



HAL
open science

Libraries of Extremely Localized Molecular Orbitals. II. Comparison with the Pseudoatoms Transferability

Benjamin Meyer, Benoît Guillot, Manuel Ruiz-Lopez, Christian Jelsch,
Alessandro Genoni

► **To cite this version:**

Benjamin Meyer, Benoît Guillot, Manuel Ruiz-Lopez, Christian Jelsch, Alessandro Genoni. Libraries of Extremely Localized Molecular Orbitals. II. Comparison with the Pseudoatoms Transferability. Journal of Chemical Theory and Computation, 2016, 12 (3), pp.1068-1081. 10.1021/acs.jctc.5b01008 . hal-02367785

HAL Id: hal-02367785

<https://hal.science/hal-02367785>

Submitted on 18 Nov 2019

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

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

This document is the Accepted Manuscript version of a Published Work that appeared in final form in the Journal of Chemical Theory and Computation, copyright © American Chemical Society after peer review and technical editing by the publisher.

To access the final edited and published work see <https://pubs.acs.org/doi/10.1021/acs.jctc.5b01008>.

Meyer, B., Guillot, B., Ruiz-Lopez, M. F., Jelsch, C., & Genoni, A. (2016). Libraries of extremely localized molecular orbitals. 2. Comparison with the pseudoatoms transferability. *Journal of chemical theory and computation*, 12(3), 1068-1081.

Libraries of Extremely Localized Molecular Orbitals. II.

Comparison with the Pseudoatoms Transferability

Benjamin Meyer ^(1,2), Benoît Guillot ^(3,4) *, Manuel F. Ruiz-Lopez ^(1,2), Christian Jelsch ^(3,4), 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;

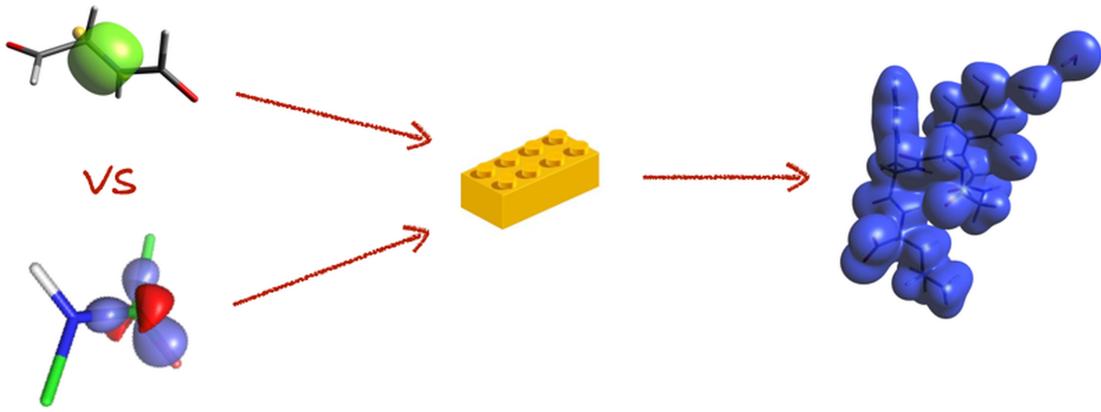
- Benoît Guillot, CNRS and Université de Lorraine, Laboratoire CRM2 - UMR 7036, Boulevard des Aiguillettes, BP 70239, 54506 Vandoeuvre-lès-Nancy, France. E-mail : Benoit.Guillot@univ-lorraine.fr; Phone: +33 (0)3 83 68 48 99; Fax+33 (0)3 83 68 43 00;

Abstract

Due to both technical and methodological difficulties, determining and analyzing charge densities of very large molecular systems represents a serious challenge that, in the crystallographers community, has been mainly tackled observing that the so-called pseudoatoms of the electron density multipole expansions are reliably transferable from molecule to molecule. This has led to the construction of pseudoatoms databanks that have allowed to successfully refining crystallographic structures of macromolecules, while taking into account their corresponding reconstructed electron distributions. A recent alternative/complement to the previous approach is represented by techniques based on Extremely Localized Molecular Orbitals (ELMOs) that, due to their strict localization on small molecular fragments (e.g., atoms, bonds and functional groups), are also in principle exportable from system to system.

The ELMOs transferability has been already tested in detail and, in this paper, it has been compared to the one of the pseudoatoms. To accomplish this task, electron distributions obtained both through the transfer of pseudoatoms and through the transfer of Extremely Localized Molecular Orbitals have been analyzed, especially taking into account topological properties and similarity indexes. The obtained results indicate that all the considered reconstruction methods give completely reasonable and similar charge densities and, consequently, the new ELMOs libraries will probably represent new useful tools not only for refining crystal structures, but also for computing approximate electronic properties of very large molecules.

TOC Graphic



1. Introduction

As well known, the Hohenberg & Kohn theorem¹, which has established the existence of a functional relation between the ground state charge distributions and the ground state wave functions of the many-electron systems, has raised the electron density to the role of prominent physical quantity. In fact, due to this theorem, it has become clear that, at least in principle, the only electron density would be sufficient to fully define all the ground state properties of any electronic system. For this reason, research efforts devoted to theoretically and experimentally obtain accurate charge distributions have become more and more important and they have contributed to the creation of a well-defined and mature domain of science.²⁻⁴

In particular, the determination of electron densities from experimental X-ray diffraction data has known a tremendous improvement and, to accomplish this task, several theoretical models have been proposed over the years. In this context, it is worth mentioning the wave function-based approaches⁵⁻²⁰, the maximum entropy methods²¹ and, above all, the well-known multipole models,²²⁻²⁴ which, until now, remain unquestionably the most successful and widely used techniques to reconstruct charge distributions from experimental diffraction measurements. They basically consist in writing the global electron density as a sum of atomic multipolar expansions (known as “pseudoatoms”) depending on suitable parameters that are usually refined through a “least squares fitting” against experimentally collected structure factors. Their main advantage relies on the fact that, unlike the traditional “Independent Atom Model”, the aspherical deformation of the atomic density functions, which can be ascribed to bonding and non-bonding interactions, can be fully taken into account.

Nowadays, the multipole models allow charge density studies of high-resolution structures, not only providing reliable atomic coordinates and thermal parameters (i.e., Anisotropic Displacement Parameters (ADPs)), but also giving important insights into the electrostatic properties and the intra- and the inter-molecular interactions that characterize the systems under investigation. This would be particularly relevant for biological molecules. In fact, the obtained results would be of invaluable help for a better understanding of the chemical and biological role played by the macromolecules and, consequently, they could also lead to the rational design of new drug molecules. Nevertheless, although impressive high-resolution charge density studies of quite large systems have been recently performed,²⁵⁻³² the application of the multipole models to large biomolecules still remains a challenge in most of the cases. In fact, the refinement of macromolecular charge densities usually faces non-negligible problems, such as the limited resolution of the X-ray diffraction data and the high thermal motion of atoms.

In order to circumvent the previous difficulties, crystallographers interested in charge density studies have resorted to the observation that multipole parameters of atoms characterized by similar chemical environments are reliably transferable from molecule to molecule.³³ This has led to the construction of different libraries of pseudoatoms that have been successfully used not only to instantaneously obtain electron distributions of very large molecular systems, but also to obtain more accurate structural refinements (namely, atomic positions and thermal parameters) and approximate electrostatic properties of several biomolecules.

The seminal paper for the creation of pseudoatoms databanks dates back to 1995 when Pichon-Pesme and coworkers have reconstructed the charge distribution of a peptide backbone using multipolar parameters previously obtained from the

experimental electron density refinements of small peptides and amino acids.³⁴ This has led the Nancy group to the construction of the Experimental Library of Multipolar Atom Model (ELMAM),^{25,34-36} which contains all the possible chemically unique pseudoatoms in the twenty natural amino acids found in proteins. The database has been afterwards improved (ELMAM2)^{37,38} and now it covers also the most common functional groups found in organic molecules. The main strategy to build these databanks consisted in averaging the values of multipole parameters of chemically equivalent pseudoatoms, which had been previously obtained from high-quality experimental charge density refinements. This enabled to label both ELMAM and ELMAM2 as “experimental pseudoatoms libraries”.

Two theoretical libraries of pseudoatoms have been afterwards proposed independently by two other research groups: the University at Buffalo Pseudoatom Databank (UBDB)³⁹⁻⁴³ and the Invariom database⁴⁴⁻⁴⁸. Both of them are based on the idea of building pseudoatoms libraries starting from charge densities refinements that exploit theoretical structure factors resulting from proper gas-phase *ab initio* calculations. In the former, the stored multipole parameters have been determined averaging over results obtained from refinements of theoretical electron distributions computed at B3LYP / 6-31G(d,p) level for crystallographic geometries deposited in the Cambridge Structural Database (CSD). In the latter, each pseudoatom derives only from a single multipole model refinement that uses the structure factors associated with a gas-phase charge distribution previously computed on the optimized geometry of the chosen model molecule for the pseudoatom in exam (B3LYP/D95++(3df,3pd) level).

