Crystal structure determination from powder diffraction:

case of non-molecular compounds

Radovan Černý

Laboratoire de Cristallographie, Université de Genève, 24, quai Ernest-Ansermet,
CH-1211 Genève 4, Switzerland
Radovan.Cerny@unige.ch

Keywords

Powder diffraction, structure determination, global optimization, simulated annealing

Abstract

Methods of structure determination from powder diffraction of non-molecular compounds (inorganics, extended solids, intermetallic compounds etc.) are discussed. Direct space methods do not need powder pattern decomposition, and are based on a global optimization of a structural model to improve the agreement between the observed and calculated diffraction patterns. The success of the method depends very much on a proper modeling of the structure from building blocks. The modeling from larger building blocks improves the convergence of the global optimization algorithm by a factor of up to ten. The correctness of the building block like its rigidity, deformation, bonding distances and ligand identity must be examined carefully. Dynamical Occupancy Correction implemented in the direct space program FOX has shown to be useful when merging excess atoms, and even larger building blocks like coordination polyhedra. It allows also joining smaller blocks into larger ones in the case when the connectivity is not *a priori* evident from the structural model.

Introduction

Powder diffraction using X-rays and neutrons plays a major role in the search of new materials which are not available in the form of single crystals. Moreover, most of the industrial applications of inorganic and organic compounds are in the form of polycrystalline materials (for example metal hydrides for storage and battery applications, metallic and intermetallic compounds in industry, thin films, organic compounds in pharmaceutical industry etc.). Structure determination from powder diffraction (SDPD) is more difficult than structure determination on single crystals, because the available data are a projection of a three-dimensional diffraction pattern onto one dimension (radial distance from the reciprocal space origin), and consequently the diffraction peaks overlap. The extraction of structure factor amplitudes can be further complicated by a broadening (often anisotropic) due to the crystal lattice defects. Two alternative solutions of this problem exist when trying to solve a crystal structure from powder data: Either we try to improve the decomposition of the observed powder pattern into individual peaks, or we try to model the observed pattern as a whole. Consequently the methods of SDPD can be divided into two groups according to the working space as can be found in[1] and references therein:

- *Reciprocal space methods (RSM)*: They use procedures developed for single crystal data like direct methods or Patterson synthesis, and optimized for powder data. They need structure factor amplitudes obtained by powder pattern decomposition.

- *Direct space methods (DSM)*: Different algorithms for a search in the direct space of structural parameters are used, and an agreement factor between the observed and calculated powder diffraction data is evaluated, and the structural model is optimized to improve the agreement.

In this talk we will concentrate on the second case and on the compounds which do not contain isolated molecules (extended solids), i.e. most of inorganic compounds. The molecular compounds like organics, or hybrids like coordination polymers and metal organic frameworks, are in principle treated by the same approach. The molecules are described by internal coordinates, knowledge of molecule conformation obtained by other methods is actively used as well as organic structure databases and energy minimization of molecular crystals.

Some Definitions (from A. Le Bail talk on ESCA-9, Egypt 2004):

Sometimes the "direct space methods" (not to be confused with the direct methods) are called "global optimization methods" or "model building methods", and even sometimes "real space methods". "Direct space" was the definition retained in the pioneering papers. "Direct space" as opposed to "reciprocal space" has an adequate crystallographic structural sense, and should be preferred to "real space", which, opposed to "imaginary" would call to mind both parts of the diffusion factors. "Global optimization" has a large sense and designates the task of finding the absolutely best set of parameters in order to optimize an objective function, a task not at all limited to crystallography.

Under the name direct space methods we will not understand the interpretation of electron density or Patterson maps by search for molecular fragments, even if they work in direct space and use the global optimization algorithm like genetic algorithm. These methods still needs the structure amplitudes, i.e. decomposition of a powder pattern, which is avoided by direct space methods mentioned here.

