



HAL
open science

Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data

Alessandro Genoni

► **To cite this version:**

Alessandro Genoni. Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-ray Diffraction Data. *Journal of Physical Chemistry Letters*, 2013, 4 (7), pp.1093-1099. 10.1021/jz400257n . hal-02196463

HAL Id: hal-02196463

<https://hal.univ-lorraine.fr/hal-02196463>

Submitted on 10 May 2020

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

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

This document is the Accepted Manuscript version of a Published Work that appeared in final form in *The Journal of Physical Chemistry Letters*, 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/jz400257n>.

Molecular Orbitals Strictly Localized on Small Molecular Fragments from X-Ray Diffraction Data

Alessandro Genoni^{*,†,‡}

*CNRS, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France, and Université
de Lorraine, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France*

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

Phone: +33 (0)3 83 68 43 77. Fax: +33 (0)3 83 68 43 71

*To whom correspondence should be addressed

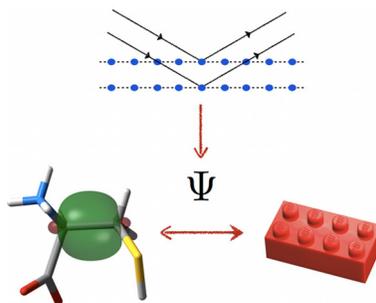
†CNRS

‡Université de Lorraine

Abstract

Nowadays, the electron density is recognized as a fundamental property that contains most of the information concerning the electronic structure of molecules and, therefore, its determination from high-resolution X-Ray diffraction data is becoming more and more important. In this context we propose a new strategy for the charge density analysis, strategy in which the chemical interpretability of the multipole model is combined with the quantum mechanical rigor of the wavefunction-based approaches. In particular, the novel technique aims at extracting Molecular Orbitals strictly localized on small molecular fragments (e.g., atoms, bonds or functional groups) from a set of measured structure factors amplitudes. Preliminary tests have shown that their determination is really straightforward and, given their reliable transferability, we envisage the possibility of constructing new Extremely Localized Molecular Orbitals databases as alternative to the existing pseudoatoms libraries.

TOC Graphic



Keywords: Extremely Localized Molecular Orbitals, experimental wavefunction, electron density, transferability, quantum chemistry, X-Ray diffraction.

It is well known that the electron density is of paramount importance to fully define the ground state properties of many-electron systems. Its prominence derives from the Hohenberg & Kohn theorem¹ which has shown the existence of a functional relation between the ground state electron distribution and the ground state wave function of electronic systems. This is one of the reasons why the determination of the electron density from high-resolution X-ray diffraction measurements has become a mature field of science over the years.^{2,3}

In this context, the multipole model⁴⁻⁶ is by far the most widely used technique and it consists in fitting the X-Ray diffraction data to a model electron distribution written as a sum of aspherical atomic density functions (also known as pseudoatoms) depending on proper refinable parameters. Different features have made this strategy really appealing. For instance, we can mention its linear scaling behavior and the generally high observations/parameters ratio which leads to unique and rapidly-convergent fits. Moreover, the success of this approach is also deeply associated with the fact that it allows easy chemical interpretations of the molecular electron densities. In fact, due to the intrinsic atomic partitioning in the pseudoatom definition, the charge density of a molecule can be simply seen as the sum of atomic electron distributions that are deformed by the presence of bonding and non-bonding interactions. This important peculiarity, along with the plausible idea that chemically equivalent atoms (namely, identical atoms with a similar local connectivity) provide identical contributions to the electron density, has prompted different research groups to study the transferability of pseudoatoms from small molecules to large biomolecules with identical subunits.^{7,8} The results of these investigations have led to the construction of pseudoatoms libraries, obtained either from experimental electron densities⁸⁻¹⁰ or from theoretical electron distributions,¹¹⁻¹⁶ that are currently used to refine charge densities of macromolecules and to predict electrostatic properties of large systems.

Nevertheless, notwithstanding its great success, the multipole model exhibits some drawbacks: a) the fitted electron density may be characterized by unphysical negative regions; b) the refinable coefficients in the multipole expansion may have a strong correlation with the Anisotropic Displacement Parameters (ADPs); c) the number of properties that can be computed directly from the

fitted charge distribution is limited because the exact functional relation between the ground state electron density and the ground state wave function is generally unknown. All these problems can be solved by another family of strategies: the so called wavefunction- or orbital- based methods for the electron density determination and analysis. Pioneered by Clinton and coworkers,¹⁷⁻²⁴ these techniques consist in fitting X-ray scattering data to an assumed form for the density matrix²³⁻²⁹ or the wave function³⁰⁻³⁷ of the system under investigation and provide a two-center orbital product-based representation of the electron distribution. Therefore, unlike the multipole model, the wave function-based strategies allow to obtain electron densities with intrinsic quantum mechanical features, but, unfortunately, this results into the loss of the easy chemical interpretability of the charge distribution, which is one of the strengths of the multipole model.

In this letter we present a new technique for the determination of molecular electron densities, technique in which we combine the quantum mechanical rigor of the orbital-based approaches with the ease of chemical interpretation typical of the multipole model. In particular, the new strategy aims at extracting from X-Ray diffraction data a single Slater determinant built up with Molecular Orbitals strictly localized (namely, without tails) on molecular fragments (e.g., atoms, bonds or functional groups) and it can be seen as a combination of the experimental wave function approach proposed by Jayatilaka³¹⁻³⁷ with the method developed by Stoll for the *a priori* determination of Extremely Localized Molecular Orbitals (ELMOs).³⁸

