Libraries of Extremely Localized Molecular Orbitals. 1. Model Molecules Approximation and Molecular Orbitals Transferability

Benjamin Meyer, Benoît Guillot, Manuel F. Ruiz-Lopez, Alessandro Genoni

Libraries of Extremely Localized Molecular Orbitals. I. Model Molecules Approximation and Molecular Orbitals Transferability

Benjamin Meyer [(1,2)], Benoît Guillot [(3,4)], Manuel F. Ruiz-Lopez [(1,2)], Alessandro Genoni [(1,2)] *

(1) CNRS, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France.

(2) Université de Lorraine, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France.

(3) CNRS, Laboratoire CRM2, UMR 7036, Vandoeuvre-lès-Nancy, F-54506, France.

(4) Université de Lorraine, Laboratoire CRM2, UMR 7036, Vandoeuvre-lès-Nancy, F-54506, France.

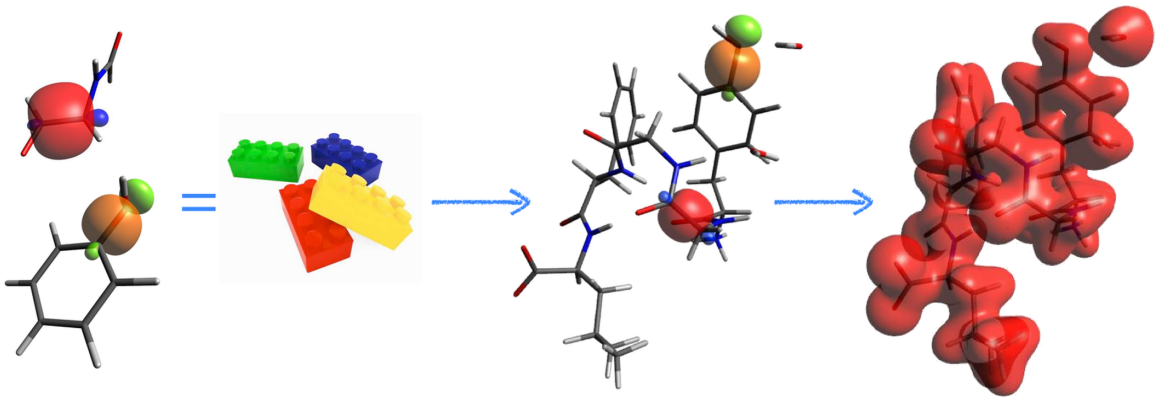
* Correspondence to:

Alessandro Genoni, CNRS and Université de Lorraine, Laboratoire SRSMC - UMR 7565, Boulevard des Aiguillettes, BP 70239, 54506 Vandoeuvre-lès-Nancy, France. E-mail: Alessandro.Genoni@univ-lorraine.fr; Phone: +33 (0)3 83 68 43 77; Fax: +33 (0)3 83 68 43 71.

Abstract

Despite more and more remarkable computational *ab initio* results are nowadays continuously obtained for large macromolecular systems, the development of new linear scaling techniques is still an open and stimulating field of research in theoretical chemistry. In this family of methods, an important role is occupied by those strategies based on the observation that molecules are generally constituted by recurrent functional units with well-defined intrinsic features. In this context, we propose to exploit the notion of Extremely Localized Molecular Orbitals (ELMOs) that, due to their strict localization on small molecular fragments (e.g., atoms, bonds or functional groups), are in principle transferable from one molecule to another. Accordingly, the construction of orbitals libraries to almost instantaneously build up approximate wave functions and electron densities of very large systems becomes conceivable. In this paper, the ELMOs transferability is further investigated in detail and, furthermore, suitable rules to construct model molecules for the computation of ELMOs to be stored in future databanks are also defined. The obtained results confirm the reliable transferability of the ELMOs and show that electron densities obtained from the transfer of Extremely Localized Molecular Orbitals are very close to the corresponding Hartree-Fock ones. These observations prompt us to construct new ELMOs databases that could represent an alternative/complement to the already popular pseudoatoms databanks both for determining electron densities and for refining crystallographic structures of very large molecules.

TOC Graphic



1. Introduction

Nowadays, both the steady advances in computer technologies and the tremendous efforts committed to the improvement of standard quantum chemistry methods have allowed electronic structure calculations on larger and larger systems that were previously treated by means of classical force fields. Nevertheless, although *ab initio* computations at Hartree-Fock (HF) and Density Functional Theory (DFT) levels with moderate-size basis-sets are now routine for molecules consisting of few hundreds of atoms[1-5], a completely quantum mechanical treatment of larger systems (e.g., very large biomolecules) remains a challenge.

To tackle this problem, over the years, many research groups have developed linear-scaling strategies (N-methods), which are approximate quantum-mechanical techniques where the computational cost scales linearly with the increasing size of the system[6,7]. Among them a prominent position is occupied by the well know "Divide & Conquer" method originally devised by Yang in 1991[8,9] and currently used in the framework of DFT, HF and semi-empirical approaches[10-14]. This strategy is essentially based on the preliminary partition of large molecules in small overlapping fragments characterized by local Hamiltonians that are projections of the global one on the different subsystems. An electronic problem is thus solved for each resulting subunit and, afterwards, the overall electron density (or density matrix) of the target molecule is properly assembled. In this context, the Molecular Tailoring Approach introduced by Gadre and coworkers[15-17] follows a similar philosophy. In fact, after the subdivision of a parent system into fragments, calculations at HF (or MP2) level are performed to obtain subunit density matrices that, at a later stage, are properly combined to obtain the global density matrix.

Another important family of linear-scaling methods is represented by the fragment interaction techniques. Although these approaches are based on a partition into different subunits as well, they distinguish from the previous ones in the exploitation of the theory of molecular interactions that allows writing the total energy of the systems as a sum of monomer energies and intermolecular interaction energies corresponding to dimers and larger molecular assemblies. In this group of methods it is worth mentioning the MFCC (Molecular Fractionation with Conjugated Caps) approach[18-24] originally devised by Zhang and Zhang and more recently improved by Li and coworkers through an energy-corrected formalism (EC-MFCC technique) [25]. However, the most representative strategies of this family of techniques are probably the Fragment Molecular Orbital (FMO) approach [26-30] and the Kernel Energy Method (KEM).[31-38] Including both the effects of the electrostatic field of the whole system and the exchange interactions with the other fragments in all the subunits calculations, the former has rapidly grown in few years and, nowadays, it is recognized as a useful technique to investigate the electronic structure of very large and complex molecules and molecular clusters. At the same time, the latter, which subdivides large molecules into small fragments capped with hydrogen atoms and which mainly relies on an equation that fully considers at least self- and two-body interaction terms, has turned out to be an extremely reliable strategy for the study of macromolecules, ranging from small proteins to graphene and extended aromatics even in presence of external electric fields. The importance and reliability of the Kernel Energy Method has been also recently confirmed by recent investigations[39,40] showing how this technique can be also used to accurately reconstruct electron densities, electrostatic potentials, QTAIM (Quantum Theory of Atoms in Molecules[41]) localization and delocalization matrices and QTAIM charges of very large molecules.

Finally, a third important group of N-methods is represented by all those approaches that exploit the transferability principle starting from the observation that molecular systems are generally composed of units (e.g., functional groups) that keep their main features in different environments. A noteworthy example is the pioneering Molecular Electron Density Lego Assembler (MEDLA) technique[42-44] proposed by Walker and Mezey that, using a properly constructed database of fuzzy electron densities associated with small molecular fragments, have successfully reconstructed *ab initio* quality charge distributions of very large systems with a sensitively reduced computational cost. Afterwards, the method has been also extended by Exner and Mezey giving rise to the Adjustable Density Matrix Assembler (ADMA) approach[45,46], which is based on a library of fragment density matrices and which also allowed the fast calculation of electron densities and electrostatic potentials of macromolecules. Other techniques closely related to those devised by Mezey and coworkers are undoubtedly the Transferable Atom Equivalents (TAE) method[47,48] developed by Breneman *et al.* and the real-space approach, initially introduced by Bader[49-51] and more recently reconsidered by Matta.[52]

In this context, a further alternative is represented by the possibility of defining Molecular Orbitals (MOs) strictly localized on small molecular subunits (e.g., atoms, bonds or functional groups) and, consequently, easily transferable to other molecules containing the same fragment. Unfortunately, the canonical Hartree-Fock orbitals are not suitable to this purpose due to their completely delocalized nature that does not disappear even after the application of traditional localization procedures[53-57]. In fact, orthogonalization tails remain outside the main localization regions and, although quite small, their deletion entails a dramatic increase of the molecular electronic energy.

This drawback can be overcome only considering Molecular Orbitals variationally determined under the constraint of expanding them only in subsets of basis functions centered on atoms belonging to pre-selected fragments. These orbitals are indeed tail-free Extremely Localized Molecular Orbitals (ELMOs) and are suitable to be transferred from a molecule to another.

To obtain ELMOs several researchers have successfully proposed similar techniques[58-68], each of them strictly connected to the original McWeeny group function method[69-71] that can be considered as the first attempt of decomposing a global electronic wave function into functions describing smaller subsets of electrons. Among all these strategies, the method introduced by Stoll and coworkers[62] in 1980 represents an important turning point since it was the first generalization of the Hartree-Fock equations to determine Extremely Localized Molecular Orbitals. The Stoll equations have been afterwards implemented by Fornili *et al.*[67] in a robust algorithm and, more recently, they have been also successfully extended in order to reproduce experimental structure factors[72-75] following the Jayatilaka X-ray constrained wave function approach[76-82], which originates from the original and pioneering work of Clinton and Massa.[83-87]

Since the ELMOs are expected to be transferable, the construction of datasets of Extremely Localized Molecular Orbitals that cover all the possible functional units of a particular class of molecules represents an appealing perspective in order to almost instantaneously obtain wave functions or electron densities of extended systems. However, it is necessary to observe that the use of transferred ELMOs might not be the most suitable choice to accurately study highly polarizable and/or delocalized macromolecules (e.g., functionalized graphene nanoribbons). Therefore, further test

calculations in this direction will be of paramount importance to completely assess the capabilities of the proposed technique.

At the moment, our main goal is to build up proper libraries of ELMOs both to promptly determine electron distributions and to refine crystallographic structures of macromolecular systems. Of course, these libraries may be considered as alternatives/complements to the already existing and successful databases of pseudoatoms, which have been independently developed by different research groups in the framework of the multipole model approaches in crystallography[88-97]. To this aim, the ELMOs and the pseudoatoms transferabilities have been recently compared in detail showing that, indeed, the Extremely Localized Molecular Orbitals can be reliably used for the two main goals just mentioned above.[98]

For the sake of completeness, it is worth mentioning that a database of ELMOs has been already constructed.[99] Nevertheless, this library is limited only to the very small STO-4G basis-set and, more importantly, the associated transfer protocol gives a final geometry of the target system that is always slightly different from the input one. This obviously makes this dataset and the corresponding transfer protocol unsuitable to our purpose of refining crystallographic structures.

The main objective of this paper is to present the results of our preliminary investigations for the construction of new ELMOs libraries. In particular, at first, we will focus on the studies performed to define the suitable level of approximation in the model molecules definition that will be used to compute the Extremely Localized Molecular Orbitals of the databanks. At the same time, this analysis will also allow us to further assess the ELMOs transferability to large systems since, until now, this problem has been examined only taking into account quite small molecules.[67,100,101] Finally, at a second stage, we will analyze both the effects of using localized

molecular orbitals and the influence of the adopted basis-sets in the reconstruction of macromolecular electron densities.

The paper is organized as follows. At first, we will briefly review the theory of the ELMO method and the strategy adopted to transfer the Extremely Localized Molecular Orbitals. At a later stage, we will dedicate a section to describe all the computational details, especially focusing on the description of the model molecules approximations taken into account in our investigation. Afterwards, we will present and analyze in detail the results obtained for a case study, namely the Leu-enkephalin pentapeptide. Finally, in the last section, we will draw some conclusions and discuss some perspectives for the future development of ELMOs libraries.

