Quantum Mechanics / Extremely Localized Molecular Orbital Embedding Technique: Theoretical Foundations and Further Validation

Giovanni Macetti, Alessandro Genoni

Quantum Mechanics / Extremely Localized Molecular Orbital Embedding Technique: Theoretical Foundations and Further Validation

Giovanni Macetti, Alessandro Genoni

Quantum Mechanics / Extremely Localized Molecular Orbital Embedding Technique: Theoretical Foundations and Further Validation

Giovanni Macetti[a], Alessandro Genoni[a,*]

[a] *Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, 57078 Metz, France*

Abstract

Embedding techniques are nowadays considered the go-to-methods to have an optimal trade-off between chemical accuracy and computational cost in modelling large molecular systems. Several efficient strategies of this kind have been developed over the years, from QM/MM (quantum mechanics/molecular mechanics) techniques to more recent and promising density functional theory embedding approaches. Along this line, we have recently proposed the QM/ELMO (quantum mechanics/extremely localized molecular orbital) method. This strategy describes the chemically important region of an extended system at a fully quantum mechanical level, while the rest is treated through frozen extremely localized molecular orbitals (ELMOs) properly transferred from recently assembled libraries or *ad hoc* model molecules. In this work, after reviewing the theoretical bases of the novel technique, we will show and discuss the results of further validation tests, with a particular focus on basis-set dependence and computational cost.

Keywords:
Embedding methods, QM/QM' strategies, Extremely Localized Molecular Orbitals (ELMOs), QM/ELMO technique, basis-set dependence

*Corresponding author
Email address: Alessandro.Genoni@univ-lorraine.fr (Alessandro Genoni)

1. Introduction

Nowadays, despite great and continuous advances in computer technologies, the development of fast and very accurate computational techniques for the description of large and complex molecular systems remains a crucial research field in quantum chemistry. For this reason, approximate methods characterized by an optimal compromise between chemical accuracy and computational effort have been developed over the years.

Within this context, a prominent place is certainly occupied by the embedding strategies [1, 2], namely methods in which the most important region of the investigated system (namely, the chemically active region) is treated at a (possibly high) quantum mechanical level of theory, while the rest is described through a lower level technique. Pioneering and archetypical examples belonging to this category are the QM/MM (quantum mechanics/molecular mechanics) approaches [3, 4, 5, 6, 7], where a quantum chemical strategy is used for the chemically crucial subunit and molecular mechanics for the remaining part. A similar technique is Morokuma's ONIOM method [8, 9], where the (macro)molecule under exam is generally subdivided into multiple fragments (typically two or three) that, if necessary, can be all treated quantum mechanically, thus giving rise to the first type of QM/QM' technique.

Pertaining to modern fully quantum mechanical embedding methods, it is also worth mentioning (i) the density matrix embedding techniques and (ii) the density functional theory embedding approaches. The strategies belonging to the former category introduce suitable quantum baths to mimic the effects of the environment [10, 11, 12, 13], while those in the latter group generally exploit density functional theory (DFT) as a starting point and have known a rapid and large development in the last twenty years [14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30].

In this theoretical framework, we have recently proposed the new QM/ELMO (quantum mechanics/extremely localized molecular orbital) method [31, 32, 33, 34], a fully quantum mechanical embedding approach based on extremely localized molecular orbitals (ELMOs) [35, 36, 37] and their intrinsic transferability. ELMOs are functions strictly localized on small molecular fragments (e.g, atoms, bonds or functional groups). Due to this absolute localization, they can be easily transferred from one molecule to another reliably reconstructing approximate wavefunctions and electron densities of large systems, as shown through several preliminary investigations [36, 37, 38, 39, 40] and

also through the use of the recently assembled ELMO libraries [41, 42, 43]. Therefore, exploiting this property and following the QM/MM philosophy, in the QM/ELMO strategy [31, 32, 33, 34] the chemically crucial region of the examined system is treated at fully quantum mechanical level, while the environment is described by means of frozen ELMOs transferred from the above-mentioned databanks or from proper model molecules.

Initially introduced at Hartree-Fock level [31], the QM/ELMO method has been afterwards extended to DFT, correlated post-HF techniques [32], and also to methods for the treatment of excited states [33]. In all cases, it has been shown that, by only including a limited number of atoms in the quantum mechanical subsystem, the QM/ELMO results agree with the corresponding fully QM ones within the limit of chemical accuracy. It has been also observed that the QM/ELMO computations are characterized by significant savings in terms of computational time, especially when post-Hartree-Fock approaches are taken into account [32, 33]. Finally, the QM/ELMO strategy has also been recently coupled with the emerging Hirshfeld atom refinement (HAR) technique [44, 45, 46] of quantum crystallography [47, 48, 49, 50, 51, 52] for the accurate determination of hydrogen atom positions also for crystals characterized by strong intermolecular interactions [53].

