Journal of
Applied Crystallography
ISSN 0021-8898

Received 28 October 2010
Accepted 22 December 2010

EPCryst: a computer program for solving crystal structures from powder diffraction data

Xiaodi Deng* and Cheng Dong*

National Laboratory for Superconductivity, Institute of Physics, and Beijing National Laboratory for Condensed Matter Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China. Correspondence e-mail: dengxiaodi@ssc.iphy.ac.cn, chengdon@aphy.iphy.ac.cn

Recently, a new algorithm has been proposed to generate a full set of trial structure models (TSMs) by enumerating all possible equivalent position combinations (EPCs) based on the unit-cell content and space-group information [Deng & Dong (2009). *J. Appl. Cryst.* **42**, 953–958]. Using this algorithm, a new computer program named *EPCryst* has been developed for crystal structure determination from powder diffraction data. It is designed to solve a crystal structure as efficiently as possible. Rather than applying a time-consuming global optimization procedure on each TSM, *EPCryst* firstly tries to eliminate improbable TSMs. Several methods (such as the statistical analysis method and direct-space heavy-atom method) are designed to achieve this goal. Usually, a lot of improbable TSMs can be eliminated and only a few promising TSMs are preserved for global optimization by grid search or simulated annealing. These methods can greatly increase the efficiency of structure solution in direct space. Bond-length checking is available in *EPCryst* as a tool for crystal structure validation. *EPCryst* was successfully applied to determine several inorganic crystal structures using powder diffraction data obtained from PowBase [Le Bail (2000). http://sdpd.univ-lemans.fr/powbase/].

1. Introduction

The direct-space method has been used widely and successfully for crystal structure determination from powder diffraction (SDPD) data in recent years (Černý, 2008; David & Shankland, 2008). The process involves finding the best agreements between calculated and observed diffraction patterns by using global optimization methods. Several global optimization algorithms have been successfully implemented in SDPD programs, such as Monte Carlo search (Harris *et al.*, 1994), simulated annealing (SA) (Newsam *et al.*, 1992; Adreev *et al.*, 1997) or parallel tempering (Favre-Nicolin & Černý, 2002; Earl & Deem, 2005), genetic algorithm (Shankland *et al.*, 1997; Kariuki *et al.*, 1997; Harris *et al.*, 1998; Feng & Dong, 2007), and particle swarm algorithm (Csoka & David, 1999; Feng *et al.*, 2009). To date, the majority of SDPD programs using the direct-space method have been designed for organic molecular crystals, because molecules with known connectivity lead to an easy parameterization for global optimization (David & Shankland, 2008). Building trial models of inorganic crystals is more difficult than for molecular crystals, although some progress has been made in this direction, especially for zeolites (Deem & Newsam, 1989; Falcioni & Deem, 1999) and inorganic crystals with known polyhedral connectivity (Favre-Nicolin & Černý, 2002).

A general method is to describe every atom in the crystal with fractional coordinates (x, y, z) in the asymmetric unit of the unit cell (David *et al.*, 2002). This will need $3N$ parameters for a crystal with N independent atoms. The complexity of the global optimization problem is exponentially related to the number of parameters, and it becomes computationally intractable when N grows larger. In addition, some atoms often occupy special positions in inorganic crystals, and these atoms must be handled by some time-consuming algorithms such as the 'dynamical occupancy correction' used in the *FOX* program (Favre-Nicolin & Černý, 2002).

Recently, we have proposed an algorithm to generate trial structure models (TSMs) from the unit-cell content and space-group information, and TSMs are generated by enumerating every possible equivalent position combination (EPC) (Deng & Dong, 2009). This EPC algorithm is particularly suitable for dealing with inorganic crystals because information about the molecular connectivity or coordination polyhedron is not required for model generation. It also has a significant advantage because it facilitates the structure solution in direct space by using the 'divide and conquer' strategy. Using the EPC algorithm, a $3N$-parameter global optimization problem is naturally divided into a set of smaller ones with fewer parameters. Consequently, the efficiency of the direct-space search can be greatly increased, and some previously intractable structures could be solved effectively.

This paper describes a new computer program, *EPCryst*, for crystal structure solution from X-ray and neutron powder diffraction data using TSMs generated by the EPC algorithm. The program searches for the structural parameters of each TSM by global optimization to find the best agreement (*i.e.* the lowest grouped Bragg R factor, $R_{\rm Bg}$, as defined in §3.1) between calculated and observed diffraction patterns. The best TSM with determined structural parameters is selected as the structure solution. The obtained structure solutions can be used as the initial crystal data for structural refinement by the Rietveld method.

The number of TSMs generated by the EPC algorithm could be rather large when the structure is very complex or when the space group has many Wyckoff positions with the same multiplicity. In order to speed up the structure solution process, several methods

Table 1
Wyckoff positions of space group *Fmmm*.

Wyckoff position	Site symmetry	Coordinate (partial)
32*p*	1	x, y, z
16*o*	..*m*	$x, y, 0$
16*n*	.*m*.	$x, 0, z$
16*m*	*m*..	$0, y, z$
16*l*	2..	$x, \frac{1}{4}, \frac{1}{4}$
16*k*	.2.	$\frac{1}{4}, y, \frac{1}{4}$
16*j*	..2	$\frac{1}{4}, y, \frac{1}{4}$
8*i*	*mm*2	$0, 0, z$
8*h*	*m*2*m*	$0, y, 0$
8*g*	2*mm*	$x, 0, 0$
8*f*	222	$\frac{1}{4}, \frac{1}{4}, \frac{1}{4}$
8*e*	..2/*m*	$\frac{1}{4}, \frac{1}{4}, 0$
8*d*	.2/*m*.	$\frac{1}{4}, 0, \frac{1}{4}$
8*c*	2/*m*..	$0, \frac{1}{4}, \frac{1}{4}$
4*b*	*mmm*	$0, 0, \frac{1}{2}$
4*a*	*mmm*	$0, 0, 0$

were used to effectively eliminate the redundant and improbable TSMs before global optimization. These strategies will be described in subsequent sections.

