JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# X-ray Constrained Extremely Localized Molecular Orbitals: Theory and Critical Assessment of the New Technique

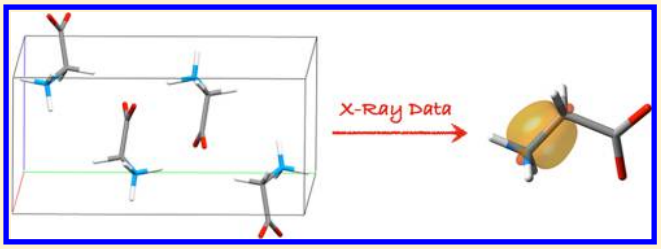
Alessandro Genoni*[,†,‡]

[†]CNRS, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France

[‡]Université de Lorraine, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France

Supporting Information

**ABSTRACT:** Following the X-ray constrained wave function approach proposed by Jayatilaka, we have devised a new technique that allows to extract molecular orbitals strictly localized on small molecular fragments from sets of experimental X-ray structure factors amplitudes. Since the novel strategy enables to obtain electron distributions that have quantum mechanical features and that can be easily interpreted in terms of traditional chemical concepts, the method can be also considered as a new useful tool for the determination and the analysis of charge densities from high-resolution X-ray experiments. In this paper, we describe in detail the theory of the new technique, which, in comparison to our preliminary work, has been improved both treating the effects of isotropic secondary extinctions and introducing a new protocol to halt the fitting procedure against the experimental X-ray scattering data. The performances of the novel strategy have been studied both in function of the basis-sets flexibility and in function of the quality of the considered crystallographic data. The tests performed on four different systems ($\alpha$-glycine, L-cysteine, (aminomethyl)phosphonic acid and N-(trifluoromethyl)formamide) have shown that the achievement of good statistical agreements with the experimental measures mainly depends on the quality of the crystal structures (i.e., geometry positions and thermal parameters) used in the X-ray constrained calculations. Finally, given the reliable transferability of the obtained Extremely Localized Molecular Orbitals (ELMOs), we envisage to exploit the novel approach to construct new ELMOs databases suited to the development of linear-scaling methods for the refinement of macromolecular crystal structures.




## 1. INTRODUCTION

The determination of wave functions from experimental data has been an open research field for many years. As well explained by Jayatilaka and Grimwood,[1] the main reason that motivated many groups to work on this subject is the desire to find a physically meaningful object able to condensate in it all the data collected from an experiment. In this sense, the wave function represents the best possible choice from the theoretical point of view because, as well-known, in quantum theory it is the fundamental entity that intrinsically contains all the information about a system. In fact, after obtaining an "experimental" wave function, many properties can be computed, in principle, even those that are not strictly related to the ones used to determine the wave function itself.

In this area of interest, it is important to mention the pioneering papers of Clinton and co-workers[2−6] that have paved the way to the development of more and more sophisticated techniques for the determination of wave functions constrained by X-ray diffraction data.[1,17−20] Among them the strategy proposed by Jayatilaka is noteworthy.[1,15−20] It consists in extracting a single Slater determinant that reproduces a collected set of reflections intensities within a predefined accuracy, and it can be considered as a practical implementation of the Henderson and Zimmerman's idea,[21] according to which, of all the possible single Slater determinants that are compatible with a given electron density, the optimal one is that which minimizes the Hartree−Fock energy. Although initially limited to the case of restricted Hartree−Fock wave functions,[1,15−19] the Jayatilaka formalism has been afterward extended also to the unrestricted case and to the relativistic scalar second-order Douglas-Kroll-Hess approach.[20]

Following this philosophy, we have combined the method proposed by Jayatilaka with the technique developed by Stoll[22] for the *a priori* determination of Extremely Localized Molecular Orbitals (ELMOs), which are molecular orbitals strictly localized on small molecular fragments such as atoms, bonds, or functional groups. Our preliminary work[23] has shown that the computation of ELMOs from X-ray diffraction data is straightforward and that the novel strategy can be considered as a new useful tool for the determination and the analysis of electron densities from X-ray experiments.

It is worthwhile to note that the techniques currently used for the determination and the analysis of charge distributions from high-resolution X-ray scattering data can be subdivided into two great categories: the multipole model and the wave function- (or orbital-) based approaches. The former fits the collected experimental structure factors amplitudes to a model

**Received:** April 10, 2013
**Published:** June 6, 2013

electron distribution written as a sum of aspherical atomic density functions (also known as pseudoatoms) that depend on proper refinable parameters.[24−26] This is the most widely used technique, and its popularity mainly relies on its linear scaling behavior and, above all, on the easy chemical interpretability of the resulting electron densities. In fact, due to the pseudoatoms definition, the charge distributions can be interpreted as sum of atomic electron densities characterized by deformations deriving from the presence of bonding and nonbonding interactions. Nevertheless, the multipole model is affected by some serious drawbacks, such as the possible presence of unphysical negative regions in the fitted electron densities and the strong statistical correlation between the Anisotropic Displacement Parameters (ADPs) and the refinable coefficients in the multipole expansions. On the contrary, the strategies belonging to the latter group, which comprises the Jayatilaka approach, do not exhibit these problems, and through the fitting of the experimental X-ray diffraction data to a predetermined form for the density matrix[7−13,27,28] or for the wave function[1,14−20] of the systems in exam, they enable to obtain quantum mechanically rigorous charge distributions. Unfortunately, this desirable feature is associated with the loss of the easy chemical interpretability, which, as already stressed, is one of the strengths of the multipole model.

This scenario has prompted us to develop a new technique that combined the quantum mechanical rigor of the orbital-based methods with the ease of chemical interpretation of the multipole model. This has resulted in our novel strategy[23] that, extracting Extremely Localized Molecular Orbitals from X-ray scattering data, allows to determine charge distributions that have intrinsic quantum mechanical features and that, at the same time, can be interpreted in terms of traditional chemical concepts. In this context, it is important to mention that techniques to recover the chemical interpretability in the framework of the orbital-based approaches have been already proposed by Jayatilaka and co-workers, which have extracted quantities such as the Electron Localization Functions (ELFs)[29] and the Electron Localizability Indicators (ELIs)[30,31] from X-ray constrained wave functions in order to study interesting chemical problems, such as the significance of ionic chemical bonding in sulfur dioxide.[32] For the sake of completeness, it is worth observing that the same philosophy that inspired the new "experimental" ELMO approach and the methods devised by Jayatilaka and co-workers also represents the foundation of the standard topological strategies for the analysis of chemical concepts.[33]

In this paper, we will present in detail the theory of the new technique, which has been also extended to the case in which the treatment of secondary extinctions is essential, and furthermore, we will present a critical assessment of the new approach, evaluating its performances both in function of the basis-sets used for the calculations and in function of the crystallographic data (namely, structure factors amplitudes, molecular geometry, ADPs, and cell parameters) considered for our computational tests.

## 2. THEORY

In this section, we will describe in detail the theory on which our novel method for the extraction of X-ray constrained Extremely Localized Molecular Orbitals (XC-ELMOs) is based. At first, we will dedicate a brief subsection to the original ELMO technique proposed by Stoll,[22] and afterward, we will present the fundamental assumptions and equations of our approach. Finally, in the last subsection, we will describe the strategy adopted for optimizing the scale factor and the secondary extinction parameter.

**Extremely Localized Molecular Orbitals.** Let us consider a $2N$-electron closed shell system, and let us subdivide it into $f$ fragments that correspond to small molecular subunits, such as atoms, bonds, or functional groups. Each fragment $i$ is characterized by a local basis set $\{|\chi_{i\mu}\rangle\}_{\mu=1}^{M_i}$ constituted by the only $M_i$ basis functions that are centered on the atomic nuclei belonging to the subunit. The generic $\alpha$-th ELMO of the $i$-the fragment can thus be expressed as

$$|\varphi_{i\alpha}\rangle = \sum_{\mu=1}^{M_i} C_{i\mu,i\alpha}|\chi_{i\mu}\rangle \tag{1}$$

and therefore, it is strictly localized on the subunit in exam. Furthermore, it is important to note that, since the overlapping fragments may share part of their local basis-sets, a natural nonorthogonality between ELMOs associated with different subunits arises, nonorthogonality that often results into nontrivial convergence problems during the ELMOs determination.[22,34,35]

The coefficients in eq 1 are determined variationally minimizing the energy associated with the ELMO wave function that has the following expression:

$$|\Psi_{\mathrm{ELMO}}\rangle = \hat{A}[\prod_{i=1}^{f}\prod_{\alpha=1}^{n_i}\varphi_{i\alpha}\overline{\varphi}_{i\alpha}] \tag{2}$$

where $\hat{A}$ is the antisymmetrizer, $n_i$ is the number of occupied ELMOs for the $i$-th fragment, and $\overline{\varphi}_{i\alpha}$ is a spinorbital with spatial part $\varphi_{i\alpha}$ and spin part $\beta$. In other words, the ELMOs are those orbitals which minimize the following functional:

$$E[\boldsymbol{\varphi}] = \langle\Psi_{\mathrm{ELMO}}|\hat{H}|\Psi_{\mathrm{ELMO}}\rangle \tag{3}$$

with $\hat{H}$ as the nonrelativistic Hamiltonian operator for the system in exam.

Defining the dual orbitals $|\tilde{\varphi}_{j\beta}\rangle$ as follows

$$|\tilde{\varphi}_{j\beta}\rangle = \sum_{k=1}^{f}\sum_{\gamma=1}^{n_k}[\mathbf{S}^{-1}]_{k\gamma,j\beta}|\varphi_{k\gamma}\rangle \tag{4}$$

with $\mathbf{S}$ as the overlap matrix of the occupied ELMOs, Stoll[22] has shown that the arbitrary variation of the functional $E[\boldsymbol{\varphi}]$ with respect to the occupied ELMO $|\varphi_{j\beta}\rangle$ (from now on indicated as $\delta_{(j\beta)}E$) is

$$\delta_{(j\beta)}E = 4\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle \tag{5}$$

where $\hat{F}$ is the Fock operator and $\hat{\rho}$ is the global density operator corresponding to

$$\hat{\rho} = \sum_{j=1}^{f}\sum_{\beta=1}^{n_j}|\tilde{\varphi}_{j\beta}\rangle\langle\varphi_{j\beta}| = \sum_{j=1}^{f}\sum_{\beta=1}^{n_j}|\varphi_{j\beta}\rangle\langle\tilde{\varphi}_{j\beta}| \tag{6}$$

The lowest value of the functional is achieved when $\delta_{(j\beta)}E$ vanishes for all $j$ and $\beta$ and, therefore, the ELMOs that minimize $E[\boldsymbol{\varphi}]$ are the ones that satisfy the following equation for each fragment:

$$(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle = 0 \tag{7}$$

which is equivalent to the eigenvalue problem

$$\hat{F}^{j}|\varphi_{j\beta}\rangle = \varepsilon_{j\beta}|\varphi_{j\beta}\rangle \tag{8}$$

where $\hat{F}^j$, which is the modified Fock operator for the $j$-th subunit, can be written as

$$\hat{F}^j = (1 - \hat{\rho} + \hat{\rho}_j^{\dagger})\hat{F}(1 - \hat{\rho} + \hat{\rho}_j) \tag{9}$$

and it clearly depends on the local density operator $\hat{\rho}_j$ defined as

$$\hat{\rho}_j = \sum_{\beta=1}^{n_j} |\tilde{\varphi}_{j\beta}\rangle\langle\varphi_{j\beta}| \tag{10}$$

Hence, in order to determine the desired Extremely Localized Molecular Orbitals for the system in exam, eq 8 must be solved self-consistently for each fragment using the proper local basis-sets defined *a priori* through the choice of the localization scheme. It is worthwhile to note that the equations associated with the different subunits are coupled because each $\hat{F}^j$ operator depends on the global density operator $\hat{\rho}$ (see eq 9).

