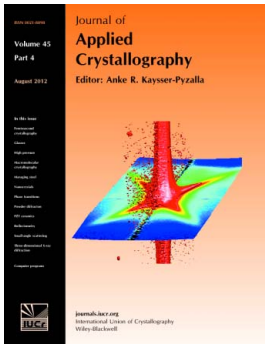
Journal of
Applied Crystallography
ISSN 0021-8898
Editor: **Anke R. Kaysser-Pyzalla**

DAJUST: a suite of computer programs for pattern matching, space-group determination and intensity extraction from powder diffraction data

Oriol Vallcorba, Jordi Rius, Carlos Frontera, Inma Peral and Carles Miravitlles

J. Appl. Cryst. (2012). **45**, 844–848



Journal of
Applied Crystallography
ISSN 0021-8898

Received 17 April 2012
Accepted 9 May 2012

DAJUST: a suite of computer programs for pattern matching, space-group determination and intensity extraction from powder diffraction data

Oriol Vallcorba,[a] Jordi Rius,[a]* Carlos Frontera,[a] Inma Peral[b] and Carles Miravitlles[a]

[a]Institut de Ciència de Materials de Barcelona, CSIC, Campus de la UAB, Bellaterra, Catalunya 08193, Spain, and [b]CELLS ALBA Synchrotron Light Facility, Cerdanyola del Vallès, Barcelona 08290, Spain. Correspondence e-mail: jordi.rius@icmab.es

DAJUST is the generic name of a software package for powder diffraction formed by the core programs *AJUST* and *SGAID* (both written in Fortran) controlled by an easy-to-use Java user interface (*DAJUST_UI*). While *AJUST* performs whole-pattern matching (cell-parameter refinement, profile fitting and intensity extraction), *SGAID* provides a list of the most probable space groups. For the extraction of the integrated intensities, *AJUST* uses the Le Bail procedure but with a different formula for refining the integrated intensities. Laboratory, synchrotron X-ray and neutron sources, and both reflection and transmission experimental geometries, are supported. Other program options include automated background estimation, asymmetry correction, and corrections for absorption, variable divergence and/or illumination. The extracted intensity data are written in text format and can be directly processed by the direct methods program *XLENS* [Rius (2011). *Acta Cryst.* A**67**, 63–67] and the multisolution direct-space structure determination program *TALP* [Vallcorba, Rius, Frontera & Miravitlles (2011). *Acta Cryst.* A**67**, C272].

1. Introduction

Two decisive steps in the solution process of crystal structures from powder diffraction data are the determination of the space group and the retrieval of the maximum amount of information on the structure factor amplitudes. This second part is normally accomplished by a procedure called pattern matching, which takes advantage of the metric (which is assumed to be known) and of the possibility of describing the diffraction peaks by a suitable profile shape function. The pattern-matching procedure delivers a set of extracted integrated intensities which can correspond to resolved reflections, or to a cluster of reflections (multiplets) when peak overlap is severe. In general, peak overlap becomes a serious problem at high 2θ angles, owing to the increasing number of reflections and to the gradual fall-off of intensities. This circumstance limits the useful d-spacing interval for direct methods, *i.e.* the intensities of resolved reflections are mostly restricted to the lower 2θ portion of the pattern. Recently, this limitation has been relaxed through the introduction of Patterson function direct methods (Rius, 2011), which can treat the intensities of overlapping peaks more effectively.

Historically, the first whole-pattern matching procedure that used the metric and the space-group symmetry as constraints was described by Pawley (1981). Basically, it consisted of a least-squares refinement with the intensities, instead of the structural parameters, as independent variables. Later on, Le Bail *et al.* (1988) developed an alternative easy-to-implement two-stage procedure. In the first stage, new estimates of the integrated intensities are obtained by iteratively applying the formula

$$I_{k,\mathrm{new}} = \sum_i \left(y_{\mathrm{obs},i}/y_{\mathrm{calc},i}\right)\Omega_{k,i} I_{k,\mathrm{old}}, \qquad (1)$$

where the summation extends over the i points of the pattern for which the value of the profile function $\Omega_{k,i}$ of the reflection k is not negligible, and where $y_{\mathrm{obs},i}$ and $y_{\mathrm{calc},i}$ are the net observed and calculated counts at $2\theta_i$, respectively. The second stage consists of the refinement of the profile parameters by the Rietveld method, thereby keeping the intensities constant. This procedure is repeated until convergence is reached.

