COMPUTER PROGRAMS

J. Appl. Cryst. (1999). **32**, 1169–1179

PowderSolve – a complete package for crystal structure solution from powder diffraction patterns

G. E. Engel,[a]* S. Wilke,[a] O. König,[a]† K. D. M. Harris[b] and F. J. J. Leusen[a] at [a]*Molecular Simulations Ltd, 230/250 The Quorum, Barnwell Road, Cambridge CB5 8RE, England, and* [b]*School of Chemistry, The University of Birmingham, Edgbaston, Birmingham B15 2TT, England. E-mail: gengel@msicam.co.uk*

(*Received 18 March 99; accepted 15 July 99*)

Abstract

Powder diffraction techniques are becoming increasingly popular as tools for the determination of crystal structures. The authors of this paper have developed a software package, named *PowderSolve*, to solve crystal structures from experimental powder diffraction patterns and have applied this package to solve the crystal structures of organic compounds with up to 18 variable degrees of freedom (defined in terms of the positions, orientations, and internal torsions of the molecular fragments in the asymmetric unit). The package employs a combination of simulated annealing and rigid-body Rietveld refinement techniques to maximize the agreement between calculated and experimental powder diffraction patterns. The agreement is measured by a full-profile comparison (using the *R* factor R_{wp}). As an additional check at the end of the structure solution process, accurate force-field energies may be used to confirm the stability of the proposed structure solutions. To generate the calculated powder diffraction pattern, lattice parameters, peak shape parameters and background parameters must be determined accurately before proceeding with the structure solution calculations. For this purpose, a novel variant of the Pawley algorithm is proposed, which avoids the instabilities of the original Pawley method. The successful application and performance of *PowderSolve* for crystal structure solution of 14 organic compounds of differing complexity are discussed.

1. Introduction

Analysis of X-ray (and neutron) diffraction data is without question the most powerful tool for the determination of crystal and molecular structures, and many of the most important discoveries of the 20th century have relied on the use of this approach. If sufficiently large single crystals of the material are available for a single-crystal diffraction experiment, powerful techniques (such as direct methods and Patterson methods) now exist to resolve the electron density and hence to determine the crystal structure from the three-dimensional X-ray diffraction pattern.

In many cases, however, the material is available only as a polycrystalline powder. For powder diffraction patterns, the reflections from different crystal planes are averaged over directions and projected onto a single variable, the diffraction angle (2θ). This makes the reconstruction of the underlying crystal structure a much more difficult and ill-conditioned task than for single-crystal diffraction patterns.

Good powder diffraction patterns still yield a wealth of information. Modern synchrotron sources give rise to extremely narrow peak widths and allow highly accurate measurements of peak positions and intensities in the experimental powder diffraction pattern. Unit-cell parameters can be obtained by indexing, using programs such as *ITO* (Visser, 1969), *TREOR*90 (Werner *et al.*, 1985) or *DICVOL*91 (Boultif & Louër, 1991). Likely symmetry groups can be identified from systematic absences in the powder diffraction pattern. The number of molecules in the asymmetric unit can be assessed from density considerations (once the symmetry group is known) or from other techniques, such as solid-state NMR (Thomas *et al.*, 1983).

Once the unit-cell parameters, symmetry group and unit-cell contents are known, two distinct approaches may be adopted to deduce the positions of the individual atoms within the cell from the powder diffraction data: traditional approaches and direct-space approaches. A detailed review of the literature up to 1996 is given elsewhere (Harris & Tremayne, 1996). We mention in passing that a variety of methods have also been developed to predict crystal structures *ab initio* without recourse to experimental powder diffraction patterns. Such methods can be used to generate initial models for subsequent Rietveld (1969) refinement; a review is given by Verwer & Leusen (1998).

Traditional methods (Hauptman & Karle, 1953; Giacovazzo, 1980; Altomare *et al.*, 1994) rely on the successful extraction of integrated Bragg intensities I_{hkl} from the experimental powder diffraction pattern. Once the integrated intensities are known, an electron-density map (assuming X-ray radiation) is constructed using the same techniques that have been developed for single-crystal diffraction data. To extract the integrated intensities, various modifications of the Pawley (Pawley, 1981) or Le Bail (Le Bail *et al.*, 1988; Altomare *et al.*, 1995) methods are commonly used. Variants of this basic idea have been applied successfully to organic systems with up to 31 non-H atoms (Knudsen *et al.*, 1998).

Traditional methods work best when the powder diffraction peaks are very sharp and narrow, and sufficiently well resolved to allow the assignment of an unambiguous intensity value to each reflection. For systems in which peak overlap is severe, the use of high-quality synchrotron radiation data is often advantageous.

Direct-space methods are characterized by direct handling of molecular fragments within the unit cell and do not require the extraction of intensity data for individual reflections from the powder diffraction pattern. Position, orientation and

† New address: Merck KGaA, Pharmaceutical Division, Frankfurter Str. 250, D-64293 Darmstadt, Germany.

© 1999 International Union of Crystallography
Printed in Great Britain – all rights reserved

conformation (degrees of freedom) of these fragments are then varied to generate 'trial' crystal structures, until optimum agreement between calculated and experimental powder diffraction patterns is achieved. In the context of the direct-space approach, which is the subject of the present work, a number of different algorithms to explore parameter space have been used: grid search (Reck *et al.*, 1988; Cirujeda *et al.*, 1995; Dinnebier *et al.*, 1995; Hammond *et al.*, 1997), genetic algorithms (Shankland *et al.*, 1997; Kariuki *et al.*, 1997; Harris, Johnston & Kariuki, 1998; Harris, Johnston, Kariuki & Tremayne, 1998; Kariuki, Calcagno *et al.*, 1999), and Monte Carlo/simulated annealing (Deem & Newsam, 1989, 1992; Newsam *et al.*, 1992; Harris *et al.*, 1994; Harris, Kariuki & Tremayne, 1998; Andreev *et al.*, 1996, 1997; Tremayne, Kariuki & Harris, 1997; David *et al.*, 1998).

We have developed a software package, *PowderSolve*, employing the Monte Carlo/simulated-annealing approach. *PowderSolve* is fully integrated within the *Cerius*^{2} molecular-modelling environment.† Algorithmically, it is based partly on the *StructureSolve* program available in the *Insight*II environment (Newsam *et al.*, 1992).† Here we demonstrate that this method is capable of overcoming the large barriers between local minima in the figure-of-merit hypersurface, which represents the quality of the fit between calculated and experimental powder diffraction patterns expressed as a function of the degrees of freedom defining the structure. By working with a figure of merit based on a full profile comparison between calculated and experimental data, *PowderSolve* uses the experimental data directly as measured, and thus avoids any ambiguities inherent in methods that rely on the prior extraction of integrated intensities.‡

Apart from its speed and efficiency, an important aspect of the present software package is its ease of use. Data preparation, indexing, peak shape analysis, structure solution, Rietveld refinement and lattice-energy calculations may all be carried out within the same package. The degrees of freedom and rigid fragments are defined intuitively by selecting atoms on the graphics screen, and all settings are automatically stored together with the structural information. The structural stability of proposed structure solutions may be easily checked using solid-state force-field or quantum-mechanical calculations, thus providing additional information to validate the proposed structure solution.

2. *PowderFit*: data preparation and Pawley refinement

In many cases, one of the most difficult aspects of the structure solution process is the determination of suitable unit-cell parameters *via* indexing of the powder diffraction pattern. For the purpose of the present paper, we assume that it is possible to obtain one or a limited number of possible lattices by using a suitable indexing approach (Visser, 1969; Werner *et al.*, 1985; Boultif & Louër, 1991; Kariuki, Belmonte *et al.*, 1999).

