

# Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules

Giovanni Macetti, Alessandro Genoni

▶ **To cite this version:**

Giovanni Macetti, Alessandro Genoni. Quantum Mechanics/Extremely Localized Molecular Orbital Method: A Fully Quantum Mechanical Embedding Approach for Macromolecules. Journal of Physical Chemistry A, 2019, 10.1021/acs.jpca.9b08882 . hal-02324013

**HAL Id: hal-02324013**

**https://hal.univ-lorraine.fr/hal-02324013**

Submitted on 10 May 2020

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

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

This document is the Accepted Manuscript version of a Published Work that appeared in final form in *The Journal of Physical Chemistry A*, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see https://pubs.acs.org/doi/10.1021/acs.jpca.9b08882.

# QM/ELMO Method: a Fully Quantum Mechanical Embedding Approach for Macromolecules.

*Giovanni Macetti*[†] *and Alessandro Genoni* [†,*]

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

**AUTHOR INFORMATION**

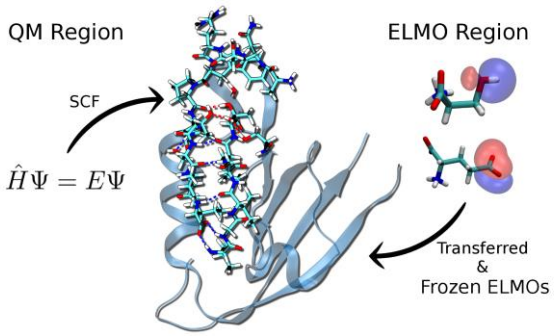
**Corresponding Author**

Alessandro Genoni, Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), 1 Boulevard Arago, F-57078 Metz, France; Phone: +33 (0)3 72 74 91 70; Fax: +33 (0)3 72 74 91 87.

E-mail: Alessandro.Genoni@univ-lorraine.fr.

**ABSTRACT.** The development of methods for the quantum mechanical study of macromolecules has been always an important challenge in theoretical chemistry. Nowadays, the techniques proposed in this context can be used to investigate very large systems and can be subdivided into two main categories: fragmentation and embedding strategies. In this paper, we introduce a new multi-scale embedding method modifying and improving the local self-consistent field (LSCF) approach originally proposed for quantum mechanics/molecular mechanics techniques. The new strategy enables to treat chemically relevant regions of large biological molecules through usual methods of quantum chemistry, while describing the remaining parts of the systems by means of frozen extremely localized molecular orbitals (ELMOs) transferred from properly constructed libraries. Test calculations have shown the correct functioning and the high reliability of the new approach, thus anticipating its possible applications to different fields of physical chemistry, such as rational drug design and structural refinements of proteins.

**TOC GRAPHICS**




**KEYWORDS.** Multi-scale methods, QM/QM' techniques, embedding, extremely localized molecular orbitals, biological molecules

## 1. Introduction

The treatment of large biological molecules at a fully quantum mechanical level has been one of the most important and difficult challenges of theoretical chemistry for a long time. Over the years, different research groups have proposed several techniques that can be generally subdivided into two main categories: fragmentation[1] and embedding methods[2-5].

The former basically consist in subdividing the large molecules under exam into smaller subunits, performing quantum mechanical calculations on the resulting subsystems and finally recombining the results to obtain approximate wave functions, electron densities or only energy values for the target molecules. In this category, it worth mentioning the divide & conquer strategy,[6-12] the molecular fractionation with conjugated caps (MFCC) technique,[13-17] the molecular tailoring approach,[18-20] the kernel energy method (KEM),[21-24] the ONIOM technique,[25-28] the well-known fragment molecular orbital (FMO) approach[29-33] and all those strategies based on the transferability principle, such as the molecular electron density LEGO assembler (MEDLA) technique,[34,35] the adjustable density matrix assembler (ADMA) method[36-38] and the recently proposed approach based on libraries of extremely localized molecular orbitals (ELMOs)[39-41].

Conversely, the latter group of techniques relies on the observation that, in most of the cases, only a small part of the system under investigation determines the properties of interest of a very large molecule (e.g., active sites of a proteins). For this reason, in the embedding methods only the chemically relevant region of the molecule is treated at a fully quantum mechanical level, while the rest of the system is seen as a source of slight perturbation and is generally treated at a lower level of theory. Of course, in this context, a prominent place is occupied by the quantum

mechanics/molecular mechanics (QM/MM) strategies,[2-4,42,43] which are particularly successful for the computational study of biological molecules. Other important embedding techniques are the frozen-density embedding theory (FDET),[5,44-46] which can be considered as a modified density functional theory (DFT) approach where a frozen electron density describes the surrounding and constraints the optimization of the electron density for the crucial region of the investigated system, and the effective group potentials (EGP) method,[47,48] which extends the successful concept of atomic effective core potentials (ECP) to groups of atoms in order to significantly reduce the computational cost of *ab initio* calculations.

In this paper, we propose a new fully quantum mechanical, multi-scale, embedding strategy (hereafter indicated as QM/ELMO technique) in which only the chemically/biologically important part of very large macromolecules is treated by means of traditional methods of quantum chemistry, while the rest of the system is described by means of transferred and frozen extremely localized molecular orbitals (ELMOs).[49-51] ELMOs are molecular orbitals strictly localized on small molecular units and, due to their extreme localization, they can be reliably transferred from molecule to molecule and considered as plausible electronic building blocks to reconstruct wave functions and electron densities of large systems.[39,40] On the basis of these features, libraries of ELMOs that cover all the elementary units of the twenty natural amino acids in all their possible protonation states have been recently constructed with the final goal of instantaneously obtaining approximate wave functions of polypeptides and proteins[41] (see also Supporting Information for more details).

Therefore, to develop the novel QM/ELMO method, we have properly modified the local self-consistent field (LSCF) approach,[52-55] which was successfully used in QM/MM methods to

describe frontier regions by means of strictly localized bond orbitals, and we have combined it with the recently constructed databanks of ELMOs.[39-41]

## 2. Theoretical details and implementation

In the new approach we subdivide the system under exam into two quantum subunits, which we will hereafter indicate as QM and ELMO regions. As shown in Figure 1, in the general case of connected subsystems, the QM and ELMO regions obviously share only the frontier atoms $\{Y_i\}$ and, consequently, the basis functions centered on the nuclei of those atoms.




