Journal of Molecular Structure 1209 (2020) 127934

Contents lists available at ScienceDirect

Journal of Molecular Structure

journal homepage: http://www.elsevier.com/locate/molstruc

Post-Hartree-Fock methods for Hirshfeld atom refinement: are they necessary? Investigation of a strongly hydrogen-bonded molecular crystal

Erna K. Wieduwilt [a], Giovanni Macetti [a], Lorraine A. Malaspina [b], Dylan Jayatilaka [c], Simon Grabowsky [d], Alessandro Genoni [a, *]

[a] Université de Lorraine & CNRS, Laboratoire de Physique et Chimie Théoriques (LPCT), UMR CNRS 7019, 1 Boulevard Arago, F-57078, Metz, France
[b] Institut für Anorganische Chemie und Kristallographie, Fachbereich 2 – Biologie/Chemie, Universität Bremen, Leobener Straße 3, 28359, Bremen, Germany
[c] University of Western Australia, School of Molecular Science, 35 Stirling Highway, Perth, WA, 6009, Australia
[d] Departement für Chemie und Biochemie, Universität Bern, Freiestrasse 3, 3012, Bern, Switzerland

ARTICLE INFO

Article history:
Received 13 December 2019
Received in revised form
24 January 2020
Accepted 18 February 2020
Available online 22 February 2020

ABSTRACT

Hirshfeld atom refinement (HAR) is a method for refining X-ray crystal structures that is able to provide bond lengths involving hydrogen atoms in statistical agreement with those derived from neutron diffraction data, provided the data reach 0.8 Å resolution. Rather than using tabulated spherical atomic structure factors, HAR uses "tailor made" aspherical atomic structure factors obtained directly from quantum chemical calculations. Despite the very good results obtained so far, which make HAR an emerging refinement method of modern crystallography, until now all the Hirshfeld atom refinements were exclusively based on Hartree-Fock (HF) or density functional theory (DFT) methods, but never on the so-called post-HF techniques of quantum chemistry. Post-HF methods exploit more sophisticated multi-determinant wavefunctions and, consequently, should provide more accurate electron densities for the refinements. For this reason, for the first time we have performed HARs based on two well-known post-HF strategies (MP2 and Coupled Cluster) combined with three different basis-sets (def2-SVP, def2-TZVP and def2-TZVPP). The obtained results have been afterwards analyzed and compared to those resulting from neutron and other Hirshfeld atom refinements, the latter relying on Hartree-Fock and DFT (BLYP and B3LYP) calculations in order to evaluate if the use of more sophisticated and expensive approaches of quantum chemistry can improve the performances of the HAR technique.

1. Introduction

Accurate molecular structures are a fundamental starting point for determining the properties of molecules, which define the suitability or "activity" of a molecule for particular desired purposes. For example, the strong connection between molecular structure and activity is fruitfully exploited in Quantitative Structure-Activity Relationship (QSAR) regression models [1], which are widely used in chemical and biological sciences, and in engineering.

In this context, X-ray diffraction is undoubtedly the main experimental technique used to determine the structure of molecules that crystallize [2,3]. Every year, thousands of crystal structures, comprised of atomic coordinates and displacement parameters, are routinely deposited in crystallographic databanks and, more importantly, the X-ray diffraction technique has led to important breakthroughs in science, e.g. the discovery and determination of the DNA structure [4].

However, despite the general applicability and utility of the X-ray diffraction method, it is often neglected that most "experimental" crystal structures rely on a very basic approximation: the independent atom model [5] (IAM), according to which the underlying model electron density is given by a sum of spherical atomic electron density distributions (EDDs) derived from calculated quantum mechanical wavefunctions for the uncharged atoms. The IAM does not take into account the distortions in the EDD around atoms that are due to chemical bonding. These chemical bonding effects comprise a contraction of the electron density near

* Corresponding author.
E-mail address: Alessandro.Genoni@univ-lorraine.fr (A. Genoni).

https://doi.org/10.1016/j.molstruc.2020.127934
0022-2860/

the nucleus, as well as a polarization of the EDD toward more electronegative atoms.

The most important consequence of the shortcomings in the IAM is the inability of this model to accurately determine the positions and the anisotropic displacement parameters (APDs) of hydrogen atoms. This is not a negligible detail because the structural data related to the hydrogen atoms are crucial in many fields of chemistry, such as the study of enzymatic and organic reactions [6,7], evaluation of thermodynamic properties [8], materials science [9], supramolecular chemistry and crystal engineering [10,11]. The intrinsic limitation of the IAM is due to the fact that X-rays are diffracted by electrons, and hydrogen atoms possess only one electron that is involved in covalent bonding and that hardly scatters X-rays. As a consequence, it is generally impossible to refine hydrogen atom positions accurately and precisely, unless some account is taken for the contraction of the electron density of the bonded hydrogen atoms compared to the EDD of the isolated atoms, as done for example by Stewart and coworkers in their 1965 seminal paper [12]. Consequently, hydrogen-element bond distances obtained with the IAM are unavoidably underestimated compared to the corresponding neutron diffraction results.

Neutron diffraction measurements can solve the problem of hydrogen atom localization because neutrons are strongly diffracted by hydrogen nuclei. The recent construction of new spallation sources indicates the importance of correctly detecting the positions of hydrogen atoms (e.g., for the determination of protonation states of amino acid side chains) [13]. However, neutron diffraction requires nuclear reactors or spallation facilities, which are rare and expensive to run. Because of these issues, X-ray diffraction remains the preferred experimental technique for the determination of crystal structures and, therefore, different strategies have been proposed to overcome the shortcomings of the IAM in obtaining structural parameters associated with the hydrogen atoms.

Concerning bond distances, the simplest technique consists in adjusting the bond lengths that involve hydrogen atoms according to tabulated average values [14,15]. This works well for hydrogen atoms in “standard” environments, but these are often hydrogen atoms of least interest. For thermal parameters, Madsen proposed the rigid-body SHADE approach that permits hydrogen ADPs to be estimated through comparisons of the ADPs in molecules with similar neutron structures [16,17]. Madsen and coworkers have also recently developed a very advanced quantum chemistry-based method called *NoMoRe* (Normal Mode Refinement) for the determination of hydrogen and non-hydrogen ADPs [8,18].

More sophisticated approaches of quantum crystallography [19–23] go beyond the simple independent atom model by introducing the asphericity of the atomic electron density distributions due to chemical bonds. For example, it was observed that, using the currently available databanks of multipole model (MM) pseudoatoms [24–32], better hydrogen atom positions can be obtained, with X–H bond lengths that differ by only a few hundredths of an Ångstrom from the corresponding and tabulated neutron values [33–37]. Another noteworthy early example of this kind of refinement is the charge density study of *syn*-1,6:8,13-biscarbonyl[14] annulene at 19 K by Destro & Merati [38]. They precisely and accurately determined the position of hydrogen atoms by exploiting the approach suggested by Stewart & Bentley in 1975 [39], which consisted in the introduction of a finite multipole expansion for the hydrogen atom obtained from its atomic electron density in molecular hydrogen. However, the corresponding hydrogen ADPs can be accurately obtained using these kinds of multipole models in only few cases, as in the charge density investigation conducted by Zhurov et al. on the explosive hexaydro-1,3,5-trinitro-1,3,5-triazine [40]. Therefore, the application of the multipolar method is not generally considered a feasible way to obtain accurate ADPs for hydrogen atoms [41].

Another recent and promising quantum crystallography method is the Hirshfeld atom refinement (HAR) [42–47] technique, which is the main subject of this paper and which strongly relies on quantum chemistry calculations (for more details, see the Methods section).

Since its development, HAR has been thoroughly tested. Importantly, it was observed that, even using X-ray diffraction data at resolutions as low as 0.8 Å, HAR permits to locate the positions of hydrogen atoms as accurately and precisely as from neutron diffraction measurements [45]. Pertaining to ADPs, a more recent investigation using high resolution X-ray diffraction data has shown that, for the non-hydrogen atoms, HAR provides values as accurate and precise as those from multipole model refinements or neutron diffraction data, while, for the hydrogen ones, it gives ADPs with slightly larger discrepancies from the neutron-measured values compared to those from the SHADE model [47,48].

Initially developed by Jayatilaka and Dittrich in 2008 [42] and implemented in the quantum crystallography [19–23] software *Tonto* [49], HAR has also been improved and extended over the years. For example, the *Tonto* code has been recently interfaced with the *Olex2* program [47] and the HAR technique has been also coupled to the X-ray constrained wavefunction (XCW) strategy [50–57] giving rise to the so-called X-ray wavefunction refinement (XWR) method [58,59] that allows to refine both structural and electronic structure parameters (i.e., coefficients of molecular orbitals) against experimental X-ray diffraction data. However, in the context of XWR, HAR has been used only in connection with Hartree-Fock and Density functional theory (DFT) XCW strategies, but, in the future, the X-ray wavefunction refinement could be extended to other types of XCW techniques, such as those based on Extremely Localized Molecular Orbitals (ELMOs) [60–65], or the more recent X-ray constrained spin-coupled approach [66,67]. Therefore it is important, as a first step towards extended XWR techniques, to develop, implement and test post-HF HARs.

So far all HARs have exploited only quantum mechanical calculations at Hartree-Fock (HF) or DFT(BLYP) level (also in the relativistic framework [68,69]). Other flavors of HAR exist, but still at HF or DFT level. One of them consists in performing Hirshfeld atom refinements along with periodic *ab initio* calculations carried out through VASP [46]. In that study it was shown that HAR gives electron densities that are in very good agreement with the experimental data and simultaneously with theoretical benchmark calculations.

Another noteworthy and newly released extension of HAR exploits the intrinsic transferability of the extremely localized molecular orbitals (ELMOs) [70–76] and consists in its coupling to the recently constructed libraries of ELMOs [77–79], which led to the novel HAR-ELMO method [80]. This technique enables to significantly reduce the intrinsically high computational cost of HAR by replacing the necessary quantum chemical calculation at each HAR-iteration with the instantaneous reconstruction of the wavefunction for the system under exam through the transfer of ELMOs from the ELMO libraries. Consequently, by applying the HAR-ELMO technique, refinements of macromolecular crystal structures (e.g., polypeptides and proteins) and of organometallic compounds are now possible at reduced computational cost [80]. Preliminary test calculations have shown that, despite the approximation introduced through the use of transferred and frozen ELMOs, accuracy and precision for structural parameters are only slightly affected. Derived properties of large systems could also be obtained, such as electrostatic potentials and electron localizability indicator (ELI-D) [81] plots or non-covalent interaction (NCI) maps exploiting the novel NCI-ELMO strategy [82]. In this framework, extension of HAR

to the recently developed multi-scale embedding technique QM/ELMO [83] is already envisaged; this will allow the refinement of important regions of proteins (e.g., active sites) exploiting higher levels of theory, but keeping the rest of the system treated by means of frozen ELMOs.

Nevertheless, notwithstanding the Hirshfeld atom refinement method has been on the scene for about ten years, until now it has never been exploited in combination with correlated post-Hartree-Fock techniques of quantum chemistry. These strategies are based on a better (multi-determinant) wavefunction *ansatz* compared to the Hartree-Fock method and provide better EDDs that might lead to improved structural parameters. To test this assertion, in this paper we present the first HARs on the strongly hydrogen-bonded crystal of L-alanine using the second-order Møller-Plesset (MP2) and Coupled Cluster with single and double excitations (CCSD) techniques.