Once the unit cell is known, the next step in the structure solution process is to refine further the unit-cell parameters; this can be performed without any knowledge of the atomic positions (Pawley, 1981). A meaningful comparison of calculated and experimental powder diffraction patterns in subsequent structure solution calculations also requires that during the calculation of powder diffraction patterns for trial structures, the parameters defining peak shape (*i.e.* peak width and possibly peak mixing parameters) accurately reflect the experimental data.

The original Pawley (1981) procedure to determine lattice parameters, peak shape parameters and background parameters requires the introduction of artificial constraints on the intensities of overlapping peaks, in order to overcome problems of ill-conditioning. A variety of more or less complicated procedures (Le Bail *et al.*, 1988; Jansen *et al.*, 1992; Sivia *et al.*, 1993; Altomare *et al.*, 1994; Shankland & Sivia, 1996) have been proposed to overcome this ill-conditioning problem inherent to the peak fitting procedure. Here, we show that a simple modification to the original Pawley procedure ensures that the method is very stable and well conditioned, even for strongly overlapping peaks. As in the original Pawley method, the whole profile can be fitted simultaneously. The modification ensures that all extracted intensities are positive.

In the Pawley procedure, the experimental powder diffraction pattern is fitted by a sum of profile functions P_{hkl}, centred at the reflection angles θ_{hkl}, and a sum of background functions B_i:

$$I_{\rm exp}(\theta) \simeq I_{\rm calc}(\theta) = \sum_{hkl} I_{hkl} P_{hkl}[(\theta - \theta_{hkl})/H(\theta_{hkl})] + \sum_i b_i B_i(\theta). \quad (1)$$

For convenience of notation, we assume that the profile functions P_{hkl} include any multiplicity and Lorentz–polarization correction factors. $H(\theta_{hkl})$ determines the full width at half-maximum of each reflection and is a function of the reflection angle θ. A common parametrization is

$$H(\theta) = U \tan^2(\theta) + V \tan(\theta) + W. \quad (2)$$

θ_{hkl} itself is a function of the lattice parameters $\{a, b, c, \alpha, \beta, \gamma\}$ *via*

$$2 \sin(\theta_{hkl}) = \lambda / d_{hkl}, \quad (3)$$

where λ is the wavelength of the radiation and d_{hkl} is the Miller spacing for the lattice planes (hkl).

In the Pawley procedure, the integrated Bragg intensities I_{hkl}, background coefficients b_i, lattice parameters and peak width parameters, such as U, V and W in equation (2), are optimized, in order to minimize the weighted profile R factor $R_{\rm wp}$:

$$R_{\rm wp} = \left[\frac{\sum_i w_i |I_{\rm exp}(\theta_i) - I_{\rm calc}(\theta_i)|^2}{\sum_i w_i |I_{\rm exp}(\theta_i)|^2} \right]^{1/2}, \quad (4)$$

where $w_i = 1/I_{\rm exp}(\theta_i)$. As described by Jansen *et al.* (1992), substituting $I_{\rm calc}$ from equation (1) into (4), for fixed values of the lattice and peak shape parameters, the optimization of equation (4) constitutes a linear least-squares problem for the intensities I_{hkl} and background coefficients b_i. The standard method of solution is to solve the resulting linear system of equations; however, if peaks overlap strongly, the least-squares matrix becomes singular, and no unique solution can be found. Even with the introduction of artificial constraints, it is common to obtain negative and wildly fluctuating values of the

† Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121–3752, USA.

‡ As stated by David *et al.* (1998), the profile comparison measure introduced by Shankland *et al.* (1997) is essentially equivalent to a full-profile comparison and therefore also avoids such ambiguities, even though it is based on extracted intensities.

integrated intensities I_{hkl} from the Pawley refinement, unless the starting values for the integrated intensities are very close to the correct values.

This arbitrariness is reduced if we impose positivity on the variable intensities I_{hkl}. This is achieved by using the structure-factor amplitudes $|F_{hkl}|$ as the basic optimization variables instead of the intensities I_{hkl}. The Bragg intensities I_{hkl} are related to the structure factors F_{hkl} *via*

$$I_{hkl} = |F_{hkl}|^2. \tag{5}$$

Instead of minimizing $R_{\rm wp}$ with respect to I_{hkl}, we minimize $R_{\rm wp}$ with respect to $|F_{hkl}|$. This ensures positivity of the I_{hkl}.

However, there is a 'price to pay' for these advantages: the least-squares problem is now nonlinear with respect to the parameters $|F_{hkl}|$ and the optimization cannot be performed by solving a linear system of equations. We therefore use a standard iterative conjugate gradient minimizer. The evaluation of the gradient

$$\frac{\partial R_{\rm wp}}{\partial |F_{hkl}|} = 2|F_{hkl}| \frac{\partial R_{\rm wp}}{\partial I_{hkl}} \tag{6}$$

involves the same sparse least-squares matrix $\partial^2 R_{\rm wp}/\partial I_{hkl}\partial I_{h'k'l'}$ as the original least-squares problem and can be performed efficiently using sparse matrix-vector multiplications. The introduction of a positivity constraint *via* the auxiliary variables $|F_{hkl}|$ makes the optimization very stable and well behaved.

As suggested by Jansen *et al.* (1992), we recommend that the peak fitting is performed as a two-step procedure. In the first step, the integrated intensities and background parameters are optimized, as described above, for fixed values of peak shape parameters, lattice parameters and zero-point shift parameters. In the second step, these parameters are adjusted with the values of the intensity and background parameters fixed. This two-step procedure is repeated until convergence is reached. For both steps, the same conjugate gradient minimizer is used; however, the evaluation of the gradient with respect to $|F_{hkl}|$ and b_i in the first step is much faster, due to the use of sparse matrix algebra, than the evaluation of the gradient with respect to the remaining parameters, which requires complete recalculation of all the peak shape functions at every point in the powder diffraction pattern.

A drawback of using conjugate gradient minimization, instead of inverting a linear system of equations, is that error estimates and correlations for the integrated intensities are not readily calculated. However, for the purpose of structure solution, such error estimates and correlations are generally not required.

We have implemented the above algorithm in the program *PowderFit*. At present, *PowderFit* allows the refinement of lattice parameters, background coefficients, zero-point shift parameters and peak width parameters. Seven profile functions have been implemented: apart from standard Gaussian and Lorentzian functions, the program also allows two modified Lorentzians, pseudo-Voigt, Pearson VII and modified Thompson–Cox–Hastings functions (Young, 1993). Where applicable, appropriate mixing parameters for these profile functions have been introduced as described by Young (1993). These mixing parameters and a simple asymmetry correction (Rietveld, 1969) can also be refined. An extension to arbitrary profile functions and more sophisticated anisotropic peak broadening functions and asymmetry corrections will be implemented in the future.

Once a suitable set of parameters has been found, *PowderFit* can also be used to explore systematic absences and thereby aid in the determination of possible space groups. This, in combination with density considerations, is generally straightforward for organic crystals, which are known to crystallize almost exclusively in a limited number of triclinic, monoclinic or orthorhombic space groups (Baur & Kassner, 1992).

