Patterson-function direct methods for structure determination of organic compounds from powder diffraction data. XVI

Jordi Rius

Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, 08193 Bellaterra, Catalonia 08193, Spain. Correspondence e-mail: jordi.rius@icmab.es

A new type of direct methods (DM) called Patterson-function DM are presented that directly explore the Patterson instead of the modulus function. Since they work with the experimental intensities, they are particularly well suited for handling powder diffraction data. These methods are based on the maximization of the sum function $S_P \propto \sum \mathbf{H}(I_\mathbf{H} - \langle I \rangle)G_{-\mathbf{H}}(\Phi)$ in terms of the Φ phases of the structure factors. The quantity accessible from the experiment is $I_\mathbf{H}$, the equidistributed multiplet intensity of reflection $\mathbf{H}$, and $\langle I \rangle$ is the average intensity taken over all non-systematically absent reflections. $G_{-\mathbf{H}}(\Phi)$ is the calculated structure-factor amplitude of the squared structure that includes the positivity and the atomicity of the density function in its definition. The S_P sum function can be optimized with the Patterson-function tangent formula (TF) using a variant of the S-FFT algorithm [Rius *et al.* (2007), *Acta Cryst.* A**63**, 131–134]. It is important that overlapped reflections also participate in the phase refinement, so that not only the resolved reflections but the whole pattern contribute decisively to the refinement. The increase in effective data resolution minimizes Fourier series termination effects and improves the accuracy of $G(\Phi)$. The Patterson-function TF has been applied to synchrotron powder data of various organic compounds. In all cases the molecules were easily identified in the respective Fourier maps. By way of illustration the method is applied to synchrotron powder data of a dimer formed by 30 symmetry-independent non-H atoms. Since single-crystal data may be regarded as overlap-free powder data, it is clear that Patterson-function DM can cope with powder and single-crystal data.

1. Introduction

While structure refinement from carefully measured powder diffraction patterns proves to be a rather routine task using the Rietveld method (Rietveld, 1967), the inherent presence in the pattern of overlapping reflections often renders difficult or even prevents the derivation of the starting structural model. Crystal structure solution from powder data is usually carried out either by (i) extracting the intensities of the resolved reflections and processing them by direct methods (DM), or (ii) minimizing the difference between observed and calculated pattern profiles as a function of the structural parameters (direct-space methods). Recently, other methods based on techniques of density-function modification combined with additional density-function information such as, for example, histogram matching (Baerlocher *et al.*, 2007) have appeared with some success. In the case of DM the bottleneck is to obtain sufficient accurate individual intensity data when peak overlap is severe. In contrast, for direct-space

methods, the principal limitation is the introduction of chemical information in the form of expected building units. This extra information makes them somewhat more robust against poor data quality but at the same time makes them less general than DM. In the application of DM to powder data it is often observed that the calculated Fourier peaks appear shifted from their ideal positions, thus complicating map interpretation. This is principally caused by Fourier series termination effects owing to limited data resolution. Recently, Altomare *et al.* (2008) have developed algorithms attempting to reduce the resolution bias in electron-density maps. No doubt that one of the principal causes for inaccurate Fourier maps when applying DM is the more or less straightforward adaptation of conventional DM strategies, initially developed for single-crystal (SC) data, to powder (PD) data. For the specific case of DM based on the origin-free modulus sum function (Rius, 1993), the quantity to be maximized in terms of the collectivity Φ of phases of the quasi-normalized structure factors $\mathbf{E}$ is (Rius, 2002)

$$S_{\mathrm{M}} = 2 \sum_{\mathbf{H}} \frac{E_{\mathbf{H}} - \langle E \rangle}{N^{1/2}} G_{-\mathbf{H}}(\Phi), \qquad (1)$$

where the Patterson-type function being explored, *i.e.* the modulus function, is indicated by the M subscript and where $G_{-\mathbf{H}}(\Phi)$ are the calculated amplitudes of the structure factors of the squared structure. Since peak overlap increases with Bragg angle, resolved reflections will mostly correspond to data at low or moderate resolution. It is evident that the limited number of resolved reflections will negatively affect phase refinement and, by extension, the corresponding Fourier maps especially for organic compounds.