In this paper, in addition to reviewing the QM/ELMO theory (section 2), we will discuss the results of new test calculations (section 3) that mainly aimed at further assessing capabilities and limitations of our technique. In particular, we will focus on the basis-set dependence of the results (subsection 4.1) and on the analysis of the QM/ELMO computational cost (subsection 4.2). Section 5 will be dedicated to conclusions and future perspectives.

2. Theory

The first step of the QM/ELMO method consists in dividing the system under exam into two regions: (A) the quantum mechanical (QM) subsystem and (B) the ELMO subunit. In case of covalent boundaries between the two parts, the two subsystems share only the frontier atoms and, consequently, only the basis functions associated with those atomic centers. Afterwards, extremely localized molecular orbitals are exported to the ELMO subunit. In particular, the ELMOs that are transferred to the target system are molecular orbitals contained in the currently available ELMO libraries or resulting from *ad hoc* preliminary computations on suitable model molecules that mimic the system of interest.

After the ELMOs transfer and before starting the real QM/ELMO calculations, it is necessary to carry out preliminary orthogonalizations of the transferred ELMOs and of the atomic orbitals belonging to the QM region. The orthogonalization protocol consists in three distinct steps: 1) Löwdin orthonormalization of the exported ELMOs; 2) orthogonalization of the QM basis functions with respect to the orthonormalized ELMOs; 3) canonical orthogonalization of the QM basis functions resulting from the previous step. Now, if we indicate $\boldsymbol{\chi} = \Big[|\chi_1\rangle, |\chi_2\rangle, \ldots, |\chi_M\rangle\Big]$ as the initial $1 \times M$ array of the M non-orthogonal basis functions for the whole system, and $\boldsymbol{\chi'} = \Big[|\chi'_1\rangle, |\chi'_2\rangle, \ldots, |\chi'_{M_{QM}}\rangle\Big]$ as the $1 \times M_{QM}$ array of the final M_{QM} orthogonal basis functions of the only QM region (with $M_{QM} \ll M$), the above-described three-step procedure can be summarized through the following transformation:

$$\boldsymbol{\chi'} = \boldsymbol{\chi}\,\mathbf{B} \tag{1}$$

where $\mathbf{B}$ is a $M \times M_{QM}$ transformation matrix that, as we will see below, is extremely important in the QM/ELMO self-consistent field (SCF) computations. More specific details on the derivation of matrix $\mathbf{B}$ can be found in the original papers of the QM/ELMO approach [31, 32].

The orthogonalization procedure is followed by the QM/ELMO SCF algorithm that can be summarized in this way:

1. Computation of the $M \times M$ Fock matrix $\mathbf{F}$ in the starting non-orthogonal basis-set $\boldsymbol{\chi}$, for which the generic element $F_{\mu\nu}$ has this form:

$$\begin{aligned} F_{\mu\nu} &= \Big\langle \chi_\mu | \hat{h} | \chi_\nu \Big\rangle + \\ &+ \sum_{\lambda,\sigma=1}^{M} P^{QM}_{\lambda\sigma} \left[(\chi_\mu\,\chi_\nu \,|\, \chi_\sigma\,\chi_\lambda) - \frac{x}{2}\,(\chi_\mu\,\chi_\lambda \,|\, \chi_\sigma\,\chi_\nu) \right] + \\ &+ \sum_{\lambda,\sigma \in ELMO}^{M} P^{ELMO}_{\lambda\sigma} \left[(\chi_\mu\,\chi_\nu \,|\, \chi_\sigma\,\chi_\lambda) - \frac{x}{2}\,(\chi_\mu\,\chi_\lambda \,|\, \chi_\sigma\,\chi_\nu) \right] + \\ &+ \langle \chi_\mu | \hat{v}^{xc} \Big[\mathbf{P}^{QM} + \mathbf{P}^{ELMO} \Big] | \chi_\nu \rangle = \\ &= h_{\mu\nu} + F^{QM}_{\mu\nu} + F^{ELMO}_{\mu\nu} + v^{xc}_{\mu\nu} \end{aligned} \tag{2}$$

   where $\hat{h}$ is the core one-electron Hamiltonian operator, $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$ are the QM and ELMO one-electron density matrices in the original

basis-set $\boldsymbol{\chi}$, respectively, $(\chi_\alpha\,\chi_\beta\,|\,\chi_\gamma\,\chi_\delta)$ stands for a two-electron repulsion integral, x is the fraction of exact exchange and $\langle\chi_\mu|\hat{v}^{xc}\Big[\mathbf{P}^{QM}+\mathbf{P}^{ELMO}\Big]|\chi_\nu\rangle$ is the element $v^{xc}_{\mu\nu}$ of the matrix associated with the exchange-correlation potential. If the QM region is treated at Hartree-Fock level, x becomes equal to 1 and the exchange-correlation contribution disappears.

