HAL
open science

Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density

Leonardo dos Santos, Alessandro Genoni, Piero Macchi

▶ **To cite this version:**

Leonardo dos Santos, Alessandro Genoni, Piero Macchi. Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density. Acta Crystallographica Section A : Foundations and Advances [2014-..], 2014, 70 (6), pp.532-551. 10.1107/S2053273314019652 . hal-02196465

HAL Id: hal-02196465

https://hal.univ-lorraine.fr/hal-02196465v1

Submitted on 27 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.

CrossMark

electronic reprint

Acta Crystallographica Section A
Foundations and Advances
ISSN 2053-2733

Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density

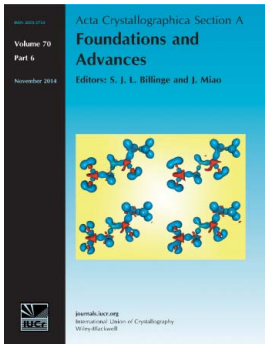
Leonardo H. R. Dos Santos, Alessandro Genoni and Piero Macchi

Acta Cryst. (2014). A**70**, 532–551

Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see http://journals.iucr.org/services/authorrights.html



Acta Crystallographica Section A: Foundations and Advances publishes articles reporting fundamental advances in all areas of crystallography in the broadest sense. This includes metacrystals such as photonic or phononic crystals, *i.e.* structures on the meso- or macroscale that can be studied with crystallographic methods. The central themes are, on the one hand, experimental and theoretical studies of the properties and arrangements of atoms, ions and molecules in condensed matter, periodic, quasiperiodic or amorphous, ideal or real, and, on the other, the theoretical and experimental aspects of the various methods to determine these properties and arrangements. In the case of metacrystals, the focus is on the methods for their creation and on the structure–property relationships for their interaction with classical waves.

Crystallography Journals Online is available from **journals.iucr.org**

Acta Crystallographica Section A
Foundations and Advances
ISSN 2053-2733

Received 30 April 2014
Accepted 30 August 2014

Unconstrained and X-ray constrained extremely localized molecular orbitals: analysis of the reconstructed electron density

Leonardo H. R. Dos Santos,[a] Alessandro Genoni[b,c]* and Piero Macchi[a]

[a]Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, Bern, CH-3012, Switzerland, [b]CNRS, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France, and [c]Université de Lorraine, Laboratoire SRSMC, UMR 7565, Vandoeuvre-lès-Nancy, F-54506, France. Correspondence e-mail: alessandro.genoni@univ-lorraine.fr

The recently developed X-ray constrained extremely localized molecular orbital (XC-ELMO) technique is a potentially useful tool for the determination and analysis of experimental electron densities. Molecular orbitals strictly localized on atoms, bonds or functional groups allow one to combine the quantum-mechanical rigour of the wavefunction-based approaches with the easy chemical interpretability typical of the traditional multipole models. In this paper, using very high quality X-ray diffraction data for the glycylglycine crystal, a detailed assessment of the capabilities and limitations of this new method is given. In particular, the effects of constraining the ELMO wavefunctions to experimental X-ray structure-factor amplitudes and the ability of the method to reproduce benchmark electron distributions have been accurately investigated. Topological analysis of the XC-ELMO electron densities and of the zero-flux surface-integrated charges and dipole moments shows that the new strategy is already reliable, provided that sufficiently flexible basis sets are used. These analyses also raise new questions and call for further improvements of the method.

© 2014 International Union of Crystallography

1. Abbreviations

General

ELMO, extremely localized molecular orbital.
XC-ELMO, X-ray constrained extremely localized molecular orbital.
IAM, independent atom model.
MM, multipole model.
$\{F_{\mathbf{h},\mathrm{obs}}\}$, $\{F_{\mathbf{h},\mathrm{calc}}\}$, set of experimentally observed or calculated structure-factor amplitudes.
$\sigma(F_{\mathbf{h},\mathrm{obs}})$, $\sigma(F^2_{\mathbf{h},\mathrm{obs}})$, standard uncertainty associated with the experimental structure-factor amplitude $F_{\mathbf{h},\mathrm{obs}}$ or the corresponding intensity $F^2_{\mathbf{h},\mathrm{obs}}$.
η, overall **h**-independent scale factor which multiplies the calculated structure-factor amplitudes.

Electron densities

$\rho^{\mathrm{MM/XC}}$, multipole-fitted (X-ray constrained) electron density; refined against experimental structure factors.
$\rho^{\mathrm{MM/P\text{-}B3LYP}}$, multipole-fitted periodic B3LYP electron density; refined against B3LYP/6-31G(2d,2p) crystal-phase calculated structure factors.
$\rho^{\mathrm{MM/ELMO}}$, multipole-fitted ELMO electron density; refined against ELMO gas-phase calculated structure factors.
$\rho^{\mathrm{MM/XC\text{-}ELMO}}$, multipole-fitted XC-ELMO electron density; refined against XC-ELMO calculated structure factors.
$\rho^{\mathrm{P\text{-}B3LYP}}$, electron density calculated directly from the periodic B3LYP/6-31G(2d,2p) computation.
ρ^{ELMO}, electron density calculated directly from an ELMO gas-phase wavefunction.
$\rho^{\mathrm{XC\text{-}ELMO}}$, electron density calculated directly from an XC-ELMO wavefunction.
ρ^{HF}, electron density calculated directly from a Hartree–Fock gas-phase wavefunction.
$\rho^{\mathrm{XC\text{-}HF}}$, electron density calculated directly from an XC-Hartree–Fock wavefunction.
ρ^{IAM}, independent atom model density; sum of spherically averaged electron densities of isolated atoms.

2. Introduction

According to the Born interpretation of quantum mechanics, the wavefunction is the fundamental entity that contains all the information of a system. In fact, all the physical quantities can be obtained as expectation values of the corresponding Hermitian operators. Moreover, Hohenberg & Kohn (1964) have shown that the ground-state electron density and the wavefunction of an electronic system can be used alternatively as full descriptors of the ground state and, therefore, it would be possible to obtain all properties of a system from its

electronic reprint

electron-density distribution, which is an observable and, unlike the wavefunction, is only a three-dimensional function of the spatial coordinates.

The prominence conferred by the Hohenberg & Kohn theorem to the electron density together with the observation that knowledge of the electron distribution and of its properties has profound consequences for the understanding and the rationalization of chemical bonding (Coppens, 1997) have significantly increased the importance of the accurate determination of electron densities in crystals by means of high-resolution X-ray diffraction experiments (Gatti & Macchi, 2012). Nevertheless, the observed diffraction intensities do not directly provide the electron density because only a finite number of structure factors can be collected and, within the kinematic approximation, their phase cannot be measured. Consequently, the electron distribution in the unit cell can be accurately reconstructed only by fitting a model against the collected X-ray diffraction intensities. The parameters of this model electron density are usually obtained minimizing the difference between experimentally observed and calculated structure factors. An alternative is represented by the maximization of the information entropy, which however would not return an electron density separated from the thermal smearing. For more discussion about this method, see for example Roversi *et al.* (1998).

In this context, the multipole models (Hirshfeld, 1971; Stewart, 1972; Hansen & Coppens, 1978) are by far the most widely adopted approaches. Since the total electron distribution is the sum of the aspherical atom-centred density functions (also known as pseudoatoms), the resulting ground-state electron density of a molecule is simply approximated by the sum of atomic electron distributions deformed by the presence of chemical interactions. This is the main reason why these models offer an easy chemical interpretability and justifies their widespread popularity. Notwithstanding this important advantage, they also exhibit a non-negligible drawback: the number of properties directly available from the model electron density is limited because the exact functional relation between the ground-state electron distribution and the ground-state wavefunction is practically unknown. Furthermore, the parameters of a multipole expansion may have a strong correlation with the atomic anisotropic displacement parameters (ADPs).

A possible solution to overcome the above-mentioned drawbacks is offered by a work which dates back to 1969, thus even before the appearance of multipole models. Clinton, Nakhleh & Wunderlich (1969), Clinton, Galli & Massa (1969), Clinton, Henderson & Prestia (1969), Clinton & Lamers (1969) and Clinton, Galli, Henderson *et al.* (1969) proposed obtaining wavefunctions or one-electron density matrices constrained to experimental or theoretical X-ray diffraction data. Since then, many researchers have drawn inspiration from Clinton's original ideas (Clinton & Massa, 1972; Clinton *et al.*, 1973; Frishberg & Massa, 1981; Massa *et al.*, 1985; Tanaka, 1988; Howard *et al.*, 1994; Hibbs *et al.*, 2005; Cassam-Chenaï, 1995; Snyder & Stevens, 1999; Gillet *et al.*, 2001; Gillet, 2007). Within this framework, the most promising method is the X-ray constrained (initially called 'experimental') wavefunction approach developed by Jayatilaka and co-workers (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001; Grimwood & Jayatilaka, 2001; Bytheway, Grimwood & Jayatilaka, 2002; Bytheway, Grimwood, Figgis *et al.*, 2002; Grimwood *et al.*, 2003; Hudák *et al.*, 2010). This method provides a single Slater determinant which, other than minimizing the corresponding energy, reproduces a set of experimentally collected X-ray structure-factor amplitudes within a predefined precision. The strategy simply exploits the Lagrange multiplier technique to variationally minimize a new functional given by the energy associated with the single Slater determinant and an additional term represented by the statistical agreement with the experimental diffraction data (namely, the experimental constraint). In particular, the Lagrange multiplier is iteratively adjusted until the desired agreement level is achieved between the structure-factor amplitudes obtained from the diffraction experiment and those calculated from the single Slater determinant *ansatz*. Therefore, the constrained wavefunction reproduces the experimental data within a given precision, and, as it possesses all the quantum-mechanical features of a wavefunction, it can also be used to compute those properties that are not directly related to the experimental structure factors used to determine the wavefunction itself (Grimwood & Jayatilaka, 2001; Grimwood *et al.*, 2003). Compared to the multipole models, the approach enables one to obtain quantum-mechanically rigorous electron densities, but it is affected by a reduced chemical interpretability because the canonical molecular orbitals are usually completely delocalized over the whole system. Therefore, they are far from the traditional and intuitive picture of a molecule constituted by atoms, bonds and functional groups, typical of the pseudoatom approaches.

One of us has recently devised a new technique to extract from X-ray diffraction data a single Slater determinant built up with molecular orbitals strictly localized (namely, without tails) on molecular fragments, such as atoms, bonds or functional groups (Genoni, 2013*a*,*b*). This novel strategy can be considered as a combination of the experimentally constrained wavefunction approach proposed by Jayatilaka with the method developed by Stoll (Stoll *et al.*, 1980) for the determination of extremely localized molecular orbitals (ELMOs). This synergy should be useful to solve the above-mentioned drawbacks associated with the X-ray constrained wavefunctions. Moreover, due to the complete absence of tails, the ELMOs are directly transferable; thus they can be computed on fragments of small model molecules and afterwards properly exported to subunits belonging to more complex systems such as macromolecules or polymers (Genoni *et al.*, 2005; Sironi *et al.*, 2009). The ELMOs' transferability is analogous to the well known pseudoatoms' transferability within the framework of the multipole models. In this view, new databases could be constructed from X-ray constrained ELMOs (XC-ELMOs) and used to complement the existing experimental (Pichon-Pesme *et al.*, 1995, 2004; Jelsch *et al.*, 1998; Zarychta *et al.*, 2007) or theoretical (Koritsanszky *et al.*, 2002; Volkov *et al.*, 2004; Dittrich *et al.*,

electronic reprint

2004, 2006; Dominiak *et al.*, 2007) libraries of pseudoatoms, successfully used both for the refinement of macromolecular crystallographic structures and for the evaluation of electrostatic properties in crystals.

In the present paper, we present a comprehensive assessment of the performance of the XC-ELMO technique. In particular, we analyse the machinery of the constraint procedure, evaluating how much the unconstrained and the X-ray constrained wavefunctions differ and to what extent the XC-ELMO electron density is able to reproduce the 'exact' crystal electron density. The latter is here approximated with two well established methods, namely the multipole model density fitted or, in analogy with the wavefunction approach, *constrained*[1] to the experimental structure factors, hereafter $\rho^{\mathrm{MM/XC}}$, and the electron density directly obtained from a periodic density functional theory (DFT) calculation with a standard hybrid functional, hereafter $\rho^{\mathrm{P\text{-}B3LYP}}$. To accomplish this task we have investigated the α-glycylglycine crystal, already known from X-ray (Biswas *et al.*, 1968) and neutron (Kvick *et al.*, 1977) diffraction experiments. Except for a qualitative deformation electron-density study by Kvick *et al.* (1979) and a spherical κ-formalism refinement by Coppens *et al.* (1979), to the best of our knowledge, no complete electron-density refinement for this system has been reported in the literature so far.

The paper is structured as follows. First, we summarize the theory of the X-ray constrained ELMO technique, previously presented in detail (Genoni, 2013*a*,*b*). We go on to describe the experimental and computational methodologies and then we analyse the results. We initially focus on the effects of the fitting procedure, comparing the residual features and the topological properties computed from unconstrained or X-ray constrained ELMO electron densities (from now on ρ^{ELMO} and $\rho^{\mathrm{XC\text{-}ELMO}}$, respectively). In a second step, we compare ρ^{ELMO} and $\rho^{\mathrm{XC\text{-}ELMO}}$ with the density from a periodic calculation ($\rho^{\mathrm{P\text{-}B3LYP}}$). In a third step, we project ρ^{ELMO} and $\rho^{\mathrm{XC\text{-}ELMO}}$ in terms of multipoles, thus obtaining $\rho^{\mathrm{MM/ELMO}}$ and $\rho^{\mathrm{MM/XC\text{-}ELMO}}$, respectively, and we compare these densities with multipole electron distributions fitted against the experimental intensities ($\rho^{\mathrm{MM/XC}}$) or against theoretical structure factors obtained from the periodic calculation ($\rho^{\mathrm{MM/P\text{-}B3LYP}}$). This procedure allows us to compare all the densities equally affected by the limitations of the multipolar model. Finally, we investigate how the choice of initial parameters (fractional coordinates and ADPs of the atoms in the unit cell) affects the XC-ELMO results and the convergence of the fitting process. In the last section, we draw some conclusions and discuss perspectives offered by this new strategy.[2]

3. Theoretical background

3.1. Extremely localized molecular orbitals

In this section we will briefly introduce the extremely localized molecular orbitals method proposed by Stoll *et al.* (1980). This strategy is strictly connected to the earlier group function method devised by McWeeny (1959, 1960, 1992) and it can be considered one of the many theoretical approaches that have been developed over the years in order to decompose the global electronic wavefunction into functions describing smaller subsets of electrons (Adams, 1961; Huzinaga & Cantu, 1971; Gilbert, 1974; Matsuoka, 1977; Stoll *et al.*, 1980; Smits & Altona, 1985; Francisco *et al.*, 1992; Ordejón *et al.*, 1993; Couty *et al.*, 1997; Fornili *et al.*, 2003; Szekeres & Surján, 2003).

Let us consider a $2N$-electron closed-shell molecule and let us introduce a localization scheme that subdivides the system in exam into f fragments (*e.g.*, atoms, bonds or functional groups) that can overlap. Because of this fragmentation, each generic subunit j is characterized by a local basis set $\beta^j = \{|\chi^j_\mu\rangle\}_{\mu=1}^{M^j}$, which is constituted by only the M^j basis functions centred on the atoms belonging to the fragment. Consequently, the extremely localized molecular orbitals describing the subunit are expanded on β^j and, therefore, the generic α-th ELMO for the jth fragment can be written as:

$$|\varphi^j_\alpha\rangle = \sum_{\mu=1}^{M^j} C^j_{\mu\alpha} |\chi^j_\mu\rangle. \tag{1}$$

Here it should be observed that predefined subunits may overlap and, consequently, share atomic orbitals, which leads to a natural non-orthogonality between ELMOs associated with different fragments.