3. *PowderSolve*: structure solution

Once the experimental powder diffraction pattern has been fitted and lattice parameters, peak shape parameters and background parameters have been determined, we employ a combination of simulated annealing and rigid-body Rietveld refinement to deduce the structural degrees of freedom, *i.e.* the positions, orientations and intramolecular torsions of the molecular fragments in the asymmetric unit. As discussed in §1, this procedure is carried out using a simulated-annealing algorithm. Simulated-annealing techniques and their application to structure solution from powder diffraction data have been described in detail elsewhere (Kirkpatrick *et al.*, 1983; van Laarhoven & Aarts, 1987; Deem & Newsam, 1989; Newsam *et al.*, 1992; Andreev *et al.*, 1997; David *et al.*, 1998).

The degrees of freedom are defined intuitively by selecting groups of atoms, such that within each group, the relative positions of atoms remain unchanged throughout the simulation (rigid bodies). Flexible torsions can be defined around the bonds that link these rigid groups. Note that it is possible to define arbitrary numbers of rigid groups (whether linked or not), and that it is possible to define rigid groups consisting of single atoms.

During each simulated-annealing step, a single degree of freedom is modified by a random amount limited by the step size for that degree of freedom. The powder diffraction pattern for the resulting structure is then calculated, and this powder pattern is compared to the experimental powder pattern, using $R_{\rm wp}$ defined in equation (4). The rate-limiting step is the evaluation of the structure-factor amplitudes $|F_{hkl}|$. We have therefore spent much effort to optimize the evaluation of these structure-factor amplitudes (see §4.5).

In our simulated-annealing method, we use an adaptive temperature schedule: the rate of cooling is controlled by the fluctuations in the figure of merit $R_{\rm wp}$. Also, the step widths determining how far the system moves in parameter space for a given simulated-annealing step are controlled individually for each degree of freedom, based on the acceptance ratio and fluctuations.

The efficiency of the method is enhanced significantly by performing a local Rietveld optimization within the parameter space defined by the degrees of freedom, whenever a promising structure solution is obtained during the calculation. By performing these intermediate structure optimizations (local quenching), we avoid having to go to very low annealing temperatures during the main simulated-annealing runs. In the standard simulated-annealing procedure, once the system is cooled to low temperatures, the thermal fluctuations are insufficient to move the system across barriers in $R_{\rm wp}$; the system would then be effectively 'frozen' in one local minimum, and further cooling would merely perform a local

optimization. The intermediate minimization operations in our method perform this local optimization more efficiently on a larger number of structures, allowing the global simulated-annealing procedure to remain at relatively high temperatures throughout the simulation. If an intermediate optimization results in a structure with lower R_{wp} than any structure found previously, this structure is written to a trajectory file and retained for future consideration as a potential structure solution. Subsequent simulated-annealing steps proceed from the structure generated prior to the intermediate optimization.

By default, the starting temperature of the calculation is 1.5 times the average fluctuations in R_{wp} for a random sequence of moves, and the end temperature is one fifth of the start temperature. We found these values to be appropriate in many cases, but in some cases a different temperature schedule may be more efficient. Note that with these default values, the temperature at the end of the simulation is still high enough for the system to overcome most barriers on the R_{wp} hypersurface. The local Rietveld optimization (with respect to the degrees of freedom in the calculation) is performed using the method of Powell (Press *et al.*, 1986), which does not require the evaluation of any gradients. Since the minimization is performed very infrequently, the additional time required is insignificant compared to the overall time spent on the global simulated-annealing calculation.

PowderSolve also allows access to the Powell optimization outside the framework of the simulated-annealing structure solution calculation. By exporting intensity information for a trial structure solution into *PowderFit*, it is even possible to post-refine peak shape parameters and background parameters after a potential structure solution has been found. Thus, in addition to its main application for structure solution, the program package may also be used as a rigid-body Rietveld refinement tool.

As an additional analysis tool, the program allows visualization of the variation of the figure of merit R_{wp} as a function of individual degrees of freedom, which is useful for assessing the relative importance and behaviour of different degrees of freedom and their influence on the overall figure of merit.

PowderSolve works with all possible space groups and common space-group settings. In its present form, it does not cope well with systems in which individual atoms or fragments are located on special positions; these can be treated by reducing symmetry to remove special positions.

4. Applications

4.1. *Structures studied*

The performance of *PowderSolve* has been validated and tested for a set of known molecular crystal structures. For most of these structures, with the exception of 4-amidinoindanone guanylhydrazone (AIGH) (Karfunkel *et al.*, 1996) (see below), direct-space methods (based on a profile *R* factor) have previously been applied successfully to solve the crystal structures from powder diffraction data. The test set was chosen to cover a wide range of molecular crystals of differing complexity. In general, the complexity and difficulty of the structure solution process for direct-space methods increases with the number of degrees of freedom that are varied. In contrast, for traditional methods of structure solution, the complexity depends more directly on the total number of atoms in the asymmetric unit.

Fig. 1 shows the chemical formulae of the compounds included in our tests. Table 1 lists important structural data and major results from our simulations. The first structure, 1-methylfluorene (Tremayne *et al.*, 1996*a*), comprises a simple rigid molecule. The next five structures, *p*-methoxybenzoic acid (Tremayne *et al.*, 1996*b*; Harris, Johnston & Kariuki, 1998), red fluorescein (Tremayne, Kariuki & Harris, 1997), *o*-thymotic acid (Kariuki *et al.*, 1997), formylurea (Harris, Johnston & Kariuki, 1998), and 4-toluenesulfonylhydrazine (Lightfoot *et*

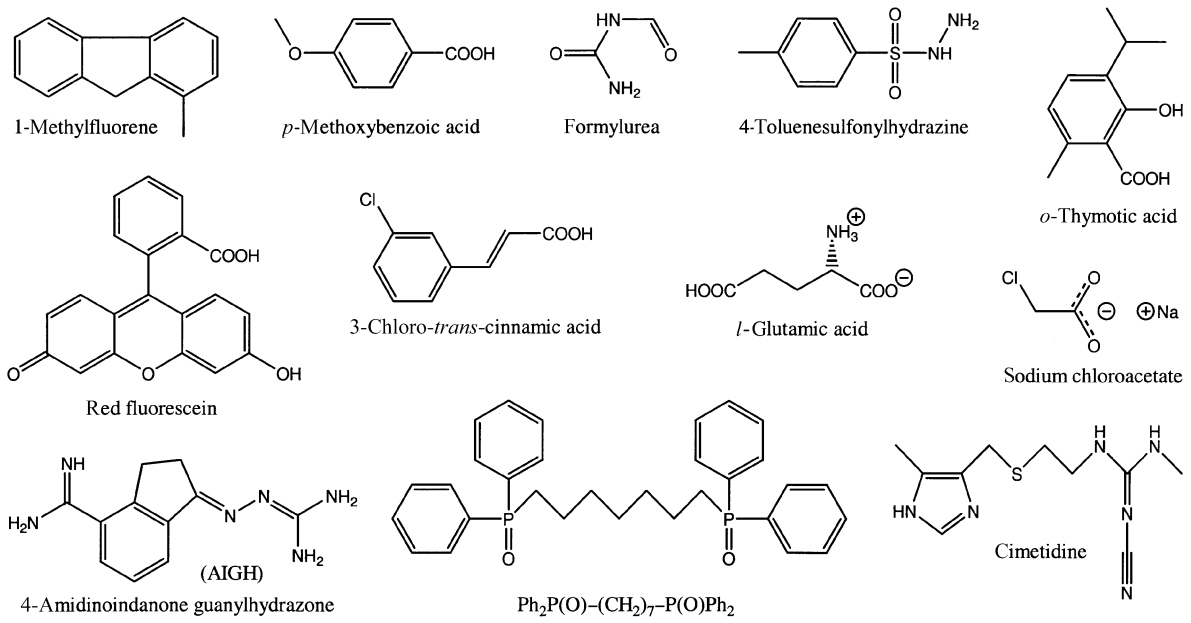


Fig. 1. Compounds considered in this work.

