



New refinement strategies for a pseudoatom databank – toward rapid electrostatic interaction energy estimations

**Sławomir Antonii Bojarowski, Barbara Gruza, Damian Trzybiński,
Radosław Kamiński, Anna Agnieszka Hoser, Prashant Kumar, Krzysztof
Woźniak and Paulina Maria Dominiak**

Acta Cryst. (2022). **B78**, 823–834



IUCr Journals

CRYSTALLOGRAPHY JOURNALS ONLINE

Author(s) of this article 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 <https://journals.iucr.org/services/authorrights.html>



New refinement strategies for a pseudoatom databank – toward rapid electrostatic interaction energy estimations

Sławomir Antonii Bojarowski,^a Barbara Gruza,^a Damian Trzybiński,^a
Radosław Kamiński,^b Anna Agnieszka Hoser,^a Prashant Kumar,^a
Krzysztof Woźniak^a and Paulina Maria Dominiak^{a*}

Received 11 February 2022

Accepted 3 September 2022

Edited by P. Macchi, Politecnico di Milano, Italy

Keywords: quantum crystallography; TAAM refinement; multipole model refinement; charge density; electrostatic interaction energy.

CCDC references: 2205347; 2205348; 2205349

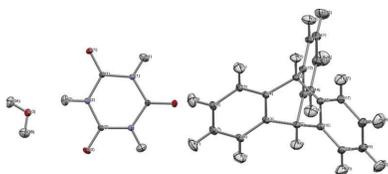
Supporting information: this article has supporting information at journals.iucr.org/b

^aBiological and Chemical Research Center, Department of Chemistry, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland, and ^bDepartment of Chemistry, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland.
*Correspondence e-mail: pdomin@chem.uw.edu.pl

Pseudoatom databanks, collections of parameters from the multipole model of electron densities for various atom types, are used to replace the Independent Atom Model with the more accurate Transferable Aspherical Atom Model (TAAM) in crystal structure refinements. The databanks are also employed to reconstruct the electron density of a molecule, crystal or biomacromolecular complex in a fast yet accurate way and compute various properties such as the energy of electrostatic interactions, for example. A even faster but similarly accurate model for estimations of electrostatic energy exists called aug-PROmol [Bojarowski, Kumar & Dominiak (2016). *ChemPhysChem*, **17**, 2455–2460]. A model analogous to aug-PROmol cannot be built from the current pseudoatom databanks, as they perform badly when truncated to the monopole level. Here, new strategies for multipole model refinements were sought, leading to better parametrization at the monopole level. This would allow the creation of a pseudoatom databank in a single route of model parametrization, which would be suitable for both crystal structure refinement and rapid electrostatic energy calculations. Here it is shown that the cumulative approach to multipole model refinements, as opposed to simultaneous or iterative refinements of all multipole model parameters (P_v , κ , P_{lm} , κ'), leads to substantially different models of electron density. Cumulative refinement of two blocks of parameters, the first with P_v and κ and then the second with P_{lm} and κ' , leads to the $P_v\kappa|P_{lm}\kappa'$ model having promising properties. The $P_v\kappa|P_{lm}\kappa'$ model is as good as the University at Buffalo DataBank (UBDB) in X-ray structure TAAM refinements and electrostatic energy estimations, especially for less polar molecules. When truncated to the monopole level, the $P_v\kappa$ model has a chance to replace aug-PROmol in fast yet accurate electrostatics energy calculations, although some improvements in κ parametrization for polar functional groups are still needed. The $P_v\kappa$ model is also a source of point charges which behave similarly to restrained electrostatic potential (RESP) charges in electrostatic interaction energy estimations.

1. Introduction

X-rays are scattered mostly by electrons. Thus, X-ray diffraction experiments provide information on electron densities, $\rho(\mathbf{r})$, in the solid state. Since the physical properties of molecules are functionals of their electron densities, accurate knowledge of $\rho(\mathbf{r})$ is crucial for a wide variety of chemical, physical and even biological studies. Parallel to the progress in instrumentation, there has been a great development in theoretical approaches to enhance the quality and the type of information to be extracted from X-ray diffraction data. As a



result, over the recent decades, more and more complex models of electron densities have been established.

One of the very first and the most robust model is the independent atom model (IAM). In IAM the electron density of the molecule is approximated in terms of a promolecule – superposition of spherical densities of isolated atoms ρ_{atom}^0 centered at \mathbf{R}_{atom} :

$$\rho_{\text{IAM}}(\mathbf{r}) = \sum_{\text{atom}} \rho_{\text{atom}}^0(\mathbf{r} - \mathbf{R}_{\text{atom}}) \quad (1)$$

The model follows the fundamental concept of chemistry: partitioning the total electron density into its atomic contributions. Even now the IAM is very successful, used in the majority of classical structural crystallography.

More advanced electron density models are also in use (De Bruyne & Gillet, 2020; Hansen & Coppens, 1978; Jayatilaka & Dittrich, 2008; Jayatilaka & Grimwood, 2001; Lübben *et al.*, 2019; Malaspina *et al.*, 2019; Stewart, 1976; Tanaka, 2018). The most widely known is the Hansen–Coppens multipole model (HCMM) (Hansen & Coppens, 1978). In this model atoms are no longer treated as spherical and charge transfer between atoms is allowed. The electron density of an atom (called pseudoatom) is defined in the HCMM as follows:

$$\rho_{\text{atom}}(\mathbf{r}) = P_c \rho_{\text{core}}(r) + P_v \kappa^3 \rho_{\text{valence}}(\kappa r) + \sum_{l=0}^{l_{\text{max}}} \sum_{m=-l}^l \kappa_l'^3 R_l(\kappa_l' r) P_{lm} d_{lm}(\theta, \varphi), \quad (2)$$

where the $\rho_{\text{core}}(r)$ and $\rho_{\text{valence}}(r)$ are the spherical electron densities of core and valence electrons, respectively, precomputed for isolated atoms and scaled to one electron. The third part of the model, built from Slater-type radial functions $R_l(r)$ and spherical harmonics $d_{lm}(\theta, \varphi)$, describes deformations of valence electron densities. P_c , P_v and P_{lm} are electron population parameters, and κ and κ' are expansion/contraction parameters. In standard applications of HCMM the following settings are used: (a) the P_v , κ , P_{lm} and κ' parameters are refined against diffraction data, (b) P_c is kept frozen at formal values characteristic for particular chemical elements, (c) the monopole term ($l=0$) in the deformation part is not taken into account, thus P_v and κ parameters are the only ones responsible for the modeling monopole level of electron density expansion, and (d) κ' parameters for particular levels of multipolar expansion (from $l=1$ to $l=4$) are kept constrained to each other thus effectively there is only one κ' parameter per pseudoatom. The diffraction data must be of high quality and subatomic resolution to allow for stable refinement of fine electron density features, proper deconvolution of electron density from atomic motions, and to maintain a statistically correct ratio of the number of reflections per refined parameter. Only a small part of the experimental diffraction data fulfills such criteria.

The parameters of HCMM have specific values for atoms with a similar chemical environment (Brock *et al.*, 1991). This observation made it possible to build pseudoatom databanks (Zarychta *et al.*, 2007; Domagała *et al.*, 2012; Nassour *et al.*, 2017; Dittrich *et al.*, 2004, 2005, 2006, 2013; Volkov *et al.*, 2007;

Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012; Kumar *et al.*, 2019). The databanks store HCMM pseudoatom parameters derived from the multipole model refinements against experimental X-ray diffraction data (Zarychta *et al.*, 2007; Domagała *et al.*, 2012; Nassour *et al.*, 2017) or theoretical simulations (Dittrich *et al.*, 2004, 2005, 2006, 2013; Volkov *et al.*, 2007; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012; Kumar *et al.*, 2019). Theoretical simulations allowed the parametrization of HCMM for a much wider range of different molecules, overcoming the limited access to high-quality experimental data of subatomic resolution.

Pseudoatom databanks have two major areas of application. They are used to replace IAM with a more accurate so-called transferable aspherical atom model (TAAM) in crystal structure refinements (Jha *et al.*, 2020; Gruza *et al.*, 2020; Dittrich *et al.*, 2013; Dadda *et al.*, 2012; Bąk *et al.*, 2011, Jarzemska *et al.*, 2012; Durka *et al.*, 2013). TAAM is parametrized with values transferred from a pseudoatom databank and allows the refinement of a crystal structure with the same number of parameters as IAM, yet achieving a better fit of the model to the data and more accurate structure. The databanks may also be used to reconstruct the electron density of a molecule, crystal or biomacromolecular complex in a fast yet accurate way and compute various properties from it, such as electrostatic potential, electric field gradient, topological bonding descriptors or energy of electrostatic interactions (Holstein *et al.*, 2012; Malińska *et al.*, 2014; Mandal *et al.*, 2020; Zarychta *et al.*, 2015, Jarzemska *et al.*, 2017; Mazur *et al.*, 2016).