Following Stoll *et al.* (1980), we assume that the wavefunction describing the system is a normalized single Slater determinant built up with extremely localized molecular orbitals (ELMO wavefunction):

$$|\Psi_{\mathrm{ELMO}}\rangle = \frac{1}{[(2N)!]^{1/2}\det[\mathbf{S}]} \times \hat{A}[\varphi^1_1\overline{\varphi}^1_1 \ldots \varphi^1_{n_1}\overline{\varphi}^1_{n_1} \ldots \varphi^f_1\overline{\varphi}^f_1 \ldots \varphi^f_{n_f}\overline{\varphi}^f_{n_f}] \tag{2}$$

where $\det[\mathbf{S}]$ is the determinant of the overlap matrix between the occupied ELMOs, $\hat{A}$ is the anti-symmetrizer, n_j is the number of occupied ELMOs for the jth fragment and $\overline{\varphi}^j_\alpha$ is a spin orbital with spatial part φ^j_α and spin part β. The coefficients in equation (1) are obtained variationally minimizing

[1] In this paper we adopt the terminology currently used in the literature. Thus, the wavefunction fitted to the X-ray structure factors is called *constrained*, and not *restrained* as would have been more intuitive for the crystallographic community. In keeping with Jayatilaka (2012), the traditional multipolar fitting can also be considered as an 'X-ray constrained density function'. Therefore, in this paper we will call the traditional multipolar model fitted to the X-ray structure factors as 'X-ray constrained multipolar model' (MM/XC). Of course, this nomenclature does not imply an additional constraint equation to be satisfied, other than the least-squares minimization functions and those already present in the multipolar fitting (*e.g.* electroneutrality condition). When discussing local symmetry constraints on the multipolar model, which implies that some parameters are not refined, we will refer to them as *rigid constraints* to avoid ambiguity.

[2] Supporting information contains additional experimental details, plots of electron densities and their Laplacians, charges and dipoles in tabular form, electrostatic potentials, residual density analysis of ELMO and XC-ELMO densities and illustrative input and output files for unconstrained ELMO calculations. It is available from the IUCr electronic archives (Reference: TD5019).

electronic reprint

the energy associated with the ELMO wavefunction, minimization that is equivalent to solving self-consistently the following modified Hartree–Fock equations for each subunit:

$$\hat{F}^j|\varphi_\alpha^j\rangle = \varepsilon_\alpha^j|\varphi_\alpha^j\rangle \tag{3}$$

with $\hat{F}^j$ as the modified Fock operator for the generic fragment j given by

$$\hat{F}^j = (1-\hat{\rho}+\hat{\rho}^{j^\dagger})\,\hat{F}(1-\hat{\rho}+\hat{\rho}^j) \tag{4}$$

where $\hat{F}$ is the usual Fock operator, $\hat{\rho}^j$ is the local density operator for the jth subunit, which depends only on the occupied ELMOs of the fragment, and $\hat{\rho}$ is the global density operator, which, on the contrary, depends on all the occupied ELMOs of the system and, for this reason, couples all the modified Hartree–Fock equations associated with the different subunits.

Finally, it is worth mentioning that, due to the non-orthogonality of the ELMOs, convergence problems and instabilities may arise when solving equation (3) (Stoll *et al.*, 1980; Smits & Altona, 1985). To overcome these important drawbacks, Fornili *et al.* (2003) have implemented a completely equivalent strategy for ELMOs determination which consists in directly minimizing the ELMO energy through a quasi-Newton procedure that exploits an approximated Hessian. In particular, this Hessian is analytically computed only at the first iteration, while it is afterwards updated exploiting a variable metric algorithm that uses the Broyden–Fletcher–Goldfarb–Shanno formula (Press *et al.*, 1992).

3.2. X-ray constrained extremely localized molecular orbitals

In order to extract extremely localized molecular orbitals from X-ray diffraction data it is necessary not only to minimize the energy associated with the ELMO wavefunction as discussed in the previous subsection, but also to reproduce a set of collected structure-factor amplitudes $\{F_{\mathbf{h},\mathrm{obs}}\}$ within a predetermined desired agreement. In other words, this is equivalent to looking for those ELMOs that minimize the following functional:

$$J[\boldsymbol{\varphi}] = E_{\mathrm{ELMO}}[\boldsymbol{\varphi}] + \lambda\big(\chi^2[\boldsymbol{\varphi}] - \Delta\big), \tag{5}$$

where $[\boldsymbol{\varphi}]$ indicates the functional dependence on the occupied ELMOs, E_{ELMO} is the energy associated with the ELMO wavefunction for the reference crystal unit, λ is the Lagrange multiplier representing the strength of the constraint associated with the experimental data, Δ is the desired agreement between theoretical and experimental values, and χ^2 is the agreement statistics between the calculated and the observed structure-factor amplitudes ($F_{\mathbf{h},\mathrm{calc}}$ and $F_{\mathbf{h},\mathrm{obs}}$, respectively), namely

$$\chi^2 = \frac{1}{N_r - N_p}\sum_{\mathbf{h}} \frac{\big(\eta|F_{\mathbf{h},\mathrm{calc}}| - |F_{\mathbf{h},\mathrm{obs}}|\big)^2}{\sigma^2_{\mathbf{h},\mathrm{obs}}} \tag{6}$$

with N_r as the number of considered experimental diffraction data, N_p as the number of adjustable parameters (in our case only the Lagrange multiplier λ), $\mathbf{h}$ as the tern of Miller indexes labelling the reflection and $\sigma_{\mathbf{h},\mathrm{obs}}$ as the standard uncertainty associated with the experimental structure-factor amplitude $F_{\mathbf{h},\mathrm{obs}}$ [*i.e.*, $\sigma_{\mathbf{h},\mathrm{obs}} = \sigma(F_{\mathbf{h},\mathrm{obs}})$]. All the calculated structure-factor amplitudes are properly multiplied by an overall $\mathbf{h}$-independent scale factor η, determined in order to minimize χ^2 as discussed in detail in a previous paper (Genoni, 2013*b*). Moreover, in this work the set of collected structure-factor amplitudes $\{F_{\mathbf{h},\mathrm{obs}}\}$ was previously corrected for secondary extinctions.

Following Jayatilaka's philosophy (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001), the crystal is a set of non-interacting units, which allows one to write the global wavefunction for the crystal as

$$|\Psi_{\mathrm{cryst}}\rangle = \prod_k |\Psi_k\rangle, \tag{7}$$

where all the unit wavefunctions $|\Psi_k\rangle$ are related to each other by means of the crystal symmetry operations and are formally identical to the ELMO wavefunction given by equation (2). Furthermore, assuming that all the non-interacting units correspond to symmetry-unique portions of the crystal unit cell, the unit-cell electron distribution can be expressed using only the electron density $\rho_0(\mathbf{r})$ associated with the ELMO wavefunction $|\Psi_0\rangle$ of the reference crystal unit, namely

$$\rho_{\mathrm{cell}}(\mathbf{r}) = \sum_{k=1}^{N_m}\rho_k(\mathbf{r}) = \sum_{k=1}^{N_m}\rho_0\big[\mathbf{R}_k^{-1}(\mathbf{r}-\mathbf{r}_k)\big], \tag{8}$$

where the N_m unit-cell electron distributions are related to the reference one through the crystal symmetry operations $\{\mathbf{R}_k, \mathbf{r}_k\}$. It is important to point out that equation (8) is exact provided that $\rho_0(\mathbf{r})$ is not obtained by means of an isolated crystal unit calculation. Nevertheless, in this case, the previous condition is not completely fulfilled since $\rho_0(\mathbf{r})$ is the electron density associated with the reference unit ELMO wavefunction $|\Psi_0\rangle$ that is obtained searching those extremely localized molecular orbitals that minimize the functional given by equation (5).

Now, defining the structure-factor operator

$$\hat{I}_{\mathbf{h}} = \textstyle\sum_{k=1}^{N_m}\exp\big[i2\pi(\mathbf{R}_k\mathbf{r}+\mathbf{r}_k)\cdot(\mathbf{Bh})\big] = \hat{I}_{\mathbf{h},R} + i\,\hat{I}_{\mathbf{h},C}, \tag{9}$$

where $\mathbf{B}$ is the reciprocal-lattice matrix and both $\hat{I}_{\mathbf{h},R}$ and $\hat{I}_{\mathbf{h},C}$ (real and imaginary parts of $\hat{I}_{\mathbf{h}}$, respectively) are Hermitian operators, it is possible to show that finding the ELMOs that minimize the functional $J[\boldsymbol{\varphi}]$ introduced above is equivalent to solving self-consistently this new eigenvalue problem for each fragment:

$$\hat{F}^{j,\mathrm{exp}}|\varphi_\alpha^j\rangle = \varepsilon_\alpha^{j,\mathrm{exp}}|\varphi_\alpha^j\rangle, \tag{10}$$

where the new modified Fock operator for the generic jth subunit, $\hat{F}^{j,\mathrm{exp}}$, can be expressed as

$$\begin{aligned}\hat{F}^{j,\mathrm{exp}} &= (1-\hat{\rho}+\hat{\rho}^{j^\dagger})\,\hat{F}(1-\hat{\rho}+\hat{\rho}^j)\\ &\quad+\lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Re}\{F_{\mathbf{h},\mathrm{calc}}\}(1-\hat{\rho}+\hat{\rho}^{j^\dagger})\,\hat{I}_{\mathbf{h},R}(1-\hat{\rho}+\hat{\rho}^j)\\ &\quad+\lambda\sum_{\mathbf{h}}K_{\mathbf{h}}\mathrm{Im}\{F_{\mathbf{h},\mathrm{calc}}\}(1-\hat{\rho}+\hat{\rho}^{j^\dagger})\,\hat{I}_{\mathbf{h},C}(1-\hat{\rho}+\hat{\rho}^j)\end{aligned} \tag{11}$$

electronic reprint

with

$$K_{\mathbf{h}} = \frac{2\eta}{N_r - N_p} \frac{\eta F_{\mathbf{h},\mathrm{calc}} - F_{\mathbf{h},\mathrm{obs}}}{\sigma^2_{\mathbf{h},\mathrm{obs}} F_{\mathbf{h},\mathrm{calc}}}. \qquad (12)$$

As for the equations for the 'theoretical' ELMOs, also the equations for the X-ray constrained extremely localized molecular orbitals are coupled since all the operators $\hat{F}^{j,\mathrm{exp}}$ depend on the global density operator $\hat{\rho}$ which is constructed with all the occupied ELMOs of the system. Furthermore, as for the theoretical case, we have also implemented an analogous quasi-Newton procedure to overcome possible convergence problems and instabilities in the resolution of equation (10) (Genoni, 2013*a*,*b*).

In order to compute a set of structure-factor amplitudes $\{F_{\mathbf{h},\mathrm{calc}}\}$ that properly takes into account the effects of the thermal vibrations, we have followed the Stewart (1969) thermal smearing scheme using the same equations proposed by Jayatilaka and co-workers (Jayatilaka, 1998; Jayatilaka & Grimwood, 2001), except for the fact that, for the calculation of the Fourier transforms of basis function pairs, an Obara–Saika scheme (Obara & Saika, 1986, 1988) with both vertical and horizontal recurrence relations (Head-Gordon & Pople, 1988) has been implemented (Genoni, 2013*a*,*b*).

Owing to experimental errors in the collected experimental structure-factor amplitudes, it is not necessary to force χ^2 equal to zero, but it is better to set the value of the desired agreement Δ in equation (5) equal to 1, so that, at the end of the calculations, the computed values are on average within one standard deviation of the experimental data. However, as already observed by Jayatilaka and co-workers (Grimwood & Jayatilaka, 2001; Whitten *et al.*, 2006; Jayatilaka *et al.*, 2009; Hudák *et al.*, 2010), the convergence towards the desired agreement is not always fast and straightforward. In order to avoid large values of λ producing only minimal improvements in the χ^2 statistics and large unphysical changes of the energy, the following termination criteria have recently been proposed (Genoni, 2013*b*):

$$\begin{cases} \chi^2 < 1 \\ \left(\dfrac{\Delta\chi^2}{\Delta\lambda}\right)_i = \dfrac{\chi^2_i - \chi^2_{i-1}}{\lambda_i - \lambda_{i-1}} > -5 \times 10^{-1} \\ \dfrac{|E^{\mathrm{el}}_{\lambda_i} - E^{\mathrm{el}}_{\lambda=0}|}{|E^{\mathrm{el}}_{\lambda=0}|} > 5 \times 10^{-4} \end{cases}. \qquad (13)$$

While the first criterion is the traditional one, namely it checks if the desired statistical agreement $\Delta = 1.0$ has been reached, the second one avoids the case that a very large Lagrange multiplier produces only very small improvements in the agreement statistics. The third criterion ensures that the XC-ELMO electronic energy does not excessively change compared to the unconstrained ELMO electronic energy.

Although these criteria are quite reliable and well defined, the termination of the fitting procedure in the X-ray constrained approaches is still an open problem and will deserve further investigations in the future.

4. Experimental and computational details

4.1. X-ray data collection and processing

α-Glycylglycine (Fig. 1*a*) was recrystallized by slow evaporation of an *n*-propanol–water mixture. We have redetermined the crystal structure by single-crystal X-ray diffraction at 100 K. The temperature was stable within ±0.5 K. Details of data-collection and refinement procedures are given in Table 1. Measurements were carried out on an Agilent SuperNova diffractometer equipped with an Mo $K\alpha$ Al-filtered microsource (Macchi *et al.*, 2011) and an Oxford Cryosystem 700 cryostream for low temperature. Data collection, reduction and cell refinement were performed using the *CrysAlis Pro* programs (Agilent Technologies, 2013). A total of 2765 image frames were obtained from 33 ω-scan sets (1.0° oscillation angle) using three different exposure times. The scan sets with low detector θ offsets were measured for 10 + 10 s, intermediate-angle images were collected for 30 + 30 s and the high-angle images were measured for 60 + 60 s. The unit-cell dimensions were determined by post-refinement of 18 946 reflections ($2.7° < \theta < 52.4°$). An analytical absorption correction was applied using a multifaceted crystal model based on expressions derived by Clark & Reid (1995) as implemented in the *SCALE*3 *ABSPACK* scaling algorithm (Agilent Technologies, 2013). A total of 38 933 reflections were collected with a mean redundancy of 5.9. The resulting

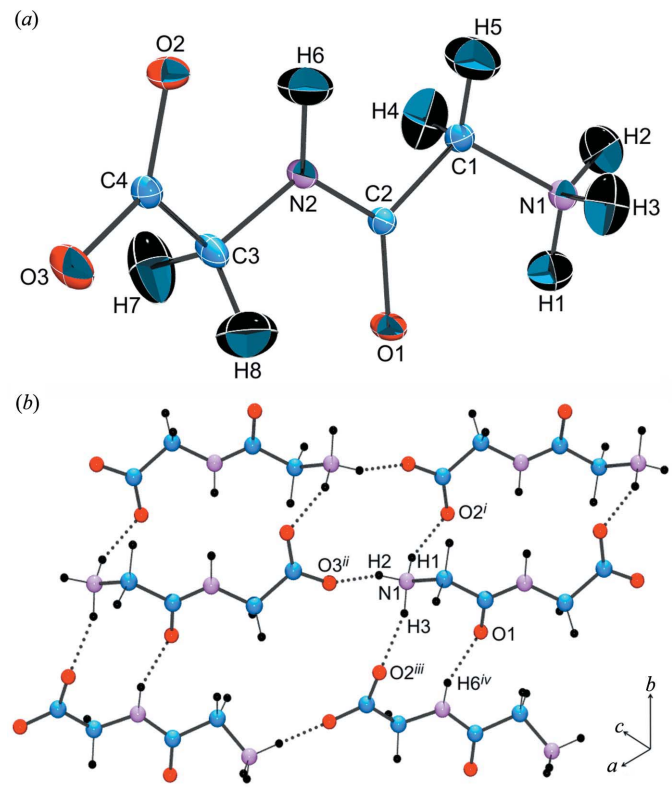


Figure 1
Experimental molecular structure of crystalline α-glycylglycine (*a*) and strongest hydrogen-bond network (*b*). Thermal ellipsoids are drawn at the 70% probability level. Symmetry codes: (i) $-x+1, y+\frac{1}{2}, -z+\frac{3}{2}$; (ii) $x+1, y, z+1$; (iii) $-x+1, -y, -z+1$; (iv) $-x+1, y+\frac{1}{2}, -z+\frac{3}{2}$.

electronic reprint

Table 1
Crystallographic details and results of IAM and multipolar refinements for glycylglycine.