As in the Jayatilaka method, we assume that each crystal unit does not interact with the other ones and we write the wave function for the whole crystal like this:

$$\Psi_{cryst} = \prod_k \Psi_k \quad (1)$$

where all the unit wave functions Ψ_k are formally identical and related to each other through the crystal symmetry operations. The further approximation of our approach consists in describing each crystal unit with an ELMO wave function, namely with a single Slater determinant built up with Extremely Localized Molecular Orbitals. For the sake of simplicity, let us consider the k -th

crystal unit and let us subdivide it into f fragments (i.e., localization centers such as atoms, bonds or functional groups), each one characterized by a proper number of ELMOs expanded on a local basis-set consisting only of the basis functions centered on the atoms belonging to the fragment. Therefore, if for the i -th fragment the local basis-set of M_i^k atomic orbitals is $\left\{|\chi_{i\mu}^k\rangle\right\}_{\mu=1}^{M_i^k}$, the α -th ELMO of the fragment can be expressed as

$$|\varphi_{i\alpha}^k\rangle = \sum_{\mu=1}^{M_i^k} C_{i\mu,\alpha}^k |\chi_{i\mu}^k\rangle \quad (2)$$

and the ELMO wavefunction for the k -th crystal unit (assuming it as a closed-shell system) as

$$\Psi_k = \hat{A} \left[\varphi_{11}^k \bar{\varphi}_{11}^k \cdots \varphi_{1n_1}^k \bar{\varphi}_{1n_1}^k \cdots \varphi_{f1}^k \bar{\varphi}_{f1}^k \cdots \varphi_{fn_f}^k \bar{\varphi}_{fn_f}^k \right] \quad (3)$$

where \hat{A} is the antisymmetrizer, n_i is the number of occupied ELMOs for the i -th fragment and $\bar{\varphi}_{i\alpha}^k$ is a spinorbital with spatial part $\varphi_{i\alpha}^k$ and spin part β .

At this stage we consider only the case in which each non-interacting unit corresponds to a symmetry-unique portion of the crystal unit-cell. This allows to write the unit-cell charge density as a sum of N_m crystal unit electron densities $\rho_j(\mathbf{r})$ that are related to a reference electron distribution $\rho_0(\mathbf{r})$ by the unit-cell symmetry operations $\{\mathbf{R}_j, \mathbf{r}_j\}$:

$$\rho_{cell}(\mathbf{r}) = \sum_{j=1}^{N_m} \rho_j(\mathbf{r}) = \sum_{j=1}^{N_m} \rho_0(\mathbf{R}_j^{-1}(\mathbf{r} - \mathbf{r}_j)) \quad (4)$$

Eq. (4) is exact provided that $\rho_0(\mathbf{r})$ is not obtained through an isolated crystal unit calculation. Nevertheless, in our new method this condition is not fulfilled because $\rho_0(\mathbf{r})$ is the electron density associated with the ELMO wave function Ψ_0 for the reference unit, wave function which is obtained looking for those Extremely Localized Molecular Orbitals that not only minimize the ELMO energy of the reference crystal unit, but that also reproduce as much as possible the set of collected experimental structure factors amplitudes $\{|F_{\mathbf{h}}^{exp}|\}$. In other words, we look for those

ELMOs that minimize the following functional:

$$J[\boldsymbol{\varphi}] = \langle \Psi_0 | \hat{H}_0 | \Psi_0 \rangle + \lambda (\chi^2 - \Delta) \quad (5)$$

where \hat{H}_0 is the non-relativistic Hamiltonian operator for the reference crystal unit, λ is the Lagrange multiplier associated with the constraint, Δ is the desired agreement between theoretical and experimental values and χ^2 is the measure of the fitting accuracy between the calculated and the observed structure factor amplitudes, namely:

$$\chi^2 = \frac{1}{N_r - N_p} \sum_{\mathbf{h}} \frac{(\eta |F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{exp}|)^2}{\sigma_{\mathbf{h}}^2} \quad (6)$$

with N_r as the number of experimental scattering data, N_p as the number of adjustable parameters (in our case only the Lagrange multiplier λ), \mathbf{h} as the term of Miller indexes labeling the reflection, $\sigma_{\mathbf{h}}$ as the error associated with each measure and η as an overall \mathbf{h} -independent scale factor that is determined to minimize the χ^2 value. Due to the presence of experimental errors in the collected data, it does not make sense to obtain χ^2 equal to zero. Therefore, an optimal value for the desired agreement Δ in Eq. (5) is one, so that the calculated structure factors are on average within one standard deviation of the experimental values. Furthermore, for the sake of completeness, it is worth mentioning that for a proper calculation of the structure factors amplitudes $\{|F_{\mathbf{h}}^{calc}|\}$, including a proper correction for the thermal vibrational effects, we use the same equations proposed by Jayatilaka and coworkers^{31,32} (see Supporting Information for further details), with the only difference that for the Fourier transforms of the basis functions pairs we have implemented an Obara-Saika scheme^{39,40} exploiting both vertical and horizontal recurrence relations.⁴¹

Now, for the sake of clarity, let us omit subscripts and superscripts corresponding to the reference crystal unit and, in order to search the ELMOs that minimize the functional defined by Eq. (5), let us consider the arbitrary variation of $J[\boldsymbol{\varphi}]$ with respect to the occupied ELMO $|\varphi_{j\beta}\rangle$ (from now indicated as $\delta_{(j\beta)}J$). Introducing the global and the local density operators, which, due to the

non-orthogonality of the ELMOs, are respectively given by

$$\hat{\rho} = \sum_{i,j=1}^f \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} [\mathbf{S}^{-1}]_{j\beta,i\alpha} |\varphi_{j\beta}\rangle \langle \varphi_{i\alpha}| \quad (7)$$

and

$$\hat{\rho}_i = \sum_{j=1}^f \sum_{\alpha=1}^{n_i} \sum_{\beta=1}^{n_j} [\mathbf{S}^{-1}]_{j\beta,i\alpha} |\varphi_{j\beta}\rangle \langle \varphi_{i\alpha}| \quad (8)$$

