# Verification of structural and electrostatic properties obtained by the use of different pseudoatom databases

**Joanna Maria Bąk,[a] Sławomir Domagała,[b] Christian Hübschle,[c] Christian Jelsch,[b] Birger Dittrich[c] and Paulina Maria Dominiak[a]\***

[a]Department of Chemistry, University of Warsaw, ul. Pasteura 1, 02-093 Warsaw, Poland, [b]Laboratoire de Cristallographie, Résonance Magnétique et Modélizations (CRM2), CNRS, UMR 7036, Institut Jean Barriol, Faculté des Sciences et Technologies, Nancy Université, BP 70239, 54506 Vandoeuvre-lès-Nancy Cedex, France, and [c]Institut für Anorganische Chemie, Georg-August-Universität Göttingen, Tammannstrasse 4, Göttingen 37077, Germany. Correspondence e-mail: pdomin@chem.uw.edu.pl

The existing pseudoatom databases (ELMAM, Invariom, UBDB and ELMAM2) enable structure refinement to be performed with the use of aspherical scattering factors computed from the transferable aspherical atom model (TAAM) as an alternative to independent atom model refinement. In addition, electrostatic properties can be estimated with the help of the databases. The quality of the structural and electrostatic properties obtained from the individual databases was tested. On the basis of a 100 K high-resolution single-crystal X-ray diffraction experiment on L-His–L-Ala dihydrate and 23 K high-resolution data for L-Ala [Destro & Marsh (1988). *J. Phys. Chem.* **92**, 966–973], the structural properties, electron-density distributions and molecular electrostatic potentials obtained from different TAAMs were compared to each other and to reference models. Experimental multipolar models and theoretical models refined against theoretical structure factors computed from periodic density functional theory (DFT) calculations were compared to the TAAMs in order to determine which model best describes the crystal-field effect. Unperturbed wavefunctions based on the MP2 and DFT calculations and properties obtained directly from these were used as a reference to judge how properly the databases reproduce the properties of isolated molecules. For Gly–L-His dihydrate, D,L-His and the above-mentioned two crystal structures, deviations of the molecular dipole moments and Coulombic intermolecular interaction energies from the reference values were examined. Root-mean-square deviations (RMSDs) and correlation coefficients were used as a quantitative measure of the quality of the analysed properties. TAAM refinements reproduce $X-H$ bond lengths optimized in theoretical periodic calculations. Structural properties obtained from different database models are similar to each other. The anisotropic displacement parameters from TAAMs are similar to the results of experimental multipolar refinement; differences are about 0.5 and 2.5% for high-resolution and low-resolution data, respectively. Differences in dipole-moment magnitudes calculated from database models are about 5%, and directions differ by up to 30°. The values of electrostatic interaction energies estimated from the individual TAAMs differ greatly from each other and from the reference values. RMSDs are about 9–15 and 22–33 kJ mol$^{-1}$ for UBDB and the other database models, respectively.

## 1. Introduction

Knowledge about the charge-density distribution in crystals allows one to study inter- and intramolecular interactions (Popelier, 2005) or molecular electrostatic properties. To obtain the electron-density distribution of molecules in the crystal environment from high-resolution X-ray data, crystallographers frequently use the multipolar model (Koritsanszky & Coppens, 2001) based on the Hansen & Coppens (1978) formalism. This model only approximates the charge-density

**Table 1**
The database construction details and selected applications.

ED: electron density. CSD: Cambridge Structural Database.

| | ELMAM or ELMAM2 | Invariom | UBDB |
|---|---|---|---|
| Structure factors (SFs) | SFs are obtained from high-resolution X-ray experiments. The two databases differ in definitions of the local coordination systems (Domagała & Jelsch, 2008) and in the way of averaging multipolar parameters | SFs are derived from theoretical density. The theoretical wavefunction is obtained from monomer optimization of the unique model compound with the B3LYP/D95++(3df,3pd) level of theory | SFs are derived from theoretical density based on the valence orbitals only. The theoretical wavefunctions are obtained from single-point calculation performed for several molecules with geometries taken from the CSD (Allen, 2002). The B3LYP/6-31G** level of theory is used |
| Multipolar parameters | Averaged over a family of chemically similar pseudoatoms | Obtained for the unique model compound – not averaged | Averaged over a family of chemically similar pseudoatoms |
| Crystal-field influence | In principle taken into account, but it is not known whether the effect can still be observed after averaging of the parameters | Not taken into account | Generally not taken into account, but geometric consequences of intermolecular interactions are to some extent included |
| H-atom treatment in the TAAM refinement | $X-H$ lengths constrained to standard neutron distances | Extended along the $X-H$ bond to theoretical distances | Extended along the $X-H$ bond to standard neutron distances |
| Main applications of the databases published so far | – TAAM refinement of proteins,<br>– deconvolution of ED from atomic motion [1],<br>– used as a starting point in refinement of the constrained multipolar parameters of proteins [2],<br>– calculation of protein electrostatic potential [3],<br>– calculation of the electrostatic interaction energies of inhibitor–enzyme complex [4] | – TAAM refinement of small molecules,<br>– increasing precision of the refined Flack parameter [5],<br>– modelling of ED of disordered solvent molecules in experimental charge-density study [6],<br>– deconvolution of ED from the atom motion and detecting disorder [7],<br>– obtaining ED from low-resolution X-ray data [8],<br>– calculation of electrostatic potential and topological atom volumes and charges for a given organic structure [9],<br>– calculation of a macromolecular electrostatic potential [10] | – TAAM refinement of small molecules,<br>– calculation of the electrostatic interaction energy of small biomolecular complexes [11] and ligand–protein complexes [12],<br>– modelling of ED of disordered methyl group in experimental charge-density study [13] |

References: [1] Jelsch *et al.* (1998); [2] Jelsch *et al.* (2000), Guillot *et al.* (2008); [3] Muzet *et al.* (2003), Liebschner *et al.* (2009); [4] Fournier *et al.* (2009); [5] Dittrich, Strumpel *et al.* (2006); [6] Dittrich *et al.* (2007); [7] Dittrich, Warren *et al.* (2009); [8] Dittrich, Hübschle *et al.* (2009); [9] Dittrich, Weber *et al.* (2009); [10] Dittrich *et al.* (2010); [11] Li *et al.* (2006); [12] Dominiak *et al.* (2007, 2009); [13] Bąk *et al.* (2009).

distribution. This is because X-ray data have limited resolution and include experimental errors. Furthermore, multipole model parameters can be highly correlated with atomic displacement parameters (ADPs) and the model itself is not flexible enough to accurately describe the electron density near the nuclear positions and in the bond region at the same time (Abramov *et al.*, 2000; Spackman & Byrom, 1996; Volkov *et al.*, 2001; Madsen *et al.*, 2004). However, the electron density obtained from experimental multipolar refinement has been used with some success to study, for example, intermolecular interactions and the electrostatic properties of several biomolecules (for recent results see, for example, Holstein *et al.*, 2010; Munshi *et al.*, 2010; Fournier *et al.*, 2009; Grabowsky *et al.*, 2008) and molecular dipole moments in the crystal (Spackman *et al.*, 2007; Koritsanszky & Coppens, 2001, and references cited therein).

The resolution of experimental data limits the number of multipolar parameters that can be refined. Therefore, the finding that multipolar parameters of atoms with similar chemical environments are transferable between different molecules (Brock *et al.*, 1991) has been used for the creation of databanks of aspherical atom parameters. With fast progress in X-ray measurements, the databanks were meant to allow more precise structure refinement and to approximate electrostatic properties of proteins and other biomolecules. There are currently four databanks available: the experimental ELMAM (Pichon-Pesme *et al.*, 1995; Zarychta *et al.*, 2007) and ELMAM2 (Domagała & Jelsch, 2008) libraries, the theoretical Invariom (Dittrich *et al.*, 2004; Dittrich, Hübschle *et al.*, 2006) database and the theoretical University at Buffalo Pseudoatom Databank (UBDB) (Koritsanszky *et al.*, 2002; Volkov, Li *et al.*, 2004; Dominiak *et al.*, 2007). The multipole parameters stored in each of the databases can be transferred to a chosen structure, thus offering the possibility of modelling the electron density more correctly than the independent atom model (IAM) can. The transferable aspherical atom model (TAAM) obtained in such a procedure can be used instead of the IAM to refine only coordinates and atomic displacements from low- ($0 < \sin\theta/\lambda < 0.7$ Å$^{-1}$) or high-resolution X-ray data. One may perform TAAM refinement with the *XDLSM* (Volkov, King & Coppens, 2006) or *MoPro* (Jelsch *et al.*, 2005) programs. The TAAM refinement signifi-

cantly improves molecular geometries, discrepancy $R$ factors and ADPs, (Jelsch *et al.*, 1998; Zarychta *et al.*, 2007; Dittrich *et al.*, 2005, 2008; Volkov *et al.*, 2007). The TAAM may also be used to estimate electrostatic properties and may serve as a starting point for high-resolution multipolar refinement. Details of the particular database construction method and other possible applications of the databases are listed in Table 1. A comparison of selected properties has already been performed for pairs of databases. The multipolar parameters and electron densities from ELMAM and UBDB were compared by Pichon-Pesme *et al.* (2004) and Volkov, Korit-sanszky, Li & Coppens (2004). Similarities between the electron densities and bond-critical-point properties obtained from UBDB and Invariom were described by Johnas *et al.* (2009).