2. Transformation of the Fock matrix to the orthogonal basis-set $\boldsymbol{\chi'}$ of the QM subsystem through the relation

$$\mathbf{F'} = \mathbf{B}^\dagger\,\mathbf{F}\,\mathbf{B} \tag{3}$$

with $\mathbf{B}$ as the matrix obtained and used in the preliminary orthogonalization procedure (see Eq. (1)).

3. Diagonalization of the Fock matrix $\mathbf{F'}$ of the QM region:

$$\mathbf{F'}\,\mathbf{C'} = \mathbf{C'}\,\mathbf{E'} \tag{4}$$

4. Transformation of the molecular orbitals coefficients obtained at step 3 to the original basis-set $\boldsymbol{\chi}$:

$$\mathbf{C} = \mathbf{B}\,\mathbf{C'} \tag{5}$$

5. Computation of the QM density matrix $\mathbf{P}^{QM}$ exploiting matrix $\mathbf{C}$ obtained at the previous step.
6. Convergence-test on energy and density matrix: if convergence is achieved, the SCF procedure stops; otherwise the cycle restarts with step 1, where the new density matrix for the QM region $\mathbf{P}^{QM}$ is used to update the Fock matrix $\mathbf{F}$ in the original basis-set.

From the analysis of Eq. (2), it is easy to observe that only two terms have to be updated during the QM/ELMO SCF cycle: $F^{QM}_{\mu\nu}$ and $v^{xc}_{\mu\nu}$, with the latter taken into account only in case of DFT/ELMO calculations. At the same time, the contributions $h_{\mu\nu}$ and $F^{ELMO}_{\mu\nu}$ remain constant throughout the SCF iterations and can be evaluated only once before starting the cycle, thus entailing an important reduction in terms of computational cost. Nevertheless, it is worth pointing out again that the starting Fock matrix $\mathbf{F}$ is initially computed in the supermolecular basis-set $\boldsymbol{\chi}$ and, for this reason, step 1 is the current rate-limiting step of the QM/ELMO SCF algorithm.

Suitable criteria to reduce the number of atomic orbitals over which initially evaluating the Fock matrix are currently under investigation.

The above-described SCF procedure is preparatory to perform post-HF/ELMO computations. In this regard, note that, for a $2N$-electron closed-shell QM region, the diagonalization of the $M_{QM} \times M_{QM}$ Fock matrix $\mathbf{F}'$ provides N occupied (with $N \ll N + N_{ELMO}$) and $M_{QM} - N$ virtual (with $M_{QM} - N \ll M - N$) molecular orbitals. As we will see in subsection 4.2, these reduced sets of occupied and, above all, of virtual orbitals allow significant reductions of the computational cost associated with the post-HF/ELMO calculations.

3. Computational Details

In order to further validate the QM/ELMO method and to study the basis-set dependence, we considered (i) the Diels-Alder reaction between 1-octene and butadiene (see Figure 1A) and (ii) the S_N2 nucleophilic substitution reaction of chlorine in 1-chlorooctane by the fluoride anion (see Figure 1B). In both cases, the geometries of reactants and products were initially optimized at B3LYP/cc-pVDZ level and the transition-states were afterwards determined exploiting the synchronous transit-guided quasi-Newton method. All these preliminary calculations were performed through the *Gaussian09* quantum chemistry package [55].

Using the previously obtained geometries of reactants, products and transition states, we then carried out fully QM and QM/ELMO calculations. For the Diels-Alder reaction we performed CCSD (Coupled Cluster with single and double substitutions) and CCSD/ELMO computations. For the nucleophilic substitution, we carried out CCSD(T) (Coupled Cluster with single and double substitutions plus perturbative triples) and CCSD(T)/ELMO calculations. Since the main goal of these tests was to assess the dependence of the results on the adopted basis-sets, all the above mentioned computations were performed using different sets of basis functions. To evaluate the dependence on correlation-consistent basis-sets, we considered the cc-pVDZ, aug-cc-pVDZ, cc-pVTZ and mixed cc-pVTZ/cc-pVDZ (the former and the latter basis-sets for the QM and ELMO regions, respectively) sets of functions. To assess the dependence on Pople-style basis-sets, we used the 6-31G(d,p), 6-31++G(d,p), 6-311G(d,p) and mixed 6-311G(d,p)/6-31G(d,p) sets of functions.