**Figure 1.** Schematic representation of the QM and ELMO regions in QM/ELMO calculations.

The ELMO subsystem is described through extremely localized molecular orbitals, which are transferred from the available ELMO libraries and kept frozen during the computations. It is very important to note that, unlike the original LSCF QM/MM approach, where all the frozen strictly localized molecular orbitals for the frontier region are expanded on atomic orbitals of the QM region, in this case the frozen ELMOs mainly use basis functions belonging to the ELMO subsystem and share with the QM part only those atomic orbitals centered on the frontier atoms. In other words, for the *i*-th transferred ELMO, we can write:

$$|\phi_i\rangle = \sum_{\nu \in ELMO} C_{\nu i} \, |\chi_\nu\rangle \quad (1)$$

The previous precaution is fundamental to overcome the linear dependency problem that affects the original LSCF QM/MM approach and that would prevent to develop a QM/ELMO method applicable to very large systems.

Given their intrinsic non-orthogonality, the ELMOs are then mutually orthonormalized using the Löwdin procedure:

$$|\phi_i^{\perp}\rangle = \sum_{j=1}^{N_{ELMO}} \left[\boldsymbol{s}^{-1/2}\right]_{ji} |\phi_j\rangle = \sum_{\nu \in ELMO} C_{\nu i}^{\perp} |\chi_\nu\rangle \quad (2)$$

with $\boldsymbol{s}$ as the overlap matrix of the transferred ELMOs $\{|\phi_i\rangle\}_{i=1}^{N_{ELMO}}$. As well known, the Löwdin orthogonalization does not significantly change the starting orbitals and, therefore, it preserves the localized nature of the ELMOs to a large extent. Therefore, after the orthogonalization procedure, the transferred and frozen molecular orbitals will only slightly delocalize over the frontier atoms $\{Y_i\}$, thus further helping in avoiding the linear dependency problems of the original LSCF QM/MM method.

Afterwards, the $M_{QM}$ basis functions of the QM region are orthogonalized against the previously orthonormalized ELMOs, by removing the projections of the former on the latter. This is equivalent to the following transformation:

$$\tilde{\boldsymbol{\chi}} = \boldsymbol{\chi}\,\mathbf{G} \quad (3)$$

where $\tilde{\boldsymbol{\chi}}$ is the $1 \times M_{QM}$ array of the basis functions belonging to the QM region after the orthogonalization (i.e., $\tilde{\boldsymbol{\chi}} = [|\tilde{\chi}_1\rangle, |\tilde{\chi}_2\rangle, \dots, |\tilde{\chi}_{M_{QM}}\rangle]$), $\boldsymbol{\chi}$ is the $1 \times M$ array of the original and complete set of basis functions for the system under exam (i.e., $\boldsymbol{\chi} = [|\chi_1\rangle, |\chi_2\rangle, \dots, |\chi_M\rangle]$) and $\mathbf{G}$ is the $M \times M_{QM}$ transformation matrix having elements

$$G_{\nu\mu} = \left[1 - \sum_{i=1}^{N_{ELMO}} \left(S_{\mu i}\right)^2\right]^{-1/2} \left[\delta_{\nu\mu} - t_{\nu\mu}\right] \quad (4)$$

with $S_{\mu i}$ as the overlap integral between the original basis function $|\chi_\mu\rangle$ and the orthogonalized ELMO $|\phi_i^{\perp}\rangle$, $\delta_{\nu\mu}$ as the usual Kronecker delta and $t_{\nu\mu}$ given by the following expression:

$$t_{\nu\mu} = \sum_{i=1}^{N_{ELMO}} C_{\nu i}^{\perp} S_{\mu i} \quad (5)$$

At this point it is worth noting that, by considering only those original basis functions that belong to the ELMO region and that are characterized by non-negligible $t_{\nu\mu}$ values, the size of the matrix $\boldsymbol{G}$ and of the array $\boldsymbol{\chi}$ might significantly decrease ($M' \times M_{QM}$ and $1 \times M'$, respectively, with $M' \ll M$), thus significantly reducing also the computational cost of the preliminary orthogonalization step. Actually, for all the QM/ELMO computations discussed in the present work, the complete set of basis functions were taken into account, but the determination of a suitable criterion to select a reduced subset of atomic orbitals will be certainly defined in future investigations. The introduction of this criterion will be also beneficial to reduce the CPU time associated with the computation of the Fock matrix, which is the current rate-limiting step of the QM/ELMO strategy (see discussion below).

Since the obtained basis functions $\tilde{\boldsymbol{\chi}}$ are not orthogonal among each other, they are afterwards subject to a canonical orthogonalization, namely we have that

$$\boldsymbol{\chi}' = \tilde{\boldsymbol{\chi}}\,\mathbf{Q} \quad (6)$$

where $\boldsymbol{\chi}'$ is the $1 \times M_{QM}$ array of the final orthonormal basis functions belonging to the QM region (i.e., $\boldsymbol{\chi}' = [|\chi'_1\rangle, |\chi'_1\rangle, \ldots, |\chi'_{M_{QM}}\rangle]$) and $\mathbf{Q}$ is the transformation matrix of elements

$$Q_{ij} = \frac{\widetilde{U}_{ij}}{\tilde{u}_j^{\,1/2}} \quad (7)$$

with $\widetilde{U}_{ij}$ as the *i*-th component of the *j*-th eigenvector and $\tilde{u}_j$ as the *j*-th eigenvalue of the matrix $\tilde{\mathbf{S}}$ of the overlap integrals between the transformed basis functions contained in the array $\tilde{\boldsymbol{\chi}}$. It is worth noting that, unlike the LSCF QM/MM approach, due to the suitable definition of the QM and ELMO regions described above, in our case all the eigenvalues of the matrix $\tilde{\mathbf{S}}$ are always significantly different from zero (i.e., always greater than $10^{-4}$). Therefore, as already anticipated, linear dependency problems do not exist in the novel QM/ELMO technique and the total number of basis functions associated with the QM region remains unchanged.

