

Localized Molecular Orbital-Based Embedding Scheme for Correlated Methods

Giovanni Macetti^{(1)*}, Erna K. Wieduwilt⁽¹⁾, Xavier Assfeld⁽²⁾, 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.

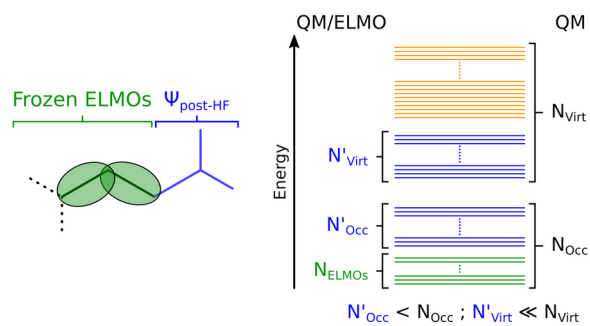
(2) Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, Boulevard des Aiguillettes, BP 70239, F-54506 Vandoeuvre-lès-Nancy, France.

ABSTRACT

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

Embedding strategies currently provide the best compromise between accuracy and computational cost in modeling chemical properties and processes of large and complex systems. In this framework, different methods have been proposed all over the years, from the very popular QM/MM approaches to the more recent and very promising density matrix and density functional embedding techniques. In this paper, we present a further development of the quantum mechanics/extremely localized molecular orbital technique (QM/ELMO) method, a recently proposed multi-scale embedding strategy in which the chemically active region of the investigated system is treated at fully quantum mechanical level, while the rest is described by frozen extremely localized molecular orbitals previously transferred from proper libraries or tailor-made model molecules. In particular, in this work we discuss and assess in detail the extension of the QM/ELMO approach to density functional theory and post Hartree-Fock techniques by evaluating its performances when it is used to describe chemical reactions, bond dissociations and intermolecular interactions. The preliminary test calculations have shown that, in the investigated cases, the new embedding strategy enables to reproduce the results of the corresponding fully quantum mechanical computations within chemical accuracy in almost all the cases, but with a significantly reduced computational cost, especially when correlated post Hartree-Fock strategies are used to describe the quantum mechanical subsystem. In light of the obtained results, we already envisage the future application of the new correlated QM/ELMO techniques to the investigation of more challenging problems, such as the modeling of enzyme catalysis, the study of excited states of biomolecules, and the refinement of macromolecular X-ray crystal structures.

TOC GRAPHICS



KEYWORDS. Embedding methods, QM/QM' techniques, extremely localized molecular orbitals, Local Self-Consistent Field approach.

I. INTRODUCTION

Nowadays the development of accurate methods to model chemical processes in large and complex systems remains one of the main goals and challenges of electronic structure theory. To achieve chemical accuracy (< 1 kcal/mol), the correlated methods of quantum chemistry¹⁻⁴ would be ideal, but their prohibitive computational scaling (M^n , with $n \geq 5$ and M the number of basis functions used in the calculation) prevents their application to very large molecules. A possible way to partially circumvent the problem is offered by Kohn-Sham density functional theory (KS-DFT),^{5,6} which is characterized by a still-affordable computational cost (M^3 and successfully works in the investigation of weakly correlated systems. On the other hand, the use of approximate and non-systematically improvable exchange-correlation functionals has also sometimes led to some failures in the calculations, such as the underestimations of barriers in chemical reactions, of band gaps of materials and of charge transfer excitation energies, but also the overestimations of binding energies for charge transfer complexes and of responses to external electric fields.⁷⁻⁹

Therefore, in this context, a good compromise between chemical accuracy and computational cost is offered by the so-called multi-scale embedding strategies, which rely on the simple and general observation that many chemical processes are characterized by local changes, namely variations that occur within very localized subsystems. On the basis of this fact, in the multi-scale embedding strategies, the system under investigation is usually partitioned in at least two regions: i) the active region, which is crucial for the problem of interest and which is treated through a high-level quantum chemical method, and ii) the surrounding environment, which embeds the chemically active subsystem and is described by means of a lower-level technique. Of course, the success of the embedding strategies naturally implies some chemical knowledge of the system under exam, which is associated with the proper

choice of the active region, but it also strongly requires both a suitable description of the environment and the use of a proper coupling between the regions into which the system has been initially partitioned.

Different multi-scale embedding methods have been proposed over the years. The simplest and most popular example is certainly represented by the quantum mechanics/molecular mechanics (QM/MM) strategies,¹⁰⁻¹⁴ which were originally conceived to investigate systems of biological interest and in which the (bio)chemically active region (e.g., the active site of a protein) is described at quantum mechanical level, while the remaining part is treated through a classical molecular mechanics force field. Nowadays, QM/MM techniques are so widely used that they became not only cornerstone approaches of theoretical chemistry, but also universally recognized methods in science, as testified by the recent 2013 Nobel Prize in Chemistry awarded for the development and contributions on this topic. Another embedding approach similar to the QM/MM techniques is the well-known and established ONIOM method originally introduced by Morokuma and coworkers.¹⁵⁻¹⁸ It consists in subdividing the systems of interest into multiple layers that are afterwards treated using different levels of theory to obtain reliable energies and geometries with a significantly lower computational effort. In this case, the layers may be even treated at different quantum mechanical levels and, for this reason, ONIOM may be also considered as the first example of QM/QM' embedding technique.

Considering fully quantum mechanical embedding methods, different approaches have been devised in the last twenty years. In this group we can even mention the fragmentation methods^{19,20} or strategies as the Divide & Conquer-based technique devised by Merz and coworkers to study large molecular systems, where a composite DFT/semiempirical Hamiltonian with a heterogeneous basis-set (Gaussian/Slater basis

functions) was exploited.²¹ However, in the context of fully quantum mechanical embedding, a prominent place is also occupied by the so-called density matrix²²⁻²⁸ and density functional embedding strategies.²⁹⁻⁵⁸ Among the former group of methods we have the density matrix embedding theory (DMET),^{22,23} where the entanglement of a fragment with the rest of the system is not described through link atoms or boundary regions, but through a rigorously designed quantum bath that implicitly takes into account the full environment. In this context another interesting development is also the bootstrap embedding (BE) strategy,^{27,28} which, unlike DMET, overcomes the problem of rigid partitioning by allowing the definition of overlapping fragments and, for this reason, enables to significantly accelerate the rate of convergence towards the exact solution as the fragments size increases. Moreover, the introduction of the recent atom-based fragment definition allowed the partitioning of molecular systems into atomic and bond subunits, thus improving the performances of the BE method when applied to molecules.²⁸ Concerning density functional embedding techniques, an example is the promising projection-based method jointly developed by the Manby and Miller research groups.⁴⁹⁻⁵⁵ In this technique the problems associated with the evaluation of the non-additive kinetic potential (NAKP) contributions, which usually emerge in DFT embedding strategies, are avoided through the introduction of a level shifting projection operator that keeps the orbitals of one subsystem orthogonal to those belonging to the other one, thus automatically enforcing the Pauli exclusion between the electrons of the various subsystems. Starting from this embedding framework, which is exact in case both subsystems are described at the very same DFT level (i.e., with the same exchange-correlation functional), the projection-based approach has been used to perform wavefunction-in-density functional theory (WF-in-DFT) embedding calculations, in which the active region of the system under exam is

treated with correlated wavefunctions (e.g., Coupled Cluster or MP2 wavefunctions), while the remaining part is described through DFT. The obtained results indicated that the strategy allows to significantly reduce the computational cost of the correlated computations on the full systems without quantitatively affecting the accuracy of the results for a large variety of chemical problems, such as, transition-metal catalysis, enzyme reactivity and battery electrolyte decomposition.⁵⁵

The recently developed QM/ELMO method,⁵⁹ which is the main topic of this paper, fully belongs to the category of the multi-scale quantum mechanical embedding approaches mentioned above. The novel technique is strongly related to the local self-consistent field (LSCF) approach successfully employed in QM/MM calculations.⁶⁰⁻⁶³ So far, its original version was only able to perform Hartree-Fock (HF) computations on active molecular regions embedded by frozen extremely localized molecular orbitals (ELMOs) that describe the remaining parts of the systems under examination. In fact, ELMOs are molecular orbitals strictly localized on small molecular units⁶⁴⁻⁷⁰ (e.g., atoms, bonds or functional groups) and, due to their extreme localization, they can be easily exported from model to target molecules in order to reconstruct approximate wavefunctions and electron densities of large systems.^{71,72} Exploiting this interesting property, ELMO libraries have been constructed and currently comprise molecular orbitals for all the possible elementary fragments of the twenty natural amino acids,⁷³ thus enabling to obtain approximate wavefunctions and electron distributions of proteins and polypeptides with a significantly reduced computational cost (see Supporting Information for more details about the ELMO theory and the ELMO libraries).

The main goal of this work is to present the extension of the QM/ELMO embedding method to density functional theory and post-HF techniques. To accomplish this task,

other than providing all the necessary theoretical details, we will show the preliminary assessment of the QM/ELMO strategy when applied to some important chemical aspects and processes, such as intermolecular interactions and, above all, the more challenging breakings and formations of chemical bonds.

II. QM/ELMO THEORY

In the first part of this section we will critically review the QM/ELMO self-consistent field (SCF) algorithm, which is the procedure followed during the QM/ELMO calculations when the quantum mechanical region is treated at Hartree-Fock or DFT level. Here, compared to the original paper on the Hartree-Fock/ELMO technique,⁵⁹ the equations will be slightly modified to also account for the extension of the approach to density functional theory. Finally, in the second subsection, we will discuss the extension of the QM/ELMO strategy to multi-determinant post-HF methods, mainly pointing out the computational advantages intrinsically offered in these cases by the QM/ELMO embedding scheme.

A. The QM/ELMO SCF algorithm

The self-consistent field algorithm for the QM/ELMO method (with the QM part treated at Hartree-Fock or DFT level) consists in the following steps.

- 1. Definition of the QM and ELMO regions.* Before starting the calculations, the system is subdivided into two quantum subunits (i.e., the QM and the ELMO regions) that, in case of covalently bonded systems, share only the frontier atoms and, in terms of basis-set, only the basis functions centered on those atoms.
- 2. Transfer of the ELMOs to the ELMO subsystem.* To describe the ELMO region, extremely localized molecular orbitals available in the current ELMO libraries or

obtained from tailor-made computations on suitable model molecules (as in this work, see the Computational Details section) are transferred only to the ELMO subsystem and are kept frozen during the calculations. Here it is worth pointing out a key difference between the original LSCF QM/MM and the new QM/ELMO approaches. While in the former the strictly localized bond orbitals describing the frontier region are formally expanded on all the atomic orbitals of the QM subunit, in the latter the transferred ELMOs are expanded on the basis functions of the ELMO subsystem, thus sharing with the basis-set of the QM part only the atomic orbitals centered on the frontier atoms. Although the previous aspect may seem only a slight difference, as we will see below, it is actually a fundamental choice to avoid linear-dependency problems that affect the LSCF approach and that would not allow to apply the QM/ELMO strategy to investigate large molecular systems.

3. Orthonormalization of the transferred ELMOs. The transferred ELMOs are mutually orthonormalized using the traditional Löwdin method that largely preserves the localized nature of the original orbitals. Therefore, the transformed ELMOs are characterized by small orthogonalization tails on the atoms that are very close to the fragments on which the starting orbitals were strictly localized. In this regards it is possible to see that, considering the square of the 2-norm (i.e., the square of the Euclidean norm) for each transformed orbital, the basis functions that contribute the most are still those corresponding to the atoms on which the ELMOs were initially absolutely localized (~98%), while the basis functions centered on the very neighboring atoms contribute only for a small percentage (~2%). For this reason, only few orthonormalized ELMOs slightly delocalize on the atoms at the frontier with the QM region and, consequently, this further reduces the possibility of incurring linear-dependency problems.

4. *Orthogonalization of the QM basis-set.* In the QM/ELMO method, the actual quantum mechanical calculation on the QM subsystem is performed on a new set of orthogonal M_{QM} basis functions χ' (with χ' corresponding to the $1 \times M_{QM}$ array $[\left|\chi'_1\right\rangle, \left|\chi'_2\right\rangle, \dots, \left|\chi'_{M_{QM}}\right\rangle]$) obtained from the original and complete set of M (with $M \gg M_{QM}$) non-orthogonal basis functions for the system under examination, which will be hereinafter indicated as χ (with χ corresponding to the $1 \times M$ array $[\left|\chi_1\right\rangle, \left|\chi_2\right\rangle, \dots, \left|\chi_M\right\rangle]$). It is worth noting that M_{QM} corresponds to the number of basis functions originally centered on the atoms of the QM region and it automatically results from the fact that χ' is obtained through the following transformation:

$$\chi' = \chi B \quad (1)$$

where \mathbf{B} is an $M \times M_{QM}$ transformation matrix, which can be expressed as

$$B = T Q \quad (2),$$

namely, as the product of two matrices, each of them corresponding to an intermediate step in the orthogonalization procedure.