2. Theory

In this section we will briefly describe the theoretical foundations for the construction of ELMOs libraries. In particular, in the first part, we will provide an essential description of the Stoll strategy[62] for the ELMOs computation, while, in the second part, for the sake of completeness, we will present the definition of ELMO-rotation, which has been introduced by Philipp and Friesner[102] and which has been adopted to transfer the ELMOs from our libraries.

Determination of Extremely Localized Molecular Orbitals. Let us consider a *2N*-electron closed-shell molecule. According to the traditional chemical concepts of core electrons, lone-pairs, covalent bonds and functional groups, let us introduce a localization scheme that subdivides the system in exam into a set of overlapping fragments as illustrated in Figure 1. As we will discuss more in detail below, the fragmentation procedure is not unique and, in some circumstances, it may appear arbitrary. Assuming for the moment that this *a priori* fragmentation can be done, each

fragment or subunit i is automatically associated with a local basis-set of atomic orbitals $\beta_i = \left\{ \left| \chi_{i\mu} \right\rangle \right\}_{\mu=1}^{M_i}$ consisting of the only basis functions centered on the atoms belonging to the fragment. Consequently, the ELMOs of the different subunits are expanded only in the corresponding local basis-sets and, for instance, the generic α-th ELMO of the i-th fragment can be simply written as follows:

$$| \varphi_{i\alpha} \rangle = \sum_{\mu=1}^{M_i} C_{i\mu,i\alpha} \left| \chi_{i\mu} \right\rangle \qquad (1)$$

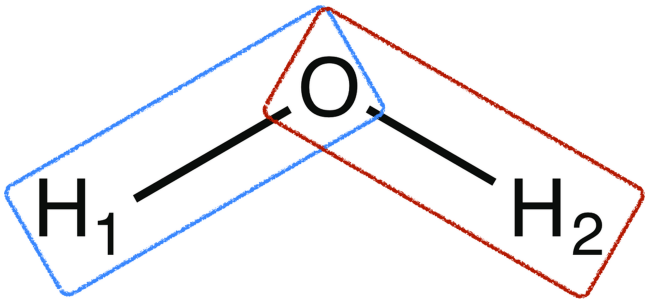


Figure 1. Localization scheme for the water molecule with the two overlapping bond fragments O-H_1 and O-H_2 explicitly shown. For the sake of clarity, the atomic fragment O, which describes core and lone-pair electrons of the oxygen atom, is not depicted.

Following Stoll[62], the wave function that describes the global system is a normalized single Slater determinant built up with Extremely Localized Molecular Orbitals (ELMO wave function):

$$|\Psi_{ELMO}\rangle = \frac{1}{\sqrt{(2N)!\, det[\boldsymbol{S}]}} \hat{A} \left[\varphi_{11}\, \overline{\varphi}_{11} \dots \varphi_{1n_1}\, \overline{\varphi}_{1n_1} \dots \varphi_{f1}\, \overline{\varphi}_{f1} \dots \varphi_{fn_f}\, \overline{\varphi}_{fn_f} \right] \quad (2)$$

where $\hat{A}$ is the antisymmetrizer, n_i is the number of occupied ELMOs for the i-th fragment, $\varphi_{i\alpha}$ is a spin-orbital with spatial part $\varphi_{i\alpha}$ and spin part α and $\overline{\varphi}_{i\alpha}$ is a spin-orbital with spatial part $\varphi_{i\alpha}$ and spin part β. Finally, $det[\boldsymbol{S}]$ is the determinant of the occupied ELMOs overlap-matrix, which is due to the ELMOs non-orthogonality

arising from the fact that the pre-defined overlapping fragments (see Figure 1) share part of their local basis-sets.

Like the Hartree-Fock MOs are obtained minimizing the energy of the single Slater determinant built up with them, in the same way the ELMOs are determined variationally minimizing the energy corresponding to the ELMO wave function:

$$E[\boldsymbol{\varphi}] = \langle \Psi_{ELMO} | \hat{H} | \Psi_{ELMO} \rangle \qquad (3)$$

where $\boldsymbol{\varphi}$ stresses the functional dependence of the energy on the occupied ELMOs and $\hat{H}$ is the non-relativistic Hamiltonian operator of the system in exam.

Now, through simple mathematical derivations, the variation of the functional $E[\boldsymbol{\varphi}]$ with respect to the arbitrary occupied ELMO $\left| \varphi_{i\beta} \right\rangle$ can be expressed as

$$\delta_{(i\beta)} E[\boldsymbol{\varphi}] = 4 \left\langle \delta\varphi_{i\beta} \left| (1 - \hat{\rho}) \, \hat{F} \, (1 - \hat{\rho} + \hat{\rho}_i) \right| \varphi_{i\beta} \right\rangle \qquad (4)$$

with $\hat{\rho}$ being the global density operator given by

$$\hat{\rho} = \sum_{i,j=1}^{f} \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} |\varphi_{i\alpha}\rangle \, [\boldsymbol{S}^{-1}]_{i\alpha,j\beta} \, \langle \varphi_{j\beta}| \qquad (5)$$

and $\hat{\rho}_i$ being the local density operator for the *i*-th fragment, which can be expressed like this:

$$\hat{\rho}_i = \sum_{j=1}^{f} \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} |\varphi_{i\alpha}\rangle \, [\boldsymbol{S}^{-1}]_{i\alpha,j\beta} \, \langle \varphi_{j\beta}| \qquad (6)$$

Since the minimum of the functional is reached when $\delta_{(i\beta)} E[\boldsymbol{\varphi}]$ vanishes for all *j* and *β*, the desired ELMOs are the orbitals that satisfy the following equation for each subunit:

$$(1 - \hat{\rho}) \, \hat{F} \, (1 - \hat{\rho} + \hat{\rho}_i) \, \left| \varphi_{i\beta} \right\rangle = 0 \qquad (7)$$

which is actually equivalent to the eigenvalue problem

$$\hat{F} \left| \varphi_{i\beta} \right\rangle = \varepsilon_{i\beta} \left| \varphi_{i\beta} \right\rangle \qquad (8)$$

where $\hat{F}_i$ is the modified Fock operator for the *i*-th fragment:

$$\hat{F}_i = \left(1 - \hat{\rho} + \hat{\rho}_i^{\dagger}\right) \hat{F} (1 - \hat{\rho} + \hat{\rho}_i) \quad (9)$$

Therefore, the Extremely Localized Molecular Orbitals of the closed-shell molecule in exam are obtained solving self-consistently the modified Hartree-Fock equations (8) for each subunit, equations that are coupled since each operator $\hat{F}_i$ depends on the global density operator $\hat{\rho}$ (see Equation (9)).

It is worth mentioning that an alternative way to determine Extremely Localized Molecular Orbitals consists in a direct minimization of the ELMO energy through a quasi-Newton procedure exploiting an approximate analytic Hessian that is evaluated only at the first iteration and that is afterwards updated by means of the Broyden-Fletcher-Goldfarb-Shanno formula.[103] This approach has been proposed and reliably implemented by Fornili and coworkers[67] to overcome convergence problems that sometimes arise in the solution of equations (8) due to the non-orthogonal nature of the ELMOs.

ELMOs Rotation. As mentioned in the introduction, our final goal will be the use of Extremely Localized Molecular Orbitals stored in suitable libraries to reconstruct the electron densities and refine crystallographic structures of very large systems. To accomplish this task, the selected ELMOs must be properly transferred (namely, rotated) from the model molecule geometries, on which they have been originally determined, to the target system geometry. The rotation matrix of each ELMO can be obtained exploiting the Philipp and Friesner strategy[102] that will be thoroughly reviewed below. Before describing this technique in detail, it is worth mentioning that this method assumes that the ELMOs are expressed in terms of Cartesian Gaussian

basis functions and, more importantly, that it does not automatically take into account the variations in the overlap between basis functions due to the changes of bond lengths and angles in the target molecule. Therefore, to overcome this drawback, all the ELMOs must be properly renormalized after their rotations.

To define a suitable rotation matrix for the coefficients of an ELMO, it is crucial to define a reference frame both for the model and for the target molecule. According to Philipp and Friesner, in order to guarantee the uniqueness of the rotation, three atoms must be used in the definition of these two reference frames. In the most common situation of a bond-type ELMO, along with the two atoms involved in the bond, it is necessary to select a third atom that is bonded to one of the other two and that represents the local dissymmetry of the bond in exam.[104] Of course, in the case of a three-center ELMO (e.g., peptide-bond, carboxylate group or phenyl ring subunits), which corresponds to the most delocalized orbital considered in this study, the three atoms are automatically chosen, while for ELMOs describing atomic fragments, the triad can be defined by the atom in exam and two other atoms, usually the ones directly bonded to it.

In order to better introduce the definition of ELMO-rotation proposed by Philipp and Friesner[102], in this subsection, the triads of atoms in the model and target molecules will be indicated as (A_1, A_2, A_3) and (A_1', A_2', A_3'), respectively. They enable to define the two associated reference frames that are given by the triads of vectors (**a**, **c**, **d**) and (**a'**, **c'**, **d'**), with **a** (**a'**) as the position vector of A_2 (A_2') relative to A_1 (A_1') and with

$$\begin{cases} \mathbf{c} = \mathbf{a} \times \mathbf{b} & (\mathbf{c}' = \mathbf{a}' \times \mathbf{b}') \\ \mathbf{d} = \mathbf{c} \times \mathbf{a} & (\mathbf{d}' = \mathbf{c}' \times \mathbf{a}') \end{cases} \quad (10)$$

where **b** (**b'**) is the position vector of A_3 (A_3') with respect to A_1 (A_1') (see Figure 2).

A fundamental step to determine the rotation matrix of the ELMO coefficients consists in obtaining the matrix **P** associated with the transformation from (**a**, **c**, **d**) to (**a'**, **c'**, **d'**). Following Philipp and Friesner, this rotation can be considered as the composition of the two consecutive transformations schematized in Figure 2, i.e. the rotation from the reference frame (**a**, **c**, **d**) to the reference orthonormal frame ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$) and the rotation from ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$) to (**a'**, **c'**, **d'**).

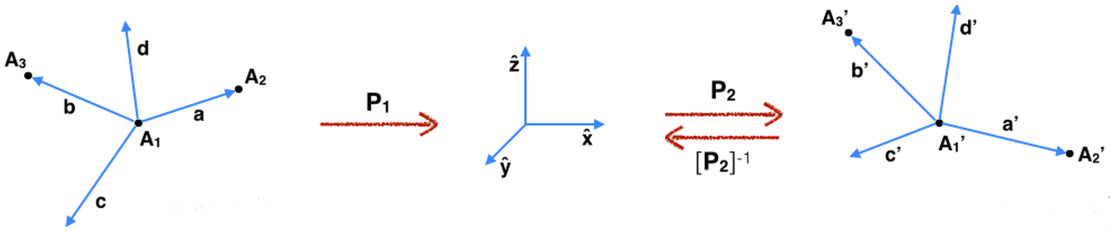


Figure 2. Definition of the reference frames for the rotation from the geometry of the model molecule to the geometry of the target system.