University at Buffalo DataBank (UBDB) of pseudoatoms (Volkov *et al.*, 2007; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012; Kumar *et al.*, 2019), recently remodeled into the databank of Multipolar Atom Types from Theory and Statistical clustering (MATTS) (Jha *et al.*, 2022), was extensively tested for its applications in crystal structure refinement (Jha *et al.*, 2020; Gruza *et al.*, 2020) and electrostatic energy estimating (Kumar *et al.*, 2014; Bojarowski *et al.*, 2017). Detailed analyses based on the S66 and S66x8 benchmark datasets (Řezáč *et al.*, 2011a,b) revealed that the combination of UBDB with the Exact Potential / Multipole Moments (EPMM) method of computing electrostatic energies (Volkov *et al.*, 2004a) leads to energies of chemical accuracy for a wide range of intermolecular distances and interaction types. The strength of the UBDB+EPMM approach lies mostly in accounting for the charge penetration contribution to the energy (Bojarowski *et al.*, 2017). The usage of the higher moments of atomic charge distributions seems to be less important for electrostatic energy estimations.

The question is, can the calculation of electrostatic energies be even faster, even at the cost of a slight loss of accuracy? One possibility is to lower the level of multipole expansion, preferably up to monopole term only (Nassour *et al.*, 2017), but to retain a continuous representation of the charge density for the proper estimation of penetration contributions. It appeared that a truncation of the multipole expansion of the UBDB model to monopole increased the level of errors in the electrostatic energy estimations to unacceptable values

(Bojarowski *et al.*, 2017). A new model of electron density, however, was introduced, called aug-PROMol (Bojarowski *et al.*, 2016). The aug-PROMol model is a hybrid model built upon HCMM, where the third part of HCMM is omitted, values of κ parameters are set to 1.0 and 1.16 for non-hydrogen and hydrogen atoms, respectively, and P_v parameters are computed from the restrained electrostatic potential (RESP) point charges (PC) (Bayly *et al.*, 1993). The RESP PC are derived externally, on-demand for a molecule in question, following procedures widely used in the field of molecular mechanics. Alternatively, RESP charges can be taken from the Invariom Point Charges database (Bojarowski *et al.*, 2018). The aug-PROMol model is spherical by design which allows to speed up energy calculation greatly by omitting time-consuming integration. The accuracy of electrostatic energy estimation by the aug-PROMol+EPMM is similar to the UBDB+EPMM approach. However, the lack of higher electron-density moments might limit the use of the aug-PROMol model in crystal structure refinement, if used instead of IAM. Verification of this hypothesis is one of the aims of this contribution.

Spherical, atom-centered, but charge-aware models were already used in experimental charged density studies. In the so-called κ -model introduced by Coppens *et al.* (1979), pseudoatom electron densities were defined only by the first two parts of equation (2). In the κ -model, populations of valence electrons, P_v , along with the κ parameters describing expansion or contraction of valence electron densities were refined from high-resolution X-ray diffraction data. With this model, atomic partial charges and molecular dipole moments were obtained for the first time from X-ray data. The κ -model was however never widely used, the general conclusion was that the κ -model is not good enough to model high-resolution X-ray data.

Ideally, one would like to have a model: (a) parametrized in one single way, (b) accurate enough for both approaches, TAAM refinement and intermolecular electrostatic interaction energy calculations, and (c) which can be truncated at the monopole level of multipole expansion to speed up energy calculations but with an only slight loss in accuracy. One way to achieve this is the complete redesign of HCMM, by adding additional radial functions (Koritsanzky *et al.*, 2012). This would allow the model to describe an electron density in more detail and may also impact the convergence of the model. Nonetheless, there is still room for improvements with the current HCMM, in the way it is being refined.

Here we designed and tested new strategies for multipole model refinements used when the UBDB databank is created. We focused the most on the refinement of monopole functions, to be more specific, on the refinement of the spherical valence density part. We aimed to enhance the role of the spherical (monopole) part in electron density modeling. In the standard UBDB/MATTS approach, spherical valence densities were refined simultaneously with angular deformation functions. Here we decided to apply the cumulative approach and refine multipole model parameters step by step. We firstly refined only spherical valence densities, allowing them to model

atomic electron densities as much as possible. In the next stage(s) of the refinement, we kept spherical valence densities fixed and refined only angular deformation functions. The second crucial decision was related to κ and κ' parameters which model expansion/contraction of electron density functions. These parameters should not significantly deviate from 1.0. They are the most difficult parameters to be refined within the Hansen–Coppens multipole model. Most importantly, they influence the values of penetration contributions to electrostatic interaction energies computed from the model. In the new refinement strategies, we refined κ and κ' parameters either simultaneously with multipole population parameters or independently from them with a cumulative approach.

The models resulting from the new strategies were then firstly validated towards their performance in electrostatic interaction energy estimation, secondly as a source of atomic X-ray scattering factors in crystal structure refinements on experimental data. As in our previous publications (Bojarowski *et al.*, 2016; 2017; 2018; Kumar *et al.*, 2014), the S66 and S66 \times 8 datasets were chosen as a benchmark for the electrostatic energy estimations. In addition to electrostatic energies obtained from the DFT-SAPT approach, the electrostatic energies estimated from the UBDB and aug-PROMol models constituted the second reference points in our validation procedure. To test models in TAAM refinements we collected subatomic high-resolution X-ray diffraction data for two molecular crystals: cyanuric acid dihydrate and triptycene. In addition for triptycene, we utilized published previously neutron diffraction data (Sanjuan-Szklarz *et al.*, 2016). Full multipole model refinement on X-ray data and structural model from neutron diffraction data served as reference points for testing the new models.

With all the above we hope to create a new databank, collecting sets of multipole model parameters for various atom types, where all parameters together perform well for the crystal structure refinement and the spherical parameters only perform well for the computation of the electrostatic energy.

2. Methods

2.1. Electrostatic interaction energy validation on the S66 and S66 \times 8 datasets

2.1.1. Parametrization of electron density models. New electron density models were obtained for 14 different molecules for geometries extracted from the S66 and S66 \times 8 datasets (Rezáč *et al.*, 2011a,b). The parametrization followed the same procedure as in the case of the UBDB construction (Volkov *et al.*, 2004b; Dominiak *et al.*, 2007; Jarzemska & Dominiak, 2012) with some changes. The valence-only theoretical structure factors were computed (Frisch *et al.*, 2009) from molecular wavefunctions obtained at the B3LYP/6-31G** (Becke, 1988; Perdew, 1986; Lee *et al.*, 1988; Kendall *et al.*, 1992) level of theory. The multipole model parameters were refined (Volkov *et al.*, 2006) against theoretical structure factors following various refinement strategies (see below).

The final averaging procedure of multipole model parameters over the atom types used to build UBDB was omitted. The multipole model parameters thus obtained constituted so-called tailored pseudoatoms, which means they were derived for and applied to the same molecules. It is worth noting, that electron density models obtained from this procedure represent electron densities of isolated molecules, not affected by intermolecular interactions.

Of many possible strategies, only two of them lead to refinements stable enough to give promising electron density models: $P_v\kappa|P_{lm}\kappa'$ and $P_v\kappa|P_{lm}\kappa'$.

In the first refinement strategy, multipole model parameters were divided into two blocks: P_v and κ parameters describing valence electron densities, and P_{lm} and κ' parameters describing angular deformations of electron densities. At the first step of the refinement, P_v and κ parameters were refined together for all atoms in the molecule, until the convergence was achieved. Then these parameters were fixed and the remaining parameters, P_{lm} ($l = 1, \dots, 4$ for non-hydrogen and $l = 1, 2$ for hydrogen atoms) and κ' (one for all levels of multipole expansion, from $l = 1$ to $l = 4$), were refined together until the convergence. The electron density model thus obtained we marked as $P_v\kappa|P_{lm}\kappa'$.

The second strategy was similar, however, including three blocks of refined parameters and thus three stages of refinement. The first stage was the same, *i.e.* refinement of P_v and κ parameters. In the second stage, the values of P_v and κ parameters were kept fixed and only population parameters of multipole functions P_{lm} ($l = 1, \dots, 4$ for non-hydrogen and $l = 1, 2$ for hydrogen atoms) were refined. Finally, the refinement of κ' parameters (one for all levels of multipole expansion, from $l = 1$ to $l = 4$) was done in the presence of fixed values of the other parameters. The electron density model thus obtained we marked as $P_v\kappa|P_{lm}\kappa'$.

