Application of simulated annealing approach for structure solution of molecular crystals from X-ray laboratory powder data

S. G. Zhukov[*, I], V. V. Chernyshev[I], E. V. Babaev[I], E. J. Sonneveld[II] and H. Schenk[II]

[I] Moscow State University, Department of Chemistry, 119899, Moscow, Russia
[II] University of Amsterdam, Laboratory of Crystallography, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands

Abstract. Simulated annealing approach was successfully applied to solve three unknown molecular structures from X-ray laboratory powder data using *a priory* known structural fragments. Some possible developments of the method are discussed.

Introduction

For crystal structure determination from powder diffraction data one can use a number of methods that have been discussed in details in recent reviews (Giacovazzo, 1996; Harris & Tremayne, 1996; Le Bail, 1998; Louër, 1999). The most known of them, implemented in available programs, are the direct methods (Jansen, Peschar, Schenk, 1992a;1992b; Altomare, Cascarano, Giacovazzo, Guagliardi, Burla, Polidori, Camalli, 1994), the maximum entropy and likelihood-ranking method (Bricogne & Gilmore, 1990), the direct grid search (Masciocchi, Bianchi, Cairati, Mezza, Pulati, Sironi, 1994; Chernyshev & Schenk, 1998) and the global optimization procedures: genetic algorithm (Kariuki, Serrano-Gonzalez, Johnston, Harris, 1997) and Monte Carlo based methods (Harris, Tremayne, Lightfoot, Bruce, 1994; Andreev, Lightfoot, Bruce, 1997). In recent years the Monte Carlo simulated annealing approach has been applied in a large variety of structural studies (Deem, Newsam, 1992; Tremayne, Kariuki, Harris, 1997; Andreev & Bruce, 1998; David, Shankland, Shankland, 1998) and the software package has been developed (Engel, Wilke, König, Harris, Leusen, 1999). Recently the new optimization method, based on the combined cost function including the potential energy and the difference between the calculated and measured diffraction patterns has been suggested and successfully tested on a number of ionic and intermetallic compounds (Putz, Schön, Jansen, 1999).

However, the structure solution from powder data is not a straightforward and routine task yet. Therefore, till now there is still a strong need for elaboration of efficient solving procedures.

The aim of this work was the development of the simulated annealing approach. It was successfully applied to the study of crystal structures from laboratory Guinier data using *a priory* known molecule fragments. Three unknown molecular structures of 2-chloro-1-(p-nitrophenacyl)-pyridinium bromide $(C_{13}H_{10}N_2O_3Cl)^+Br^-$ (compound I); 2-(p-nitrophenyl)-thiazolo[3,2-a]-pyridinium perchlorate $(C_{13}H_9N_2O_2S)^+ClO_4^-$ (compound II) and 3-(p-nitrobenzoyl)2-oxo-oxazolo[3,2-a]pyridine $C_{14}H_8N_2O_5$ (compound III) have been solved.

Method and structure solution

The simulated annealing (*SA*) procedure was based on the standard Metropolis Monte Carlo method (Metropolis, Rosenbluth, Rosenbluth, Teller, Teller, 1953). For structure solution one can use the figure-of-merit (FOM) function, such as *R*-factors, calculated over the whole powder pattern (Harris *et al.*, 1994):

$$R_{wp} = \left[\sum w_i |Y_{obs,i} - Y_{calc,j}|^2 / \sum w_i |Y_{obs,j}|^2 \right]^{1/2}$$

$$R_B = \sum |Y_{obs,i} - Y_{calc,j}| / \sum |Y_{obs,j}|$$

where $Y_{obs,i}$ – the experimental intensity of *i*-th point; $Y_{calc,i}$ – the calculated intensity; $w_i = 1/Y_{obs}$. However, this method is time consuming and, instead of it, in (David *et al.*, 1998) the FOM function was suggested, which is based on the extracted integrated intensities from the profile refinement using the Pawley method (Pawley, 1981):

$$\chi^2 = \sum_h \sum_k (I_h - c|F_h|^2) (V^{-1})_{hk} (I_k - c|F_k|^2)$$

where I_h, I_k – the Lorentz-polarization corrected extracted intensities; V_{hk} – the covariance matrix from the profile refinement; c – the scale factor; $|F_h|$, $|F_k|$ – the calculated structure factors of the trial structure. The using of the covariance matrix permits to handle the correlation between the extracted intensities of the overlapping reflections. However, in the case of the moderate overlapping it is possible to simplify the FOM function for faster computation. Following (Chernyshev & Schenk, 1998) we used the FOM function *R(X)*:

$$R(X) = \sum |X_{obs,j} - c \cdot X_{calc,j}| / \sum X_{obs,j}$$

* Correspondence author (e-mail: zhukov@struct.chem.msu.ru)

