W&M ScholarWorks

Arts & Sciences Articles

Arts and Sciences

2-2018

WinPSSP: a revamp of the computer program PSSP and its performance solving the crystal structures of small organic compounds and solids of biological and pharmaceutical interest

Silvina Pagola
College of William and Mary, spagol@wm.edu

Alekos Polymeros

Nikolaos Kourkoumelis

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

WinPSSP: a revamp of the computer program PSSP and its performance solving the crystal structures of small organic compounds and solids of biological and pharmaceutical interest

Silvina Pagola,[a]* Alekos Polymeros[b] and Nikolaos Kourkoumelis[c]

[a]Department of Applied Science, College of William and Mary – Applied Research Center, 12050 Jefferson Avenue, Newport News, VA 23606, USA, [b]Department of Physics, University of Ioannina, Ioannina 45110, Greece, and [c]Department of Medical Physics, University of Ioannina, Ioannina 45110, Greece. *Correspondence e-mail: spagol@wm.edu

The direct-space methods software *Powder Structure Solution Program* (*PSSP*) [Pagola & Stephens (2010). *J. Appl. Cryst.* **43**, 370–376] has been migrated to the Windows OS and the code has been optimized for fast runs. *WinPSSP* is a user-friendly graphical user interface that allows the input of preliminary crystal structure information, integrated intensities of the reflections and FWHM, the definition of structural parameters and a simulated annealing schedule, and the visualization of the calculated and experimental diffraction data overlaid for each individual solution. The solutions are reported as `filename.cif` files, which can be used to analyze packing motifs and chemical bonding, and to input the atomic coordinates into the Rietveld analysis software *GSAS*. *WinPSSP* performance in straightforward crystal structure determinations has been evaluated using 18 molecular solids with 6–20 degrees of freedom. The free-distribution program as well as multimedia tutorials can be accessed at http://users.uoi.gr/nkourkou/winpssp/.

1. Introduction

Even though the number of organic and organometallic crystal structures solved from X-ray powder diffraction (XRPD) deposited in the Cambridge Structural Database (Allen, 2002) has not yet reached 1% (Shankland *et al.*, 2013), direct-space methods (Černý & Favre-Nicolin, 2007; Harris, 2012) have fundamentally changed the scope of the powder diffraction technique, affording crystal structure determination. Their development started during the late 1980s (Deem & Newsam, 1989); however, they only earned full credibility much more recently (Lapidus *et al.*, 2010). This is evidenced by the growing number of crystal structures published, estimated as around 200 per year in 2009 (Le Bail *et al.*, 2009), and the availability of specialized software packages contributing to their widespread use. Among those packages are *FOX* (Favre-Nicolin & Černý, 2002), *EXPO2013* (Altomare *et al.*, 2013), *DASH* (David *et al.*, 2006), *TOPAS* (Coelho, 2000), *ESPOIR* (Le Bail, 2001), *POWDER SOLVE* (Engel *et al.*, 1999) and *ENDEAVOUR* (Putz *et al.*, 1999).

Crystal structure determination from powders has been reviewed in various articles (Louër & Langford, 1996; Harris & Tremayne, 1996; Černý & Favre-Nicolin, 2007; David & Shankland, 2008; Harris, 2012). For organic solids in particular, the applications to the economically important pharmaceutical solids (and their salts, hydrates, solvates, polymorphs and co-crystals) have occupied a prominent
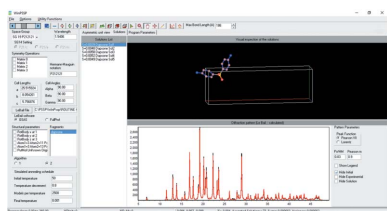
position, leading to further development of this research field (Harris & Cheung, 2003; Datta & Grant, 2004; Stephenson, 2005; Shankland *et al.*, 2013). However, crystal structure determination from powders still remains a rather time-consuming and not well automated task in comparison with its single-crystal diffraction counterpart (Le Bail *et al.*, 2009).

The feasibility of solving a crystal structure from its powder diffraction pattern considerably depends on the quality of the diffraction data, the chemical composition and the structural complexity of the material studied. Using direct methods for the analysis of single-crystal X-ray diffraction data, the only requested *a priori* information about the material is its chemical formula (Altomare *et al.*, 2007). However, these methods commonly fail for crystal structure solution of organic solids (mainly composed of C, H, N and O) from powders, in particular if high-quality diffraction data at small *d* spacings (around 1 Å) are not available. The presence of strong X-ray scatterers often facilitates crystal structure solution from powders using direct methods, or a combination of the above and direct-space approaches, since the heavy atoms are relatively easy to locate. However, organic solids are optimal candidates for the direct-space approach (Harris, 2012) owing to the known atom connectivity and shape of organic molecules, providing additional structural information which can be used to reconstruct the three-dimensional crystal structure from the powder pattern. An important advantage is that direct-space methods generally do not require high-quality diffraction data up to $d \simeq 1$ Å, which often are not available for various reasons, such as low specimen crystallinity or poor counting statistics.

The direct-space methods software *PSSP* (Pagola & Stephens, 2010) was written for the crystal structure determination of the malaria pigment (Stephens *et al.*, 2000) and the analysis of synchrotron XRPD data collected at the National Synchrotron Light Source (NSLS-I), Brookhaven National Laboratory, USA. *PSSP* has been successfully used for the crystal structure determination of various test organic solids (Pagola & Stephens, 2000, 2010), organic pigments (Pagola *et al.*, 2001), pharmaceutical solids including salts and hydrates (Huq & Stephens, 2003; Botez *et al.*, 2003; Nunes *et al.*, 2004; Sperandeo *et al.*, 2005), natural products (Pagola *et al.*, 2003; Pagola, Tracana *et al.*, 2008; García *et al.*, 2009), charge transfer salts of tetrathiafulvalene (Lapidus *et al.*, 2014; Mohamud *et al.*, 2016), and various small molecule organic materials whose diffraction data and atomic coordinates have been submitted to ICDD for inclusion in the PDF-4 Organics (http://www.icdd.com/products/pdf4-organics.htm) database (Pagola & Stephens, 2009, 2012). *PSSP* has also been used for the completion of the crystal structures of a metal–organic framework (Pagola, Pike *et al.*, 2008) and a ternary transition metal oxide (unpublished results).

2. Description of the new graphical user interface and user guidelines

The main stages in solving a crystal structure from powder diffraction data are described in several articles (Harris & Tremayne, 1996; Chernyshev, 2001) and they are enumerated below: (1) preparation and pre-treatment of specimens, (2) determination of the unit-cell parameters and space group symmetry, (3) powder pattern decomposition into integrated intensities of the Bragg reflections (though this is sometimes avoided), (4) search for the structural motif (approximate atomic coordinates), (5) structure completion, and (6) Rietveld refinement.

PSSP is used for finding approximate atomic coordinates of small-molecule organic solids, or materials containing fragments of known atom connectivity and predictable shape. *PSSP* was written in the C++ programming language; it was initially developed in the Linux OS and later in the Windows OS running in an MS-DOS prompt. One of the objectives of the present work has been to adapt this software for use by undergraduate students. Thus, this publication reports the development of a user-friendly graphical user interface (*WinPSSP*) to migrate *PSSP* to modern Windows-based operating systems. Another objective is to provide additional 'user friendliness' (Le Bail *et al.*, 2009) in the form of user guidelines and instructional multimedia towards overcoming the disadvantages in automation that characterize crystal structure solution from powders.

