Three-Layer Multiscale Approach Based on Extremely Localized Molecular Orbitals to Investigate Enzyme Reactions

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

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

* Correspondence to:

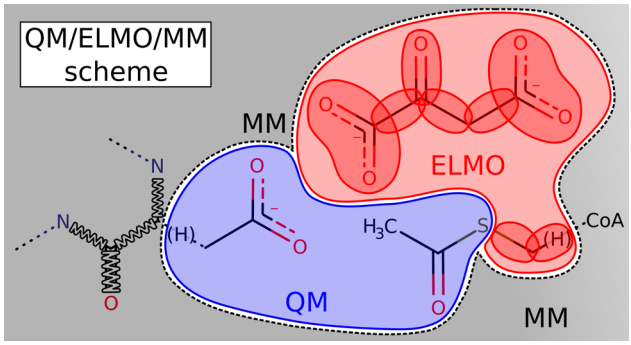
Alessandro Genoni, Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078 Metz, France. E-mail: alessandro.genoni@univ-lorraine.fr; Phone: +33 (0)3 72 74 91 70; Fax: +33 (0)3 72 74 91 87.

ABSTRACT

Quantum mechanics / molecular mechanics (QM/MM) calculations are undoubtedly the most widely used multi-scale embedding techniques to computationally investigate large biological systems and particularly enzyme reactions. In most QM/MM computations, the quantum mechanical region is treated through density functional theory (DFT), which is the quantum chemistry method that offers the best compromise in terms of accuracy and computational cost. Nevertheless, DFT calculations are also unavoidably characterized by a certain degree of variability and uncertainty associated with the choice of the functional with which the computations are performed. Therefore, if one really wants to be sure of achieving chemical accuracy through QM/MM calculations, it would be desirable to treat the QM region by means of a high-level wave function-based method (e.g., Coupled Cluster). Obviously, the drawback of this choice is the larger computational cost, which could become too high already for the relatively small QM subsystems usually considered in QM/MM computations. In this work, to overcome this drawback, we propose the coupling of the recently developed fully quantum mechanical QM/ELMO (quantum mechanics / extremely localized molecular orbital) embedding approach with molecular mechanics, namely we introduce the new three-layer QM/ELMO/MM technique. The QM/ELMO strategy is a reliable and accurate embedding method in which only the most chemically relevant part of the system is treated at quantum mechanical level, while the rest is described through frozen extremely localized molecular orbitals previously transferred from proper libraries or tailor-made model molecules. Since the QM/ELMO method reproduces the results of corresponding fully-QM computations within chemical accuracy and with a much lower computational effort, it can be also considered a suitable strategy to extend the range of applicability and accuracy of the QM/MM

scheme without increasing the computational cost. In this paper, in addition to briefly presenting the theoretical bases of the new QM/ELMO/MM technique, we will also show and discuss its validation, which was conducted by performing a series of benchmark calculations on the well-known and tested deprotonation of acetyl coenzyme A by aspartate in citrate synthase.

TOC GRAPHICS



KEYWORDS: Multi-scale embedding methods, QM/MM techniques, Extremely Localized Molecular Orbitals, QM/ELMO strategy, enzyme reactions

1. INTRODUCTION

Nowadays it is more and more clear that accurate modeling of biological macromolecules and particularly of enzyme reactions requires the use of advanced electronic structure calculations. Nevertheless, it is also well known that the cost of these computations unfavorably scales with the size of the systems under examination and, for this reason, their use to get insights into problems of biological/biochemical interest is possible only if suitable approximations are introduced.[1,2]

There are different ways of reducing the high computational scaling associated with the sophisticated methods of quantum chemistry. One possibility is offered by the so-called fragmentation techniques. In these cases, a large system is partitioned into smaller fragments and less computationally expensive quantum chemical calculations are carried out for all the subunits. Afterwards, the results obtained on the different subsystems are properly combined to obtain the quantity of interest (e.g., energy or electron density) for the investigated macromolecule. In this context, we can mention strategies such as the Divide & Conquer method,[3-8] the molecular tailoring approach[9-13] and the so-called "fragment interaction techniques" (e.g, fragment molecular orbital (FMO) approach,[14-18] kernel energy method[19-26] (KEM) and molecular fractionation with conjugated caps (MFCC) strategy[27-34]).

Another group of methods that provide wave functions and electron densities of large systems at a significantly reduced computational cost are those based on databanks of electron densities, density matrices or molecular orbitals. They are conceptually similar to the fragmentation techniques discussed above and rely on the transferability principle, namely on the observation that molecules are generally constituted of units (e.g., functional groups) whose main properties remain almost invariant in different environments. One of the first approaches of this kind was the MEDLA (molecular

electron density LEGO assembler) technique[35,36] based on libraries of fuzzy electron densities, a technique that has been afterwards extended to databanks of density matrices giving rise to the ADMA (adjustable density matrix assembler) method.[37-39] In this framework, it is also worth mentioning the new libraries of extremely localized molecular orbitals (ELMOs),[40-42] which also play a pivotal role in the method proposed in this paper. ELMOs are molecular orbitals strictly localized on small molecular fragments (i.e., atoms, bonds and functional groups).[43-45] They are characterized by a reliable transferability from molecule to molecule and can be thus considered as plausible electronic LEGO building blocks.[40,41,45-49] Exploiting this property, databanks of extremely localized molecular orbitals covering all the possible elementary fragments of the twenty natural amino acids have been constructed in order to almost instantaneously obtain approximate wave functions and electron densities of large systems.[42] Interestingly, they have been also recently used to successfully refine crystal structures of polypeptides and proteins through the novel HAR-ELMO (Hirshfeld atom refinement – extremely localized molecular orbital) approach[50] of quantum crystallography.[51-55]

A third way to investigate large biological systems at quantum mechanical level is represented by the embedding methods, where only a small region of the macrosystem is described at fully (even very-high) quantum mechanical level, while the remaining part is generally treated at a lower level of theory. The observation at the basis of this approximation is that, in large biomolecules, only a small part (e.g., the active sites in proteins) generally determines the properties of interest (such as, reaction barriers and energies in enzyme reactions), while the rest of the system only slightly influences the chemically relevant region. The most popular approaches in this category are the quantum mechanics / molecular mechanics (QM/MM) techniques.[56-60] These methods

are multiscale approaches where the fully quantum mechanical treatment of the most important part of the system under exam is combined with a force field-based description of the environment. They are widely used to investigate biochemical problems[61] and their relevance has been recently recognized by the award of the 2013 Nobel Prize in Chemistry.[62-64] Within this context, the energy of the full system can be computed in two different ways. The most widely used option corresponds to the fully Hamiltonian scheme, where the total energy is decomposed as the sum of three terms: the energy for the QM region, the energy of the MM subunit, and the energy due to the interaction between the QM and MM parts. The other possibility consists in the so-called subtractive scheme, where, after computing the energy of the whole system at MM level, the energies of the QM region obtained through a fully quantum chemical method and through a molecular mechanics force field are added and subtracted, respectively. A well-known representative example of subtractive scheme is the ONIOM strategy[65-68] proposed by Morokuma and his collaborators. In that case, the system of interest can be partitioned into multiple subunits (generally two or three) described at different levels of theory, which, if desired, can be all quantum mechanical. For this reason, the ONIOM technique may be also considered as one of the first examples of QM/QM' strategy.

More advanced methods are those based on sophisticated quantum embeddings.[69] Among them we can mention the density matrix-based strategies, such as the density matrix embedding theory (DMET) developed by Chan and collaborators[70,71] or the latest bootstrap embedding (BE) technique proposed by the van Voorhis lab.[72,73] Other relevant techniques are also the pioneering frozen-density embedding theory (FDET) initially developed by Wesolowski and Warshel,[74-77] the multilevel approaches introduced by Koch and collaborators,[78-80] and the projection-based embedding (PbE)

method devised by Miller, Manby and coworkers.[81-87] In particular, through the PbE method it is possible to perform accurate, but computationally advantageous, high-level wave function calculations embedded by DFT potentials (namely, wave function-in-DFT (WF-in-DFT) computations) by simply exploiting a suitable projection operator that enforces the orthogonality between the orbitals associated with the two subunits into which the system was initially partitioned.

However, notwithstanding the large number of techniques developed over the years to treat large biological systems quantum mechanically, in most cases the method of choice remains the QM/MM approach, especially for the study of enzyme reactions. Of course, the accuracy of the QM/MM calculations depends on several factors, such as the size of the QM region and the chosen quantum mechanical method. As one should expect, in order to obtain reliable results, the QM part must be sufficiently large to include the most important interactions in the chemically relevant region of the system. At the same time, the higher the level of theory for the QM region, the better the result is. To reach a good trade-off between accuracy and computational cost, QM/MM calcualtions are generally performed adopting a DFT level to describe the QM subunit. Nevertheless, this necessarily introduces a certain degree of variability associated with the choice of the exchange-correlation (XC) functionals to be used in the calculations, which sometimes leads to different and inconsistent results.[88-93] Therefore, it is clear that even in the framework of the QM/MM techniques, if one wants to achieve chemical accuracy, it would be desirable to resort to post-Hartree-Fock (post-HF) methods, which automatically remove the ambiguity associated with the XC functionals, are systematically improvable, and consequently provide better predictions for the biochemical reactions under exam. The obvious drawback is the large computational

cost of the post-HF techniques, which could lead to expensive QM/MM calculations already when relatively small QM regions are adopted.