2.1.2. Estimation of electrostatic energy with new models of electron densities. The electrostatic interaction energies were estimated using the EPMM method (Volkov *et al.*, 2004a) and the *XD2006* software (Volkov *et al.*, 2006) as described in our previous publications (Bojarowski *et al.*, 2016, 2017, 2018). It is a hybrid method of exact potential (EP) integration at a closer distance (5 Å) and the simple sum of multipole moments (MM) interactions at larger distances. The software allows us to use the EP or MM method separately, and it is possible to estimate the electrostatic energy at any level of the multipolar expansion as described in our previous publication (Bojarowski *et al.*, 2016; 2017; 2018).

2.1.3. Reference values of electrostatic energy. As an ultimate reference, we used quantum mechanics values obtained at the DFT-SAPT/B3LYP/aug-cc-pVTZ level of theory (Bojarowski *et al.*, 2018). To determine whether new models are better in the electrostatic energy estimation than already known models, we used the results obtained from the tailored UBDB approach (simultaneous refinement of all multipole model parameters) computed at B3LYP/6-31G** level of theory as described in §2.1.1, and from the aug-PROmol approach taken from our previous publications (Bojarowski *et al.*, 2016, 2018).

2.2. Experimental data collection and analysis

2.2.1. Crystallization, data collection, data reduction, structure solution and the IAM refinements.

Cyanuric acid. Cyanuric acid powder (8 mg) was added to a water (1.5 ml) : ethanol (1.5 ml) (1 : 1, $v : v$) mixture, and it was dissolved in a water bath at 56°C, the solution was stirred from time to time. The solution was set aside for one day to initiate crystallization at room temperature. The next day, just before the measurement, the solution was gradually poured onto the Petri dish so that the crystals grew to an optimal size and did not dry out. Single crystals with well formed faces were obtained. Due to instability both in air and in oil, crystals were mounted on a capillary tube and immediately put into a nitrogen stream at 100 K.

Single-crystal X-ray diffraction measurement was performed at 100 K on a Rigaku Oxford Diffraction SuperNova four-circle diffractometer with a molybdenum radiation source equipped with a low-temperature nitrogen gas-flow device (Oxford Cryosystems Cryostream Plus). The determination of unit-cell parameters, integration of reflection intensities and data reduction, including 3D profile fitting, following profile size changes with incidence angle, numerical absorption correction based on Gaussian integration, and empirical absorption correction implemented in SCALE3 ABSPACK scaling algorithm, were performed using *CrysAlisPro* (Rigaku Oxford Diffraction, 2016). Finally, reflections were merged with the *SORTAV* program (Blessing, 1987, 1995, 1997). Final data collection and reduction parameters are summarized in Table 1.

The structure was solved in the *Olex2* suite (Dolomanov *et al.*, 2009) using the *SHELXT* program (Sheldrick, 2015a) with the intrinsic phasing method applied. Refinement with the independent atom model (IAM) was performed with the *SHELXL* program (Sheldrick, 2015b).

Raw diffraction images and associated data are available online under the following DOI: 10.18150/MRGTVJ (Repository for Open Data, Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw, Warsaw, Poland).

Triptycene – X-ray diffraction experiment. Single crystals of triptycene were obtained from Dr Tomasz Ratajczyk from the Institute of Organic Chemistry, Polish Academy of Sciences, Poland.

A single-crystal X-ray diffraction dataset was collected on a Bruker AXS Kappa APEX II Ultra single-crystal diffractometer equipped with a CCD-type APEX II area detector, molybdenum TXS rotating anode (Mo $K\alpha$ radiation, $\lambda = 0.71073$ Å), four-circle goniometer, multi-layer optics and a low-temperature nitrogen gas-flow device by Oxford Cryosystems (700 Series Cryostream). The determination of the unit-cell parameters and the integration of raw diffraction images were performed with the *APEX3* program package (Bruker, 2016). The dataset was corrected for Lorentz, polarization and oblique incidence effects. The multi-scan absorption correction, frame-to-frame scaling and merging of reflections were carried out with the *SORTAV* program

Table 1

Parameters characterizing the studied crystals, diffraction data collection and refinement.

Crystal parameters		
Name	Cyanuric acid dihydrate	Triptycene
Formula	C ₃ H ₃ N ₃ O ₃ ·2(H ₂ O)	C ₂₀ H ₁₄
Formula weight, M_r (g mol ⁻¹)	165.12	254.31
Space group	$I2/a$ (No. 15)	$P2_12_12_1$ (No. 19)
Z	4	4
F_{000}	344	536
Crystal color and habit	Colorless, block	Colorless, block
Crystal size (mm)	0.31 × 0.23 × 0.18	0.10 × 0.16 × 0.34
Diffraction data parameters		
T (K)	100	100
a (Å)	8.6385 (1)	8.0808 (3)
b (Å)	6.6817	8.1662 (3)
c (Å)	11.6069 (1)	20.3821 (8)
β (°)	107.466 (1)	90
V (Å ³)	639.061 (10)	1345.00 (9)
d_{calc} (g cm ⁻³)	1.716	1.256
Absorption coefficient, μ (mm ⁻¹)	0.163	0.071
θ range (°)	3.51–71.10	2.69–60.68
$(\sin \theta/\lambda)_{\text{max}}$ (Å ⁻¹)	1.33	1.23
No. of reflections collected, unique	61980, 6322	110051, 20485
Completeness (%)	99.4	98.5
R_{mrg}^\dagger	0.0255	0.0389
No. of reflections with $I \geq 3\sigma(I)$	5349 (84.1)‡	16190 (77.8)‡
Refinement parameters§		
No. of reflections, parameters, restraints	6322, 198, 0	20485, 263, 14
$R(F)$ [$I \geq 3\sigma(I)$], (all data)	0.0115, 0.0172	0.0195, 0.0301
$wR_2(F^2)$ [$I \geq 3\sigma(I)$], (all data)	0.0318, 0.0341	0.0425, 0.0462
$R(F^2)$ [$I \geq 3\sigma(I)$], (all data)	0.0159, 0.0162	0.0237, 0.0259
$S(F^2)$ [$I \geq 3\sigma(I)$], (all data)	0.989, 0.979	0.965, 0.933
$\Delta\rho_{\text{min}}, \Delta\rho_{\text{max}}$ (e Å ⁻³) [$I \geq 3\sigma(I)$]	-0.14, +0.15	-0.14, +0.17
$\Delta\rho_{\text{min}}, \Delta\rho_{\text{max}}$ (e Å ⁻³) (all data)	-0.15, +0.14	-0.20, +0.19

† Definition of R_{mrg} follows Blessing's definitions used in the *SORTAV* program: $R_{\text{mrg}} = \sum_{hkl} [n/(n-1)]^{1/2} \sum_{i=1}^n |I_i - I_{\text{mean}}| / \sum_{hkl} \sum_{i=1}^n I_i$. ‡ The numbers in brackets are the percentage (%) of the 'observed' unique reflections out of the total number of unique reflections. § From multipole model refinements.

(Blessing, 1987, 1995, 1997). The collected dataset exhibited a high-resolution limit of $(\sin \theta/\lambda)_{\text{max}} = 1.23 \text{ \AA}^{-1}$. Final data collection and reduction parameters are summarized in Table 1.

The structure was solved using direct methods as implemented in the *SHELXS* program (Sheldrick, 2008), and refined with the *SHELXL* program (Sheldrick, 2015b) within the independent atom model (IAM) approximation. At this stage it was checked to see if the data were extinction free by refining extinction coefficient x using the following expression:

$$F_{\text{calc}}^* = kF_{\text{calc}} \left[1 + 0.001x F_{\text{calc}}^2 \lambda^3 / \sin(2\theta) \right]^{-1/4}.$$

The extinction coefficient refined to 0.000 (3) suggesting there is no indication of significant extinction in the data.

Raw diffraction images and associated data are available online under the following DOI: 10.18150/ZDLZTR (Repository for Open Data, Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw, Warsaw, Poland).

Triptycene – the neutron diffraction experiment. Reflection data from the single-crystal time-of-flight neutron diffraction experiment was taken from Sanjuan-Szklarz *et al.* (2016). The crystal structure (with anisotropic hydrogen-atom displacement parameters) was then re-refined in the *JANA* package

(Petříček *et al.*, 2014). The extinction was successfully modeled by the Becker–Coppens approach (Becker & Coppens, 1974, 1975).

Selected parameters: $a = 8.1019$ (13), $b = 8.1922$ (13), $c = 20.442$ (3) Å; total number of reflections: 21802; number of reflections with $I \geq 3\sigma(I)$: 21749; index ranges: $-16 \leq h \leq 6$, $-20 \leq k \leq 21$, $-56 \leq l \leq 52$; number of parameters: 316; discrepancy factors: $R(F)$ [$I \geq 3\sigma(I)$] = 0.0627, $wR(F)$ [$I \geq 3\sigma(I)$] = 0.0987, $R(F)$ (all data) = 0.0629, $wR(F)$ (all data) = 0.0989, goodness of fit (S) [$I \geq 3\sigma(I)$] = 4.39; nuclear density largest extrema: $-2.09/+4.37 \text{ fm \AA}^{-3}$; extinction parameter (type I Lorentzian isotropic): 76.3 (12). The CIF file is available in supporting information.