**X-ray Constrained ELMOs: Fundamental Assumptions.** Unlike the Stoll technique, the computation of X-ray constrained Extremely Localized Molecular Orbitals consists in looking for the ELMO wave function that not only minimizes the energy but that also provides the desired agreement with a set of collected experimental structure factors amplitudes $\{|F_{\mathbf{h}}^{\mathrm{exp}}|\}$, namely, in looking for those ELMOs that minimize this new functional:

$$\begin{aligned} J[\boldsymbol{\varphi}] &= \langle\Psi_0|\hat{H}_0|\Psi_0\rangle + \lambda(\chi^2 - \Delta) \\ &= E_0[\boldsymbol{\varphi}] + \lambda(\chi^2[\boldsymbol{\varphi}] - \Delta) \end{aligned} \tag{11}$$

with $\Psi_0$ and $\hat{H}_0$ as the ELMO wave function and the non-relativistic Hamiltonian operator for the reference crystal-unit, respectively, $\lambda$ as the Lagrange multiplier crucial for the wave function fitting process, $\Delta$ as the desired error between theoretical and experimental values, and $\chi^2$ as the agreement statistics between the calculated and the observed structure factor amplitudes $|F_{\mathbf{h}}^{\mathrm{calc}}|$ and $|F_{\mathbf{h}}^{\mathrm{exp}}|$, namely,

$$\chi^2 = \frac{1}{N_r - N_p}\sum_{\mathbf{h}}\frac{(\eta Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}| - |F_{\mathbf{h}}^{\mathrm{exp}}|)^2}{\sigma_{\mathbf{h}}^2} \tag{12}$$

where $N_r$ is the number of considered experimental scattering data, $N_p$ is the number of adjustable parameters, $\mathbf{h}$ is the tern of Miller indexes labeling the reflection, and $\sigma_{\mathbf{h}}$ is the error associated with each measure. Furthermore, $\eta$ is an overall $\mathbf{h}$-independent scale factor that is necessary to have $|F_{\mathbf{h}}^{\mathrm{calc}}|$ and $|F_{\mathbf{h}}^{\mathrm{exp}}|$ on the same scale, while $Y_{\mathbf{h}}(\varepsilon)$ is a term that takes into account the secondary extinction phenomena affecting the experimental data and that, in particular, following the Larson approach,[36,37] can be expressed in this way:

$$Y_{\mathbf{h}}(\varepsilon) = \left(1 + 2\varepsilon|F_{\mathbf{h}}^{\mathrm{calc}}|^2\frac{1 + \cos^4(2\theta_{\mathbf{h}})}{1 + \cos^2(2\theta_{\mathbf{h}})}\right)^{-1/4} \tag{13}$$

with $\theta_{\mathbf{h}}$ as the angle between the incident radiation and the diffracting plane and $\varepsilon$ as the secondary extinction parameter. Both $\eta$ and $\varepsilon$ are determined in order to minimize the $\chi^2$ value, and the adopted strategy for their optimization will be described. Finally, to complete the discussion about the measure of the statistical agreement, it is extremely important to note that, due to the experimental errors associated with the collected structure factors amplitudes, $\chi^2$ must not be forced to zero.[1,15] Therefore, the desired accuracy $\Delta$ is usually chosen equal to 1, so that the calculated values are on average within one standard deviation of the experimental data.

As mentioned, $\Psi_0$ in eq 11 is the ELMO wave function for the reference crystal-unit. This follows from assuming that the crystal is a collection of noninteracting units, namely, from assuming that the wave function of the crystal is

$$\Psi_{\mathrm{cryst}} = \prod_k \Psi_k \tag{14}$$

where all the crystal-unit wave functions $\Psi_k$ are formally identical, of the form given by eq 2, and related to each other through the crystal symmetry operations. This approximation, along with the further assumption that all the noninteracting units correspond to symmetry-unique portions of the crystal unit-cell, allows to compute the unit-cell charge distribution using the only reference electron density $\rho_0(\mathbf{r})$ associated with the wave function $\Psi_0$. In other words, we have

$$\rho_{\mathrm{cell}}(\mathbf{r}) = \sum_{j=1}^{N_m}\rho_j(\mathbf{r}) = \sum_{j=1}^{N_m}\rho_0(\mathbf{R}_j^{-1}(\mathbf{r} - \mathbf{r}_j)) \tag{15}$$

where the $N_m$ unit-cell electron distributions are related to the reference one through the symmetry operations $\{\mathbf{R}_j\mathbf{r}_j\}$. It is important to stress that, since in our case $\rho_0(\mathbf{r})$ is obtained through an isolated fragment calculation (namely, an isolated X-ray constrained ELMO computation), we introduce an approximation in eq 15, which is formally exact.

The previous assumptions are crucial for the computation of the structure factors amplitudes $\{|F_{\mathbf{h}}^{\mathrm{calc}}|\}$ that appear in eq 12. In fact, exploiting eq 15 and bearing in mind that the structure factors are Fourier transforms of the unit-cell electron density, it is possible to show that

$$F_{\mathbf{h}}^{\mathrm{calc}} = \mathrm{Tr}[\mathbf{D}_0\mathbf{I}_{\mathbf{h}}] \tag{16}$$

where $\mathbf{D}_0$ is the one-particle density matrix associated with $\Psi_0$, while $\mathbf{I}_{\mathbf{h}}$ is the matrix of the Fourier transforms of the basis functions products summed over all the equivalent unit-cell sites, namely,

$$[\mathbf{I}_{\mathbf{h}}]_{\mu\nu} = \sum_{j=1}^{N_m} e^{i2\pi\mathbf{r}_j\cdot(\mathbf{Bh})}T_{\mu\nu}[\mathbf{B}^{-1}\mathbf{R}_j^T\mathbf{Bh}]\int d\mathbf{r}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})e^{i2\pi(\mathbf{R}_j\mathbf{r})\cdot(\mathbf{Bh})} \tag{17}$$

with $\mathbf{B}$ as the reciprocal lattice matrix and with $T_{\mu\nu}[\mathbf{B}^{-1}\mathbf{R}_j^T\mathbf{Bh}]$ as a parameter that empirically takes into account the thermal motion effects on the charge density. In this paper, we have decided to evaluate this parameter following the Stewart model,[38] and therefore, it can be expressed as

$$T_{\mu\nu}[\mathbf{h}'] = \exp[-2\pi\tau_{\mu\nu}(\mathbf{Bh}')\cdot(\mathrm{U}^\mu + \mathrm{U}^\nu)\mathbf{Bh}'] \tag{18}$$

with $\mathbf{h}' = \mathbf{B}^{-1}\mathbf{R}_j^T\mathbf{Bh}$. $\tau_{\mu\nu}$ is a factor that is equal to 0.5 if the distance between the nuclei associated with the atomic orbitals $\chi_\mu(\mathbf{r})$ and $\chi_\nu(\mathbf{r})$ is lower than 2.5 au, while it is equal to 0.25 in all the other cases. Finally, $\mathbf{U}^\mu$ and $\mathbf{U}^\nu$ are the matrices containing the thermal vibration parameters (obtained from the X-ray experiment) for the atoms on which the basis functions $\chi_\mu(\mathbf{r})$ and $\chi_\nu(\mathbf{r})$ are centered, respectively. It is important to mention that another possible model for the computation of the thermal factors could be the one proposed by Jayatilaka and Dittrich.[39] They have recently shown that the partitioning of the molecular electron density in terms of Hirshfeld atoms, followed by the adoption of the usual atom-centered formalism, provides better results, although the model assumes that the atoms vibrate independently, which is often not the case in many crystals. However, the proper treatment of thermal

motion is still an open field of research and is out of the scope of the present paper. Finally, for the sake of completeness, it is important to note that an Obara−Saika scheme[40,41] with vertical and horizontal recurrence relations[42] has been implemented for the calculation of the integrals in eq 17 (see the Supporting Information for more details).

**X-ray Constrained ELMOs: the Eigenvalue Equation.** In order to look for the Extremely Localized Molecular Orbitals that minimize the functional given by eq 11, let us consider the arbitrary variation of $J[\boldsymbol{\varphi}]$ with respect to the occupied ELMO $|\varphi_{j\beta}\rangle$. Omitting all the subscripts and the superscripts corresponding to the reference crystal-unit, we have

$$\delta_{(j\beta)}J = \delta_{(j\beta)}E + \lambda\delta_{(j\beta)}\chi^2 \tag{19}$$

where $\boldsymbol{\delta}_{(j\beta)}E$ is given by eq 5, while, exploiting the relation $|F_{\mathbf{h}}^{\mathrm{calc}}| = [F_{\mathbf{h}}^{\mathrm{calc}}(F_{\mathbf{h}}^{\mathrm{calc}})^*]^{1/2}$ and neglecting the dependence of $Y_{\mathbf{h}}(\varepsilon)$ on $F_{\mathbf{h}}^{\mathrm{calc}}$, we obtain

$$\delta_{(j\beta)}\chi^2 = \frac{\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{(\eta Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}| - |F_{\mathbf{h}}^{\mathrm{exp}}|)Y_{\mathbf{h}}(\varepsilon)}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{\mathrm{calc}}|} \times \{(F_{\mathbf{h}}^{\mathrm{calc}})^*\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}} + F_{\mathbf{h}}^{\mathrm{calc}}\delta_{(j\beta)}(F_{\mathbf{h}}^{\mathrm{calc}})^*\} \tag{20}$$

Now, defining the structure factor operator

$$\hat{I}_{\mathbf{h}} = \sum_{j=1}^{N_m} e^{i2\pi(\mathbf{R}_j\boldsymbol{r}+\mathbf{r}_j)\cdot(\mathbf{Bh})} = \hat{I}_{\mathbf{h},R} + i\hat{I}_{\mathbf{h},C} \tag{21}$$

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 get

$$F_{\mathbf{h}}^{\mathrm{calc}} = 2\sum_{i=1}^{f}\sum_{\alpha=1}^{n_i}\langle\varphi_{i\alpha}|\hat{I}_{\mathbf{h}}|\tilde{\varphi}_{i\alpha}\rangle \tag{22}$$

and

$$(F_{\mathbf{h}}^{\mathrm{calc}})^* = 2\sum_{i=1}^{f}\sum_{\alpha=1}^{n_i}\langle\tilde{\varphi}_{i\alpha}|\hat{I}_{\mathbf{h}}^{\dagger}|\varphi_{i\alpha}\rangle \tag{23}$$

Furthermore, taking advantage of the following relation established by Stoll[22] between $|\delta_{(j\beta)}\tilde{\varphi}_{i\alpha}\rangle$ and $|\delta\varphi_{j\beta}\rangle$

$$|\delta_{(j\beta)}\tilde{\varphi}_{i\alpha}\rangle = (1-\hat{\rho})|\delta\varphi_{j\beta}\rangle[\mathbf{S}^{-1}]_{j\beta,i\alpha} - |\tilde{\varphi}_{j\beta}\rangle\langle\delta\varphi_{j\beta}|\tilde{\varphi}_{i\alpha}\rangle \tag{24}$$

we obtain

$$\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}} = 2\{\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h}}|\tilde{\varphi}_{j\beta}\rangle + \langle\tilde{\varphi}_{j\beta}|\hat{I}_{\mathbf{h}}(1-\hat{\rho})|\delta\varphi_{j\beta}\rangle\} \tag{25}$$

and

$$\delta_{(j\beta)}(F_{\mathbf{h}}^{\mathrm{calc}})^* = (\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}})^* \tag{26}$$

Using eq 26 in eq 20, $\delta_{(j\beta)}\chi^2$ becomes

$$\delta_{(j\beta)}\chi^2 = \frac{2\eta}{N_r - N_p}\sum_{\mathbf{h}}\frac{(\eta Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}| - |F_{\mathbf{h}}^{\mathrm{exp}}|)Y_{\mathbf{h}}(\varepsilon)}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{\mathrm{calc}}|} \times \mathrm{Re}\{(F_{\mathbf{h}}^{\mathrm{calc}})^*\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}}\} \tag{27}$$

and exploiting the fact that we assume to work with real orbitals and that the structure factor operator can be decomposed into its real and imaginary parts, $\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}}$ can be expressed as

$$\delta_{(j\beta)}F_{\mathbf{h}}^{\mathrm{calc}} = 4\{\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle + i\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\} \tag{28}$$

Now, substituting eqs 23 and 28 into eq 27, we have

$$\delta_{(j\beta)}\chi^2 = 4\{\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Re}\{F_{\mathbf{h}}^{\mathrm{calc}}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle + \sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Im}\{F_{\mathbf{h}}^{\mathrm{calc}}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\} \tag{29}$$

with

$$K_{\mathbf{h}} = \frac{2\eta Y_{\mathbf{h}}(\varepsilon)}{N_r - N_p}\frac{\eta Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}| - |F_{\mathbf{h}}^{\mathrm{exp}}|}{\sigma_{\mathbf{h}}^2|F_{\mathbf{h}}^{\mathrm{calc}}|} \tag{30}$$

and therefore, the arbitrary variation of $J[\boldsymbol{\varphi}]$ given by eq 19 becomes

$$\delta_{(j\beta)}J = 4\{\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{F}|\tilde{\varphi}_{j\beta}\rangle + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Re}\{F_{\mathbf{h}}^{\mathrm{calc}}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},R}|\tilde{\varphi}_{j\beta}\rangle + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Im}\{F_{\mathbf{h}}^{\mathrm{calc}}\}\langle\delta\varphi_{j\beta}|(1-\hat{\rho})\hat{I}_{\mathbf{h},C}|\tilde{\varphi}_{j\beta}\rangle\} \tag{31}$$

Since the minimum of the functional is reached when $\delta_{(j\beta)}J$ vanishes for all $j$ and $\beta$, the desired XC-ELMOs are the ones that satisfy the following equation for each fragment:

$$\{(1-\hat{\rho})\hat{F} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Re}\{F_{\mathbf{h}}^{\mathrm{calc}}\}(1-\hat{\rho})\hat{I}_{\mathbf{h},R} + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Im}\{F_{\mathbf{h}}^{\mathrm{calc}}\}(1-\hat{\rho})\hat{I}_{\mathbf{h},C}\}|\tilde{\varphi}_{j\beta}\rangle = 0 \tag{32}$$

As already shown in our preliminary work,[23] it is possible to prove that eq 32 is equivalent to

$$\hat{F}^{j,\mathrm{exp}}|\varphi_{j\beta}\rangle = \varepsilon_{j\beta}^{\mathrm{exp}}|\varphi_{j\beta}\rangle \tag{33}$$

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

$$\hat{F}^{j,\mathrm{exp}} = (1-\hat{\rho}+\hat{\rho}_j^{\dagger})\hat{F}(1-\hat{\rho}+\hat{\rho}_j) + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Re}\{F_{\mathbf{h}}^{\mathrm{calc}}\}(1-\hat{\rho}+\hat{\rho}_j^{\dagger})\hat{I}_{\mathbf{h},R}(1-\hat{\rho}+\hat{\rho}_j) + \lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Im}\{F_{\mathbf{h}}^{\mathrm{calc}}\}(1-\hat{\rho}+\hat{\rho}_j^{\dagger})\hat{I}_{\mathbf{h},C}(1-\hat{\rho}+\hat{\rho}_j) \tag{34}$$

Hence, in order to compute X-ray constrained ELMOs, the eigenvalue problem given by eq 33 must be solved self-consistently for each fragment, and as for the "theoretical" ELMOs, for each subunit, only the space spanned by the local basis-set defined *a priori* is involved. Also, for the XC-ELMOs, the equations associated with the different fragments are coupled because each operator $\hat{F}^{j,\mathrm{exp}}$ depends on the global density operator (see eq 34).