T is the $M \times M_{QM}$ matrix that brings from original basis-set χ to the set of the M_{QM} basis functions $\tilde{\chi}$ (with $\tilde{\chi} = [\left|\tilde{\chi}_1\right\rangle, \left|\tilde{\chi}_2\right\rangle, \dots, \left|\tilde{\chi}_{M_{QM}}\right\rangle]$) and corresponds to projecting out the Löwdin orthonormalized ELMOs (see step 3 of the algorithm) from the original atomic orbitals of the QM region. The generic $T_{\nu\mu}$ element of matrix T can be expressed as follows:

$$T_{\nu\mu} = \left[1 - \sum_{i=1}^{N_{ELMO}} (S_{\mu i})^2 \right]^{-1/2} \left[\delta_{\nu\mu} - \sum_{i=1}^{N_{ELMO}} C_{\nu i}^\perp S_{\mu i} \right] \quad (3)$$

where $S_{\mu i}$ is the overlap integral between the original basis function $|\chi_\mu\rangle$ and the i -th orthonormalized ELMO, $\delta_{v\mu}$ is the usual Kronecker delta and C_{vi}^\perp is the coefficient associated with the original basis function $|\chi_v\rangle$ for the i -th orthonormalized ELMO. Q is the matrix corresponding to the canonical orthogonalization of the basis functions $[|\tilde{\chi}_1\rangle, |\tilde{\chi}_2\rangle, \dots, |\tilde{\chi}_{M_{QM}}\rangle]$ and brings from the “intermediate” basis-set $\tilde{\chi}$ to the final basis-set χ' . The generic Q_{ij} element of matrix Q is given by:

$$Q_{ij} = \frac{\tilde{V}_{ij}}{\tilde{v}_j^{1/2}} \quad (4)$$

where \tilde{V}_{ij} is the i -th component of the j -th eigenvector and \tilde{v}_j is the j -th eigenvalue of the overlap matrix \tilde{S} between the transformed basis functions $[|\tilde{\chi}_1\rangle, |\tilde{\chi}_2\rangle, \dots, |\tilde{\chi}_{M_{QM}}\rangle]$. It is worth noting that, since only few orthonormalized ELMOs present small orthogonalization tails on the atoms at the frontier between the QM and ELMO regions (see steps 2 and 3 above), in our case all the eigenvalues $\{\tilde{v}_j\}$ of the overlap matrix \tilde{S} are always significantly different from zero ($\tilde{v}_j > 10^{-4} \forall j$). This means that, unlike the original LSCF approach, the number of basis functions corresponding to the QM region does not change and, therefore, linear-dependency problems are avoided.

5. Computation of the Fock matrix in the original basis-set. The QM/ELMO method described by the present algorithm consists in performing a HF or DFT calculation on the QM region in presence of transferred, orthonormalized and frozen extremely localized molecular orbitals that describe the ELMO subsystem. Therefore, the Fock matrix for the QM region is to be determined and the first step to accomplish this task

consists in computing the $M \times M$ Fock matrix F in the original basis-set (namely in the basis-set before steps 3 and 4), which can be expressed as follows:

$$F_{\mu\nu} = \langle \chi_\mu | \hat{h}^{core} | \chi_\nu \rangle + \sum_{\lambda, \sigma=1}^M P_{\lambda\sigma}^{QM} \left[(\chi_\mu \chi_\nu | \chi_\sigma \chi_\lambda) - \frac{1}{2} x (\chi_\mu \chi_\lambda | \chi_\sigma \chi_\nu) \right] + \sum_{\lambda, \sigma \in ELMO} P_{\lambda\sigma}^{ELMO} \left[(\chi_\mu \chi_\nu | \chi_\sigma \chi_\lambda) - \frac{1}{2} x (\chi_\mu \chi_\lambda | \chi_\sigma \chi_\nu) \right] + h_{\mu\nu} + F_{\mu\nu}^{QM} + F_{\mu\nu}^{ELMO} + v_{\mu\nu}^{XC} \quad (5),$$

where \hat{h}^{core} is the standard core one-electron Hamiltonian operator, P^{QM} and P^{ELMO} are the QM and ELMO one-electron reduced density matrices in the original basis-set, respectively, $(\chi_\alpha \chi_\beta | \chi_\gamma \chi_\delta)$ represents a two-electron repulsion integral, x is the fraction of exact exchange, and $\langle \chi_\mu | \hat{v}^{XC} [P^{QM} + P^{ELMO}] | \chi_\nu \rangle$ is the element (μ, ν) of the exchange-correlation potential matrix, which depends on the global one-electron reduced density matrix P given by the sum of P^{QM} and P^{ELMO} . Of course, if a QM/ELMO calculation is performed at Hartree-Fock level (hereafter, HF/ELMO calculation), x is set equal to 1 and the contribution of the exchange-correlation potential disappears. Moreover, it is important to note that, in case of a QM/ELMO computation at DFT level (namely, DFT/ELMO computation), the transferred and orthonormalized ELMOs are currently treated in the algorithm as frozen Kohn-Sham orbitals. Therefore, a DFT/ELMO calculation should be more correctly labeled as a DFT/DFT-ELMO calculation. However, for the sake of simplicity, the former convention will be used throughout the text.

Analyzing equation (5), the contributions that change and are regularly updated during the self-consistent field iterations are the contribution of the QM region to the Fock matrix (namely, elements $F_{\mu\nu}^{QM}$ in equation (5)) and, in case of a DFT calculation, the exchange-correlation term ($v_{\mu\nu}^{XC}$ elements). Concerning the latter, it is important to note that, although P^{ELMO} remains constant during the iterations, due to the non-

linearity of the exchange-correlation potential, it is not possible to separately evaluate the terms $v_{\mu\nu}^{XC}[P^{QM}]$ and $v_{\mu\nu}^{XC}[P^{ELMO}]$, but we have to compute $v_{\mu\nu}^{XC}[P]$ depending on the global density matrix; moreover to accomplish this task the method currently exploits a full-molecule grid to compute the full density and the exchange-correlation matrix elements at each iteration. For a future improvement of the method, we already envisage the possibility of specifying in input the range of the grid that needs to be updated at each iteration.

On the contrary, the core and ELMO contributions to the Fock matrix ($h_{\mu\nu}$ and $F_{\mu\nu}^{ELMO}$ elements, respectively) are constant and can be evaluated only once before starting the real SCF cycle, which enables to reduce the computational cost. Nevertheless, since the Fock matrix for the QM region is initially evaluated in the full (or supermolecular) basis-set, at the moment this is the rate-limiting step of the QM/ELMO SCF algorithm. Although in the present work all kind of QM/ELMO computations were performed using the supermolecular basis, we have already envisaged a separate study aiming at determining a suitable criterion to reduce the number of basis functions in the initial computation of the Fock matrix, as, for instance, the criterion based on the net Mulliken populations that was successfully proposed in the framework of the projection-based method.^{52,55}

6. Computation of the Fock matrix in the basis-set of the QM subsystem. Exploiting the transformation matrix B given by equation (2), the Fock matrix for the QM subsystem in the full basis-set is converted to the Fock matrix for the QM subsystem

in the basis-set $\chi' = \left[\left| \chi_1' \right\rangle, \left| \chi_2' \right\rangle, \dots, \left| \chi_{M_{QM}}' \right\rangle \right]$ through the following matrix relation:

$$F' = B^\dagger F B \quad (6)$$

with B^\dagger as the transpose of matrix B . F' is obviously an $M_{QM} \times M_{QM}$ and, since M_{QM} is generally much lower than M , this enables to significantly reduce the computational cost of the QM/ELMO calculations (see next step), especially if very large systems are investigated.

7. *Diagonalization of the Fock matrix.* The Fock matrix F' is diagonalized:

$$F' C' = C' E' \quad (7)$$

For the sake of simplicity, let us assume to work with a $2N$ electron closed-shell system and to perform a restricted Hartree-Fock/ELMO (RHF/ELMO) or a restricted DFT/ELMO calculation. The diagonalization will provide N doubly occupied molecular orbitals and $M_{QM} - N$ virtual orbitals, which will be used to perform post-HF/ELMO calculations, such as CCSD(T)/ELMO or MP2/ELMO computations (see subsection II.B).

8. *Transformation of the molecular orbitals to the supermolecular basis.* The matrix of the molecular orbitals coefficients C' resulting from the diagonalization of the Fock matrix F' is transformed back in the original and supermolecular basis $\chi = [|\chi_1\rangle, |\chi_2\rangle, \dots, |\chi_M\rangle]$. This is done by exploiting again the transformation matrix B initially obtained at step 4:

$$C = B C' \quad (8)$$

8. *Computation of the QM one-electron density matrix.* Using matrix C obtained at the previous step, the one-electron density matrix for the QM subsystem in the full basis-set is computed as follows:

$$P_{\lambda\sigma}^{QM} = 2 \sum_{i=1}^N C_{\sigma i}^\dagger C_{\lambda i} \quad (9)$$

It is also worth noting that, the ELMO one-electron density matrix is calculated in the full supermolecular basis-set, but from the coefficients $\{C_{\mu i}^{\perp}\}$ of the Löwdin orthonormalized extremely localized molecular orbitals of the ELMO region:

$$P_{\lambda\sigma}^{ELMO} = 2 \sum_{i=1}^N C_{\sigma i}^{\perp} \cdot C_{\lambda i}^{\perp} (10).$$

However, as already pointed out above, since the ELMOs remain frozen during the QM/ELMO computation, the determination of the ELMO one-electron density matrix through equation (10) is performed only once, namely, immediately after the preliminary orthonormalization of the extremely localized molecular orbitals transferred to the ELMO subsystem.

After step 8, convergence is checked: if it is achieved, the cycle ends and final properties (e.g., the final energy (see below)) are computed; otherwise a new iteration starts from step 5 and the QM one-electron density matrix obtained through equation (9) is used to update the Fock matrix F in the full basis-set.

For the sake of completeness, the electronic energy resulting from an HF/ELMO calculation is given by:

$$E_{HF/ELMO} = \frac{1}{2} \sum_{\mu, \nu=1}^M P_{\nu\mu}^{QM} (2h_{\mu\nu}^{core} + F_{\mu\nu}^{QM}) + \frac{1}{2} \sum_{\mu, \nu=1}^M P_{\nu\mu}^{ELMO} (2h_{\mu\nu}^{core} + F_{\mu\nu}^{ELMO}) + \frac{1}{2} \left(\sum_{\mu, \nu=1}^M P_{\nu\mu}^{QM} F_{\mu\nu}^{ELMO} + \sum_{\mu, \nu=1}^M P_{\nu\mu}^{ELMO} F_{\mu\nu}^{QM} \right) (11)$$

where the matrices P^{QM} , P^{ELMO} , h^{core} are identical to those in equation (5). F^{QM} and F^{ELMO} are also analogous to those in relation (5) when the fraction of exact exchange is set equal to 1. It is easy to observe that the total HF/ELMO electronic energy can be decomposed into three contributions, namely a purely QM (first term on the right-hand side of equation (11)), a purely ELMO (second term) and mixed QM/ELMO contribution (third term).

In case of a DFT/ELMO computation, the expression of the electronic energy becomes:

$$E_{DFT/ELMO} = \frac{1}{2} \sum_{\mu, \nu=1}^M P_{\nu\mu}^{QM} (2h_{\mu\nu}^{core} + F_{\mu\nu}^{QM}) + \frac{1}{2} \sum_{\mu, \nu=1}^M P_{\nu\mu}^{ELMO} (2h_{\mu\nu}^{core} + F_{\mu\nu}^{ELMO}) + \frac{1}{2} \left(\sum_{\mu, \nu=1}^M P_{\nu\mu}^{QM} F_{\mu\nu}^{ELMO} + \sum_{\mu, \nu=1}^M P_{\nu\mu}^{ELMO} F_{\mu\nu}^{QM} \right) + I$$

Also in this situation P^{QM} , P^{ELMO} and h^{core} correspond to those in equation (5), while, unlike relation (11), F^{QM} and F^{ELMO} are characterized by a fraction of exact exchange that is related to the chosen functional ($0 \leq x \leq 1$, with x always defined in equation (5)). $E_{XC}[P^{QM} + P^{ELMO}]$ is the exchange-correlation energy, which also depends on the chosen exchange-correlation functional and on the global density matrix. Due to the non-linearity of the exchange-correlation functional, we cannot exactly separate the QM and ELMO contributions to $E_{XC}[P^{QM} + P^{ELMO}]$ and, consequently, in case of a DFT/ELMO computation, the DFT/ELMO electronic energy cannot be decomposed into terms that can be classified as purely DFT, ELMO or DFT/ELMO.

Finally, it is worth noting that the HF/ELMO and DFT/ELMO algorithms were implemented through a suitable modification of the Hartree-Fock and DFT subroutines within the Gaussian09 quantum chemistry package.⁷⁴

B. Post-HF/ELMO methods

The required preliminary step for a post-HF/ELMO calculation is obviously a converged Hartree-Fock/ELMO computation, which, in case of a closed-shell system, provides a wavefunction having the following form (restricted case):

$$\Psi_{\text{RHF/ELMO}} = \frac{1}{\sqrt{(2(N+N_{\text{ELMO}}))!}} \hat{A} \left(\phi_1^\perp \phi_1^\perp \dots \phi_{N_{\text{ELMO}}}^\perp \phi_{N_{\text{ELMO}}}^\perp \varphi_1 \varphi_1 \dots \varphi_N \varphi_N \right) (12),$$

where \hat{A} is the usual antisymmetrizer, $\{\phi_i^\perp\}_{i=1}^{N_{\text{ELMO}}}$ are the orthonormalized, frozen ELMOs obtained at step 3 of the QM/ELMO SCF algorithm, and $\{\varphi_i\}_{i=1}^N$ are the molecular orbitals resulting from the RHF/ELMO calculation and describing the QM subsystem; the overbar symbol is used to indicate a spinorbital with spin-part β , while its absence indicates a spinorbital with spin-part α .

As one should expect, the substitutions from occupied to virtual orbitals for the subsequent post-HF/ELMO computations do not involve the frozen occupied ELMOs $\{\phi_i^\perp\}_{i=1}^{N_{\text{ELMO}}}$, but only the molecular orbitals $\{\varphi_i\}_{i=1}^N$. Therefore, since N is generally much lower than $N+N_{\text{ELMO}}$, the computational cost is already significantly reduced. The cost of the post-HF/ELMO calculation can be obviously further lowered if only valence occupied molecular orbitals are considered for the substitutions (frozen core approximation), as it was the case for all the post-HF computations performed in the present investigation (see Computational Details section).

Moreover, and more importantly, it is worth reminding that, as seen in the previous subsection (particularly, step 7 of the algorithm), the number of virtual orbitals resulting from an HF/ELMO calculation is quite low compared to the size of the full basis-set for the investigated system. For example, from an RHF/ELMO calculation we obtain only $M_{\text{QM}} - N$ virtual orbitals and not $M - N$, where $M \gg M_{\text{QM}}$. Considering that the computational effort to perform a post-HF computation depends more importantly on the number of virtual molecular orbitals than on the number of occupied ones, the cost of the post-HF/ELMO computations is already automatically and significantly reduced (much more than for the lower number of occupied

molecular orbitals), without introducing an additional criterion to properly truncate the number of atomic orbitals in the preliminary HF/ELMO step.