Table 1. *Summary of results of the structure solution from X-ray powder diffraction patterns for 14 molecular crystals, using PowderSolve*

The table lists the space group, the number of non-H atoms per asymmetric unit, the total number of degrees of freedom (DOF) and the number of torsional degrees of freedom, the number of simulated-annealing steps per run, the R_{wp} and R_p factors between the calculated and experimental powder diffraction patterns for the best structure solutions, the success rate taken from ten independent runs, and the time per run on an SGI O2 workstation with an R5000 180 MHz processor.

	Space group	No. of non-H atoms	Total DOF	Torsional DOF	Steps (×1000)	R_{wp} (%)	R_p (%)	Success (%)	Time (min)
1-Methylfluorene [a]	$P2_1/n$	14	6	0	70	12.9	9.7	100	3.9
p-Methoxybenzoic acid [b]	$P2_1/a$	11	8	2	100	9.4	7.3	100	3.9
Red fluorescein [c]	$Pn2_1a$	25	7	2	80	14.8	11.4	100	7.3
o-Thymotic acid [d]	$P2_1/n$	14	8	2	100	11.7	9.1	90	6.2
Formylurea [e]	$Pna2_1$	6	7	2	80	10.3	7.8	100	1.4
4-Toluenesulfonylhydrazine [f]	$P2_1/n$	12	8	2	100	9.5	6.8	100	4.2
3-Chloro-*trans*-cinnamic acid [g]	$P2_1/a$	12	9	3	140	22.5	17.9	90	6.0
l-Glutamic acid (α phase) [g]	$P2_12_12_1$	10	10	4	300	15.9	12.5	30	9.3
l-Glutamic acid (β phase) [g]	$P2_12_12_1$	10	10	4	300	15.4	12.2	30	8.9
AIGH (α phase) [h]	$P\bar{1}$	17	10	4	300	21.8	16.9	100	10.3
AIGH (β phase) [h]	$P2_1/c$	17	10	4	300	20.6	16.2	50	13.8
Sodium chloroacetate [i]	$P2_1/a$	6	10	1	300	18.3	14.1	50	4.5
Cimetidine [j]	$P2_1/n$	17	14	8	4800	12.3	9.2	30	220
$Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$ [k]	$P2_1/n$	37	18	12	73600	4.4	3.2	30	11400

References: (*a*) Tremayne *et al.* (1996*a*); (*b*) Tremayne *et al.* (1996*b*); Harris, Johnston & Kariuki (1998); (*c*) Tremayne, Kariuki & Harris (1997); (*d*) Kariuki *et al.* (1997); (*e*) Harris, Johnston & Kariuki (1998); (*f*) Lightfoot *et al.* (1993); (*g*) Kariuki *et al.* (1996); (*h*) Karfunkel *et al.* (1996); (*i*) Elizabé *et al.* (1997); (*j*) Cernik *et al.* (1991); (*k*) Kariuki, Calcagno *et al.* (1999).

al., 1993) represent crystal structures of small molecules with some degree of flexibility. Significantly more complex are molecules with several connected intramolecular torsions. The crystal structures of 3-chloro-*trans*-cinnamic acid (Kariuki *et al.*, 1996), the α and β phases of *l*-glutamic acid (Kariuki *et al.*, 1998), and the α and β phases of AIGH (Karfunkel *et al.*, 1996) have been chosen as representatives of such systems. Sodium chloroacetate (Elizabé *et al.*, 1997) is an example of a simple salt, which illustrates the applicability of our approach to structures with more than one molecular fragment in the asymmetric unit. The last two structures, cimetidine (Cernik *et al.*, 1991) and $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$ (Kariuki, Calcagno *et al.*, 1999), are two of the most complex molecular crystal structures solved from powder X-ray diffraction patterns so far. Both molecules are highly flexible (8 and 12 intramolecular torsions, respectively) and have long chains of connected torsions.

4.2. *Simulation setup*

The aim of this validation study was not only to establish the applicability of the program to solve the crystal structures of a wide range of molecular crystals, but also to test systematically the reliability and speed of the program. This allows us to gain some understanding of how speed and reliability of direct-space methods depend on the number and type of degrees of freedom included in the calculation, and how to optimize the setup of the calculations. Such knowledge is particularly important for designing procedures that allow routine and reliable solution of crystal structures.

The majority of experimental powder X-ray diffraction patterns used to solve the crystal structures in Table 1 were collected using conventional laboratory diffractometers. Diffraction patterns recorded using synchrotron X-ray radiation were available only for fluorescein (Tremayne, Kariuki & Harris, 1997), sodium chloroacetate (Elizabé *et al.*, 1997) and cimetidine (Cernik *et al.*, 1991). Details of the data collection procedures are given in the original literature cited. Although the availability of synchrotron data may be desirable in many cases, it is by no means essential for successful structure solution by direct-space methods employing the weighted profile *R* factor. High-quality laboratory powder X-ray diffraction patterns are sufficient for solving the crystal structures of molecules as complex as $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$ (Kariuki, Calcagno *et al.*, 1999). Clearly, direct-space methods employing the profile *R* factor depend more directly on having a good definition of profile parameters, rather than high resolution *per se*.

We have generally not attempted to re-index the powder diffraction patterns. Unit-cell parameters and space groups were taken from the published work. Except for these data, no other information was used to assist in the determination of the crystal structures.

Once the unit-cell parameters and space group were known, *PowderFit* was applied to determine more accurate unit-cell parameters, profile parameters and background parameters, as described in §2. Since the time per simulated-annealing step increases linearly with the number of reflections included in the calculation, it is important to optimize the 2θ range of reflections included in the calculations. We have found that it is adequate to restrict the high-angle limit of 2θ values such that only the first 100 to 200 reflections are included in the calculations (see Table 2). In general, this range still contains many strongly overlapping peaks. For each structure, Table 2 lists the number of reflections used as well as the best R_{wp} factor obtained for the fit of the experimental powder diffraction pattern using *PowderFit*.

In the next step, the molecular fragments forming the asymmetric unit of the crystal are constructed. Since torsions are the only intramolecular degrees of freedom varied during

Table 2. *Summary of results of profile fits for 14 powder diffraction patterns using PowderFit*

The R_{wp} and R_p factors [see equations (4) and (8)] measure the quality of the fit; the R_B factor [see equation (7)] shows how well the extracted intensities agree with the calculated intensities for the known structures corresponding to these powder diffraction patterns. Two R_B factors are shown: the column R_B is calculated for the number of reflections shown in the first column; the column R_B^{50} is calculated for the first 50 reflections only.

