Journal of
Applied
Crystallography
ISSN 0021-8898

Received 20 July 2012
Accepted 21 September 2012

TALP: a multisolution direct-space strategy for solving molecular crystals from powder diffraction data based on restrained least squares

Oriol Vallcorba,* Jordi Rius, Carlos Frontera and Carles Miravitlles

Institut de Ciència de Materials de Barcelona (CSIC), Campus de la UAB, Bellaterra, Catalunya 08193, Spain. Correspondence e-mail: ovallcorba@icmab.es

TALP is a new direct-space strategy for *ab initio* crystal structure determination of molecular crystals from powder diffraction data. The strategy is based on a preliminary exploration stage, which considers atomic overlap, followed by a subsequent stage of local incremental scans, both coupled to fast restrained least-squares minimizations with the atomic coordinates as refined parameters. The observed intensities are extracted from the powder pattern by a three-step procedure [Vallcorba, Rius, Frontera, Peral & Miravitlles (2012). *J. Appl. Cryst.* **45**, 844–848], and the molecular model and distance restraints are derived from molecular mechanics calculations or from similar reported structures. The solution process consists of several independent trials, each one resulting in a crystal structure proposal with an associated figure of merit. TALP has been tested on laboratory X-ray powder diffraction data of 14 molecular compounds of known crystal structure and of variable complexity. In most cases, the crystal structure is solved in a short time (less than an hour), even for calculated models. For the most complex structures (*e.g.* 13 torsion angles), the general scan is assisted by a rotation function, which provides a ranked list of most probable model orientations. In this way only the positional and conformation parameters need to be explored.

1. Introduction

Crystal structure determination from powder diffraction data (SDPD) has proved to be a powerful method to characterize a wide range of materials. With this technique crystal structures from inorganic substances or molecular organic compounds can be solved when no single crystals are available. For SDPD two main approaches exist: (*a*) the direct determination of the structure factor phases from the observed amplitudes, known as direct methods, and (*b*) the optimization of the structure model in the unit cell by minimizing the discrepancies between observed and calculated powder diffraction patterns, known as direct-space methods. The former approach requires data at atomic resolution and, consequently, is mostly applied to inorganic compounds. The latter approach, however, tolerates data at lower resolution and hence is normally used for solving molecular compounds, *e.g.* organic samples measured with laboratory X-ray sources. In this second approach the introduction of a molecular model compensates for the limited resolution (Rius & Miravitlles, 1988). Commonly used strategies and software for SDPD have been the subject of several reviews (Harris *et al.*, 2001; Černý & Favre-Nicolin, 2007; David & Shankland, 2008). In direct-space methods the solution process can be formulated as an optimization procedure of a cost function expressed in terms of parameters describing the conformation of the molecular model and its position in the unit cell. Most global optimization methods in SDPD are based on Monte Carlo/simulated annealing searches [implemented in computer programs such as *FOX* (Favre-Nicolin & Černý, 2002), *DASH* (David *et al.*, 2006), *ESPOIR* (Le Bail, 2001), *PowderSolve* (Engel *et al.*, 1999), *TOPAS* (Coelho, 2000) or *PSSP* (Pagola & Stephens, 2010)], on genetic algorithms (Harris *et al.*, 2004; Kariuki *et al.*, 1997; Shankland *et al.*, 1997) or on evolutionary optimizations (Chong & Tremayne, 2006). The definition of the molecular model is extremely important in direct-space methods since it restricts the number of possible solutions. Most previously cited SDPD programs choose the Z matrix to describe the model (Shankland, 2005, and references therein) so that the parameters to refine are reduced to the position of the molecule and its torsion angles. However, there is an alternative description with only the atomic coordinates as parameters but making use of geometrical restraints as in a conventional restrained Rietveld refinement (Baerlocher, 1993; Favre-Nicolin & Černý, 2004). In this last approach, the number of parameters to refine is larger than with torsion angles but, as a counterpart, the model is easier to handle and the conformational changes during the minimization are not limited to torsional ones.

The success of a specific direct-space strategy depends on many factors, the principal ones being (i) the quality of the diffraction data (resolution, peak overlap, preferred orientation and other instrumental aspects), (ii) the complexity of the

structure (number of torsion angles, number of molecules in the asymmetric unit, eventual existence of partial occupations, presence of heavy atoms or of solvent molecules) and (iii) the model description (accuracy and completeness).

In this work, a direct-space multisolution strategy (TALP) is described for the solution of crystal structures of molecular compounds from powder X-ray diffraction data. This strategy basically combines a general scan and a local incremental scan both coupled to fast least-squares (FLS) minimizations carried out with an adapted version of the restrained Rietveld refinement program *RIBOLS* (Rius, 2012), developed and used by our research group for longer than 20 years. The quantity being minimized in terms of the atomic coordinates is

$$M = S_Y + kS_R, \tag{1}$$

where S_Y is the residual measuring the differences between observed and calculated net intensities over all **H** reflections in the powder pattern:

$$S_Y = \sum_{\mathbf{H}} w_{\mathbf{H}}(y_{\mathrm{obs},\mathbf{H}} - y_{\mathbf{H}})^2, \tag{2}$$

with $w_{\mathbf{H}}$ equal to the reciprocal of the measured pattern intensity at $2\theta_{\mathbf{H}}$. The observed intensities in equation (2), y_{obs}, are obtained with the *DAJUST* software (Vallcorba *et al.*, 2012) according to the following three-step extraction procedure:

(*a*) The intensity $y_{\mathrm{obs},\mathbf{h}}$ at the peak maximum of each resolved **h** reflection is stored together with its multiplicity $j_{\mathbf{h}}$ and the value of the profile function Ω at $2\theta_{\mathbf{h}}$, $\Omega_{\mathbf{h}}$.

(*b*) The total contributions of the resolved reflections are subtracted from the observed powder pattern to produce a difference pattern containing only the information of the nonresolved reflections.