In this work, only Møller-Plesset/ELMO and Coupled Cluster/ELMO strategies have been developed and discussed. As the SCF QM/ELMO algorithms, they have been implemented by modifying the original Møller-Plesset and Coupled Cluster routines of Gaussian09.⁷⁴

III. COMPUTATIONAL DETAILS

A. Investigated systems and preliminary steps

In order to start testing performances and efficiency of the newly developed QM/ELMO method, we considered three chemically relevant situations: a nucleophilic substitution reaction S_N2 , the dissociation of a bond, and the case of a system characterized by hydrogen-bond intermolecular interactions.

Concerning the S_N2 reaction, we studied the nucleophilic substitution of bromine in the 1-bromodecane molecule by the chloride anion (see the corresponding reaction scheme in Figure 1A). As a first step, reactants and products geometries were optimized at B3LYP level of theory (cc-pVDZ basis-set) using the Gaussian09 software package.⁷⁴ Starting from these geometries, the transition state (TS) of the S_N2 reaction was detected using the synchronous transit-guided quasi-Newton method (QST2 keyword in Gaussian09).⁷⁵ The reaction energy profiles were then reconstructed using the TS geometry as initial point and integrating the mass-weighted intrinsic reaction coordinate (IRC) in both directions until minima were reached.⁷⁶⁻⁷⁸

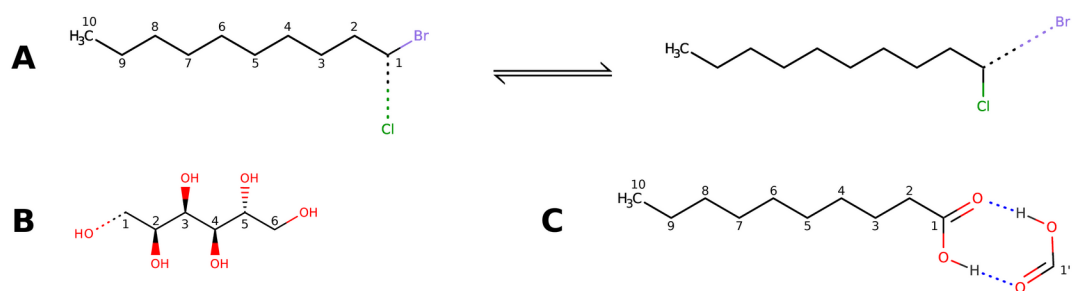


Figure 1. Schematic representation of the chemically relevant processes and properties investigated in this work: (A) S_N2 nucleophilic substitution of bromine in the 1-bromodecane molecule by the chloride anion, (B) dissociation of the terminal C1-OH bond in sorbitol, (C) intermolecular interactions in the formic acid - decanoic acid dimer.

As a second case study, we considered the cleavage of the covalent terminal C-OH bond in sorbitol to study the capabilities of the new method to correctly retrieve molecular dissociation energies. The sorbitol geometry was initially optimized using the same level of theory and basis-set employed for the S_N2 reaction (B3LYP/cc-pVDZ). The dissociation energy profile was afterwards constructed scanning the C1-O internal coordinate (see Figure 1B for the labels) from 0.98 Å to 5.88 Å, with 1.43 Å as the equilibrium distance.

Finally, we decided to assess the capabilities of the QM/ELMO strategy in correctly modeling intermolecular interactions. To accomplish this task, we considered the formic acid - decanoic acid dimer and, in particular, we analyzed the corresponding interaction energy profiles as a function of the distance between the monomers. The initial structure of the complex was properly designed to have two hydrogen bonds between the carboxylic groups of the two monomers (see Figure 1C). As for the other systems, the initial geometry was optimized at B3LYP level with the cc-pVDZ basis-set, in this case properly checking that the two hydrogen bonds remained in the converged structure. The energy profile was then constructed scanning the internal coordinate represented by the distance between the two carboxylic carbon atoms (C1

and C1', see Figure 1C) from 3.29 Å to 7.79 Å, with 3.79 Å as the equilibrium distance.

In the three cases described above, vibrational frequencies were calculated to assess the correct nature of the detected stationary points (minima or transition states). Furthermore, sample geometries were extracted from the obtained curves and used to perform QM and QM/ELMO calculations (see below).

B. QM/ELMO and QM calculations

QM/ELMO and QM calculations were performed on all the geometries extracted from the energy profiles obtained for the three case studies mentioned above.

As already discussed in the Theory section, the QM/ELMO computations were carried out by exploiting a properly modified version of the Gaussian09 package.⁷⁴ For each of the three systems described in the previous subsection, we tried to employ four different levels of theory to treat the quantum mechanical region in the calculations: i) Hartree-Fock (HF), ii) density functional theory with the B3LYP exchange-correlation functional (DFT-B3LYP or, hereinafter, also indicated simply as B3LYP), iii) second-order Møller-Plesset perturbation theory (MP2), and iv) Coupled Cluster. Concerning the Coupled Cluster level, when possible (see below), we opted for the CCSD(T) (Coupled Cluster with single and double substitutions plus perturbative triples) method, otherwise we limited to the CCSD (Coupled Cluster with single and double substitutions) one. In particular, for the study of the S_N2 reaction and of the intermolecular interactions, we used the HF, B3LYP, MP2 and CCSD(T) techniques. For the C-OH bond breaking in sorbitol, we only used the HF, B3LYP and CCSD levels of theory. The reason behind this choice has to be sought in the known shortcoming of CCSD(T) in describing bonds breaking, where the inclusion of

perturbative triples leads to a wrong estimation of the dissociation energy and, above all, to the asymptotic behavior breakdown far from equilibrium.⁷⁹⁻⁸¹ The same anomalous asymptotic behavior was also observed at MP2 level and, for this reason, also the MP2 method was not exploited for the computations on sorbitol.

Furthermore, to explore the effect of the ELMO embedding, we carried out QM/ELMO calculations progressively changing the size of the QM regions. For the S_N2 reaction, we considered from two to eight CH_2 alkyl moieties in the QM region together with the chlorine and bromine atoms; for the C-OH bond dissociation in sorbitol, we included from two to five carbon atoms (along with the corresponding substituents), namely from two to five alcohol moieties; finally, in the study of the intermolecular interactions, we considered from three to nine carbon atoms or, more precisely, we considered QM regions constituted by formic acid plus the carboxylic group and a variable number of alkyl groups (from one to seven) of decanoic acid. Hereinafter, for all the case studies, the syntax QM(*N*)/ELMO will indicate QM/ELMO calculations where the QM active region comprises *N* carbon atoms and corresponding substituents as described above.

To evaluate the approximation introduced through the new approach, all the QM/ELMO results were then compared to the benchmark ones obtained from the corresponding, standard and fully quantum mechanical calculations at the same level of theory. In particular, in this work the fully QM calculations were performed using two different basis-sets: cc-pVDZ and the aug-cc-pVDZ. The results obtained with the former basis-set were considered as references for the corresponding QM/ELMO computations with the same standard set of basis functions. The outcomes of the calculations with the aug-cc-pVDZ basis-set were taken as benchmark values of QM/ELMO computations carried out with a the mixed basis-set aug-cc-pVDZ/cc-

pVDZ, which allowed us to evaluate the performances and the stability of our approach when different sets of basis functions are used for the QM and ELMO regions. In particular, in the case of the mixed basis-set, all the atoms belonging to the ELMO region plus the frontier ones were treated with the cc-pVDZ set of functions, while the remaining atoms of the QM subsystem were described at aug-cc-pVDZ level.

Finally, for the sake of completeness, it is worth noting that all kind of post-HF computations (both QM and QM/ELMO) were carried out exploiting the frozen core approximation.

C. ELMOs calculations

The extremely localized molecular orbitals employed in the QM/ELMO computations were previously computed on proper model molecules by exploiting a modified version⁶⁵ of the GAMESS-UK quantum chemistry package⁸² that implements the Stoll equations⁶⁴ (see Supporting Information for more details about the ELMO theory). In particular, for the study of the S_N2 reactions, the ELMOs describing the fragments of the alkyl groups were calculated on an optimized structure of decane; concerning the bond dissociation, the necessary ELMOs were determined on the minimum geometry of the investigated sorbitol molecule; finally, for the study of the intermolecular interactions, the ELMOs were computed by exploiting an optimized geometry of the isolated decanoic acid (since the formic acid always belongs to the QM region, see previous subsection). All the geometry optimizations of the model molecules were carried out at B3LYP/cc-pVDZ level.

Before each QM/ELMO calculation, the pre-computed ELMOs were transferred from the model molecules on which they were determined to the target molecular

geometries using the *ELMOdb* program,⁷³ which is the software associated with the recently constructed ELMO libraries⁷³ and which implements the strategy originally proposed by Philipp and Friesner^{71,83} for the rotation of strictly localized molecular orbitals (see Supporting Information for more details).

IV. BENCHMARK CALCULATIONS

In this section we will show and analyze the results of the test calculations carried out to preliminarily assess the capabilities of the developed QM/ELMO techniques. As already anticipated above, we will analyze the performances of the new methods when they are applied to model an S_N2 reaction (subsection IV.A), the dissociation of a bond (subsection IV.B), and intermolecular interactions (subsection IV.C). Finally, in subsection IV.D, we will discuss the computational cost associated with the performed Coupled Cluster/ELMO calculations.

A. S_N2 reaction

To start assessing the performances of the new QM/ELMO techniques in modeling the S_N2 substitution reaction between the chloride anion and CH₃(CH₂)₉Br (see Figure 1A), for each level of theory (i.e., HF, DFT-B3LYP, MP2 and CCSD(T)) and basis-set (i.e., cc-pVDZ and mixed aug-cc-pVDZ/cc-pVDZ) taken into account, we decided to determine the discrepancies $\Delta \Delta E_{TS-R}$ and $\Delta \Delta E_{TS-P}$, namely the discrepancies of the energy variations $\Delta E_{TS-R} = E_{TS} - E_{reactants}$ and $\Delta E_{TS-P} = E_{TS} - E_{products}$ computed at QM/ELMO level from the values of the same quantities computed through the corresponding fully quantum mechanical method. These deviations have been monitored by gradually increasing the size of the QM subsystem in the QM/ELMO

computations, as mentioned in the Computational Details section. The results obtained for the discrepancies of the ΔE_{TS-R} and ΔE_{TS-P} quantities are reported in the left and right panels of Figure 2, respectively. Top and bottom panels of Figure 2 refer to the results for basis-sets cc-pVDZ and aug-cc-pVDZ/cc-pVDZ, respectively.

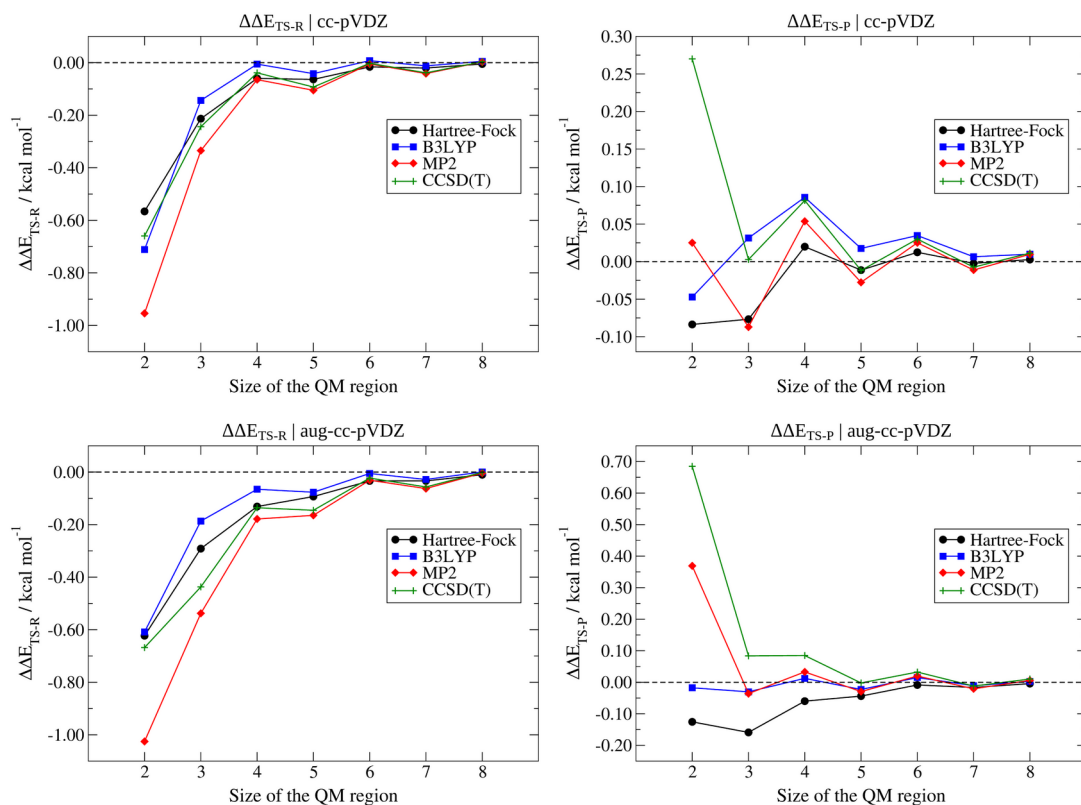


Figure 2. S_N2 reaction: discrepancies of the energy variations ΔE_{TS-R} (left panels) and ΔE_{TS-P} (right panels) computed at QM/ELMO levels (QM = HF, B3LYP, MP2 and CCSD(T)) from those obtained through the corresponding fully quantum mechanical computations (top panels: cc-pVDZ basis-set; bottom panels: mixed aug-cc-pVDZ/cc-pVDZ basis-set). The variation of the discrepancies as a function of the QM region size (i.e., number of alkyl groups) is explicitly shown.