Finally, it is worth mentioning that, due to the nonorthogonal nature of the Extremely Localized Molecular Orbitals, convergence problems and instabilities in the resolution of eqs 33 might arise, especially near the functional minimum. Therefore, to solve this problem, following a strategy already used by Fornili et al. in the determination of "theoretical" ELMOs,[35] we have implemented a second-order technique consisting in a direct minimization of the functional $J[\boldsymbol{\varphi}]$.[23] In particular, we have developed a quasi-Newton procedure in which an approximate analytic Hessian is computed only at the first iteration and, afterward, it is updated exploiting a variable metric algorithm that uses the Broyden−Fletcher−Goldfarb−Shanno formula.[43]

**X-ray Constrained ELMOs: $\eta$ and $\varepsilon$ Optimization.** As already mentioned, both $\eta$ and $\varepsilon$ are determined in order to minimize the $\chi^2$ value. To accomplish this task, we have devised a strategy that is schematically depicted in Figure 1 and that

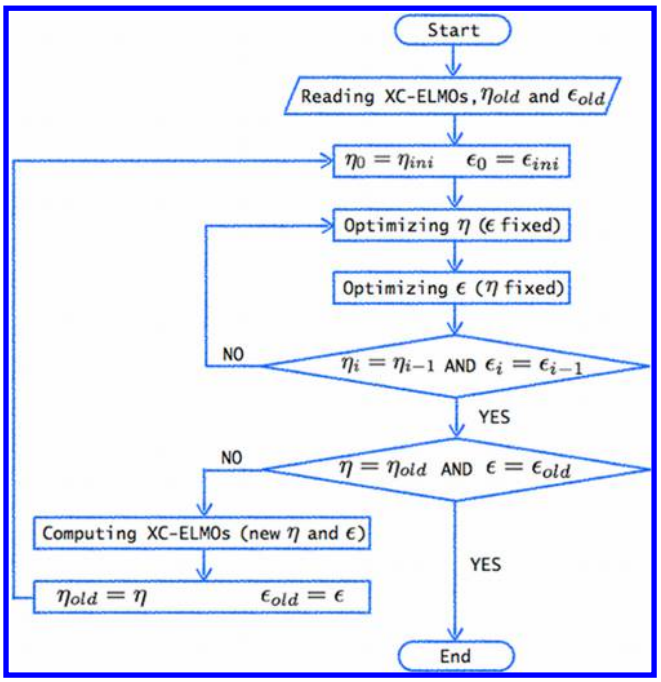



**Figure 1.** Strategy for optimizing the scale factor $\eta$ and the secondary extinction parameter $\varepsilon$.

basically consists in an optimization characterized by two nested cycles. In fact, for a given set of XC-ELMOs (read as guess Molecular Orbitals or determined at the previous iteration of the external cycle), an internal optimization of $\eta$ and $\varepsilon$ is performed. If the obtained values do not significantly differ from the ones associated with the previous set of X-ray constrained ELMOs, convergence is reached, otherwise the new parameters are used to compute new XC-ELMOs and a new optimization-iteration (external cycle) starts.

The internal optimization is essentially a two-step cycle. The first step consists in finding the $\eta$ value that minimizes the $\chi^2$ agreement statistics, keeping fixed the parameter $\varepsilon$ to the starting value or to the value obtained at the previous iteration of the internal cycle. In other words, we look for the scale factor that fulfills the following relations

$$\begin{cases} \dfrac{\partial \chi^2}{\partial \eta} = 0 \\ \dfrac{\partial^2 \chi^2}{\partial \eta^2} > 0 \end{cases} \tag{35}$$

It is easy to show that the second derivative is an always positive quantity and, therefore, the desired $\eta$ value is obtained simply setting the first derivative equal to zero:

$$\eta = \left[\sum_{\mathbf{h}} \frac{Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}||F_{\mathbf{h}}^{\mathrm{exp}}|}{\sigma_{\mathbf{h}}^2}\right]\left[\sum_{\mathbf{h}} \left(\frac{Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}|}{\sigma_{\mathbf{h}}}\right)^2\right]^{-1} \tag{36}$$

The value obtained from eq 36 is afterward kept fixed during the minimization of $\chi^2$ with respect to $\varepsilon$. To perform this minimization, a variant of the one-dimensional Brent method[43,44] is exploited, a variant that is characterized by a limited use of the information provided by the first derivative

$$\frac{\partial \chi^2}{\partial \varepsilon} = -\frac{\eta}{N_r - N_p} \sum_{\mathbf{h}} \frac{|F_{\mathbf{h}}^{\mathrm{calc}}|^3 P(\theta_{\mathbf{h}})}{\sigma_{\mathbf{h}}^2} (\eta Y_{\mathbf{h}}(\varepsilon)|F_{\mathbf{h}}^{\mathrm{calc}}| - |F_{\mathbf{h}}^{\mathrm{exp}}|)(Y_{\mathbf{h}}(\varepsilon))^5 \tag{37}$$

where

$$P(\theta_{\mathbf{h}}) = \frac{1 + \cos^4(2\theta_{\mathbf{h}})}{1 + \cos^2(2\theta_{\mathbf{h}})} \tag{38}$$

For the sake of completeness, it is important to note that when the secondary extinctions treatment is not required, $\varepsilon$ is imply set equal to zero and its optimization in the internal cycle is skipped.

## 3. ASSESSMENT OF THE TECHNIQUE

**Preliminary Information.** The strategies described in the Theory section for the determination of X-ray constrained ELMOs have been implemented modifying the version 8 of the GAMESS-UK quantum chemistry package,[45] which has been also used to perform all the calculations that will be discussed.

In order to assess the capabilities of the new technique, we have studied four different systems (see Table 1). In particular,

**Table 1. Crystallographic Structures Considered for the Computational Tests**[a]

| system | high quality | secondary extinctions |
|---|---|---|
| $\alpha$-glycine | yes ($\chi^2$ = 1.19) | yes |
| L-cysteine | yes ($\chi^2$ = 2.00) | no |
| AMPA | no ($\chi^2$ = 4.42) | yes |
| TFMF | no ($\chi^2$ = 11.70) | no |

[a]The $\chi^2$ values represent the statistical agreement between the experimental and the calculated structure factors amplitudes associated with the published structures.

we have considered two high-quality crystal structures (namely, structures with high-quality geometry positions and ADPs): the one of the $\alpha$-glycine at 23 K[46] and the one of the L-cysteine (orthorhombic phase I) at 30 K.[47] It is important to note that while for the latter the associated experimental structure factors amplitudes are not affected by secondary extinctions, in the former anisotropic secondary extinction phenomena play a crucial role. Furthermore, two other crystallographic structures of lower quality have been taken into account: the one associated with the (aminomethyl)phosphonic acid (AMPA)[48] and the one corresponding to the N-(trifluoromethyl)formamide (TFMF).[49] Also in this case, only the experimental structure factors amplitudes for the AMPA crystal are affected by secondary extinctions. Finally, since one of the goals of our investigation was also to study how the performances of the proposed technique change in function of the basis-set quality, all the computations on the systems discussed above have been carried out using the STO-3G, 6-31G, 6-31G**, and cc-pVDZ basis sets.

For each system, we have performed single point unconstrained calculations at the Restricted Hartree−Fock (RHF), B3LYP, and ELMO levels using the molecular geometries obtained from the X-ray diffraction experiments. Unlike our preliminary work,[23] for the ELMO calculations, we have adopted less partitioning localization patterns, namely, we have defined


dx.doi.org/10.1021/ct400293m | J. Chem. Theory Comput. 2013, 9, 3004−3019

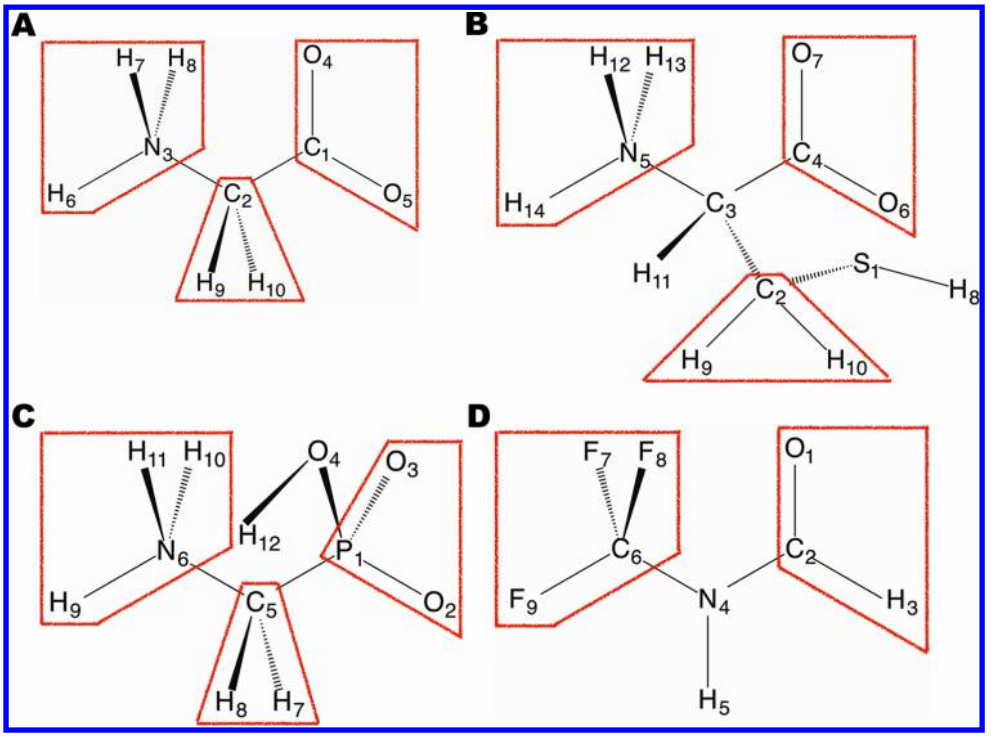



**Figure 2.** Localization schemes for the (A) $\alpha$-glycine, (B) L-cysteine, (C) (aminomethyl)phosphonic acid, and (D) N-(trifluoromethyl)-formamide systems. The red frames show fragments corresponding to functional groups.

atomic fragments to describe the core electrons, whereas, for the valence electrons, when possible, we have tried to choose subunits corresponding to large functional groups (see Figure 2). For instance, in the L-cysteine case (see Figure 2B), we have defined a fragment for the protonated amino group, one for the carboxylic group, and another one for the subunit $H_9-C_2-H_{10}$. The other fragments, which are not highlighted in the figure, correspond to the remaining single bonds. Furthermore, it is worth noting that the atomic subunit corresponding to the sulfur atom is used to describe both its core electrons and its valence lone-pairs. We have adopted the same criterion for the oxygen atom $O_4$ and the nitrogen atom $N_4$ in the AMPA and TFMF cases, respectively.