Even if these three pseudoatoms libraries share the same objective and overall philosophy, each of them presents its own specific pros and cons. The great advantage

of the experimental ELMAM and ELMAM2 consists in the possibility of automatically include, in an average way, the effects of the chemical environment as found in crystalline state, especially the ones due to directional intermolecular interactions such as hydrogen bonds. Nevertheless, the stored averaged multipole parameters are determined only after suitable high-resolution experimental X-ray diffraction data collections and careful electron density models refinement. Therefore, the main shortcoming associated with the experimental pseudoatoms libraries is their lack of flexibility, since the inclusion of possible missing “pseudoatom types” relies on the availability of compatible sets of experimental data. The just mentioned drawback is obviously overcome by the use of the *ab initio* computations on which the UBDB and Invariom databanks are based. In fact, in these cases, adding a missing pseudoatom needed for a particular application can be done either by selecting appropriate CSD entries for the UBDB approach, or “designing” an *ad hoc* compound containing the pseudoatom type of interest in the case of the Invariom databank. Nevertheless, for both the two theoretical libraries, the effect of the intermolecular interactions is missing, although, in the UBDB case, the fact of considering crystal geometries as starting point might alleviate this shortage.

The three databases, which have already been used for about ten/fifteen years, have been extensively tested and compared.⁴⁹⁻⁵¹ The main conclusion is that they usually provide similar results, especially considering atomic positions and thermal parameters obtained from structural refinements of large molecules. The same general agreement has been also observed for the reconstructed electron distributions, with the largest dissimilarities detected in the valence regions of the polar atoms. These discrepancies, which obviously arise from the different strategies adopted in the construction of the databanks, are probably the cause of the larger differences that

were observed for the electrostatic properties, especially for the calculated electrostatic intermolecular interaction energies.

Finally, for the sake of completeness, it is worth mentioning that a new experimental library of pseudoatoms has been also recently proposed. The aim of the new Supramolecular Synthon-Based Fragments Approach (SBFA)⁵²⁻⁵³ is to provide an alternative database that, including the effects of the intermolecular interactions between synthons, can be suitably used for crystal engineering purposes.

Although nowadays the pseudoatoms databanks seem well-established and quite powerful tools to elude the main hindrances associated with charge density studies of macromolecules, a completely new alternative way to accomplish this task has been recently envisaged: the possibility of exploiting the so-called Extremely Localized Molecular Orbitals (ELMOs). These orbitals⁵⁴⁻⁵⁶ are variationally obtained under the constraint of being expanded exclusively on “fragment basis-sets” determined according to a user-defined and chemically meaningful localization scheme. For this reason, they are strictly localized (namely, without small tails outside the main localization region) on small molecular units, such as atoms, bonds or functional groups and, consequently, they are easily transferable from molecule to molecule, provided that, as in the pseudoatoms case, the environment of the fragment in exam is chemically equivalent in the two different systems. Therefore, the construction of ELMOs libraries represents a reasonable and tempting alternative to the pseudoatoms databases, both to reconstruct electron densities and to refine crystallographic structures of large macromolecules. The novelty of the new approach consists in the fact that, after the transfer, not only the charge distribution, but also an approximate wave function for the system in exam will be available almost instantaneously, which will obviously allow the computation of electronic properties that, at the moment,

cannot be determined using only the electron density since the knowledge of an exact density functional is still missing. For this reason, the new technique based on the transfer of Extremely Localized Molecular Orbitals can be perfectly placed in the framework of the linear scaling methods^{57,58}, among which it is worth mentioning the Divide & Conquer strategy^{59,60}, the LEGO-type approaches introduced by Mezey⁶¹⁻⁶⁴, the Fragment Molecular Orbital (FMO) technique^{65,66}, the Kernel Energy Method⁶⁷⁻⁷⁴ (KEM), the Transferable Atom Equivalent (TAE) strategy^{75,76} and other real-space approaches.⁷⁷⁻⁸⁰

Although the ELMOs transferability had been already thoroughly investigated through several test calculations on small molecules,^{55,56,81,82} a very detailed study has recently confirmed that the ELMOs are indeed reliably transferable also to very large systems, as large polypeptides or proteins.⁸³ However, before constructing complete libraries of Extremely Localized Molecular Orbitals and developing novel ELMO-based refinement techniques, a comparison between the ELMOs and the pseudoatoms transferability would be desirable. In this paper, to accomplish this task, we will carefully compare the electron distributions obtained both through the transfer of Extremely Localized Molecular Orbitals and through of the transfer of pseudoatoms (in particular, considering the ELMAM2 and the UBDB libraries). This will be done mainly analyzing the topological properties of the considered electron distributions, but more global descriptors, such as net integrated charges and similarity indexes, will be also taken into account.

The paper is organized as follows. At first, we will briefly present the ELMO and the multipole model methods. Afterwards, we will focus on the computational details of our investigation, while, in a later section, we will show and discuss the obtained results. In the last part, we will draw the final conclusions of our study.

2. Theory

To conduct our investigation we have mainly used two theoretical methods: the strategy proposed by Stoll⁵⁴ to obtain Extremely Localized Molecular Orbitals and the Hansen & Coppens multipole model technique²⁴ with its extension to the related pseudoatoms libraries.

Concerning the ELMOs, in this section we only remind that they can be simply obtained (i) defining *a priori* a localization scheme that subdivides the investigated system into “chemical” fragments and (ii) solving a proper eigenvalue equation for each defined subunit.

Furthermore, as already mentioned in the introduction, due to their strict localization, the ELMOs are easily transferable from molecule to molecule. However, the Extremely Localized Molecular Orbitals can be really exported only if it is possible to define a reference frame both for the model molecule (namely, the molecule on which the ELMO is determined) and for the target molecule (namely, the molecule to which we want to transfer the previously computed localized orbital). As proposed by Philipp and Friesner⁸⁴ this is done by defining two corresponding triads of atoms (one associated with the model molecule and one associated with the target molecule) that, if necessary, fully take into account the local dissymmetry of the fragment in exam.⁸⁵

More details about the theory of the Extremely Localized Molecular Orbitals and the strategy for their transfer can be found in the Supporting Information and in our preliminary study about the ELMOs transferability.⁸³

Given its wide spread in the charge density studies, also for the Hansen & Coppens multipolar model²⁴ we will sketch only its main features. First of all, it is worth reminding that this method models the molecular (or crystal) static electron density as

a sum of aspherical pseudoatoms, whose charge density $\rho_{atom}(\mathbf{r})$ is described by an atom-centered multipole expansion

$$\rho_{atom}(\mathbf{r}) = \rho_{core}(r) + P_{val}\kappa^3\rho_{val}(r) + \sum_{l=1}^4 \kappa'^3 R_l(\kappa'r) \sum_{m=-l}^{+l} P_{lm} Y_{lm}(\theta, \varphi) \quad (1)$$

where ρ_{core} and ρ_{val} are spherically averaged free-atom Hartree-Fock core and valence electron densities, the latter being normalized to one electron and weighted by an atomic valence population P_{val} . The deformation part of the valence electron density is modelled by radial Slater-type functions R_l , and angular real spherical harmonics Y_{lm} populated by multipole parameters P_{lm} . Finally, κ and κ' are dimensionless expansion-contraction coefficients applied to the spherical and multipolar parts of the valence electron density, respectively. During an experimental charge density study, the P_{val} , $\{P_{lm}\}$, κ and κ' parameters (usually referred as *multipolar parameters*) are refined against X-ray data to fit the experimental deformation density. The sum over l indices (Equation (1)) depends on the level of the multipolar expansion, which determines the number of refined multipole populations P_{lm} needed to model properly the deformation electron density of various chemical species. An octupolar level ($l = 3$) is usually applied for C, N and O atoms while a hexadecapolar one ($l = 4$) is generally considered for heavier atom types. Hydrogen atoms are modelled either by a bond-oriented dipolar function ($l = 1$) or by a quadrupolar expansion ($l = 2$).

In the ELMAM2 and UBDB libraries, the transferable pseudoatoms are stored as sets of multipolar parameters. They have been obtained by least squares fitting against experimental (for ELMAM2) or theoretical (for UBDB) high-resolution structure factors. The parameters were subsequently averaged by chemical atom types^{34,42}, whose definitions are based on the chemical nature of the atoms, their hybridation

state and the nature of their neighbors. In both libraries, the number of multipole parameters P_{lm} needed to describe the electron density of a given pseudoatom is minimized by accounting for the local symmetry of the corresponding atom type.⁸⁶

3. Computational Details

Target System and Methods. As in our previous paper regarding the ELMOs transferability,⁸³ our target system consisted in the Leu-enkephalin polypeptide (Tyr-Gly-Gly-Phe-Leu) and three interacting water molecules (see Figure 1). The geometry used for all our calculations has been obtained by means of an X-ray diffraction experiment conducted at 100 K (1.15 \AA^{-1} resolution).⁸⁷ Also in this case, it is worth reminding that the hydrogen atoms positions have been initially determined from the collected experimental data and, in order to correct for the well-known bias deriving from the X-ray diffraction measurements, they have been afterwards properly optimized elongating all the X-H bond lengths to match the average values resulting from neutron diffraction experiments.⁸⁸

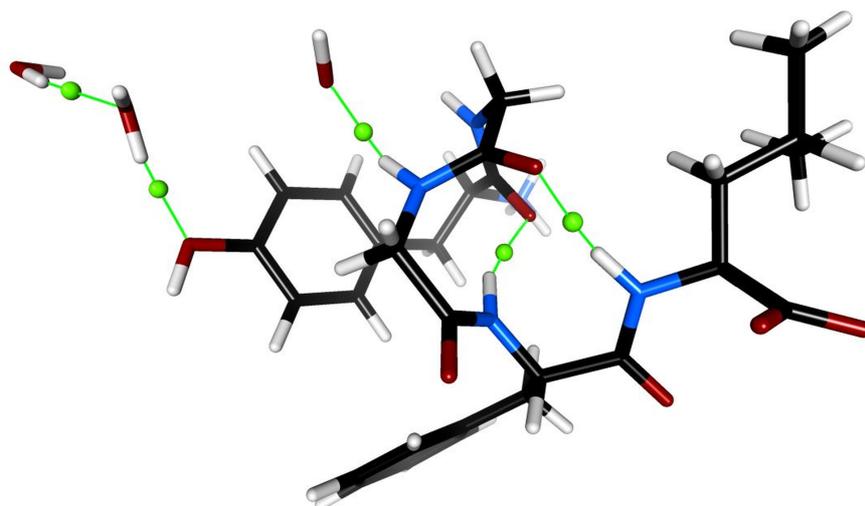


Figure 1. Leu-enkephalin pentapeptide and interacting water molecules. The five hydrogen bond interactions and the corresponding bond critical points are explicitly shown (green spheres).

