



W&M ScholarWorks

---

Arts & Sciences Articles

Arts and Sciences

---

2010

## PSSP, a computer program for the crystal structure solution of molecular materials from X-ray powder diffraction data

Silvina Pagola  
*William & Mary*, [spagol@wm.edu](mailto:spagol@wm.edu)

Silvina Pagola

Peter W. Stephens

Follow this and additional works at: <https://scholarworks.wm.edu/aspubs>

---

### Recommended Citation

Pagola, S., & Stephens, P. W. (2010). PSSP, a computer program for the crystal structure solution of molecular materials from X-ray powder diffraction data. *Journal of Applied Crystallography*, 43(2), 370-376.

This Article is brought to you for free and open access by the Arts and Sciences at W&M ScholarWorks. It has been accepted for inclusion in Arts & Sciences Articles by an authorized administrator of W&M ScholarWorks. For more information, please contact [scholarworks@wm.edu](mailto:scholarworks@wm.edu).

# *PSSP*, a computer program for the crystal structure solution of molecular materials from X-ray powder diffraction data

Silvina Pagola<sup>a,b,\*</sup> and Peter W. Stephens<sup>c</sup>

<sup>a</sup>Department of Applied Science, College of William and Mary, Williamsburg, VA 23187-8795, USA, <sup>b</sup>Applied Research Center, 12050 Jefferson Avenue, Newport News, VA 23606, USA, and <sup>c</sup>Department of Physics and Astronomy, SUNY at Stony Brook, NY 11794-3800, USA. Correspondence e-mail: spagol@wm.edu

This work describes the computer program *PSSP* (powder structure solution program) for the crystal structure solution of molecular solids from X-ray powder diffraction data. This direct-space structure solution program uses the simulated annealing global optimization algorithm to minimize the difference between integrated intensities calculated from trial models and those extracted in a Le Bail fit of the experimental pattern, using a cost function for dealing with peak overlap through defined intensity correlation coefficients, computationally faster to calculate than  $R_{wp}$ . The methodology outlined is applicable to organic solids composed of moderately complex rigid and flexible molecules, using diffraction data up to relatively low resolution. *PSSP* performance tests using 11 molecular solids with six to 20 degrees of freedom are analyzed.

© 2010 International Union of Crystallography  
Printed in Singapore – all rights reserved

## 1. Introduction

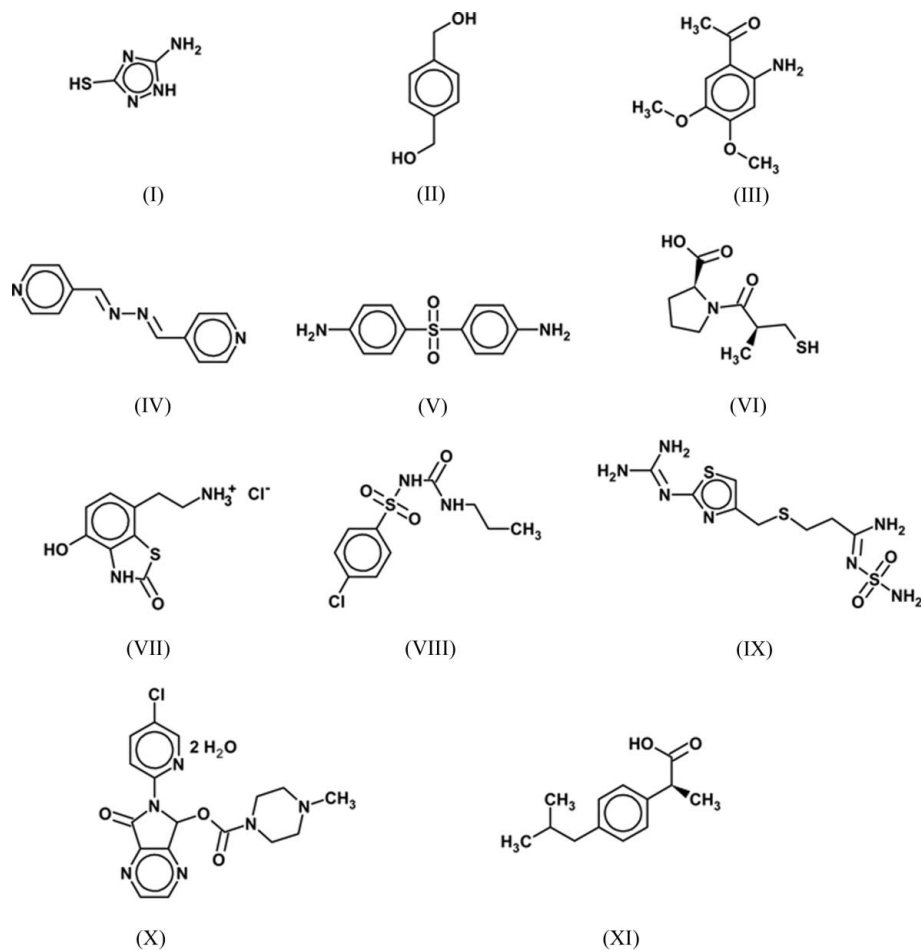
Single-crystal X-ray diffraction is routinely used for crystal structure analysis. However, many materials are only available as micro-crystalline powders, and single crystals of sufficient size and quality for structure determination cannot be prepared. Single-crystal and powder forms may have different properties, or the structural properties of the powder form may be those of main interest. The information contained in a powder diffraction pattern is intrinsically more limited, since the three-dimensional intensity information of single-crystal diffraction data is compressed into one dimension, the  $2\theta$  angle. The overlap of the powder diffraction peaks, depending on the crystal symmetry and the complexity of the structure, makes uncertain the assignment of diffraction intensities to individual Bragg reflections, traditionally used in the determination of crystal structures. Nevertheless, there have been many recent advances in the methodologies used in structure determination from powder diffraction (SDPD), and the complexity of the structures that have been thus elucidated is growing rapidly.