The paper is organized as follows: after a section where we will briefly review the main algorithmic features of HAR and where we will describe in detail the computational techniques and procedures used in this investigation, we will show and discuss the obtained results focusing on crystallographic figures of merits, residual and deformation densities, bond lengths involving hydrogen atoms, hydrogen ADPs and comparisons of electron densities. The final section of the paper will be dedicated to draw brief and general conclusions.

2. Theory and computational methods

The HAR algorithm and its coupling to post-Hartree-Fock methods. The Hirshfeld atom refinement is illustrated in Fig. 1. It is a self-consistent refinement procedure consisting in the following steps.

1. Single-point quantum mechanical calculation on a molecular system that must correspond at least to the asymmetric unit of the crystal structure under examination. The starting geometry at the first iteration is generally the result of a preliminary IAM refinement.
2. Partitioning of the obtained molecular electron density into atomic contributions exploiting the Hirshfeld stockholder partitioning method [84,85].
3. Fourier transform of the atomic densities obtained at point 2 to obtain the atomic scattering factors and subsequent computation of the global structure factors.
4. Least-squares minimization of the statistical agreement (χ^2) between calculated and experimental structure factor magnitudes with respect to a scale factor η (see equation (1) below) and the atomic structural parameters (i.e., atomic coordinates and ADPs). The χ^2 statistical agreement is given by this expression:

$$\chi^2 = \frac{1}{N_r - N_p} \sum_{\boldsymbol{h}} \frac{\left(\eta \left|F_{\boldsymbol{h}}^{calc}\right| - \left|F_{\boldsymbol{h}}^{exp}\right|\right)^2}{\sigma_{\boldsymbol{h}}^2} \tag{1}$$

where N_r is the number of reflections, N_p the number of parameters in the model, $\left|F_{\boldsymbol{h}}^{calc}\right|$ and $\left|F_{\boldsymbol{h}}^{exp}\right|$ respectively the calculated and experimental structure factor magnitudes for the reflection characterized by the Miller indices triad $\boldsymbol{h} = (h,k,l)$, $\sigma_{\boldsymbol{h}}$ the experimental uncertainty associated with the experimental structure factor magnitude $\left|F_{\boldsymbol{h}}^{exp}\right|$, and η an overall ($\boldsymbol{h}$-independent) factor that puts $\left|F_{\boldsymbol{h}}^{calc}\right|$ on the same scale of $\left|F_{\boldsymbol{h}}^{exp}\right|$.

5. Check of convergence on the atomic structural parameters. If convergence is reached, the iterative procedure is terminated and electronic properties of the system may be computed exploiting the wavefunction obtained on the converged geometry. Otherwise, the new geometry is used to start a new iteration and, particularly, to perform a new quantum mechanical calculation as indicted at point 1.

As mentioned in the Introduction, the iterative self-consistent procedure described above can be carried out through the quantum crystallographic software *Tonto* [49]. However, so far, only the Hartree-Fock strategy and the DFT functionals BLYP and B3LYP are available in *Tonto*. Therefore, to perform Hirshfeld atom refinements in combination with post-Hartree-Fock methods, *Tonto* must be coupled with an external quantum chemistry package, here *Gaussian09* [86]. This is done within the software *lamaGOET* [87] (http://www.tinyurl.com/lamaGOET; https://github.com/lomalaspina/lamaGOET) and can be used generally for HARs exploiting other software for the wavefunction calculation. In our

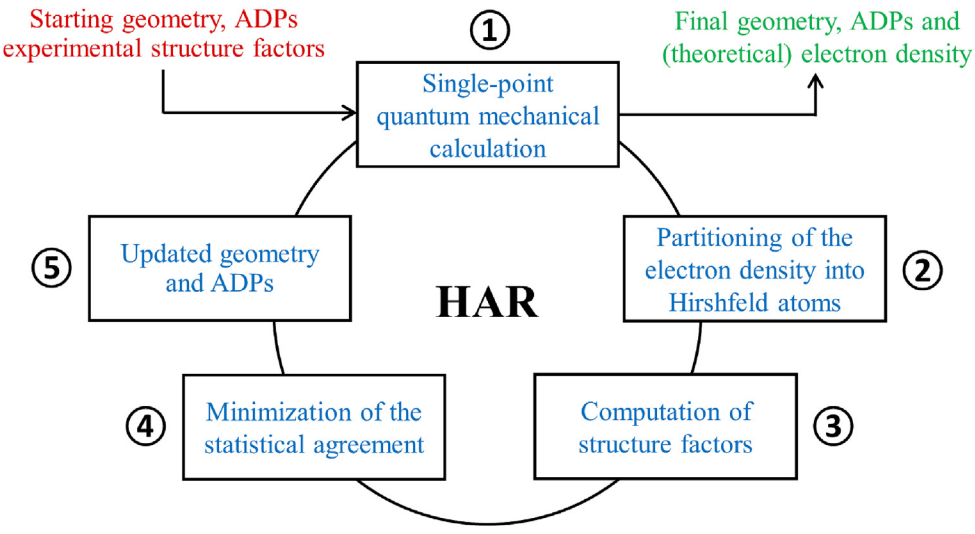

Fig. 1. Schematic representation of the Hirshfeld atom refinement algorithm.

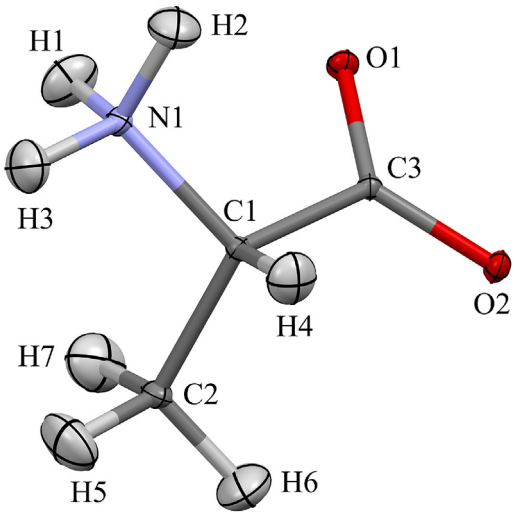

Fig. 2. Neutron structure of L-alanine with atomic labels [80]. Anisotropic displacement parameters (ADPs) shown with 50% probability.

case it was used i) to pass the molecular geometry at the current iteration to *Gaussian09*, ii) to submit the quantum mechanical calculation and, finally, iii) to provide the obtained EDD (in the form of the one-electron density matrix) to *Tonto*, which afterwards executed steps 2–5 of the HAR cycle. To obtain compatible and comparable results, in this study all the HARs including the traditional HAR-HF, HAR-BLYP and HAR-B3LYP refinements were also performed by interfacing the *Tonto* and *Gaussian09* software programs through the *lamaGOET* interface mentioned above.

X-ray data. The compound considered in our investigation is the amino acid L-alanine (see Fig. 2). This molecule was selected for different reasons: (i) it is reasonably small to allow high-level wavefunction calculations (e.g. Coupled-Cluster); (ii) X-ray and neutron diffraction data were available at the same very low temperature (23 K); (iii) the X-ray dataset is known to be of such a high quality that it is often used for benchmarking new refinement strategies. In particular, the X-ray data were collected by Destro and coworkers up to a resolution of 0.46 Å [88], while the reference neutron data were collected up to a resolution of 0.50 Å [80]. For all our refinements of X-ray data, we selected only those reflections for which $\left|F_{\boldsymbol{h}}^{exp}\right| > 4\sigma_{\boldsymbol{h}}$.

Hirshfeld atom refinements. In this work the Hirshfeld atom refinements were performed at different levels of theory: Hartree-Fock, BLYP, B3LYP, MP2 and CCSD. In combination with these five quantum chemical methods, we used the basis-sets def2-SVP, def2-TZVP and def2-TZVPP, which were chosen according to a basis-set ranking for HAR proposed by Fugel et al. [47], who classified def2-SVP as adequate, def2-TZVP as excellent and def2-TZVPP as a possible benchmark for Hirshfeld atom refinements. We also performed an IAM refinement of the available X-ray data exploiting the software *Tonto*. Finally, the structural parameters resulting from the refinements of the neutron diffraction data were used as benchmark values to evaluate and compare the quality of the different X-ray diffraction refinements mentioned above (HARs and IAM). The bond lengths (involving hydrogen atoms) obtained by Destro and coworkers through different multipole model (MM) refinements [88,89] on the collected data were also considered for comparison.

3. Results and discussion

To evaluate the effect of using different quantum mechanical methods in the HAR machinery, we decided to compare all the performed refinements on the basis of the following quantities: a) structure factor-based quantities, such as crystallographic figures of merit along with deformation and residual densities; b) bond lengths involving hydrogen atoms; and c) anisotropic displacement parameters. Finally, in the last subsection, to rationalize the performances of the different HARs, we will also consider and discuss observed differences in the electron densities resulting from the quantum mechanical calculations at the basis of the different refinements.

Structure factor-based quantities. In this first subsection we will compare the results of different refinements on the basis of structure factor-based descriptors. First of all, we will focus on the very general differences between all the HARs and the traditional IAM refinement. Afterwards, we will use the same descriptors to investigate more precisely the discrepancies between the different types of performed HARs, trying to evaluate both the influence of the adopted quantum chemical method and the effect of the used basis-set on the obtained results.

As is well known, the quality of a measurement and the capability of a model to describe features in the experimental data are usually measured in terms of crystallographic R-values and of the χ^2 agreement statistics. An optimal agreement between calculated and measured structure factors would lead to a χ^2 value of 1, (meaning that each structure factor is reproduced within one experimentally estimated standard deviation) although, in practice, χ^2 often converges at higher values. Therefore, in order to have experimental data in reasonable agreement with the theoretical model underlying the refinement, one should obtain a χ^2 value as close as possible to 1 and the lowest possible crystallographic R-values. Here, the χ^2 agreement statistics and the R-values for the refinements of L-alanine are reported in Table 1 and Table S1, respectively. In Table 1 it is easy to observe that χ^2 significantly decreases when HAR is used instead of the traditional independent atom model. For example, only considering the case of the def2-SVP basis-set, we have reductions that range from 3.630 at Hartree-Fock level up to 3.898 at B3LYP level. Similar trends have been also observed for the crystallographic R-values (see Table S1 in the Supporting Information).

The reason why all the different HARs give lower values for the figures of merit than the IAM refinement lies in the fact that high-resolution and high-quality X-ray diffraction data normally contain more information than those that one can model through the simple independent atom model. In fact, visualizing the so-called residual density after an IAM refinement, it is immediately possible to identify regions where the electron density is poorly modeled. In almost all the cases, these regions correspond to chemical bonds and lone pairs (see Fig. S1 for the L-alanine case considered in this study). On the contrary, after a Hirshfeld atom refinement, the residual density should be statistically distributed in the unit-cell and the unmodelled residuals of electron distribution in correspondence of chemical bonds and lone pairs should disappear (see again Fig. S1 for L-alanine). This happens because

Table 1
χ^2 agreement statistics for the Hirshfeld atom and independent atom model refinements performed on L-alanine.