History

The first successful attempt to solve a crystal structure by an automatic (not manual!) localization of a building block (rigid molecule) in the direct space can be seen in the program RISCON[2] which was then modified for powder data as P-RISCON[3]. The optimization algorithm used was the constrained least-squares refinement, which is limited to structures not bigger than 10 independent atoms and resulted in only approximate atomic positions.

The authors of ref. [4] were one step from being the first using a true global optimization algorithm – simulated annealing (SA) for structure solution from powder data. However, they did not believed to the power of the method: “At present the method is not efficient enough for use in most practical problems of *ab-initio* structure determination.” The authors have used SA for structure prediction based on the optimization of the crystal potential energy. So, the first use of a global optimization algorithm (SA) in the structure solution from powder data belongs to Newsam *et al*.[5] even if the structure solved in the paper was known and small (benzene). Later on the direct space method of structure solution from powder data developed rapidly, using different algorithms like Monte Carlo (MC) search[6] and genetic algorithm (GA)[7]. An essential step forward was achieved by applying the description of structural blocks by internal coordinates like bond distances, angles and torsion angles by[8] allowing so a direct stereo-chemical interpretation and/or constraining of optimized structural parameters. Since that, the list of programs dealing with direct space methods of structure solution from powder (but also single crystal) X-ray and/or neutron diffraction data continues to grow. For review see http://www.cristal.org/ or http://www.ccp14.ac.uk/.

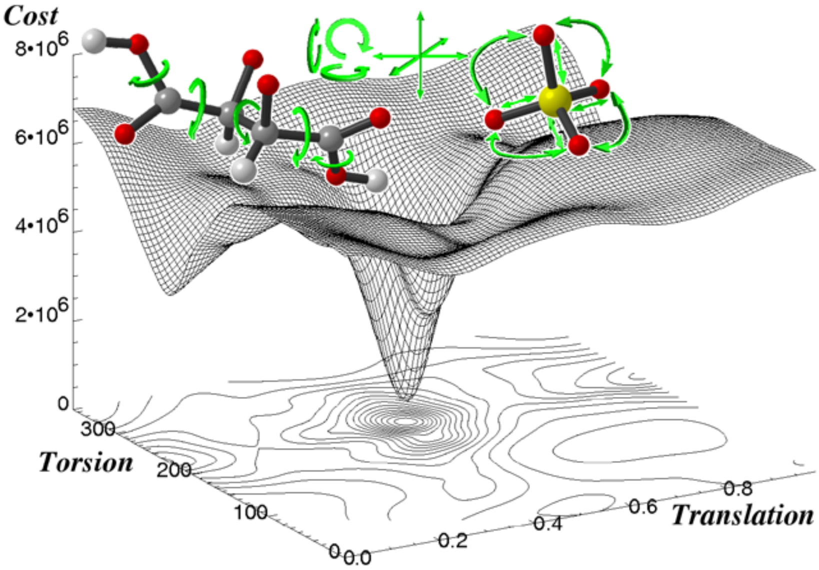

Figure 1. Solving a structure ab-initio in direct space implies describing the structure through a number (N) of Degrees of Freedom (DoF): translation and rotation of the molecule or polyhedron, and internal DoF like torsion angles, bond length and bond angles. These parameters must then be randomly changed in order to find the minimal cost (usually the best agreement between the calculated and experimental powder pattern). This corresponds to exploring a N-dimensional hypersurface until the global minimum is found. The surface represented here corresponds to a 2D cut of the hypersurface corresponding to the variation of one torsion angle and one translation

.Principles

The direct space methods are based on location of building blocks in the elementary cell by using random or systematic moves and/or modifications of the blocks, and comparison of calculated and observed diffraction patterns and/or other cost functions (CF) like crystal energy, atomic coordination etc (Figure 1). Based on a "fitness" of the current structural model the decisions are taken how to improve the model. Generally said, it is a global optimization problem of a great complexity, where the algorithm must explore a hypersurface (see Figure 1) which describes the “cost” of the model as a function of all structural parameters (see chapter 15.6 in ref. [1]), and find its global minimum. A flow chart representing a typical implementation, like in program FOX [9], of the global optimization approach to the crystal structure solution from powder diffraction data is given in Figure 2.