The currently available SDPD methodologies are based on reciprocal-space methods, direct-space methods and, recently, a dual-space method, the charge-flipping algorithm (Oszlányi & Sütő, 2004, 2008). The reciprocal-space methods include algorithms traditionally applied to single-crystal X-ray diffraction (such as Patterson and direct methods) which have been successfully extended to powder diffraction (*e.g.* Altomare *et al.*, 2007). The direct-space approach is based on generating candidate structures and comparing their diffraction patterns with the experimental diffraction data. This is applicable in materials such as inorganic compounds containing coordination polyhedra (Černý, 2006), *e.g.* zeolites (Deem & Newsam, 1992), and in molecular solids (Harris *et al.*, 1994), for which prior structural knowledge significantly reduces the required search space. This permits one to obtain the crystal structure solution from lower-quality intensity data than are generally needed by reciprocal-space methods. There are different realizations of this idea, princi-

pally based on the way that candidate structures are generated and sorted using different algorithms. A recent list of software applying direct-space methodology in SDPD mentions 15 different computer programs (Černý & Favre-Nicolin, 2007).

In this article we describe the computer program *PSSP* (powder structure solution program). The program is only designed to determine the approximate atomic coordinates of molecular solids or other solid types containing moieties of predictable geometry. The user supplies a set of integrated intensities obtained after whole powder pattern decomposition using the Le Bail algorithm (Le Bail *et al.*, 1988; Le Bail, 2005), implemented in computer programs such as *GSAS* (Larson & Von Dreele, 2004), *FULLPROF* (Rodríguez-Carvajal, 1990), *TOPAS* (Coelho, 2007), *BGMN* (Bergmann *et al.*, 1998) and *XND* (Baldinozzi *et al.*, 1999), and the user must carry out a final Rietveld refinement of the model obtained. This paper also describes the structure solution methodology and the results obtained in the structure solution runs of 11 trial molecular compounds of known or unknown crystal structures, with six to 20 degrees of freedom (d.o.f.), shown in Fig. 1. The synchrotron diffraction data of the first three compounds – (I), 3-amino-5-mercapto-1,2,4-triazole (a metabolite of the pesticide amitrole), (II), 1,4-benzenedimethanol (Shan & Jones, 2001), and (III), 2-amino-4,5-dimethoxyacetophenone (a chemical bird repellent added to contaminated water; Clark & Shah, 1993) – have been analyzed previously (Pagola & Stephens, 2000). Additional experimental details are included in the deposited CIF.<sup>1</sup> For testing purposes, we have also used *PSSP* to re-determine the structures of the following compounds (Florence *et al.*, 2005) from laboratory X-ray powder diffraction data shared at <http://www.powderdata.info>: (IV), *N,N'*-bis[1-pyridin-4-ylmeth-(*E*)-ylidene]hydrazine; (V), dapsone; (VI), captopril; (VII), 2-(4-hydroxy-2-oxo-2,3-dihydro-1,3-benzothiazol-7-yl)ethylammonium chloride; (VIII),

<sup>1</sup> Supplementary data for this paper are available from the IUCr electronic archives (Reference: HX5102). Services for accessing these data are described at the back of the journal.



**Figure 1**  
The molecular structures of (I)–(XI).

chlorpropamide; (IX), famotidine; (X), zopiclone dihydrate; and (XI), *S*-ibuprofen.

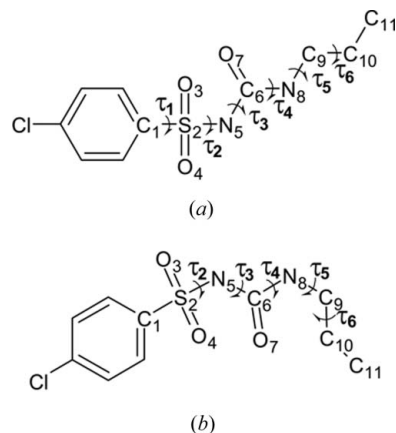
## 2. Methodology

For organic molecular solids, a large amount of structural information, *e.g.* bond lengths and angles, can be predicted from the chemical structure of the molecule. Accordingly, the crystal structure of a molecular solid can be described by specifying the lattice parameters and space-group symmetry, the number, location and orientation of the molecules in the unit cell, and the molecular conformations. Under the assumption that bond lengths and angles are known, the molecular conformation is specified by a few torsion angles ( $\tau_i$ ) about single bonds. For a solid composed of rigid molecules, the molecular locations may be specified as three fractional coordinates ( $x_m, y_m, z_m$ ) of some reference point on the molecule in the unit cell, and the molecular orientations as three Eulerian angles ( $\varphi, \chi, \omega$ ) of rotation about that reference point, with respect to some chosen axes. In many cases, the asymmetric unit of low-symmetry organic solids will be described by one molecule located at a general position of the space group under study. This may be extended to the structures of salts, where the counter-ion must be independently located, and to structures with more than one identical molecule in inequivalent crystallographic sites ( $Z' > 1$ ).