	No. of reflections	R_{wp} (%)	R_p (%)	R_B (%)	R_B^{50} (%)
1-Methylfluorene	189	8.4	6.1	30.5	14.0
p-Methoxybenzoic acid	152	7.1	5.3	19.0	9.2
Red fluorescein	107	13.4	10.1	24.0	10.4
o-Thymotic acid	195	8.0	6.3	85.8	32.6
Formylurea	69	6.4	4.7	23.4	19.4
4-Toluenesulfonylhydrazine	159	5.5	4.0	23.0	12.0
3-Chloro-*trans*-cinnamic acid	161	14.8	9.9	50.3	28.8
l-Glutamic acid (α phase)	82	11.6	7.4	23.3	20.1
l-Glutamic acid (β phase)	81	11.3	7.3	22.3	18.8
AIGH (α phase)	130	9.0	6.3	89.2	40.5
AIGH (β phase)	128	12.4	8.9	74.1	51.3
Sodium chloroacetate	101	15.0	11.5	37.2	27.8
Cimetidine	122	7.8	5.2	16.3	11.7
$Ph_2P(O)–(CH_2)_7–P(O)Ph_2$	190	2.6	2.0	55.8	45.3

the calculation, it is important to generate molecular fragments which reflect all other (fixed) intramolecular geometric parameters as accurately as possible. Often the initial structure of molecular fragments may be obtained using standard values for bond lengths and angles (see *e.g.* Kariuki *et al.*, 1997; Tremayne, Kariuki & Harris, 1997; Tremayne, Kariuki, Harris *et al.*, 1997). Alternatively, a molecular-modelling package such as *Cerius*[2] can be used to sketch and minimize the molecules using an appropriate force-field-based or a quantum-mechanics-based method. This latter approach provides an effective way of generating a highly accurate initial molecular geometry. In all approaches, the constraints on bond lengths and angles can be removed once a promising crystal packing arrangement has been found, and the structure can be further refined using a Rietveld method.

Every structure solution calculation is started from a randomly generated initial structure (with random initial values for each degree of freedom). The number of steps in the simulated-annealing calculation has been chosen to increase exponentially with the total number of degrees of freedom included in the calculation. Since simulated annealing is a stochastic procedure, there is no guarantee that the global minimum will actually be located in a given run with a finite number of steps. Thus, it is a good strategy to repeat the calculation several times from different starting structures. If the same structure or very similar structures are found repeatedly, it is a strong indication that these represent the global minimum. Table 1 lists the number of steps per run chosen for the test structures, as well as the success rate found for ten independent runs.

4.3. *Results*

For all the test examples listed in Table 1, the program was able to find a solution that was the same or very close to the known crystal structure given in the literature. In the following discussion, our results are described in more detail.

In the case of small rigid or partly flexible molecules (see Table 1 from 1-methylfluorene to formylurea), the correct crystal structure is found routinely. The complexity of the global optimization clearly increases if the molecules become more flexible (*i.e.* as the number of degrees of freedom increases) and, in particular, if the intramolecular torsions are connected and form long chains. This is shown for the α and β phases of *l*-glutamic acid and for AIGH. In these cases, the success rates with our standard setup for the calculations are typically lower than for the previous examples.

Similarly, if there is more than one independent fragment in the asymmetric unit, as for an organic salt such as sodium chloroacetate (or indeed for a structure with two or more independent molecules of the same type in the asymmetric unit), the complexity of the global optimization is increased. Our tests seem to indicate, however, that cases with long chains of connected intramolecular torsions represent a greater challenge than cases with two or more rigid (or partly flexible) molecular fragments in the asymmetric unit. We conjecture that as a result of strong coupling between the torsional degrees of freedom in long flexible chains, the correct location of a minimum in the R_{wp} hypersurface requires several degrees of freedom in order to achieve simultaneously their correct values. In addition, for long chains, the number of similar chain conformations increases, resulting in a large number of local minima spread over the R_{wp} hyperspace.

Typically for these flexible molecules it is found that the R_{wp} hypersurface is very flat with narrow but deep minima. Thus, a large number of trial structures with high R_{wp} are generated before eventually an appropriate minimum is found and the R_{wp} factor drops sharply. This is illustrated in Fig. 2, which shows the distribution of R_{wp} values for all structures generated in a simulated-annealing calculation. The calculation spends most of the time at R_{wp} values close to the maximum. The probability of sampling low R_{wp} values is clearly enhanced by the use of simulated annealing, although the high 'plateau' in the R_{wp} hypersurface is still sampled frequently.

For flexible long-chain molecules, usually several low-R_{wp} solutions are found, which correspond to one of the many similar conformations. Examples for molecules with such long chains are cimetidine and $Ph_2P(O)–(CH_2)_7–P(O)Ph_2$. Fig. 3 illustrates this behaviour for the case of $Ph_2P(O)–(CH_2)_7–P(O)Ph_2$. The figure shows the crystal structure found in the previous work (Kariuki *et al.*, 1999) and compares it to the best solution found during our present work. The R_{wp} factors

considering the whole measured powder diffraction pattern up to $2\theta = 50^\circ$ are 4.98% for the solution of Kariuki, Calcagno *et al.* (1999) and 4.83% for the best solution in our present work. Note that we used a more accurate background description than is provided in the released version of *PowderFit* to obtain such low R_{wp} values for this compound. From Fig. 3, although both crystal structures show the same packing motif, it can be seen that several torsions in the $(CH_2)_7$ chain have different values. These small differences illustrate the difficulty in locating the crystal packing which globally optimizes R_{wp} for large flexible molecules.

To resolve the small differences in the structure solutions from *PowderSolve* compared to the published structure for this compound, we performed a force-field-based energy minimization on both structures, keeping the unit-cell parameters fixed. The *COMPASS* force-field (Sun, 1998) was used for this optimization. Interestingly, the structures minimized from the two different starting points are identical. The energetically optimized structure is extremely close to the solution described by Kariuki, Calcagno *et al.* (1999). This illustrates how force-field-based calculations can provide additional information in cases in which the powder pattern alone does not contain sufficient information to distinguish unambiguously between similar structure solutions.

4.4. *Influence of quality of diffraction patterns*

We now consider how the quality of the experimental powder diffraction pattern influences the possibility for successful structure solution. Since most of the diffraction data for our test structures were recorded using laboratory diffractometers, our test results verify that high-quality laboratory powder X-ray diffraction patterns are sufficient for successful structure solution of even highly flexible molecules.

There are two crucial steps in structure solution from powder X-ray diffraction patterns: indexing the diffraction pattern (including space-group determination) and locating the crystal structure that represents the global optimum of agreement between the calculated and experimental powder diffraction patterns. It is not clear from the outset which of these steps, indexing or global optimization, provides the more severe limitation in the case of lower quality powder diffraction patterns. The crystal structure of formylurea represents a good testing case to investigate the second aspect, *i.e.* how the broadening of the experimental powder pattern influences the prospects for locating the optimum crystal structure.

The optimum structure solution for formylurea at the end of a structure solution calculation (*i.e.* without additional Rietveld refinement of parameters not included in the structure solution process) has an R_{wp} factor of 10.3%, but a second solution with a different conformation of the molecule has a higher R_{wp} factor of 12.0%. Using the experimental powder diffraction pattern, the optimal crystal structure is located every time with our standard setup. If we artificially broaden (convolute) the diffraction pattern with Gaussian functions of increasing peak widths, we find that the difference in the R_{wp} factor between these two structure solutions decreases as the peaks become broader. As a consequence, starting at a broadening (half width) of 0.6°, the second solution (or solutions with an intermediate molecular conformation) is sometimes found at the end of a standard run with 80 000 simulated-annealing steps. But even for the broad overlapping peaks obtained by broadening the experimental powder X-ray diffraction pattern with a Gaussian function of half width 1.2°, the correct solution is still found in two out of five simulated-annealing runs. The difference in R_{wp} between the two lowest lying structurally distinct local minima drops from 1.7% for no broadening (*i.e.* the experimentally recorded data) to 1.5% for a broadening of 0.6° and to 0.2% for a broadening of 1.2°. Even for larger broadening, the correct structure remains