2. Methodology

Generally speaking, solving a crystal structure with N parameters by the direct-space method corresponds to finding the global minimum of an N-dimensional hypersurface. Usually, the shape of the hypersurface is very complex, and exploring the whole hypersurface to find the global minimum requires appropriate algorithms with carefully choose parameters (Černý, 2008; Favre-Nicolin & Černý, 2002). From this point of view, model generation by the EPC algorithm is a process of dividing the whole hypersurface into a set of smaller sections, and each generated TSM corresponds to a hypersurface section. These sections can be completely described in lower-dimensional spaces whenever some atoms are located in special positions, because the number of atomic coordinate parameters in special positions is always less than that in general positions.

Fig. 1 illustrates the main flow of *EPCryst*. The structure solution process consists of three main steps. The first step is to generate TSMs by the EPC algorithm based on the unit-cell content and space-group symmetry. In the second step, Monte Carlo tests are carried out on the generated TSMs to eliminate the improbable TSMs (see §2.2 for details). The third and final step is to solve the crystal structure by global optimization for the TSMs preserved in the second step. At every step of the structure solution, *EPCryst* is designed to give sufficient flexibility in the use of various strategies aiming to eliminate the bad TSMs and accelerate the global optimizations.

2.1. Elimination of equivalent TSMs

Reducing the number of generated TSMs is a very effective way to simplify the structure solving process. Equivalent TSMs are TSMs that describe identical crystal structures in different ways, and the only difference between two equivalent TSMs is the different choice of the unit-cell origin, so two equivalent TSMs can interconvert into each other by a coordinate translation. Equivalent TSMs commonly occur when some different Wyckoff positions of a space group have the same site symmetry.

Take the La_2CuO_4 (space group *Fmmm*) crystal structure as an example. There are four formula units in the unit cell, and the corresponding Wyckoff positions are listed in Table 1. The EPC algorithm will generate in total 372 TSMs when no additional restraints are used. However, many of the generated TSMs are practically equivalent. Table 2 gives an example of two equivalent TSMs taken from the 372 TSMs. Referring to Table 1, we can see that TSM-*b* is equivalent to TSM-*a* because it can be obtained by a constant coordinate translation $(0, 0, \frac{1}{2})$ on TSM-*a*, and this translation transform can be described as follows:

Table 2
Two equivalent TSMs.

Atoms	TSM-*a*†	TSM-*b*†
La	8*i*	8*i*
Cu	4*a*	4*b*
O	16*o*	16*o*

† TSM-*a*: La(8*i*) + Cu(4*a*) + O(16*o*); TSM-*b*: La(8*i*) + Cu(4*b*) + O(16*o*).

$$8i + \left(0, 0, \tfrac{1}{2}\right) = \left(0, 0, z + \tfrac{1}{2}\right) \equiv 8i, \quad (1)$$

$$4a + \left(0, 0, \tfrac{1}{2}\right) = \left(0, 0, \tfrac{1}{2}\right) \equiv 4b, \quad (2)$$

$$16o + \left(0, 0, \tfrac{1}{2}\right) = \left(x, y, \tfrac{1}{2}\right) \equiv 16o. \quad (3)$$

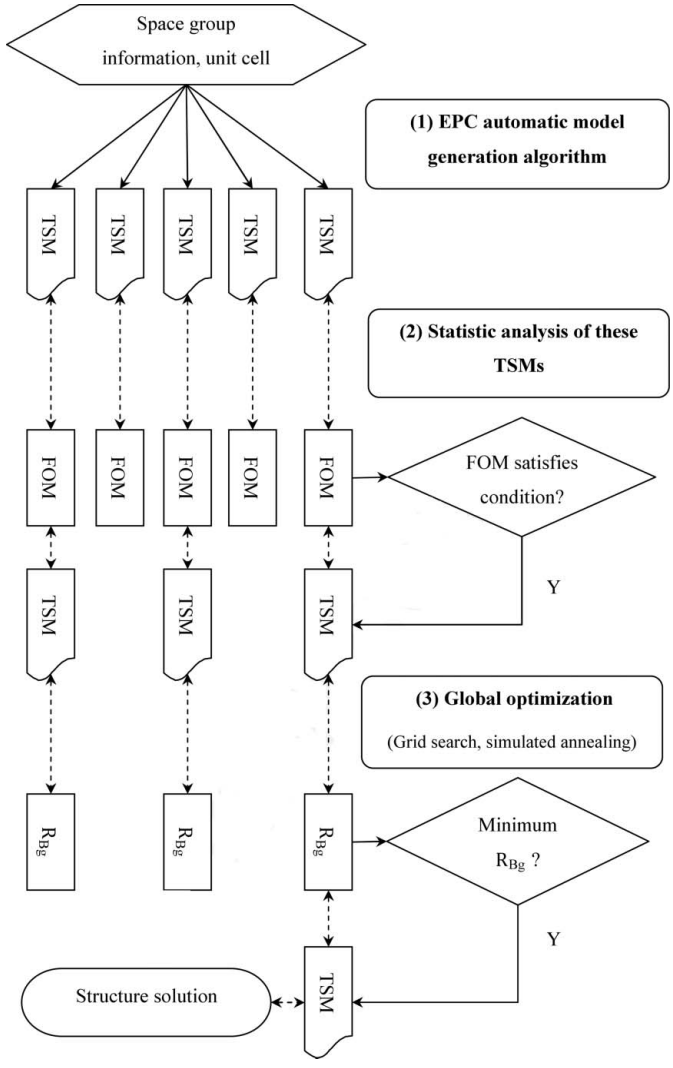

Figure 1
Flow chart of *EPCryst*.

Because the four Cu atoms can be located only in either the '4*a*' position or the '4*b*' position, it is easy to see that the above 372 TSMs are actually 186 pairs of equivalent TSMs. The global optimization search will find the same crystal structure for equivalent TSMs. We can save a lot of time if we can eliminate the redundant equivalent TSMs. In this example, we can put Cu atoms in either the '4*a*' or the '4*b*' position before model generation so that the number of generated TSMs will be halved to 186.