Compound	I	II	III
Formula	$(C_{13}H_{10}N_2O_3Cl)^+ \cdot Br^-$	$(C_{13}H_9N_2O_2S)^+ \cdot ClO_4$	$C_{14}H_8N_2O_5$
Sp.gr.	*Fdd*2	*Pbca*	*P*2$_1$/*n*
a, Å	30.77(1)	16.65(1)	13.64(1)
b, Å	35.04(1)	24.09(1)	22.28(1)
c, Å	5.130(3)	7.518(3)	3.917(3)
β, deg.			90.63(2)
V, Å^3	5531	3016	1190
Z	16	8	4
Absorption coef. μ (cm^{-1})	59.2	38.6	10.8
Calculated density d_x (g cm^{-1})	1.72	1.57	1.65
F(000)	2848	1456	608

Table 1. Crystal data of 2-chloro-1-(p-nitro-phenacyl)-pyridinium bromide (I); 2-(p-nitro-phenyl)-thiazolo[3,2-a]-pyridinium perchlorate (II); 3-(p-nitrobenzoyl)2-oxo-oxazolo[3,2-a]pyridine (III).

where *j*-summation is over the number of the extracted intensities, including the single and overlapping reflections; $X_{obs} = \sum m_i F^2_{obs,i}$; $X_{calc} = \sum y_i m_i F^2_{calc,i}$ with *i*-summation over the overlapping reflections; m_i – the multiplicity of the *i*-th reflection; y_i – the texture correction coefficient; *c* – the scale factor. The values of X_{obs} could be easily obtained using the Pawley or Le Bail profile refinement (Pawley, 1981; Le Bail, Duroy, Fourquet, 1988). Thus, the problem of the extracting of the intensity of each overlapping reflection is avoided and the peak overlapping is intrinsically taken care of. This approach permits a large reduction of computing time in comparison with the use of the whole pattern profile.

During the *SA* run several rigid molecular fragments were displaced independently with the random translational and rotational shifts $\{X_iY_iZ_i\psi_i\varphi_i\chi_i\} \rightarrow \{X_iY_iZ_i\psi_i\varphi_i\chi_i\} + \alpha_{R;\Phi}\xi_i$, where ξ_i – a random number between –1 and 1; $\alpha_{R;\Phi}$ – the maximum possible shift of the translation or rotation. For each random configuration the FOM function was calculated after a rapid scale factor minimization. Following (Metropolis, et al., 1953) the trial structure was accepted with the probability $P = \exp\{-(R_{trial} - R_{prev})/T\}$ and rejected with the probability $(1 - P)$, where R_{trial}, R_{prev} – the values of the FOM function of the trial and previously accepted configurations, *T* – the analogy of the slowly decreasing temperature.

The choice of the appropriate value of α is important for the uniform exploring of the phase space. In (Serra, Stanton, Kais, Bleil, 1997) it was shown that the different types of the random shift distribution leads to a different exploration of the phase space. Therefore we used $\alpha(T)$ with normal distribution $N(\alpha, \sigma)$, where the standard de-

viation σ decreases with decreasing temperature. As a result large shifts were permitted at high temperatures and at low temperatures the variation of α was negligible.

The studied powders were synthesised following the standard methods (Babaev, 2000). The X-ray powder patterns were measured using an evacuated Guinier-Johanson camera (Enraf-Nonius FR552) with quartz monochromatized CuKα_1 radiation. Digitised intensities were collected with a LS18 microdensitometer from 4.0 to 83.0 degrees 2θ in steps of 0.01°. Cell parameters of the three compounds were determined by indexing of the powder patterns with the use of the ITO program (Visser, 1969) and space groups were derived from systematic extinctions (see Table 1). The Pawley profile refinement, *SA* procedure and Rietveld refinement were performed with the updated MRIA program (Zlokazov & Chernyshev, 1992). The split-type pseudo-Voight peak profile function (Toraya, 1986) and March-Dollase texture formalism (Dollase, 1986) were used for powder pattern processing.