Bearing in mind that each generic vector ***k*** can be simply expressed in the orthonormal reference frame ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$) as

$$\mathbf{k} = k_x\hat{\mathbf{x}} + k_y\hat{\mathbf{y}} + k_z\hat{\mathbf{z}} \quad (11),$$

the matrix $\mathbf{P}_1$ associated with the former transformation is the directional cosine matrix, namely

$$\mathbf{P}_1 = \begin{pmatrix} \frac{d_x}{|\mathbf{d}|} & \frac{d_y}{|\mathbf{d}|} & \frac{d_z}{|\mathbf{d}|} \\ \frac{c_x}{|\mathbf{c}|} & \frac{c_y}{|\mathbf{c}|} & \frac{c_z}{|\mathbf{c}|} \\ \frac{a_x}{|\mathbf{a}|} & \frac{a_y}{|\mathbf{a}|} & \frac{a_z}{|\mathbf{a}|} \end{pmatrix} \quad (12),$$

while the matrix $\mathbf{P}_2$ corresponding to the latter rotation is the inverse (or equivalently the transpose, since rotation matrices are orthogonal matrices) of the matrix associated with the transformation from (**a'**, **c'**, **d'**) to ($\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{z}}$) (see Figure 2), namely

$$\mathbf{P}_2 = \begin{pmatrix} \frac{\mathrm{d'_x}}{|\mathbf{d}'|} & \frac{\mathrm{c'_x}}{|\mathbf{c}'|} & \frac{\mathrm{a'_x}}{|\mathbf{a}'|} \\ \frac{\mathrm{d'_y}}{|\mathbf{d}|'} & \frac{\mathrm{c'_y}}{|\mathbf{c}'|} & \frac{\mathrm{a'_y}}{|\mathbf{a}'|} \\ \frac{\mathrm{d'_z}}{|\mathbf{d}'|} & \frac{\mathrm{c'_z}}{|\mathbf{c}'|} & \frac{\mathrm{a'_z}}{|\mathbf{a}'|} \end{pmatrix} \quad (13)$$

Therefore, the rotation matrix that aligns the reference frame of the model molecule (**a**, **c**, **d**) to the one of the target system (**a'**, **c'**, **d'**) is:

$$\mathbf{P} = \mathbf{P}_2\mathbf{P}_1 = \begin{pmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{pmatrix} \quad (14)$$

Using the transformation matrix **P**, now we can define the rotation matrices for the different types of basis functions. The simplest case is represented by the *s* functions that, due to their spherical symmetry, are invariant to rotations. Therefore, the matrix that transforms the *s* functions and the related coefficients is simply given by:

$$\mathbf{S} = S_{11} = \mathbf{I} = 1 \quad (15)$$

On the contrary, the matrix **P** obtained through equation (14) can be used to directly transform the ELMO coefficients corresponding to the *p* functions, since, during the rotation, the latter behave as the coordinates *x*, *y* and *z*. Finally, the matrices for rotating basis functions (and the associated coefficients) with angular momentum greater than 1 can be easily expressed in terms of **P**.

For example, taking into account the normalization factors, the rotated Cartesian Gaussian basis function d'_{x^2} can be written as

$$d'_{x^2} = \frac{1}{\sqrt{3}} \frac{p'_x\, p'_x}{s'} \quad (16)$$

where, as just mentioned above, the rotated basis function s' is simply the original *s* function while the rotated Cartesian Gaussian function p'_x is given by:

$$p'_x = P_{xx}\, p_x + P_{yx}\, p_y + P_{zx}\, p_z \quad (17)$$

with p_x, p_y and p_z as the starting Cartesian Gaussian p functions and with P_{ij} as the elements of the rotation matrix **P**.

Hence, substituting equation (17) into equation (16), we easily obtain:

$$d'_{x^2} = \frac{1}{\sqrt{3}} \left[P_{xx}^2 \frac{p_x p_x}{s} + P_{yx}^2 \frac{p_y p_y}{s} + P_{zx}^2 \frac{p_z p_z}{s} + 2\, P_{xx} P_{yx} \frac{p_x p_y}{s} + 2\, P_{xx} P_{zx} \frac{p_x p_z}{s} + 2\, P_{yx} P_{zx} \frac{p_y p_z}{s} \right] \quad (18)$$

Now, bearing in mind that for the three Cartesian Gaussian basis functions d_{x^2}, d_{y^2} and d_{z^2}

$$\frac{p_k\, p_k}{s} = \sqrt{3}\, d_{k^2} \quad (k = x, y, z) \qquad (19)$$

and that for d_{xy}, d_{xz} and d_{yz}

$$\frac{p_h\, p_k}{s} = d_{hk} \quad (h, k = x, y, z \text{ and } h \neq k) \quad (20),$$

equation (18) becomes

$$d'_{x^2} = P_{xx}^2\, d_{x^2} + P_{yx}^2\, d_{y^2} + P_{zx}^2\, d_{z^2} + \frac{2}{\sqrt{3}} P_{xx} P_{yx}\, d_{xy} + \frac{2}{\sqrt{3}} P_{xx} P_{zx}\, d_{xz} + \frac{2}{\sqrt{3}} P_{yx} P_{zx}\, d_{yz} \quad (21)$$

where the coefficients multiplying the six starting Cartesian basis functions correspond to the elements of the first column of the rotation matrix for the d functions. Therefore, using the same strategy for d'_{y^2}, d'_{z^2}, d'_{xy}, d'_{xz} and d'_{yz}, we obtain the following 6×6 matrix:

$$\mathbf{D} = \begin{pmatrix} P_{xx}^2 & P_{xy}^2 & P_{xz}^2 & \sqrt{3}P_{xx}P_{xy} & \sqrt{3}P_{xx}P_{xz} & \sqrt{3}P_{xy}P_{xz} \\ P_{yx}^2 & P_{yy}^2 & P_{yz}^2 & \sqrt{3}P_{yx}P_{yy} & \sqrt{3}P_{yx}P_{yz} & \sqrt{3}P_{yy}P_{yz} \\ P_{zx}^2 & P_{zy}^2 & P_{zz}^2 & \sqrt{3}P_{zx}P_{zy} & \sqrt{3}P_{zx}P_{zz} & \sqrt{3}P_{zy}P_{zz} \\ \frac{2}{\sqrt{3}}P_{xx}P_{yx} & \frac{2}{\sqrt{3}}P_{xy}P_{yy} & \frac{2}{\sqrt{3}}P_{xz}P_{yz} & P_{xx}P_{yy}+P_{xy}P_{yx} & P_{xx}P_{yz}+P_{xz}P_{yx} & P_{xy}P_{yz}+P_{xz}P_{yy} \\ \frac{2}{\sqrt{3}}P_{xx}P_{zx} & \frac{2}{\sqrt{3}}P_{xy}P_{zy} & \frac{2}{\sqrt{3}}P_{xz}P_{zz} & P_{xx}P_{zy}+P_{xy}P_{zx} & P_{xx}P_{zz}+P_{xz}P_{zx} & P_{xy}P_{zz}+P_{xz}P_{zy} \\ \frac{2}{\sqrt{3}}P_{yx}P_{zx} & \frac{2}{\sqrt{3}}P_{yy}P_{zy} & \frac{2}{\sqrt{3}}P_{yz}P_{zz} & P_{yx}P_{zy}+P_{yy}P_{zx} & P_{yx}P_{zz}+P_{yz}P_{zx} & P_{yy}P_{zz}+P_{yz}P_{zy} \end{pmatrix} \quad (22)$$

Now, exploiting the rotation matrices (**S**, **P**, **D**, **F**, *etc.*) for the different types of basis functions (*s*, *p*, *d*, *f*, *etc.*), assembling the global rotation matrix **R** for the ELMO coefficients is straightforward. In fact, since during a rotation only the coefficients corresponding to basis functions on the same atom and of the same type (same angular momentum and same principle quantum number) are combined among each other, **R** has a block diagonal structure, with each block being the rotation matrix for a particular subset of basis functions (see Figure 3).

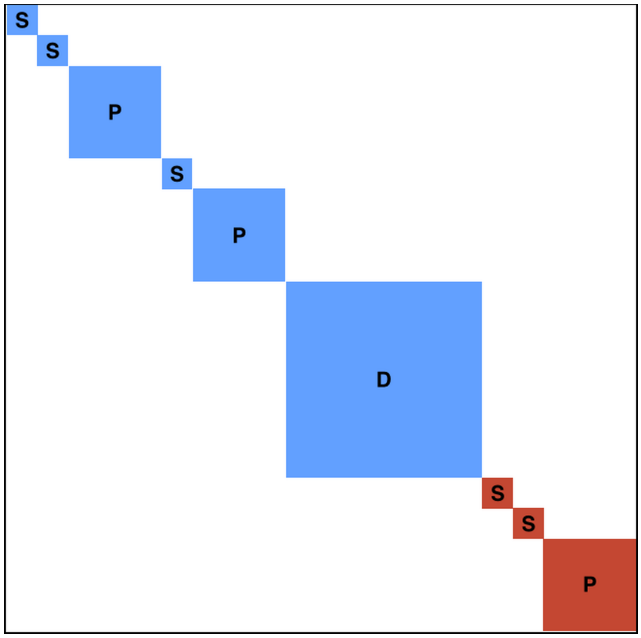


Figure 3. Block-diagonal structure of the rotation matrix **R** for an ELMO localized on a C-H bond when the split-valence basis-set 6-31G(d,p) is used. The blue blocks transform the carbon atom basis functions (and the related coefficients), while the red ones rotate the atomic orbitals (and the corresponding coefficients) of the hydrogen atom. The **S**, **P** and **D** blocks are 1×1, 3×3 and 6×6 sub-matrices, respectively, while the overall rotation matrix **R** has dimension 20×20.

Finally, it is worth noting that a transformation matrix $\mathbf{R}^{i\alpha}$ is constructed for each occupied ELMO $|\varphi_{i\alpha}\rangle$. All these matrices are afterwards assembled in a rank-three

tensor that is applied to the coefficients matrix **C** of the starting ELMOs (namely, the ones corresponding to the model molecules) to obtain the coefficients $\mathbf{C}'$ of the transferred (rotated) ELMOs:

$$\mathbf{C}' = \mathfrak{R}\,\mathbf{C} \qquad (23)$$

3. General strategy and computational details

General strategy. To assess the feasibility of constructing ELMOs libraries, we have performed our calculations on a crystallographic structure of the Leu-enkephalin pentapeptide. At first, we have investigated the ELMOs transferability and the accuracy of the model molecules approximations (see subsection below) used for the computation of the ELMOs to be included in the databases. In particular, to accomplish these tasks, we have compared the wave function and the electron density resulting from a preliminary variational-ELMO calculation (i.e., an ELMO calculation with Extremely Localized Molecular Orbitals directly determined on the molecule in exam) to those obtained by means of transferred-ELMO calculations (i.e., ELMO calculations with Extremely Localized Molecular Orbitals transferred from proper model systems) based on the three different model molecules approximations taken into account.

At a later stage, we have also studied the effects of the Molecular Orbitals localization on the reconstruction of the electron density. To achieve this goal, considering different basis-sets, we have determined charge distributions through the simple transfer of ELMOs from model molecules properly designed according to the Nearest Functional Group Approximation (see one of the next subsections for further details). These electron densities have been afterwards compared to those obtained from corresponding calculations carried out at the Hartree-Fock and B3LYP levels. Since

different basis-sets have been taken into account, all the performed computations have also allowed us to investigate the effects of using different sets of basis functions when the electron distributions are determined through the transfer of strictly localized Molecular Orbitals.

In the next subsections, more details both about the investigated system and about the computational strategies used to achieve our predetermined goals will be given.

Description of the target system. All our calculations have been performed on the Leu-enkephalin pentapeptide (Tyr-Gly-Gly-Phe-Leu). In particular, we have considered a crystallographic structure that was determined at 100 K with a resolution of 1.15 $Å^{-1}$ and that also comprises three water molecules involved in distinct intermolecular interactions with the polypeptide[105]. The positions of the hydrogen atoms, including those of the three water molecules, have been obtained on the basis of experimental electron density peaks, and then refined against experimental data. For our study, these positions were afterward modified to account for the positional bias due to the X-ray diffraction method, which actually locates hydrogen atoms bonding electrons instead of hydrogen atoms nuclei. For this reason, X-H bond lengths were elongated to fit average values observed in neutron diffraction experiments.[106]

Model molecules approximations. As already mentioned, the preliminary step for the construction of ELMOs libraries consists in performing ELMO calculations on suitable model molecules. The choice of these model systems obviously has a crucial influence on the quality of the databases and of the target molecule description. Therefore, a detailed analysis of this issue has been conducted.

For this purpose, to create the desired model molecules we have considered three levels of approximations that are schematically represented in Figure 4: a) the *Nearest Atom Approximation* (NAA), according to which the model system is simply obtained capping the fragment of interest with hydrogen atoms (see Figure 4B); b) the *Nearest Bond Approximation* (NBA), for which the model molecule is constructed considering the subunit in exam and its nearest neighbor bonds properly capped with hydrogen atoms (see Figure 4C); c) the *Nearest Functional Group Approximation* (NFGA), which consists in constructing the model system taking into account the fragment of interest and its nearest neighbor functional groups properly capped with hydrogen atoms (see Figure 4D). This scheme has been also extended to the cases in which the subunit of interest is involved in an intermolecular interaction (e.g., a hydrogen-bond), as illustrated in Figure 5.