The same geometries and the same localization schemes have been used for all the X-ray constrained ELMO calculations (with and without secondary extinctions treatment), for which we have also considered the unit-cell parameters, the ADPs, and the experimental structure factors amplitudes associated with the different crystallographic structures. Except for the $\alpha$-glycine, whose data have been supplied by Destro, in all the other cases we have used the reflections intensities deposited with the corresponding papers. Experimental values characterized by $|F_h^{exp}| < 3\sigma_h$ have been rejected, which resulted in the selection of 3621, 1482, 7269, and 820 structure factors amplitudes for the $\alpha$-glycine, L-cysteine, AMPA, and TFMF systems, respectively. Furthermore, starting from $\lambda = 0$, all the X-ray constrained ELMO calculations have been carried out with ever-increasing values of the Lagrange multiplier, generally using different $\lambda$-steps for the different systems and basis-sets taken into account (see Table 2). Concerning this aspect, it was also particularly important to define a termination criterion for the XC-ELMO calculations. Since the goal of our computations is to obtain calculated structure factors amplitudes that are, on average, within one standard deviation of the experimental data, an obvious choice is to stop when the $\chi^2$ value starts being lower than 1.0. Unfortunately, as already observed by Jayatilaka and co-workers in the framework of the X-ray constrained wave function approach,[16,20,50,51] the convergence toward the desired agreement may be extremely slow, especially when the starting crystallographic data are not of very high quality. Therefore, bearing in mind that, in those situations, large values of $\lambda$ generally correspond to only minimal improvements in the $\chi^2$ statistics along with large unphysical changes in the energy, we have decided to add two further termination criteria. One considers the incremental ratio of $\chi^2$ with respect to $\lambda$

$$\left(\frac{\Delta\chi^2}{\Delta\lambda}\right)_i = \frac{\chi_i^2 - \chi_{i-1}^2}{\lambda_i - \lambda_{i-1}} \tag{39}$$

where $\chi_i^2$ and $\lambda_i$ represent the agreement statistics and the Lagrange multiplier values at the $i$-th step, respectively, and it consists in checking if the quantity defined by eq 39 is lower than 0.5 or not. This criterion allows to avoid undesirable situations in which the experimental constraint acquires a greater and greater weight in the functional given by eq 11, but the statistical agreement is only slightly affected. Finally, the third termination criterion, which is essentially energetic, verifies whether the change in the electronic energy with respect to the unconstrained ELMO wave function is, in absolute value, higher than the 0.05% of the unconstrained ELMO electronic energy itself. This last criterion is based on the observation that

**Table 2. $\lambda$-Step Values Used for the Different XC-ELMO Calculations**

| system | STO-3G | 6-31G | 6-31G** | cc-pVDZ |
|---|---|---|---|---|
| $\alpha$-glycine | 0.005 | 0.005 | 0.005 | 0.005 |
| L-cysteine | 0.2 | 0.1 | 0.05 | 0.05 |
| AMPA | 0.2 | 0.1 | 0.05 | 0.05 |
| TFMF | 0.2 | 0.1 | 0.05 | 0.05 |

Table 3. $\chi^2$ Agreement Statistics and Energy Values Corresponding to All the Unconstrained and Constrained Calculations Performed on the $\alpha$-Glycine[a]

| | no extinctions treatment[b] | | | extinctions treatment[b] | | |
|---|---|---|---|---|---|---|
| method and basis-set | $\lambda_{max}$ | $\chi^2$ | energy (au) | $\lambda_{max}$ | $\chi^2$ | energy (au) |
| | | | STO-3G | | | |
| ELMO | | 39.34 | −278.94 | | 26.31 | |
| RHF | | 39.21 | −278.96 | | 26.24 | |
| B3LYP | | 36.28 | −280.47 | | 23.37 | |
| XC-ELMO | 0.080 | 27.32 | −278.71 | 0.070 | 12.60 | −278.70 |
| | | | 6-31G | | | |
| ELMO | | 17.59 | −282.60 | | 6.87 | |
| RHF | | 17.32 | −282.64 | | 6.70 | |
| B3LYP | | 17.24 | −284.28 | | 6.07 | |
| XC-ELMO | 0.195 | 12.00 | −282.37 | 0.610 | 2.86 | −282.37 |
| | | | 6-31G** | | | |
| ELMO | | 15.59 | −282.74 | | 4.83 | |
| RHF | | 15.41 | −282.78 | | 4.73 | |
| B3LYP | | 15.26 | −284.38 | | 3.91 | |
| XC-ELMO | 0.180 | 10.43 | −282.50 | 0.370 | 2.21 | −282.64 |
| | | | cc-pVDZ | | | |
| ELMO | | 15.30 | −282.75 | | 4.55 | |
| RHF | | 15.13 | −282.79 | | 4.46 | |
| B3LYP | | 15.04 | −284.39 | | 3.74 | |
| XC-ELMO | 0.175 | 10.44 | −282.52 | 0.360 | 2.22 | −282.66 |

[a]The $\lambda_{max}$ value is shown for the constrained wave functions. [b]The scale factor $\eta$ and, when necessary, the secondary extinction parameter $\varepsilon$ for the ELMO, RHF, and B3LYP cases have been optimized using the density matrices obtained from the corresponding unconstrained calculations.

a constrained wave function, to be physically acceptable, must not have an energy that is very different from the unconstrained one.[16] Hence, when one of the three criteria just described is satisfied, the X-ray constrained ELMO calculations end. However, although the protocol for the termination of the fitting procedure here proposed is well-defined and seems reliable, it is important to observe that this still remains an open problem and further investigations in future studies will be needed.

In the following subsections we will analyze in detail the results obtained for the four systems in exam. In particular, we will study how the achievement of the desired agreement ($\chi^2 = 1$) is influenced both by the "crystallographic starting point" (namely, crystal structure quality and/or presence of secondary extinction phenomena) and by the quality of the basis-sets used for the calculations. Furthermore, we will also examine the effects of the wave function fitting on the electron density both performing topological analyses and computing real-space similarity indexes that involve the obtained charge distributions.

**Performances of the Method.** In Table 3, we have reported the $\chi^2$ and energy values associated with all the computations performed on the $\alpha$-glycine system. It is easy to observe that, among all the strategies taken into account, the new X-ray constrained ELMO technique always provides the best statistical agreements with the experimental data. Furthermore, as expected, the description improves as the basis-set quality increases, especially with sufficiently large and flexible basis-sets, such as the 6-31G** and the cc-pVDZ ones. However, in this case, since secondary extinction phenomena strongly affect the experimental diffraction data, it is worth noting that, without any correction, even the 6-31G** and the cc-pVDZ basis-sets give very poor results ($\chi^2 = 10.43$ and $\chi^2 = 10.44$, respectively). The situation significantly improves performing X-ray constrained ELMO calculations that include secondary extinctions treatments (from now on indicated in the text as XC-ELMO-Ext calculations). Of course, also in this situation, we observe an improvement of the statistical agreement as the basis-sets become larger and more flexible. In particular, using the 6-31G** and the cc-pVDZ basis-sets, we get $\chi^2 = 2.21$ and $\chi^2 = 2.22$, respectively. Nevertheless, these values are still far from the desired accuracy, and this is probably due to the fact that the secondary extinctions, which seriously affect the strong diffraction intensities associated with the $\alpha$-glycine crystals, are essentially anisotropic, while the Larson model implemented in our algorithm is able to treat properly only the isotropic cases.

To better investigate the previous aspect, in Figure 3, the normalized residuals $||F_h^{calc}| - |F_h^{exp}||/\sigma_h$ corresponding to the XC-ELMO/cc-pVDZ and XC-ELMO-Ext/cc-pVDZ calculations have been plotted against the experimental structure factors amplitudes $|F_h^{exp}|$. In Figure 3A, which represents the situation in which the secondary extinctions are completed neglected, it is easy to observe that, although most of the deviations are lower than $5\sigma_h$, the poorly modeled reflections are generally very strong, which confirms that the inadequate description provided by the XC-ELMO technique is mainly due to the complete neglect of the secondary extinction phenomena. When the Larson correction is applied, the number of large deviations drops from 59 to 26, but in Figure 3B, it is possible to see that the strong reflections are still improperly treated. Therefore, we believe that the insufficient correction of the secondary extinctions is indeed the reason why the statistical agreement with the experimental data of the $\alpha$-glycine remains far from 1.

As for the $\alpha$-glycine, the XC-ELMO strategy always provides the best $\chi^2$ values also for the L-cysteine and the description improves with the basis-set flexibility (see Table 4). In particular, it is worthwhile to note that, in the 6-31G** and cc-pVDZ cases (both with and without secondary extinctions corrections), the new technique allows to reach the desired statistical agreement. The success of the method is probably

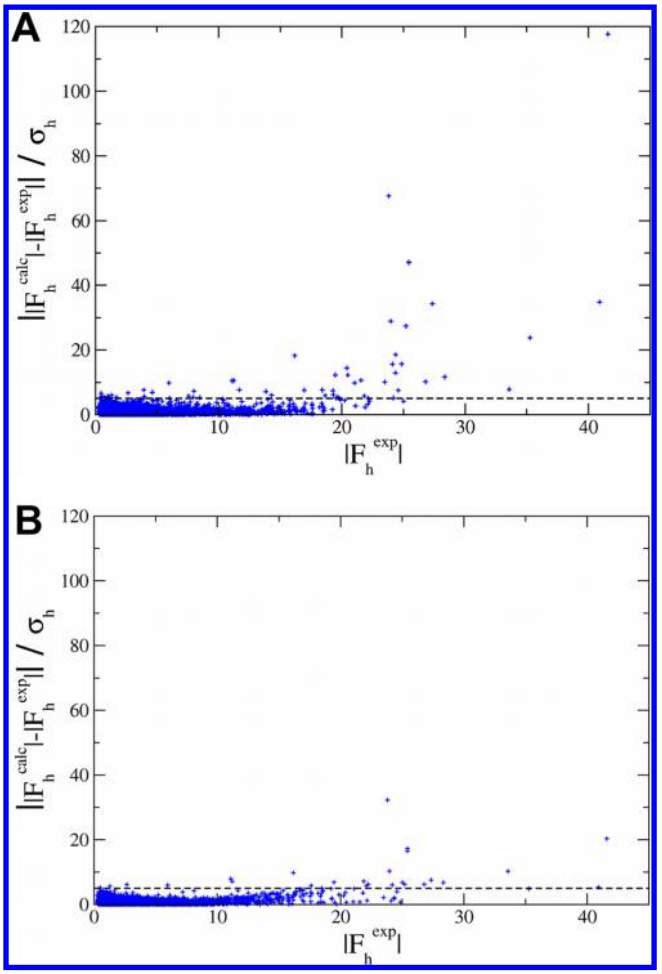



**Figure 3.** Normalized residuals associated with the (A) XC-ELMO/cc-pVDZ and (B) XC-ELMO-Ext/cc-pVDZ wave functions versus the corresponding experimental data for the α-glycine. In B, the calculated structure factors amplitudes have been corrected to take into account the secondary extinctions effects.

related to the high quality of the crystallographic structure and also to the absence of the secondary extinctions that seriously influence the strong reflections of the α-glycine crystal. This very last aspect can be clearly seen both from Table 4, where the results obtained with secondary extinctions corrections are essentially identical to the ones obtained without any treatment of the extinction phenomena, and from Figure 4, which shows the distribution of the normalized residuals of the structure factors amplitudes for the XC-ELMO/cc-pVDZ and the XC-ELMO-Ext/cc-pVDZ calculations in function of the collected experimental data. In fact, in Figure 4A, which corresponds to the case of the X-ray constrained ELMO wave function obtained without extinctions treatment, all the reflections are very well modeled (namely, all the normalized residuals are lower than 5), and the situation does not change when the Larson correction is applied (see Figure 4B).

The same main trends observed for the α-glycine and L-cysteine can be easily noted also for the (aminomethyl)-phosphonic acid (see Table 5). In fact, the best statistical agreements are obtained with the X-ray constrained ELMO technique and, above all, with the larger and more flexible basis-sets 6-31G** and cc-pVDZ. Moreover, it is worth noting that, including the Larson correction in all the XC-ELMO calculations, the description further improves (for instance, from

**Table 4. $\chi^2$ Agreement Statistics and Energy Values Corresponding to All the Unconstrained and Constrained Calculations Performed on the L-Cysteine[a]**

| method and basis-set | no extinctions treatment[b] $\lambda_{max}$ | $\chi^2$ | energy (au) | extinctions treatment[b] $\lambda_{max}$ | $\chi^2$ | energy (au) |
|---|---|---|---|---|---|---|
| STO-3G | | | | | | |
| ELMO | | 6.15 | −710.49 | | 5.94 | |
| RHF | | 6.08 | −710.52 | | 5.87 | |
| B3LYP | | 5.51 | −712.88 | | 5.32 | |
| XC-ELMO | 1.000 | 2.95 | −709.96 | 1.000 | 2.77 | −709.96 |
| 6-31G | | | | | | |
| ELMO | | 4.44 | −718.95 | | 4.36 | |
| RHF | | 4.28 | −719.01 | | 4.20 | |
| B3LYP | | 4.27 | −721.60 | | 4.16 | |
| XC-ELMO | 0.700 | 1.14 | −718.60 | 0.700 | 1.14 | −718.61 |
| 6-31G** | | | | | | |
| ELMO | | 4.24 | −719.14 | | 4.15 | |
| RHF | | 4.11 | −719.21 | | 4.03 | |
| B3LYP | | 4.04 | −721.75 | | 3.92 | |
| XC-ELMO | 0.400 | 0.99 | −718.89 | 0.400 | 0.99 | −718.90 |
| cc-pVDZ | | | | | | |
| ELMO | | 3.95 | −719.16 | | 3.88 | |
| RHF | | 3.87 | −719.23 | | 3.81 | |
| B3LYP | | 3.81 | −721.77 | | 3.72 | |
| XC-ELMO | 0.400 | 0.99 | −718.93 | 0.400 | 0.99 | −718.93 |

[a]The $\lambda_{max}$ value is shown for the constrained wave functions. [b]The scale factor η and, when necessary, the secondary extinction parameter ε for the ELMO, RHF, and B3LYP cases have been optimized using the density matrices obtained from the corresponding unconstrained calculations.

$\chi^2$ = 3.74 to $\chi^2$ = 3.47 in the 6-31G case), and this is probably due to the fact that, as already discussed, the reflections intensities of the AMPA crystals are affected by secondary extinctions. Unfortunately, even applying the Larson correction, the $\chi^2$ values remain far from the desired agreement (2.74 and 2.79 for the 6-31G** and the cc-pVDZ basis sets, respectively).