To solve this problem, Mulholland and coworkers have recently devised an interesting solution. They coupled the projection-based embedding approach with molecular mechanics,[94,95] thus leading to the new multi-level WF-in-DFT/MM strategy in which i) a very accurate wave function method (e.g., Coupled Cluster with single and double substitutions plus perturbative triples, CCSD(T)) is used to treat the most chemically important part of the system, ii) DFT is exploited to describe those subunits that do not directly participate in the reaction but that may influence the electronic structure of the chemically relevant region, and iii) molecular mechanics is exploited to deal with the protein environment and its longer range effects. The technique has been tested on benchmark enzyme reactions and it seems really promising since it always provided reliable reaction barriers and energies at a moderate computational cost. A similar philosophy has been recently followed also by Koch *et al*., who interfaced their multilevel methods with molecular mechanics to study electronic excited states in solutions.[96]

In line with the strategies just mentioned in the previous paragraph, in this paper we propose the new three-layer approach QM/ELMO/MM, which results from the combination of our recently developed QM/ELMO technique[97-101] with molecular mechanics following a fully Hamiltonian scheme. QM/ELMO is a quantum mechanical embedding method where the most important region of the system can be treated with any traditional approach of quantum chemistry, while the remaining part is described through frozen extremely localized molecular orbitals previously transferred from the above-mentioned ELMO libraries[42] or from suitable model molecules. Recent test calculations have clearly shown that the QM/ELMO technique gives results that are

within chemically accuracy with respect to the outcomes of corresponding fully QM computations for both ground and excited states, even when small quantum mechanical regions are taken into account.[98,99] In other words, the new QM/ELMO approach keeps the accuracy of the corresponding quantum chemical methods but with a significantly lower computational effort. This is especially true when high-level wave function techniques (such as, Coupled Cluster or Equation-of-Motion Coupled Cluster) are used.[98,99] For this reason, the QM/ELMO strategy can be considered as a perfect fully quantum mechanical embedding method to be coupled with force fields. In this way, it will be possible to improve the accuracy of QM/MM calculations, but without raising their cost in terms of CPU time, and, at the same time, the range of applicability of QM/MM methods will be automatically extended.

Here, after presenting the main theoretical aspects at the basis of the QM/ELMO/MM technique, the validation of the new approach will be shown and discussed. In particular, after comparing the results of QM/ELMO/MM and traditional QM/MM computations for a well-characterized enzyme reaction, we will examine the effect of varying the level of theory to describe the quantum mechanical region. The computational cost of the method and comparisons to the WF-in-DFT/MM technique will be also considered. In the final section of the paper, we will draw general conclusions and we will discuss possible future applications and perspectives of the new strategy.

2. THEORY

2.1 General strategy. Before performing a QM/ELMO/MM calculation, the system must be subdivided into three subsystems: the QM, ELMO and MM regions (see Figure 1). If the boundaries between the QM and ELMO subunits correspond to covalent

bonds, the two regions share only the frontier atoms (indicated as E and E′ in Figure 1) and the frontier covalent bonds are described through properly transferred and frozen extremely localized molecular orbitals (as in the parent fully quantum mechanical QM/ELMO approach). On the contrary, if covalent boundaries exist between the ELMO and MM subsystems, the link atom strategy is used (see again Figure 1).

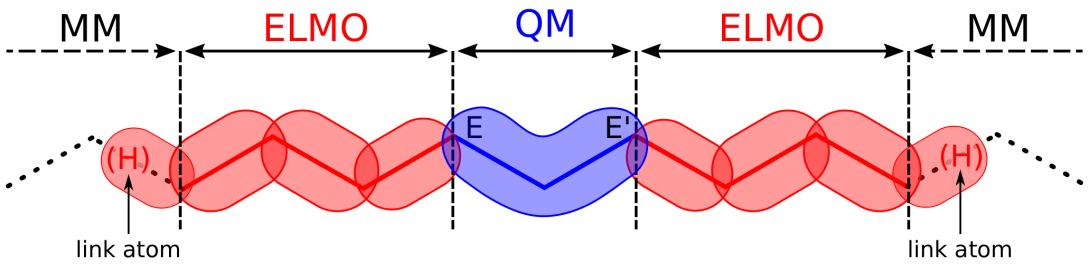


Figure 1. Schematic representation of the QM, ELMO and MM regions in the QM/ELMO/MM method. *E* and *E'* are the frontier atoms between the QM and ELMO subsystems.

As a consequence of the partitioning into the three different subunits, the total energy can be simply written as follows:

$$E = E_{QM/ELMO} + E_{MM} + E_{QM/ELMO/MM} \qquad (1).$$

$E_{QM/ELMO}$ is the purely quantum mechanical energy associated with the QM and ELMO regions. E_{MM} is the purely classical energy corresponding to the MM subsystem, namely the energy associated with geometrical terms, van der Waals repulsion-dispersion and electrostatic interactions from a standard force field. Finally, $E_{QM/ELMO/MM}$ is a hybrid term that comprises different classical energy contributions due to the interactions of the MM region with the QM and ELMO subunits (namely, electrostatic, van der Waals and, in case of covalent boundaries between the ELMO and MM subsystems, bonded interactions). In our method, the last term is evaluated as in usual QM/MM approaches that exploit the link atom strategy. In the following subsection, we will discuss the non-conventional $E_{QM/ELMO}$ term.

The QM/ELMO/MM method outlined above has been implemented by properly coupling an in-house modified version of the *Gaussian09* quantum chemistry software[102] with the Molecular Dynamics (MD) package *AMBER 2016*[103]. The former deals with the quantum contributions ($E_{QM/ELMO}$ term in equation (1), see below and Supporting Information for more details), while the latter handles the classical force-field contributions (E_{MM} and $E_{QM/ELMO/MM}$ terms in equation (1)).

2.2 Energy of the QM/ELMO subsystem. To compute the energy corresponding to the QM and ELMO regions, it is necessary to resort to the QM/ELMO approach. Since this technique has been already introduced and thoroughly described in previous papers, here we will simply limit to recall its essential features. Interested readers can find more specific details in the Supporting Information or in the seminal works about the QM/ELMO strategy.[97,98,101]

Overall, the QM/ELMO method can be seen as composed of three/four different parts: i) transfer to the ELMO region of the necessary extremely localized molecular orbitals that are contained in the current libraries or that are obtained through calculations on tailor-made model molecules; ii) preliminary orthogonalization of molecular orbitals and basis functions for the real calculation; iii) QM/ELMO self-consistent field (SCF) algorithm; iv) subsequent post-HF computation (if necessary).

The orthogonalization consists in three distinct steps (see more details in the Supporting Information) that lead to the definition of a transformation matrix **B** that plays a crucial role in the QM/ELMO SCF iterations.

Afterwards, the QM/ELMO SCF cycle starts with the construction of the Fock matrix in the supermolecular basis-set, namely in the original basis-set for the whole QM/ELMO region. In particular, in case of a QM/ELMO/MM computation, the Fock matrix **F** in the original basis has this form:

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

with $\hat{h}^{core}$ as the usual core one-electron Hamiltonian operator, $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$ as the QM and ELMO one-electron reduced density matrices in the original basis-set, respectively, $(\chi_\alpha \, \chi_\beta | \, \chi_\gamma \, \chi_\delta)$ as a generic two-electron repulsion integral, x as the fraction of exact exchange, and $\langle \chi_\mu | \hat{v}^{XC} [\mathbf{P}^{QM} + \mathbf{P}^{ELMO}] | \chi_\nu \rangle$ as the $v_{\mu\nu}^{XC}$ element of the exchange-correlation potential matrix, which depends on the global one-electron reduced density matrix **P** given by the sum of $\mathbf{P}^{QM}$ and $\mathbf{P}^{ELMO}$. Obviously, when a Hartree-Fock/ELMO/MM (HF/ELMO/MM) calculation is carried out, x becomes equal to 1 and the exchange-correlation contribution disappears. Furthermore, and more importantly for this work, unlike the original version of the QM/ELMO approach,[97-99] here we also have an additional one-electron term (i.e., the fifth one in the right-hand sides of equation (2), namely $F_{\mu\nu}^{MM} = \sum_{K \in MM} \left\langle \chi_\mu \middle| \frac{q_K}{R_{iK}} \middle| \chi_\nu \right\rangle$) that accounts for the electrostatic interactions between the classical point charges of the MM subunit and the electrons of the QM and ELMO regions.

The Fock matrix in the supermolecular basis-set is then transformed to the reduced orthogonal basis of the QM region by exploiting the transformation matrix **B** resulting from the above-mentioned orthogonalization procedure. The new reduced Fock matrix in the orthogonal basis is diagonalized and the obtained molecular orbitals are finally

transformed back to the original set of basis functions to compute the QM one-electron density matrix $\mathbf{P}^{QM}$, which is used to update the Fock matrix **F** (see the second and fourth term in the right-hand sides of equation (2)) in the following SCF iteration. The cycle is iterated until convergence is reached in energy and density matrix.

Finally, in case of a post-HF/ELMO/MM calculation, the global wavefunction describing the QM and ELMO regions is a linear combination of Slater determinants obtained through excitations from occupied to virtual orbitals of the only QM subsystem, namely from occupied to virtual orbitals resulting from the diagonalization of the reduced Fock matrix in the orthogonal basis (see previous paragraph). In other words, the substitutions from the frozen extremely localized molecular orbitals are not taken into account, thus already leading to a decrease of the computational cost. However, in our approach the most significant reduction of the computational effort originates from the use of a smaller and more compact set of virtual molecular orbitals (i.e., the virtual orbitals of the only QM region). To this regard, a detailed analysis of the computational cost associated with Coupled Cluster/ELMO/MM calculations will be discussed in subsection 4.2 of this paper.