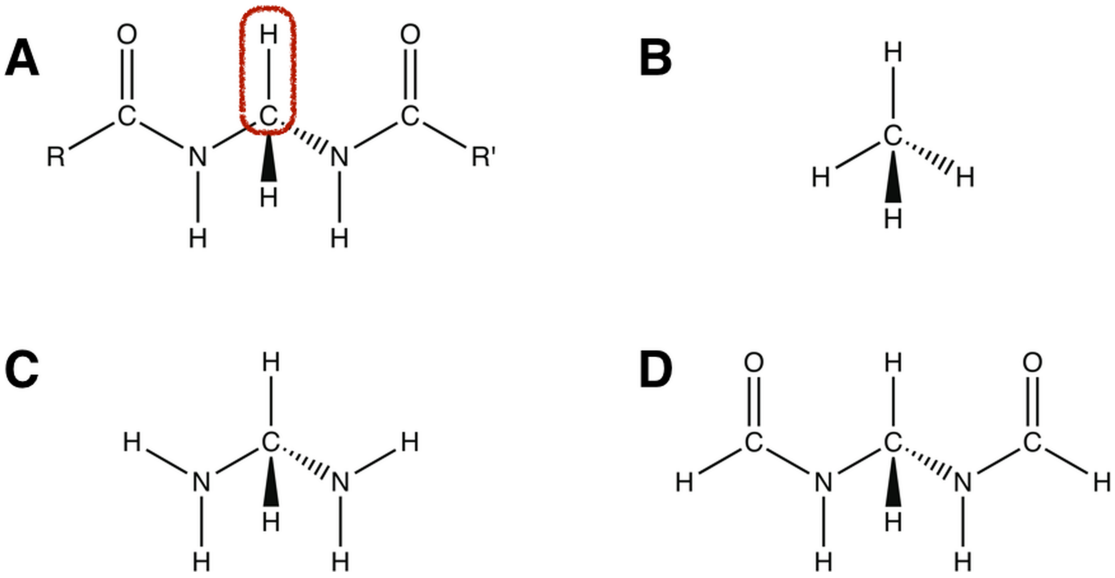


Figure 4. Model molecules approximations: A) target molecule with subunit in exam framed in red; B) Nearest Atom Approximation; C) Nearest Bond Approximation; D) Nearest Functional Group Approximation.

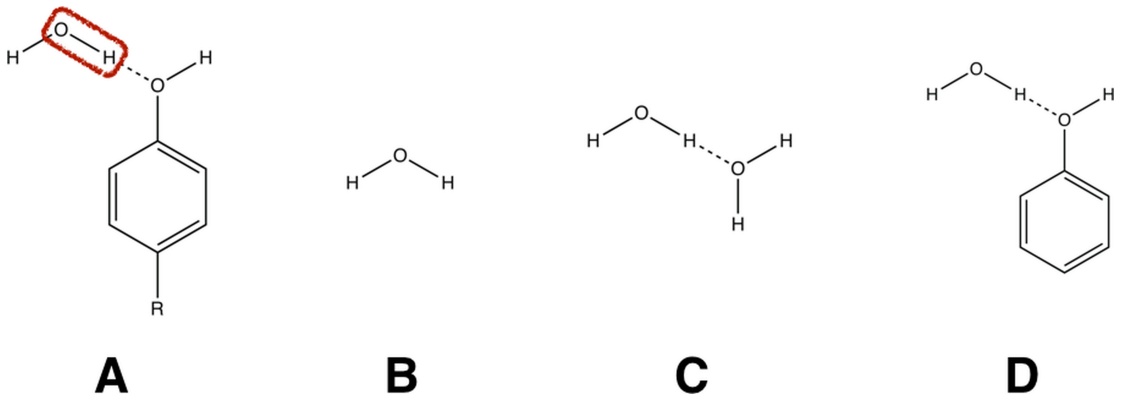


Figure 5. Extension of the model molecules approximations to the case of subunits involved in intermolecular interactions: A) target molecule with the fragment of interest framed in red; B) Nearest Atom Approximation; C) Nearest Bond Approximation; D) Nearest Functional Group Approximation

Computational methods. In the first part of our investigations, in order to assess both the ELMOs transferability and the accuracy of the different model molecules approximations, the variational- and the transferred-ELMO calculations have been carried out using the 6-31G basis-set (436 basis functions for the system in exam). The choice of this small basis-set in the first part of our study is due to the current limitations of the available ELMO program, which was originally conceived to perform variational-ELMO computations on small molecules in order to obtain Extremely Localized Molecular Orbitals to be exported to larger systems.

Larger and more flexible basis-sets (6-311G, 6-31G(d,p) and 6-311G(d,p)) have been used to study the effects of the Molecular Orbitals localization on the reconstruction of the electron density, when charge distributions obtained by transferring ELMOs determined on properly designed model molecules (see the Nearest Functional Group Approximation) have been compared to those resulting from calculations carried out at Hartree-Fock and B3LYP levels. Furthermore, given the nature of the system taken into account, we have decided to perform these preliminary test calculations without

introducing diffuse basis functions since we expect that their use does not provide results sensitively different from those obtained with the above mentioned basis-sets.

Concerning each approximation described in the previous subsection, we have constructed all the necessary model molecules for the calculation of the ELMOs to be transferred to the Leu-enkephalin polypeptide. In particular, within the NAA, NBA and NFGA, we have considered 23, 31 and 25 model systems, respectively. For all these model molecules a geometry optimization at the B3LYP/6-311G(d,p) level has been performed and, afterwards, using the optimized geometries, the sets of ELMOs for the transfer have been determined carrying out variational-ELMO calculations with the basis-sets 6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p).

It is worth noting that, for all the variational- and transferred-ELMO computations performed in our studies, we have always adopted a localization scheme almost corresponding to the Lewis structure of the molecules. Hence, we have defined both atomic fragments for the description of core and lone-pair electrons associated with each atom, and diatomic subunits for the treatment of the bond electron pairs. The only exceptions are represented by the following three-atom fragments which are necessary to take into account the two resonant Lewis structures of the system: a) the subunit O-C-O that describes the σ and π electrons of the carboxylate group (see Figure 6A); b) the subunits N-C-O to treat the eight electrons of the peptide groups, that is, the two σ bonds, the CO π bond and the delocalized lone-pair of the nitrogen atom (see Figure 6B); c) the subunits C-C-C to describe each of the delocalized π electron-pairs of the phenyl rings (see Figure 6C).

Finally, all the ELMO calculations have been performed using the version 8 of the GAMESS-UK quantum chemistry package[107] that has been properly modified to introduce the ELMO technique introduced by Stoll, while all the other computations

(Hartree-Fock and DFT) have been carried out by means of the Gaussian09[108] suite of programs. Finally, the ELMOs have been transferred exploiting an in-house code that implements the rotation strategy proposed by Philipp & Friesner and previously described in the Theory section.

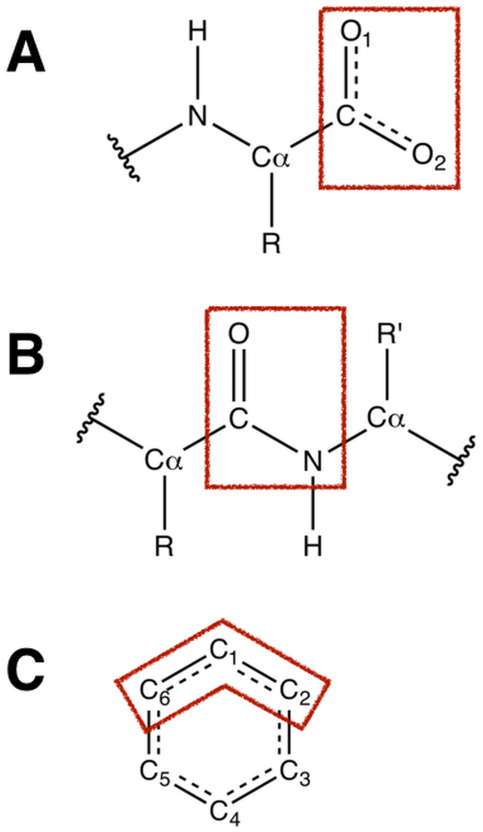


Figure 6: Three-atom fragments used to describe (A) the σ and π electrons of the carboxylate group, (B) the electrons involved in the peptide bonds and (C) the delocalized π electron pairs of the phenyl rings.

Comparison of the Electron Densities. A fundamental tool to achieve most of the goals of the present paper is the critical comparison of the electron densities obtained by means of different methods and approximations.

This has been done directly comparing the topological properties resulting from QTAIM analyses[41], which in our case have been performed using the AIMAll

package[109] (version 13.11.04). In particular, we have focused on the values of the electron density and of its Laplacian at the bond critical points, from now on also indicated as $\rho(\boldsymbol{r}_b)$ and $\nabla^2\rho(\boldsymbol{r}_b)$, respectively. Furthermore, we have compared the net atomic charges obtained through the integration of the different electron distributions over the QTAIM atomic basins.

In this context, it is important to observe that, since our target polypeptide molecule is characterized by a large number of atoms and bond critical points. Therefore, in the next section, in order to simplify the analyses and discussions of the obtained results, the comparisons of the QTAIM properties will be presented in terms of average values. In particular, for $\rho(\boldsymbol{r}_b)$ and $\nabla^2\rho(\boldsymbol{r}_b)$ we will consider their Mean Absolute Relative Variations (MARVs) with respect to the reference data, namely

$$MARV(X) = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{X_i - X_{i,ref}}{X_{i,ref}}\right| \qquad (24)$$

with N as the number of data taken into account. For the net integrated atomic charges, since small charges may lead to very large relative variations, we will use the Mean Absolute Deviations (MADs)

$$MAD(Q) = \frac{1}{N}\sum_{i=1}^{N}\left|Q_i - Q_{i,ref}\right| \qquad (25)$$

Finally, in order to obtain more global comparisons between electron densities, two real-space similarity indexes have been also computed: the Real-Space R value[110] (RSR) and the Walker-Mezey index[43] $L(a, a')$. While the former is simply given by

$$RSR\left(\rho_x, \rho_y\right) = 100\,\frac{\sum_{i=1}^{n_p}\left|\rho_x(\boldsymbol{r}_i) - \rho_y(\boldsymbol{r}_i)\right|}{\sum_{i=1}^{n_p}\left|\rho_x(\boldsymbol{r}_i) + \rho_y(\boldsymbol{r}_i)\right|} \qquad (26),$$

where n_p is the number of grid points, and the complete similarity is obtained when RSR=0, the latter allows a point-by-point comparison of two charge distributions within sets of points defined by the density shells

$$S(\rho_x, a, a') = \{\boldsymbol{r}: a \leq \rho_x \leq a'\} \quad (27)$$

and

$$S\left(\rho_y, a, a'\right) = \left\{\boldsymbol{r}: a \leq \rho_y \leq a'\right\} \quad (28)$$

In particular, following Walker and Mezey[43], the similarity index is defined as

$$L(a, a') = 100\ \frac{L^*\left(\rho_x, \rho_y, a, a'\right) + L^*\left(\rho_y, \rho_x, a, a'\right)}{2} \quad (29)$$

where

$$L^*\left(\rho_x, \rho_y, a, a'\right) = 1 - \left[\sum_{\boldsymbol{r} \in S(\rho_x, a, a')} \frac{\left|\rho_x(\boldsymbol{r}) - \rho_y(\boldsymbol{r})\right|}{max\left(\rho_x(\boldsymbol{r}), \rho_y(\boldsymbol{r})\right)}\right] \Big/ n\left(S(\rho_x, a, a')\right) \quad (30)$$

and

$$L^*\left(\rho_y, \rho_x, a, a'\right) = 1 - \left[\sum_{\boldsymbol{r} \in S\left(\rho_y, a, a'\right)} \frac{\left|\rho_x(\boldsymbol{r}) - \rho_y(\boldsymbol{r})\right|}{max\left(\rho_x(\boldsymbol{r}), \rho_y(\boldsymbol{r})\right)}\right] \Big/ n\left(S\left(\rho_y, a, a'\right)\right) \quad (31)$$

with $n\left(S(\rho_x, a, a')\right)$ and $n\left(S\left(\rho_y, a, a'\right)\right)$ as the number of grid points belonging to the density shells $S(\rho_x, a, a')$ and $S\left(\rho_y, a, a'\right)$, respectively. Unlike the RSR value, the complete similarity is obtained when $L(a, a')$ is equal to 100.