The orthogonalization procedure that leads from the original basis functions $\boldsymbol{\chi}$ to the final orthogonal basis functions $\boldsymbol{\chi}'$ can be actually summarized like this:

$$\boldsymbol{\chi}' = \boldsymbol{\chi}\,\mathbf{B} \quad (8)$$

with $\mathbf{B}$ as the global $M \times M_{QM}$ transformation matrix, which can be expressed as

$$\mathbf{B} = \mathbf{G}\,\mathbf{Q} \quad (9)$$

Now, the new QM/ELMO technique proposed in this paper basically consists in performing a full Hartree-Fock (HF) calculation on a QM subsystem in presence of a frozen set of transferred and orthogonalized extremely localized molecular orbitals describing the ELMO region. Therefore, in case of a closed-shell QM subunit consisting of *2N* electrons, the generic element

$F'_{\mu'\nu'}$ of the Fock matrix for the QM region in the orthogonal basis $\boldsymbol{\chi}'$ can be expressed in this way:

$$F'_{\mu'\nu'} = \langle \chi'_\mu | \hat{h}^{core} | \chi'_\nu \rangle + \sum_{i=1}^{N} \left[ 2\,(\chi'_\mu\, \chi'_\nu | \varphi'_i\, \varphi'_i) - (\chi'_\mu\, \varphi'_i | \varphi'_i\, \chi'_\nu) \right]$$

$$+ \sum_{i=1}^{N_{ELMO}} \left[ 2\,(\chi'_\mu\, \chi'_\nu | \phi_i^\perp\, \phi_i^\perp) - (\chi'_\mu\, \phi_i^\perp | \phi_i^\perp \chi'_\nu) \right] \quad (10)$$

where $\hat{h}^{core}$ is the usual one-electron core Hamiltonian operator, $|\phi_i^\perp\rangle$ is the *i*-th orthogonalized ELMO defined in equation (2) and $|\varphi'_i\rangle$ is the *i*-th occupied molecular orbital for the QM part expanded on the orthogonal basis functions $\boldsymbol{\chi}'$:

$$|\varphi'_i\rangle = \sum_{\mu'=1}^{M_{QM}} C'_{\mu' i}\, |\chi'_\mu\rangle \quad (11).$$

Explicitly using equations (2) and (11) in equation (10), the expression for the element of the Fock matrix of the QM subsystem can be rewritten like this:

$$F'_{\mu'\nu'} = \langle \chi'_\mu | \hat{h}^{core} | \chi'_\nu \rangle + \sum_{\lambda',\sigma'=1}^{M_{QM}} P^{QM'}_{\lambda'\sigma'} \left[ (\chi'_\mu\, \chi'_\nu | \chi'_\sigma\, \chi'_\lambda) - \frac{1}{2} (\chi'_\mu\, \chi'_\lambda | \chi'_\sigma\, \chi'_\nu) \right]$$

$$+ \sum_{\lambda,\sigma \in ELMO} P^{ELMO}_{\lambda\sigma} \left[ (\chi'_\mu\, \chi'_\nu | \chi_\sigma\, \chi_\lambda) - \frac{1}{2} (\chi'_\mu\, \chi_\lambda | \chi_\sigma\, \chi'_\nu) \right]$$

$$= h^{core\,\prime}_{\mu'\nu'} + F^{QM\prime}_{\mu'\nu'} + F^{ELMO\prime}_{\mu'\nu'} \quad (12)$$

where $\mathbf{P}^{QM\prime}$ is the one-electron density matrix for the QM region expressed in the orthogonal basis $\boldsymbol{\chi}'$, namely

$$P_{\lambda'\sigma'}^{QM\prime} = 2\sum_{i=1}^{N} C_{\sigma' i}^{\prime *}\, C_{\lambda' i}^{\prime} \qquad (13),$$

while $\mathbf{P}^{ELMO}$ is the one-electron density matrix for the ELMO region expressed in the original basis, namely

$$P_{\lambda\sigma}^{ELMO} = 2\sum_{i=1}^{N_{ELMO}} C_{\sigma i}^{\perp *}\, C_{\lambda i}^{\perp} \qquad (14).$$

As already indicated in the seminal paper of the original LSCF method, the matrix element $F'_{\mu'\nu'}$ given by equation (12) does not change if the basis functions $\boldsymbol{\chi}'$ are transformed back to the basis functions $\boldsymbol{\chi}$, provided that the one-electron density matrix is also transformed accordingly. In the original basis-set, the element of the Fock matrix for the QM region naively becomes

$$\begin{aligned} F_{\mu\nu} = \langle \chi_\mu | \hat{h}^{core} | \chi_\nu \rangle &+ \sum_{\lambda,\sigma=1}^{M} P_{\lambda\sigma}^{QM} \left[ (\chi_\mu\, \chi_\nu | \chi_\sigma\, \chi_\lambda) - \frac{1}{2}(\chi_\mu\, \chi_\lambda | \chi_\sigma\, \chi_\nu) \right] \\ &+ \sum_{\lambda,\sigma \in ELMO} P_{\lambda\sigma}^{ELMO} \left[ (\chi_\mu\, \chi_\nu | \chi_\sigma\, \chi_\lambda) - \frac{1}{2}(\chi_\mu\, \chi_\lambda | \chi_\sigma\, \chi_\nu) \right] \\ &= h_{\mu\nu}^{core} + F_{\mu\nu}^{QM} + F_{\mu\nu}^{ELMO} \qquad (15), \end{aligned}$$

while the relation existing between the Fock matrix $\mathbf{F}'$ expressed in the basis $\boldsymbol{\chi}'$ and the Fock matrix $\mathbf{F}$ expressed in the basis $\boldsymbol{\chi}$ is simply

$$\mathbf{F}' = \mathbf{B}^{\dagger}\mathbf{F}\,\mathbf{B} \quad (16)$$

with $\mathbf{B}$ as the transformation matrix defined in equation (9) and $\mathbf{B}^{\dagger}$ its transpose.

The algorithm of the QM/ELMO method is quite analogous to the one for the original LSCF method and it is shown in the flow chart depicted in Figure 2. It is easy to observe that the computation of the contribution to the Fock matrix due to the transferred and frozen ELMOs is carried out only once before starting the real self-consistent field (SCF) cycle and, afterwards, this contribution is kept frozen. Conversely, at each SCF-iteration, only the $\mathbf{F}^{QM}$ contribution is updated and, furthermore, the diagonalization of the Fock matrix is performed on the reduced and orthogonal basis-set for the QM region, which allows a reduction of the global computational cost. Throughout the SCF procedure, the transferred ELMOs are kept frozen and not polarized by the quantum mechanical part. For this reason, the currently proposed QM/ELMO approach can be classified as a multi-scale electrostatic embedding technique.

However, it is also worth noting that, in the current preliminary implementation of the algorithm, at each iteration, the Fock matrix $\mathbf{F}^{QM}$ is initially computed in the space of the complete set of *M* basis functions for the investigated system. This obviously makes this step the real rate-limiting step of the algorithm. To overcome this drawback we plan to introduce a suitable criterion based on equation (5) to select a reduced number of basis functions, which will limit the size of $\mathbf{F}^{QM}$ and, consequently, will be crucial to further speed up the calculations.