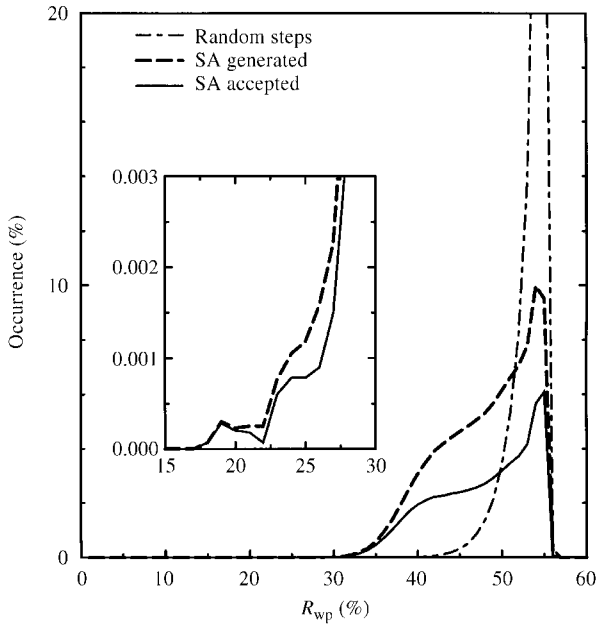


Fig. 2. Distribution of R_{wp} factors of structures generated during a random generation of 10^6 trial structures for the β phase of *l*-glutamic acid (dot–dashed line) and during a simulated-annealing run of 3×10^5 steps (dashed line). Compared to the Monte Carlo procedure at fixed temperature, simulated annealing preferentially samples those parts of phase space with lower R_{wp}.

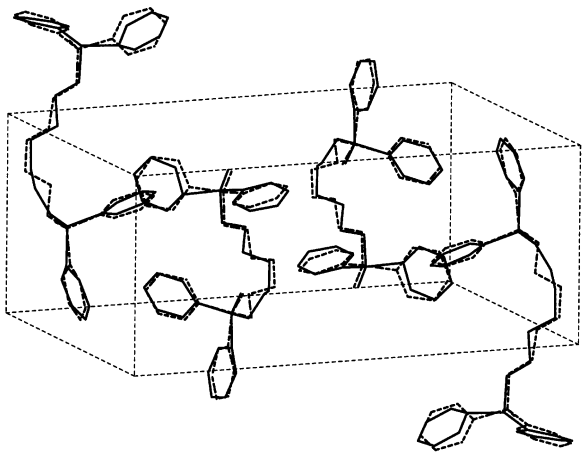


Fig. 3. Crystal structure of $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$. The figure compares the crystal structure found by Kariuki, Calcagno *et al.* (1999) (full lines) to the best solution found during our present work (dashed lines).

slightly lower in R_{wp} than other local minima, but the discrimination and therefore the ability of the program to locate this global minimum is reduced.

Although further investigations are necessary, this simple test indicates that the actual structure solution for a known unit cell and space group is not substantially affected by the width of the peaks in the diffraction pattern, except perhaps in the case of severe line broadening. However, indexing a powder diffraction pattern with broad peaks is much more difficult than indexing a high-resolution powder diffraction pattern, and this is probably the limiting aspect for the application of laboratory powder diffraction data in structure determination.

4.5. *Speed*

The speed of the calculation determines the extent of parameter space that can be explored within an acceptable period of time using the simulated-annealing method, and thereby determines the chance of finding the global optimum structure solution. Additionally, in many cases it may be necessary to carry out a series of independent calculations to test different potential space groups and/or unit-cell choices.

Using pre-calculations wherever possible and employing an efficient algorithm for the calculation of the structure factors, *PowderSolve* evaluates 100 to 1000 trial structures per second on a standard SGI O2 workstation with an R5000 processor at 180 MHz. For calculation of the structure factors, the calculation time depends linearly on the number of atoms used to calculate the structure factor and on the number of reflections in the calculated powder diffraction pattern. For all structures tested, the evaluation rate for the structure factors on the SGI O2 is nearly constant: 300 structures per second per 50 atoms† and per 100 reflections. Except for cimetidine and $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$, the structure solution calculations took less than 15 min. The complete solution of the crystal structure of $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$ took about four days on one 225 MHz R10000 processor of an SGI Octane machine, indicating that systems of up to about 18 degrees of freedom can be solved in realistic periods of time using modern techniques for crystal structure solution from powder diffraction data.

4.6. *Number of simulated-annealing steps*

Simulated annealing is based on a stochastic process and is guaranteed to find the global optimum only for an infinitely long run. In practice, there are two strategies to optimize the simulation: either to perform a small number of independent, long simulated-annealing runs, or to perform a large number of relatively short, independent simulated-annealing runs. In either case, each simulated-annealing run should start at a new randomly generated point in parameter space. During our test runs, we have found that both of these strategies are applicable, but the most consistent success has been obtained using about ten relatively long runs. Since the parameter space expands exponentially as the number of degrees of freedom increases, the number of simulated-annealing steps necessary to achieve a reasonable success rate should also increase exponentially. We have investigated in more detail the dependence of the number of simulated-annealing steps needed to locate the optimal structure on the number of degrees of freedom. The gray diamonds in Fig. 4 show the average number of simulated-annealing steps that were necessary to locate the correct crystal structure, as a function of the total number of degrees of freedom included in the calculation. Although the statistics of this graph are not converged, in particular for the cases with a large number of degrees of freedom, the plot illustrates the exponential increase of simulated annealing steps necessary to find the optimal solution. The graph provides a rough estimate of the number of steps necessary to solve the crystal structure for a given number of degrees of freedom. Based on this estimate, *PowderSolve* automatically proposes to the user the recommended length of the simulated annealing runs. This number, which has also been used for the test runs, is indicated by the black squares in Fig. 4. For nearly all the test structures, calculations using the proposed number of steps find the optimal crystal structure with a reasonable success rate (see Table 1). In the case of highly flexible molecules, such as cimetidine and $Ph_2P(O)$–$(CH_2)_7$–$P(O)Ph_2$, the proposed number of steps was not sufficient and it was necessary to double this number. Work is in progress to obtain more accurate estimates, taking into account not only the number of degrees of freedom, but also the nature of the degrees of freedom for a given structure.

5. Extended applications of *PowderFit*

In this section we discuss potential applications of *PowderFit*, the peak fitting program, in the context of traditional methods of structure solution based on the use of extracted peak intensities. *PowderFit* was designed to establish appropriate values of parameters, such as peak widths and lattice parameters, required by the direct-space structure solution program *PowderSolve*, but it is legitimate to ask how accurately the novel Pawley algorithm employed by *PowderFit* extracts integrated intensities of reflections (although they are not required by *PowderSolve*, but are potentially useful for other applications).

As an initial step in this direction, we have investigated how well the extracted intensities from the set of powder diffraction patterns investigated in §4 match calculated intensities for the

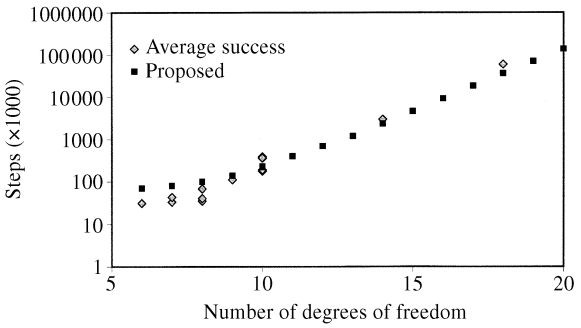


Fig. 4. Dependence of the number of simulated-annealing steps on the total number of variable degrees of freedom defining the crystal structure. Grey diamonds show the average number of simulated-annealing steps necessary to locate the correct crystal structure; the black squares indicate the number of steps proposed by *PowderSolve* using an empirical formula.