An overview of an ideal application of *PSSP* for crystal structure determination and the modifications implemented in *WinPSSP* are summarized in Fig. 1. The main stages of structure solution are discussed in more detail as follows.

(1) *Preparation and pre-treatment of specimens*. Crystal structure determination from powders is greatly facilitated by using high-quality powder diffraction data. This means a small FWHM of the peaks, good counting statistics, small 2θ zero error, sample shift and transparency errors, and limited preferred orientation effects. Inspection of the sample under a microscope can show large plates or needles, giving rise to preferred orientation effects in the powder pattern. These should be considerably diminished or avoided by sample preparation techniques or by using transmission geometry combined with continuous sample spinning (McCusker *et al.*, 1999). Laboratory XRPD data with good counting statistics from a carefully prepared specimen are often sufficient for crystal structure determination. Furthermore, high-resolution synchrotron XRPD stations afford highly intense and monochromatic X-rays, low instrumental contribution to the peak widths, and thus diminished peak overlap, facilitating crystal structure determination and allowing the elucidation of more complex structures (Le Bail *et al.*, 2009).

(2) *Determination of the unit-cell parameters and space group symmetry*. Prior to the start of the atomic coordinate search, the unit-cell parameters and crystal system [(1) in Fig. 1] must be determined using indexing software such as *McMAILLE* (Le Bail, 2004), *ITO* (Visser, 1969), *TREOR* (Werner *et al.*, 1985), *N-TREOR* (Altomare *et al.*, 2000) or *DICVOL* (Boultif & Louër, 2004, 2014; Louër & Boultif, 2006). Then the space group symmetry [(3) in Fig. 1] is determined from the observation of the systematic absences in an initial Le Bail fit (Le Bail, 2005) using a space group without extinctions in the crystal system found (*e.g.* $P\bar{1}$ for

triclinic cases, *P*2 for monoclinic and *P*222 for orthorhombic structures). This is less cumbersome than it may seem to unexperienced users, since translation, inversion, twofold screw axes and glide planes are the sole symmetry operators generating close packing in organic solids (Gavezzotti, 1998). Thus, organic materials most often crystallize in one of five space groups (Brock & Dunitz, 1994): $P2_1/c$ (No. 14), $P\bar{1}$ (No. 2), $P2_1$ (No. 4), $P2_12_12_1$ (No. 19) and $C2/c$ (No. 15), corresponding to the monoclinic, triclinic and orthorhombic crystal systems.

The '*Reflection conditions, diffraction symbols and possible space groups*' tables (Table 3.1.4.1) for space group determination in ch. 3 of *International Tables for Crystallography*, Vol. A (Hahn, 2005), are typically used. Alternatively, the programs *CHEKCELL* (Laugier & Bochu, 2010) or *ExtSym* (Markvardsen *et al.*, 2008) can be used to distinguish between possible space groups. However, *ExtSym* does not support *GSAS* (Larson & von Dreele, 2004) constant wavelength XRPD data.

Occasionally only a small set of possible space groups can be determined, and each of them must be tried separately for structure solution (for example, the indexing programs may find the unit cell corresponding to any of the three settings of the space group No. 14, $P2_1/c$). Once a space group has been selected, it is used again in a Le Bail fit (Le Bail, 2005) that will furnish the reflections list read by *PSSP* and *WinPSSP*.

Furthermore, the chemical composition of the solid must be (at least partially) known, and the three-dimensional shape and geometry of the molecules or fragments of known connectivity [(4) in Fig. 1] can be obtained from the Cambridge Structural Database or from molecular building and geometry optimization programs, such as *MOPAC* (Stewart, 1990), *Arguslab* (Rodrigues *et al.*, 2015) and others. In our experience, it works well also to complete or partially modify chemically similar fragments extracted from the Cambridge Structural Database.

Once the molecular formula and shape are known, the number of formula units (molecules) in the unit cell (*Z*) can be estimated from the unit-cell volume, assigning 17–20 Å^3 per non-hydrogen atom in the molecular formula of typical organic solids. Concurrently, the calculated density must be around 1–1.5 g cm^{-3}. It is useful to keep in mind that, in these types of low-symmetry molecular solids, the asymmetric unit often consists of one molecule on the general position of the space group, in particular if the molecule itself does not possess inversion or other point group symmetry. In addition, only five space groups account for ~75% of organic crystal structures (Brock & Dunitz, 1994), $P2_1/c$, $P\bar{1}$, $P2_1$, $P2_12_12_1$ and $C2/c$, and thus commonly $Z = 4$, 2, 2, 4 and 4, respectively, according to the multiplicity of the general position of the respective space groups. Therefore, by combining the above information one can estimate the most probable *Z* and the number of crystallographically independent molecules (*Z′*) in the unit cell, and further support the choice of space group symmetry.

(3) *Powder pattern decomposition into integrated intensities of the Bragg reflections*. The results of the Le Bail fit (Le Bail, 2005) tabulated in the `filename.rfl` files written by *GSAS* through the 'ASCII reflection file' option, or the `filename.hkl` files from *FULLPROF* (Roisnel & Rodríguez-Carvajal, 2001) using HKL = 1 in the `filename.pcr` file, are used to input the 'experimental' powder diffraction data into *PSSP* or *WinPSSP* [(7) in Fig. 1]. At present many users intending to solve crystal structures from powders are typically trained in Rietveld analysis (Young, 2002), and the use of the above software for whole pattern decomposition allows the calculation of a high-quality set of integrated intensities and FWHMs, through the already familiar procedure for the refinement of the lattice parameters, 2θ zero error, background intensities and peak profile parameters (McCusker *et al.*, 1999). In addition, *GSAS* (Larson & von Dreele, 2004) provides all Rietveld refinement capabilities needed to carry out the Rietveld fit of organic structures after atomic coordinates are available (soft bond distance and bond angle restraints and rigid body fits).

The *S* agreement factor in *PSSP* and *WinPSSP* uses a set of correlation factors between the integrated intensities of a small number of neighbor reflections (in 2θ) to overcome the loss of information arising from the peak overlap in the powder pattern. This is due to the projection of the three-dimensional reciprocal lattice (integrated intensities *versus hkl*) into one dimension (2θ angle or *d* spacings): for example, the equipartitioning of intensity among exactly overlapped
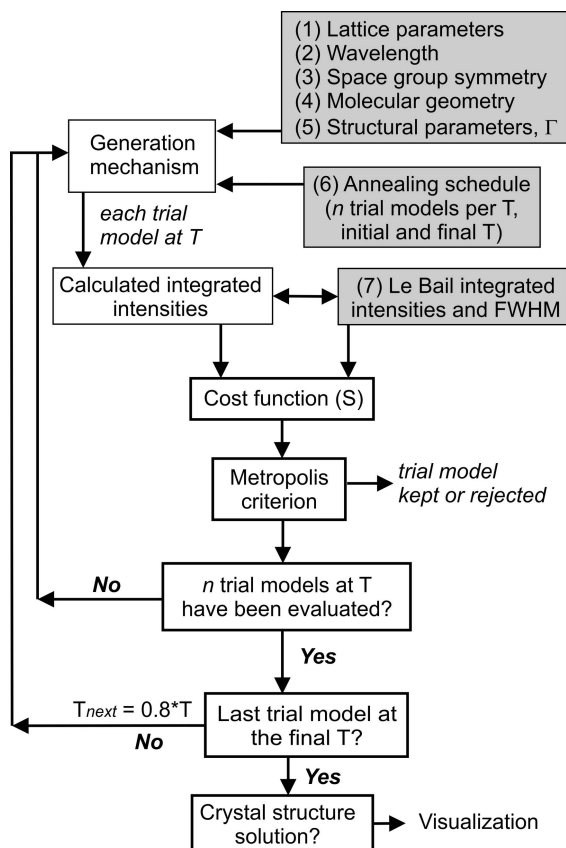

Figure 1
Schematic overview of the direct-space methodology implemented in *PSSP* and *WinPSSP*. The information shaded in grey is the input data.

reflections resulting from the Le Bail algorithm (Le Bail, 2005), if initial arbitrary all-equal 'calculated' structure factors are used.