Crystal data	
Chemical formula	$C_4H_8N_2O_3$
M_r (g mol^{-1})	132.12
Crystal system, space group	Monoclinic, $P2_1/c$
Temperature (K)	100.0 (5)
a, b, c (Å)	7.9798 (1), 9.5201 (1), 7.7643 (1)
β (°)	106.151 (1)
V (Å^3)	566.56 (1)
Z	4
$F(000)$	280
θ range (°) for cell measurement	2.7–52.4
μ (mm^{-1})	0.13
Crystal size (mm)	0.20 × 0.17 × 0.07
Data collection	
Diffractometer	Agilent SuperNova diffractometer
Radiation type	Mo $K\alpha$
Absorption correction	Analytical (Clark & Reid, 1995)
T_{min}, T_{max}	0.974, 0.991
No. of measured, independent and observed [$F > 3\sigma(F)$] reflections	38933, 6597, 5467
Redundancy	5.9
R_{int}†	0.029
θ values (°)	θ_{max} = 52.6, θ_{min} = 2.1
$(\sin\theta/\lambda)_{max}$ (Å^{-1})	1.118
Range of h, k, l	$h = -17\rightarrow17$, $k = -21\rightarrow21$, $l = -17\rightarrow17$
Spherical, independent atom refinement	
Refinement on	F^2 (for $F > 0$)
$R[F > 3\sigma(F)]$, R_{all}, wR, S, χ^2‡	2.82, 4.15, 4.16, 1.06, 2.60
No. of parameters	107
$(\Delta/\sigma)_{max}$	0.001
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})	0.54, −0.37
Extinction coefficient	0.007 (4)
Multipole refinement	
Refinement on	F^2 [for $F > 3\sigma(F)$]
$R[F > 3\sigma(F)]$, R_{all}, wR, S, χ^2‡	1.97, 3.30, 2.14, 1.03, 0.91 (MM1) 1.72, 2.93, 1.71, 0.80, 0.57 (MM2) 1.70, 2.91, 1.68, 0.80, 0.55 (MM3)
No. of parameters§	271
$(\Delta/\sigma)_{max}$§	0.00001
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e Å^{-3})§	0.14, −0.14
Extinction coefficient§	0.315 (15)

† $R_{int} = \sum_{\mathbf{h}} |F^2_{\mathbf{h},obs} - \langle F^2_{\mathbf{h},obs}\rangle| / \sum_{\mathbf{h}} F^2_{\mathbf{h},obs}$ (summation is carried out only where more than one symmetry-equivalent reflection is averaged). ‡ $R(F) = 100 \sum_{\mathbf{h}} ||F_{\mathbf{h},obs}| - |F_{\mathbf{h},calc}|| / \sum_{\mathbf{h}} |F_{\mathbf{h},obs}|$, $wR(F) = 100[\sum_{\mathbf{h}} w_{\mathbf{h}}(|F_{\mathbf{h},obs}| - |F_{\mathbf{h},calc}|)^2 / \sum_{\mathbf{h}} w_{\mathbf{h}} F^2_{\mathbf{h},obs}]^{1/2}$, $S = [\sum_{\mathbf{h}} w_{\mathbf{h}}(F^2_{\mathbf{h},obs} - F^2_{\mathbf{h},calc})^2/(N-P)]^{1/2}$ with $w_{\mathbf{h}} = 1/\sigma^2_{\mathbf{h},obs}$, N as the number of reflections and P as the number of parameters. χ^2 is given by equation (6), see main text. Both in the spherical model and in the multipole model refinements the calculated structure-factor magnitudes are properly multiplied by a scale factor η and by an additional factor $Y_{\mathbf{h}}(\varepsilon)$ that corrects for secondary extinctions. For the spherical model, $Y_{\mathbf{h}}(\varepsilon) = [1 + 0.001\varepsilon F^2_{\mathbf{h},calc}\lambda^3/\sin(2\theta_{\mathbf{h}})]^{-1/4}$, while, for the multipole model, $Y_{\mathbf{h}}(\varepsilon)$ has been chosen following the Becker & Coppens (1974) equations. § For MM2 model.

data were merged using *SORTAV* (Blessing, 1987) giving 6597 independent reflections until a resolution of $\sin\theta_{max}/\lambda = 1.12$ Å^{-1} (θ_{max} = 52.6° for Mo $K\alpha$ radiation).

4.2. Spherical, independent atom refinement

Atomic coordinates and ADPs were taken from Biswas *et al.* (1968) and refined on F^2 applying the IAM and using all data with full-matrix least-squares as implemented in *SHELXL* (Sheldrick, 2008), within the *WinGX* package (Farrugia, 1999). The weighting scheme $w_{\mathbf{h}} = [\sigma^2(F^2_{\mathbf{h},obs}) + (AP)^2 + BP]^{-1}$ was applied, where $P = F^2_{\mathbf{h},obs}/3 + 2F^2_{\mathbf{h},calc}/3$ and A = 0.0606, B = 0.0360. All non-hydrogen atoms were refined anisotropically. X—H distances and hydrogen ADPs were fixed according to experimental neutron diffraction data (Kvick *et al.*, 1977). As recommended by Blessing (1995), the H-atom ADPs were scaled against the ADPs obtained for the heavier atoms from the two experiments. An isotropic extinction parameter was also refined according to the empirical expression implemented in *SHELXL*, where $F_{\mathbf{h},calc}$ is multiplied by $\eta[1 + 0.001\varepsilon F^2_{\mathbf{h},calc}\lambda^3/\sin(2\theta_{\mathbf{h}})]^{-1/4}$, with η as the overall scale factor and ε as the extinction parameter, which was refined to 0.007 (4).

4.3. Multipole refinement

The multipole refinement was performed using the *XD2006* program suite (Volkov *et al.*, 2006) and the Hansen–Coppens formalism (Hansen & Coppens, 1978) that is able to model the deformation fraction of $\rho(\mathbf{r})$ using atom-centred multipole functions (also known as pseudoatoms) having the form

$$\rho_{atom}(\mathbf{r}) = \rho_{core}(\mathbf{r}) + P_v\kappa\rho_{valence}(\kappa\,\mathbf{r}) + \Delta\rho(\mathbf{r}), \qquad (14)$$

where

$$\Delta\rho(\mathbf{r}) = \sum_{l=0}^{l_{max}} \kappa'_l R_l(\kappa'_l r) \sum_{m=-l}^{+l} P_{lm} Y_{lm}(\theta, \varphi). \qquad (15)$$

The core and spherical valence density (ρ_{core} and $\rho_{valence}$, respectively) are composed of Hartree–Fock wavefunctions expanded in Slater-type orbitals (Clementi & Roetti, 1974). The valence shell is either contractible or expandable by the use of the κ parameter. Single-zeta orbitals with energy-optimized Slater exponents are used for the radial part of the deformation terms (Clementi & Raimondi, 1963). The additional radial parameters κ'_l are defined for each angular momentum l in order to deal with the radial expansion or contraction of the deformation density. P_v and P_{lm} represent the population parameters for the valence and deformation density multipoles, respectively. The angular functions Y_{lm} are density-normalized real spherical harmonics (Paturle & Coppens, 1988). Several multipole models were tested, as described in §5.

The function minimized in the least-squares procedure was $\sum_{\mathbf{h}} w_{\mathbf{h}}[F^2_{\mathbf{h},obs} - \eta Y_{\mathbf{h}}(\varepsilon)F^2_{\mathbf{h},calc}]^2$, where the statistical weight $w_{\mathbf{h}} = 1/\sigma^2(F^2_{\mathbf{h},obs})$ was applied and where only those reflections characterized by $F_{\mathbf{h},obs} > 3\sigma_{\mathbf{h},obs}(F_{\mathbf{h},obs})$ were included. Furthermore, η is the overall scale factor while $Y_{\mathbf{h}}(\varepsilon)$ is the correction for secondary extinctions according to the model proposed by Becker & Coppens (1974). This model depends on the isotropic extinction parameter ε, which was refined to 0.315 (15), corresponding to a mosaicity spread of 10″.

The multipole expansion was truncated at the hexadecapole level (l_{max} = 4) for all the non-H atoms, while bond-directed dipoles and quadrupoles were applied to all the H atoms. The κ and κ'_l parameters were initially set to proper reference values (Volkov *et al.*, 2001) and posteriorly refined. A single κ'_l

electronic reprint

was refined for all the l values belonging to a defined set while κ'_l for the H atoms was not refined. The X—H distances were initially set to neutron diffraction averages (Allen & Bruno, 2010) but then freely refined. A high-order refinement with $\sin\theta/\lambda \geq 0.7$ Å^{-1} was performed for the non-H atoms to obtain accurate positional and displacement parameters. In the next step, the H-atom ADPs were estimated by the *SHADE* routine (Madsen, 2006) and the obtained values were used as fixed parameters in the subsequent refinements.

4.4. *Ab initio* periodic calculation and multipole refinement of theoretical structure factors

In order to provide an additional and reliable benchmark for the XC-ELMO calculations, a single-point periodic calculation was performed on the final multipole model geometry without including thermal smearing, using the *CRYSTAL09* package (Dovesi *et al.*, 2009) at the DFT level using the B3LYP functional. Since in *CRYSTAL09* the Bloch functions are expanded as a linear combination of Gaussian-type functions centred on the atoms of the periodic system, the calculation was carried out using the standard 6-31G(2d,2p) basis set.

In order to mimic the experimental refinement as much as possible, the calculation of theoretical structure factors was limited to the $\sin\theta/\lambda = 1.2$ Å^{-1} resolution and, afterwards, a static multipole model refinement (*i.e.*, ADPs set to zero) of the theoretical data was performed without optimizing atomic positions, thus eliminating an important source of correlation between parameters. As for the experimental refinement, the multipole expansion was truncated at the hexadecapole level for the non-H atoms and only bond-directed dipoles and quadrupoles were refined for the H atoms. Both κ and κ'_l parameters were refined independently for each heavy atom, while only one κ was refined for all the H atoms. The rigid local symmetry constraints and chemical equivalences were also consistent with the experimental refinement (see §5).

4.5. Unconstrained and X-ray constrained ELMO and Hartree–Fock calculations

The X-ray constrained ELMO strategy has been implemented modifying version 8 of the *GAMESS-UK* quantum-chemistry package (Guest *et al.*, 2005), which has been used to perform all the unconstrained ($\lambda = 0$) and constrained ($\lambda \neq 0$) ELMO calculations that will be discussed hereafter.

In order to assess the fitting effects on the ELMO electron density of the glycylglycine molecule, ELMO and XC-ELMO calculations have been performed using the 6-31G, 6-31G(d,p) and cc-pVDZ basis sets and considering both the IAM and the MM experimental molecular geometries obtained from the X-ray diffraction experiment. Electron-density-related properties derived from the constrained ELMO wavefunctions were afterwards compared to those corresponding to the unconstrained ELMO wavefunctions.

For all the ELMO calculations, the adopted localization scheme almost corresponds to the Lewis structure of the molecule, with atomic fragments, which describes the core electrons and the lone pairs associated with each atom, and with bond subunits, which describe all the electron pairs between each couple of nuclei. The only exceptions are represented by two three-atom fragments: one for the σ and the π electrons of the amide group O1—C2—N2 (comprising also the electrons for the delocalized lone pair of the nitrogen atom) and another one for the σ and the π electrons of the carboxylic group O2—C4—O3. The same localization scheme has been used for all the XC-ELMO calculations for which the unit-cell parameters and the ADPs associated with the different refinement models were also taken into account. Concerning the experimental structure-factor amplitudes used to constrain the ELMO wavefunctions, only those characterized by $F_{\mathbf{h},\mathrm{obs}} > 3\sigma_{\mathbf{h},\mathrm{obs}}(F_{\mathbf{h},\mathrm{obs}})$ were selected (overall 5467 reflections). As anticipated, the set of amplitudes $\{F_{\mathbf{h},\mathrm{obs}}\}$ was previously corrected for secondary extinctions. Furthermore, the scale factor η was properly optimized during the XC-ELMO computations.

For the sake of comparison, using the *TONTO* package (Jayatilaka & Grimwood, 2003), we have computed unconstrained and X-ray constrained Hartree–Fock electron densities (ρ^{HF} and $\rho^{\mathrm{XC\text{-}HF}}$, respectively) for the three selected basis sets. These densities are to be compared with the corresponding ELMO and XC-ELMO densities.[3]

4.6. Topological analyses

The quantum theory of atoms in molecules (QTAIM) (Bader, 1990) has been exploited to properly analyse all the obtained charge distributions. In particular, the *TOPXD* module (Volkov *et al.*, 2000) was used to partition and integrate the atomic basins of all the multipole-fitted electron densities, namely, $\rho^{\mathrm{MM/XC}}$, $\rho^{\mathrm{MM/P\text{-}B3LYP}}$, $\rho^{\mathrm{MM/ELMO}}$ and $\rho^{\mathrm{MM/XC\text{-}ELMO}}$. The *TOPOND98* software (Gatti, 1998) was used to perform the topological analysis of the periodic B3LYP/6-31G(2d,2p) electron density $\rho^{\mathrm{P\text{-}B3LYP}}$, while the analyses of all the ρ^{ELMO}, $\rho^{\mathrm{XC\text{-}ELMO}}$, ρ^{HF} and $\rho^{\mathrm{XC\text{-}HF}}$ electron distributions were performed with the *AIMAll* software (Keith, 2013).

Given the complexity associated with the definition of zero-flux surfaces in the electron density and with the subsequent integration (Popelier, 1998), the numerical integration error of the atomic basins, which is defined by

$$L(\Omega) = -\frac{1}{4}\int_{\Omega} \nabla^2\rho(\mathbf{r})\,\mathrm{d}\mathbf{r} \qquad (16)$$

and which should be zero for an ideal integration (Bader, 1990), was also carefully monitored. In this work, the values of $L(\Omega)$ approximately ranged from 3×10^{-6} to 2×10^{-3} a.u.