3. COMPUTATIONAL DETAILS

3.1 Investigated enzyme reaction. To assess capabilities and performances of the novel QM/ELMO/MM technique, we considered the deprotonation of acetyl coenzyme A (Ac-CoA) by citrate synthase, which represents the first step in the mechanism of the reaction that leads to the formation of citrate and coenzyme A from Ac-CoA and oxaloacetate (first reaction of the Krebs cycle). In fact, after an aspartate residue (i.e., Asp375) deprotonates the α-carbon of acetyl coenzyme A, the resulting enolate intermediate acts as a nucleophile and attacks the carbonyl carbon of oxaloacetate

(OAA), thus giving the citryl-CoA intermediate. Finally, through a hydrolysis reaction, citryl-CoA is converted to citrate and coenzyme A.

To test our method on the deprotonation of Ac-CoA by Asp375 in citrate synthase, we used geometries previously optimized at QM/MM level (particularly, at B3LYP/6-31+G(d) // CHARMM27 level) by van der Kamp and coworkers.[104] These geometries resulted from the definition of a suitable reaction coordinate ($RC = d(C_{Ac-CoA} - H) - d(O_{Asp375} - H)$) able to represent the reaction path accurately. In particular, we employed geometries corresponding to RC between -1.4 Å and 1.4 Å with a step of 0.1 Å.

3.2 QM/ELMO/MM and QM/MM calculations. For our validation tests, we performed QM/ELMO/MM calculations on the geometries mentioned in the previous subsection, with the QM region treated at RHF, B3LYP, M06-2X, MP2 (second-order Møller-Plesset perturbation theory), CCSD and CCSD(T) level, the ELMO subsystem described by means of properly transferred extremely localized molecular orbitals (see subsection 3.3 for more details), and the MM subunit treated with the ff14SB force field[105] for all the protein atoms and with the Generalized Amber Force Field[106] (GAFF) for the atoms of the other molecules (Ac-CoA and oxaloacetate).

To evaluate how the dimensions of the QM and ELMO regions influence the QM/ELMO/MM calculations, four possible partitioning schemes were adopted (see Figure 2):

i) scheme S1, with the QM region corresponding to the carboxylate group of Asp375 plus the acetyl group of Ac-CoA; the ELMO subsystem consisting of the whole OAA molecule, the remaining part of the Asp375 sidechain capped with a hydrogen link atom, and the $C(2)-S(3)-C(4)H_2$ moiety of acetyl coenzyme A

(also capped with a hydrogen link atom); and the MM part comprising the rest of the system (for the atomic labels, see again Figure 2);

ii) scheme S2, with the QM region corresponding to the sidechain of Asp375 capped with a hydrogen link atom plus the methyl-thioester subunit of Ac-CoA; the ELMO subsystem consisting of the whole OAA molecule, and the $S(3)-C(4)H_2$ moiety of acetyl coenzyme A capped with a hydrogen link atom; and the MM part including the remaining part of the system;

iii) scheme S3, with the QM region corresponding to the Asp375 sidechain capped with a hydrogen link atom, the methyl-thioester subunit plus the additional fragment $S(3)-C(4)H_2$ of Ac-CoA (also capped with a hydrogen link atom); the ELMO subsystem consisting of the whole OAA molecule; and with the MM part comprising the rest of the system.

iv) scheme S3-MM, with the QM region as in S3, but with the oxaloacetate molecule completely treated at molecular mechanics level; in other words, the calculations with this scheme corresponded to standard QM/MM computations with the largest QM subsystem among those adopted for the QM/ELMO/MM cases.

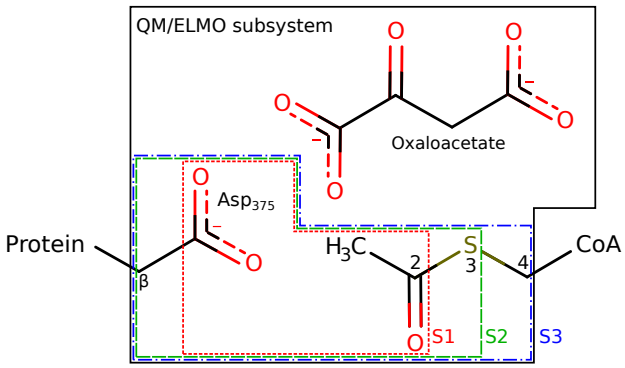


Figure 2. Partitioning schemes adopted in the performed QM/ELMO/MM calculations. The black (continuous) frame contains the whole QM/ELMO subsystem; the red (small dashed) frame contains the QM region for partitioning scheme S1; the green (large dashed) frame contains the QM region for partitioning scheme S2; the blue (dotted and dashed) frame contains the QM region for partitioning schemes S3 and for the QM/MM calculation with scheme S3-MM (oxaloacetate in the MM subsystem).

As we will discuss in detail in the next section, the calculations in this study were carried out using three basis-sets of increasing size: cc-pVDZ, aug-cc-pVDZ and cc-pVTZ, with the first one that was adopted to perform the preliminary benchmark computations and to evaluate the best partitioning schemes among those described above.

All the QM/ELMO/MM and QM/MM computations were carried out by exploiting our in-house modified version of *Gaussian09*[102] that implements the QM/ELMO strategy[97-101] and that is coupled with *AMBER 2016*[103] for the treatment of the classical terms in equation (1).

3.3 ELMOs calculation and transfer. The extremely localized molecular orbitals used in all the QM/ELMO/MM calculations of this study (cc-pVDZ, aug-cc-pVDZ and cc-pVTZ basis-sets) were transferred from the recently constructed ELMO libraries[42] or from tailor-made model molecules on which the desired ELMOs were previously computed by exploiting the Stoll technique[43] implemented[44] in a modified version of the GAMESS-UK quantum chemistry suite of programs[107] (more details about the ELMO theory and libraries are provided in the Supporting Information).

In particular, the extremely localized molecular orbitals used to describe the ELMO regions associated with the sidechain of Asp375 were directly transferred from our ELMO databanks, while the extremely localized molecular orbitals employed to treat the ELMO regions corresponding to acetyl coenzyme A were obtained through preliminary ELMO calculations on a model molecule (see Figure S3 in the Supporting information) that properly mimics the chemical environment of the fragments in the ELMO subsystem. The ELMOs for oxaloacetate were directly determined on the

isolated OAA molecule. Before performing the necessary ELMO computations, the geometries of all the model molecules were optimized at B3LYP/cc-pVDZ level.

The transfers of the ELMOs for all the QM/ELMO/MM calculations were carried out by exploiting the *ELMOdb* software[42] associated with the ELMO databanks, a program that also implements the Philipp and Friesner strategy[40,108] to rotate strictly localized bond orbitals (more details about this technique are also given in the Supporting Information).

4. TEST CALCULATIONS

In this section we will present and discuss the results of the test calculations that were performed to assess performances and capabilities of the QM/ELMO/MM method. At first, by analyzing the results of some benchmark computations, we will evaluate how the size of the QM and ELMO regions influence the accuracy of the results (subsection 4.1). Afterwards, we will analyze the computational cost of the QM/ELMO/MM technique at Coupled Cluster level, especially as a function of the size of the quantum mechanical and ELMO subsystems (subsection 4.2). Finally, we will focus on the impact of the chosen quantum chemical level of theory to describe the QM subunit (subsection 4.3) and we will compare our results to those obtained through other methods (subsection 4.4).

4.1 Benchmark calculations. To start validating the new QM/ELMO/MM approach, we computed potential energy profiles for the examined reaction considering different levels of theory for the quantum mechanical subsystem (RHF, B3LYP, M06-2X, MP2, CCSD and CCSD(T)), using the cc-pVDZ basis-set for the whole QM/ELMO region, and adopting the partitioning schemes S1, S2 and S3 described in the section dedicated to the computational details (in particular, see subsection 3.2). These schemes were

chosen in such a way that the size of the QM subunit gradually increased from S1 to S3, while the one of the ELMO subsystem gradually decreased, in a sort of squeezebox trend. Therefore, the size of the global QM/ELMO region was constant among the partitioning schemes and, consequently, also the dimension of the MM region remained unchanged. In this way it was possible to evaluate the importance of the fully-QM region and to determine its proper size for the enzyme reaction under investigation. The references for our computations were the corresponding QM/MM calculations (with QM = RHF, B3LYP, M06-2X, MP2, CCSD and CCSD(T)) that used a partitioning scheme as the one with the largest QM region for the QM/ELMO/MM computations (scheme S3), but with also the oxaloacetate molecule treated at fully quantum mechanical level. Hereinafter, we will refer to these QM/MM calculations also as full-QM/MM calculations.

First of all, we focused on reaction barriers (ΔE_{act}) and reaction energies (ΔE_r), for which, in Tables 1 and 2, we have respectively reported the absolute discrepancies of the QM/ELMO/MM results (indicated as $\Delta\Delta E_x(S1)$, $\Delta\Delta E_x(S2)$, $\Delta\Delta E_x(S3)$, with $x = act$ or r) from the corresponding full-QM/MM reference values (indicated as $\Delta E_x(QM/MM)$, again with $x = act$ or r).