The correlation factors (Pagola & Stephens, 2010) are calculated once at the beginning of a structure solution run assuming a Gaussian peak profile (so far, a valid approximation) using the experimentally determined FWHM. This methodology affords a fast evaluation of the agreement factor S, which is faster to calculate than R_{wp}, using a still accurate representation of the experimental powder diffraction data.

(4) *Search for the structural motif (approximate atomic coordinates)*. In order to search for the atomic coordinates using direct-space methods, a large number of trial models are generated, locating differently the above defined asymmetric unit in the unit cell, then calculating the total unit cell contents through the application of the space group symmetry operations, as well as the integrated intensities of the reflections corresponding to each trial crystal structure. Thus, the crystal structure solution must be parameterized by the user [(5) in Fig. 1]. Any trial structural model can be generated from a set of structural parameters Γ ($x_m, y_m, z_m, \varphi, \chi, \omega, \tau_1 \ldots \tau_i$), where x_m, y_m, z_m are fractions of the unit-cell parameters defining the position of the fragment in the unit cell; φ, χ, ω are the Eulerian angles defining the orientation of the fragment; and $\tau_1 \ldots \tau_i$ are a variable number of torsion angles defining the conformation of flexible sections of the molecules. Typically, all possible torsional angles are explored as rotations of groups of atoms around single bonds (Pagola & Stephens, 2010). This is shown in Fig. 2 and in the examples available as supporting information, illustrating how to set up the structural parameters of some of the solids discussed in this publication. Occasionally some torsion angles are kept fixed owing to expected intramolecular hydrogen bonding in molecules (Etter, 1991). In this work, examples are given in the crystal structure of the antibacterial levofloxacin anhydrate, wherein the –OH of the carboxylic acid group is hydrogen bonded to the O atom of the keto group adjacent to it, as well as in the crystal structures of the orange and dark-red pigments of 5-methyl-2-[(2-nitro-4-methylphenyl)amino]-3-thiophenecarbonitrile (see Fig. 4 in §3).

Once the crystal structure has been described through a set of structural parameters Γ, a global optimization method must be used to find the Γ values that lead to the best agreement between the experimental powder diffraction pattern and the calculated powder diffraction data (corresponding to the structure 'solution') among the potentially infinite trial models. In *PSSP* and *WinPSSP*, this task is performed by simulated annealing (Kirkpatrick *et al.*, 1983). Random numbers are given to the structural parameters Γ ($x_m, y_m, z_m, \varphi, \chi, \omega, \tau_1 \ldots \tau_i$) from an arbitrarily chosen initial model, while performing the minimization of the agreement factor, S, which is a quantitative expression of the agreement between the correlated 'experimental' and the 'trial model' diffraction intensities.

Any practical implementation of the simulated annealing method requires the definition of a simulated annealing schedule [(6) in Fig. 1], consisting of the initial and final values of the control parameter or temperature (T), the decrement rate for it, and the number of trial models to evaluate at each temperature. During previous years a default simulated annealing schedule consisting of 49 temperatures has been successfully used, with initial $T = 50$, final $T = 0.001$ and decrement rate 0.8. Only the number of trial models evaluated at each temperature has been varied, according to the number of structural parameters to search for.

The input of the above information in *WinPSSP* (see Fig. 1) is rapidly and intuitively done by the user through the 'Asymmetric unit view' panel. The lattice parameters (in Å and °) and the wavelength (in Å) can be typed by the user, and the space group selected from a drop-down menu, or alternatively they can be read from a filename.cif file. The Cartesian or crystallographic coordinates of all fragments in the asymmetric unit can be read from an ASCII text file [*e.g.* filename.xyz from *ORTEP* for Windows (Farrugia, 2012)] or from a filename.cif file, respectively. The atom occupancy factors can be given in the filename.cif file; otherwise they are set to 1 as default. The asymmetric unit will then be displayed on the GUI. Typical examples of the above files are included in the supporting information file.

At this point the structural parameters must be defined. Each molecule or ion in the asymmetric unit must be defined as a 'fragment'. For fragments that consist of more than one
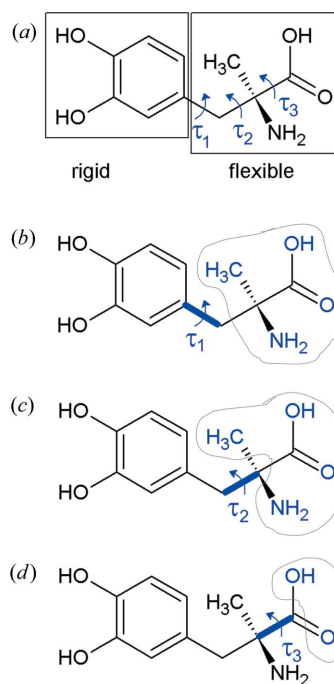

Figure 2
(*a*) Identification of the rigid and flexible sections of an L-α-methyldopa molecule (note that the molecular shape of the correct enantiomorph must be provided for crystal structure solution from powders). Three torsions τ_1, τ_2 and τ_3 that will define the molecular conformation in the solid state are shown with arrows. (*b*) The C—C bond highlighted with a bold line defines the direction of an axis around which the atoms enclosed in the grey figure will be rotated τ_1 degrees. (*c*) and (*d*) correspond to the application of the τ_2 and τ_3 torsional angles, respectively. Atom selection to define the axis and a free-hand drawing tool for enclosing and selecting the atoms to rotate have been implemented in *WinPSSP*, so that it is not necessary to keep track of atom numbering.

atom, using the icons at the top left of the main panel (shown in Fig. 3a), it is possible to define three Eulerian angles, the torsional degrees of freedom (if the fragment contains flexible sections) and three positional parameters. The parameterization of the crystal structure will be automatically displayed under 'Structural parameters' (Fig. 3a). The graphic representation of the fragments in the asymmetric unit reduces user mistakes, such as missing atoms, incorrect number of molecules, wrong molecular shape or enantiomorph, missing unit-cell parameters, incorrect atom labeling in the definition of the rotational degrees of freedom *etc.*

The Le Bail fit results to reconstruct the experimental pattern are read from the GUI panel through a *GSAS* `filename.rfl` or a *FULLPROF* `filename.hkl` file, as previously. Moreover, a script to interface *GSAS-II* (Toby & Von Dreele, 2013) with *WinPSSP* is in our immediate plans.

Once the reflections list has been read, the number of reflections to use must be chosen. Most commonly, 50–300 reflections are used for crystal structure solution, depending on the quality of the diffraction data and the number of degrees of freedom (or structural parameters to search for). For optimal performance, the intensity of the unread reflections should not be higher than 10–15% of the maximum.

