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

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

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

[121] 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 *21st Century Challenges in Chemical Crystallography II - Structural Correlations and Data Interpretation*. *Structure and Bonding*; Mingos, D. M. P., Rathby, P., Eds.; Springer: Berlin, Heidelberg; vol 186, pp 65-144. DOI: 10.1007/430_2020_62.

[122] Macchi, P. The connubium between crystallography and quantum mechanics. *Crystallogr. Rev.* **2020**, *26*, 209-268.

[123] Wieduwilt, E. K.; Macetti, G.; Genoni, A. Climbing Jacob's Ladder of Structural Refinement: Introduction of a Localized Molecular Orbital-Based Embedding for Accurate X-ray Determinations of Hydrogen Atoms Positions. *J. Phys. Chem. Lett.* **2021**, *12*, 463-471.

[124] 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.

[125] Assfeld, X.; Rivail, J.-L.; Quantum Chemical Computations on Parts of Large Molecules: the Ab Initio Local Self Consistent Field Method. *Chem. Phys. Lett.* **1996**, *263*, 100-106.

[126] Ferré, N.; Assfeld, A.; Rivail, J.-L. Specific Force Field Parameters Determination for the Hybrid Ab Initio QM/MM LSCF Method. *J. Comput. Chem.* **2002**, *23*, 610-624.

[127] 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. et al. *Gaussian 09*, Revision D.01; Gaussian, Inc., Wallingford, CT, USA, 2009.

[128] Wieduwilt, E. K.; Macetti, G.; Malaspina, L. A.; Jayatilaka, D.; Grabowsky, S.; Genoni, A. Post-Hartree-Fock methods for Hirshfeld atom refinement: are they necessary? Investigation of a strongly hydrogen-bonded molecular crystal. *J. Mol. Struct.* **2020**, *1209*, 127934.

[129] Chodkiewicz, M. L.; Woińska, M.; Woźniak, K. Hirshfeld atom like refinement with alternative electron density partitions. *IUCrJ* **2020**, *7*, 1199-1215.

[130] Hirshfeld, F. L. XVII. Spatial Partitioning of Charge Density. *Isr. J. Chem.* **1977**, *16*, 198-201.

[131] Hirshfeld, F. L. Bonded-Atom Fragments for Describing Molecular Charge Densities. *Theoret. Chim. Acta* **1977**, *44*, 129-138.

[132] Köhler, C.; Lübben, J.; Krause, L.; Hoffmann, C.; Herbst-Irmer, R.; Stalke, D. Comparison of different strategies for modelling hydrogen atoms in charge density analyses. *Acta Crystallogr., Sect. B* **2019**, *75*, 434-441.

[133] Malaspina, L. A.; Hoser, A. A.; Edwards, A. J.; Woińska, M.; Turner, M. J.; Price, J. R.; Sugimoto, K.; Nishibori, E.; Bürgi, H.-B.; Jayatilaka, D. et al. Hydrogen atoms in bridging positions fromquantum crystallographic refinements: influenceof hydrogen atom displacement parameters ongeometry and electron density. *CrystEngComm* **2020**, *22*, 4778-4789.

[134] Madsen, A. Ø.; Sørensen, H. O.; Flensburg, C.; Stewart, R. F.; Larsen, S. Modeling of the nuclear parameters for H atoms in X-ray charge-density studies. *Acta Cryst. A* **2004**, *60*, 550-561.

[135] Jayatilaka, D; Grimwood, D. J. Tonto: A Fortran Based Object-Oriented System for Quantum Chemistry and Crystallography. In *Computational Science – ICCS 2003*; Sloot, P. M. A., Abramson, D., Bogdanov, A. V., Dongarra, J. J., Zomaya, A. Y., Gorbachev, Y. E., Eds.; Springer-Verlag: Berlin & Heidelberg, 2003; Chapter 4, pp 142-151.

[136] Madsen, A. Ø.; Mason, S.; Larsen, S. A neutron diffraction study of xylitol: derivation of mean square internal vibrations for H atoms from a rigid-body description. *Acta Cryst. B* **2003**, *59*, 653-663.

[137] Ryde, U.; Olsen, L.; Nilsson, J. Quantum chemical geometry optimizations in proteins using crystallographic raw data. *J. Comput. Chem.* **2002**, *23*, 1058-1070.