χ^2	def2-SVP	def2-TZVP	def2-TZVPP
MP2	1.421	1.331	1.336
CCSD	1.418	1.296	1.302
HF	1.630	1.425	1.437
BLYP	1.364	1.411	1.414
B3LYP	1.362	1.318	1.323
IAM	5.260		

HAR accounts for aspherical features of the electron density associated with chemical bonding and non-covalent interactions. This is also reflected in the fractal distribution of the residual density (i.e., in the so-called Meindl-Henn plots [90], here obtained through the *jnk2RDA* software), which, in the ideal case, should follow a narrow parabolic distribution, with a balance between positive and negative values of the residual density and a maximum at null residual density.

Analyzing in detail the residual density maps in the plane of the carboxylate group of L-alanine for the different refinements carried out in the present study (see Fig. S1), it is evident that all the HAR plots show much less features than the IAM one, although, also for the HARs, the residuals do not seem completely randomly distributed. Furthermore, it is also worth pointing out that minimum and maximum values in the obtained HAR residual densities are significantly lower (in absolute value) than the IAM ones (see Table S2). For the sake of comparison, in Table S2, we have also reported the corresponding values obtained by Destro et al. through different multipole model refinements [88,89], values that are always larger than those resulting from all performed HARs, regardless of the underlying quantum chemical method and/or basis-set.

Examining the fractal distribution plots in Fig. 3, it is possible to see that the one associated with the IAM refinement does not show a parabolic shape. In fact, "shoulders" are present in the negative as well as in the positive region, and the plot is not symmetric. Conversely, for all the considered HARs, the plots are much narrower (note the different scale on the x-axis for the Meindl-Henn plots associated with the HARs) and show a symmetric parabolic trend.

To further highlight the differences between aspherical and independent atom models, deformation densities are also used. They reveal if the aspherical modeling of the electron distribution in the region of chemical bonds and lone pairs is chemically meaningful, which must be the case if we use a model based on an *ab initio* method of quantum chemistry as in HAR. Therefore, for the sake of completeness, in Fig. S2 of the Supporting Information, we have reported all the deformation densities in the plane of the carboxylate group of L-alanine. They were obtained from all the performed HARs. From the analysis of the plots, it emerges that all the Hirshfeld atom refinements led to models that are correct and sound from the chemical point of view, regardless of the underlying quantum chemical method or the adopted set of basis functions.

Always considering the quantities analyzed above, we will discuss in detail the effects of the basis-set choice on the HARs of the L-alanine crystal structure. With the only exception of the BLYP case, for all the other quantum chemistry methods considered in this study, the χ^2 value clearly decreases when passing from the split-valence basis-set def2-SVP to the triple-zeta basis-sets def2-TZVP and def2-TZVPP (see Table 1). For example, only considering def2-SVP and def2-TZVP, $\Delta\chi^2$ amounts to −0.090, −0.122, −0.205, +0.047 and −0.044 for MP2, CCSD, Hartree-Fock, BLYP and B3LYP, respectively. Concerning the results obtained with the triple-zeta basis-sets, the χ^2 variations are less pronounced, although def2-TZVPP always provides slightly larger values for the statistical agreement, with $\Delta\chi^2$ that ranges from +0.003 to +0.012. The same trends can be observed for the crystallographic R-values reported in Table S1 of the Supporting Information.

Deformation densities, residual densities and Meindl-Henn plots further confirm the higher similarity between the Hirshfeld atom refinements carried out with triple-zeta basis-sets. In fact, although the discrepancies among the deformation densities in the plane of the carboxylate group of L-alanine (Fig. S2) are quite small, the differences between the def2-SVP and the triple-zeta basis-sets are slightly more evident than those resulting from the use of the

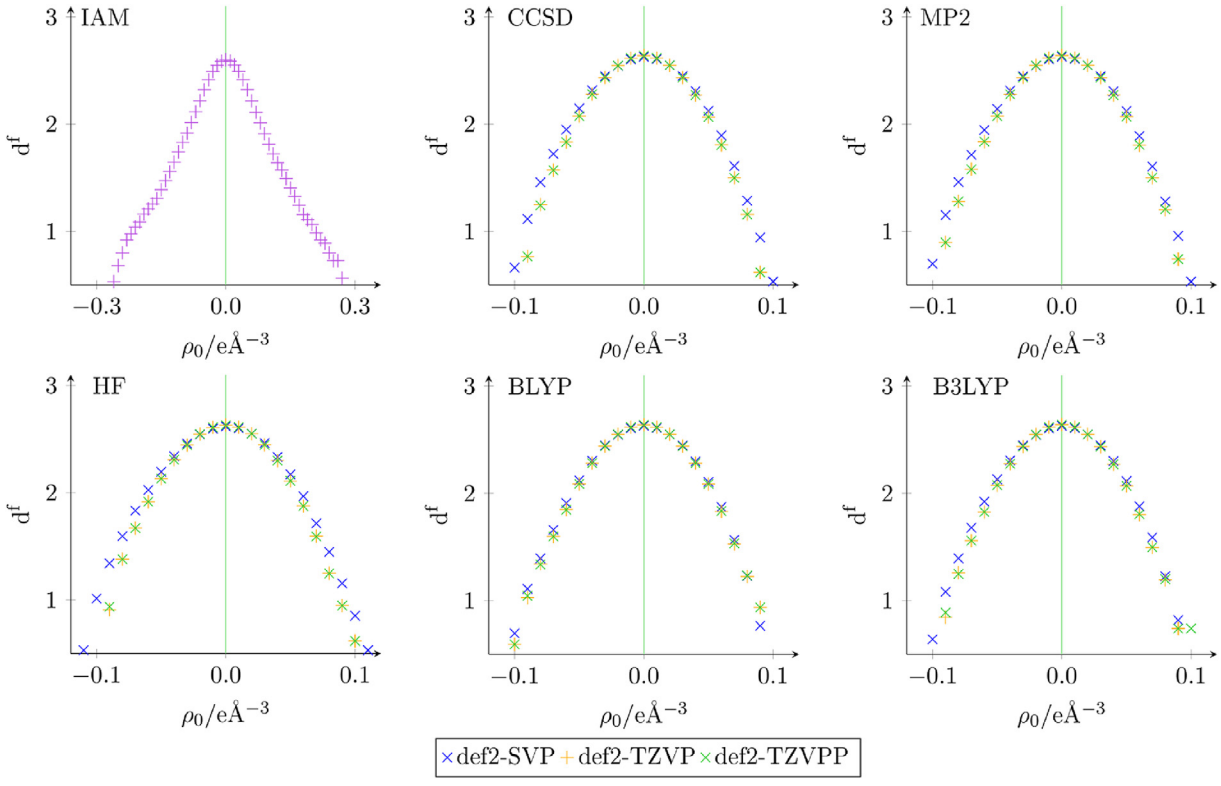

Fig. 3. Meindl-Henn fractal dimension distribution corresponding to the Hirshfeld atom and independent atom model refinements performed on L-alanine.

def2-TZVP and def2-TZVPP sets of basis functions. In analogous way, also the residual density maps (Fig. S1) show more features for refinements with def2-SVP than for the refinements with the triple-zeta basis-sets. This is further supported by the residual density distribution in the Meindl-Henn plots (Fig. 3), which is slightly broader for the refinements with def2-SVP than for those with the def2-TZVP or def2-TZVPP sets of functions. On the contrary, the differences between the residual density maps and between the Meindl-Henn plots corresponding to the HARs carried out with the triple-zeta basis-sets are very small.

On the basis of the results described above, it is possible to provisionally conclude that, overall, the use of triple-zeta basis-sets leads to higher agreements with experimental structure factors than the use of the double-zeta basis-set def2-SVP. However, the additional polarization functions in the def2-TZVPP basis-set do not significantly improve the figures of merit compared to the refinements with def2-TZVP. This is probably due to the fact that, considering the elements that are present in L-alanine, the only difference between the two triple-zeta basis-sets consists in the number and type of polarization functions for the hydrogen atoms. This is also the reason why the deformation and residual densities in the plane of the carboxylate group are barely affected when passing from def2-TZVP to def2-TZVPP. Concerning the fact that higher χ^2 and R-values are obtained for the def2-TZVPP basis-sets, we conclude that the extra polarization functions on the hydrogen atoms do not improve the agreement with the collected experimental data because the experiment is not sensitive to polarization of hydrogen atoms. Improving the basis-set does not always guarantee better results compared to experiments. It does guarantee only better results relative to the "exact results" within a particular theoretical method.

As explained in the Introduction, the main goal of this study is to evaluate the performances of HAR when coupled with different methods of quantum chemistry, especially when coupled with correlated post-Hartree-Fock techniques. In this paragraph, we will start discussing this aspect in relation to the structure factor-based quantities considered above. Analyzing again Tables 1 and S1, we can observe that, for the def2-SVP basis-set, the refinements using DFT functionals provide final χ^2 and R-values that are lower than those obtained through refinements based on correlated post-HF methods. In fact, while HAR-B3LYP and HAR-BLYP yield χ^2 values of about 1.36, the corresponding CCSD and MP2 Hirshfeld atom refinements give values of the statistical agreement that approximately amount to 1.42. Again, a different trend is observed for the triple-zeta basis-sets, with the HAR-CCSD, HAR-MP2 and HAR-B3LYP providing better figures of merit compared to the HAR-BLYP and HAR-HF cases.

Pertaining to the deformation densities (see Fig. S2), the maps in the plane of the carboxylate group of L-alanine are almost identical for all the HARs coupled with post-Hartree-Fock methods (MP2 and CCSD) and with HF, whereas the two HAR-DFT refinements provide slightly different deformation densities, especially in the regions of the oxygen atoms.

At the chosen contour levels, the plots of the residual densities (Fig. S1) are similar for all the considered quantum chemical methods except for HAR-HF, which shows systematically larger regions of negative residual densities (red contour lines) for all three adopted basis-sets. This trend is also observed in the Meindl-Henn plots (see Fig. 3 and Fig. S3), which are very similar for all the refinements, with the only exception of the HAR-HF ones that provide slightly broader plots compared to those associated with the DFT and post-HF strategies.

Finally, in conclusion of this section, if we simultaneously consider the effects of the used quantum chemical methods and those of the adopted basis-sets, we can see that the highest agreements between experimental and theoretical structure factors are achieved for HAR-CCSD in combination with the def2-TZVP and def2-TZVPP sets of basis functions, although HAR-B3LYP and HAR-MP2 with the triple-zeta basis-sets also lead to similar χ^2 values. On the whole, the agreement between experimental and theoretical structure factors is higher for HARs that are based on post-Hartree-Fock methods and triple-zeta basis-sets, which is also confirmed by lower R-values and narrower Meindl-Henn plots.

Bond lengths involving hydrogen atoms. As explained in the Introduction, the main advantage of HAR over IAM derives from the fact that the former allows the determination of bond lengths involving hydrogen atoms that are usually in very good agreement with those obtained through neutron diffraction measurements. This is often shown for averaged bond lengths. However, since in L-alanine only four C—H and three N—H bonds are present, here we decided to base our discussions and conclusions on the direct comparison of the obtained HAR bond lengths with the corresponding neutron diffraction derived ones, as shown in Figs. 4 and 5 for the C—H and N—H bond lengths, respectively. Note that, in case of a perfect agreement with the neutron results, all the values would lie on the diagonal of the depicted plots. For the sake of completeness, in Tables S3 and S4 of the Supporting Information, we have also reported averaged values, averaged mean absolute deviations and averaged ratios between X-ray and neutron bond lengths.

