



# Towards solution and refinement of organic crystal structures by fitting to the atomic pair distribution function

Dragica Prill,<sup>a</sup> Pavol Juhás,<sup>b</sup> Simon J. L. Billinge<sup>b,c\*</sup> and Martin U. Schmidt<sup>a\*</sup>

Received 5 June 2015

Accepted 24 November 2015

Edited by A. Altomare, Institute of Crystallography – CNR, Bari, Italy

This paper is dedicated to Professor Dr Ernst Egert on the occasion of his 65th birthday.

**Keywords:** pair distribution function; organic crystal structures; structure refinement; structure solution.

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

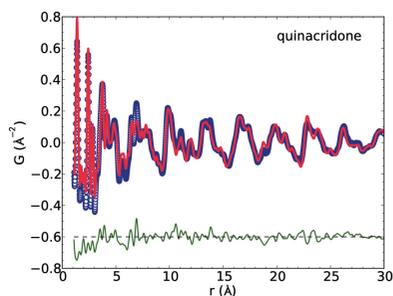
<sup>a</sup>Institute of Inorganic and Analytical Chemistry, Goethe University, Max-von-Laue-Strasse 7, 60438 Frankfurt am Main, Germany, <sup>b</sup>Condensed Matter Physics and Materials Science Department, Brookhaven National Laboratory, Upton, New York 11973, USA, and <sup>c</sup>Department of Applied Physics and Applied Mathematics, Columbia University, New York 10027, USA. \*Correspondence e-mail: sb2896@columbia.edu, m.schmidt@chemie.uni-frankfurt.de

A method towards the solution and refinement of organic crystal structures by fitting to the atomic pair distribution function (PDF) is developed. Approximate lattice parameters and molecular geometry must be given as input. The molecule is generally treated as a rigid body. The positions and orientations of the molecules inside the unit cell are optimized starting from random values. The PDF is obtained from carefully measured X-ray powder diffraction data. The method resembles ‘real-space’ methods for structure solution from powder data, but works with PDF data instead of the diffraction pattern itself. As such it may be used in situations where the organic compounds are not long-range-ordered, are poorly crystalline, or nanocrystalline. The procedure was applied to solve and refine the crystal structures of quinacridone ( $\beta$  phase), naphthalene and allopurinol. In the case of allopurinol it was even possible to successfully solve and refine the structure in  $P1$  with four independent molecules. As an example of a flexible molecule, the crystal structure of paracetamol was refined using restraints for bond lengths, bond angles and selected torsion angles. In all cases, the resulting structures are in excellent agreement with structures from single-crystal data.

## 1. Introduction

### 1.1. The pair distribution function (PDF)

Understanding the arrangement of atoms and molecules within a material is fundamental for numerous applications. The crystal structure of well crystallized compounds can be determined by single-crystal structure X-ray analysis or by structure determination from powder data. These methods fail for nanocrystalline or amorphous materials. However, X-ray powder diffractograms of nanocrystalline and amorphous materials still contain valuable information when measured over a sufficiently wide range of reciprocal space. This information can be extracted from the powder pattern through the analysis of the atomic pair distribution function. The pair distribution function (PDF) was originally developed for the investigation of inorganic materials containing no long-range order such as liquids and glasses (Debye & Menke, 1930; Warren, 1990; Egami & Billinge, 2012). The PDF, also known as the ‘radial distribution function’ (RDF), gives the probability of finding pairs of atoms separated by a distance  $r$ . It is weighted by the scattering power of the two atoms and summed over all atom pairs. The PDF is related to the widely known Patterson function. The Patterson function is derived from single-crystal diffraction data by a Fourier transform of the reflection intensities (Patterson, 1934, 1935). The



Patterson map shows the lengths and the directions of the crystallographically averaged interatomic vectors. The PDF is the Fourier transform of the scattered intensity from an isotropic sample such as a powder or amorphous material and gives a distribution of interatomic distances like the Patterson function, except that directional information is lost and, because both diffuse and Bragg intensities are sampled in the PDF transform, the PDF is sensitive to non-periodic components of the structure. In this sense, the PDF can be thought of as an orientationally averaged generalized Patterson function. Different variants of the PDF or RDF are in use. Here we use the function  $G(r)$  which is experimentally obtained by a sine Fourier transform of the corrected and normalized total scattering function  $S(Q)$ ,

$$G(r) = \frac{2}{\pi} \int_{Q_{\min}}^{Q_{\max}} Q[S(Q) - 1] \sin(Qr) \, dQ \quad (1)$$

$$= 4\pi r[\rho(r) - \rho_0\gamma_0(r)], \quad (2)$$

with the magnitude of the scattering vector  $Q = 4\pi \sin \theta/\lambda$ . Here,  $\rho(r)$  is the microscopic pair density,  $\rho_0$  the average number density and  $\gamma_0$  is the autocorrelation function of the particle shape (Farrow & Billinge, 2009).

The  $S(Q)$  function is gained by removing the self-scattering from the coherent scattered intensity per atom,  $I(Q)$ , and dividing by the average scattering power per atom,  $f(Q)$ , according to (Egami & Billinge, 2012)

$$S(Q) = \frac{I(Q) - \langle f(Q)^2 \rangle}{\langle f(Q)^2 \rangle} + 1. \quad (3)$$

The resulting pair distribution function  $G(r)$  is defined as a deviation from an average electron density. Hence a value of  $G(r) = 0$  does not correspond to a probability of zero, but to the average probability of finding two atoms with a distance  $r$ . Values of  $G(r)$  above and below zero denote distances where there is a higher, or lower, probability of finding an atom pair than implied by the average number density.

The PDF is derived from the total scattering powder pattern, which includes not only the Bragg peaks, but also the diffuse scattering. To distinguish which part of the ‘background’ in the powder pattern originates from the diffuse scattering, and which part is caused by the instrument or by the capillary, an empty capillary is measured under identical conditions, and the background is subtracted.

## 1.2. PDF analysis of organic compounds

Most PDF studies to date deal with inorganic materials such as liquids, glasses, amorphous, nanocrystalline or disordered compounds (Billinge & Kanatzidis, 2004; Billinge, 2008; Brühne *et al.*, 2008; Young & Goodwin, 2011). Recently, the PDF method has been gaining traction in the study of organic materials, especially of nanocrystalline and amorphous organic compounds (Sheth *et al.*, 2005; Newman *et al.*, 2008; Nollenberger *et al.*, 2009; Schmidt *et al.*, 2009; Billinge *et al.*, 2010; Dykhne *et al.*, 2011; Davis *et al.*, 2013; Terban *et al.*, 2015;

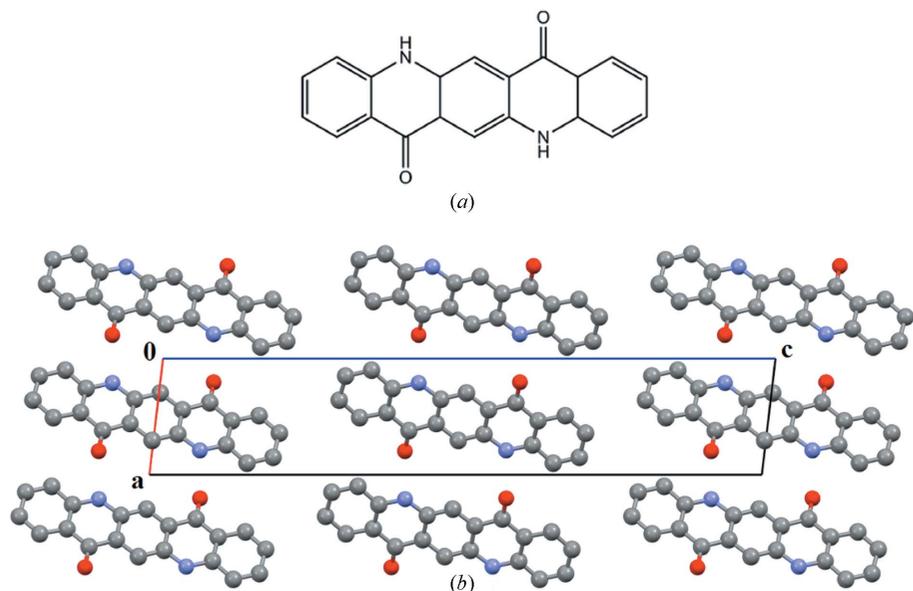
Juhás *et al.*, 2006; Cliffe *et al.*, 2010; Temleitner, 2014). However, the PDF analysis of organic compounds is at present mainly restricted to the use of the PDF curve as a fingerprint in the comparison of the local structures of crystalline, nanocrystalline and amorphous samples. This allows one to identify packing patterns or polymorphs and to determine the ordering length (Schmidt *et al.*, 2009). Lattice parameters can also be refined (Wolf *et al.*, 2012).