Two algorithms of the global optimization have found larger application in the SDPD:

Simulated annealing and Parallel Tempering

SA and PT algorithms are both based on MC sampling, earlier known as “statistical sampling” (for review, see ref. [10]). The first, and now widely used, algorithm of the MC sampling is based on Boltzmann distribution, and is known as Metropolis algorithm[11]. The MC sampling as applied in SDPD is also called Reverse Monte Carlo[12], because the system is modified by random changes under the constraint of observed data, like diffraction pattern. A flow chart of the Metropolis algorithm applied to SDPD is given in Figure 3.

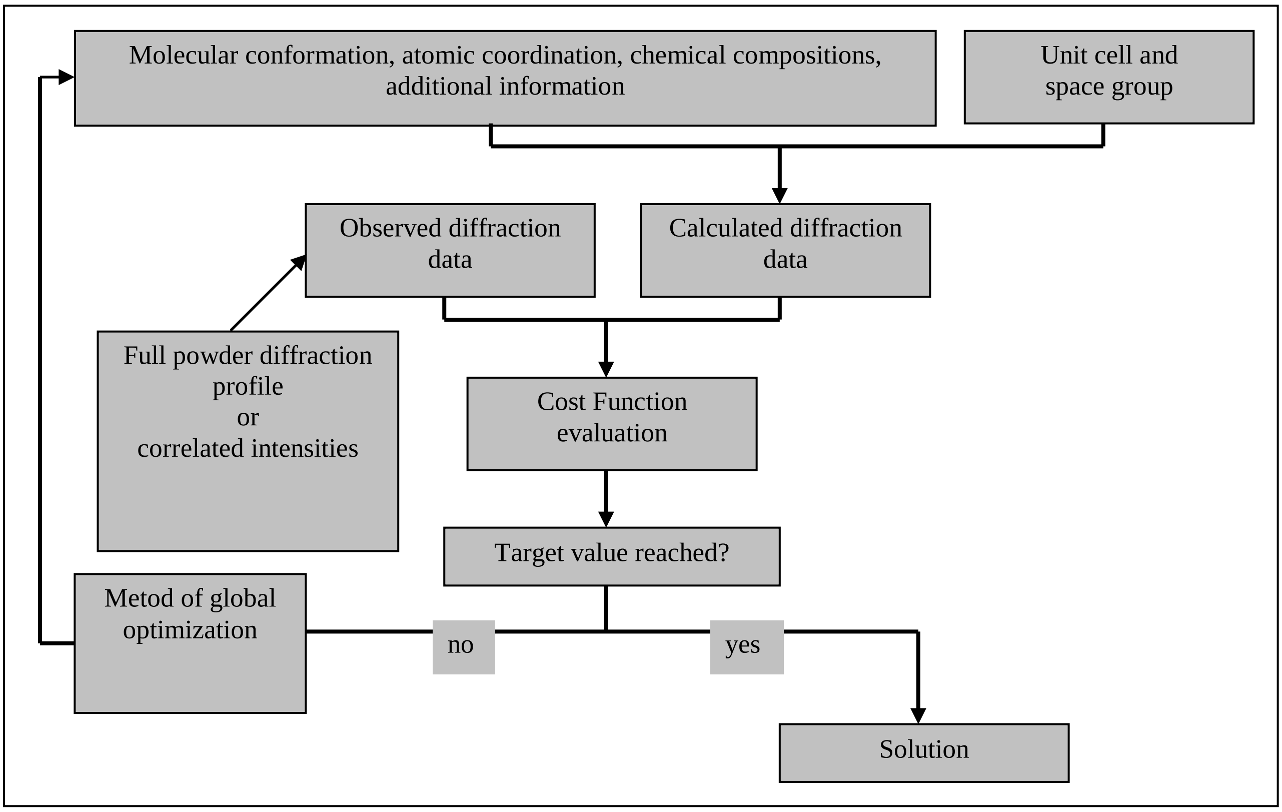

Figure 2. A typical flow chart of a direct space method as applied to SDPD, case of the program FOX[9].