For the sake of completeness, all the similarity indexes have been evaluated considering three-dimensional grids with a 0.083131 bohr step-size for each direction. All of them have been obtained using the *Cubegen* utility of the Gaussian09 package.[108]

4. Results and Discussion

Accuracy of the model molecules approximations. As already mentioned in the previous section, in order to evaluate the accuracy of the three considered model molecules approximations, the results obtained by means of a fully variational-ELMO calculation have been compared to the ones resulting from transferred-ELMO computations associated with the NAA, NBA and NFGA.

At first, using the variational-ELMO energy as a reference, we have evaluated the energy increase associated with the use of wave functions obtained through the transfer of ELMOs from the constructed model molecules to the target system. In Table 1 we can easily observe that, when all the occupied ELMOs are simultaneously transferred to the Leu-enkephalin polypeptide, the most sophisticated approximation (NFGA) entails, by far, the lowest energy increase (see the ΔE_{Full} values in Table 1), which is indeed more than halved compared to the NAA case.

Furthermore, in order to investigate more in detail the reliability of the three approximations, we have also decided to estimate the energetic effects associated with the transfer of each single ELMO. To accomplish this task we froze all the variationally determined Extremely Localized Molecular Orbitals, except the one in exam that was substituted by an ELMO properly transferred from the desired model molecule. Therefore, computing the energy associated with these new wave functions, for each model molecules approximation we have been able to roughly evaluate the energy increases ΔE corresponding to the different single-ELMO transfers. Due to the large number of occupied Molecular Orbitals (163), in Table 1, for the sake of clarity, we have reported only average values $\langle \Delta E \rangle$ that have been calculated both over the complete set of ELMOs ($\langle \Delta E \rangle_{Overall}$) and over particular subsets of them, such as, for example, the ELMOs localized on atomic fragments ($\langle \Delta E \rangle_{Atoms}$) or the ELMOs

localized on bond fragments ($\langle \Delta E \rangle_{Bonds}$). The obtained results show that the NFGA globally provides the lowest energy increase per ELMO, a trend that is also confirmed considering the average ΔE values associated with the transfer of Molecular Orbitals localized only on atoms or bonds (see the $\langle \Delta E \rangle_{Atoms}$ and $\langle \Delta E \rangle_{Bonds}$ values).

Table 1. Comparison of the Nearest Atom, Nearest Bond and Nearest Functional Group Approximations: full and average energy variations (in kcal/mol) associated with the transfer of ELMOs. The variational ELMO/6-31G calculation is used as reference.

Energy Change	NAA	NBA	NFGA
ΔE_{Full}	149.58	118.41	58.52
$\langle \Delta E \rangle_{Overall}$	0.92	0.73	0.36
$\langle \Delta E \rangle_{Atoms}$	0.20	0.20	0.04
$\langle \Delta E \rangle_{Bonds}$	1.33	1.04	0.56
$\langle \Delta E \rangle_{C}$	0.02	0.02	0.03
$\langle \Delta E \rangle_{N}$	0.05	0.05	0.00
$\langle \Delta E \rangle_{O}$	0.41	0.41	0.07
$\langle \Delta E \rangle_{C-H}$	0.70	0.63	0.44
$\langle \Delta E \rangle_{C-C}$	2.55	0.95	0.61
$\langle \Delta E \rangle_{C-N}$	5.62	5.40	0.69
$\langle \Delta E \rangle_{C-O}$	0.16	0.23	0.23
$\langle \Delta E \rangle_{N-H}$	3.93	3.40	0.36
$\langle \Delta E \rangle_{O-H}$	4.96	3.03	1.77
$\langle \Delta E \rangle_{(C-C-C)_{ar}}$	0.50	0.42	0.52
$\langle \Delta E \rangle_{O-C-O}$	0.42	0.29	0.29
$\langle \Delta E \rangle_{N-C-O}$	0.84	0.93	1.07

From a more detailed analysis, we can also observe that the three different approximations are almost equivalent if we take into account carbon and nitrogen fragments, while the NFGA provides a slightly better description for the transfer of ELMOs localized only on oxygen atoms.

Concerning the bond-fragments, the Nearest Functional Group Approximation gives the lowest energy changes in almost all the cases. In particular, for the C-N and N-H bonds, the $\langle \Delta E \rangle$ value significantly decreases from 5.62 and 3.93 kcal/mol (NAA) to 0.69 and 0.36 kcal/mol (NFGA), respectively. Finally, in this context, the only situation in which the NFGA provides a larger energy variation is represented by the fragment C-O of the phenol group (tyrosine residue of the Leu-enkephalin peptide) for which the Nearest Atom Approximation is the best one. However, it is worth noting that the difference between the NFGA and the NAA values amounts only to 0.07 kcal/mol and, therefore, in this case, the three approximations can be considered practically equivalent.

To conclude our analysis of the energetic effects associated with the ELMOs transfer, in Table 1 we can easily see that, for the three-atom fragments, the NFGA provides average ΔE values that can be also quite larger than the ones corresponding to the NAA and the NBA. In particular, a non-negligible discrepancy is observed for the N-C-O fragment for which we have a difference of 0.23 kcal/mol between the NAA and the NFGA. Our investigations have revealed that this fact is actually related to the greater complexity of the NFGA model molecules that, although successful and useful in most of the situations, may sometimes be characterized by non-covalent interactions (e.g., intra- or inter-molecular interactions) that are not present in the target system. Therefore, in those situations, undesired interactions artificially perturb the ELMOs to be transferred and, consequently, the NAA and the NBA become superior to the NFGA, although less sophisticated chemical environments of the fragments of interest are taken into account.

At a second stage, the accuracy of the model molecules approximations has been evaluated comparing the topological properties of the obtained charge distributions.

Considering the electron density at the bond critical points, the three approximations reproduce quite well the reference values associated with the variational ELMO/6-31G wave function, with the NFGA slightly better than the other ones (see Table 2).

This is confirmed by the more detailed inspection of the mean absolute relative variations (MARVs) for the different types of bond critical points (BCPs). In fact, the largest MARV for the NAA, NBA and NFGA amounts only to 7.35%, 7.17% and 2.29%, respectively, and, furthermore, in most of the situations, the Nearest Functional Group Approximation provides the lowest discrepancies. In particular, it is easy to see that, for the $(C\text{-}N)_{peptide}$, $(C\text{-}N)_{term}$ and $C_{\alpha}\text{-}N$ bonds, the average relative variations significantly decrease from 4.99%, 4.02% and 7.35% (NAA) to 1.31%, 1.42% and 2.29% (NFGA), respectively.

In only one case the NFGA gives a quite significantly larger discrepancy compared to the NAA ($(C\text{-}O)_{phenol}$ bond critical point). This can be probably ascribed to the inadequacy of the NFGA model molecule that is not able to properly mimic the network of hydrogen-bond interactions in which the subunit C-O is directly or indirectly involved.

Concerning the Laplacian of the electron density at the BCPs, as expected, larger relative variations have been observed (see Table 2). Nevertheless, considering the extreme sensitivity of this topological property, the obtained MARVs are actually relatively small, which is a further evidence of the ELMOs transferability. Furthermore, also in this case the Nearest Functional Group Approximation generally gives the best agreements with the reference variational values. This is particularly evident for the $(C\text{-}N)_{peptide}$ bond critical points, for which the mean absolute relative variation significantly drops from 32.21% and 28.22% (NAA and NBA, respectively) to 5.28% (NFGA).

Table 2. Comparison of the Nearest Atom, Nearest Bond and Nearest Functional Group Approximations: Mean Absolute Relative Variations (in percentage) of the values of the electron density and of its Laplacian at the bond critical points. The variational ELMO/6-31G values are used as reference.

	ρ_{BCP}			$\nabla^2\rho_{BCP}$		
Bond	NAA	NBA	NFGA	NAA	NBA	NFGA
Overall	1.67	1.62	1.18	5.34	5.14	2.99
C-H	1.32	1.12	0.93	4.29	3.84	3.36
C-C	0.83	0.61	0.43	1.03	1.20	0.89
$(C\text{-}C)_{ar}$	0.31	0.30	0.28	0.77	0.77	0.91
$(C\text{-}N)_{peptide}$	4.99	4.40	1.31	32.21	28.22	5.28
$(C\text{-}N)_{term}$	4.02	6.00	1.42	28.43	65.43	14.23
Cα-N	7.35	7.17	2.29	21.95	21.36	5.97
$(C\text{-}O)_{term}$	0.34	0.61	0.26	2.76	3.50	2.39
$(C\text{-}O)_{peptide}$	0.66	0.13	0.31	17.08	9.40	8.16
$(C\text{-}O)_{phenol}$	0.73	2.18	2.06	18.37	25.04	16.99
N-H	0.62	0.93	0.96	2.38	2.36	0.87
O-H	1.78	1.36	0.96	0.78	0.61	0.53

To complete the comparison of the topological properties, the net integrated atomic charges have been also taken into account. In Table 3 we can easily note that, for each atom-type, all the three approximations provide charges that are very close to those resulting from the variational ELMO/6-31G calculation. Moreover, once again, the NFGA gives the best agreements, with the only exception observed for the nitrogen atom for which the NFGA Mean Absolute Deviation (MAD) is smaller than the one associated with the NBA, but larger than the one corresponding to the NAA.

Table 3. Comparison of the Nearest Atom, Nearest Bond and Nearest Functional Group Approximations: Mean Absolute Deviations (in e) of the net integrated atomic charges. The variational ELMO/6-31G values are used as reference.

Atom	NAA	NBA	NFGA
Overall	0.056	0.048	0.037
C	0.075	0.056	0.043
N	0.051	0.061	0.053
O	0.042	0.039	0.035
H	0.047	0.043	0.032

Finally, as already mentioned in the section dedicated to the computational details, in order to obtain a more global quantity to evaluate the accuracy of the considered model molecules approximations, proper point-by-point-similarity indexes have been used to compare the NAA, NBA and NFGA electron densities to the reference charge distribution. In Table 4 the obtained values show that, despite a great similarity among all the considered charge densities, the Nearest Functional Groups Approximation allowed the reconstruction of an electron distribution that is globally more similar to the variational one (see, in particular, the *RSR* and the $L(0.001,10)$ indexes). Moreover, since the Walker-Mezey index enables to investigate the similarity in different electron density ranges by changing the a and a' values (see Equations (27)-(31)), we can also observe that, as expected, the greater global accuracy of the NFGA is more evident in regions far from the nuclei (see the index $L(0.001,0.01)$) than in regions close to the nuclei (index *L(0.1,10)*).

Summarizing, the results shown in this subsection have further confirmed the ELMOs transferability. This has been particularly evinced from the great similarity of the NAA, NBA and NFGA electron densities with the reference (namely, variational)

charge distribution, great similarity that has been found both in the quite small relative (and absolute) deviations of the analyzed topological properties (see Tables 2 and 3) and in the very small/large values of the RSR/Walker-Mezey indexes (see Table 4).

Table 4. Comparison of the Nearest Atom, Nearest Bond and Nearest Functional Group Approximations: values of the Real-Space R and Walker-Mezey similarity indexes (in percentage) with the variational ELMO/6-31G charge distribution used as reference. For the Walker-Mezey indicator, the electron densities are compared within the a and a' limits expressed in e/bohr^3.

Similarity Index	NAA	NBA	NFGA
RSR	0.74	0.70	0.51
L(0.001,10)	96.15	96.34	97.21
L(0.001,0.01)	94.60	94.85	96.01
L(0.01,0.1)	97.12	97.28	98.01
L(0.1,10)	98.55	98.64	98.95

Of course, all the collected results have also importantly shown that, among the three considered model molecules approximations, the NFGA is the one that globally allows the best reproduction of the reference electron distribution. Nevertheless, in some circumstances, the Nearest Functional Group Approximation might not be optimal. In fact, some of the sophisticated NFGA model molecules might be sometimes characterized by undesired non-covalent intra-molecular interactions that do not allow us to properly reproduce the real environment of the fragment of interest in the target system. However, it is worth pointing out that, in those situations, the NFGA is only slightly less accurate than the other two approximations, while, on the contrary, in almost all the situations in which it perfectly works, it represents a significant improvement compared to the NAA and the NBA.