It is straightforward to calculate the PDF from structural models, but hitherto the fit of a crystal structure (or of a model of the local structure) to the PDF data has not been practically feasible for organic compounds. The major difficulty is to maintain the geometry of the molecule during the refinement. Most of the presently available PDF software programs refine structures by moving individual atoms, which is useful for inorganic compounds but inappropriate for organic compounds which consist of defined molecules. For a refinement of organic structures one needs to introduce constraints or restraints on bond lengths, bond angles and torsion angles in order to maintain rigid groups of atoms, often called rigid-body constraints in crystallographic programs.

Another issue in molecular refinement is to distinguish between intramolecular and intermolecular atom–atom distances, which is at present not implemented in most PDF software. In the experimental PDF the peaks have a certain peak width that originates not only from the finite  $2\theta$  range of the experiment, but also from the interatomic distance distribution between a given pair of atoms, coming from vibrations and disorder in the material. Since intramolecular forces are much stronger than intermolecular ones, the intramolecular vibrational displacements are generally much smaller than the intermolecular ones. Correspondingly the PDF peaks arising from intramolecular atom–atom distances are much sharper than those from intermolecular ones (Rademacher *et al.*, 2012; Prill *et al.*, 2015). In an earlier paper we introduced an easy and successful approach to account for this effect using different atomic displacement parameters (ADPs) for intra- and intermolecular atom–atom distances (Prill *et al.*, 2015).

## 1.3. Towards solution and refinement of organic crystal structures by fitting to the PDF

In this work we have developed a new modelling approach, using the new *DiffPy-CMI* software (Juhás *et al.*, 2015) to make a first step towards a full refinement of organic crystal structures to PDF data. As a first approach, the molecules are treated as completely rigid bodies (ignoring internal degrees of freedom). Lattice parameters, positions and spatial orientations of the molecules are varied in order to obtain a good agreement between the calculated and measured PDF data. Two cases are investigated: (i) a refinement of a given structure is carried out, by starting from given values for lattice parameters, molecular position and orientation. (ii) It will be shown that this procedure allows one not only to refine, but even to solve crystal structures by using random starting



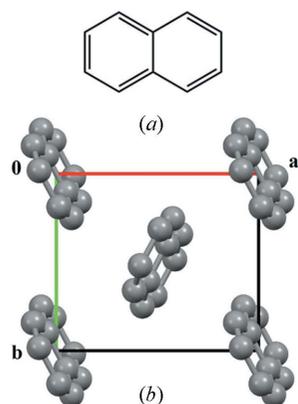
**Figure 1**  
 (a) Structural formula of quinacridone. (b) Crystal structure of  $\beta$ -quinacridone viewed along the [010] direction. Hydrogen atoms are omitted for clarity.

values for the molecular positions and orientations, together with approximate values for the lattice parameters.

Further developments of the approach are performed to include molecular flexibility. Therefore all atoms are refined individually and the molecular geometry is held together by using restraints (soft constraints) for all bond lengths, all bond angles and planar groups or other selected torsion angles. These atomic positions are refined together with the lattice parameters.

#### 1.4. Application examples

Quinacridone, naphthalene and allopurinol were used as examples of rigid molecules. Paracetamol was chosen as an example of a flexible molecule. All crystal structures are known from single-crystal X-ray diffraction analyses.

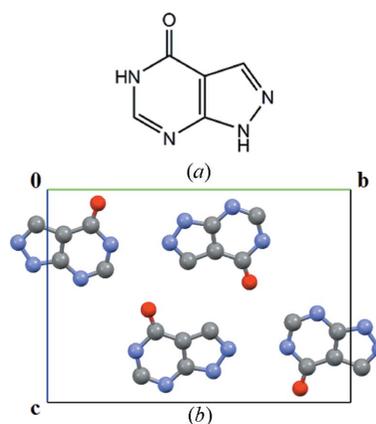


**Figure 2**  
 (a) Structural formula of naphthalene. (b) Crystal structure of naphthalene viewed along the [001] direction. Hydrogen atoms are omitted for clarity.

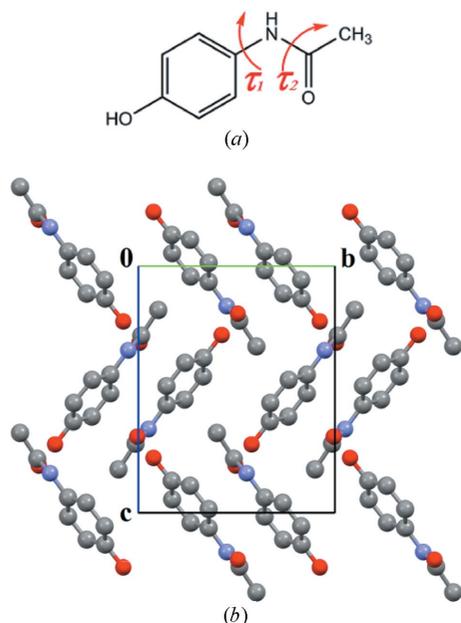
*Quinacridone* ( $C_{20}H_{12}N_2O_2$ , Pigment Violet 19, Fig. 1) is an important industrial pigment for red to violet shades. Four phases are known to date ( $\alpha^I$ ,  $\alpha^{II}$ ,  $\beta$  and  $\gamma$ ) (Paulus *et al.*, 2007). The reddish-violet  $\beta$  phase is used for the colouration of plastics and coatings (Herbst & Hunger, 2004). In this phase each molecule is bound to two neighbouring molecules by two  $N-H \cdots O=C$  hydrogen bonds each, which results in a crystal structure of infinite chains of linked molecules. The chains run in [110] and  $[\bar{1}\bar{1}0]$  directions. In the  $b$  direction the chains are stacked on top of each other, resulting in two-dimensional stacks parallel to (001). In the [001] direction neighbouring stacks are connected by weak van der Waals interactions between CH groups only (Paulus *et al.*, 2007). The space group is  $P2_1/c$ , with molecules located on inversion centres.

*Naphthalene* ( $C_{10}H_8$ , Fig. 2) is the simplest polycyclic aromatic hydrocarbon. Its main use is as a precursor in the synthesis of other chemicals such as phthalic acid anhydride, 2-naphthol and tetralin. Naphthalene was – together with anthracene – the first organic compound to have its structure determined by X-ray structure analysis (Bragg, 1921). Bragg erroneously assumed that the two rings possess a chair conformation, because this conformation of the  $C_6$  rings was found not only in diamond (Bragg & Bragg, 1913), but had also been found in graphite by Hull (1917), though Debye & Scherrer had determined the correct structure of graphite with planar rings in 1916 (Debye & Scherrer, 1916).