Also in this situation, to better understand the reasons for this drawback, we have considered the plots of the normalized residuals $||F_h^{calc}| - |F_h^{exp}||/\sigma_h$ in function of the experimental structure factors amplitudes, both for the XC-ELMO/cc-pVDZ (see Figure 5A) and for the XC-ELMO-Ext/cc-pVDZ (see Figure 5B) wave function. Without any treatment of secondary extinctions, 82 structure factors amplitudes are not properly modeled (normalized residuals greater than 5), and from Figure 5A, it is possible to observe that a part of them corresponds to strong reflections. When the treatment for the secondary extinctions is applied, the number of poorly modeled amplitudes reduces to 50, and it is worthwhile to note that all of them are associated with low-intensity reflections (see Figure 5B). Since it is well-known that the secondary extinctions effect is generally greater for strong reflections, it is possible to conclude that, for the AMPA crystal, the secondary extinction phenomena have been taken into account. However, it is also important to observe that many of the improperly modeled strong reflections observed in Figure 5A are also low-angle reflections whose corresponding calculated amplitudes are generally affected by errors in the ADPs. Given that the crystal structure used for all the computations performed on the AMPA system has been obtained through an Independent Atom Model (IAM) refinement, we believe that the large

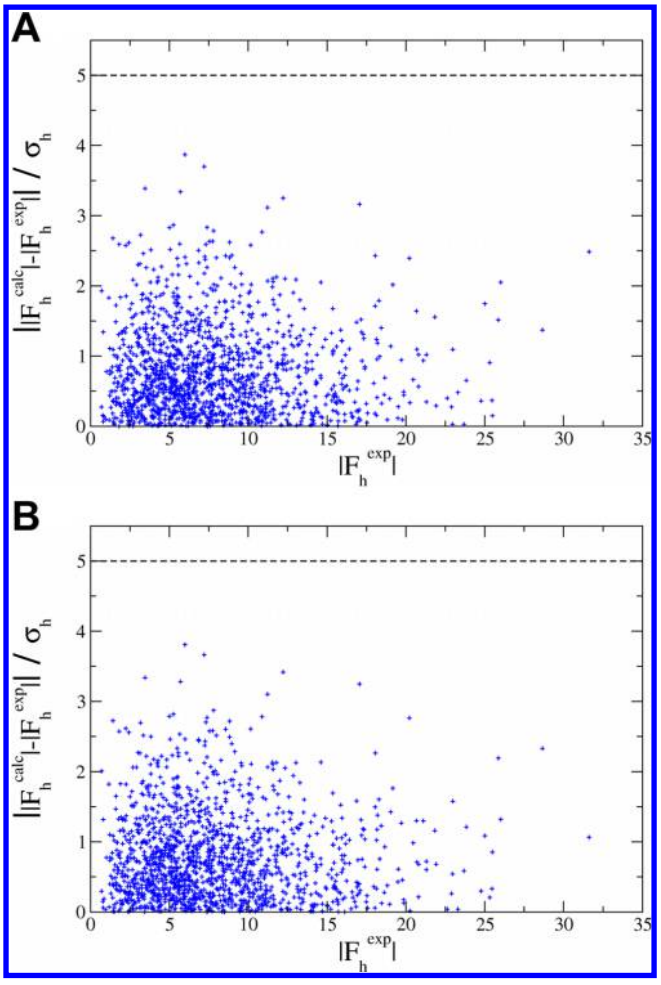



**Figure 4.** Normalized residuals associated with the (A) XC-ELMO/cc-pVDZ and (B) XC-ELMO-Ext/cc-pVDZ wave functions versus the corresponding experimental data for the L-cysteine. In B, the calculated structure factors amplitudes have been corrected to take into account the secondary extinctions effects.

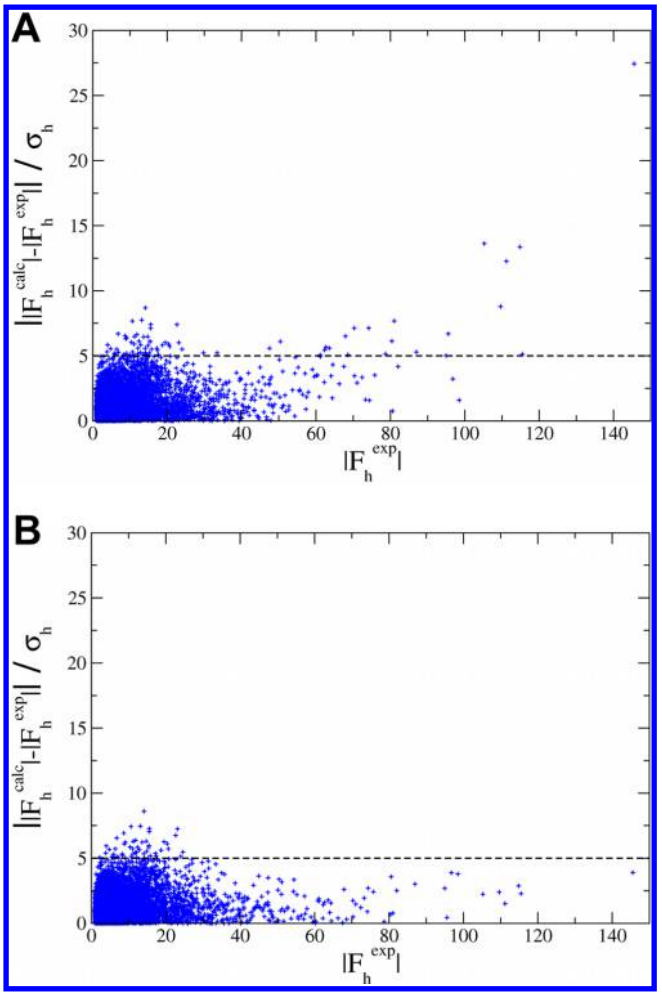



**Figure 5.** Normalized residuals associated with the (A) XC-ELMO/cc-pVDZ and (B) XC-ELMO-Ext/cc-pVDZ wave functions versus the corresponding experimental data for the (aminomethyl)phosphonic acid. In B, the calculated structure factors amplitudes have been corrected to take into account the secondary extinctions effects.

discrepancies initially observed for the strong reflections are also due to the low-quality of the ADPs used in the XC-ELMO calculations and that, consequently, the following introduction of the extra parameter $\varepsilon$ for the secondary extinctions treatment has improperly concealed this problem. To overcome this drawback, namely to disentangle the secondary extinction effects from the poor modeling of the thermal parameters, we would need to perform a further refinement of the crystal structure. For instance, following Jayatilaka and Dittrich, we could perform a Hirshfeld atom refinement[39] or we could even develop and apply a new ELMO-based refinement technique in which the XC-ELMOs previously determined on the low-quality structure are kept frozen, while the molecular geometry and the thermal parameters are optimized against the experimental structure factor amplitudes. The refined geometry positions and ADPs will eventually provide improved agreements with the experimental data and will be better starting points for further XC-ELMO calculations that will certainly lead to $\chi^2$ values closer to the desired accuracy.

Finally, in Table 6, the $\chi^2$ and the energy values for the N-(trifluoromethyl)formamide system are shown. As for the previous cases, we observe that the XC-ELMO wave functions provide the best statistical agreements, which improve with basis-sets of better quality. Nevertheless, in this case, even the best description ($\chi^2$ = 3.53 for the 6-31G** basis-set) is quite far from the desired value. Furthermore, it is important to note that the XC-ELMO results obtained with the treatment of the secondary extinctions are practically identical to the ones obtained without any correction. This can be simply explained considering the fact that the diffraction data used for the calculations are not affected by secondary extinctions (see Table 1). This is further confirmed by the plots of the normalized residuals distributions for the cc-pVDZ X-ray constrained ELMO wave functions. In fact, all the 23 poorly modeled amplitudes (see Figure 6A) essentially correspond to weak reflections, and as expected, the situation remains unchanged when the Larson correction is included (see Figure 6B). Therefore, we believe that the nonsatisfactory results provided by the X-ray constrained ELMO approach are strictly connected to the very low-quality of the starting TFMF crystallographic structure, which is indeed the responsible for the "outliers" observed in Figure 6.

**Fitting Effects on the Electron Density.** In order to study the effects of the wave function fitting on the electron density, we have performed topological analyses of the obtained cc-pVDZ charge distributions, excluding only the ones corresponding to the X-ray constrained ELMO wave functions that

Table 5. $\chi^2$ Agreement Statistics and Energy Values Corresponding to All the Unconstrained and Constrained Calculations Performed on the (Aminomethyl)Phosphonic Acid[a]

| | no extinctions treatment[b] | | | extinctions treatment[b] | | |
|---|---|---|---|---|---|---|
| method and basis-set | $\lambda_{max}$ | $\chi^2$ | energy (au) | $\lambda_{max}$ | $\chi^2$ | energy (au) |
| | | | STO-3G | | | |
| ELMO | | 21.77 | −652.58 | | 18.88 | |
| RHF | | 20.82 | −652.62 | | 18.47 | |
| B3LYP | | 18.51 | −654.80 | | 16.30 | |
| XC-ELMO | 0.400 | 13.62 | −651.95 | 0.400 | 11.68 | −651.97 |
| | | | 6-31G | | | |
| ELMO | | 8.28 | −660.67 | | 7.73 | |
| RHF | | 7.97 | −660.73 | | 7.44 | |
| B3LYP | | 7.93 | −663.14 | | 7.34 | |
| XC-ELMO | 0.900 | 3.74 | −660.14 | 0.900 | 3.47 | −660.18 |
| | | | 6-31G** | | | |
| ELMO | | 7.41 | −661.00 | | 6.82 | |
| RHF | | 7.21 | −661.07 | | 6.63 | |
| B3LYP | | 7.38 | −663.42 | | 6.73 | |
| XC-ELMO | 0.950 | 3.01 | −660.48 | 0.750 | 2.74 | −660.59 |
| | | | cc-pVDZ | | | |
| ELMO | | 7.24 | −660.99 | | 6.66 | |
| RHF | | 7.06 | −661.06 | | 6.51 | |
| B3LYP | | 7.30 | −663.41 | | 6.67 | |
| XC-ELMO | 0.900 | 3.07 | −660.52 | 0.700 | 2.79 | −660.62 |

[a]The $\lambda_{max}$ value is shown for the constrained wave functions. [b]The scale factor $\eta$ and, when necessary, the secondary extinction parameter $\varepsilon$ for the ELMO, RHF, and B3LYP cases have been optimized using the density matrices obtained from the corresponding unconstrained calculations.

Table 6. $\chi^2$ Agreement Statistics and Energy Values Corresponding to All the Unconstrained and Constrained Calculations Performed on the N-(trifluoromethyl)formamide[a]

| | no extinctions treatment[b] | | | extinctions treatment[b] | | |
|---|---|---|---|---|---|---|
| method and basis-set | $\lambda_{max}$ | $\chi^2$ | energy (au) | $\lambda_{max}$ | $\chi^2$ | energy (au) |
| | | | STO-3G | | | |
| ELMO | | 12.11 | −497.58 | | 11.51 | |
| RHF | | 11.99 | −497.64 | | 11.37 | |
| B3LYP | | 11.06 | −499.76 | | 10.53 | |
| XC-ELMO | 0.400 | 6.56 | −497.07 | 0.400 | 6.56 | −497.11 |
| | | | 6-31G | | | |
| ELMO | | 10.04 | −504.28 | | 10.03 | |
| RHF | | 9.83 | −504.34 | | 9.83 | |
| B3LYP | | 9.53 | −506.76 | | 9.53 | |
| XC-ELMO | 0.400 | 4.50 | −503.84 | 0.400 | 4.47 | −503.84 |
| | | | 6-31G** | | | |
| ELMO | | 9.75 | −504.47 | | 9.73 | |
| RHF | | 9.44 | −504.55 | | 9.43 | |
| B3LYP | | 9.19 | −506.91 | | 9.18 | |
| XC-ELMO | 0.450 | 3.53 | −504.01 | 0.450 | 3.53 | −504.01 |
| | | | cc-pVDZ | | | |
| ELMO | | 9.75 | −504.50 | | 9.74 | |
| RHF | | 9.47 | −504.57 | | 9.45 | |
| B3LYP | | 9.22 | −506.95 | | 9.22 | |
| XC-ELMO | 0.450 | 3.63 | −504.04 | 0.450 | 3.62 | −504.05 |

[a]The $\lambda_{max}$ value is shown for the constrained wave functions. [b]The scale factor $\eta$ and, when necessary, the secondary extinction parameter $\varepsilon$ for the ELMO, RHF, and B3LYP cases have been optimized using the density matrices obtained from the corresponding unconstrained calculations.

do not include secondary extinctions corrections. In particular, using the AIMAll package (Version 12.11.09),[52] we have determined the main properties of all the bond critical points and we have computed the integrated net charges associated with all the atoms.