(*c*) The intensity $y_{\mathrm{obs},\mathbf{k}}$ at the maximum of each nonresolved **k** reflection is stored together with the products $j_{\mathbf{l}}\,\Omega_{\mathbf{l},\mathbf{k}}$, where $\Omega_{\mathbf{l},\mathbf{k}}$ is the profile contribution, at $2\theta_{\mathbf{k}}$, of the neighboring reflection **l**.

The corresponding calculated intensities in equation (2) are computed

(*a*) for a resolved reflection **h** at $2\theta_{\mathbf{h}}$ with the expression

$$y_{\mathbf{h}} = cj_{\mathbf{h}}\Omega_{\mathbf{h}}|F_{\mathbf{h}}|^2, \tag{3}$$

where $\Omega_{\mathbf{h}}$ is the value of the profile function at the reflection center, $j_{\mathbf{h}}$ and $F_{\mathbf{h}}$ are the multiplicity and the structure factor, respectively, and c is a scaling factor, and

(*b*) for an arbitrary nonresolved reflection (**k**) at $2\theta_{\mathbf{k}}$ with

$$y_{\mathbf{k}} = cj_{\mathbf{k}}\Omega_{\mathbf{k}}|F_{\mathbf{k}}|^2 + c\sum_{\mathbf{l}(\mathbf{k})} j_{\mathbf{l}}\Omega_{\mathbf{l},\mathbf{k}}|F_{\mathbf{l}}|^2, \tag{4}$$

where the summation extends over all nonresolved reflections **l** contributing to **k**.

The three-step extraction procedure is an improved version of the one already used by our group to perform fast Rietveld refinements of molecular compounds (Rius *et al.*, 1990). The criterion for distinguishing between resolved and nonresolved reflections is based on the test calculations performed by Rius *et al.* (1996), which allowed the conclusion that two peaks may be safely regarded as resolved if the angular distance between their centers is larger than 0.5 times their FWHM. The observed intensities and the remaining reflection information are stored in a QCK file by *DAJUST*, as described by Vallcorba *et al.* (2012).

Similarly to S_Y, the residual S_R in equation (1), involving the restraints, is defined by

$$S_R = \sum_j \sigma_j^{-2}\,(d_{\mathrm{obs},j} - d_j)^2, \tag{5}$$

where $d_{\mathrm{obs},j}$ and d_j are, respectively, the expected interatomic distances (with estimated variances σ_j^2) and the calculated ones derived from the model. The factor k in equation (1) controls the relative weight of the restraints residual S_R in M.

2. Description of the molecular model

The model should be chemically complete and the model description must allow all possible conformational changes. Any error in the model description will increase the residual values and in extreme cases will render it impossible to reach or identify the true solution. In the TALP strategy the molecular model is specified by the atomic coordinates and held together during the FLS refinements by distance restraints, which may correspond to bond distances, bond angles and other interatomic distances. Additional restraints such as molecular planes can be added. The model definition is completed by selecting the free rotation bonds (defined below) allowing conformational changes during the general and local incremental scans.

2.1. Bond distance and bond angle restraints

Usually, σ values associated with bond distances are taken as approximately 1% of the value of the corresponding bond length. Bond angle restraints are introduced in the form of distance restraints between the three atoms involved. To further illustrate this latter case, let the restraint on the bond angle $\alpha = 109.5^\circ$ (subtended by C6—C7—C8) be introduced in the form of two bond distance restraints C6—C7 and C7—C8 plus the additional C6···C8 interatomic distance, the latter

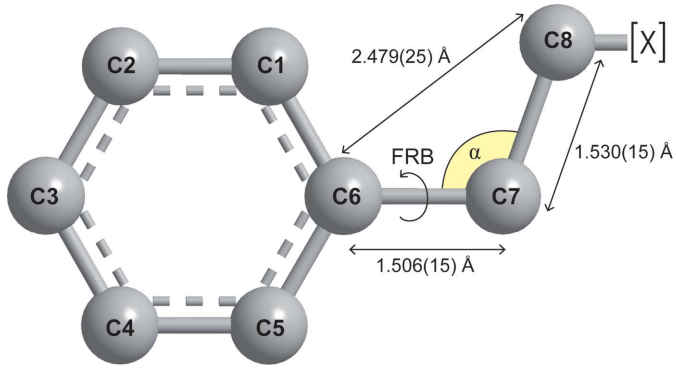

Figure 1
Bond length restraints (C6—C7 and C7—C8), bond angle restraint (C6—C7—C8) and free rotation bond (C6—C7) for the ethylbenzene fragment discussed in the text. All restraints needed to fix the geometry of this fragment are listed in Table 1.

Table 1
Restraints and FRB necessary for the model description in Fig. 1.

Bond distance	Bond angle	FRB
C1—C2, C2—C3, C3—C4, C4—C5, C5—C6, C6—C7, C7—C8	C1···C3, C2···C4, C3···C5, C4···C6, C5···C1, C1···C7, C5···C7, C6···C8	C6—C7

Note: If atom C8 is bound to atom *X* (Fig. 1), the C8—*X* bond distance and C7—C8—*X* bond angle restraints as well as the C7—C8 FRB have to be added.

with an estimated σ equal to the absolute value of the C6···C8 length increment obtained with $\alpha - 2^\circ$ and $\alpha + 2^\circ$ (Fig. 1).

2.2. Free rotation bonds

In general, changes in molecular conformations during the general and the local incremental scans are accomplished through rotations around certain selected bonds of the molecule, which are denoted free rotation bonds (FRBs). For the sequence of bound atoms C5—C6—C7—C8 represented in Fig. 1, if C6—C7 represents an FRB then any rotation around C6—C7 will always fulfill the two bond angle restraints C5—C6—C7 and C6—C7—C8, *i.e.* five distance restraints (C6—C7 is shared) are necessary to refine the conformation around C6—C7 during the FLS refinement. For illustrative purposes, the restraints and FRBs needed to define the complete fragment in Fig. 1 are listed in Table 1.

For each *j* restraint in the residual S_R, the tolerance between expected and calculated distances is controlled by the corresponding weight, $w_j = \sigma_j^{-2}$. This restraints-based approach allows the complete definition of the model and keeps the conformational freedom of the molecule. In other words, the use of atomic coordinates provides to each atom its own freedom of movement within the paths imposed by the geometrical restraints.