The wrong molecular structure of naphthalene was later corrected (Banerjee, 1929). However, the crystal symmetry ( $P2_1/a$ ,  $Z = 2$ ) and the molecular packing were correct from the beginning. The molecules are situated on inversion centres and form a herringbone arrangement, which is today known to



**Figure 3**  
 (a) Structural formula of allopurinol. (b) Crystal structure of allopurinol viewed along the [100] direction. Hydrogen atoms are omitted for clarity.



**Figure 4**  
(a) Structural formula of paracetamol. (b) Crystal structure of paracetamol viewed along the [100] direction. Hydrogen atoms are omitted for clarity.

be present in many other aromatic compounds too. Naphthalene represents a typical organic van der Waals crystal, where the molecules are held together by van der Waals interactions only.

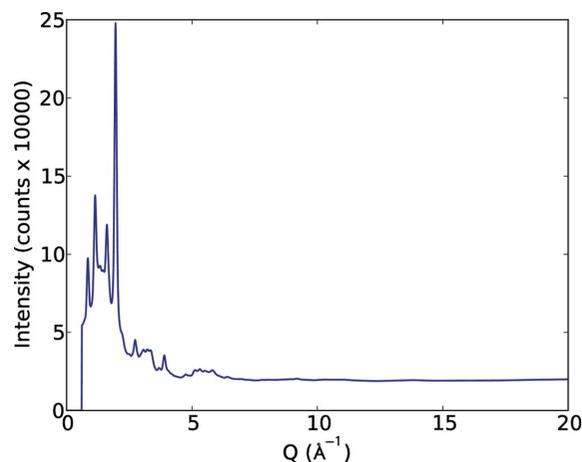
*Allopurinol* ( $C_5H_4N_4O$ , Fig. 3) is a xanthine oxidase inhibitor and is primarily used to treat hyperuricemia. The compound crystallizes in the space group  $P2_1/c$ , with the molecule on the general position. Each allopurinol molecule is bound to six neighbouring molecules by  $N-H \cdots N$  and  $C-H \cdots O$  hydrogen bonds. Because of such bonding, the molecules form layers parallel to the  $(\bar{1}02)$  plane. Neighbouring layers are connected by a high number of close van der Waals contacts supported by Coulomb interactions between the partial charges of the atoms (Prusiner & Sundaralingam, 1972).

*Paracetamol* ( $C_8H_9NO_2$ , Fig. 4) is an active pharmaceutical ingredient which is widely used as a mild analgesic and antipyretic. Its crystal structure was determined by Bouhmaida *et al.* (2009) from single-crystal data. In contrast to the three examples described above, the paracetamol molecule is not planar. The methyl amide group is rotated around the Ph–N bond by about  $22.60^\circ$  out of the benzene plane. The molecules are connected by  $N-H \cdots O=C$  hydrogen bonds between the amide groups to chains along [100]. These chains are connected by  $O-H \cdots O=C$  hydrogen bonds resulting in zigzag sheets parallel to (010) (Bouhmaida *et al.*, 2009).

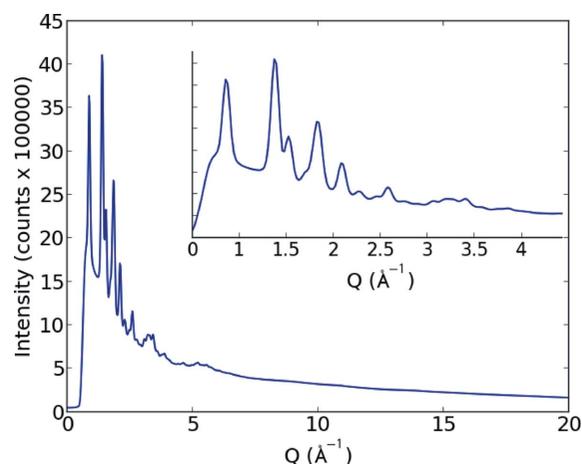
## 2. Experimental

### 2.1. X-ray powder diffraction and data preparation

The quinacridone in its  $\beta$  phase was obtained from Clariant GmbH [trade name Hostaperm(R) Red Violet ER02].



**Figure 5**  
Observed powder pattern of  $\beta$ -quinacridone.



**Figure 6**  
Observed powder pattern of naphthalene. The inset shows the low- $Q$  region.

Naphthalene, allopurinol and paracetamol were purchased from Sigma Aldrich (99% purity). All compounds were used without further purification. X-ray powder diagrams of all four samples were measured at 300 K at the X17A beamline of the National Synchrotron Light Source at Brookhaven National Laboratory, using a two-dimensional Perkin Elmer amorphous silicon detector. Samples were carefully ground in a mortar and then packed in cylindrical polyimide capillaries 1 mm in diameter. The capillaries were sealed with clay at both ends. A monochromatic incident X-ray beam 0.5 mm by 0.5 mm in size was used, conditioned using a Si (311) monochromator to have an energy of 67.42 keV ( $\lambda = 0.1839 \text{ \AA}$ ). The detector was mounted orthogonal to the beam path with a sample-to-detector distance of 204.2 mm, as calibrated with a  $LaB_6$  standard sample. Multiple scans were performed on each sample to achieve a total exposure time of 30 min. The two-dimensional diffraction data were integrated and converted to intensity versus  $2\theta$  using the software *FIT2D* (Hammersley *et al.*, 1996) (Figs. 5, 6, 7 and 8). For the PDF analysis the data were corrected and normalized (Egami & Billinge, 2012) using the program *PDFgetX3* (Juhás *et al.*, 2013) to obtain the total

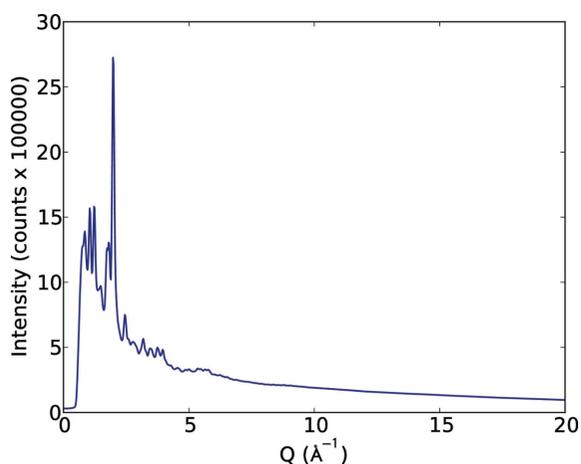


Figure 7  
Observed powder pattern of allopurinol.

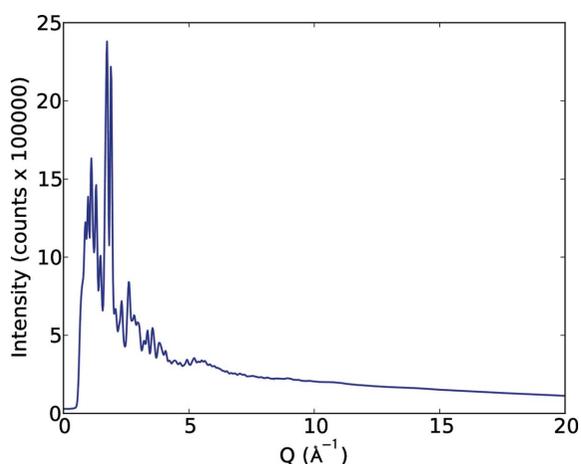


Figure 8  
Observed powder pattern of paracetamol.

scattering structure function,  $F(Q)$ , and  $G(r)$ . The data were truncated at a finite maximum value of the momentum transfer  $Q_{\max}$ , which was optimized to avoid large termination effects whilst maximizing the signal-to-noise ratio. The values  $Q_{\max} = 19.0 \text{ \AA}^{-1}$  for quinacridone,  $Q_{\max} = 19.5 \text{ \AA}^{-1}$  for naphthalene,  $Q_{\max} = 18.5 \text{ \AA}^{-1}$  for allopurinol and  $Q_{\max} = 18.7 \text{ \AA}^{-1}$  for paracetamol were found to be optimal.