Elimination of the equivalent TSMs is realized by imposing restraints on the occupancy of the Wyckoff positions. We use ($R_{\rm min}$, $R_{\rm max}$) pairs to impose a restraint on the numbers of atoms in a specific Wyckoff position (Deng & Dong, 2009). Here, $R_{\rm min}$ and $R_{\rm max}$ represent the minimum and maximum numbers of atoms in the Wyckoff positions, respectively. We can control the model generation by editing the values of the ($R_{\rm min}$, $R_{\rm max}$) pairs. For example, to fix the Cu atoms in the 4*a* position, we set both $R_{\rm min}$ and $R_{\rm max}$ for the 4*a* positions to 1. Simultaneously, we must set both $R_{\rm min}$ and $R_{\rm max}$ for the 4*b* positions to 0 so that Cu atoms are forbidden from occupying the 4*b* positions. *EPCryst* provides the input interfaces (Fig. 2) for editing restraints on each Wyckoff position for every atom.

Besides the restraints used in the EPC algorithm and described above for elimination of the equivalent TSMs, other restraints can also be imposed to shorten the list of generated TSMs. The additional constrains can be specified on the basis of the chemical and physical information, such as the unit-cell dimensions, atomic radii and bonding characteristics of the atoms. Heavy atoms can be fixed in positions obtained by the Patterson method or the heavy-atom method in direct space (see §2.4). Moreover, it is possible to place the restraints by analogy with chemically similar materials (David & Shankland, 2008) or structure predictions (Abrahams, 2010; Donnay *et al.*, 1964; Le Bail, 2005).

2.2. Statistical analysis

The number of TSMs is sometimes rather large, even though the specified restraints are imposed during the model generation. A statistical analysis method is introduced to eliminate TSMs that have little chance of giving the correct structure.

From the above discussion, we know that each TSM corresponds to a section of the entire hypersurface. The statistical analysis method is based on Monte Carlo sampling of the hypersurface sections, *i.e.* probing the possible solutions by a stochastic search. Firstly, a series of Monte Carlo tests are performed and their *R* factors are calculated for every TSM with random structural parameters. The number of Monte Carlo tests can be specified by the user or generated automatically by the program according to the scale of the problem. Then, an empirical figure of merit (FOM), representing the goodness of a TSM, is calculated from the statistical data of these *R* factors. Finally,

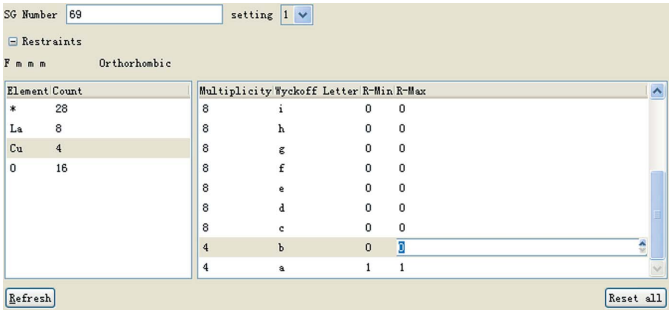

Figure 2
Wyckoff position editing window of *EPCryst*.

the promising TSMs with higher FOMs are selected for subsequent structure solution. Improbable TSMs with appreciably lower FOMs can be eliminated.

The Monte Carlo method is widely used in global optimization, and it is used in *EPCryst* to find the most promising TSMs. If the number of Monte Carlo tests is sufficiently large, the correct solution can be obtained. However, the computing time required to carry out a Monte Carlo search for the structure solution in all TSMs is often too long. Therefore, we use the statistical data from a certain number of Monte Carlo tests in the hope that they can reflect approximately the properties of the corresponding hypersurface sections. *EPCryst* uses three important statistical data, the minimum $R_{\rm min}$, the average $\overline{R}$ and the standard deviation σ of the *R* factors for the stochastic structures generated in the Monte Carlo test. Obviously, $R_{\rm min}$ is the most important factor because the correct structure would be obtained if it is sufficiently small, for example, less than 0.05. However, the average $\overline{R}$ and the standard deviation σ also contain valuable information about the hypersurface section. It is highly likely that the correct solution is found in a TSM with relatively small $\overline{R}$ and large σ because the probability that the global minimum resides on a rough hypersurface is higher than that on a smooth one. Therefore, we define an empirical FOM factor to describe the relative goodness of a TSM in the Monte Carlo test,

$$\mathrm{FOM} = a\frac{1}{R_{\rm min}} + b\frac{\sigma}{\overline{R}}, \qquad (4)$$

where $a = 10$ and $b = 100$ are constants.

Fig. 3 illustrates three typical profiles of hypersurfaces for the La_2CuO_4 TSMs with two parameters. The data are obtained by using a grid search method (§3.3) with a searching resolution of 0.05 Å. It can be seen that the FOM factor is smaller when the hypersurface is smoother. The empirical FOM is a good indicator of the goodness of the corresponding TSMs. The TSMs with relatively large FOM factors will have a higher probability of finding the correct solution, and many improbable TSMs with a FOM below a user-defined threshold value can be eliminated. Because the variation range of FOM factors is problem-dependent, the FOM threshold value set to eliminate the TSMs should be changed correspondingly. In our experiments, the TSM corresponding to the correct structure almost always appears in the top part of the TSM list sorted in descending FOM order so that only a few TSMs require subsequent global optimization.