As long as the prior information is correct, all chemically possible structures can be defined with a few parameters,  $\{P_i\} = (x_m, y_m, z_m, \varphi, \chi, \omega, \tau_1, \dots, \tau_i)$  for a single molecule in the irreducible cell. An example of torsion angle definitions is shown in Fig. 2. The correct structure can then be determined by searching the parameter space and finding the values that give the best agreement between the experimental diffraction data and the calculated diffraction pattern. For the case of a molecule of  $N$  atoms, this is a much smaller space of possible solutions than the  $3N$  dimensions that would be required if no chemical information were included. However, for the examples given here, it is still a space of six to 20 dimensions and it is relatively easy to find problems having higher complexity, especially when there is more than one independent molecule in the asymmetric unit ( $Z' > 1$ ).

The simulated annealing algorithm is based on analogies between classical statistical mechanics (behavior of systems with many degrees of freedom in thermal equilibrium) and a global optimization problem (to find the extremum of a function that depends on many parameters; Kirkpatrick *et al.*, 1983). In this case, it is applied to find the minimum in the differences between the calculated and observed diffraction intensities. To apply it to SDPD, it is necessary to define the following:

- (i) A generation mechanism that postulates trial structures by modifying a given configuration. In the present case, this consists of changing one or more of the structural parameters  $\{P_i\}$ .
- (ii) A cost function that ranks the quality of each configuration (analogous to the energy of a given configuration of a physical system). In the present case, this is a numerical measure  $S$ , defined in



**Figure 2**  
In order to search the intramolecular degrees of freedom for the molecule of (VIII) (chlorpropamide), (a) atoms O<sub>3</sub>, O<sub>4</sub>, N<sub>5</sub>, C<sub>6</sub>, O<sub>7</sub>, N<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub> and C<sub>11</sub> are rotated  $\tau_1^\circ$  around an axis with a direction coincident with that of the C<sub>1</sub>–S<sub>2</sub> bond, and (b) atoms C<sub>6</sub>, O<sub>7</sub>, N<sub>8</sub>, C<sub>9</sub>, C<sub>10</sub> and C<sub>11</sub> are rotated  $\tau_2^\circ$  around an axis with a direction coincident with the S<sub>2</sub>–N<sub>5</sub> bond, and so on.

§2.3, of the agreement between the computed diffraction pattern of the candidate structure and the experimental one.

(iii) A control parameter (analogous to the temperature) that determines the fraction of unfavorable local steps accepted in the search.

Simulated annealing is essentially a series of searches applying the Metropolis criterion (Metropolis *et al.*, 1953) at decreasing values of the control parameter or temperature (Aarts & Korst, 1989). If carried out with enough trial structures at each temperature, it is likely that a solution with a low value of the cost function will be found. Simulated annealing has been widely used in SDPD and has had the largest impact (David & Shankland, 2008).

Among the crystal structures solved with *PSSP* can be mentioned the malaria pigment (Pagola *et al.*, 2000); the disordered structure of ranitidine hydrochloride (Huq & Stephens, 2003); a new isolated natural product involving the stereochemistry determination of four chiral centers with reference to a steroidal fragment of known chirality (García *et al.*, 2009); and amcinonide, a glucocorticoid with an uncommonly large number of atoms in the asymmetric unit (142, including H; S. Pagola & P. W. Stephens, unpublished results).

### 2.1. Generation of candidate solutions

The generation mechanism must be a compromise between using the partial success from the previous trial and a willingness to move elsewhere in the configuration space to avoid false minima. In this work, each trial configuration is described as a set of parameters  $\{P_i\}$ , and a trial step is defined by a series of increments  $\{\Delta_i\}$ . Each parameter  $P_i$  has a natural range  $R_i$ , generally  $0-360^\circ$  for Eulerian angles and unrestricted rotations, and  $0-1$  for fractional coordinates, with periodic boundary conditions. Small steps allow for the improvement of a nearly correct structure, but room must be left for major reconfigurations. We experimented with several algorithms and settled upon alternations between the following two steps: (i) change all parameters together by random amounts within their allowed range, according to a power-law distribution that favors small steps, and (ii) choose one of the parameters  $P_i$  at random, and change it to a random value anywhere within its domain. Additional details are provided in the supplementary information. The program keeps the parameters of the trial model with the lowest value of the cost function ( $S_{\min}$ ) obtained at any determined temperature and uses them as the starting point in the search under the next decreased temperature.

*PSSP* generates its trial structures by starting with a chemically acceptable molecular geometry in Cartesian coordinates (ångströms), furnished by the user. Torsions are applied relative to this starting configuration, by rotating a specified group of atoms about a direction coincident with that of the bond between two atoms. This may be illustrated by the example of chlorpropamide (see Fig. 2). One may perform these rotations in any order.