The number of trial models per temperature ('models/T') must be then provided (this can also be automatically estimated from the number of structural parameters defined, Γ, by selecting 'Options' and 'Suggest number of trial models/T' through the drop-down menu), according to our performance results discussed in next section. The remaining parameters defining the simulated annealing schedule are seldom modified. The calculation can be started through the 'Options' and 'Run' drop-down menu. The program then prompts for the number of solutions to calculate (typically 5–20) and a common root for their names, which will be followed by a sequential number to distinguish between solutions. Then the calculation process schematized in Fig. 1 is carried out.

In general one must carry out several simulated annealing runs for various reasons, such as (1) to determine the lowest achievable S for the data quality, molecular geometry and calculation conditions used; (2) to check the repeatability of the packing motif within the lowest-S structure solutions; and (3) to check the credibility of the chemical bonding in the solution (hydrogen bonding motifs, π–π interactions, van der Waals forces *etc.*). Thus, each tentative crystal structure solution must be visualized and, if time allows, the coordinates input into *GSAS* (Larson & von Dreele, 2004) for the calculation of R_{wp} and χ^2 (*e.g.* only refining the scale factor and background). Then, on the basis of those results, one can select the best set of atomic coordinates to start a Rietveld fit.

Fig. 3(b) shows a typical view of the main panel once the 'Solutions' tab has been hit. The space labeled 'Visual inspection of the solutions' at the top right of the screen shows the location of the asymmetric unit in the unit cell for the selected solution from the 'Solutions List' on its left side. At the bottom right of the screen, the area labeled 'Diffraction pattern' displays an overlay of the experimental powder pattern (reconstructed from the Le Bail fit file) and the calculated pattern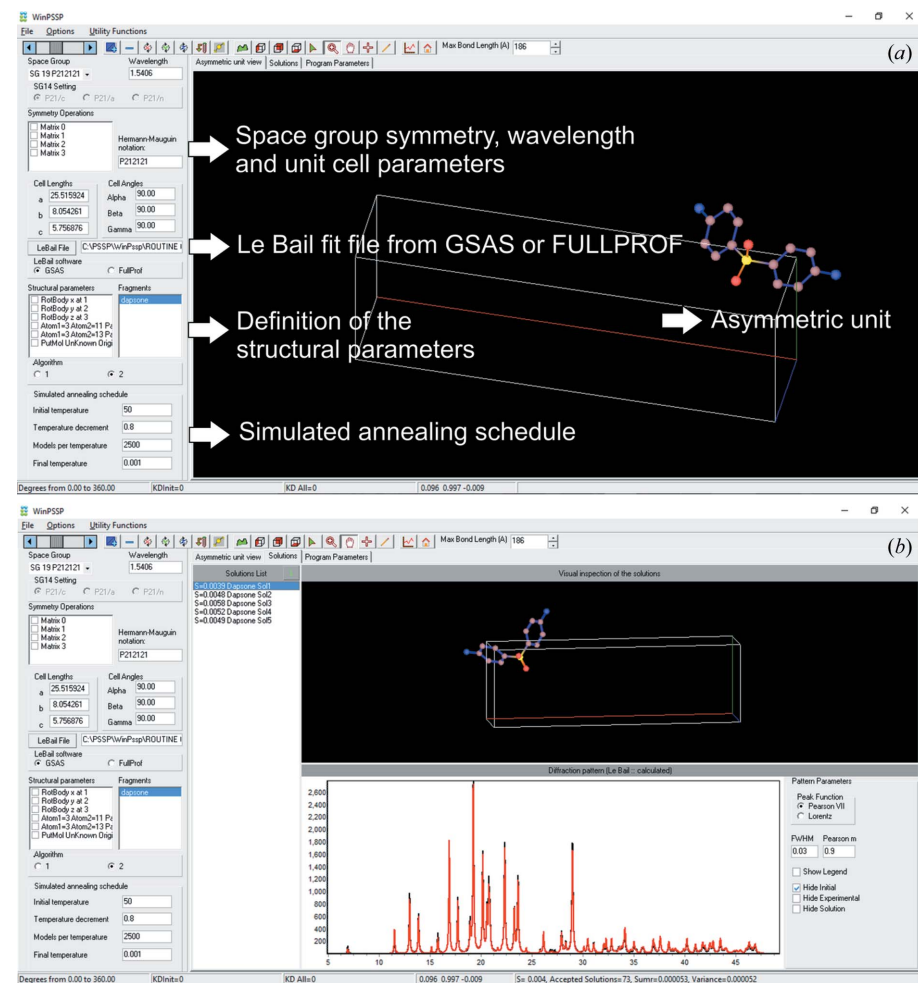 corresponding to the solution selected. The atomic coordinates for each solution are automatically written to the `filename.cif` files, and the `filename.out` files summarize the results of each individual simulated

Figure 3
(a) Initial 'Asymmetric unit view' panel in *WinPSSP*, showing the unit-cell parameters read, the space group symmetry, the wavelength, the asymmetric unit, and the definition of the structural parameters and the simulated annealing schedule for the crystal structure solution of dapsone. (b) The 'Solutions' panel, showing the asymmetric unit location in a structural model selected from five solutions, and the overlay of its calculated diffraction pattern (red line) and the reconstructed 'experimental' diffraction data (black line) using Pearson VII peak profiles.

annealing run. The agreement factors (S) are displayed on the left side of the 'Solutions List' area of the GUI panel, and at the bottom of the 'Diffraction pattern' section.

(5)–(6) *Structure completion and Rietveld refinement.* Once a satisfactory and complete (or almost complete) set of approximate atomic coordinates has been obtained, a Rietveld fit (Young, 2002) can be carried out. For additional information, we direct the reader to the guidelines for Rietveld analysis (McCusker *et al.*, 1999) and the discussion of the meaning of the Rietveld agreement factors given by Toby (2006).

The `filename.cif` files resulting from *WinPSSP* runs can be used to input the coordinates into *GSAS* (Larson & von Dreele, 2004). *GSAS* has the capability of defining rigid bodies from the initial atomic coordinates. The changes in the flexible parts of the organic molecules can be refined using soft bond length and angle restraints. Often the target values can be chosen by selecting chemically related compounds from the Cambridge Structural Database. The hydrogen atoms should be added using single-crystal software such as *CRYSTALS* (Betteridge *et al.*, 2003) or *WinGX* (Farrugia, 2012). Their positions can be refined subject to bond length and angle restraints, and the variation in their atomic displacement parameters can be constrained to 1.2 or 1.5 times the values of those corresponding to the bonded non-H atom (riding hydrogen model).

Perhaps a fact not often emphasized is that various Rietveld fits are often required until an acceptable solution is obtained. It is necessary to perform Fourier difference calculations to investigate possible hydration or solvation, since it is not uncommon to find missing water molecules in the structure solution (García *et al.*, 2009); and it is always necessary to check the validity of the Rietveld-refined solution with *PLATON* (Spek, 2009).

Moreover, the details in the final molecular shape obtained from the Rietveld fit should be further evaluated, for example by comparison with chemically similar structures in the Cambridge Structural Database (Pagola & Stephens, 2012). Further treatment of the results such as optimization by DFT methods (Avila *et al.*, 2008) is beneficial, in particular when ambiguities in the molecular structure remain to be elucidated from the diffraction experiment. Finally, the uncertainties in the atomic coordinates should be corrected, for example using the procedure described by Scott (1983).