Historically, prior to the introduction of the Rietveld method, the structural parameters, X, were refined by least-squares methods using as experimental data groups of over-lapping intensities (Rietveld, 1966; Will, 1979), *i.e.* by mini-mizing a residual of the type

$$R_1 = \sum_{i=1}^{N_{\mathrm{m}}} w_i \left[\sum_{\mathbf{k}(i)} j_{\mathbf{k}} E_{\mathbf{k}}^2 - \sum_{\mathbf{k}(i)} j_{\mathbf{k}} E_{\mathbf{k}}^2(X) \right]^2, \qquad (2)$$

in which $j_{\mathbf{k}}$ is the multiplicity of reflection $\mathbf{k}$ in multiplet i and N_{m} is the number of multiplets (clusters of unresolved reflections) in the pattern. By using the total intensity of each multiplet as a single data unit, crystal structures were refined accurately without considering the individual peak profiles during the refinement. It is clear that this practice is only useful when N_{m} is much larger than the number of structural parameters. In present days, integrated intensities can be extracted by means of whole-pattern matching methods like the iterative Le Bail algorithm (Le Bail *et al.*, 1988) which contributes to reduce the impact of this limitation. Instead of X, one can select in (2) the collectivity of phases of the structure factors (Φ) as variables. In contrast to phase refinement with S_{M} where only the resolved reflections participate in the summation, phase refinement with R_1 involves as many sum terms as there are multiplets in the pattern, *i.e.* the whole pattern takes part. In this way the effective data resolution is enhanced, which contributes to minimize Fourier series termination effects. Unfortunately, R_1, when expressed in terms of Φ, is difficult to manipulate; however, the closely related Patterson-function sum function can be used instead. This sum function constitutes the basis of the Patterson-function DM that are described more exten-sively in §3.

2. Intensity information from a powder diffraction pattern

Let the structure factor of an arbitrary $\mathbf{H}$ reflection (not systematically absent by space-group symmetry) be

$$\mathbf{F}_{\mathbf{H}} = F_{\mathbf{H}} \exp(i\varphi_{\mathbf{H}}) = \sum_{j=1}^{N} f_j(\mathbf{H}) \exp(i2\pi\mathbf{H}\mathbf{r}_j), \qquad (3)$$

where f_j is the form factor of atom j, $\mathbf{r}_j$ is the position vector of atom j, and N is the total number of atoms in the unit cell. The structure factor can be corrected for atom scattering fall-off through division by

$$\left[\sum_{p=1}^{N} f_p^2(\mathbf{H}) \right]^{1/2} \qquad (4)$$

to give the so-called quasi-normalized structure factor

$$\mathbf{E}_{\mathbf{H}} = E_{\mathbf{H}} \exp(i\varphi_{\mathbf{H}}) = \sum_{j=1}^{N} f_j^s(\mathbf{H}) \exp(i2\pi\mathbf{H}\mathbf{r}_j). \qquad (5)$$

In view of equations (3), (4) and (5) and for crystal structures containing equal atoms with the same thermal factors, the decay correction is exact. For this special case the sharpened scattering factor simplifies to

$$f^s(\mathbf{H}) = N^{-1/2}, \qquad (6)$$

so that $\mathbf{E}_{\mathbf{H}}$ may be regarded as the structure factor of a hypothetical crystal structure made up of point-like atoms of height $N^{-1/2}$. The structure factor of the squared crystal structure

$$\mathbf{G}_{\mathbf{H}} = G_{\mathbf{H}} \exp(i\psi_{\mathbf{H}}) \qquad (7)$$

is obtained by squaring $N^{-1/2}$, so that for organic compounds the very simple relationship between $G_{\mathbf{H}}$ and $E_{\mathbf{H}}$ follows,

$$G_{\mathbf{H}}^2 = E_{\mathbf{H}}^2/N. \qquad (8)$$

In the present application of Patterson-function DM to powder data the smallest unit of intensity information is the total intensity of each group of unresolved reflections (multiplet). For an arbitrary ith multiplet, the total intensity is

$$D_i = \sum_{\mathbf{k}(i)} j_{\mathbf{k}} E_{\mathbf{k}}^2, \qquad (9)$$

and the sum of the multiplicities of all symmetry-independent reflections is given by

$$n_i = \sum_{\mathbf{k}(i)} j_{\mathbf{k}}. \qquad (10)$$

If $\mathbf{H}$ is an arbitrary reflection of this multiplet, then, in view of (9) and (10), the equidistributed intensity for $\mathbf{H}$ is

$$I_{\mathbf{H}} = D_i/n_i, \qquad (11)$$

so that $\langle I \rangle$, the average taken over all space-group allowed reflections, is equal to $\langle E^2 \rangle = 1$.