## 2.2. Structural models

The structural models used for the calculations were taken from the Cambridge Structural Database (CSD) (Allen, 2002). The corresponding CSD reference codes are QNACRD07 for quinacridone (Paulus *et al.*, 2007), NAPHTA11 for naphthalene (Cruickshank, 1957), ALOPUR for allopurinol (Prusiner & Sundaralingam, 1972) and HXACAN27 for paracetamol (Bouhmaida *et al.*, 2009). The structures were determined using single-crystal X-ray diffraction. Hydrogen atoms were not taken into consideration in the PDF calculations since, due to their weak scattering power, they have only a minor effect on the PDF curves.

## 3. Method development

### 3.1. Description of molecular structure, position and orientation

**3.1.1. Rigid molecules.** Most crystallographic software, including most PDF programs, describe a crystal structure by its space group, lattice parameters and fractional coordinates of all atoms inside the asymmetric unit. This allows an easy refinement of all atoms individually. For molecular compounds, it is more appropriate to move all atoms of a molecule simultaneously in order to avoid bond breaking or distortions of the molecule. Therefore we use the molecular structure, the molecular position and its spatial orientation to describe the crystal structure.

In a first approach the molecule is treated as a rigid body. The position is given as the centre of the molecule ( $X_m$ ,  $Y_m$ ,  $Z_m$ ). The spatial orientation is described by quaternions ( $Q_0$ ,  $Q_1$ ,  $Q_2$ ,  $Q_3$ ) (Fritzer, 2001). For structures with molecules on crystallographic inversion centres, the molecular positions were fixed, and the refinements were performed with a complete molecule in one of the corresponding subgroups (*e.g.* in  $P2_1$ ,  $Z = 2$  instead of  $P2_1/a$ ,  $Z = 2$ ). Additionally, calculations were performed with up to four symmetrically independent molecules per asymmetric unit (*e.g.* in  $P1$ ,  $Z = Z' = 4$  instead of  $P2_1/c$ ,  $Z = 4$ ,  $Z' = 1$ ). In this case, each molecule was described by its own set of positional and orientational variables.

**3.1.2. Flexible molecules.** For flexible molecules, a different approach was used. Instead of describing the molecule as a rigid body, all atoms are refined individually. Restraints (soft constraints) were used for all bond lengths, all bond angles and all torsion angles except those that should be refined freely. For paracetamol, a CSD search revealed that the torsion angle  $\tau_1$  between the phenyl ring and the amide group (Fig. 4a) shows a broad distribution (see Fig. S1 in the supporting information). The amide group generally adopts a planar *trans* conformation, with a torsion angle  $\tau_2$  (C–N–C=O) between the PhNH and the COCH<sub>3</sub> group close to 0° (see Fig. S2). Hence this torsion angle was included in the set of restraints. The rotation of the OH group out of the phenyl ring plane, and the rotation of the methyl group could be ignored since all calculations were performed without taking hydrogen atoms into account.

### 3.2. Calculating the PDF of molecular systems

For a known structural model, the PDF is calculated using the relation (Farrow *et al.*, 2011)

$$G_c(r) = \frac{1}{Nr} \sum_i \sum_{j \neq i} \left[ \frac{f_i f_j}{\langle f \rangle^2} \delta(r - r_{ij}) \right] - 4\pi r \rho_0. \quad (4)$$

The sum iterates over all pairs of atoms  $i$  and  $j$  separated by distance  $r_{ij}$  within the structure model. The scattering powers of atoms  $i, j$  are  $f_i, f_j$ ,  $\langle f \rangle$  is the average atomic scattering factor and  $N$  the number of atoms in the structural model. Equation (4) generates peaks at every position  $r$  where the structural

model has pairs of atoms separated by this distance. In order to model the peak broadening due to the ADP (which we note includes static disorder effects as well as thermal motion) the  $\delta(r - r_{ij})$  function in equation (4) is convolved with a Gaussian-like profile function. To account for the finite  $Q_{\max}$  range in the measurement, the curve is also convolved with the Fourier transform of a step function terminated at  $Q_{\max}$  (Proffen & Billinge, 1999).

In order to model the simulated PDF curve as accurately as possible, we used the method described in our previous paper (Prill *et al.*, 2015). This procedure uses two different isotropic ADPs,  $B_{\text{intra}}$  and  $B_{\text{inter}}$ , for intra- and intermolecular atom pairs, which results in an excellent modelling of both the sharp intramolecular and broad intermolecular PDF peaks (Prill *et al.*, 2015). Here we take advantage of *pyobjcryst*, an extension to the *DiffPy-CMI* framework that makes available functionality from *ObjCryst++* (Favre-Nicolin & Cerný, 2002) within *DiffPy-CMI* that implements the molecule-in-crystal structure representation that we make use of. *DiffPy-CMI*, including *pyobjcryst*, is an open-source program that is available from <http://www.diffpy.org/>.

All calculations were performed with the program *DiffPy-CMI* (Juhás *et al.*, 2015), which is a powerful, extensible software framework for structure fitting that is operated through a series of user-written Python scripts. The program *DiffPy-CMI* searches for the structure that is most consistent with the experimental PDF by varying user-chosen parameters in the model and using a user-chosen regression method, here least-square minimization. The agreement of the calculated and experimental PDF is described by the residual function

$$R_w = \left\{ \frac{\sum_i^N [G_{\text{obs}}(r_i) - G_{\text{calc}}(r_i; \mathbf{p})]^2}{\sum_i^N G_{\text{obs}}^2(r_i)} \right\}^{1/2} \quad (5)$$

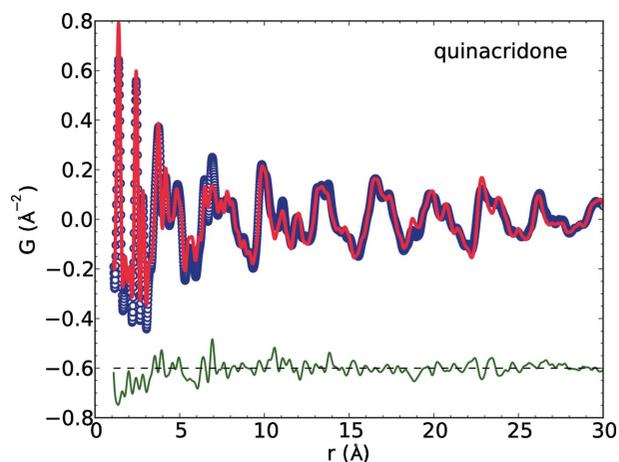
where  $G_{\text{obs}}$  is the experimental PDF,  $G_{\text{calc}}$  the PDF calculated from the model and  $\mathbf{p}$  is the list of parameters in the model. In the PDF refinements done during the course of this work, the lattice parameters, the two isotropic displacement parameters, the scale factor and the variables for molecular position and orientation, or all atomic positions, respectively, were refined. As starting values we chose lattice parameters close to the experimental values. For the molecular positions and orientations two sets of calculations were performed: (i) the calculations started from values close to the experimental ones. This corresponds to a structure refinement based on a given structural model. (ii) The calculations were started with random values assigned automatically for the molecular position and orientation. The method of an automatic and straightforward designation of these parameters corresponds to a structure solution with subsequent refinement. For flexible molecules only structure refinements based on a given structural model were performed hitherto. Further developments are currently in progress.

## 4. Results

### 4.1. Quinacridone

**4.1.1. Refinement of a given structural model by fitting to the PDF curve.** The crystal structure of quinacridone was fitted to the PDF. Since the quinacridone molecule is situated on a crystallographic inversion centre in  $P2_1/c$ ,  $Z = 2$ , the position of the molecule was kept fixed. A simultaneous refinement of molecular orientation, lattice parameters ( $a$ ,  $b$ ,  $c$  and  $\beta$ ) and all other variables leads quickly and easily to a good agreement between experimental and simulated PDF curves (Fig. 9). The calculated and experimentally determined structures match very well (Fig. 10). This shows that a refinement of a given structural model of an organic compound to a PDF curve is feasible.