The algorithm was implemented by modifying the Hartree-Fock routines of the Gaussian09 quantum chemistry package,[56] which was afterwards used also to perform all the quantum mechanical calculations described in the paper, including the ONIOM computations that also exploit molecular mechanics as lower-level method (see Supporting Information for a complete description of all the computational details).

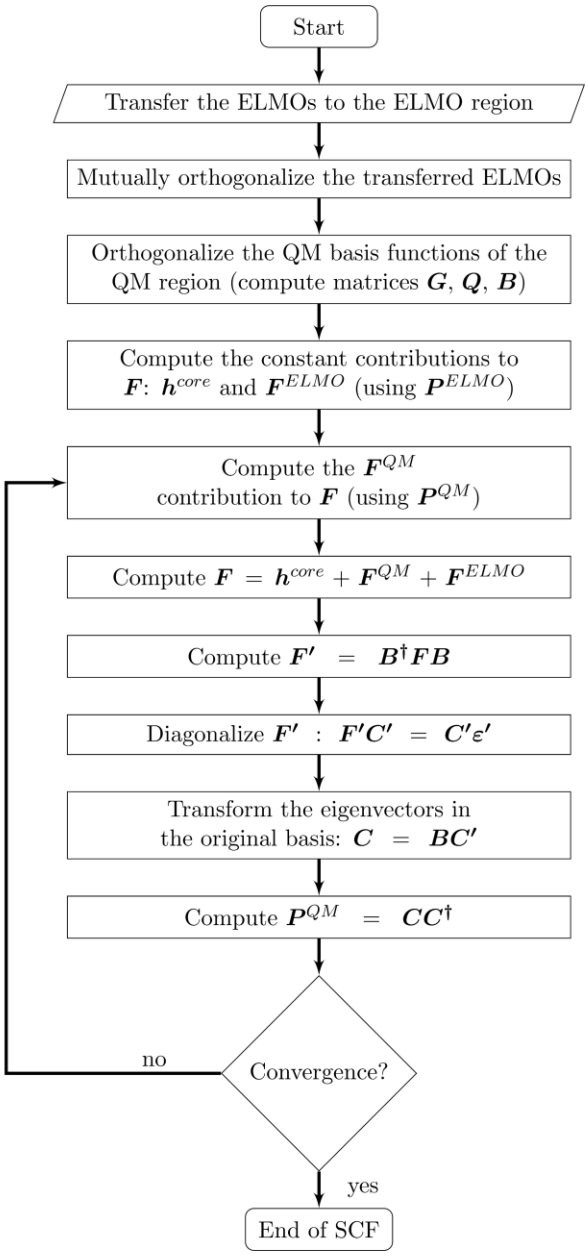



**Figure 2.** Flow chart of the QM/ELMO algorithm implemented in the modified version of Gaussian09.

## 3. Test calculations and discussion

To test its correct functioning, the new technique was initially applied to model homopeptides constituted of 100 serine residues ($Ser_{100}$) in three different conformations: $\alpha$-helix, $\beta$-sheet and globular. For each of them, we carried out different QM/ELMO calculations (cc-pVDZ basis-set) by gradually increasing the size of the QM region and we compared the obtained absolute energies to those resulting from the corresponding full Hartree-Fock/cc-pVDZ computations. The collected results are shown in Table 1. Since the transferred ELMOs are kept frozen during the calculations, the energy values associated with the QM/ELMO method are systematically higher than the Hartree-Fock ones, always leading to positive energy variations ($\Delta E$). The extent of the discrepancies decreases when the number of residues in the quantum mechanical region increases and it converges to zero when the whole system is treated at Hartree-Fock level. This point also confirms that the new strategy is variational and that the HF energy is the lower bound. Similar results were obtained also for the $\alpha$-helix and $\beta$-sheet conformations of the $Trp_{45}$ polypeptide (see Table S1 in the Supporting Information).

A slight but important difference has been observed for the globular conformation when compared to the $\alpha$-helix and $\beta$-sheet ones. In fact, for an equal number of residues treated at fully quantum mechanical level, the energy difference with the Hartree-Fock reference value is always greater than those observed in the $\alpha$-helix and $\beta$-sheet cases (see Table 1). The discrepancy between the energy differences decreases when more amino acids are treated at Hartree-Fock level. This can be explained by considering that, due to its very compact structure, the globular conformation is characterized by a larger number of non-covalent interactions that the transferred and frozen ELMOs can only partially describe. This is also the reason why the inclusion of a

larger number of residues in the QM subsystem entails larger "corrections" compared to the case of the $\alpha$-helix and $\beta$-sheet structures.

**Table 1.** Hartree-Fock energies ($E_{HF}$) and their differences with respect to the QM/ELMO energies ($\Delta E_{QM(N)/ELMO} = E_{QM(N)/ELMO} - E_{HF}$), as resulting from the calculations performed on the $Ser_{100}$ polypeptide.[(a)]

| | α-helix | β-sheet | Globular |
|---|---|---|---|
| $E_{HF}$ | -32148.9857 | -32148.4900 | -32147.6523 |
| $\Delta E_{QM(2)/ELMO}$ | 13.0062 | 12.7134 | 14.1009 |
| $\Delta E_{QM(6)/ELMO}$ | 12.4778 | 12.1953 | 13.5367 |
| $\Delta E_{QM(10)/ELMO}$ | 11.9477 | 11.6773 | 12.9783 |
| $\Delta E_{QM(18)/ELMO}$ | 10.8862 | 10.6412 | 11.8353 |
| $\Delta E_{QM(26)/ELMO}$ | 9.8249 | 9.6048 | 10.6917 |
| $\Delta E_{QM(50)/ELMO}$ | 6.6406 | 6.4961 | 7.1958 |
| $\Delta E_{QM(75)/ELMO}$ | 3.3297 | 3.2578 | 3.6299 |

[(a)] The number of residues ($N$) included in the QM region for the QM/ELMO calculations are reported in parentheses. All the values are expressed in $E_h$.

To complete the analysis of the results obtained on the considered homopeptides, we also evaluated the energy differences between the different conformations, always using the $\alpha$-helix conformer as reference (see Tables S2 and S3 in the Supporting Information for the $Ser_{100}$ and $Trp_{45}$ polypeptides, respectively). As one should expect, in all cases we observe a systematic trend in gradually recovering the Hartree-Fock energy difference between the conformers when the number of residues in the QM region increases. However, in this regard, it is also worth

pointing out that the new QM/ELMO approach, as any other multi-scale embedding technique (e.g., QM/MM strategies), was conceived to mainly study local variations in specific regions of macromolecules (e.g., conformational changes of a ligand in the binding pocket of an enzyme) and not to evaluate energy variations associated with very global conformational changes.