3. *WinPSSP* performance in 'routine' structure solution cases

The organic solids shown in Fig. 4 were selected as 'routine' or typical crystal structure solution cases used to evaluate the program performance. An obvious question seems to be, why do we define these cases as 'routine'? To support our choices we refer to the work of Le Bail *et al.* (2009) and Shankland *et al.* (2013).

The complexity of a structure solution problem is described by Le Bail *et al.* (2009) as given by 'sample/instrumental' and 'structural' factors. Some powder specimen characteristics

Table 1
Total and internal degrees of freedom (DoF), number of crystallographically independent fragments, Z', space group symmetry and number of atomic coordinates for the compounds **I** to **XVII**.

Compound	DoF		Cryst. indep. fragments	Z'	Space group	Coordinates (without/with H)
	Total	Internal				
I	6	0	1	1	$P2_12_12_1$	11/19
II	6	0	1	1	$P2_1/c$	7/11
III	7	1	1	1	$P\bar{1}$	12/24
IV	7	1	1	1	$P2_1/c$	16/24
V	7	1	1	1	$P2_1/c$	18/27
VI[a]	7	1	1	1	$P2_12_12_1$	18/27
VI[b]	7	1	1	1	$P\bar{1}$	18/27
VII	8	2	1	1	$P2_12_12_1$	17/29
VIII	9	3	1	1	$P2_1/c$	12/23
IX	9	3	1	1	$P2_1/c$	18/26
X	9	3	1	1	$P2_12_12_1$	15/28
XI	10	4	1	1	$P2_12_12_1$	20/34
XII	12	0	2	1	$P\bar{1}$	22/26
XIII	14	2	2	2	$C2$	54/92
XIV	14	2	2	1	$C2/c$	22/37
XV	14	2	2	1	$P2_1/c$	35/58
XVI	18	9	2	1	$P2_1/c$	24/56
XVII	20	8	2	2	$P2_1$	30/66

VI[a] corresponds to the orange polymorph and **VI**[b] to the dark-red polymorph.

complicating structure determination are small crystallite size and strain (Bataille *et al.*, 2006), while instrumental factors are eased by using monochromatic radiation and synchrotron (high-resolution) XRPD, which diminishes the loss of information due to the peak overlap among reflections close in 2θ. The 'structural' complexity factor is mainly determined by the number of degrees of freedom (DoF) at the structure solution stage and the number of atomic coordinates to refine by the Rietveld method. For organic solids in particular, the structural complexity is related to the number of atomic coordinates to refine by the Rietveld method as well, since we must not forget that the initially assumed molecular geometry is only known to a very good approximation. It is often the case that rigid body refinements using such initial molecular shapes do not yield satisfactory Rietveld agreement factors, and the Rietveld fits are substantially improved by using soft bond and angle restraints. It is within the crystallographer's judgement to determine which results are physically meaningful. Furthermore, occasionally the crystallographer is surprised by hydration water, or crystallographic disorder can substantially complicate the structure determination process at the parameterization and previous stages.

In a recent review of the crystal structure complexity of pharmaceutical solids solved to date (Shankland *et al.*, 2013) the authors mentioned that the majority of crystal structures solved had around 13 DoF, though cases of $Z' = 5$, 142 atomic coordinates and DoF = 30 have been reported. We will use DoF, the number of crystallographically independent fragments and the number of atomic coordinates as indicators of structural complexity, as shown in Table 1.

One of the main objectives of this research area is to optimize the global optimization algorithms to find the structure solution fast and efficiently. Among global optimization

methods, the simulated annealing algorithm (Kirkpatrick *et al.*, 1983) has had the largest impact, owing to its high efficacy and ease of implementation (Shankland *et al.*, 2013). Simulated annealing has been implemented in several direct-space software packages (Coelho, 2000; Altomare *et al.*, 2003; Flor-

ence *et al.*, 2005), although several other global optimization methods such as Monte Carlo (Harris & Tremayne, 1996; Brodski *et al.*, 2003), parallel tempering (Falcioni & Deem, 1999; Favre-Nicolin & Černý, 2002), genetic algorithms (Harris & Cheung, 2004; Harris, 2009) and hybrid approaches
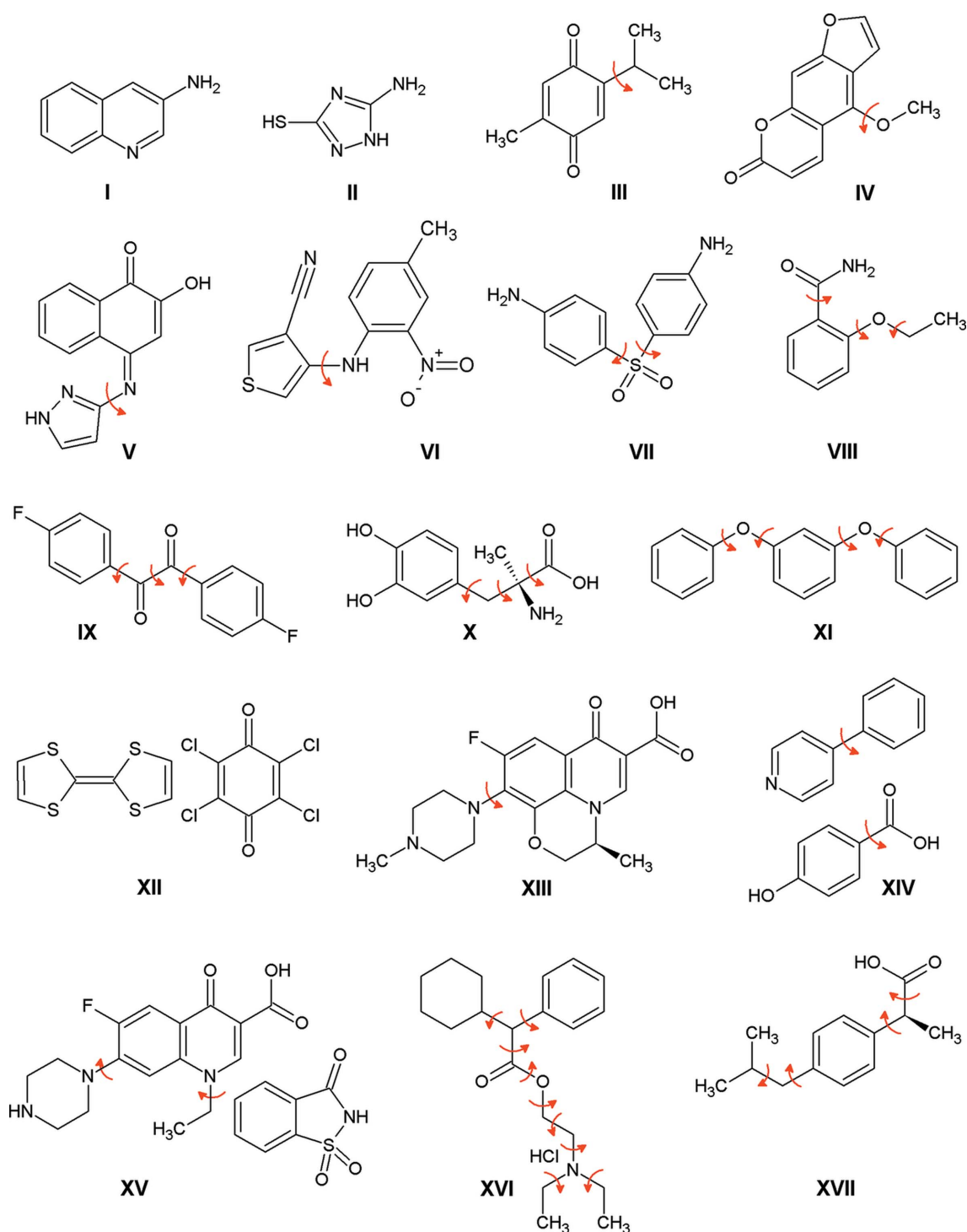

