



## A theoretical-electron-density databank using a model of real and virtual spherical atoms

**Ayoub Nassour, Slawomir Domagala, Benoit Guillot, Theo Leduc, Claude Lecomte and Christian Jelsch**

*Acta Cryst.* (2017). **B73**, 610–625



**IUCr Journals**

CRYSTALLOGRAPHY JOURNALS ONLINE

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>

# A theoretical-electron-density databank using a model of real and virtual spherical atoms

Ayoub Nassour,<sup>a</sup> Slawomir Domagala,<sup>b</sup> Benoit Guillot,<sup>a</sup> Theo Leduc,<sup>a</sup> Claude Lecomte<sup>a</sup> and Christian Jelsch<sup>a\*</sup>

<sup>a</sup>Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2), CNRS, UMR 7036, Institut Jean Barriol, Faculté des Sciences et Technologies, Université de Lorraine, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France, and <sup>b</sup>Department of Chemistry, University of Warsaw, ul. Pasteura 1, 02-093 Warsaw, Poland. \*Correspondence e-mail: christian.jelsch@univ-lorraine.fr

Received 28 February 2017

Accepted 2 June 2017

Edited by J. Benedict, University of Washington, USA

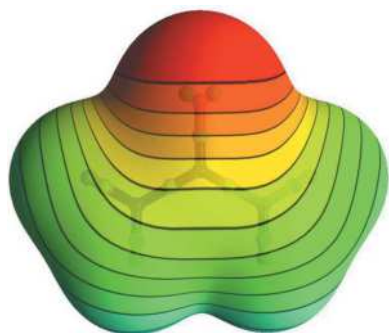
**Keywords:** bonding density; electron lone pairs; electrostatic potential; charges.

**Supporting information:** this article has supporting information at journals.iucr.org/b

A database describing the electron density of common chemical groups using combinations of real and virtual spherical atoms is proposed, as an alternative to the multipolar atom modelling of the molecular charge density. Theoretical structure factors were computed from periodic density functional theory calculations on 38 crystal structures of small molecules and the charge density was subsequently refined using a density model based on real spherical atoms and additional dummy charges on the covalent bonds and on electron lone-pair sites. The electron-density parameters of real and dummy atoms present in a similar chemical environment were averaged on all the molecules studied to build a database of transferable spherical atoms. Compared with the now-popular databases of transferable multipolar parameters, the spherical charge modelling needs fewer parameters to describe the molecular electron density and can be more easily incorporated in molecular modelling software for the computation of electrostatic properties. The construction method of the database is described. In order to analyse to what extent this modelling method can be used to derive meaningful molecular properties, it has been applied to the urea molecule and to biotin/streptavidin, a protein/ligand complex.

## 1. Introduction

The coupling of the multipolar atom model (Hansen & Coppens, 1978; Koritsanszky & Coppens, 2001) of the charge-density distribution and high-resolution data leads to more accurate molecular geometries and atomic anisotropic displacement parameters. Although this model is now commonly used for small molecules, its applications to macromolecules are reduced by several factors: (i) macromolecular X-ray data have limited resolution, (ii) the number of reflections available is generally not sufficient for an aspherical-atom refinement, (iii) the quality of X-ray data for macromolecules is generally lower than for small molecules. To overcome these limitations, Brock *et al.* (1991) proposed the new idea of transferability of deformation-electron-density parameters between different molecules as an alternative to a full multipolar refinement. It is based on the hypothesis that the atomic deformation electron density due to covalent chemical bonding in a given chemical group is similar in different molecules and crystal structures. The electron-density parameters can be transferred to a crystal structure without any multipolar refinement, so offering the opportunity of modelling the electron density more precisely than using the independent-atom model (IAM). Therefore, several libraries of transferable multipolar parameters were developed: the experimental ELMAM (Pichon-Pesme *et al.*,



1995; Zarychta *et al.*, 2007) and ELMAM2 (Domagała & Jelsch, 2008; Domagała *et al.*, 2011, 2012) libraries, the theoretical Invariom database (Dittrich *et al.*, 2006, 2013) and the theoretical University at Buffalo Pseudoatom Databank (UBDB; Koritsanszky *et al.*, 2002; Volkov, Koritsanszky & Coppens, 2004; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012).

The multipole parameters stored in each of the databases can be transferred to compatible – containing recognizable atom types – crystal structures obtained at atomic or lower atomic resolution. A number of applications of these databanks to the refinement of X-ray crystal structures and the modelling of their electrostatic properties have been reported (Pichon-Pesme *et al.*, 1995; Volkov, Koritsanszky & Coppens, 2004; Dittrich *et al.*, 2006; Zarychta *et al.*, 2007; Bąk *et al.*, 2011). The results of these studies indicate that density modelling beyond the spherical-atom formalism can lead to significantly improved thermal atomic displacement parameters and molecular geometry, and allows the computation of reliable electrostatic properties of peptides and biological molecules.

Database transfer to biomacromolecules such as proteins and nucleic acids allows the analysis of their electrostatic potential and of the accurate electrostatic interaction energy with ligands (Dominiak *et al.*, 2007; Fournier *et al.*, 2009; Kumar & Dominiak, 2016) and the topological analysis of hydrogen bonds (Zarychta *et al.*, 2015; Liebschner *et al.*, 2011; Held & van Smaalen, 2014).

In the present work, another method of aspherical electron-density modelling, and consequently of building such a transferable parameter databank, is presented. The electron-density model is based on a collection of atomic and virtual charges, located respectively on nuclei positions and on the positions of expected deformation-electron-density maxima, *i.e.* on covalent bonds and at electron lone-pair sites. This approach allows modelling of the electron density as an alternative to the classical Hansen & Coppens (1978) multipolar atom model.

Few analogue modelling applications were already reported in the 1970s literature: cyanuric acid (Scheringer, Kutoglu, Hellner *et al.*, 1978; Dietrich *et al.*, 1979), urea (Scheringer, Kutoglu, Mullen & Hellner, 1978; Mullen & Hellner, 1978), thiourea (Scheringer, Kutoglu, Mullen & Hellner, 1978), diborane (Mullen & Hellner, 1977), decaborane (Dietrich & Scheringer, 1978) and silicon (Scheringer, 1980). More recently, Afonine *et al.* (2004, 2007) reintroduced the approach using ‘interatomic scatterers’ and applied it to macromolecular refinement at subatomic resolution. The charge-density analysis of 2-methyl-1,3-cyclopentanedione, a compound displaying a resonance-assisted hydrogen bond, as well as the 1,4-bis(5-hexyl-2-thienyl)butane-1,4-dione molecule (Ahmed *et al.*, 2016) were determined with the spherical charge model from synchrotron diffraction data (Nassour *et al.*, 2014).

Following this principle, we have built a database of real and virtual spherical-atom parameters (VIR) for molecules containing C, H, N, O, S, P, Cl and F atoms. This database is

therefore applicable to the construction of the electron density of proteins, peptides, nucleic acids and small organic compounds. Additional scattering centres, further called virtual atoms, were introduced in the molecular structures. The parameters describing both the real and virtual atoms were refined against theoretical structure factors, obtained from *ab initio* calculations with the density functional theory method at the B3LYP/6-31G(d,p) level of theory on crystal structures of the small molecules. Real atoms presenting the same connectivity and covalent bonding were considered to be chemically equivalent and their parameters were averaged to build the library. The same rules were applied to the associated virtual atoms located on covalent bonds and on electron lone-pair sites.

The validity of the resulting molecular-electron-density model is discussed in terms of deformation electron density and electrostatic properties (dipole moment, potential and interaction energies in molecular dimers) of the urea crystal structure. Moreover, to assess the usability of the VIR library parameters in the study of macromolecular system properties, electrostatic potential computation and protein–ligand hydrogen-bond topological analysis (Bader, 1990) were conducted on a subatomic resolution biotin/streptavidin complex (PDB code: 2F01). Results obtained using VIR library parameters were then compared to those issued from the multipolar ELMAM2 and UBDB electron-density distributions.

## 2. Methods

### 2.1. The electron-density model

The multipolar atom model, developed by Hansen & Coppens (1978), is now the most widely used in charge-density analysis. The model describes the molecular electron density as a sum of pseudo-atomic densities composed of a spherical part and a ‘multipolar’ part. In this model, the electron density of one atom is described as

$$\rho_{\text{atom}}(\mathbf{r}) = \rho_{\text{core}}(r) + P_{\text{val}}\kappa^3\rho_{\text{val}}(\kappa r) + \sum_{l=0}^{l_{\text{max}}} \kappa_l^3 R_{nl}(\kappa' r) \sum_{m=0}^l P_{lm\pm} y_{lm\pm}(\theta, \varphi), \quad (1)$$

where  $\rho_{\text{core}}$  and  $\rho_{\text{val}}$  are spherically averaged Hartree–Fock core and valence electron densities, respectively. The third multipolar term contains a sum of real spherical harmonic functions  $y_{lm\pm}(\theta, \varphi)$ , modulated by radial Slater functions  $R_{nl}$  to take into account aspherical deformations. The coefficient  $P_{\text{val}}$  is the atomic valence population and the  $P_{lm\pm}$  parameters are multipole populations for the aspherical valence electron density.  $\kappa$  and  $\kappa'$  are scaling parameters, which determine the expansion/contraction of the spherical and multipolar valence electron densities, respectively. The atomic electron density obtained from the multipole model deviates from the spherical model by an accumulation of electrons on the covalent bonds and on the electron lone-pair positions.

Based on these considerations, an alternative empirical model was developed in order to reproduce results of quality

comparable to those of the ‘multipolar’ model. It is based on the principle of adding extra spherical charges on the positions of the deformation–electron–density peaks. Hence, the molecular electron density is considered as a superposition of real and virtual spherical atoms,

$$\rho(\mathbf{r}) = \sum_{\text{atom}} (\rho_{\text{core}}(r) + P_{\text{val}}\kappa^3\rho_{\text{val}}(\kappa r)) + \sum_{\text{vir}} P_{\text{vir}}\kappa_{\text{vir}}^3\rho_{\text{vir}}(\kappa_{\text{vir}}r). \quad (2)$$

The real atoms (H, C, O, N, Cl, S, P and F in the current database) are treated spherically and are described as the first and second terms of equation (1), which are similar to those of the Hansen–Coppens equation. The third term now corresponds to the electron density  $\rho_{\text{vir}}$  generated by the virtual spherical atoms placed on the covalent bonds and on the lone-pair sites. The spherical deformation electron density is then modelled by the  $P_{\text{val}}/\kappa$  parameters, while the aspherical deformation features are taken into account by the parameters associated with the VIR model: a spherical valence population  $P_{\text{vir}}$  and an expansion/contraction coefficient  $\kappa_{\text{vir}}$ .

## 2.2. Fitting Slater-type orbital functions for virtual atoms

Analytical expressions for atomic wavefunctions by Clementi & Roetti (1974) have been used extensively in X-ray charge-density refinements applied to real atoms in the evaluation of electrostatic and topological properties of the charge density. In the present work, the same type of radial wavefunction was adopted for the VIR atom, expressed as a linear combination of three Slater-type functions:

$$\Phi(r) = \sum_{j=1,3} C_j N_j r^{n_j} \exp(-\zeta_j r), \quad (3)$$

where  $N_j = (2\zeta_j)^{n_j+1/2} / \sqrt{(2n_j)!}$  is the Slater-type orbital normalization constant and  $n_j = n - 1$  where  $n$  is the principal quantum number of the basis function. Here,  $n_j = 0, 1$  and  $2$  for  $j = 1, 2$  and  $3$ , respectively.  $C_j$  and  $\zeta_j$  are the adjustable parameters.