3. The TALP structure solution strategy

The TALP strategy (hereafter simply TALP) relies on the availability of (*a*) a file containing the $y_{\mathrm{obs},\mathbf{H}}$ intensities and the peak profile information (see §1); (*b*) the calculated pattern from a model-free pattern matching (*e.g.* LeBail fit) together with the profile parameters used for its calculation; and (*c*) a starting molecular model described in terms of the atomic coordinates, the restraints and the FRBs. The molecular model is obtained either from calculations (*e.g.* molecular mechanics) or from similar reported structures, *e.g.* using *Mogul* (Bruno *et al.*, 2004).

TALP consists of several independent trials (Fig. 2), each one giving at its end a crystal structure proposal with a figure of merit involving all measured points of the powder pattern. This final figure of merit, called Q_{trial}, is defined by the quotient

$$Q_{\mathrm{trial}} = R_{\mathrm{wp,fp}}/R_{\mathrm{wp,model\_free}}, \qquad (6)$$

where $R_{\mathrm{wp,model\_free}}$ is the residual value obtained in the model-free whole-pattern matching and where $R_{\mathrm{wp,fp}}$ is the value at the end of the restrained fixed-profile Rietveld refinement using as observed data the calculated pattern from the model-free pattern matching. In the absence of impurities, the value of Q must be very close to 1 for correct solutions.

As shown in Fig. 2, one trial is divided into three consecutive stages: the general scan, where an exhaustive exploration of the position (including the orientation) and of the conformation of the molecule in the unit cell is performed, a local incremental scan, where the exploration concentrates around the most promising solution found in the general scan, followed by the fixed-profile Rietveld refinement of the final model (the 'fixed-profile' adjective has been introduced to emphasize that the profile parameters are not refined, *i.e.* those found during the initial model-free pattern matching are employed). In both scans, general and local, the models are refined with FLS. The quality of the intermediate models is measured with

$$\chi_{\mathrm{TALP}} = \left[S_Y/(N_{\mathbf{H}} + N_{\mathrm{rest}} - P)\right]^{1/2}, \qquad (7)$$

wherein $N_{\mathbf{H}}$ is the total number of reflections, N_{rest} the total number of restraints and P the number of refined parameters.

3.1. General scan

By default, random (positional, rotational and torsional) increments are applied to the initial model during the general scan (random exploration mode). For complicated models showing a fragment with a certain conformational rigidity, the general scan may be assisted by the rotation function (assisted exploration mode). The details of both exploration modes are briefly discussed.

3.1.1. Random exploration mode (REM). This is the usual exploration mode because of its versatility. It refines N_{rnd} randomly generated models (*i.e.* random in position, orientation and conformation) with FLS using χ_{TALP} as figure of merit. Of all N_{rnd} refined models, the one with the best χ_{TALP} value passes over to the local scan stage. Before applying FLS, each model is checked for the presence of molecular overlap.

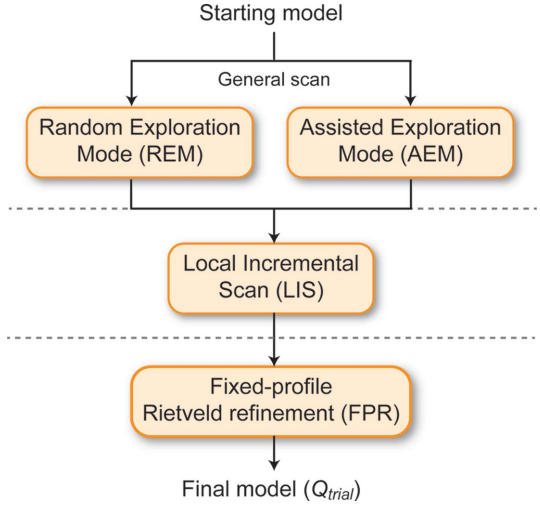

Figure 2
Flow-sheet of one TALP trial.

This is especially important when an FRB moves large parts of the molecule, or if there is more than one molecule in the asymmetric unit. The amount of tolerated molecular overlap is controlled by the global overlap criterion (GOC).

GOC is completely general and can handle crystal structures with $Z' > 1$. To describe its philosophy, the simplest case with one molecule in the asymmetric unit is discussed. A starting molecular model with minimum internal overlap is selected. A cube of length equal to 1.5 Å, with density function values equal to unity (zero outside), is centered at each atomic position, and at each point **r** the density values of all contributing cubes are added. The resulting density function $g_{\rm imol}$ (imol denote the initial molecule) is squared and integrated over the molecular volume:

$$O_{\rm imol} = \int_V g_{\rm imol}^2 \, {\rm d}V. \tag{8}$$

For each randomly generated model, the same procedure is applied to a single molecule ($O_{\rm mol}$) to estimate the $q_{\rm intra}$ ($= O_{\rm mol}/O_{\rm imol}$) ratio measuring the degree of internal molecular overlap. If $q_{\rm intra}$ is greater than ~1.1, internal overlap is assumed and a new model is generated. Otherwise, the integral

$$O_{\rm all} = \int_V g_{\rm all}^2 \, {\rm d}V \tag{9}$$

is calculated ($g_{\rm all}$ refers to all symmetry-related molecules) and the amount of intermolecular overlap is estimated according to $q_{\rm inter} = O_{\rm all}/(ZO_{\rm mol})$. If $q_{\rm inter}$ exceeds 1.05, this model is abandoned and a new one generated.

In the general scan, the number of FLS refinement cycles ($N_{\rm FLS}$) lies between 10 and 30. Both $N_{\rm rnd}$ and $N_{\rm FLS}$ largely depend on the model complexity ($\simeq$ number of FRBs). The idea of local minimizations of random models used in REM is similar to the one employed by Shankland *et al.* (2010) as a benchmark method for global optimization problems in SDPD.