The properties estimated from different TAAM models have limited accuracy. The aim of the current study was to test the quality of the structural and electrostatic properties obtained by the use of the databases of aspherical pseudo-atoms: ELMAM, ELMAM2, Invariom and UBDB. For the crystal structures of L-alanine (marked as ALA), L-His–L-Ala dihydrate (marked as HA), D,L-histidine (marked as HIS) and Gly–L-His dihydrate (marked as GH), the properties of the TAAM models were compared with several reference values. The structures of these molecules are shown in Fig. 1. We compared experimental database models to the results of two experimental multipolar refinements against high-resolution X-ray data of ALA (Destro & Marsh, 1988; Destro *et al.*, 2008) and HA (this work), and multipolar models fitted to structure factors calculated from the theoretical periodic wavefunction. To judge to what extent the databases reproduce the properties of isolated molecules, we used properties obtained directly from theoretical molecular densities, which are free from the use of the multipolar model.
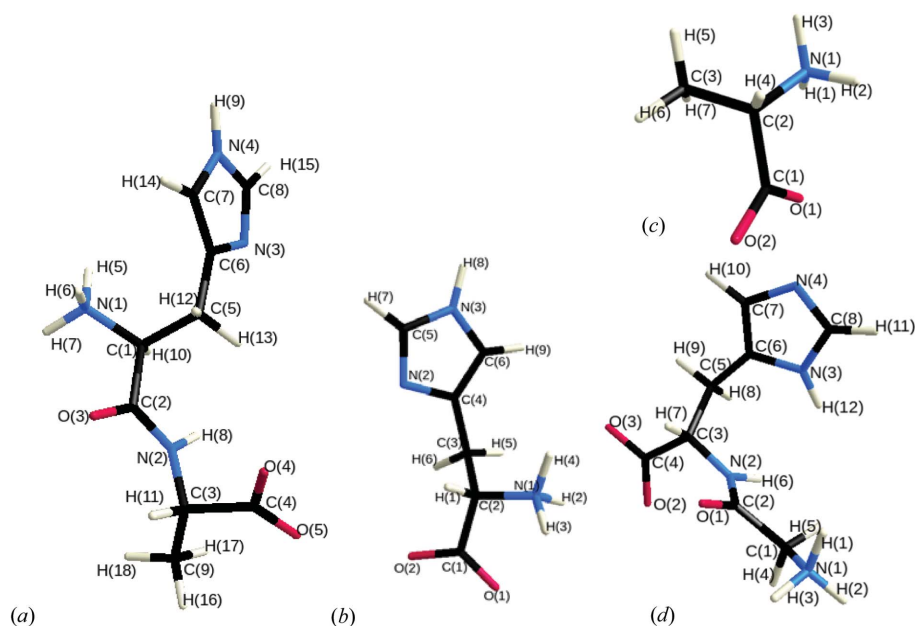
## 2. Experimental and computational methods

### 2.1. Periodic wavefunctions, theoretical structure factors and theoretical multipole models

We used the *CRYSTAL06* program (Dovesi *et al.*, 2008) to optimize geometries and calculate the periodic wavefunctions of HA, ALA, HIS and GH crystals. Starting geometries were taken from the following experimental data: our own for HA [the structure was reported by Steiner (1996) for the first time)], Destro & Marsh (1988) for ALA, Coppens *et al.* (1999) for HIS and Cheng *et al.* (2005) for GH. In all the calculations, the density functional theory (DFT) method at the B3LYP (Becke, 1993; Lee *et al.*, 1988) level with the DZP basis set (Dunning, 1970) was applied. Because of *CRYSTAL06* program limitations, diffuse functions were not used, but the DZP basis set has been shown to provide reliable and consistent results in studies involving intermolecular interactions (Spackman & Mitchell, 2001). The lattice parameters were kept fixed at the X-ray values. The shrinking factors (IS) along the reciprocal-lattice vectors were set at 4 and 8. The truncation parameters were set as ITOL1 − ITOL4 = 6 and ITOL5 = 19. The exponents of the polarization functions were not scaled because of a large difference between ITOL4 and ITOL5 as suggested by Spackman & Mitchell. A level shifter value of 0.6 Hartree and a FMIXING function of 30% were used. The optimized geometries were used as reference in comparison to different X-ray structure refinements.

Four sets of static theoretical structure factors were generated by the XFAC command in *CRYSTAL06* from the calculated wavefunctions, up to the resolution $\sin\theta/\lambda = 1.2 \text{ Å}^{-1}$. On the basis of theoretical structure factors the CR theoretical multipolar models were obtained for each compound using the *XDLSM* module from the *XD2006* (Volkov, Macchi *et al.*, 2006) package. Local atomic site-symmetry constraints were applied to multipolar population parameters. Multipoles up to the hexadecapolar level for non-H atoms, and bond-directed dipoles and quadrupoles for H atoms were refined together with individual $\kappa$ and $\kappa'$ parameters for each atom. Atomic positions and scale factors were not refined. Refinements were carried out on $F$. Atomic scattering factors were based on the atomic wavefunctions of Clementi & Roetti (1974). For the deformation functions single-$\zeta$ exponents corresponding to weighted averages over the $s$- and $p$-shell values given by Clementi & Raimondi (1963) were used.

### 2.2. Experimental multipole refinements

Multipole refinements were performed against our own 100 K experi-



**Figure 1**
The molecular structures and labelling schemes of: (*a*) L-His–L-Ala (HA), (*b*) D,L-His (HIS), (*c*) L-Ala (ALA), (*d*) Gly L-His (GH).

**Table 2**
Details of the particular types of TAAM refinements.

| Definition of the type of refinement | TAAM $P_v$, $P_{lmp}$, $\kappa$ and $\kappa'$ transferred from | | | | |
| | ELMAM [1] | ELMAM2 [2] | Invariom [3] | UBDB [4] | CR |
|---|---|---|---|---|---|
| Overall scale factor, positions of all atoms, ADPs for non-H atoms and IDPs for H atoms were refined against high-resolution experimental data | EL | EL2 | IM | UB | TH |
| 0.7 – overall scale factor, positions of all atoms, ADPs for non-H atoms and IDPs for H atoms were refined against low-resolution experimental data ($\sin\theta/\lambda < 0.7$ Å$^{-1}$) | EL0.7 | EL20.7 | IM0.7 | UB0.7 | TH0.7 |
| RB – H atoms were shifted along $X$—H directions taken from the particular 0.7 refinement to the distances recommended by the given database (average neutron or theoretical distances), overall scale factor, positions and ADPs for non-H atoms and IDPs for H atoms were refined against high-resolution experimental data | EL_RB | EL2_RB | IM_RB | UB_RB | TH_RB |
| CR –$P_v$, $P_{lmp}$, $\kappa$ and $\kappa'$ were transferred to geometries optimized in theoretical periodic calculations (CRs). The models were used to compare deformation densities and to calculate the Coulombic energy of interaction (Ee) between selected dimers | EL_CR | EL2_CR | IM_CR | UB_CR | CR |

References: [1] Pichon-Pesme et al. (1995), Zarychta et al. (2007); [2] Domagała & Jelsch (2008); [3] Dittrich, Hübschle et al. (2006); [4] Dominiak et al. (2007).

mental data for HA and against those collected by Destro & Marsh (1988) for ALA at 23 K. Details of crystallization, data collection and data reduction for HA are given in the supplementary materials.[1] Two kinds of multipole refinements on $F$ were carried out for each compound with the *XDLSM* module from the *XD2006* (Volkov, Macchi et al., 2006) package, with $I/\sigma_I > 2$ cutoff:

MM_HO – in which starting atomic positions for all atoms and ADPs for non-H atoms were found by 'mixed' refinements as proposed by Hoser et al. (2009) (high-order refinement of non-H atoms and low-order refinement of H atoms) with the $\sin\theta/\lambda$ cutoff at 0.7 Å$^{-1}$;

MM_TH – in which atomic positions for all atoms and ADPs for non-H atoms were taken from the TH_RB refinement described below.

In both refinement procedures H-atom positions were shifted to the theoretical $X$—H distances obtained from theoretical geometry optimization of the crystal structure. ADPs for H atoms were generated by the *SHADE* program (simple hydrogen anisotropic displacement estimator; Madsen, 2006). The $\kappa$ and $\kappa'$ parameters for all atoms were taken from the theoretical multipolar model (CR) and kept fixed. Local atomic site-symmetry constraints (the same as in CR) were applied to multipolar population parameters. Multipoles up to the hexadecapolar level for non-H atoms, and bond-directed dipoles and quadrupoles for H atoms were refined. Multipole expansion was increased successively in the course of the refinement. All the positions and ADPs were kept fixed during refinement of multipolar populations. Finally, the positions and ADPs of non-H atoms were refined together with multipolar populations (only in the case of the MM_HO type of refinements). The same atomic scattering

factors were applied as in the case of theoretical multipolar refinements.

### 2.3. TAAM refinements

TAAM refinements are based on the pseudoatom multipole model. Pseudoatom parameters are transferred from an independent source such as a database or theoretical calculations. In TAAM refinement, an overall scale factor, atomic positions and ADPs are refined against experimental data, in the presence of pseudoatom parameters which are kept fixed.