2.3. Direct-space heavy-atom method

The direct-space heavy-atom method is different from the conventional heavy-atom method (Harker, 1936). The conventional heavy-atom method locates the positions of heavy atoms by analyzing the Patterson (1934, 1935) function, which is the Fourier transform of the measured intensities. Because heavy atoms generally make a dominant contribution to the structure factors of each reflection, the relative intensities of the reflections are also sensitively related to the positions of the heavy atoms. As a result, we can find the positions of the heavy atoms by using the direct-space approach, *i.e.* by finding the best agreement between observed and calculated diffraction patterns where merely the heavy atoms are taken into account for the pattern calculation. No Fourier transform is needed here because the measured intensities are used directly. This direct-space heavy-atom method has been used previously. For example, only the S atom was considered in the first stage of the structure determination of *p*-$CH_3C_6H_4SO_2NHNH_2$ by Monte Carlo methods (Harris *et al.*, 1994). The direct-space heavy-atom method can be easily employed as a

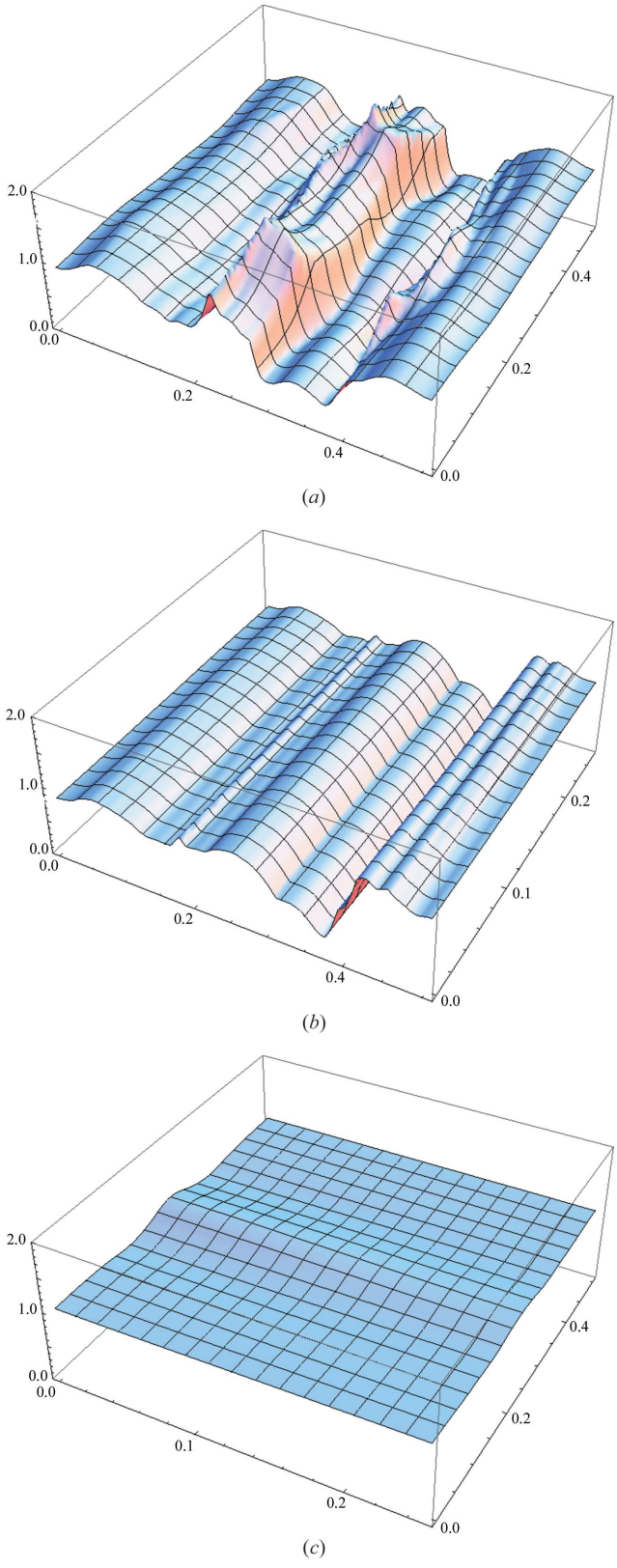

Figure 3
Typical TSM hypersurface shapes of La_2CuO_4: (*a*) hypersurface with a lot of peaks and valleys, EPC = La(8*i*) + Cu(4*a*) + O(8*i*) + O(8*e*), $\sigma = 0.28344$, $R_{min} = 0.08035$, $\overline{R} = 0.842197$, FOM = 158.111; (*b*) hypersurface smoother than (*a*), EPC = La(8*i*) + Cu(4*a*) + O(16*l*), $\sigma = 0.239084$, $R_{min} = 0.205321$, $\overline{R} = 0.85756$, FOM = 76.5838; (*c*) hypersurface that is almost a plane, EPC = La(8*e*) + Cu(4*a*) + O(16*m*), $\sigma = 0.035195$, $R_{min} = 1.00388$, $\overline{R} = 1.07098$, FOM = 13.2476.

Table 3
TSMs generated for the partial structure of La_2CuO_4 (O atoms are excluded) and their Bragg factor (R_{Bg}) obtained by grid search.

TSM	Number of parameters	R_{Bg}†
La(8*i*) + Cu(4*a*)	1	0.209
La(8*h*) + Cu(4*a*)	1	0.909
La(8*g*) + Cu(4*a*)	1	0.909
La(8*f*) + Cu(4*a*)	0	1.221
La(8*e*) + Cu(4*a*)	0	1.085
La(8*d*) + Cu(4*a*)	0	1.121
La(8*c*) + Cu(4*a*)	0	1.116

† See §3.1.

Table 4
Numbers of TSMs and DOFs for La_2CuO_4.

	Atoms	TSMs	DOF range	Total number of TSMs
Normal	La + Cu + O	186	0–3	186
Heavy-atom method	Step 1: La + Cu	7	0–1	37
	Step 2: La† + Cu† + O	30	1–3	

† Atoms are fixed in specified positions.

two-stage process in *EPCryst*. In the first stage, one would input only the heavy atoms in the unit cell and run *EPCryst* to solve the partial structure to determine the positions of the heavy atoms. In the second stage, one would input all atoms in the unit cell and fix the heavy atoms in the positions determined in the first stage, and then rerun the program to solve the complete structure. Using this method helps to reduce the total number of generated TSMs substantially and to solve complex structures efficiently.