The spherical electron density is then obtained as the square of the orbital function,

$$\rho(\mathbf{r}) = \Phi^2(\mathbf{r}). \quad (4)$$

The wavefunction parameters of the VIR atoms were optimized by minimizing a cost function using the Marquardt–Levenberg algorithm built into the *GNUPLLOT4.0* software (Racine, 2006),

$$F(\gamma) = \sum_i |\rho_{\text{vir}}(r_i) - \rho_{\text{sph}}^{\text{res}}(r_i)|^2, \quad (5)$$

where  $F(\gamma)$  represents the deviations between the corresponding data computed from Slater-type functions and the reference Fourier residual data.  $\rho_{\text{vir}}(r)$  is the virtual electron density determined with the parameterization  $\gamma(C_j, \zeta_j, j = 1, 3)$ . The residual Fourier electron density was calculated on the  $C_\alpha$ – $C_\beta$  bond of a centrosymmetric crystal, DL-histidine (Coppens *et al.*, 1999; Dadda *et al.*, 2012) from a spherical-atom (IAM) modelling against theoretical structure-factor amplitudes, which were computed from first-principles calcu-

**Table 1**

Coefficients of the Slater radial functions modelling the bond virtual atom electron density.

The value of the normalization factor  $N_j$  is given in equation (7). The wavefunction  $\varphi$  has units bohr<sup>-3/2</sup>,  $\zeta_j$  is in bohr<sup>-1</sup>, while  $N_j$  is in bohr<sup>-(n<sub>j</sub>+1/2)</sup>.

$n_j$	$\zeta_j$	$c_j$
1	6.2644	0.0736
2	6.0620	0.1395
3	3.7385	0.8768

lations. The bonding electron density does not contribute to the diffraction at ultra-high resolution; therefore, the Fourier synthesis was truncated at 0.7 Å resolution. As the residual electron density showed a parabolic shape (see Fig. 1) around its maximum, the  $\rho_{\text{vir}}(r)$  function was imposed to have a zero derivative at  $r = 0$ .

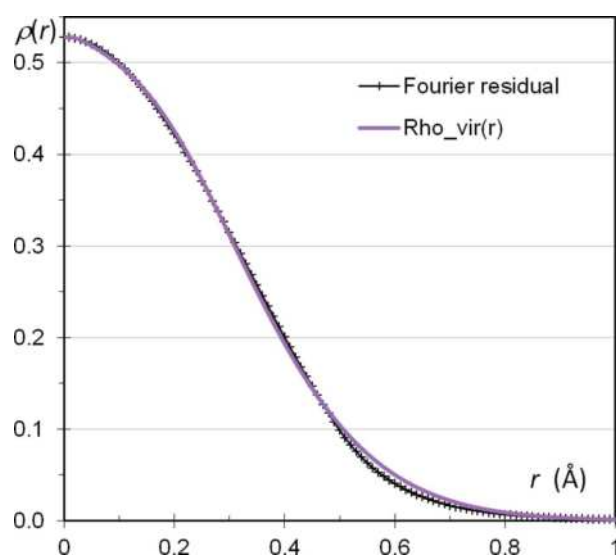
The converged wavefunctions that give the minimum need to be normalized. The global dimensionless normalization factor  $N$  is obtained using the relationship

$$N = \int_0^\infty r^2 \left\{ \sum_j N_j C_j r^{n_j-1} \exp(-\zeta_j r) \right\}^2 dr \quad (6)$$

with

$$N_j = (2\zeta_j)^{n_j+1/2} / \sqrt{(2n_j)!}. \quad (7)$$

The factor  $N^{-1/2}$  is applied to scale the coefficients  $C_j$  so that the final expansion of  $\rho(r)$  is perfectly normalized. A linear combination of three Slater-type functions was deemed necessary to correctly model the residual electron density. The coefficients and orbital exponents for the VIR atom are listed in Table 1. The resulting virtual-atom electron density is



**Figure 1**  
Comparison between the VIR electron density and the IAM Fourier residual density ( $e \text{ \AA}^{-3}$ ) at the midpoint of the  $C_\alpha$ – $C_\beta$  bond of DL-histidine. Black curve: Fourier residual electron density  $\Delta\rho(r)$ ; purple curve: three Slater functions model fitted to  $\Delta\rho(r)$  in VIR formalism.

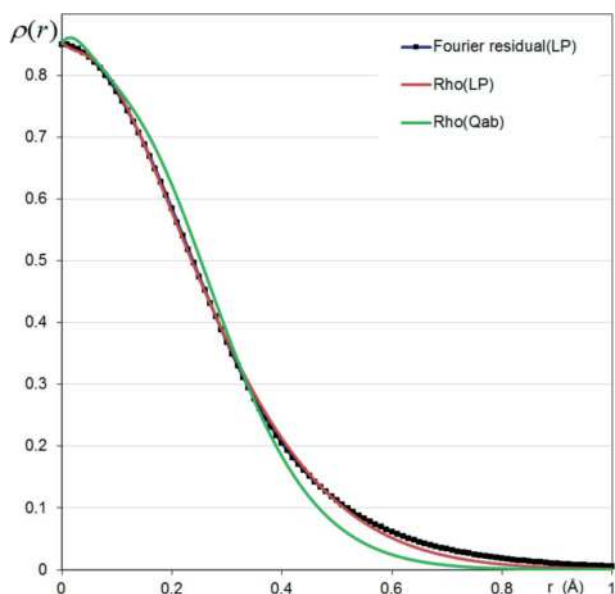
shown in Fig. 1. Clearly, the fitted radial curve  $\rho_{\text{vir}}(r)$  is in good agreement with the residual density on the covalent bond.

A similar method was used to parameterize the Slater-type function which corresponds to the VIR electron lone pairs. In this case, the residual deformation density was calculated on the lone pair of the imidazole non-protonated N atom. The radial function was evaluated along a direction lying on the imidazole plane and perpendicular to the direction joining the lone-pair maximum to the N atom. The resulting electronic density of the bond and lone-pair virtual atoms are compared in Fig. 2. The electron-density curve of the lone pair is similar to that of a contracted bonding density using a fitted  $\kappa = 1.197$  coefficient. The virtual atoms designed to model the electron lone pairs were then assigned the same wavefunction description as the covalent-bond virtual atoms but, as they are more contracted in space, refined to larger  $\kappa$  values.

The great advantage of the present fitting procedure lies in its simplicity. There is no need to calculate the wavefunction self-consistently, as in the Hartree–Fock or density functional theory (DFT) method and it is unnecessary to solve a generalized eigenvalue problem as was done for real atoms (Clementi & Roetti, 1974). The  $Q_{AB}$  virtual atoms refer, in this paper, to the VIR modelling of the deformation density between the covalently bound atoms  $A$  and  $B$ . The  $Q_{LP}$  virtual atoms take into account the density located on an expected lone-pair site.

### 2.3. Electron-density-map computation

The static deformation electron density is defined for the multipolar atom model (Coppens, 1997) as the difference



**Figure 2**  
Comparison between the bond and lone-pair electron density ( $\text{e } \text{\AA}^{-3}$ ) on the N $\delta$ 1 of DL-histidine. The bonding electron density virtual-atom model was fitted to the lone-pair electron density, resulting in a fitted  $\kappa$  value of 1.197. Blue: Fourier residual electron density  $\Delta\rho(r)$  on the lone pair; red: three Slater function model fitted to  $\Delta\rho(r)$  on the lone pair; green: bond virtual-atom model contracted ( $\kappa_{\text{vir}} = 1.197$ ) to fit  $\Delta\rho(r)$  on the lone pair.

between the molecular electron density and the superposition of spherical independent atoms. For the virtual spherical atom modelling, it is obtained from the sum

$$\Delta\rho(\mathbf{r}) = \sum_{\text{atoms}} [P_{\text{val}}\kappa^3\rho_{\text{val}}(\kappa r) - N_{\text{val}}\rho_{\text{val}}(r)] + \sum_{\text{vir}} P_{\text{vir}}\kappa_{\text{vir}}^3\rho_{\text{vir}}(\kappa_{\text{vir}}r), \quad (8)$$

where  $N_{\text{val}}$  is the valence population of the neutral atoms.

### 2.4. Theoretical structure factors

To construct the databank, periodic B3LYP density functional theory (DFT) calculations were performed using the *CRYSTAL09* program (Dovesi *et al.*, 2005, 2013). A set of 38 published crystal structures (Table S3 of the supporting information) obtained from X-ray diffraction experiments were used as the starting geometry. Geometry optimization for H atoms has been carried out *via* DFT (Hohenberg & Kohn, 1964) in conjunction with the B3LYP hybrid functional (Lee *et al.*, 1988; Becke, 1993) using the 6-31G(d,p) basis set (Hariharan & Pople, 1973). The level of accuracy in evaluating the Coulomb and exchange integrals is controlled by five parameters, for which values of (ITOL $i = 6, i = 1, 4$ ) and ITOL5 = 17 were used. The shrinking factor of the reciprocal space was set to 4, corresponding to 30  $k$  points in the irreducible Brillouin zone at which the Hamiltonian matrix was diagonalized. Upon energy convergence ( $10^{-8}$  Hartree), the periodic wavefunction based on the optimized geometry was obtained. The coordinates of the H atoms were relaxed, but the unit-cell parameters were kept fixed. The structure factors were then computed by analytical Fourier transform of the resulting electron density up to a resolution of  $1.25 \text{ \AA}^{-1}$ .

### 2.5. Restraints and constraints

The refinement of the crystal structure using VIR modelling against the theoretical structure-factor data alone can be unstable. To obtain a reliable model, it is for example necessary to constrain or restrain the bond virtual atoms  $Q_{AB}$  to stay on the  $A-B$  covalent bond, notably for  $X-H$  bonds and weak bonding electron densities on *e.g.* C–O and C–S simple bonds. Therefore, the set of restraints and constraints was designed and incorporated into the *MoPro* program to stabilize the refinements (for the whole set of constraints and restraints used, see Table 2). While constraints simply reduce the number of variable parameters, restraints can be considered as additional data. The used stereochemical restraints and constraints, implemented in the *MoPro* software (Jelsch *et al.*, 2005), concern notably interatomic distances, angles, linearity of  $A-Q_{AB}-B$  segments and planarity. Distance and angle similarity restraints are also available. The restraints use quadratic functions which are defined as the squared difference between the actual and ideal value of restrained parameters. The weight of the restraint is taken as the square inverse of the corresponding restraint sigma (*i.e.* tolerance) value.

The positions of the virtual atoms had to be stabilized by stereochemical restraints or constraints. The bond virtual



Table 2

Lists of restraints and constraints applied in the charge-density refinements using the VIR model. The first column refers to the *MoPro* keywords.

Type	Description	Sigma
Restraints		
DISTAN	Distance H– $Q_{HX}$ between atoms	0.01 Å
PLANAR	Planarity (dimensionless), electron lone pairs for $sp^2$ -hybridized atoms	0.001
SIMDIS	Similarity of $X-Q_{LP}$ distances for two or more atom pairs	0.01 Å
SIMANG	Similarity of $X-Y-Q_{LP}$ angles for two or more atomic triplets	0.1°
LINEAR	Colinearity of three atoms ( $A-Q_{AB}-B$ )	0.001 Å
KAPPA1	$\kappa \simeq 1.16$ for H atoms	0.01
Constraints		
CONKAP	Atoms have the same $\kappa$ ( $\kappa_{vir}$ ) values	–
CONVAL	Atoms have same valence populations $P_{val}$ and $P_{vir}$	–
COLINR	Colinearity of 3 atoms ( $A-Q_{AB}-B$ )	–

atoms were refined and prevented to deviate from the  $A-B$  bond segment by using linearity constraints for  $A-Q_{AB}-B$  triplets. However, linearity constraints were not applied to epoxy groups where the bonding density is known to lie slightly outside for the C–C bond (Grabowsky *et al.*, 2008).