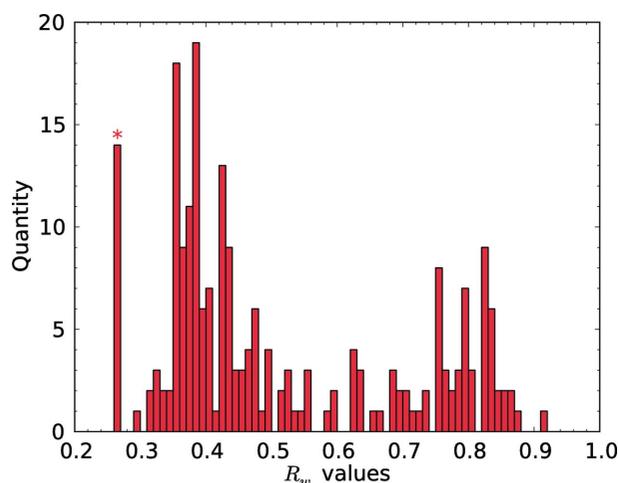
**4.1.2. Structure solution by fitting to the PDF curve.** In a second run, the calculations started with random values for the molecular orientation. This corresponds to a structure solution with given lattice parameters, as done similarly in real-space methods. The molecular position was kept on the inversion centre in  $P2_1/c$ . The molecular orientation, lattice parameters ( $a$ ,  $b$ ,  $c$  and  $\beta$ ) and all other variables were simultaneously refined. From 220 different random starting orientations, the correct structure was found 14 times with an  $R_w$  value of 0.26. The reproducibility was excellent. The  $R_w$  values of all 14 correct structures differ by less than 0.000001 whereas the best incorrect structure had an  $R_w$  value higher by 0.03 (Fig. 11). The fit of the correct structure (Fig. 12) was as good as when



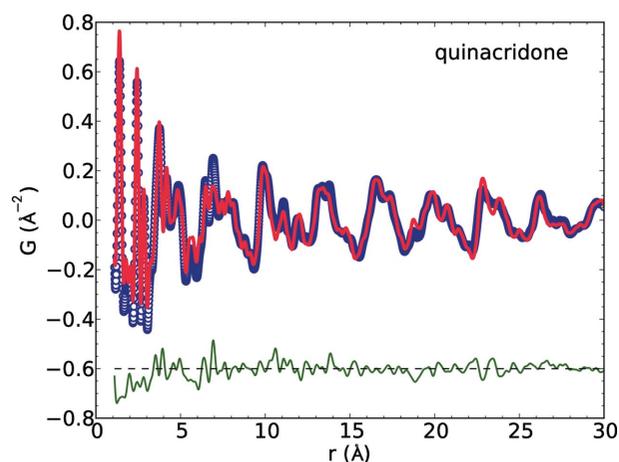
**Figure 9**  
Refinement of the crystal structure of quinacridone by fitting to the PDF data, starting from the experimental structure. Experimental PDF (blue), calculated PDF (red) and difference curve (green).



**Figure 10**  
Overlay of the single-crystal structure of quinacridone (blue) and the structure obtained by fitting to the PDF curve (red), starting from the experimental structure, viewed along the [010] direction.



**Figure 11**  
Distribution of  $R_w$  values for quinacridone calculation, starting from random values for the molecular orientation. The star denotes the correct solutions.



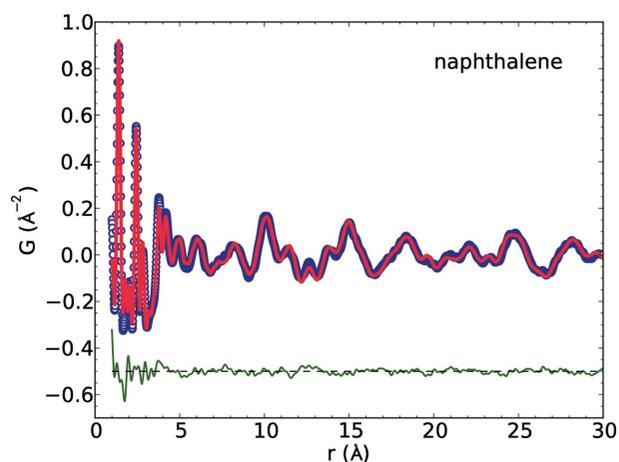
**Figure 12**  
Solution and refinement of the crystal structure of quinacridone by fitting to the PDF data, starting from random values. Experimental PDF (blue), calculated PDF (red) and difference curve (green).



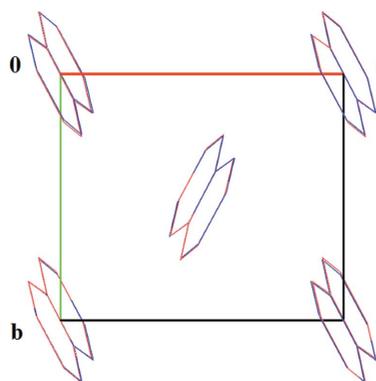
**Figure 13**  
Overlay of the single-crystal structure of quinacridone (blue) and the structure obtained by fitting to the PDF curve (red), starting from random values, viewed along the  $[010]$  direction.

starting from a correct orientation. Also the resulting crystal structures were identical (Fig. 13). This demonstrates that organic crystal structures of rigid molecules can not only be *refined*, but even be *solved* by fitting to the PDF curve, given a known unit cell and space group.

The refinements went so well that, for all following refinements, we did not give the correct molecular orientation as input, but started directly with random values for the molecular orientation.



**Figure 14**  
Naphthalene. Structure determination in  $P2_1/a$ ,  $Z = 2$ . Experimental PDF (blue), calculated PDF (red) and difference curve (green).

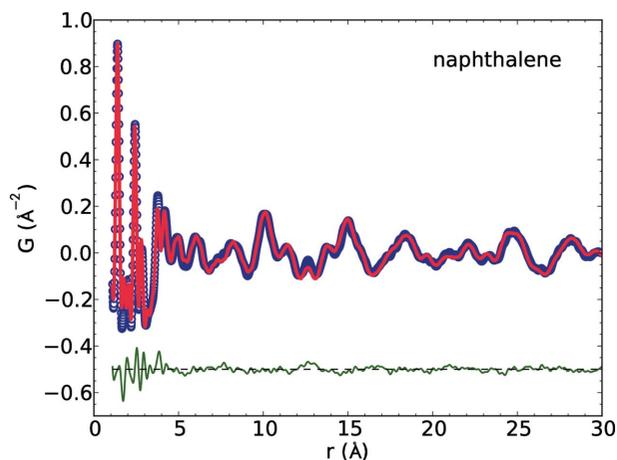


**Figure 15**  
Overlay of the single-crystal structure of naphthalene (blue) and the structure determined by fitting to PDF data in  $P2_1/a$ ,  $Z = 2$  (red), viewed along the  $[001]$  direction.