[3] XC-HF wavefunctions were computed at the same $\lambda_{\max}$ values as the corresponding XC-ELMO wavefunctions, except for the 6-31G basis set, for which self-consistent field (SCF) convergence could not be achieved for $\lambda \geq 0.40$ (see Table 2).

electronic reprint

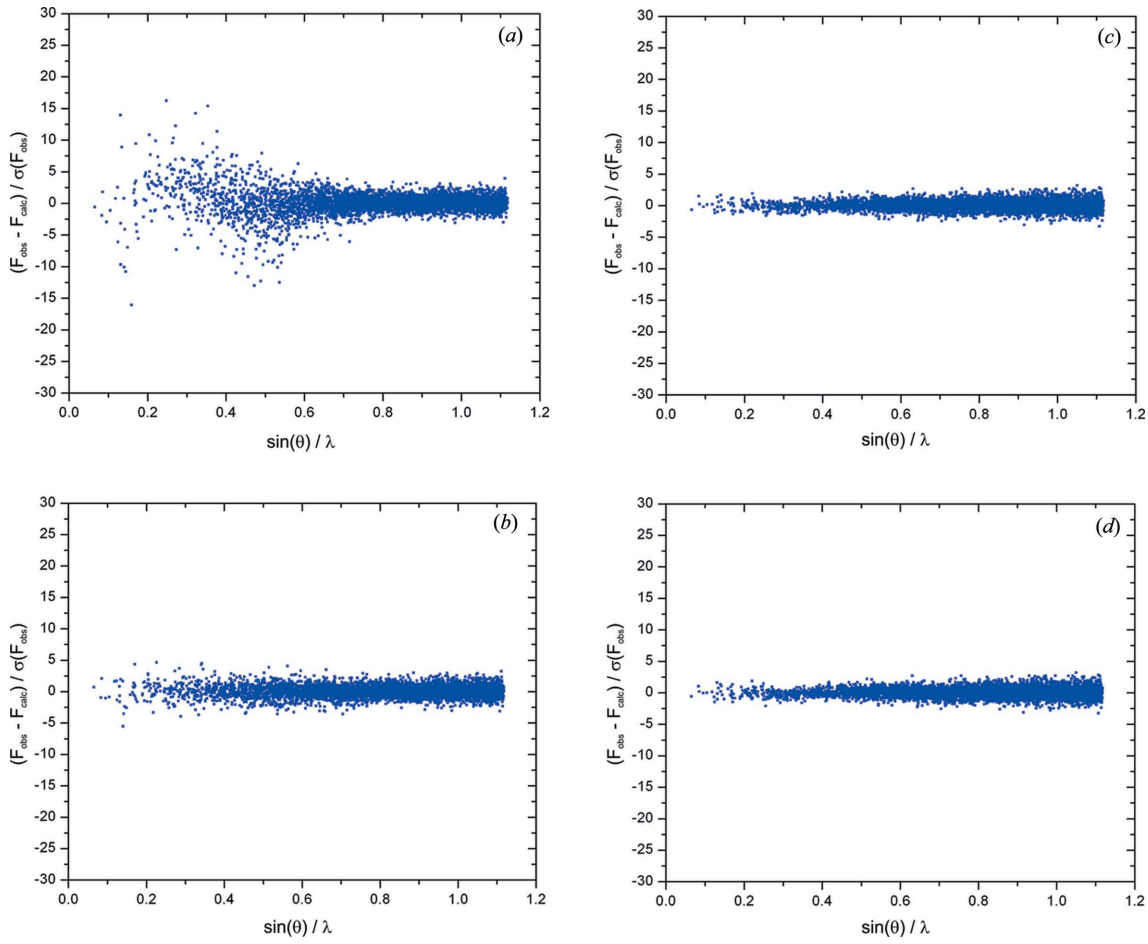


Figure 2
Normalized residuals of the structure-factor amplitudes *versus* the scattering resolution for (*a*) the IAM and the different multipole models (*b*) MM1, (*c*) MM2 and (*d*) MM3 refined against the experimental set of structure factors.

5. Fitting effects on the multipole model

In this section, we analyse the effects of fitting several Hansen & Coppens multipole models against the experimentally collected structure factors. In other words, we investigate step-by-step the relaxation of the local symmetries of atoms from spherical and neutral to aspherical and charged. After the spherical atom refinement (§4.2), an initial multipole model (MM1) was refined using several rigid constraints on the local symmetry of the atomic density functions (*mm*2 for O1, O2, O3 and C4; *m* for C1, C2, C3 and N2; 3*m* for N1; see Fig. 1*a* for atomic labels) and chemical equivalences (O2 = O3, C1 = C3). These constraints imply sp^2 hybridization for N2, C4 and for all the O atoms, and sp^3 hybridization for N1, C1 and C3 (however, a 'perfect' hybridization includes here an asymmetric polarization induced by chemical bonds to atoms of different electronegativity). The H atoms were treated with cylindrical symmetry ($m\infty$) and those bonded to equivalent heavier atoms were also treated as equivalent.

The atomic symmetries were progressively reduced, using the statistical parameters *R* and *wR* and the residual density maps as guides to select the best refinement model. The chemical equivalences were then removed and the following local symmetries were used (MM2 model): *mm*2 for C4, 3*m* for N1 and *m* for all the other non-H atoms, and $m\infty$ for all the H's. This implies sp^2 hybridization for C4, sp^3 for N1, C1 and C3, but a mixed character for all the O atoms and N2. In the last cycle, the coordinates and the ADPs for all the non-H atoms, the coordinates for the H atoms, the κ and κ'_l parameters, the multipole populations, the extinction parameter ε and the overall scale factor η were refined together, for an overall optimization of 271 parameters. A satisfactory deconvolution of thermal motion from the deformation electron-density distribution has been obtained, as shown by the Hirshfeld rigid-bond test (Hirshfeld, 1976). In fact, the largest difference of mean-squares displacement amplitudes (DMSDA) was 5×10^{-4} Å^2, which is lower than the limit of 0.001 Å^2 suggested by Hirshfeld. Additionally, the comparison between the final ADPs of the MM2 model with those previously published for the neutron diffraction on glycylglycine (Kvick *et al.*, 1977) shows a mean absolute difference of 0.0008 (14) Å^2, taking into account however that the temperatures of the two experiments are different (100 K for the current X-ray diffraction *versus* 82 K for the neutron diffraction). The neutron and X-ray geometries are in good agreement, but, as expected, larger discrepancies occur for positions of H atoms: in fact, X—H distances are on average 0.04 (2) Å shorter in the MM2 model with respect to the

electronic reprint

Table 2
Statistical agreements and energy values corresponding to all the unconstrained and constrained calculations performed on glycylglycine considering the geometry and the ADPs resulting from the MM refinement.

$\chi^2 = [1/(N_r - N_p)] \sum_{\mathbf{h}} w_{\mathbf{h}}(\eta|F_{\mathbf{h},\mathrm{calc}}| - |F_{\mathbf{h},\mathrm{obs}}|)^2$, $\%R(F) = 100[\sum_{\mathbf{h}}|\eta|F_{\mathbf{h},\mathrm{calc}}| - |F_{\mathbf{h},\mathrm{obs}}||/\sum_{\mathbf{h}}|F_{\mathbf{h},\mathrm{obs}}|]$, $\%wR(F) = 100(\chi^2/\sum_{\mathbf{h}} w_{\mathbf{h}}F^2_{\mathbf{h},\mathrm{obs}})^{1/2}$ with $w_{\mathbf{h}} = 1/\sigma^2_{\mathbf{h},\mathrm{obs}}(F_{\mathbf{h},\mathrm{obs}})$.

Basis set	χ^2	$\%R(F)$	$\%wR(F)$	Energy (a.u.)	$\lambda_{\max}$	χ^2	$\%R(F)$	$\%wR(F)$	Energy (a.u.)
ELMO calculations†					XC-ELMO calculations				
6-31G	2.56	2.62	3.44	−489.161	0.40	1.27	2.17	2.43	−489.060
6-31G(d,p)	1.74	2.38	2.84	−489.385	0.12	1.00	2.04	2.15	−489.357
cc-pVDZ	1.66	2.36	2.78	−489.394	0.12	0.98	2.03	2.13	−489.368
Hartree–Fock calculations†					XC-Hartree–Fock calculations				
6-31G	2.11	2.47	3.13	−489.310	0.38‡	1.15	2.05	2.17	−489.227
6-31G(d,p)	1.41	2.25	2.56	−489.549	0.12	0.85	1.93	1.98	−489.528
cc-pVDZ	1.31	2.20	2.47	−489.574	0.12	0.83	1.93	1.97	−489.556

† The scale factors η have been optimized using the density matrices obtained from the corresponding unconstrained calculations. ‡ χ^2 reaches an asymptotic value above $\lambda = 0.3$ but no SCF convergence is found for $\lambda \geq 0.40$.

neutron diffraction geometry. The maximum and minimum residual peaks are +0.14 and −0.14 e Å^{-3}, with the residual density maps showing only few and small discrepancies that could not be removed by any deformation model.

For the sake of completeness, we have also performed a multipole refinement up to hexadecapole level for all non-H atoms and up to the quadrupole level for the H atoms without imposing any local symmetry constraint (MM3), namely we have refined all the corresponding multipoles, accounting 423 parameters overall. The final statistical agreements (see Table 1) are very close to those obtained in the last refinement cycle of the symmetry-constrained MM2 model described above.

Fig. 2 shows scatter plots of the difference between experimental and calculated structure-factor amplitudes normalized by the experimental standard deviations as a function of the $\sin\theta/\lambda$ resolution for the IAM and the three MM refinements. It is obvious that the multipolar refinement improves the agreement between measured and calculated structure factors compared to the IAM (Fig. 2*a*), even when rigid local symmetry constraints are extensively applied (MM1 model, Fig. 2*b*). The progressive reduction of the local symmetries and the removal of equivalences (MM2) further improve the agreement (Fig. 2*c*), whereas a model without any local symmetry constraint (MM3) does not further reduce the normalized residuals (Fig. 2*d*). In fact, the number of reflections computed within $\pm 1\sigma$ from the observed ones is 3336, 3888, 4236 and 4276, for IAM, MM1, MM2 and MM3, respectively. The Hamilton significance test (Hamilton, 1965) also indicates that removing the local symmetry constraints from MM2 does not lead to a significant model improvement, despite using 152 additional parameters. Therefore, the models MM2 and MM3 are not statistically different and, from now on, we only refer to the results obtained using the local symmetry constrained multipole model MM2 just described. A similar conclusion can be reached refining the P-B3LYP structure factors with these two models, meaning also that the theoretical crystal density shows such a local symmetry for all the atoms.

Attempts to refine core deformations were not carried out, as this would be beyond the scope of this paper. Such a study would require even higher resolution and, at present, has been applied only on smaller and more symmetric crystals (Fischer *et al.*, 2011).

Based on the difference density maps, the residual density analysis and the normal probability plot provided in the supporting information, one can easily conclude that both the measured intensities and the refined multipole models are of extremely good quality and therefore could be a very good benchmark for the XC-ELMO calculations.

6. Fitting effects on the ELMO wavefunctions

6.1. Agreement statistics and energy

The agreement statistics and the energies for all the X-ray constrained computations on glycylglycine, using the geometry and the ADPs from MM2 refinement, are reported in Table 2. For each basis set, the XC-ELMO calculations are of course in better agreement with the measured intensities. For the 6-31G basis set, however, the desired agreement cannot be reached ($\chi^2 = 1.27$ for $\lambda_{\max} = 0.40$, with an asymptote above 1.0), whereas it is quite smoothly obtained

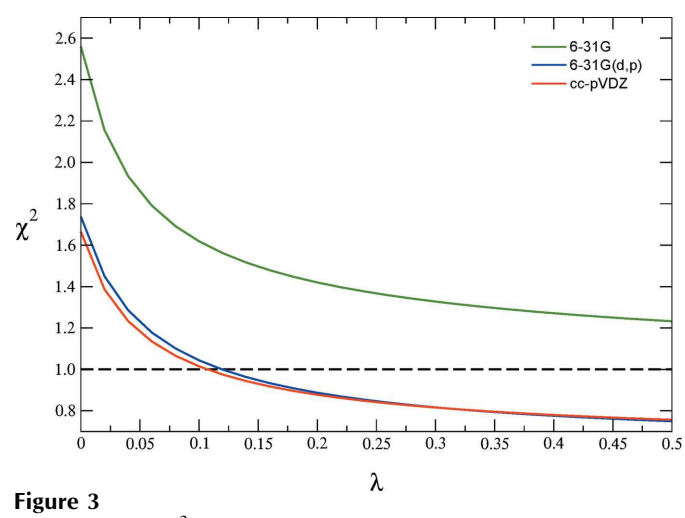


Figure 3
The variation of χ^2 agreement statistic with the Lagrange multiplier λ for the XC-ELMO 6-31G (green), 6-31G(d,p) (blue) and cc-pVDZ (red) calculations, using geometry and ADPs from the multipole model MM2.

electronic reprint

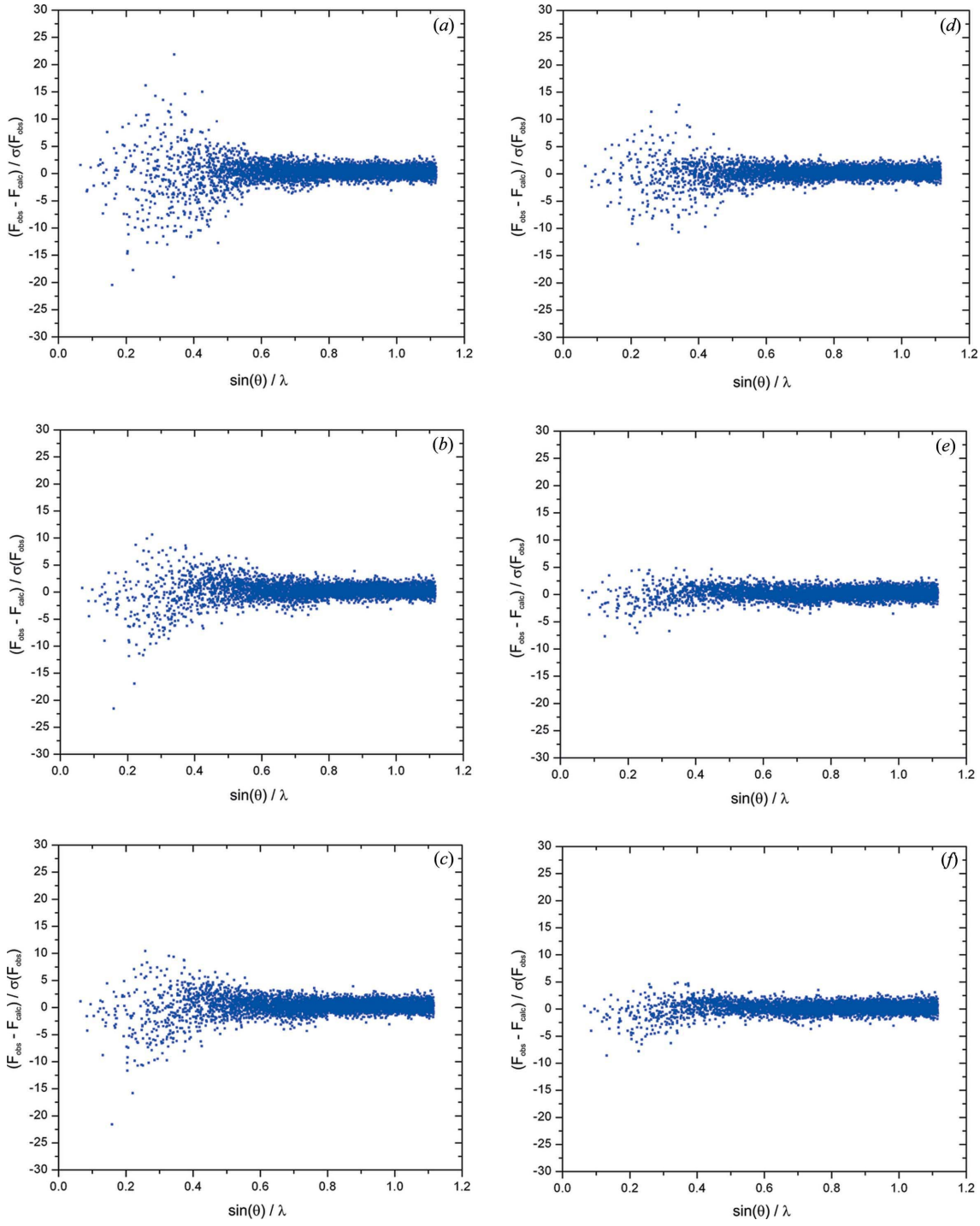


Figure 4
Normalized residuals of the structure-factor amplitudes *versus* the scattering resolution for the unconstrained ELMO (*a*) 6-31G, (*b*) 6-31G(d,p) and (*c*) cc-pVDZ wavefunctions and for the XC-ELMO (*d*) 6-31G, (*e*) 6-31G(d,p) and (*f*) cc-pVDZ wavefunctions.

for the larger and more flexible basis sets 6-31G(d,p) and cc-pVDZ ($\chi^2 \leq 1$ for λ as large as 0.12) (see Fig. 3). Here we point out that for the 6-31G basis set the second termination criterion in equation (13) is satisfied, while in the other two cases we have observed the fulfilment of the more traditional condition $\chi^2 < 1$.

χ^2 rapidly decreases as λ increases, showing that even a weak constraint to the X-ray data is sufficient to improve the agreement significantly. The asymptotic value of χ^2 is slightly above the limit obtained for the multipolar models (0.57 and 0.55 for MM2 and MM3, respectively) and smaller than for the IAM (2.60). The agreement indexes R and wR shown in Table 2 mirror the behaviour of the χ^2 statistics, but they are much less sensitive to λ and, therefore, less useful to compare the quality of the different constrained wavefunctions. On the other hand, R and wR can be used for comparison against the multipole models (see Table 1), which of course give better agreements because they do not have to satisfy an energy

electronic reprint

minimization criterion and, above all, because they are not subject to the termination criterion $\chi^2 \leq 1$ imposed by equation (13). Furthermore, R and wR confirm that XC-ELMO performs much better than an IAM refinement.