For all the refinements of L-alanine, HAR performs better for the C—H bond lengths. In fact, comparing the plots in Fig. 4 to those in Fig. 5, the C—H bond lengths from HAR are closer to the diagonal than the N—H ones, which were underestimated by all our refinements. The reason is probably related to the fact that all the hydrogen atoms bonded to nitrogen are involved in hydrogen bonds in the crystal, which are not properly taken into account by the single-molecule quantum chemical calculations underlying all the performed Hirshfeld atom refinements. Improvement might be obtained by using a cluster of point charges and dipoles or by performing cluster calculations, as shown for example by Fugel et al. [47] for the treatment of systems characterized by strong hydrogen bonds. Another possibility might consist in exploiting the recently developed QM/ELMO approach [83] to carry out embedded HARs, where a quantum chemical calculation is performed on the molecule of interest embedded in an environment described by frozen molecular orbitals. Nevertheless, HAR is anyway in much better agreement with the neutron data than the IAM refinement, for which the corresponding bond lengths are reported in the captions of Figs. 4 and 5 for the C—H and N—H bond lengths, respectively.

In Figs. 4 and 5, we have also reported the C—H and N—H bond lengths obtained by Destro and coworkers through a multipole model refinement of the collected data [89] that exploited hydrogen ADPs computed with a model based on approximate vibrational modes [91]. The MM-derived C—H and N—H bond lengths are completely comparable to those resulting from the performed HARs, with overlapping standard deviations in all cases. However, it is worth noting that, considering the original multipole model refinement of the collected data with isotropic treatment of the thermal motion for the hydrogen atoms [88], C—H and N—H bond lengths are sometimes significantly longer than the corresponding neutron and HAR values (see the MM bond length values given in the captions of Figs. 4 and 5).

Concerning the effect of the basis-set choice in the refinement, from Fig. 4 it is easy to see that three C—H bond lengths are shorter for refinements with def2-SVP than for those with def2-TZVP, whereas the C2—H7 bond length becomes longer (see Fig. 2 for the atom labels). From Fig. 5 it is not possible to observe a general trend for the individual N—H bond lengths, except for the fact that the

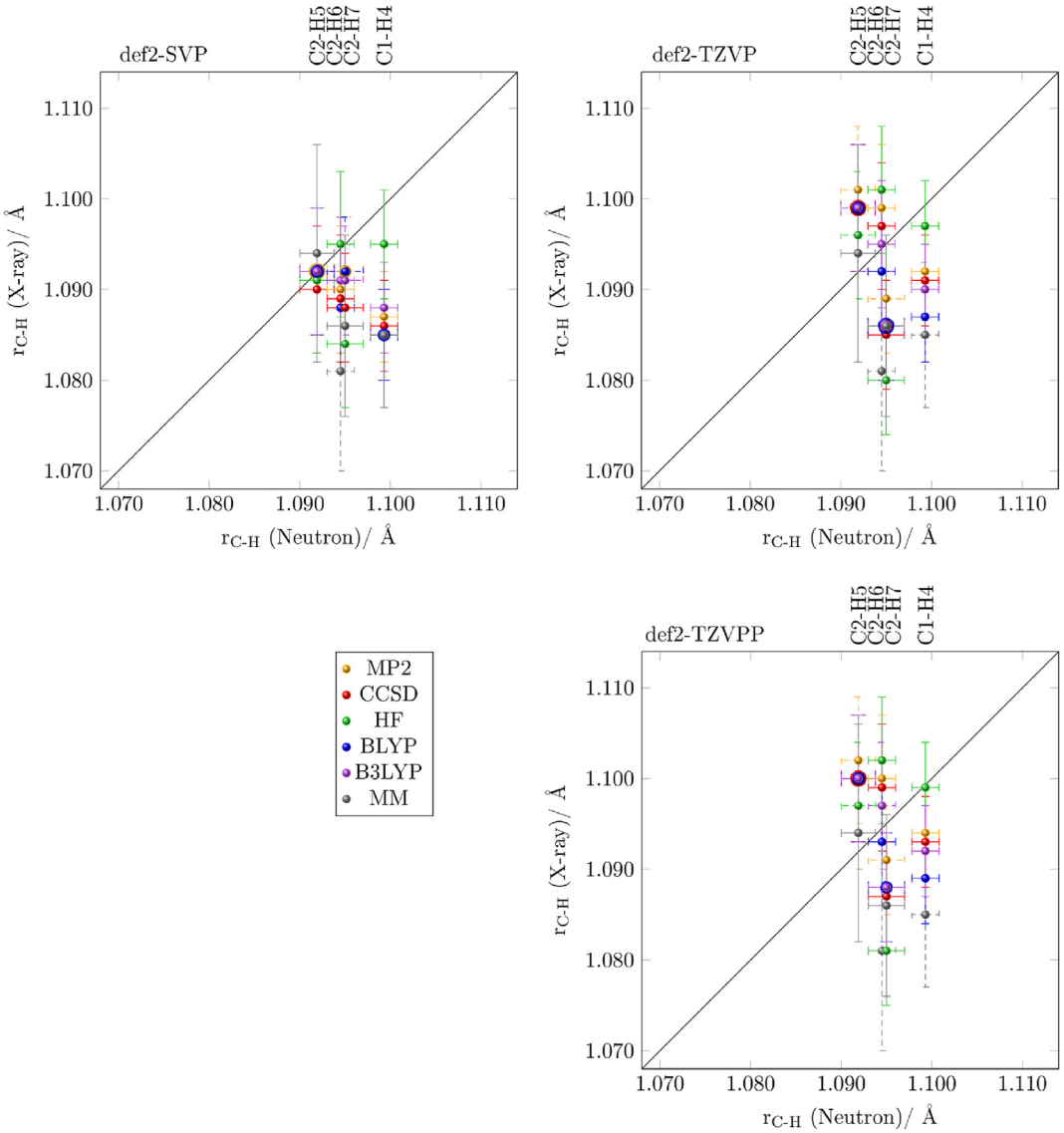

Fig. 4. C−H bond lengths in L-alanine for HARs using the def2-SVP (top left), def2-TZVP (top right) and def2-TZVPP (bottom right) basis-sets plotted against neutron C−H bond lengths. C−H bond lengths obtained by Destro and coworkers [89] through a multipole model refinement based on computed hydrogen ADPs are also shown. Vertical bars indicate the experimental errors for the X-ray (HAR or MM) bond lengths, while the horizontal ones represent the uncertainties associated with the neutron results. Bond lengths (in Å) from the IAM refinement: C1−H4: 0.999(9), C2−H5: 0.973(12), C2−H6: 0.997(11), C2−H7: 1.021(11). Bond lengths (in Å) from the MM refinement with isotropic treatment of hydrogen atoms performed by Destro et al. [88]: C1−H4: 1.125(9), C2−H5: 1.086(9), C2−H6: 1.118(9), C2−H7: 1.082(7).

refinement with triple-zeta basis-set leads to larger discrepancies between the methods than the def2-SVP refinements. Furthermore, both Figs. 4 and 5 indicate that basis-sets def2-TZVP and def2-TZVPP give very similar trends, although for the latter all bond lengths are slightly longer.

Now, we will analyze the effects of the quantum mechanical methods underlying the different refinements. First of all, it is worth bearing in mind that one should be careful with drawing conclusions according to the absolute values without considering the experimental errors. This is particularly true in the present study since, in almost all the cases, the bond lengths overlap within their standard deviations. However, despite this fact, it is still interesting to note some trends. For the most polarized bonds, namely all the N−H bonds and the C−H bond lengths that are in short contact with oxygen atoms (C1−H4 and C2−H6), we systematically observe that HF gives the longest bond lengths, which are nearly always the closest to the neutron-diffraction derived reference values. Furthermore, for the refinements adopting the triple-zeta basis-sets, we also noticed that the DFT methods give the shortest bond lengths, while MP2 and CCSD provide bond lengths that lie between the HF and DFT ones. Anyway, although the obtained results allowed us to define some trends, in almost all the situations, the bond lengths overlap within their experimental error. Therefore, at this stage, the main conclusion is that the differences between single- and multi-determinant HAR bond lengths are not statistically significant.

Atomic displacement parameters. As explained in the Introduction, while in standard IAM refinements the atomic displacement parameters of hydrogen atoms are often refined isotropically, HAR allows for anisotropic refinements. The obtained parameters can be thus compared to ADPs from neutron refinements. Since each atomic thermal tensor is described by six unique parameters, in this case the statistical analysis in terms of averaged mean absolute deviations and averaged ratios between X-ray and neutron