The refinement of virtual atoms on the covalent bonds between two non-H atoms resulted in deformation density peaks located in the same region as for the multipolar atom model. However, the refinement of  $Q_{HX}$  virtual atoms yielded generally bonding deformation density peaks which were not in accordance with the multipolar atom models or the IAM Fourier residual maps. Therefore, the H– $Q_{HX}$  distances were restrained to the typical value observed between the H-atom nuclei and the maximum of the bonding density peak visible in the IAM Fourier residual maps, as determined in Dadda *et al.* (2012). This way, the H– $Q_{HX}$  distance target values were set to 0.27, 0.34, 0.20 and 0.37 Å for H–O, H–N, H–S and H–C virtual atoms, respectively. The position of oxygen, nitrogen and sulfur electron lone-pair virtual atoms were stabilized by using distance restraints with target values as observed in IAM Fourier residual-electron-density maps [ $d(O-Q_{LP}) = 0.28$  Å,  $d(N-Q_{LP}) = 0.30$  Å and  $d(S-Q_{LP}) = 0.60$  Å].

Distance similarity restraints (SIMDIS) were also applied when several  $d(X, Q_{LP})$ ,  $X = O, S$  and  $N$  of the same type were present in the molecule. In addition, when two lone pairs were present for an atom ( $X = O$  or  $S$ ), their positions were restrained by similarity to the  $A-X-Q_{LP}$  angles (SIMANG), where  $A$  are atoms bonded to the  $X$  atom. The similarity restraints avoid introducing target values in addition to the diffraction data set. Planarity restraints (PLANAR) were also applied to the lone pairs on  $sp^2$  or aromatic atoms, e.g.  $>C=O$  carbonyl or a nitrogen belonging to a planar cycle.

The kappa parameters of the H atoms were restrained to a value of 1.16 (1) (Stewart, 1976) during the refinements. To reduce the number of the least-squares variables and improve the convergence of the refinements, chemical-equivalence constraints were applied to the virtual and H atoms. The

chemical equivalences are detected in an automatic procedure based on the chemical connectivity and geometry (Domagała *et al.*, 2011). In the case of the VIR modelling, the equivalence constraint applies to two types of parameters only: the valence populations  $P_{val}$  and  $P_{vir}$  and the expansion/contraction coefficients  $\kappa$  and  $\kappa_{vir}$  (Table 2). H atoms belonging to the same  $CH_2$ ,  $CH_3$  or  $NH_3$  groups were constrained to have identical valence populations and  $\kappa$  parameters. The same equivalence constraints were applied to the virtual atoms belonging to these groups. For instance, in the water molecules HOH, the two H atoms, the two  $Q_{OH}$  and the two  $Q_{LP}$  virtual atoms were each constrained to be identical. The two external electron lone pairs of O atoms on carboxylate, nitro or phosphate  $H_2PO_4^-$  groups were constrained to be equivalent. On the other side, the two internal lone pairs were also considered equivalent.

## 2.6. Real and virtual spherical atoms refinement

The function minimized in the crystallographic least-squares refinement is a summation over the reflections and over all applied restraints components [equation (9)]. The coordinates of real atoms were kept fixed to the position obtained from the DFT calculations while, for virtual atoms, the coordinates were allowed to refine, accounting for the above described restraints. To determine the optimal set of parameters  $\beta$  ( $P_{val}$ ,  $\kappa$ ,  $P_{vir}$ ,  $\kappa_{vir}$ ,  $xyz_{vir}$ ), we tried to match the structure factors supplied by first-principles calculations with those calculated during refinement. The model parameters are optimized by minimizing the cost function

$$\chi^2(\beta) = \sum_{hkl} W_{hkl} (F_{calc}(\beta) - F_{ref})^2 + \sum_j W_j (R_j(\beta) - R_{ref})^2. \quad (9)$$

The first term represents the contributions of the structure-factor amplitudes,  $F(\beta)$  is the structure factor obtained from the parameterization  $\beta$  ( $P_{val}$ ,  $\kappa$ ,  $P_{vir}$ ,  $\kappa_{vir}$ ,  $xyz_{vir}$ ) and  $F_{ref}$  is the reference structure factor derived from *ab initio* calculations. The second term is made of the additional restraints;  $R_j(\beta)$  is calculated from the molecular parameterization and  $R_{ref}$  is the target value. The weight of restraints is  $W_j = 1/\sigma^2$ , where  $\sigma$  is the allowed standard deviation of the refined parameter with respect to the target value.

The least-squares refinement of each molecule against theoretical structure factor was performed using all reflections up to  $s = 1.25 \text{ \AA}^{-1}$ . The refinement strategy was as follows:

(i) Initial positions for all virtual atoms as well as the concomitant restraints and constraints can be automatically generated within the *MoPro* program. The bond virtual atoms were initially placed in the middle of the bonds. The virtual atoms modelling the S- and O-atom electron lone pairs were initially placed in a trigonal geometry ( $Q_{LP1}-O-Q_{LP2} = 120^\circ$ ) for  $sp^2$  atoms (carbonyl  $C = O$ , nitro  $NO_2$ ) and in a tetrahedral geometry for the COH, CSH, CSC, POH and HOH  $sp^3$  groups. The nitrogen lone-pair virtual atoms were positioned in the aromatic ring plane in the bisecting direction

of the  $X-N-Y$  triplet. All the initial distances between the  $Q_{LP}$  and the carrier atom were the same as the restraint target.

(ii) The atomic displacement parameters were all fixed to zero and the scale factor was set to unity.

(iii) The initial valence populations  $P_{vir}$  of all virtual atoms were set to zero. The valence populations of real atoms  $P_{val}$  and of virtual atoms  $P_{vir}$  had to be the first parameter to be refined, as the initial virtual atoms with  $P_{vir} = 0$  had no contribution to the diffraction.

(iv) Each type of parameter ( $P_{val} + P_{vir}, \kappa + \kappa_{vir}, xyz_{vir}$ ) was refined successively. Valence electron populations and expansion/contraction parameters were refined for all atoms. The positions  $xyz$  were refined only for the virtual atoms. The procedure was repeated until convergence.

## 2.7. Parameters averaging for the VIR library building

The published crystal structures of organic compounds, including amino acids, oligopeptides and nucleic acid bases, were refined in the real and virtual spherical-atom formalism to build the library (the list in the supporting information). The charge-density distribution of each crystal structure was refined against theoretical structure factors obtained from first-principles calculations.

The strategy for generating the atom types was similar, as previously described for the multipolar database ELMAM2 (Domagała *et al.*, 2012). The electron-density parameters (valence populations and kappa parameters) for a given real or virtual atom type were averaged and stored in the database. Additionally, the distance between the  $A$  and  $Q$  atom for the  $Q_{AB}$  bond charges, and the distance to the carrier atom  $X$  and the angle  $Q_{LP1}-X-Q_{LP2}$  were stored for the  $Q_{LP}$  lone-pair atoms.

The distances between the considered atom and the first neighbours were compared. Contrarily to multipoles, the VIR modelling can accommodate variations in bond angles, as the  $Q_{AB}$  atoms remain by construction on the covalent bonds. For C, S, F and N atom types, only the first shell of neighbours was exactly compared. O is a polar atom type; its chemical properties and electron density are more subject to significant variations when the second neighbour shell is varied. For example, hydroxyl O atoms in phenols and alcohols have quite a different charge density regarding the electron lone pairs (Ahmed *et al.*, 2013); therefore, the second-shell atoms were also taken into account. As H atoms can be polar or apolar, their second-shell neighbours were also compared. Regarding the virtual atom,  $Q_{LP}$ 's associated with lone pairs were treated as their carrier atoms, while for bond  $Q_{AB}$  atoms, the chemical nature of  $A$  and  $B$  atom types were taken into account.

The electron-density parameters for each real and virtual atom type were represented as the weighted mean over all atoms contributing to one atom type. The average values over all other parameters (*i.e.* distances, angles) were computed. The resulting average electron-density parameters of the different atom types and geometrical parameters were written in a formatted database file.

## 2.8. Application to urea theoretical charge density

The charge density derived from the present library was compared with that of the previously published experimental multipolar ELMAM2 database (Domagała *et al.*, 2012) and with the theoretical UBDB library (Jarzemska & Dominiak, 2012) for the urea molecule. The coordinates of C, O and N atoms were taken from a refinement *versus* experimental X-ray data of Birkedal *et al.* (2004). The H-atom positions were obtained after their optimization with the *CRYSTAL09* software (Dovesi *et al.*, 2005, 2013).

The multipolar parameters ( $P_{val}, P_{lm}, \kappa, \kappa'$ ) were transferred from the ELMAM2 and UBDB databases to the atoms of the urea molecule, while keeping the coordinates fixed. After charge-density transfers, the urea molecule was neutralized, using the charge-scaling procedure of Faerman & Price (1990).

The THEO\_MUL charge-density refinement of urea was performed using the Hansen–Coppens (Hansen & Coppens, 1978) multipolar atom model, implemented in the *MoPro* package (Guillot *et al.*, 2001; Jelsch *et al.*, 2005). Multipoles were developed up to the hexadecapolar level for the non-H atom. For H atoms, one dipole and quadrupole were modelled. Refinements were performed *versus* the theoretical structure-factor amplitudes  $F_{hkb}$  using a unitary weighting scheme ( $\sigma_F = 1$ ). The multipolar refinement was carried out in the following manner:

(i) the atomic positions were kept fixed to the values obtained from the relaxed geometry; the scale factor was fixed to the absolute value (1.0);

(ii) the atomic displacement parameters were set to zero;

(iii) the only restraints imposed on the electron-density distribution were the following: the  $\kappa$  parameters of H atoms were restrained to a value of 1.16 (1) (Stewart, 1976);

(iv) the valence and multipole populations,  $\kappa$  and  $\kappa'$  parameters were refined sequentially, until convergence.

The properties derived from the charge-density model such as electron density, Fourier maps, electrostatic potential and electrostatic interaction energy, and topology of critical points were calculated with the *VMoPro* module of the *MoProSuite* software. The energy is obtained by integration of the product  $\rho_1(\mathbf{r})V_2(\mathbf{r})$  over the molecular volume. To accelerate the energy calculation over multipolar atoms, the Buckingham summation method based on interacting atomic moments (Buckingham, 1967; Volkov, Li *et al.*, 2004) is applied in *VMoPro* for distant atoms ( $d > 5 \text{ \AA}$ ) which do not have overlapping electron density. The energy from fitted point charges (FITQ) is computed by a Coulombic summation  $q_1q_2/r_{12}$  over atom pairs. The electrostatic lattice energy is computed in *VMoPro* by an automatic procedure which adds successive shells of molecules around the reference molecule in the crystal. A layer  $N$  is defined as translations ( $t_x, t_y, t_z$ ) of the unit cell, with the condition  $\max(|t_x|, |t_y|, |t_z|) = N$ . The energy contribution of successive layers  $N$  was observed to decrease proportionally to  $N^{-3}$ . The summation was stopped at  $N = 10$  layers of unit cells as convergence was reached (contribution  $< 0.05 \text{ kJ mol}^{-1}$  for layer 10).

## 2.9. The streptavidin/biotin complex

In the PDB deposited complex, two stereoisomers of biotin are bound to a streptavidin tetramer in a disordered manner, so that the two configurations of the ligand overlap in the binding sites, with occupancies of  $\sim 0.6/0.4$  for biotin and epibiotin, respectively. For our calculations, we selected the major conformation (*i.e.* the biotin ligand) associated with one of the monomers in which H atoms were added using the *MolProbity* server (Chen *et al.*, 2010) and, in the case of hydroxyl groups, manually checked and adjusted to comply with the most probable hydrogen-bonding possibilities. H-atom positions were subsequently adjusted so that their covalent bond lengths correspond to the average values observed by neutron diffraction (Allen & Bruno, 2010). Electron-density parameters of the VIR and ELMAM2 databases were then transferred to the resulting protein/ligand structure. In both electron-density models, in order to perform electrostatic potential computations using a neutral system, the charge of the biotin ligand was adjusted to its formal value of  $-1e$  (owing to the presence of the carboxylate group), and the protein monomer to a total charge of  $+1e$ .

The properties studied for this application (electrostatic potential and topology of the hydrogen bonds between the protein and the ligand) were also computed using the *VMoPro* module, part of the *MoProSuite* package (Jelsch *et al.*, 2005).