2.2.2. Multipole model refinements. *Cyanuric acid dihydrate.* Multipole model refinement on experimental data was performed in the *WinXD2016* package (Volkov *et al.*, 2016) with using the Hansen–Coppens multipolar model (Hansen & Coppens, 1978). The initial atomic coordinates for all atoms, the anisotropic atomic displacement parameters for non-hydrogen atoms and the isotropic atomic displacement parameters for hydrogen atoms were taken from the IAM refinement. The local Cartesian coordinate systems, the local symmetry constraints, and the initial population and the expansion/contraction parameters of the multipole model were defined by the *LSDB* program (Volkov *et al.*, 2004b)

combined with UBDB2011 (Jarzemska & Dominiak, 2012). Each atom was assigned core and spherical valence scattering factors derived from Su & Coppens (1998) atomic wavefunctions. $X-H$ bond lengths were constrained to neutron-normalized distances ($d_{O-H} = 0.958 \text{ \AA}$, $d_{N-H} = 1.030 \text{ \AA}$) (Allen & Bruno, 2010). The values of the anisotropic atomic displacement parameters for the hydrogen atoms were estimated using the SHADE 3.0 server (Madsen, 2006). For non-hydrogen atoms, anisotropic atomic displacement parameters were refined in harmonic approximation, except for O2 and O3 oxygen atoms, for which anharmonic approximation was applied. Gram-Charlier coefficients (Johnson, 1969; Kuhs, 1983; Scheringer, 1985) up to the fourth order were used, while the physical reliability of the anharmonic model was confirmed by the probability density function computed with *XDPDF* from the *WinXD2016* suite. The multipole expansion was truncated at the hexadecapole ($l_{\max} = 4$) and quadrupole ($l_{\max} = 2$) levels for all non-hydrogen and hydrogen atoms, respectively. Only parameters allowed by local and crystallographic symmetry were refined. κ and κ' parameters for hydrogen atoms were kept fixed at the UBDB-transferred values. The general multipole model refinement strategy consisted of the following steps: the refinement of (i) scale factor (which was also refined in all other stages); (ii) atomic coordinates and ADPs for all atoms; (iii) introduction of $X-H$ constraints, refinement of atomic coordinates and ADPs for non-hydrogen atoms; (iv) *SHADE* estimation of anisotropic hydrogen atom ADPs; (v) atomic coordinates and ADPs refinement for non-hydrogen atoms including the third- and fourth-order anharmonic parameters for the O2 and O3 oxygen atoms; (vi) atomic coordinates and ADPs for non-hydrogen atoms, multipole population and expansion/contraction parameters in a stepwise manner; (vii) update of *SHADE* estimation of anisotropic hydrogen atom ADPs; (viii) all parameters simultaneously.

Refinements were carried out on F^2 on all data up to $\sin \theta/\lambda \leq 1.33 \text{ \AA}^{-1}$, without any $F^2/\sigma(F^2)$ cut-off, using the statistical weights [*i.e.* for the i -th reflection $w_i = 1/\sigma(F^2)_i$].

The correctness of the multipole model refinement was judged from refinement statistics [$R(F)$, $wR_2(F^2)$ and goodness of fit $S(F^2)$], differences of mean-square displacement amplitudes (DMSDA) for non-hydrogen atoms and residual density analysis. The largest DMSDA value of $-4 \times 10^{-4} \text{ \AA}^2$ was observed for the $C2=O2$ bond, which is satisfactory. A comparison of F_o^2 and F_c^2 (Fig. 1S) shows that they scale to each other with the same proportion across all resolution bins and intensity values, and the differences between them (intensity residuals) have close to normal distribution. Residual electron density analysis (Meindl & Henn, 2008) shows that the derived model is characterized by a rather flat and featureless residual density distribution (Figs. 2S and 4S).

At this stage, it was checked once again, if the data are extinction free by refining isotropic type 2 extinction (Becker & Coppens, 1974), assuming the absorption weighted path length $TBAR = 0.24 \text{ mm}$. The extinction coefficient was refined to 0.013 (3), with only five reflections requiring correction larger than 1%, the largest correction was 2.4% and was applied to the reflection 404. The coordinates, atomic displacement parameters and multipole model parameters were refined to values which differ by less than one standard deviation when compared to values obtained with no extinction correction. It was concluded that the extinction in the data is very weak and was not taken into account in the final analyses.

All final refinement statistics are summarized in Table 1. The *ORTEP* plot for the asymmetric unit is given in Fig. 1. The CIF file is given in supporting information or can be retrieved free of charge from the Cambridge Crystallographic Data Centre (Groom *et al.*, 2016).

Triptycene. Multipole model refinement was carried out using the *MoPro* suite (Guillot *et al.*, 2001; Jelsch *et al.*, 2005)

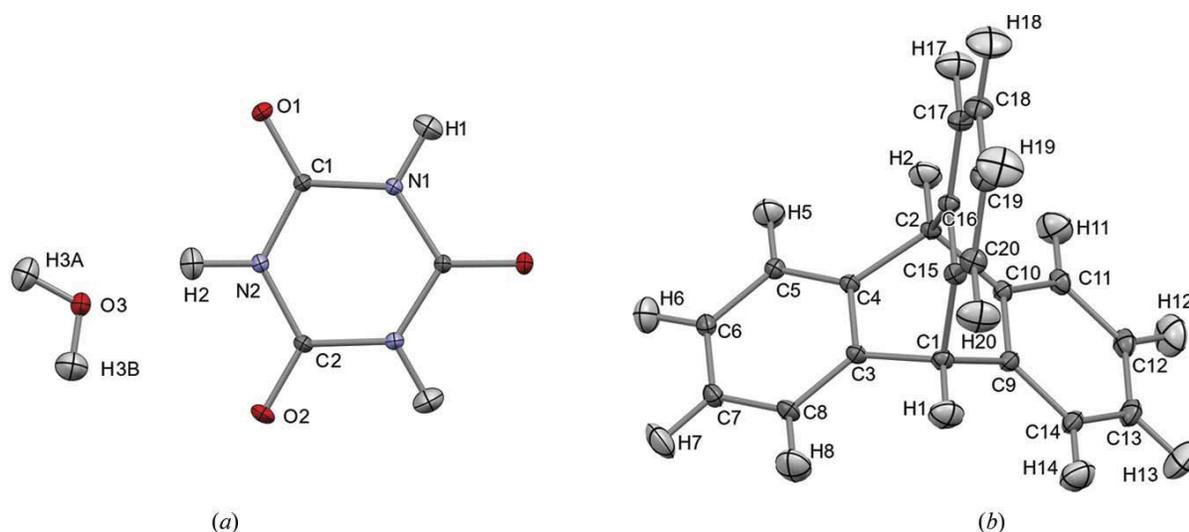


Figure 1

ORTEP plot for cyanuric acid dihydrate (a) and triptycene (b) after multipole model refinement on experimental X-ray data drawn with *Mercury* (Macrae *et al.*, 2008). Ellipsoids are drawn at the 50% probability level.

with the use of the Hansen–Coppens multipole model (Hansen & Coppens, 1978). The initial atomic coordinates for all atoms, the anisotropic atomic displacement parameters for non-hydrogen atoms, and the isotropic atomic displacement parameters for hydrogen atoms were taken from the IAM refinement. The local Cartesian coordinate systems, the local symmetry constraints, the initial population, and the contraction/expansion parameters of the multipole model were defined by the LSDb (Volkov *et al.*, 2004b) combined with the UBDB2011 databank (Jarzemska & Dominiak, 2012). Each atom was assigned core and spherical valence scattering factors derived from the Su & Coppens (1998) atomic wavefunctions. The X–H bond lengths were restrained with $\sigma = 0.001 \text{ \AA}$ to distances derived from neutron diffraction data. The values of the anisotropic atomic displacement parameters for the hydrogen atoms were transferred from the neutron diffraction data with appropriate scale. For non-hydrogen atoms, anisotropic atomic displacement parameters (anisotropic ADPs) were refined in the harmonic approximation. The multipole expansion was truncated at the hexadecapole ($l_{\max} = 4$) and quadrupole ($l_{\max} = 2$) levels for all non-hydrogen and hydrogen atoms, respectively. Only parameters allowed by local and crystallographic symmetry were refined. Constraints resulting from chemical similarities were applied to multipole model parameters. κ parameters for hydrogen and κ' parameters for all atoms were kept fixed at the UBDB-transferred values. The general multipole model refinement strategy consisted of the following steps: the refinement of (i) scale factor (which was also refined in all other stages); (ii) atomic coordinates; (iii) atomic coordinates and ADPs; (iv) SHADE estimation of anisotropic hydrogen atom ADPs (which was also updated in-between other stages); (v) multipole population parameters in a stepwise manner; (vi) all multipole population parameters and structural parameters simultaneously; (vii) block refinement of non-hydrogen atom κ parameters (first block), step (vi) (second block); (viii) all parameters from step (vii) together; (ix) anisotropic hydrogen ADPs transfer from neutron diffraction data; (x) all parameters simultaneously.