First of all, let us consider the results obtained with the cc-pVDZ basis-set. From the analysis of the top-left panel of Figure 2, it is immediately clear that, regardless of the size of the QM subsystem, all the QM/ELMO methods provide energy variations that

agree with those resulting from the corresponding fully QM computations within chemical accuracy, namely the discrepancies are always lower than 1.0 kcal/mol. Moreover, all the QM/ELMO techniques show the expected trend: larger discrepancies from the benchmark values when only few alkyl moieties are fully treated at quantum mechanical level and much lower differences when larger QM subunits are considered. Just considering the CCSD(T)/ELMO calculations as an example, in absolute value the initial discrepancy (two CH₂ moieties in the QM region) amounts to 0.66 kcal/mol, it decreases to 0.24 kcal/mol when the QM subsystem is increased by one alkyl unit and then it remains lower than 0.1 kcal/mol for all the other QM/ELMO computations, obviously converging towards the benchmark CCSD(T) result as the QM subunit becomes larger and larger. Concerning the top-right panel of Figure 2, we observe exactly the same trends highlighted in the top-left one. The only difference is that the deviations of the ΔE_{TS-P} values oscillate between positive and negative values as the size of QM subsystem increases.

Pertaining to the QM/ELMO results obtained with the mixed aug-cc-pVDZ/cc-pVDZ basis-set (bottom panels of Figure 2), we observe exactly the same qualitative trends of the obtained $\Delta\Delta E$ values as a function of the size of the QM region. Furthermore, also for the mixed basis-set, in practically all the cases, the discrepancies with respect to the fully QM calculations are lower than 1.0 kcal/mol, with the only exception being the $\Delta\Delta E_{TS-R}$ value computed at MP2/ELMO level with the smallest active region (-1.03 kcal/mol). Finally, it is possible to notice that, for the QM/ELMO computations with the smallest QM subsystems, the deviations from the fully quantum chemical results are generally larger than those observed with the cc-pVDZ basis-set, although there are also cases in which the opposite is true. The results tend to coincide already when 3 or 4 alkyl moieties are included in the QM region and this

is might due to the fact that, as the active subsystem becomes larger, the inhomogeneous-basis-set-frontier is moved away from the subunit of the system where chemistry is really taking place. However, it is worth pointing out again that practically all the $\Delta\Delta E$ values related to the aug-cc-pVDZ/cc-pVDZ basis-set are within the limit of chemical accuracy, thus suggesting that the use of mixed basis-sets to describe the QM and ELMO subsystems in QM/ELMO computations does not significantly impact on the quality of the results.

To investigate the performances of the different QM/ELMO approaches more in detail, we also considered the full S_N2 reaction profiles. Also in these cases, we did that for all the possible levels of theory considered in this study and varying the extension of the quantum mechanical region. For the sake of simplicity, we limited this analysis to the cc-pVDZ basis-set. All the obtained reaction profiles were afterwards compared point-by-point to the corresponding ones resulting from purely quantum chemical calculations. Here we report only the results obtained at Hartree-Fock and CCSD(T) levels (see Figures 3 and 4, respectively), while the analogous graphs obtained at B3LYP and MP2 levels are provided in the Supporting Information (see Figures S3 and S4).

Analyzing Figure 3, we can notice that, at Hartree-Fock level, the $\Delta\Delta E$ discrepancies with the benchmark value are lower than 1.0 kcal/mol for every point along the reaction coordinate and for each size of the QM subsystem. Furthermore, as observed also for the ΔE_{TS-R} and ΔE_{TS-P} quantities, we can see that the deviations reduce when the QM region becomes larger. In particular, starting from the HF/ELMO computations with at least five alkyl groups included in the active region, the differences with the correct Hartree-Fock values are lower than 0.1 kcal/mol for the complete reaction profile.

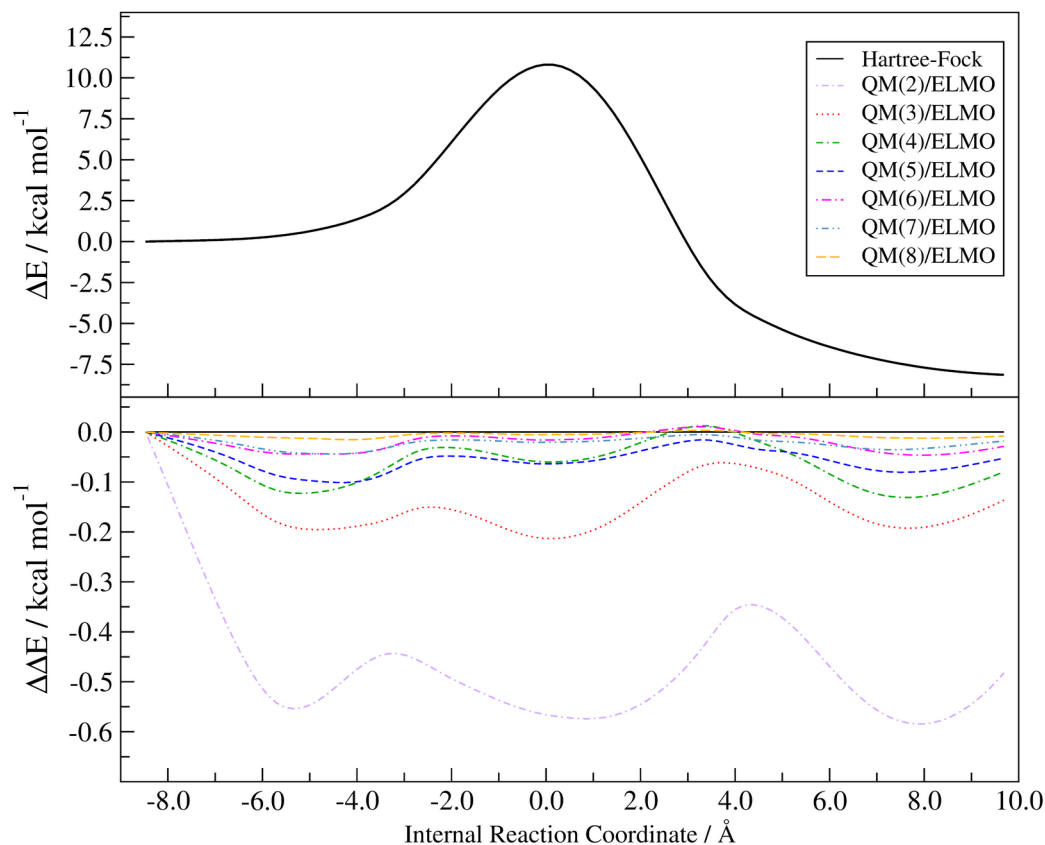


Figure 3. S_N2 reaction energy profile at HF level (top panel) and deviations from it when HF/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves refer to the cc-pVDZ basis-set.

Pertaining to the QM/ELMO computations at CCSD(T) level (see Figure 4), we have very similar results. In almost all the cases, the discrepancies with the reference CCSD(T) values are lower than 1.0 kcal/mol for each point of the reaction profile. An exception is represented by some points of the reaction profile obtained with a QM region including only two alkyl groups. However, already from the CCSD(T)/ELMO calculations with three CH₂ moieties in the QM subsystem, the differences constantly remain below 0.3 kcal/mol and the situation further improves as larger QM regions are taken into account. Completely analogous trends have been obtained through the B3LYP/ELMO and MP2/ELMO calculations (see Figures S3 and S4 in the

Supporting Information), with the largest discrepancies that are observed when the smallest active region is considered and that systematically decrease when the active region becomes larger.

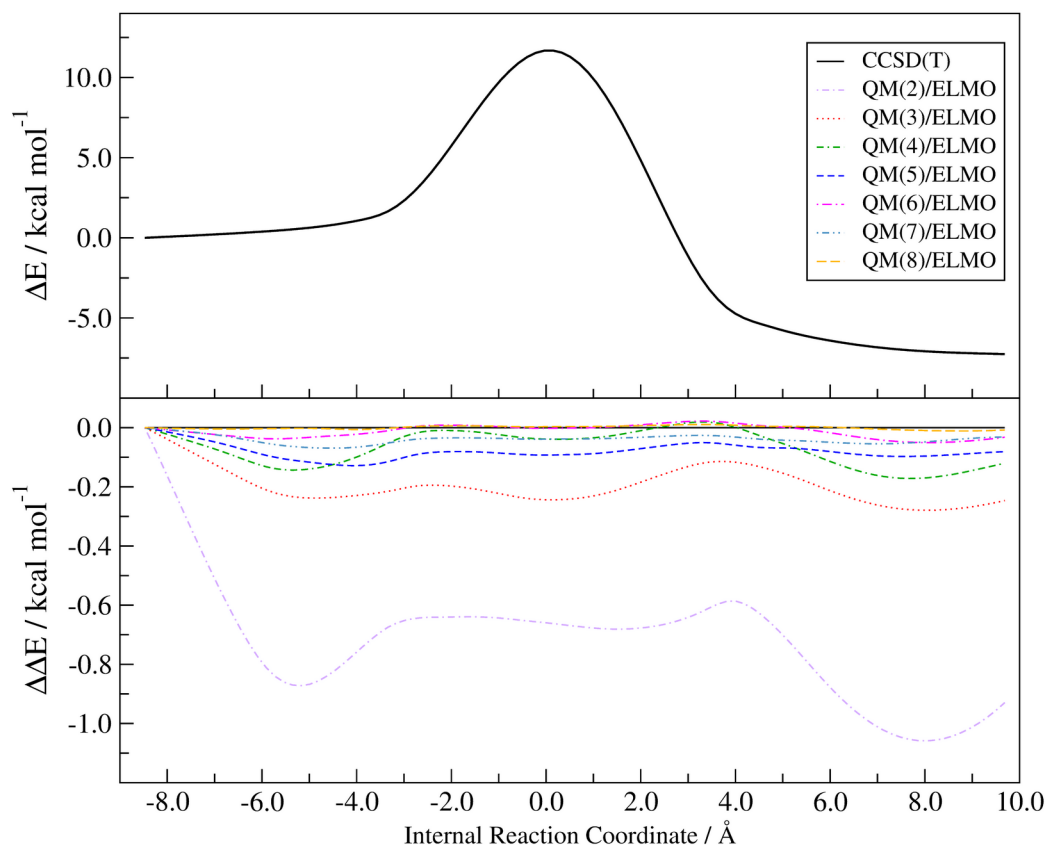


Figure 4. S_N2 reaction energy profile at CCSD(T) level (top panel) and deviations from it when CCSD(T)/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves refer to the cc-pVDZ basis-set.

To complete the previous analysis, in Figure 5, we reported the maximum and average absolute deviations observed for all the reaction profiles obtained through QM/ELMO calculations. The depicted histograms basically confirm what was just pointed out above. Concerning the HF/ELMO and B3LYP/ELMO computations, the largest discrepancies always remain within chemical accuracy, while, for the MP2/ELMO and CCSD(T)/ELMO calculations, the maximum errors are larger than 1 kcal/mol

only for the smallest QM region. However, in all the situations the maximum errors rapidly decrease as the extension of the active subunit increases. The same trend as a function of the size of the QM subsystem is also observed for the average absolute discrepancies associated with the different QM/ELMO techniques. In all cases, the average absolute errors are always lower than 1 kcal/mol. For the sake of completeness, it is worth pointing out that, in this case, the range over which the average discrepancies were computed corresponds to the one having reactants and products as boundaries along the reaction path (see subsection III.A for the exact range).

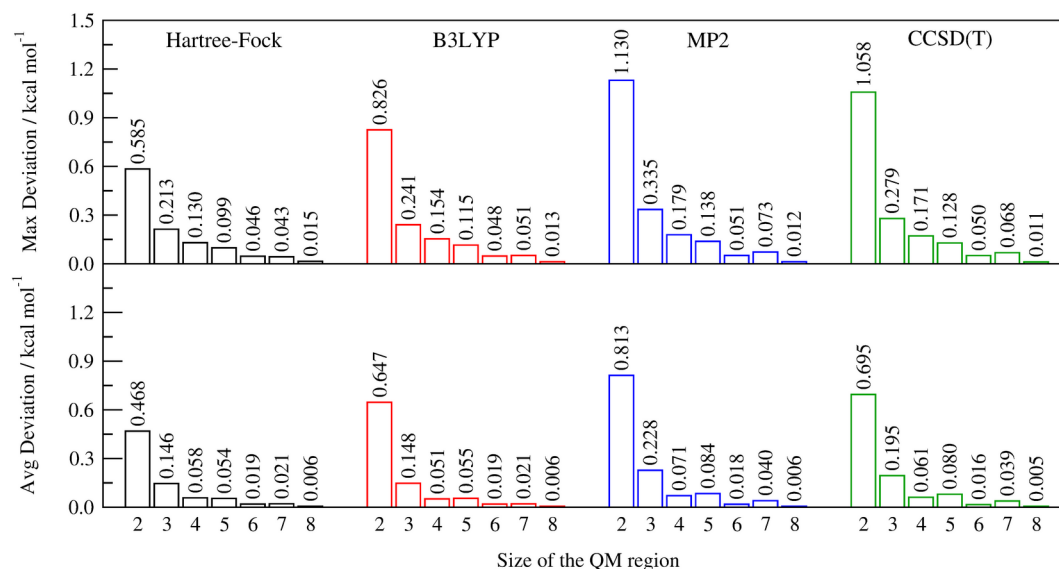


Figure 5. Maximum and average absolute deviations between corresponding QM/ELMO and fully QM S_N2 reaction energy profiles (cc-pVDZ basis-set). The variation of the deviations as a function of the QM region size is explicitly shown.

B. Bond dissociation

The second chemical process that we took into account to evaluate the capabilities of the newly developed QM/ELMO methods was the dissociation of the terminal OH

group (position 1) from sorbitol (see Figure 1B). As mentioned in the Computational Details section, in this case only three quantum mechanical levels of theory were considered for our calculations (Hartree-Fock, DFT-B3LYP and CCSD) with two different basis-sets (cc-pVDZ and mixed aug-cc-pVDZ/cc-pVDZ). For each combination of level of theory and basis-set, we evaluated the difference of the dissociation energy determined with the QM/ELMO approach with respect to the same quantity computed with the corresponding fully-QM technique. As done above, the discrepancies in the dissociation energies were determined for different sizes of the active region in the QM/ELMO computations (see Figure 6).