Fig. 4 shows the normalized differences between experimental and ELMO or XC-ELMO structure-factor amplitudes as a function of the resolution. Almost all the structure-factor amplitudes computed from the 6-31G(d,p) and cc-pVDZ XC-ELMO wavefunctions (Figs. 4*e* and 4*f*) are within $\pm 5\sigma$ from the experimental values without any obvious resolution dependence. The number of reflections within the $\pm 1\sigma$ range is 4149 and 4240 for the 6-31G(d,p) and the cc-pVDZ basis sets, respectively, contrasting with 3663 and 3764 reflections within $\pm 1\sigma$ for the unconstrained ELMO 6-31G(d,p) and cc-pVDZ wavefunctions, respectively. On the other hand, for the XC-ELMO/6-31G wavefunction, for which $\chi^2 = 1.27$, only the structure-factor amplitudes calculated at resolution $\sin\theta/\lambda > 0.5$ Å^{-1} agree with the experimental values within $\pm 5\sigma$, while many low-angle structure factors significantly exceed the 5σ limit (Fig. 4*d*). Moreover, the XC-ELMO/6-31G plot shows a distribution of normalized residuals quite similar to that associated with the unconstrained ELMO wavefunctions (3689 and 3954 reflections are within the $\pm 1\sigma$ range for the unconstrained and the X-ray constrained ELMO 6-31G wavefunctions, respectively), meaning that the X-ray constraining procedure was actually not particularly effective, despite the better agreement indexes. In this respect, the comparison between Figs. 4(*d*), 4(*b*) and 4(*c*) is extremely elucidative because it shows that an X-ray constrained wavefunction is not better than an unconstrained one in the absence of polarization functions in the basis set. This demonstrates that the ELMO/6-31G wavefunction is definitely not flexible enough to fit the experimental data. On the contrary, if sufficient variational flexibility is present in the basis set, even unconstrained ELMO wavefunctions better reproduce the experimentally collected structure factors.

The constraint to strictly localize the molecular orbitals on molecular subunits introduces additional approximations over the usual Hartree–Fock method, which enables full delocalization of the canonical molecular orbitals. In order to quantify this effect, we have computed unconstrained and X-ray

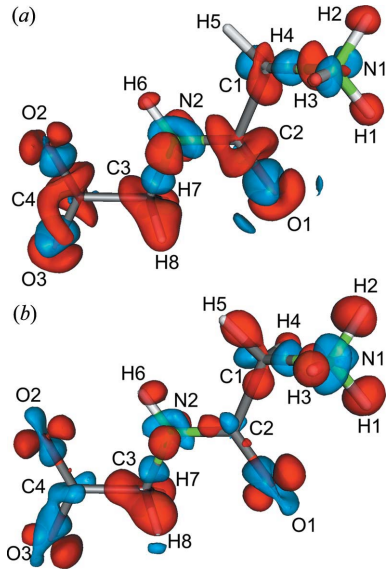


Figure 5
Three-dimensional plots of the electron-density difference $\rho^{\text{XC-ELMO}} - \rho^{\text{ELMO}}$ for (*a*) the 6-31G and (*b*) 6-31G(d,p) basis sets. The isosurface value is set to 0.005 a.u., with negative isosurfaces in red and positive isosurfaces in blue.

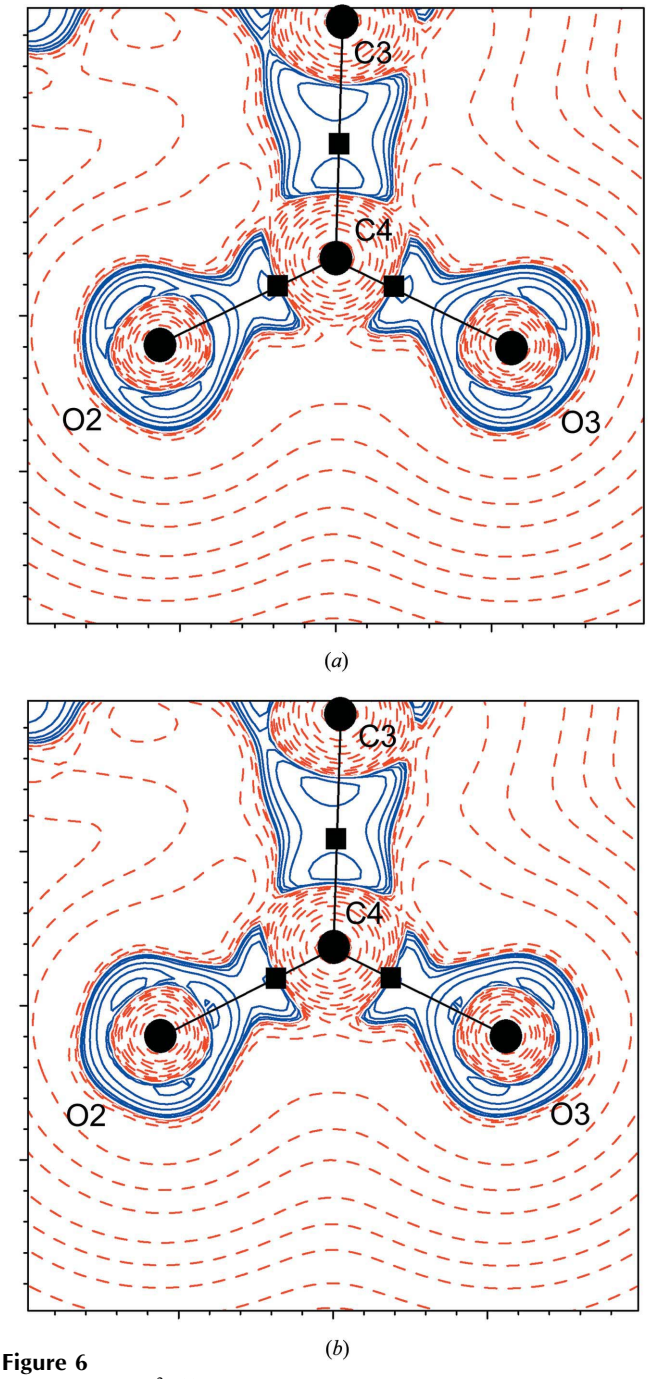


Figure 6
Plots of $L = -\nabla^2\rho(\mathbf{r})$ for (*a*) the ELMO/6-31G(d,p) and (*b*) XC-ELMO/6-31G(d,p) wavefunctions of glycylglycine in the carboxylate plane. Contours are drawn at intervals of $(\pm 2, \pm 4, \pm 8) \times 10^n$ e Å^{-5} ($n = -3$ to 3). Blue lines denote regions of charge concentration ($L > 0$) and red lines denote regions of charge depletion ($L < 0$).

electronic reprint

Table 3
Bond-critical-point data for all glycylglycine electron densities resulting from unconstrained and constrained ELMO calculations (data for the $\rho^{\text{P-B3LYP}}$ density are also shown).

For each bond critical point A–B, d_A is its distance from the nucleus A in Å, $\rho(\mathbf{r}_b)$ is its electron-density value in a.u. and $\nabla^2\rho(\mathbf{r}_b)$ is its Laplacian value in a.u.

	$\rho^{\text{P-B3LYP}}_{6-31\text{G(2d,2p)}}$	$\rho^{\text{ELMO}}_{6-31\text{G}}$	$\rho^{\text{ELMO}}_{6-31\text{G(d,p)}}$	$\rho^{\text{ELMO}}_{\text{cc-pVDZ}}$	$\rho^{\text{XC-ELMO}}_{6-31\text{G}}$	$\rho^{\text{XC-ELMO}}_{6-31\text{G(d,p)}}$	$\rho^{\text{XC-ELMO}}_{\text{cc-pVDZ}}$
O1−C2							
d_A	0.825	0.799	0.833	0.835	0.806	0.835	0.837
$\rho(\mathbf{r}_b)$	0.397	0.386	0.406	0.401	0.398	0.405	0.404
$-\nabla^2\rho(\mathbf{r}_b)$	0.409	0.922	0.346	0.243	0.917	0.158	0.123
O2−C4							
d_A	0.837	0.811	0.849	0.850	0.819	0.851	0.851
$\rho(\mathbf{r}_b)$	0.380	0.368	0.388	0.383	0.375	0.389	0.387
$-\nabla^2\rho(\mathbf{r}_b)$	0.532	0.908	0.387	0.330	0.842	0.217	0.229
O3−C4							
d_A	0.834	0.807	0.845	0.846	0.813	0.846	0.847
$\rho(\mathbf{r}_b)$	0.385	0.374	0.395	0.389	0.381	0.393	0.390
$-\nabla^2\rho(\mathbf{r}_b)$	0.519	0.952	0.413	0.342	0.888	0.223	0.194
N1−C1							
d_A	0.879	0.957	1.012	1.004	0.907	1.000	0.981
$\rho(\mathbf{r}_b)$	0.252	0.218	0.240	0.235	0.233	0.249	0.246
$-\nabla^2\rho(\mathbf{r}_b)$	0.642	0.326	0.271	0.307	0.521	0.502	0.575
N2−C2							
d_A	0.836	0.798	0.885	0.879	0.854	0.890	0.888
$\rho(\mathbf{r}_b)$	0.350	0.335	0.363	0.355	0.327	0.346	0.344
$-\nabla^2\rho(\mathbf{r}_b)$	1.282	1.110	1.153	1.213	0.959	0.846	0.927
N2−C3							
d_A	0.874	0.957	0.995	0.989	0.897	0.988	0.975
$\rho(\mathbf{r}_b)$	0.263	0.227	0.248	0.243	0.250	0.259	0.256
$-\nabla^2\rho(\mathbf{r}_b)$	0.712	0.259	0.203	0.253	0.629	0.429	0.528
C1−C2							
d_A	0.747	0.777	0.786	0.788	0.757	0.749	0.751
$\rho(\mathbf{r}_b)$	0.256	0.242	0.277	0.270	0.240	0.267	0.262
$-\nabla^2\rho(\mathbf{r}_b)$	0.560	0.509	0.862	0.772	0.505	0.770	0.703
C3−C4							
d_A	0.757	0.787	0.797	0.802	0.776	0.764	0.771
$\rho(\mathbf{r}_b)$	0.254	0.240	0.275	0.268	0.238	0.266	0.260
$-\nabla^2\rho(\mathbf{r}_b)$	0.551	0.503	0.867	0.778	0.505	0.771	0.705
N1−H1							
d_A	0.808	0.771	0.792	0.804	0.776	0.798	0.807
$\rho(\mathbf{r}_b)$	0.322	0.322	0.354	0.339	0.316	0.336	0.326
$-\nabla^2\rho(\mathbf{r}_b)$	1.849	-1.590	2.094	1.925	1.553	1.900	1.825
N1−H2							
d_A	0.785	0.741	0.763	0.774	0.741	0.768	0.776
$\rho(\mathbf{r}_b)$	0.352	0.346	0.378	0.364	0.342	0.353	0.346
$-\nabla^2\rho(\mathbf{r}_b)$	2.083	1.710	2.232	2.001	1.679	1.937	1.837
N1−H3							
d_A	0.804	0.763	0.784	0.797	0.768	0.793	0.803
$\rho(\mathbf{r}_b)$	0.325	0.324	0.356	0.342	0.324	0.338	0.330
$-\nabla^2\rho(\mathbf{r}_b)$	1.873	1.581	2.085	1.855	1.600	1.887	1.802
C1−H4							
d_A	0.708	0.696	0.692	0.697	0.695	0.709	0.723
$\rho(\mathbf{r}_b)$	0.269	0.268	0.297	0.290	0.267	0.287	0.284
$-\nabla^2\rho(\mathbf{r}_b)$	0.893	0.817	1.163	1.142	0.813	1.114	1.134
C1−H5							
d_A	0.692	0.680	0.674	0.684	0.676	0.690	0.703
$\rho(\mathbf{r}_b)$	0.287	0.282	0.312	0.304	0.276	0.298	0.294
$-\nabla^2\rho(\mathbf{r}_b)$	1.036	0.921	1.290	1.255	0.878	1.188	1.197
N2−H6							
d_A	0.789	0.750	0.768	0.779	0.759	0.779	0.788
$\rho(\mathbf{r}_b)$	0.331	0.332	0.363	0.348	0.332	0.349	0.339
$-\nabla^2\rho(\mathbf{r}_b)$	1.899	1.610	2.093	1.815	1.640	1.942	1.814
C3−H7							
d_A	0.678	0.641	0.629	0.639	0.647	0.655	0.667
$\rho(\mathbf{r}_b)$	0.289	0.298	0.327	0.318	0.285	0.300	0.297
$-\nabla^2\rho(\mathbf{r}_b)$	1.041	1.019	1.396	1.330	0.973	1.168	1.179
C3−H8							
d_A	0.646	0.631	0.619	0.638	0.620	0.630	0.642
$\rho(\mathbf{r}_b)$	0.317	0.316	0.345	0.335	0.289	0.303	0.301
$-\nabla^2\rho(\mathbf{r}_b)$	1.277	1.172	1.565	1.489	1.013	1.187	1.205

electronic reprint

constrained Hartree–Fock electron densities ($\rho^{\rm HF}$ and $\rho^{\rm XC\text{-}HF}$, respectively) for the three selected basis sets and compared them with the corresponding ELMO and XC-ELMO densities. Both unconstrained and X-ray constrained Hartree–Fock electron densities produce better statistical agreements than the corresponding ELMO and XC-ELMO electron distributions. Moreover, the gap between $\rho^{\rm HF}$ and $\rho^{\rm ELMO}$ is roughly constant as a function of λ (see Table 2). These results suggest that $\rho^{\rm HF}$ and $\rho^{\rm ELMO}$ are equally able to incorporate the information from the experimental structure-factor amplitudes.

In Table 2 the energies associated with all the unconstrained and constrained wavefunctions are also reported. As already observed (Grimwood & Jayatilaka, 2001; Hudák *et al.*, 2010; Genoni, 2013*a*), the energies of constrained wavefunctions are always higher, in keeping with what is expected in a variational procedure when a constraint is added without introducing a new variational parameter.

6.2. Electron-density distribution and its topology

Fig. 5 shows three-dimensional plots of $\rho^{\rm XC\text{-}ELMO} - \rho^{\rm ELMO}$ for the 6-31G and the 6-31G(d,p) basis sets [the distribution associated with the cc-pVDZ case is shown only in the supporting information because it is very similar to the 6-31G(d,p) one]. While for the 6-31G(d,p) basis set the main consequences of the fitting consist of a large redistribution of the electron density around the nuclei and only small rearrangements in the bonding regions, the 6-31G fitting procedure entails significant changes of the electron density both in the core and in the bonding domains. This is especially evident for the oxygen atoms, for which a depletion of electron density in the lone-pair regions is noteworthy. Other important fitting effects are the reduction of electron density associated with the C—H bonds and the shifting of electronic charge from the H atoms to the N atoms. Shifts of electron density from C3 to C4 and, analogously, from C2 to N2 and from N1 to C1 are also observed.

Topological properties at the bond critical points of $\rho^{\rm ELMO}$ and $\rho^{\rm XC\text{-}ELMO}$ are gathered in Table 3 and they are compared to the results obtained from the periodic B3LYP/6-31G(2d,2p) computation ($\rho^{\rm P\text{-}B3LYP}$). The properties obtained from the XC-ELMO wavefunctions are in general similar to those obtained from the corresponding unconstrained ELMO calculations, especially the electron density at the bond critical points, $\rho(\mathbf{r}_{\rm b})$. Upon closer inspection, we see that the XC-ELMO $\rho(\mathbf{r}_{\rm b})$ generally approach the P-B3LYP limit for the more complete and polarized basis sets.