The entire molecule is then rotated by the three Eulerian angles ( $\varphi$ ,  $\chi$ ,  $\omega$ , generally adjustable parameters) and transformed into crystallographic fractional coordinates with its origin shifted to a point, generally specified by three further adjustable parameters ( $x_m$ ,  $y_m$ ,  $z_m$ ). This process is repeated for each molecular fragment that might be in the irreducible cell. A single atom or ion is the particular case of a fragment not rotated, merely set at a location within the unit cell. For a molecular fragment with mirror, inversion or rotational symmetry located at a special position of the space group, one can apply suitable restricted rotation and translation parameters, or modify the fragment accordingly. The program in its current form is

not able to move a molecule from one special position to another. Further details are given in the supplementary material.

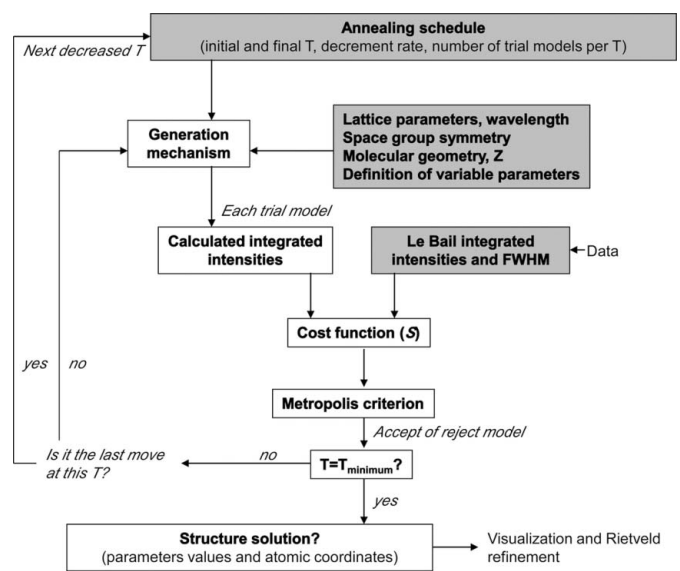
### 2.2. Program sequential operations

Fig. 3 is a block diagram of the operations carried out by *PSSP* in sequential order. The shaded blocks contain the information provided to the program through a control file. This includes unit-cell parameters, radiation wavelength, space-group symmetry operations, specification of the asymmetric unit contents (the geometry of the molecules in Cartesian coordinates and their number), a list of the structural parameters to be searched and their variation ranges, a pointer to the experimental data file (Le Bail fit), with integrated intensities and FWHM required for the calculation of the cost function  $S$  (defined in the following section), and an annealing schedule (initial temperature, decrement rate, final temperature and number of trial structures evaluated at each temperature). Each of these inputs is described in detail in the supplementary material. The search is started at the initial temperature, after assigning random numbers to the set of structural parameters. For each trial model evaluated, the calculated integrated intensities are compared with the Le Bail ones, using the cost function  $S$ , and the Metropolis algorithm (Metropolis *et al.*, 1953) is applied. After the evaluation of all trial models required at the current temperature, the search is re-started at a decreased temperature from the best model so far obtained, and the process is repeated until the minimum temperature is reached.

An idea of the results of a *PSSP* run is given in Fig. 4, showing the increasing agreement between Le Bail and trial model integrated intensities of (V) (dapsone) at four selected decreasing values of the temperature parameter ( $T$ ).

### 2.3. Treatment of overlapping diffraction peaks

The overlap of nearby powder diffraction peaks is a ubiquitous problem in the extraction of intensities. The effectiveness of direct-space methods comes in part from the fact that it is not necessary to partition a measured pattern into individual peak intensities, but rather a computed profile is compared with the experimental powder



**Figure 3** Schematic procedure of structure solution from X-ray powder diffraction data using *PSSP*. The grey blocks contain the information entered in to the program through the input file.

diffraction pattern. The computed profile should have peak shapes as close as possible to the measured pattern. However, it is computationally expensive to generate a full powder pattern for each candidate structure. David *et al.* (1998, 2006) have addressed this problem by using the correlation coefficients in a Pawley fit (each intensity varied as a separate least-squares parameter). We have developed a different approach which can work from the Le Bail fits (Le Bail *et al.*, 1988; Le Bail, 2005).

In general, it will be necessary to use data with substantial overlap of nearby peaks, and so the Le Bail reduction of the powder profile to a list of integrated intensities appears to reduce the information content substantially. We have approached this problem by assuming that the Le Bail fitted profile is an accurate representation of the data, even though its extracted intensities may not have apportioned the intensity correctly to each reflection. One can calculate an agreement factor between the trial model and Le Bail profiles without further loss of information. We define

$$S = \int d2\theta \frac{[I_{\text{Le Bail}}(2\theta) - I_{\text{model}}(2\theta)]^2}{[I_{\text{Le Bail}}(2\theta)]^2}, \quad (1)$$

where  $I_{\text{Le Bail}}(2\theta)$  and  $I_{\text{model}}(2\theta)$  are the calculated profiles of the Le Bail fit and the structural model, respectively. (We write this as an integral over continuous  $2\theta$ , even though powder data are most often collected in discrete steps.) This looks like a weighted profile  $R$  factor, with unit weights for all data and with a particular normalization.  $S$  can be efficiently computed without constructing the entire profile, since the shape of each reflection has been determined in the Le Bail fit. Specifically, let  $f_{hkl}(2\theta)$  be the profile function for the  $hkl$  reflection, normalized to unit integral, and let  $A_{hkl}$  and  $B_{hkl}$  be the inte-

grated intensities of the  $hkl$  reflection from the Le Bail fit and the model under test, respectively. Then  $S$  can be expressed as

$$S = \frac{\int d2\theta [\sum_{hkl} A_{hkl} f_{hkl}(2\theta) - \sum_{hkl} B_{hkl} f_{hkl}(2\theta)]^2}{[\int d2\theta \sum_{hkl} A_{hkl} f_{hkl}(2\theta)]^2}. \quad (2)$$