In this study, the pentapeptide electron density has been reconstructed both through the transfer of pseudoatoms and through the transfer of Extremely Localized Molecular Orbitals. In the former case, we have used both the experimental database ELMAM2 exploiting the MoPro software^{89,90} (version June 2015) and the theoretical databank UBDB (version 2012) making use of the computer program LSDB.^{40,42,43} No “missing atom types” were signaled during the transfer procedure, using either databases. As it involves independently averaged atomic valence populations P_{val} (Equation (1)), the transfer can lead to a slight deviation from the expected formal charge of the studied compound. In our case, the pentapeptide and the water molecules are expected to be neutral, with a total of 240 valence electrons in the whole complex. Using the ELMAM2 library, the transfer led to a charge deficit of 0.0681 electrons, which was subsequently corrected by uniformly adding +0.00079 e to each atomic valence populations of the 86 transferred pseudoatoms. When transferring multipole parameters from the UBDB databank, the deviation from neutrality is larger, with a charge deficit of 0.9699 e. In this case, the non-uniform “*sigmaPv*” charge scaling method was used as implemented in the LSDB program, leading to atomic valence populations shifts ranging from +0.00072 e (for the hydrogen atoms of the water molecules) to +0.032 e (for the carbon atoms of the leucine side-chain). These neutralization procedures were performed on the whole system, without ensuring the neutrality of the individual molecules. Nevertheless, the resulting (valence populations-based) charges of the water molecules and of the pentapeptide turned out to be very small, with -0.00062 e / +0.0019 e, and -0.0039e / +0.0117 e respectively for ELMAM2 and UBDB, reflecting the robustness of the multipolar pseudoatoms transferability principle.

Concerning the Extremely Localized Molecular Orbitals, the transfer has been performed following a strategy already described in our first paper,⁸³ namely the desired ELMOs have been determined on suitable model molecules and afterwards properly exported to the target system using an in-house program that implements the rotation technique introduced by Philipp and Friesner.⁸⁴ For the sake of clarity and completeness, it is worthwhile to point out that, unlike our previous study,⁸³ in the present investigation the model molecules for the preliminary ELMO calculations have not been designed considering only one approximation. In particular, both the *Nearest Bond Approximation* (NBA) and the *Nearest Functional Group Approximation* (NFGA) have been used in order to obtain the best possible description using transferred ELMOs. While the former approximation provides model molecules consisting in the fragment of interest (namely, the fragment on which the desired ELMO is localized) and in its nearest neighbor bonds properly capped with hydrogen atoms, the latter gives larger model systems composed of the fragment in exam and of its suitably hydrogen-capped nearest functional groups. The reason for choosing this dual model molecules approximation relies on the fact that the generally more reliable NFGA approximation may sometimes lead to model molecules characterized by undesired intra-molecular interactions that are completely missing in the target system and that consequently artificially perturb the ELMOs of interest.⁸³ Therefore, to overcome this drawback, in this study we have decided to also resort to the NBA, which, although less sophisticated, in some circumstances is able to provide superior descriptions compared to the Nearest Functional Group Approximation.

All the ELMO calculations have been performed exploiting the version 8 of the GAMESS-UK quantum chemistry package⁹¹ that has been properly modified to

implement the Stoll equations.^{54,55} In particular, the ELMOs have been computed considering six different basis-sets. The results obtained at the 6-311G, 6-311G(d,p) and 6-311+G(2d,2p) levels will be thoroughly reported and commented in the text, while, for the sake of completeness, the analogous results resulting from computations with the 6-31G, 6-31G(d,p) and 6-311+G(d,p) sets of basis functions will be provided in the Supplementary Information (See Tables S3-S19 and Figures S3, S5 and S7).

Comparison of the electron densities. The topological properties resulting from the QTAIM (Quantum Theory of Atoms in Molecules) analyses⁹² of the electron densities obtained at the different levels of approximation (ELMAM2, UBDB and ELMO) have been taken into account in order to compare the ELMOs and the pseudoatoms transferability. All the analyses of the ELMO electron densities have been carried out exploiting the AIMAll package⁹³ (version 13.11.04), while the ones of the pseudoatoms charge distributions have been performed using the MoPro software.^{89,90} In this context, we have mainly focused on the values of the electron density and of its Laplacian ($\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$, respectively) both at the covalent and at the non-covalent bond critical points (BCPs). In the former case, given the obvious large number of covalent interactions in the pentapeptide, the QTAIM properties will be reported only as average values for the different types of bonds occurring in the molecule. On the contrary, the $\rho(\mathbf{r}_b)$ and $\nabla^2\rho(\mathbf{r}_b)$ values obtained for all the non-covalent bond critical points will be explicitly shown.

Furthermore, to have a global view of the similarities/dissimilarities between the obtained values of the above mentioned topological properties, for each couple (A, B) of reconstruction methods, the following index has been also computed:

$$R_{A-B}(X) = 100 \frac{\sum_{i=1}^{n_{BCP}} |X_A(\mathbf{r}_i) - X_B(\mathbf{r}_i)|}{\sum_{i=1}^{n_{BCP}} |X_A(\mathbf{r}_i) + X_B(\mathbf{r}_i)|} \quad (2)$$

where n_{BCP} is the number of considered bond critical points and X is the topological property in exam (for instance, $\rho(\mathbf{r}_b)$ at the covalent or at the non-covalent BCPs). Of course, the complete similarity is obtained when $R_{A-B}(X) = 0$.

Only for the five critical points corresponding to hydrogen-bond interactions (see Figure 1) three other topological properties have been also considered: the kinetic energy density $G(\mathbf{r}_b)$, the potential energy density $V(\mathbf{r}_b)$ and the positive curvature $\lambda_3(\mathbf{r}_b)$ of the electron distribution. The obtained values for these properties have been compared to the ones resulting from the following three empirical exponential relations:

$$G(\mathbf{r}_b) = 12(2) \times 10^3 \exp[-2.73(9) d(\text{H} \cdots \text{O})] \quad (3)$$

$$V(\mathbf{r}_b) = -50.0(1.1) \times 10^3 \exp[-3.6 d(\text{H} \cdots \text{O})] \quad (4)$$

$$\lambda_3(\mathbf{r}_b) = 0.41(8) \times 10^3 \exp[-2.4(1) d(\text{H} \cdots \text{O})] \quad (5)$$

which have been established by Espinosa and coworkers analyzing topological and structural data from accurate electron density studies that involve X-H \cdots O (X = C, N, O) hydrogen bonds.^{94,95} In this regard it is worthwhile to point out that the universal validity of these relations is still an open problem in the charge density community, as also shown by the recent study performed by Spackman⁹⁶, who has investigated and discussed the reliability of the Espinosa-Molins-Lecomte relation to estimate interaction energies between molecules in crystals. In this paper equations (3)-(5) have been mainly used to have a further reference for our comparisons, without the aim of exploiting them to compute intermolecular interaction energies.

A comparison of the charges obtained integrating the different electron distributions over the resulting QTAIM atomic basins has been also performed. However, given the large number of atoms for the system in exam, we have decided not to consider the individual atomic charges. Instead, we have computed and compared the global

charges for the main “functional groups” of the polypeptide backbone (i.e., the amino group, the carboxylic group and the four peptide groups) and the ones for the side-chains (including in each case the C α -H bond).

Finally, more global comparisons of the charge distributions have been performed taking into account two real-space similarity indexes: the Walker-Mezey similarity indicator⁶² and the Real-Space R (RSR) value.⁹⁷ The former, which is usually indicated as $L(a, a')$, allows point-by-point comparisons of two charge distributions in density shells enclosed by the values a and a' and the perfect agreement is obtained when $L=100$. The latter is defined as

$$RSR(\rho_x, \rho_y) = 100 \frac{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) - \rho_y(\mathbf{r}_i)|}{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) + \rho_y(\mathbf{r}_i)|} \quad (6)$$

where n_p is the number of density-grid points, and the complete similarity is reached when $RSR=0$. It is worth noting that the RSR value differs only for a different grid sampling from the index R , which was just introduced above (see equation (2)) to evaluate the global agreement between topological properties values associated with different reconstruction strategies of the electron density.