Finally, in Tables 2 and S4 we also reported the computational costs of the performed QM/ELMO calculations. In its current preliminary implementation, the new approach already entails a significant reduction in terms of CPU time, mainly because the global Fock matrix is diagonalized in the reduced and orthogonal basis-set associated with the QM region. However, as already mentioned above, it is also important to note that the computational performances of the technique can be further and significantly improved through the introduction of a criterion to reduce the size of the basis functions space in which the $\mathbf{F}^{QM}$ matrix is constructed.

**Table 2.** CPU time associated with the SCF cycles of the Hartree-Fock and QM/ELMO calculations performed on the $Ser_{100}$ homopeptides. The Hartree-Fock times are given in seconds, while the QM/ELMO ones are expressed as fraction of the corresponding Hartree-Fock reference value. The number of SCF iterations is also reported in parenthesis.

| Method | α-helix | β-sheet | Globular |
|---|---|---|---|
| *Hartree-Fock* | 390623.4 (20) | 328668.5 (21) | 439674.1 (15) |
| $QM(2)/ELMO$ | 0.206 (15) | 0.133 (16) | 0.307 (16) |
| $QM(6)/ELMO$ | 0.230 (16) | 0.136 (15) | 0.337 (17) |
| $QM(10)/ELMO$ | 0.244 (17) | 0.149 (16) | 0.364 (17) |
| $QM(18)/ELMO$ | 0.280 (17) | 0.161 (16) | 0.481 (17) |
| $QM(26)/ELMO$ | 0.280 (16) | 0.151 (16) | 0.558 (19) |
| $QM(50)/ELMO$ | 0.434 (18) | 0.322 (18) | 0.658 (18) |
| $QM(75)/ELMO$ | 0.628 (18) | 0.471 (15) | 0.871 (19) |

After the preliminary test on the correct functioning of the technique, we started assessing the capability of the QM/ELMO approach in reproducing protein-ligand interaction energies. To accomplish this task we have considered the complex formed by the CFTR Associated Ligand (CAL) PDZ domain with the polypeptide iCAL36 (PDB entry: 4E34).[57] Figure 3 shows the three-dimensional structure of the complex, also highlighting the hydrogen bonds formed between the ligand and the protein. Two QM/ELMO calculations (cc-pVDZ basis-set) were performed on this complex: i) one with a QM region that includes all the residues of the polypeptide and those residues of the PDZ domain forming hydrogen bonds or $\pi-\pi$ interactions with the ligand (quantum mechanical region QM1), and ii) one with a larger QM subsystem obtained by adding to the region QM1 those residues of the PDZ domain involved in indirect or hydrophobic interactions with iCAL36 (quantum mechanical region QM2). The complete list of residues included in the QM regions is given in the Supporting Information. The obtained interaction energies (with and without counterpoise correction[58] to account for the basis-set superposition error (BSSE)) were compared to those obtained at full Hartree-Fock/cc-pVDZ level (benchmark values) and to those resulting from different kind of ONIOM calculations, for which we exploited the same partitionings adopted in the QM/ELMO computations to define the higher and lower level regions. For the ONIOM calculations, the lower level regions were treated both at quantum mechanics and molecular mechanics level (see Table 3 and Supporting Information for more details).

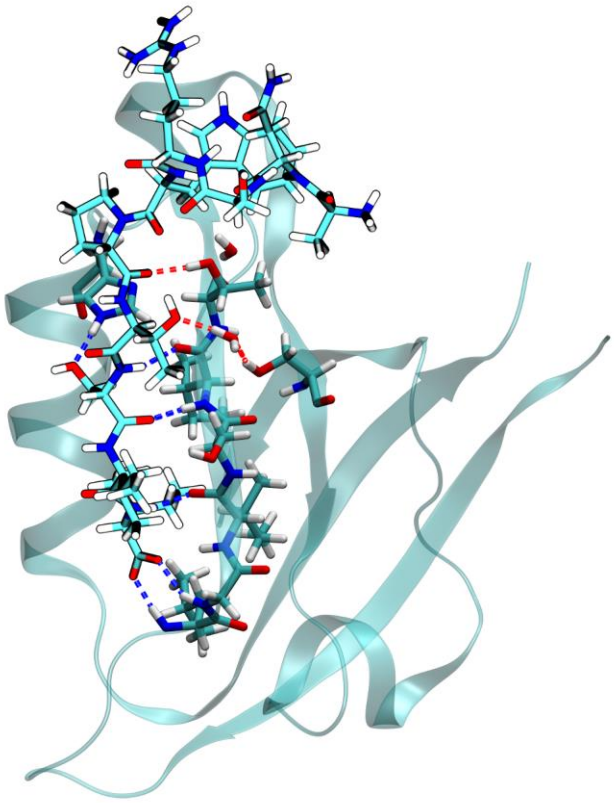


**Figure 3.** Three-dimensional structure of the protein:ligand complex, with the region of the most important intermolecular contacts in "licorice representation". Blue dashed lines indicate hydrogen bonds.

In Table 3 we can observe that the QM/ELMO method provides global interaction energies that are in a very good agreement with those resulting from the full Hartree-Fock calculations. Despite the large number of intermolecular contacts (i.e., 8 strong hydrogen bonds and other hydrophobic interactions; see Figure S2), discrepancies are in absolute value always lower than 5 kcal/mol, independently of the chosen QM region and of the introduction of the counterpoise correction. Furthermore, and more importantly, it is also worth noting that, using the larger quantum mechanical region QM2, the QM/ELMO interaction energies are closer (around 1 kcal/mol) to the Hartree-Fock ones. Concerning the comparison to ONIOM, it is immediately clear that the QM/ELMO strategy gives better results compared to all the performed QM:MM

ONIOM-type calculations, regardless of the use of electronic embedding. Conversely, a less clear trend is observed when the QM:QM' results are taken into account. For example, considering the HF/cc-pVDZ:HF/6-31G case, the ONIOM technique provides a better agreement with the Hartree-Fock benchmark when the quantum mechanical region QM1 is considered, while the QM/ELMO approach outperforms ONIOM when the larger region QM2 is exploited. The opposite behavior is actually observed for the HF/cc-pVDZ:HF/3-21G calculations. However, notwithstanding the observed and unavoidable small discrepancies, it is worth stressing that the QM/ELMO results are completely comparable to those obtained by means of QM:QM' ONIOM calculations, which further confirms the reliability of the proposed approach.