In Tables 7 and 8, we have reported the results obtained for the $\alpha$-glycine and, focusing on the XC-ELMO-Ext charge distribution, it is possible to observe that, except for the bonds in which the $C_2$ atom is involved, the positions of the bond critical points do not change significantly compared to the unconstrained ELMO electron density. Furthermore, considering the values of the electron distributions at the bond critical points ($\rho(\mathbf{r}_b)$), it is worthwhile to note the case of the carboxylic group. In fact, for the XC-ELMO-Ext charge density,

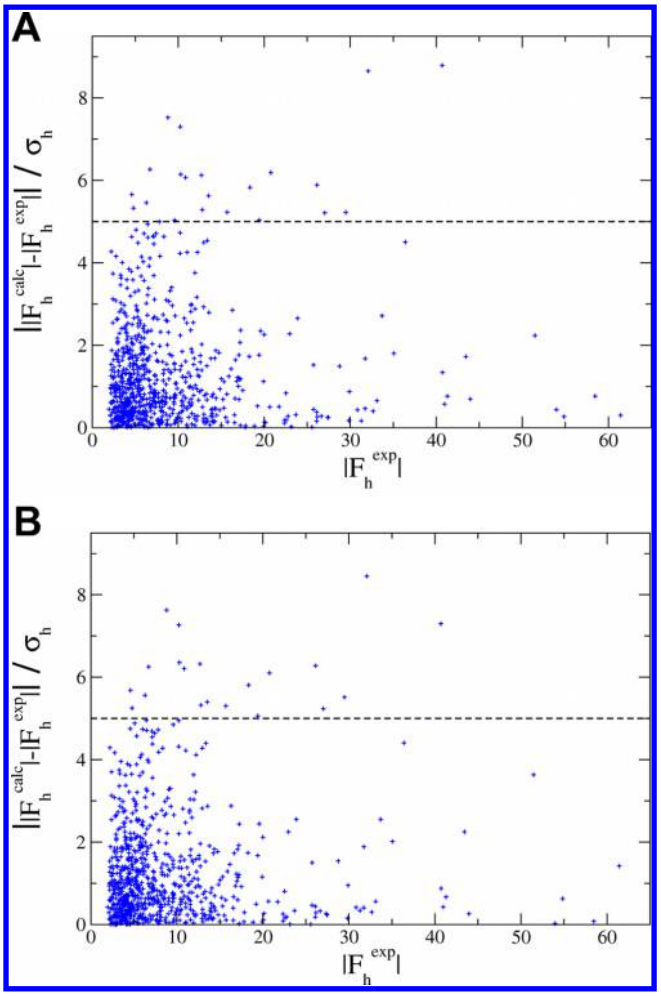



**Figure 6.** Normalized residuals associated with the (A) XC-ELMO/cc-pVDZ and (B) XC-ELMO-Ext/cc-pVDZ wave functions versus the corresponding experimental data for the N-(trifluoromethyl)-formamide. In B, the calculated structure factors amplitudes have been corrected to take into account the secondary extinctions effects.

we observe that the two C−O bonds are characterized by different $\rho(\mathbf{r}_b)$ values (0.392 and 0.373), which is in contrast with the quasi-equivalence observed for all the unconstrained calculations. Another important feature of the XC-ELMO-Ext charge distribution is the decrease of the $\rho(\mathbf{r}_b)$ values for all the N−H bonds (and, to a lower extent, for all the C−H bonds) compared to the unconstrained ELMO case. This is in line with the results just sketched in the B3LYP situation, and it is also associated with an increase of the net charges for almost all the hydrogen atoms.

In Table 9, for the L-cysteine, we can note that, for all the C−H and N−H bonds, the XC-ELMO-Ext electron density is characterized by large decreases of the $\rho(\mathbf{r}_b)$ values (greater than 15%). Furthermore, as for the α-glycine, the carboxylic group represents an interesting case. Nevertheless, here, the fitting entails large decreases in the amount of charge at the C−O critical points (10.2% and 7.3%), which eventually result into identical $\rho(\mathbf{r}_b)$ values. For the sake of completeness, it is also worth observing that large increases in the electron density at the $N_5-C_3$ and $S_1-C_2$ bond critical points are also observed (12.8% and 9.4%, respectively). Moreover, for the $N_5-C_3$ case, it is possible to note that, as a result of the fitting, the net charge of the nitrogen atom increases by 0.25 au while the

**Table 7. Topological Analysis of cc-pVDZ Charge Distributions for the α-Glycine[a]**

| | methods | | | |
|---|---|---|---|---|
| BCPs properties | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| | | $C_1-C_2$ | | |
| $d_A$ | 0.682 | 0.683 | 0.703 | 0.761 |
| $\rho(\mathbf{r}_b)$ | 0.259 | 0.258 | 0.247 | 0.247 |
| | | $N_3-C_2$ | | |
| $d_A$ | 1.018 | 1.013 | 0.962 | 0.944 |
| $\rho(\mathbf{r}_b)$ | 0.223 | 0.225 | 0.230 | 0.241 |
| | | $O_4-C_1$ | | |
| $d_A$ | 0.850 | 0.850 | 0.843 | 0.848 |
| $\rho(\mathbf{r}_b)$ | 0.378 | 0.378 | 0.373 | 0.392 |
| | | $O_5-C_1$ | | |
| $d_A$ | 0.850 | 0.850 | 0.842 | 0.847 |
| $\rho(\mathbf{r}_b)$ | 0.379 | 0.380 | 0.374 | 0.373 |
| | | $N_3-H_6$ | | |
| $d_A$ | 0.805 | 0.803 | 0.797 | 0.829 |
| $\rho(\mathbf{r}_b)$ | 0.327 | 0.328 | 0.321 | 0.305 |
| | | $N_3-H_7$ | | |
| $d_A$ | 0.808 | 0.807 | 0.801 | 0.814 |
| $\rho(\mathbf{r}_b)$ | 0.331 | 0.332 | 0.325 | 0.315 |
| | | $N_3-H_8$ | | |
| $d_A$ | 0.805 | 0.804 | 0.798 | 0.812 |
| $\rho(\mathbf{r}_b)$ | 0.332 | 0.333 | 0.327 | 0.315 |
| | | $C_2-H_9$ | | |
| $d_A$ | 0.677 | 0.682 | 0.697 | 0.734 |
| $\rho(\mathbf{r}_b)$ | 0.295 | 0.293 | 0.286 | 0.288 |
| | | $C_2-H_{10}$ | | |
| $d_A$ | 0.686 | 0.692 | 0.706 | 0.731 |
| $\rho(\mathbf{r}_b)$ | 0.289 | 0.288 | 0.281 | 0.289 |

[a]For each bond critical point (BCP) A−B, its distance from the nucleus A ($d_A$, in Å) and its electron density value ($\rho(\mathbf{r}_b)$, in au) are shown.

**Table 8. Topological Analysis of cc-pVDZ Charge Distributions for the α-Glycine: Integrated Net Atomic Charges (in au)**

| | methods | | | |
|---|---|---|---|---|
| atoms | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| $C_1$ | 2.09 | 2.08 | 1.74 | 1.84 |
| $C_2$ | 0.43 | 0.40 | 0.27 | 0.26 |
| $N_3$ | −1.32 | −1.31 | −1.17 | −1.30 |
| $O_4$ | −1.42 | −1.40 | −1.22 | −1.38 |
| $O_5$ | −1.37 | −1.36 | −1.18 | −1.31 |
| $H_6$ | 0.51 | 0.50 | 0.47 | 0.60 |
| $H_7$ | 0.57 | 0.56 | 0.52 | 0.55 |
| $H_8$ | 0.56 | 0.54 | 0.51 | 0.59 |
| $H_9$ | −0.03 | −0.02 | 0.02 | 0.08 |
| $H_{10}$ | −0.02 | 0.00 | 0.03 | 0.06 |

one corresponding to the carbon atom decreases by 0.08 au (see Table 10). Therefore, a possible consequence of the fitting procedure consists in a redistribution of electronic charge from $N_5$ to $C_3$. Concerning the positions of the critical points, significant changes are observed only for the bonds in which the $C_2$ and $C_3$ atoms are involved.

The effects of the wave function fitting have been studied also for the (aminomethyl)phosphonic acid, and in Table 11, considering the XC-ELMO-Ext charge distribution, it is immediately observable that all the bonds that do not involve


dx.doi.org/10.1021/ct400293m | J. Chem. Theory Comput. 2013, 9, 3004−3019

Table 9. Topological Analysis of cc-pVDZ Charge Distributions for the L-Cysteine[a]

| BCPs properties | methods | | | |
|---|---|---|---|---|
| | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| | | $S_1-C_2$ | | |
| $d_A$ | 0.939 | 0.942 | 0.949 | 1.018 |
| $\rho(\mathbf{r}_b)$ | 0.185 | 0.185 | 0.178 | 0.203 |
| | | $C_3-C_2$ | | |
| $d_A$ | 0.780 | 0.785 | 0.777 | 0.789 |
| $\rho(\mathbf{r}_b)$ | 0.257 | 0.255 | 0.246 | 0.253 |
| | | $C_3-C_4$ | | |
| $d_A$ | 0.848 | 0.845 | 0.827 | 0.799 |
| $\rho(\mathbf{r}_b)$ | 0.257 | 0.255 | 0.244 | 0.249 |
| | | $N_5-C_3$ | | |
| $d_A$ | 1.017 | 1.013 | 0.960 | 0.946 |
| $\rho(\mathbf{r}_b)$ | 0.222 | 0.223 | 0.228 | 0.250 |
| | | $C_4-O_6$ | | |
| $d_A$ | 0.403 | 0.403 | 0.410 | 0.396 |
| $\rho(\mathbf{r}_b)$ | 0.388 | 0.389 | 0.384 | 0.348 |
| | | $C_4-O_7$ | | |
| $d_A$ | 0.410 | 0.410 | 0.417 | 0.403 |
| $\rho(\mathbf{r}_b)$ | 0.375 | 0.375 | 0.370 | 0.348 |
| | | $S_1-H_8$ | | |
| $d_A$ | 0.636 | 0.636 | 0.665 | 0.653 |
| $\rho(\mathbf{r}_b)$ | 0.222 | 0.222 | 0.222 | 0.227 |
| | | $C_2-H_9$ | | |
| $d_A$ | 0.558 | 0.563 | 0.605 | 0.619 |
| $\rho(\mathbf{r}_b)$ | 0.366 | 0.365 | 0.360 | 0.311 |
| | | $C_2-H_{10}$ | | |
| $d_A$ | 0.631 | 0.634 | 0.653 | 0.646 |
| $\rho(\mathbf{r}_b)$ | 0.350 | 0.349 | 0.343 | 0.292 |
| | | $C_3-H_{11}$ | | |
| $d_A$ | 0.544 | 0.555 | 0.599 | 0.598 |
| $\rho(\mathbf{r}_b)$ | 0.399 | 0.397 | 0.390 | 0.319 |
| | | $N_5-H_{12}$ | | |
| $d_A$ | 0.650 | 0.650 | 0.648 | 0.645 |
| $\rho(\mathbf{r}_b)$ | 0.537 | 0.540 | 0.531 | 0.451 |
| | | $N_5-H_{13}$ | | |
| $d_A$ | 0.683 | 0.680 | 0.681 | 0.685 |
| $\rho(\mathbf{r}_b)$ | 0.483 | 0.486 | 0.477 | 0.409 |
| | | $N_5-H_{14}$ | | |
| $d_A$ | 0.694 | 0.691 | 0.691 | 0.707 |
| $\rho(\mathbf{r}_b)$ | 0.456 | 0.457 | 0.448 | 0.377 |

[a]For each bond critical point (BCP) A-B, its distance from the nucleus A ($d_A$, in Å) and its electron density value ($\rho(\mathbf{r}_b)$, in au) are shown.

hydrogen atoms are characterized by an increase of the corresponding $\rho(\mathbf{r}_b)$ values. In particular, it is interesting to note that the decreases in the amount of charge at the bond critical points for the three P−O bonds are also associated with quite large variation in the net charges of the four atoms involved (see Table 12): a decrease for the phosphorus atom and an increase for the oxygen atoms. Therefore, one of the fitting consequences is probably a redistribution of electron density from the oxygen atoms (especially $O_2$ and $O_3$) to the phosphorus one. On the contrary, for the bonds that involve hydrogen atoms, the amount of charge at the bond critical points significantly decreases, a feature that is only sketched in the B3LYP electron density. For the N−H bonds of the amino group, this decrease gets along with a large increase in the net charge for the nitrogen atom and with a decrease in the net charges for the corresponding hydrogen atoms (especially $H_9$

Table 10. Topological Analysis of cc-pVDZ Charge Distributions for the L-Cysteine: Integrated Net Atomic Charges (in au)

| atoms | methods | | | |
|---|---|---|---|---|
| | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| $S_1$ | 0.37 | 0.41 | 0.29 | 0.08 |
| $C_2$ | 0.26 | 0.25 | 0.06 | 0.30 |
| $C_3$ | 0.51 | 0.46 | 0.30 | 0.43 |
| $C_4$ | 2.09 | 2.08 | 1.75 | 1.84 |
| $N_5$ | −1.27 | −1.24 | −1.18 | −1.02 |
| $O_6$ | −1.40 | −1.38 | −1.21 | −1.29 |
| $O_7$ | −1.41 | −1.39 | −1.21 | −1.38 |
| $H_8$ | −0.36 | −0.38 | −0.27 | −0.24 |
| $H_9$ | −0.19 | −0.18 | −0.07 | −0.07 |
| $H_{10}$ | −0.01 | 0.00 | 0.05 | 0.02 |
| $H_{11}$ | −0.18 | −0.16 | −0.04 | −0.11 |
| $H_{12}$ | 0.53 | 0.52 | 0.51 | 0.45 |
| $H_{13}$ | 0.55 | 0.54 | 0.53 | 0.49 |
| $H_{14}$ | 0.50 | 0.49 | 0.48 | 0.50 |