For this study, HA and ALA structures refined by the IAM method in *SHELX* (Sheldrick, 2008) against low-resolution experimental data ($0 < \sin\theta/\lambda < 0.7$ Å$^{-1}$) were used as a starting point for TAAM refinements. In the IAM refinement, the overall scale factor, all atom positions, ADPs for non-H atoms and isotropic displacement parameters (IDPs) for H atoms were refined. Next, pseudoatom parameters ($P_v$, $P_{lmp}$, $\kappa$ and $\kappa'$) were transferred from one of the four databases or from theoretical multipolar models (CR). The transfer of the pseudoatom parameters from the particular database and setting of the local coordinate systems (characteristic for the database) were accomplished with the *Invariomtool* (Hübschle et al., 2007), *LSDB* (version 10 October 2006; Volkov, Li et al., 2004) program and *MoPro* (Jelsch et al., 2005) package. To ensure the molecular electroneutrality, the monopole populations of the pseudoatoms were scaled *a posteriori* by distributing the same fraction of charge difference over all the atoms of the molecule in the case of ELMAM and Invariom databases, or by using the Faerman & Price (1990) method in the case of the ELMAM2 and UBDB databases. For the Invariom and UBDB databases, several methods of rescaling are available to the user, but these did not improve the quality of the properties analysed here. TAAM refinements reported in this paper were performed with the *XDLSM* program, but similar geometries resulting

---

[1] Supplementary data for this paper are available from the IUCr electronic archives (Reference: CN5023). Services for accessing these data are described at the back of the journal.

**Table 3**
Refinement statistics calculated for the high-resolution ($\sin\theta/\lambda \leq 1.24$ Å$^{-1}$ for HA and $\sin\theta/\lambda \leq 1.08$ Å$^{-1}$ for ALA) diffraction data.

| | HA | | | ALA | | |
|---|---|---|---|---|---|---|
| | $R(F)$ (%) | $wR(F)$ (%) | min; max of residual density (e Å$^{-3}$) | $R(F)$ (%) | $wR(F)$ (%) | min; max of residual density (e Å$^{-3}$) |
| TH_RB | 2.0 | 1.7 | −0.21; 0.26 | 2.0 | 2.3 | −0.22; 0.17 |
| EL_RB | 2.0 | 1.7 | −0.20; 0.24 | 2.3 | 2.6 | −0.30; 0.21 |
| EL2_RB | 1.9 | 1.5 | −0.23; 0.25 | 2.0 | 2.3 | −0.22; 0.18 |
| IM_RB | 2.0 | 1.4 | −0.17; 0.28 | 2.1 | 2.4 | −0.22; 0.20 |
| UB_RB | 1.9 | 1.4 | −0.15; 0.25 | 2.1 | 2.3 | −0.22; 0.18 |
| MM_HO | 1.8 | 1.3 | −0.15; 0.21 | 2.0 | 2.0 | −0.18; 0.21 |
| MM_TH | 1.9 | 1.3 | −0.18; 0.23 | 2.0 | 2.0 | −0.21; 0.18 |
| IAM (high resolution) | 2.8 | 2.7 | −0.30; 0.50 | 2.9 | 3.3 | −0.29; 0.45 |

from TAAM refinements were obtained using the *MoPro* package. Details of the particular refinement procedures carried out are given in Table 2; details of the atom types used are listed in Table S1 in the supplementary material.

## 2.4. Electrostatic properties and Coulombic energy of interaction between dimers

To compute Coulombic interaction energies (Ees) two different methods were used. In the first method Ees were computed directly from the wavefunctions. On the basis of geometries optimized in periodic calculations (as for CRs), single-point calculations of unperturbed wavefunctions were performed for eight monomer molecules: ALA, HA, two HA water molecules, HIS, GH and two GH water molecules. In the calculations, several levels of theory and different basis sets, as implemented in the *GAUSSIAN03* program (Frisch *et al.*, 2004), were used: MP2/aug-cc-pVDZ marked as MP2, BLYP/ DZP marked as DZP, B3LYP/D95++3df3pd marked as D95, B3LYP/6-31Gdp marked as 6-31Gdp. Next, Ees of interactions between two monomer densities in relative orientations as found in the crystal were computed using the *SPDFG* program (Volkov, King & Coppens, 2006). In the second method Ee



**Figure 2**
The $X-H$ bond lengths (Å) in the HA structure obtained from IAM and different TAAM refinements (against high- or low-resolution data) compared to structural parameters from the reference theoretical geometries (CR). Standard deviations from all refinements do not exceed 0.02 Å. Horizontal lines indicate standard neutron distances (Allen & Bruno, 2010).

values were calculated from multipolar models with the exact potential/multipole moment (EPMM; Volkov, Koritsanszky & Coppens, 2004) method as implemented in *XDPROP* of *XD2006*. A distance of 6 Å was used as the boundary between the exact evaluation of the potential and the Buckingham approximation.
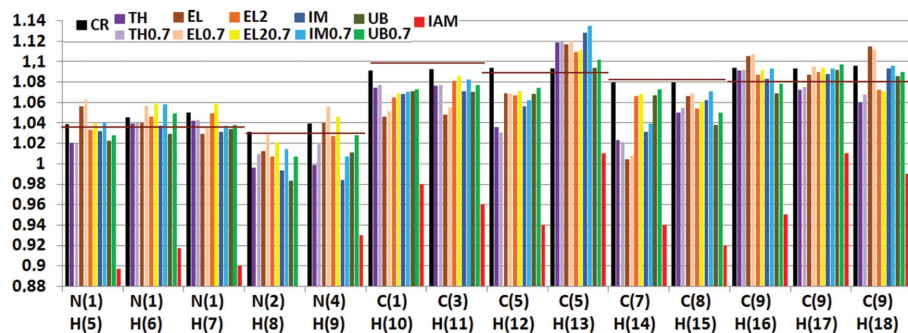
The *XDPROP* module of the *XD2006* package was used to calculate dipole-moment vectors and molecular electrostatic potentials (MEPs) from the multipolar models. Electrostatic properties from the MP2, DZP, D95 and 6-31Gdp models were calculated directly from wavefunctions in the *GAUSSIAN03* program.

## 3. Results and discussion

The TAAM refinements were performed on the experimental data for HA and ALA. The multipolar parameters were transferred from the different databases: ELMAM, ELMAM2, Invariom and UBDB, or from theoretical models. The results of the above refinements were compared to those from IAM and experimental high-resolution multipole refinements (MM_HO, MM_TH). To give an overview of the quality of the data sets and differences among models, refinement statistics are summarized in Table 3. All statistics, *i.e.* $R(F)$, $wR(F)$ and residual densities, are comparable for models where multipole parameters were used (MM_HO, MM_TH and all TAAM refinements) and much better than for the IAM refinement against high-resolution data, the $R$ factor being lower by about 1%. A similar decrease in the $R$ factor was observed in previous studies regarding databases (Jelsch *et al.*, 1998; Zarychta *et al.*, 2007; Dittrich, Hübschle *et al.*, 2006; Volkov *et al.*, 2007).

### 3.1. Structural parameters

It is a well known phenomenon that $X-H$ bond lengths obtained from standard IAM refinement of X-ray data are severely underestimated. Although the authors of each database recommend restraining $X-H$ bond distances during the TAAM refinement, we decided to check to what extent each of the databases is able to recover the proper position of H atoms when these are freely refined. Because of the lack of good neutron data for both crystal structures, we chose geometries obtained from periodic structure optimization as reference values, mostly to have reliable estimates of H-atom positions. Another possibility was to take standard neutron distances as a reference, but distances resulting from periodic optimization are more specific for the studied systems and still close to the standard ones. The $X-H$ bond lengths and the $Y-X-H$ angles obtained from the different TAAM refinements and from the conventional

**Table 4**
Root-mean-square deviations (RMSDs) of $X$—H bond lengths (Å) and $Y$—$X$—H angles (°) obtained in different structural refinements of HA and ALA relative to structural parameters from the reference theoretical geometries (CR).

|  | TH | EL | EL0.7 | EL2 | EL20.7 | IM | IM0.7 | UB | UB0.7 | IAM |
|---|---|---|---|---|---|---|---|---|---|---|
| $X$—H | 0.03 | 0.04 | 0.04 | 0.02 | 0.02 | 0.03 | 0.03 | 0.03 | 0.02 | 0.12 |
| $X$—H—$Y$ | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |

IAM refinement together with the reference values are presented in Fig. 2 and in Fig. 1S in the supplementary material. The root-mean-square deviations (RMSDs) of the H—$X$ bond lengths and $X$—H—$Y$ angles, relative to structural parameters from the reference theoretical geometries, are summarized in Table 4.

All TAAM refinements lead to $X$—H bond lengths similar to the reference distances, with RMSDs not exceeding 0.04 Å in the worst case (Table 4). Standard deviations in all refinements do not exceed 0.02 Å. $X$—H bond lengths obtained from the conventional IAM are much shorter – the RMSD equals 0.12 Å. For both molecules studied, TAAM refinements did not alter the $X$—H—$Y$ angle values in comparison to IAM. The angle RMSDs are around 2° for all refinements.