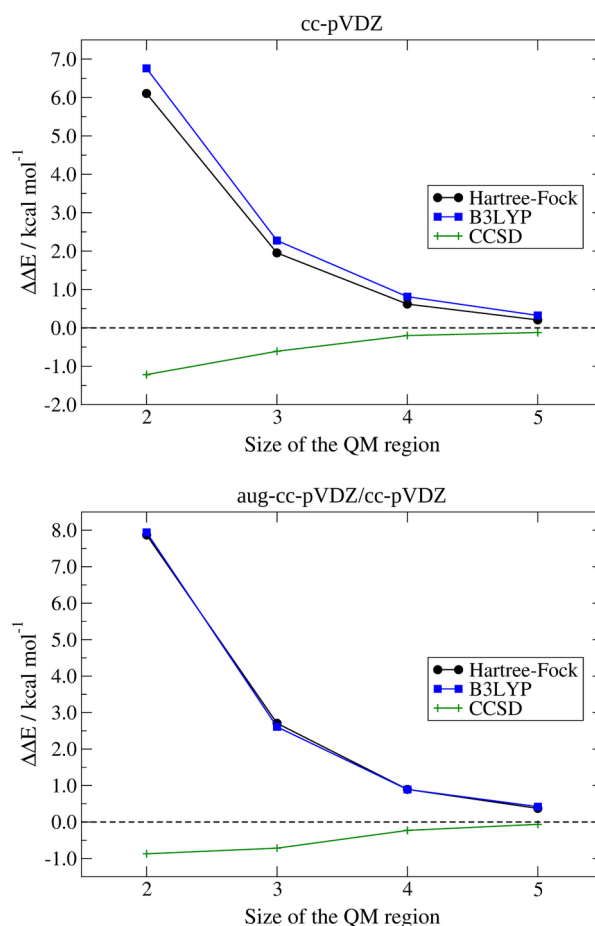


Figure 6. Dissociation of the terminal C1-OH bond in sorbitol: discrepancies of the dissociation energy values computed at QM/ELMO levels (QM = HF, B3LYP and CCSD) from those obtained through the corresponding fully quantum mechanical computations (top

panel: cc-pVDZ basis-set; bottom panel: mixed aug-cc-pVDZ/cc-pVDZ basis-set). The variation of the discrepancies as a function of the QM region size (i.e., number of alcohol moieties) is explicitly shown.

Considering the cc-pVDZ basis-set (top panel of Figure 6), unlike what we observed for the S_N2 reaction, the HF/ELMO and B3LYP/ELMO methods are characterized by discrepancies that are larger than 1 kcal/mol when two or three alcohol moieties are included in the QM part. On the contrary, the CCSD/ELMO calculations provided better results, with deviations from the corresponding CCSD benchmark that are slightly greater than 1 kcal/mol only when two carbon atoms (and corresponding substituents) are considered in the active region. However, also in this case, for all types of QM/ELMO computations, the discrepancies with the benchmark values decrease as the QM subunit becomes larger. Finally, for the sake of completeness, we note that, while the CCSD/ELMO method underestimates the dissociation energy computed at CCSD level, the HF/ELMO and B3LYP/ELMO techniques always overestimate the corresponding benchmark results. These trends might be due to the size of the QM region adopted for the calculations.

Concerning the mixed aug-cc-pVDZ/cc-pVDZ basis-set (bottom panel of Figure 6), from the qualitative point of view, we have the same trends observed in the cc-pVDZ case. In particular, as already observed above, the $\Delta\Delta E$ values for the HF/ELMO and B3LYP/ELMO calculations start being lower than 1 kcal/mol when 4 alcohol moieties are included in the QM region. For completeness, we can also notice that HF/ELMO and B3LYP/ELMO discrepancies computed with the inhomogeneous basis-set practically overlap. For the CCSD/ELMO calculations we also have a trend analogous to the corresponding cc-pVDZ one, but the deviation with respect to the full CCSD computation is already within the chemical accuracy limit when only 2 carbon atoms

(and corresponding substituents) belong to the QM region. The previous observations confirm that, at least from the energetic point of view, the use of different sets of basis functions for the QM and ELMO subsystems does not significantly influence the quality of the results.

For a more detailed analysis, the full dissociation energy profiles were also computed for all the levels of theory mentioned above with the cc-pVDZ basis-set, always gradually increasing the dimension of the active region in the QM/ELMO calculations. Point-by-point comparisons of the obtained profiles were performed against those resulting from the corresponding and traditional QM computations. In Figure 7, we reported the results obtained at Hartree-Fock level. For the two smallest QM regions (two and three alcohol moieties included), we can observe that the deviation of the HF/ELMO profile from the fully Hartree-Fock one increases with the C-O distance and discrepancies become larger than 1.0 kcal/mol. This is probably due to the fact that we are breaking a polar bond. In order to better describe this bond cleavage, it is necessary to take into account a larger QM region. In fact, although the same qualitative trend is observed for the HF/ELMO computations that involve four and five alcohol moieties, in those cases, the discrepancies with the benchmark values remain below the maximum limit for chemical accuracy. Exactly the same behavior can be observed for the B3LYP/ELMO calculations (see Figure S5 in the Supporting Information). On the contrary, better results were obtained through the calculations at CCSD level. In fact, in Figure 8, we can notice that, except for some points exceeding chemical accuracy for the QM/ELMO computations with the smallest active subsystem, all the other CCSD/ELMO profiles show deviations from the fully CCSD one that are lower than 0.61 kcal/mol. Also in this situation, the differences clearly

and significantly decrease when four or five carbon atoms (with corresponding substituents) are included in the QM part of the QM/ELMO computations.

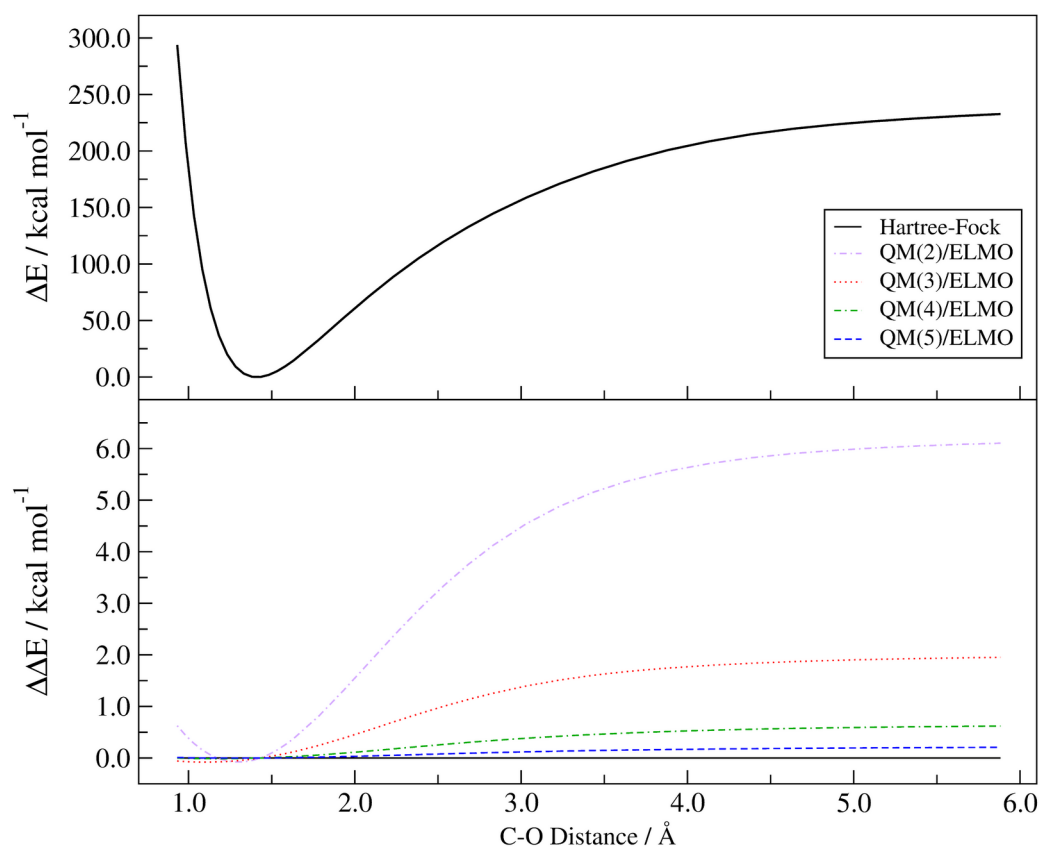


Figure 7. Dissociation energy profile of the terminal C-OH bond in sorbitol at HF level (top panel) and deviations from it when HF/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves refer to the cc-pVDZ basis-set.

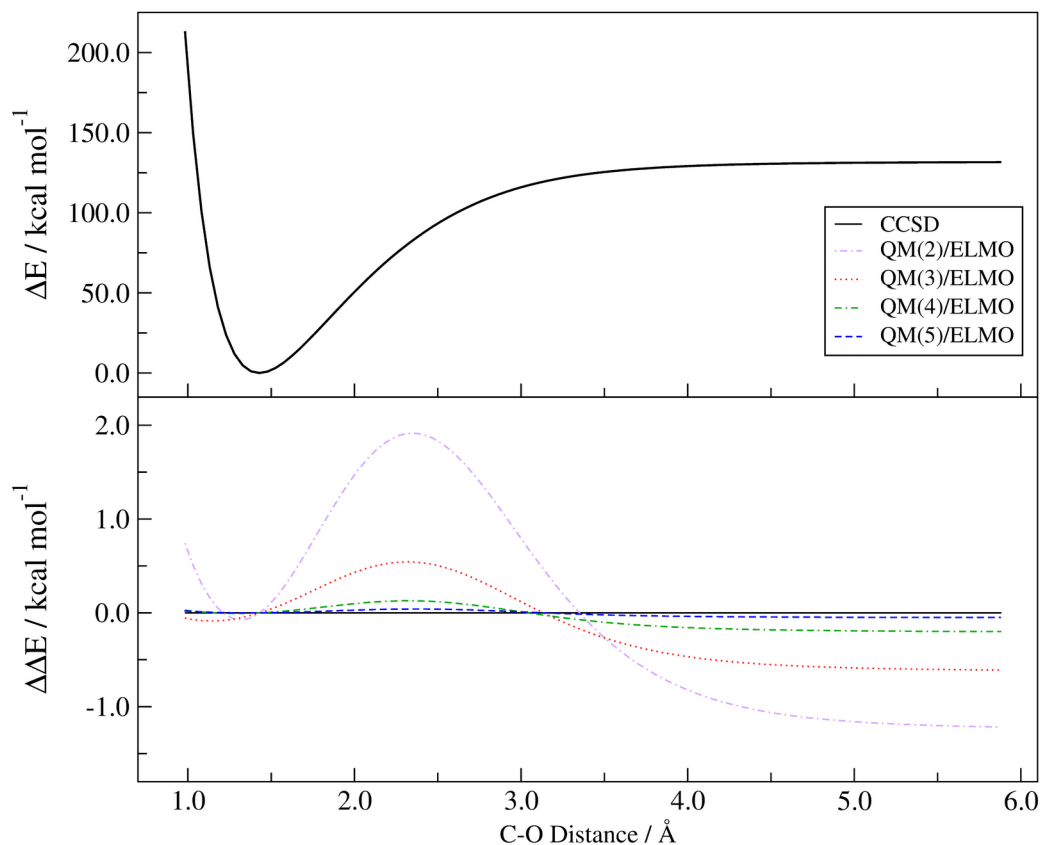


Figure 8. Dissociation energy profile of the terminal C-OH bond in sorbitol at CCSD level (top panel) and deviations from it when CCSD/ELMO calculations are performed with QM regions of different size (bottom panel) ; all the curves refer to the cc-pVDZ basis-set.

To better highlight and somehow summarize the trends discussed in the previous paragraph, in Figure 9, we show the maximum and average absolute discrepancies obtained for all the reaction profiles calculated at QM/ELMO levels. In agreement with what was pointed out above, at Hartree-Fock and B3LYP levels the maximum absolute deviation from the benchmark values is greater than 1.0 kcal/mol when the two smallest QM region are employed in the calculations, but it constantly decreases by increasing the size of the active subsystem. In the CCSD case, the maximum absolute error is greater than 1.0 kcal/mol only when two carbon atoms are included in the quantum mechanical part, but in all the other cases the maximum absolute

discrepancy is below the threshold required for chemical accuracy. We also calculated average discrepancies. However, unlike the S_N2 reaction, in this situation, there were no clear boundaries in the energy profile to define the range over which computing the average deviations. For this reason we decided to consider the profile obtained at CCSD level as a reference and, as upper boundary, we took the point where the energy variation was small enough to be considered as a point belonging to the plateau region (see again section III.A for the exact range) and practically corresponding to the dissociation limit. For the sake of consistency and compare all the results on the same footing, this range was used for all the QM levels of theory used in our computations. Obviously, the choice of the range may influence the values obtained for the average discrepancies. Therefore, for the selected range, at Hartree-Fock level, only the QM/ELMO computations with the smallest QM region are above the chemical accuracy limit, while, at B3LYP level, also the one with three alcohol moieties treated with DFT is slightly greater than 1.0 kcal/mol. Finally, as one should expect from the analysis of the dissociation energy profiles, all the average absolute discrepancies remain within chemical accuracy for the CCSD/ELMO calculations, with the largest average absolute difference amounting to 0.75 kcal/mol and corresponding to the computations with only two carbon atoms (and corresponding substituents) in the active QM region.