Effects of the Molecular Orbitals localization. To assess the level of approximation introduced by using Extremely Localized Molecular Orbitals in the reconstruction of electron densities, ELMO-NFGA electron distributions (namely, electron distributions corresponding to wave functions obtained transferring ELMOs from the NFGA model molecules) have been compared to corresponding Hartree-Fock (HF) charge densities considering four different basis-sets (6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p)). Furthermore, to have a reference, comparisons between B3LYP and HF electron densities have been also taken into account.

As in the previous subsection, at first, we focused on the main topological properties associated with the obtained charge densities. In Table 5, we have reported the mean absolute relative variations (with respect to the Hartree-Fock references) obtained at the ELMO-NFGA and B3LYP levels for the values of the electron density at the bond critical points. We can easily observe that, overall and for all the basis-sets, the transfers of ELMOs provide quite small discrepancies with respect to the Hartree-Fock computations. Furthermore, these discrepancies are of the same order of magnitude of the ones arising from the DFT calculations. The previous observations are confirmed if we consider the MARVs computed for specific types of bond critical points. In fact, the largest discrepancies at the ELMO-NFGA level have been obtained for the C_α-N bonds, with mean absolute relative variations only in the amount of 6.11%, 6.25%, 7.02% and 6.47% for the basis-sets 6-31G, 6-311G, 6-31G(d,p) and 6-311G(d,p), respectively. In analogous way, also the DFT calculations provide the largest MARVs for the C_α-N bonds, but slightly smaller than the ones associated with the Nearest Functional Group Approximation.

Table 5. Effects of the Molecular Orbitals Localization: Mean Absolute Relative Variations (in percentage) of the values of the electron density at the bond critical points. Hartree-Fock values are used as reference.

Bond	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
Overall	1.83	1.34	1.90	1.33	2.16	2.76	1.99	2.74
C-H	1.08	0.59	1.19	0.38	1.20	2.09	1.12	2.21
C-C	0.42	0.51	1.68	2.21	1.57	4.08	2.01	5.21
$(C\text{-}C)_{ar}$	0.74	0.22	1.08	1.24	0.83	2.86	0.53	3.24
$(C\text{-}N)_{peptide}$	1.37	0.34	1.33	0.86	1.55	1.77	1.55	1.62
$(C\text{-}N)_{term}$	2.08	4.01	2.25	3.26	2.19	1.65	1.83	2.13
Cα-N	6.11	4.68	6.25	4.63	7.02	5.45	6.47	5.07
$(C\text{-}O)_{term}$	1.78	1.19	1.68	1.00	2.04	1.76	2.13	1.25
$(C\text{-}O)_{peptide}$	0.95	1.15	0.94	0.90	1.16	1.79	1.24	1.16
$(C\text{-}O)_{phenol}$	0.89	0.81	0.97	1.05	0.58	0.19	0.45	0.31
N-H	1.07	0.80	0.63	0.17	1.07	1.64	0.86	1.63
O-H	1.24	0.92	0.69	0.51	1.11	1.06	1.54	0.44

Concerning the values of the Laplacian at the bond critical points, as expected, we have obtained larger discrepancies as in the previous subsection (see Table 6). However, at least for the basis-sets without polarization functions (i.e., 6-31G and 6-311G), the ELMO-NFGA and B3LYP methods provide absolute relative variations of the same order of magnitude, both globally and for almost all the types of bond critical points. Unfortunately, for the 6-31G(d,p) and 6-311G(d,p) basis-sets, the trends are much less clear. In fact, while in some cases the transfer of ELMOs gives much larger deviations than the DFT method (e.g., $(C\text{-}O)_{term}$ BCP for the 6-31G(d,p)

basis-set), in other circumstances, the exactly opposite behavior is observed (e.g., $(C\text{-}N)_{term}$ BCP for the 6-311G(d,p) basis-set).

Table 6. Effects of the Molecular Orbitals Localization: Mean Absolute Relative Variations (in percentage) of the values of the Laplacian of the electron density at the bond critical points. Hartree-Fock values are used as reference.

Bond	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
Overall	6.44	5.42	7.62	8.05	14.90[(a)]	12.19[(a)]	10.01[(b)]	16.28[(b)]
C-H	3.18	1.31	2.33	2.87	2.86	7.75	2.00	9.02
C-C	4.85	4.28	11.17	14.51	7.95	17.94	9.58	21.27
$(C\text{-}C)_{ar}$	4.26	2.58	5.13	8.33	4.62	13.11	3.12	13.79
$(C\text{-}N)_{peptide}$	5.81	8.77	6.16	11.24	17.49	10.14	15.85	7.39
$(C\text{-}N)_{term}$	32.78	26.86	39.86	34.55	35.92	70.94	39.19	155.04
Cα-N	29.63	16.54	30.16	11.17	45.47	49.20	43.43	61.44
$(C\text{-}O)_{term}$	26.59	25.60	21.84	7.51	232.43	15.34	77.76	32.57
$(C\text{-}O)_{peptide}$	14.76	29.35	15.39	5.09	108.58[(c)]	10.83[(c)]	53.36	44.54
$(C\text{-}O)_{phenol}$	58.69	10.33	86.32	93.98	//	//	//	//
N-H	0.70	0.67	3.25	6.46	2.18	6.10	3.41	12.44
O-H	0.51	0.51	3.21	8.34	1.64	8.31	1.83	6.89

(a) The values related to the C48-O39 and C12-O5 BCPs have not been considered; (b) the value related to the C12-O5 BCP has not been considered; (c) the value related to the C48-O39 BCP has not been considered.

Furthermore, as also pointed out in Table 6, it is worth noting that two bond critical points have been excluded from the computation of the mean average relative variations due to their surprisingly too small $\nabla^2\rho(\boldsymbol{r}_b)$ values obtained at Hartree-Fock level. They are the C48-O39 BCP (only for the 6-31G(d,p) basis-set), which

corresponds to the carbonyl group of a peptide bond (for the labels specification see the Leu-enkephalin target geometry deposited in the Supporting Information), and the C12-O5 BCP (for the basis-sets 6-31G(d,p) and 6-311G(d,p)), which is associated with the C-O bond of the phenol group in the tyrosine residue. For the former, the Laplacian values at HF and B3LYP levels are too small for a covalent bond critical point (-0.0036 e/Å^5 and -0.3597 e/Å^5, respectively), while the one obtained transferring ELMOs is completely reasonable (-3.8758 e/Å^5). In the latter case, only the Hartree-Fock method has provided too small Laplacian values (-0.3671 e/Å^5 and 0.0180 e/Å^5 for the 6-31G(d,p) and the 6-311G(d,p) basis-sets, respectively), with one of them even positive, which is not acceptable for a bond interaction that should be covalent. On the contrary, the DFT and ELMO values are correctly negative both at the 6-31G(d,p) (-4.8117 e/Å^5 and -4.6644 e/Å^5, respectively) and at the 6-311G(d,p) level (-6.7886 e/Å^5 and -4.4229 e/Å^5).

To complete the analysis of the main topological properties, the net integrated atomic charges associated with the considered electron distributions have been also compared. In Table 7, for each basis-set the ELMO-NFGA charges are generally very close to the Hartree-Fock ones, with the mean absolute deviations (MADs) obtained at the ELMO level always lower than 0.1 e. On the contrary, the DFT charges show more significant variations, especially for the nitrogen and the oxygen atoms, for which the largest MADs amount to 0.32 e and 0.19 e (6-311G(d,p) basis-set).

For the sake of completeness, the simple Mean Relative Variations (with respect to the Hartree-Fock references) of the $\rho(\boldsymbol{r}_b)$ and $\nabla^2\rho(\boldsymbol{r}_b)$ values have been also computed (see Tables S1 and S4 in the Supporting Information). These signed variations, along with the maximum and minimum deviations (see Tables S2, S3, S5 and S6 in the Supporting Information), have allowed to notice that, in most of the

cases, the ELMO-NFGA strategy overestimates the electron density at the BCPs, while it generally underestimates the Laplacian at the BCPs with respect to the reference Hartree-Fock values. Less clear trends can be observed from the mean, maximum and minimum deviations for the net integrated atomic charges (see Tables S7-S9 in the Supporting Information). However, the obtained results confirm that the ELMO-NFGA method provides charges that are very close to the Hartree-Fock ones.

Table 7. Effects of the Molecular Orbitals Localization: Mean Absolute Deviations (in e) of the net integrated atomic charges. Hartree-Fock values are used as reference.

Atom	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
Overall	0.041	0.051	0.042	0.065	0.046	0.102	0.050	0.098
C	0.054	0.063	0.055	0.079	0.064	0.121	0.077	0.121
N	0.054	0.134	0.056	0.200	0.046	0.282	0.048	0.319
O	0.045	0.106	0.043	0.140	0.046	0.163	0.039	0.185
H	0.031	0.020	0.031	0.023	0.034	0.039	0.035	0.036

In order to have more global comparisons, similarity indexes have been computed also in this situation. The obtained results have been collected in Table 8 and they clearly show that the degree of similarity between the transferred-ELMO electron distributions and the corresponding Hartree-Fock ones is quite high. Moreover, considering all the $L(a, a')$ values, we can observe that the ELMO/HF similarity always increases closer to the nuclei, even if it is worth noting that, since the Walker-Mezey index measures the similarity only in a percentage sense, large absolute differences are anyway probably present in the core regions, which are indeed characterized by very high electron densities. Finally, from Table 8 it is also possible to note that, for each basis-set and for each selected range of electron density values

(see the different $L(a, a')$ values in Table 8), the DFT/HF degree of similarity is completely comparable to the ELMO/HF one.

Table 8. Effects of the Molecular Orbitals Localization: values of the Real-Space R and Walker-Mezey similarity indexes (in percentage). For the Walker-Mezey indicator, the electron density is compared within the a and a' expressed in e/bohr^3. Hartree-Fock densities are used as reference.

Similarity Index	6-31G		6-311G		6-31G(d,p)		6-311G(d,p)	
	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP	NFGA	B3LYP
RSR	0.69	0.74	0.71	0.91	0.75	0.90	0.81	1.02
L(0.001, 10)	96.09	95.89	95.78	96.17	96.02	96.02	95.47	96.22
L(0.001, 0.01)	94.34	94.20	93.85	94.67	94.49	94.60	93.73	94.90
L(0.01, 0.1)	97.24	96.95	97.09	97.43	97.01	97.06	96.59	97.49
L(0.1, 10)	98.65	98.65	98.68	98.03	98.46	98.03	98.42	97.54

To summarize, almost all the results shown in this subsection mainly reveal that charge distributions obtained by transferring Extremely Localized Molecular Orbitals from model molecules are in agreement with the corresponding Hartree-Fock electron densities and that the extent of the observed discrepancies is generally comparable to the extent of the differences detected in the DFT/HF comparisons. Significant deviations have been observed only for the Laplacian of the electron density at the bond critical points (especially for polarized basis-sets), but this might be also due to the extreme sensitivity of the topological property in exam.

Effects of the basis-sets. The effects of reconstructing an electron density through Extremely Localized Molecular Orbitals expanded in different basis-sets have been also investigated. As in the previous subsections, this task has been accomplished comparing the main topological properties associated with the examined electron

distributions. However, unlike the previous studies, where mean absolute relative variations and mean absolute deviations have been considered, in this case the Mean Deviations (MDs) of the considered properties will be used:

$$MD(X) = \frac{1}{N}\sum_{i=1}^{N} X_i - X_{i,ref} \qquad (32),$$

Concerning the $\rho(\boldsymbol{r}_b)$ and $\nabla^2\rho(\boldsymbol{r}_b)$ values, all the results are collected in Table 9, where, first of all, we can easily see that the 6-311G and 6-31G electron distributions are practically equivalent. Larger deviations have been instead observed for the other two cases, in which the effects of introducing polarization functions in the calculations have been investigated. In particular, the comparison of the 6-31G(d,p) and 6-31G charge distributions and the one of the 6-311G(d,p) and 6-311G electron densities have been examined and the results show that the consequences of adding polarization function to a double-zeta basis-set are roughly equivalent to the effects of adding the same type of basis functions to a triple-zeta basis-set.