with \mathbf{S} as the overlap matrix of the occupied ELMOs, and defining the structure factor operator

$$\hat{I}_{\mathbf{h}} = \sum_{j=1}^{N_m} e^{i2\pi(\mathbf{R}_j\mathbf{r}+\mathbf{r}_j)\cdot(\mathbf{B}\mathbf{h})} = \hat{I}_{\mathbf{h},R} + i \hat{I}_{\mathbf{h},C} \quad (9)$$

where both $\hat{I}_{\mathbf{h},R}$ and $\hat{I}_{\mathbf{h},C}$ (real and imaginary part of $\hat{I}_{\mathbf{h}}$, respectively) are hermitian, we obtain:

$$\begin{aligned} \delta_{(j\beta)}J &= 4 \left\{ \langle \delta\varphi_{j\beta} | (1-\hat{\rho})\hat{F}(1-\hat{\rho}+\hat{\rho}_j) | \varphi_{j\beta} \rangle + \right. \\ &\quad \left. + \lambda \sum_{\mathbf{h}} K_{\mathbf{h}} \operatorname{Re} \left\{ F_{\mathbf{h}}^{calc} \right\} \langle \delta\varphi_{j\beta} | (1-\hat{\rho})\hat{I}_{\mathbf{h},R}(1-\hat{\rho}+\hat{\rho}_j) | \varphi_{j\beta} \rangle + \right. \\ &\quad \left. + \lambda \sum_{\mathbf{h}} K_{\mathbf{h}} \operatorname{Im} \left\{ F_{\mathbf{h}}^{calc} \right\} \langle \delta\varphi_{j\beta} | (1-\hat{\rho})\hat{I}_{\mathbf{h},C}(1-\hat{\rho}+\hat{\rho}_j) | \varphi_{j\beta} \rangle \right\} \quad (10) \end{aligned}$$

with

$$K_{\mathbf{h}} = \frac{2\eta}{N_r - N_p} \frac{\eta |F_{\mathbf{h}}^{calc}| - |F_{\mathbf{h}}^{exp}|}{\sigma_{\mathbf{h}}^2 |F_{\mathbf{h}}^{calc}|} \quad (11)$$

Since the lowest value of the functional is achieved if $\delta_{(j\beta)}J$ vanishes for all j and β , we have that the ELMOs that minimize J ("experimental" ELMOs) are the ones that satisfy the following equation for each fragment:

$$\begin{aligned} &\left\{ (1-\hat{\rho})\hat{F}(1-\hat{\rho}+\hat{\rho}_j) + \lambda \sum_{\mathbf{h}} K_{\mathbf{h}} \operatorname{Re} \left\{ F_{\mathbf{h}}^{calc} \right\} (1-\hat{\rho})\hat{I}_{\mathbf{h},R}(1-\hat{\rho}+\hat{\rho}_j) \right. \\ &\quad \left. + \lambda \sum_{\mathbf{h}} K_{\mathbf{h}} \operatorname{Im} \left\{ F_{\mathbf{h}}^{calc} \right\} (1-\hat{\rho})\hat{I}_{\mathbf{h},C}(1-\hat{\rho}+\hat{\rho}_j) \right\} | \varphi_{j\beta} \rangle = 0 \quad (12) \end{aligned}$$

Now, adding the following quantities to both hand sides of Eq. (12)

$$|Q1\rangle = \hat{\rho}_j^\dagger \hat{F} (1 - \hat{\rho} + \hat{\rho}_j) |\varphi_{j\beta}\rangle \quad (13)$$

$$|Q2\rangle = \lambda \sum_{\mathbf{h}} K_{\mathbf{h}} \left[\text{Re}\{F_{\mathbf{h}}^{calc}\} \hat{\rho}_j^\dagger \hat{I}_{\mathbf{h},R} (1 - \hat{\rho} + \hat{\rho}_j) + \text{Im}\{F_{\mathbf{h}}^{calc}\} \hat{\rho}_j^\dagger \hat{I}_{\mathbf{h},C} (1 - \hat{\rho} + \hat{\rho}_j) \right] |\varphi_{j\beta}\rangle \quad (14)$$

and defining the hermitian operators

$$\hat{F}^j = (1 - \hat{\rho} + \hat{\rho}_j^\dagger) \hat{F} (1 - \hat{\rho} + \hat{\rho}_j) \quad (15)$$

$$\hat{I}_{\mathbf{h},R}^j = K_{\mathbf{h}} \text{Re}\{F_{\mathbf{h}}^{calc}\} (1 - \hat{\rho} + \hat{\rho}_j^\dagger) \hat{I}_{\mathbf{h},R} (1 - \hat{\rho} + \hat{\rho}_j) \quad (16)$$

$$\hat{I}_{\mathbf{h},C}^j = K_{\mathbf{h}} \text{Im}\{F_{\mathbf{h}}^{calc}\} (1 - \hat{\rho} + \hat{\rho}_j^\dagger) \hat{I}_{\mathbf{h},C} (1 - \hat{\rho} + \hat{\rho}_j) \quad (17)$$

we have

$$\hat{F}^{j,exp} |\varphi_{j\beta}\rangle = \sum_{\gamma=1}^{n_j} \langle \varphi_{j\gamma} | \hat{F}^{j,exp} | \varphi_{j\beta} \rangle |\varphi_{j\gamma}\rangle \quad (18)$$

where $\hat{F}^{j,exp}$, which is the modified Fock operator for the j -th fragment, is given by:

$$\hat{F}^{j,exp} = \hat{F}^j + \lambda \sum_{\mathbf{h}} \hat{I}_{\mathbf{h},R}^j + \hat{I}_{\mathbf{h},C}^j \quad (19)$$

Finally, applying a unitary transformation that only mixes the occupied ELMOs of each fragment among themselves and given the invariance of $\hat{F}^{j,exp}$ to that transformation, Eq. (18) becomes:

$$\hat{F}^{j,exp} |\varphi_{j\beta}\rangle = \varepsilon_{j\beta} |\varphi_{j\beta}\rangle \quad (20)$$

Therefore, the "experimental" ELMOs are obtained solving Eq. (20) self-consistently for each fragment using the proper local basis-set deriving from the localization scheme defined *a priori*. It is important to observe that, although solved separately, the equations associated with the different subunits are coupled because each $\hat{F}^{j,exp}$ operator depends on the global density operator, as it can be seen from Eqs. (15), (16) and (17).