## 3. Application to urea

### 3.1. Validation of the VIR library using urea

One goal of this work is to compare refinements and electron-density-derived properties using the VIR library to those obtained using various models, including the multipolar ones. The charge densities and related electronic properties of the urea molecule obtained using the parameters stored in the new library (DB\_VIR) are then compared with those calculated from the virtual (THEO\_VIR) and multipolar modelling (THEO\_MUL) against theoretical structure factors, to transferred experimental ELMAM2 and theoretical UBDB databases.

Moreover, to assess the performances of the VIR modelling in the computation of electrostatic properties, a model of fitted atomic point charges was developed (FITQ model). These point charges were fitted against the electrostatic potential computed from the THEO\_MUL model in the region at a distance from 0 to 2 Å outside the van der Waals surface of the urea molecule. The FITQ charges were fitted using *VMoPro* by minimizing the sum of squares  $(V_{\text{THEO\_MULT}} - V_{\text{FITQ}})^2$ .

To summarize, a total of six different models of charge distribution are considered:

DB\_VIR: transferred parameters from the present VIR databank.

ELMAM2: transferred multipole parameters from the ELMAM2 library.

UBDB: transferred multipole parameters from the UBDB library.

**Table 3**

Crystallographic agreement factors obtained for the different models of the urea molecule after refinement *versus* theoretical  $F_{hkl}$  diffraction data at  $d = 0.4$  Å.

The minimum and maximum residual electron-density peaks are retrieved from Fourier residual maps computed using the reflections up to a resolution of  $d = 0.7$  Å.

Model	$R(F)$	$wR2(F)$	Minimum, maximum, r.m.s. residual density ( $e \text{ \AA}^{-3}$ )	No. of parameters
IAM	0.019	0.031	-0.36, 0.70, 0.12	0
DB_VIR	0.013	0.017	-0.33, 0.30, 0.062	0
THEO_VIR	0.012	0.016	-0.36, 0.29, 0.060	24
THEO_MUL	0.009	0.009	-0.12, 0.090, 0.022	49
ELMAM2	0.013	0.015	-0.11, 0.074, 0.019	0
UBDB	0.010	0.012	-0.16, 0.12, 0.023	0

THEO\_VIR: charge distribution refined against theoretical structure factors in the VIR formalism.

THEO\_MUL: charge distribution refined against theoretical structure factors in the multipole formalism.

FITQ: model of point charges fitted against the electrostatic potential obtained with the THEO\_MUL model.

The improvement of residual electron densities over the independent-atom model (IAM) was assessed for the urea crystal structure published by Birkedal *et al.* (2004).

### 3.2. Residual density

To give a general idea of the quality of the different crystallographic refinements, the  $R$  factors and residual-electron-density statistics are given for five models as well as for the IAM one in Table 3. The values show that the IAM has the highest  $R$  factor, while the THEO\_MUL refinement shows the best statistics. As expected, when compared with the IAM approach, the refined (THEO\_VIR or THEO\_MUL) and database-transferred (DB\_VIR, ELMAM2 or UBDB) models, which take into account the deformation of the molecular electron density, show improved crystallographic statistics. The three models based on multipoles show lower  $R$  factors than their VIR counterparts. This can be explained by the larger flexibility of the multipolar-atom model for which the number of parameters is higher compared with the VIR modelling. For a given modelling (MUL or VIR), direct charge-density refinement yields, as expected, lower  $R$  factors than the databank transfer counterpart. Since UBDB is from a theoretical origin, it can also be expected to yield better  $R$  factors *versus* theoretical  $F_{hkl}$ 's than the experimental ELMAM2 databank. Analogous behaviour of the  $R$  factors was observed in an earlier study for a crystal structure with experimental data at atomic resolution (Dadda *et al.*, 2012).

The quality of the charge-density models was also judged by analysing the Fourier residual-electron-density maps of urea shown in Figs. 3, S1 and S2. To highlight the contribution of the deformation density, which has limited contribution to the diffraction at ultra-high resolution, Fourier maps at 0.6 Å were computed. In all residual maps apart from the IAM, most of the bonding and lone-pair electron density is modelled. The bonds and lone-pair regions do not show significantly more



**Table 4**

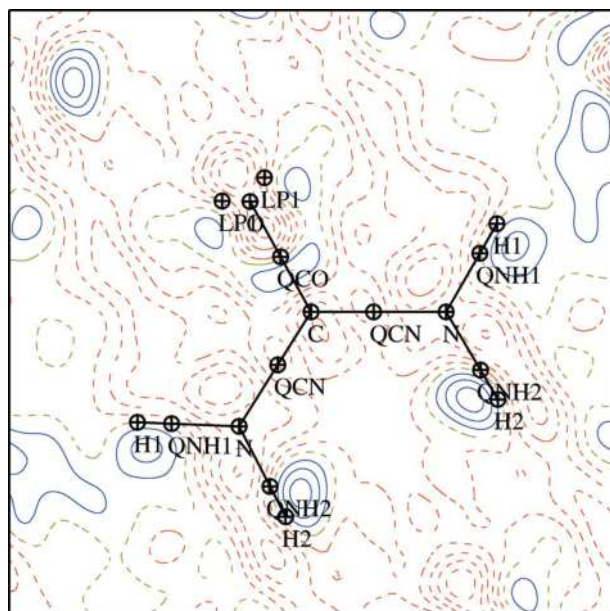
Correlation coefficients calculated for pairs of static deformation-electron-density maps computed on a three-dimensional volume around the urea molecule (2 Å margin around atoms).

Root-mean-square values of the maps are in italics on the diagonal elements of the table.

Model	DB_VIR	THEO_VIR	THEO_MUL	ELMAM2	UBDB
DB_VIR	<i>0.0396</i>	0.978	0.812	0.798	0.663
THEO_VIR		<i>0.0421</i>	0.728	0.789	0.562
THEO_MUL			<i>0.0373</i>	0.902	0.830
ELMAM2				<i>0.0395</i>	0.759
UBDB					<i>0.0368</i>

residual electron density for the transferred models compared to their refined counterpart. All peaks observed on the multipolar residual maps are low except for a few peaks reaching  $0.1 \text{ e } \text{Å}^{-3}$ . The VIR maps show, however, generally more residual density compared with the multipolar one. Owing to the nature of the additional VIR charges, the positive bonding and lone-pair electron-density accumulations are well reduced. However, modelling of electron-density depletion is limited to spherically distributions around scattering centres.

The residual peaks are globally sharper and higher in Fourier maps computed at  $0.4 \text{ Å}$  compared with  $0.6 \text{ Å}$  resolution. All the models (IAM, VIR, multipolar) show significant electron depletion around the nuclei of non-H atoms, especially at the higher resolution (Fig. S2). This is due to the different functions used in modelling the electron density in

**Figure 3**

Residual Fourier electron-density map in the plane of the urea molecule for the DB\_VIR model at  $0.6 \text{ Å}$  resolution after refinement *versus* theoretical  $F_{hkl}$  data. Contour level:  $\pm 0.05 \text{ e } \text{Å}^{-3}$ . Blue solid and red dashed lines denote positive and negative contours, respectively. The additional VIR atoms are visible on each covalent bond and close to the O atom. The maps for the other models as well as the maps computed at  $0.4 \text{ Å}$  are shown in the supporting information.

the theoretical calculation and in the crystallographic refinement. The structure factors obtained from theoretical data in *CRYSTAL06* software are derived from calculations using Gaussian functions. On the other hand, the electron density is modelled using Slater functions in the *MoPro* software. This results in a significant discrepancy in the core electron density located around the atom nuclei which can be corrected by using an additional expansion/contraction coefficient  $\kappa_{\text{core}}$  applied to the core electrons (Ahmed *et al.*, 2016).

The root-mean-square (r.m.s.) of the residual electron density is reduced by about 80% for the multipolar model compared to IAM. For the THEO\_VIR and DB\_VIR models, the decrease is only 50 and 48%, respectively (see Table 3). The DB\_VIR library yields results in between the multipolar and IAM models as it represents a significantly simpler formalism of the charge density compared with the Hansen & Coppens (1978) model.

### 3.3. Static deformation electron densities

To evaluate the performance of the VIR library, the static deformation electron density (DED) was computed by different methods for the urea molecule (see Fig. 4). The DED obtained from the DB\_VIR library shows that the electron-density peaks on the covalent bonds and on the lone-pair sites are correctly reproduced and compare well with those deduced from the THEO\_VIR model, the THEO\_MUL model and the ELMAM2 database. The UBDB DED is, however, lower for the C–N bonds bonding density in urea compared with the other models (Fig. 4e).

The difference between the static deformation-electron-density maps (see Fig. S3) highlights the dissimilarities between the models. One discrepancy is found around the carbonyl O atom between DB\_VIR and the two multipolar models ELMAM2 and THEO\_MUL, the lone-pair electron density being stronger in the first model.

The correlation coefficient of the DED was computed from three-dimensional maps for each pair of models (Table 4). The strongest correlation is observed between the DB\_VIR databank and THEO\_VIR model as both are obtained from theoretical data and use the same formalism to describe the DED. The two VIR models are closer to the THEO\_MUL model than to ELMAM2, as the latter originates from experimental data. The transferred multipolar ELMAM2 DED correlates very well with the THEO\_MUL one. The lowest correlations are found between UBDB and the two VIR models. Globally, the r.m.s. values of the DED show similar magnitudes, between  $0.0368$  and  $0.0421 \text{ e } \text{Å}^{-3}$  for the different models.

For all models, the DED peaks are centred on the C–N bond, but are closer to H on the N–H bonds. On the C=O bonds, the DED is centred for the VIR models but is closer to the C atom for the multipolar models. The bonding electron density is generally of comparable height in DB\_VIR and THEO\_VIR models. For example, the C–N bond shows DED peak heights at  $0.70$ ,  $0.65$ ,  $0.50$  and  $0.55 \text{ e } \text{Å}^{-3}$  for

**Table 5**

Correlation coefficients between the electrostatic potentials  $V$  calculated from the six electron-density models applied on the experimental crystal structure.

The statistics on  $V$  ( $e \text{ \AA}^{-1}$ ) are computed in a volume from a 0–2  $\text{\AA}$  distance outside of the van der Waals surface of urea. The r.m.s.( $V$ ) values are given in italics on the diagonal of the table.

Model	DB_VIR	THEO_VIR	THEO_MULT	ELMAM2	UBDB	FITQ
DB_VIR	<i>0.0466</i>	0.9954	0.9945	0.9969	0.9811	0.9983
THEO_VIR		<i>0.0548</i>	0.9845	0.9874	0.9684	0.9912
THEO_MULT			<i>0.0587</i>	0.9963	0.9830	0.9966
ELMAM2				<i>0.0624</i>	0.9909	0.9976
UBDB					<i>0.0530</i>	0.9830
FITQ						<i>0.0588</i>

DB\_MUL, THEO\_VIR, THEO\_MUL and ELMAM2, respectively.

The DB\_VIR and THEO\_VIR maps show high negative/positive DED peaks on the atomic nuclei of the C/N atoms. On the other hand, the electron lone pairs show the strongest DED in the DB\_VIR and THEO\_VIR maps. Overall, the VIR databank represents DEDs well correlated to the multipolar ones; lone pairs are particularly well modelled by the VIR model.