† Our algorithm makes use of inversion symmetry and centring operations in the evaluation of the structure factors, so the number of atoms quoted here is the number of atoms per unit cell reduced by inversions and centring operations.

corresponding crystal structures. Note that for the structures considered, exact accidental overlap was rare due to their low symmetry, and since we are interested particularly in the ability of *PowderFit* to extract intensities for strongly overlapping peaks, no equipartitioning was employed: the intensities are those resulting from the conjugate gradient optimization.

Table 2 shows the Bragg R factor R_B and profile R factor R_p for 12 trial structures used as a test for the structure solution. R_B and R_p are defined as

$$R_B = \frac{\sum_{hkl} |I_{hkl}(\text{true}) - I_{hkl}(\text{fit})|}{\sum_{hkl} I_{hkl}(\text{true})} \tag{7}$$

and

$$R_p = \frac{\sum_i |I_{\text{exp}}(\theta_i) - I_{\text{calc}}(\theta_i)|}{\sum_i I_{\text{exp}}(\theta_i)}. \tag{8}$$

We find that for most structures, the values of R_B are in the range 30–40% if calculated over the full range of the experimental powder diffraction pattern. If we consider only the first 50 reflections, the R_B values are as low as 10–20%. As one might expect, the use of synchrotron X-ray powder diffraction data (fluorescein, sodium chloroacetate and cimetidine) seems to allow a more accurate extraction of integrated intensities, but other factors such as the presence or absence of preferred orientation, temperature factors, *etc*., are also important. The powder diffraction pattern for the compound AIGH has very broad peaks; in that case, peak overlap prevents an accurate extraction of intensities even for the low-angle peaks.

Fig. 5 shows the case of cimetidine, for which the extraction of integrated intensities from the powder diffraction pattern works well. It is clear that the relative deviations of extracted intensities from calculated intensities are larger for small peaks. Work is in progress to assess whether the extracted intensities from *PowderFit* are sufficiently accurate to be used in traditional methods for structure solution.

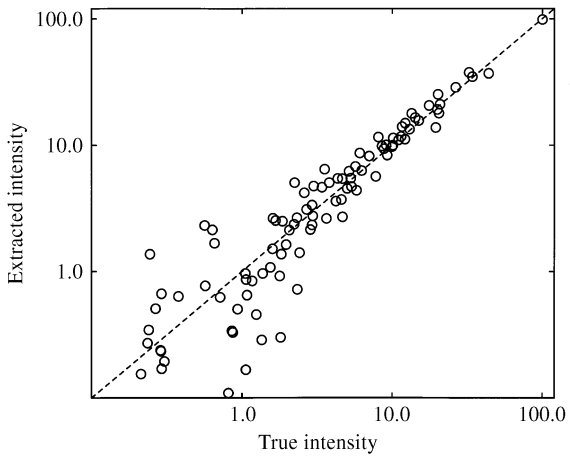


Fig. 5. Bragg intensities extracted from an experimental powder diffraction pattern of cimetidine using *PowderFit*, *versus* intensities calculated from the correct crystal structure (determined following Rietveld refinement).

6. Conclusions

The past few years have witnessed the development of many new algorithms and methods for crystal structure determination from powder diffraction data. Both traditional and direct-space methods for structure solution have now been applied successfully to solve the crystal structures of fairly complex compounds.

In this work, a carefully optimized implementation of a direct-space structure solution method has been presented, which is fully integrated within the *Cerius*^{2} modelling package. The structure optimization is based on a Monte Carlo/simulated-annealing technique.

It has been demonstrated that for up to about ten degrees of freedom, this approach is capable of locating the positions and orientations of molecular fragments to match an experimental powder diffraction pattern within a matter of minutes (as with all structure solution methods, it is assumed that the unit-cell parameters and the space group have been determined beforehand, and in addition, the unit-cell contents must be provided). The structure solution of a compound with 18 degrees of freedom took a few days on an SGI workstation. Note that this number of degrees of freedom is similar to the largest number of degrees of freedom solved to date from powder diffraction data using global optimization methods (Le Bail, 1993–1999; Kariuki, Calcagno *et al.*, 1999).

Speed and reliability of the program are, of course, not the only requirements which have driven this development. Ultimately, if structure solution from powder diffraction data is to become a mainstream analytical technique, it is necessary to provide the laboratory scientist with a software package that enables him or her to perform all stages of structure determination within a common environment: model definition, indexing, profile fitting, structure solution, Rietveld refinement and tests for structural stability based on lattice-energy calculations. An important additional component in this process, reported here, has been the development of *PowderFit* to perform the peak shape analysis of an experimental powder diffraction pattern. The robustness and reliability of this method are achieved *via* a simple enhancement of the Pawley procedure.

Ease of use should not distract from the fact that a number of bottlenecks in structure determination from powder diffraction data still remain.

In the structure solution process itself, the exponential increase of the size of the search space as a function of the number of degrees of freedom means that there may be a limit to the complexity of systems that can be tackled, regardless of the type of optimization algorithm used. Nevertheless, it should still be possible to increase the currently demonstrated limit of 18 degrees of freedom somewhat, by using global optimization techniques of improved efficiency.

One possible way to overcome this apparent barrier starts from the observation that the structures sampled by direct-space structure solution calculations include a vast number of structures which could theoretically be excluded on the basis of stability and energy arguments; for example, most currently used direct-space methods do not exclude structures in which molecules overlap. A future challenge is to find ways of effectively reducing the range of parameter space to be explored by taking such structural considerations into account, without compromising speed and the probability of accessing the region of parameter space which contains the correct

structure solution (excluding structures on the basis of stability or energy may result in the removal of pathways towards this region of parameter space).

Another possible approach to overcome the complexity constraints may lie in a combination of traditional and direct-space methods. Such an approach may benefit from the improved Pawley procedure proposed in this paper.

A second bottleneck exists at the indexing stage. While we have demonstrated that the quality of the powder diffraction pattern is not of critical importance at the structure solution stage, provided the search is conducted with the correct unit cell and space group, unambiguously determining this unit cell and likely space groups in the first place frequently remains a difficult task which often requires a high-resolution powder diffraction pattern. Further progress and developments in strategies for indexing powder diffraction patterns are clearly required (Kariuki, Belmonte *et al.*, 1999).

In conclusion, we have developed a powerful and easy-to-use software package, *PowderSolve*, for crystal structure solution from powder diffraction data, which has been validated by successfully solving the crystal structures of 14 compounds of differing complexity. Both structure solution and subsequent rigid-body Rietveld refinement may be carried out by the same program. Current research is aimed at ways of further improving the optimization strategy, by including energy terms and constraints on the variable degrees of freedom to guide the structure solution calculation towards packing arrangements that are structurally and energetically sound.†‡

We thank B. M. Kariuki and M. Tremayne for recording some of the experimental powder diffraction patterns studied here. Many members of staff at MSI have contributed to this development; in particular, we would like to thank C. M. Freeman and N. A. Austin for many valuable suggestions throughout this development.

References

Altomare, A., Burla, M. C., Cascarano, G., Giacovazzo, A., Guagliardi, A., Moliterni, A. G. G. & Polidori, G. (1995). *J. Appl. Cryst.* **28**, 842–846.