Unfortunately, as in the case of the "theoretical" ELMOs, convergence problems in the resolution of modified Hartree-Fock equations (in our case Eqs. (20)) might arise near the functional minimum^{38,42} and, therefore, to overcome this drawback, we have devised an algorithm which uses the information provided by exact first and approximate second derivatives of J with respect to the ELMOs coefficients (see Supporting Information for their expressions). In particular, following a strategy successfully used by Fornili *et al.* for the calculation of the "theoretical" ELMOs,⁴² we have adopted a quasi-Newton procedure where an approximate analytic Hessian is calculated only at the first iteration and, afterwards, it is updated exploiting the Broyden-Fletcher-Goldfarb-Shanno formula.⁴³

Starting from an existing code for the ELMOs calculation, both the self-consistent resolution of Eqs. (20) and the quasi-Newton algorithm just described above have been implemented modifying the version 8 of the GAMESS-UK package,⁴⁴ which has been also used to perform all the other calculations that will be mentioned below.

The capabilities of the new technique have been afterwards tested considering the crystal structure of the L-cysteine (orthorhombic phase I) determined at 30 K.⁴⁵ To accomplish this task we have performed single point unconstrained calculations at the Restricted Hartree-Fock (RHF), B3LYP and ELMO levels using the Dunning cc-pVDZ basis-set and the molecular geometry obtained from the X-Ray diffraction experiment (see Figure 1). In particular, for the ELMO computation we have adopted a localization scheme almost corresponding to the Lewis structure of the molecule. In fact, we have defined both atomic fragments, which describe the core electron and the lone pairs associated with each atom, and bond subunits, which describe each "electronic couple" between two nuclei, with the only exception of a three-atom fragment for the σ and π electrons of the carboxylic group (see Figure 1). We have used the same geometry, basis-set and localization pattern for all the "experimental" ELMOs computations, for which we have also considered the unit cell parameters, the ADPs and the experimental structure factors amplitudes deposited by Moggach and coworkers with their paper.⁴⁵ It is important to observe that experimental values characterized by $|F_{\mathbf{h}}^{exp}| < 3\sigma_{\mathbf{h}}$ have been rejected and this resulted in the selection

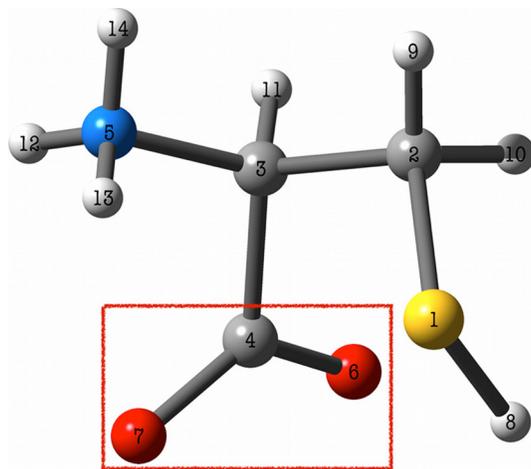


Figure 1: Molecular geometry obtained from the X-ray scattering experiment. The red frame shows the three-atom fragment defined for the σ and π electrons of the carboxylic group.

of 1482 structure factors amplitudes all corresponding to low Fourier components, namely all corresponding to the so called valence electron density. This is the reason why we have decided to perform two different types of "X-ray constrained" ELMO calculations that, from now on, will be indicated as ELMO-XC/OPT and ELMO-XC/FRZ. While the former consist in a full determination of all the Extremely Localized Molecular Orbitals against the experimental scattering data, in the latter the ELMOs describing the core electrons are the ones obtained from the unconstrained ELMO calculation and are kept frozen during the computations. Both the ELMO-XC/OPT and the ELMO-XC/FRZ calculations have been performed with successively larger values for the Lagrange multiplier λ , in particular from $\lambda = 0$ (unconstrained ELMO computation) to $\lambda = 0.7$ with 0.05 λ -steps.

After performing the calculations, we have analyzed the effect of λ on the χ^2 agreement statistics. As we can observe from Figure 2, the trends for the two types of "X-Ray constrained" ELMO calculations are essentially identical and characterized by an initial sharp decrease in the χ^2 value, which means that only a small amount of fitting can significantly improve the agreement with the experimental data. The only small difference consists in the fact that the convergence towards the desired agreement ($\chi^2 = 1.0$) is slightly slower in the ELMO-XC-FRZ case. In fact, in the ELMO-XC/OPT calculations the χ^2 value starts being lower than 1.0 approximately from λ equal

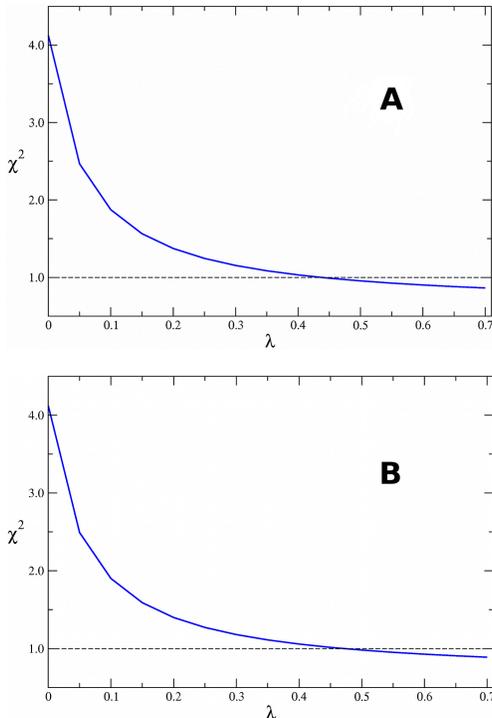


Figure 2: Variation of the χ^2 agreement statistics in function of the Lagrange multiplier λ for (A) the ELMO-XC/OPT and the (B) ELMO-XC/FRZ calculations.