The analysis of the Laplacian at the bond critical points, $\nabla^2\rho(\mathbf{r}_{\rm b})$, shows much larger discrepancies. As previously discussed (Macchi *et al.*, 1998; Bytheway, Grimwood & Jayatilaka, 2002), this is a consequence of the intrinsic nature of the Laplacian, especially for polar bonds. Fig. 6 depicts the plots of $L = -\nabla^2\rho(\mathbf{r})$, which show regions of electron-density concentration ($L > 0$) and depletion ($L < 0$), in the carboxylate group plane of glycylglycine. The constrained or unconstrained plots are qualitatively very similar, showing regions of electronic charge concentration along the C—C and C—O bonds as well as in the lone-pair regions of the oxygen atoms. The atomic graphs of C4, O2 and O3 have the expected trigonal arrangement of three charge concentration maxima in both constrained and unconstrained cases. Nevertheless, because polar bond critical points lie close to nodal surfaces of the Laplacian, small changes in the position of these bond critical points may lead to large changes of $\nabla^2\rho(\mathbf{r}_{\rm b})$. This explains the large differences between constrained and unconstrained $\nabla^2\rho(\mathbf{r}_{\rm b})$ values found in Table 3 and stresses the importance of correctly locating the critical point along the corresponding bond path (Bytheway, Grimwood & Jayatilaka, 2002).

The topological properties computed for both the $\rho^{\rm HF}$ and $\rho^{\rm XC\text{-}HF}$ electron densities are deposited in the supporting information (Table S9). The comparison with the corresponding ELMO and XC-ELMO densities indicates that the ELMO approximation does not introduce significant changes in the electron distributions, in keeping with the small differences between the agreement indices reported in Table 2. For example, the difference in electron density at the bond critical points between $\rho^{\rm ELMO}$ and $\rho^{\rm HF}$ and between $\rho^{\rm XC\text{-}ELMO}$ and $\rho^{\rm XC\text{-}HF}$ are usually less than 5% for the 6-31G basis set and even less for the larger 6-31G(d,p) and cc-pVDZ basis sets.

6.3. QTAIM atomic charges and dipoles

Bar graphs of the net atomic charges are shown in Fig. 7. Overall, charges from ELMO and XC-ELMO wavefunctions are qualitatively similar, addressing negatively charged O and N atoms, positively charged C and H atoms of the amino group and basically neutral methylenic H atoms. For the O, N and C atoms, the XC-ELMO/6-31G charges are relatively close to the ELMO/6-31G results with the largest discrepancy around 0.13 a.u. The differences are even smaller for the polarized and more flexible basis sets. For the H atoms, the relative changes are larger, but these charges are very small, so the largest absolute differences are around 0.04 a.u. Furthermore, the results obtained from the $\rho^{\rm XC\text{-}ELMO}$ and $\rho^{\rm P\text{-}B3LYP}$ densities are similar for all the basis sets, with the XC-ELMO charges slightly larger in absolute values. Surprisingly, the best agreement with the $\rho^{\rm P\text{-}B3LYP}$ results is found for the charges calculated using the smaller 6-31G basis set. However, the results previously discussed indicate that this better agreement must be just incidental.

Atomic dipole moments measure the displacement of the centroid of an atomic charge density from its nucleus and can be calculated from the integration of the dipolar density function $\mathbf{r}\rho(\mathbf{r})$ inside the atomic basin Ω (Bader, 1990):

$$\boldsymbol{\mu}(\Omega) = -\int_{\Omega} \mathbf{r}\rho(\mathbf{r})\,{\rm d}\mathbf{r}. \qquad (17)$$

Magnitudes of atomic dipole moments are given as bar graphs in Fig. 8. For O atoms, ELMO/6-31G and XC-ELMO/6-31G dipole moments are markedly underestimated compared to the more polarized basis sets and P-B3LYP. For these atoms, the extra *d* functions are therefore vital to describe the internal polarization. Using $\rho^{\rm P\text{-}B3LYP}$ as a benchmark, we see

electronic reprint

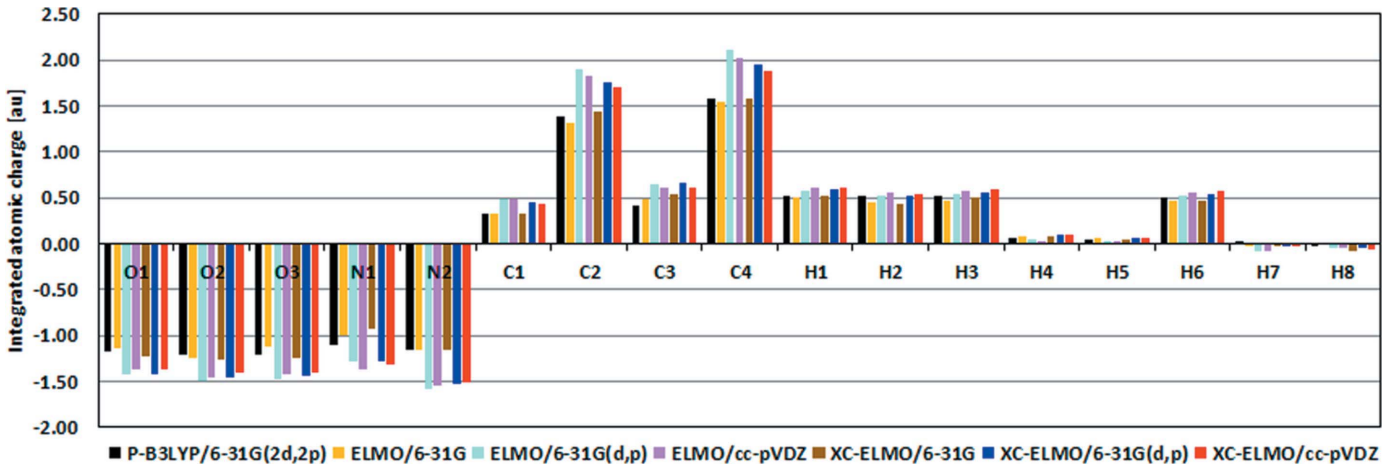


Figure 7
Integrated atomic charges (a.u.) from unconstrained (ρ^{ELMO}) and constrained ($\rho^{\mathrm{XC\text{-}ELMO}}$) ELMO densities and from the *ab initio* periodically calculated density ($\rho^{\mathrm{P\text{-}B3LYP}}$).

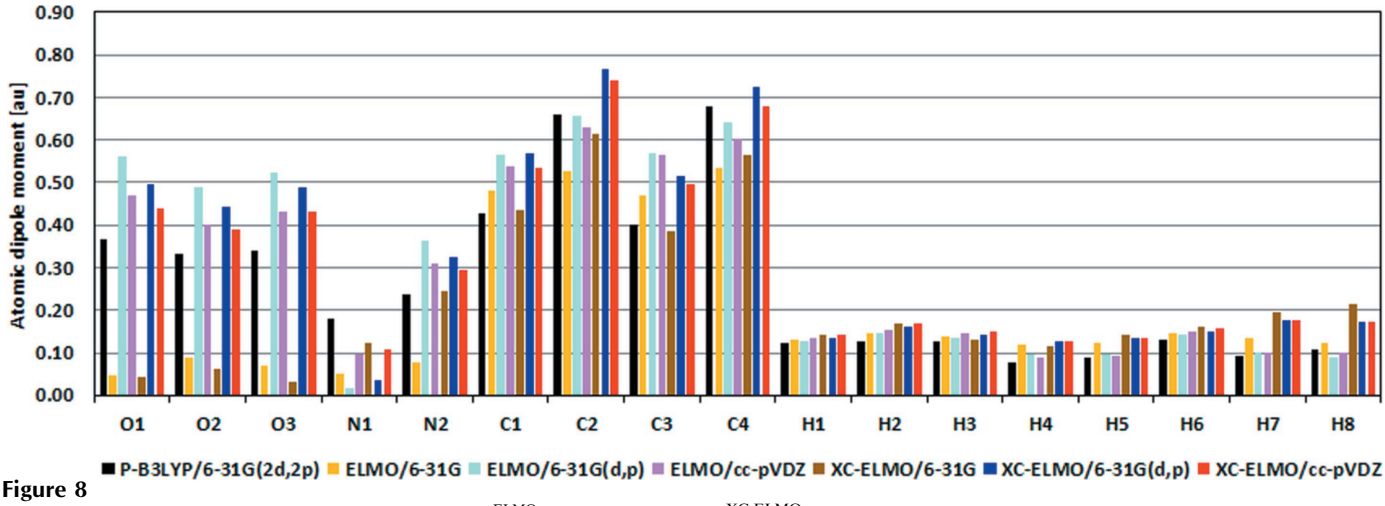


Figure 8
Atomic dipole moments (a.u.) from unconstrained (ρ^{ELMO}) and constrained ($\rho^{\mathrm{XC\text{-}ELMO}}$) ELMO densities and from the *ab initio* periodically calculated density ($\rho^{\mathrm{P\text{-}B3LYP}}$).

that the X-ray constraining improves the dipole moments of O atoms for the 6-31G(d,p) and the cc-pVDZ ELMO wavefunctions, although they remain considerably overestimated. The largest discrepancies between unconstrained and constrained calculations are always observed for the carbonylic oxygen O1, with decrements of 0.06 and 0.03 a.u. for the 6-31G(d,p) and cc-pVDZ basis sets, respectively.

On the contrary, for the N atoms, the XC-ELMO/6-31G wavefunction gives a dipole moment closer to the one associated with $\rho^{\mathrm{P\text{-}B3LYP}}$. For N1 and N2, the X-ray constraining is helpful, whereas the polarization functions are not so necessary. The largest difference between ELMO and XC-ELMO dipole moments is 0.07 a.u. for N1, the amino nitrogen involved as 'donor' atom in some hydrogen bonds with neighbouring molecules.

For C and H atoms, the polarized basis sets usually perform better and the dipole moment magnitude improves after the fitting procedure. The directions of the atomic dipole moments are substantially similar for all methods, with a maximum difference smaller than 2°.

A more important analysis is that of the molecular dipole moment, which can be expressed in terms of two contributions: the integration of the dipolar density function over atomic basins, equation (17), and the net atomic charge $q(\Omega)$. Therefore,

$$\boldsymbol{\mu}_{\mathrm{MOL}} = \sum_{\Omega}\left[\boldsymbol{\mu}(\Omega) + q(\Omega)\mathbf{R}_{\Omega}\right], \qquad (18)$$

where $\mathbf{R}_{\Omega}$ is the position vector of the nucleus of atom Ω with respect to an arbitrary origin. As previously shown, both terms are equally important in the description of a molecular dipole (Bader *et al.*, 1987; Laidig & Bader, 1990). Magnitudes of the molecular dipole moments for glycylglycine are given as bar graphs in Fig. 9(*a*). Spackman *et al.* (2007) have shown that typical enhancements of molecular dipole moments from the gas phase to crystals are within 10–40%, depending on the polarizability of the molecule and its specific packing in the solid state. The gas-phase B3LYP/6-31G(2d,2p) molecular dipole moment of glycylglycine is 9.445 a.u. whereas the P-B3LYP molecular dipole moment, using the same basis set,

electronic reprint

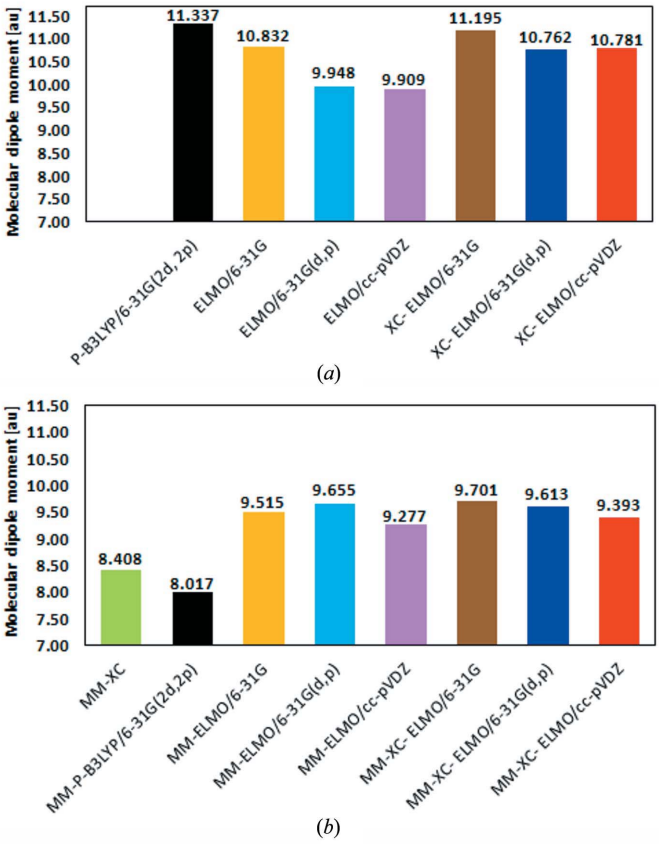


Figure 9
Molecular dipole moments (a.u.) from unconstrained ($\rho^{\rm ELMO}$) and constrained ($\rho^{\rm XC\text{-}ELMO}$) ELMO densities and from the *ab initio* periodically calculated density ($\rho^{\rm P\text{-}B3LYP}$). Dipole moments calculated from both (*a*) the primary densities and (*b*) the multipole-projected densities are shown.

is 11.337 a.u., corresponding to an enhancement of 20%. In our work, the difference between $\rho^{\rm P\text{-}B3LYP}$ and $\rho^{\rm XC\text{-}ELMO}$ dipole moments may inform us on the ability of XC-ELMO wavefunctions to account for intermolecular crystal-field effects. In fact, from Fig. 9(*a*), we see that the X-ray constraining procedure makes the molecular dipole moment quite close to the P-B3LYP value, but the increase is smaller than 10% for all the basis sets. The value obtained with the 6-31G basis set is closer to the P-B3LYP one, but this again seems a consequence of the fortuitous agreement on atomic charges described before. In reality, we learn from Fig. 9(*a*) that, as the basis sets become more complete and flexible, the dipole moments obtained through the X-ray constrained wavefunctions converge to a value that is *circa* 5% smaller than the P-B3LYP one.

7. Fitting effects on the multipole-model-projected ELMO electron densities

In order to avoid potential ambiguities due to the multipolar expansion used to model the benchmark experimental electron density, we have also projected the P-B3LYP, all the ELMO and XC-ELMO electron densities in terms of Hansen & Coppens multipoles, using the very same MM2 model

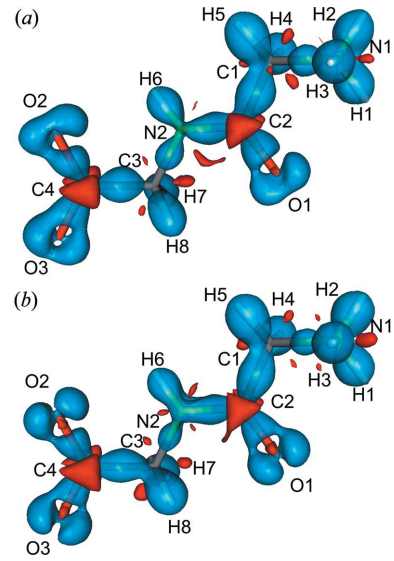


Figure 10
Three-dimensional plots of the deformation electron densities (*a*) $\rho^{\rm MM/XC} - \rho^{\rm IAM}$ and (*b*) $\rho^{\rm MM/P\text{-}B3LYP} - \rho^{\rm IAM}$ for glycylglycine. The isosurface value is set to 0.02 a.u., with negative isosurfaces in red and positive isosurfaces in blue.

discussed in §5. This will enable us to compare all the densities, affected in the same way by the inherent limitations of the multipolar expansion (Fischer *et al.*, 2011). Of course, all the atomic coordinates and ADPs (for XC-ELMO) were kept fixed to those from the MM2 refinement against experimental data.

7.1. Deformation density

Fig. 10 shows the deformation densities $\rho^{\rm MM/XC} - \rho^{\rm IAM}$ and $\rho^{\rm MM/P\text{-}B3LYP} - \rho^{\rm IAM}$ for glycylglycine. As expected, the plots clearly show an accumulation of electron density in all the covalent bonds and in the lone-pair domains of the oxygen atoms. Electron-density depletions are mainly concentrated around the nuclei.