The *DAJUST* software package discussed here consists of the two main programs *AJUST* and *SGAID*, along with the user interface *DAJUST_UI*. The procedure for extracting the intensities implemented in *AJUST* is essentially the two-stage Le Bail procedure but differs in the formula used for refining the intensities. *AJUST* has been used and improved by our research group over the past 20 years and satisfactorily applied to many crystal structure solutions (*e.g.* Guirado *et al.*, 1998). Since it is intended for crystal structure solution, it only handles a single crystalline phase. Should another crystalline phase of known crystal structure be present, *AJUST* can work with the difference pattern supplied by a conventional Rietveld refinement program, *e.g. FULLPROF* (Rodríguez-Carvajal, 1993).

To assist users in space-group determination, a new functionality (the program *SGAID*) has been added to *DAJUST*, following the trend of other programs like *EXPO* (Altomare *et al.*, 2004) or *EXTSYM* (Markvardsen *et al.*, 2008). Also, a simple graphical interface (*DAJUST_UI*) has been created to make both programs friendlier to the final user. The resulting *DAJUST* software generates the input intensity files required by the two structure determination programs developed in our laboratory, namely the Patterson function direct methods code *XLENS* (Rius, 2011) and the multisolution direct-space code *TALP* (Vallcorba *et al.*, 2011). The purpose of this

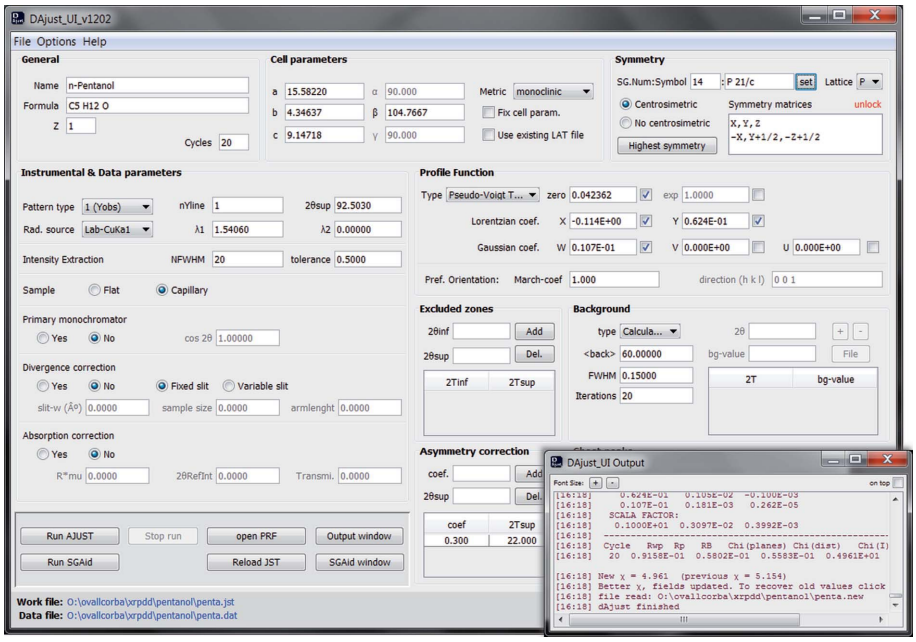


Figure 1
A general view of the *DAJUST* user interface for 1-pentanol data.

contribution is the description of the principal features of the *DAJUST* suite of computer programs.

2. Technical description, availability and requirements

The *DAJUST* software consists of the *AJUST* and *SGAID* programs, both written in Fortran (Lahey–Fujitsu Fortran 7.2), controlled by a Java user interface (*DAJUST_UI*). Currently, only the Windows version of *DAJUST* is available, which can be executed on any PC with the Java Runtime Environment installed (version 1.6.0_10 or higher). The software can be downloaded from the authors' institution web site, http://www.icmab.es/crystallography/software (subject to the conditions of use specified therein), along with a user manual with a step-by-step guide and an example file. A release of the Linux version of the software is planned.