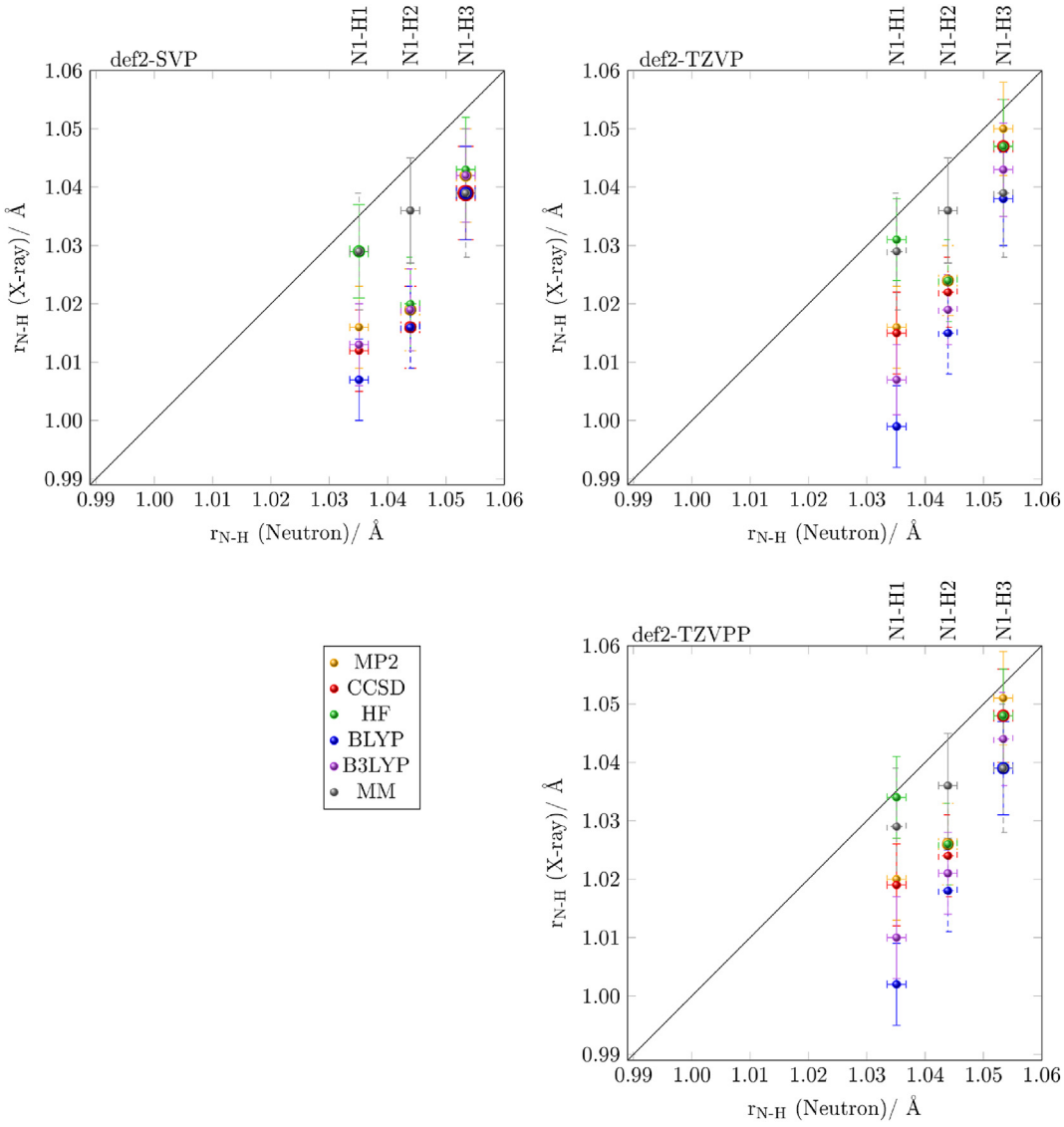

Fig. 5. N–H bond lengths in L-alanine for HARs using the def2-SVP (top left), def2-TZVP (top right) and def2-TZVPP (bottom right) basis-sets plotted against neutron N–H bond lengths. N–H bond lengths obtained by Destro and coworkers [89] through a multipole model refinement based on computed hydrogen ADPs are also shown. Vertical bars indicate the experimental errors for the X-ray (HAR or MM) bond lengths, while the horizontal ones represent the uncertainties associated with the neutron results. Bond lengths from the IAM refinement (in Å): N1–H1: 0.935(11), N1–H2: 0.917(10), N1–H3: 1.042(13). Bond lengths (in Å) from the MM refinement with isotropic treatment of hydrogen atoms performed by Destro et al. [88]: N1–H1: 1.044(10), N1–H2: 1.018(9), N1–H3: 1.082(10).

ADPs is more reliable than for the previously discussed bond lengths. Therefore, in Tables 2 and 3 we present the results of this analysis for hydrogen atoms bonded to carbon and nitrogen, respectively. The same quantities for non-hydrogen ADPs are given in Table S5 of the Supporting Information. In Fig. 6, we show a visual comparison of the crystal structures of L-alanine resulting from the neutron refinement and from the HARs with the basis-set def2-TZVPP (the analogous structures for the other basis-sets are given in Fig. S4 of the Supporting information).

From a visual point of view (Fig. 6), the HAR ADPs of hydrogen atoms are longer and flatter than the corresponding neutron ones. The different method and the basis-set used for HAR do have an effect on the sizes and the orientation of the thermal parameters, but the longer and flatter shape generally remains. As seen for the figures of merit and for the hydrogen bond lengths, larger discrepancies are again observed between the def2-SVP and triple-zeta basis-sets, whereas the hydrogen ADPs from refinements obtained with the def2-TZVPP set of basis functions are very similar to those resulting from the refinements with def2-TZVP (see Fig. S4).

What was discussed above is also reflected in the statistical analyses of the hydrogen ADPs (see Tables 2 and 3), which reveal that the mean differences between neutron and HAR ADPs are non-zero, although standard deviations are also quite large. Furthermore, considering the ratio between the diagonal-elements of HAR and neutron ADPs, they are generally above 1.0, which means that, on average, HAR ADPs are larger than neutron ADPs. Finally, comparing Tables 2 and 3, we can see that the HAR ADPs are in better agreement with the neutron ones if the hydrogen atom is bonded to carbon, whereas the agreement is worse if HAR ADPs for hydrogen atoms bonded to nitrogen are taken into account. The trends just described above are general and not influenced by the choice of the quantum chemical method and basis-set adopted for the Hirshfeld atom refinements. However, also in this case, we

Table 2
Statistical analysis of HAR ADPs for hydrogen atoms bonded to carbon: (i) mean ratios of the diagonal elements ($<U^{ii}_{HAR}/U^{ii}_{Neutron}>$); (ii) mean absolute differences of the diagonal terms ($<|U^{ii}_{HAR} - U^{ii}_{Neutron}|>$); (iii) mean absolute differences of the non-diagonal elements ($<|U^{ij}_{HAR} - U^{ij}_{Neutron}|>$ with $i \neq j$). The first column for each basis-set refers to the value of the quantity, the second column to the corresponding population standard deviation upon averaging.

	def2-SVP		def2-TZVP		def2-TZVPP	
$<U^{ii}_{HAR}/U^{ii}_{Neutron}>$						
MP2	1.02	0.17	1.07	0.24	1.10	0.24
CCSD	1.09	0.16	1.14	0.21	1.16	0.22
HF	1.21	0.21	1.28	0.18	1.29	0.19
BLYP	1.02	0.20	1.17	0.31	1.18	0.31
B3LYP	1.03	0.16	1.16	0.26	1.17	0.26
$<|U^{ii}_{HAR} - U^{ii}_{Neutron}|>$ /Å^2						
MP2	0.0035	0.0019	0.0049	0.0045	0.0051	0.0048
CCSD	0.0033	0.0026	0.0049	0.0048	0.0053	0.0050
HF	0.0043	0.0032	0.0064	0.0038	0.0066	0.0042
BLYP	0.0038	0.0030	0.0068	0.0066	0.0069	0.0067
B3LYP	0.0032	0.0023	0.0059	0.0059	0.0060	0.0061
$<|U^{ij}_{HAR} - U^{ij}_{Neutron}|>$ /Å^2 (with $i \neq j$)						
MP2	0.0033	0.0026	0.0029	0.0025	0.0031	0.0025
CCSD	0.0033	0.0027	0.0028	0.0025	0.0029	0.0025
HF	0.0049	0.0031	0.0040	0.0024	0.0041	0.0022
BLYP	0.0032	0.0026	0.0039	0.0025	0.0040	0.0025
B3LYP	0.0031	0.0024	0.0032	0.0025	0.0033	0.0026

Table 3
Statistical analysis of HAR ADPs for hydrogen atoms bonded to nitrogen: (i) mean ratios of the diagonal elements ($<U^{ii}_{HAR}/U^{ii}_{Neutron}>$); (ii) mean absolute differences of the diagonal terms ($<|U^{ii}_{HAR} - U^{ii}_{Neutron}|>$); (iii) mean absolute differences of the non-diagonal elements ($<|U^{ij}_{HAR} - U^{ij}_{Neutron}|>$ with $i \neq j$). The first column for each basis-set refers to the value of the quantity, the second column to the corresponding population standard deviation upon averaging.

	def2-SVP		def2-TZVP		def2-TZVPP	
$<U^{ii}_{HAR}/U^{ii}_{Neutron}>$						
MP2	1.23	0.68	1.24	0.59	1.30	0.61
CCSD	1.32	0.70	1.33	0.63	1.39	0.65
HF	1.45	0.97	1.48	0.88	1.52	0.88
BLYP	1.21	0.46	1.37	0.42	1.41	0.43
B3LYP	1.22	0.55	1.34	0.49	1.38	0.49
$<|U^{ii}_{HAR} - U^{ii}_{Neutron}|>$ /Å^2						
MP2	0.0075	0.0065	0.0068	0.0055	0.0071	0.0057
CCSD	0.0080	0.0064	0.0073	0.0057	0.0077	0.0061
HF	0.0109	0.0086	0.0093	0.0083	0.0094	0.0086
BLYP	0.0064	0.0036	0.0065	0.0057	0.0072	0.0058
B3LYP	0.0067	0.0049	0.0064	0.0051	0.0067	0.0054
$<|U^{ij}_{HAR} - U^{ij}_{Neutron}|>$ /Å^2 (with $i \neq j$)						
MP2	0.0074	0.0065	0.0085	0.0079	0.0086	0.0081
CCSD	0.0071	0.0063	0.0081	0.0075	0.0081	0.0077
HF	0.0076	0.0068	0.0079	0.0078	0.0079	0.0077
BLYP	0.0069	0.0057	0.0082	0.0084	0.0083	0.0084
B3LYP	0.0063	0.0060	0.0078	0.0079	0.0077	0.0080

believe that performing HARs with cluster calculations or carrying out embedded HARs based on the QM/ELMO approach might improve the results for the ADPs of the hydrogen atoms bonded to nitrogen.

Analyzing Tables 2 and 3 more in details and in relation to the effects of the basis-set choice, we can only observe a possible trend for the diagonal and non-diagonal ADPs of hydrogen atoms bonded to carbon and for the non-diagonal ADPs of hydrogen atoms bonded to nitrogen: the discrepancies with neutron ADPs are smaller for refinements that adopted the def2-SVP basis-set than for those that used triple-zeta sets of basis functions. Furthermore, the differences associated with the triple-zeta basis-sets refinements are generally very similar, especially for ADPs of hydrogen atoms bonded to carbons.

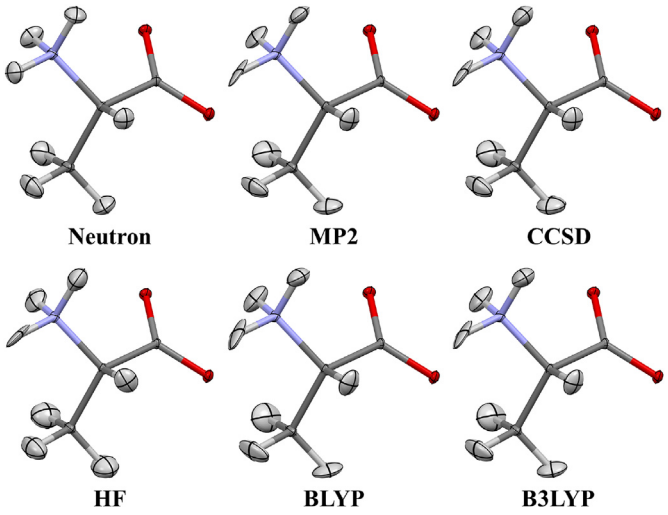

Fig. 6. Crystal structures of L-alanine obtained from the refinement of neutron data [80] and from Hirshfeld atom refinements of X-ray data [88] exploiting the basis-set def2-TZVPP. Anisotropic displacement parameters shown with 50% probability. The orientation of the molecules correponds to the one in Fig. 2.

Let us now focus on the effects of the underlying quantum mechanical method. Again from Tables 2 and 3, we can see that, when basis-set def2-SVP is used, the agreement with neutron ADPs improves for HARs exploiting post-Hartree-Fock and DFT methods compared to HAR-HF. However, for hydrogen atoms bonded to carbon, we cannot really establish a clear ranking among the different correlated quantum chemical techniques (namely, DFT and post-HF ones). For hydrogen atoms bonded to nitrogen, we can only say that the DFT functionals taken into account give lower discrepancies than the two considered post-Hartree-Fock methods, especially compared to CCSD.

Concerning triple-zeta basis-sets, we can notice different trends for the ADPs of hydrogen atoms bonded to carbon or nitrogen. For the former, the worst agreement with neutron ADPs is again observed for the HF-based refinements, although BLYP gives similar results and, for the mean absolute difference of the diagonal ADPs, it even provides the worst results. Moreover, correlated post-HF techniques (MP2 and CCSD) yield the lowest discrepancies, particularly for the mean absolute differences of both diagonal and off-diagonal ADPs. This is less clear for the ratio of the diagonal parameters. For the ADPs of hydrogen atoms bonded to nitrogen we need to distinguish between diagonal and non-diagonal thermal parameters. In the first case, HAR-HF is again the worst method in providing ADPs close to the neutron ones. The post-HF and DFT-based HARs improve the situation, even if it is again difficult to find a clear trend. Concerning non-diagonal parameters, B3LYP seems the best, but the discrepancies with the other methods are so small that clear conclusions cannot be drawn.