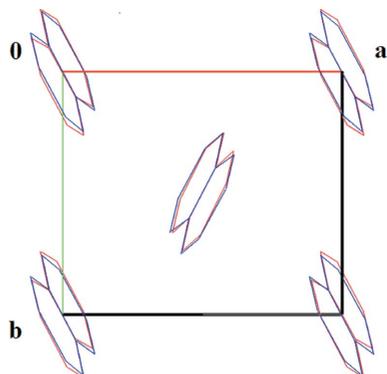
## 4.2. Naphthalene

**4.2.1. Structure determination in  $P2_1/a$ ,  $Z = 2$ .** The crystal structure of naphthalene was solved by fitting to the experimental PDF in its experimental space group  $P2_1/a$ ,  $Z = 2$ . The molecule is situated on an inversion centre; hence the molecular position was kept fixed and the calculations were performed with a complete molecule in the subgroup  $P2_1$ . For the orientation, random starting values were used. The molecular orientation, lattice parameters ( $a$ ,  $b$ ,  $c$  and  $\beta$ ) and all other variables were simultaneously refined. The calculation was performed 240 times. The correct structure was found 79 times, with very smooth difference curves and  $R_w$  values between 0.159 and 0.160 (Fig. 14). The resulting structure agrees excellently with the structure determined from single-crystal data (Fig. 15). The best incorrect structure had a significantly higher  $R_w$  value of 0.24.

**4.2.2. Structure determination in  $P1$ ,  $Z = 2$ .** In addition, a second calculation was performed in space group  $P1$  with  $Z' = 2$ . Random values were used for the positions and orientations of the two symmetrically independent molecules. The lattice parameters  $a$ ,  $b$ ,  $c$  and  $\beta$  were taken from the single-crystal data, but were refined simultaneously with the para-



**Figure 16**  
Naphthalene. Structure determination in  $P1$ ,  $Z = 2$ . Experimental PDF (blue), calculated PDF (red) and difference curve (green).

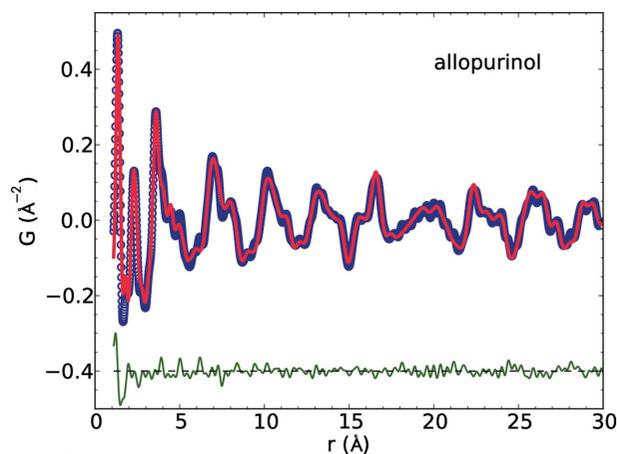


**Figure 17**  
Overlay of the single-crystal structure of naphthalene (blue) and the structure determined by fitting to PDF data in  $P1$ ,  $Z = 2$  (red), viewed along the  $[001]$  direction.

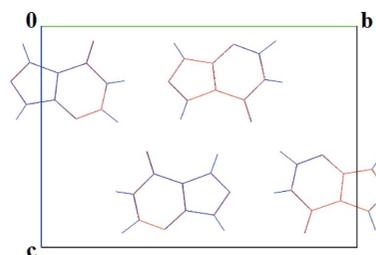
meters for molecular position and orientation. The calculation was performed 280 times and the correct structure was found seven times. The fit (Fig. 16) in  $P1$  was slightly worse than in the calculations in  $P2_1/a$  with the molecule fixed on the inversion centre. In  $P1$ , the  $R_w$  value was 0.17 (instead of 0.16), which is still well below the  $R_w$  value of 0.20 for the lowest incorrect structure. Again, the correct refined structure and the structure from single-crystal data match very well (Fig. 17).

### 4.3. Allopurinol

**4.3.1. Structure determination in  $P2_1/c$ ,  $Z' = 1$ .** At first, the crystal structure of allopurinol was determined in its experimental space group  $P2_1/c$  with the molecule on the general position. Random starting values were used for the molecular position and orientation, whereas the starting values for lattice parameters were taken from the single-crystal structure. All parameters were refined simultaneously. The molecular geometry was kept fixed. In 329 runs, the correct structure was found 12 times with smooth difference curves and an  $R_w$  value of 0.18 with mutual differences of less than 0.0001 (Fig. 18). The resulting structure matches very well the structure from single-crystal data (Fig. 19). The second best result corresponded to an incorrect structure with an  $R_w$  value of 0.26.



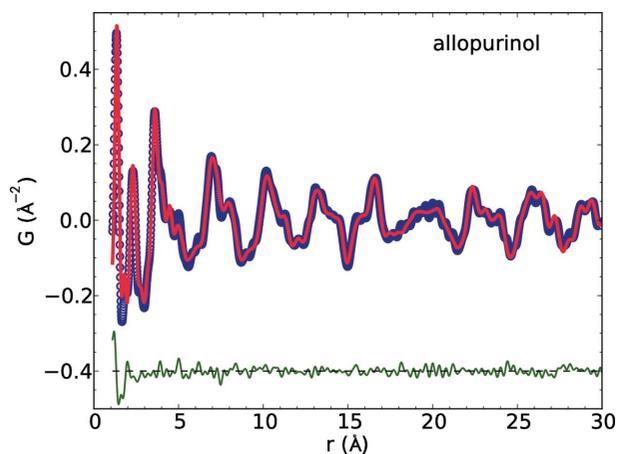
**Figure 18**  
Allopurinol. Structure determination in  $P2_1/c$ ,  $Z' = 1$ . Experimental PDF (blue), calculated PDF (red) and difference curve (green).



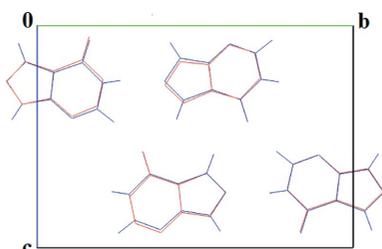
**Figure 19**  
Overlay of the single-crystal structure of allopurinol (blue) and the structure determined by fitting to PDF data in  $P2_1/c$  (red), viewed along the  $[100]$  direction.

**4.3.2. Structure determination in  $P1$ ,  $Z = 4$ .** In order to test the ability of the PDF fit, we attempted to solve the structure in  $P1$  with four independent molecules. Independent variables were used for each molecular position and orientation. The molecular geometry was kept fixed, but random values were used for the molecular orientations and positions. The lattice parameters ( $a$ ,  $b$ ,  $c$  and  $\beta$ , starting from single-crystal structure values) were refined simultaneously with the parameters for molecular positions and orientations. The calculation was performed 534 times. The two structures with the lowest  $R_w$  values (0.1933 and 0.1947) corresponded to the correct structure. The difference curve (Fig. 20) is smooth, as for the calculations in  $P2_1/c$ . Despite the high number of independent molecules, the resulting structure is very close to having  $P2_1/c$  symmetry, and matches very well the experimental single-crystal structure (Fig. 21).

With somewhat worse  $R_w$  values (0.22), other structures in  $P1$  appeared, which exhibited a similar molecular packing, but in which one or two of the four allopurinol molecules were rotated by  $180^\circ$ . Very similar phenomena (low success rate, inverted molecules) are also frequently observed when structures with several independent molecules are solved from powder data by real-space methods, or when such crystal structures are predicted by lattice-energy minimizations. Also in these cases structures are obtained which exhibit the 'correct' molecular packing, but contain some molecules with wrong orientations. Apparently, the flipping of a planar



**Figure 20**  
Allopurinol. Structure determination in  $P1$ ,  $Z' = 4$ . Experimental PDF (blue), calculated PDF (red) and difference curve (green).