### 3.4. Urea electrostatic potential

The molecular electrostatic potential (ESP) is an important property that can be derived from the electron-density

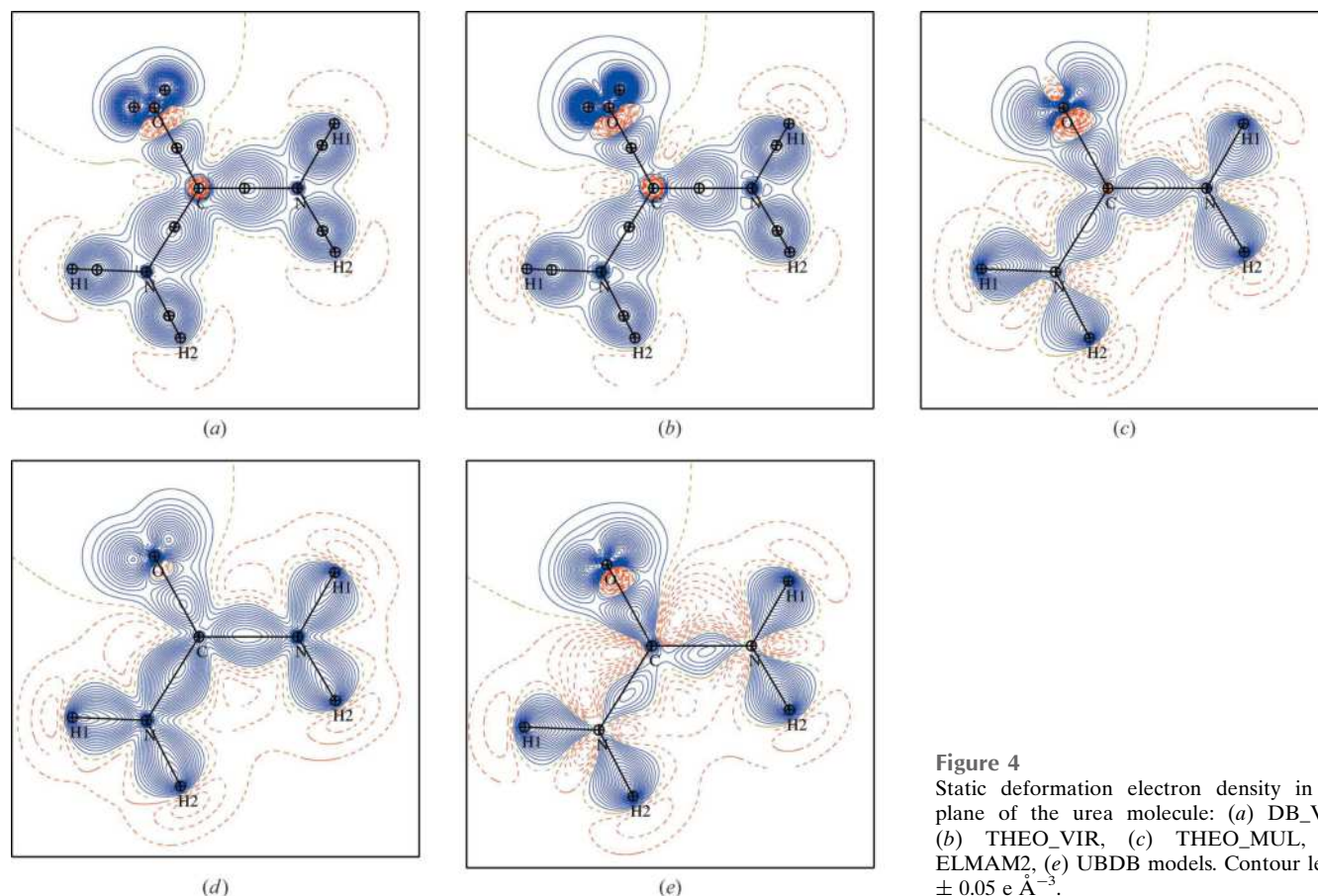
distribution and is an important tool for investigating molecular interactions.

The ESPs of all the examined models illustrate positive and negative potential values at the molecular 0.001 a.u. electron-density surface attributed, respectively, to the H and O atoms of urea (Fig. 5). All methods represent similarly the sign of the ESP around all the surface of the molecule. The agreement between the results of all methods is quite satisfactory, especially when experimental transferred DB\_VIR and ELMAM2, and theoretical THEO\_VIR and THEO\_MULT are respectively compared. Upon

visual inspection of the colour-mapped electron-density surfaces, the UBDB ESP shows the less extended negative electrostatic potential around the O atom.

To compare the models quantitatively, the ESP was computed in a volume around the van der Waals surface of the urea molecule, at a distance ranging from 0 to 2  $\text{\AA}$  to the surface. The r.m.s. values of the ESP are indicated in Table 5. A good quantitative agreement is observed between the r.m.s. ESP values for the six models. The urea multipolar ESPs, except UBDB, show higher r.m.s. values compared with the VIR ESPs. On the other hand, the ESPs transferred from the database are slightly lower than their refined counterparts.

The correlation coefficients between the electrostatic potential obtained from the different charge-density models



**Figure 4**  
Static deformation electron density in the plane of the urea molecule: (a) DB\_VIR, (b) THEO\_VIR, (c) THEO\_MUL, (d) ELMAM2, (e) UBDB models. Contour level:  $\pm 0.05 e \text{ \AA}^{-3}$ .

**Table 6**

Electrostatic interaction energy ( $\text{kJ mol}^{-1}$ ) for two closest interacting urea dimers present in the crystal packing, computed using the different models.

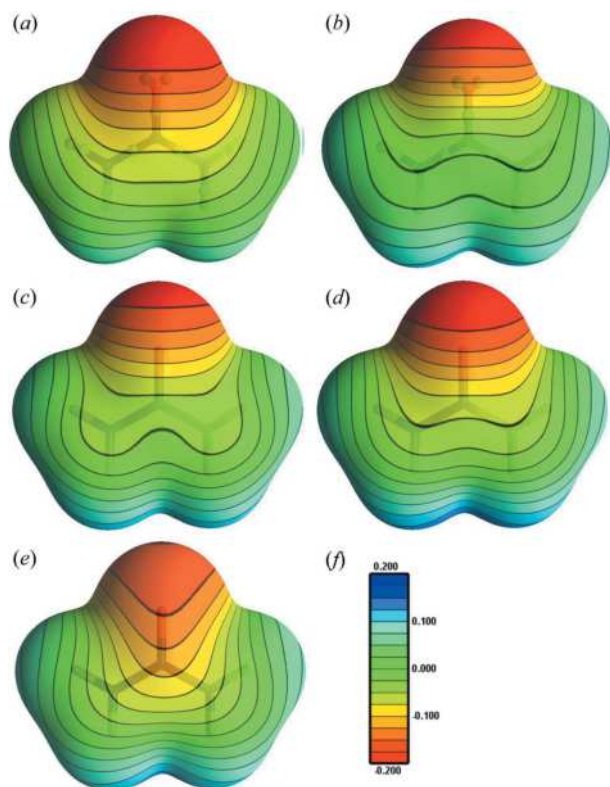
The crystal lattice energies ( $E_{\text{es}}$ ) are also computed.

Model dimer: symmetry	DB VIR	THEO VIR	THEO MUL	ELMAM2	UBDB	FITQ
(1): $x, y, z - 1$	-43	-54	-53	-65	-51	-48
(2): $y, -x + 1, -z + 1$	-36	-49	-51	-55	-48	-33
Lattice $E_{\text{es}}$	-106	-135	-153	-145	-144	-93

are also listed in Table 5. The different models display generally a very high correlation ( $r > 0.9684$ ). The pair of models displaying the lowest correlations is UBDB/THEO\_VIR. The fitted point charges yield, in the case of urea, an ESP in very good accordance with the other models outside of the molecular van der Waals surface.

### 3.5. Electrostatic interaction energy

The new databank (DB\_VIR model) is a tool to describe DEDs as well as ESPs. Gavezzotti (2013) underlines the importance of atomic charges in intermolecular electrostatic interaction energies ( $E_{\text{es}}$ ). This property is an invaluable tool for understanding molecular interactions.



**Figure 5**  
Electrostatic potential ( $\text{e Å}^{-1}$ ) mapped on the 0.001 a.u. electron-density surface of the urea molecule using: (a) DB\_VIR, (b) THEO\_VIR, (c) THEO\_MUL, (d) ELMAM2 and (e) UBDB models. Part (f) shows the legend of the mapped colours ( $\text{e Å}^{-1}$ ). The view was generated with the *MoProViewer* program (Guillot, 2011).

In order to test the transferability of the electron-density parameters (*i.e.* in the case of DB\_VIR, valence populations and kappa parameters), the  $E_{\text{es}}$  values have been computed (Table 6), for the two different kinds of urea hydrogen-bonded dimers found in the crystal packing, using the different models (Fig. 6).

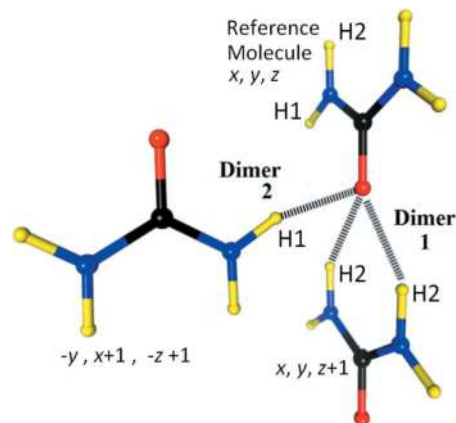
In the first head-to-tail dimer (hereafter noted dimer 1), there are two  $\text{C}=\text{O} \cdots \text{H}_2-\text{N}$  hydrogen bonds, which are symmetry-equivalent. The two molecules are related by a translation along  $\mathbf{c}$  and the bifurcated acceptor hydrogen bond displays an  $\text{O} \cdots \text{H}$  distance of 2.14 Å and  $\text{C}=\text{O} \cdots \text{H}$  angles of 145.5°.

The urea dimer 2 presents a peculiar hydrogen bond as the  $\text{H}_1 \cdots \text{O}$  direction is largely out of the O-atom lone-pair plane (the angle between  $\text{H}_1 \cdots \text{O}$  and the  $sp^2$  orbitals plane is 104.3°), which is unusual for such a short  $\text{O} \cdots \text{H}$  distance (Ahmed *et al.*, 2013). Indeed, the  $\text{O} \cdots \text{H}$  distance is 1.95 Å and the  $\text{C}=\text{O} \cdots \text{H}$  angle is 104.3°. This second dimer is obtained by application of the  $\bar{4}$  symmetry axis to the reference molecule so that the two urea planes are perpendicular.

In the crystal geometry, the  $E_{\text{es}}$  values of the two dimers show the same trends for the six models, as dimer 1 always has the strongest energy. The  $E_{\text{es}}$  is larger in magnitude for dimer 1 despite a longer  $\text{O} \cdots \text{H}$  distance, as it involves two hydrogen bonds compared with dimer 2 (Fig. 6).

In terms of  $E_{\text{es}}$  estimations obtained with the different models, it appears that DB\_VIR and the fitted point charges FITQ energies give the smallest values in magnitude. The model DB\_VIR yields the weakest energy at these distances for dimer 1. On the other hand, FITQ yields the weakest interaction energy for dimer 2, as, in the short  $\text{O} \cdots \text{H}$  hydrogen bond, the overlap of the electron densities is not taken into account. Conversely, for both dimers, ELMAM2 yields the largest interaction energy with -65 and -55  $\text{kJ mol}^{-1}$ . These  $E_{\text{es}}$  values are respectively 12 and 4  $\text{kJ mol}^{-1}$  stronger than those calculated with THEO\_MULT.