In the TAAM refinements, a more realistic atomic scattering model based on the detailed electron distribution is used. Thus, in such refinements, contributions of bonding electron density and systematic errors to the refined atomic positions and ADPs are minimized. In the case of ADPs, it has been shown already (Volkov *et al.*, 2007; Zarychta *et al.*, 2007; Dittrich *et al.*, 2008) that any TAAM refinement leads to ADPs smaller than the ones from IAM refinement and closer



**TH0.7 minus IAM**   **TH_RB minus UB0.7**   **TH_RB minus EL2_RB**

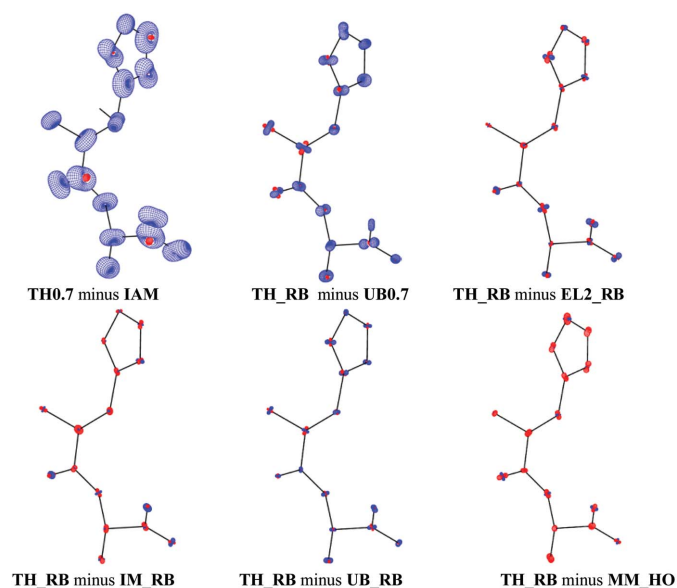**TH_RB minus IM_RB**   **TH_RB minus UB_RB**   **TH_RB minus MM_HO**

**Figure 3**
The *PEANUT* representations of the differences in ADPs of non-H atoms for the HA molecule. A scale of ten was used for the representation of differences in root-mean-square deviation surfaces. An overestimation of the ADPs with respect to the TH0.7 or TH_RB model, respectively, appears in blue.

in magnitude to ADPs from multipolar models. Also, the difference between IAM and TAAM results is more pronounced when refinements are performed against experimental data of low resolution (Volkov *et al.*, 2007; Dittrich *et al.*, 2008). Our study confirms all those findings (see Fig. 3, Fig. 2S in the supplementary material and Table 4 as an example). ADPs from TAAM refinements against high-order data differ in overall size only by about 0.5% from the multipolar MM_HO refinement, whereas IAM refinement against data of the same resolution leads to ADPs larger by about 2.5%. In addition, ADPs from TAAM refinements are less affected by data truncation (to $\sin\theta/\lambda < 0.7$ Å$^{-1}$) than those from IAM ($\sim$3 and $\sim$13% larger ADPs when compared to multipolar refinement results, respectively).

In order to better quantify the analysis, we decided to use the following descriptor (known as the $L_2$ norm in matrix space; Murshudov *et al.*, 1999):

$$
\mathrm{RMSD}[U_{ij}(y)] = \left[ \frac{1}{n} \sum_{i=1}^{n} \frac{1}{3} \left( \left\{ \sum_{i=1}^{3} [U_{ii}(x) - U_{ii}(y)] \right\}^2 \right. \right.
$$
$$
\left. \left. + 2 \left\{ \sum_{i=1; i \neq j}^{3} \sum_{j=1}^{3} [U_{ij}(x) - U_{ij}(y)] \right\}^2 \right) \right]^{1/2},
$$

which helps to measure precisely not only overall differences in magnitude, but also discrepancies in the shape of ADPs between the particular models (Table 5). The obtained values of RMSD($U_{ij}$) indicate that the ADPs from the TAAM refinements against high-resolution data are on average very similar in shape to those from high-order and multipolar refinements, the discrepancies being lower than three standard deviations. The differences are small but systematic; similar differences were found in the previous study (Bąk *et al.*, 2009). The accuracy of the ADPs from TAAM refinement against truncated data is lower: RMSD($U_{ij}$) are two to three times higher than for high-resolution data.

It has been proposed that ADPs from TAAM refinements may serve as a starting point for multipolar refinements (*e.g.* Dittrich, Hübschle *et al.*, 2009; Guillot *et al.*, 2008) instead of ones from high-order refinement. Since ADPs usually correlate strongly with the overall scale factor and some multipole parameters, the results of multipolar refinements depend on the ADPs used, even though compared sets of ADPs are very similar [RMSD($U_{ij}$) < 0.0003 Å$^2$]. This phenomenon is well illustrated when MM_HO and MM_TH refinements are compared. The discrepancies between ADPs obtained from these two refinements usually do not exceed three standard deviations except for the oxygen ADPs, which differ by up to six estimated standard deviations. However, electron density, and properties derived from these models, differ from each other, as we will show in the next paragraphs.

The quality of ADPs may also be examined by the Hirshfeld (1976) rigid-bond test. The results of the test for some models of HA are given in Table 2S in the supplementary material. ADPs from all TAAM models refined against high-resolution data and ADPs from high-order refinement fulfil the test ($z_{A,B}^2 < 10 \times 10^{-4}$ Å$^2$). The lowest values of $z_{A,B}^2$ are found for the

**Table 5**
RMSDs ($U_{ij}$) ($\text{Å}^2$) and $\Delta U_{ii}$ (%) resulting from sets of ADPs obtained from different refinements of HA.

Standard deviations of $U_{ij}$ from the MM_HO refinement do not exceed $1.1 \times 10^{-4}$ $\text{Å}^2$.
$\Delta U_{ii}$ (%) $= (1/n) \sum_{i=1}^{n} \{| \sum_{i=1}^{3} (1/3) U_{ii}(x) - \sum_{i=1}^{3} (1/3) U_{ii}(y)|/[\sum_{i=1}^{3} (1/3) U_{ii}(x)]\} \times 100$.

| | IAM (high resolution) | IAM | EL_RB | EL0.7 | EL2_RB | EL20.7 | IM_RB | IM0.7 | UB_RB | UB0.7 | MM_HO |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $U_{ij}$ from the MM_HO refinement as reference: | | | | | | | | | | | |
| $\Delta U_{ii}$ (%) | 2.6 | 12.7 | 0.6 | 2.6 | 0.4 | 2.4 | 0.4 | 2.8 | 0.8 | 3.3 | |
| RMSD($U_{ii}$) $\times 10^4$ ($\text{Å}^2$) | 25 | 30 | 3 | 9 | 3 | 12 | 2 | 6 | 3 | 8 | |
| $U_{ij}$ from the TH refinement as reference: | | | | | | | | | | | |
| $\Delta U_{ii}$ (%) | 2.1 | 11.8 | 0.2 | 2.2 | 0.3 | 2.0 | 0.7 | 2.4 | 0.4 | 2.8 | 0.4 |
| RMSD($U_{ii}$) $\times 10^4$ ($\text{Å}^2$) | 25 | 30 | 3 | 12 | 4 | 9 | 2 | 6 | 2 | 7 | 4 |

**Table 6**
Averaged linear correlation coefficients (with e.s.d.) calculated for particular pairs of deformation maps (a), (b), (c) and (d) as defined in Fig. 4.

Coefficients have been averaged using the formula $(\sum_{n} r^2/n)^{1/2}$ where $r$ is the correlation coefficient and $n$ is the number of coefficients to be averaged. Underlined coefficients indicate very strong ($r \geq 0.95$) correlations.

| | CR | EL_CR | EL2_CR | IM_CR | UB_CR | MM_TH |
|---|---|---|---|---|---|---|
| CR | 1.00 | | | | | |
| EL_CR | 0.84 (3) | 1.00 | | | | |
| EL2_CR | 0.94 (2) | 0.90 (3) | 1.00 | | | |
| IM_CR | 0.94 (2) | 0.87 (3) | 0.96 (1) | 1.00 | | |
| UB_CR | 0.97 (1) | 0.80 (4) | 0.93 (2) | 0.94 (1) | 1.00 | |
| MM_TH | 0.92 (3) | 0.93 (3) | 0.93 (1) | 0.91 (2) | 0.90 (2) | 1.00 |
| MM_HO | 0.92 (1) | 0.89 (3) | 0.91 (6) | 0.89 (1) | 0.89 (3) | 0.960 (5) |

TH_RB and IM_RB refinement, and the highest for EL_RB. It appears that ADPs from TH and IM refinements against high-resolution data are a better starting point for experimental multipolar refinement than ADPs from high-order refinement.
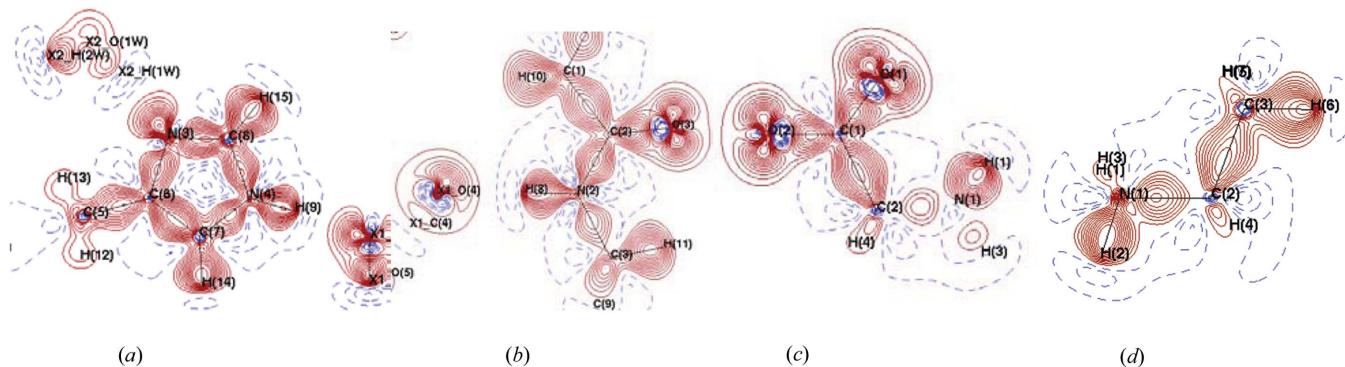
### 3.2. Deformation density

In order to compare electron densities, we focused on deformation densities. Examples of deformation maps are given in Fig. 4. For every pair of the studied models, we computed correlations of the density values in the grid points from successive two-dimensional deformation maps (Fig. 4).