For the *SA* minimization we constructed the initial molecular models from the chemically stable and rigid fragments. Preliminary information about the possible molecular structure was obtained using the IR and NMR spectroscopy measurements. In the case of the compound I the p-nitro-benzoyl and the (2-Cl-pyridine-1)methyl groups with the Br atom have been taken to generate the random configurations; for the compound II we used the p-nitro-phenyl and the thiazolo[3,2-a]pyridyl groups and the ClO$_4$ anion, and for compound III two independent units: the 2-oxo-oxazolo[3,2-a]pyridine and the p-nitro-benzoyl groups have been used. The molecular fragments are shown in Fig. 1.
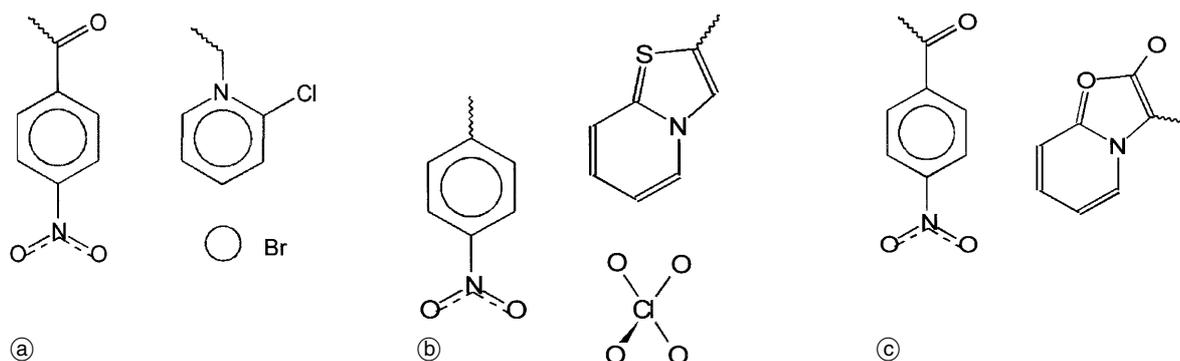
Fig. 1. Molecular fragments used for simulated annealing minimization. **a)** 2-chloro-1-(p-nitrophenacyl)-pyridinium bromide (compound I); **b)** 2-(p-nitrophenyl)-thiazolo[3,2-a]-pyridinium perchlorate (compound II); **c)** 3-(p-nitrobenzoyl)2-oxo-oxazolo[3,2-a]pyridine (compound III).
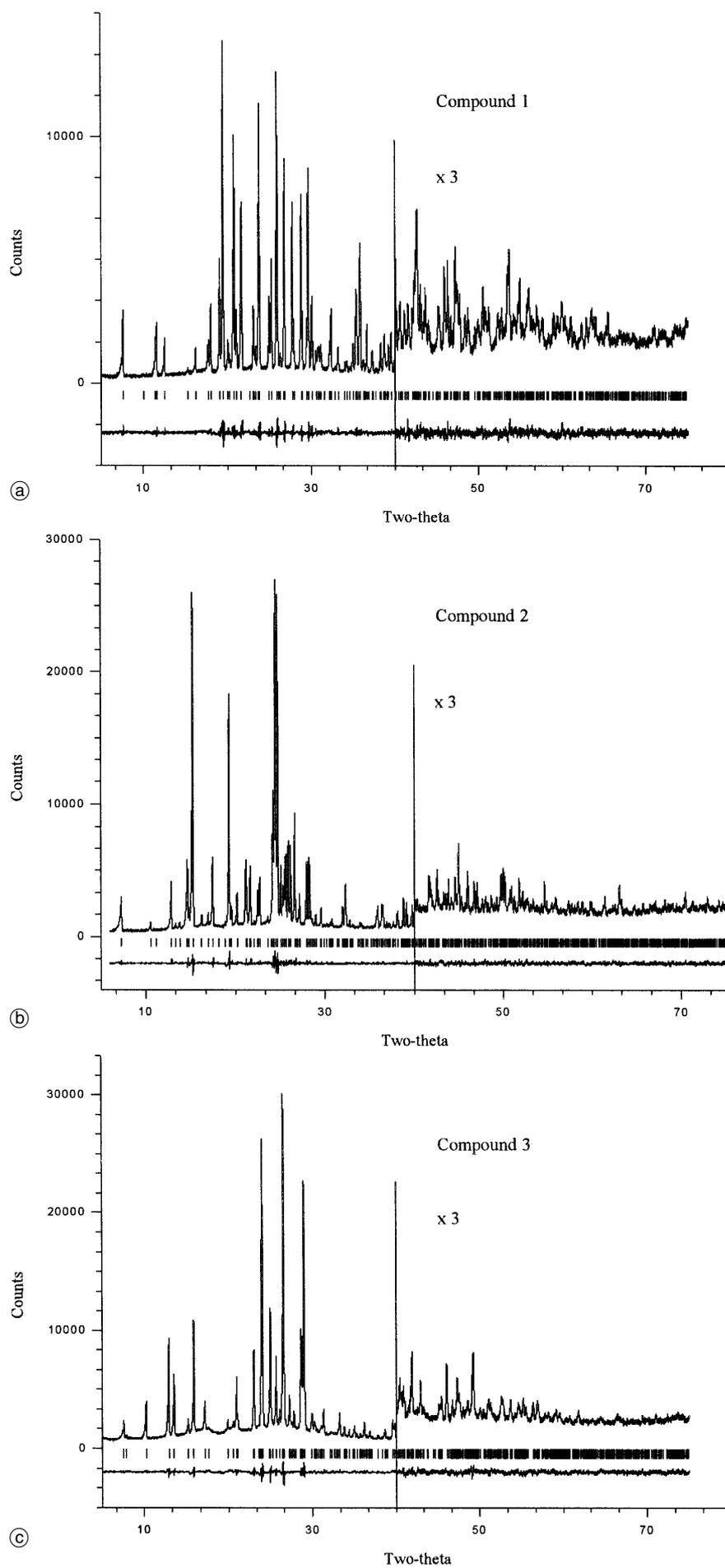
Fig. 2. Rietveld plots of **a**) 2-chloro-1-(p-nitro-phenacyl)-pyridinium bromide (compound I); **b**) 2-(p-nitrophenyl)-thiazolo[3,2-a]-pyridinium perchlorate (compound II); **c**) 3-(p-nitroben-zoyl)2-oxo-oxazolo[3,2-a]pyridine (compound III).