In summary, for the hydrogen ADPs we can conclude that HAR-HF seems to be slightly less accurate than the other refinements. However, it is important to bear in mind that, although trends among the different quantum mechanical methods can be observed, all the discussed differences between the different HARs are quite small, especially if compared with the corresponding standard deviations.

For the sake of completeness, we also analyze the results obtained for the non-hydrogen ADPs (see Table S5). It is possible to see that, for the diagonal parameters and independently of the chosen basis-set, HAR-HF always provides the best agreements and the post-HF methods always outperform the DFT ones. For the non-

diagonal terms, HAR-CCSD is always the best refinement in terms of agreement with neutron ADPs, while HAR-HF is the worst (in some cases together with HAR-BLYP). However, despite some rankings can be established, also in this case it is important to be aware of the large standard deviations.

Differences in the electron densities. Finally, in this subsection we will try to rationalize the results of the different Hirshfeld atom refinements on the basis of the electron densities obtained from the combinations of the quantum mechanical methods and basis-sets considered in this study. To accomplish this task, we calculated all the electron densities on the common starting structure of all the performed HARs, namely the IAM geometry. The obtained electron distributions have been afterwards compared using different references. To evaluate the effects of the quantum chemical methods, we considered the Hartree-Fock electron density as reference point for each basis-set; to assess the influence of the basis-sets choice, we used the def2-SVP charge distribution as reference point for each method. The reference densities were then subtracted from all the other densities. The resulting difference electron densities for the basis-sets comparison are depicted in Fig. 7, whereas the ones for the methods comparison are shown in Fig. 8. It is important to note that the reason for choosing the IAM structure in all the electron density calculations is that the first iteration of the HAR cycle is generally the most crucial one. This is because it directs the refinement towards the final converged structure. The (aspherical) electron density computed at that step on the initial geometry has the largest influence on the refinement.

From Fig. 7 we can observe that the choice of a double- or triple-zeta basis-set has an influence on the electron density of the whole molecule, even in the domains of the hydrogen atoms. It is especially worth noting that the def2-TZVP and def2-TZVPP difference densities are very similar, which explains the previously observed similarities between the HARs exploiting triple-zeta basis-sets.

On the contrary, as it can be seen in Fig. 8, the choice of the quantum chemical method mainly affects the electron densities of the non-hydrogen atoms. MP2 and CCSD electron densities differ from the Hartree-Fock ones almost only at the carboxylic group, whereas the ones derived from DFT calculations are different for all the non-hydrogen atoms. This is consistent with what we discussed for the non-hydrogen ADPs, where, especially for the diagonal terms, we noticed similarities between the post-HF HARs (MP2 and CCSD) and between the DFT ones (B3LYP and BLYP). The observed trends might be also related to the fact that DFT functionals sometimes deviate (even significantly) from the exact density, as recently pointed out by Medvedev and coworkers [92].

The fact that significant differences in the electron densities are not observed for the hydrogen atoms explains why we did not observe large discrepancies in the structural parameters of the same atoms obtained by means of the different HARs (see discussions above) and why it was sometimes difficult to establish trends.

4. Conclusions

In this paper we report the first Hirshfeld atom refinements based on aspherical atomic form factors from *ab initio* second order Møller-Plesset (MP2) and Coupled Cluster singles and doubles (CCSD) wavefunctions. These refinements were carried out on the strongly hydrogen-bonded L-alanine crystal structure using HF- and DFT-based HARs for comparison.

Focusing on the hydrogen atoms, for the largest basis-set (which provides the most meaningful MP2 and CCSD wave functions) we found that hydrogen atom bond lengths from the post-HF methods are not systematically improved relative to neutron diffraction results; however, the MP2 and CCSD results are very similar to each other, and all of the post-HF and DFT methods are roughly grouped together. These results are mirrored in plots of the electron densities for each of the wavefunctions, which show qualitative

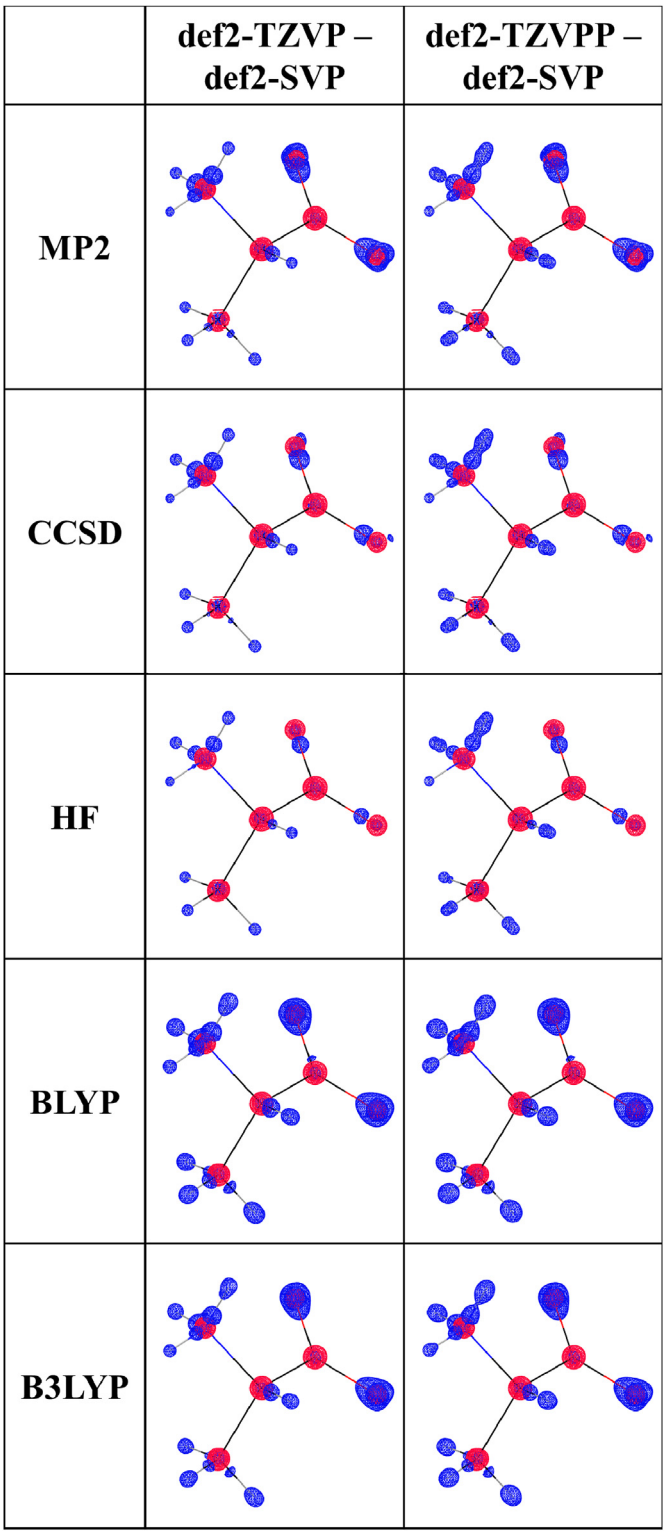

Fig. 7. Comparison of the electron densities obtained with the different basis-sets (references: def2-SVP electron distributions) for each quantum chemical method taken into account (indicated in the leftmost column). All the electron densities were computed on the IAM structure. Contour level: 0.02 e bohr^{-3}; colors in the difference density plots: blue (positive) and red (negative); the orientation of the molecules corresponds to the one in Fig. 2.

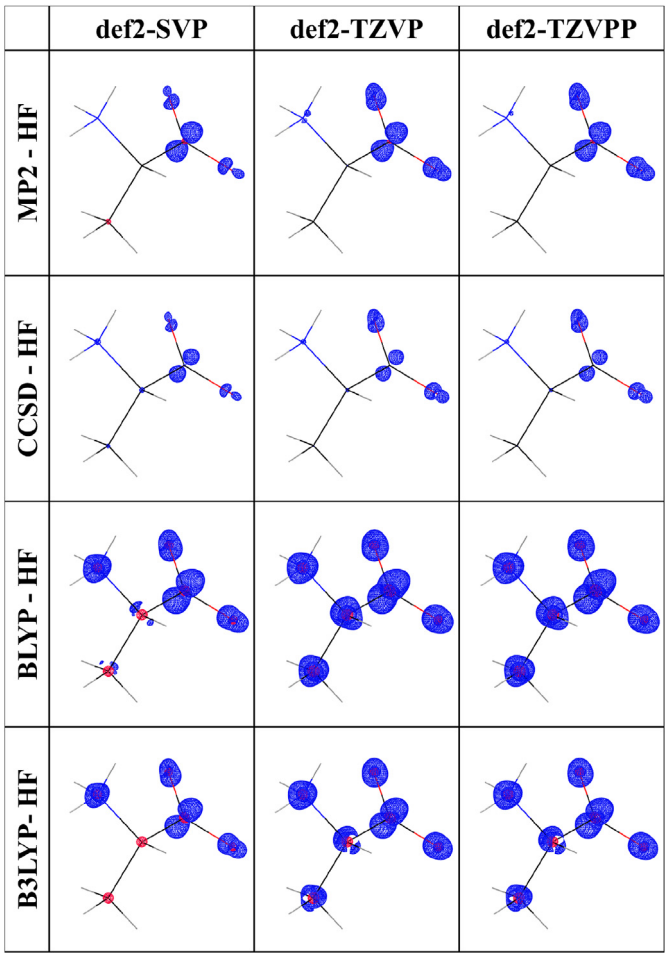

Fig. 8. Comparison of the electron densities obtained with the different quantum mechanical methods (references: Hartree-Fock electron distributions) for each adopted basis-set (indicated in the top row). All the electron densities were computed on the IAM structure. Contour level: 0.02 e bohr^{-3}; colors in the difference density plots: blue (positive) and red (negative); the orientation of the molecules corresponds to the one in Fig. 2.

similarities, but with the DFT methods being the most different to the HF electron densities. The HAR bond lengths involving hydrogen atoms are much less precise than the corresponding neutron-measured values. For hydrogen ADPs, we observed that the HAR results are systematically larger than the neutron ones. However, concerning the effects of the different quantum chemical calculations on bond lengths and ADPs, it was difficult to establish general trends since they change depending on the chemical environment of the hydrogen atoms.

Therefore, from the analysis of our results, we conclude that some differences can be observed when different quantum chemical methods or different basis-sets are used in the calculations underlying the HARs. Nevertheless, we have also noticed that these differences are actually too small to draw clear conclusions. In particular, despite trends could be observed in some cases for the structural parameters (bond lengths and ADPs), the differences were smaller than the corresponding standard deviations. Obviously, this does not allow us to conclude that HARs based on post-HF methods provide more accurate structural results than the more traditional HAR-HF. Given these results and also in light of their large computational cost, post-HF computations may not be necessary for HAR determinations of bond lengths involving hydrogen atoms or hydrogen ADPs. However, before drawing definitive conclusion on the necessity of using post-HF in the framework of HAR, follow-up studies may be required, for example on compounds that are influenced significantly by intramolecular electron correlation effects. In particular, it will be interesting to study the influence of post-HF methods in HARs of molecular and periodic network structures containing heavy elements, which are known to be more susceptible to electron correlation effects. In this regard, we have just recently shown that there are still significant unexplained residual electron-density effects close to heavy elements even if anharmonicity and relativistic effects are accounted for in HARs and HAR-ELMOs [80,93]. Furthermore, it could be also interesting to reconsider the performances of post-HF methods in HARs when the crystal field effects are taken into account by the underlying theoretical model, for instance through cluster or QM/ELMO computations.