The electrostatic energy of a urea molecule in the crystal lattice was also computed with *VMoPro* and is in the range



**Figure 6**  
View of the two dimers found in the crystal packing of urea.

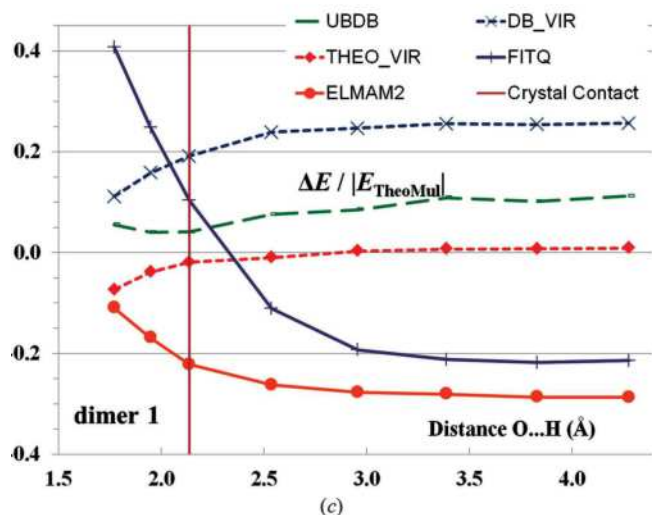
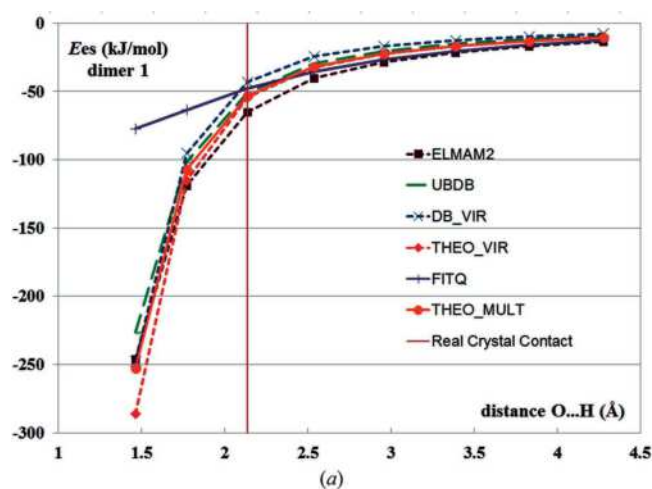


−93 to −153 kJ mol<sup>−1</sup>. For comparison, the lattice  $E_{es}$  value computed by Gavezzotti (2003) is −96 kJ mol<sup>−1</sup> and the reported experimental enthalpy of sublimation which also incorporates polarization and van der Waals energy is −88 kJ mol<sup>−1</sup>. The lattice  $E_{es}$  magnitudes follow globally the same trend as the two dimers; the fitted point charges yield, however, the smallest energy, followed by the DB\_VIR model.

The electrostatic energy dependency with distance was also assessed for the different models in Fig. 7. The second molecule of dimer 1 was further translated along the **c** axis, while in dimer 2 the translation was along **a** + **b**, which is perpendicular to one molecule and parallel to the other.

As expected, at short distances all the electrostatic energy values increase in magnitude. As shown in Figs. 7(a) and (b) in the short-range zone, the three multipolar and the two virtual-atom models follow the same trend, while a wide gap is observed for the energy computed from point charges. In the crystal geometry, for dimer 2, which has the shortest hydrogen bond, the FITQ  $E_{es}$  value is already underestimated, when compared with THEO\_MUL, by 18 kJ mol<sup>−1</sup>.

The agreement of the relative  $E_{es}$  values can be analyzed in the  $E_{es}/E_{es}(\text{THEO\_MUL})$  graphs shown in Figs. 7(c) and (d).



THEO\_MUL was used as the reference as it is the most elaborated model and was refined *versus* the theoretical structure factors of the urea crystal structure. The  $\Delta E/|E|$  values appear generally stable at distances longer than the dimers found in the crystal. UBDB and THEO\_VIR yield consistent results as they show less than 10% deviation with THEO\_MUL. At distances longer than the crystal dimers, DB\_VIR provides the weakest bonding energies while ELMAM2 displays the strongest, for both dimers. At long distances the FITQ model shows larger  $E_{es}$  values than THEO\_MUL. At short distances, the electrostatic energy is generally in good accordance within 23% deviation, except for the FITQ point charges which yield a significantly underestimated  $E_{es}$  as this model does not account for penetration of atomic electron clouds (Bojarowski *et al.*, 2016).

### 3.6. Urea dipole and quadrupole moments

For comparison, the dipole moment of urea was computed by two quantum-mechanical methods: (i) density functional theory using B3LYP/6-31G(d,p) and (ii) second-order Møller & Plesset (1934) MP2. The dipole moments were obtained by

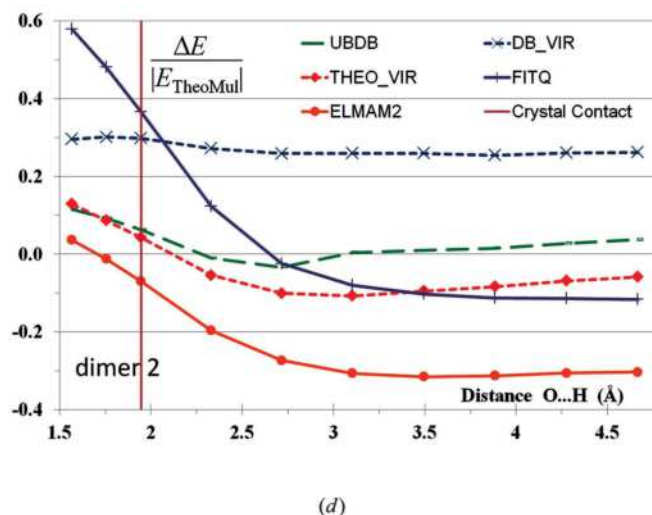
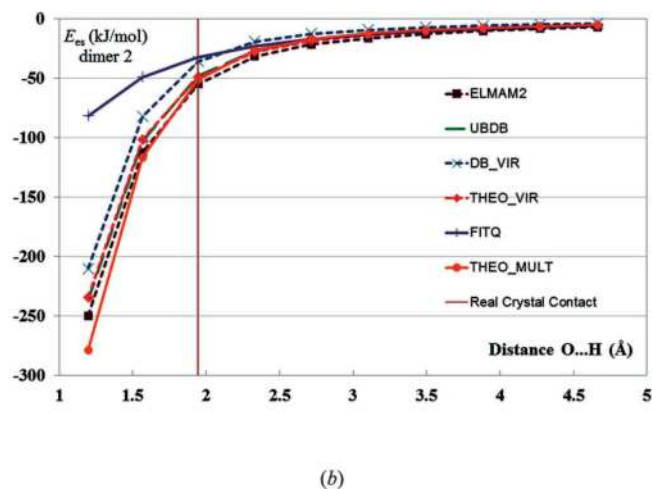


Figure 7

Electrostatic interaction energy for the different models of the two urea dimers as a function of O...H distance for dimer 1 (a) and dimer 2 (b). The relative discrepancy  $(E_{es} - E_{\text{TheoMul}})/|E_{\text{TheoMul}}|$  is also shown in (c) and (d).



Table 7

Molecular dipole moment  $\mu$  (D) of urea for the different models and percentage of enhancement with respect to the B3LYP theoretical calculation of the isolated molecule *in vacuo*.

Quadrupole moment ( $e \text{ \AA}^{-2}$ ) eigenvalues at the molecule centre of mass were also computed. The  $Q_{xx}$  eigenvector is collinear to the C=O bond,  $Q_{yy}$  is collinear to the N1–N2 direction and  $Q_{zz}$  is perpendicular to the molecular plane.

Method	DB_VIR	THEO_VIR	THEO_MUL	ELMAM2	UBDB	FITQ	B3LYP/MP2 <i>in vacuo</i>
$\mu$	4.44	5.10	5.57	5.24	4.71	5.52	4.30/4.33
$\Delta\mu/\mu$	+3%	+19%	+30%	+22%	+10%	+28%	reference/+0.7%
$Q_{xx}$	-0.67	-1.17	-0.74	-0.71	-0.56	-0.59	
$Q_{yy}$	1.56	1.97	1.98	2.15	1.60	1.76	
$Q_{zz}$	-0.89	-0.80	-1.24	-1.44	-1.04	-1.18	

*ab initio* calculations [6-31G(d,p) and MP2] and were computed directly from wavefunctions with the GAUSSIAN09 program package (Frisch *et al.*, 2009). The B3LYP/6-31G(d,p) model was treated as a reference to corroborate the results from the other models. The urea dipole moments calculated from the different approaches are shown in Table 7.

The experimental dipole moment of urea has been estimated to be 3.83 D (Brown *et al.*, 1975) in the gas phase, 4.2 D (Gilkerson & Srivastava, 1960) in an aqueous solution and 4.66 D (Lefebvre, 1973) in the solid state.

The MP2 and B3LYP methods carried on an isolated urea molecule yield dipole moments of 4.33 and 4.30 D, which are in good agreement with the experimental value in the gas phase. The dipole moments determined in the crystal are, as expected, higher than those computed in the gas phase.

obtained from DFT calculation using the B3LYP functional is defined as  $\% \Delta\mu = 100 \times (\mu - \mu_{\text{theo\_gas}})/\mu_{\text{theo\_gas}}$ . It is equal to +3, +19, +30, +22 and +10% for DB\_VIR, THEO\_VIR, THEO\_MUL, ELMAM2 and UBDB, respectively. The dipole moment has been observed to be very sensitive to the modelling and refinement strategy (Poulain-Paul *et al.*, 2012). The enhancement obtained from THEO\_VIR indicates a similar tendency as seen for the ELMAM2 databank. The largest enhancement is noticed for the THEO\_MUL model, which has notably a larger dipole moment than the THEO\_VIR model. The results are compatible with the generally maximum enhancement accepted of about +75% as fixed in the exhaustive study on enhancement of dipole moments for a great number of molecules by Spackman *et al.* (2007). Overall, the dipole moment predicted by the DB\_VIR

database is in good agreement with the other models and the experimentally measured value (Lefebvre, 1973).

The quadrupole moments computed for all models yield the eigenvectors, which are directed according to the urea molecular symmetries. The eigenvalues for a given direction are always of the same sign for all models. The largest positive eigenvalue  $Q_{yy}$  is associated with the direction N1...N2, while  $Q_{zz}$  is the largest negative eigenvalue in the plane perpendicular to the urea molecule. The models display generally similar values but there are notable exceptions. The  $Q_{xx}$  value is an outlier for the THEO\_VIR model as its magnitude is much larger compared with other models.

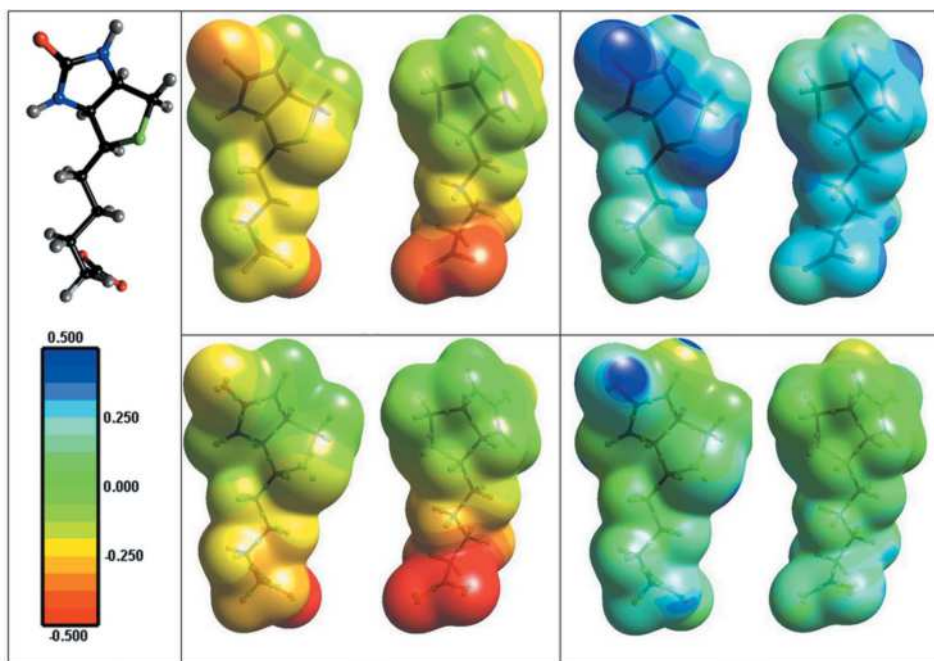


Figure 8

Analysis of the electrostatic potential in the biotin/streptavidin complex. A representation of the biotin molecule is shown on the left. The surface of biotin at the electron-density level 0.001 a.u. is coloured according to the ESP values ( $e \text{ \AA}^{-1}$ ). The ESP was generated using the ELMAM2 (first row) and DB\_VIR (second row) libraries. The two left-most surfaces (front and rear view) present the electrostatic potential generated by the biotin molecule and the two right-most ones present the electrostatic potential generated by the streptavidin protein. Images were generated with *MoProViewer* (Guillot, 2011).