Let us initially focus on the reaction barriers (see Table 1). For all the quantum mechanical levels chosen to treat the QM region, partitioning scheme S1 (namely, the one with the smallest QM subsystem) provided the least accurate results, with deviations from the full-QM/MM reference values that are in absolute values always greater than 1.0 kcal/mol (chemical accuracy limit). However, the situation greatly improved already when the intermediate S2 pattern was considered, being all the absolute discrepancies within the chemical accuracy threshold. The largest and smallest deviations were observed when the M06-2X (0.70 kcal/mol) and RHF (0.09 kcal/mol)

levels were used, respectively, while all the correlated post-HF strategies gave $\Delta\Delta E_{act}$ values lower than 0.3 kcal/mol. Then, except for the Hartree-Fock case, the description further improved with the S3 partitioning scheme. In particular, when the M06-2X level is considered, the discrepancy decreases to 0.23 kcal/mol, while, when the MP2, CCSD and CCSD(T) methods are used to describe the quantum mechanical region, the absolute deviations from the full-QM/MM results reduce to very small values that are even lower than 0.1 kcal/mol. Finally, as already mentioned in subsection 3.2, to evaluate the importance of treating the whole OAA molecule at ELMO level, we have also performed additional QM/MM calculations with partitioning scheme S3-MM, where the QM subsystem corresponded to the one used in the QM/ELMO/MM computations with scheme S3, but with oxaloacetate included in the MM region. In Table 1, it is possible to notice that this approximation always led to worse results compared to the corresponding QM/ELMO/MM cases based on the S3 and S2 (with the exception of the DFT cases) subdivision patterns. This indicates the need of describing the oxaloacetate molecule at least at an approximate quantum mechanical level, such as the one given by transferred and frozen extremely localized molecular orbitals considered in this study.

Table 1. Reaction barriers obtained at the reference full-QM/MM levels ($\Delta E_{act}(QM/MM)$) and absolute discrepancies from these values ($\Delta\Delta E_{act}(SX)$), as resulting from QM/ELMO/MM and QM/MM calculations with different partitioning schemes.[a,b]

QM method	$\Delta E_{act}(QM/MM)$	$\Delta\Delta E_{act}(S1)$	$\Delta\Delta E_{act}(S2)$	$\Delta\Delta E_{act}(S3)$	$\Delta\Delta E_{act}(S3-MM)$
RHF	19.01	2.67	0.09	0.23	0.91
B3LYP	7.36	3.09	0.52	0.19	0.46
M06-2X	6.40	3.18	0.70	0.23	0.59
MP2	8.52	1.59	0.29	0.03	0.46
CCSD	12.07	2.37	0.19	0.07	0.46
CCSD(T)	10.55	2.12	0.27	0.05	0.42

[a] All energy values in kcal/mol; [b] all calculations with basis-set cc-pVDZ.

The results obtained for the reaction energies are given in Table 2. Unlike the reaction barriers, the trends are less linear. First of all, the QM/ELMO/MM calculations with partitioning scheme S1 provided absolute deviations from the full-QM/MM values that are greater than 1.0 kcal/mol for the Hartree-Fock and DFT cases, but lower than the chemical accuracy limit for the post-HF strategies. Concerning the intermediate partitioning scheme S2, the absolute deviations increase for the calculations that adopted RHF, MP2 or Coupled Cluster levels for the quantum mechanical region, while the description significantly improved in the B3LYP and M06-2X cases. Finally, except when the RHF method was used for the QM subsystem, all the other QM/ELMO/MM calculations with scheme S3 provided $\Delta\Delta E_r$ discrepancies always lower than 1.0 kcal/mol, with the lowest one obtained at M06-2X level. If the oxaloacetate molecule is described only at MM level (scheme S3-MM), the absolute deviations are again always larger than the QM/ELMO/MM ones with scheme S3 and, more importantly, even almost always greater than those obtained by adopting the S2 partitioning, despite a larger size of the QM region in the S3-MM pattern.

Table 2. Reaction energies obtained at the reference full-QM/MM levels ($\Delta E_r(QM/MM)$) and absolute discrepancies from these values ($\Delta\Delta E_r(SX)$), as resulting from QM/ELMO/MM and QM/MM calculations with different partitioning schemes.[a,b]

QM level	$\Delta E_r(QM/MM)$	$\Delta\Delta E_r(S1)$	$\Delta\Delta E_r(S2)$	$\Delta\Delta E_r(S3)$	$\Delta\Delta E_r(S3-MM)$
RHF	10.03	1.03	1.04	1.15	1.92
B3LYP	6.35	2.46	0.51	0.15	0.74
M06-2X	3.73	2.00	0.58	0.08	1.32
MP2	6.48	0.44	0.84	0.34	0.93
CCSD	8.31	0.52	1.09	0.50	1.17
CCSD(T)	7.91	0.83	0.96	0.38	0.96

[a] All energy values in kcal/mol; [b] all calculations with basis-set cc-pVDZ.

For a more global analysis, the complete QM/ELMO/MM reaction energy profiles were also compared to the corresponding reference full-QM/MM ones. They are depicted in Figure 3 for all the quantum mechanical methods and all the partitioning schemes that we considered in our benchmark calculations.

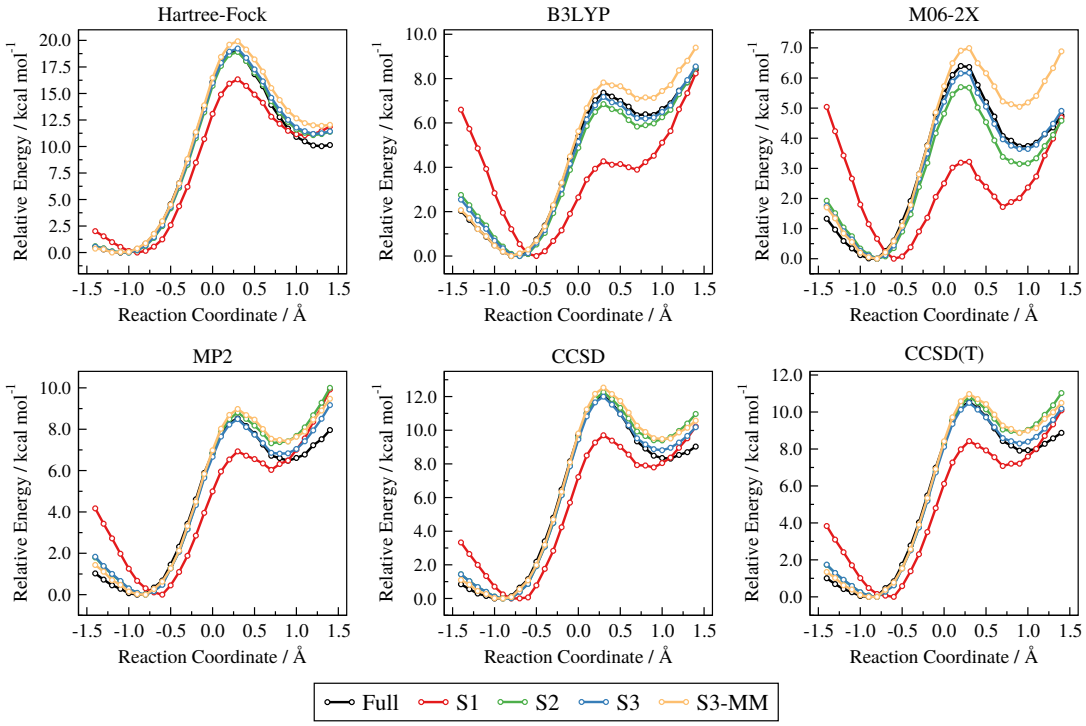


Figure 3. Potential energy profiles for the deprotonation of acetyl coenzyme A by citrate synthase as obtained from QM/ELMO/MM calculations (with partitioning schemes S1, S2 and S3) and QM/MM computations (Full and with partitioning scheme S3-MM) using basis-set cc-pVDZ. For the sake of completeness, all the potential energy profiles with the same scale for the y-axis are reported in Figure S4 of the Supporting Information.

Already from a qualitative point of view, we can notice that, in all the investigated cases, the energy profiles obtained with scheme S1 are those that deviate the most from the reference QM/MM ones. Furthermore, the situation significantly and overall improved when the S2 and, above all, the S3 schemes were adopted. Larger deviations are observed for the profiles resulting from the QM/MM computations with the S3-MM subdivision pattern. All these observations are qualitatively consistent with the results obtained for reaction barriers and energies reported in Tables 1 and 2, respectively. In

particular, the obtained profiles reveal again that the introduction of the intermediate ELMO layer allows reliable reductions of the regions that need to be treated at fully quantum mechanical level, leading to results that are completely comparable to those obtained through standard QM/MM computations but with much larger QM subsystems. This aspect will be further supported by the analysis of the computational costs that will be presented more in details in the next subsection.

For a more quantitative comparison, we also computed maximum and average absolute deviations for all the determined reaction energy profiles, always using the reference full-QM/MM ones as benchmarks. For the sake of completeness, it is also worth pointing out that the values were obtained by considering the range between the minimum points along the full-QM/MM profiles.