to 0.45, while in the ELMO-XC-FRZ computations this occurs roughly from λ equal to 0.5. The extreme similarity of the trends shown in Figures 2A and 2B is a consequence of the fact that the experimental structure factors amplitudes considered in our calculations mainly refer to the valence electron density and, therefore, optimizing the "core" ELMOs against the experimental data, as in the ELMO-XC/OPT calculations, does not significantly improve the agreement statistics. Finally, as already mentioned above, it is meaningless to force the fitting much beyond $\chi^2 = 1.0$ and, in the following analyses, unless otherwise stated, we will refer to the ELMO-XC/OPT and ELMO-XC/FR wave functions as those constrained ELMO wave functions obtained for $\lambda = 0.45$ and $\lambda = 0.5$, respectively.

In Table 2 we have reported the χ^2 and energy values corresponding to all the performed calculations and, as expected, the unconstrained ELMO wave function provides the worst statistics agreement with the experimental data. This agreement improves using the RHF and B3LYP methods, even if the χ^2 values are still far from an acceptable experimental error range, which, as

already observed, is reached through the ELMO-XC/OPT and the ELMO-XC/FRZ wave functions. Before analyzing in detail the fitting effects on the electron density, it is worth noting that

Table 1: χ^2 agreement statistics and energy values for the analyzed unconstrained and constrained wave functions.

Method	χ^2	Energy (a.u.)
ELMO	4.12	-719.10
RHF	3.87	-719.23
B3LYP	3.81	-721.77
ELMO-XC/OPT	0.99	-718.84
ELMO-XC/FRZ	0.98	-718.82

in Table 2 the energies associated with the constrained ELMO wave functions are higher than all the other ones, especially than the one corresponding to the unconstrained ELMO wave function. This is in line with what has been already observed by Jayatilaka and coworkers,^{33,37} and it can be simply explained as follows. In a "theoretical" ELMO calculation we look for those ELMOs coefficients that correspond to a minimum point on the energy hypersurface. When we extract an "experimental" ELMO wave function from X-ray scattering data, we introduce an additional constraint (see Eq. (5)) without providing a new variational parameter and, therefore, the coefficients of the "experimental" ELMOs are the ones that correspond to the minimum point on the functional hypersurface, which is different from the minimum energy point.

To further study the effects of the wave function fitting, we have compared the unconstrained ELMO electron density with two ELMO-XC/OPT charge distributions: the converged one and the intermediate one obtained for $\lambda = 0.20$ (indicated as ELMO-XC/OPT_{0.20}), which shows how the charge density changes during the fitting procedure. Furthermore, to better distinguish the effects of the crystal environment, we have also considered the comparison of the unconstrained ELMO electron density with the RHF and B3LYP ones. From Figures 3 and 4 (see also Supporting Information for two-dimensional plots relative to other molecular planes) it is possible to note that one of the main fitting consequences consists in a large redistribution of the electron density around the nuclei, such as the evident depletion of charge in the lone-pairs region of the oxygen atoms. This decrease in electron density is common to the RHF and B3LYP methods, but in the