**Table 3.** Interaction energies calculated at Hartree-Fock level ($E_{int,HF}$, with and without counterpoise (CP) correction for BSSE) and their differences with respect to the values obtained through the other methods used in this study ($\Delta E_{int,X} = E_{int,X} - E_{int,HF}$, where $X$ indicates the method).[(a)]

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

[(a)] All the energies are expressed in kcal/mol.

Moreover, unlike the ONIOM approach, the novel QM/ELMO method has also the non-negligible advantage of providing an approximate wave function and electron density for the whole system under investigation. Therefore, the obtained QM/ELMO electron distributions and the reference Hartree-Fock charge density have been compared point-by-point in the crucial region for the protein-ligand interaction by exploiting three different similarity indexes: the traditional root-mean-square deviation (RMSD), the real-space R value[59] (RSR) and the Walker-Mezey index[35] $L(\rho_x, \rho_y, a, a')$. More details about these similarity descriptors are given in the Supporting Information; here, for the sake of clarity and completeness, it is only worth mentioning that complete similarities are observed for values of RMSD, RSR and $L$ equal to 0, 0 and 100, respectively. The obtained values for the similarity indexes are collected in Table 4, where, for the sake of comparison, the values associated with the comparison between the Hartree-Fock and pure ELMO electron densities are also shown. Both QM/ELMO calculations provide electron distributions that are in better agreement with the reference Hartree-Fock one compared to the ELMO case. Furthermore, as one should expect, the agreement improves as a larger QM subsystem is taken into account in the QM/ELMO calculations.

**Table 4.** Values of the similarity indexes corresponding to the comparison of the the Hartree-Fock, ELMO and QM/ELMO electron densities in the region of intermolecular interactions for the protein:ligand complex.[(a)]

| | RMSD | RSR | $L(0.001,10.0)$ | $L(0.001,0.1)$ | $L(0.1,10.0)$ |
|---|---|---|---|---|---|
| HF Vs. ELMO | 0.00212 | 0.00812 | 96.09 | 95.66 | 98.31 |
| HF Vs. QM/ELMO (QM1) | 0.00153 | 0.00448 | 97.67 | 97.40 | 99.06 |
| HF Vs. QM/ELMO (QM2) | 0.00115 | 0.00271 | 98.52 | 98.34 | 99.44 |

[(a)] Acronyms are described in the text.

All these aspects are also graphically confirmed in Figure 4, where we plotted the three-dimensional differences between the ELMO, QM/ELMO and Hartree-Fock charge distributions. For the QM/ELMO residual densities there are no relevant differences between the compared electron densities in the intermolecular interaction region where important hydrogen bonds are formed (compare Figures 4B and 4C to Figure 4A). On the contrary, small residuals start appearing only within the subsystem treated at ELMO level in the QM/ELMO computations. Furthermore, in agreement with the values of the similarity indexes reported in Table 4, we can also observe that the region without electron density residuals clearly expands when a larger QM subsystem is exploited in the QM/ELMO calculations (compare Figure 4C to Figure 4B). In light of these results we believe that the QM/ELMO approach could be fruitfully exploited in the framework of quantum crystallography[60-64] for more accurate refinements of crucial regions of protein crystal structures (e.g., coupling with the quantum chemistry-based Hirshfeld Atom Refinement technique[65,66]).

Finally, also to show the potential usefulness of the QM/ELMO approach in the context of drug design studies, we have computed the electrostatic potentials associated with the obtained Hartree-Fock and the QM/ELMO wave functions for the protein under exam. From visual inspections of Figure 5 and from the analyses of the computed root-mean-square deviations (see the reported RMSD values in Figure 5), it is possible to observe that, also for the electrostatic potentials, the newly developed technique provides results in excellent agreement with the Hartree-Fock benchmark.

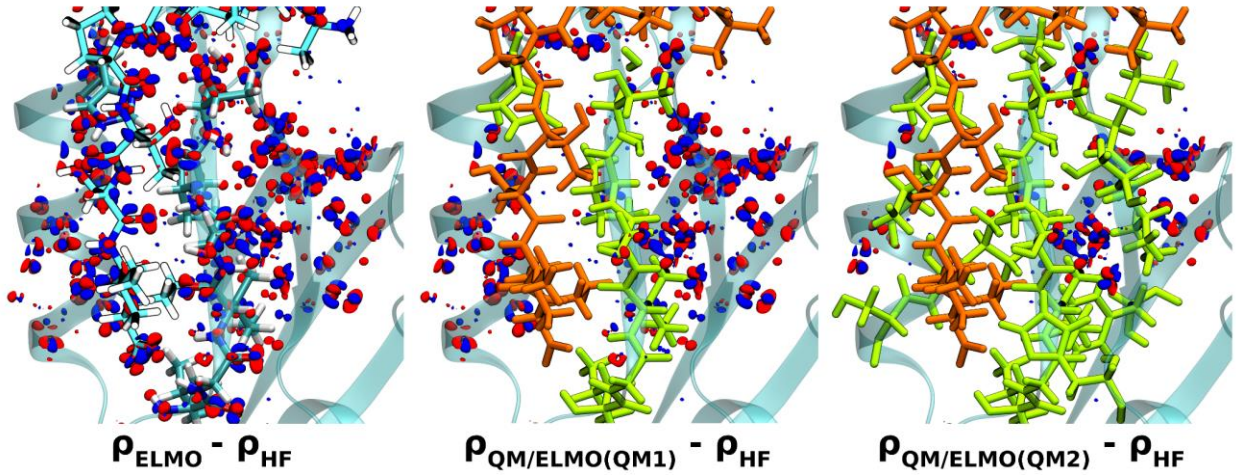



**Figure 4.** Three-dimensional residual densities obtained comparing the ELMO, QM/ELMO (QM1) and QM/ELMO (QM2) electron distributions to the Hartree-Fock one. The isosurface-values are set to 0.01 e/bohr$^3$, with positive and negative isosurfaces in blue and red, respectively; in the ELMO case (left panel) the "licorice representation" is only used to highlight the region of the most important intermolecular contacts, while, for the QM/ELMO comparisons (center and right panels), it is used to indicate the QM regions, with the ligand molecule colored in orange and the protein residues in lime green.