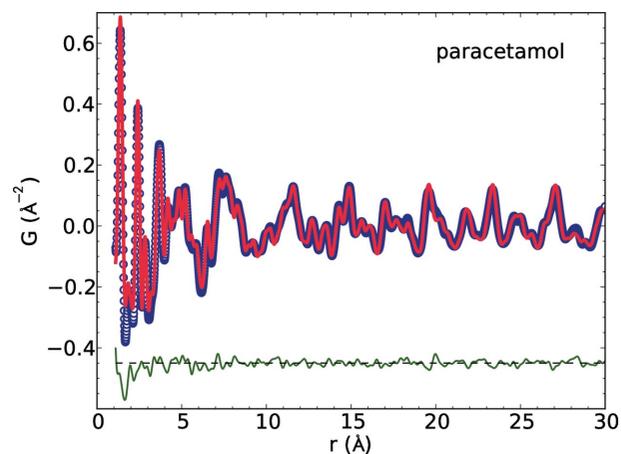


**Figure 21**  
Overlay of the single-crystal structure of allopurinol (blue) and the structure determined by fitting to PDF data in  $P1$ ,  $Z' = 4$  (red), viewed along the  $[100]$  direction. Despite the fact that the refinement was performed with four independent molecules in  $P1$ , the agreement is excellent.

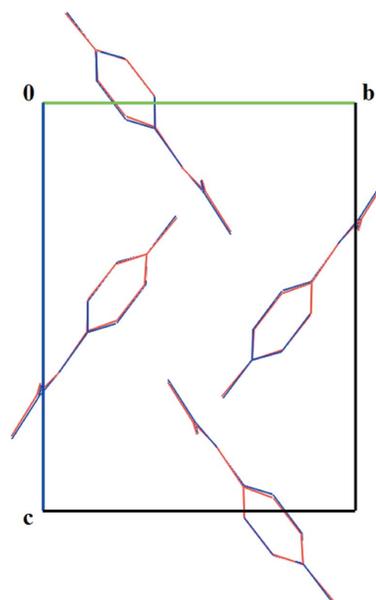
aromatic molecule by  $180^\circ$  does not have a strong effect – neither on the powder pattern, nor on the PDF, nor on the lattice energy.

#### 4.4. Paracetamol

**4.4.1. Refinement of a given structural model by fitting to the PDF curve using restraints for the molecular geometry.** As an example of a fit with a flexible molecule, the crystal structure of paracetamol was chosen. The molecule is located on the general position in  $P2_1/n$ . All atomic positions were refined together with the lattice parameters ( $a$ ,  $b$ ,  $c$  and  $\beta$ ), scale factor and two isotropic displacement parameters using restraints for bond lengths, bond angles and torsion angles. The first run started from the experimental structure. In the subsequent runs the minimization started with a considerably distorted molecular geometry. In all cases the refinement converged well with an  $R_w$  value of 0.17 (Fig. 22). The resulting structure agrees well with the experimental one (Fig. 23). This indicates that organic crystal structures can be refined by a fit to the PDF curve even if the molecule is flexible. We have also tried this approach for solving the structure of paracetamol, using random starting values for the position and orientation of the molecule as well as for the intramolecular degree of freedom. First results were promising, and further development is in progress.



**Figure 22**  
Paracetamol. Refinement in  $P2_1/n$  with a flexible molecule using restraints. Experimental PDF (blue), calculated PDF (red) and difference curve (green).



**Figure 23**  
Overlay of the single-crystal structure of paracetamol (blue) and the structure refined by a fit to PDF data in  $P2_1/n$  with a flexible molecule using restraints (red), viewed along the  $[100]$  direction.

## 5. Discussion

For obtaining a good fit between simulated and experimental PDF curves, two points turned out to be crucial: (i) the experimental PDF data must be reliable and free of artefacts. This requires a careful measurement of the powder patterns and a careful background correction procedure. (ii) In order to correctly simulate the mixture of sharp and broad peaks in the PDF curve, it is critical to distinguish between intramolecular and intermolecular atom–atom distances. The recently developed procedure which uses two different  $B$  values for intramolecular and intermolecular distances (Prill *et al.*, 2015) turned out, again, to give very good results. Under these conditions, the fit of organic crystal structures works very well. The correct solutions always had considerably lower

$R_w$  values than all incorrect ones, even in the challenging case of allopurinol in space group  $P1$  with four independent molecules.

The fit of the molecular positions and orientations to the PDF data described here resembles a rigid-body Rietveld refinement. Whereas the Rietveld refinement performs a fit to the powder pattern itself, our method uses its Fourier transform instead. The structure solution procedure described here – fitting molecular position and orientation starting from random values – resembles the ‘real-space methods’ for solving crystal structures from powder diffraction data. Like our method, real-space methods start from a given molecular geometry and given lattice parameters and optimize the positions and orientations of the molecules, until the simulated pattern matches the experimental one as closely as possible. Again, the real-space method performs a fit to the powder pattern itself, whereas our method fits to its Fourier transform. In principle, a powder pattern contains the same amount of information as its Fourier transform. Hence the orientation and translation fit to the PDF curve should work as well as the real-space methods. This is proven by the example of allopurinol in  $P1$  with four independent molecules. A structure solution with  $Z' = 4$  in  $P1$ , requiring 21 structural parameters ( $4 \times 3$  for orientation,  $4 \times 3$  for position, minus 3 for the floating origin), is quite challenging for real-space methods (Fernandes *et al.*, 2007). Nevertheless, the fit to the PDF curve in  $P1$  with  $Z' = 4$  was successful, although we additionally refined the four lattice parameters ( $a$ ,  $b$ ,  $c$ ,  $\beta$ ), bringing the total to 25 independent structural parameters, which were refined simultaneously. This shows that it should also be possible to successfully solve structures with counter ions, hydrates and solvates. The first tests on paracetamol as a flexible molecule show that intramolecular degrees of freedom should, in principle, not be a major problem. However, based on experience with the real-space methods, we expect that the difficulties increase exponentially with increasing number of intramolecular degrees of freedom.

A procedure, similar to the PDF fit and the real-space methods, is also possible in Patterson space. For example, Egert developed a method, called PATSEE, for structure solution of organic compounds from single-crystal data in Patterson space (Wagner *et al.*, 2001). In a Patterson map, the near-origin peaks correspond to intramolecular atom–atom vectors, whereas the peaks further away originate from intermolecular atom–atom vectors. This feature is used to determine the molecular orientation and the molecular position in two separate steps. In the first step, a Patterson map of a given molecule is calculated, and the orientation of this pattern – and thereby the orientation of the molecule – is fitted to the near-origin peaks of the experimental Patterson map. In the second step the remaining peaks are used to determine the molecular positions and the space-group symmetry. This method also works for structure determination from powder data. Depending on the data quality it is even possible to use only a part of the molecule, *e.g.* a sufficiently large rigid unit, in the search. A prerequisite is a powder of good crystallinity which allows the extraction of reliable

reflection intensities for the calculation of the three-dimensional Patterson map. In contrast, the method developed here is designed for samples of low crystallinity which do not allow for a reliable extraction of a sufficient number of reflection intensities so that the three-dimensional Patterson function cannot be derived. The near-origin peaks in the PDF curve, which originate from intramolecular distances only, cannot be used to determine the orientations of the molecules. Hence the orientations have to be determined from the intermolecular distances, together with the molecular positions, by fitting to the full PDF curve.

Patterson, real-space and Rietveld methods were developed many years ago. Consequently, the corresponding software is well developed and user friendly; and there is much experience on how to handle difficult problems. In contrast, the fitting of organic structures to the PDF curve is in its infancy and not much experience is available. In principle, a fit to the PDF curve should be as powerful as Patterson, real-space and Rietveld methods. At present, the PDF fit requires knowledge of the lattice parameters (at least approximately), like Patterson, real-space and Rietveld methods. If the powder diagram cannot be indexed, lattice parameters can, *e.g.*, be determined by electron diffraction (see *e.g.* Gorelik *et al.*, 2009; Schmidt *et al.*, 2009). The next development step should be to extend the PDF-fit method to cases where the lattice parameters are not known in advance.

The main advantage of performing structure solution and refinement in PDF space instead of in diffraction space is that the PDF also contains information from the diffuse scattering, which is ignored in the real-space and Rietveld refinement procedures. Hence, working in PDF space lends its use to all materials which exhibit diffuse scattering, especially for poorly crystalline, nanocrystalline, disordered and amorphous samples.