Refinements were carried out on F^2 on all data up to $\sin \theta/\lambda \leq 1.23 \text{ \AA}^{-1}$, without any $F^2/\sigma(F^2)$ cut-off, using statistical weights [*i.e.* for the i -th reflection $w_i = 1/\sigma(F^2)_i^2$].

Correctness of the multipole model refinement was judged from refinement statistics [$R(F)$, $wR_2(F^2)$ and goodness of fit $S(F^2)$], differences of mean-square displacement amplitudes (DMSDA) for non-hydrogen atoms, and residual density analysis. The largest DMSDA value of $-5 \times 10^{-4} \text{ \AA}^2$ was observed for the C1–C15 and C2–C16 bonds, which is satisfactory. A comparison of F_o^2 and F_c^2 (Fig. 1S) shows that they scale to each other with the same proportion across all resolution bins and intensity values, and the differences between them (intensity residuals) have close to normal distribution. Residual electron density analysis (Meindl & Henn, 2008) shows that the derived model is characterized by a rather flat and featureless residual density distribution (Figs. 2S and 4S).

All the final refinement statistics are summarized in Table 1. The ORTEP plot for the asymmetric unit is given in Fig. 1. The CIF file is present in the supporting information or can be retrieved free of charge from the Cambridge Crystallographic Data Centre (Groom *et al.*, 2016).

2.2.3. TAAM refinements. TAAM refinements were carried out on F^2 , without any $F^2/\sigma(F^2)$ cut-off, using statistical weights [*i.e.* for the i -th reflection $w_i = 1/\sigma(F^2)_i^2$] in the XD2016 package. Datasets of two resolutions were used: truncated at the atomic resolution $d_{\min} = 0.83 \text{ \AA}$ ($\sin \theta/\lambda \leq 0.60 \text{ \AA}^{-1}$) and of subatomic high-resolution resulting from the experiment – the same as for multipole model refinement. Atomic coordinates and atomic displacement parameters (anisotropic for non-hydrogen and isotropic for hydrogen atoms) for all atoms were simultaneously refined with no restraints. In addition, for O2 and O3 oxygen atoms in cyanuric acid dihydrate anharmonic approximation was used with Gram–Charlier coefficients (Kuks, 1983; Johnson, 1969; Scheringer, 1985) up to the fourth order. The number of unique reflections at atomic resolution and of refined parameters in TAAM refinements for cyanuric acid dihydrate was 581 and 104, respectively, and for triptycene 2427 and 237, respectively.

TAAM was parametrized with the tailored UBDB approach, $P_v\kappa|P_{lm}\kappa'$ and $P_v\kappa|P_{lm}| \kappa'$ models, following procedures as described in §2.1.2, based on molecular geometries taken from multipole models refined on experimental data. For refinements on low-resolution data, TAAM parametrized with multipole model parameters transferred from multipole model refinements (MMR) on high-resolution data was also tested. In addition, the aug-PRomol model was constructed on the same geometries following the procedure described by Bojarowski *et al.* (2016).

3. Results

3.1. Refinement strategies

Following the arguments mentioned in the *Introduction*, one can imagine several possible refinement strategies to obtain new models of electron densities. The most strict cumulative approach would be to refine each level of multipole expansion separately, one after another, *i.e.* refining higher multipole parameters in the presence of previously refined and frozen parameters of lower multipoles. With the Hansen–Coppens multipole model that would be the following consecutive steps: P_v and κ , P_{1m} and κ'_1 , P_{2m} and κ'_2 , P_{3m} and κ'_3 , and P_{4m} and κ'_4 . One can be even more restrictive, and at each level of multipole expansion refine firstly population parameters, then freeze them and refine the κ'_i parameters of that level. These two strategies turned out to be very unstable during the refinement of the Hansen–Coppens multipole model. There were often problems with achieving refinement convergence, or values of refined parameters were far from the acceptable (*i.e.* κ and κ' parameters not close to one). Only two strategies lead to stable refinements and

promising electron density models: the $P_{\nu}\kappa|P_{lm}\kappa'$ model where P_{ν} and κ parameters were refined firstly, then P_{lm} ($l = 1, \dots, 4$) and κ' (as one parameter for all levels of multipole expansion, $\kappa' = \kappa'_1 = \kappa'_2 = \kappa'_3 = \kappa'_4$); and the $P_{\nu}\kappa|P_{lm}\kappa'$ model where the second stage was divided into two: firstly P_{lm} , then κ' parameters.

3.2. Electrostatic energy estimations from new electron density models

The first model combined with the EPMM method of electrostatic energy calculation, $P_{\nu}\kappa|P_{lm}\kappa'+EPMM$, always showed a similar trend in computed electrostatic energy values as tailored UBDB+EPMM and to the reference values (REF) (Fig. 2). This is intuitive, as the refinement procedure is the closest to that of UBDB. Nevertheless, RMSE for the $P_{\nu}\kappa|P_{lm}\kappa'+EPMM$ method is larger, by up to ~50% compared to the UBDB+EPMM method, in almost all analyzed subgroups of the S66x8 dataset.

When comparing the two tested models, the most important conclusion seems to be the enormous influence of the κ' value on the accuracy of the electrostatic energy estimation. The $P_{\nu}\kappa|P_{lm}\kappa'$ model generates electrostatic energies which are much further away from the reference values, with RMSE double, or even more, than for the first model. Apparently, population and expansion/contraction parameters strongly

depend on each other and therefore they must be refined simultaneously (or at least iteratively, not cumulatively).

The second important finding is the overestimation of electrostatic energy. While the main reason for the need to develop further the UBDB method was the underestimation of the energy, the new models overestimate electrostatic interactions.

Interestingly, when comparing the energies obtained from models truncated to the monopole level ($l = 0$), a significant improvement in the estimation of electrostatic energy is seen with the new model(s) over the tailored UBDB at $l = 0$. Please note that both new models, $P_{\nu}\kappa|P_{lm}\kappa'$ and $P_{\nu}\kappa|P_{lm}\kappa'$, when truncated to monopole level simplify to the same model $P_{\nu}\kappa$. With the application of the MM method of calculations, the $P_{\nu}\kappa$ model already at $l = 0$ gives results very close to the results of the RESP charges. At the same time, UBDB truncated at $l = 0$ is much worse, and only at $l = 4$ approaches a similar level of accuracy. When the EPMM method is used, the results are slightly inferior for the $P_{\nu}\kappa$ model but still promising. The $P_{\nu}\kappa$ model is significantly worse only for short distances in electrostatic dominated dimers. Knowing the construction of the aug-PROmol model (Bojarowski *et al.*, 2018), it can be concluded that expansion/contraction parameters are very important and their fine-tuning may lead in the future to even better results.