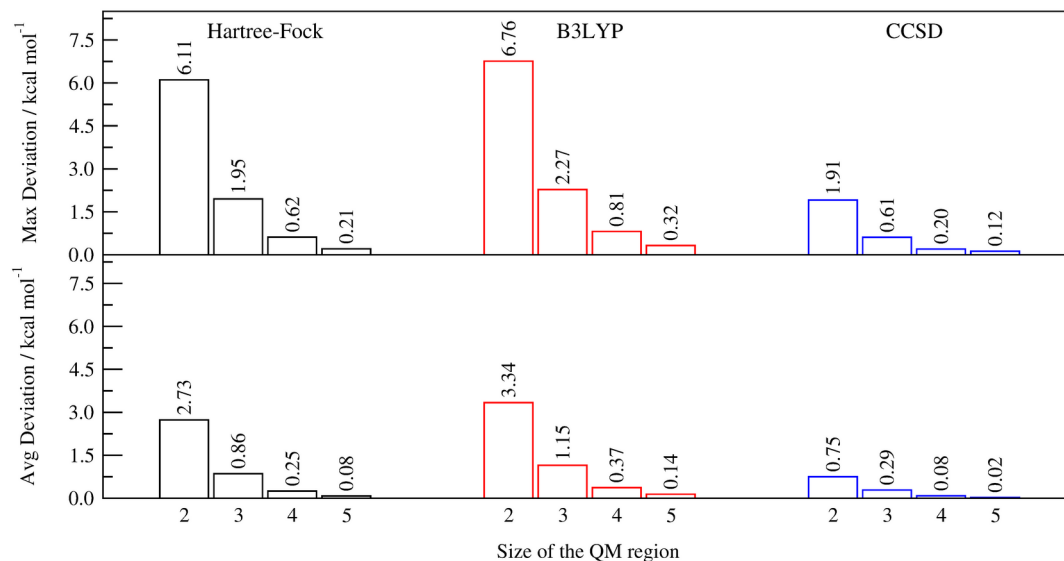


Figure 9 Maximum and average absolute deviations between corresponding QM/ELMO and fully QM dissociation energy profiles for the terminal C-OH bond in sorbitol (cc-VDZ basis-set). The variation of the deviations as a function of the QM region size is explicitly shown.

C. Intermolecular interactions

To complete our preliminary test calculations, we also decided to assess the performances of the QM/ELMO approaches in describing intermolecular interactions. To accomplish this task, we particularly considered the two hydrogen bond interactions occurring between decanoic and formic acids (see Figure 1C). As in the S_N2 reaction investigation, also in this case study, for our calculations we considered four levels of theory (Hartree-Fock, DFT-B3LYP, MP2 and CCSD(T)) with two different basis-sets (cc-pVDZ and mixed aug-cc-pVDZ/cc-pVDZ). For all the combinations of levels of theory and basis-sets, we evaluated the intermolecular interaction energies ($E_{\text{int}} = E(R_{\text{eq}}) - E(R_{\infty})$) at QM/ELMO level and the obtained values were compared to those resulting from the corresponding, fully QM calculations. As indicated in the Computational Details section, this was done for different extensions

of the quantum mechanical subsystem in the QM/ELMO computations. Figure 10 summarizes the obtained results.

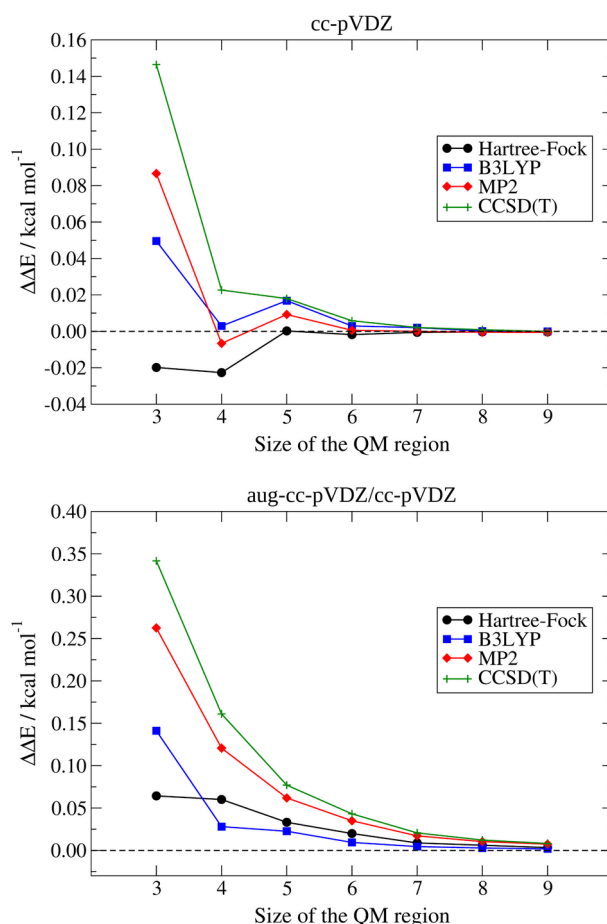


Figure 10. Intermolecular interactions in the formic acid - decanoic acid dimer: discrepancies of the interaction energy values computed at QM/ELMO levels (QM = HF, B3LYP and CCSD(T)) from those obtained through the corresponding fully quantum mechanical computations (top panel: cc-pVDZ basis-set; bottom panel: mixed aug-cc-pVDZ/cc-pVDZ basis-set). The variation of the discrepancies as a function of the QM region size (i.e., number of carbon atoms and corresponding substituents) is explicitly shown.

Pertaining to the cc-pVDZ basis-set (top panel of Figure 10), we can observe that all the QM/ELMO calculations provide interaction energies that are in excellent agreement with the corresponding fully quantum mechanical ones, irrespective of the QM region size. In particular, we can notice that all the discrepancies for the

interaction energies are below 1.0 kcal/mol and, from the second smallest active region, all the differences with the benchmark values even drop below 0.03 kcal/mol. As usual, the results constantly improve as the size of the QM subsystem increases and they converge to the results of the benchmark computations.

Promising results were also obtained in the case of the mixed basis-set aug-cc-pVDZ/cc-pVDZ. In fact, also in this situation, all the $\Delta\Delta E$ values are always lower than 1.0 kcal/mol, with the largest discrepancy observed for the CCSD/ELMO calculation performed with the smallest QM region (0.34 kcal/mol). Furthermore, for QM subsystems larger than 4 carbon atoms (and corresponding substituents), the energy deviations from the results of the fully QM computations are always lower than 0.1 kcal/mol, regardless of the quantum chemical method used for the active region. All of these observations seem to confirm that the use of different basis-sets for the QM and ELMO subsystems does not change the quality of the QM/ELMO results, which could have an important repercussion on calculations performed on larger systems.

At a later stage we considered the complete interaction energy profiles obtained at the QM/ELMO levels with active subunits having different dimensions. Also in this case, without losing generality, we limited our analysis only to the cc-pVDZ basis-set. In analogy with the procedure followed for the S_N2 reaction and the bond dissociation, the QM/ELMO energy profiles were compared point-by-point to the corresponding QM ones. Here we reported the results associated with the HF/ELMO and CCSD(T)/ELMO computations (see Figures 11 and 12, respectively).

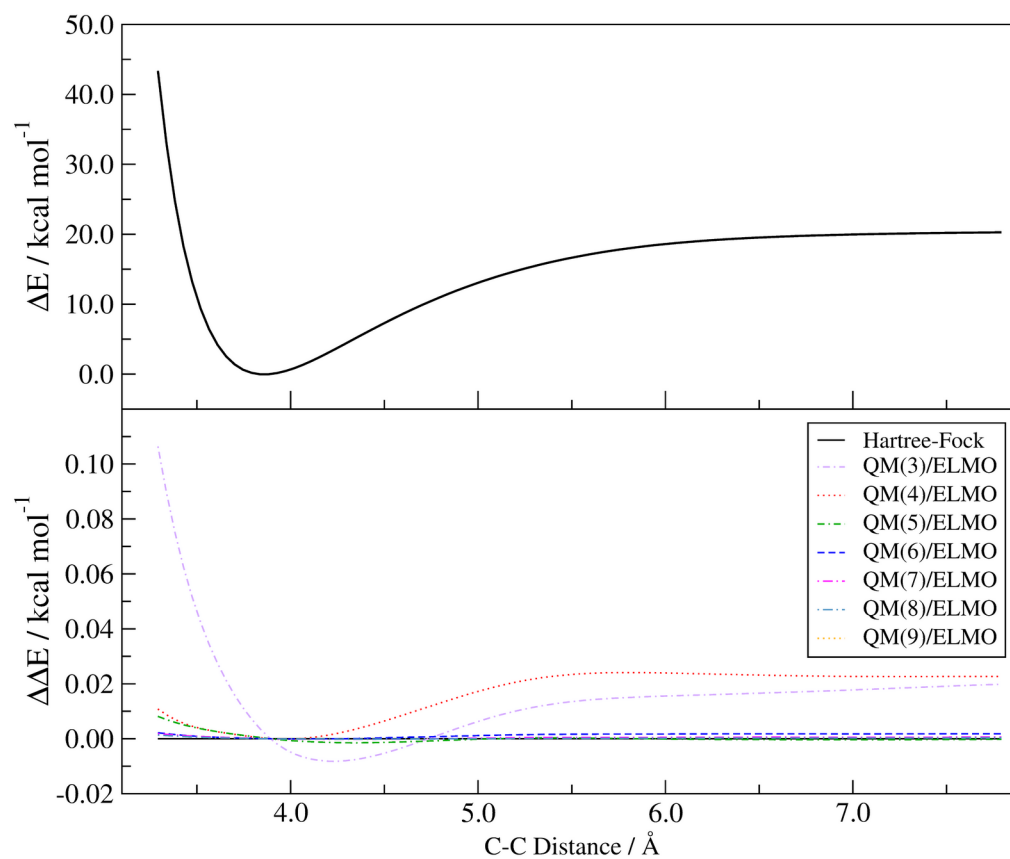


Figure 11. Interaction energy profile for the formic acid - decanoic acid dimer at HF level (top panel) and deviations from it when HF/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves refer to the cc-pVDZ basis-set.

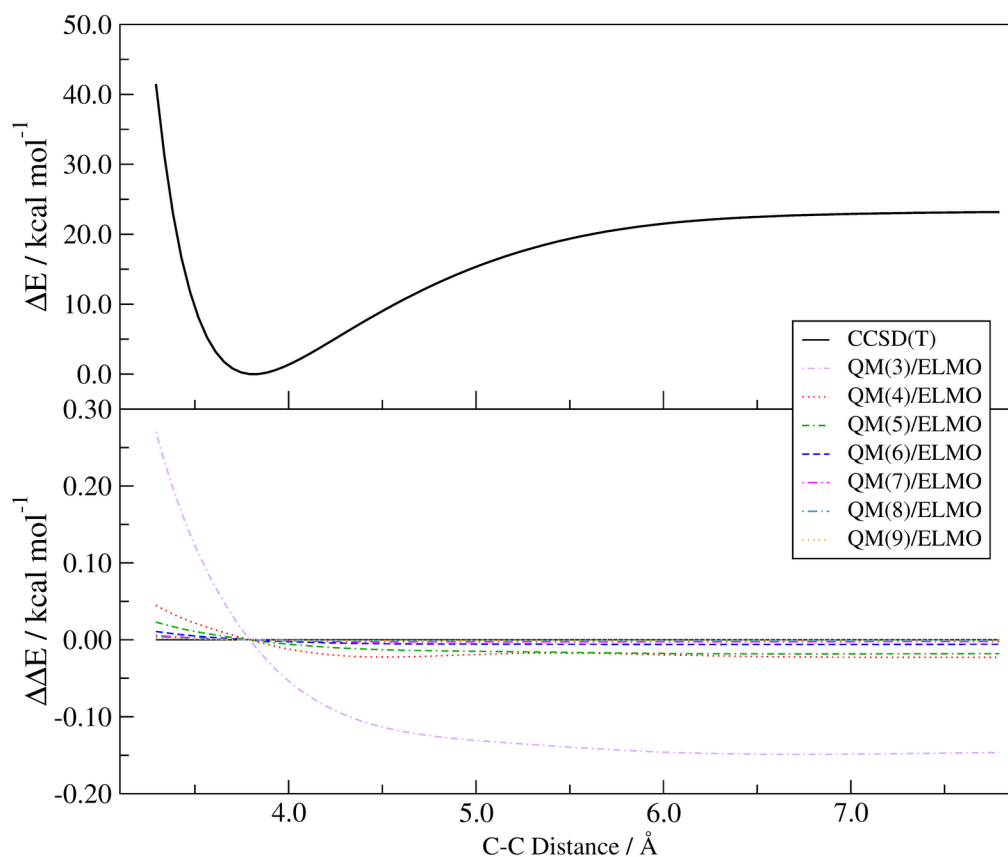


Figure 12. Interaction energy profile for the formic acid - decanoic acid dimer at CCSD(T) level (top panel) and deviations from it when CCSD(T)/ELMO calculations are performed with QM regions of different size (bottom panel); all the curves refer to the cc-pVDZ basis-set.

Concerning the calculations at Hartree-Fock level (see Figure 11), we can see that, even with the smallest QM region, the differences with the corresponding fully QM calculations remain extremely small along the whole interaction energy profile and become very negligible when the QM regions comprise five carbon atoms. A very similar trend can be observed for the computations at CCSD(T)/ELMO level (see Figure 12), where, for all the profiles, the discrepancies from the CCSD(T) benchmark always remain within the limit of chemical accuracy. We can only observe that, in this case, the discrepancies are slightly larger than those observed at Hartree-

Fock level. Completely analogous results were obtained through the B3LYP/ELMO and MP2/ELMO computations (see Figures S6 and S7 in the Supporting Information).