3. Program functionality and features

The main user interface window (Fig. 1) is essentially composed of text boxes, buttons and combo boxes distributed into different panels to facilitate the input of data by users. It creates or modifies the input file for *AJUST* (extension `.jst`) and controls the execution of the *AJUST* and *SGAID* programs. *AJUST* performs the whole-pattern matching (cell-parameter refinement, profile fitting and intensity extraction), while *SGAID* provides a ranked list of the most probable space groups. The only requirements are prior knowledge of the lattice type, the approximate unit-cell parameters and the powder data file with the extension `.dat` (or `.net` if the background has been removed). Two formats are supported: (i) a list of 2θ and y_{obs} values written in two columns; and (ii) a sequence of y_{obs} values. In both cases, the initial 2θ value, the 2θ increment and the final 2θ value must be given in the first line. *SGAID* requires the refined metric (including the lattice type) and the refined profile parameters, so that its execution must follow a preliminary *AJUST* run.

The powder diffraction data file must be in the same folder as the instruction file (`.jst`) generated or modified by the user interface. After running *AJUST*, the following output files are written to this working folder: (i) a summary of the refinement (`.lst`); (ii) an updated instructions file (`.new`), which will be loaded automatically by the user interface after the first run or if the figure of merit for the current refinement is better than the previous one; (iii) a file (`.prf`) for the graphical display of the refinement using *WinPLOTR* (Roisnel & Rodriguez-Carvajal, 2000); and (iv) two files (`.pow` and `.qck`) containing the extracted intensity data, to be used by the programs *XLENS* and *TALP* for crystal structure solution from powder data using direct methods and direct-space methods, respectively. More information about the internal structure of these files is given in §5.

AJUST estimates the background by applying the smoothing procedure described by Brückner (2000). For ill-behaved backgrounds, the background intensities can be calculated by linear interpolation of selected points (which can be visually estimated using *e.g. WinPLOTR*, and stored in a file that is later read by *DAJUST_UI*). *DAJUST* can treat laboratory, synchrotron and neutron data measured in reflection and transmission geometries. In the case of laboratory data measured with $K\alpha_{1,2}$ radiation, the $K\alpha_2$ contribution is eliminated using the Keating (1959) algorithm. Other instrumental parameters that *AJUST* can handle are the absorption correction when the sample is included in a capillary or when the sample is a thin layer with a diffracting substrate (*e.g.* a silver membrane filter), the variable divergence correction, and the illumination correction for flat sample holders (Bragg–Brentano geometry).

An example of application of *DAJUST* to 1-pentanol data is illustrated in Figs. 1–3 ($C_5H_{12}O$, $Z = 8$, $P2_1/c$, $a = 15.592$ (9), $b = 4.349$ (1), $c = 9.157$ (1) Å, $\beta = 104.7$ (7)°; Ramírez-Cardona *et al.*, 2005). In the user's guide for the program, a step-by-step description of the

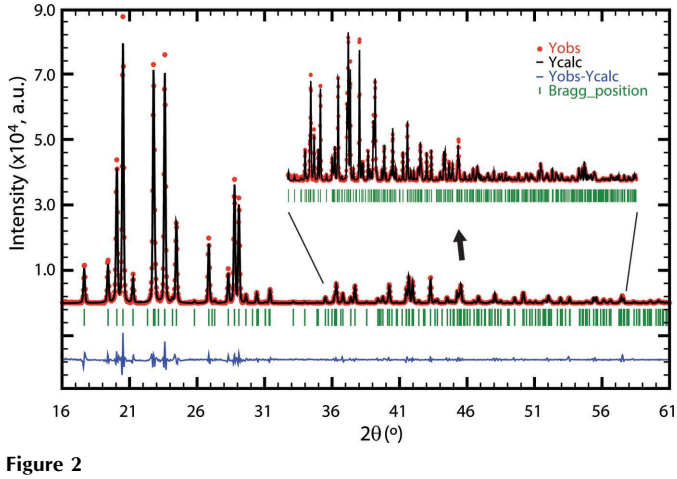


Figure 2
The observed (points) and calculated (line) patterns, along with the difference profile (bottom), for a whole-pattern matching exercise using *DAJUST* (1-pentanol data).

application is given. Running *DAJUST* and *SGAID* in this example takes less than 10 s using an Intel Core i7 (2.80 GHz) CPU.