This can be computed as

$$S = \frac{\sum_{hkl,h'k'l'} (A_{hkl} - B_{hkl}) F_{hkl,h'k'l'} (A_{h'k'l'} - B_{h'k'l'})}{(\sum_{hkl} A_{hkl})^2} \quad (3)$$

by making use of the overlap coefficients  $F_{hkl,h'k'l'}$  defined as

$$F_{hkl,h'k'l'} = \int d2\theta f_{hkl}(2\theta) f_{h'k'l'}(2\theta). \quad (4)$$

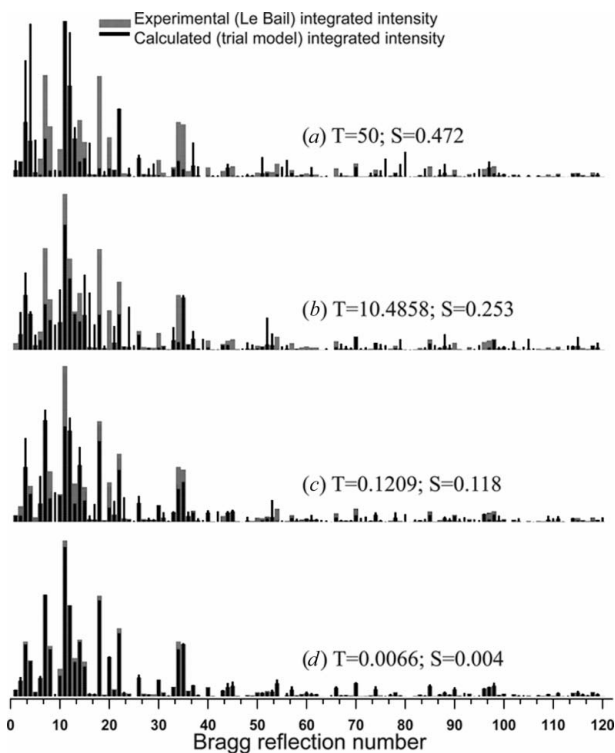
There are several noteworthy points about this formulation.  $F_{hkl,h'k'l'}$  is large only for closely separated peaks, and so the sum in the numerator contains only a few times more terms than the number of reflections under consideration.  $F_{hkl,h'k'l'}$  can be calculated once and tabulated at the beginning of a search for a structure solution, to speed the calculation of the  $S$  factor for a candidate solution. In our work, we have limited the consideration of overlapping peaks to a band of five to ten distinct  $(h'k'l')$  values above each  $(hkl)$ . For simplicity, we have used an analytical approximation based on treating the overlapping peaks as if they were pure Gaussians, so that

$$F_{hkl,h'k'l'} = \frac{1}{(2\pi)^{1/2} \sigma} \exp\left[-\frac{(2\theta_{hkl} - 2\theta_{h'k'l'})^2}{2\sigma^2}\right], \quad (5)$$

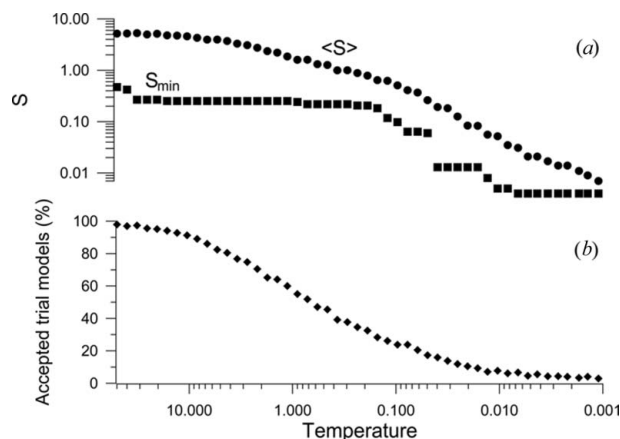
where

$$\sigma^2 = \frac{\text{FWHM}_{hkl}^2 - \text{FWHM}_{h'k'l'}^2}{8 \ln 2}. \quad (6)$$

These approximations may shift the estimate of  $S$  calculated using equations (3) and (5) from its correct value obtained in equation (1), but they do not appear to degrade the ability of the algorithm to find a satisfactory structure solution. Equation (3) is similar to the formulation by David *et al.* (1998), based on a least-squares (Pawley) fit of all integrated intensities in which the place of the overlap coefficients is taken by elements of the correlation matrix. In both cases, the formulation makes use of all the information in a profile, even if it has been reduced to estimates of the integrated intensities and the overlap or correlation coefficients (David, 2004).



**Figure 4** The first 120 experimental Le Bail (grey) and trial model (black) integrated intensities for the crystal structure solution of (V) (dapsone) and the lowest  $S$  factors obtained at (a)  $T = 50$ , (b)  $T = 10.4858$ , (c)  $T = 0.1209$  and (d)  $T = 0.0066$ .