For a given pair of models, we then calculated average values of the four correlation coefficients (see Table 6).

Despite the overall similarity in shape of each deformation density (see Fig. 3S), the differential density maps (see Fig. 5) reveal that most of the models differ significantly from each other. The differential maps correspond very well to the computed correlation coefficients. Only for strongly correlating maps ($r > 0.95$) are the observed differences relatively small, mostly below 0.2 e $\text{Å}^{-3}$. Such strong correlations are observed just between the CR model and the UB_CR model, the EL2_CR and IM_CR models, and the MM_TH and MM_HO models (multipolar parameters of the last two have also been transferred on the CR coordinates).

All TAAM densities differ from the ones from experimental MM models. The differences are largest at the very positions of some nuclei, in the X—H bonds and valence density of oxygen atoms. It appears that the distribution of the density in the X—H bonds in MM models is different to that in any TAAM models (including even experimental database models). This shows that the accuracy of the H-atom description, even if improved by aspherical scattering factors as seen in Fig. 2, is still not satisfactory in X-ray diffraction. The sharp, high peaks in differential electron density at nuclei positions result most probably from Fourier truncation error. In all the databases (and experimental model) multipolar parameters of pseudoatoms were obtained through the fit to the density in the Fourier space. With finite Fourier expansion, rapid changes in electron-density distribution expected in the



(a)  (b)  (c)  (d)

**Figure 4**
Definitions of deformation density maps which were chosen for the correlation analysis, here computed for the CR model as an example: (a) histidine (HA) ring plane; (b) peptide bond plane of HA; (c) plane of the carboxylate group of ALA; (d) main body of ALA. The contour interval is equal to 0.05 e $\text{Å}^{-3}$. Positive values appear as red continuous lines; negative values appear as blue dashed lines.
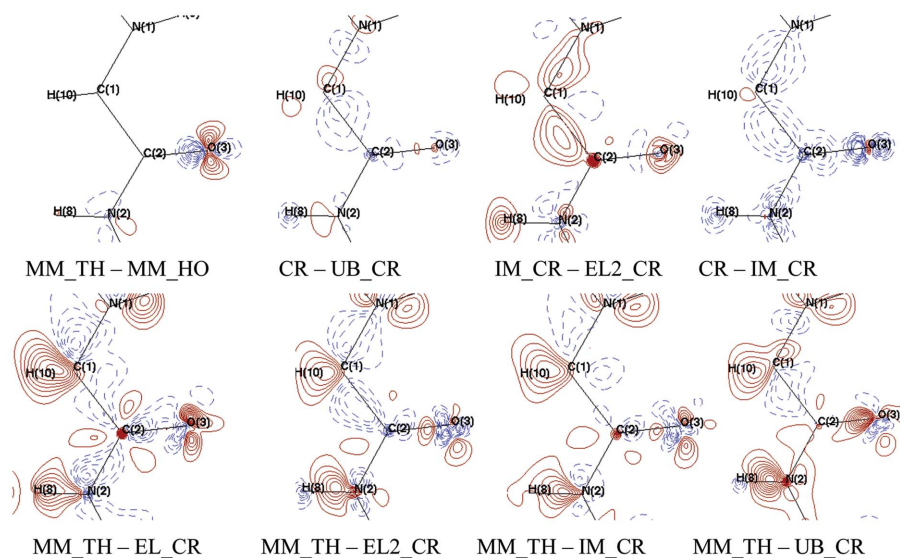
**Figure 5**
Differential density maps computed in the peptide bond plane of HA for selected pairs of models. Contour interval and line styles are as in Fig. 4.

close vicinity of the nucleus cannot be accurately described. Inflexibility of the multipolar model makes the situation even worse.

Valence densities of two experimental MM models differ from each other mostly around oxygen atoms. The differences reach values of 0.35 and $-0.4$ e $\mathrm{\AA}^{-3}$. This is the consequence of using slightly different ADPs as a starting point for the multipole refinement.

The weakest correlations are generally observed for EL_CR. This might be caused by the limited number and quality of the diffraction data sets used for the construction of this first version of the database, or by the definition of the local coordinate systems of atoms (Fig. 4S in the supplementary material). The choice of the local coordinate systems of atoms seems to be of high importance for obtaining an adequate model of molecular electron density from database pseudoatoms. The local geometry of a particular atom, especially valence angles, varies from one molecule to another. On the other hand, pseudoatom densities taken from the database are rigid and, in practice, the orientation of the first axis of the local coordinate system determines which directions all the lobes of deformation density are pointing at. This is visible in Fig. 3S in the supplementary material, where the database model wrongly describes the density of the ring bonds.

## 3.3. Electrostatic potential

Molecular electrostatic potential mapped on an isodensity surface (MEPS) provides a visual method to understand intermolecular association and electrostatic complementarity. Politzer and co-workers (Politzer *et al.*, 2005; Murray *et al.*, 2000) have shown that quantitative information could be derived from MEPSs and used to predict several molecular properties. We decided to check whether such quantitative information can be obtained from database models. Another

interesting question is whether MEPSs derived from experimental databases (EL and EL2 models) will exhibit enhanced polarization due to the influence of crystal field when compared to databases built on isolated-molecule densities (IM and UB models).

MEPSs of all examined models show positive or negative potential values characteristic for particular types of chemical functional groups (see Fig. 6 and Fig. 5S in the supplementary material). To quantitatively compare MEPSs, we computed the following isosurface quantities ($\rho = 0.00675$ e $\mathrm{\AA}^{-3}$; Murray *et al.*, 2000): positive and negative average potential values ($V_{av}^{+}$, $V_{av}^{-}$), average deviation from the overall potential $\pi$, positive and negative variances of the electrostatic potential on the surface ($\sigma^{2+}$, $\sigma^{2-}$). Detailed definitions of these quantities are given in the supplementary material. The results are summarized in Table 7.

The obtained values vary greatly among the models for all these quantities. The discrepancies between the different models of the same molecule are of the same magnitude as the differences between results obtained from the same model from two different molecules. In general, the $\pi$ values are large and similar to those calculated by Murray *et al.* (2000) for amino acid zwitterions, and this indicates the high local polarity of the molecules under study. The average values of the positive and negative surface potentials ($V_{av}^{+}$, $V_{av}^{-}$) from our study are closer to the values obtained for zwitterionic ($\sim$25 and $-45$ kcal mol$^{-1}$) than for the non-ionic form ($\sim$15 and $-20$ kcal mol$^{-1}$) of amino acids (Murray *et al.*, 2000). It seems that the analysed models and quantities computed from them can be used to distinguish between very different groups of compounds. The precision is, however, too small to make a quantitative distinction between two compounds of the same kind.

It is also difficult to find any systematic effects due to polarization when models based on periodic densities
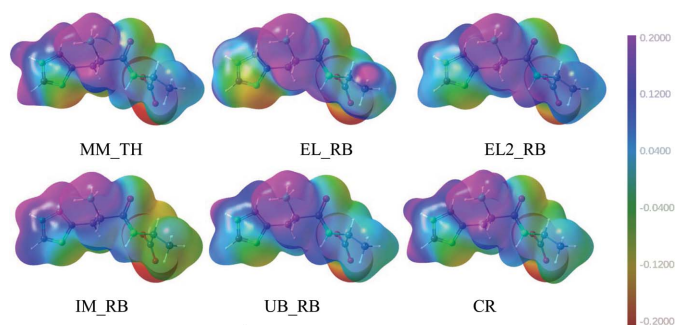


**Figure 6**
Electrostatic potential (e $\mathrm{\AA}^{-1}$) of HA mapped on an electron-density isosurface at $\rho = 0.0135$ e $\mathrm{\AA}^{-3}$.