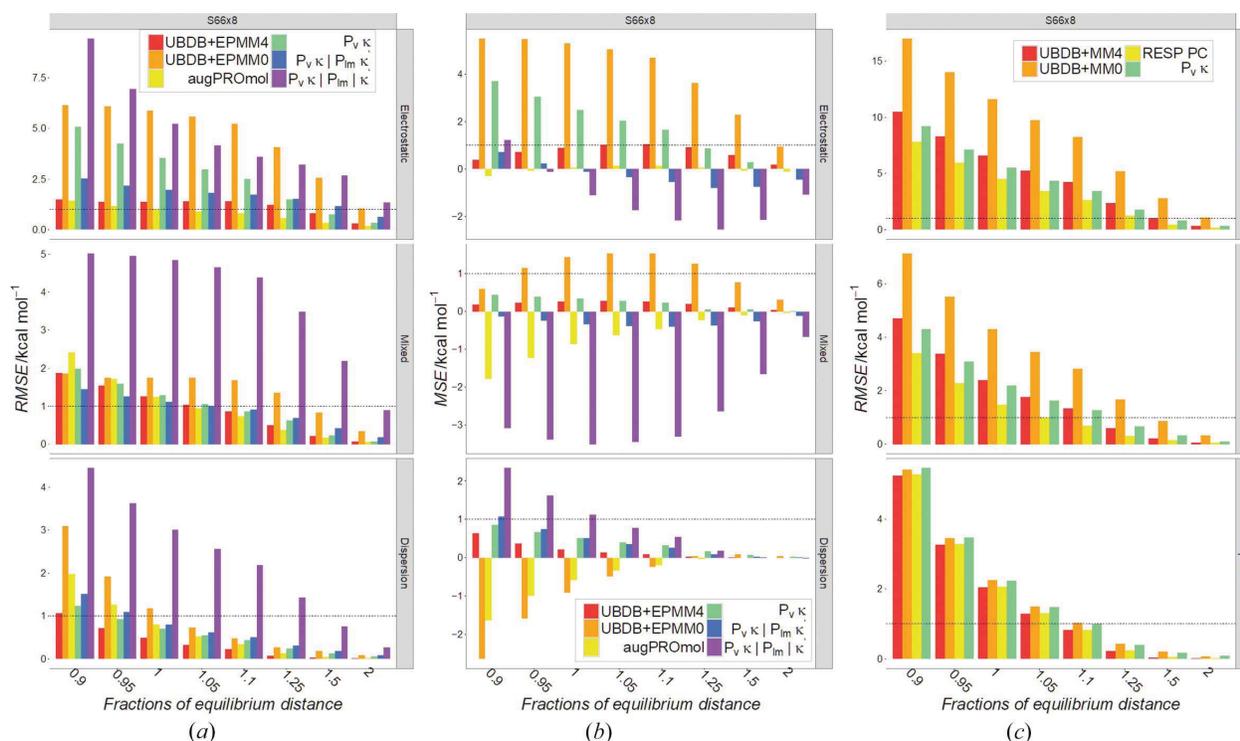


Figure 2 Visualization of the root mean square error (RMSE) and the mean signed error (MSE) of the electrostatic interaction energies (kcal mol^{-1}) computed with the new models ($P_{\nu}\kappa|P_{lm}\kappa'$, $P_{\nu}\kappa|P_{lm}\kappa'$ and $P_{\nu}\kappa$) at different intermolecular distances in three subgroups of the S66x8 dataset. The electrostatic subgroup contains dimers in which electrostatic interactions dominate, whereas in the dispersion subgroup dispersion interaction dominates, and in the mixed subgroup contributions from electrostatic and dispersion interactions are of similar magnitude. Results from other approaches (tailored UBDB and aug-PROmol) are given to allow direct comparisons. The energies were computed by applying EPMM or MM methods, either to full models, up to $l = 4$ (EPMM4, MM4) or truncated at $l = 0$ (EPMM0, MM0). Corresponding reference energies were obtained at the DFT-SAPT/B3LYP/aug-cc-pVTZ level of theory. The dashed horizontal lines along the x axis correspond to the value of $1.0 \text{ kcal mol}^{-1}$.

Table 2

Fitting statistics computed for all reflections from refinements on experimental X-ray diffraction data with the usage of various electron density models.

	Cyanuric acid dihydrate				Triptycene			
	$R(F)$ (%)	$wR_2(F^2)$ (%)	$\Delta\rho_{\min}$ ($e \text{ \AA}^{-3}$)	$\Delta\rho_{\max}$ ($e \text{ \AA}^{-3}$)	$R(F)$ (%)	$wR_2(F^2)$ (%)	$\Delta\rho_{\min}$ ($e \text{ \AA}^{-3}$)	$\Delta\rho_{\max}$ ($e \text{ \AA}^{-3}$)
Atomic resolution								
IAM†	2.7	9.6	−0.25	0.21	2.7	7.3	−0.20	0.12
aug-PROmol	3.1	9.1	−0.37	0.28	2.7	7.5	−0.19	0.13
TAAM($P_{\nu,\kappa} P_{lm}\kappa'$)	1.5	4.8	−0.17	0.11	1.4	3.8	−0.09	0.09
TAAM($P_{\nu,\kappa} P_{lm} \kappa'$)	1.6	5.4	−0.16	0.14	1.4	3.9	−0.09	0.09
TAAM(UBDB)	0.9	2.8	−0.08	0.07	1.1	3.2	−0.09	0.07
TAAM(MMR)	0.8	2.4	−0.07	0.05	1.1	3.3	−0.08	0.07
Subatomic resolution								
IAM†	3.0	8.1	−0.33	0.65	4.2	8.2	−0.25	0.50
aug-PROmol	3.4	9.5	−0.47	0.83	4.2	8.2	−0.28	0.51
TAAM($P_{\nu,\kappa} P_{lm}\kappa'$)	2.3	5.6	−0.29	0.39	3.4	5.3	−0.21	0.28
TAAM($P_{\nu,\kappa} P_{lm} \kappa'$)	2.6	6.7	−0.30	0.58	3.4	5.4	−0.22	0.27
TAAM(UBDB)	1.9	4.1	−0.23	0.21	3.1	4.7	−0.20	0.22

† Refined in *SHELXL* in harmonic approximation only.

3.3. Crystal structure refinements on experimental X-ray diffraction data

To validate the performance of tested models as a source of better atomic X-ray scattering factors we focused on two aspects of X-ray structure refinement: (a) how well the model can be fitted to experimental data, and (b) how accurate structures can be obtained with the model. In addition, we tested the models in two situations, firstly in refinements against subatomic high-resolution data, and secondly in standard types of refinements with the atomic resolution of data ($d_{\min} = 0.83 \text{ \AA}$). As an ultimate reference we used results from refined multipole model (MMR) on high-resolution data, and in the case of triptycene, from neutron diffraction. We also had several TAAM refinement reference points. Firstly, we did TAAM refinements with tailored UBDB parametrization, TAAM(UBDB). Secondly, we did TAAM refinements with multipole model parameters taken from multipole model refinements, TAAM(MMR).

To check the quality of the fitting we focused on the following fitting statistics: $R(F)$, $wR_2(F^2)$, and maximum and minimum residual peaks, see Table 2. From all fitting statistics, it is clear that tested models $P_{\nu,\kappa}|P_{lm}\kappa'$ and $P_{\nu,\kappa}|P_{lm}|\kappa'$ can fit the experimental data much better than standard IAM. This is true for both standard atomic and subatomic high-resolution datasets. $R(F)$ lowers by ~ 1 percentage point, and a reduction in absolute values of extrema for residual electron densities is observed. From comparing values of fitting statistics for $P_{\nu,\kappa}|P_{lm}\kappa'$ and $P_{\nu,\kappa}|P_{lm}|\kappa'$, it can be concluded that the first model fits the data slightly better, but the difference is very small. None of the tested models, however, fits the data as well as the tailored UBDB model. UBDB allows for further lowering of the $R(F)$ factor, by additional ~ 0.5 percentage point, and further lowering of residual extrema, especially in the case of cyanuric acid dihydrate. In fact, the tailored UBDB model fits the data almost as well as the MMR, the latter gives better statistics only in the case of refinements on high-resolution data for cyanuric acid dihydrate. It is worth remem-

bering that with full multipole model refinements, a much larger number of parameters is fitted to the data than in the case of TAAM or IAM refinements. The aug-PROmol model, which was very successful in electrostatic interaction energy estimations, performs very badly in X-ray crystal structure refinements. It can fit the experimental data at a level similar to the standard IAM (for triptycene) or even slightly worse (for cyanuric acid dihydrate).

Besides the fitting statistics, it is important to validate the quality of the obtained structural model. In the case of X-ray diffraction data, the most sensitive parameters are the X –H bond lengths and atomic displacement parameters. With cyanuric acid dihydrate structure, we have in addition a unique opportunity to observe the performance of various models in deconvolution of static electron density from an anharmonic motion of atoms. To compare the performance in atomic displacement modeling we focused on the analysis of U_{eq} for non-H atoms and U_{iso} for hydrogen atoms, as well as on averaged absolute values of DMSDA analysis for bonds between non-H atoms. Following our previous experience on the sensitivity of atomic displacement parameters to the applied electron density models and data resolution, we focused only on atomic resolution refinements, since for these refinements the effects of replacing one scattering model with another are much larger than for subatomic resolution data (Sanjuan-Szklarz *et al.*, 2020).

Surprisingly, it is not so clear anymore if the tested new models are better or not than IAM in obtaining more accurate crystal structures (see Table 3). Positions of hydrogen atoms do improve with new models. The RMSDs for X –H bond lengths are in the range 0.02 – 0.04 \AA when compared to reference mean values from neutron diffraction (Allen & Bruno, 2010) or neutron diffraction data collected directly for triptycene. In the case of IAM, the RMSDs were 0.14 \AA and 0.10 \AA for cyanuric acid dihydrate and triptycene, respectively. The improvement with new models is of the same level as for the tailored UBDB approach. It is only slightly worse than from the TAAM(MMR) refinement, which arrived at an

Table 3

Statistics for selected structural parameters from refinements on experimental X-ray diffraction data with the usage of various electron density models.

DMSDA - differences of mean-squares displacement amplitudes along interatomic vectors (bonds between non-hydrogen atoms) averaged for all bonds. RMSD - root-mean-square difference when compared with results from referential multipole model refinements.