[138] Ryde, U.; Nilsson, J. Quantum refinement – a combination of quantum chemistry and protein crystallography. *J. Mol. Struct.: THEOCHEM* **2003**, *632*, 259-275.

[139] Ryde, U. Accurate metal-site structures in proteins obtained by combining experimental data and quantum chemistry. *Dalton Trans.* **2007**, 607-625.

[140] Yu, N.; Hayik, S. A.; Wang, B.; Liao, N.; Reynolds, C. H.; Merz, K. M., Jr. Assigning the protonation states of key aspartates in β-secretase using QM/MM X-ray structure refinement. *J. Chem. Theory Comput.* **2006**, *2*, 1057-1069.

[141] Li, X.; He, X.; Wang, B.; Merz, K. M., Jr. Conformational Variability of Benzamidinium-Based Inhibitors. *J. Am. Chem. Soc.* **2009**, *131*, 7742-7754.

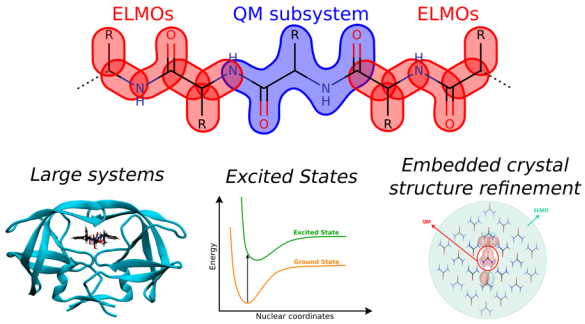
[142] Li, X.; Hayik, S. A.; Merz, K. M., Jr. QM/MM X-ray refinement of zinc metalloenzymes. *J. Inorg. Biochem.* **2010**, *104*, 512-522.

[143] 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.

[144] 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.

[145] 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. et al. Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy. *J. Phys. Chem. Lett.* **2019**, *10*, 6973-6982.

TOC GRAPHICS



BIOGRAPHIES (AND PHOTOGRAPHS)



Giovanni Macetti got his Philosophiae Doctor (Ph.D.) title in Chemistry in January 2019 at the University of Milan under the supervision of Prof. Leonardo Lo Presti and Dr. Carlo Gatti defending a Thesis entitled “Topological Descriptors Enabling Novel Dissections of Electron Position and Spin Properties in Complex Molecular Systems”. Since then, he has been working as a CNRS Post-Doctoral researcher at the Laboratory of Theoretical Physics and Chemistry (LPCT) of the University of Lorraine in Dr. Alessandro Genoni’s group. His current research activities mainly focus on the development and application of quantum chemistry methods for molecular and crystal systems.



Erna Katharina Wieduwilt earned her Master of Science degree in Chemistry from the University of Bremen (Germany) in 2018, where she worked on different projects in the field of quantum crystallography in the group directed by Dr. Simon Grabowsky, also spending a research period with Prof. Dylan Jayatilaka at the University of Western Australia. Since October 2018 she has been a Ph.D. student in Dr. Alessandro Genoni's group at the Laboratory of Theoretical Physics and Chemistry (LPCT) of the University of Lorraine. In her thesis, she is focusing on the development and application of quantum mechanics-based methods aiming at accurately refining crystal structures and at analyzing non-covalent interactions in macromolecules.



Alessandro Genoni earned his Ph.D. in Chemical Sciences in 2006 from the University of Milan (Italy) working under the supervision of Prof. Maurizio Sironi. After two postdoctoral experiences, one in Prof. Kenneth Merz's group at the Quantum Theory Project of the University of Florida (Gainesville, Florida, USA) and the other one in Dr. Giorgio Colombo's group at the ICRM Institute of the Italian CNR (Milan, Italy), in 2011 he joined the *Centre National de le Recherche Scientifique* (CNRS). He currently works as CNRS researcher at the Laboratory of Theoretical Physics and Chemistry (LPCT) of the University of Lorraine (Metz & Nancy, France) and his activity mainly focuses on the development of quantum chemistry methods for large systems and of new computational techniques of modern quantum crystallography (https://alessandrogenoni.weebly.com/).