Altomare, A., Cascarano, G., Giacovazzo, A., Guagliardi, A., Burla, M. C., Polidori, G. & Camalli, M. (1994). *J. Appl. Cryst.* **27**, 435–436.

Andreev, Y. G., Lightfoot, P. & Bruce, P. G. (1996). *J. Chem. Soc. Chem. Commun.* pp. 2169–2170.

Andreev, Y. G., Lightfoot, P. & Bruce, P. G. (1997). *J. Appl. Cryst.* **30**, 294–305.

Baur, W. H. & Kassner, D. (1992). *Acta Cryst.* B**48**, 356–369.

Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987–993.

Cernik, R. J., Cheetham, A. K., Prout, C. K., Watkin, D. J., Wilkinson, A. P. & Willis, B. T. M. (1991). *J. Appl. Cryst.* **24**, 222--226.

Cirujeda, J., Ochando, L. E., Amigó, J. M., Rovira, C., Rius, J. & Veciana, J. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 55–57.

† In response to a referee's comment, we would like to point out that *PowderSolve* can be used to locate the positions of individual atoms (by defining one-atom 'molecular fragments') without prior knowledge of the molecular structure. This approach works successfully for cimetidine (51 degrees of freedom), but fails in many other cases, in which unphysical atomic arrangements with R_{wp} lower than the correct solution are often found.

‡ A report describing the post-deadline structure solution of the Structure Determination by Powder Diffractometry (SDPD) Round Robin sample II (tetracycline hydrochloride) can be found at http://pcb4122.univ-lemans.fr/SDPD/part5.txt.

David, W. I. F., Shankland, K. & Shankland, N. (1998). *J. Chem. Soc. Chem. Commun.* pp. 931–932.

Deem, M. W. & Newsam, J. M. (1989). *Nature (London)*, **342**, 260–262.

Deem, M. W. & Newsam, J. M. (1992). *J. Am. Chem. Soc.* **114**, 7189–7198.

Dinnebier, R. E., Stephens, P. W., Carter, J. K., Lommen, A. N., Heiney, P. A., McGhie, A. R., Brard, L. & Smith, A. B. III (1995). *J. Appl. Cryst.* **28**, 327–334.

Elizabé, L., Kariuki, B. M., Harris, K. D. M., Tremayne, M., Epple, M. & Thomas, J. M. (1997). *J. Phys. Chem. B*, **101**, 8827–8831.

Giacovazzo, C. (1980). *Direct Methods in Crystallography*. London: Academic Press.

Hammond, R. B., Roberts, K. J., Docherty, R. & Edmondson, M. (1997). *J. Phys. Chem. B*, **101**, 6532–6536.

Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (1998). *Acta Cryst.* A**54**, 632–45.

Harris, K. D. M., Johnston, R. L., Kariuki, B. M. & Tremayne, M. (1998). *J. Chem. Res. (S)*, 390–391.

Harris, K. D. M., Kariuki, B. M. & Tremayne, M. (1998). *Mater. Sci. Forum*, **278–291**, 32–37.

Harris, K. D. M. & Tremayne, M. (1996). *Chem. Mater.* **8**, 2554–2570.

Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am Chem. Soc.* **116**, 3543–3547.

Hauptman, H. & Karle, J. (1953). *The Solution of the Phase Problem. I. The Centrosymmetric Crystal. ACA Monograph*, No. 3. New York: Polycrystal Book Service.

Jansen, J., Peschar, R. & Schenk, H. (1992). *J. Appl. Cryst.* **25**, 231–243.

Karfunkel, H. R., Wu, Z. J., Burkhard, A., Rihs, G., Sinnreich, S., Buerger, H. M. & Stanek, J. (1996). *Acta Cryst.* B**52**, 555–561.

Kariuki, B. M., Belmonte, S. A., McMahon, M. I., Johnston, R. L., Harris, K. D. M. & Nelmes, R. J. (1999). *J. Synchrotron Rad.* **6**, 87–92.

Kariuki, B. M., Calcagno, P., Harris, K. D. M., Philp, D. & Johnston, R. L. (1999) *Angew. Chem.* **38**, 831–835.

Kariuki, B. M., Johnston, R. L., Harris, K. D. M., Psallidas, K. P., Ahn, S. & Serrano-González, H. (1998). *Commun. Math. Comput. Chem.* **38**, 123–135.

Kariuki, B. M., Serrano-González, H., Johnston, R. L. & Harris, K. D. M. (1997). *Chem. Phys. Lett.* **280**, 189–195.

Kariuki, B. M., Zin, D. M. S., Tremayne, M. & Harris, K. D. M. (1996). *Chem. Mater.* **8**, 565–569.

Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). *Science*, **220**, 671–680.

Knudsen, K. D., Pattison, P., Fitch, A. N. & Cernik, R. J. (1998). *Angew. Chem. Int. Ed. Engl.* **37**, 2340–2343.

Laarhoven, P. J. M. van & Aarts, E. H. L. (1987). *Simulated Annealing: Theory and Applications*. Amsterdam: D. Riedel.

Le Bail, A. (1993–1999). http://www.cristal.Org/iniref.html.

Le Bail, A., Duroy, H. & Forquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.

Lightfoot, P., Tremayne, M., Glidewell, C., Harris, K. D. M. & Bruce, P. G. (1993). *J. Chem. Soc. Perkin Trans. 2*, pp. 1625–1630.

Newsam, J. M., Deem, M. W. & Freeman, C. M. (1992). *Accuracy in Powder Diffraction II: NIST Spec. Pub.* **846**, 80–91.

Pawley, G. S. (1981). *J. Appl. Cryst.* **14**, 357–361.

Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986). *Numerical Recipes*, pp. 298–300. Cambridge University Press.

Reck, G., Kretschmer, R.-G., Kutschabsky, L. & Pritzkow, W. (1988). *Acta Cryst.* A**44**, 417–421.

Rietveld, H. M. (1969). *J. Appl. Cryst.* **2**, 65–71.

Shankland, K., David, W. I. F. & Csoka, T. (1997). *Z. Kristallogr.* **212**, 550–552.

Shankland, K. & Sivia, D. S. (1996). *Int. J. Pharmaceut.* **139**, 9–14.

Sivia, D. S., David, W. I. F., Knight, K. S. & Gull, S. F. (1993). *Physica D*, **66**, 234–242.

Sun, H. (1998). *J. Phys. Chem. B*, **102**, 7338–7364.

Thomas, J. M., Klinowski, J., Ramdas, S., Hunter, B. K. & Tennakoon, D. T. B. (1983). *Chem. Phys. Lett.* **102**, 158–162.

Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1996*a*). *J. Mater. Chem.* **6**, 1601–1604.

Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1996*b*). *J. Appl. Cryst.* **29**, 211–214.
Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1997). *Angew. Chem. Int. Ed. Engl.* **36**, 770–772.
Tremayne, M., Kariuki, B. M., Harris, K. D. M., Shankland, K. & Knight, K. S. (1997). *J. Appl. Cryst.* **30**, 968–974.
Verwer, P. & Leusen, F. J. J. (1998). *Rev. Comput. Chem.* **12**, 327–365.
Visser, J. W. (1969). *J. Appl. Cryst.* **2**, 89–95.
Werner, P.-E., Eriksson, L. & Westdahl, M. (1985). *J. Appl. Cryst.* **18**, 367–370.
Young, R. A. (1993). *The Rietveld Method*, pp. 9–10. Oxford University Press.