Finally, we observed that the agreement between the calculated and observed X-ray structure factors parallels what is generally accepted for the wavefunction types, namely that, for the largest basis set, the agreement gets better in the order HF, BLYP, MP2 ~ B3LYP, CCSD, which indicates that such structure factors may be used to test wavefunction quality. Therefore, the present work represents a starting point that, in the future, could be continued and could lead to the construction of databases of high-quality diffraction datasets. These databanks would enable not only to evaluate the performances of new refinement strategies but also to benchmark wavefunction methods and DFT functionals by comparison of experimental and calculated structure factors. In fact, although there are currently only very few measurements which have the capability of testing wavefunctions and functionals through reciprocal space, our results indicate that such data may be valuable for purposes beyond structural refinements.

CRediT authorship contribution statement

Erna K. Wieduwilt: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing - original draft, Visualization. **Giovanni Macetti:** Methodology, Validation, Writing - review & editing. **Lorraine A. Malaspina:** Software, Investigation, Writing - review & editing. **Dylan Jayatilaka:** Software, Writing - review & editing. **Simon Grabowsky:** Methodology, Writing - review & editing, Funding acquisition. **Alessandro Genoni:** Conceptualization, Methodology, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

A.G., E.K.W. and G.M. acknowledge the French Research Agency (ANR) for financial support of the Young Researcher Project *QuMacroRef* through Grant No. ANR-17-CE29-0005-01. S.G. thanks the German Research Foundation (DFG) for funding within the projects GR 4451/1-1 and GR 4451/2-1.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.molstruc.2020.127934.

References

[1] K. Roy, Advances in QSAR Modeling, Applications in Pharmaceutical, Chemical, Food, Agricultural and Environmental Sciences, Springer, Dordrecht, Netherlands, 2017.
[2] J. Dunitz, X-ray Analysis and the Structure of Organic Molecules, *second ed.*, Verlag Helvetica Chimica Acta, Zürich, 1995.
[3] C. Giacovazzo, Fundamentals of Crystallography, *second ed.*, Oxford University Press, Oxford, UK, 2002.

[4] J.D. Watson, F.H.C. Crick, Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid, Nature 171 (1953) 737–738.
[5] A.H. Compton, The distribution of electrons in atoms, Nature 95 (1915) 343–344.
[6] B. Halliwell, J. Gutteridge, Free Radicals in Biology and Medicine, *third ed.*, Oxford University Press, Oxford, UK, 1999.
[7] J.T. Hynes, J. Klinman, H.-H. Limbach, R. Schowen, Hydrogen Transfer Reactions, vols. 1–4, Wiley-VCH, Weinheim, 2007.
[8] A.A. Hoser, A.Ø. Madsen, Dynamic quantum crystallography: lattice dynamical models refined against diffraction data. II. Applications to L-alanine, naphthalene and xylitol, Acta Crystallogr. A 73 (2017) 102–114.
[9] P. Jena, Materials for hydrogen storage: past, present, and future, J. Phys. Chem. Lett. 2 (2011) 206–211.
[10] G. Desiraju, Crystal Engineering – the Design of Organic Solids, Elsevier, Amsterdam, 1989.
[11] G.R. Desiraju, T. Steiner, The Weak Hydrogen Bond in Structural Chemistry and Biology, Oxford University Press, Oxford, UK, 1999.
[12] R.F. Stewart, E.R. Davidson, W.T. Simpson, Coherent X-ray scattering for the hydrogen atom in the hydrogen molecule, J. Chem. Phys. 42 (1965) 3175–3187.
[13] M.P. Blakeley, S.S. Hasnain, S.V. Antonyuk, Sub-atomic resolution X-ray crystallography and neutron crystallography: promise, challenges and potential, IUCrJ 2 (2015) 464–474.
[14] F.H. Allen, D.G. Watson, L. Brammer, A.G. Orpen, R. Taylor, International Tables for Crystallography, 1st online Ed., C, International Union of Crystallography, Chester, UK, 2006, pp. 790–811. Chapter 9.5.
[15] F.H. Allen, I.J. Bruno, Bond lengths in organic and metal-organic compounds revisited: X-H bond lengths from neutron diffraction data, Acta Crystallogr. B 66 (2010) 380–386.
[16] A.Ø. Madsen, SHADE web server for estimation of hydrogen anisotropic displacement parameters, J. Appl. Crystallogr. 39 (2006) 757–758.
[17] A.Ø. Madsen, A.A. Hoser, SHADE3 server: a streamlined approach to estimate H-atom anisotropic displacement parameters using periodic ab initio calculations or experimental information, J. Appl. Crystallogr. 47 (2014) 2100–2104.
[18] A.A. Hoser, A. Madsen, Dynamic quantum crystallography: lattice-dynamical models refined against diffraction data. I. Theory, Acta Crystallogr. A 72 (2016) 206–214.
[19] S. Grabowsky, A. Genoni, H.-B. Bürgi, Quantum crystallography, Chem. Sci. 8 (2017) 4159–4176.
[20] A. Genoni, L. Bučinský, N. Claiser, J. Contreras-García, B. Dittrich, P.M. Dominiak, E. Espinosa, C. Gatti, P. Giannozzi, J.-M. Gillet, D. Jayatilaka, P. Macchi, A.Ø. Madsen, L.J. Massa, C.F. Matta, K.M. Merz Jr., P.N.H. Nakashima, H. Ott, U. Ryde, K. Schwarz, M. Sierka, S. Grabowsky, Quantum crystallography: current developments and future perspectives, Chem. Eur J. 24 (2018) 10881–10905.
[21] R.F. Novara, A. Genoni, S. Grabowsky, What is quantum crystallography? ChemViews (2018) https://doi.org/10.1002/chemv.201800066.
[22] L. Massa, C.F. Matta, Quantum crystallography: a perspective, J. Comput. Chem. 39 (2017) 1021–1028.
[23] V. Tsirelson, Early days of quantum crystallography: a personal account, J. Comput. Chem. 39 (2017) 1029–1037.
[24] V. Pichon-Pesme, C. Lecomte, H. Lachekar, On building a data bank of transferable experimental electron density parameters: application to polypeptides, J. Phys. Chem. 99 (1995) 6242–6250.
[25] B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte, C. Jelsch, On the application of an experimental multipolar pseudo-atom library for accurate refinement of small-molecule and protein crystal structures, Acta Crystallogr. A 63 (2007) 108–125.
[26] S. Domagała, B. Fournier, D. Liebschner, B. Guillot, C. Jelsch, An improved experimental databank of transferable multipolar atom models – ELMAM2. Construction details and applications, Acta Crystallogr. A 68 (2012) 337–351.
[27] T. Koritsanszky, A. Volkov, P. Coppens, Aspherical-atom scattering factors from molecular wave functions. 1. Transferability and conformation dependence of atomic electron densities of peptides within the multipole formalism, Acta Crystallogr. A 58 (2002) 464–472.
[28] P.M. Dominiak, A. Volkov, X. Li, M. Messerschmidt, P. Coppens, A theoretical databank of transferable aspherical atoms and its application to electrostatic interaction energy calculations of macromolecules, J. Chem. Theor. Comput. 3 (2007) 232–247.
[29] A. Volkov, M. Messerschmidt, P. Coppens, Improving the scattering-factor formalism in protein refinement: application of the university at buffalo aspherical-atom databank to polypeptide structures, Acta Crystallogr. D 63 (2007) 160–170.
[30] B. Dittrich, T. Koritsanszky, P. Luger, A simple approach to nonspherical electron densities by using invarioms, Angew. Chem. Int. Ed. 43 (2004) 2718–2721.
[31] B. Dittrich, C.B. Hübschle, P. Luger, M.A. Spackman, Introduction and validation of an invariom database for amino-acid, peptide and protein molecules, Acta Crystallogr. D 62 (2006) 1325–1335.
[32] B. Dittrich, C.B. Hübschle, K. Pröpper, F. Dietrich, T. Stolper, J.J. Holstein, The generalized invariom database (GID), Acta Crystallogr. B 69 (2013) 91–104.
[33] B. Dittrich, C.B. Hübschle, M. Messerschmidt, R. Kalinowski, D. Girnt, P. Luger, The invariom model and its application: refinement of D,L-serine at different temperatures and resolution, Acta Crystallogr. A 61 (2005) 314–320.
[34] B. Dittrich, M. Weber, R. Kalinowski, S. Grabowsky, C.B. Hübschle, P. Luger, How to easily replace the independent atom model - the example of bergenin, a potential anti-HIV agent of traditional Asian medicine, Acta Crystallogr. B 65 (2009) 749–756.
[35] E. Bendeif, C. Jelsch, The experimental library multipolar atom model refinement of L-aspartic acid, Acta Crystallogr. C 63 (2007) o361–o364.
[36] J.M. Bąk, S. Domagała, C. Hübschle, C. Jelsch, B. Dittrich, P.M. Dominiak, Verification of structural and electrostatic properties obtained by the use of different pseudoatom databases, Acta Crystallogr. A 67 (2011) 141–153.
[37] N. Dadda, A. Nassour, B. Guillot, N. Benali-Cherif, C. Jelsch, Charge-density analysis and electrostatic properties of 2-carboxy-4-methylanilinium chloride monohydrate obtained using a multipolar and a spherical-charges model, Acta Crystallogr. A 68 (2012) 452–463.
[38] R. Destro, F. Merati, Bond lengths, and beyond, Acta Crystallogr. B 51 (1995) 559–570.
[39] R.F. Stewart, J. Bentley, Generalized x-ray scattering factors in diatomic molecules, J. Chem. Phys. 63 (1975) 3786–3793.
[40] V.V. Zhurov, E.A. Zhurova, A.I. Stash, A.A. Pinkerton, Importance of consideration of anharmonic motion in charge-density: a comparison of variable-temperature studies on two explosives, RDX and HMX, Acta Crystallogr. A 67 (2011) 160–173.
[41] A.A. Hoser, P.M. Dominiak, K. Woźniak, Towards the best model for H atoms in experimental charge-density refinement, Acta Crystallogr. A 65 (2009) 300–311.
[42] D. Jayatilaka, B. Dittrich, X-ray structure refinement using aspherical atomic density functions obtained from quantum mechanical calculations, Acta Crystallogr. A 64 (2008) 383–393.
[43] S.C. Capelli, H.-B. Bürgi, B. Dittrich, S. Grabowsky, D. Jayatilaka, Hirshfeld atom refinement, IUCrJ 1 (2014) 361–379.
[44] M. Woińska, D. Jayatilaka, M.A. Spackman, A.J. Edwards, P.M. Dominiak, K. Woźniak, E. Nishibori, K. Sugimoto, S. Grabowsky, Hirshfeld atom refinement for modeling strong hydrogen bonds, Acta Crystallogr. A 70 (2014) 483–498.
[45] M. Woińska, S. Grabowsky, P.M. Dominiak, K. Woźniak, D. Jayatilaka, Hydrogen atoms can be located accurately and precisely by x-ray crystallography, Sci. Adv. 2 (2016), e1600192.
[46] M.E. Wall, Quantum crystallographic charge density of urea, IUCrJ 3 (2016) 237–246.
[47] M. Fugel, D. Jayatilaka, E. Hupf, J. Overgaard, V.R. Hathwart, P. Macchi, M.J. Turner, J.A.K. Howard, O.V. Dolomanov, H. Puschmann, B.B. Iversen, H.-B. Bürgi, S. Grabowsky, Probing the accuracy and precision of Hirshfeld atom refinement with HARt interface with Olex2, IUCrJ 5 (2018) 32–44.
[48] C. Köhler, J. Lübben, L. Krause, C. Hoffmann, R. Herbst-Irmer, D. Stalke, Comparison of different strategies for modelling hydrogen atoms in charge density analyses, Acta Cystallogr. B 75 (2019) 434–441.
[49] D. Jayatilaka, D.J. Grimwood, Tonto: a fortran based object-oriented system for quantum chemistry and crystallography, in: P.M.A. Sloot, D. Abramson, A.V. Bogdanov, J.J. Dongarra, A.Y. Zomaya, Y.E. Gorbachev (Eds.), *Computational Science - ICCS* 2003, Springer-Verlag: Berlin & Heidelberg, 2003, pp. 142–151.
[50] D. Jayatilaka, Wave function for beryllium from X-ray diffraction data, Phys. Rev. Lett. 80 (1998) 798–801.
[51] D. Jayatilaka, D.J. Grimwood, Wavefunctions derived from experiment. I. Motivation and theory, Acta Crystallogr. A 57 (2001) 76–86.
[52] D.J. Grimwood, D. Jayatilaka D, Wavefunctions derived from experiment. II. A wavefunction for oxalic acid dihydrate, Acta Crystallogr. A 57 (2001) 87–100.
[53] I. Bytheway, D. Grimwood, D. Jayatilaka, Wavefunctions derived from experiment. III. Topological analysis of crystal fragments, Acta Crystallogr. A 58 (2002) 232–243.
[54] I. Bytheway, D.J. Grimwood, B.N. Figgis, G.S. Chandler, D. Jayatiaka, Wavefunctions derived from experiment. IV. Investigation of the crystal environment of ammonia, Acta Crystallogr. A 58 (2002) 244–251.
[55] D.J. Grimwood, I. Bytheway, D. Jayatilaka, Wavefunctions derived from experiment. V. Investigation of electron densities, electrostatic potentials, and electron localization functions for noncentrosymmetric crystals, J. Comput. Chem. 24 (2003) 470–483.
[56] D. Jayatilaka, Using wave functions to get more information out of diffraction experiments, in: C. Gatti, P. Macchi (Eds.), Modern Charge-Density Analysis, Springer, Dordrecht, Netherlands, 2012, pp. 213–257.
[57] M. Hudák, D. Jayatilaka, L. Peraínova, S. Biskupic, J. Kozísek, L. Bučinský, X-ray constrained unrestricted Hartree–Fock and Douglas–Kroll–Hess wavefunctions, Acta Crystallogr. A 66 (2010) 78–92.
[58] S. Grabowsky, P. Luger, J. Buschmann, T. Schneider, T. Schirmeister, A.N. Sobolev, D. Jayatilaka, The significance of ionic bonding in sulfur dioxide: bond orders from X-ray diffraction data, Angew. Chem. Int. Ed. 51 (2012) 6776–6779.
[59] M. Woińska, D. Jayatilaka, B. Dittrich, R. Flaig, P. Luger, K. Wozńiak, P.M. Dominiak, S. Grabowsky, Validation of X-ray wavefunction refinement, ChemPhysChem 18 (2017) 3334–3351.
[60] A. Genoni, Molecular orbitals strictly localized on small molecular fragments from X-ray diffraction data, J. Phys. Chem. Lett. 4 (2013) 1093–1099.
[61] A. Genoni, X-ray constrained extremely localized molecular orbitals: theory and critical assessment of the new technique, J. Chem. Theor. Comput. 9 (2013) 3004–3019.
[62] L.H.R. Dos Santos, A. Genoni, P. Macchi, Unconstrained and X-ray constrained