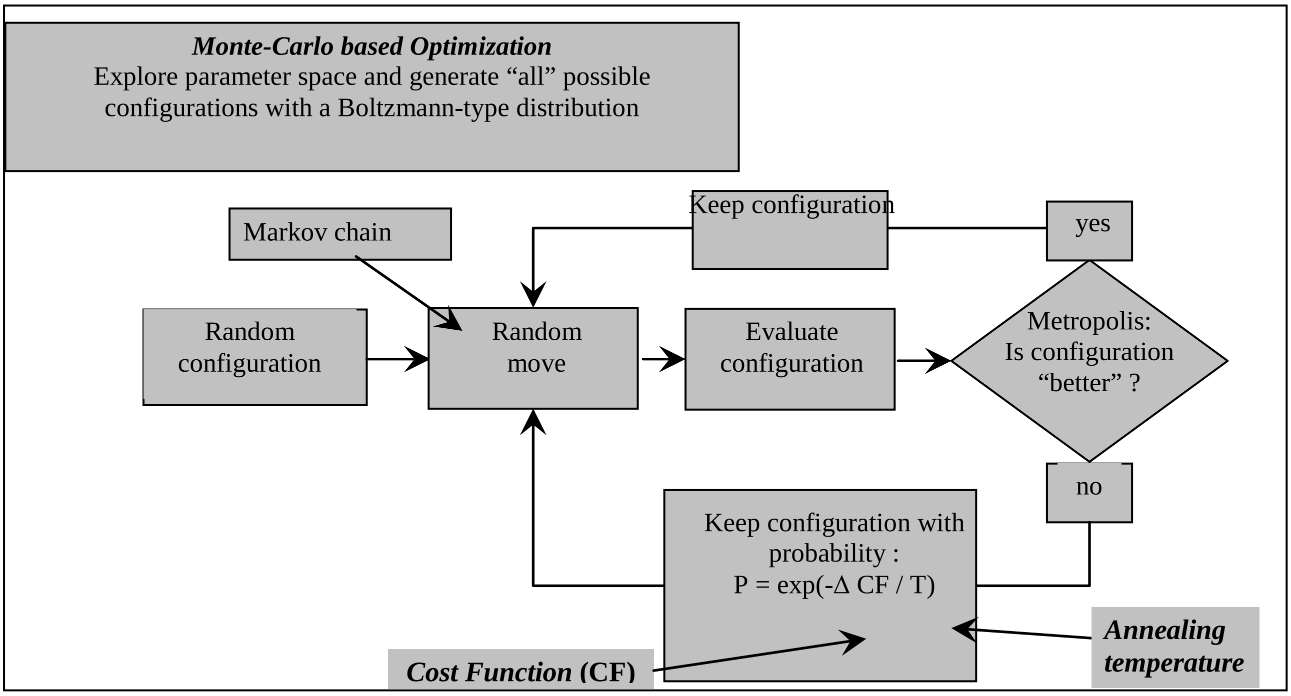

Figure 3. Flow chart of the Metropolis algorithm[11] of the simulated annealing as applied to SDPD.

The modification of SA called parallel tempering algorithm (PT) was first used in the SDPD in ref. [13]. The principal advantage of PT algorithm within the SDPD compared to SA algorithm is its generality for any type of problem; no parameters like annealing rate, starting temperature, are required. The algorithm is also generally able to escape from the local minima in the parameter space[9].