4. Methodology

4.1. *AJUST*

4.1.1. Iterative intensity refinement. As mentioned before, the two-stage pattern-matching procedure implemented in *AJUST* is similar to that of Le Bail *et al.* (1988) but uses a different recursive formula for the estimation of integrated intensities. If the profile contribution ($\Omega_{k,i}$) of a reflection k at $2\theta_i$ is known approximately, the best values for the integrated intensities are those that minimize the expression

$$M = \sum_i w_i \left[y_{\mathrm{obs},i} - \left(y_{\mathrm{bkg},i} + \sum_k j_k I_k \Omega_{k,i} \right) \right]^2, \tag{2}$$

with the k summation extending over all reflections with non-negligible contributions at point i and where $y_{\mathrm{obs},i}$ is the number of counts at $2\theta_i$, $y_{\mathrm{bkg},i}$ is the background intensity at $2\theta_i$, w_i is a weighting factor inversely proportional to the measured number of counts at $2\theta_i$ and j_k is the multiplicity of reflection k.

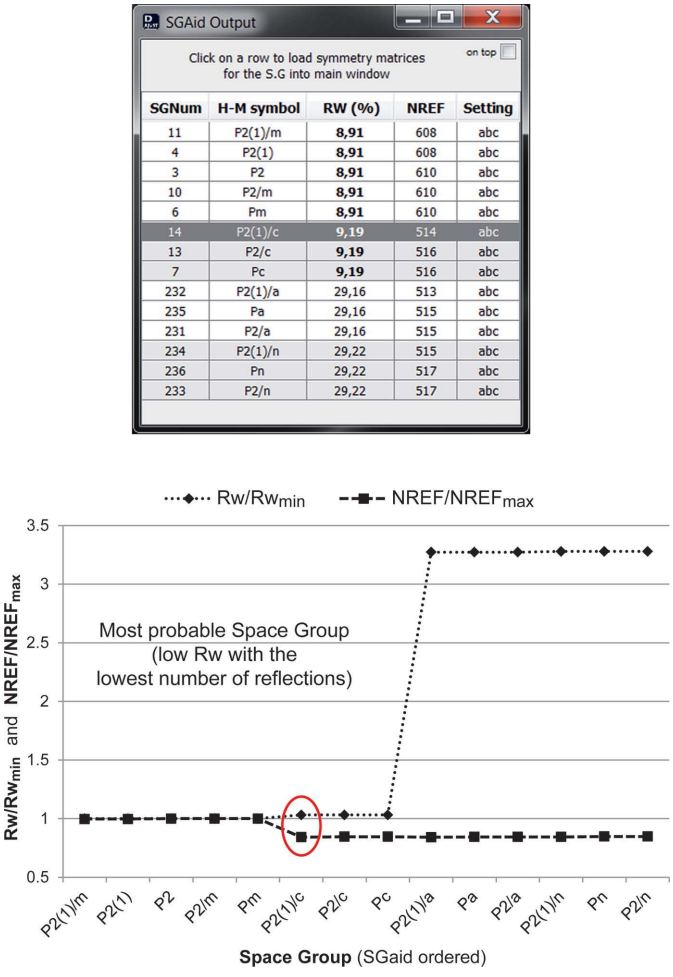


Figure 3
(Top) The *SGAID* output window, showing the table of possible space groups ranked according to increasing R_w values and the number of reflections (NREF). (Bottom) A graphical representation for a clearer interpretation. The circled region corresponds to the most probable space group (1-pentanol data).

As shown by Rius *et al.* (1996), the minimization of M as a function of the intensities is carried out using the recursive formula

$$I_{k,\mathrm{new}} = I_{k,\mathrm{old}} + c_k \sum_i w_i \Omega_{k,i} \left(y_{\mathrm{obs},i} - y_{\mathrm{calc},i} \right), \tag{3}$$

with

$$y_{\mathrm{calc},i} = y_{\mathrm{bkg},i} + \sum_k j_k I_k \Omega_{k,i} \tag{4}$$

and

$$c_k = \frac{1}{\sum_i w_i \Omega_{k,i}^2}, \tag{5}$$

until convergence is achieved. The minimization process starts with $I_{k,\mathrm{old}} = 0$ and the condition $I_{k,\mathrm{new}} > 0$ is applied.