Similar conclusions can be drawn considering the net integrated atomic charges (see Table 10). In fact, while only very small changes have been obtained for the 6-311G/6-31G comparison, more significant and similar variations have been observed when comparing polarized to un-polarized basis-sets.

Finally, it is worth noting that the same trends detected at the ELMO-NFGA level have been also noticed when the basis-set effects have been investigated for the Hartree-Fock and B3LYP methods (see Tables 11-14). In particular, the mean deviations computed for these strategies are quantitatively comparable to the ELMO-NFGA ones (see Tables 9-10), especially at Hartree-Fock level, for which the agreement with the ELMO results is significantly high.

Table 9. Basis-set effect: mean deviations of the values of the electron density (in e/Å^3) and of its Laplacian (in e/Å^5) at the bond critical points when ELMO-NFGA calculations are performed with different basis-sets.

Bond	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	0.01	-0.74	0.19	-8.23	0.15	-6.23
C-C	0.04	-1.58	0.21	-7.55	0.17	-6.35
$(C-C)_{ar}$	0.01	-1.62	0.19	-7.04	0.14	-5.24
$(C-N)_{peptide}$	-0.01	2.11	0.17	2.62	0.13	0.71
$(C-N)_{term}$	0.01	0.23	0.15	-1.53	0.11	1.92
Cα-N	0.01	0.50	0.17	-3.19	0.12	-0.47
$(C-O)_{term}$	0.00	0.80	0.14	11.48	0.13	6.46
$(C-O)_{peptide}$	0.00	-0.62	0.15	13.36	0.13	8.71
$(C-O)_{phenol}$	0.00	1.96	0.14	5.08	0.12	3.36
N-H	0.01	2.35	0.21	-11.81	0.17	-11.94
O-H	0.03	-8.70	0.19	-15.67	0.16	-16.38

Table 10. Basis-set effect: mean deviations of the net integrated atomic charges (in e) when ELMO-NFGA calculations are performed with different basis-sets.

Atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.020	0.164	0.135
N	0.016	-0.384	-0.312
O	0.017	-0.246	-0.203
H	0.008	-0.005	-0.006

Table 11. Basis-set effect: mean deviations of the values of the electron density (in e/Å^3) and of its Laplacian (in e/Å^5) at the bond critical points when Hartree-Fock calculations are performed with different basis-sets.

Bond	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	0.01	-0.46	0.18	-7.89	0.15	-6.20
C-C	0.01	-0.81	0.19	-6.69	0.17	-5.96
$(C-C)_{ar}$	0.01	-1.40	0.19	-6.68	0.16	-5.49
$(C-N)_{peptide}$	-0.01	1.95	0.16	2.82	0.13	0.79
$(C-N)_{term}$	0.01	0.50	0.15	-0.97	0.12	1.35
Cα-N	0.01	0.45	0.15	-1.24	0.12	0.34
$(C-O)_{term}$	0.01	0.00	0.13	13.56	0.11	9.10
$(C-O)_{peptide}$	0.00	-0.41	0.14	14.65	0.12	10.07
$(C-O)_{phenol}$	0.00	1.96	0.15	5.77	0.12	4.19
N-H	0.01	2.18	0.21	-11.02	0.16	-12.07
O-H	0.04	-10.05	0.19	-15.12	0.13	-14.08

Table 12. Basis-set effect: mean deviations of the net integrated atomic charges (in e) when Hartree-Fock calculations are performed with different basis-sets.

Atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.024	0.159	0.132
N	0.030	-0.359	-0.304
O	0.014	-0.245	-0.207
H	0.009	-0.005	-0.003

Table 13. Basis-set effect: mean deviations of the values of the electron density (in e/Å^3) and of its Laplacian (in e/Å^5) at the bond critical points when B3LYP calculations are performed with different basis-sets.

Bond	$\langle X_{6-311G} - X_{6-31G} \rangle$		$\langle X_{6-31G(d,p)} - X_{6-31G} \rangle$		$\langle X_{6-311G(d,p)} - X_{6-311G} \rangle$	
	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$	ρ_{BCP}	$\nabla^2\rho_{BCP}$
C-H	-0.01	0.30	0.13	-5.61	0.11	-4.45
C-C	-0.02	0.40	0.11	-4.03	0.11	-3.91
$(C\text{-}C)_{ar}$	-0.02	-0.31	0.12	-4.01	0.11	-3.73
$(C\text{-}N)_{peptide}$	-0.02	2.32	0.13	-1.44	0.12	-3.25
$(C\text{-}N)_{term}$	0.00	0.17	0.12	-4.56	0.10	-3.88
Cα-N	-0.01	1.06	0.13	-5.11	0.11	-4.25
$(C\text{-}O)_{term}$	0.00	-3.16	0.12	8.58	0.11	5.00
$(C\text{-}O)_{peptide}$	0.01	-4.58	0.12	9.42	0.12	5.81
$(C\text{-}O)_{phenol}$	0.01	-1.33	0.13	1.96	0.11	1.31
N-H	0.00	4.42	0.15	-8.10	0.12	-8.57
O-H	0.04	-6.85	0.15	-10.51	0.11	-12.84

Table 14. Basis-set effect: mean deviations of the net integrated atomic charges (in e) when B3LYP calculations are performed with different basis-sets.

Atom	$\langle Q_{6-311G} - Q_{6-31G} \rangle$	$\langle Q_{6-31G(d,p)} - Q_{6-31G} \rangle$	$\langle Q_{6-311G(d,p)} - Q_{6-311G} \rangle$
C	-0.041	0.101	0.090
N	0.097	-0.211	-0.185
O	0.048	-0.188	-0.162
H	0.004	0.003	0.001

5. Conclusions

In this paper, a preliminary study about the feasibility of constructing libraries of extremely localized molecular orbitals has been presented. The two main goals were the further investigation of the ELMOs transferability, which had been never examined in detail considering sufficiently large target molecules, and the definition of simple and suitable rules with which to construct model molecules for the computation of the ELMOs to be stored in the desired databanks.

For this purpose, the obtained results have confirmed that the ELMOs are indeed reliably transferable from a molecule to another. Furthermore, we have also shown that, among the three model molecules approximations devised and considered in our study, the Nearest Functional Group Approximation (NFGA) is globally and by far the best one both from the energetic point of view and in terms of electron densities reconstruction. Nevertheless, it has been noticed that the sophisticated NFGA model molecules are sometimes characterized by undesired non-covalent interactions (i.e., missing interactions in the target system) that artificially perturb the ELMOs to be transferred. Therefore, from these observations, it is evident that the Nearest Functional Group Approximation, although reliable in most of the situations, can be considered only as a zero-order approximation. For this reason, we envisage the possibility of relaxing the ELMOs immediately after the transfer through proper linear scaling techniques that suitably take into account the real chemical environment of the target system.

In the paper it has been also shown that the electron densities obtained from the transfer of ELMOs are generally very similar to the corresponding Hartree-Fock ones and it has been interestingly seen that the extent of the detected discrepancies is comparable to the size of the differences between corresponding B3LYP and Hartree-

Fock charge distributions. This evidence also encourages the construction of Kohn-Sham ELMOs libraries that might allow the reconstruction of DFT-quality electron densities. To accomplish this task, we are planning the development of a new DFT-ELMO method starting from the original ideas already introduced by Burresi and Sironi.[111]

Finally, we have observed that, independently of the method chosen to obtain the desired electron densities (e.g., transferring ELMOs from a library or performing computation at Hartree-Fock or DFT levels), changing basis-set (especially introducing polarization functions) has approximately the same effect.

All these results, along with the recent finding that the ELMOs transferability is as reliable as the one of the multipole models pseudoatoms[98], further encourage the construction of new databanks of Extremely Localized Molecular Orbitals that will hopefully provide another useful tool both to reconstruct electron densities and to refine crystallographic structures of large macromolecules.

Acknowledgments

Alessandro Genoni and Benoît Guillot gratefully thank the Institute Jean Barriol of the University of Lorraine for financial support (Axe R12 Project - Years 2013 & 2014).

Supporting Information Available

Crystal geometry of the target polypeptide Leu-enkephalin. Tables S1-S3: Mean, maximum and minimum relative variations of the electron density at the bond critical points using the Hartree-Fock values as reference. Tables S4-S6: Mean, maximum and minimum relative variations of the Laplacian of the electron density at the bond

critical points with respect to the corresponding Hartree-Fock values. Tables S7-S9: Mean, maximum and minimum deviations of the net integrated atomic charges using the Hartree-Fock values as reference. This material is available free of charge via the Internet at http://pubs.acs.org/.

References

1. Van Alsenoy, C.; Yu, C.-H.; Peeters, A.; Martin, J. M. L.; Schäfer, L. Ab Initio Geometry Determinations of Proteins.1. Crambin. *J. Phys. Chem. A* **1998**, *102*, 2246-2251.
2. Scuseria, G. E. Linear Scaling Density Functional Calculations with Gaussian Orbitals. *J. Phys. Chem. A*. **1999**, *103*, 4782-4790.
3. Sato, F.; Yoshihiro, T.; Era, M; Kashiwagi, H. Calculation of all-electron wavefunction of hemoprotein cytochrome c by density functional theory *Chem. Phys. Lett.* **2001**, *341*, 645-651.
4. Inaba, T.; Tahara, S.; Nisikawa, N.; Kashiwagi, H.; Sato, F. All-electron density functional calculation on insulin with quasi-canonical localized orbitals *J. Comput. Chem.* **2005**, *26*, 987-993.
5. Inaba, T.; Sato, F. Development of parallel density functional program using distributed matrix to calculate all-electron canonical wavefunction of large molecules. *J. Comput. Chem.* **2007**, *28*, 984-995.
6. Li, X.-P.; Nunes, R. W.; Vanderbilt, D. Density-matrix electronic-structure method with linear system-size scaling. *Phys. Rev. B* **1993**, *47*, 10891-10894.
7. Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71,* 1085-1123.
8. Yang, W. Direct calculation of electron density in density-functional theory. *Phys.*

Rev. Lett. **1991**, *66*, 1438-1441.

9. Yang, W. Direct calculation of electron density in density-functional theory: Implementation for benzene and a tetrapeptide. *Phys. Rev. A* **1991**, *44*, 7823-7826.
10. 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.
11. Dixon, S. L.; Merz, K. M., Jr. Semiempirical molecular orbital calculations with linear system size scaling. *J. Chem. Phys.* **1996**, *104*, 6643-6649.
12. Dixon, S. L.; Merz, K. M., Jr. Fast, accurate semiempirical molecular orbital calculations for macromolecules. *J. Chem. Phys.* **1997**, *107*, 879-893.
13. He, X.; Merz, K. M., Jr. Divide and Conquer Hartree-Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405-411.
14. Merz, K. M., Jr. Using Quantum Mechanical Approaches to Study Biological Systems. *Acc. Chem. Res.* **2014**, *47*, 2804-2811.
15. Gadre, S. R.; Shirsat, R. N.; Limaye, A. C. Molecular Tailoring Approach for Simulation of Electrostatic Properties. *J. Phys. Chem.* **1994**, *98*, 9165-9169.
16. 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.
17. Babu, K.; Ganesh V.; Gadre S. R.; Ghermani N. E. Tailoring approach for exploring electron densities and electrostatic potentials of molecular crystals. *Theor. Chem. Acc.* **2004**, *111*, 255-263.
18. 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.

19. 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.
20. 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.
21. 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.
22. 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.
23. He, X.; Zhang, J. Z. H. A new method for direct calculation of total energy of protein. *J. Chem. Phys.* **2005**, 122, 031103.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.