To evaluate all the similarity values, three-dimensional grids with a 0.083131 bohr step-size for each direction have been considered. While the ELMO grids have been obtained using the *Cubegen* utility of the Gaussian09 package⁹⁸, the ones corresponding to the pseudoatoms electron distributions have been determined through the VMoPro software included in the MoPro package.

Results and Discussion

As already mentioned in this paper, the comparison of the ELMOs and the pseudoatoms transferability has been mainly performed considering the topological properties of the reconstructed charge distributions.

At first, we have considered the electron density at the covalent bond critical points. In Table 1, where we have reported the average values obtained for each type of covalent bond, it is possible to see that the different methods generally provide reasonable and quite similar results. In particular, using ELMAM2 as reference strategy, we can observe that, within each reconstruction technique, the largest discrepancies amount to $-0.13 \text{ e}/\text{\AA}^3$ (UBDB), $-0.30 \text{ e}/\text{\AA}^3$ (ELMO/6-311G), $0.18 \text{ e}/\text{\AA}^3$ (ELMO/6-311G(d,p)) and $-0.16 \text{ e}/\text{\AA}^3$ (ELMO/6-311+G(2d,2p)). Furthermore, concerning the ELMO transfer, we can observe that, exploiting the 6-311G basis-set, the $\rho(\mathbf{r}_b)$ values are lower than the ELMAM2 ones for almost all the covalent bond types. However, the electron density values at the covalent BCPs generally increase as larger and more flexible sets of basis functions (i.e., 6-311G(d,p) and 6-311+G(2d,2p)) are used. Finally, in Table 1, it is also interesting to note that for less-polar bonds (e.g., C-H and C-C) the “un-polarized” basis-set 6-311G allows obtaining $\rho(\mathbf{r}_b)$ values that are the closest to the ones obtained after the transfer of pseudoatoms. On the contrary, when more polar bonds are taken into account (e.g., (C-O)_{term} and O-H), the closest ELMO values to the ELMAM2 and UBDB ones are those obtained with the 6-311+G(2d,2p) basis-set.

Regarding the Laplacian of the electron density at the covalent bond critical points (see Table 2), an expected much larger variability has been observed, with the largest discrepancies with respect to the reference ELMAM2 values amounting to $18.6 \text{ e}/\text{\AA}^5$ (UBDB), $10.9 \text{ e}/\text{\AA}^5$ (ELMO/6-311G), $15.5 \text{ e}/\text{\AA}^5$ (ELMO/6-311G(d,p)) and $-14.0 \text{ e}/\text{\AA}^5$

(ELMO/6-311+G(2d,2p)). However, although it is difficult to find a consensus, for almost all the bond types, the different methods provide values that are of the same order of magnitude and, above all, the lowest discrepancies are not necessarily always between the ELMAM2 and the UBDB results. Furthermore, it is worth noting that, within all the different strategies, all the Laplacian values are negative, as it is expected for covalent bond interactions.

Table 1. Average values^(a) of the electron density at the covalent bond critical points (in $e/\text{\AA}^3$) for each bond type after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). The corresponding standard deviations are reported in the Supporting Information (see Table S1).

Bond Type	Pseudoatoms Transfer		ELMOs Transfer		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	1.84	1.89	1.81	1.96	2.01
C-C	1.61	1.66	1.62	1.79	1.77
(C-C) _{ar}	2.12	2.08	2.02	2.16	2.20
(C-N) _{peptide}	2.27	2.26	2.15	2.27	2.35
(C-N) _{term}	1.62	1.56	1.32	1.54	1.46
C α -N	1.73	1.74	1.66	1.79	1.84
(C-O) _{term}	2.66	2.55	2.46	2.58	2.65
(C-O) _{peptide}	2.65	2.70	2.59	2.73	2.81
(C-O) _{phenol}	2.10	2.00	1.88	1.99	2.06
N-H	2.26	2.19	2.15	2.32	2.37
O-H	2.52	2.39	2.33	2.49	2.51

(a) For (C-N)_{term} and (C-O)_{phenol} only one bond critical point is considered.

Table 2. Average values^(a) of the Laplacian of the electron density at the covalent bond critical points (in $e/\text{\AA}^5$) for each bond type after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). The corresponding standard deviations are reported in the Supporting Information (see Table S2).

Bond Type	Pseudoatoms Transfer		ELMOs Transfer		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
C-H	-17.9	-20.2	-19.7	-25.9	-28.6
C-C	-9.0	-11.3	-12.9	-19.2	-17.2
(C-C) _{ar}	-18.4	-17.3	-19.5	-24.8	-26.0
(C-N) _{peptide}	-22.1	-23.7	-20.9	-18.0	-29.8
(C-N) _{term}	-6.9	-9.6	-1.8	-6.4	-6.8
C α -N	-9.4	-10.1	-14.1	-16.0	-23.4
(C-O) _{term}	-29.8	-24.2	-21.0	-14.3	-24.7
(C-O) _{peptide}	-22.1	-24.5	-20.8	-12.4	-23.8
(C-O) _{phenol}	-17.6	-17.1	-8.2	-4.2	-14.3
N-H	-37.0	-28.5	-32.6	-44.5	-46.8
O-H	-58.0	-39.4	-47.1	-64.0	-69.0

(a) For (C-N)_{term} and (C-O)_{phenol} only one bond critical point is considered.

At a second stage, non-covalent bond critical points have been considered. First of all, it is important to observe that the topological analyses of the five electron distributions taken into account have detected the same non-covalent interactions, namely five intra- and inter-molecular hydrogen bonds and other eight weak interactions. All of them are correctly characterized by a BCP with a positive value of the Laplacian of the electron density.

Analyzing the values of the electron distributions at the non-covalent bond critical points (see Table 3), we can immediately note a reduced variability compared to the covalent case. In fact, considering only the five hydrogen bond interactions and using again the ELMAM2 values as a reference, the largest discrepancies observed for each method are $-0.070 \text{ e}/\text{\AA}^3$ (UBDB), $-0.058 \text{ e}/\text{\AA}^3$ (ELMO/6-311G), $-0.081 \text{ e}/\text{\AA}^3$ (ELMO/6-311G(d,p)) and $-0.088 \text{ e}/\text{\AA}^3$ (ELMO/6-311+G(2d,2p)). Furthermore, considering the other eight weak interactions, the discrepancies further decrease and a very good agreement among the different reconstruction techniques is observed.

The same general trend is detected for the Laplacian of the electron density (see Table 4). In fact, a reduced variability can be easily seen and the largest differences with respect to the ELMAM2 values amount only to $0.91 \text{ e}/\text{\AA}^5$ (for UBDB), $1.72 \text{ e}/\text{\AA}^5$ (ELMO/6-311G), $1.57 \text{ e}/\text{\AA}^5$ (ELMO/6-311G(d,p)) and $1.17 \text{ e}/\text{\AA}^5$ (ELMO/6-311+G(2d,2p)). Moreover, also in this case, considering the eight weak interactions, the agreement among the five reconstruction strategies increases.

A more global indication of the similarities/dissimilarities between the values of the topological properties associated with the charge distributions obtained through the different methods can be found in Tables 5-8, where the values of the agreement index R (see equation (2)) are reported.

Considering the electron density at the covalent bond critical points (Table 5), it is easy to note a similar overall agreement among the ELMAM2, UBDB and ELMO/6-311G electron distributions, while, as expected, a quite large similarity has been observed between the ELMO/6-311G(d,p) and ELMO/6-311+G(2d,2p) charge densities. A completely analogous trend is observed for the Laplacian of the electron density at the BCPs (Table 6) with two main groups of charge distributions detected by the values of the index R : the one constituted by the ELMAM2, UBDB and

ELMO/6-311G electron distributions and the one comprising the other two ELMO charge densities.

Table 3. Values of the electron density at the non-covalent bond critical points (in $e/\text{\AA}^3$) after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). Both hydrogen bonds and weak interactions are considered. For the labels specification see the Leu-enkephalin target geometry deposited in the Supporting Information.

Non-Covalent Interaction	Pseudoatoms Transfer		ELMOs Transfer		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37...Ow84	0.191	0.191	0.186	0.174	0.163
Ow81-Hw83...O5	0.238	0.176	0.189	0.168	0.158
Ow78-Hw80...Ow81	0.134	0.132	0.113	0.103	0.103
N58-H74...O25	0.224	0.154	0.166	0.143	0.136
N38-H57...O6	0.165	0.166	0.145	0.133	0.131
C11...H35	0.024	0.028	0.030	0.030	0.029
C13...O84	0.040	0.041	0.046	0.046	0.045
H17...H50	0.022	0.024	0.020	0.021	0.021
H19...C42	0.064	0.069	0.065	0.063	0.064
H20...C46	0.017	0.018	0.023	0.023	0.022
H21...O81	0.029	0.032	0.035	0.036	0.037
H28...H65	0.021	0.025	0.022	0.023	0.023
H56...O32	0.052	0.052	0.048	0.047	0.049

Concerning the values of the topological properties at the non-covalent bond critical points, larger global similarities (namely, lower values of the agreement index) have

been observed between the different electron densities (see Tables 7 and 8), which have been already noted analyzing Tables 3 and 4. Furthermore, in these cases, an optimal agreement between all the ELMO electron distributions has been noted, with the index R always lower than or equal to 0.5%.