	Cyanuric acid dihydrate				Triptycene			
	DMSDA $\times 10^4$ (\AA^2)	RMSD $U_{\text{eq}} \times 10^4$ (\AA^2)	$U_{\text{iso}} \times 10^4$ (\AA^2)	$d(X-H)^\dagger$ (\AA)	DMSDA $\times 10^4$ (\AA^2)	RMSD $U_{\text{eq}} \times 10^4$ (\AA^2)	$U_{\text{iso}} \times 10^4$ (\AA^2)	$d(X-H)^\dagger$ (\AA)
Atomic resolution								
IAM‡	22	30	70	0.14	11	50 (60)§	100 (110)§	(0.10)§
aug-PRomol	85	110	250	0.17	11	40 (60)	200 (200)	(0.10)
TAAM($P_v\kappa P_{lm}\kappa'$)	38	20	80	0.03	5	10 (20)	30 (40)	(0.03)
TAAM($P_v\kappa P_{lm} \kappa'$)	44	20	110	0.04	7	10 (20)	60 (70)	(0.02)
TAAM(UBDB)	3	10	30	0.04	4	10 (20)	30 (40)	(0.02)
TAAM(MMR)	7	10	40	0.01	4	10 (20)	10 (30)	(0.01)

‡ Refined in harmonic approximation only in *SHELXL*. † Compared to mean values from neutron diffraction (Allen & Bruno, 2010). § In brackets compared to values from neutron diffraction results for triptycene.

RMSD of 0.01 Å. Note that the X–H bond lengths were constrained (cyanuric acid dihydrate) or restrained (triptycene) to neutron data during MMR. On the other hand, both new models have some problems in the refinement of atomic displacement parameters. The first model, $P_v\kappa|P_{lm}\kappa'$, is visibly better than IAM and as good as UBDB only in the case of triptycene. For triptycene, the RMSD of U_{eq} of non-hydrogen atoms and U_{iso} for hydrogen are approximately three times smaller than for IAM when compared to reference neutron diffraction data. DMSDA improves as well. However, for cyanuric acid dihydrate the $P_v\kappa|P_{lm}\kappa'$ model is slightly worse in U_{iso} refinement of hydrogen atoms than IAM, and for non-hydrogen atoms is somewhere in between IAM and UBDB in quality of U_{eq} , but is worse than IAM in the refinement of atomic displacement anisotropy. The second model, $P_v\kappa|P_{lm}|\kappa'$, performs even worse than the first one, even for triptycene. This is especially visible for non-hydrogen atoms and refining correct anisotropy of non-hydrogen atom displacements.

The aug-PRomol model performs the worst. It does not improve the positions of hydrogen atoms, it makes them even worse than IAM in the case of cyanuric acid dihydrate. Similarly, it either fails to improve atomic displacement parameters or actually makes them worse. The model performs especially badly in the refinement of atomic displacements in the presence of anharmonic motion.

4. Discussion

From the two tested models, resulting from two slightly different strategies of cumulative refinement of the multipole model, the second one, $P_v\kappa|P_{lm}|\kappa'$, is of no use either in electrostatic interaction energy estimations or in X-ray crystal structure refinements. In the latter area of applications, the improper treatment of κ' parameters has the largest consequences in modeling atomic displacement parameters. Thus it might be concluded that cumulative refinement of P_{lm} and κ' parameters of the multipole model leads to much worse values than simultaneous refinement. Therefore such an approach to

the multipole model refinement should never be used and whenever simultaneous refinement of P_{lm} and κ' is not stable, care must be taken to assure that iterative refinement converged.

The first tested model, $P_v\kappa|P_{lm}\kappa'$, resulting from cumulative refinement of two blocks of parameters, P_v and κ and P_{lm} and κ' , has promising properties. It seems to model quite well non-polar molecules or non-polar chemical fragments. For non-polar molecule triptycene, it was better than IAM and only slightly worse than UBDB in X-ray crystal structure refinement. For molecular dimers in which dispersive intermolecular interactions dominate, or dispersive interactions are as similarly important as electrostatic interactions, it estimates electrostatic interaction energies computed with the EPMM approach similarly as well as UBDB, and for molecules at short distances better than aug-PRomol. For polar molecules, however, the $P_v\kappa|P_{lm}\kappa'$ model has some problems with the proper refinement of atomic displacement parameters (probably enhanced by the presence of anharmonic motion) and is slightly worse than UBDB in the estimation of electrostatic energy.

The $P_v\kappa|P_{lm}\kappa'$ model has interesting properties when truncated to the monopole level – the $P_v\kappa$ model. The charges derived from the $P_v\kappa$ model behave similarly to the RESP charges in electrostatic interaction energy estimations from point charges (the MM method). In this respect, the $P_v\kappa$ model is much better than the UBDB model. The UBDB model approaches the quality of energy estimations seen for the RESP charges only at full expansion, up to octupole level to be more precise, as we showed in Bojarowski *et al.* (2017). This indirectly shows that electrostatic potential at the molecular surface computed from the point charges derived from the $P_v\kappa$ model is already very good. The $P_v\kappa$ model is also a promising model to replace the aug-PRomol approach. When combined with the EPMM method of electrostatic energy calculations, it is as good as aug-PRomol for mixed dimers, a bit better for dispersion-dominated dimers and worse only for electrostatic-dominated dimers. Apparently, some improvements in κ parametrization for polar groups are still needed.

5. Conclusions

A cumulative approach to multipole model refinements, as opposed to the simultaneous refinement of all multipole model parameters, leads to substantially different models of electron density. The consequences of the differences are clearly visible in both electrostatic properties derived from the models and the descriptions of atomic displacements resulting from the TAAM refinements based on those models.

The $P_{\nu\kappa}|P_{lm\kappa'}$ model is very promising and, after some improvements, it has a chance to be no worse than UBDB in X-ray crystal structure refinements and at the same time, with the same route of model parameterization, it may deliver the $P_{\nu\kappa}$ model not worse than the aug-PROMOL model in speed and accuracy of electrostatic energy estimation. The $P_{\nu\kappa}|P_{lm\kappa'}$ model by design, similarly to the UBDB and aug-PROMOL models, does not account for charge density polarization due to intermolecular interactions, in contrast to multipole modeling of the experimental X-ray diffraction data (Spackman, 2018). Thus, the electrostatic interaction energies obtained with a combination of the model and the EPMM method will correspond to the first-order electrostatic energy contribution to the total interaction energy computed within the framework of the SAPT theory.

In addition, the aug-PROMOL model, although it is at the moment still the best spherical model for electrostatic interaction energy estimations, is not suitable for crystal X-ray structure refinement. It neither fits the X-ray data nor leads to a structural model better than IAM. It may be even worse than IAM in some cases.

Acknowledgements

We thank Dr Anna Makal, University of Warsaw, Poland, and Oxford Diffraction division of Rigaku, Poland, for help in preliminary high-resolution X-ray diffraction measurements for cyanuric acid dihydrate; and Dr Tomasz Ratajczyk, Polish Academy of Sciences, Poland, for providing crystals of tripylene.

Funding information

The following funding is acknowledged: the National Science Centre, Poland (grant No. 2015/19/N/ST4/00010 to Sławomir Antonii Bojarowski); the PL-GRID Infrastructure (grant No. ubdb2018 to Paulina Maria Dominiak; grant No. ubdb2019 to Paulina Maria Dominiak).

References

Allen, F. H. & Bruno, I. J. (2010). *Acta Cryst.* **B66**, 380–386.
 Bąk, J. M., Domagała, S., Hübschle, C., Jelsch, C., Dittrich, B. & Dominiak, P. M. (2011). *Acta Cryst.* **A67**, 141–153.
 Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. (1993). *J. Phys. Chem.* **97**, 10269–10280.
 Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.
 Becker, P. J. & Coppens, P. (1974). *Acta Cryst.* **A30**, 148–153.
 Becker, P. J. & Coppens, P. (1975). *Acta Cryst.* **A31**, 417–425.
 Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.
 Blessing, R. H. (1995). *Acta Cryst.* **A51**, 33–38.
 Blessing, R. H. (1997). *J. Appl. Cryst.* **30**, 421–426.