For each basis-set, the results of the fully QM calculations were used as benchmark values for the corresponding QM/ELMO computations, which

were carried out by gradually increasing the size of the QM subsystem to also investigate the influence of the ELMO embedding region. For the Diels-Alder reaction, in the QM regions we considered from 1 to 4 alkyl groups along with the terminal carbon-carbon double bond in 1-octene and the whole butadiene molecule. Concerning the nucleophilic substitution, the QM subsystems comprised from 2 to 5 CH_2 moieties together with the chlorine and fluorine atoms. The following quantities were compared: (i) the reaction barriers $\Delta E_{act} = E_{TS} - E_{reactants}$ (where TS stands for transition state), and (ii) the global reaction energy $\Delta E_r = E_{products} - E_{reactants}$. For the sake of precision, when QM/ELMO computations were performed with mixed basis-sets (cc-pVTZ/cc-pVDZ or 6-311G(d,p)/6-31G(d,p)), the corresponding benchmark values were those resulting from standard fully QM calculations entirely performed with the basis-set for the QM region (i.e., cc-pVTZ and 6-311G(d,p)).

The ELMOs used in the QM/ELMO calculations were previously computed for all the considered basis-sets on the butane molecule (geometry preliminarily optimized at B3LYP/cc-pVDZ level) by exploiting a modified version of the *GAMESS-UK* quantum chemical software [54], where the Stoll equations [35] were implemented [36]. The obtained extremely localized molecular orbitals were transferred to the target molecules according to the Philipp and Friesner rotation strategy [41, 56] through the *ELMOdb* program associated with the ELMO databanks [43]. Finally, all the QM/ELMO computations were carried out using an in-house variant of *Gaussian09* [55], where the QM/ELMO approach was coded.

4. Results and Discussion

4.1. Basis-set dependence

The results obtained for the Diels-Alder reaction are shown and summarized in Figure 2. Let us consider the outcomes of the calculations performed with the correlation-consistent basis-sets. Concerning the reaction barriers (see Figure 2A), the CCSD/ELMO calculations provided results that are in optimal agreement with the corresponding full CCSD ones regardless of the considered set of basis functions, with discrepancies that start being lower than the chemical accuracy threshold (1 kcal/mol) when only two alkyl groups of 1-octene are included in the quantum mechanical subsystem. Furthermore, the computations with the cc-pVDZ, cc-pVTZ and mixed cc-pVTZ/cc-pVDZ basis-sets gave very similar trends as a function of the QM region size. A slightly slower convergence is observed for the calculations

with the aug-cc-pVDZ set of basis functions, which can certainly be ascribed to the larger extent of electronic structure delocalization when diffuse basis functions are introduced in the computations. Anyway, for all the basis-sets, the situation improves when the size of the QM subsystem increases, with deviations that are always lower than 0.3 kcal/mol when three CH_2 moieties are treated quantum mechanically.

Analogous trends can be also noticed for the overall reaction energies (see Figure 2B), although larger discrepancies were almost always obtained compared to the case of the above-analyzed reaction barriers when identical QM regions were used. We can see again that the cc-pVDZ, cc-pVTZ and cc-pVTZ/cc-pVDZ basis-sets provided very similar results, with deviations that drop below the chemical accuracy limit when only two alkyl groups of 1-octene are considered in the chemically crucial subsystem. Convergence is again slower for the CCSD/ELMO computations performed with the aug-cc-pVDZ set of basis functions. In fact, the discrepancy is still greater than 1.0 kcal/mol when two CH_2 moieties are included in the quantum chemical region (1.3 kcal/mol), while it starts being lower than 1.0 kcal/mol when three alkyl groups are treated quantum mechanically (0.47 kcal/mol). However, also in this case, for all the basis-sets the description progressively improves as larger and larger QM subunits are employed in the embedding calculations.

Let us now analyze the results obtained through the Pople-style basis-sets. Pertaining to the reaction barriers (see Figure 2C), irrespective of the chosen set of basis functions, the observed $\Delta\Delta E_{act}$ values drop below the chemical accuracy limit already when two CH_2 groups are included in the quantum chemical subsystem. Moreover, as observed for the correlation-consistent basis-sets, the QM/ELMO calculations without diffuse functions provided smaller deviations. For example, when three alkyl groups were described in a fully quantum chemical way, the CCSD/ELMO computations with basis-sets 6-31G(d,p), 6-311G(d,p) and 6-311G(d,p)/6-31G(d,p) gave discrepancies lower than 0.18 kcal/mol, while the CCSD/ELMO calculations with diffuse basis functions (6-31++G(d,p) basis-set) provided a deviation of 0.29 kcal/mol. Completely similar trends can be also noticed for the overall reaction energies resulting from the computations with the Pople basis-sets (see Figure 2D). Furthermore, unlike what we observed for the correlation-consistent sets of basis functions (see Figure 2B), in this case chemical accuracy was always achieved with only two alkyl moieties in the quantum mechanical region, also when the 6-31++G(d,p) basis-set was adopted.