#### 4. Application to a protein/ligand complex

In order to assess the applicability of the VIR library to protein-

ligand complexes, we built a real and spherical-atom model of the electron density for the structure of the streptavidin protein bound to biotin and to (+)-epi-biotin solved at 0.85 Å resolution (PDB code: 2F01; Le Trong *et al.*, 2006). This complex is known to present a particularly tight binding, exploited in biotechnology applications such as surface or membrane functionalization (Schiestel *et al.*, 2004; Roberts *et al.*, 2004).

#### 4.1. Electrostatic potentials on the surface of biotin

Electrostatic potential maps have been computed using ELMAM2 and DB\_VIR models, accounting separately for the contributions of biotin and protein atoms, and mapped on a 0.001 a.u. total electron-density surface of the biotin ligand (Fig. 8). These representations correspond on one hand to the ESP generated by biotin atoms (including virtual atoms in the case of the DB\_VIR model), and on the other hand to the potential arising from protein streptavidin atoms as being felt by the ligand. Comparison of maps calculated for the two models reveals a qualitative agreement, where essentially the regions of maximum and minimum electrostatic potential are reproduced.

As expected, these regions correspond to the most negatively charged moieties of biotin (*i.e.* O atoms of ketone and carboxylate groups). The ESP generated by the protein atoms is highest near the H atoms in the binding pocket, forming strong hydrogen bonds with the biotin ligand, for example between Tyr 30 hydroxyl H and ketone biotin O atoms.

Correlation coefficients between ESP values on the ~76 000 surface voxels reach 90 and 88% when ELMAM2 and DB\_VIR models are compared using, respectively, biotin and protein atoms. However, other comparison points lead to mixed results, especially when the ESP generated by the streptavidin monomer is considered. For instance, average ESP values on the surface obtained from the ELMAM2 and DB\_VIR models are identical in the case of biotin ( $-0.21 \text{ e } \text{Å}^{-1}$ ), but differ more for the ESP generated by the protein atoms ( $+0.13 \text{ e } \text{Å}^{-1}$ ,  $+0.28 \text{ e } \text{Å}^{-1}$ ).

Similarly, a good agreement can be observed between minimum values of the ESP generated by biotin atoms [ $V_{\min}(\text{DB\_VIR}) = -0.46 \text{ e } \text{Å}^{-1}$ ,  $V_{\min}(\text{ELMAM2}) = -0.59 \text{ e } \text{Å}^{-1}$ ], while a discrepancy in the sign can be noted in the case of the ESPs generated by streptavidin [ $V_{\min}(\text{DB\_VIR}) = -0.18 \text{ e } \text{Å}^{-1}$ ,  $V_{\min}(\text{ELMAM2}) = +0.04 \text{ e } \text{Å}^{-1}$ ]. The DB\_VIR model qualitatively reproduces the main features of the ESP obtained by a multipolar model of the electron density in the regions of the molecules that present the most electrophilic or nucleophilic characters.

#### 4.2. Critical points of the biotin/streptavidin complex

The ability of DB\_VIR modelling to allow the characterization of hydrogen-bond interactions by means of QTAIM (quantum theory of atoms in molecules) analyses was examined for a protein–ligand complex. For this purpose, a hydrogen-bond critical point (CP) search was carried out, focusing on interactions between residues lining the strepta-

vidin binding pocket and biotin atoms, using both ELMAM2 and DB\_VIR modelling of the molecular electron densities. It appears that CPs were similarly reproduced using either of the two models, with the single exception of a weak C–H···O interaction found using only the DB\_VIR model (Fig. 9, Table 8), resulting in a qualitatively very similar topological description of the interaction patterns between the protein and its ligand.

To describe the results in a more quantitative manner, a comparison of CP positions was first performed by computing distances between equivalent hydrogen-bond CPs (Table 8). Separation distances between CPs range between 0.03 and 0.16 Å, with an average value of 0.07 (3) Å. Moreover, a general trend can be drawn as the smallest distances between CPs correspond to the strongest hydrogen bonds (on the basis of the electron-density value at the critical-point position) and the largest one to the weakest. At a second stage, electron-density and Laplacian values at the hydrogen-bond CPs obtained with the two models were compared. The results show that the DB\_VIR model is able to reproduce electron densities on CPs obtained with a traditional multipolar electron density within a deviation ranging from less than 1% up to 16%. In this case, unlike for the CPs, no obvious correlations appear between the strength of the considered hydrogen bond and the electron-density discrepancies.

Regarding the Laplacian of the electron density at the CPs, as expected much larger variability is observed: Laplacian values obtained with the DB\_VIR modelling are almost systematically larger than their ELMAM2 counterparts, reaching in some cases more than twice their values. However, it is worth noting that both electron-density and Laplacian values at the CPs positions obtained with the DB\_VIR modeling are within usual values characterizing hydrogen-bond interactions in organic molecules. Weak positive Laplacian values of a few  $\text{e } \text{Å}^{-5}$  units and electron-density magnitudes around  $0.1 \text{ e } \text{Å}^{-3}$  were reported, for instance, by Espinosa *et al.* (1999).

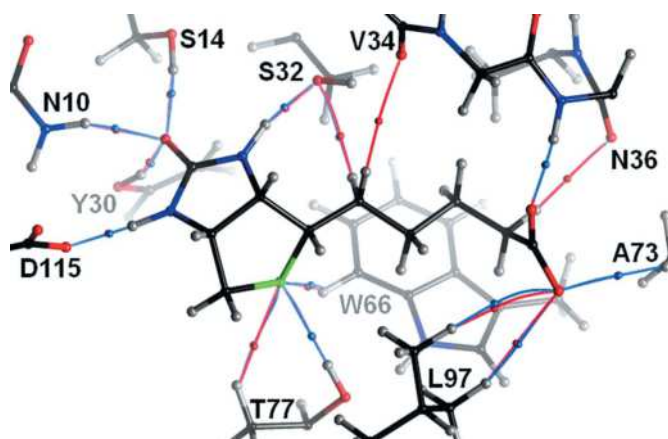


Figure 9  
View of the hydrogen bonds between biotin and protein streptavidin in the complex. The bond paths and critical points are shown in blue for ELMAM2 and in red for DB\_VIR models.

## 5. Conclusion

The new VIR library, applicable to common organic molecules, amino acids, oligopeptides and nucleic acid bases, has been constructed from periodic *ab initio* calculations of small molecules *via* theoretical structure factors. The VIR crystallographic refinements have benefitted from the application of constraint/restraints developed in the *MoPro* program. Database transfer is an essential tool for the analysis of the charge density, electrostatic potential and electrostatic interaction energy of organic compounds as well as biological macromolecules. The charge density derived from the VIR databank has been compared with THEO\_VIR, THEO\_MUL, ELMAM2 and UBDB models for the urea molecule.

The crystallographic refinement statistics (Table 3) are improved and the residual maps are clearer with any other tested charge-density models compared with the IAM. The VIR model, however, has fewer parameters to model the electron density and the crystallographic statistics are intermediate between the multipolar-atom models and the IAM approach. The weakest correlation of static deformation-electron-density maps was found between the THEO\_VIR and UBDB models.

The electrostatic potential distributions around the urea molecule were calculated and all the electron-density models exhibit highly correlated ESPs. The electrostatic interaction energies of urea dimers in the crystal estimated from the different models agree generally well, except for point charges at short distances. The DB\_VIR model shows globally a slight attenuation in the magnitude of the electrostatic interaction energies of urea compared with THEO\_VIR and to the multipolar models.

The DB\_VIR database reproduces well the dipole moment of the urea molecule with respect to experimental data in the solid state. Finally, the VIR library allows evaluation of the electrostatic interaction energies while avoiding the multipolar development. It can be used to calculate the molecular-electron-density distribution and its properties. The VIR library has the potential to be of interest to the scientific community for predicting new crystal structures (Dong *et al.*, 2015; Stavrou *et al.*, 2016) by coupling it with a code based on the concepts of evolutionary algorithms. Fast electrostatic energy calculations between sets of spherical charges can be achieved by implementing the methodology described by Bojarowski *et al.* (2016). In their study, accurate  $E_{es}$  values could be obtained using the *Amber* program point charges by applying a correcting factor at a short distance to account for the penetration of atomic electron densities. We also note that the VIR database could be used in existing force fields as it has a simpler formalism than the multipolar model. Computational approaches including off-centre charges (electron lone-pair sites) in molecular-mechanics calculations have been already applied (Dixon & Kollman, 1997). The increase in the number of charge sites results in overall improvement of the estimated energy and the angular dependence of hydrogen bonds is better retrieved. The Databank\_VIR\_2017 file is

**Table 8**

Topological properties at the streptavidin–biotin hydrogen-bond critical points: electron density and Laplacian.

First lines: ELMAM2 model, second lines: DB\_VIR model. *nf*: not found.  $\Delta\rho/\rho_{ELMAM2}$  and  $\Delta L/L_{ELMAM2}$  are the relative discrepancies of the electron density  $\rho$  and Laplacian  $L = \nabla^2(\rho)$  between the two models.  $\Delta d$  are the distances between the two CPs. Hessian eigenvalues are shown in the supporting information.

Protein atom	Biotin atom	$\rho$ (e Å <sup>-3</sup> )	$\Delta\rho/\rho$ (%)	$L = \nabla^2(\rho)$ (e Å <sup>-5</sup> )	$\Delta L/L$ (%)	$\Delta d$ (CPs) (Å)
HD21_10	O3	0.135	-6	1.47	-13	0.07
		0.151		1.89		
HGS_14	O3	0.297	4	1.42	-39	0.06
		0.275		3.22		
HH_30	O3	0.288	7	1.35	-37	0.05
		0.249		2.95		
OD2_115	HN1	0.295	0	2.04	-24	0.03
		0.296		3.31		
OG_32	HN2	0.180	0	1.46	-20	0.04
		0.181		2.19		
OG_32	H71	0.045	-6	0.68	-2	0.08
		0.051		0.71		
O_34	H72	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>	<i>nf</i>
		0.017		0.26		
O_36	H102	0.040	0	0.49	-7	0.04
		0.041		0.57		
H_36	O11	0.235	-1	0.28	-23	0.03
		0.239		0.33		
HD11_97	O12	0.024	1	0.14	-8	0.05
		0.011		0.16		
HD21_97	O12	0.012	3	0.14	-7	0.16
		0.011		0.16		
HB1_73	O12	0.016	-7	0.26	-1	0.05
		0.018		0.27		
HG21_77	S1	0.045	-5	0.45	-9	0.10
		0.049		0.54		
HZ2_66	S1	0.045	-5	0.58	2	0.08
		0.050		0.56		
HG1_77	S1	0.140	9	0.69	-29	0.09
		0.117		1.25		

available on the *MoPro* download website (<http://crm2.univ-lorraine.fr/lab/fr/software/mopro/download-mopro/>).

## 6. Related literature

References cited in the supporting information include: Bacon *et al.* (1964), Bacon & Jude (1973), Benabicha *et al.* (2000), Cohen *et al.* (2007), Dahaoui, Jelsch *et al.* (1999), Dahaoui, Pichon-Pesme *et al.* (1999), Domagała *et al.* (2011), Destro *et al.* (1988, 2000), Espinosa *et al.* (1996), Fronczek *et al.* (2002), Guillot *et al.* (2001), Guille & Clegg (2006), Hübschle *et al.* (2008), Harding & Long (1968), Kalinowski *et al.* (2007), Kulkarni *et al.* (1998), Lai & Marsh (1972), Munshi & Row (2000), Munshi *et al.* (2010), Narasimhamurthy *et al.* (1976), Tret'yak *et al.* (1987), Scheins *et al.* (2004), Saraswathi & Vijayan (2001), Sugawara *et al.* (1994), Tonogaki *et al.* (1993) and Zarychta *et al.* (2011).

## Acknowledgements

GENCI-CINES is thanked for providing access to computing facilities.