Table 4. Values of the Laplacian of the electron density at the non-covalent bond critical points (in $e/\text{\AA}^5$) after the different transfers. Both hydrogen bonds and weak interactions are considered. For the labels specification see the Leu-enkephalin target geometry deposited in the Supporting Information.

Non-Covalent Interaction	Pseudoatoms Transfer		ELMOs Transfer		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37...Ow84	1.96	2.43	3.09	2.96	2.70
Ow81-H83...O5	1.22	1.92	2.87	2.69	2.30
Ow78-H80...Ow81	1.42	2.33	3.14	2.99	2.59
N58-H74...O25	1.13	1.41	1.87	1.74	1.63
N38-H57...O6	1.43	1.87	2.50	2.36	2.19
C11...H35	0.27	0.28	0.32	0.29	0.30
C13...O84	0.57	0.59	0.59	0.58	0.56
H17...H50	0.24	0.23	0.30	0.26	0.26
H19...C42	0.64	0.63	0.76	0.69	0.75
H20...C46	0.26	0.27	0.27	0.27	0.27
H21...O81	0.50	0.53	0.55	0.53	0.52
H28...H65	0.25	0.24	0.32	0.29	0.29
H56...O32	0.67	0.62	0.71	0.67	0.70

Table 5. Values of the agreement index R (%) between the different reconstruction methods (ELMAM2, UBDB and ELMO with basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)) considering as topological property the electron density at the covalent bond critical points.

	ELMAM2	UBDB	ELMO 6-311G	ELMO 6-311G(d,p)	ELMO 6-311+G(2d,2p)
ELMAM2	0.0	1.3	1.6	2.1	2.7
UBDB		0.0	1.6	1.8	2.5
ELMO/6-311G			0.0	3.3	3.9
ELMO/6-311G(d,p)				0.0	0.9
ELMO/6-311+G(2d,2p)					0.0

Table 6. Values of the agreement index R (%) between the different reconstruction methods (ELMAM2, UBDB and ELMO with basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)) considering as topological property the Laplacian of the electron density at the covalent bond critical points.

	ELMAM2	UBDB	ELMO 6-311G	ELMO 6-311G(d,p)	ELMO 6-311+G(2d,2p)
ELMAM2	0.0	7.1	7.5	16.8	17.0
UBDB		0.0	5.6	16.4	13.0
ELMO/6-311G			0.0	13.0	14.5
ELMO/6-311G(d,p)				0.0	7.0
ELMO/6-311+G(2d,2p)					0.0

Table 7. Values of the agreement index R (%) between the different reconstruction methods (ELMAM2, UBDB and ELMO with basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)) considering as topological property the electron density at the non-covalent bond critical points.

	ELMAM2	UBDB	ELMO 6-311G	ELMO 6-311G(d,p)	ELMO 6-311+G(2d,2p)
ELMAM2	0.0	0.7	1.0	1.2	1.2
UBDB		0.0	0.7	0.8	0.8
ELMO/6-311G			0.0	0.3	0.5
ELMO/6-311G(d,p)				0.0	0.2
ELMO/6-311+G(2d,2p)					0.0

Table 8. Values of the agreement index R (%) between the different reconstruction methods (ELMAM2, UBDB and ELMO with basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)) considering as topological property the Laplacian of the electron density at the non-covalent bond critical points.

	ELMAM2	UBDB	ELMO 6-311G	ELMO 6-311G(d,p)	ELMO 6-311+G(2d,2p)
ELMAM2	0.0	1.0	2.1	1.7	1.5
UBDB		0.0	1.3	0.9	0.8
ELMO/6-311G			0.0	0.4	0.5
ELMO/6-311G(d,p)				0.0	0.4
ELMO/6-311+G(2d,2p)					0.0

As already mentioned in the previous section, for the detected hydrogen bond interactions, three other topological properties have been also considered: the kinetic energy density, the potential energy density and the positive curvature of the electron distribution at the BCPs. Their values have been compared to those obtained using the empirical relations (7)-(9) established by Espinosa.

Table 9. Values of the kinetic energy density at the hydrogen bond critical points (in $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-3}$) obtained through the Espinosa empirical relation (see equation (3)) and after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). The experimental distances $\text{H}\cdots\text{O}$ are also reported in \AA .

Hydrogen-Bond Interaction	$d(\text{H}\cdots\text{O})$	Espinosa	Pseudoatoms Transfer		ELMOs Transfer		
			ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 \cdots Ow84	1.920	63.4	55.4	63.9	75.2	70.7	64.2
Ow78-H80 \cdots Ow81	1.892	68.6	54.4	59.5	76.4	70.3	61.5
N58-H74 \cdots O25	2.098	39.0	31.5	36.3	42.1	38.7	36.7
Ow81-H83 \cdots O5	1.931	61.5	48.0	48.8	67.8	61.2	53.0
N38-H57 \cdots O6	1.986	53.0	41.6	49.6	57.9	53.7	50.3

Considering the kinetic energy density (see Table 9), it is immediately possible to observe that all the ELMO values are always greater than the corresponding ELMAM2 and UBDB ones. Now, bearing in mind that, according to Abramov⁹⁹, the kinetic energy density at the bond critical points of closed-shell interactions can be expressed as

$$G(\mathbf{r}_b) = \frac{3}{10} (3\pi^2)^{2/3} \rho^{5/3}(\mathbf{r}_b) + \frac{1}{6} \nabla^2 \rho(\mathbf{r}_b) \quad (11),$$

the trend observed in Table 9 can be mainly related to the fact that the same behavior is similarly detected for the values of the Laplacian (Table 4) and to the fact that, as already mentioned above, the electron density values do not show a large variability (Table 3). Further analyzing Table 9, we can also clearly see that the ELMO charge density reconstructions generally provide the best agreements with the $G(\mathbf{r}_b)$ values resulting from Equation (3). In particular, except for the N31-H37 \cdots Ow84 interaction (for the labels specification see the Leu-enkephalin target geometry deposited in the

Supporting Information), the ELMO/6-311G(d,p) method always gives the smallest discrepancies with respect to Espinosa’s reference values.

Clear trends are less easily detectable in Table 10, where the values of the potential energy density at the hydrogen bond BCPs are shown. However, as for the kinetic energy density, the ELMOs transfer seems quite successful and, especially for the 6-311G(d,p) and 6-311+G(2d,2p) basis-sets, it allows obtaining very good agreements with the corresponding empirical reference values. Finally, in Table 11, we have reported the positive curvatures of the electron distributions at the bond critical points. All the methods give values that are very close to the Espinosa ones, but, unlike the kinetic and the potential energy densities just considered above, in this case, the transfers of pseudo-atoms (namely, both ELMAM2 and UBDB) generally provide smaller discrepancies compared to the transfers of ELMOs.

Table 10. Values of the potential energy density at the hydrogen bond critical points (in $\text{kJ}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-3}$) obtained through the Espinosa empirical relation (see equation (4)) and after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). The experimental distances $\text{H}\cdots\text{O}$ are also reported in \AA .

Hydrogen-Bond Interaction	d(H \cdots O)	Espinosa	Pseudoatoms Transfer		ELMOs Transfer		
			ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 \cdots Ow84	1.920	-49.8	-57.4	-61.6	-66.1	-60.8	-55.0
Ow78-H80 \cdots Ow81	1.892	-55.2	-70.0	-55.7	-67.3	-59.2	-52.5
N58-H74 \cdots O25	2.098	-26.3	-32.2	-34.2	-33.4	-29.9	-29.0
Ow81-H83 \cdots O5	1.931	-47.8	-62.9	-45.3	-57.5	-49.0	-43.4
N38-H57 \cdots O6	1.986	-39.3	-44.1	-48.3	-47.7	-43.1	-40.9

Table 11. Values of the positive curvature λ_3 of the electron density at the hydrogen bond critical points (in $e/\text{\AA}^5$) obtained through the Espinosa empirical relation (see equation (5)) and after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). The experimental distances H \cdots O are also reported in \AA .

Hydrogen-Bond Interaction	d(H \cdots O)	Espinosa	Pseudoatoms Transfer		ELMOs Transfer		
			ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
N31-H37 \cdots Ow84	1.920	-49.8	-57.4	-61.6	-66.1	-60.8	-55.0
Ow78-H80 \cdots Ow81	1.892	-55.1	-70.0	-55.7	-67.3	-59.2	-52.5
N58-H74 \cdots O25	2.098	-26.3	-32.2	-34.1	-33.4	-29.9	-29.0
Ow81-H83 \cdots O5	1.931	-47.8	-62.9	-45.2	-57.5	-49.0	-43.4
N38-H57 \cdots O6	1.986	-39.3	-44.1	-48.3	-47.7	-43.1	-40.9

To further investigate the ELMOs and the pseudo-atoms transferability, we have also taken into account the net integrated atomic charges obtained through the integration of the different electron distributions over the QTAIM atomic basins. In this case, rather than considering the simple atomic charges, the global charges for the main functional groups of the polypeptide backbone and for the five side-chains have been compared.