4.1.2. The profile parameters. Once the new intensity estimates have been obtained using the recursive formula of equation (3), the next step is to fit the profile by applying the Rietveld method (with the intensities kept constant). The parameters describing the profile are of two types: (i) the unit-cell parameters and the origin shift, which determine the position of the diffraction peaks; and (ii) the parameters describing the peak shape functions. Two peak shape functions are supported by *AJUST*, as follows.

(*a*) The first peak shape function is a pseudo-Voigt peak shape function with variable mixing factor and variable breadth. It is known that the shape of a measured peak mostly follows a Voigt function, h, which is defined as the convolution of a Gaussian (G) with a Cauchy (C) function,

$$h(\Delta 2\theta; w_g, w_c) = G(\Delta 2\theta; w_g) \otimes C(\Delta 2\theta; w_c), \tag{6}$$

where

$$G(\Delta 2\theta; w_g) = \frac{2}{w_g} \left(\frac{\ln 2}{\pi} \right)^{1/2} \exp\left[\frac{-4 \ln 2}{w_g^2} (\Delta 2\theta)^2 \right], \tag{7}$$

and

$$C(\Delta 2\theta; w_c) = \frac{2}{\pi w_c} \left[1 + \frac{4}{w_c^2} (\Delta 2\theta)^2 \right]^{-1}, \tag{8}$$

with w_g and w_c being the full widths at half-maximum of G and C, respectively. The mathematical expression of this operation is complicated, so that, in practice, it is approximated by a pseudo-Voigt profile

$$h_{\mathrm{PV}}(\Delta 2\theta; w, \eta) = \eta C(\Delta 2\theta; w) + (1 - \eta) G(\Delta 2\theta; w), \tag{9}$$

where η is a mixing factor controlling the relative weights of the Cauchy (C) and Gaussian (G) contributions, and w is the full width at half-maximum of both C and G. Obviously, since w_g and w_c vary with 2θ, both w and η are also variable. *AJUST* uses the mathematical equations derived by Thompson *et al.* (1987) to express the (w, η) pairs in terms of (w_g, w_c) ones, so that h can be written in the form

$$h(\Delta 2\theta; w_g, w_c) \cong \eta(w_g, w_c) C[\Delta 2\theta; w(w_g, w_c)] + [1 - \eta(w_g, w_c)] G[\Delta 2\theta; w(w_g, w_c)]. \tag{10}$$

The advantage of expressing the pseudo-Voigt profile as a function of (w_g, w_c) instead of (w, η) is that the former have an immediate physical meaning. In *DAJUST*, the dependencies of the refined physical parameters Y, X, W, V and U on w_c and w_g are given by

$$w_c = Y/\cos\theta + X \tan\theta, \tag{11}$$

and

$$w_g^2 = U \tan^2\theta + V \tan\theta + W. \tag{12}$$

The physical meaning of these parameters can be found in textbooks (*e.g.* Dinnebier & Billinge, 2008).

(*b*) The second peak shape function is the Pearson-VII function, which can be used as an alternative to the pseudo-Voigt when the latter fails to describe the peak shape accurately (Hall *et al.*, 1977). It is calculated using

$$P_{\rm VII}(\Delta 2\theta) = I_0 \frac{0.76642}{w}\left[1 + 4\left(2^{1/m} - 1\right)\left(\frac{\Delta 2\theta}{w}\right)^2\right]^{-m}, \qquad (13)$$

where $w = w_1 \tan^2\theta + w_2 \tan\theta + w_3$ is the full width at half-maximum of the peak, expressed in terms of θ and of the coefficients w_1, w_2 and w_3 in the polynomial expression (Cagliotti *et al.*, 1958), m is the flank exponent, and 0.76642 is a normalizing constant. *AJUST* handles the peak asymmetry using the expression proposed by Allmann (2003), which modifies the contribution of peak k at point $2\theta_i$ according to

$$y_{i,\rm corr} = y_i\left\{1 - \frac{P(2\theta_i - 2\theta_k)^3}{\left[(w/2)^2 + (2\theta_i - 2\theta_k)^2\right]^{1.5}}\right\}. \qquad (14)$$

This asymmetry correction is simple to apply and always yields positive y values for $|P| \leq 1$. The asymmetry parameter P can be made 2θ dependent.