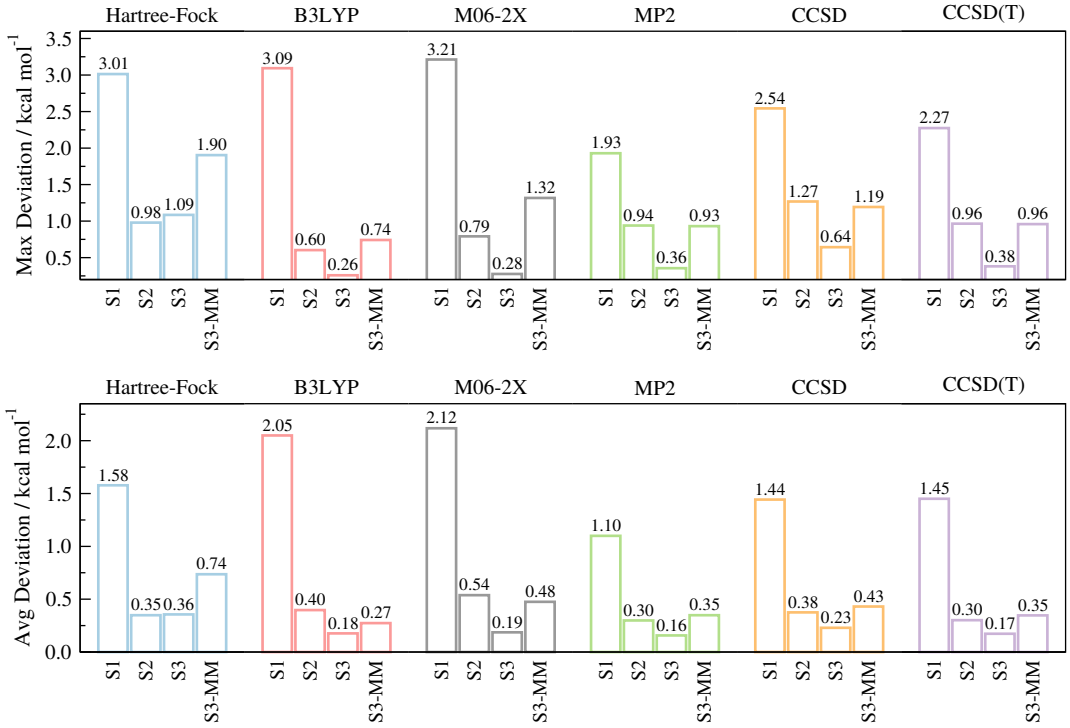


Figure 4. Maximum and average absolute deviations of the QM/ELMO/MM (with partitioning schemes S1, S2 and S3) and QM/MM (with partitioning scheme S3-MM) reaction energy profiles from the corresponding full-QM/MM ones used as references (cc-pVDZ basis-set).

In Figure 4, we can see that for scheme S1 we have maximum absolute discrepancies that are always much larger than 1.0 kcal/mol, as one could also expect from a visual inspection of Figure 3. However, the adoption of schemes S2 and S3 in the QM/ELMO calculations allowed significant reductions of the deviations from the reference profiles, with maximum discrepancies that are almost always within the chemical accuracy limit. Moreover, except for the Hartree-Fock case, the maximum absolute deviation systematically decreases from S1 to S2 and from S2 to S3. In Figure 4 we can also see that the results worsened by treating the oxaloacetate molecule at MM level. In fact, the S3-MM absolute deviations are always larger than those observed for the S3 scheme, and, notwithstanding the larger size of the QM subsystem, they are generally comparable to those obtained by adopting the S2 pattern (with the only exception of the Hartree-Fock level).

Similar trends are observed for the average absolute discrepancies. The largest values are again those associated with the reaction energy profiles obtained with partitioning scheme S1. Furthermore, the situation always improved when schemes S2 and S3 were used, with the latter giving the smallest average deviations for all the considered QM methods, except in the RHF case. Also for the average discrepancies, we observe that the treatment of OAA at MM level (scheme S3-MM) globally worsened the results, leading to values that are generally comparable to those resulting from the QM/ELMO/MM calculations carried out with subdivision pattern S2 (again exception for the Hartree-Fock case).

Concerning the profiles depicted in Figure 3, we can also observe that, except for those obtained by adopting the too crude S1 partitioning scheme, all the curves are smooth and completely reliable, with minimum and transition state points that practically coincides with the full-QM/MM ones in almost all the situations. However,

notwithstanding these very good results, in the future the description could be further improved through the implementation of the analytic QM/ELMO/MM gradient, which would enable optimizations of reaction paths along well-identified reaction coordinates. The implementation of the analytic gradient, which is not straightforward due to the non-variational and non-orthogonal nature of the employed frozen ELMOs, would also pave the way to QM/ELMO/MM Molecular Dynamics simulations, thus significantly broadening scope and applications of the technique proposed in this work. Nevertheless, only based on the results discussed in this section, it is worth observing that, in its present form, the new QM/ELMO/MM approach could be already profitably exploited for free energy calculations through the so-called dual level method.[109-111] This is a strategy in which a low-level approach (i.e., semiempirical or standard QM/MM with the QM region treated at Hartree-Fock or DFT level) is used to produce the sampling, while a higher-level technique (e.g., our CCSD(T)/ELMO/MM method) can be afterwards exploited to get accurate energies on selected snapshots of the previous sampling and to consequently introduce free-energy corrections through perturbation theory. It was shown that this method provides results that are practically as accurate as those resulting from *ab initio* MD simulations but obtained with a much lower computational cost.

In conclusion of this section, the benchmark calculations have shown that, by adopting quite small QM regions, the QM/ELMO/MM computations generally reproduce accurately (i.e., with differences within chemical accuracy) the results of traditional QM/MM calculations performed with a much larger QM subsystem. Furthermore, it was also seen that, although the oxaloacetate molecule does not really participate in the considered reaction, it is crucial to treat it at quantum mechanical level and, to this purpose, the ELMO description represents an excellent way to reach a very good

compromise between chemical accuracy and computational cost (see also the detailed analysis in the next subsection). All these results clearly indicate that the QM/ELMO technique is a perfect strategy to extend the range of applicability (and above all the range of accuracy) of the QM/MM approaches.

4.2 Analysis of the computational cost. In this subsection we will discuss the computational cost of the QM/ELMO/MM calculations when the quantum mechanical region was described through the CCSD(T) technique, which was the most accurate and computationally expensive wave function strategy considered in the present investigation.

At first, using as reference the full-CCSD(T)/MM calculation (which adopted the fragmentation pattern based on scheme S3, but with the oxaloacetate molecule included in the QM region), we considered the CCSD(T)/ELMO/MM computations performed with partitioning schemes S1, S2 and S3 and the CCSD(T)/MM calculation carried out with the subdivision pattern S3-MM. All the computations were performed with basis-set cc-pVDZ on the transition state geometry for the considered reaction, which was identical for all the considered cases (structure at reaction coordinate RC = + 0.3 Å).

For each performed calculation, in Table 3 we reported the number of active occupied molecular orbitals and the number of virtual molecular orbitals that were used, the taken CPU time and its percentage referred to the CPU time for the full-CCSD(T)/MM computation.

Analyzing Table 3, it is possible to see that all the CCSD(T)/ELMO/MM calculations are characterized by large reductions in the number of active occupied and virtual molecular orbitals. As already anticipated in the Theory section, this has an important effect on the overall computational cost. In fact, even considering the CCSD(T)/ELMO/MM calculation with the largest QM region (i.e., the calculation with

partitioning scheme S3), the global CPU time is only about 2.5% of the one associated with the reference full-CCSD(T)/MM computation. This reduction is in line with the typical o^3v^4 scaling of the CCSD(T) method, where o and v are the number of occupied and virtual molecular orbitals used in the calculation, respectively.

Table 3. Number of active occupied molecular orbitals (N_{occ}), number of virtual molecular orbitals (N_{virt}), and CPU times associated with the CCSD(T)/ELMO/MM and CCSD(T)/MM calculations (cc-pVDZ basis-set) performed on the transition-state structure for the deprotonation of acetyl coenzyme A by citrate synthase.[(a)]

Calculation	N_{occ}	N_{virt}	CPU time (dd:hh:mm:ss)	%
CCSD(T)/ELMO/MM (S1)	16	83	00:00:26:07	0.31
CCSD(T)/ELMO/MM (S2)	23	119	00:01:41:18	1.19
CCSD(T)/ELMO/MM (S3)	27	144	00:03:33:18	2.51
CCSD(T)/MM (S3-MM)	27	144	00:03:20:22	2.36
Full-CCSD(T)/MM	52	255	05:21:39:42	100.00

[(a)] The recorded CPU times were obtained by performing parallel calculations on 8 Intel Xeon Gold 6130 2.1 GHz processors.

We can also notice that the computational cost of the CCSD(T)/ELMO/MM calculation with the S3 partitioning is only slightly larger than the one for the CCSD(T)/MM computation exploiting scheme S3-MM. The discrepancy can be ascribed to the greater number of total basis functions used to construct the starting Fock matrix (see equation (2)) in the S3 case (inclusion of the atomic orbitals centered on the atoms of oxaloacetate). However, on the basis of the superior results generally obtained with the QM/ELMO/MM computations by adopting the S3 partitioning scheme, this slightly larger computational cost is completely acceptable. At the same time, the CCSD(T)/ELMO/MM calculation with scheme S2 is characterized by a CPU time that is about half of the one for the CCSD(T)/MM computation with scheme S3-MM. This

result is also remarkable if we take in to account the fact that these two calculations practically provide very similar results (see subsection 4.1).

Therefore, considering the analysis of the computational cost and the results of the benchmark calculations discussed in the previous subsection, we can confirm that the QM/ELMO technique is indeed a convenient tool to extend the range of applicability and accuracy of the traditional QM/MM methods. In particular, we can conclude that both S2 and S3 seem suitable partitioning schemes and, therefore, as we will see in subsection 4.3, they were adopted in all the computations that we performed to investigate the impact of the chosen level of theory for the quantum mechanical region in the QM/ELMO/MM strategy.

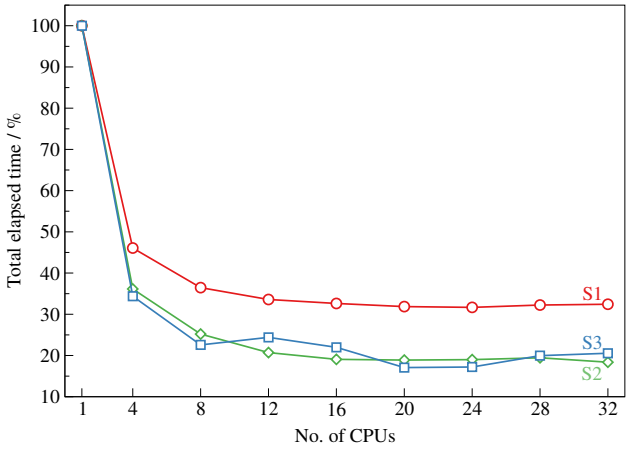


Figure 5. Trend of the total elapsed times as a function of the CPUs used in the QM/ELMO/MM calculations (basis-set cc-pVDZ) carried out on the transition-state structure for the deprotonation of acetyl coenzyme A by citrate synthase with partitioning schemes S1, S2 and S3. The elapsed times are reported as percentages of the timings recorded for the computations performed with only one single CPU core. All calculations were carried out by exploiting Intel Xeon Gold 6130 2.1 GHz processors.