The inspection of Table 12 shows that the global electroneutrality of the polypeptide is conserved for each considered method. However, it is worth pointing out that, while no electroneutrality corrections are imposed after the transfer of the ELMOs, in the case of the pseudo-atoms transfer the global molecular neutrality is preserved through a charge redistribution over all the atoms of the molecule (see the Computational Details Section), even if the charge redistribution per atom is usually quite small and decreases with the dimension of the system.

From Table 12, we can also observe that the obtained charges always agree with the expected electrostatic nature of the corresponding fragments. For instance, in all the cases, the N-terminal ammonium and the C-terminal carboxylate groups of the polypeptide exhibit a positive and a negative global charge, respectively. Furthermore, while the peptide bonds show a consistent negative charge among all the methods, all the hydrophobic side-chains of the pentapeptide are positively charged. Nevertheless, from a more detailed analysis it seems that the ELMOs provide quite systematic overestimations. In fact, for almost all the fragments taken into account, the global charges obtained transferring the ELMOs are more positive or more negative compared to the corresponding ELMAM2 and the UBDB values, with the two largest discrepancies observed for the two glycine $C\alpha H_2$ moieties. These differences are probably mainly due to the strong localization of the molecular orbitals and, secondly, to the lack of charge relaxation after the transfer of the ELMOs. If the former drawback is unavoidable to guarantee the orbitals transferability, the latter could be partially overcome in the future developing and applying an automatic procedure that allows a sort of charge redistribution after the ELMOs transfer taking into account the environment of each orbital.

Table 12. Net QTAIM integrated charges (in e) for the backbone “functional groups” and for the five side-chains (including the C α -H α bonds) of the Leu-enkephalin polypeptide. The water molecules charges and the global molecular charge of the system after the transfers are also shown. R refers to the generic side chains of the amino acids.

Fragment	Pseudoatoms Transfer		ELMOs Transfer		
	ELMAM2	UBDB	6-311G	6-311G(d,p)	6-311+G(2d,2p)
Ammonium (N-term)	0.244	0.401	0.367	0.361	0.311
H-N-C=O (Tyr-Gly)	-0.483	-0.418	-0.489	-0.604	-0.515
H-N-C=O (Gly-Gly)	-0.448	-0.430	-0.531	-0.660	-0.557
H-N-C=O (Gly-Phe)	-0.434	-0.426	-0.594	-0.726	-0.622
H-N-C=O (Phe-Leu)	-0.435	-0.422	-0.550	-0.659	-0.570
Carboxylate (C-term)	-0.695	-0.673	-0.992	-0.964	-0.972
C α H α -R (Tyr)	0.321	0.395	0.697	0.687	0.739
C α H $_2$ (Gly 1)	0.391	0.257	0.526	0.655	0.547
C α H $_2$ (Gly 2)	0.386	0.250	0.588	0.725	0.622
C α H α -R (Phe)	0.501	0.563	0.572	0.690	0.583
C α H α -R (Leu)	0.651	0.530	0.438	0.535	0.452
Water 1	0.031	0.010	-0.008	-0.005	-0.010
Water 2	-0.002	-0.005	-0.002	-0.002	0.001
Water 3	-0.025	-0.026	-0.015	-0.018	-0.005
Global	0.004	0.006	0.007	0.015	0.005

To evaluate the similarities between the obtained topological integrated charges, we have also evaluated the agreement index R (see equation (2)) for each couple of methods used to reconstruct the pentapeptide electron density. To accomplish this task, we have considered the net integrated QTAIM charges obtained for the backbone “functional groups” and for the five side-chains of the Leu-enkephalin polypeptide (see Table 12). The results have been reported in Table 13. As already

seen for the topological properties at the non covalent BCPs, also in this case, it is easy to observe that the agreement is better if the comparison is conducted within the pseudoatoms or the ELMO techniques (for instance, ELMAM2 vs. UBDB or ELMO/6-311G vs. ELMO/6-311G(d,p), respectively), while it is smaller comparing results associated with methods belonging to different transfer “philosophies” (e.g., ELMAM2 vs. ELMO/6-311G). Therefore, this can be considered as a confirmation of what we have already observed in Table 12, with the ELMO charges generally having a larger magnitude compared to the ones associated with the pseudoatoms electron densities.

Table 13. Values of the agreement index R (%) between the different reconstruction methods (ELMAM2, UBDB and ELMO with basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)) considering as topological property the net QTAIM integrated charges obtained for the backbone “functional groups” and for the five side-chains of the Leu-enkephalin polypeptide.

	ELMAM2	UBDB	ELMO		
			6-311G	6-311G(d,p)	6-311+G(2d,2p)
ELMAM2	0.0	8.3	15.7	20.5	16.5
UBDB		0.0	16.5	21.5	18.3
ELMO/6-311G			0.0	7.4	2.3
ELMO/6-311G(d,p)				0.0	6.5
ELMO/6-311+G(2d,2p)					0.0

Finally, to provide more global comparisons, we have computed the values of the RSR and Walker-Mezey similarity indexes (see the Computational Details Section) between the different charge distributions.

At first, we have considered the ELMAM2 electron density as a reference. As expected, in Table 14 it is easy to observe that, for the RSR and the global $L(0.001)$,

1000) indexes, the electron distributions associated with the ELMO method are the least similar to the ELMAM2 one. Furthermore, considering other charge density ranges, namely modifying the values a and a' in the Walker-Mezey index, we have been able to compare the electron densities in different regions. In particular, we have noted that, while in core and covalent domains (namely, $L(3, 1000)$ and $L(1, 3)$ indexes, respectively) the similarity between the ELMO methods and ELMAM2 remains quite significant, in regions far from the nuclei it decreases (see the indexes $L(0.01, 1)$ and $L(0.001, 0.01)$). However, it is worth reminding that the Walker-Mezey indicator measures the similarity in a relative sense and, since in domains far from the nuclei, the electron density values are very small, quite large relative differences between the charge distributions can be certainly observed even if these discrepancies are nearly negligible in magnitude. For the sake of completeness, at a later stage, we have also determined the values of the global similarity index RSR between all the possible couples of reconstructed electron densities. The obtained results are collected in Table 15 where we can easily confirm the main trend already observed for the agreement index R of the net integrated charges, namely a larger similarity between charge distributions resulting from similar transfer methods and a reduced likeness when comparing electron densities obtained transferring pseudoatoms to the electron distributions obtained transferring ELMOs. Moreover, it is easy to note that, compared to ELMAM2, the UBDB library provides a charge distribution that is closer to the ELMO ones. This is probably due to the fact that the structure factors used to construct the multipole model-based UBDB databank originate from quantum mechanical computations. Finally, an expected significant global similarity between the ELMO/6-311G(d,p) and the ELMO/6-311+G(2d,2p) electron densities is also observed.

Table 14. Point-by-point comparison of the UBDB, ELMO/6-311G, ELMO/6-311G(d,p) and ELMO/6-311+G(2d,2p) electron densities with the ELMAM2 charge distribution: values of the Real-Space R and of the Walker-Mezey similarity $L(a, a')$ indexes (%). For the Walker-Mezey indicator, the electron density is compared within the a and a' limits that are expressed in $e/\text{\AA}^3$.

Similarity Index	UBDB	ELMO		
		6-311G	6-311G(d,p)	6-311+G(2d,2p)
RSR	1.3	2.3	2.0	2.0
$L(0.001,1000)$	91.4	83.1	83.2	82.8
$L(0.001,0.01)$	87.0	71.1	71.3	71.3
$L(0.01,1)$	93.4	88.2	88.2	87.9
$L(1,3)$	97.7	96.2	96.6	96.4
$L(3,1000)$	97.9	96.2	97.2	97.4

Table 15. Point-by-point comparison of the ELMAM2, UBDB, ELMO/6-311G, ELMO/6-311G(d,p) and ELMO/6-311+G(2d,2p) electron densities: values of the Real-Space R similarity index (%).

	ELMAM2	UBDB	ELMO 6-311G	ELMO 6-311G(d,p)	ELMO 6-311+G(2d,2p)
ELMAM2	0.0	1.3	2.3	2.0	2.0
UBDB		0.0	1.9	1.7	1.7
ELMO/6-311G			0.0	1.3	1.5
ELMO/6-311G(d,p)				0.0	0.6
ELMO/6-311+G(2d,2p)					0.0

Conclusions

Due to their extreme localization on small molecular fragments as atoms, bonds or functional groups, the ELMOs are easily transferable orbitals from a molecule to another. For this reason, they can be considered as a possible alternative/complement

to the pseudoatoms of the multipole models, whose reliable exportability has been largely exploited both for reconstructing the electron densities and for refining the crystallographic structures of large molecular systems. In this context, the present paper can be considered as a further preliminary study to the construction of new libraries of Extremely Localized Molecular Orbitals and its main goal was the comparison of the ELMOs and the pseudoatoms transferability.