Table 2. Simulated annealing (*SA*) data of structure solution of: 2-chloro-1-(p-nitrophenacyl)-pyridinium bromide (I); 2-(p-nitrophenyl)-thiazolo[3,2-a]-pyridinium perchlorate (II); 3-(p-nitrobenzoyl)2-oxo-oxazolo[3,2-a]pyridine (III).

Com-pound	N of X_{obs}	N of non-hydrogen atoms	N of degrees of freedom	N of SA runs	$R(X)$ (%) of solution
I	70–100	20	15	11	48
II	100–120	23	18	16	39
III	70–100	21	12	7	52

Table 3. The results of full-pattern decomposition and Rietveld refinements of: 2-chloro-1-(p-nitrophenacyl)-pyridinium bromide (I); 2-(p-nitrophenyl)-thiazolo[3,2-a]-pyridinium perchlorate (II); 3-(p-nitrobenzoyl)2-oxo-oxazolo[3,2-a]pyridine (III).

	Compound I	Compound II	Compound III
Full pattern decomposition[a]			
R_{wp}, %	7.6	7.3	5.6
R_{exp}, %	3.2	2.6	2.5
χ^2	5.6	7.9	4.8
Rietveld refinement			
R_{wp}, %	8.0	8.1	5.7
R_{exp}, %	3.2	2.6	2.5
χ^2	6.1	9.9	5.3
Texture correction,			
direction	001	010	001
parameter r	0.78	0.91	0.83

a: R_{wp}, R_{exp} are defined in (Young, Wiles, 1982)

The number of *SA* runs to find the actual structure solution varied from 7 to 16 depending on the degrees of freedom. At each temperature the total number of possible random steps was equal $N_{tot} = 20000 \cdot N_{units}$. However, the temperature was decreased compulsorily if the number of successful trials has exceeded $0.3 \cdot N_{tot}$. Usually 70–120 X_{obs} extracted intensities in total were used for the *SA* calculation. Each *SA* run took one or several hours on a Pentium-III computer depending on the number of X_{obs} and the degrees of freedom. The details of *SA* procedures for the studied compounds are summarised in Table 2.