## Funding information

The following funding is acknowledged: CNRS and Université de Lorraine (grant No. 2009-X2009085106).

## References

- Afonine, P. V., Grosse-Kunstleve, R. W., Adams, P. D., Lunin, V. Y. & Urzhumtsev, A. (2007). *Acta Cryst.* **D63**, 1194–1197.
- Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.* **D60**, 260–274.
- Ahmed, M., Jelsch, C., Guillot, B., Lecomte, C. & Domagała, S. (2013). *Cryst. Growth Des.* **13**, 315–325.
- Ahmed, M., Nassour, A., Noureen, S., Lecomte, C. & Jelsch, C. (2016). *Acta Cryst.* **B72**, 572–580.
- Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* **B66**, 380–386.
- Bacon, G. E., Curry, N. A. & Wilson, S. A. (1964). *Proc. R. Soc. London Ser. A*, **279**, 98–110.
- Bacon, G. E. & Jude, R. J. (1973). *Z. Kristallogr.* **138**, 19–40.
- Bader, R. F. W. (1990). *Atoms in Molecules: A Quantum Theory*. Oxford University Press.
- Bağ, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* **A67**, 141–153.
- Becke, A. D. (1993). *J. Chem. Phys.* **98**, 5648–5652.
- Benabicha, F., Pichon-Pesme, V., Jelsch, C., Lecomte, C. & Khmou, A. (2000). *Acta Cryst.* **B56**, 155–165.
- Birkedal, H., Madsen, D., Mathiesen, R. H., Knudsen, K., Weber, H.-P., Pattison, P. & Schwarzenbach, D. (2004). *Acta Cryst.* **A60**, 371–381.
- Bojarowski, S. A., Kumar, P. & Dominiak, P. M. (2016). *Chem. Phys. Chem.* **17**, 2455–2460.
- Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* **B47**, 789–797.
- Brown, R. D., Godfrey, P. D. & Storey, J. (1975). *J. Mol. Spectrosc.* **58**, 445–450.
- Buckingham, A. D. (1967). *Advances in Chemical Physics*, edited by J. O. Hirschfelder, pp. 107–142. Hoboken: John Wiley and Sons, Inc.
- Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* **D66**, 12–21.
- Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables*, **14**, 177–478.
- Cohen, D. E., Benedict, J. B., Morlan, B., Chiu, D. T. & Kahr, B. (2007). *Cryst. Growth Des.* **7**, 492–495.
- Coppens, P. (1997). *X-ray Charge Densities and Chemical Bonding*. New York: Oxford University Press, Inc.
- Coppens, P., Abramov, Y., Carducci, M., Korjov, B., Novozhilova, I., Alhambra, C. & Pressprich, M. R. (1999). *J. Am. Chem. Soc.* **121**, 2585–2593.
- Dadda, N., Nassour, A., Guillot, B., Benali-Cherif, N. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 452–463.
- Dahaoui, S., Jelsch, C., Howard, J. A. K. & Lecomte, C. (1999). *Acta Cryst.* **B55**, 226–230.
- Dahaoui, S., Pichon-Pesme, V., Howard, J. A. K. & Lecomte, C. (1999). *J. Phys. Chem. A*, **103**, 6240–6250.
- Destro, R., Marsh, R. E. & Bianchi, R. (1988). *J. Phys. Chem.* **92**, 966–973.
- Destro, R., Roversi, P., Barzaghi, M. & Marsh, R. E. (2000). *J. Phys. Chem. A*, **104**, 1047–1054.
- Dietrich, H. & Scheringer, C. (1978). *Acta Cryst.* **B34**, 54–63.
- Dietrich, H., Scheringer, C., Meyer, H., Schulte, K.-W. & Schweig, A. (1979). *Acta Cryst.* **B35**, 1191–1197.
- Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* **D62**, 1325–1335.
- Dittrich, B., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* **B69**, 91–104.
- Dixon, R. W. & Kollman, P. A. (1997). *J. Comput. Chem.* **18**, 1632–1646.
- Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 337–351.
- Domagała, S. & Jelsch, C. (2008). *J. Appl. Cryst.* **41**, 1140–1149.
- Domagała, S., Munshi, P., Ahmed, M., Guillot, B. & Jelsch, C. (2011). *Acta Cryst.* **B67**, 63–78.
- Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
- Dong, H., Oganov, A. R., Zhu, Q. & Qian, G.-R. (2015). *Sci. Rep.* **5**, 9870.
- Dovesi, R., Orlando, R., Civalleri, B., Roetti, C., Saunders, V. R. & Zicovich-Wilson, C. M. (2005). *Z. Kristallogr.* **220**, 571–573.
- 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. (2013). *CRYSTAL09 User’s Manual*. University of Turin, Turin, Italy.
- Espinosa, E., Lecomte, C. & Molins, E. (1999). *Chem. Phys. Lett.* **300**, 745–748.
- Espinosa, E., Lecomte, C., Molins, E., Veintemillas, S., Cousson, A. & Paulus, W. (1996). *Acta Cryst.* **B52**, 519–534.
- Faerman, C. H. & Price, S. L. (1990). *J. Am. Chem. Soc.* **112**, 4915–4926.
- Fournier, B., Bendeif, E. E., Guillot, B., Podjarny, A., Lecomte, C. & Jelsch, C. (2009). *J. Am. Chem. Soc.* **131**, 10929–10941.
- Frisch, M. J. *et al.* (2009). *GAUSSIAN09*, Revision A.02. Gaussian Inc. Wallingford, CT, USA.
- Fronczek, F. R., Kim, K. K. & Strongin, R. M. (2002). Private communication.
- Gavezzotti, A. (2003). *CrystEngComm*, **5**, 429–438.
- Gavezzotti, A. (2013). *CrystEngComm*, **15**, 4027–4035.
- Gilkeson, W. R. & Srivastava, K. K. (1960). *J. Phys. Chem.* **64**, 1485–1487.
- Grabowsky, S., Weber, M., Buschmann, J. & Luger, P. (2008). *Acta Cryst.* **B64**, 397–400.
- Guille, K. & Clegg, W. (2006). *Acta Cryst.* **C62**, o515–o517.
- Guillot, B. (2011). *Acta Cryst.* **A67**, C511–C512.
- Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
- Guillot, R., Muzet, N., Dahaoui, S., Lecomte, C. & Jelsch, C. (2001). *Acta Cryst.* **B57**, 567–578.
- Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
- Harding, M. M. & Long, H. A. (1968). *Acta Cryst.* **B24**, 1096–1102.
- Hariharan, P. C. & Pople, J. A. (1973). *Theor. Chim. Acta*, **28**, 213–222.
- Held, J. & van Smaalen, S. (2014). *Acta Cryst.* **D70**, 1136–1146.
- Hohenberg, P. & Kohn, W. (1964). *Phys. Rev.* **136**, B864–B871.
- Hübschle, C. B., Dittrich, B., Grabowsky, S., Messerschmidt, M. & Luger, P. (2008). *Acta Cryst.* **B64**, 363–374.
- Jarzembska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* **A68**, 139–147.
- Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
- Kalinowski, R., Dittrich, B., Hübschle, C. B., Paulmann, C. & Luger, P. (2007). *Acta Cryst.* **B63**, 753–767.
- Koritsanszky, T. S. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1627.
- Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* **A58**, 464–472.
- Kulkarni, G. U., Kumaradhas, P. & Rao, C. N. R. (1998). *Chem. Mater.* **10**, 3498–3505.
- Kumar, P. & Dominiak, P. M. (2016). *Biophys. J.* **110**, 380.
- Lai, T. F. & Marsh, R. E. (1972). *Acta Cryst.* **B28**, 1982–1989.
- Le Trong, I., Aubert, D. G. L., Thomas, N. R. & Stenkamp, R. E. (2006). *Acta Cryst.* **D62**, 576–581.
- Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.
- Lefebvre, J. (1973). *Solid State Commun.* **13**, 1873–1875.
- Liebschner, D., Jelsch, C., Espinosa, E., Lecomte, C., Chabrière, E. & Guillot, B. (2011). *J. Phys. Chem. A*, **115**, 12895–12904.
- Møller, C. & Plesset, M. S. (1934). *Phys. Rev.* **46**, 618–622.
- Mullen, D. & Hellner, E. (1977). *Acta Cryst.* **B33**, 3816–3822.
- Mullen, D. & Hellner, E. (1978). *Acta Cryst.* **B34**, 1624–1627.



- Munshi, P., Jelsch, C., Hathwar, V. R. & Row, T. N. G. (2010). *Crystal growth Des.* **10**, 1516–1526.
- Munshi, P. & Row, T. N. G. (2006). *Acta Cryst.* **B62**, 612–626.
- Narasimhamurthy, M. R., Venkatesan, K. & Winkler, F. (1976). *J. Chem. Soc. Perkin Trans. 2*, pp. 768–771.
- Nassour, A., Kubicki, M., Wright, J., Borowiak, T., Dutkiewicz, G., Lecomte, C. & Jelsch, C. (2014). *Acta Cryst.* **B70**, 197–211.
- Pichon-Pesme, V., Lecomte, C. & Lachezar, H. (1995). *J. Phys. Chem.* **99**, 6242–6250.
- Poulain-Paul, A., Nassour, A., Jelsch, C., Guillot, B., Kubicki, M. & Lecomte, C. (2012). *Acta Cryst.* **A68**, 715–728.
- Racine, J. (2006). *J. Appl. Econ.* **21**, 133–141.
- Roberts, J. L., Moretti, P. A., Darrow, A. L., Derian, C. K., Vadas, M. A. & Pitson, S. M. (2004). *Anal. Biochem.* **331**, 122–129.
- Saraswathi, N. T. & Vijayan, M. (2001). *Acta Cryst.* **B57**, 842–849.
- Scheins, S., Dittrich, B., Messerschmidt, M., Paulmann, C. & Luger, P. (2004). *Acta Cryst.* **B60**, 184–190.
- Scheringer, C. (1980). *Acta Cryst.* **A36**, 205–210.
- Scheringer, C., Kutoglu, A., Hellner, E., Hase, H. L., Schulte, K.-W. & Schweig, A. (1978). *Acta Cryst.* **B34**, 2162–2165.
- Scheringer, C., Kutoglu, A., Mullen, D. & Hellner, E. (1978). *Acta Cryst.* **A34**, 475–476.
- Schiestel, T., Brunner, H. & Tovar, E. M. G. (2004). *J. Nanosci. Nanotech.* **4**, 504–511.
- Spackman, M. A., Munshi, P. & Dittrich, B. (2007). *ChemPhysChem*, **8**, 2051–2063.
- Stavrou, E., Lobanov, S. V., Dong, H. F., Oganov, A. R., Prakapenka, V. B., Konopkova, Z. & Goncharov, A. F. (2016). *Chem. Mater.* **28**, 6925–6933.
- Stewart, R. F. (1976). *Acta Cryst.* **A32**, 565–574.
- Sugawara, Y., Iimura, Y., Iwasaki, Y., Urabe, H. & Saito, H. (1994). *J. Biomol. Struct. Dyn.* **11**, 721.
- Tonogaki, M., Kawata, T., Ohba, S., Iwata, Y. & Shibuya, I. (1993). *Acta Cryst.* **B49**, 1031–1039.
- Tret'yak, S. M., Mitkevich, V. V. & Sukhodub, L. F. (1987). *Kristallografiya*, **32**, 1268.
- Volkov, A., Koritsanszky, T. & Coppens, P. (2004). *Chem. Phys. Lett.* **391**, 170–175.
- Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004). *J. Phys. Chem. A*, **108**, 4283–4300.
- Zarychta, B., Lyubimov, A., Ahmed, M., Munshi, P., Guillot, B., Vrielink, A. & Jelsch, C. (2015). *Acta Cryst.* **D71**, 954–968.
- Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* **A63**, 108–125.
- Zarychta, B., Zaleski, J., Kyzioł, J., Daszkiewicz, Z. & Jelsch, C. (2011). *Acta Cryst.* **B67**, 250–262.