Let us consider a practical example to solve the La_2CuO_4 structure using the direct-space heavy-atom method. The Wyckoff positions 4*a* and 4*b* are equivalent for Cu atoms as we have discussed in §2.1. Therefore, we can put the Cu atoms in position 4*a*. When La, Cu and O atoms are considered together, 186 TSMs (with one–three parameters) will be generated. Using the direct-space heavy-atom method, we can exclude O atoms in the first stage because O is a weaker scattering atom than La and Cu. Only seven TSMs were generated when we input just the La and Cu atoms. The positions of the La and Cu atoms were determined using a grid search with a searching resolution of 0.01 Å. Table 3 lists the TSMs and the corresponding search results. From this table, we can easily find that the correct positions of the La and Cu atoms are 8*i* and 4*a*, respectively. Then, all atoms (including the O atoms) were considered for the TSM generation, with the La and Cu atoms fixed in the 8*i* and 4*a* positions, respectively. In total, 30 TSMs were generated. From Table 4, we can see that the numbers of TSMs and degrees of freedom (DOFs) are greatly reduced by using the direct-space heavy-atom method.

2.4. Bond-length checking

Bond lengths, bond angles and connectivity (Spek, 2009) can be calculated to validate a crystal structure. Bond-length checking is available in the current version of *EPCryst*, and it is helpful to choose candidate TSMs or validate the final solution. Users can edit the corresponding radii of atoms and ions to define appropriate bond lengths, and specify the relative variable ranges of the bond lengths. Bond-length checking in *EPCryst* can be applied when needed, *e.g.* after the statistical analysis or during the two-step search as an auxiliary tool to eliminate incorrect TSMs. Certainly, the principal use

of bond-length checking is to validate the final structural solutions obtained by the global optimization process.

3. Global optimization

Global optimization methods are used to search the structural parameters that minimize $R_{\rm Bg}$ for each TSM, and the TSM that gives the lowest $R_{\rm Bg}$ is selected as the structure solution. In current versions of *EPCryst*, two kinds of global optimization methods have been implemented: grid search and simulated annealing. Since a modular programming technique is used in the design of *EPCryst*, additional global optimization methods can be easily implemented in the future. Fig. 4 shows the global optimization windows of *EPCryst*.

3.1. Grouped Bragg factor

The Bragg *R* factor (Young, 1993) is often used to measure the agreement between the reflection intensities calculated from a crystallographic model and those measured experimentally, and it is defined as

$$R_{\rm B} = \sum_i |I_{oi} - I_{ci}| / \sum_i I_{oi}, \tag{5}$$

where $I_{\rm o}$ and $I_{\rm c}$ are defined as the observed and the calculated intensity, respectively. To deal with the problem of overlapping peaks, peaks with $\Delta 2\theta < \delta\,{\rm FWHM}$ are considered as a group in *EPCryst*, and the value of δ can be specified by the user. The observed and calculated intensities for the grouped peaks can be calculated by the following formulas:

$$I_{\rm go} = \sum_{\Delta 2\theta < \delta\,{\rm FWHM}} I_{oi}, \tag{6}$$

$$I_{\rm gc} = \sum_{\Delta 2\theta < \delta\,{\rm FWHM}} I_{ci}. \tag{7}$$

The Bragg *R* factor for the grouped peaks can then be written as

$$R_{\rm Bg} = \sum_i |I_{{\rm go}i} - I_{{\rm gc}i}| / \sum_i I_{{\rm go}i}. \tag{8}$$

3.2. Searching in the asymmetric unit

To increase the efficiency of the global optimization method, the search process is bounded in the asymmetric unit. However, the shape of the asymmetric unit may be very complex (Koch & Fischer, 1974) and it is not easy to encode the exact boundaries of the complex asymmetric units. Sometimes searching in the asymmetric unit may fail to find the best solution. In this situation, we must force the program to search in the whole unit cell manually.