29. 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.
30. 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.
31. Huang L.; Massa, L.; Karle, J. Quantum kernels and quantum crystallography: Applications in biochemistry. In *Quantum Biochemistry: Electronic Structure and Biological Activity*; Matta, C. F. Ed.; Wiley-VCH: Weinheim, 2010; Chapter 1, pp 3-60.
32. Huang, L.; Massa, L.; Karle, J. Kernel energy method applied to vesicular stomatitis virus nucleoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1731-1736.
33. Huang, L.; Massa, L.; Karle, J. The Kernel Energy Method: Application to a tRNA. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1233-1237.
34. Huang, L.; Massa, L.; Karle, J. Kernel energy method illustrated with peptides. *Int. J. Quantum Chem.* **2005**, *103*, 808-817.
35. Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochemistry* **2005**, *44*, 16747-16752.

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

37. 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.

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

39. 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.

40. 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.

41. Bader, R. F. W. Atoms in Molecules: A Quantum Theory; Oxford University Press: Oxford, 1990.

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

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

44. Mezey, P. G. Macromolecular density matrices and electron densities with adjustable nuclear geometries. *J. Math. Chem.* **1995**, *18*, 141-168.

45. 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.

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

47. Breneman, C. M.; Thompson, T. R.; Rhem, M.; Dung, M. Electron density modeling of large systems using the transferable atom equivalent method. *Comput. Chem.* **1995**, *19*, 161-179.

48. Breneman, C. M.; Rhem, M. QSPR Analysis of HPLC column capacity factors for a set of high-energy materials using electronic van der Waals surface property descriptors computed by transferable atom equivalent method. *J. Comput. Chem.* **1997**, *18*, 182-197.

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

50. Bader, R. F. W.; Martín, F. J. Interdeterminancy of basin and surface properties of an open system. *Can. J. Chem.* **1998**, *76*, 284-291.

51. Martín, F. J. "Theoretical Synthesis of Macromolecules from Transferable Functional Groups". Ph.D. Thesis; McMaster University: Hamilton, 2001.

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

53. Pipek, J.; Mezey, P. G. A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions. *J. Chem. Phys.* **1989**, *90*, 4916–4926.

54. Boys, S. F. Construction of Some Molecular Orbitals to be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.

55. Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300–302.

56. Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–465.

57. Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. II. *J. Chem. Phys.* **1965**, *43*, S97–S116.

58. Adams, W. H. On the Solution of the Hartree-Fock Equations in Terms of Localized Orbitals *J. Chem. Phys.* **1961**, *34*, 89-102.

59. Huzinaga, S.; Cantu, A. A. Theory of Separability of Many-Electron Systems. *J. Chem. Phys.* **1971** *55*, 5543–5549.

60. Gilbert, T. L. Multiconfiguration self-consistent-field theory for localized orbitals. II. Overlap constraints, Lagrangian multipliers, and the screened interaction field *J. Chem. Phys.* **1974**, *60*, 3835-3844.

61. Matsuoka, O. Expansion methods for Adams–Gilbert equations. I. Modified Adams–Gilbert equation and common and fluctuating basis sets. *J. Chem. Phys.* **1977**, *66*, 1245-1254.

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

63. Smits, G. F.; Altona, C. Calculation and properties of non-orthogonal, strictly local molecular-orbitals. *Theor. Chim. Acta* **1985**, *67*, 461–475.

64. Francisco, E.; Martín Pendás, A.; Adams, W. H. Generalized Huzinaga Building-Block Equations for Nonorthogonal Electronic Groups – Relation to the Adams-Gilbert Theory. *J. Chem. Phys.* **1992**, *97*, 6504-6508.

65. Ordejón, P.; Drabold, D.; Grumbach, M.; Martin, R. Unconstrained Minimization Approach for Electronic Computations that Scales Linearly with System Size.

Phys. Rev. B **1993**, *48*, 14646–14649.

66. Couty, M.; Bayse, C. A.; Hall, M. B. Extremely localized molecular orbitals (ELMO): a non-orthogonal Hartree-Fock method. *Theor. Chem. Acc.* **1997**, *97*, 96–109.
67. 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.
68. Szekeres, Z.; Surján, P. R. Direct determination of fragment localized molecular orbitals and the orthogonality constraint. *Chem. Phys. Lett.* **2003**, *369*, 125–130.
69. McWeeny, R. The Density Matrix in Many-Electron Quantum Mechanics. 1. Generalized Product Functions – Factorization and Physical Interpretation of the Density Matrices. *Proc. R. Soc. London Ser. A* **1959**, *253*, 242–259.
70. McWeeny, R. Some Recent Advances in Density Matrix Theory. *Rev. Mod. Phys.* **1960**, *32*, 335–369.
71. McWeeny, R. In *Methods of Molecular Quantum Mechanics*, 2nd ed.; Academic Press: London, 1992; Chapter 14, pp 485-495.
72. Genoni, A.; Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data. *J. Phys. Chem. Lett.* **2013**, *4*, 1093-1099.
73. Genoni, A.; X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique. *J. Chem. Theory Comput.* **2013**, *9*, 3004-3019.
74. Dos Santos, L. H. R.; Genoni, A.; Macchi, P. Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density. *Acta Crystallogr., Sect. A* **2014**, *70*, 532-551.

75. Genoni, A.; Meyer, B. X-Ray Constrained Wave Functions: Fundamentals and Effects of the Molecular Orbitals Localization. *Adv. Quantum Chem.*, doi: 10.1016/bs.aiq.2015.05.008.

76. Jayatilaka, D. Wave Function for Beryllium from X-ray Diffraction Data. *Phys. Rev. Lett.* **1998**, *80*, 798−801.

77. Jayatilaka, D.; Grimwood, D. J. Wavefunctions Derived from Experiment. I. Motivation and Theory. *Acta Crystallogr., Sect. A* **2001**, *57*, 76−86.

78. Grimwood, D. J.; Jayatilaka, D. Wavefunctions Derived from Experiment. II. A Wavefunction for Oxalic Acid Dihydrate. *Acta Crystallogr., Sect. A* **2001**, *57*, 87−100.

79. Bytheway, I.; Grimwood, D.; Jayatilaka, D. Wavefunctions Derived from Experiment. III. Topological Analysis of Crystal Fragments. *Acta Crystallogr., Sect. A* **2002**, *58*, 232−243.

80. Bytheway, I.; Grimwood, D. J.; Figgis, B. N.; Chandler, G. S.; Jayatiaka, D. Wavefunctions Derived from Experiment. IV. Investigation of the Crystal Environment of Ammonia. *Acta Crystallogr., Sect. A* **2002**, *58*, 244−251.

81. Grimwood, D. J.; Bytheway, I.; Jayatilaka, D. Wavefunctions Derived from Experiment. V. Investigation of Electron Densities, Electrostatic Potentials, and Electron Localization Functions for Noncentrosymmetric Crystals. *J. Comput. Chem.* **2003**, *24*, 470−483.

82. Hudák, M.; Jayatilaka, D.; Peraínova, L.; Biskupic, S.; Kozísek, J.; Bucinský, L. X-ray Constrained Unrestricted Hartree−Fock and Douglas−Kroll−Hess Wavefunctions. *Acta Crystallogr., Sect. A* **2010**, *66*, 78−92.

83. Clinton, W. L.; Galli, A. J.; Massa, L. J. Direct determination of pure-state density matrices. II. Construction of constrained idempotent one-body densities.

Phys. Rev. **1969**, *177*, 7-12.

84. Clinton, W. L.; Massa, L. J. Determination of the electron density matrix from x-ray diffraction data. *Phys. Rev. Lett.* **1972**, *29*, 1363-1366.
85. Clinton, W. L.; Frishberg, C. A.; Massa, L. J.; Oldfield, P. A. Methods for obtaining an electron-density matrix from x-ray data. *Int. J. Quantum Chem.* **1973**, *7*, 505-514.
86. Frishberg, C.; Massa, L. J. Idempotent density matrices for correlated systems from x-ray-diffraction structure factors. *Phys. Rev. B* **1981**, **24**, 7018-7024.
87. Massa, L.; Huang, L.; Karle, J. Quantum crystallography and the use of kernel projector matrices. *Int. J. Quantum. Chem: Quantum Chem. Symp.* **1995**, *29*, 371-384.
88. Pichon-Pesme, V.; Lecomte, C.; Lachekar, H. On Building a Data Bank of Transferable Experimental Electron Density Parameters: Application to Polypeptides. *J. Phys. Chem.* **1995**, *99*, 6242−6250.
89. Jelsch, C.; Pichon-Pesme, V.; Lecomte, C.; Aubry, A. Transferability of Multipole Charge-Density Parameters: Application to Very High Resolution Oligopeptide and Protein Structures. *Acta Crystallogr., Sect. D* **1998**, *54*, 1306−1318.
90. Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. On the Application of an Experimental Multipolar Pseudo-Atom Library for Accurate Refinement of Small-Molecule and Protein Crystal Structures. *Acta Crystallogr., Sect. A* **2007**, *63*, 108−125.
91. Koritsanszky, T.; Volkov, A.; Coppens, P. Aspherical-Atom Scattering Factors from Molecular Wave Functions. 1. Transferability and Conformation Dependence of Atomic Electron Densities of Peptides within the Multipole Formalism. *Acta Crystallogr., Sect. A* **2002**, *58*, 464−472.

92. Volkov, A.; Li, X.; Koritsanszky, T.; Coppens, P. Ab Initio Quality Electrostatic Atomic and Molecular Properties Including Intermolecular Energies from a Transferable Theoretical Pseudoatom Databank. *J. Phys. Chem. A* **2004**, *108*, 4283−4300.

93. Dominiak, P. M.; Volkov, A.; Li, X.; Messerschmidt, M.; Coppens, P. A Theoretical Databank of Transferable Aspherical Atoms and Its Application to Electrostatic Interaction Energy Calculations of Macromolecules. *J. Chem. Theory Comput.* **2007**, *3*, 232−247.

94. Dittrich, B.; Koritsanszky, T.; Luger, P. A Simple Approach to Nonspherical Electron Densities by Using Invarioms. *Angew. Chem., Int. Ed.* **2004**, *43*, 2718−2721.

95. Dittrich, B.; Hübschle, C. B.; Messerschmidt, M.; Kalinowski, R.; Grint, D.; Luger, P. The Invariom Model and Its Application: Refinement of D,L-Serine at Different Temperatures and Resolution. *Acta Crystallogr., Sect. A* **2005**, *61*, 314−320.

96. Dittrich, B.; Hübschle, C. B.; Luger, P.; Spackman, M. A. Introduction and Validation of an Invariom Database for Amino-Acid, Peptide and Protein Molecules. *Acta Crystallogr., Sect. D* **2006**, *62*, 1325−1335.

97. Dittrich, B.; Hübschle, C. B.; Pröpper, K.; Dietrich, F.; Stolper, T.; Holstein, J. J. The generalized invariom database (GID). *Acta Crystallogr., Sect. B* **2013**, *69*, 91-104.

98. Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Jelsch, C.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. II. Comparison with the Pseudoatoms Transferability. *J. Chem. Theory Comput.*, submitted.

99. 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.

100. Sironi, M.; Famulari, A.; Raimondi, M.; Chiesa, S. The transferability of extremely localized molecular orbitals. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 47-54.

101. 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.

102. 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.

103. Press,W. H.; Teukolsky, S. A.; Vetterling,W. T.; Flannery, B. P. Numerical Recipes in Fortran 77: The Art of Scientific Computing, 2nd ed.; Cambridge University Press: New York, 1992; pp 387−448.

104. 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.

105. Wiest, R.; Pichon-Pesme, V.; Bénard, M.; Lecomte, C. Electron distribution in peptides and related molecules. Experimental and theoretical study of Leu-Enkephalin Trihydrate. *J. Phys. Chem.* **1994**, *98*, 1351-1362.

106. Allen, F.H.; Bruno, I.J. Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data. *Acta Crystallogr., Sect. B* **2010**, *66*, 360−366.

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. Gaussian 09, Revision D.01, 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, Inc., Wallingford CT, 2009.

109. Keith, T. A. AIMAll, Version 13.11.04; TK Gristmill Software: Overland Park, KS, 2012; available online: aim.tkgristmill.com.

110. 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.

111. Burresi, E.; Sironi, M. Determination of extremely localized molecular orbitals in the framework of density functional theory. *Theor. Chem. Acc.* **2004**, *112*, 247-253.