**Table 7**
Computed MEPS surface quantities.

Units: $V_{av}^+$, $V_{av}^-$ and $\pi$ are in kcal mol$^{-1}$; $\sigma^{2+}$ and $\sigma^{2-}$ are in (kcal mol$^{-1}$)$^2$.

| | $V_{av}^+$ | | $V_{av}^-$ | | $\pi$ | | $\sigma^{2+} \times 10$ | | $\sigma^{2-} \times 10$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | HA | ALA | HA | ALA | HA | ALA | HA | ALA | HA | ALA |
| CR | 39 | 35 | −36 | −43 | 37 | 38 | 202 | 140 | 205 | 265 |
| EL_RB | 45 | 49 | −43 | −55 | 44 | 52 | 331 | 331 | 265 | 375 |
| EL2_RB | 30 | 37 | −35 | −45 | 32 | 40 | 115 | 143 | 203 | 276 |
| IM_RB | 50 | 55 | −43 | −42 | 46 | 48 | 364 | 408 | 254 | 265 |
| UB_RB | 33 | 47 | −34 | −50 | 34 | 48 | 155 | 265 | 201 | 375 |
| MM_TH | 52 | 31 | −45 | −31 | 49 | 31 | 364 | 110 | 331 | 143 |
| MM_HO | 52 | 33 | −47 | −35 | 50 | 34 | 375 | 132 | 353 | 165 |

(EL_RB, EL2_RB, MM_HO, MM_TH and CR) are compared to the ones derived from isolated molecules (IM_RB and UB_RB).

### 3.4. Molecular dipole moment

This is another important molecular property. To reliably judge the quality of dipole moments obtained from the databases, we extended our analysis to two additional crystal structures (GH and HIS) and included additional theoretical models (MP2, DZP, D95 and 6-31Gdp). In particular, D95 and 6-31Gdp were treated as the most appropriate reference models to validate the outcome from the Invariom and UBDB databases, respectively. The results from the ELMAM and ELMAM2 databases were, in the first instance, compared to experimental multipolar models. Molecular dipole moments will obviously depend on atomic positions. To make the results independent of the geometry, we used the CR coordinates for all theoretical calculations and some of the database models (models with the _CR suffix). Estimates from TAAM refinements with constrained $X$—H bonds (models with the _RB suffix) were compared to the results from experimental multipolar models. Properties of the MP2, DZP, D95 and 6-31Gdp models were computed directly from wavefunctions. Properties of other models were estimated from multipolar models of charge density.

For all molecules, the magnitude and orientation of dipole moments are dominated by the positions of the carboxylate and ammonium groups, as one may expect for zwitterionic amino acids. Magnitudes of dipole moments reproduced by database models differ from each other and from experimental and theoretical models by 2–3 D on average (see Table 8). The directions of dipole moments derived from the databases are even more dissimilar (see Fig. 7 and Fig. 6S in the supplementary material). Directions estimated from the theoretical databases (IM and UB) deviate from their reference theoretical models by up to 15°, whereas dipole model vectors computed directly from MP2, DZP, D95 and 6-31Gdp models have nearly the same directions. Experimental databases (EL and EL2 models) seem to deviate from experimental multipolar models (MM_TH and MM_HO) by up to 30°, but to draw more general conclusions about dipole-moment dissimilarity more data points are needed.

It is worth noting here that such small differences in dipole-moment vectors may lead to large differences in electrostatic interaction energy. If electrostatic interaction energy between two HA molecules is represented only by molecular dipole–dipole interaction, for molecules related by $1 - x, \frac{1}{2} + y, -z$ symmetry, the energy difference equals about 45 and 23 kJ mol$^{-1}$ in the case of dipoles taken from EL_CR and MM_TH, and from EL2_CR and UB_CR models, respectively.

It seems that systematic enhancement of molecular dipole moments due to the crystal field is not visible in the CR, EL and EL2 models. This is in agreement with findings by Spackman *et al.* (2007) who showed that for zwitterionic molecules, which are already highly polarized, the enhancement is very small.

### 3.5. Coulombic intermolecular interaction energy (Ee)

This is a property that is very sensitive to differences in electron-density models. We decided to check whether the Ees estimated from database models can be used in quantitative analyses.

It has been found already that UBDB, when combined with the EPMM method (UBDB + EPMM), reproduces quite well the electrostatic interaction energies resulting from exact theoretical calculations for dimers represented as a superposition of isolated molecular densities. The RMSD between the UBDB + EPMM method and the ADF/BLYP/TZP results equals 8 kJ mol$^{-1}$ for six $\alpha$-glycine, two *N*-acetylglycine and three dimers of L-(+)-lactic acid (Volkov, Koritsanszky & Coppens, 2004). For nucleic base dimers (Czyżnikowska *et al.*, 2010) the discrepancy is slightly larger: the RMSD between the UBDB + EPMM and high-level (MP2/aug-cc-pVDZ) *ab initio* methods is 16 kJ mol$^{-1}$ for the set comprising more than 200 Watson–Crick base pairs.

In the current study, we calculated Ees by application of the EPMM method to densities obtained either from one of the four database models, from experimental multipolar models
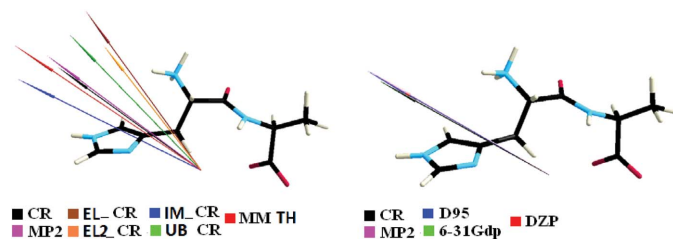
**Table 8**
Molecular dipole-moment magnitudes (D).

| | MP2 | DZP | D95 | 6-31Gdp | CR | EL_CR (EL_RB) | EL2_CR (EL2_RB) | IM_CR (IM_RB) | UB_CR (UB_RB) | MM_TH (MM_HO) |
|---|---|---|---|---|---|---|---|---|---|---|
| HA | 24 | 24 | 24 | 23 | 21 | 25 (22) | 20 (17) | 27 (26) | 23 (21) | 28 (29) |
| ALA | 12 | 11 | 11 | 11 | 11 | 14 (14) | 11 (11) | 12 (12) | 12 (12) | 8 (9) |
| GH | 24 | 22 | 24 | 22 | 19 | 21 | 21 | 25 | 22 | |
| HIS | 15 | 14 | 15 | 14 | 14 | 16 | 15 | 15 | 15 | |

# research papers



**Figure 7**
The *MoleCoolQt* representations of the molecular dipole-moment vectors obtained for HA from the different models. The vector colours indicate the particular model.

or from the CR model. We also computed Ees directly from theoretical molecular densities (exact evaluation *via* numerical quadrature). Theoretical densities for isolated molecules have been computed at several different levels of theory (MP2/aug-cc-pVDZ; BLYP/DZP; B3LYP/6-31Gdp; B3LYP/D95++3df,3pd). A broad range of attractive interactions was taken into account; Ee values are in the range of $-1$ to $-222$ kJ mol$^{-1}$, according to the MP2 results (see Table 3S in the supplementary material).

Since differences in the geometry of molecules (and dimers) may influence the values of the Coulombic energy of interactions between them, we first checked how models with the _RB suffix differ from the _CR models. It appears that the difference is small – the RMSD is 6 kJ mol$^{-1}$. Therefore, our further analysis of Ees was based on the models with the _CR suffix only. Two experimental models, MM_HO and MM_TH, differ from each other by 8 kJ mol$^{-1}$ on average. For the further analysis we chose to use MM_TH models because these models seem to be more correct, as Hirshfeld's rigid-bond test has shown.

The experimental refinement MM_TH, the theoretical pseudoatom model (CR), and possibly the EL and EL2 models are supposed to bring some information about the influence of the nearest molecules on the studied Coulombic interaction energy (*e.g.* due to the polarization effect). However, Ees from the experimental MM_TH model differ significantly from the CR, EL and EL2 models (see Fig. 8 or Table 4S in the supplementary material), with RMSDs equal to 37, 53 and 55 kJ mol$^{-1}$, respectively. The Ees estimated from CR, EL and EL2 models are closer to the Ees resulting from theoretical calculation for isolated molecules (MP2, for example), than from MM_TH. The large discrepancy between the energy values obtained from MM_TH and the other models might be explained in two different ways. Firstly, the polarization effect is seen only in the experimental refinement results, and this is not visible either in EL, EL2 or CR model results. Secondly, the polarization effect in the studied crystals is small and the differences between models are due to experimental or model errors, or theory limitations. Further studies are needed to resolve this issue.

Therefore, we compared results obtained from all databases to the Ees computed from high-level theoretical methods to see to what extent database models are able to reproduce electrostatic energies of interactions between isolated mole-

cules. Our analysis was based mainly on interactions between 17 different dimers of dipeptides (HA, GH) or amino acids (HIS, ALA). Sometimes, interactions of dipeptides with water molecules were also taken into account, and the number of dimers was extended to 24. We decided to include in the analysis Ees computed from the DZP model, because in the DZP and CR models the same level of theory was used. We also studied the D95 and 6-31Gdp models to provide reference Ee values for comparison with the results of the Invariom and UBDB databases (D95++3df3pd and 6-31Gdp basis sets were used in construction of the theoretical databases). As a final test we used Ees computed from the MP2/aug-cc-pVDZ molecular densities (the MP2 model). The MP2 model is known to give good estimates of electrostatic energy (Volkov, King & Coppens, 2006), so it was used by us to judge how reliable, in general, are the results obtained from the other studied models.