3. Patterson-function direct methods

3.1. Definition of the Patterson-function sum function

Patterson-function DM are based on the maximization, in terms of Φ, of the sum function

$$S_{\mathrm{P}} = 2 \sum_{\mathbf{H}} \frac{I_{\mathbf{H}} - \langle I \rangle}{N} G_{-\mathbf{H}}(\Phi). \qquad (12)$$

In this expression, $(I_{\mathbf{H}} - \langle I \rangle)/N$ is the Fourier coefficient of the best approximation to the origin-free Patterson function (P') derivable from powder data and $G_{-\mathbf{H}}(\Phi)$ is the calculated structure-factor amplitude of the modulus function (M) of the squared structure. The $\mathbf{G}_{-\mathbf{H}}(\Phi)$ structure factor can be

expressed as a function of Φ by applying the convolution theorem to $\rho^2 = \rho \cdot \rho$, i.e.,

$$G_{-\mathbf{H}}(\Phi)\exp\left[i\psi_{-\mathbf{H}}(\Phi)\right] = \frac{1}{V}\sum_{\mathbf{h}}E_{-\mathbf{h}}E_{\mathbf{h}-\mathbf{H}}\exp\left[i(\varphi_{-\mathbf{h}}+\varphi_{\mathbf{h}-\mathbf{H}})\right],$$

(13)

wherein V is the unit-cell volume. Notice that $\rho^2 = \rho \cdot \rho$ can be used because the true density function is assumed to be positive and made of equal atoms. In the present approach ρ corresponds to the Fourier synthesis

$$\rho(\mathbf{r}) = \frac{1}{V}\sum_{\mathbf{H}}w_{\mathbf{H}}E_{\mathbf{H}}\exp(i\varphi_{\mathbf{H}})\exp(-i2\pi\mathbf{H}\mathbf{r}),\qquad(14)$$

where the weight $w_{\mathbf{H}}$ is 1.0 for experimental or extrapolated $E_{\mathbf{H}}$ values greater than $E_{\lim}$ ($\simeq 1.10$) and 0.0 otherwise. In (13), $\psi_{-\mathbf{H}}(\Phi)$ is the phase associated with $\mathbf{G}_{-\mathbf{H}}(\Phi)$ and its value is derived from the available φ estimates. By isolating $G_{-\mathbf{H}}(\Phi)$ in (13) and by introducing the resulting expression in (12), it follows that

$$S_{\mathrm{P}} = \frac{2}{N}\sum_{\mathbf{H}}(I_{\mathbf{H}}-\langle I\rangle)$$

$$\times\left(\frac{1}{V}\sum_{\mathbf{h}}E_{-\mathbf{h}}E_{\mathbf{h}-\mathbf{H}}\exp\left\{i\left[\psi_{\mathbf{H}}(\Phi)+\varphi_{-\mathbf{h}}+\varphi_{\mathbf{h}-\mathbf{H}}\right]\right\}\right),\ (15)$$

where the difference $I_{\mathbf{H}}-\langle I\rangle$ can be either positive or negative.

3.2. The Patterson-function tangent formula

At the maximum, the partial derivatives of S_{P} vanish. By differentiating (15) and by equating to zero and isolating the $\sin\varphi_{\mathbf{h}}/\cos\varphi_{\mathbf{h}}$ quotient in the resulting expression, the desired Patterson-function tangent formula (TF) is obtained. It can be written in compact form as the angular part of $Q_{\mathbf{H}}$, the complex quantity between brackets in (16),

$$\varphi_{\mathbf{h}}(\mathrm{new}) = \text{phase of }\left(\frac{1}{V}\sum_{\mathbf{H}}(I_{\mathbf{H}}-\langle I\rangle)E_{\mathbf{h}-\mathbf{H}}\right.$$

$$\left.\times\exp\left\{i\left[\psi_{\mathbf{H}}(\Phi)+\varphi_{\mathbf{h}-\mathbf{H}}\right]\right\}\right).\qquad(16)$$