**Figure 5** (a)  $S$  and  $S_{\text{min}}$ , and (b) percentage of accepted trial models versus temperature (at logarithmic scale), for a simulated annealing run of (V) (dapsone).

**Table 1**

External and internal molecular degrees of freedom (d.o.f.),  $\chi^2_{Le Bail}$ , minimum and maximum  $S$  factors of solutions, and  $\chi^2/\chi^2_{Le Bail}$  after ten cycles of Rietveld refinement, for the models counted as structure solutions of compounds (I)–(III).

Compound	d.o.f.		$\chi^2_{Le Bail}$	$S$	$\chi^2/\chi^2_{Le Bail}$
	External	Internal			
(I)	6	0	1.59	0.028–0.500	1.68–9.37
(II)	6	2	1.54	0.034–0.080	3.06–10.0
(III)	6	3	2.22	0.046–0.140	4.51–8.77

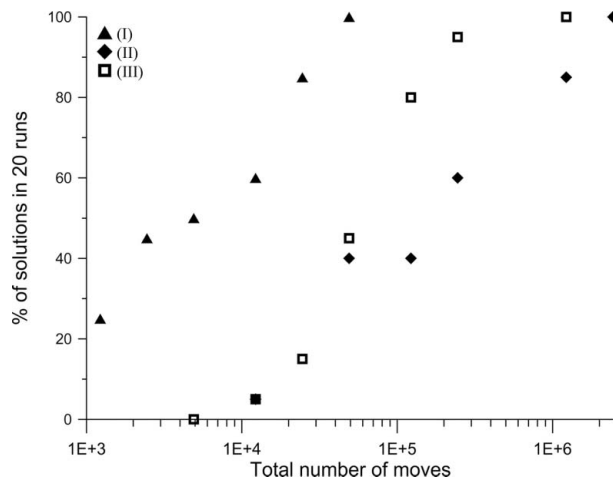
**2.4. Simulated annealing schedules**

General annealing schedule considerations are given by Aarts & Korst (1989). In this specific application of the simulated annealing algorithm, the following parameters gave rise to an adequate annealing schedule. The initial temperature value used is typically 50, so under this condition almost all moves are accepted. The decrement factor for the following temperatures is typically 0.8, thus

$$T_{+1} = 0.8T. \tag{7}$$

The final temperature value depends on the structural complexity. Whereas a final temperature of  $\sim 0.01$  may be sufficient to obtain refinable solutions for simple molecules, 0.001 may be required to solve more complex structures. For simple molecules even the first 40 reflections have been successfully used for molecular location, whereas 150 reflections or more can be necessary to solve molecular solids of larger complexities. Fig. 5 shows the progress of a typical simulated annealing run, through plots of (a) average and minimum  $S$  values versus  $T$ , and (b) the percentage of structures accepted versus temperature.

Since the simulated annealing algorithm is an approximation and convergence to a global minimum is asymptotic, it is not always possible to obtain the crystal structure solution, and so in general several independent runs are necessary. It is therefore worthwhile to develop an empirical understanding of the speed and efficiency in finding the solution with which the program works. An additional way of characterizing our simulated annealing application to SDPD is outlined in the following sections, presenting *PSSP* performance tests on 11 molecular solids of six to 20 d.o.f.



**Figure 6** Percentage of correct crystal structure solutions obtained in 20 runs versus the total number of moves used per run (at logarithmic scale) for compounds (I)–(III).

**Table 2**

External and internal molecular degrees of freedom (d.o.f.),  $\chi^2_{Le Bail}$ , minimum and maximum  $S$  factors of structure solutions, and  $\chi^2/\chi^2_{Le Bail}$  after ten cycles of Rietveld refinement, for the models counted as structure solutions of compounds (IV)–(XI).

Compound	d.o.f.		$\chi^2_{Le Bail}$	$S$	$\chi^2/\chi^2_{Le Bail}$
	External	Internal			
(IV)	6	1	2.48	0.016–0.027	4.04–8.80
(V)	6	2	3.33	0.003–0.025	2.11–8.14
(VI)	6	4	3.30	0.009–0.030	3.59–9.62
(VII)	6 + 3	2	1.70	0.018–0.025	4.48–7.56
(VIII)	6	6	4.77	0.011–0.023	5.44–9.98
(IX)	6	7	2.02	0.008–0.010	2.35–3.75
(X)	6 + 3 + 3	4	2.80	0.015–0.024	4.54–8.50
(XI)	6 + 6	4 + 4	2.13	0.009–0.013	7.66–9.97