Figure 4
Molecular structures of (**I**) 3-aminoquinoline; (**II**) 3-amino-5-mercapto-1,2,4-triazole; (**III**) thymoquinone; (**IV**) bergapten; (**V**) 2-hydroxy-*N*-[3(5)-pyrazolyl]-1,4-naphthoquinone-4-imine; (**VI**) 5-methyl-2-[(2-nitro-4-methylphenyl)amino]-3-thiophenecarbonitrile (orange and dark-red polymorphs); (**VII**) dapsone; (**VIII**) 2-ethoxybenzamide; (**IX**) 4,4′-difluorobenzyl; (**X**) L-α-methyldopa; (**XI**) 1,3-diphenoxybenzene; (**XII**) tetrathiafulvalene chloranil (black polymorph); (**XIII**) levofloxacin (anhydrous); (**XIV**) 4-hydroxybenzoic phenylpyridine 1:1 co-crystal; (**XV**) norfloxacinium saccharinate; (**XVI**) drofenine hydrochloride; and (**XVII**) *S*-ibuprofen. Torsional degrees of freedom are indicated with arrows.

computer programs

Table 2
The compounds studied, their total and internal DoF, the number of reflections used for structure solution with *WinPSSP*, χ^2 of the Le Bail fit, the maximum S of the crystal structure solutions, and the $\chi^2_{\text{solution}}/\chi^2_{\text{Le Bail}}$ ratio after 20 cycles of Rietveld refinement (with only scale factor and background refined).

Compound	DoF Total	Internal	Reflections	$\chi^2_{\text{Le Bail}}$	S	$\chi^2_{\text{solution}}/\chi^2_{\text{Le Bail}}$
I	6	0	75	1.517	0.099	3.6
II	6	0	120	1.771	0.029	4.9
III	7	1	52	9.723	0.12	11.4
IV	7	1	56	3.406	0.089	11.1
V	7	1	120	2.675	0.0225	1.4
VI[a]	7	1	114	1.888	0.0478	4.3
VI[b]	7	1	120	17.21	0.0234	1.6
VII	8	2	120	3.141	0.0182	8.4
VIII	9	3	82	1.391	0.1307	7.6
IX	9	3	86	1.791	0.039	1.4
X	9	3	86	1.412	0.031	1.6
XI	10	4	70	1.283	0.0995	2.8
XII	12	0	247	1.551	0.12	7.4
XIII	14	2	250	1.226	0.046	11.2
XIV	14	2	75	1.217	0.0184	4.1
XV	14	2	103	6.209	0.084	14.9
XVI	18	9	156	1.641	0.04	6.5
XVII	20	8	122	2.128	0.012	12

VI[a] corresponds to the orange polymorph and **VI**[b] to the dark-red polymorph.

(Lanning *et al.*, 2000; Johnston *et al.*, 2002) have been successfully used. Thus, given the existing algorithms, available software and computer calculation capability, most commonly our chances of success using direct-space methods are not limited by computational resources; but besides getting workable diffraction data, they rather rely on our ability to correctly 'parameterize' the structure [defining the lattice, space group, asymmetric unit composition, shape of the fragments and the structural parameters Γ ($x_m, y_m, z_m, \varphi, \chi, \omega, \tau_1 \ldots \tau_i$)] leading to the crystal structure solution.

The definition of the structural parameters to solve the structures of the solids represented in Fig. 4 is straightforward, and we chose them as 'routine' cases for which DoF varies from 6 to 20, the number of crystallographically independent fragments to locate is 1 or 2, and the number of coordinates (including H) varies between 11 and 92. Except for anhydrous levofloxacin, the space group symmetry belongs to the most frequently found space groups. This information is summarized in Table 1.

Furthermore, in practice crystal structure determination of organic materials from powders can be far from routine when ambiguities exist in the definition of the structural parameters. Two versatile and widely studied material types which typically conform to 'non-routine' cases include metal–organic frameworks and large organic co-crystals (including those of pharmaceuticals). For the former, the '*a priori*' atom connectivity information is limited in comparison with purely organic solids, whereas for the latter the number of crystallographically independent fragments in the asymmetric unit can be large. Disordered crystal structures are also typically difficult to parameterize; various potentially good structural

models must be calculated and their validity assessed by the Rietveld method, which can be a time-consuming task. For large structures composed of rigid and flexible sections or of many crystallographically independent fragments, owing to the limitations imposed by the large number of DoF, often one has to try locating parts of the asymmetric unit successively, re-extracting integrated intensities after certain fragments have been confidently located, and relocating others until the model is complete (Pagola & Stephens, 2012; Altheimer *et al.*, 2013). The authors plan to discuss some of those 'non-routine' cases in further detail in a future publication, as well as to improve the efficiency of *WinPSSP* to systematically solve structures posing the same difficulties.

The XRPD patterns of (**VII**) dapsone and (**XVII**) *S*-ibuprofen were collected from a laboratory diffractometer with monochromatic Cu $K\alpha_1$ radiation and their Le Bail fit results were reported by Pagola & Stephens (2010). The remaining high-resolution X-ray powder diffraction patterns were collected from glass capillaries at the NSLS-I, Brookhaven National Laboratory, USA, including that of the 4-hydroxybenzoic phenylpyridine 1:1 co-crystal (Lapidus *et al.*, 2010). The wavelengths are reported in Table S1. The Le Bail fits were carried out with *GSAS* (Larson & Von Dreele, 2004) in all cases.

For the evaluation of the *WinPSSP* performance for crystal structure determination, 20 simulated annealing runs were carried out using a fixed number of trial models/T for each solid of Fig. 4. For the smallest compounds those could be as low as 100 models/T, while up to 1.75×10^6 models/T were used for the largest compounds. For each material, two additional sets of 20 simulated annealing runs were then carried out, each new set at a constant but different value of trial models/T.

The maximum S values of the models considered as crystal structure solutions are shown in Table 2. While Pagola & Stephens (2010) regarded a model as a solution if $\chi^2_{\text{Rietveld}} < 10 \chi^2_{\text{Le Bail}}$, using *WinPSSP* and synchrotron diffraction data we chose the cut-off S values to distinguish the solutions from other models on the basis of the best achievable S for the set of conditions used. In most cases, the above criterion (Pagola & Stephens, 2010) is also satisfied or almost so.

Graphs of the percentage of solutions obtained *versus* models/T used are shown in the supporting information (Figs. S1–S18). The analysis of the results for the 18 solids studied indicates that the number of models/T must increase approximately exponentially to obtain a linear increase in the percentage of solutions.

For each of the 18 compounds, the fits of the above data (shown in Figs. S1–S18) were then used to calculate the models/T giving rise to 50% of crystal structure solution success in 20 simulated annealing runs. We considered these values reasonable estimations of the number of trial models required on average to solve these 'routine' crystal structures using our simulated annealing implementation.