The Patterson-function tangent formula (16) can be applied in parallel mode by means of a slightly modified version of the full-symmetry S-FFT algorithm (Rius *et al.*, 2007; Rius & Frontera, 2007). This S-FFT variant is described in Fig. 1. In a first step, the two Fourier syntheses, $\rho = \mathrm{FT}^{-1}[w_{\mathbf{H}}E_{\mathbf{H}}\exp(i\varphi_{\mathbf{H}})]$ and $\delta = \mathrm{FT}^{-1}[I_{\mathbf{H}}-\langle I\rangle\exp(i\psi_{\mathbf{H}})]$, are calculated. In a second step, the $\rho \cdot \delta$ product is performed with $Q_{\mathbf{H}}$ being the corresponding Fourier coefficients (new structure factors). Calculation of $\rho \cdot \delta$ in direct space is very convenient since it allows information on the unit-cell contents to be introduced by computing $Q_{\mathbf{H}}$ with the N largest peaks in $\rho \cdot \delta$. Notice that (16) requires the $E_{\mathbf{H}}$ amplitudes to be known. This forces the amplitudes of overlapped reflections to be updated from cycle to cycle (Fig. 1). This is done using the expression
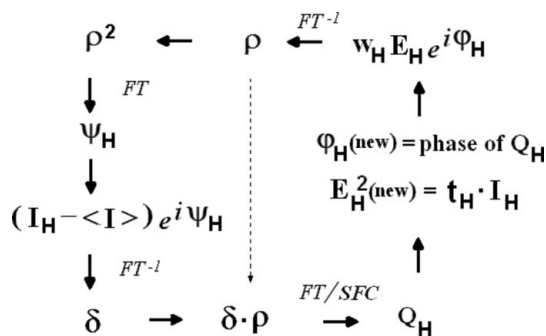

Figure 1
Iterative S-FFT phase-refinement procedure slightly modified for Patterson-function TF refinement with powder data. Initial phase values (upper right corner) are combined with weighted experimental and extrapolated amplitudes to give the initial ρ values (upper right corner). Information on the multiplet intensities is introduced via the $(I_{\mathbf{H}}-\langle I\rangle)$ coefficients. New structure-factor estimates are alternatively obtained by Fourier inverting $\delta \cdot \rho$ (FT) or directly from the N top-ranked Fourier peaks (SFC) of $\delta \cdot \rho$. For overlapped reflections, $E_{\mathbf{H}}^2$ values are updated every cycle [see equation (17) for the meaning of $t_{\mathbf{H}}$] while keeping the global intensity of each multiplet constant.

$$E_{\mathbf{H}}^2(\mathrm{new}) = \left[\frac{n_i Q_{\mathbf{H}}^2}{\sum_{\mathbf{k}}j_{\mathbf{k}(i)}Q_{\mathbf{k}(i)}^2}\right]I_{\mathbf{H}} = t_{\mathbf{H}}I_{\mathbf{H}}\qquad(17)$$

in order to satisfy the total intensity of each multiplet.

4. The Patterson-function TF applied to powder data of organic compounds

In recent years considerable efforts have been devoted to the solution of crystal structures of inorganic compounds from powder diffraction data, so that nowadays many of them can be determined straightforwardly if they are crystalline enough. Principal causes for this success compared with organic compounds are that inorganic ones (i) tend to have smaller unit cells and/or higher metrics (less accidental peak overlap), (ii) contain stronger scatterers with higher contrast (smaller effective number of atoms) and (iii) are more stable (higher consistency of the measured data set over the whole 2θ range). From the single-crystal experience it is known that direct methods work well with organic compounds if data sets reach 1.1–1.2 Å resolution. In the case of powder diffraction this resolution limit corresponds to the high-angle portion of the pattern, where peak overlap is most severe and intensity statistics poorer. It is clear that any inconsistency in the data (*e.g.* variation of unit-cell dimensions or radiation damage during data acquisition) and poor intensity statistics will negatively affect the peak-overlap treatment. To overcome these experimental limitations the combination of modern fast-readout solid-state microstrip detectors with synchrotron radiation is decisive. However, accurate data collection is not the only difficulty. Phasing algorithms capable of handling the very severe peak overlap at high-angle regions are also necessary to cope with data from organic compounds. As confirmed by some tests performed on synchrotron powder diffraction data of various organic compounds, the Patterson-
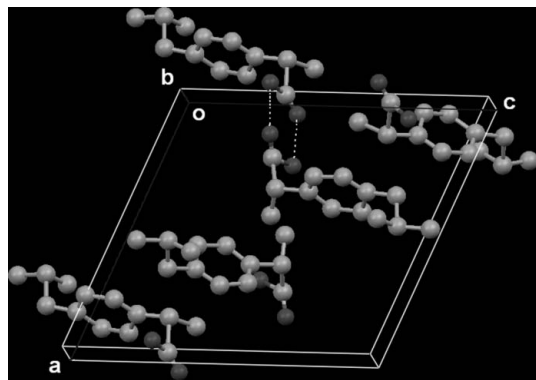
Figure 2
Perspective view along **b** of the monoclinic unit cell of ibuprofen (Freer *et al.*, 1993) with atomic positions taken directly from the direct-methods Fourier map. There are two symmetry-independent molecules forming carboxylic acid dimers. Image created using *Mercury2.2* (Cambridge Crystallographic Data Centre).