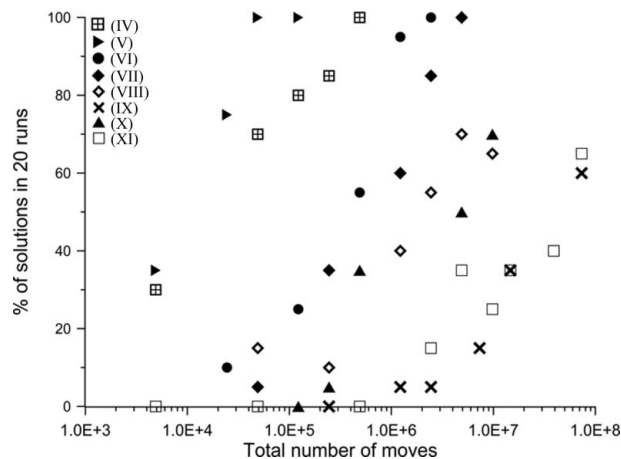
**3. PSSP performance tests**

**3.1. Compounds (I)–(III)**

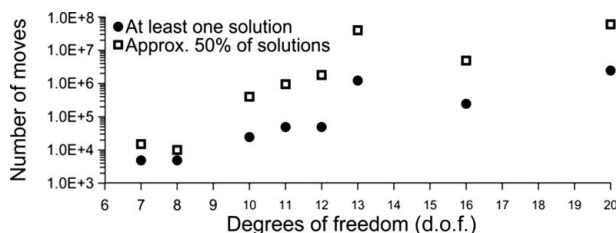
The synchrotron patterns of compounds (I)–(III) (d.o.f. = 6–9) were indexed and the space groups determined, as reported by Pagola & Stephens (2000). More recently, Le Bail fits were carried out using *GSAS* (Larson & Von Dreele, 2004). The molecular geometries were optimized using the program *ArgusLab* (Thompson, 2004), using semiempirical methods. For (I), (II) and (III), the first 120, 126 and 102 reflections, respectively, were used for structure solution.

Refinable solutions show, as a guideline, values of  $S < 0.1$ . In analogy with the previously reported work on compounds (IV)–(XI) (Florence *et al.*, 2005), the models for which  $\chi^2/\chi^2_{Le Bail} < 10$  after ten cycles of Rietveld refinement with *GSAS* (Larson & Von Dreele, 2004), with only the background coefficients and the scale factor refined, were counted as structure solutions. The models were examined on an individual basis using the program *Mercury* (Version 1.4.2; Macrae *et al.*, 2006). Table 1 and Fig. 6 summarize the results for these runs.

The lowest number of total moves,  $\sim 4900$  ( $\sim 100$  models per temperature using a typical annealing schedule with 49 temperatures), was required to solve the structure of the rigid molecule (I) in 50% of the runs, whereas the more flexible (II) and (III) required of the order of 60 000–200 000 moves (around 1250–4000 models per temperature). For (I), 1000 moves per  $T$  were sufficient to solve the structure in all runs. A few models in each run for which  $\chi^2/\chi^2_{Le Bail} < 10$  were found to be a different packing, and the range of  $S$  values of the



**Figure 7** Percentage of correct crystal structure solutions obtained in 20 runs versus the total number of moves used per run (at logarithmic scale) for compounds (IV)–(XI).



**Figure 8**

Estimated minimum number of moves required to produce at least one refinable structure solution in 20 runs, and to solve the structure in around 50% of the runs, versus the molecular degrees of freedom for compounds (IV)–(XI).

correct solutions is rather large (see Table 1). The corresponding values of this  $S$  range, when analyzed for individual runs, decrease when the total number of moves in the run increases.

In all cases, Rietveld refinements were completed using *GSAS* (Larson & Von Dreele, 2004). Details of the refinements and crystal structures of (I), (II) and (III) are included in the deposited CIF.

### 3.2. Compounds (IV)–(XI)

The molecular geometries of compounds (IV)–(XI) were taken from the respective CIFs, as cited by Florence *et al.* (2005). The first 103, 120, 66, 108, 80, 121, 113 and 122 reflections were used in the structure solutions of (IV)–(XI), respectively. Le Bail fits of their laboratory patterns were calculated with *GSAS* (Larson & Von Dreele, 2004), in some cases using microstrain broadening profile parameters (Stephens, 1999).

The efficiency of *PSSP* in finding the structure solutions of these solids with seven to 20 d.o.f. is depicted in Fig. 7 and summarized in Table 2. The models for which  $\chi^2/\chi_{Le\ Bail}^2 < 10$  after ten cycles of Rietveld refinement with *GSAS* (Larson & Von Dreele, 2004), with the background coefficients and the scale factor refined, were counted as structure solutions. All solutions counted showed values of  $S < 0.030$ . This relatively low value is likely due to the use of more accurate molecular geometries in the structure solution search than for (I)–(III).

Fig. 7 shows that around 1000–2500 moves per temperature (49 000–122 500 trial models for the typical annealing schedule described above using 49 temperatures) can sometimes be sufficient to solve in all runs the structures of compounds containing one rigid fragment and small bonded groups subjected to torsions, such as (V). A first attempt to solve the structure of (IV) indicated that the molecules were located over inversion centers. The asymmetric unit was then modified to describe only half the molecule, and (IV) was solved using one internal d.o.f. without restrictions in the ‘half molecule’ positions search. Compounds (VII) and (VIII) (d.o.f. = 11 and 12, respectively) required around 700 000 and  $2 \times 10^6$  trial models in total, respectively, to obtain the crystal structure solution in 50% of the runs. *PSSP* found the solution of the very flexible molecule (IX) (d.o.f. = 13) in 60% of the runs with  $7.35 \times 10^7$  moves, without using modal torsional constraints (Florence *et al.*, 2005, and citations therein). *S*-Ibuprofen [(XI), d.o.f. = 20, two independent molecules in the asymmetric unit] was solved in 35% of cases with  $1.47 \times 10^7$  moves (300 000 trial models per temperature), 40% with  $3.92 \times 10^7$  moves (800 000 models per  $T$ ) and 65% of the runs with 1 500 000 models per  $T$ .