The ELMAM, ELMAM2 and Invariom databases reproduce the theoretical Ees with RMSDs in the range 26–33 kJ mol$^{-1}$ or 22–24 kJ mol$^{-1}$ when compared to the results of MP2 or DZP calculations, respectively (see Table 9). The largest individual discrepancy in Ees between the DZP and databases results is as high as 70 kJ mol$^{-1}$ (see Fig. 9). The UB model leads to smaller discrepancies – the RMSDs are 15 and 9 kJ mol$^{-1}$ when compared to MP2 or DZP results, respectively. It seems that an RMSD equal to 15 kJ mol$^{-1}$ is too high for the method when fine electrostatic energy results are important, but also pure quantum mechanical methods may lead to discrepancies of a similar level depending on the method and basis set used.

The Ee values obtained from MP2 and D95 differ systematically from the results of DZP and 6-31Gdp, on average by about 14 kJ mol$^{-1}$ (see Table 10). This agrees well with findings by Volkov, King & Coppens (2006) that addition of diffuse functions into the basis set lowers values of Ee. The Ees estimated from all database models, even IM, are closer to the results of the DZP or 6-31Gdp models, in which
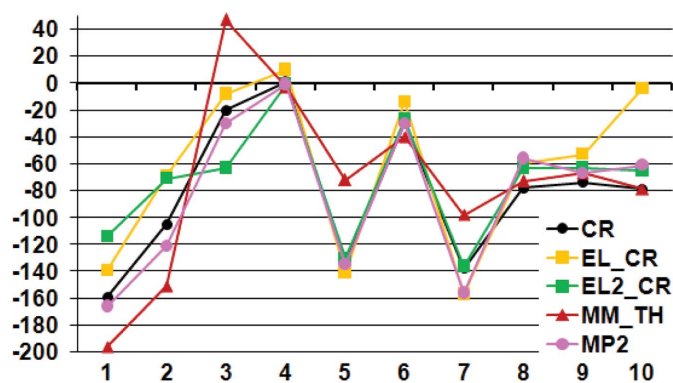


**Figure 8**
Coulombic interaction energies (in kJ mol$^{-1}$) in HA and ALA dimers estimated from the different models. Symmetry codes defining a particular dimer are as follows for HA–HA: (1) $1-x, \frac{1}{2}+y, -z$; (2) $2-x, \frac{1}{2}+y, -z$; (3) $1+x, y, z$; (4) $1+x, y, 1+z$; and for ALA–ALA: (5) $x, y, 1+z$; (6) $\frac{3}{2}-x, -y, \frac{1}{2}+z$; (7) $\frac{1}{2}+x, \frac{1}{2}-y, 1-z$; HA–H$_2$O: (8) $2-x, \frac{1}{2}+y, 1-z$; (9) $1+x, 1+y, -1+z$; (10) $1+x, 1+y, z$.

**Table 9**
RMSDs in Ees (kJ mol$^{-1}$) between the results from different models.

Number of dimers = 17. Values in italics are RMSD values computed for 24 dimers (including interactions with water molecules).

|         | CR | EL_CR | EL2_CR | IM_CR | UB_CR | 6-31Gdp | DZP | D95 |
|---------|-----|-------|--------|-------|-------|---------|-----|-----|
| 6-31Gdp | 14 | 26 | 26 | 22 | 12 |   |   |   |
| DZP | 12 | 26 | 22 | 24 | 9 | 5 |   |   |
| D95 | 19 | 31 | 25 | 30 | 15 | 13 | 13 |   |
| MP2 | 19 *18* | 33 *33* | 26 *23* | 30 *27* | 17 *15* | 14 | 14 | 3 |

**Table 10**
The linear correlation coefficients for Ees.

Underlined coefficients indicate very strong ($r \geq 0.95$) correlation. Number of dimers = 17.

|        | MP2 | D95 | DZP | 6-31Gdp | EL_CR | EL2_CR | IM__CR | UB_CR |
|--------|------|------|------|---------|-------|--------|--------|-------|
| EL_CR | 0.92 | 0.91 | 0.92 | 0.92 |   |   |   |   |
| EL2_CR | 0.93 | 0.94 | 0.93 | 0.91 | 0.82 |   |   |   |
| IM_CR | 0.92 | 0.94 | 0.93 | 0.94 | 0.86 | 0.79 | 1.00 |   |
| UB_CR | 0.99 | 0.99 | 0.99 | 0.98 | 0.95 | 0.93 | 0.92 |   |
| CR | <u>0.97</u> | <u>0.98</u> | <u>0.98</u> | <u>0.97</u> | 0.89 | 0.94 | 0.91 | <u>0.98</u> |

diffuse functions were not included, than to Ees from D95 and MP2.

All database models, as well as theoretical DZP and 6-31Gdp, underestimate the Ees of strong interactions. In spite of large differences among Ees obtained from theoretical models, the results still correlate highly with each other ($r > 0.99$). The Ees from the UBDB models also show a high correlation ($r = 0.99$) with the theoretical results (see Table 10). The Ees from experimental multipolar models, MM_TH or MM_HO, do not correlate either with periodic theoretical models ($r = 0.84$), or with the EL and EL2 models ($r = 0.68$ and 0.53, respectively; see Table 5S in the supplementary material). The highest correlation ($r = 0.93$) with experimental results was found for the IM model.

## 4. Conclusions

The refinements utilizing multipolar parameters stored in the ELMAM, Invariom, UBDB and ELMAM2 databases (TAAM refinements) reproduce very well the geometries optimized in theoretical periodic calculations, especially $X-H$ bond lengths (RMSDs < 0.04 Å). There are no significant differences in geometries resulting from individual databases. The quality of coordinates and thermal displacement parameters obtained from TAAM refinements is significantly better compared with that from IAM refinement. TAAM refinement based on any of the studied databases could be routinely used to improve the quality of the refined structures.
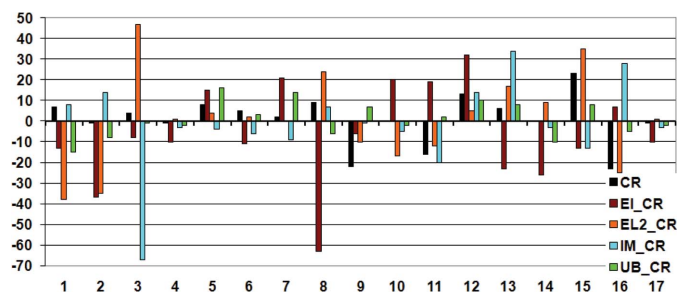
**Figure 9**
The visualization of the differences in the Ee values (kJ mol$^{-1}$) between DZP and the given model. Symmetry codes defining dimers (1)–(7) are the same as in Fig. 7 and dimers (8)–(17) are as follows: HIS–HIS: (8) $1 + x, y, z$; (9) $-x, 1 - y, -z$; (10) $x, \frac{1}{2} - y, \frac{1}{2} + z$; (11) $-x, \frac{1}{2} + y, \frac{1}{2} - z$; (12) $1 - x, -\frac{1}{2} + y, \frac{1}{2} - z$; (13) $1 - x, 1 - y, 1 - z$; (14) $-1 + x, \frac{1}{2} - y, \frac{1}{2} + z$; GH–GH: (15) $-1 + x, y, z$; (16) $1 + x, 1 + y, z$; (17) $1 - x, \frac{1}{2} + y, 1 - z$.

The differences between ADPs obtained from the TAAM refinements and experimental multipolar refinement are small, about 0.5%, which is equivalent to three standard deviations of $U_{ij}$. Such ADPs are an equally good (or sometimes better) starting point for the experimental multipolar refinement as ADPs from high-order refinement. However, the uncertainty of ADPs obtained from the analysed TAAM refinements or high-order refinement can lead to large differences in refined scale factor, deformation densities and, consequently, in electrostatic properties derived from experimental multipolar refinements.

The deformation maps differ significantly among databases, experimental and theoretical periodic models. The differences are seen mostly at the very positions of some nuclei, in the $X-$H bonds and valence density of oxygen atoms. Generally, the weakest correlations of deformation maps, with all other models, were found for the ELMAM database.

The MEPSs calculated from the database models can be used in qualitative study, to make a distinction between different groups of compounds. The quantitative information, as proposed by the Politzer group (Politzer *et al.*, 2005; Murray *et al.*, 2000), cannot be obtained from MEPSs estimated from database models, because of the low precision of such MEPSs.

Database models reproduce dipole-moment magnitudes of isolated molecules with RMSDs equal to 2–3 D. The directions of dipole moments estimated from databases vary within 30°. Such differences in dipole-moment vectors lead to large differences in electrostatic interaction energy.

The Ees calculated from experimental multipolar models differ substantially from the results of the ELMAM, ELMAM2 and periodic theoretical models, as well as theoretical models of isolated molecules. Our results did not help to explain whether the polarization effect is seen only in the experimental refinement results, or whether the polarization effect in the studied crystals (zwitterionic molecules) is small and the differences are due to experimental or model errors. To answer the question of whether the ELMAM and/or ELMAM2 databases include some information about the influence of crystal packing, further investigations are needed.