Evolutionary theory – genetic algorithm

GA form a subset of broader classes of global-optimisation strategies called *population-based methods,* and *evolutionary algorithms*. The concept of GA follows the old idea of minimizing human efforts in solving difficult scientific and technical problems by learning from the nature**.** The genetic computation proceeds in the space of (usually binary coded) variables. It mimics the evolution of living organisms represented by the points in this space (trial solutions). In the beginning, a population of individuals (called also chromosomes, agents...), which may represent trial solutions of the optimisation task, is generated. Next generations are successively created using simplified principles of plant or animal (Darwinian) evolution. The calculation is terminated by application of a suitable stop condition. The basic genetic operators used in formation of each new population include selection, crossover and mutation. GA was first used for SDPD in refs. [7] and [14]. A flow chart of the genetic algorithm applied to SDPD is given in Figure 4.

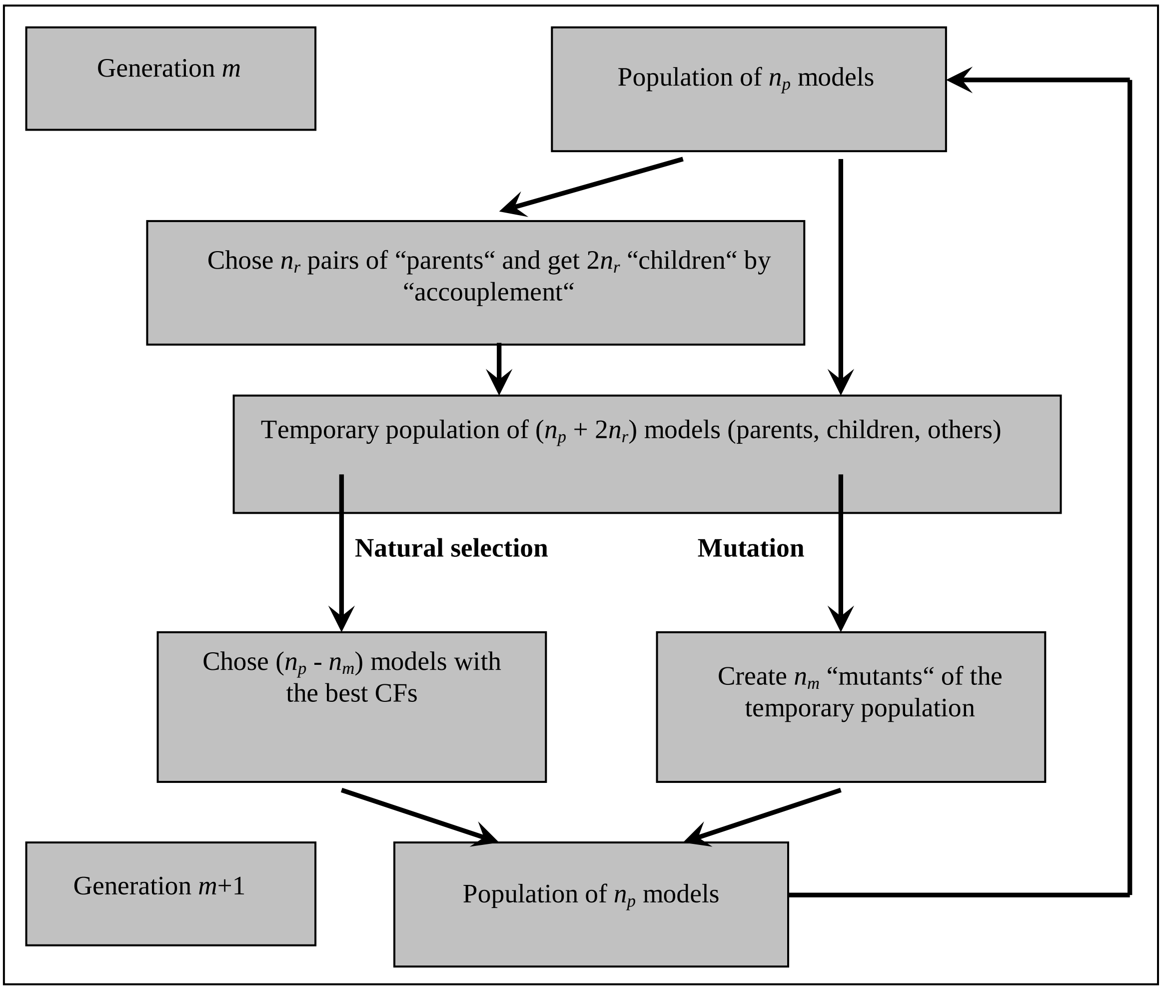

Figure 4. A general flow chart of the genetic algorithm as applied to SDPD.