This information is also represented in Fig. 8, showing (with logarithmic scale) the minimum numbers of moves required to obtain at least one refinable solution in 20 runs, and the estimated number of trial models required to obtain the solution in around 50% of the

runs, resulting in an approximate exponential dependence with the d.o.f., as reported by other research groups (Favre-Nicolin & Černý, 2002).

### 4. Computational requirements, availability and documentation

The *PSSP* source code is written in C++ for Windows and Linux platforms. *PSSP* is available from the authors upon request at spagol@wm.edu. The guidelines required to write input files and to run the program are included in the supplementary material.

At present, crystal structure solutions can be obtained in times of the order of less than one minute to a few hours, depending on computer calculation capabilities and crystal structure complexity. As examples, some typical computer times necessary to solve structures using an Intel Core2 Dual processor, 2.00 GHz each, 2038 MB RAM memory, running under the Windows Vista operating system, are as follows: to solve (IV), each run of 2500 moves per temperature with 103 reflections takes around 13 s; for (XI), a run of 200 000 moves per temperature and 122 reflections takes approximately 1 h.

This research was performed at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (contract No. DE-AC02-98CH10886). SP thanks the Physics Department, College of William and Mary, for funding.

### References

- Aarts, E. & Korst, J. (1989). *Simulated Annealing and Boltzmann Machines*. New York: J. Wiley and Sons.
- Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2007). *J. Appl. Cryst.* **40**, 344–348.
- Baldinozzi, G., Grebille, D. & Bézar, J.-F. (1999). *Proceedings of Aperiodic '97*, pp. 297–301. Singapore: World Scientific.
- Bergmann, J., Friedel, P. & Kleeberg, R. (1998). *IUCr Commission on Powder Diffraction Newsletter*, No. 20, pp. 5–8.
- Černý, R. (2006). *Croat. Chem. Acta*, **79**, 319–326.
- Černý, R. & Favre-Nicolin, V. (2007). *Z. Kristallogr.* **222**, 105–113.
- Clark, L. & Shah, P. S. (1993). *J. Wildl. Manage.* **57**, 657–664.
- Coelho, A. (2007). *TOPAS-Academic*. Version 4.1, Coelho Software, Brisbane, Australia, <http://www.topas-academic.net>.
- David, W. I. F. (2004). *J. Appl. Cryst.* **37**, 621–628.
- David, W. I. F. & Shankland, K. (2008). *Acta Cryst.* **A64**, 52–64.
- David, W. I. F., Shankland, K. & Shankland, N. (1998). *Chem. Commun.* pp. 931–932.
- David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2006). *J. Appl. Cryst.* **39**, 910–915.
- Deem, M. W. & Newsam, J. M. (1992). *J. Am. Chem. Soc.* **114**, 7189–7198.
- Favre-Nicolin, V. & Černý, R. (2002). *J. Appl. Cryst.* **35**, 734–743.
- Florence, A. J., Shankland, N., Shankland, K., David, W. I. F., Pidcock, E., Xu, X., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D. & Frampton, C. S. (2005). *J. Appl. Cryst.* **38**, 249–259.
- García, M. E., Pagola, S., Navarro-Vázquez, A., Phillips, D. D., Gayathri, C., Krakauer, H., Stephens, P. W., Nicotra, V. E. & Gil, R. R. (2009). *Angew. Chem. Int. Ed.* **48**, 5670–5674.
- Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543–3547.
- Huq, A. & Stephens, P. W. (2003). *J. Pharm. Sci.* **92**, 244–249.
- Kirkpatrick, S., Gellatt, C. D. Jr & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
- Larson, A. C. & Von Dreele, R. B. (2004). *GSAS*. Report LAUR 86-748. Los Alamos National Laboratory, New Mexico, USA.
- Le Bail, A. (2005). *Powder Diffr.* **20**, 316–326.
- Le Bail, A., Duroy, H. & Fourquet, J. L. (1988). *Mater. Res. Bull.* **23**, 447–452.
- Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M. & van de Streek, J. (2006). *J. Appl. Cryst.* **39**, 453–457.

- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). *J. Chem. Phys.* **21**, 1087–1092.
- Oszlányi, G. & Sütő, A. (2004). *Acta Cryst.* **A60**, 134–141.
- Oszlányi, G. & Sütő, A. (2008). *Acta Cryst.* **A64**, 123–134.
- Pagola, S. & Stephens, P. W. (2000). *Mater. Sci. Forum*, **321–324**, 40–45.
- Pagola, S., Stephens, P. W., Bohle, D. S., Kosar, A. D. & Madsen, S. K. (2000). *Nature (London)*, **404**, 307–310.
- Rodríguez-Carvajal, J. (1990). *FULLPROF*. Abstracts of the Satellite Meeting on Powder Diffraction of the XV Congress of the IUCr, p. 127, Toulouse, France.
- Shan, N. & Jones, W. (2001). *Acta Cryst.* **E57**, o1233–o1234.
- Stephens, P. W. (1999). *J. Appl. Cryst.* **32**, 281–289.
- Thompson, M. A. (2004). *ArgusLab*. Version 4.0.1. Planaria Software LLC, Seattle, Washington, USA, <http://www.arguslab.com>.