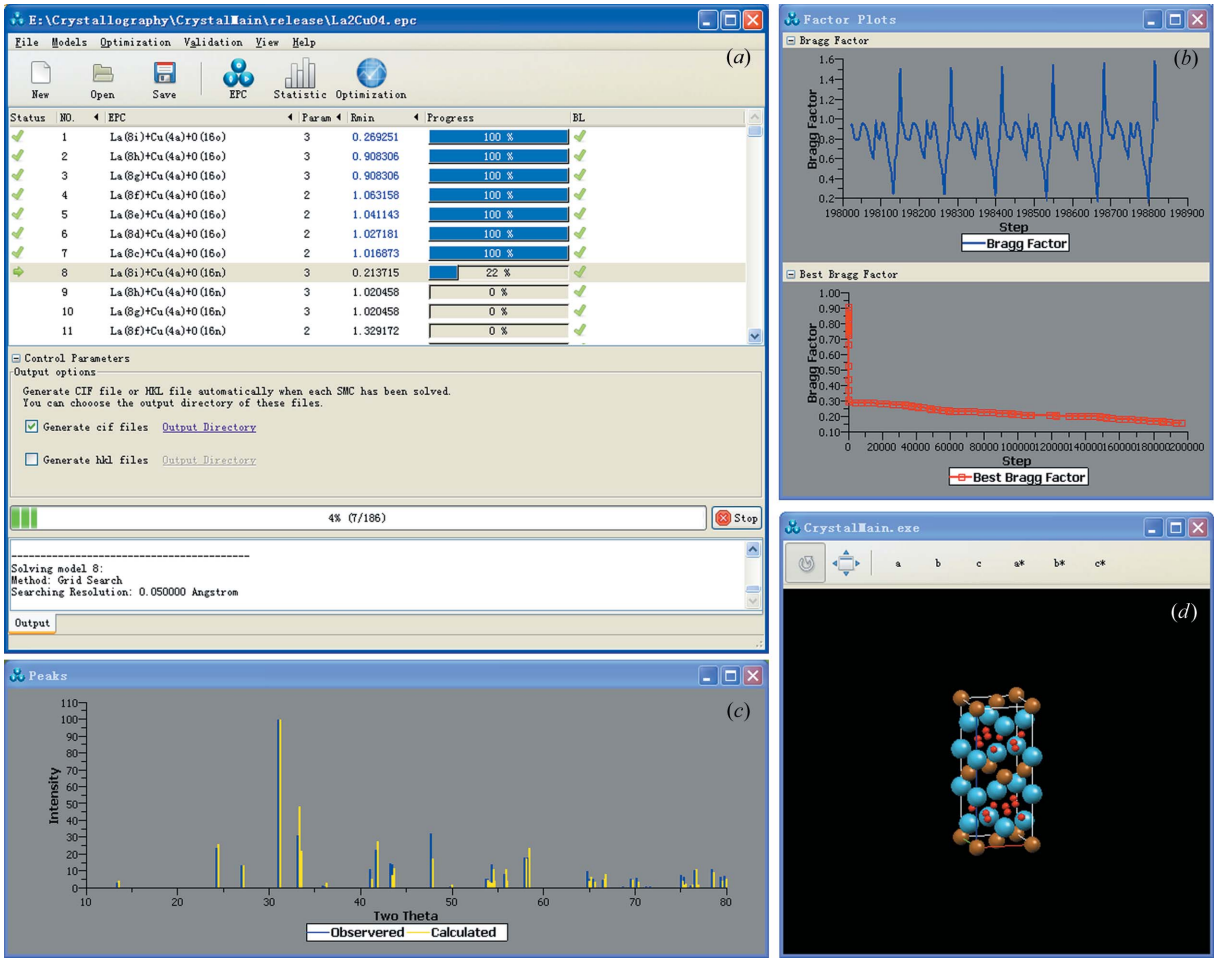

Figure 4
The 'global optimization' windows of *EPCryst*, consisting of (*a*) one main window that lists all TSMs and the corresponding information (EPC, parameters, solving progress *etc.*), (*b*) a window that shows the calculated and observed patterns, (*c*) a window that shows Bragg factors and the best Bragg factors of the current solving TSM, and (*d*) a window that shows the best crystal structure in three dimensions. All windows work dynamically.

Table 5
Solved structures.

	Space group	Atoms	DOF†	Z	TSMs	DOF‡ range	Method
La_2CuO_4	*Fmmm*	7	21	4	186	0–3	GS
In_2O_3	$Ia\bar{3}$	5	15	16	17	3–5	GS + SA
Y_2O_3	$Ia\bar{3}$	5	15	16	17	3–5	GS + SA
K_2TiF_6	$P\bar{3}m1$	9	27	1	48	1–4	GS
K_2NaAlF_6	$Fm\bar{3}m$	10	30	4	4	0–1	GS
$PbSO_4$	*Pnma*	6	18	4	57	6–12	SA

† DOF of model when considering each atom in the asymmetric unit cell independently. ‡ DOF distribution of TSMs generated by EPC algorithm.

3.3. Grid search

The grid search method is a determinative global optimization method. The whole parameter space is divided into grids, and each grid point is tested to find the global minimum. *EPCryst* uses the search resolution τ in ångströms to define the grid size. Each dimension of the hypersurface will be divided into N_i parts:

$$N_i = \begin{cases} a/\tau & \text{if dimension } i \text{ represents } x, \\ b/\tau & \text{if dimension } i \text{ represents } y, \\ c/\tau & \text{if dimension } i \text{ represents } z, \end{cases} \quad (9)$$

where a, b and c are the cell parameters.

The computing cost of the grid search is strongly dependent on the search resolution and the number of parameters. The grid search will become computationally intractable as the number of parameters increases. So, the grid search method usually applies to TSMs with less than six parameters with a rational choice of the search resolution, and it is the default method to solve structures with less than three parameters in *EPCryst*.

3.4. Simulated annealing

Both a basic and an adaptive simulated annealing algorithm for global optimization have been implemented in *EPCryst*. The user can specify the initial temperature, final temperature and annealing rate μ of basic simulated annealing (http://objcryst.sourceforge.net/ObjCryst/; Pagola & Stehens, 2010). The annealing temperature changes according to the following formula:

$$T_i + 1 = \mu T_i. \quad (10)$$

For most problems, temperatures in the range [10, 0.001] will be sufficient. The annealing rate μ is usually set to be around 0.9.

In contrast to basic simulated annealing, in the adaptive simulated annealing approach (Engel *et al.*, 1999) the annealing parameters are set automatically. At the initialization stage, a set of Monte Carlo tests are applied to the TSMs, and a standard deviation σ is calculated from the test results. The initial temperature is set as 1.5 times the standard deviation, and the final temperature is set as 0.2 times the initial temperature. The temperature T and atomic displacement amplitudes are automatically updated during the adaptive simulated annealing process. The temperature of the next step is updated to the standard deviation of the Monte Carlo runs at the current temperature, and the displacement amplitude is updated to keep the acceptance ratio near 0.5. If the displacement satisfies $d_{\rm new} = d_{\rm old} + rm$ (where r is a uniform random number within the range $[-1, 1]$ and m is the displacement range) then the displacement range m and the acceptance ratio p have the following relationship (Corana *et al.*, 1987):

$$\begin{cases} m_{\rm new} = m_{\rm old} g(p), & \\ g(p) = 1 + c(p - 0.6)/0.4 & \text{if } p > 0.6, \\ g(p) = [1 + c(0.4 - p)/0.6]^{-1} & \text{if } p < 0.4, \\ g(p) = 1 & \text{otherwise}, \end{cases} \quad (11)$$

where c is a scaling parameter. According to the acceptance ratio calculated at each temperature, the displacement amplitude will be updated automatically.

3.5. Mixed use of grid search and simulated annealing

The TSMs generated using the EPC algorithm often have different numbers of parameters, and some of them are suitable for gird search while others are not. In such situations, mixed use of a grid search and simulated annealing is a more appropriate approach. *EPCryst* can automatically choose the appropriate method.