Modeling non-molecular structure

The modelling of non-molecular compounds is by nature complex: this is due to the fact that they are extended solids, and therefore no 'obvious' group of atoms (like a molecule) can be isolated to be used as a *building unit* (BU) for the entire structure. This does not induce a particular difficulty for *ab initio* structure solution methods working entirely in reciprocal space (RSM). However, it becomes a difficulty for the methods working entirely in direct space (DSM) and using *a priori* chemical knowledge about the compound. The structure solution algorithm working in direct space has to know how to define basic structural blocks of the crystal, BUs, which are then manipulated (optimized) by the algorithm. Moreover, the presence of high-order symmetries leads to a fragmentation of these BUs, as any group of atom (e.g. an octahedron) can be located on a symmetry element, so that only part of the atoms of a symmetry-constrained BU are truly independent. Finally, the actual arrangement in space between BU and relatively to the symmetry elements of the unit cell, are *a priori* unknown.

The basic principles to be followed when choosing the BU are the following:

- the model must describe the structure with few parameters (degrees of freedom or DoF), so that the parameter space to explore is as small as possible,
- the basic building units must allow to exploit any a priori chemical knowledge about the material, to provide an chemically unambiguous solution,
- all BUs must be constructed so that they can easily be linked to other BUs, in order to form an extended solid. This includes bonding with themselves in the presence of a symmetry element.

Building units

The choice of a BU is easy when isolated molecules are present in the crystal like in organic or coordination compounds where the molecules are easily described by a finite number of the internal molecular coordinates (bonding distances, angles and torsion angles . A molecule is naturally a well defined BU that can be moved in the unit cell, and deformed by modifying torsion angles. No sharing of atoms between the molecules occurs.

The situation becomes rather unclear when one tries to model the crystal structure of a non-molecular compound. The definition of BU for a crystal like iron or sodium chloride is not unambiguous: should it be one atom, the first coordination sphere of each atom, a structural sheet of finite thickness, the asymmetric unit ? Pauling[15] was at the origin of the concept of BU when he defined the coordination polyhedra for the description of structures, and stated the Pauling’s rules. Coordination polyhedra, typically the first coordination spheres, of selected atoms, which will generate the whole inorganic crystal structure, can always be identified. These polyhedra will share corners, edges or faces in close packed crystals and frameworks but can be also chosen as isolated as in ionic crystals containing complex ions. A wide range of polyhedra can be defined, and these are naturally described using again bond distances, angles and torsion (dihedral) angles. The type of atomic coordination in the non-molecular crystals can be very often predicted from their chemical formula, cationic valence state, known structure of a similar compounds and all available spectroscopic methods, locally sensitive, like NMR, EXAFS, Mössbauer spectroscopy etc. Definition of BU is easy for tetrahedral frameworks like zeolites, where the concept of primary building units (PBU) and secondary building units (SBU) is used. In the case of zeolites a PBU is TO_4 tetrahedron (T being a tetrahedral atom), and SBUs are various fragment of the zeolite framework built from PBU with the aim to use as few SBU as possible, ideally just one, to build up the whole framework[16].