To accomplish this task, we have thoroughly analyzed the charge distributions obtained transferring both pseudoatoms (from the experimental ELMAM2 and the theoretical UBDB databanks) and Extremely Localized Molecular Orbitals (from suitably constructed model molecules) to the crystallographic structure of the relatively large polypeptide Leu-enkephalin. The results show that, although the unavoidable and expected presence of discrepancies, all the different reconstruction methods provide completely reasonable and similar electron densities. In particular, we have noted that the ELMOs transfer is quite successful in reproducing the topological properties at the non-covalent bond critical points. Nevertheless, it has been observed that the use of Extremely Localized Molecular Orbitals generally leads to an overestimation of the integrated charges associated with the subunits of the system. However, as already mentioned, this drawback, which is closely related to the strictly localized nature of the orbitals, might be overcome through a proper method to relax the transferred ELMOs taking into account the environment of the fragment on which they are localized. Although possible solutions in this direction have been already proposed,^{81,82,100} new linear scaling strategies to accomplish this task are desirable in the near future. In particular, we currently envisage the development of an original technique that allows the perturbation of the transferred ELMOs taking into

account the chemical environment of the target system and keeping as much as possible the localized nature of the orbitals.

In conclusion, the outcomes of the present study have shown that the transferability of the Extremely Localized Molecular Orbitals is as reliable as the one of the pseudoatoms and that it enables to obtain completely acceptable reconstructions of electron distributions. Therefore, this result, along with the recent validation of the ELMOs transferability to large systems, further encourages the construction of new libraries of Extremely Localized Molecular Orbitals with the final purpose of refining crystallographic structures and computing approximate properties of 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. Details about the theory of the Extremely Localized Molecular Orbitals. Tables S1-S2: standard deviations associated with the average values of the electron density and of its Laplacian at the covalent bond critical points after the transfer of pseudoatoms (ELMAM2 and UBDB databanks) and ELMOs (basis-sets 6-311G, 6-311G(d,p) and 6-311+G(2d,2p)). Tables S3-S11: topological properties associated with the electron distributions obtained after the transfer of ELMOs (basis-sets 6-31G, 6-31G(d,p) and 6-311+G(d,p)). Table S12: net integrated charges associated with the electron distributions resulting from the transfer of ELMOs (basis-sets 6-31G, 6-31G(d,p) and

6-311+G(d,p)). Tables S13-S17: values of the agreement index R for the different topological properties taken into account and considering the ELMAM2, UBDB, ELMO/6-31G, ELMO/6-31G(d,p) and ELMO/6-311+G(d,p) methods. Tables S18-S19: values of the similarity indexes (RSR and Walker-Mezey) corresponding to the comparison of the ELMAM2, UBDB, ELMO/6-31G, ELMO/6-31G(d,p) and ELMO/6-311+G(d,p) charge distributions. Figure S1: sectional view of ELMAM2 electron density isosurfaces corresponding to values selected to compute the Walter-Mezey similarity indexes. Figures S2-S7: kinetic energy densities, potential energy densities and positive curvatures of the electron density at the hydrogen bond critical points. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

References

1. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev. B* **1964**, *136*, 864–871.
2. Coppens, P. *X-Ray Charge Densities and Chemical Bonding*; Oxford University Press: New York, 1997.
3. Koritsanszky, T. S.; Coppens, P. Chemical Applications of X-ray Charge-Density Analysis. *Chem. Rev.* **2001**, *101*, 1583–1628.
4. Gatti, C.; Macchi P. *Modern Charge-Density Analysis*; Springer: Heidelberg, 2012.
5. Clinton, W. L.; Massa, L. J. Determination of the Electron Density Matrix from X-ray Diffraction Data. *Phys. Rev. Lett.* **1972**, *29*, 1363–1366.
6. Clinton, W. L.; Frishberg, C. A.; Massa, L. J.; Oldfield, P. A. Methods for Obtaining an Electron-Density Matrix from X-ray Diffraction Data. *Int. J. Quantum Chem. Symp.* **1973**, *7*, 505–514.

7. Frishberg, C.; Massa, L. J. Idempotent Density Matrices for Correlated Systems from X-ray Diffraction Structure Factors. *Phys. Rev. B* **1981**, *24*, 7018–7024.
8. Tanaka, K. X-ray Analysis of Wavefunctions by the Least-Squares Method Incorporating Orthonormality. I. General Formalism. *Acta Crystallogr., Sect. A* **1988**, *44*, 1002–1008.
9. Howard, S. T.; Huke, J. P.; Mallinson, P. R.; Frampton, C. S. Density Matrix Refinement for Molecular Crystals. *Phys. Rev. B* **1994**, *49*, 7124–7136.
10. Snyder, J. A.; Stevens, E. D. A Wavefunction and Energy of the Azide Ion in Potassium Azide Obtained by a Quantum-Mechanically Constrained Fit to X-ray Diffraction Data. *Chem. Phys. Lett.* **1999**, *313*, 293–298.
11. Hibbs, D. E.; Howard, S. T.; Huke, J. P.; Waller, M. P. A New Orbital-Based Model for the Analysis of Experimental Molecular Charge Densities: An Application to (Z)-N-Methyl-C-phenylnitrene. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1772–1778.
12. Waller, M. P.; Howard, S. T.; Platts, J. A.; Piltz, R. O.; Willock, D. J.; Hibbs, D. E. Novel Properties from Experimental Charge Densities: An Application to the Zwitterionic Neurotransmitter Taurine. *Chem.-Eur. J.* **2006**, *12*, 7603–7614.
13. Jayatilaka, D. Wave Function for Beryllium from X-ray Diffraction Data. *Phys. Rev. Lett.* **1998**, *80*, 798–801.
14. Jayatilaka, D.; Grimwood, D. J. Wavefunctions Derived from Experiment. I. Motivation and Theory. *Acta Crystallogr., Sect. A* **2001**, *57*, 76–86.
15. Grimwood, D. J.; Jayatilaka, D. Wavefunctions Derived from Experiment. II. A Wavefunction for Oxalic Acid Dihydrate. *Acta Crystallogr., Sect. A* **2001**,

- 57, 87–100.
16. Hudák, M.; Jayatilaka, D.; Peráinova, 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.
 17. Genoni, A. Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data, *J. Phys. Chem. Lett.* **2013**, *4*, 1093–1099.
 18. 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.
 19. 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.
 20. 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.
 21. Roversi, P.; Irwin, J. J.; Bricogne, G. Accurate Charge-Density Studies as an Extension of Bayesian Crystal Structure Determination, *Acta Crystallogr., Sect. A* **1998**, *54*, 971–996.
 22. Hirshfeld, F. L. Difference Densities by Least-Squares Refinement: Fumaramic Acid. *Acta Crystallogr., Sect. B* **1971**, *27*, 769–781.
 23. Stewart, R. F. Electron Population Analysis with Rigid Pseudoatoms. *Acta Crystallogr., Sect. A* **1976**, *32*, 565–574.
 24. Hansen, N. K.; Coppens, P. Testing Aspherical Atom Refinements on Small-Molecule Data Sets. *Acta Crystallogr., Sect. A* **1978**, *34*, 909–921.

25. 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.
26. Jelsch, C.; Teeter, M. M.; Lamzin, V.; Pichon-Pesme, V.; Blessing, R. H.; Lecomte, C. Accurate Protein Crystallography at Ultra-High Resolution: Valence Electron Distribution in Crambin. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3171-3176.
27. Housset, D.; Benabicha, F.; Pichon-Pesme, V.; Jelsch, C.; Maierhofer, A.; David, S.; Fontecilla-Camps, J. C.; Lecomte, C. Towards the Charge-Density Study of Proteins: a Room-Temperature Scorpion-Toxin Structure at 0.96 Å Resolution as a First Test Case. *Acta Crystallogr., Sect. D* **2000**, *56*, 151-160.
28. Lecomte, C.; Guillot, B.; Muzet, N.; Pichon-Pesme, V.; Jelsch, C. Ultra-High-Resolution X-ray Structure of Proteins. *Cell. Mol. Life Sci.* **2004**, *61*, 774-782.
29. Afonine, P. V.; Grosse-Kunstleve, R. W.; Adams, P. D.; Lunin, V. Y.; Urzhumtsev, A. On Macromolecular Refinement at Subatomic Resolution with Interatomic Scatterers. *Acta Crystallogr., Sect. D* **2007**, *63*, 1194-1197.
30. Dittrich, B.; Koritsánszky, T.; Volkov, A.; Mebs, S.; Luger, P. Novel Approaches to the Experimental Charge Density of Vitamin B₁₂. *Angew. Chem. Int. Ed.* **2007**, *46*, 2935-2938.
31. Guillot, B.; Jelsch, C.; Podjarny, A.; Lecomte, C. Charge-Density Analysis of a Protein Structure at Subatomic Resolution: the Human Aldose Reductase Case. *Acta Crystallogr., Sect. D* **2008**, *64*, 567-588.
32. Johnas, S. K. J.; Dittrich, B.; Meents, A.; Messerschmidt, M.; Weckert, E. F. Charge-Density Study on Cyclosporine A. *Acta Crystallogr., Sect. D* **2009**,