The parameter space is also gridded for simulated annealing, *i.e.* Monte Carlo moves are adjusted to the grid points, so that the results are comparable to that of the grid search. The grid size (search resolution) does not affect the problem scale of simulated annealing. When used alone, the search resolution of the simulated annealing should be set as fine as possible.

3.6. Two-step search strategy

A global optimization search can be easily realized by a two-step process using *EPCryst*. The first step is to perform a coarse grid search (or rapid simulated annealing) to select promising TSMs with smaller $R_{\rm Bg}$ factors, and the second step is to determine the accurate atomic parameters of the selected TSMs by a fine grid search (or elaborate simulated annealing). The main purpose of the two-step search strategy is to speed up the computationally expensive global optimization process. As we have discussed in §2.2, each TSM corresponds to a section of the whole hypersurface, and the search is carried out for the global minimum of the hypersurface. The first step is to find the approximate local minimum of each hypersurface section, and the second step is to search carefully on sections having a high probability of finding the global minimum. Usually, the two-step strategy is more efficient than applying detailed global optimization to all TSMs.

4. Applications

EPCryst has advantages in solving inorganic structures with orthorhombic or higher symmetry since the numbers of independent parameters or DOFs will decrease considerably in such cases. *EPCryst* works particularly well when the number of generated TSMs is small (say less than 100) and the number of parameters is not greater than ten. We downloaded some experimental data from the powder pattern database PowBase (Le Bail, 2000) to test the *EPCryst* program. We used the program *Fullprof* (Rodriguez-Carvajal, 1990) to extract intensities by full-profile fitting. Table 5 summaries these tested crystal structures, and the structure solution procedures and results are described in the following sections. For detailed information about the solved structures and more examples, please visit http://www.epcryst.com/examples.

4.1. La_2CuO_4

There are seven atoms in an asymmetric unit of the La_2CuO_4 cell. The EPC algorithm generates 186 TSMs when restraints are used as described in §2.1. The numbers of parameters of these TSMs vary from zero to three, so the grid search method is used to solve the structure.

FOM factors were used to choose candidate TSMs for global optimization during the statistical analysis. Fig. 5 shows a typical FOM distribution of the TSMs. Because the statistical analysis is based on the Monte Carlo process, sometimes the FOM threshold should be chosen carefully. TSMs either with FOM > 80 or with FOM > 20 will be fine in this situation. After statistical analysis, a grid search was used to solve the chosen TSMs using a search resolution of 0.05 Å. The obtained atomic positions and coordinates are Cu in position 4*a* (0, 0, 0), La in 8*i* (0, 0, 0.361), and O in 8*i* (0, 0, 0.175) and 8*e* (0.25, 0.25, 0). This result is very close to the data from the Inorganic Crystal Structure Database (ICSD; Bergerhoff *et al.*, 1983). The greatest deviation is the z coordinate for O in the 8*i* position ($z_{8i} = 0.182$ in the ICSD data). The correct structure solution can also be obtained by the direct-space heavy-atom method described in §2.4.

4.2. In_2O_3, Y_2O_3

In_2O_3 and Y_2O_3 have the same space group and similar chemical composition, so the same number of TSMs (merely 17) was generated for both compounds. The numbers of parameters are in the range from three to five. The grid search method is perfectly suitable for TSMs with three parameters. However, fine grid searches for TSMs with four or five parameters require a much longer time. Therefore, the correct solutions were successfully obtained by the mixed use of the grid search and simulated annealing methods. The obtained structural data for In_2O_3 and Y_2O_3 are approximately equal to those from the ICSD. Taking In_2O_3 as an example, only minor differences exist between the coordinates of O obtained, (0.387, 0.161, 0.379), and those from the ICSD, (0.3905, 0.1529, 0.3832). These two structures can be solved by using the direct-space heavy-atom method.

4.3. K_2TiF_6, K_2NaAlF_6

K_2TiF_6 and K_2NaAlF_6 crystals have very high cubic symmetry, so the benefit of using *EPCryst* is very obvious. Most of the TSMs have only one or two parameters. Grid search method can quickly give the solution. For K_2NaAlF_6 only a few seconds are needed to find the structure solution. The x coordinate of F in the 24*e* position is the only variable to be determined, and it was found to be 0.221, which is quite close to the value from the ICSD ($x_{24e} = 0.2225$).

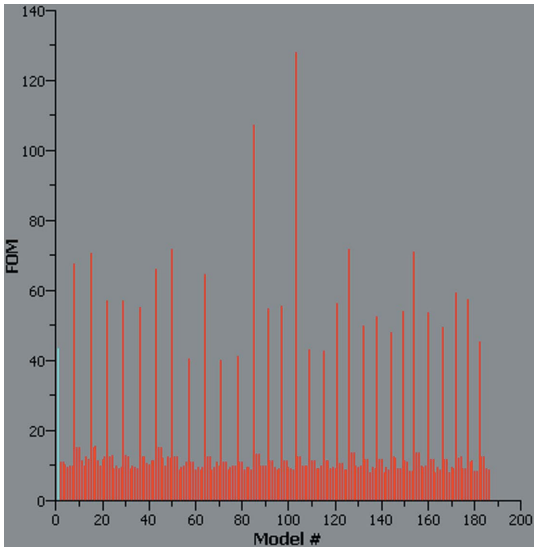

Figure 5
A typical FOM distribution of La_2CuO_4 TSMs.

4.4. $PbSO_4$

Both X-ray and neutron pattern data of $PbSO_4$ (Hill, 1992) were used to solve the crystal structure. In total 57 TSMs were generated with parameter numbers in the range from six to 12. Simulated annealing was used to find the solution. Because the O atom is a weaker X-ray scatterer than the Pb and S atoms, it is difficult to find accurate coordinates for the O atoms using only the X-ray data. Therefore, both X-ray and neutron data are used to solve the $PbSO_4$ structure.