Special attention was paid to the preferred orientation effect. Because it is difficult to avoid it completely during the experiments we used several data sets for structure solution with March-Dollase texture correction applied along several probable directions and different texture coefficients.

The results of the final Rietveld refinements of the crystal structures are given in Table 3. The details of the crystal structures are to be published in a separate papers. In Fig. 2a, b, c the Rietveld plots for the compounds 1, 2 and 3 are shown.

Concluding remarks

The simulated annealing method is efficient for the crystal structure determination from X-ray powder data using *a priory* known structural fragments. Even without the precise knowledge of the possible molecular structure it gives the correct positions of the molecular groups. A useful addition could be made if the structural fragments are al-lowed to be partly flexible including a certain degree of flexibility in bond length, bond angles and torsion angles (see Andreev, et al., 1997; Andreev & Bruce, 1998). However, the increasing of the number of degrees of freedom may lead to good values of the FOM function but not to the correct structure, hence, the number of simulated annealing steps and runs should be increased to find the structure. It seems to be worthwhile to accompany the simulated annealing run with an intermediate Rietveld refinement to exclude false minima at low "temperatures" (see Engel *et al.*, 1999).

One of the crucial points is the choice of the simulated annealing scheme, *i.e.* the random shift distribution and the type of the acceptance probability. It was shown (Serra, et al., 1997) that the exploring of phase space essentially depends on the distribution type of the random shift vector. The use of the suggested Tsallis distribution may improve the exploration of the phase space during the random walk. Another point is the acceptance criterion, which is usually taken from the Metropolis algorithm, but could also be based on another probability type for speeding up the minimization process. This will be studied in our future work.

The systematic errors, for instance the noticeable texture, may lead to the failure of the structure determination. If texture free measurements are impossible then the experimental data could be corrected using a trial and error method or by applying the special procedures (Peschar, Schenk, Čapkova, 1995).

References

Altomare, A.; Cascarano, G.; Giacovazzo, C.; Guagliardi, A.; Burla, M. C.; Polidori, G.; Camalli, M.: SIRPOW92- a program for automatic solution of crystal structures by direct methods optimized for powder data. J. Appl. Crystallogr. **27** (1994) 435–436.

Andreev, Y. G.; Lightfoot, P.; Bruce, P. G.: A general Monte Carlo approach to structure solution from powder-diffraction data: application to poly(ethylene oxide)$_3$:LiN(SO$_2$CF$_3$)$_2$. J. Appl. Crystallogr. **30** (1997) 294–305.

Andreev, Y. G.; Bruce, P. G.: Solving crystal structures of molecular solids without single crystals: a simulated annealing approach. J. Chem. Soc., Dalton Trans. (1998) 4071–4080.

Babaev, E. V.: Fused munchnones in recyclization tandems. J. Heterocyclic Chem. **37** (2000) 1–8.

Bricogne, G.; Gilmore, J.: A multisolution method of phase determination by combined maximization of entropy and likelihood. I. Theory, algorithms and strategy. Acta Crystallogr. A**46** (1990) 284–297.

Chernyshev, V. V.; Schenk, H.: A grid search procedure of positioning a known molecule in an unknown crystal structure with the use of powder diffraction data. Z. Kristallogr. **213** (1998) 1–3.

David, W. I. F.; Shankland, K.; Shankland, N.: Routine determination of molecular crystal structures from powder diffraction data. J. Chem. Soc., Chem. Commun. **8** (1998) 931–932.

Deem, M. W.; Newsam, J. M.: Framework crystal structure solution by simulated annealing: test application to known zeolite structures. J. Am. Chem. Soc. **114** (1992) 7198–7207.