4.1.3. Criterion of fit. The two-stage refinement procedure continues until no significant differences between the old and new intensity estimates are observed. The square root of the goodness of fit

$$\chi = R_{\rm wp}/R_{\rm exp}, \qquad (15)$$

is calculated after each *DAJUST* cycle. $R_{\rm wp}$ and $R_{\rm exp}$ are defined by

$$R_{\rm wp} = \left[\frac{\sum_i w_i\left(y_{{\rm obs},i} - y_{{\rm calc},i}\right)^2}{\sum_i w_i\left(y_{{\rm obs},i} - y_{{\rm bkg},i}\right)^2}\right]^{1/2} \qquad (16)$$

and

$$R_{\rm exp} = \left[\frac{N - P}{\sum_i w_i\left(y_{{\rm obs},i} - y_{{\rm bkg},i}\right)^2}\right]^{1/2}, \qquad (17)$$

where

$$w_i = \frac{1}{\sum_i\left(y_{{\rm obs},i} - y_{{\rm bkg},i}\right)}, \qquad (18)$$

N and P are the number of measured intensity points and the number of refined parameters, respectively, and $y_{\rm bkg}$ is the intensity (counts) of the background at $2\theta_i$.

4.2. *SGAID*

Before executing *SGAID*, a previous run of *AJUST* is necessary, entering the highest symmetry compatible with the metric and the lattice type. This first run serves to refine the profile parameters and to confirm the entered lattice type for the subsequent *SGAID* run. *SGAID* systematically checks all space groups that are compatible with the metric and the lattice type. The output consists of the space groups ranked according to increasing residuals ($R_{\rm w}$) and, for each one, the number of (allowed) reflections involved in the $R_{\rm w}$ calculation (Fig. 3). Normally, the true space group combines a low $R_{\rm w}$ (highlighted in bold in the table) with the lowest number of reflections. In case of ambiguity, the space-group symmetry can easily be introduced in *DAJUST_UI* to generate a `.prf` file for visual inspection (*e.g.* using *WinPLOTR*). Once the true space group has been established and entered into *DAJUST_UI*, the subsequent *AJUST* execution will supply the files with the extracted integrated intensities.

5. Data files with extracted intensities

The profile intensity at a certain point of the powder pattern results from the sum of the contributions of the neighbouring peaks. A neighbouring reflection should be considered a possible contributor to this point if the absolute angular distance between point and reflection is less than approximately 25w for Cauchy peaks or 4w for Gaussian peaks. Besides this contribution range, another important parameter for the intensity extraction is the overlap criterion. Two reflections are considered as overlapping if, because of their proximity, it is not possible to estimate their individual intensities. In *AJUST*, the criterion for considering two arbitrary reflections, *e.g.* h and k, as overlapping is $|2\theta_h - 2\theta_k| <$ tol w. The default tolerance parameter tol is 0.5 (Rius *et al.*, 1996). *AJUST* creates two different files (`.pow` and `.qck`) with the extracted intensities. These files differ not only in their format but also in the calculated quantities.

5.1. File .pow

The `.pow` file is an ASCII file with one line per reflection (h) containing the following:

(i) the Miller indices of h;

(ii) the square root of the extracted integrated intensity including the multiplicity factor ($D_h^{1/2}$);

(iii) the multiplicity (j_h);

(iv) the total number of reflections (NVEI) forming the multiplet to which h belongs (NVEI = 1 for a resolved reflection);

(v) an indicator variable called NGRUP, which is 1 to indicate that h begins a new multiplet or 0 if h still belongs to the same multiplet as the previous reflection in the file;

(vi) the calculated $2\theta_h$ in degrees;

(vii) the d spacing of h in ångström;

(viii) the full width at half-maximum (w) of the h peak in degrees.

In the case of resolved reflections, it holds that

$$\left(D_h^{1/2}\right)^2 = K j_h |F_h|^2, \qquad (19)$$

where $|F_h|$ is the structure factor modulus of the h reflection and K is a suitable scaling constant. However, for multiplets, if k refers to all the reflections in one multiplet, the global intensity of this multiplet is given by

$$\sum_k \left(D_k^{1/2}\right)^2 = K \sum_k j_k |F_k|^2. \qquad (20)$$

The file `.pow` contains the intensity information necessary for the Patterson-function direct methods program *XLENS* (Rius, 2011).