3.1.2. Assisted exploration mode (AEM). In this mode, the rotation function assists the exploration. First of all, that part of the model having fixed or known geometry must be identified, since it will constitute the search model. The rotation function, Rot, was first introduced by Rossmann & Blow (1962) and takes advantage of the fact that Patterson peaks corresponding to intramolecular interatomic vectors depend only on the orientation of the molecule, not on its position in the unit cell. For the special case discussed here, it takes the simple form

$${\rm Rot}(\mathbf{\Omega}) = \int_V P'_{\rm obs} \, P_{\rm mod}(\mathbf{\Omega}) \, {\rm d}V, \tag{10}$$

where $P'_{\rm obs}$ is the Patterson function (with suppressed origin peak) derived from the $y_{{\rm obs},\mathbf{H}}$ values and $P_{\rm mod}$ is the Patterson function of the search model rotated by $\mathbf{\Omega}$. The rotation function measures the coincidence between the two functions. A high Rot value means a high coincidence and, consequently, that this particular $\mathbf{\Omega}$ rotation is a probable solution. TALP automatically eliminates the Rot solutions that are equivalent by crystallographic symmetry (Rius & Miravitlles, 1987). In AEM, the rotation to be applied to the complete starting model (not just to the search model) is randomly selected from the best Rot solutions. In this way, the solution process is simplified since only the positional and conformational parameters need to be explored.

3.2. Local incremental scan (LIS)

In this second stage, the best model obtained in the general scan is further optimized by applying increments of variable size. Depending on the size of the increments two types of iteration may be distinguished:

(*a*) Coarse increment iteration (CI). This starts with the model with lowest $\chi_{\rm TALP}$ found in the general scan and explores nearby positions, orientations and conformations by applying increments of limited size to this model. The updated model is refined by FLS and the final $\chi_{\rm TALP}$ is computed. If the new $\chi_{\rm TALP}$ value is lower, the starting model for the next iteration is replaced by the new one. The maximum sizes of the increments decrease as the CI iteration proceeds ($N_{\rm CI}$ = total number of iteration cycles). This procedure reduces the local minima sticking problem of least-squares minimizations. In each iteration cycle all increments are applied simultaneously to (i) the (x, y, z) coordinates of the center of the molecule (Δt), (ii) the spin angle ($\Delta\psi$) around the randomly selected spin axis of the molecule and (iii) each FRB angle ($\Delta\tau$). The initial and final values for the maximum increment sizes are $\Delta t_{\rm max}$ = 1.5–0.5 Å, $\Delta\psi_{\rm max}$ = 20–5° and $\Delta\tau_{\rm max}$ = 35–10°. The increments are obtained from normal distributions defined by the corresponding σ values (equal to $\Delta t_{\rm max}$, $\Delta\psi_{\rm max}$ and $\Delta\tau_{\rm max}$) centered at the starting model values.

(*b*) Fine increment iteration (FI). This proceeds as the CI part but with Δt, $\Delta\psi$ and $\Delta\tau$ being constant and equal to 0.05 Å, 1° and 2°, respectively. The number of FI iteration cycles is 500 and the number of FLS cycles is 5.

3.3. Fixed-profile Rietveld refinement (FPR)

The model resulting from the local incremental scan undergoes 30 cycles of fixed-profile restrained Rietveld refinement using the complete powder pattern to obtain more accurate structural parameters and to derive the corresponding $Q_{\rm trial}$ figure of merit. In this final refinement stage, besides the bond distance and angle restraints, plane restraints may be considered. Logically, this is only the case for molecules having clear planar regions.

4. Solving crystal structures with TALP

TALP has been implemented in the form of a computer program and tested on known and unknown crystal structures of molecular compounds of variable degrees of difficulty (Fig. 3). Data for the test compounds (1)–(9) were taken from transmission geometry experiments by Florence *et al.* (2005): carbamazepine β polymorph, (1), paracetamol form I polymorph, (2), captopril, (3), zopiclone dihydrate, (4), salbutamol, (5), (*S*)-ibuprofen, (6), famotidine, (7), capsaicin, (8), and verapamil hydrochloride, (9). Data for diphenhydramine

Table 2
Crystallographic data for compounds (1)–(14).

The CSD code identifies the CIF used to generate the molecular model and the restraints (see Table 4); V is the unit-cell volume; d_{min} is the smallest useful d spacing of the powder diffraction data.

Compound number	Formula	Z	Space group	V (Å^3)	d_{min} (Å)	CSD code (reference)†
(1)	$C_{15}H_{12}N_2O$	4	$P2_1/c$	1166.4	1.43	CBMZPN10 (*a*)
(2)	$C_8H_9NO_2$	4	$P2_1/n$	771.6	1.44	HXACAN07 (*b*)
(3)	$C_9H_{15}NO_3S$	4	$P2_12_12_1$	1077.4	1.43	MCPRPL (*c*)
(4)	$C_{17}H_{17}ClN_6O_3{\cdot}2H_2O$	4	$P2_1/c$	1874.6	1.44	UCUVET (*d*)
(5)	$C_{13}H_{21}NO_3$	8	$Pbca$	2822.8	1.44	BHHPHE (*e*)
(6)	$C_{13}H_{18}O_2$	8	$P2_1$	1244.7	1.43	JEKNOC10 (*f*)
(7)	$C_8H_{15}N_7O_2S_3$	4	$P2_1/c$	1421.8	1.44	FOGVIG03 (*g*)
(8)	$C_{18}H_{27}NO_3$	4	$P2_1/c$	1077.4	1.43	FABVAF01 (*h*)
(9)	$C_{27}H_{39}N_2O_4{\cdot}Cl$	2	$P\overline{1}$	1382.1	1.43	CURHOM (*i*)
(10)	$C_{17}H_{22}NO{\cdot}Cl$	4	$Pna2_1$	1639.6	1.35	JEMJOA (*j*)
(11)	$C_{21}H_{14}N_4O_2S$	8	$Pbca$	3470.9	1.64	–
(12)	$C_{16}H_{22}N_6$	3	$P3$	1208.0	2.01	–
(13)	$C_7H_{11}Cl_2N_2OZn$	4	$P2_1/a$	1135.0	1.20	–
(14)	$C_{14}H_{22}Cd_2Cl_4N_4O_2$	4	$C2/c$	2336.6	1.20	–