The concept of PBU and SBU may be used also for modeling an extended solid for structure solution purpose. In DSM the PBU are carefully selected from crystal chemistry knowledge of the compound under question, and are used as rigid or semi-rigid blocks which are then localized by the global optimization algorithm within the crystals structure (unit cell) of the compound. The SBU may be used in the same way, i.e. as blocks constructed from PBU or as additional information (restraint) in the global optimization algorithm. Various SBU are possible like isolated polyhedral clusters, infinite polyhedral chains, sheets or frameworks. A correct choice of PBU is important for correct description of *a priori* unknown structure, and may become decisive for the convergence of the global optimization algorithm.

The final choice is not trivial, particularly due to the presence of special positions. The use of a dynamical occupancy correction (DOC)[9] simplifies the problem as it is no longer necessary to manually adjust the occupancy of atoms falling in a special position, and also makes two identical elements fully overlapping half-occupied to be “seen” by the diffraction as a single atom. DOC has revealed to be very powerful in the case when the exact composition of the studied compound is a priori not known exactly, like hydrogen in metal hydrides obtained by the hydrogen absorption in a metallic matrix or even larger building blocks like coordination polyhedral (for examples see ref. [9]).

Theoretically, this means that it is possible to add *more* atoms than initially deemed necessary, expecting the DOC to artificially "merge" the excess atoms. In practice, adding too many atoms will slow down the optimization, as more atoms have to fall in the correct position to find the correct structure. This also implies that if some atoms are shared between polyhedra, a choice should be made on where the bridging atoms should be used.

Structure validation – help of theoreticians

The crystal structure solved from powder diffraction data must be validated. The first step is done in the reciprocal space as it is the evaluation how well the structural model explains the observed powder diffraction data. It is the necessary but not sufficient condition . There are a number of statistics which can be used to assert how good is the agreement between observed and calculated diffraction data, the most commonly used being the χ^2 and R_{wp}[17]. However, these must be used with a care. A "good" R_{wp} value will depend a lot on the quality and type of powder pattern recorded — a 12% may be very good for a synchrotron powder pattern with a very low noise, but would be very bad with neutron or laboratory data with a high background. It is also possible to compare the finally obtained R_{wp} to the value obtained using profile-fitting[18]. In practice, the best tool to evaluate how good a fit is still the eye of an experienced crystallographer[19].

The second step of the validation is the analysis of the crystal structure itself in the direct space:

– Are there atoms "colliding" each other? If DSM repeatedly finds a global solution with atoms overlapping, an anti-bump penalty (see for example ref. [9]) can be used to avoid this solution. Note that this should only happen if the quality of the powder pattern is insufficient, and therefore includes insufficient information to "lock" the DSM algorithm to the correct solution.

– Do the atomic environments make sense? A penalty based on bond valence sum calculation[4] can point out unlikely configuration. Again, if valences are known a priori, this can be used as a criterion for the DSM optimization of the structure.

The third step should be a post-experimental optimization of the structure based on crystal energy minimization like DFT. It helps to validate new structures, locate light atoms (hydrogen), especially when using high-pressure diffraction data, and even correct the symmetry and some structural details.The post-experimental DFT-optimization of the inorganic crystal structures allows achieving the global minimum in the Rietveld refinement which is not easily visible from the diffraction data only, and can be thus highly recommended.

Conclusion and perspectives

Structure determination of non-molecular compounds from powder diffraction data has undergone an intensive development in last 30 years. First applying the reciprocal space methods, which were optimized to work with lower quality data obtained from powder diffraction patterns (like programs SIRPOW-EXPO[20], DOREES-POWSIM[21] and XLENS[22]), and with no important difference when applied to molecular or non-molecular crystals. Following the pioneering works of Pannetier *et al*.[4] and mainly of Newsam *et al*.[5] the direct space method has rapidly evolved, and continues still to be developed, as an user-friendly tool for SDPD of non-molecular crystals. The main principles are again the same as for molecular crystals, however, some specials tools were necessary to develop for treatment of special crystallographic positions, sharing of atoms between different building blocks like coordination polyhedra, and for correct optimization of disordered atomic positions. Current (known) limits of direct space methods are around 30-100 independent atoms. The success depends on the quality of the diffraction data and even more on the amount of the additional chemical information (knowledge about structural building blocks) injected into the structure solution process. The same is true for the reciprocal space methods. However, the use of the additional information, like atomic coordination, interatomic distances, angles, is easier and natural when working in direct space, the space where this information comes from. Among the current challenges and prospects of SDPD belongs active modeling of preferred orientation, active evolution of the structural model during the optimization, improvement of the optimization algorithm, and speeding up the calculation. All these developments may proceed towards an automatic SDPD connected with the structure prediction. Actual state of knowledge, however, still requires active interaction of an experienced crystallographer.