The outcomes of the test calculations for the S_N2 reaction are reported in

Figure 3. Concerning the barriers obtained through the use of the correlation-consistent basis-sets (see Figure 3A), for the cc-pVDZ, aug-cc-pVDZ and cc-pVTZ sets of basis functions, the deviations drop below 1.0 kcal/mol (in absolute value) when the third CH_2 group of the alkyl chain is included in the QM subsystem. If only two alkyl moieties are treated quantum mechanically, the discrepancies are much larger than the chemically accuracy limit. This can be rationalized by inspecting the geometries of reactants and transition state, where a non-negligible interaction can be noticed between the fluoride anion and one of the hydrogen atoms of the third CH_2 group in the alkyl chain (see and compare the distances reported in Figure 1B). The CCSD(T)/ELMO computations with mixed basis-set cc-pVTZ/cc-pVDZ provided a slightly worse performance, with chemical accuracy that is achieved only when the QM subsystem comprises at least four alkyl moieties. Anyway, for all the different sets of basis functions, the description generally improves as larger and larger QM regions are considered in the CCSD(T)/ELMO calculations.

We obtained analogous trends for the overall reaction energies (see Figure 3B). In fact, chemical accuracy is reached when at least three alkyl groups are treated in a fully quantum mechanical way. However, in this case, this is also true for the computations carried out with mixed basis-set cc-pVTZ/cc-pVDZ. Finally, if we compare the deviations obtained with the cc-pVDZ set of basis functions to those obtained with the aug-cc-pVDZ one, the latter are in absolute value systematically smaller than the former. This can be explained with the fact that in both reactants and products we have an anion (the fluoride anion in the reactants and the chloride anion in the products), which is clearly better described when diffuse basis functions are used.

Pertaining to the Pople basis-sets, for both reaction barriers (see Figure 3C) and overall reaction energies (see Figure 3D), two alkyl groups in the QM region are not enough to reach the desired chemical accuracy limit. In fact, in all cases in which two CH_2 subunits are treated quantum mechanically, the discrepancies from the reference fully QM values are quite large. For basis-sets 6-31G(d,p), 6-311G(d,p) and 6-31++G(d,p), the deviations start being lower than 1.0 kcal/mol in absolute value when the quantum chemical subsystem comprises at least three alkyl moieties, while four CH_2 groups are necessary for the mixed set of basis functions 6-311G(d,p)/6-31G(d,p). Finally, for the global reaction energy ΔE_r, the QM/ELMO calculations performed with the basis-set including diffuse functions (i.e., basis-set 6-31++G(d,p)) provided smaller absolute discrepancies compared to the computations carried

out with the corresponding basis-set without diffuse functions (i.e., basis-set 6-311G(d,p)).

For the sake of completeness, it is worth mentioning that the deviations from the standard fully QM calculations observed in our test computations are probably imputable to two factors: i) the lack of polarization of the environment described through the frozen ELMOs and ii) the lack of electron correlation between the QM and ELMO regions. The former problem might be solved by introducing a technique to polarize the ELMOs after the transfer, while the second is a source of error inherent to our method and can be partially limited by moderately increasing the size of the QM region.

4.2. Computational Cost

In Table 1 we assembled the number of active occupied molecular orbitals (N_{occ}), the number of virtual molecular orbitals (N_{virt}) and the computational cost (referred to that of the corresponding, standard fully QM computation) for the CCSD/ELMO calculations performed on the transition state of the investigated Diels-Alder reaction, and for the CCSD(T)/ELMO computations carried out on the products configuration of the analyzed nucleophilic substitution. In both cases we considered the cc-pVTZ basis-set, which was the largest set of basis functions used in our study.

Note that the number of occupied and, above all, of virtual molecular orbitals used in the QM/ELMO calculations are significantly lower compared to those associated with the corresponding standard QM computations. This is particularly true when the size of the quantum mechanical subsystem remains quite small. As anticipated in the Theory section, the reduction in the number of virtual molecular orbitals impacts on the overall computational cost. Concerning the Diels-Alder reaction, if we consider the QM subsystem comprising only two alkyl groups (which was enough to reach chemical accuracy), the recorded time is about 16% of the one for the full calculation. An even more advantageous decrease of the computational effort is observed for the nucleophilic substitution, for which the CPU time corresponding to the QM/ELMO calculation with three CH_2 moieties in the QM region (which was again enough to achieve chemical accuracy) is only about 3% of the one associated with the standard CCSD(T) computation. Although the computational cost increases with the size of the quantum mechanical subunit, it generally remains quite small also when the dimensions of the QM regions approach those of the corresponding full systems.