† References: (*a*) Himes *et al.* (1981); (*b*) Nichols & Frampton (1998); (*c*) Fujinaga & James (1980); (*d*) Shankland *et al.* (2001); (*e*) Beale & Grainger (1972); (*f*) Freer *et al.* (1993); (*g*) Florence *et al.* (2003); (*h*) David *et al.* (1998); (*i*) Carpy *et al.* (1985); (*j*) Glaser & Maartmann-Moe (1990).

hydrochloride, (10), and (*Z*)-3-methyl-*N*-(7-nitroacridin-3-yl)-2,3-dihydro-1,3-benzothiazol-2-imine, (11), were collected on a PANalytical X'Pert PRO MPD in transmission geometry (Vallcorba *et al.*, 2011); for tris(3,5-dimethylpyrazol-1-yl)-methane, (12), data were measured with an INEL cylindrical position-sensitive detector (CPS120) using Debye–Scherrer geometry (Ochando *et al.*, 1997); and for compounds $[ZnCl_2(C_7H_{11}N_2O)]$, (13), and $[CdCl_2(C_7H_{11}N_2O)]_2$, (14), data were acquired on a Siemens D5000 diffractometer in Bragg–Brentano geometry (Guerrero *et al.*, 2011). In all cases, the QCK intensity data file necessary for TALP was derived from a preliminary whole-pattern matching with *DAJUST* (Vallcorba *et al.*, 2012). Additional crystallographic details for compounds (1)–(14) are listed in Table 2.

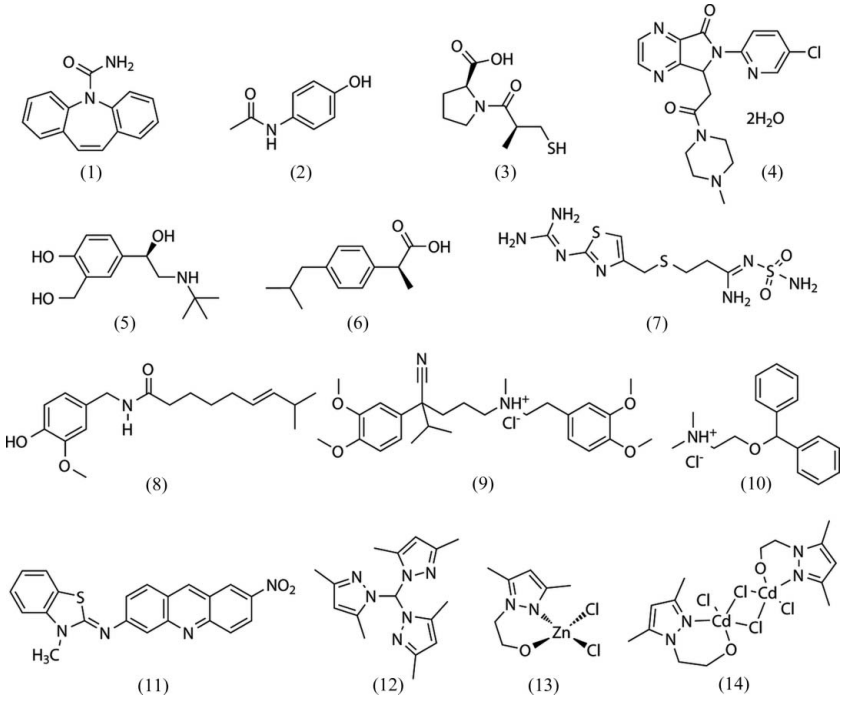

Figure 3
Molecular structures of compounds (1)–(14) used as test examples.

Table 3
Description of pre-fixed experimental conditions (*A*–*E*) used by the TALP strategy.

N_{rnd} is the number of randomly generated models in the scan stage; N_{FLS} (REM/AEM) is the number of FLS cycles applied to models generated by REM or AEM; N_{CI} (N_{FI}) is the number of CI (FI) cycles; N_{FLS} (CI) [N_{FLS} (FI)] is the number of FLS cycles at each CI (FI).

	A	*B*	*C*	*D*	*E*
N_{rnd}	1500	3000	6000	10000	20000
N_{FLS} (REM/AEM)	20	20	30	30	30
N_{CI}	3000	6000	8000	15000	30000
N_{FLS} (CI)	15	15	15	15	15
N_{FI}	200	200	200	200	200
N_{FLS} (FI)	5	5	5	5	5

Initial model coordinates and values for distance restraints were retrieved from crystal structures deposited in the Cambridge Structural Database (CSD; Version 5.33; Allen, 2002) or derived from molecular mechanics calculations with CS (*Chem3D Pro*; CambridgeSoft, 2012). H atoms have been omitted during structure solution and torsion angles are considered as (random) FRBs each time REM/AEM generates a new model. No additional conformational information is used in the solution process.

The TALP experimental conditions must be necessarily different depending on the overall complexity of the crystal structure. This complexity basically comprises the number of FRBs (conformational freedom), the number of refined parameters (molecular size), the number of symmetry-independent molecules in the asymmetric unit (Z') and the quality of the diffraction data. As the complexity increases, more minimizations and more time are required. Table 3 contains the parameters of the five experimental conditions that have been pre-fixed in the TALP strategy implementation and are applied depending on the structural complexity, thereby always seeking the best compromise between speed and number of solutions. To solve the crystal structures, only reflections up to 3 Å of *d*-spacing resolution are considered, except for the fine increment iteration and the FPR where all available data are used.

The crystal structures of compounds (1)–(14) have been dealt with satisfactorily by TALP and the results are summarized in Table 4. The Q_{trial} figure of merit [equation (6)], the main indicator for candidate solutions, is normally below 1.5 for correct solutions. Q_{trial} values below 1.0 indicate a high

Table 4
Summary of the application of TALP to structures (1)–(14).