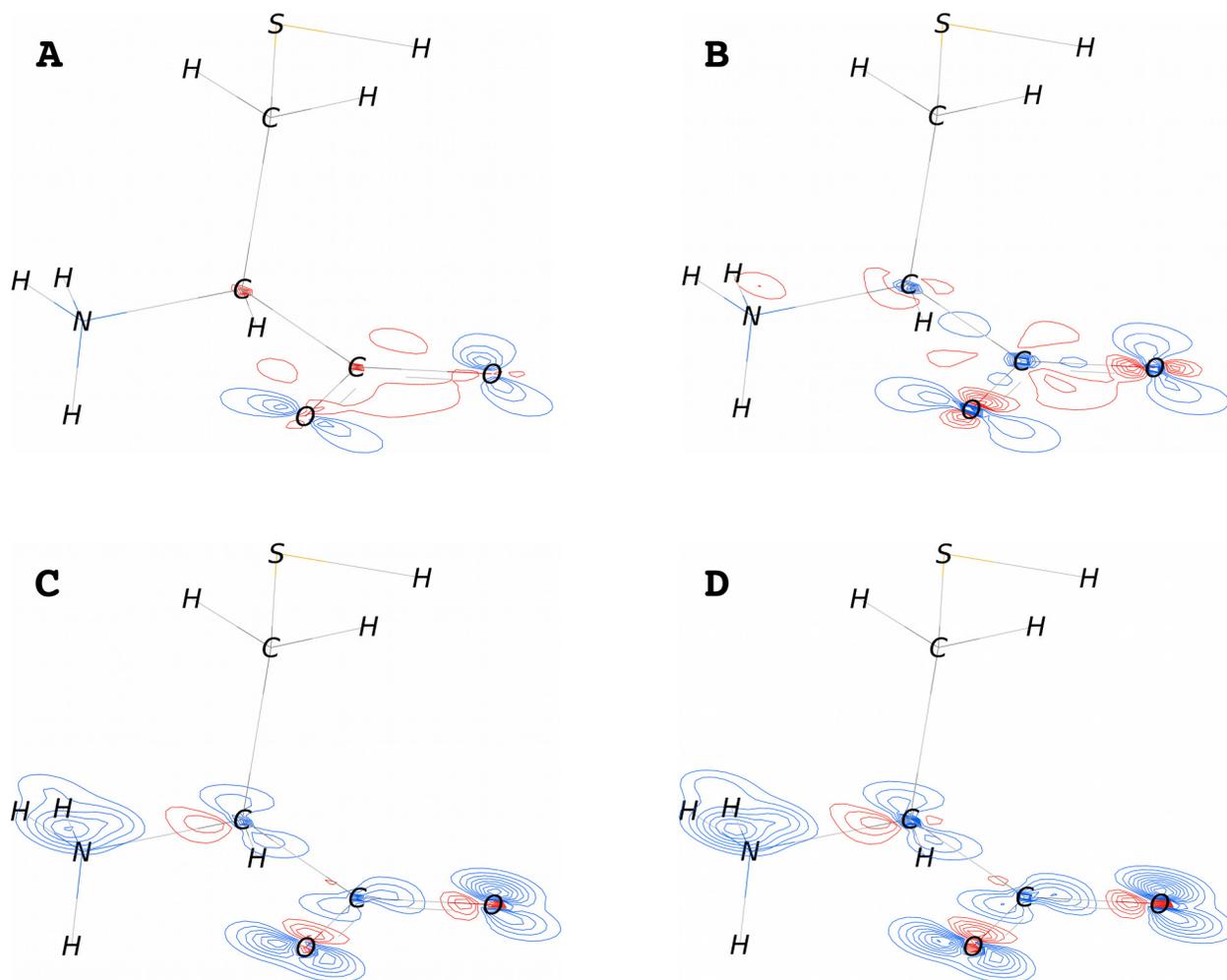


Figure 3: Two-dimensional plots of electron density differences in the plane of the carboxylic group: (A) $\rho(RHF)-\rho(ELMO)$, (B) $\rho(B3LYP)-\rho(ELMO)$, (C) $\rho(ELMO-XC/OPT_{0.20})-\rho(ELMO)$, (D) $\rho(ELMO-XC/OPT)-\rho(ELMO)$. The contours are at linear increments of 0.01 a.u. from -0.1 a.u. to 0.1 a.u., with negative contours in blue and positive contours in red.

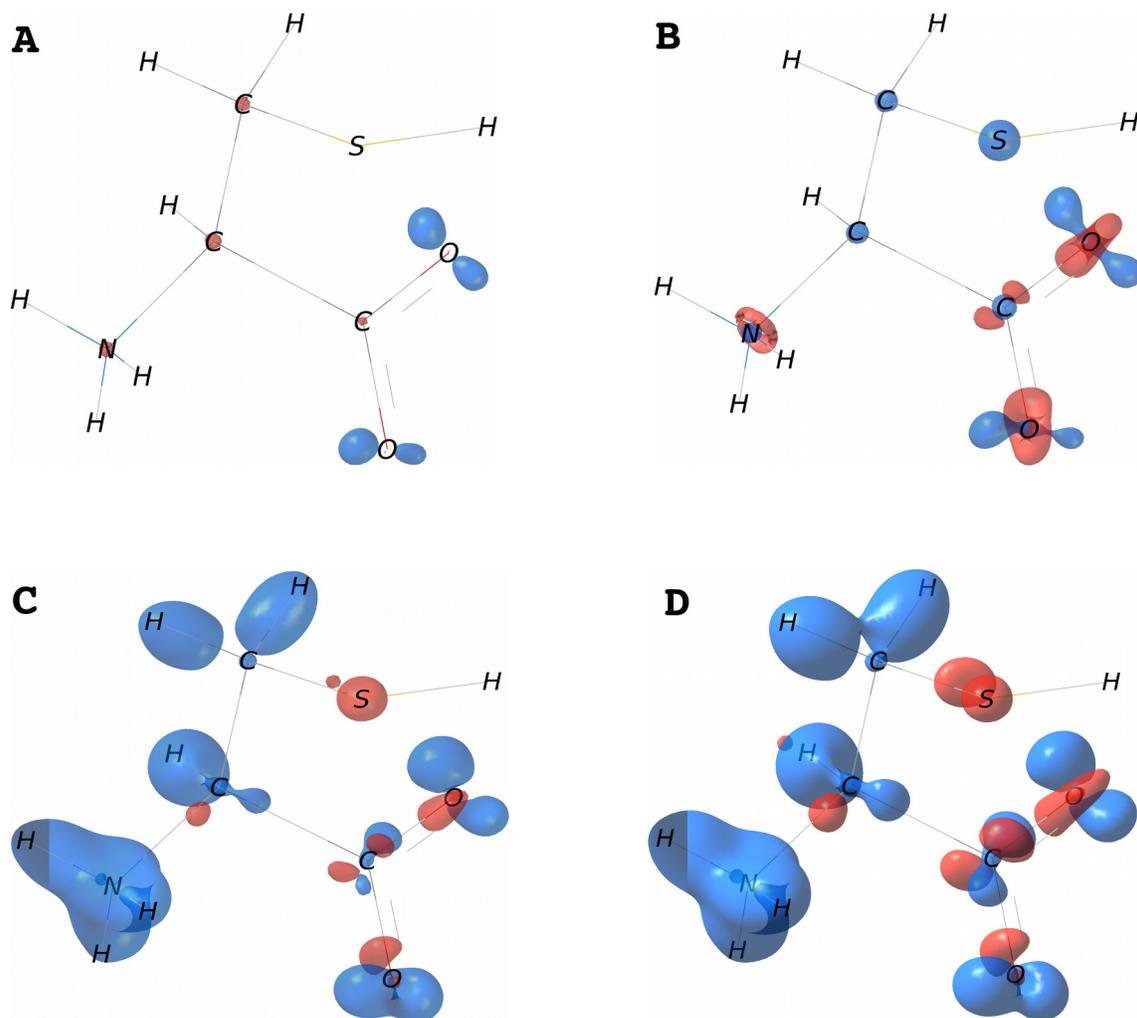


Figure 4: Three-dimensional plots of electron density differences: (A) $\rho(\text{RHF})-\rho(\text{ELMO})$, (B) $\rho(\text{B3LYP})-\rho(\text{ELMO})$, (C) $\rho(\text{ELMO-XC/OPT}_{0.20})-\rho(\text{ELMO})$, (D) $\rho(\text{ELMO-XC/OPT})-\rho(\text{ELMO})$. The isosurface value is set to 0.025 a.u., with negative isosurfaces in blue and positive isosurfaces in red.