The latter results were then used to visualize how the number of trial models required to solve the structures varies according to the increase in the DoF of the compounds. This is

represented in Fig. 5, which shows that the models/T required for around 50% of crystal structure solution success increases approximately exponentially with the DoF of the structures. The relationship found is

$$\text{Models per temperature for 50\% solutions} = 8.368 \exp(0.678 \text{DoF}). \quad (1)$$

This equation has been used in *WinPSSP* to automatically calculate a default number of trial models/T according to the DoF defined by the user in the 'Structural parameters' section. Note that the models/T calculated from equation (1) is only a reasonable estimate, and in practice that number may still vary by around one order of magnitude above and below the value calculated using equation (1), as the dispersion of the estimated quantities for the individual compounds (see Fig. 5) is rather large {*e.g.* for structures with DoF = 7 the calculated models/T from equation (1) is 963, whereas in practice the required models/T estimated from the individual fits vary between 313 models/T to solve bergapten (**IV**) and 8389 models/T to solve the dark-red polymorph of 5-methyl-2-[(2-nitro-4-methylphenyl)amino]-3-thiophenecarbonitrile (**VI**[b]). This seems reasonable owing to the stochastic character of the method, the finite number of models used and simulated annealing runs carried out, and other factors such as the initial user choices for structure solution.

The results of the individual runs afforded also the estimation of the computer calculation time that is taken to solve these structures in *WinPSSP* (excluding indexing, space group determination, and Le Bail and Rietveld fits). As expected, the time required to obtain a structural model also increases approximately exponentially with the DoF of the structure. From equation (1), 1,3-diphenoxybenzene (DoF = 10)
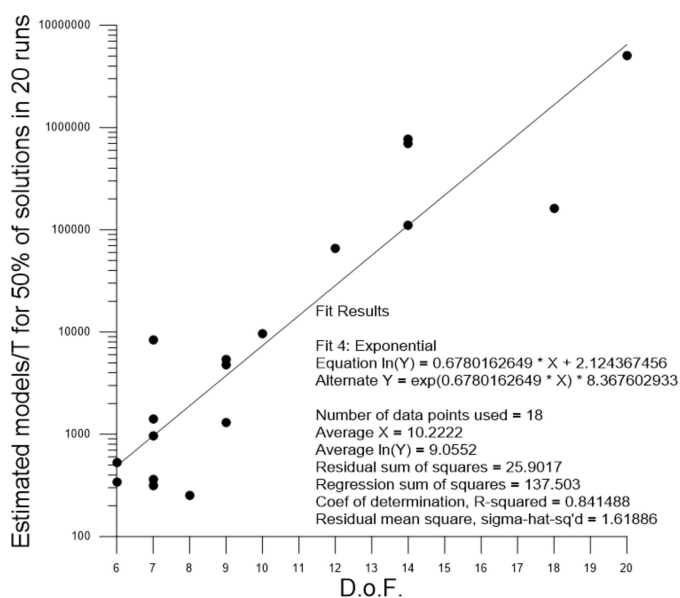
requires around 7364 models/T for structure solution. One simulated annealing run (all other conditions equal to those in Fig. 5) takes around 60 s in a 2.60 GHz Pentium dual-core CPU with 64 bit Windows 10 OS. The situation is different for *S*-ibuprofen, DoF = 20 and $Z' = 2$. In this case, equation (1) results in 6.5×10^6 models/T. In this and similar cases, the computational resources may become a limiting factor. The largest number of models/T used to test *S*-ibuprofen in *WinPSSP* has been 1.75×10^6, taking around 16 870 s (or around 4.7 h) for each simulated annealing run and resulting in 45% of solutions in 20 simulated annealing runs.

4. Computational requirements, availability and documentation

WinPSSP has been written in Delphi programming language. The existing C++ code (Pagola & Stephens, 2010) has been ported to a dynamic link library (DLL) file compiled using the free Code::Blocks IDE (The Code::Blocks Team, 2016) coupled with GCC MinGW/Cygwin (GNU Project, 1987), and it has been kept largely unmodified, except for the necessary calls for the DLL and some adjustments for speed improvements. The program has been tested in the Windows Vista, Windows 7 (32 bit and 64 bit) and 64 bit Windows 10 OS.

WinPSSP is a free-distribution program available from http://users.uoi.gr/nkourkou/winpssp/. It is compiled as a portable application, and as such it has many advantages compared with similar programs: (i) it can run from any folder or USB stick, (ii) portability and compatibility issues among users are diminished (advantageous for teaching courses), and (iii) it requires no installation setup, administrative privileges or access to the registry.

In addition to the information in this publication, examples of files are provided in the supporting information for this article and at http://users.uoi.gr/nkourkou/winpssp/, where users can also find instructional videos.

5. Conclusion

WinPSSP is a freely distributed computer program with demonstrated utility towards the crystal structure determination of typical small organic molecular solids from powders (20 or fewer DoF, $Z' \leq 2$ and frequently found space group symmetry). Eighteen examples have been presented and the performance of *WinPSSP* was discussed. The user guidelines have been elaborated in detail, improving the 'user-friendliness' of the software and contributing to the systematization of the procedure to solve 'routine' organic structures from powders.

Using this software, a typical organic crystal structure can be considered solved in time scales of the order of minutes to hours or days (in particular if more than one Le Bail fit is required). Challenges are still encountered in the parameterization of disordered structures, certain structures with asymmetric unit fragments in special positions, and unexpected hydrates or solvates, and where limitations are imposed

Figure 5
The number of models per temperature necessary to obtain 50% of solutions in 20 simulated annealing runs as a function of the total DoF for the compounds shown in Fig. 4.