Model is the initial molecular model and restraints derived from single-crystal coordinates (CIF) or from molecular mechanics (MEC) by using *Chem3D Pro*; FRB is the number of free rotation bonds in the model; N_{par} is the number of refined parameters (atomic coordinates, overall displacement parameter, scale factor); N_{rest} is the number of distance restraints; Cond. is the pre-fixed experimental condition used (Table 3); N_{sol} is the number of solutions out of 100 trials; r.m.s.d. is the root mean square deviation between the best solution and reference structure; t_{trial} is the average time per trial; $t_{sol} = 100(t_{trial})/N_{sol}$ is the estimated time for reaching one solution on an Intel Core i7 (2.80 GHz) CPU.

Code	Model	FRB	N_{rest}	N_{par}	Cond.	N_{sol}	Q_{trial}	R.m.s.d. (Å)	t_{trial} (min)	t_{sol} (min)
(1)	CIF	1	64	56	*A*	100	1.081	0.064	1.5	1.5
(1)	MEC	1	64	56	*A*	99	1.112	0.048	1.5	1.5
(2)	CIF	2	37	35	*A*	100	0.961	0.060	0.5	0.5
(2)	MEC	2	37	35	*A*	100	1.200	0.057	0.5	0.5
(3)	CIF	4	42	44	*A*	51	0.929	0.105	1.6	3.1
(3)	MEC	4	42	44	*A*	50	0.950	0.091	1.6	3.2
(4)	CIF	4	92	89	*A*†	18	1.145	0.127	4.0	22.2
(4)	MEC	4	92	89	*A*†	19	1.443	0.168	4.0	21.1
(5)	CIF	5	50	53	*A*	69	0.994	0.138	1.5	2.2
(5)	MEC	5	50	53	*A*	74	0.950	0.151	1.5	2.0
(6)	CIF	8	86	91	*A*	16	1.116	0.154	3.6	22.5
(6)	MEC	8	86	91	*A*	20	1.094	0.103	3.6	18.0
(7)	CIF	9	47	62	*D*	17	1.273	0.072	11.1	65.3
(7)	MEC	9	47	62	*D*	19	1.309	0.066	11.1	58.4
(8)	CIF	11	66	68	*A*	13	0.953	0.323	2.6	20.0
(8)	CIF	11	66	68	*A*‡	35	0.953	0.323	2.6	7.4
(8)	MEC	11	66	68	*A*‡	36	0.953	0.323	2.6	7.2
(9)	CIF	13	99	104	*D*	0	–	–	37.2	–
(9)	CIF	13	99	104	*D*‡	13	1.469	0.141	37.2	286.2
(9)	MEC	13	99	104	*D*‡	0	–	–	37.2	–
(10)	CIF	6	59	44	*A*	100	0.942	0.290	1.4	1.4
(10)	MEC	6	59	44	*A*	100	1.008	0.281	1.4	1.4
(11)	MEC	3	110	91	*A*	85	1.976	0.113§	3.7	4.4
(12)	MEC	3	57	68	*A*	100	0.738	0.127§	1.0	1.0
(13)	MEC	0	43	41	*A*	100	0.998	0.397§	1.2	1.2
(14)	MEC	0	43	41	*A*	100	0.817	0.105§	1.3	1.3

† The two missing water molecules were found in a subsequent straightforward TALP run. ‡ AEM used in the general scan stage. § No single-crystal data available; r.m.s.d. calculated by comparing TALP with *RIBOLS* solutions.

R_{wp} in the model-free pattern matching, which is principally due to profile/background problems or to the presence of small impurity peaks. The final proof of the correctness of a solution obviously comes from the visual inspection of the found crystal structure and from the discrepancies in the Rietveld fit. If a reference structure is available, the quality of the model can be objectively estimated by calculating the root mean square deviation (r.m.s.d.) between refined and (error-free) reference models.

In Table 4 the experimental conditions have been adjusted to reach a solution in the shortest time (t_{sol}). t_{sol} is calculated as $t_{trial}(N_{trial}/N_{sol})$, where N_{trial} is the total number of TALP trials (here 100), t_{trial} the average time per trial and N_{sol} the total number of solutions. For (1)–(14) the softest conditions (*A*) give the lowest t_{sol} in all cases except for (7) and (9), where stronger conditions (*D*) are needed to obtain a representative N_{sol} value. For structures with up to 11 torsion angles [(1)–(8), (10)–(14)] TALP finds the correct crystal structure in times varying from 0.5 min to 1 h. Besides their conformational flexibility, these structures do show other difficulties, such as *e.g.* $Z' = 2$ in the case of (7), and the presence of crystallization solvents for (4), salts for (10), atoms in special positions for (12), and heavy atoms like Zn and Cd for (13) and (14). For the most complex test structure solved with TALP, *i.e.* (9), a longer time and use of AEM is needed to reach the solution. Inclusion of the FPR stage at the end of each TALP trial delivers accurate models as confirmed by the small r.m.s.d. values.

Logically, the number of correct solutions increases when more demanding experimental conditions are selected. This is illustrated in Fig. 4(*a*) where the evolution of N_{sol} as a function of the different conditions *A*, *B*, *C*, *D* and *E* is plotted. In general, however, this is at the cost of longer t_{trial} and t_{sol} times, as can be seen in Fig. 4(*b*). For compounds (3) and (5), all 100 trials reach the correct solution but t_{sol} increases almost exponentially when going from conditions *A* to *E*. For more complex structures, (6), (7) and (8), the evolution of t_{sol} is more linear, *i.e.* the longer t_{trial} times tend to be compensated by higher N_{sol} values.