ELMO-XC cases it is more distinct and it is also combined with a simultaneous increase in the charge distribution close to the same oxygen atoms. Another important unique effect of the fitting to the crystal environment is represented by the reduction of charge in correspondence of all the C-H and N-H bonds. A shift of charge from the nitrogen atom to the C_α atom and, in analogous way, from the carbon atom to the oxygen atoms in the carboxylic group are also noteworthy and it is possible to observe that they are just sketched in the B3LYP plots. For the sake of completeness, it is important to note that in proximity of the sulfur atom the ELMO-XC charge redistributions are characterized by an increase in electron density, which is completely missing in the RHF case and which is even replaced by a charge depletion if we consider the B3LYP method.

Finally, in order to have a more quantitative picture of the fitting effects, we have computed the values of some similarity indexes between the unconstrained ELMO charge distribution and the other electron densities just mentioned above. In particular, we have considered two indexes that allow to compare point-by-point two electron distributions in the real space: the Real-Space R value (*RSR*), which is defined as

$$RSR(\rho_x, \rho_y) = 100 \frac{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) - \rho_y(\mathbf{r}_i)|}{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) + \rho_y(\mathbf{r}_i)|} \quad (21)$$

where n_p is the number of electron density grid points, and the Walker-Mezey similarity index $L(\rho_x, \rho_y, a, a')$, which compares two charge densities in the space bound by the isosurfaces characterized by the values a and a' (see Supporting Information for further details).⁴⁶ For the sake of clarity, it is important to note that a complete similarity corresponds to values of *RSR* and *L* equal to zero and 100, respectively. From Table 2 it is easy to observe that, in agreement with Figures 3 and 4, the electron distributions associated with the constrained ELMO wave functions are the least similar to the unconstrained ELMO charge density. Furthermore, since the Walker-Mezey index allows to study the electron densities similarity in different regions simply changing the iso-surface values a and a' , it is important to note that the discrepancies between the constrained and

unconstrained ELMO charge distributions are more pronounced far from the nuclei (see the index $L(0.001, 0.01)$ in Table 2) than in the region close to nuclei (see the index $L(0.1, 10)$).

Table 2: Values of the similarity indexes associated with the comparison of the ELMO electron density with the RHF, B3LYP, ELMO-XC/OPT_{0.20} and ELMO-XC/OPT charge distributions.

Comparison Vs. ELMO	<i>RSR</i>	$L(0.001, 10)$	$L(0.1, 10)$	$L(0.01, 0.1)$	$L(0.001, 0.01)$
RHF	0.55	96.85	98.85	97.55	95.80
B3LYP	1.04	94.25	97.74	95.67	92.31
ELMO-XC/OPT _{0.20}	1.65	91.30	96.23	93.89	88.08
ELMO-XC/OPT	2.15	89.12	95.11	92.16	85.26

To summarize, in this letter we have proposed a new wavefunction-based technique for the determination and the analysis of molecular electron densities from high-resolution X-ray scattering data. Our preliminary tests on a crystal structure of the cysteine at 30 K have shown that to obtain Molecular Orbitals strictly localized on small molecular fragments (e.g., atoms, bonds or functional groups) reproducing a set of measured structure factors amplitudes within a predefined accuracy is straightforward. The determination of Extremely Localized Molecular Orbitals from X-Ray diffraction data allows to easily analyze the molecular charge distributions in terms of usual chemical concepts (as in the multipole model) and, at the same time, to obtain quantum mechanically rigorous electron densities (as in the orbital-based approaches) that can be used to compute many important properties. Furthermore, given the reliable transferability of the Extremely Localized Molecular Orbitals,⁴⁷⁻⁵⁰ the construction of "experimental" ELMOs databases can be seen as a possible alternative to the existing pseudoatoms libraries.

Acknowledgement

The author thanks Enrique Espinosa, Emmanuel Aubert, Antonio Monari and Toru Shiozaki for helpful discussions. Maurizio Sironi is also gratefully acknowledged for providing the ELMO code.

Supporting Information Available

Details about the computation of the structure factors, expressions of the functional first and second derivatives with respect to the ELMOs coefficients and details about the Walker-Mezey similarity index. Figure S1: two-dimensional plots of electron density differences in the N-C $_{\alpha}$ -C $_{\beta}$ plane. Figure S2: two-dimensional plots of electron density differences in the C $_{\beta}$ -S-H plane. 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, NY, U.S.A., 1997.
- (3) Koritsanszky, T. S.; Coppens, P. Chemical Applications of X-ray Charge-Density Analysis. *Chem. Rev.* **2001**, *101*, 1583–1628.
- (4) Hirshfeld, F. L. Difference Densities by Least-Squares Refinement: Fumaramic Acid. *Acta Cryst. B* **1971**, *27*, 769–781.
- (5) Stewart, R. F. Electron Population Analysis with Rigid Pseudoatoms. *Acta Cryst. A* **1976**, *32*, 565–574.
- (6) Hansen, N. K.; Coppens, P. Testing Aspherical Atom Refinements on Small-Molecule Data Sets. *Acta Cryst. A* **1978**, *34*, 909–921.
- (7) 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 Cryst. B* **1991**, *47*, 789–797.
- (8) 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.
- (9) 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 Cryst. D* **1998**, *54*, 1306–1318.
- (10) 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 Cryst. A* **2007**, *63*, 108–125.