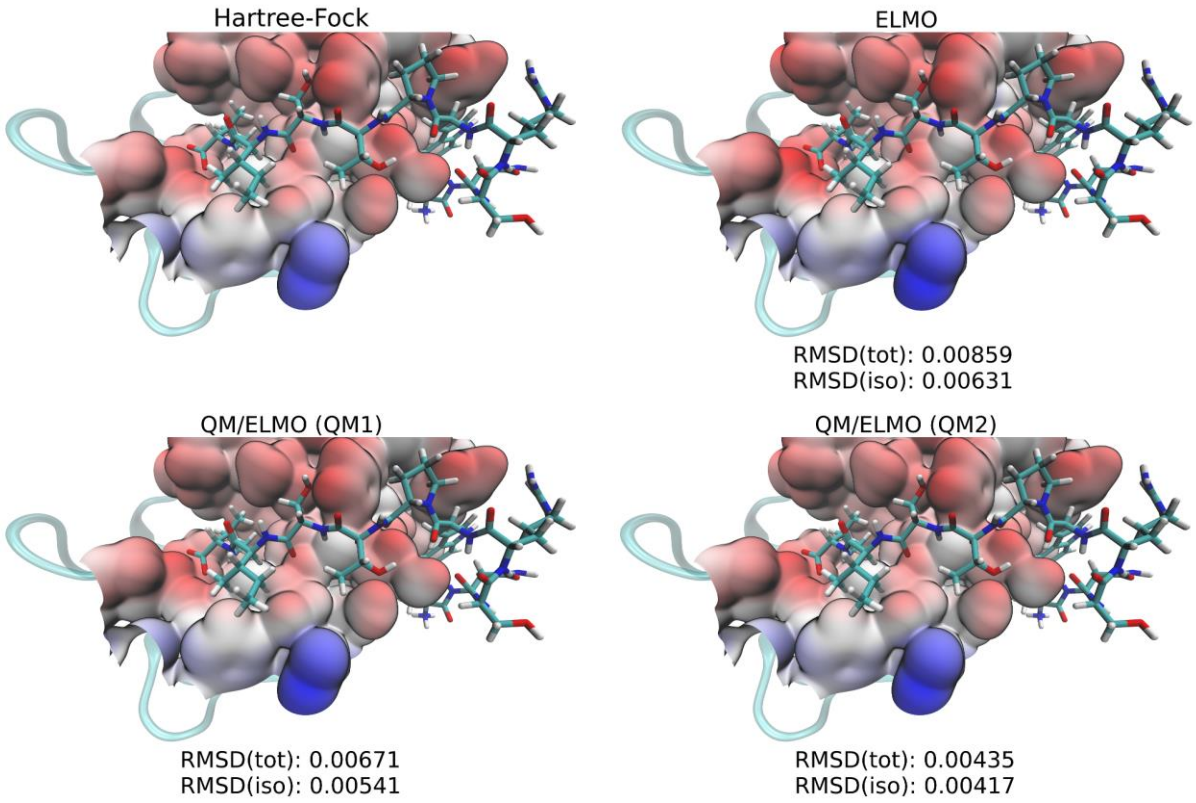



**Figure 5.** Electrostatic potentials (ESPs) plotted onto the corresponding 0.001 e/bohr$^3$ electron density isosurfaces (BWR color-map in the range [-0.332, 0.035] (values in $E_h$)). Below the plots of the ELMO and QM/ELMO electrostatic potentials, their RMSD values with respect to the Hartree-Fock ESP are also shown (see section S3.3 of this Supporting Information for more details about the RMSD(tot) and RMSD(iso) indexes).

## 4. Conclusions

To summarize, exploiting the recently constructed libraries of extremely localized molecular orbitals and properly modifying the LSCF approach, in this paper we have proposed a new multi-scale electrostatic embedding method that allows the investigation of large biological molecules at a fully quantum mechanical level. Preliminary test calculations have shown that the new

QM/ELMO technique works correctly and provides results in very good agreement with those obtained through fully quantum mechanical computations, both in the evaluation of intermolecular interaction energies and in the determination of electron density distributions and electrostatic potentials. Although further and more detailed test calculations will be necessary to completely assess the capabilities of the method and although the technique can be still significantly improved (e.g., introduction of criteria to further speed up the SCF iterations and treatment of the QM region with DFT or post Hartree-Fock strategies), the obtained results are already quite promising and we envisage the future application of the QM/ELMO approach to different fields of physical chemistry, such as enzyme catalysis, *in silico* drug design[67-69] and structural refinement of biological molecules through the techniques of quantum crystallography[60-64].

## ASSOCIATED CONTENT

**Supporting Information**. Details about all the performed computations, details about Extremely Localized Molecular Orbitals and their libraries (with Figure S1 showing plots of ELMOs), details about all the similarity indexes used to compare the obtained molecular electron densities and electrostatic potentials, Table S1 reporting the Hartree-Fock and QM/ELMO energy values resulting from the test calculations performed on $Trp_{45}$ in the $\alpha$-helix and $\beta$-sheet conformations, Tables S2 and S3 showing the Hartree-Fock and QM/ELMO energy differences between the conformations of the $Ser_{100}$ and $Trp_{45}$ homopeptides, Table S4 reporting the computational cost of the Hartree-Fock and QM/ELMO calculations carried out on $Trp_{45}$, Table S5 providing the list of all the residues included in the quantum mechanical (high-level) regions (QM1 and QM2) for the QM/ELMO (ONIOM) calculations performed on the protein:ligand complex. (PDF)

This material is available free of charge via the Internet at http://pubs.acs.org/.

## AUTHOR INFORMATION

### Notes

The authors declare no competing financial interests.

## ACKNOWLEDGMENT

We gratefully acknowledge the French Research Agency (ANR) for financial support of the Young Researcher Project *QuMacroRef* through Grant No. ANR-17-CE29-0005-01. We also thank Xavier Assfeld and Manuel F. Ruiz-López for helpful discussions.

## REFERENCES

(1) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643-5677.

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

(3) 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.

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

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

(6) Yang, W. Direct Calculation of electron Density in Density-Functional Theory. *Phys. Rev. Lett.* **1991**, *66*, 1438-1441.

(7) Yang, W. Direct Calculation of Electron Density in density-Functional Theory: Implementation for Benzene and a Tetrapeptide. *Phys. Rev. A* **1991**, *44*, 7823-7826.

(8) 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.