Dollase, W. A.: Correction of intensities for preferred orientation in powder diffractometry: application of the March model. J. Appl. Crystallogr. **19** (1986) 267–272.

Engel, G. E.; Wilke, S.; König, O.; Harris, K. D. M.; Leusen, F. J. J.: PowderSolve – a complete package for crystal solution from powder diffraction patterns. J. Appl. Crystallogr. **32** (1999) 1169–1179.

Giacovazzo, C.: Direct methods and powder data: state of art and perspectives. Acta Crystallogr. A**52** (1996) 331–339.

Harris, K. D. M.; Tremayne, M.; Lightfoot, P.; Bruce, P. G.: Crystal structure determination from powder diffraction data by Monte Carlo methods. J. Am. Chem. Soc. **116** (1994) 3543–3547.

Harris, K. D. M.; Tremayne, M.: Crystal structure determination from powder diffraction data. Chem. Mater. **8** (1996) 2554–2570.

Jansen, J.; Peschar, R.; Schenk, H.: On the determination of accurate intensities from powder diffraction data. I. Whole-pattern fitting with a least-squares procedure. J. Appl. Crystallogr. **25** (1992a) 231–236.

Jansen, J.; Peschar, R.; Schenk, H.: On the determination of accurate intensities from powder diffraction data. II. Estimation of intensities of overlapping reflections. J. Appl. Crystallogr. **25** (1992b) 237–243.

Kariuki, B. M.; Serrano-Gonzalez, H.; Johnston, R. L.; Harris, K. D.M.: The application of genetic algorithm for solving crystal structures from powder diffraction data. Chem. Phys. Lett. **280** (1997) 189–195.

Le Bail, A.; Duroy, H.; Fourquet, J. L.: *Ab-initio* structure determination of LiSbWO$_6$ by XRay powder diffraction. Mater. Res. Bull. **23** (1988) 447–452

Le Bail, A.: Structure determination from powder diffraction – database. http://www.cristal.org/iniref.html. (1998).

Louër, D.: Advances in Powder Diffraction Analysis. In: *Crystallography Across the Sciences*. (Ed. H. Schenk.) p. 922–933. International Union of Crystallogrphy. 1998.

Masciocchi, N.; Bianchi, R.; Cairati, P.; Mezza, G.; Pulati, T.; Sironi, A.: P-RISCON: a real scavenger for crystal structure determination from powder diffraction data. J. Appl. Crystallogr. **27** (1994) 426–429.

Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E.: Equation of state calculations by fast computing machines. J. Chem. Phys. **21** (1953) 1087–1092.

Pawley, G. S.: Unit-cell refinement from powder diffraction scans. J. Appl. Crystallogr. **14** (1981) 357–361.

Peschar, R.; Schenk, H.; Čapkova, P.: Preferred-orientation correction and normalization procedure for ab-initio structure determination from powder data. J. Appl. Crystallogr. **28** (1995) 127–140.

Putz, H.; Schön, J. C.; Jansen, M.: Combined method for *ab initio* structure solution from powder diffraction data. J. Appl. Crystallogr. **32** (1999) 864–870.

Serra, P.; Stanton, A. F.; Kais, S.; Bleil, R. E.: Comparison study of pivot methods for global optimization. J. Chem. Phys. **106** (1997) 7170–7177.

Toraya, H.: Whole-powder-pattern fitting without reference to a structural model: application to X-ray powder diffractometer data. J. Appl. Crystallogr. **19** (1986) 440–447.

Tremayne, M.; Kariuki, B. M.; Harris, K. D. M.: Structure determination of a complex organic solid from X-ray powder diffraction data by a generalized Monte Carlo method: the crystal structure of red fluorescein. Angew. Chem. Int. Ed. Engl. **36** (1997) 770–772.

Visser, J. W.: A fully automatic program for finding the unit cell from powder data. J. Appl. Crystallogr. **2** (1969) 89–95.

Young, R. A.; Wiles, D. B.: Profile shape functions in Rietveld refinement. J. Appl. Crystallogr. **15** (1982) 430–438.

Zlokazov, V. B.; Chernyshev, V. V.: MRIA- a program for a full profile analysis of powder multiphase neutron-diffraction time-of-flight (direct and Fourier) spectra. J. Appl. Crystallogr. **25** (1992) 447–451.