5.2. File .qck

The `.qck` file is an ASCII file which can be best described by first considering only the resolved reflections of the pattern, since in this case the `.qck` file only contains a single line for each h reflection. Each line begins with a number (code) assigned to each h (going from 1 to the total number of reflections), followed by the Miller indices of h, the calculated $2\theta_h$ value (in degrees), the NREST parameter, which indicates the number of neighbouring reflections contributing to h (for resolved reflections it is obviously 0), the observed profile intensity $y_{\rm obs}$ at $2\theta_h$, the integrated area D_h, the multiplicity j_h, the value of the profile function at the reflection centre Ω_h, the full width at half-maximum w_h (in degrees), the Lorentz fraction η_h in the

pseudo-Voigt function, the Pearson-VII exponent m (-5.0 indicates that the pseudo-Voigt function is used), and, finally, at $2\theta_h$, the intensity without background $y_{obs} - y_{bkg}$, the background intensity y_{bkg}, the Lorentz polarization factor and the asymmetry coefficient. Once this information is stored, the partial pattern calculated with the contributions of the resolved reflections is removed from the observed pattern.

We now consider the difference pattern, so the first line for each non-resolved reflection has the same structure as the single line for resolved reflections. However, NREST, the number of neighbouring reflections contributing to h, is here different from zero. The first line is then followed by NREST lines, each containing the code for h, the code for the neighbouring reflection k, the calculated 2θ value of k (in degrees) and the value of the profile contribution of k at $2\theta_h$.

The `.qck` file stores the complete information of the powder diffraction pattern in a very efficient way and is the input file for the direct-space structure determination program *TALP* (Vallcorba *et al.*, 2011).

OV thanks the CSIC for a contract (No. PIE 201060E068). IP is supported by the Ramon y Cajal programme from the Spanish Ministerio de Ciencia e Innovación. The authors also thank the Spanish Ministerio de Ciencia e Innovación Tecnológica (project No. MAT2009-07967 and Consolider NANOSELECT No. CSD2007-00041) and the Generalitat de Catalunya for financial support.

References

Allmann, R. (2003). *Röntgenpulverdiffraktometrie*, p. 157. Heidelberg: Springer Verlag.
Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., da Silva, I., Giacovazzo, C., Moliterni, A. G. G. & Spagna, R. (2004). *J. Appl. Cryst.* **37**, 957–966.
Brückner, S. (2000). *J. Appl. Cryst.* **33**, 977–979.
Cagliotti, G., Paoletti, A. & Ricci, F. (1958). *Nucl. Instrum.* **3**, 223–228.
Dinnebier, R. E. & Billinge, S. J. L. (2008). *Powder Diffraction: Theory and Practice*. Cambridge: Royal Society of Chemistry.
Guirado, F., Galí, S., Chinchón, S. & Rius, J. (1998). *Angew. Chem. Int. Ed.* **37**, 72–75.
Hall, M. M., Veeraraghavan, V. G., Rubin, H. & Winchell, P. G. (1977). *J. Appl. Cryst.* **10**, 66–68.
Keating, D. T. (1959). *Rev. Sci. Instrum.* **30**, 725–727.
Le Bail, A., Duroy, H. & Fourquet, J. (1988). *Mater. Res. Bull.* **23**, 447–452.
Markvardsen, A. J., Shankland, K., David, W. I. F., Johnston, J. C., Ibberson, R. M., Tucker, M., Nowell, H. & Griffin, T. (2008). *J. Appl. Cryst.* **41**, 1177–1181.
Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.
Ramírez-Cardona, M., Ventolà, L., Calvet, T., Cuevas-Diarte, M. A., Rius, J., Amigó, J. M. & Reventós, M. M. (2005). *Powder Diffr.* **20**, 311–315.
Rius, J. (2011). *Acta Cryst.* A**67**, 63–67.
Rius, J., Sañé, J., Miravitlles, C., Amigó, J. M., Reventós, M. M. & Louër, D. (1996). *An. Quim. Int.* **92**, 223–227.
Rodríguez-Carvajal, J. (1993). *Physica B*, **192**, 55–69.
Roisnel, T. & Rodriguez-Carvajal, J. (2000). *Mater. Sci. Forum*, **378–381**, 118–123.
Thompson, P., Cox, D. E. & Hastings, J. B. (1987). *J. Appl. Cryst.* **20**, 79–83.
Vallcorba, O., Rius, J., Frontera, C. & Miravitlles, C. (2011). *Acta Cryst.* A**67**, C272.