4.1. AEM for compounds (8) and (9)

For compounds (8) and (9), both containing long atomic chains anchored on terminal aryl groups with 11 and 13 torsion angles, respectively, application of AEM represents a significant improvement. In the case of (8), the search fragment used in the rotation function calculations was the

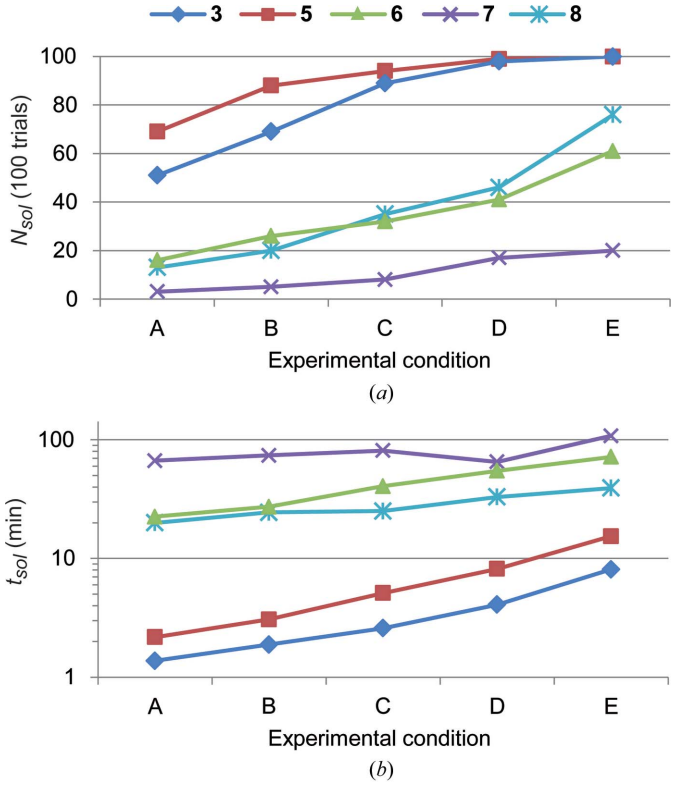

Figure 4
Effect of preset experimental conditions (*A*, *B*,*C*, *D*, *E*) of TALP on (*a*) the number of solutions (N_{sol}) out of 100 trials and (*b*) the estimated minimum time for obtaining a solution (t_{sol}) for test compounds (3), (5), (6), (7) and (8) using REM.

terminal phenyl ring with the direct bonded atoms (Fig. 5). The rotation function found the four symmetry-related fragment orientations that were used to produce the starting models for the general scan. By using the same pre-fixed conditions (*A*), AEM leads to 35 correct solutions whereas REM yields only 13. For compound (9) the difference in behavior between AEM and REM is more pronounced. As search fragment for the rotation function calculation one of the terminal phenyl rings with the immediate substituents was selected (Fig. 5). Combination of the *D* pre-fixed conditions with AEM leads to 13 correct (plus seven nearly correct) solutions, whereas REM only gives one nearly correct solution. It is instructive to look in more detail at one of the seven nearly correct solutions ($1.5 < Q_{\rm trial} < 2.5$). As can be seen in Fig. 6, these are characterized by having the second terminal phenyl ring of the molecule rotated by 180°, which can be easily noticed because of slight distortions in the resulting crystal structure. In these cases, a simple 180° rotation of the terminal phenyl ring followed by 20 cycles of FPR leads to the correct solution.

According to these preliminary results, AEM seems to be particularly useful for crystal structures with several FRBs distributed in long atomic chains. In these cases, the search for the correct conformation is more effective as the origin of the long organic chain always starts pointing in the correct direction. Further developments of AEM should consider the local symmetry of the search fragment in order to increase the efficiency of the rotation function.

4.2. Molecular model influence

To evaluate the impact of the model accuracy on the success ratio, duplicate experiments have been carried out with initial molecular models and restraints from (i) a reference crystal structure (CIF) and (ii) a molecular mechanics calculation (MEC). As expected, solutions from less accurate models have slightly higher $Q_{\rm trial}$ values compared to solutions obtained with the true model. This can be seen in Table 4 for compound (4), whose accurate model is difficult to calculate owing to its complexity. However, for most test compounds, almost equivalent $Q_{\rm trial}$ and r.m.s.d. values are obtained with data of both sources. This also indicates that the variance assigned to each restraint and the relative weight of $S_{\rm R}$ in the *M* function are flexible enough to allow small discrepancies between the initial molecular model and the final correct solution.

In contrast, the Rot calculation in AEM is very sensitive to the fragment accuracy. While for (8) the calculated fragment is obtained very accurately (phenyl ring) and no differences are observed in comparison with the true model, in the case of (9) correct solutions are only found when the model from the CIF is used.

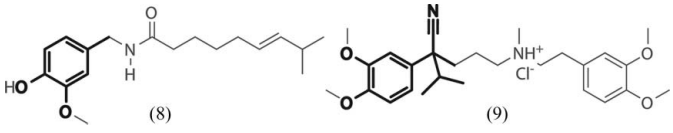

Figure 5
Search fragments of compounds (8) and (9) (bold) used for the respective rotation function calculations in AEM.

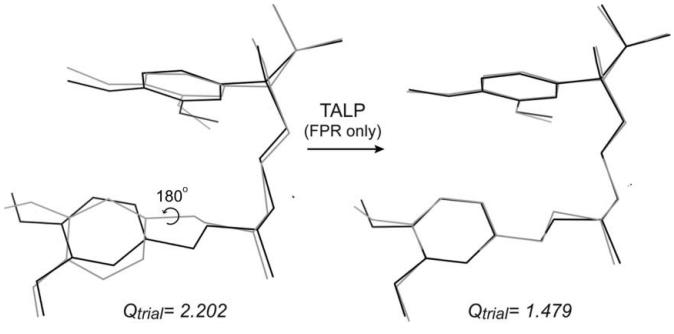

Figure 6
Superposition of the reference crystal structure of (9) (black) with a nearly correct solution (left, grey) and with the final solution after a 180° ring rotation and the subsequent fixed-profile Rietveld refinement (right, grey).