Firstly, we used X-ray data and the direct-space heavy-atom method (§3.2) to locate the positions of Pb and S. In total, seven TSMs with parameter numbers in the range from zero to four were generated with input 'PbS'. After applying the grid search method with a search resolution of 0.1 Å, the Wyckoff positions and approximate coordinates of Pb and S can be determined: both the Pb and the O atoms should be at 4*c*.

Secondly, we used neutron data to locate the positions of the light O atoms. Fixing the Pb and S atoms in the Wyckoff positions obtained above, only nine TSMs were generated and the numbers of parameters are in the range from seven to 12. Then, the coordinates of the Pb and S atoms were also fixed to that obtained above from the X-ray data; the numbers of parameters decreased to the range from three to eight. Finally, we used simulated annealing to find the structure solution. The obtained solution had almost the same coordinates for Pb, (0.188, 0.25, 0.167), and S, (0.059, 0.25, 0.667), as that from the ICSD. The coordinates of O showed slight shifts, but the SO_4 tetrahedron can be clearly observed.

5. The *EPCryst* program

EPCryst is written in the C programming language and the graphical user interface (GUI) is built using the GTK+ library. The GTK+ library is a feature-rich toolkit for creating GUIs with cross-platform compatibility. The core of *EPCryst* is carefully designed to separate from the GUI part, which makes it highly reusable and extendable. One very important feature of *EPCryst* is its interactivity, and the user has full control of the structure solving process from model generation to global optimization.

EPCryst uses the extracted intensities for speed considerations, so a full-profile fitting program should be used to prepare the input file with the extracted intensities. *EPCryst* can read the output file format (with control code JLHK or HKL set to 1) of the *Fullprof* program (Rodriguez-Carvajal, 1990) directly. The output files from other full-profile fitting programs must be converted into an internal format described in the user manual.

Both Windows and Linux versions of the *EPCryst* program are now available from the authors upon request (dixiaodeng@gmail.com or dengxiaodi@ssc.iphy.ac.cn). Examples and details about the use of *EPCryst* are fully documented in the user manual, which can be downloaded from http://www.epcryst.com.

This project was supported by the National Natural Science Foundation of China under grant Nos. 20571083 and 20871119.

References

Abrahams, S. C. (2010). *Acta Cryst.* B**66**, 173–183.
Adreev, Yu. G., Lightfoot, P. & Burce, P. G. (1997). *J. Appl. Cryst.* **30**, 294–305.

Bergerhoff, G., Hundt, R., Sievers, R. & Brown, I. D. (1983). *J. Chem. Inf. Comput. Sci.* **23**, 66–69.
Černý, R. (2008). *Chem. Met. Alloys*, **1**, 120–127.
Corana, A., Marchesi, M., Martini, C. & Ridella, S. (1987). *ACM Trans. Math. Softw.* **13**, 262–280.
Csoka, T. & David, W. I. F. (1999). *Acta Cryst.* A**55** Suppl., Abstract No. P08.03.012.
David, W. I. F. & Shankland, K. (2008). *Acta Cryst.* A**64**, 52–64.
David, W. I. F., Shankland, K., McCusker, L. B. & Baerlocher, Ch. (2002). *Structure Determination from Powder Diffraction Data*, pp. 256–261. Oxford University Press.
Deem, M. W. & Newsam, J. M. (1989). *Nature* (*London*), **342**, 260–262.
Deng, X. & Dong, C. (2009). *J. Appl. Cryst.* **42**, 953–958.
Donnay, G., Morimoto, N. & Takeda, H. (1964). *Acta Cryst.* **17**, 1369–1373.
Earl, D. J. & Deem, M. W. (2005). *Phys. Chem. Chem. Phys.* **7**, 3910–3916.
Engel, G. E., Wilke, S., König, O., Harris, K. D. M. & Leusen, F. J. J. (1999). *J. Appl. Cryst.* **32**, 1169–1179.
Falcioni, M. & Deem, M. W. (1999). *J. Chem. Phys.* **110**, 1754–1766.
Favre-Nicolin, V. & Černý, R. (2002). *J. Appl. Cryst.* **35**, 734–743.
Feng, Z. J. & Dong, C. (2007). *J. Appl. Cryst.* **40**, 583–588.
Feng, Z. J., Dong, C., Jia, R. R., Deng, X. D., Cao, S. X. & Zhang, J. C. (2009). *J. Appl. Cryst.* **42**, 1189–1193.
Harker, D. (1936). *J. Chem. Phys.* **4**, 381–390.
Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. (1998). *Acta Cryst.* A**54**, 632–645.
Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543–3547.
Hill, R. J. (1992). *J. Appl. Cryst.* **25**, 589–610.
Kariuki, B. M., Serrano-González, H., Johnston, R. L. & Harris, K. D. M. (1997). *Chem. Phys. Lett.* **280**, 189–195.
Koch, E. & Fischer, W. (1974). *Acta Cryst.* A**30**, 490–496.
Le Bail, A. (2000). PowBase, http://sdpd.univ-lemans.fr/powbase/.
Le Bail, A. (2005). *J. Appl. Cryst.* **38**, 389–395.
Newsam, J. M., Deem, M. W. & Freeman, C. M. (1992). *Accuracy in Powder Diffraction II*, NIST Special Publication No. 846, pp. 80–91. Gaithersburg: NIST.
Pagola, S. & Stephens, P. W. (2010). *J. Appl. Cryst.* **43**, 370–376.
Patterson, A. L. (1934). *Phys. Rev.* **46**, 372–376.
Patterson, A. L. (1935). *Z. Kristallogr. A*, **90**, 517–542.
Rodriguez-Carvajal, J. (1990). Abstracts of the Satellite Meeting on Powder Diffraction of the XV Congress of the International Union of Crystallography, 19–28 July, Toulouse, France, p. 127.
Shankland, K., David, W. I. F. & Csoka, T. (1997). *Z. Kristallogr.* **212**, 550–552.
Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.
Young, R. A. (1993). *The Rietveld Method.* Oxford University Press.