- 65, 284-293.
33. Brock, C. P.; Dunitz, J. D.; Hirshfeld, F. L. Transferability of Deformation Densities Among Related Molecules: Atomic Multipole Parameters from Perylene for Improved Estimation of Molecular Vibrations in Naphthalene and Anthracene. *Acta Crystallogr., Sect. B* **1991**, *47*, 789–797.
 34. 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.
 35. 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.
 36. Lecomte, C.; Jelsch, C.; Guillot, B.; Fournier, B.; Lagoutte, A. Ultrahigh-Resolution Crystallography and Related Electron Density and Electrostatic Properties in Proteins. *J. Synchrotron Rad.* **2008**, *15*, 202-203.
 37. Domagała, S.; Munshi, P.; Ahmed, M.; Guillot, B.; Jelsch, C. Structural Analysis and Multipole Modelling of Quercetin Monohydrate. A Quantitative and Comparative Study. *Acta Crystallogr., Sect. B* **2011**, *67*, 63-78.
 38. Domagała, S.; Fournier, B.; Liebschner, D.; Guillot, B.; Jelsch, C. An Improved Experimental Databank of Transferable Multipolar Atom Models – ELMAM2. Construction Details and Applications. *Acta Crystallogr., Sect. A* **2012**, *68*, 337-351.
 39. 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.
40. 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.
41. Li, X.; Volkov, A. V.; Szalewicz, K.; Coppens, P. Interaction Energies between Glycopeptide Antibiotics and Substrates in Complexes Determined by X-ray Crystallography: Application of a Theoretical Databank of Aspherical Atoms and a Symmetry-Adapted Perturbation Theory-Based Set of Interatomic Potentials. *Acta Crystallogr., Sect. D* **2006**, *62*, 639-647.
42. 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.
43. Volkov, A.; Messerschmidt, M.; Coppens, P. Improving the Scattering-Factor Formalism in Protein Refinement: Application of the University at Buffalo Aspherical-Atom Databank to Polypeptide Structures. *Acta Crystallogr., Sect. D* **2007**, *63*, 160-170.
44. Dittrich, B.; Koritsanszky, T.; Luger, P. A Simple Approach to Nonspherical Electron Densities by Using Invarioms. *Angew. Chem., Int. Ed.* **2004**, *43*, 2718–2721.
45. 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.

46. 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.
47. Dittrich, B.; Hübschle, C. B.; Holstein, J. J.; Fabbiani, F. P. A. Towards Extracting the Charge Density from Normal-Resolution Data. *J. Appl. Cryst.* **2009**, *42*, 1110-1121.
48. 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.
49. Pichon-Pesme, V.; Jelsch, C.; Guillot, B.; Lecomte, C. A Comparison between Experimental and Theoretical Aspherical-Atom Scattering Factors for Charge-Density Refinement of Large Molecules. *Acta Crystallogr., Sect. A* **2004**, *60*, 204-208.
50. Volkov, A.; Koritsanszky, T.; Li, X.; Coppens, P. Response to the Paper A Comparison between Experimental and Theoretical Aspherical-Atom Scattering Factors for Charge-Density Refinement of Large Molecules, by Pichon-Pesme, Jelsch, Guillot & Lecomte (2004). *Acta Crystallogr., Sect. A* **2004**, *60*, 638-639.
51. Bąk, J. M.; Domagała, S.; Hübschle, C.; Jelsch, C.; Dittrich, B.; Dominiak, P. M. Verification of Structural and Electrostatic Properties Obtained by the Use of Different Pseudoatom Databases. *Acta Crystallogr., Sect. A* **2011**, *67*, 141-153.
52. Hathwar, V. R.; Thakur, T. S.; Dubey, R.; Pavan, M. S.; Row, T. N. G.; Desiraju, G. R. Extending the Supramolecular Synthon Based Fragment Approach (SBFA) for Transferability of Multipole Charge Density Parameters

- to Monofluorobenzoic Acids and their Cocrystals with Isonicotinamide: Importance of C-H \cdots O, C-H \cdots F, and F \cdots F Intermolecular Regions. *J. Phys. Chem. A* **2011**, *115*, 12852-12863.
53. Hathwar, V. R.; Thakur, T. S.; Row, T. N. G.; Desiraju, G. R. Transferability of Multipole Charge Density Parameters for Supramolecular Synthons: A New Tool for Quantitative Crystal Engineering. *Cryst. Growth Des.* **2011**, *11*, 616-623.
54. Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theor. Chim. Acta* **1980**, *57*, 169–178.
55. 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.
56. Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685-698.
57. Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085-1123.
58. Merz, K. M., Jr. Using Quantum Mechanical Approaches to Study Biological Systems. *Acc. Chem. Res.* **2014**, *47*, 2804-2811.
59. Yang, W. Direct calculation of electron density in density-functional theory. *Phys. Rev. Lett.* **1991**, *66*, 1438-1441.
60. He, X.; Merz, K. M., Jr. Divide and Conquer Hartree-Fock Calculations on Proteins. *J. Chem. Theory Comput.* **2010**, *6*, 405-411.
61. Walker, P. D.; Mezey, P. G. Molecular electron density Lego approach to

- molecule building. *J. Am. Chem. Soc.* **1993**, *115*, 12423-12430.
62. Walker, P. D.; Mezey, P. G. Ab Initio Quality Electron Densities for Proteins: A MEDLA Approach. *J. Am. Chem. Soc.* **1994**, *116*, 12022-12032.
63. 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.
64. Exner, T. E.; Mezey, P. G. Ab initio quality properties for macromolecules using the ADMA approach *J. Comput. Chem.* **2003**, *24*, 1980-1986.
65. 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.
66. 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.
67. 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.
68. Huang, L.; Massa, L.; Karle, J. Kernel energy method applied to vesicular stomatitis virus nucleoprotein. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 1731-1736.
69. Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to DNA. *Biochemistry* **2005**, *44*, 16747-16752.
70. Huang, L.; Massa, L.; Karle, J. Kernel energy method: Application to insulin. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 12690-12693.

71. 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.
72. Huang, L.; Massa, L.; Matta, C. F. A graphene flake under external electric fields reconstructed from field-perturbed kernels. *Carbon* **2014**, *76*, 310-320.
73. 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.
74. 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.
75. 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.
76. 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.
77. Chang, C.; Bader, R. F. W. Theoretical construction of a polypeptide. *J. Phys. Chem.* **1992**, *96*, 1654-1662.
78. Bader, R. F. W.; Martín, F. J. Interdeterminacy of basin and surface properties of an open system. *Can. J. Chem.* **1998**, *76*, 284-291.
79. 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.
80. Scheins, S; Messerschmidt, M.; Luger, P. Submolecular partitioning of morphine hydrate based on its experimental charge density at 25 K. *Acta Crystallogr., Sect. B* **2005**, *61*, 443-448.
81. Genoni, A; Sironi, M. A Novel Approach to Relax Extremely Localized Molecular Orbitals: the Extremely Localized Molecular Orbital-Valence Bond Method. *Theor. Chem. Acc.* **2004**, *112*, 254-262.
82. Genoni, A.; Fornili, A.; Sironi, M. Optimal Virtual Orbitals to Relax Eave Functions Built Up with Transferred Extremely Localized Molecular Orbitals. *J. Comp. Chem.* **2005**, *26*, 827-835.
83. Meyer, B.; Guillot, B.; Ruiz-Lopez, M. F.; Genoni, A. Libraries of Extremely Localized Molecular Orbitals. I. Definition of the Model Molecules Approximation and Assessment of the Molecular Orbitals Transferability. *J. Chem. Theory Comput.*, submitted.
84. 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.
85. 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.
86. Domagała, S.; Jelsch, C. Optimal Local Axes and Symmetry Assignment for Charge Density Refinements. *J. Appl. Crystallogr.* **2008**, *41*, 1140-1149.
87. 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.
88. 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.
89. Guillot, B.; Viry, L.; Guillot, R.; Lecomte, C.; Jelsch, C. Refinement of Proteins at Subatomic Resolution with MoPro. *J. Appl. Crystallogr.* **2001**, *34*, 214-223.
90. Jelsch, C.; Guillot, B.; Lagoutte, A.; Lecomte, C. Advances in Protein and Small Molecules Charge Density Refinement Methods, Using Software MoPro. *J. Appl. Crystallogr.* **2005**, *38*, 38-54.
91. 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.
92. Bader, R. F. W. *Atoms in Molecules: A Quantum Theory*; Oxford University Press: Oxford, 1990.
93. Keith, T. A. AIMAll, Version 13.11.04; TK Gristmill Software: Overland Park, KS, 2012; available online: aim.tkgristmill.com.
94. Espinosa, E.; Molins, E.; Lecomte, C. Hydrogen Bond Strengths Revealed by Topological analyses of Experimentally Observed Electron Densities. *Chem. Phys. Lett.* **1998**, *285*, 170-173.
95. Espinosa, E.; Souhassou, M.; Lachekar, H.; Lecomte, C. Topological Analysis of the Electron Density in Hydrogen Bonds. *Acta Crystallogr., Sect. B* **1999**, *55*, 563-572.
96. Spackman, M. A. How Reliable Are the Intermolecular Interaction Energies

- Estimated form Topological Analysis of Experimental Electron Densities?
Cryst. Growth Des. **2015**, *15*, 5624-5628.
97. 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.
98. 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.
99. Abramov, Yu. A. On the Possibility of Kinetic Energy Density Evaluation from the Experimental Electron-Density Distribution. *Acta Crystallogr., Sect. A* **1997**, *53*, 264-272.
100. Genoni, A.; Merz, K. M., Jr.; Sironi, M. A Hylleraas Functional Based Perturbative Technique to Relax the Extremely Localized Molecular Orbital

Wavefunction. *J. Chem. Phys.* **2008**, *129*, 054101.