Table 11. Topological Analysis of cc-pVDZ Charge Distributions for the (Aminomethyl)phosphonic Acid[a]

| BCPs properties | methods | | | |
|---|---|---|---|---|
| | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| | | $P_1-O_2$ | | |
| $d_A$ | 0.598 | 0.598 | 0.602 | 0.596 |
| $\rho(\mathbf{r}_b)$ | 0.211 | 0.211 | 0.210 | 0.232 |
| | | $P_1-O_3$ | | |
| $d_A$ | 0.603 | 0.603 | 0.608 | 0.603 |
| $\rho(\mathbf{r}_b)$ | 0.205 | 0.204 | 0.203 | 0.229 |
| | | $P_1-O_4$ | | |
| $d_A$ | 0.620 | 0.621 | 0.626 | 0.617 |
| $\rho(\mathbf{r}_b)$ | 0.169 | 0.166 | 0.169 | 0.198 |
| | | $C_5-P_1$ | | |
| $d_A$ | 1.156 | 1.156 | 1.140 | 1.147 |
| $\rho(\mathbf{r}_b)$ | 0.150 | 0.148 | 0.149 | 0.177 |
| | | $C_5-N_6$ | | |
| $d_A$ | 0.472 | 0.474 | 0.524 | 0.539 |
| $\rho(\mathbf{r}_b)$ | 0.216 | 0.216 | 0.221 | 0.244 |
| | | $C_5-H_7$ | | |
| $d_A$ | 0.606 | 0.607 | 0.630 | 0.627 |
| $\rho(\mathbf{r}_b)$ | 0.368 | 0.368 | 0.362 | 0.295 |
| | | $C_5-H_8$ | | |
| $d_A$ | 0.561 | 0.562 | 0.604 | 0.635 |
| $\rho(\mathbf{r}_b)$ | 0.379 | 0.378 | 0.373 | 0.298 |
| | | $N_6-H_9$ | | |
| $d_A$ | 0.691 | 0.688 | 0.688 | 0.662 |
| $\rho(\mathbf{r}_b)$ | 0.463 | 0.464 | 0.456 | 0.413 |
| | | $N_6-H_{10}$ | | |
| $d_A$ | 0.677 | 0.675 | 0.677 | 0.659 |
| $\rho(\mathbf{r}_b)$ | 0.495 | 0.497 | 0.488 | 0.407 |
| | | $N_6-H_{11}$ | | |
| $d_A$ | 0.734 | 0.732 | 0.731 | 0.737 |
| $\rho(\mathbf{r}_b)$ | 0.407 | 0.408 | 0.401 | 0.367 |
| | | $O_4-H_{12}$ | | |
| $d_A$ | 0.773 | 0.772 | 0.765 | 0.792 |
| $\rho(\mathbf{r}_b)$ | 0.372 | 0.370 | 0.366 | 0.351 |

[a]For each bond critical point (BCP) A−B, its distance from the nucleus A ($d_A$, in Å) and its electron density value ($\rho(\mathbf{r}_b)$, in au) are shown.


dx.doi.org/10.1021/ct400293m | J. Chem. Theory Comput. 2013, 9, 3004−3019

**Table 12. Topological Analysis of cc-pVDZ Charge Distributions for the (Aminomethyl)phosphonic Acid: Integrated Net Atomic Charges (in au)**

| | methods | | | |
|---|---|---|---|---|
| atoms | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| $P_1$ | 3.85 | 3.85 | 3.54 | 3.42 |
| $O_2$ | −1.60 | −1.60 | −1.47 | −1.48 |
| $O_3$ | −1.61 | −1.60 | −1.48 | −1.50 |
| $O_4$ | −1.51 | −1.47 | −1.38 | −1.48 |
| $C_5$ | 0.07 | 0.04 | −0.17 | −0.02 |
| $N_6$ | −1.26 | −1.24 | −1.17 | −0.82 |
| $H_7$ | −0.04 | −0.04 | 0.03 | −0.02 |
| $H_8$ | −0.15 | −0.15 | −0.04 | −0.05 |
| $H_9$ | 0.51 | 0.50 | 0.49 | 0.34 |
| $H_{10}$ | 0.56 | 0.54 | 0.53 | 0.39 |
| $H_{11}$ | 0.54 | 0.52 | 0.51 | 0.52 |
| $H_{12}$ | 0.64 | 0.65 | 0.61 | 0.69 |

and $H_{10}$). This can be explained hypothesizing that another effect of the wave function fitting consists in a shift of electronic charge from $N_6$ to $H_9$, $H_{10}$ and $H_{11}$. For the sake of completeness, it is also worth noting that, both in the B3LYP and, to a larger extent, in the XC-ELMO-Ext charge density, the positions of the bond critical points remain generally unchanged, except for almost all the bonds involving the $C_5$ atom.

We have also examined the fitting effects on the electron density of the N-(trifluoromethyl)formamide (see Tables 13 and 14). Comparing the XC-ELMO-Ext charge distribution to

**Table 13. Topological Analysis of cc-pVDZ Charge Distributions for the N-(trifluoromethyl)formamide[a]**

| | methods | | | |
|---|---|---|---|---|
| BCPs properties | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| | | $O_1-C_2$ | | |
| $d_A$ | 0.824 | 0.824 | 0.819 | 0.819 |
| $\rho(\mathbf{r}_b)$ | 0.408 | 0.409 | 0.405 | 0.374 |
| | | $C_2-H_3$ | | |
| $d_A$ | 0.662 | 0.658 | 0.670 | 0.637 |
| $\rho(\mathbf{r}_b)$ | 0.340 | 0.338 | 0.328 | 0.315 |
| | | $C_2-N_4$ | | |
| $d_A$ | 0.455 | 0.446 | 0.482 | 0.432 |
| $\rho(\mathbf{r}_b)$ | 0.323 | 0.313 | 0.313 | 0.281 |
| | | $N_4-H_5$ | | |
| $d_A$ | 0.738 | 0.737 | 0.733 | 0.823 |
| $\rho(\mathbf{r}_b)$ | 0.379 | 0.385 | 0.381 | 0.288 |
| | | $N_4-C_6$ | | |
| $d_A$ | 0.825 | 0.848 | 0.816 | 0.833 |
| $\rho(\mathbf{r}_b)$ | 0.329 | 0.325 | 0.318 | 0.344 |
| | | $C_6-F_7$ | | |
| $d_A$ | 0.437 | 0.439 | 0.462 | 0.447 |
| $\rho(\mathbf{r}_b)$ | 0.277 | 0.278 | 0.285 | 0.269 |
| | | $C_6-F_8$ | | |
| $d_A$ | 0.437 | 0.439 | 0.463 | 0.447 |
| $\rho(\mathbf{r}_b)$ | 0.276 | 0.277 | 0.284 | 0.283 |
| | | $C_6-F_9$ | | |
| $d_A$ | 0.431 | 0.432 | 0.451 | 0.432 |
| $\rho(\mathbf{r}_b)$ | 0.280 | 0.279 | 0.287 | 0.270 |

[a]For each bond critical point (BCP) A−B, its distance from the nucleus A ($d_A$, in Å) and its electron density value ($\rho(\mathbf{r}_b)$, in au) are shown.

**Table 14. Topological Analysis of cc-pVDZ Charge Distributions for the N-(trifluoromethyl)formamide: Integrated Net Atomic Charges (in au)**

| | methods | | | |
|---|---|---|---|---|
| atoms | ELMO | RHF | B3LYP | XC-ELMO-Ext |
| $O_1$ | −1.26 | −1.27 | −1.11 | −1.06 |
| $C_2$ | 1.88 | 1.85 | 1.52 | 1.71 |
| $H_3$ | −0.01 | −0.02 | 0.01 | 0.02 |
| $N_4$ | −1.51 | −1.47 | −1.22 | −1.74 |
| $H_5$ | 0.52 | 0.52 | 0.48 | 0.75 |
| $C_6$ | 2.42 | 2.43 | 2.09 | 2.27 |
| $F_7$ | −0.68 | −0.68 | −0.59 | −0.59 |
| $F_8$ | −0.68 | −0.68 | −0.59 | −0.60 |
| $F_9$ | −0.69 | −0.69 | −0.60 | −0.77 |

the unconstrained ELMO one, it is easy to see that all but the C−F bonds are characterized by large changes in the $\rho(\mathbf{r}_b)$ values. In particular, it is interesting to note that the decreases of the $\rho(\mathbf{r}_b)$ values for the $O_1-C_2$ and $C_2-H_3$ bonds are associated with an increase of net charge for the $O_1$ and the $H_3$ atoms and with a decrease of net charge for the $C_2$ atom. Another fitting effect on the TFMF charge distribution probably consists in a redistribution of electronic charge from $H_5$ to $N_4$. In fact, for the $N_4-H_5$ bond, which is characterized by a very large decrease of $\rho(\mathbf{r}_b)$ (24%), we have that the net charge of the nitrogen atom decreases by 0.23 au, whereas the one corresponding to the hydrogen atom increases by the same amount. Moreover, among all the bond critical points, the $N_4-H_5$ one is also characterized by a position that has undergone the most significant change as a result of the fitting process (11.4%).

From a careful analysis of Tables 8, 10, 12, and 14, it is also possible to observe that, in the majority of cases, the integrated net atomic charges associated with the XC-ELMO-Ext electron distributions are the closest to the B3LYP ones (for a qualitative comparison, see also the bar-graphs of the atomic charges in the Supporting Information), and as shown in Table 15, for

**Table 15. Average Values of the Absolute Differences (in au) Between the B3LYP Integrated Net Atomic Charges and the ELMO, RHF, and XC-ELMO-Ext Ones (cc-pVDZ Basis-set)**

| system | α-glycine | L-cysteine | AMPA | TFMF |
|---|---|---|---|---|
| ELMO | 0.13 | 0.13 | 0.11 | 0.16 |
| RHF | 0.11 | 0.12 | 0.10 | 0.16 |
| XC-ELMO-Ext | 0.09 | 0.10 | 0.10 | 0.16 |

the α-glycine and the L-cysteine, this also goes along with the fact that the average value of the absolute differences between the B3LYP and XC-ELMO-Ext charges is the lowest. These observations might be interpreted as an indication that the X-ray constrained wave functions are able to introduce part of the electron correlation. Nevertheless, more cagily, we believe that the obtained results can be considered a starting point for more exhaustive studies involving the comparisons of X-ray constrained electron densities with different correlated charge distributions (e.g., charge distributions obtained from Coupled Cluster calculations or from various type of Density Functionals).

Finally, to provide a further quantitative measure of the fitting effects, for each system under investigation, we have computed two real-space similarity indexes between the unconstrained ELMO electron density and the other electron


dx.doi.org/10.1021/ct400293m | J. Chem. Theory Comput. 2013, 9, 3004−3019

distributions already taken into account for the topological analyses. In particular, we have considered the Real-Space R value (RSR)[53] and the Walker−Mezey index $L(a,a')$.[54] The former is simply defined as

$$\mathrm{RSR}(\rho_x, \rho_y) = 100\frac{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) - \rho_y(\mathbf{r}_i)|}{\sum_{i=1}^{n_p} |\rho_x(\mathbf{r}_i) + \rho_y(\mathbf{r}_i)|} \quad (40)$$

with $n_p$ as the number of charge density grid points, and the complete similarity corresponds to RSR = 0. The latter compares point-by-point two charge distributions in the space bound by the isosurfaces characterized by the values $a$ and $a'$ (see the Supporting Information for the complete definition) and one has the complete similarity when its value is equal to 100. As expected, it is possible to observe that, for all the systems, the XC-ELMO-Ext electron density is the least similar charge distribution to the unconstrained ELMO one (see Tables 16 and 17). Furthermore, bearing in mind that changing the values $a$ and $a'$ we can compare the electron distributions in different regions, we have noted that, while in the core domains the $L(a,a')$ values between the ELMO and the XC-ELMO-Ext charge densities are still quite high, far from the nuclei they decrease (see the Supporting Information for further data). However, we have also to observe that the Walker−Mezey index measures the similarity in a percentage sense and, since in the core regions the electron density values are very high, quite large absolute differences between the charge distributions can be certainly observed in proximity of the nuclei. Therefore, we can conclude that the fitting procedure proposed and studied in this paper entails significant redistributions of electronic charge not only in the valence domains but also in the core ones.

**Table 16. Real-Space R Values Corresponding to the Comparison of the ELMO Charge Distribution with the RHF, B3LYP, and XC-ELMO-Ext Electron Densities for All the Systems in Exam (cc-pVDZ Basis-set)**

| system | α-glycine | L-cysteine | AMPA | TFMF |
|---|---|---|---|---|
| RHF | 0.26 | 0.27 | 0.38 | 0.38 |
| B3LYP | 1.09 | 0.86 | 1.01 | 0.96 |
| XC-ELMO-Ext | 1.46 | 2.03 | 2.10 | 2.83 |

**Table 17. Values of the Walker−Mezey Similarity index $L(0.001,10)$ Corresponding to the Comparison of the ELMO Charge Distribution with the RHF, B3LYP, and XC-ELMO-Ext Electron Densities for All the Systems in Exam (cc-pVDZ Basis-set)**

| system | α-glycine | L-cysteine | AMPA | TFMF |
|---|---|---|---|---|
| RHF | 97.94 | 97.84 | 96.67 | 97.19 |
| B3LYP | 94.07 | 95.02 | 93.05 | 94.27 |
| XC-ELMO-Ext | 92.84 | 89.24 | 89.61 | 87.25 |

## 4. CONCLUSIONS AND PERSPECTIVES

In this paper, we have described a new strategy that, combining the X-ray constrained wave function approach introduced by Jayatilaka with the ELMO technique developed by Stoll, enables to determine Molecular Orbitals that are strictly localized on small molecular fragments (i.e., atoms, bonds or functional groups) and that, at the same time, reproduce sets of collected X-ray diffraction data within a predefined accuracy. Therefore, since the novel method allows to obtain quantum mechanically rigorous electron densities that can be interpreted in terms of traditional chemical concepts, the strategy can be considered as a new useful tool for the determination and the analysis of charge distributions from X-ray experiments. It is worthwhile to note that, in comparison to our preliminary work, the possible effects of isotropic secondary extinctions have been properly taken into account through the introduction of an additional parameter, and furthermore, a well-defined and reliable protocol to establish the termination of the fitting procedure has been introduced.