To complete the analysis of the computational cost associated with the QM/ELMO/MM technique we have afterwards considered the scalability with respect to the number of computer cores used in the calculations. In particular, always considering partitioning schemes S1, S2 and S3 for the transition-state geometry, we performed series of

CCSD(T)/ELMO/MM computations with basis-set cc-pVDZ by gradually increasing the number of CPUs (on a single node) from 1 to 32. The elapsed times (shown as percentages of the elapsed times recorded with only one single processor) are reported in Figure 5 as a function of the used cores (for the sake of completeness, raw elapsed time are also reported in Table S1 of the Supporting Information). For all the three subdivision patterns, we notice a significant decrease until 8 employed CPUs. After that the total elapsed times tend to stabilize and they practically remain constant when the parallel computations are performed with more than 12 cores. Interestingly, it is also possible to observe that the reduction of the elapsed time due to the use of multiple processors is more significant with partitioning schemes S2 and S3, which are those characterized by larger fully QM regions.

Table 4. Number of atoms and number of basis functions in the QM/ELMO region with CPU times corresponding to CCSD(T)/ELMO/MM calculations (cc-pVDZ basis-set) performed on the transition state structure for the deprotonation of acetyl coenzyme A by citrate synthase by gradually enlarging the ELMO subsystem associated with partitioning schemes S2 and S3.[(a)]

ELMO regions	No. of atoms (QM/ELMO)	No. of basis functions (QM/ELMO)	CPU time (dd:hh:mm:ss)	%
S2 Reference	29	329	00:01:41:17.8	100.0
4.0 Å	105	1059	00:13:56:30.6	825.8
5.0 Å	136	1494	00:18:43:52.2	1109.5
6.0 Å	202	1954	01:13:09:37.4	2201.1
S3 Reference	29	329	00:03:33:18.2	100.0
4.0 Å	105	1059	00:11:46:54.1	331.4
5.0 Å	136	1494	00:23:28:22.2	660.3
6.0 Å	202	1954	01:23:43:20.0	1342.4

[(a)] Recorded CPU times obtained by carrying out parallel calculations on 8 Intel Xeon Gold 6130 2.1 GHz processors.

Finally, we also took into account how the CPU time varies when the QM region remains fixed, but the ELMO subsystem expands at the expense of the MM part. To accomplish this task, starting from partitioning schemes S2 and S3, the ELMO regions were gradually enlarged by including surrounding atoms within 4 Å, 5 Å and 6 Å from the abstracted hydrogen atom of acetyl coenzyme A in the transition state structure. Also in this case, all the calculations were carried out on the transition state geometry by exploiting the CCSD(T) level of theory for the QM region and adopting the cc-pVDZ basis-set for the QM and ELMO subsystems. The results are reported in Table 4, where we can observe that the CPU time always increases as the ELMO region becomes larger and larger. Considering that the computational effort strictly related to the CCSD(T) procedure remains practically unchanged for all the CCSD(T)/ELMO/MM computations that derive from scheme S2 and for those that derive from scheme S3, the observed differences in the CPU time are ascribable to larger computational costs for the preliminary SCF cycle and for the four-index transformation due to the greater numbers of basis functions centered on the atoms of the QM and ELMO regions. In fact, as mentioned in the Theory section (see subsection 2.2), the Fock matrix is initially constructed over the whole supermolecular basis-set (see equation (2)) and, at the moment, this represents one of the bottlenecks of the QM/ELMO and QM/ELMO/MM techniques. To overcome this drawback, we are currently trying to introduce a suitable truncation criterion similar to the one already proposed by Manby, Miller and their collaborators[82,84] to solve the same problem in the framework of the projection-based embedding approach. This will allow a reduction in the number of basis functions for the construction of the initial Fock matrix with a consequent decrease of the computational cost associated with the SCF cycle and with the four-index transformation for the post-HF calculations, also when the ELMO

regions become quite large. The QM/ELMO/MM strategy is already an attempt in this direction because the introduction of the third molecular mechanics layer can be also seen as a way to discard some basis functions that are not strictly necessary to reach chemical accuracy for the system/process under investigation.

However, although it is true that the computational cost increases as the size of ELMO subsystem becomes larger at the expense of the MM part, it is worth pointing out that the use of pre-computed ELMOs is still more computationally advantageous compared to the case of embedding methods that need a preliminary standard QM calculation on the whole quantum mechanical region given by the union of the high- and low-level QM subunits. For example, this is what happens in the WF-in-DFT/MM approach, where a standard DFT computation on the whole QM region is necessary to preliminarily determine the Kohn-Sham orbitals that are afterwards localized and assigned to the different QM subsystems. As one can imagine, the computational cost of this preparatory procedure significantly increases as the size of the low-level QM subunit becomes larger, certainly much more than in the preliminary steps of our QM/ELMO/MM technique where we have an instantaneous transfer of pre-calculated extremely localized molecular orbitals from properly assembled databanks.

4.3 Influence of the QM level of theory. Another important aspect to consider in our test calculations was the effect of the level of theory for the QM subsystem. To address this point, we decided to carry out QM/ELMO/MM computations exploiting different quantum chemical methods for the QM subunit (RHF, B3LYP, M06-2X, MP2, CCSD, and, when possible, CCSD(T)), and three different basis-sets for the whole QM/ELMO region (cc-pVDZ, aug-cc-pVDZ and cc-pVTZ). As anticipated at the end of the first part of subsection 4.2, these computations were performed by adopting partitioning

schemes S2 and S3. The obtained reaction barriers and energies are reported in Table 5.

Table 5. Reaction barriers (ΔE_{act}) and reaction energies (ΔE_r) for the deprotonation of acetyl coenzyme A by citrate synthase, as obtained from QM/ELMO/MM calculations (partitioning schemes S2 and S3) with different quantum mechanical levels and basis-sets.[(a)]

Basis-set / QM method	ΔE_{act}		ΔE_r	
	Scheme S2	Scheme S3	Scheme S2	Scheme S3
cc-pVDZ				
RHF	18.92	19.24	11.07	11.18
B3LYP	6.85	7.18	5.84	6.20
M06-2X	5.70	6.17	3.15	3.64
MP2	8.81	8.49	7.32	6.82
CCSD	12.26	12.00	9.40	8.81
CCSD(T)	10.81	10.50	8.88	8.29
aug-cc-pVDZ				
RHF	21.72	22.18	12.63	13.14
B3LYP	9.75	10.32	8.19	8.88
M06-2X	8.33	9.10	4.66	5.64
MP2	11.42	11.12	9.55	9.09
CCSD	14.96	14.76	11.26	10.94
CCSD(T)	13.20	12.91	10.70	10.32
cc-pVTZ				
RHF	20.36	21.42	11.08	12.10
B3LYP	8.74	9.75	6.77	7.91
M06-2X	7.78	8.93	3.94	5.29
MP2	9.24	9.54	6.94	7.12
CCSD	12.93	13.40	8.73	9.07

[(a)] All energy values in kcal/mol.

If for each basis-set we consider as references the results obtained with the highest level of theory for the QM region (CCSD(T) for cc-pVDZ and aug-cc-pVDZ; CCSD for cc-pVTZ), we can observe that the QM/ELMO/MM calculations performed with the

Hartree-Fock method always overestimates both the reaction barrier (by a factor between 1.5 and 2) and, to a lower extent, the global reaction energy. This is probably due to the complete absence of Coulomb electron correlation in the Hartree-Fock description. On the contrary, we can see that the adoption of the MP2, B3LYP and M06-2X levels of theory for the QM subsystem always led to an underestimation of the Coupled Cluster results. Finally, for basis-sets cc-pVDZ and aug-cc-pVDZ, we were also able to compare the CCSD/ELMO/MM calculations to the CCSD(T)/ELMO/MM ones and we can notice that the reaction barriers and energies resulting from the former are always larger than those obtained through the latter. It is important to observe that all these trends are independent of the chosen partitioning scheme (S2 or S3) and, above all, they are completely in line with trends already observed in previous computational studies on the examined reaction.[94,104]

The performed calculations also allowed the evaluation of the basis-set choice for the whole QM/ELMO subsystem. From Table 5 we can see that, regardless of the level of theory for the fully quantum mechanical region, the reaction barriers and energies obtained with the cc-pVDZ basis-set are almost always lower than those resulting from computations with larger basis-sets aug-cc-pVDZ and cc-pVTZ. Furthermore, the cc-pVTZ values are generally in-between the cc-pVDZ and aug-cc-pVDZ ones, with the only exception for the reaction energies computed at MP2 and CCSD levels with the S2 subdivision pattern. However, within the different basis-sets, the trends discussed in the previous paragraph are always valid.