extremely localized molecular orbitals: analysis of the reconstructed electron density, Acta Crystallogr. A 70 (2014) 532–551.
[63] A. Genoni, B. Meyer, X-ray constrained wave functions: fundamentals and effects of the molecular orbitals localization, Adv. Quant. Chem. 73 (2016) 333–362.
[64] A. Genoni, A first-prototype multi-determinant X-ray constrained wavefunction approach: the X-ray constrained extremely localized molecular orbital-valence bond method, Acta Crystallogr. A 73 (2017) 312–316.
[65] N. Casati, A. Genoni, B. Meyer, A. Krawczuk, P. Macchi, Exploring charge density analysis in crystals at high pressure: data collection, data analysis and advanced modelling, Acta Crystallogr. B 73 (2017) 584–597.
[66] A. Genoni, D. Franchini, S. Pieraccini, M. Sironi, X-ray constrained Spin-Coupled wavefunction: a new tool to extract chemical information from X-ray diffraction data, Chem. Eur J. 24 (2018) 15507–15511.
[67] A. Genoni, G. Macetti, D. Franchini, S. Pieraccini, M. Sironi, X-ray constrained spin-coupled technique: theoretical details and further assessment of the method, Acta Crystallogr. A 75 (2019) 778–797.
[68] L. Bučinský, D. Jayatilaka, S. Grabowsky, Importance of relativistic effects and electron correlation in structure factors and electron density of diphenyl mercury and triphenyl bismuth, J. Phys. Chem. A 120 (2016) 6650–6669.
[69] L. Bučinský, D. Jayatilaka, S. Grabowsky, Relativistic quantum crystallography of diphenyl- and dicyanomercury. Theoretical structure factors and Hirshfeld atom refinement, Acta Crystallogr. A 75 (2019) 705–717.
[70] H. Stoll, G. Wagenblast, H. Preuss, On the use of local basis sets for localized molecular orbitals, Theor. Chim. Acta 57 (1980) 169–178.
[71] A. Fornili, M. Sironi, M. Raimondi, Determination of extremely localized molecular orbitals and their application to quantum mechanics/molecular mechanics methods and to the study of intramolecular hydrogen bonding, J. Mol. Struct. (THEOCHEM) 632 (2003) 157–172.
[72] A. Genoni, M. Sironi, A novel approach to relax extremely localized molecular orbitals: the extremely localized molecular orbital-valence bond method, Theor. Chem. Acc. 112 (2004) 254–262.
[73] A. Genoni, A. Fornili, M. Sironi, Optimal virtual Orbitals to relax wave functions built up with transferred extremely localized molecular orbitals, J. Comput. Chem. 26 (2005) 827–835.
[74] A. Genoni, M. Ghitti, S. Pieraccini, M. Sironi, A novel extremely localized molecular orbitals based technique for the one-electron density matrix computation, Chem. Phys. Lett. 415 (2005) 256–260.
[75] M. Sironi, A. Genoni, M. Civera, S. Pieraccini, M. Ghitti, Extremely localized molecular orbitals: theory and applications, Theor. Chem. Acc. 117 (2007) 685–698.
[76] M. Sironi, M. Ghitti, A. Genoni, G. Saladino, S. Pieraccini, DENPOL: a new program to determine electron densities of polypeptides using extremely localized molecular orbitals, J. Mol. Struct. (THEOCHEM) 898 (2009) 8–16.
[77] B. Meyer, B. Guillot, M.F. Ruiz-Lopez, A. Genoni, Libraries of extremely localized molecular orbitals. 1. Model molecules approximation and molecular orbitals transferability, J. Chem. Theor. Comput. 12 (2016) 1052–1067.
[78] B. Meyer, B. Guillot, M.F. Ruiz-Lopez, C. Jelsch, A. Genoni, Libraries of extremely localized molecular orbitals. 2. Comparison with the pseudoatoms transferability, J. Chem. Theor. Comput. 12 (2016) 1068–1081.
[79] B. Meyer, A. Genoni, Libraries of extremely localized molecular orbitals. 3. Construction and preliminary assessment of the new databanks, J. Phys. Chem. A 122 (2018) 8965–8981.
[80] L.A. Malaspina, E.K. Wieduwilt, J. Bergmann, F. Kleemiss, B. Meyer, M.F. Ruiz-López, R. Pal, E. Hupf, J. Beckmann, R.O. Piltz, A.J. Edwards, S. Grabowsky, A. Genoni, Fast and accurate quantum crystallography: from small to large, from light to heavy, J. Phys. Chem. Lett. 10 (2019) 6973–6982.
[81] M. Kohout, A measure of electron localizability, Int. J. Quant. Chem. 97 (2004) 651–658.
[82] D. Arias-Olivares, E.K. Wieduwilt, J. Contreras-García, A. Genoni, NCI-ELMO: a new method to quickly and accurately detect noncovalent interactions in biosystems, J. Chem. Theor. Comput. 15 (2019) 6456–6470.
[83] G. Macetti, A. Genoni, Quantum mechanics/extremely localized molecular orbital method: a fully quantum mechanical embedding approach for macromolecules, J. Phys. Chem. A 123 (2019) 9420–9428.
[84] F.L. Hirshfeld, Xvii. Spatial partitioning of charge density, Isr. J. Chem. 16 (1977) 198–201.
[85] F.L. Hirshfeld, Bonded-atom fragments for describing molecular charge densities, Theor. Chim. Acta 44 (1977) 129–138.
[86] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford, CT, USA, 2009.
[87] L. A. Malaspina, A. Genoni, D. Jayatilaka, M. J. Turner, K. Sugimoto, E. Nishibori, S. Grabowsky, lamaGOET: an interface for quantum crystallography, J. Appl. Cryst., in preparation.
[88] R. Destro, R.E. Marsh, R. Bianchi, A low-temperature (23 K) study of L-alanine, J. Phys. Chem. 92 (1988) 966–973.
[89] R. Destro, R. Soave, M. Barzaghi, Physicochemical properties of zwitterionic L- and DL-alanine crystals from their experimental and theoretical charge densities, J. Phys. Chem. B 112 (2008) 5163–5174.
[90] K. Meindl, J. Henn, Foundations of residual-density analysis, Acta Crystallogr. A 64 (2008) 404–418.
[91] P. Roversi, R. Destro, Approximate anisotropic displacement parameters for H atoms in molecular crystals, Chem. Phys. Lett. 386 (2004) 472–478.
[92] M.G. Medvedev, I.S. Bushmarinov, J. Sun, J.P. Perdew, K.A. Lyssenko, Density functional theory is straying from the path toward the exact functional, Science 355 (2017) 49–52.
[93] S. Grabowsky, A. Genoni, S.P. Thomas, D. Jayatilaka, The advent of quantum crystallography: form and structure factors from quantum mechanics for advanced structure refinement and wavefunction fitting, in: D.M.P. Mingos, P.R. Raithby (Eds.), *21st Century Challenges in Chemical Crystallography*. Structure and Bonding, Springer, 2020. Submitted for publication.