Bojarowski, S. A., Kumar, P. & Dominiak, P. M. (2016). *Chem-PhysChem*, **17**, 2455–2460.
 Bojarowski, S. A., Kumar, P. & Dominiak, P. M. (2017). *Acta Cryst.* **B73**, 598–609.
 Bojarowski, S. A., Kumar, P., Wandtke, C. M., Dittrich, B. & Dominiak, P. M. (2018). *J. Chem. Theory Comput.* **14**, 6336–6345.
 Brock, C. P., Dunitz, J. D. & Hirshfeld, F. L. (1991). *Acta Cryst.* **B47**, 789–797.
 Bruker (2016). APEX3. Bruker AXS Inc., Madison, Wisconsin, USA.
 Coppens, P., Guru Row, T. N., Leung, P., Stevens, E. D., Becker, P. J. & Yang, Y. W. (1979). *Acta Cryst.* **A35**, 63–72.
 Dadda, N., Nassour, A., Guillot, B., Benali-Cherif, N. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 452–463.
 De Bruyne, B. & Gillet, J.-M. (2020). *Acta Cryst.* **A76**, 1–6.
 Dittrich, B., Hübschle, C. B., Luger, P. & Spackman, M. A. (2006). *Acta Cryst.* **D62**, 1325–1335.
 Dittrich, B., Hübschle, C. B., Messerschmidt, M., Kalinowski, R., Girnt, D. & Luger, P. (2005). *Acta Cryst.* **A61**, 314–320.
 Dittrich, B., Hübschle, C. B., Pröpper, K., Dietrich, F., Stolper, T. & Holstein, J. J. (2013). *Acta Cryst.* **B69**, 91–104.
 Dittrich, B., Koritsánszky, T. & Luger, P. (2004). *Angew. Chem. Int. Ed.* **43**, 2718–2721.
 Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. (2009). *J. Appl. Cryst.* **42**, 339–341.
 Domagała, S., Fournier, B., Liebschner, D., Guillot, B. & Jelsch, C. (2012). *Acta Cryst.* **A68**, 337–351.
 Dominiak, P. M., Volkov, A., Li, X., Messerschmidt, M. & Coppens, P. (2007). *J. Chem. Theory Comput.* **3**, 232–247.
 Durka, K., Jarzemska, K. N., Kamiński, R., Luliński, S., Serwatowski, J. & Woźniak, K. (2013). *Cryst. Growth Des.* **13**, 4181–4185.
 Frisch, M. J. *et al.* (2009). GAUSSIAN09. Gaussian Inc., Wallingford, CT, USA.
 Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* **B72**, 171–179.
 Gruza, B., Chodkiewicz, M. L., Krzeszczakowska, J. & Dominiak, P. M. (2020). *Acta Cryst.* **A76**, 92–109.
 Guillot, B., Viry, L., Guillot, R., Lecomte, C. & Jelsch, C. (2001). *J. Appl. Cryst.* **34**, 214–223.
 Hansen, N. K. & Coppens, P. (1978). *Acta Cryst.* **A34**, 909–921.
 Holstein, J. J., Hübschle, C. B. & Dittrich, B. (2012). *CrystEngComm*, **14**, 2520–2531.
 Jarzemska, K. N. & Dominiak, P. M. (2012). *Acta Cryst.* **A68**, 139–147.
 Jarzemska, K. N., Kamiński, R., Durka, K. & Kubsik, M. (2017). *Cryst. Growth Des.* **17**, 6836–6851.
 Jarzemska, K. N., Kubsik, M., Kamiński, R., Woźniak, K. & Dominiak, P. M. (2012). *Cryst. Growth Des.* **12**, 2508–2524.
 Jayatilaka, D. & Dittrich, B. (2008). *Acta Cryst.* **A64**, 383–393.
 Jayatilaka, D. & Grimwood, D. J. (2001). *Acta Cryst.* **A57**, 76–86.
 Jelsch, C., Guillot, B., Lagoutte, A. & Lecomte, C. (2005). *J. Appl. Cryst.* **38**, 38–54.
 Jha, K. K., Gruza, B., Kumar, P., Chodkiewicz, M. L. & Dominiak, P. M. (2020). *Acta Cryst.* **B76**, 296–306.
 Jha, K. K., Gruza, B., Sypko, A., Kumar, P., Chodkiewicz, M. L. & Dominiak, P. M. (2022). *J. Chem. Inf. Model.* **62**, 3752–3765.
 Johnson, C. K. (1969). *Acta Cryst.* **A25**, 187–194.
 Kendall, R. A., Dunning, J. T. H. & Harrison, R. J. (1992). *J. Chem. Phys.* **96**, 6796–6806.
 Koritsánszky, T., Volkov, A. & Chodkiewicz, M. (2012). *Struct. Bond.* **147**, 1–25.
 Kuhs, W. F. (1983). *Acta Cryst.* **A39**, 148–158.
 Kumar, P., Bojarowski, S. A., Jarzemska, K. N., Domagała, S., Vanommeslaeghe, K. Jr, MacKerell, A. D. Jr & Dominiak, P. M. (2014). *J. Chem. Theory Comput.* **10**, 1652–1664.
 Kumar, P., Gruza, B., Bojarowski, S. A. & Dominiak, P. M. (2019). *Acta Cryst.* **A75**, 398–408.
 Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785–789.

- Lübben, J., Wandtke, C. M., Hübschle, C. B., Ruf, M., Sheldrick, G. M. & Dittrich, B. (2019). *Acta Cryst.* **A75**, 50–62.
- Macrae, C. F., Sovago, I., Cottrell, S. J., Galek, P. T. A., McCabe, P., Pidcock, E., Platings, M., Shields, G. P., Stevens, J. S., Towler, M. & Wood, P. A. (2020). *J. Appl. Cryst.* **53**, 266–235.
- Madsen, A. Ø. (2006). *J. Appl. Cryst.* **39**, 757–758.
- Malaspina, L. A., Wieduwilt, E. K., Bergmann, J., Kleemiss, F., Meyer, B., Ruiz-López, M. F., Pal, R., Hupf, E., Beckmann, J., Piltz, R. O., Edwards, A. J., Grabowsky, S. & Genoni, A. (2019). *J. Phys. Chem. Lett.* **10**, 6973–6982.
- Malińska, M., Jarzemska, K. N., Goral, A. M., Kutner, A., Woźniak, K. & Dominiak, P. M. (2014). *Acta Cryst.* **D70**, 1257–1270.
- Mandal, S. K., Guillot, B. & Munshi, P. (2020). *CrystEngComm*, **22**, 4363–4373.
- Mazur, L., Jarzemska, K. N., Kamiński, R., Hoser, A. A., Madsen, A. Ø., Pindelska, E. & Zielińska-Pisklak, M. (2016). *Cryst. Growth Des.* **16**, 3101–3112.
- Meindl, K. & Henn, J. (2008). *Acta Cryst.* **A64**, 404–418.
- Nassour, A., Domagała, S., Guillot, B., Leduc, T., Lecomte, C. & Jelsch, C. (2017). *Acta Cryst.* **B73**, 610–625.
- Perdew, J. P. (1986). *Phys. Rev. B*, **33**, 8822–8824.
- Petříček, V., Dušek, M. & Palatinus, L. (2014). *Z. Kristallogr. Cryst. Mater.* **229**, 345–352.
- Řezáč, J., Riley, K. E. & Hobza, P. (2011a). *J. Chem. Theory Comput.* **7**, 3466–3470.
- Řezáč, J., Riley, K. E. & Hobza, P. (2011b). *J. Chem. Theory Comput.* **7**, 2427–2438.
- Rigaku Oxford Diffraction (2016). *CrysAlis PRO*, Version 1.171.39.43d. Rigaku Oxford Diffraction, Yarnton, England.
- Sanjuan-Szklarz, W. F., Hoser, A. A., Gutmann, M., Madsen, A. Ø. & Woźniak, K. (2016). *IUCrJ*, **3**, 61–70.
- Sanjuan-Szklarz, W. F., Woińska, M., Domagała, S., Dominiak, P. M., Grabowsky, S., Jayatilaka, D., Gutmann, M. & Woźniak, K. (2020). *IUCrJ*, **7**, 920–933.
- Scheringer, C. (1985). *Acta Cryst.* **A41**, 79–81.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Sheldrick, G. M. (2015a). *Acta Cryst.* **A71**, 3–8.
- Sheldrick, G. M. (2015b). *Acta Cryst.* **C71**, 3–8.
- Spackman, M. A. (2018). *CrystEngComm*, **20**, 5340–5347.
- Stewart, R. F. (1976). *Acta Cryst.* **A32**, 565–574.
- Su, Z. & Coppens, P. (1998). *Acta Cryst.* **A54**, 646–652.
- Tanaka, K. (2018). *Acta Cryst.* **A74**, 345–356.
- Volkov, A., Koritsanszky, T. & Coppens, P. (2004a). *Chem. Phys. Lett.* **391**, 170–175.
- Volkov, A., Li, X., Koritsanszky, T. & Coppens, P. (2004b). *J. Phys. Chem. A*, **108**, 4283–4300.
- Volkov, A., Macchi, P., Farrugia, L., Gatti, C., Mallinson, P., Richter, T. & Koritsanszky, T. (2016). *XD2016*. University at Buffalo, State University of New York, NY, USA; University of Milan, Italy; University of Glasgow, UK; CNR-ISTM, Milan, Italy; Middle Tennessee State University, TN, USA; and Freie Universität, Berlin, Germany.
- 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.* **D63**, 160–170.
- 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.