5. Conclusions and Perspectives

In this work, we have illustrated the recently developed multiscale embedding approach QM/ELMO. In addition to briefly going through its theoretical bases, we have shown and discussed the results of new computations that were mainly performed to further assess the capabilities of the technique as a function of the QM region size and to evaluate the dependence of the results on the adopted basis-sets. The tests have confirmed that, in all cases and for all the considered sets of basis functions, the inclusion of only a small number of atoms in the QM region is generally enough to reproduce the outcomes of corresponding and much more computationally expensive fully quantum mechanical calculations. Moreover, all the considered basis-sets provided results that are more or less identical. However, it seems that adopting a mixed basis-set for the QM and ELMO regions is not significantly advantageous, while the use of diffuse basis functions is convenient when anionic species are involved in the computations.

Despite the promising results obtained through the current version of the QM/ELMO technique, algorithmic and methodological improvements are also possible. The next step will consist in defining a reliable criterion to truncate the basis-set over which the starting Fock matrix of the QM/ELMO SCF cycle is constructed. This approximation will further speed up the calculations at Hartree-Fock and DFT levels. Finally, since ELMOs are transferred to the ELMO regions and remain frozen during the QM/ELMO calculations, we envisage to develop a polarizable QM/ELMO approach in which the transferred extremely localized molecular orbitals relax in response to the electron density of the quantum mechanical subsystem. This will be useful not only for ground state calculations, but also to account for polarization response effects in QM/ELMO computations of excited states.

6. Acknowledgements

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

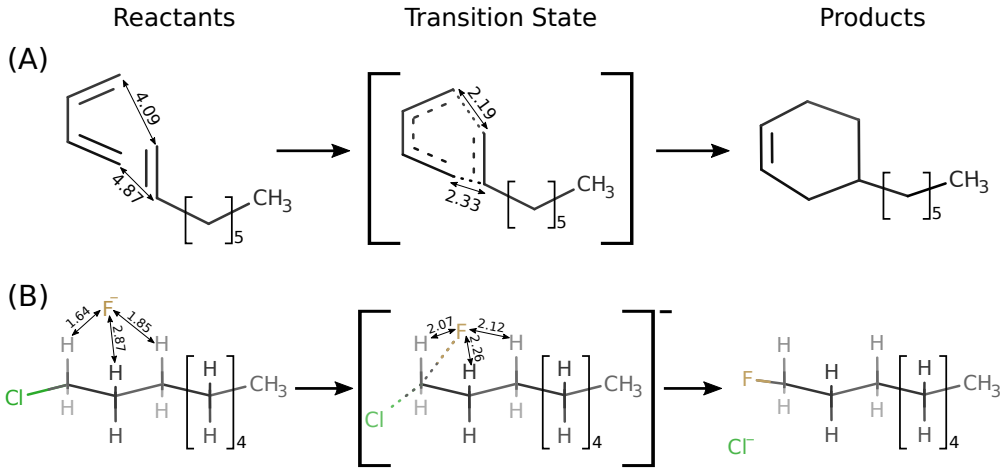

Figure 1: Schematic representation of the reactions investigated in this work: (A) reactants, transition state and products for the Diels-Alder reaction between butadiene and 1-octene; (B) reactants, transition state and products for the S_N2 reaction between 1-chlorooctane and the fluoride anion. All the reported interatomic distances are expressed in Å.