- (11) 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 Cryst A* **2002**, *58*, 464–472.
- (12) 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.
- (13) 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.
- (14) Dittrich, B.; Koritsanszky, T.; Luger, P. A Simple Approach to Nonspherical Electron Densities by Using Invarioms. *Angew. Chem. Int. Ed.* **2004**, *43*, 2718–2721.
- (15) 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 Cryst. A* **2005**, *61*, 314–320.
- (16) 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 Cryst. D* **2006**, *62*, 1325–1335.
- (17) Clinton, W. L.; Nakhleh, J.; Wunderlich, F. Direct Determination of Pure-State Density Matrices. I. Some Simple Introductory Calculations. *Phys. Rev.* **1969**, *177*, 1–6.
- (18) 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*, 7–13.
- (19) Clinton, W. L.; Henderson, G. A.; Prestia, J. V. Direct Determination of Pure-State Density Matrices. III. Purely Theoretical Densities Via an Electrostatic-Virial Theorem. *Phys. Rev.* **1969**, *177*, 13–18.

- (20) Clinton, W. L.; Lamers, G. B. Direct Determination of Pure-State Density Matrices. IV. Investigation of Another Constraint and Another Application of the P Equations. *Phys. Rev.* **1969**, *177*, 19–27.
- (21) Clinton, W. L.; Galli, A. J.; Henderson, G. A.; Lamers, G. B.; Massa, L. J.; Zarur, J. Direct Determination of Pure-State Density Matrices. V. Constrained Eigenvalue Problems. *Phys. Rev.* **1969**, *177*, 27–33.
- (22) Clinton, W. L.; Massa, L. J. The Cusp Condition: Constraint on the Electron Density Matrix. *Int. J. Quantum Chem.* **1972**, *6*, 519–523.
- (23) Clinton, W. L.; Massa, L. J. Determination of the Electron Density Matrix from X-Ray Diffraction Data. *Phys. Rev. Lett.* **1972**, *29*, 1363–1366.
- (24) 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.
- (25) Frishberg, C.; Massa, L. J. Idempotent Density Matrices for Correlated Systems from X-Ray Diffraction Structure Factors. *Phys. Rev. B* **1981**, *24*, 7018–7024.
- (26) 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.
- (27) 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.
- (28) 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.

- (29) 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.
- (30) Tanaka, K. X-Ray Analysis of Wavefunctions by the Least-Squares Method Incorporating Orthonormality. I. General Formalism. *Acta Cryst. A* **1988**, *44*, 1002–1008.
- (31) Jayatilaka, D. Wave Function for Beryllium from X-Ray Diffraction Data. *Phys. Rev. Lett.* **1998**, *80*, 798–801.
- (32) Jayatilaka, D.; Grimwood, D. J. Wavefunctions Derived from Experiment. I. Motivation and Theory. *Acta Cryst. A* **2001**, *57*, 76–86.
- (33) Grimwood, D. J.; Jayatilaka, D. Wavefunctions Derived from Experiment. II. A Wavefunction for Oxalic Acid Dihydrate. *Acta Cryst. A* **2001**, *57*, 87–100.
- (34) Bytheway, I.; Grimwood, D.; Jayatilaka, D. Wavefunctions Derived from Experiment. III. Topological Analysis of Crystal Fragments. *Acta Cryst. A* **2002**, *58*, 232–243.
- (35) 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 Cryst. A* **2002**, *58*, 244–251.
- (36) 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.
- (37) Hudák, M.; Jayatilaka, D.; Peraínova, L.; Biskupic, S.; Kozísek, J.; Bucinský, L. X-Ray Constrained Unrestricted Hartree-Fock and Douglas-Kroll-Hess Wavefunctions. *Acta Cryst. A* **2010**, *66*, 78–92.
- (38) Stoll, H.; Wagenblast, G.; Preuss, H. On the Use of Local Basis Sets for Localized Molecular Orbitals. *Theoret. Chim. Acta* **1980**, *57*, 169–178.

- (39) Obara, S.; Saika, A. Efficient Recursive Computation of Molecular Integrals over Cartesian Gaussian Functions. *J. Chem. Phys.* **1986**, *84*, 3963–3974.
- (40) Obara, S.; Saika, A. General Recurrence Formulas for Molecular Integrals over Cartesian Gaussian Functions. *J. Chem. Phys.* **1988**, *89*, 1540–1559.
- (41) Head-Gordon, M.; Pople, J. A. A Method for Two-Electron Gaussian Integral and Integral Derivative Evaluation Using Recurrence Relations. *J. Chem. Phys.* **1988**, *89*, 5777–5786.
- (42) 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.
- (43) Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. *Numerical Recipes: the Art of Scientific Computing*; Cambridge University Press: New York, NY, U.S.A., 1986.
- (44) 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.
- (45) Moggach, S. A.; Clark, S. J.; Parsons, S. L-Cysteine-I at 30 K. *Acta Cryst. E* **2005**, *61*, o2739–o2742.
- (46) Walker, P. D.; Mezey, P. G. Ab Initio Quality Electron Densities for Proteins: A MEDLA Approach. *J. Am. Chem. Soc.* **1994**, *116*, 12022–12032.
- (47) 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.
- (48) Genoni, A.; Ghitti, M.; Pieraccini, S.; Sironi, M. A Novel Extremely Localized Molecular

Orbitals Based Technique for the One-Electron Density Matrix Computation. *Chem. Phys. Lett.* **2005**, *415*, 256–260.

- (49) Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely Localized Molecular Orbitals: Theory and Applications. *Theor. Chem. Acc.* **2007**, *117*, 685–698.
- (50) 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.