4.4 Comparison to other methods. More importantly, to better assess the reliability of the results obtained through the novel QM/ELMO/MM approach, we also decided to compare the reaction barriers and energies resulting from our calculations at the highest levels of theory (i.e., CCSD(T) for cc-pVDZ and aug-cc-pVDZ, and CCSD for

cc-pVTZ) to reference values obtained in previous investigations of the proton abstraction of acetyl coenzyme A by citrate synthase. These reference values were compatible with the experimentally determined activation free energy barrier associated with the overall first step of the Krebs cycle (i.e., formation of citrate and coenzyme A from acetyl coenzyme A and oxaloacetate through the catalytic action of citrate synthase).[104,112-114] In particular, as also done by Bennie *et al.*,[94] our main references were the activation and reaction energies computed at LCCSD(T0)/aug-cc-pVDZ//MM level,[104] with LCCSD(T0) as a local Coupled Cluster method. In that case, the QM region corresponded to the fully quantum mechanical subsystem considered in partitioning scheme S3 plus the oxaloacetate molecule, and the MM subunit to the remaining part of the system (exactly as for our full-QM/MM reference). The obtained activation and reaction energies were 13.2 kcal/mol and 8.4 kcal/mol, respectively. Now, if we consider our CCSD(T)/ELMO/MM calculations with the cc-pVDZ set of basis functions, the reaction barriers are underestimated, while the global reaction energies are fully in line with the LCCSD(T0) values. On the contrary, taking into account the aug-cc-pVDZ computations, which are actually those that are directly comparable to the reference LCCSD(T0)/MM ones, the obtained activation energies are in optimal agreement with the benchmark results, while the computed reaction energies are larger by 1.9-2.3 kcal/mol. This worse behavior for the reaction energies can be ascribed to the poorer ELMO description for the electron density of the oxaloacetate molecule, which probably does not correctly model the stabilization of the deprotonated acetyl coenzyme A by the oxaloacetate ketone. In fact, in the current version of the QM/ELMO strategy, the embedding of the fully QM region is only approximate since it is given by transferred and frozen extremely localized molecular orbitals. To improve the results, a future methodological development could consist in

introducing a relaxation of the surrounding ELMO electron distribution that takes into account the polarization due to the ground and excited states of the chemically active region. This could be accomplished by exploiting the transfer and use of virtual extremely localized molecular orbitals that are already available in the current version of the ELMO libraries. To complete the analysis of the results obtained through the calculations at the highest level of theory within the different basis-sets, we can notice that, in the cc-pVTZ case, the CCSD/ELMO/MM outcomes are in very good agreement with the reference LCCSD(T0)/MM ones, independently of the chosen partitioning scheme and for both the activation and the reaction energy.

For a further and important comparison, we also considered the results of previous CCSD(T)-in-DFT/MM calculations with basis-set aug-cc-pVDZ and with a partitioning scheme intermediate between S2 and S3, as reported by Bennie *et al.*[94] These computations gave activation barriers between 11.4 and 12.1 kcal/mol and reaction energies within 1.0 kcal/mol from the reference LCCSD(T0)/MM values. If now we consider our CCSD(T)/ELMO/MM calculations performed with basis-set aug-cc-pVDZ and partitioning schemes S2 and S3 (see again Table 5), we can notice that the obtained values for activation and reaction energies are completely compatible with the CCSD(T)-in-DFT/MM ones. However, always using the LCCSD(T0)/MM results as references, we can see that our new technique outperforms the WF-in-DFT/MM method in the determination of the reaction barriers. The opposite is true for the reaction energies, for which the CCSD(T)-in-DFT/MM results are better than those obtained through the CCSD(T)/ELMO/MM approach. This is again probably related to the fact that the QM/ELMO/MM strategy uses frozen pre-computed extremely localized molecular orbitals. In fact, if on the one hand pre-calculated ELMOs are certainly advantageous from the computational point of view (as we will also stress below), on

the other hand they provide a worse stabilization of the deprotonated acetyl coenzyme A compared to tailor-made frozen Kohn-Sham orbitals used in the WF-in-DFT/MM computations.

Based on all the previous comparisons, we can conclude that the proposed QM/ELMO/MM approach overall provide completely reasonable results especially when the fully quantum mechanical region is treated through a very high level of theory. In particular, also in this case, it is worth pointing out again that the QM/ELMO/MM results are completely in line with those obtained through standard QM/MM calculations characterized by a larger computational cost. Concerning the comparison with similar WF-in-DFT/MM embedding computations (based on similar partitioning schemes), we got analogous values for reaction barriers and energies, but with the non-negligible advantage that, in our new approach, it is not necessary to perform a preliminary DFT calculation on the QM system and localize the obtained Kohn-Sham molecular orbitals to define the WF and DFT regions. In our case, we use pre-determined ELMOs that are directly transferred to the low-QM subsystem (i.e., the ELMO subsystem), thus reducing the cost associated with the preliminary-step to the real fully QM embedding computation.

5. CONCLUSIONS

In this work we have proposed a novel multi-layer embedding technique for the computational investigation of macromolecules and biochemical reactions. In the new method, the most important region of the system under exam can be treated at a very high quantum mechanical level of theory, an intermediate subsystem through a buffer of frozen extremely localized molecular orbitals previously transferred from databanks or proper model molecules, and the remaining part by means of a classical force field.

The validation of the new technique, which in this investigation was conducted on the well-studied deprotonation of acetyl coenzyme A by citrate synthase, proved that the novel QM/ELMO/MM method reproduces the results obtained through traditional QM/MM computations, but using much smaller QM subsystems and, therefore, with a significantly lower computational cost. This is possible thanks to the cheap but crucial embedding of the fully quantum mechanical region provided by frozen ELMOs. Moreover, we have also seen that the obtained values for activation and reaction energies are in very good agreement with those resulting from previous computational studies of the investigated process, especially with those obtained by means of WF-in-DFT/MM calculations. The advantage of our novel approach is that no preliminary DFT computations are necessary, but only suitable pre-computed extremely localized molecular orbitals stored in a databank are exploited to reliably describe the low-level QM region.

On the basis of the obtained results, we imagine that the new strategy will allow an easier exploitation of high-level and accurate wave function-based techniques (e.g., Coupled Cluster) in the framework of QM/MM-type calculations. Therefore, we envisage the use of the new approach in the modeling of large biosystems. To this purpose, in the future we will apply the novel three-layer embedding technique both in other computational studies of enzyme reactions and in new structural refinements of macromolecules within the context of modern quantum crystallography. In particular, considering the promising results recently obtained through the HAR-ELMO method,[50] we imagine that the novel QM/ELMO/MM strategy could be profitably coupled with the emerging Hirshfeld atom refinement technique[115-120] to obtain very accurate structural details (e.g., hydrogen atom positions) in chemically important regions of biological systems, such as in active or binding sites of proteins. Finally, as already

mentioned in the discussion of the obtained results, further methodological improvements will be also necessary and possible. For example, we are currently planning the introduction of a polarizable ELMO subunit in order to have better embeddings for the fully quantum mechanical regions of the investigated systems. Furthermore, despite in its current version our new three-layer embedding method can be already exploited within the dual level approach[109-111], one of the next development steps will consist in the implementation of the QM/ELMO/MM analytic gradient that will enable optimizations of reaction paths along reaction coordinates and the coupling with Molecular Dynamics for the computation of free energies.

ASSOCIATED CONTENT

Supporting Information. Details about theory, transfer and libraries of extremely localized molecular orbitals (with Figure S1 showing examples of ELMOs and Figure S2 giving the schematic representation of the reference frames and atomic triads required for the ELMOs rotation). Details about the QM/ELMO method. Figure S3 depicting the model molecule used to compute the extremely localized molecular orbitals describing the fragments of acetyl coenzyme A in the performed QM/ELMO/MM calculations. Figure S4 showing all the QM/ELMO/MM and QM/MM reaction energy profiles (cc-pVDZ basis-set) with the same scale for the y-axis. Table S1 with the elapsed times associated with the CCSD(T)/ELMO/MM calculations performed with different numbers of computer cores.

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGEMNTS

We sincerely thank Prof. Marc W. van der Kamp (University of Bristol, U.K.) for providing us the QM/MM optimized geometries along the reaction path, which were used to test the new QM/ELMO/MM technique. We also gratefully thank Dr. Manuel F. Ruiz-López (CNRS & University of Lorraine, France) for helpful discussions. We acknowledge the French Research Agency (ANR) for financial support of this work through the Young Investigator Project *QuMacroRef* (Grant No. ANR-17-CE29-0005-01). The High-Performance Computing Center *EXPLOR* of the University of Lorraine is thanked for providing computing time through the projects 2019CPMXX0966, 2019CPMXX0886 and 2019CPMXX1332. Dr. 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] Gordon, M. S.; Slipchenko, L. V. Introduction: Calculations on Large Systems. *Chem. Rev.* **2015**, *115*, 5605-5606.

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

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

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

[5] Yang, W.; Lee, T.-S. A density‐matrix divide‐and‐conquer approach for electronic structure calculations of large molecules. *J. Chem. Phys.* **1995**, *103*, 5674-5678.

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

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

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

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

[10] Babu, K.; Gadre, S. R. Ab initio quality one-electron properties of large molecules: Development and testing of molecular tailoring approach. *J. Comput. Chem.* **2003**, *24*, 484-495.

[11] Ganesh, V.; Dongare, R. K.; Balanarayan, P.; Gadre, S. R. Molecular Tailoring Approach for Geometry Optimization of Large Molecules: Energy Evaluation and Parallelization Strategies. *J. Chem. Phys.* **2006**, *125*, 104109.

[12] Sahu, N.; Gadre, S. R. Molecualr Tailoring Approach: A Route for *ab initio* Treatment of

Large Clusters. *Acc. Chem. Res.* **2014**, *47*, 2739-2747.