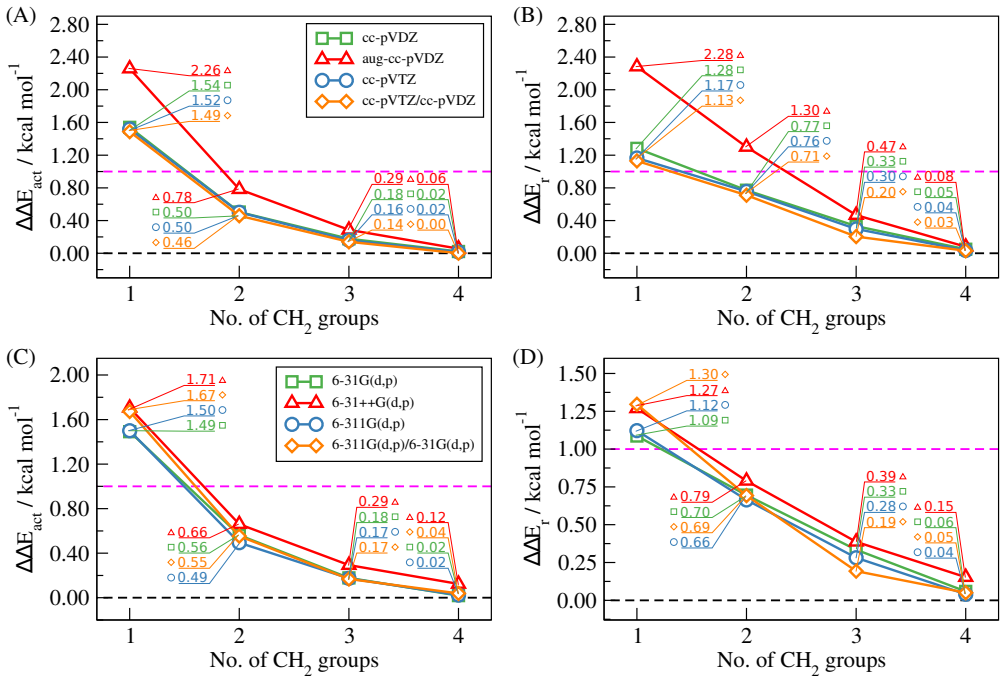

Figure 2: Diels-Alder reaction between butadiene and 1-octene: QM/ELMO-QM discrepancies of the reaction barriers ΔE_{act} (A and C) and of the reaction energies ΔE_r (B and D) computed at CCSD/ELMO level from those resulting from standard CCSD calculations ((A) and (B): results obtained with correlation-consistent basis-sets; (C) and (D): results obtained with Pople-style basis-sets). The variation of the absolute deviations is shown as a function of the number of CH_2 groups of the alkyl chain that are included in the QM region. The magenta-dashed line indicates the chemical accuracy limit.

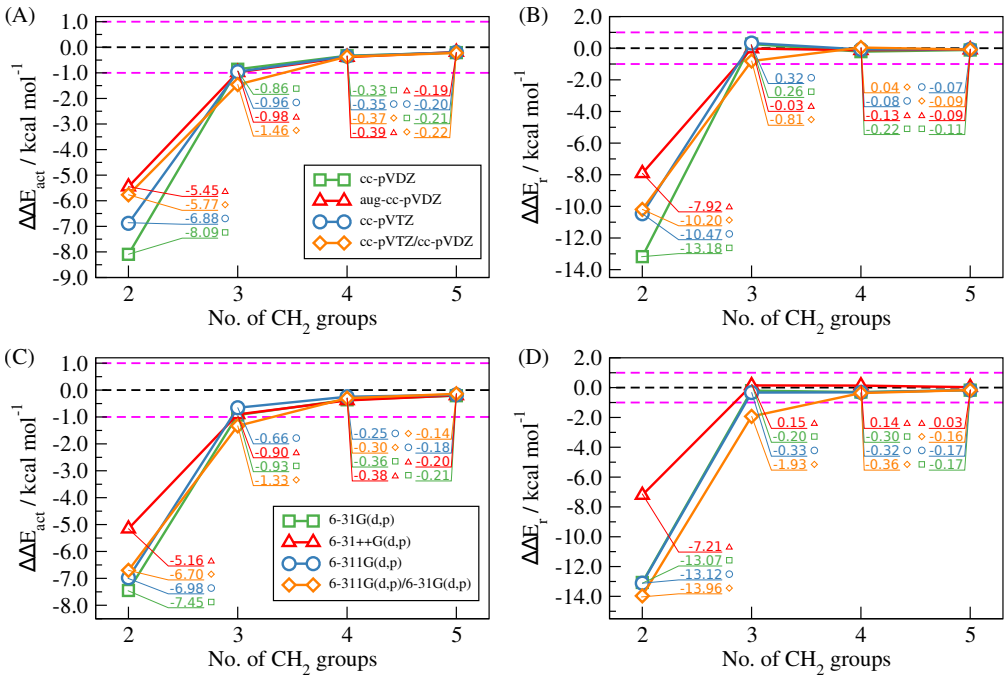

Figure 3: S_N2 reaction between 1-chlorooctane and the fluoride anion: QM/ELMO-QM discrepancies of the reaction barriers ΔE_{act} (A and C) and of the reaction energies ΔE_r (B and D) computed at CCSD(T)/ELMO level from those resulting from standard CCSD(T) calculations ((A) and (B): results obtained with correlation-consistent basis-sets; (C) and (D): results obtained with Pople-style basis-sets). The variation of the absolute deviations is shown as a function of the number of CH_2 groups of the alkyl chain that are included in the QM region. The magenta-dashed lines indicate the chemical accuracy limits.

Table 1: Number of active occupied molecular orbitals (N_{occ}), number of virtual molecular orbitals (N_{virt}) and computational cost for i) the CCSD/ELMO and CCSD calculations (cc-pVTZ basis-set) performed on the transition-state geometry for the Diels-Alder reaction between butadiene and 1-octene, and ii) the CCSD(T)/ELMO and CCSD(T) computations (cc-pVTZ basis-set) carried out on the products configuration of the S_N2 reaction between 1-chlorooctane and the fluroide anion.

Calculations[a]	N_{occ}	N_{virt}	CPU time (%)[b,c]
Diels-Alder reaction			
QM(1)/ELMO	19	384	7.72
QM(2)/ELMO	22	445	15.98
QM(3)/ELMO	25	506	17.15
QM(4)/ELMO	28	567	24.93
S_N2 reaction			
QM(2)/ELMO	13	183	1.71
QM(3)/ELMO	16	244	2.72
QM(4)/ELMO	19	305	5.84
QM(5)/ELMO	22	366	10.55

[a] The acronym QM(N)/ELMO indicates that N alkyl groups were included in the QM region for the CCSD/ELMO (Diels-Alder reaction) or CCSD(T)/ELMO (S_N2 reaction) calculations.