5. Conclusions

The direct-space global optimization strategy TALP is a useful tool for solving crystal structures of molecular compounds from powder data. Thanks to the three-step intensity extraction procedure performed by *DAJUST* and also to the use during the least-squares refinements of profile intensities at Bragg peak centers only (properly speaking this is only the case during general scans and CI/FI iterations), least-squares calculations are very fast with practically no control parameters. The fact that the refined variables are the atomic coordinates and that the molecular geometry is introduced in the form of restraints makes the strategy very versatile. In addition, incorporation of GOC into the general scans allows the rejection of less probable initial models; however, as a result of the fuzzy-logic characteristic of this criterion a certain atomic overlap is tolerated. For the test structures (1)–(14) simple combinations of general and local scans with least-squares minimizations have efficiently found the global minimum. In the last stage of the refinement process, FPR has proved to be a very efficient step to obtain structures almost identical to the reported ones since the corresponding calculated r.m.s.d. values are around 0.15 Å. At present the method can cope either with models having 13 FRBs or with crystal structures containing more than one symmetry-independent molecule but with fewer FRBs. In the case of crystal structures containing molecules formed by a known fragment with long chains attached to it, the preliminary exploration with the rotation function (AEM) is more effective. As already known from the Patterson search experience, it is better to introduce a small but accurate search fragment rather than a large approximate one.

The TALP strategy has been implemented in a Fortran computer program, which is controlled by a user-friendly Java graphical user interface. It can be downloaded (like *DAJUST*) from the authors' institution web site (http://www.icmab.es/crystallography/software) subject to the conditions of use specified therein. All test calculations have been performed with this software.

The authors thank the Spanish 'Ministerio de Ciencia e Innovación' (projects MAT2009-07967 and Consolider NANOSELECT CSD2007-00041) and the 'Generalitat de Catalunya' SQR(2009) for financial support. OV also thanks CSIC for a contract (PIE 201060E068).

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.

Baerlocher, C. (1993). *The Rietveld Method*, edited by R. A. Young. Oxford University Press.

Beale, J. P. & Grainger, C. T. (1972). *Cryst. Struct. Commun.* **1**, 71–74.

Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.

CambridgeSoft (2012). *Chem3D Pro*. Version 12. PerkinElmer Informatics, Cambridge, MA, USA.

Carpy, A., Léger, J.-M. & Melchiorre, C. (1985). *Acta Cryst.* C**41**, 624–627.

Černý, R. & Favre-Nicolin, V. (2007). *Z. Kristallogr.* **222**, 105–113.

Chong, S. Y. & Tremayne, M. (2006). *Chem. Commun.* pp. 4078–4080.

Coelho, A. A. (2000). *J. Appl. Cryst.* **33**, 899–908.

David, W. I. F. & Shankland, K. (2008). *Acta Cryst.* A**64**, 52–64.

David, W. I. F., Shankland, K., 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.

Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.

Favre-Nicolin, V. & Černý, R. (2002). *J. Appl. Cryst.* **35**, 734–743.

Favre-Nicolin, V. & Černý, R. (2004). *Z. Kristallogr.* **2 19**, 847–856.

Florence, A. J., Baumgartner, B., Weston, C., Shankland, N., Kennedy, A. R., Shankland, K. & David, W. I. (2003). *J. Pharm. Sci.* **92**, 1930–1938.

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.

Freer, A. A., Bunyan, J. M., Shankland, N. & Sheen, D. B. (1993). *Acta Cryst.* C**49**, 1378–1380.

Fujinaga, M. & James, M. N. G. (1980). *Acta Cryst.* B**36**, 3196–3199.

Glaser, R. & Maartmann-Moe, K. (1990). *J. Chem. Soc. Perkin Trans. 2*, pp. 1205–1210.

Guerrero, M., Pons, J., Ros, J., Font-Bardia, M., Vallcorba, O., Rius, J., Branchadell, V. & Merkoçi, A. (2011). *CrystEngComm*, **13**, 6457–6470.

Harris, K. D. M., Habershon, S., Cheung, E. Y. & Johnston, R. L. (2004). *Z. Kristallogr.* **219**, 838–846.

Harris, K. D. M., Tremayne, M. & Kariuki, B. M. (2001). *Angew. Chem. Int. Ed.* **40**, 1626–1651.

Himes, V. L., Mighell, A. D. & De Camp, W. H. (1981). *Acta Cryst.* B**37**, 2242–2245.

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

Le Bail, A. (2001). *Mater. Sci. Forum*, **378–381**, 65–70.

Nichols, G. & Frampton, C. S. (1998). *J. Pharm. Sci.* **87**, 684–693.

Ochando, L. E., Rius, J., Louër, D., Claramunt, R. M., Lopez, C., Elguero, J. & Amigó, J. M. (1997). *Acta Cryst.* B**53**, 939–944.

Pagola, S. & Stephens, P. W. (2010). *J. Appl. Cryst.* **43**, 370–376.

Rius, J. (2012). *RIBOLS.* Institut de Ciència de Materials de Barcelona (ICMAB–CSIC), Catalunya, Spain.

Rius, J. & Miravitlles, C. (1987). *J. Appl. Cryst.* **20**, 261–264.

Rius, J. & Miravitlles, C. (1988). *J. Appl. Cryst.* **21**, 224–227.

Rius, J., Miravitlles, C., Molins, E., Crespo, M. & Veciana, J. (1990). *Mol. Cryst. Liq. Cryst.* **187**, 155–163.

Rius, J., Sañé, J., Miravitlles, C., Amigó, J. M., Reventós, M. M. & Louër, D. (1996). *An. Quim. Int.* **92**, 223–227.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Shankland, K. (2005). *IUCr Commission on Crystallographic Computing Newsletter* No. 5, pp. 92–102.

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

Shankland, N., David, W. I. F., Shankland, K., Kennedy, A. R., Frampton, C. S. & Florence, A. J. (2001). *Chem. Commun.* pp. 2204–2205.

Shankland, K., Markvardsen, A. J., Rowlatt, C., Shankland, N. & David, W. I. F. (2010). *J. Appl. Cryst.* **43**, 401–406.

Vallcorba, O., Latorre, S., Alcobé, X., Miravitlles, C. & Rius, J. (2011). *Acta Cryst.* C**67**, o425–o427.

Vallcorba, O., Rius, J., Frontera, C., Peral, I. & Miravitlles, C. (2012). *J. Appl. Cryst.* **45**, 844–848.