Ees computed from all database models are closer to the results of the BLYP/DZP theoretical model of isolated molecules, than to the ones obtained from periodic systems. The smallest differences in the Ee values were found for the UBDB database, with RMSDs of Ees equal to 9 and 15 kJ mol$^{-1}$, when compared to BLYP/DZP and MP2/aug-cc-pVDZ, respectively. The RMSDs of Ees estimated from other

TAAM models are about 22–33 kJ mol$^{-1}$, when compared to theoretical models. In the case of each database, the Ee values correlate with almost the same strength with the theoretical results, independently from the basis set and level of theory used. The highest correlations were found for the UBDB results ($r = 0.99$ for all theoretical models).

## References

Abramov, Y. A., Volkov, A., Wu, G. & Coppens, P. (2000). *J. Phys. Chem. B*, **104**, 2183–2188.

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* B**66**, 380–386.

Bąk, J. M., Dominiak, P. M., Wilson, C. C. & Woźniak, K. (2009). *Acta Cryst.* A**65**, 490–500.

Becke, A. D. J. (1993). *Chem. Phys.* **98**, 5648–5652.

Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* B**47**, 789–797.

Cheng, F., Sun, H., Yong, Z., Dushyant, M. & Okifield, E. (2005). *J. Am. Chem. Soc.* **127**, 12544–12554.

Clementi, E. & Raimondi, D. L. (1963). *J. Chem. Phys.* **38**, 2686–2692.

Clementi, E. & Roetti, C. (1974). *At. Data Nucl. Data Tables*, **14**, 177–478.

Coppens, P., Abraamov, Y., Carducci, M., Korjov, B., Novozhilova, I., Alhambra, C. & Pressprich, M. R. (1999). *J. Am. Chem. Soc.* **121**, 2585–2593.

Czyżnikowska, Ż, Góra, R. W., Zaleśny, R., Lipkowski, P., Jarzembska, K. M., Dominiak, P. M. & Leszczyński, J. (2010). *J. Phys. Chem. B*, **114**, 9629–9644.

Destro, R. & Marsh, R. E. (1988). *J. Phys. Chem.* **92**, 966–973.

Destro, R., Soave, R. & Barzaghi, M. (2008). *J. Phys. Chem. B*, **112**, 5163–5174.

Dittrich, B., Bond, C. S., Kalinowski, R., Spackman, M. A. & Jayatilaka, D. (2010). *CrystEngComm*, doi:10.1039/c001072c.

Dittrich, B., Hübschle, C. B., Holstein, J. J. & Fabbiani, F. P. A. (2009). *J. Appl. Cryst.* **42**, 1110–1121.

Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* D**62**, 1325–1335.

Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst.* A**61**, 314–320.

Dittrich, B., Koritsanszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed.* **43**, 2718–2721.

Dittrich, B., Koritsanszky, T., Volkov, A., Mebs, S. & Luger, P. (2007). *Angew. Chem. Int. Ed.* **46**, 2935–2938.

Dittrich, B., McKinnon, J. J. & Warren, J. E. (2008). *Acta Cryst.* B**64**, 750–759.

Dittrich, B., Strumpel, M., Schäfer, M., Spackman, M. A. & Koritsánszky, T. (2006). *Acta Cryst.* A**62**, 217–223.

Dittrich, B., Warren, J. E., Fabbiani, F. P. A., Morgenroth, W. & Corry, B. (2009). *Phys. Chem. Chem. Phys.* **11**, 2601–2609.

Dittrich, B., Weber, M., Kalinowski, R., Grabowsky, S., Hübschle, C. B. & Luger, P. (2009). *Acta Cryst.* B**65**, 749–756.

Domagała, S. & Jelsch, C. (2008). *J. Appl. Cryst.* **41**, 1140–1149.

Dominiak, P. M., Volkov, A., Dominiak, A. P., Jarzembska, K. N. & Coppens, P. (2009). *Acta Cryst.* D**65**, 485–499.

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., Orlandol, R., Zicovich-Wilson, C. M., Pascale, F., Civalleri, B., Doll, K., Harrison, N. M., Bush, I. J., D'Arco, Ph. & Llunell, M. (2008). *CRYSTAL06* 1.0. Version 1_0_2. University of Turin, Italy.

Dunning, T. H. (1970). *J. Chem. Phys.* **53**, 2823–2833.

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.* (2004). *GAUSSIAN03*. Gaussian Inc., Pittsburgh, Pennsylvania, USA.

Grabowsky, S., Pfeuffer, T., Morgenroth, W., Paulmann, C., Schirmeister, T. & Luger, P. (2008). *Org. Biomol. Chem.* **6**, 2295–2307.

Guillot, B., Jelsch, C., Podjarny, A. & Lecomte, C. (2008). *Acta Cryst.* D**64**, 567–588.

Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* A**34**, 909–921.

Hirshfeld, F. L. (1976). *Acta Cryst.* A**32**, 239–244.

Holstein, J. J., Luger, P., Kalinowski, R., Mebs, S., Paulman, C. & Dittrich, B. (2010). *Acta Cryst.* B**66**, 568–577.

Hoser, A. A., Dominiak, P. M. & Woźniak, K. (2009). *Acta Cryst.* A**65**, 300–311.

Hübschle, C. B. (2010). *MoleCoolQt – a molecule viewer for charge density related science*, http://www.molecoolqt.de. Georg-August-University Göttingen, Germany.

Hübschle, C. B. & Luger, P. (2006). *J. Appl. Cryst.* **39**, 901–904.

Hübschle, C. B., Luger, P. & Dittrich, B. (2007). *J. Appl. Cryst.* **40**, 623–627.

Hummel, W., Hauser, J. & Bürgi, H. B. (1990). *J. Mol. Graph.* **8**, 214–218.

Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.

Jelsch, C., Pichon-Pesme, V., Lecomte, C. & Aubry, A. (1998). *Acta Cryst.* D**54**, 1306–1318.

Jelsch, C., Teeter, T. T., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.

Johnas, S. K. J., Dittrich, B., Meents, A., Messerschmidt, M. & Weckert, E. F. (2009). *Acta Cryst.* D**65**, 284–293.

Koritsanszky, T. S. & Coppens, P. (2001). *Chem. Rev.* **101**, 1583–1627.

Koritsanszky, T., Volkov, A. & Coppens, P. (2002). *Acta Cryst.* A**58**, 464–472.

Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.

Li, X., Volkov, A. V., Szalewicz, K. & Coppens, P. (2006). *Acta Cryst.* D**62**, 639–647.

Liebschner, D., Elias, M., Moniot, S., Fournier, B., Scott, K., Jelsch, C., Guillot, B., Lecomte, C. & Chabriere, E. (2009). *J. Am. Chem. Soc.* **131**, 7879–7886.

Madsen, A. Ø. (2006). *J. Appl. Cryst.* **39**, 757–758.

Madsen, A. Ø., Sørensen, H. O., Flensburg, C., Stewart, R. F. & Larsen, S. (2004). *Acta Cryst.* A**60**, 550–561.

Munshi, P., Jelsch, C., Hathwar, V. R. & Row, T. N. G. (2010). *Cryst. Growth Des.* **10**, 1516–1526.

Murray, J. S., Peralta-Inga, Z. & Politzer, P. (2000). *Int. J. Quant. Chem.* **80**, 1216–1223.

Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.

Muzet, N., Guillot, B., Jelsch, C. & Lecomte, C. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 8742–8747.

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.

Politzer, P., Ma, Y., Lane, P. & Concha, M. C. (2005). *Int. J. Quant. Chem.* **105**, 341–347.

Popelier, P. (2005). *Struct. Bond.* **115**, 1–56.

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Spackman, M. A. & Byrom, P. G. (1996). *Acta Cryst.* B**52**, 1023–1035.

Spackman, M. A. & Mitchell, A. S. (2001). *Phys. Chem. Chem. Phys.* **3**, 1518–1523.

Spackman, M. A., Munshi, P. & Dittrich, B. (2007). *Chem. Phys. Chem.* **8**, 2051–2063.

Steiner, Th. (1996). *Acta Cryst.* C**52**, 2554–2556.

Volkov, A., Abramov, Y. A. & Coppens, P. (2001). *Acta Cryst.* A**57**, 272–282.

Volkov, A., King, H. F. & Coppens, P. (2006). *J. Chem. Theory Comput.* **2**, 81–89.

Volkov, A. V., Koritsanszky, T. & Coppens, P. (2004). *Chem. Phys. Lett.* **391**, 170–175.

Volkov, A., Koritsanszky, T., Li, X. & Coppens, P. (2004). *Acta Cryst.* A**60**, 638–639.

Volkov, A., Li, X., Koritsanszky, T. S. & Coppens, P. J. (2004). *Phys. Chem. A*, **108**, 4283–4300.

Volkov, A., Macchi, P., Farrugia, L. J., Gatti, C., Mallinson, P., Richter, T. & Koritsanszky, T. (2006). *XD2006*. Middle Tennessee State University, USA, Università di Milano and CNR-ISTM Milano, Italy, University of Glasgow, Scotland, State University of New York at Buffalo, USA, and Freie Universität Berlin, Germany.

Volkov, A., Messerschmidt, M. & Coppens, P. (2007). *Acta Cryst.* D**63**, 160–170.

Zarychta, B., Pichon-Pesme, V., Guillot, B., Lecomte, C. & Jelsch, C. (2007). *Acta Cryst.* A**63**, 108–125.