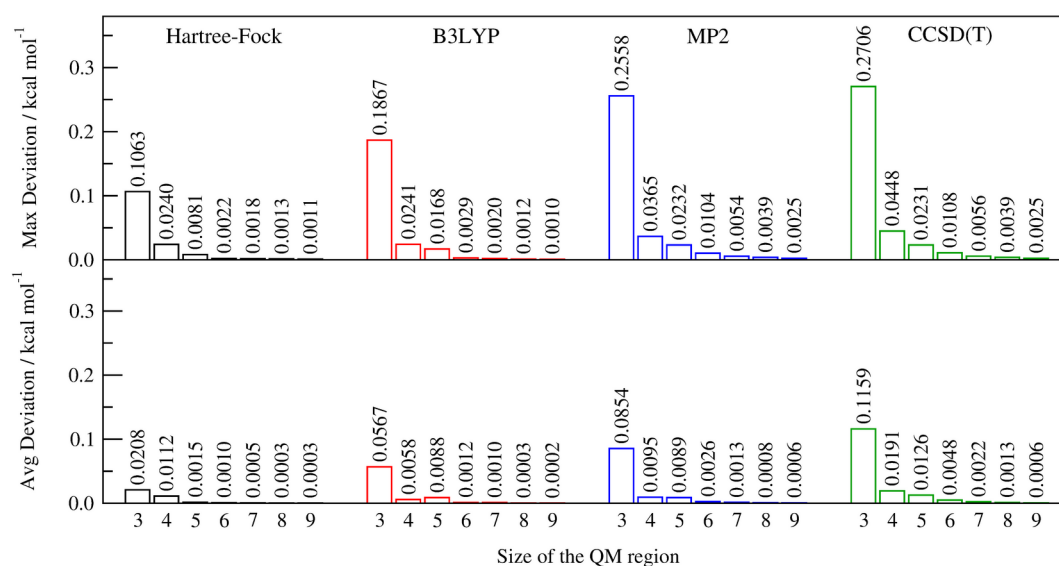


Figure 13 Maximum and average absolute deviations between corresponding QM/ELMO and fully QM interaction energy profiles for the formic acid - decanoic acid dimer (cc-pVDZ basis-set). The variation of the deviations as a function of the QM region size is explicitly shown.

To complete the previous analysis, we also considered the maximum and average absolute deviations along the energy profiles (see Figure 13). To compute the average discrepancies, we chose an averaging-range (unique for all the levels of theory) by adopting a philosophy analogous to the one used in the case of the C-O bond cleavage in sorbitol, with the upper limit corresponding to a point belonging to the plateau region in the CCSD(T) interaction energy curve. The obtained values confirm what was observed above: regardless of the theoretical level chosen for the QM/ELMO computations, the results are in excellent agreement with the corresponding traditional ones, with maximum discrepancies that are well-below the threshold of chemical accuracy. The largest error was observed at CCSD(T)/ELMO level with the smallest

active region and amounts to 0.27 kcal/mol. As usual and as one should expect, in all the cases, the maximum and average absolute deviations decrease as the QM subsystem becomes larger. Analyzing in more detail the results, we also notice that the HF/ELMO computations are the ones that provide the lowest maximum and average discrepancies with respect to the corresponding benchmark values, while the MP2/ELMO and CCSD(T)/ELMO calculations are those that give the largest deviations, although, as already pointed out above, these deviations are always significantly below 1.0 kcal/mol.

D. Computational cost

In this section, we focus on the computational cost of the QM/ELMO method at Coupled Cluster level, which is by far the most computationally expensive of all the QM/ELMO techniques developed so far. This will allow us to better evaluate the reduction of the computational effort by introducing the QM/ELMO embedding approximation in the post-HF methods. To this purpose, in Table 1 we reported the number of occupied (frozen and active) molecular orbitals, the number of virtual molecular orbitals and the CPU times associated with the CCSD(T)/ELMO and fully CCSD(T) computations carried out with the cc-pVDZ basis-set on the transition state structure for the investigated S_N2 reaction (see above).

From the analysis of Table 1 it clearly emerges that, in post-HF/ELMO calculations, the number of involved active occupied molecular orbitals and the number of virtual orbitals are much lower than the same numbers in fully post-HF computations, especially if the QM region remains relatively small. In particular, as we already anticipated in the Theory section, the number of virtual orbitals significantly reduces. For example, if we consider the QM subsystem comprising only three alkyl groups,

which is already a QM subsystem that allowed the achievement of chemical accuracy for all the points along the reaction energy profile at CCSD(T)/ELMO level, we can observe that the number of virtual orbitals only amounts to 85 against 236 for the traditional quantum mechanical calculation.

Table 1. Number of (frozen and active) occupied molecular orbitals (N_{occ}), number of virtual molecular orbitals (N_{virt}) and timings associated with the CCSD(T)/ELMO and CCSD(T) calculations (cc-pVDZ basis-set) performed on the transition-state structure for the S_N2 reaction between 1-bromodecane and chloride anion.^(a)

Calculations	N_{occ}		N_{virt}	CPU time (s)	%
	Frozen	Active			
QM(2)/ELMO	49	18	64	3962.6	1.3
QM(3)/ELMO	46	21	85	6413.0	2.1
QM(4)/ELMO	43	24	106	11207.8	3.7
QM(5)/ELMO	40	27	127	18194.1	5.9
QM(6)/ELMO	37	30	148	33806.8	11.0
QM(7)/ELMO	34	33	169	65108.8	21.2
QM(8)/ELMO	31	36	190	109481.7	35.7
Full QM (10)	24	43	236	306446.5	100.0

^(a) The recorded timings were obtained by performing parallel calculations on 16 Intel Xeon Gold 6130 2.1 GHz processors.

The previous observation has a direct consequence on the computational cost of the CCSD(T)/ELMO computations. In Table 1 we can indeed observe significant savings in terms of CPU time. For instance, always considering the active region that comprises three alkyl groups, we can notice that the CCSD(T)/ELMO calculation takes only 2.1% of the time taken by the corresponding and traditional CCSD(T) computation. As one should expect, the computational cost of the CCSD(T)/ELMO calculations increases as larger QM region are used, but in the case reported in Table

1, the recorded timings never exceed 36% of the CPU time for the full computation. Analogous trends were observed for calculations performed on the minimum energy structures of sorbitol and of the formic acid - decanoic acid dimer (see Tables S1 and S2 in the Supporting Information).

V. CONCLUSIONS

In this work we have presented the extension of the recently proposed quantum mechanics/extremely localized molecular orbital method to correlated techniques (mainly, DFT, Møller-Plesset perturbation theory and Coupled Cluster approach) in order to allow more accurate treatments of the quantum mechanical regions of the investigated systems. Already from a preliminary theoretical analysis of the QM/ELMO equations, it clearly appears that the novel post-HF/ELMO strategies are characterized by a significantly lower computational cost. In fact, compared to the corresponding fully correlated techniques, the QM/ELMO approach intrinsically allows to use not only a reduced number of occupied molecular orbitals, but also, and more importantly, a reduced number of virtual molecular orbitals with which constructing the different terms of the multi-determinant wavefunctions.

The computational advantages of the post-HF/ELMO techniques were confirmed by our preliminary test calculations, which were carried out to further evaluate the performances of the new strategies when they are exploited to describe typical chemical processes and features, such as chemical reactions, bond dissociations and intermolecular interactions. The results have also shown that, despite the simplification introduced through the description of the chemical environment by means of transferred and frozen extremely localized molecular orbitals, the QM/ELMO methods allow to reproduce the results of the corresponding fully

quantum mechanical calculations within chemical accuracy also when quite small QM subsystems are used. Of course, for a more complete and definitive assessment on the potentialities of the presented techniques, we plan to apply them to additional test cases comprising bond-rearrangements reactions (e.g., Claisen and Diels-Alder reactions) and some additional intermolecular interactions covering both dispersion-bound and hydrogen-bonded complexes. A further and more detailed analysis of the basis-set dependence is also envisaged to completely characterize the performances of the QM/ELMO strategy.

However, considering the preliminary results collected in this study and the advantages of the recently constructed ELMO libraries, we could already anticipate future applications of the novel QM/ELMO techniques to more challenging problems related to large biomolecules. In this context, one possibility is to exploit our new embedding strategies to the computational study of enzyme catalysis. To accomplish this task, we are also planning to further extend the QM/ELMO method by including a third Molecular Mechanics level in the calculations, thus giving rise to a QM/ELMO/MM strategy that will enable to treat larger and larger biochemical systems with a relatively low computational effort.

Another envisaged direction is the use of the new embedding strategy to study excited states of macromolecules, such as proteins. To this purpose, the simplest idea is to exploit the QM/ELMO technique by modeling the QM subsystem with well-established wavefunction-based methods for excited states, such as Multi-Reference Configuration Interaction (MRCI),^{84,85} Complete Active Space Self-Consistent Field (CASSCF),^{86,87} CASPT2,⁸⁸⁻⁹⁰ Equation-Of-Motion Coupled Cluster (EOM-CC),⁹¹⁻⁹³ and Linear Response Coupled Cluster (LR-CC),^{94,95} just to cite a few. In this context, another tantalizing perspective could also be the coupling of the basic embedding

Hartree-Fock/ELMO method with the single-determinant approaches for excited states (MOM (Maximum Overlap Method) and IMOM (Initial Maximum Overlap Method)) proposed by Gill and coworkers.⁹⁶⁻⁹⁸

Moreover, in the near future, the newly developed QM/ELMO techniques could be also used in connection with modern methods of quantum crystallography.⁹⁹⁻¹⁰³ Their more immediate application could consist in the coupling with the promising Hirshfeld atom refinement (HAR) strategy¹⁰⁴⁻¹⁰⁶ in order to determine accurate and precise crystal structures of metalloproteins, as an improvement of the recently proposed HAR-ELMO (Hirshfeld atom refinement – extremely localized molecular orbital) approach¹⁰⁷ that already allowed fast, accurate and precise refinements of polypeptide and small protein crystal structures. Another possibility in the framework of quantum crystallography could also be the development and implementation of a novel X-ray restrained/constrained QM/ELMO technique, which will represent the first variant for quite large systems of the X-ray restrained/constrained wavefunction strategies developed by Jayatilaka and others¹⁰⁸⁻¹²² to obtain wavefunctions that fit experimental X-ray diffraction data upon simultaneous minimization of the molecular electronic energy.

ASSOCIATED CONTENT

Supporting Information. Details about the theory, transfer and libraries of extremely localized molecular orbitals (with Figure S1 showing examples of plotted ELMOs and Figure S2 giving the schematic representation of reference frames and atomic triads required for the rotation of ELMOs). Figures S3 and S4 showing the energy profiles of the examined S_N2 reaction at DFT-B3LYP/cc-pVDZ and MP2/cc-pVDZ levels, respectively, and the corresponding differences between fully QM and QM/ELMO

computations. Figure S5 depicting the dissociation energy profile of sorbitol at DFT-B3LYP/cc-pVDZ level with the associated differences between the fully QM and QM/ELMO calculations. Figures S6 and S7 showing the interaction energy profiles for the formic acid – decanoic acid dimer at DFT-B3LYP/cc-pVDZ and MP2/cc-pVDZ levels, respectively, and the corresponding differences between fully QM and QM/ELMO computations. Tables S1 and S2 reporting the computational timings of the Coupled Cluster/ELMO calculations (cc-pVDZ basis-set) performed on the minimum energy structures of sorbitol and of the formic acid - decanoic acid dimer. The Supporting Information is available free of charge on the ACS Publications website.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

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

REFERENCES

- ¹ Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 619-622.
- ² Sherril, C. D.; Schaefer, H. F. The Configuration Interaction Method: Advances in Highly Correlated Approaches. *Adv. Quantum Chem.* **1999**, *34*, 143-269.
- ³ Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291-352.
- ⁴ Olivares-Amaya, R.; Hu, W.; Nakatani, N.; Sharma, S.; Yang, J.; Chan, G. K.-L. The ab-initio density matrix renormalization group in practice. *J. Chem. Phys.* **2015**, *142*, 034102.
- ⁵ Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864-B871.
- ⁶ Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133-A1138.
- ⁷ Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792-794.
- ⁸ Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289-320.
- ⁹ Medvedev, M. G.; Bushmarinov, I. S.; Sun, J.; Perdew, J. P.; Lyssenko, K. A. Density functional theory is straying from the path toward the exact functional. *Science* **2017**, *355*, 49-52.
- ¹⁰ 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.
- ¹¹ 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.

- ¹² 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.
- ¹³ Gao, J. Hybrid Quantum and Molecular Mechanical Simulations: An Alternative Avenue to Solvent Effects in Organic Chemistry. *Acc. Chem. Res.* **1996**, *29*, 298-305.
- ¹⁴ Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198-1229.
- ¹⁵ 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(tBu)₃)₂ + H₂ Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357-19363.
- ¹⁶ Humbel, S.; Sieber, S.; Morokuma, K. The IMOMO Method: Integration of Different Levels of Molecular Orbital Approximations for Geometry Optimization of Large Systems: Test for *n*-Butane Conformation and S_N2 Reaction: RCl+Cl⁻. *J. Chem. Phys.* **1996**, *105*, 1959-1967.
- ¹⁷ Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) method. *J. Comput. Chem.* **2000**, *21*, 1419-1432.
- ¹⁸ 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 Application. *Chem. Rev.* **2015**, *115*, 5678-5796.
- ¹⁹ Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: a route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632-672.
- ²⁰ Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643-5677.
- ²¹ Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. Quantum Mechanical/Quantum Mechanical Methods. I. A Divide and Conquer Strategy for Solving the Schrödinger Equation

for Large Molecular Systems Using a Composite Density Functional-Semiempirical Hamiltonian. *J. Chem. Phys.* **2000**, *113*, 5604-5613.

²² Knizia, G.; Chan, G. K.-L. Density matrix embedding: A simple alternative to dynamical mean-field theory. *Phys. Rev. Lett.* **2012**, *109*, 186404.

²³ Knizia, G.; Chan, G. K.-L. Density matrix embedding: A strong coupling quantum embedding theory. *J. Chem. Theory Comput.* **2013**, *9*, 1428–1432.

²⁴ Bukik, I. W.; Scuseria, G. E.; Dukelsky, J. Density matrix embedding from broken symmetry lattice mean fields. *Phys. Rev. B* **2014**, *89*, 035140.

²⁵ Bulik, I. W.; Chen, W.; Scuseria, G. E. Electron correlation in solids via density embedding theory. *J. Chem. Phys.* **2014**, *141*, 035140.

²⁶ 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.

²⁷ Welborn, M.; Tsuchimochi, T.; Van Voorhis, T. Bootstrap embedding: An internally consistent fragment-based method. *J. Chem. Phys.* **2016**, *145*, 074102.