function TF constitutes a promising option (Rius, 2010*a*). To illustrate its capability to treat peak overlap, the application to (*S*)-(+)-ibuprofen has been selected (Freer *et al.*, 1993). The unit cell is monoclinic [$a = 12.4627$ (1), $b = 8.0251$ (1), $c = 13.5293$ (2) Å, $\beta = 112.9509$ (8)°, $V = 1246$ Å^3] and contains two symmetry-independent molecules giving rise to a cyclic hydrogen-bonded dimer with formula $C_{26}H_{16}O_4$ (Fig. 2). The powder data, kindly supplied by F. Gozzo, were collected at the PD station of the Materials Science beamline of the Swiss Light Source using the Mythen-II microstrip one-dimensional detector (Schmitt *et al.*, 2004). The intensities were extracted by whole-pattern least-squares refinement using the recursive formula of Rius *et al.* (1996) as implemented in the *DAJUST* program (Rius, 2010*b*). Relevant parameters are: $\lambda = 1.00097$ (3) Å; $d_{\min} = 1.10$ Å; $\sigma(2\theta$ zero-offset) = 0.0001°; Lorentzian peak profiles overlap criterion: $<0.5 \times$ FWHM (full width at half-maximum). Peak breadths (FWHM) are 0.014, 0.023, 0.031 and 0.040° for respective *d*-spacings at >10, 2.5, 1.5 and 1.1 Å (Fig. 3). Extracted multiplet intensities were processed by a new *XLENS* version using the Patterson-
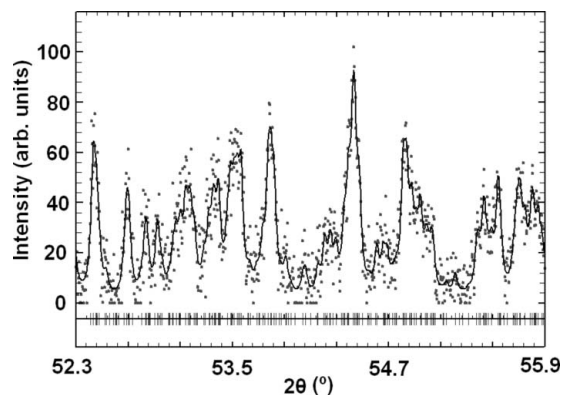
Table 1
Bond lengths (Å) obtained by application of the Patterson-function TF to powder (PD) data (second column) and from least-squares refinement with single-crystal (SC) data (third column).

To simplify the comparison, bonds have been grouped into types. It can be seen that (i) the average C—C bond lengths derived from PD are longer by a factor of approximately 1.035 (this result may be attributed to the non-inclusion of H atoms in DM); (ii) single and double C—O bonds of the carboxyl groups are correctly assigned with similar average lengths, (1.50 + 1.09)/2 = 1.30 Å for PD and (1.32 + 1.21)/2 = 1.27 Å for SC; (iii) distances between hydrogen-bonded O atoms are practically coincident.