(9) Dixon, S. L.; Merz, K. M., Jr. Semiempirical Molecular Orbital Calculations with Linear System Size Scaling. *J. Chem. Phys.* **1996**, *104*, 6643-6649.

(10) Dixon, S. L.; Merz, K. M., Jr. Fast, Accurate Semiempirical Molecular Orbital Calculations for Macromolecules. *J. Chem. Phys.* **1997**, *107*, 879-893.

(11) Gogonea, V.; Westerhoff, L. M.; Merz, K. M., Jr. Quantum Mechanical/Quantum Mechanical Methods. I. A Divide and Conquer Strategy for Solving the Schrödinger Equation for Large Molecular Systems Using a Composite Density Functional-Semiempirical Hamiltonian *J. Chem. Phys.* **2000**, *113*, 5604-5613.

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

(13) 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.

(14) 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.

(15) 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.

(16) He, X.; Zhang, J. Z. H. A New Method for Direct Calculation of Total Energy of Protein. *J. Chem. Phys.* **2005**, 122, 031103.

(17) 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.

(18) Babu, K.; Gadre, S. R. Ab Initio Quality One-Electron Properties of Large Molecules: Development and Testing of Molecular Yailoring Approach. *J. Comput. Chem.* **2003**, *24*, 484-495.

(19) 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.

(20) Sahu, N.; Gadre, S. R. Molecualr Tailoring Approach: A Route for *Ab Initio* Treatment of Large Clusters. *Acc. Chem. Res.* **2014**, *47*, 2739-2747.

(21) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method Illustrated with Peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808-817.

(22) Huang, L.; Massa, L.; Karle, J. Kernel Energy Method: Application to Insulin. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12690-12693.

(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.; 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.

(25) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A Multilayered Integrated MO+MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels-Alder Reactions and Pt(P(tBu)3)2 + H2 Oxidative Addition. *J. Phys. Chem.* **1996**, *100*, 19357-19363.

(26) 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.

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

(28) 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.

(29) Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Fragment Molecular Orbital Method: an Approximate Computational Method for Large Molecules. *Chem. Phys. Lett.* **1999**, *313*, 701-706.

(30) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. Fragment Molecular Orbital Method: Application to Polypeptides. *Chem. Phys. Lett.* **2000**, *318*, 614-618.

(31) Fedorov, D. G.; Kitaura, K. Theoretical Development of the Fragment Molecular Orbital (FMO) Method. In *Modern Methods for Theoretical Physical Chemistry and Biopolymers*; Starikov, E. B., Lewis, J. P., Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; Chapter 1, pp 3-38.

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

(33) Pruitt, S. R.; Bertoni, C.; Brorsen, K. R.; Gordon, M. S. Efficient and Accurate Fragmentation Methods. *Acc. Chem. Res.* **2014**, *47*, 2786-2794.

(34) Walker, P. D.; Mezey, P. G. Molecular Electron Density Lego Approach to Molecule Building. *J. Am. Chem. Soc.* **1993**, *115*, 12423-12430.

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

(36) 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.

(37) Exner, T. E.; Mezey, P. G. Ab Initio Quality Properties for Macromolecules Using the ADMA Approach. *J. Comput. Chem.* **2003**, *24*, 1980-1986.

(38) 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.

(39) 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.

(40) 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.

(41) 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.

(42) 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.

(43) 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.

(44) Wesolowski, T. A.; Warshel, A. Frozen Density Functional Approach for Ab Initio Calculations of Solvated Molecules. *J. Phys. Chem.* **1993**, *97*, 8050-8053.

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

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

(47) Poteau, R.; Ortega, I.; Alary, F.; Ramirez Solis, A.; Barthelat, J.-C.; Daudey, J.-P. Effective Group Potentials. 1. Method. *J. Phys. Chem. A* **2001**, *105*, 198-205.

(48) Poteau, R.; Alary, F.; Abou El Makarim, H; Heully, J.-L.; Barthelat, J.-C.; Daudey, J.-P. Effective Group Potentials. 2. Extraction and Transferability for Chemical Groups Involved in Covalent or Donor-Acceptor Bonds. *J. Phys. Chem. A* **2001**, *105*, 206-214.

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

(50) 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.

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

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

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

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

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

(56) 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.

(57) Amacher, J. F.; Cushing, P. R.; Bahl, C. D.; Beck, T.; Madden, D. R. Stereochemical Determinants of C-terminal specificity in PDZ Peptide-binding Domains. *J. Biol. Chem.* **2013**, *288*, 5114-5126.

(58) Boys, S. F.; Bernardi, F. Calculation of Small Molecular interactions by Differences of Separate Total Energies – Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553-566.

(59) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in these Models. *Acta Crystallogr., Sect. A* **1991**, *47*, 110−119.

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

(61) Novara, R. F.; Genoni, A.; Grabowsky, S. What is Quantum Crystallography? *ChemViews* **2018**, DOI: 10.1002/chemv.201800066.

(62) Grabowsky, S.; Genoni, A.; Bürgi, H.-B. Quantum Crystallography. *Chem. Sci.* **2017**, *8*, 4159-4176.

(63) Massa, L.; Matta, C. F. Quantum Crystallography: A Perspective. *J. Comput. Chem.* **2017**, *39*, 1021-1028.

(64) Tsirelson, V. Early Days of Quantum Crystallography: A Personal Account. *J. Comput. Chem.* **2017**, *39*, 1029-1037.

(65) Capelli, S.C.; Bürgi, H.-B.; Dittrich, B.; Grabowsky, S.; Jayatilaka, D. Hirshfeld Atom Refinement. *IUCrJ* **2014**, *1*, 361-379.

(66) 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.

(67) Morra, G; Genoni, A.; Neves, M. A. C.; Merz, K. M., Jr.; Colombo, G. Molecular Recognition and Drug-Lead identification: What Can Molecular Simulations Tell Us? *Curr. Med. Chem.* **2010**, *17*, 25-41.

(68) Legnani, L.; Compostella, F.; Sansone, F.; Toma, L. Cone calix[4]arenes with orientable glycosylthioureido groups at the upper rim: an in-depth analysis of their symmetry properties. *J. Org. Chem.* **2015**, *80*, 7412-7418.

(69) Ferraro, M.; D'Annessa, I.; Moroni, E.; Morra, G.; Paladino, A.; Rinaldi, S.; Compostella, F.; Colombo, G. Allosteric Modulators of HSP90 and HSP70: Dynamics Meets Function through Structure-Bases Drug Design. *J. Med. Chem.* **2019**, *62*, 60-87.