The synchrotron powder pattern of naphthalene used here contained only about six peaks and some humps; nevertheless, with given lattice parameters, it was possible to solve and refine the crystal structure by PDF fit, even with two independent molecules. With classical approaches, *e.g.* real-space methods, it would be hardly possible to solve a structure with  $Z' = 2$  from these data. Also a Rietveld refinement in  $P1$ ,  $Z' = 2$  would be highly unreliable with these data. The investigation of nanocrystalline organic samples by PDF fit is in progress. We expect that the PDF fit will still work for nanocrystalline samples and other cases which cannot be treated with classical approaches.

## 6. Conclusion

This work has shown that it is possible to solve and refine organic crystal structures by fitting to the pair distribution function. The molecular geometry and approximate values for the lattice parameters are required as inputs. The structure of allopurinol could even be successfully solved in  $P1$  with four symmetrically independent molecules per unit cell. A first test on paracetamol indicates that flexible molecules could also be treated successfully. There are many cases of interest which

cannot be treated by classical crystallographic methods, but can be analysed by PDF fit. This initial paper describes the method and demonstrates the power of a PDF fit for organic compounds. We are convinced that information on the solid-state structures of poorly crystalline, nanocrystalline and amorphous organic compounds can in future be gained by fitting structural models to the PDF data.

## Acknowledgements

The authors acknowledge the participation of Jan von Cosel and Benedikt Kloss (both Goethe University) on this project. Data collection, development of the *DiffPy-CMI* modelling software and PDF simulations were supported by Laboratory Directed Research and Development (LDRD) Program 12-007 (Complex Modelling) at Brookhaven National Laboratory (BNL), which is supported by US Department of Energy through contract SC0012704. X-ray experiments were carried out at the National Synchrotron Light Source (NSLS) beamline X17A, at BNL. BNL is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences, DE-SC00112704.

## References

- Allen, F. H. (2002). *Acta Cryst.* **B58**, 380–388.
- Banerjee, K. (1929). *Nature*, **125**, 456–457.
- Billinge, S. J. L. (2008). *J. Solid State Chem.* **181**, 1695–1700.
- Billinge, S. J. L., Dykhne, T., Juhás, P., Božin, E., Taylor, R., Florence, A. J. & Shankland, K. (2010). *CrystEngComm*, **12**, 1366–1368.
- Billinge, S. J. L. & Kanatzidis, M. G. (2004). *Chem. Commun.* pp. 749.
- Bouhmaida, N., Bonhomme, F., Guillot, B., Jelsch, C. & Ghermani, N. E. (2009). *Acta Cryst.* **B65**, 363–374.
- Bragg, W. L. (1921). *Proc. Phys. Soc.* **34**, 33–50.
- Bragg, W. & Bragg, W. (1913). *Nature*, **91**, 557.
- Brühne, S., Gottlieb, S., Assmus, W., Alig, E. & Schmidt, M. U. (2008). *Cryst. Growth Des.* **8**, 489–493.
- Cliffe, M. J., Dove, M. T., Drabold, D. A. & Goodwin, A. L. (2010). *Phys. Rev. Lett.* **104**, 125501.
- Cruickshank, D. W. J. (1957). *Acta Cryst.* **10**, 504–508.
- Davis, T., Johnson, M. & Billinge, S. J. L. (2013). *Cryst. Growth Des.* **13**, 4239–4244.
- Debye, P. & Menke, H. (1930). *Phys. Z.* **31**, 797–798.
- Debye, P. & Scherrer, P. (1916). *Phys. Z.* **17**, 277–283.
- Dykhne, T., Taylor, R., Florence, A. & Billinge, S. J. L. (2011). *Pharm. Res.* **28**, 1041–1048.
- Egami, T. & Billinge, S. J. L. (2012). *Underneath the Bragg Peaks: Structural Analysis of Complex Materials*, 2nd ed. Amsterdam: Elsevier.
- Farrow, C. L. & Billinge, S. J. L. (2009). *Acta Cryst.* **A65**, 232–239.
- Farrow, C. L., Shaw, M., Kim, H.-J., Juhás, P. & Billinge, S. J. L. (2011). *Phys. Rev. B*, **84**, 134105.
- Favre-Nicolin, V. & Cerný, R. (2002). *J. Appl. Cryst.* **35**, 734–743.
- Fernandes, P., Shankland, K., Florence, A. J., Shankland, N. & Johnston, A. (2007). *J. Pharm. Sci.* **96**, 1192–1202.
- Fritzer, H. P. (2001). *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **57**, 1919–1930.
- Gorelik, T., Schmidt, M. U., Brüning, J., Bekő, S. L. & Kolb, U. (2009). *Cryst. Growth Des.* **9**, 3898–3903.
- Hammersley, A. P., Svensson, S. O., Hanfland, M., Fitch, A. N. & Hausermann, D. (1996). *High. Press. Res.* **14**, 235–248.
- Herbst, W. & Hunger, K. (2004). *Industrial Organic Pigments*, 3rd ed. Weinheim: Wiley VCH.
- Hull, A. W. (1917). *Phys. Rev.* **10**, 661–696.
- Juhás, P., Cherba, D. M., Duxbury, P. M., Punch, W. F. & Billinge, S. J. L. (2006). *Nature*, **440**, 655–658.
- Juhás, P., Davis, T., Farrow, C. L. & Billinge, S. J. L. (2013). *J. Appl. Cryst.* **46**, 560–566.
- Juhás, P., Farrow, C., Yang, X., Knox, K. & Billinge, S. J. L. (2015). *Acta Cryst.* **A71**, 562–568.
- Newman, A., Engers, D., Bates, S., Ivanisevic, I., Kelly, R. C. & Zografi, G. (2008). *J. Pharm. Sci.* **97**, 4840–4856.
- Nollenberger, K., Gryczke, A., Meier, C., Dressman, J., Schmidt, M. U. & Brühne, S. (2009). *J. Pharm. Sci.* **98**, 1476–1486.
- Patterson, A. L. (1934). *Phys. Rev.* **46**, 372–376.
- Patterson, A. L. (1935). *Z. Kristallogr.* **90**, 517–542.
- Paulus, E. F., Leusen, F. J. J. & Schmidt, M. U. (2007). *CrystEngComm*, **9**, 131–143.
- Prill, D., Juhás, P., Schmidt, M. U. & Billinge, S. J. L. (2015). *J. Appl. Cryst.* **48**, 171–178.
- Proffen, Th. & Billinge, S. J. L. (1999). *J. Appl. Cryst.* **32**, 572–575.
- Prusiner, P. & Sundaralingam, M. (1972). *Acta Cryst.* **B28**, 2148–2152.
- Rademacher, N., Daemen, L. L., Chronister, E. L. & Proffen, T. (2012). *JACM*, **45**, 482–488.
- Schmidt, M. U. *et al.* (2009). *Acta Cryst.* **B65**, 189–199.
- Sheth, A. R., Bates, S., Muller, F. X. & Grant, D. J. W. (2005). *Cryst. Growth Des.* **5**, 571–578.
- Temleitner, L. (2014). *J. Mol. Liq.* **197**, 204–210.
- Terban, M. W., Johnson, M., DiMichiel, M. & Billinge, S. J. L. (2015). *Nanoscale*, **7**, 5480–5487.
- Wagner, K., Hirschler, J. & Egert, E. (2001). *Z. Kristallogr.* **216**, 565–572.
- Warren, B. E. (1990). *X-ray Diffraction*. New York: Dover.
- Wolf, A. K., Brühne, S., Glinemann, J., Hu, C., Kirchner, M. T. & Schmidt, M. U. (2012). *Z. Kristallogr.* **227**, 113–121.
- Young, C. A. & Goodwin, A. L. (2011). *J. Mater. Chem.* **21**, 6464–6476.