Bond type	PD	SC
C—C (single bond) (14×)	1.56 (9)	1.51 (3)
C—C (phenyl ring) (12×)	1.43 (8)	1.38 (2)
C—O (~single bond) (2×)	1.50 (3)	1.32 (1)
C—O (~double bond) (2×)	1.09 (6)	1.21 (1)
O···O (hydrogen bonds) (2×)	2.66 (1)	2.65 (1)

function TF (Rius, 2010*c*). The chemical composition constraint (number of non-H atoms in the unit cell) was applied every second refinement cycle. Relevant parameters for phase refinement are: $B_{\text{overall}} = 5.2$ Å^2; $d_{\min} = 1.10$ Å^2; total number of reflections, 1009; number of trials, 25; maximum number of cycles per trial, 50; interval of the correlation coefficients of the multiplets for true solutions, 0.92–0.89; for incorrect ones, <0.85; success rate, 7 trials out of 25. All correct solutions develop the complete structural model. Details of the model extracted from the Fourier map (Fig. 2) are listed in Table 1.

In conclusion, Patterson-function DM can be successfully applied to multiplet intensities extracted from good quality powder patterns yielding atomic distances that are accurate enough to allow molecular identification in Fourier maps. For organic compounds the restrictiveness of the squaring method (all atoms must be approximately equal) is not a problem; on the contrary, it helps to overcome the information loss owing to peak overlap.

Figure 3
Whole-profile refinement without structural model for intensity extraction. The portion of the ibuprofen powder pattern centred at $d \simeq 1.10$ Å shows peak overlap and intensity statistics. Only the pattern information above $d_{\min} = 1.10$ Å ($<2\theta \simeq 54°$) was processed by *XLENS* (Rius, 2010*c*). Plot created with *WinPLOTR* (Roisnel & Rodriguez-Carvajal, 2001).

References

Altomare, A., Cuocci, C., Giacovazzo, C., Kamel, G. S., Moliterni, A. & Rizzi, R. (2008). *Acta Cryst.* A**64**, 326–336.
Baerlocher, Ch., McCusker, L. B. & Palatinus, L. (2007). *Z. Kristallogr.* **222**, 47–53.
Freer, A. A., Bunyan, J. M., Shankland, N. & Sheen, D. B. (1993). *Acta Cryst.* C**49**, 1378–1380.
Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Math. Res. Bull.* **23**, 447–452.
Rietveld, H. M. (1966). *Acta Cryst.* **20**, 508–513.
Rietveld, H. M. (1967). *Acta Cryst.* **22**, 151–152.
Rius, J. (1993). *Acta Cryst.* A**49**, 406–409.
Rius, J. (2002). *Structure Determination from Powder Diffraction Data*, edited by W. David, K. Shankland, L. McCusker & Ch. Baerlocher, ch. 13, pp. 219–232. IUCr Monographs on Crystallography 13. Oxford Science Publications.
Rius, J. (2010*a*). *S-FFT based direct methods applied to HRPD data of molecular compounds*. Keynote lecture at the 2nd Meeting of the

Italian and Spanish Crystallographic Associations, 30 June – 3 July 2010, Oviedo, Spain.

Rius, J. (2010b). *DAJUST. A Computer Program for Extracting Integrated Intensities from Powder Patterns.* Institut de Ciència de Materials de Barcelona, CSIC, Spain.

Rius, J. (2010c). *XLENS. A Computer Program for Solving Crystal Structures from Diffraction Data by Direct Methods.* CSIC, Spain.

Rius, J., Crespi, A. & Torrelles, X. (2007). *Acta Cryst.* A**63**, 131–134.

Rius, J. & Frontera, C. (2007). *J. Appl. Cryst.* **40**, 1035–1038.

Rius, J., Sañé, J., Miravitlles, C., Amigó, J. M., Reventós, M. M. & Louër, D. (1996). *Anal. Quím. Int. Ed.* **92**, 223–227.

Roisnel, T. & Rodriguez-Carvajal, J. (2001). *Mater. Sci. Forum*, **378–381**, 118–123.

Schmitt, B., Broennimann, Ch., Eikenberry, E. F., Huelsen, G., Toyokawa, H., Horisberger, R., Gozzo, F., Patterson, B., Schulze-Briese, C. & Tomikazi, T. (2004). *Nucl. Instrum. Methods Phys. Res. A*, **518**, 436–439. (Upgraded version developed by B. Schmitt & A. Bergamaschi.)

Will, G. (1979). *J. Appl. Cryst.* **12**, 483–485.