²⁸ Ye, H.-Z.; Van Voorhis, T. Atom-Based Bootstrap Embedding For Molecules. *J. Phys. Chem. Lett.* **2019**, *10*, 6368-6374.

²⁹ Senatore, G.; Subbaswamy, K. Density dependence of the dielectric constant of rare-gas crystals. *Phys. Rev. B* **1986**, *34*, 5754–5757.

³⁰ Cortona, P. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B* **1991**, *44*, 8454–8458.

³¹ Wesolowski, T. A.; Warshel, A. Frozen density functional approach for ab-initio calculations of solvated molecules. *J. Phys. Chem.* **1993**, *97*, 8050–8053.

³² Wesolowski, T. A. Embedding a Multideterminantal Wave Function in an Orbital-Free Environment. *Phys. Rev. A* **2008**, *77*, 012504.

³³ Pernal, K.; Wesolowski, T. A. Orbital-Free Effective Embedding Potential: Density-Matrix Functional Theory Case. *Int. J. Quantum Chem.* **2009**, *109*, 2520-2525.

- ³⁴ Wesolowski, T. A.; Shedge, S.; Zhou, X. Frozen-Density Embedding Strategy for Multilevel Simulations of Electronic Structure. *Chem. Rev.* **2015**, *115*, 5891-5928.
- ³⁵ Henderson, T. M. Embedding wave function theory in density functional theory. *J. Chem. Phys.* **2006**, *125*, 014105.
- ³⁶ Iannuzzi, M.; Kirchner, B.; Hutter, J. Density functional embedding for molecular systems. *Chem. Phys. Lett.* **2006**, *421*, 16–20.
- ³⁷ 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.
- ³⁸ Roncero, O.; de Lara-Castells, M. P.; Villarreal, P.; Flores, F.; Ortega, J.; Paniagua, M.; Aguado, A. An inversion technique for the calculation of embedding potentials. *J. Chem. Phys.* **2008**, *129*, 184104.
- ³⁹ 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.
- ⁴⁰ Huang, C.; Carter, E. A. Potential-functional embedding theory for molecules and materials. *J. Chem. Phys.* **2011**, *135*, 194104.
- ⁴¹ Huang, C.; Pavone, M.; Carter, E. A. Quantum mechanical embedding theory based on a unique embedding potential. *J. Chem. Phys.* **2011**, *134*, 154110.
- ⁴² Elliott, P.; Cohen, M. H.; Wasserman, A.; Burke, K. Density functional partition theory with fractional occupations. *J. Chem. Theory Comput.* **2009**, *5*, 827–833.
- ⁴³ Elliott, P.; Burke, K.; Cohen, M. H.; Wasserman, A. Partition density-functional theory. *Phys. Rev. A* **2010**, *82*, 024501.
- ⁴⁴ Genova, A.; Ceresoli, D.; Pavanello, M. Periodic subsystem density-functional theory. *J. Chem. Phys.* **2014**, *141*, 174101.
- ⁴⁵ Mi, W.; Pavanello, M. Nonlocal Subsystem Density Functional Theory. *J. Phys. Chem. Lett.*

2020, *11*, 272-279.

⁴⁶ 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.

⁴⁷ 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.

⁴⁸ 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.

⁴⁹ 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.

⁵⁰ 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.

⁵¹ 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.

⁵² 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.

⁵³ 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.

⁵⁴ Welborn, M.; Manby, F. R.; Miller, T. F., III. Even-handed subsystem selection in projection-based embedding. *J. Chem. Phys.* **2018**, *149*, 144101.

⁵⁵ 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.

⁵⁶ 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.

⁵⁷ Chulhai, D. V.; Goodpaster, J. D. Improved Accuracy and Efficiency in Quantum Embedding through Absolute Localization. *J. Chem. Theory Comput.* **2017**, *13*, 1503-1508.

⁵⁸ 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.

⁵⁹ 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.

⁶⁰ 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.

⁶¹ 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.

⁶² Loos, P.-F.; Assfeld, X. Core-Ionized and Core-Excited States of Macromolecules. *Int. J. Quantum Chem.* **2007**, *107*, 2243-2252.

⁶³ Monari, A.; Rivail, J.-L.; Assfeld, X. Theoretical Modeling of Large Molecular Systems. Advances in the Local Self Consistent Field Method for Mixed Quantum Mechanics/Molecular Mechanics Calculations. *Acc. Chem. Res.* **2013**, *46*, 596-603.

⁶⁴ Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169-178.

⁶⁵ 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.

⁶⁶ 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.

⁶⁷ 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.

⁶⁸ 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.

⁶⁹ Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685-698.

⁷⁰ 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.

⁷¹ 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.

⁷² 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.

⁷³ 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.

⁷⁴ 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.

⁷⁵ Peng, C.; Schlegel, H. B. Combining STQN Methods to Find Transition States. *Isr. J. Chem.* **1993**, *33*, 449–454.

⁷⁶ Fukui, K. The Path of Chemical Reactions - The IRC Approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.

⁷⁷ Hratchian, H. P.; Schlegel, H. B. Finding Minima, Transition States, and Following Reaction Pathways in Ab Initio Potential Energy Surfaces. In *Theory and Applications of Computational Chemistry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G., Eds.; Elsevier: Amsterdam, 2005; Chapter 10, pp 195-249.

⁷⁸ Maeda, S.; Harabuchi, Y.; Ono, Y.; Taketsugu, T.; Morokuma, K. Intrinsic Reaction Coordinate: Calculation, Bifurcation, and Automated Search. *Int. J. Quantum Chem.* **2015**, *115*, 258–269.

⁷⁹ Taube, A. G.; Bartlett, R. J. Improving upon CCSD(T): ACCSD(T). I. Potential Energy Surfaces. *J. Chem. Phys.* 2008, *128*, 044110.

⁸⁰ Taube, A. G.; Bartlett, R. J. Improving upon CCSD(T): ACCSD(T). II. Stationary Formulation and Derivatives. *J. Chem. Phys.* 2008, *128*, 044111.

⁸¹ Taube, A. G. Alternative Perturbation Theories for Triple Excitations in Coupled-Cluster Theory. *Mol. Phys.* **2010**, *108* (21–23), 2951–2960.

⁸² 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.

- ⁸³ 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.
- ⁸⁴ Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley and Sons: New York, 2000.
- ⁸⁵ Buenker, R. J.; Peyerimhoff, S. D. CI method for the study of general molecular potentials. *Theor. Chim. Acta* **1968**, *12*, 183-199.
- ⁸⁶ Roos, B. O.; Taylor, P. R.; Siegbahn, P. E. M. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **1980**, *48*, 157-173.
- ⁸⁷ Roos, B. O. The complete active space SCF method in a fock-matrix-based super-CI formulation. *Int. J. Quantum Chem.* **1980**, *18*, 175-189.
- ⁸⁸ Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. Second-order perturbation theory with a CASSCF reference function. *J. Phys. Chem.* **1990**, *94*, 5483-5488.
- ⁸⁹ Andersson, K.; Malmqvist, P. A.; Roos, B. O. Second-order perturbation theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218-1226.
- ⁹⁰ Roos, B. O.; Andersson, K.; Fülcher, M. P.; Malmqvist, P. A.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M. Multiconfigurational perturbation theory: applications in electronic spectroscopy. *Adv. Chem. Phys.* **2007**, *93*, 219-331.
- ⁹¹ Rowe, D. J. Equations-of-Motion Method and the Extended Shell Model. *Rev. Mod. Phys.* **1968**, *40*, 153-166.
- ⁹² Mukherjee, D.; Mukherjee, P. K. A response-function approach to the direct calculation of the transition-energy in a multiple-cluster expansion formalism. *Chem. Phys.* **1979**, *39*, 325-335.
- ⁹³ Sekino, H.; Bartlett, R. J. A linear response, coupled-cluster theory for excitation energy. *Int. J. Quantum Chem.* **1984**, *26*, 255-565.

- ⁹⁴ Dalgaard, E.; Monkhorst, H. J. Some aspects of the time-dependent coupled-cluster approach to dynamic response functions. *Phys. Rev. A* **1983**, *28*, 1217-1222.
- ⁹⁵ Koch, H.; Jørgensen, P. Coupled cluster response functions. *J. Chem. Phys.* **1990**, *93*, 3333-3344.
- ⁹⁶ 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.
- ⁹⁷ Barca, G. M. J.; Gilbert, A. T. B.; Gill, P. M. W. Communication: Hartree-Fock description of excited states of H₂. *J. Chem. Phys.* **2014**, *141*, 111104.
- ⁹⁸ 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.
- ⁹⁹ Genoni, A.; Bučinský, L.; Claiser, N.; Contreras-García, J.; Dittrich, B.; Dominiak, P. M.; Espinosa, E.; Gatti, C.; Giannozzi, P.; Gillet, J.-M.; Jayatilaka, D.; Macchi, P.; Madsen, A. Ø.; Massa, L. J.; Matta, C. F.; Merz, K. M., Jr.; Nakashima, P. N. H.; Ott, H.; Ryde, U.; Schwarz, K.; Sierka, M.; Grabowsky, S. Quantum Crystallography: Current Developments and Future Perspectives. *Chem. Eur. J.* **2018**, *24*, 10881-10905.
- ¹⁰⁰ Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159-4176.
- ¹⁰¹ Novara, R. F.; Genoni, A.; Grabowsky, S. What is Quantum Crystallography? *ChemViews* **2018**, DOI:10.1002/chemv.201800066.
- ¹⁰² Massa, L.; Matta, C. F. Quantum Crystallography: A perspective. *J. Comput. Chem.* **2018**, *39*, 1021-1028.
- ¹⁰³ Tsirelson, V. Early days of quantum crystallography: A personal account. *J. Comput. Chem.* **2018**, *39*, 1029-1037.

- ¹⁰⁴ 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.
- ¹⁰⁵ Capelli, S.C.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld Atom Refinement. *IUCrJ* **2014**, *1*, 361-379.
- ¹⁰⁶ Woźńska, 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.
- ¹⁰⁷ Malaspina, L. A.; Wieduwilt, E. K.; Bergmann, J.; Kleemiss, F.; Meyer, B.; Ruiz-López, M.-F.; Pal, R.; Hupf, E.; Beckmann, J.; Piltz, R. O.; Edwards, A. J.; Grabowsky, S.; Genoni, A. Fast and Accurate Quantum Crystallography: from Small to Large, from Light to Heavy. *J. Phys. Chem. Lett.* **2019**, *10*, 6973-6982.
- ¹⁰⁸ Jayatilaka, D. Wave Function for Beryllium from X-ray Diffraction Data. *Phys. Rev. Lett.* **1998**, *80*, 798-801.
- ¹⁰⁹ Jayatilaka, D.; Grimwood, D. J. Wavefunctions Derived from Experiment. I. Motivation and Theory. *Acta Crystallogr., Sect. A* **2001**, *57*, 76-86.
- ¹¹⁰ Grimwood, D. J.; Jayatilaka, D. Wavefunctions Derived from Experiment. II. A Wavefunction for Oxalic Acid Dihydrate. *Acta Crystallogr., Sect. A* **2001**, *57*, 87-100.
- ¹¹¹ Bytheway, I.; Grimwood, D.; Jayatilaka, D. Wavefunctions Derived from Experiment. III. Topological Analysis of Crystal Fragments. *Acta Crystallogr., Sect. A* **2002**, *58*, 232-243.
- ¹¹² Bytheway, I.; Grimwood, D. J.; Figgis, B. N.; Chandler, G. S.; Jayatiaka, D. Wavefunctions Derived from Experiment. IV. Investigation of the Crystal Environment of Ammonia. *Acta Crystallogr., Sect. A* **2002**, *58*, 244-251.
- ¹¹³ Grimwood, D. J.; Bytheway, I.; Jayatilaka, D. Wavefunctions Derived from Experiment. V. Investigation of Electron Densities, Electrostatic Potentials, and Electron Localization Functions for Noncentrosymmetric Crystals. *J. Comput. Chem.* **2003**, *24*, 470-483.

- ¹¹⁴ Hudák, M.; Jayatilaka, D.; Perainova, L.; Biskupic, S.; Kozisek, J.; Bucinsky, L. X-ray Constrained Unrestricted Hartree–Fock and Douglas–Kroll–Hess Wavefunctions. *Acta Crystallogr., Sect. A* **2010**, *66*, 78–92.
- ¹¹⁵ Genoni, A.; Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data. *J. Phys. Chem. Lett.* **2013**, *4*, 1093–1099.
- ¹¹⁶ Genoni, A.; X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique. *J. Chem. Theory Comput.* **2013**, *9*, 3004–3019.
- ¹¹⁷ Dos Santos, L. H. R.; Genoni, A.; Macchi, P. Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density. *Acta Crystallogr., Sect. A* **2014**, *70*, 532–551.
- ¹¹⁸ Genoni, A.; Meyer, B. X-Ray Constrained Wave Functions: Fundamentals and Effects of the Molecular Orbitals Localization. *Adv. Quantum Chem.* **2016**, *73*, 333–362.
- ¹¹⁹ Genoni, A. A first-prototype multi-determinant X-ray constrained wavefunction approach: the X-ray constrained extremely localized molecular orbital-valence bond method, *Acta Crystallogr., Sect. A* **2017**, *73*, 312–316.
- ¹²⁰ Casati, N.; Genoni, A.; Meyer, B.; Krawczuk, A.; Macchi, P. Exploring charge density analysis in crystals at high pressure: data collection, data analysis and advanced modelling. *Acta Crystallogr., Sect. B* **2017**, *73*, 584–597.
- ¹²¹ Genoni, A.; Franchini, D.; Pieraccini, S.; Sironi, M. X-ray Constrained Spin-Coupled Wavefunction: a New Tool to Extract Chemical Information from X-ray Diffraction Data. *Chem. Eur. J.* **2018**, *24*, 15507–15511.
- ¹²² Genoni, A.; Macetti, G.; Franchini, D.; Pieraccini, S.; Sironi, M. X-ray constrained spin-coupled technique: theoretical details and further assessment of the method. *Acta Crystallogr., Sect. A* **2019**, *75*, 778–797.