In Fig. 11, we show the deformation density plots ($\rho^{\rm MM/ELMO} - \rho^{\rm IAM}$) and ($\rho^{\rm MM/XC\text{-}ELMO} - \rho^{\rm IAM}$) using the 6-31G and 6-31G(d,p) basis sets. At the ELMO/6-31G level (Fig. 11*a*) the electron accumulation in the bonding regions is significantly smaller, especially for the N2—C3 and all the C—O bonds. The experimental constraint gives only a slight improvement (Fig. 11*b*). This result confirms that constraining an ELMO wavefunction to experimental structure factors leads to a meaningful electron-density distribution only if the basis set is sufficiently flexible. In fact, the deformation densities derived from the unconstrained and constrained ELMO/6-31G(d,p) wavefunctions (Figs. 11*c* and 11*d*) are significantly closer to the $\rho^{\rm MM/XC}$ and $\rho^{\rm MM/P\text{-}B3LYP}$ deformation densities.

The $\rho^{\rm MM/ELMO}$ deformation density calculated with the polarized 6-31G(d,p) basis set (Fig. 11*c*) is already very similar to the $\rho^{\rm MM/P\text{-}B3LYP}$ deformation density (Fig. 10*b*), although only the latter takes into account crystalline environment effects through a fully periodic approach.

electronic reprint

Table 4
Bond-critical-point data corresponding to all the unconstrained and constrained multipole-model-projected electron densities for the shortest hydrogen bonds in the crystal structure of glycylglycine (data for the $\rho^{\text{P-B3LYP}}$ density are also shown).

For each bond critical point H···O, d_{H} is its distance from the nucleus H in Å, $\rho(\mathbf{r}_{\text{b}})$ is its electron-density value in a.u., $\nabla^2\rho(\mathbf{r}_{\text{b}})$ is its Laplacian value in a.u. and λ_3 is its positive curvature in a.u.

	$\rho^{\text{P-B3LYP}}_{6-31\text{G(2d,2p)}}$	$\rho^{\text{MM/XC}}$	$\rho^{\text{MM/P-B3LYP}}_{6-31\text{G(2d,2p)}}$	$\rho^{\text{MM/ELMO}}_{6-31\text{G}}$	$\rho^{\text{MM/ELMO}}_{6-31\text{G(d,p)}}$	$\rho^{\text{MM/ELMO}}_{\text{cc-pVDZ}}$	$\rho^{\text{MM/XC-ELMO}}_{6-31\text{G}}$	$\rho^{\text{MM/XC-ELMO}}_{6-31\text{G(d,p)}}$	$\rho^{\text{MM/XC-ELMO}}_{\text{cc-pVDZ}}$
H1···O2^{i}									
d_{H}	0.713	0.680	0.695	0.679	0.663	0.661	0.675	0.665	0.663
$\rho(\mathbf{r}_{\text{b}})$	0.024	0.030 (3)	0.033	0.032	0.029	0.029	0.031	0.030	0.030
$\nabla^2\rho(\mathbf{r}_{\text{b}})$	0.081	0.109 (1)	0.102	0.105	0.098	0.097	0.109	0.098	0.097
λ_3	0.14	0.20	0.20	0.20	0.18	0.18	0.20	0.19	0.18
H2···O3^{ii}									
d_{H}	0.610	0.650	0.644	0.641	0.624	0.624	0.651	0.628	0.630
$\rho(\mathbf{r}_{\text{b}})$	0.038	0.043 (4)	0.041	0.041	0.037	0.038	0.042	0.039	0.040
$\nabla^2\rho(\mathbf{r}_{\text{b}})$	0.130	0.115 (1)	0.103	0.126	0.116	0.107	0.116	0.109	0.103
λ_3	0.24	0.26	0.24	0.26	0.24	0.23	0.25	0.24	0.23
H3···O2^{iii}									
d_{H}	0.627	0.666	0.679	0.652	0.647	0.644	0.641	0.646	0.642
$\rho(\mathbf{r}_{\text{b}})$	0.038	0.036 (4)	0.038	0.037	0.033	0.033	0.039	0.033	0.034
$\nabla^2\rho(\mathbf{r}_{\text{b}})$	0.107	0.111 (1)	0.110	0.134	0.129	0.126	0.116	0.119	0.117
λ_3	0.22	0.23	0.23	0.25	0.22	0.22	0.24	0.22	0.22
H6···O1^{iv}									
d_{H}	0.904	0.964	0.968	0.950	0.944	0.946	0.942	0.953	0.954
$\rho(\mathbf{r}_{\text{b}})$	0.014	0.011 (1)	0.011	0.011	0.010	0.011	0.012	0.011	0.011
$\nabla^2\rho(\mathbf{r}_{\text{b}})$	0.046	0.047 (1)	0.047	0.048	0.044	0.045	0.051	0.046	0.047
λ_3	0.07	0.07	0.07	0.07	0.06	0.07	0.08	0.07	0.07

Symmetry codes: (i) $-x+1, y+\frac{1}{2}, -z+\frac{3}{2}$; (ii) $x+1, y, z+1$; (iii) $-x+1, -y, -z+1$; (iv) $-x+1, y-\frac{1}{2}, -z+\frac{3}{2}$.

A finer analysis shows that Fig. 11(*d*) [deformation density for $\rho^{\text{MM/XC-ELMO}}$ with 6-31G(d,p) basis set] is closer to Fig. 10(*a*) ($\rho^{\text{MM/XC}}$) than to Fig. 10(*b*) ($\rho^{\text{MM/P-B3LYP}}$), in particular for the lone-pair domains of the O atoms. This means that some features of the deformation densities are genuinely due to the X-ray constraint rather than to a crystal-field effect (at least if calculated at the B3LYP level). This is particularly encouraging because it shows the ability of XC-ELMO to extract information from experimental intensities.

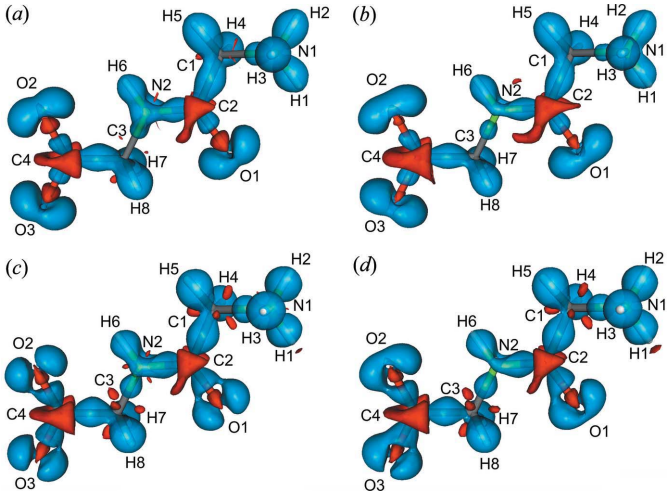


Figure 11
Deformation electron-density plots for glycylglycine obtained from unconstrained ($\rho^{\text{MM/ELMO}} - \rho^{\text{IAM}}$) and constrained ($\rho^{\text{MM/XC-ELMO}} - \rho^{\text{IAM}}$) ELMO wavefunctions. (*a*) ELMO/6-31G; (*b*) XC-ELMO/6-31G; (*c*) ELMO/6-31G(d,p); (*d*) XC-ELMO/6-31G(d,p). The isosurface value is set to 0.02 a.u., with negative isosurfaces in red and positive isosurfaces in blue.

7.2. Topological properties, charges and dipoles

A full topological analysis of all the electron-density distributions projected to the Hansen & Coppens MM2 multipole model is given in the supporting information. The results are in good agreement with the topological features shown in Table 3, *i.e.* the multipole model projections are not significantly different from the corresponding non-projected electron distributions: $\rho(\mathbf{r}_{\text{b}})$ differ by less than 5%, the bond-critical-point positions are on average within 0.08 a.u. As expected, $\nabla^2\rho(\mathbf{r}_{\text{b}})$ change more significantly. In general, $\rho^{\text{MM/XC-ELMO}}$ are closer to the $\rho^{\text{MM/XC}}$ and $\rho^{\text{MM/P-B3LYP}}$ benchmark densities than $\rho^{\text{MM/ELMO}}$ and, more importantly, the agreement is better when larger basis sets 6-31G(d,p) and cc-pVDZ are used.

Having multipolar projected electron densities, it is possible to thoroughly analyse the hydrogen bonds and comment on the performances of the XC-ELMO approach, in particular verifying if there is any improvement upon the fitting. Fig. 1(*b*) shows the pattern of strongest hydrogen bonds in the crystal structure of glycylglycine and Table 4 collects their topological features. All these bonds are of N—H···O type, with the H···O distance ranging from 1.77 (2) to 1.98 (2) Å.

Topological features of hydrogen bonds from experimental electron-density analyses have often been used to classify various types of interactions (Rozas *et al.*, 2000; Espinosa *et al.*, 1999). However, Spackman (1999) observed that for many medium–weak hydrogen bonds, the electron density, its topology and the local energy densities can be well approximated even using the pro-molecule, that is the electron-density distribution calculated from the IAM. As a matter of fact, the topological properties of hydrogen bonds calculated

electronic reprint

Table 5
Statistical agreements and energy values corresponding to all the unconstrained and constrained calculations performed on glycylglycine considering the geometry and the ADPs resulting from the IAM refinement.

$\chi^2 = [1/(N_r - N_p)] \sum_{\mathbf{h}} w_{\mathbf{h}}(\eta|F_{\mathbf{h},\mathrm{calc}}| - |F_{\mathbf{h},\mathrm{obs}}|)^2$, $\%R(F) = 100[\sum_{\mathbf{h}}|\eta|F_{\mathbf{h},\mathrm{calc}}| - |F_{\mathbf{h},\mathrm{obs}}||/\sum_{\mathbf{h}}|F_{\mathbf{h},\mathrm{obs}}|]$, $\%wR(F) = 100(\chi^2/\sum_{\mathbf{h}} w_{\mathbf{h}}F^2_{\mathbf{h},\mathrm{obs}})^{1/2}$ with $w_{\mathbf{h}} = 1/\sigma^2_{\mathbf{h},\mathrm{obs}}(F_{\mathbf{h},\mathrm{obs}})$.

Basis set	χ^2	$\%R(F)$	$\%wR(F)$	Energy (a.u.)	$\lambda_{\max}$	χ^2	$\%R(F)$	$\%wR(F)$	Energy (a.u.)
ELMO calculations†					XC-ELMO calculations				
6-31G	3.04	2.83	3.75	−489.158	0.42	1.45	2.29	2.59	−489.036
6-31G(d,p)	2.03	2.52	3.06	−489.384	0.18	0.98	2.04	2.14	−489.335
cc-pVDZ	2.01	2.51	3.05	−489.396	0.18	0.98	2.04	2.13	−489.350

† The scale factors η have been optimized using the density matrices obtained from the corresponding unconstrained calculations.

from all the electron-density models for glycylglycine are rather similar. Wavefunctions computed with the 6-31G basis set provide slightly larger electron densities than the 6-31G(d,p) and the cc-pVDZ ones. However, the effect of X-ray constraining is extremely small. This observation, along with the small dipole moment enhancements discussed in §6.3, suggests that the long-range crystal interactions do not polarize the electron density in the hydrogen-bond bond-critical-point regions. This is quite different from what happens in the electron lone pairs and intramolecular bonding regions, where the effects due to the crystal field or the X-ray constraining are more evident (see the previous section). Grimwood & Jayatilaka (2001) have briefly discussed the effects of long-range interactions on the electron density of oxalic acid dihydrate by means of a relatively large molecular cluster calculation. They concluded that, even when considering the electron distributions of intra-molecular bonds and lone pairs, the influence of the crystalline environment is not significant, at least when the long-range interactions are modelled at DFT level. However, in our opinion, much more work is needed before finding a more general understanding. The crystal environment effect on ELMO and XC-ELMO wavefunctions will be extensively analysed in a future publication.

Integrated atomic charges calculated from the Hansen & Coppens multipole-model-projected electron densities differ by 5–10% from those calculated from the respective primary densities, corresponding to an average absolute value of about 0.1 a.u. The largest differences are observed for the oxygen atoms at the 6-31G level. In this respect, the influence of the basis set on the integrated charges is much more pronounced than the influence of the multiple projection itself (see Fig. 7 and Table S6). For all the basis sets, the atomic charges obtained from X-ray constrained calculations are closer to the $\rho^{\mathrm{MM/XC}}$ ones.

The atomic dipole moments from the multipole-projected ELMO and XC-ELMO electron densities are in reasonable agreement with those from the respective primary densities using the 6-31G(d,p) and cc-pVDZ basis sets (differences in the 10–20% range), whereas larger discrepancies (up to 80%) are observed for the 6-31G basis set. As for the non-projected densities, the multipole-projected atomic dipole moments calculated for O atoms using the 6-31G basis set are markedly underestimated compared to the values determined using the larger basis sets or P-B3LYP. Again, the $\rho^{\mathrm{MM/XC\text{-}ELMO}}$ electron distributions generally give more accurate atomic dipole moment magnitudes.

Molecular dipole moment magnitudes obtained from the multipole-projected electron densities are shown in Fig. 9(*b*). Both unconstrained and constrained ELMO dipole moments are larger than the $\rho^{\mathrm{MM/XC}}$ and $\rho^{\mathrm{MM/P\text{-}B3LYP}}$ values. Compared to the values obtained from the primary densities (Fig. 9*a*), the multipole projection produces an underestimation of the molecular dipole moments of about 20%, corresponding to *circa* 1.5 a.u. The underestimation due to multipolar projection is even larger for P-B3LYP.

8. Influence of fractional coordinates and ADPs on the X-ray constrained calculations

Since the current version of the X-ray constrained ELMO strategy does not allow one to refine atomic positions and thermal parameters, all the XC-ELMO wavefunctions considered in the previous sections have been carried out using parameters refined from the multipole model MM2 (§5).

On the contrary, if XC-ELMO computations are performed using the IAM coordinates and ADPs, the statistical agreements and the energy values (Table 5) do not sensitively change, but the convergence towards $\chi^2 = 1$ is much slower, as it occurs for a larger $\lambda_{\max}$. On the other hand, since the desired agreement is anyway reached (at least for the more flexible basis sets), the small biases in the initial coordinates and ADPs have been artificially 'absorbed' into the wavefunction and this would affect the electron density and its topology: $\rho(\mathbf{r}_{\mathrm{b}})$ are not much affected (with changes smaller than 0.03 a.u.); the distances of the bond critical points from the nuclei differ by less than 5%; $\nabla^2\rho(\mathbf{r}_{\mathrm{b}})$ change more significantly (up to 15%, corresponding to about 0.05 a.u.). The same holds true for atomic dipole moment magnitudes, the largest differences being of the order of 3%, representing absolute differences around 0.04 a.u.

It is important to point out that the IAM was refined against a high-resolution data set and therefore non-hydrogen atomic positions and ADPs are already quite accurate, certainly more so than those of typical crystal structure determinations. Significantly different results are expected in the case of an IAM refined against structure factors up to a lower resolution, for which a multipolar model would also not be reliable.

electronic reprint

9. Conclusions and perspectives

We have demonstrated that XC-ELMO is a new and potentially useful tool for the determination and analysis of experimental electron densities. All the X-ray constrained wavefunction methods use the X-ray data in order to capture, at least in part, the effects due to the electron correlation and the crystal environment. XC-ELMO has the additional advantage of resuming the atomistic interpretation typical of the pseudoatom approaches, since the orbitals are one electron function accounting for the electron distributions of atoms, bonds or functional groups, depending on the localization scheme.

This work is part of a long project aimed at studying the efficiency of XC-ELMO wavefunctions in molecular crystals and testing the transferability of the ELMO to larger systems in order to devise new strategies for refining crystallographic structures and electron densities of macromolecules, such as proteins or polymers. In this first step, we have reported a detailed comparison between unconstrained and X-ray constrained ELMO wavefunctions, using traditional multipolar electron density or periodic density functional as benchmarks. The main conclusions can be summarized as follows:

(i) Sufficiently flexible basis sets are fundamental to obtaining a meaningful fitting of the wavefunction. In fact, all calculations confirm that the desired agreement with the experimental data is reached only if polarized basis functions are used.