Our computational tests have shown that the calculation of the XC-ELMOs is really straightforward. Unfortunately, as already observed for the X-ray constrained wave function strategy, good statistical agreements with the experimental X-ray scattering measures are reached provided that sufficiently flexible basis-sets and high-quality crystallographic data are used. In particular, we have observed that the presence of anisotropic secondary extinctions and the use of low-quality molecular geometries and ADPs prevent the achievement of the desired accuracy. These drawbacks open the way to further improvements of the method, such as the inclusion of more advanced models for the treatment of anisotropic secondary extinction phenomena and the possibility of developing new refinement techniques in which the XC-ELMO wave function approach is exploited to optimize both the Atomic Displacement Parameters and the geometry positions.

From a careful study of the obtained X-ray constrained electron densities, we have also noted that the fitting procedure entails significant redistributions of electronic charge both in the valence and in the core regions. Moreover, a vague indication that part of the electron correlation is included in the X-ray constrained wave functions has been deduced, but more extensive and detailed studies on this subject will be necessary in the future to better clarify this aspect.

Finally, it is important to stress that the current version of the technique allows to easily determine and analyze electron distributions associated with high-resolution crystallographic structures in terms of ELMOs, and, consequently, in terms of traditional chemical concepts. Considering this aspect and the intrinsic reliable transferability of the Extremely Localized Molecular Orbitals,[55−58] another possible future perspective consists in constructing databases of XC-ELMOs obtained through X-ray constrained ELMO calculations that use very high-resolution experimental crystallographic data. These new data sets, which can be thought as an alternative to the existing pseudoatoms libraries[59−67] proposed in the framework of the multipole model, will be eventually used for devising novel linear-scaling strategies able to refine macromolecular structures and to predict electrostatic properties of very large systems.

## ■ ASSOCIATED CONTENT

### Supporting Information

Details about the Obara−Saika scheme for the computation of Fourier transform integrals and details about the Walker−Mezey similarity index. Figures S1−S4: bar-graphs of the integrated net atomic charges computed at the ELMO, RHF, B3LYP, and XC-ELMO-Ext levels for all the systems in exam (cc-pVDZ basis-set). Tables S1−S4: values of the Walker−Mezey similarity index in different real-space regions and corresponding to the comparison of the ELMO charge distribution with the RHF, B3LYP, and XC-ELMO-Ext electron densities for all the investigated systems (cc-pVDZ basis-set). This material is available free of charge via the Internet at http://pubs.acs.org/.

## AUTHOR INFORMATION


### Corresponding Author

*Phone: +33 (0)3 83 68 43 77. Fax: +33 (0)3 83 68 43 71. E-mail: Alessandro.Genoni@univ-lorraine.fr.


### Notes

The author declares no competing financial interest.

## ACKNOWLEDGMENTS

The author thanks Enrique Espinosa, Emmanuel Aubert, Antonio Monari, and Toru Shiozaki for helpful discussions. Maurizio Sironi and Riccardo Destro are also gratefully acknowledged for providing the ELMO code and the X-ray diffraction data associated with the $\alpha$-glycine crystal, respectively.

## REFERENCES


(1) Jayatilaka, D.; Grimwood, D. J. *Acta Crystallogr., Sect. A* **2001**, *57*, 76−86.
(2) Clinton, W. L.; Nakhleh, J.; Wunderlich, F. *Phys. Rev.* **1969**, *177*, 1−6.
(3) Clinton, W. L.; Galli, A. J.; Massa, L. J. *Phys. Rev.* **1969**, *177*, 7−13.
(4) Clinton, W. L.; Henderson, G. A.; Prestia, J. V. *Phys. Rev.* **1969**, *177*, 13−18.
(5) Clinton, W. L.; Lamers, G. B. *Phys. Rev.* **1969**, *177*, 19−27.
(6) Clinton, W. L.; Galli, A. J.; Henderson, G. A.; Lamers, G. B.; Massa, L. J.; Zarur, J. *Phys. Rev.* **1969**, *177*, 27−33.
(7) Clinton, W. L.; Massa, L. J. *Phys. Rev. Lett.* **1972**, *29*, 1363−1366.
(8) Clinton, W. L.; Frishberg, C. A.; Massa, L. J.; Oldfield, P. A. *Int. J. Quantum Chem. Symp.* **1973**, *7*, 505−514.
(9) Frishberg, C.; Massa, L. J. *Phys. Rev. B* **1981**, *24*, 7018−7024.
(10) Massa, L.; Goldberg, M.; Frishberg, C.; Boehme, R. F.; Placa, S. J. L. *Phys. Rev. Lett.* **1985**, *55*, 622−625.
(11) Howard, S. T.; Huke, J. P.; Mallinson, P. R.; Frampton, C. S. *Phys. Rev. B* **1994**, *49*, 7124−7136.
(12) Cassam-Chenaï, P. *Int. J. Quantum Chem.* **1995**, *54*, 201−210.
(13) Snyder, J. A.; Stevens, E. D. *Chem. Phys. Lett.* **1999**, *313*, 293−298.
(14) Tanaka, K. *Acta Crystallogr., Sect. A* **1988**, *44*, 1002−1008.
(15) Jayatilaka, D. *Phys. Rev. Lett.* **1998**, *80*, 798−801.
(16) Grimwood, D. J.; Jayatilaka, D. *Acta Crystallogr., Sect. A* **2001**, *57*, 87−100.
(17) Bytheway, I.; Grimwood, D.; Jayatilaka, D. *Acta Crystallogr., Sect. A* **2002**, *58*, 232−243.
(18) Bytheway, I.; Grimwood, D. J.; Figgis, B. N.; Chandler, G. S.; Jayatiaka, D. *Acta Crystallogr., Sect. A* **2002**, *58*, 244−251.
(19) Grimwood, D. J.; Bytheway, I.; Jayatilaka, D. *J. Comput. Chem.* **2003**, *24*, 470−483.
(20) Hudák, M.; Jayatilaka, D.; Peraínova, L.; Biskupic, S.; Kozísek, J.; Bucinský, L. *Acta Crystallogr., Sect. A* **2010**, *66*, 78−92.
(21) Hederson, G. A.; Zimmermann, R. K. *J. Chem. Phys.* **1976**, *65*, 619−622.
(22) Stoll, H.; Wagenblast, G.; Preuss, H. *Theor. Chim. Acta* **1980**, *57*, 169−178.
(23) Genoni, A. *J. Phys. Chem. Lett.* **2013**, *4*, 1093−1099.
(24) Hirshfeld, F. L. *Acta Crystallogr., Sect. B* **1971**, *27*, 769−781.
(25) Stewart, R. F. *Acta Crystallogr., Sect. A* **1976**, *32*, 565−574.
(26) Hansen, N. K.; Coppens, P. *Acta Crystallogr., Sect. A* **1978**, *34*, 909−921.
(27) Hibbs, D. E.; Howard, S. T.; Huke, J. P.; Waller, M. P. *Phys. Chem. Chem. Phys.* **2005**, *7*, 1772−1778.
(28) Waller, M. P.; Howard, S. T.; Platts, J. A.; Piltz, R. O.; Willock, D. J.; Hibbs, D. E. *Chem.—Eur. J.* **2006**, *12*, 7603−7614.
(29) Jayatilaka, D.; Grimwood, D. *Acta Crystallogr., Sect. A* **2004**, *60*, 111−119.
(30) Grabowsky, S.; Jayatilaka, D.; Mebs, S.; Luger, P. *Chem.—Eur. J.* **2010**, *16*, 12818−12821.
(31) Grabowsky, S.; Weber, M.; Jayatilaka, D.; Chen, Y.-S.; Grabowski, M. T.; Brehme, R.; Hesse, M.; Schirmeister, T.; Luger, P. *J. Phys. Chem. A* **2011**, *115*, 12715−12732.
(32) Grabowsky, S.; Luger, P.; Buschmann, J.; Schneider, T.; Schirmeister, T.; Sobolev, A. N.; Jayatilaka, D. *Angew. Chem., Int. Ed.* **2012**, *51*, 6776−6779.
(33) Matta, C. F.; Arabi, A. A. *Future Med. Chem.* **2011**, *3*, 969−994.
(34) Smits, G. F.; Altona, C. *Theor. Chim. Acta* **1985**, *67*, 461−475.
(35) Fornili, A.; Sironi, M.; Raimondi, M. *J. Mol. Struct. (Theochem)* **2003**, *632*, 157−172.
(36) Larson, A. C. In *Crystallographic Computing*; Ahmed, F. R., Ed.; Munskgaard: Copenhagen, Denmark, 1970; pp 291−294.
(37) Kennard, O.; Wampler, D. L.; Coppola, J. C.; Motherwell, W. D. S.; Watson, D. G.; Larson, A. C. *Acta Crystallogr., Sect. B* **1971**, *27*, 1116−1123.
(38) Stewart, R. F. *J. Chem. Phys.* **1969**, *51*, 4569−4577.
(39) Jayatilaka, D.; Dittrich, B. *Acta Crystallogr., Sect. A* **2008**, *64*, 383−393.
(40) Obara, S.; Saika, A. *J. Chem. Phys.* **1986**, *84*, 3963−3974.
(41) Obara, S.; Saika, A. *J. Chem. Phys.* **1988**, *89*, 1540−1559.
(42) Head-Gordon, M.; Pople, J. A. *J. Chem. Phys.* **1988**, *89*, 5777−5786.
(43) Press,W. H.; Teukolsky, S. A.; Vetterling,W. T.; Flannery, B. P. *Numerical Recipes in Fortran 77: The Art of Scientific Computing*, 2nd ed.; Cambridge University Press: New York, 1992; pp 387−448.
(44) Brent, R. P. *Algorithms for Minimization without Derivatives*; Prentice-Hall: Englewood Cliffs, NJ, 1973; pp 61−80.
(45) 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. *Mol. Phys.* **2005**, *103*, 719−747.
(46) Destro, R.; Roversi, P.; Barzaghi, M.; Marsh, R. E. *J. Phys. Chem. A* **2000**, *104*, 1047−1054.
(47) Moggach, S. A.; Clark, S. J.; Parsons, S. *Acta Crystallogr., Sect. E* **2005**, *61*, o2739−o2742.
(48) Janicki, R.; Starynowicz, P. *Acta Crystallogr., Sect. B* **2010**, *66*, 559−567.
(49) Perpétuo, G. J.; Buschmann, J.; Luger, P.; Lentz, D.; Dreissig, D. *Acta Crystallogr., Sect. B* **1999**, *55*, 70−77.
(50) Whitten, A. E.; Jayatilaka, D.; Spackman, M. *J. Chem. Phys.* **2006**, *125*, 174505.
(51) Jayatilaka, D.; Munshi, P.; Turner, M. J.; Howard, J. A. K.; Spackman, M. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 7209−7218.
(52) Keith, T. A. *AIMAll*, Version 12.11.09; TK Gristmill Software: Overland Park, KS, 2012; available online: aim.tkgristmill.com.
(53) Jones, T. A.; Zou, J. Y.; Cowan, S. W.; Kjeldgaard, M. *Acta Crystallogr., Sect. A* **1991**, *47*, 110−119.
(54) Walker, P. D.; Mezey, P. G. *J. Am. Chem. Soc.* **1994**, *116*, 12022−12032.
(55) Genoni, A.; Fornili, A.; Sironi, M. *J. Comput. Chem.* **2005**, *26*, 827−835.
(56) Genoni, A.; Ghitti, M.; Pieraccini, S.; Sironi, M. *Chem. Phys. Lett.* **2005**, *415*, 256−260.
(57) Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. *Theor. Chem. Acc.* **2007**, *117*, 685−698.
(58) Sironi, M.; Ghitti, M.; Genoni, A.; Saladino, G.; Pieraccini, S. *J. Mol. Struct. (Theochem)* **2009**, *898*, 8−16.
(59) Pichon-Pesme, V.; Lecomte, C.; Lachekar, H. *J. Phys. Chem.* **1995**, *99*, 6242−6250.
(60) Jelsch, C.; Pichon-Pesme, V.; Lecomte, C.; Aubry, A. *Acta Crystallogr., Sect. D* **1998**, *54*, 1306−1318.
(61) Zarychta, B.; Pichon-Pesme, V.; Guillot, B.; Lecomte, C.; Jelsch, C. *Acta Crystallogr., Sect. A* **2007**, *63*, 108−125.
(62) Koritsanszky, T.; Volkov, A.; Coppens, P. *Acta Crystallogr., Sect. A* **2002**, *58*, 464−472.
(63) Volkov, A.; Li, X.; Koritsanszky, T.; Coppens, P. *J. Phys. Chem. A* **2004**, *108*, 4283−4300.
(64) Dominiak, P. M.; Volkov, A.; Li, X.; Messerschmidt, M.; Coppens, P. *J. Chem. Theory Comput.* **2007**, *3*, 232−247.

(65) Dittrich, B.; Koritsanszky, T.; Luger, P. *Angew. Chem., Int. Ed.* **2004**, *43*, 2718−2721.
(66) Dittrich, B.; Hübschle, C. B.; Messerschmidt, M.; Kalinowski, R.; Grint, D.; Luger, P. *Acta Crystallogr., Sect. A* **2005**, *61*, 314−320.
(67) Dittrich, B.; Hübschle, C. B.; Luger, P.; Spackman, M. A. *Acta Crystallogr., Sect. D* **2006**, *62*, 1325−1335.