[b] Percentage referred to the CPU time associated with the corresponding, standard fully QM calculation.

[c] All the recorded timings were obtained by performing parallel computations on 8 Intel Xeon Gold 6130 2.1 GHz processors.

References

[1] Gordon, M. S.; Slipchenko L. V. Introduction: Calculations on Large Systems. *Chem. Rev.* **2015**, *12*, 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] 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.

[4] 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.

[5] 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.

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

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

[8] 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 DielsAlder Reactions and $Pt(P(t\text{-}Bu)_3)_2$ + H_2 Oxidative Addition. *J. Phys. Chem.* **1996**, *50*, 19357-19363.

[9] 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.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, *115*, 5678-5796.

[10] Knizia, G.; Chan, G. K.-L. Density Matrix Embedding: A Strong-Coupling Quantum Embedding Theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428-1432.

[11] Bulik, I. W.; Chen, W.; Scuseria, G. E. Electron correlation in solids via density embedding theory. *J. Chem. Phys.* **2014**, *141*, 054113.

[12] Fornace, M. E.; Lee, J.; Miyamoto, K.; Manby, F. R.; Miller, III, T. F. Embedded Mean-Field Theory. *J. Chem. Theory Comput.* **2015**, *11*, 568-580.

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

[14] Cortona, P. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B* **1991**, *44*, 18454-8458.

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

[16] Wesołowski, T. A. Embedding a multideterminantal wave function in an orbital-free environment. *Phys. Rev. A* **2008**, *77*, 012504.

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

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

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

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

[21] Elliott, P.; Cohen, M. H.; Wasserman, A.; Burke, K. Density Functional Partition Theory with Fractional Occupations. *J. Chem. Theory Comput.* **2009**, *5*, 827-833.

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

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

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

[25] Manby, F. R.; Stella, M.; Goodpaster, J. D.; Miller, III, T. F. A Simple, Exact Density-Functional-Theory Embedding Scheme. *J. Chem. Theory Comput.* **2012**, *8*, 2564-2568.

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

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

[28] 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.

[29] 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.

[30] Marrazzini, G.; Giovannini, T.; Scavino, M.; Egidi, F.; Cappelli, C.; Koch, H. Multilevel Density Functional Theory. *J. Chem. Theory Comput.* **2021**, *17*, 791-803.

[31] 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.

[32] 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.

[33] 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.

[34] Macetti, G.; Wieduwilt, E. K.; Genoni, A. QM/ELMO: a Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals. *J. Phys. Chem. A* **2021**, DOI: 10.1021/acs.jpca.0c11450.

[35] Stoll, H.; Wagenblast, G.; Preuβ, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theoret. Chim. Acta* **1980**, *57*, 169-178.

[36] 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.

[37] Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely localized molecular orbitals: theory and applications. *Theor. Chem. Acc.* **2007**, *117*, 685-698.

[38] 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.

[39] Genoni, A.; Merz, Jr., K. M.; Sironi, M. A Hylleraas Functional Based Perturbative Technique to Relax the Extremely Localized Molecular Orbital Wavefunction. *J. Chem. Phys.* **2008**, *129*, 054101.

[40] 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.

[41] 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.

[42] 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.

[43] 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.

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

[45] Capelli, S. C.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld atom refinement. *IUCrJ* **2014**, *1*, 361-379.

[46] Kleemiss, F.; Dolomanov, O. V.; Bodensteiner, M.; Peyerimhoff, N.; Midgley, L.; Bourhis, L. J.; Genoni, A.; Malaspina, L. A.; Jayatilaka, D.; Spencer, J. L.; White, F.; Grundkötter-Stock, B.; Steinhauer, S.; Lentz, D.; Puschmann, H.; Grabowsky, S. Accurate crystal structures and chemical properties from NoSpherA2. *Chem. Sci.* **2021**, *12*, 1675-1692.

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

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

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

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

[51] 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., Raithby, P. R., Eds.; Springer: Berlin & Hidelberg, Germany, 2020; Vol. 186; pp. 65-144.

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

[53] 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 Atom Positions. *J. Phys. Chem. Lett.* **2021**, *12*, 463-471.

[54] Guest, M. F.; Bush, I. J.; Van Dam, H. J. J.; Sherwood, P.; Thomas, J. M. H.; Van Lenthe, J. H.; Havenith, R. W. A.; Kendrick, J. The GAMESS-UK electronic structure package: algorithms, developments and applications. *Mol. Phys.* **2005**, *103*, 719-747.

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

[56] 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.