References

1. W.I.F David, K. Shankland, L.B. McCusker and Ch. Baerlocher. *Structure Determination from Powder Diffraction Data*, Oxford University Press, IUCr. Monographs on Crystallography 13, 2002.
2. R. Bianchi, C.M. Gramaccioli, T. Pilati and M. Simonetta. *Acta Cryst.* A**37** (1981) 65-71.
3. N. Masciocchi, R. Bianchi, P. Cairati, G. Mezza, T. Pilati and A. Sironi. *J. Appl. Cryst.* **27** (1994) 426-429.
4. J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal and V. Caignaert. Nature, **346** (1990) 343-345.
5. J.W. Newsam, M.W. Deem and C.M. Freeman. *Accuracy in Powder Diffraction II*, NIST Special Publication, **846** (1992) 80-91.
6. K.D.M. Harris, M. Tremyane, P. Lightfoot and P.G. Bruce. *J. Am. Chem. Soc.* **116** (1994) 3543-3547.
7. K. Shankland, W.I.F. David and T. Csoka. *Z. Kristallographie* **212** (1997) 550-552.
8. Y.G. Andreev, P. Lightfoot and P.G. Bruce. *J. Appl. Cryst*. **30** (1997) 294-305.
9. V. Favre-Nicolin and R. Černý. *J. Appl. Cryst*. **35** (2002) 734-743.
   V. Favre-Nicolin and R. Černý. *Z. Kristallogr*. **219** (2004) 847-856.
   R. Černý R. and V. Favre-Nicolin. *Zeitschrift für Kristallographie*, **222** (2007) 105-113.
10. M.E.J. Newman and G.T. Barkema. *Monte Carlo Methods in Statistical Physics*, Claredon Press, Oxford, 1999.
11. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller. *J. Chem. Phys.* **21** (1953) 1087-1092.
12. R.L. McGreevy and L. Pusztai. *Molec. Simul*. **1** (1988) 359- 367.
13. M. Falcioni and M.W. Deem. *J. Chem. Phys*. **110** (1999) 1754-1766.
14. B.M. Kariuki, H. Serrano-González, R.L. Johnston and K.D.M. Harris. *Chem. Phys. Lett.* **280** (1997) 189-195.
15. Pauling, L. *J. Am. Chem. Soc.* **22**(1929)123.
16. Meier W.M., Olson D.H. Atlas of Zeolite Structure Types, Butterworths, London, 1987.
17. Toby, B.: R factors in Rietveld analysis: How good is good enough? *Powder Diffraction* **21** (2006) 67–70.

18 David, W. I. F.: On the equivalence of the Rietveld method and the correlated integrated intensities method in powder diffraction. *J. Appl. Cryst.* **37** (2004) 621–628.

19 McCusker, L. B.; Von Dreele, R. B.; Cox, D. E.; Loue¨r, D.; Scardi, P.: Rietveld refinement guidelines. *J. Appl. Cryst.* **32** (1999) 36–50.

20. A. Altomare, R. Caliandro, M. Camalli, C. Cuocci, C. Giacovazzo, A.G.G. Moliternia and R. Rizzi. *J. Appl. Cryst.*, **37** (2004) 1025-1028.
21. J. Jansen, R. Peschar and H. Schenk. *J. Appl. Cryst.* **25** (1992) 237-243.
22. J. Rius. *Powder Diffraction* **14** (1999) 267-273.