(ii) The fractional coordinates and ADPs used for XC-ELMO have a strong influence on the convergence: the more accurate the initial parameters are, the faster the convergence is. The inaccuracy of the initial set of coordinates and ADPs is absorbed into the molecular orbital coefficients, thus affecting the electron density. A strategy for the direct refinement of atomic coordinates and ADPs in the framework of the XC-ELMO strategy is currently in preparation.

(iii) When $\chi^2 = 1$ is reached, the constrained ELMO wavefunction is of course in much better agreement with the X-ray data than the unconstrained one, but in less good agreement than a standard multipole model, for which there is not a strict 'control' on the desired precision. In fact, while XC-ELMO only partially uses (through λ) the information contained in the X-ray intensities, the multipole models MM2 and MM3 fully exploit the experimental observations through global least-square refinements of the available diffraction data, which leads to χ^2 values much lower than 1.0. Therefore, for a definitive and fair comparison with the traditional multipole models, we should push the X-ray constrained computations beyond the usual $\chi^2 = 1.0$ limit (namely, we should consider larger λ). Whitten *et al.* (2006) have proposed pursuing the fitting until the weighted residual wR values approach those obtained in the multipole refinement of the same X-ray structure-factor magnitudes. This idea could be reformulated using χ^2 as a criterion, but, unfortunately, this is partially hampered by the problem of determining the optimal value for the Lagrange multiplier λ (see final discussion in §3.2). However, also in this case, theoretical approaches to overcome this last important drawback are under investigation.

(iv) The multipolar models seem to be much more sensitive to the valence electron density than the XC-ELMO strategy. Fig. 2 shows that the two most flexible multipolar models reproduce the low-angle diffracted intensities much better than the high-angle ones, whereas this is not true for the XC-ELMO technique that seems more 'tempered'. This might be interpreted as an over-fitting of the low-angle data by the multipolar models, which, in fact, converge to lower χ^2, or otherwise as the evidence of a too restricted model that could be improved by a more flexible treatment of the core electrons, as suggested by Fischer *et al.* (2011). A model with strictly hybridized atoms, MM1, is instead closer to $\chi^2 = 1$, and shows a more uniform agreement with the observed structure factors, like the XC-ELMO calculations (Fig. 4). However, one would normally consider MM1 as too rigid (*i.e.*, insufficient to exhaust the present data quality) and many indicators address the more flexible MM2 (or the statistically equivalent MM3) as more reliable. Once more, this observation prompts a revision of the current recipes for the X-ray constrained wavefunction calculations that could be too biased by the quantum-mechanical part of the functional in equation (5) and not sufficiently influenced by the experimental data, the second part of the functional.

(v) Many properties of the XC-ELMO reconstructed electron density suggest that the constrained wavefunction approaches the 'exact' electron density in the crystal, using as a benchmark the experimental multipole model or the periodic calculations at density functional level. However, some atomic charges and molecular dipole moments are not properly reproduced. In particular the XC-ELMO calculations generally underestimate the molecular dipole moment and this might be ascribed to the inability of XC-ELMO to include *all* the effects of the crystalline environment, in particular the polarization of the molecule. Interestingly, the multipolar model also underestimates the dipole moment in this case.

(vi) For the intermolecular interactions, here represented by medium-strength N—H···O hydrogen bonds, there is limited perturbation by the crystal packing; therefore it is not possible to judge the efficiency of the X-ray constrained procedure.

We want to stress that, although X-ray constrained methods have been known and used for more than a decade, until now there have been no reports of such an accurate comparison with respect to traditional multipolar expansions refined against X-ray intensities or theoretical calculations with periodic boundary conditions. Therefore, we believe that our results could be important not only to appreciate the advantages of XC-ELMO, but more generally to understand the necessities and pitfalls of all kinds of X-ray constrained wavefunction calculations.

In view of these results, we plan to investigate further the XC-ELMO technique in order to better analyse the ability of the XC-ELMO wavefunctions to include electronic correlation and crystal-field effects into the electron density. This may

electronic reprint

enable us to establish better criteria for the best Lagrange multiplier, which is a crucial parameter in the X-ray constrained wavefunction strategies. Moreover, we will analyse the performances of the XC-ELMO wavefunctions on a broad spectrum of molecules, including metal complexes and stronger hydrogen-bond adducts.

LHRDS and PM thank the Swiss National Science Foundation (project SNF-141271) for financial support. AG thanks the CNRS and the University of Lorraine (PEPS Mirabelle project 'KiteComb') and the Institute Jean Barriol of the University of Lorraine (Axe-R12 project) for financial support.

References

Adams, W. H. (1961). *J. Chem. Phys.* **34**, 89.
Agilent Technologies (2013). *CrysAlisPro*. Version 1.171.36.28 (release 01-02-2013 CrysAlis171.NET). Yarnton, UK.
Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* B**66**, 380–386.
Bader, R. F. W. (1990). *Atoms in Molecules: a Quantum Theory*. Oxford University Press.
Bader, R. F. W., Larouche, A., Gatti, C., Carroll, M. T., MacDougall, P. J. & Wiberg, K. B. (1987). *J. Chem. Phys.* **87**, 1142.
Becker, P. J. & Coppens, P. (1974). *Acta Cryst.* A**30**, 129–147.
Biswas, A. B., Hughes, E. W., Sharma, B. D. & Wilson, J. N. (1968). *Acta Cryst.* B**24**, 40–50.
Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
Blessing, R. H. (1995). *Acta Cryst.* B**51**, 816–823.
Bytheway, I., Grimwood, D. J., Figgis, B. N., Chandler, G. S. & Jayatilaka, D. (2002). *Acta Cryst.* A**58**, 244–251.
Bytheway, I., Grimwood, D. J. & Jayatilaka, D. (2002). *Acta Cryst.* A**58**, 232–243.
Cassam-Chenaï, P. (1995). *Int. J. Quantum Chem.* **54**, 201–210.
Clark, R. C. & Reid, J. S. (1995). *Acta Cryst.* A**51**, 887–897.
Clementi, E. & Raimondi, D. L. (1963). *J. Chem. Phys.* **38**, 2686.
Clementi, E. & Roetti, C. (1974). *At. Nucl. Data Tables*, **14**, 177–478.
Clinton, W. L., Frishberg, C. A., Massa, L. J. & Oldfield, P. A. (1973). *Int. J. Quantum Chem. Symp.* **7**, 505–514.
Clinton, W. L., Galli, A. J., Henderson, G. A., Lamers, G. B., Massa, L. J. & Zarur, J. (1969). *Phys. Rev.* **177**, 27–33.
Clinton, W. L., Galli, A. J. & Massa, L. J. (1969). *Phys. Rev.* **177**, 7–13.
Clinton, W. L., Henderson, G. A. & Prestia, J. V. (1969). *Phys. Rev.* **177**, 13–18.
Clinton, W. L. & Lamers, G. B. (1969). *Phys. Rev.* **177**, 19–27.
Clinton, W. & Massa, L. (1972). *Phys. Rev. Lett.* **29**, 1363–1366.
Clinton, W. L., Nakhleh, J. & Wunderlich, F. (1969). *Phys. Rev.* **177**, 1–6.
Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. New York: Oxford University Press.
Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* A**35**, 63–72.
Couty, M., Bayse, C. A. & Hall, M. B. (1997). *Theor. Chem. Acc.* **97**, 96–109.
Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* D**62**, 1325–1335.
Dittrich, B., Koritsánszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed. Engl.* **43**, 2718–2721.
Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
Dovesi, R., Saunders, V. R., Roetti, C., Orlando, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, P. & Llunell, M. (2009). *CRYSTAL09 User's Manual*. Torino: University of Torino.
Espinosa, E., Souhassou, M., Lachekar, H. & Lecomte, C. (1999). *Acta Cryst.* B**55**, 563–572.
Farrugia, L. J. (1999). *J. Appl. Cryst.* **32**, 837–838.
Fischer, A., Tiana, D., Scherer, W., Batke, K., Eickerling, G., Svendsen, H., Bindzus, N. & Iversen, B. B. (2011). *J. Phys. Chem. A*, **115**, 13061–13071.
Fornili, A., Sironi, M. & Raimondi, M. (2003). *J. Mol. Struct. (THEOCHEM)*, **632**, 157–172.
Francisco, E., Martín Pendás, A. & Adams, W. H. (1992). *J. Chem. Phys.* **97**, 6504.
Frishberg, C. & Massa, L. (1981). *Phys. Rev. B*, **24**, 7018–7024.
Gatti, C. (1998). *TOPOND*. CNR-CSRSRC, Milan, Italy.
Gatti, C. & Macchi, P. (2012). Editors. *Modern Charge-Density Analysis*. New York: Springer.
Genoni, A. (2013*a*). *J. Phys. Chem. Lett.* **4**, 1093–1099.
Genoni, A. (2013*b*). *J. Chem. Theory Comput.* **9**, 3004–3019.
Genoni, A., Ghitti, M., Pieraccini, S. & Sironi, M. (2005). *Chem. Phys. Lett.* **415**, 256–260.
Gilbert, T. L. (1974). *J. Chem. Phys.* **60**, 3835.
Gillet, J.-M. (2007). *Acta Cryst.* A**63**, 234–238.
Gillet, J.-M., Becker, P. & Cortona, P. (2001). *Phys. Rev. B*, **63**, 235115.
Grimwood, D. J., Bytheway, I. & Jayatilaka, D. (2003). *J. Comput. Chem.* **24**, 470–483.
Grimwood, D. J. & Jayatilaka, D. (2001). *Acta Cryst.* A**57**, 87–100.
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. (2005). *Mol. Phys.* **103**, 719–747.
Hamilton, W. C. (1965). *Acta Cryst.* **18**, 502–510.
Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.
Head-Gordon, M. & Pople, J. A. (1988). *J. Chem. Phys.* **89**, 5777.
Hibbs, D. E., Howard, S. T., Huke, J. P. & Waller, M. P. (2005). *Phys. Chem. Chem. Phys.* **7**, 1772–1778.
Hirshfeld, F. L. (1971). *Acta Cryst.* B**27**, 769–781.
Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.
Hohenberg, P. & Kohn, W. (1964). *Phys. Rev. B*, **136**, 864–871.
Howard, S., Huke, J., Mallinson, P. & Frampton, C. (1994). *Phys. Rev. B*, **49**, 7124–7136.
Hudák, M., Jayatilaka, D., Perašínová, L., Biskupič, S., Kožíšek, J. & Bučinský, L. (2010). *Acta Cryst.* A**66**, 78–92.
Huzinaga, S. & Cantu, A. A. (1971). *J. Chem. Phys.* **55**, 5543–5549.
Jayatilaka, D. (1998). *Phys. Rev. Lett.* **80**, 798–801.
Jayatilaka, D. (2012). *Modern Charge-Density Analysis*, edited by C. Gatti & P. Macchi, pp. 213–257. New York: Springer.
Jayatilaka, D. & Grimwood, D. J. (2001). *Acta Cryst.* A**57**, 76–86.
Jayatilaka, D. & Grimwood, D. J. (2003). *Computational Science – ICCS 2003*, pp. 142–151. *Lecture Notes in Computer Science*, Vol. 2660. Berlin: Springer.
Jayatilaka, D., Munshi, P., Turner, M. J., Howard, J. A. & Spackman, M. A. (2009). *Phys. Chem. Chem. Phys.* **11**, 7209–7218.
Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.
Keith, T. A. (2013). *AIMAll*. Version 13.11.04. Overland Park KS, USA: TK Gristmill Software (http://aim.tkgristmill.com/).
Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* A**58**, 464–472.
Kvick, Å., Al-Karaghouli, A. R. & Koetzle, T. F. (1977). *Acta Cryst.* B**33**, 3796–3801.
Kvick, A., Koetzle, T. F. & Stevens, E. D. (1979). *J. Chem. Phys.* **71**, 173.
Laidig, K. E. & Bader, R. F. W. (1990). *J. Chem. Phys.* **93**, 7213.
Macchi, P., Bürgi, H.-B., Chimpri, A. S., Hauser, J. & Gál, Z. (2011). *J. Appl. Cryst.* **44**, 763–771.
Macchi, P., Proserpio, D. M. & Sironi, A. (1998). *J. Am. Chem. Soc.* **120**, 13429–13435.
Madsen, A. Ø. (2006). *J. Appl. Cryst.* **39**, 757–758.
Massa, L., Goldberg, M., Frishberg, C., Boehme, R. F. & La Placa, S. J. (1985). *Phys. Rev. Lett.* **55**, 622–625.
Matsuoka, O. (1977). *J. Chem. Phys.* **66**, 1245.

electronic reprint

McWeeny, R. (1959). *Proc. R. Soc. London Ser. A*, **253**, 242–259.
McWeeny, R. (1960). *Rev. Mod. Phys.* **32**, 335–369.
McWeeny, R. (1992). *Methods of Molecular Quantum Mechanics*. London: Academic Press.
Obara, S. & Saika, A. (1986). *J. Chem. Phys.* **84**, 3963.
Obara, S. & Saika, A. (1988). *J. Chem. Phys.* **89**, 1540.
Ordejón, P., Drabold, D., Grumbach, M. & Martin, R. (1993). *Phys. Rev. B*, **48**, 14646–14649.
Paturle, A. & Coppens, P. (1988). *Acta Cryst.* A**44**, 6–8.
Pichon-Pesme, V., Jelsch, C., Guillot, B. & Lecomte, C. (2004). *Acta Cryst.* A**60**, 204–208.
Pichon-Pesme, V., Lecomte, C. & Lachekar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.
Popelier, P. (1998). *Comput. Phys. Commun.* **108**, 180–190.
Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in Fortran 77: the Art of Scientific Computing*, 2nd ed. New York: Cambridge University Press.
Roversi, P., Irwin, J. J. & Bricogne, G. (1998). *Acta Cryst.* A**54**, 971–996.
Rozas, I., Alkorta, I. & Elguero, J. (2000). *J. Am. Chem. Soc.* **122**, 11154–11161.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sironi, M., Ghitti, M., Genoni, A., Saladino, G. & Pieraccini, S. (2009). *J. Mol. Struct. (THEOCHEM)*, **898**, 8–16.
Smits, G. F. & Altona, C. (1985). *Theor. Chim. Acta*, **67**, 461–475.
Snyder, J. A. & Stevens, E. D. (1999). *Chem. Phys. Lett.* **313**, 293–298.
Spackman, M. A. (1999). *Chem. Phys. Lett.* **301**, 425–429.
Spackman, M. A., Munshi, P. & Jayatilaka, D. (2007). *Chem. Phys. Lett.* **443**, 87–91.
Stewart, R. F. (1969). *J. Chem. Phys.* **51**, 4569.
Stewart, R. F. (1972). *J. Chem. Phys.* **57**, 1664.
Stoll, H., Wagenblast, G. & Preuss, H. (1980). *Theor. Chim. Acta*, **57**, 169–178.
Szekeres, Z. & Surján, P. R. (2003). *Chem. Phys. Lett.* **369**, 125–130.
Tanaka, K. (1988). *Acta Cryst.* A**44**, 1002–1008.
Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.
Volkov, A., Abramov, Y., Coppens, P. & Gatti, C. (2000). *Acta Cryst.* A**56**, 332–339.
Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.
Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. & Koritsanszky, T. (2006). *XD2006 – A Computer Program Package for Multipole Refinement, Topological Analysis of Charge Densities and Evaluation of Intermolecular Interaction Energies from Experimental and Theoretical Structure Factors*. University at Buffalo, State University of New York, NY, USA; University of Milano, Italy; University of Glasgow, UK; CNRISTM, Milano, Italy; Middle Tennessee State University, TN USA.
Whitten, A. E., Jayatilaka, D. & Spackman, M. A. (2006). *J. Chem. Phys.* **125**, 174505.
Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* A**63**, 108–125.

electronic reprint