[13] Singh, G.; Nandi, A.; Gadre, S. R. Breaking the bottleneck: Use of molecular tailoring approach for the estimation of binding energies at MP2/CBS limit for large water clusters. *J. Chem. Phys.* **2016**, *144*, 104102.

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

[15] Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment molecular orbital method: application to polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614-618.

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

[17] Nakano, T.; Mochizuki, Y.; Fukuzawa, K.; Amari, S.; Tanaka, S. Developments and applications of ABINIT-MP software based on the fragment molecular orbital method. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; Chapter 2, pp 39-52.

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

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

[20] Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to insulin. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12690-12693.

[21] Huang, L.; Massa, L.; Karle, J. The Kernel Energy Method: Application to a tRNA. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1233-1237.

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

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

[24] Huang, L.; Massa, L.; Matta, C. F. A graphene flake under external electric fields reconstructed from field-perturbed kernels. *Carbon* **2014**, *76*, 310-320.

[25] Timm, M. J.; Matta, C. F.; Massa, L.; Huang, L. The localization-delocalization matrix and the electron density-weighted connectivity matrix of a finite graphene flake reconstructed from kernel fragments. *J. Phys. Chem. A* **2014**, *118*, 11304-11316.

[26] Huang, L.; Matta, C. F.; Massa, L. The kernel energy method (KEM) delivers fast and accurate QTAIM electrostatic charge for atoms in large molecules *Struct. Chem.* **2015**, *26*, 1433-1442.

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

[28] Zhang, D. W.; Xiang, Y.; Zhang, J. Z. H. New Advance in Computational Chemistry: Full Quantum Mechanical ab Initio Computation of Streptavidin-Biotin Interaction Energy. *J. Phys. Chem. B* **2003**, *107*, 12039-12041.

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

[30] Xiang, Y.; Zhang, D. W.; Zhang, J. Z. H. Fully quantum mechanical energy optimization for protein-ligand structure. *J. Comput. Chem.* **2004**, *25*, 1431-1437.

[31] Mey, Y.; Zhang, D. W.; Zhang, J. Z. H. New method for direct linear-scaling calculation of electron density of proteins. *J. Phys. Chem. A* **2005**, *109*, 2-5.

[32] He, X.; Zhang, J. Z. H. A new method for direct calculation of total energy of protein. *J. Chem. Phys.* **2005**, 122, 031103.

[33] He, X.; Zhang, J. Z. H. The generalized molecular fractionation with conjugate

caps/molecular mechanics method for direct calculation of protein energy. *J. Chem. Phys.* **2006**, 124, 184703.

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

[35] Walker, P. D.; Mezey, P. G. Molecular electron density Lego approach to molecule building. *J. Am. Chem. Soc.* **1993**, *115*, 12423-12430.

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

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

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

[39] Szekeres, Z.; Exner, T.; Mezey, P. G. Fuzzy Fragment Selection Strategies, Basis Set Dependence and HF-DFT Comparisons in the Applications of the ADMA Method of Macromolecular Quantum Chemistry. *Int. J. Quantum Chem.* **2005**, *104*, 847-860.

[40] 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.

[41] 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.

[42] 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.

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

[44] 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.

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

[46] 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.

[47] 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.

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

[49] 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.

[50] 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.

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

[52] 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.

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

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

[55] Grabowsky, S.; Genoni, A.; Thomas, S. P.; Jayatilaka, D. The Advent of Quantum Crystallography: Form and Structure Factors from Quantum Mechanics for Advanced Structure Refinement and Wavefunction Fitting. In *21st Century Challenges in Chemical Crystallography 2 - Structural Correlations and Data Interpretation. Structure and Bonding*; Mingos, D. M. P., Rathby, P., Eds.; Springer: Berlin and Heidelberg, Germany, 2020; Vol. *186*, pp 65-144.

[56] 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.

[57] 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.

[58] 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.

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

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

[61] Amaro, R. E.; Mulholland, A. J. Multiscale methods in drug design bridge chemical and biological complexity in the search for cures. *Nature Reviews Chemistry* **2018**, *2*, 0148.

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

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

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

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

[66] 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.

[67] Vreven, T.; Morokuma, K. On the Application of the IMOMO (Integrated Molecular Orbital + Molecular Orbital) method. *J. Comput. Chem.* **2000**, *21*, 1419-1432.

[68] 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.

[69] Sun, Q.; Chan, G. K.-L. Quantum Embedding Theories. *Acc. Chem. Res.* **2016**, *49*, 2705-2712.

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

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

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

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

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

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

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

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

[78] Saether, S., Kjaergaard, T.; Koch, H.; Høyvik, I.-M. Density-Based Multilevel Hartree-Fock Model. *J. Chem. Theory Comput.* **2017**, *13*, 5282-5290.

[79] Folkestad, S. D.; Kjønstad, E. F.; Goletto, L.; Koch, H. Multilevel CC2 and CCSD in Reduced Orbital Spaces: Electronic Excitations in Large Molecular Systems. *J. Chem. Theory Comput.* **2021**, *17*, 714-726.

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

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

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

[87] 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.

[88] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory. *Science* **2008**, *321*, 792-794.

[89] Cramer, C. J.; Truhlar, D. G. Density Functional Theory for Transition Metals and Transition Metal Chemistry. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757-10816.

[90] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory. *Chem. Rev.* **2012**, *112*, 289-320.

[91] Jakobsen, S.; Kristensen, K.; Jensen, F. Electrostatic Potential of Insulin: Exploring the Limitations of Density Functional Theory and Force Field Methods. *J. Chem. Theory Comput.* **2013**, *9*, 3978-3985.

[92] Pribram-Jones, A.; Gross, D. A.; Burke, K. DFT: A Theory Full of Holes? *Annu. Rev. Phys. Chem.* **2015**, *66*, 283-304.

[93] 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.

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

[95] Ranaghan, K. E.; Shchepanovska, D.; Bennie, S. J.; Lawan, N.; Macrae, S. J.; Zurek, J.; Manby, F. R.; Mulholland, A. J. Projector-Based Embedding Eliminates Density Functional

Dependence for QM/MM Calculations of Reactions in Enzymes and Solutions. *J. Chem. Inf. Model.* **2019**, *59*, 2063-2078.

[96] Goletto, L.; Giovannini, T.; Folkestadt, S. D.; Koch, H. Combining multilevel Hartree-Fock and multilevel coupled cluster approaches with molecular mechanics: a study of electronic excitations in solutions. *Phys. Chem. Chem. Phys.* **2021**, *23*, 4413-4426.

[97] 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.

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

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

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

[101] Macetti G.; Wieduwilt, E. K.; Genoni, A. QM/ELMO: A Multi-Purpose Fully Quantum Mechanical Embedding Scheme Based on Extremely Localized Molecular Orbitals. *J. Phys. Chem. A* **2021**, *125*, 2709-2726.

[102] 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.

[103] Case, D. A.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Luchko, T.; Luo, R.; Madej, B.; Mermelstein, D.; Merz, K. M.; Monard, G.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Simmerling, C. L.; Botello-Smith, W. L.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; Xiao. L.; Kollman. P.A. *AMBER 2016*; University of California San Francisco: San Francisco, CA, USA, 2016.

[104] van der Kamp, M. W.; Zurek, J.; Manby, F. R.; Harvey, J. N.; Mulholland, A. J. Testing High-Level QM/MM Methods for Modeling Enzyme Reactions: Acetyl-CoA Deprotonation in Citrate Synthase. *J. Phys. Chem. B* **2010**, *114*, 11303-11314.

[105] Maier, J. A.; Martinez, C.; Kasavajhala, L.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

[106] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

[107] 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.

[108] 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.

[109] Retegan, M.; Martins-Costa, M.; Ruiz-López, M. F. Free-energy calculations using dual-level Born-Oppenheimer molecular dynamics. *J. Chem. Phys.* **2010**, *133*, 064103.

[110] Martins-Costa, M. T. C.; Ruiz-Lopez, M. F. Amino Acid Capture by Aqueous Interfaces. Implications for Biological Uptake. *J. Phys. Chem. B* **2013**, *117*, 12469-12474.

[111] Martins-Costa, M. T. C.; Ruiz-López, M. F. Highly accurate computation of free energies in complex systems through *horsetail* QM/MM molecular dynamics combined with free-energy perturbation theory. *Theor. Chem. Acc.* **2017**, *136*, 50.

[112] Mulholland, A. J.; Lyne, P. D. ; Karplus, M. Ab Initio QM/MM Study of the Citrate Synthase Mechanism. A Low-Barrier Hydrogen Bond Is not Involved. *J. Am. Chem. Soc.* **2000**, *122*, 534-535.

[113] van der Kamp, M. W.; Perruccio, F.; Mulholland, A. J. Substrate polarization in enzyme catalysis: QM/MM analysis of the effect of oxaloacetate polarization on acetyl-CoA enolizaton in citrate synthase. *Proteins: Struct., Funct., Genet.* **2007**, *69*, 521-535.

[114] van der Kamp, M. W.; Perruccio, F.; Mulholland, A. High-level QM/MM modelling predicts an arginine as the acid in the condensation reaction catalysed by citrate synthase. *Chem. Commun.* **2008**, 1874-1876.

[115] 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.

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

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

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

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

[120] Kleemiss, F.; Dolomanov, O. V.; Bodensteiner, M.; Peyerimhoff, N.; Midgley, L.; Borhis, L. J.; Genoni, A.; Malaspina, L. A.; Jayatilaka, D.; Spencer, J. L.; White, F.; Grrundkötter-Stock, B.; Steinhauer, S.; Lentz, D.; Puschmann, H.; Grabowsky, S. Accurate Crystal Structures and Chemical Properties from NoSpherA2. *Chem. Sci.* **2021**, *12*, 1675-1692.