computer programs

by the computational resources required for crystal structures with a large number of DoF.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Altheimer, B. D., Pagola, S., Zeller, M. & Mehta, M. A. (2013). *Cryst. Growth Des.* **13**, 3447–3453.
Altomare, A., Caliandro, R., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2003). *J. Appl. Cryst.* **36**, 230–238.
Altomare, A., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A. G. G. & Rizzi, R. (2007). *J. Appl. Cryst.* **40**, 344–348.
Altomare, A., Cuocci, C., Giacovazzo, C., Moliterni, A., Rizzi, R., Corriero, N. & Falcicchio, A. (2013). *J. Appl. Cryst.* **46**, 1231–1235.
Altomare, A., Giacovazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E. (2000). *J. Appl. Cryst.* **33**, 1180–1186.
Avila, E. E., Mora, A. J., Delgado, G. E., Contreras, R. R., Fitch, A. N. & Brunelli, M. (2008). *Acta Cryst.* B**64**, 217–222.
Bataille, T., Mahé, N., Le Fur, E., Pivan, J.-Y. & Louër, D. (2006). *Z. Kristallogr. Suppl.* **2006**, 9–14.
Betteridge, P. W., Carruthers, J. R., Cooper, R. I., Prout, K. & Watkin, D. J. (2003). *J. Appl. Cryst.* **36**, 1487.
Botez, C. E., Stephens, P. W., Nunes, C. & Suryanarayanan, R. (2003). *Powder Diffr.* **18**, 214–218.
Boultif, A. & Louër, D. (2004). *J. Appl. Cryst.* **37**, 724–731.
Boultif, A. & Louër, D. (2014). *Powder Diffr.* **29**, S7–S12.
Brock, C. P. & Dunitz, J. D. (1994). *Chem. Mater.* **6**, 1118–1127.
Brodski, V., Peschar, R. & Schenk, H. (2003). *J. Appl. Cryst.* **36**, 239–243.
Černý, R. & Favre-Nicolin, V. (2007). *Z. Kristallogr.* **222**, 105–113.
Chernyshev, V. V. (2001). *Russ. Chem. Bull. Int. Ed.* **50**, 2273–2292.
Coelho, A. A. (2000). *J. Appl. Cryst.* **33**, 899–908.
Datta, S. & Grant, D. J. W. (2004). *Nat. Rev. Drug Discov.* **3**, 42–57.
David, W. I. F. & Shankland, K. (2008). *Acta Cryst.* A**64**, 52–64.
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. (1989). *Nature*, **342**, 260–262.
Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.
Etter, M. C. (1991). *J. Phys. Chem.* **95**, 4601–4610.
Falcioni, M. & Deem, M. W. (1999). *J. Chem. Phys.* **110**, 1754–1766.
Farrugia, L. J. (2012). *J. Appl. Cryst.* **45**, 849–854.
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.
Gavezzotti, A. (1998). *Crystallogr. Rev.* **7**, 5–121.
GNU Project (1987). *GCC, the GNU Compiler Collection*, https://gcc.gnu.org/.
Hahn, T. (2005). Editor. *International Tables for Crystallography*, Vol. A, *Space-Group Symmetry*. Heidelberg: Springer.
Harris, K. D. M. (2009). *Comput. Mater. Sci.* **45**, 16–20.
Harris, K. D. M. (2012). *Top. Curr. Chem.* **315**, 133–178.
Harris, K. D. M. & Cheung, E. Y. (2003). *Org. Process Res. Dev.* **7**, 970–976.
Harris, K. D. M. & Cheung, E. Y. (2004). *Chem. Soc. Rev.* **33**, 526–538.
Harris, K. D. M. & Tremayne, M. (1996). *Chem. Mater.* **8**, 2554–2570.
Huq, A. & Stephens, P. W. (2003). *J. Pharm. Sci.* **92**, 244–249.
Johnston, J. C., David, W. I. F., Markvardsen, A. J. & Shankland, K. (2002). *Acta Cryst.* A**58**, 441–447.
Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983). *Science*, **220**, 671–680.
Lanning, O. J., Habershon, S., Harris, K. D. M., Johnston, R. L., Kariuki, B. M., Tedesco, E. & Turner, G. W. (2000). *Chem. Phys. Lett.* **317**, 296–303.
Lapidus, S. H., Naik, A., Wixtrom, A., Massa, N. E., Ta Phuoc, V., del Campo, L., Lebègue, S., Ángyán, J. G., Abdel-Fattah, T. & Pagola, S. (2014). *Cryst. Growth Des.* **14**, 91–100.
Lapidus, S. H., Stephens, P. W., Arora, K. K., Shattock, T. R. & Zaworotko, M. J. (2010). *Cryst. Growth Des.* **10**, 4630–4637.
Larson, A. C. & von Dreele, R. B. (2004). *General Structure Analysis System* (*GSAS*). Report LAUR 86–748. Los Alamos National Laboratory, New Mexico, USA.
Laugier, J. & Bochu, B. (2010). *Chekcell*, http://www.iucr.org/resources/other-directories/software/chekcell.
Le Bail, A. (2001). *Mater. Sci. Forum*, **378–381**, 65–70.
Le Bail, A. (2004). *Powder Diffr.* **19**, 249–254.
Le Bail, A. (2005). *Powder Diffr.* **20**, 316–326.
Le Bail, A. *et al.* (2009). *Powder Diffr.* **24**, 254–262.
Louër, D. & Boultif, A. (2006). *Z. Kristallogr.* **23**, 225–230.
Louër, D. & Langford, J. I. (1996). *Rep. Prog. Phys.* **59**, 131–234.
Markvardsen, A. J., Shankland, K., David, W. I. F., Johnston, J. C., Ibberson, R. M., Tucker, M., Nowell, H. & Griffin, T. (2008). *J. Appl. Cryst.* **41**, 1177–1181.
McCusker, L. B., Von Dreele, R. B., Cox, D. E., Louër, D. & Scardi, P. (1999). *J. Appl. Cryst.* **32**, 36–50.
Mohamud, S., Ta Phuoc, V., del Campo, L., Massa, N. E. & Pagola, S. (2016). *Synth. Met.* **214**, 71–75.
Nunes, C., Suryanarayanan, R., Botez, C. E. & Stephens, P. W. (2004). *J. Pharm. Sci.* **93**, 2800–2809.
Pagola, S., Benavente, A., Raschi, A., Romano, E., Molina, M. A. A. & Stephens, P. W. (2003). *AAPS PharmSciTech*, **5**, 28.
Pagola, S., Pike, R. D., deKrafft, K. & Tronic, T. A. (2008). *Acta Cryst.* C**64**, m134–m136.
Pagola, S. & Stephens, P. W. (2000). *Mater. Sci. Forum*, **321–324**, 40–45.
Pagola, S. & Stephens, P. W. (2009). *Acta Cryst.* C**65**, o583–o586.
Pagola, S. & Stephens, P. W. (2010). *J. Appl. Cryst.* **43**, 370–376.
Pagola, S. & Stephens, P. W. (2012). *CrystEngComm*, **14**, 5349–5354.
Pagola, S., Stephens, P. W., He, X. & Byrn, S. R. (2001). *Mater. Sci. Forum*, **378–381**, 789–794.
Pagola, S., Tracana, M. I., Amani, S. M., Gonzalez, A. M., Raschi, A. B., Romano, E., Benavente, A. M. & Stephens, P. W. (2008). *Nat. Prod. Commun.* **3**, 759–764.
Putz, H., Schön, J. C. & Jansen, M. (1999). *J. Appl. Cryst.* **32**, 864–870.
Rodrigues, R. P., Andrade, S. F., Mantoani, S. P., Eifler-Lima, V. L., Silva, V. B. & Kawano, D. F. (2015). *J. Chem. Educ.* **92**, 827–835.
Roisnel, T. & Rodríquez-Carvajal, J. (2001). *Mater. Sci. Forum*, **378–381**, 118–123.
Scott, H. G. (1983). *J. Appl. Cryst.* **16**, 159–163.

Shankland, K., Spillman, M. J., Kabova, E. A., Edgeley, D. S. & Shankland, N. (2013). *Acta Cryst.* C**69**, 1251–1259.

Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.

Sperandeo, N. R., Karlsson, A., Cuffini, S., Pagola, S. & Stephens, P. W. (2005). *AAPS PharmSciTech*, **6**, 82.

Stephens, P. W., Pagola, S., Bohle, D. S., Kosar, A. D. & Madsen, S. K. (2000). *Nature*, **404**, 307–310.

Stephenson, G. A. (2005). *Rigaku J.* **22**, 2–15.

Stewart, J. J. P. (1990). *J. Comput. Aided Mol. Des.* **4**, 1–103.

The Code::Blocks Team (2016). *Code::Blocks*, Release 16.01, Rev. 10702, http://www.codeblocks.org.

Toby, B. H. (2006). *Powder Diffr.* **21**, 67–70.

Toby, B. H. & Von Dreele, R. B. (2013). *J. Appl. Cryst.* **46**, 544–549.

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. (2002). Editor. *The Rietveld Method.* Oxford University Press.