Solving Structures from Powder Data in Direct Space - State of the Art

Armel Le Bail - *Université du Maine, Laboratoire des oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France. Email : alb@cristal.org*

Abstract - To a model of atoms arranged in a cell (in direct space), even a bad one, corresponds a calculated powder pattern which may be compared to some experimental data. Solving the structure consists in the successful global optimization of the model characteristics and of the fit to the experimental data. More than ten computer programs are now available which may perform the job, they will be shortly described.

Introduction

The final step in a structure determination by powder diffractometry (SDPD) is always realized by the Rietveld method application. Going to this last step is not possible without at least an approximate model to be improved by the Rietveld refinement and eventually completed by Fourier difference synthesis. How can be obtained this starting approximate (or sometimes complete) model is the question considered here, having yet recorded a powder pattern, established that the structure is unknown, indexed the powder pattern, proposed a space group, and possessing some chemical knowledge about the sample.

Chemical knowledge is indispensable to the application of the direct space methods since they consist in placing atoms, either independent of as a whole molecule, at some positions in the cell, generally wrong positions at the beginning of the process, and moving them by translations (as well as rotations for a molecule) up to obtain a satisfying fit to the powder pattern or to a mathematical representation of that pattern. Going from wrong atomic positions to the final roughly correct ones is made by a process called global optimization which can be realized by different but finally similar procedures: Monte Carlo (MC), Monte Carlo with simulated annealing (SA) or/and with parallel tempering (PT), genetic algorithm (GA). These processes present a similarity in the use of random number sequences: atoms and molecules realize a random walk.

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 [1-4]. "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.

Computer programs

Direct space computer programs have incredibly proliferated during the ten last years (see the table below) and have extended the SDPD feasibility limits to larger problems, especially for molecular compounds. Irrespectively to the number of atoms, a molecule can be located easily in a cell, as a rigid body, corresponding to 3 positional and 3 orientational degrees of freedom (DoF), by checking the fit quality on, say, the first 50 peaks of the diffraction pattern. But the number of DoFs will increase by one for every added free torsion angle, and more complications arise if several independent molecules have to be located altogether or/and if water molecules or chlorine/sulphur/etc atoms are involved. For inorganic compounds, in principle an atom in general position corresponds to 3 DoFs (the *xyz* atomic coordinates), however, chemistry may say if some polyhedra are to be expected, then an octahedron for instance, instead of corresponding to 7x3=21 DoFs when described by the atomic coordinates, can be translated and rotated as a whole octahedron, corresponding to only 6 DoFs. Most of these computer programs are also able to start

from a complete set of independent atoms, at random at the beginning, and then will try to find their positions, moving them while matching to the data. The question of an atom moving close to a special position or then moving back to a general position is adressed by a modification of the chemical composition (in the program FOX). Combinations of molecules together with independent atoms are of course possible. The main difficulty may come finally at the Rietveld refinement stage, if the powder pattern quality becomes too low compared to the number of parameters, then it will be necessary to apply some constraints and/or restraints. To this list of programs may be added a few others which have special abilities for zeolites (ZEFSA-II [16], FOCUS [17], GRINSP [18]).

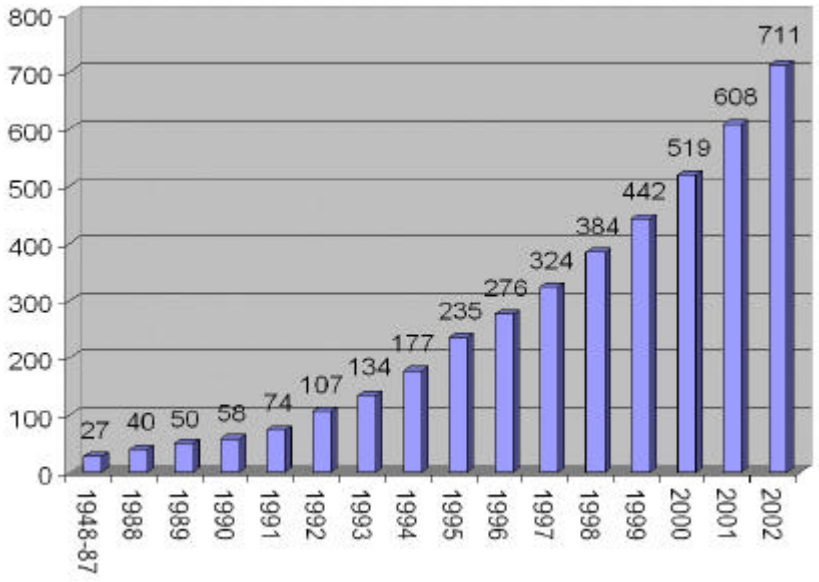
The winners of a second recent SDPD Round Robin in 2002 [19] on some organometallic and inorganic samples were the programs FOX and TOPAS (the first Round Robin in 1998 had seen the success of DASH). The conclusion is that SDPD is still by no means a routine task.

Selection of programs applying direct space methods of structure solution from powder diffraction data

Program	**Access**	**GO**	**Data**	**Example**	**DoF**	**Ref**
DASH	C	SA	P	Capsaicin	16	[5]
EAGER	A	GA	WP	$Ph_2P(O)(CH_2)_7P(O)Ph_2$	18	[6]
ENDEAVOUR	C	SA	I	Ag_2PdO_2	45	[7]
ESPOIR	O	MC	L	Gormanite	54	[8]
FOX	O	SA	WP	$Al_2(CH_3PO_3)_3$	24	[9]
OCTOPUS	A	MC	WP	Red Fluorescein	7	[10]
POWDERSOLVE	C	MC	WP	Docetaxel	29	[11]
PSSP	O	SA	L	Malaria Pigment Beta Haematin	14	[12]
SAFE	A	SA	WP	$C_{32}N_3O_6H_{53}$	23	[13]
SA	A	SA	WP	$(CH_2CH_2O)_6$:$LiAsF_6$	79	[14]
TOPAS	C	SA	WP	Caffeine Anhydrous	93	[15]

Access : C = Commercial with academic prices, O = Open access, A = contact the authors
GO = Global Optimization : MC = Monte Carlo, SA = MC+Simulated Annealing, GA = Genetic Algorithm
Data : P = Pawley, L = Le Bail, I = Integrated intensities, WP = Whole Pattern
DoF = degrees of freedom corresponding to the example
Ref : Reference corresponding to the example (not necessarily to the program publication)

These computer programs are obtaining more and more success, surpassing in number the solutions by traditional approaches (Patterson or Direct methods as applied in computer programs like SHELXS - etc - or adapted to powder data in EXPO). Nevertheless, the number of SDPD per year remains quite small (close to 100, to be compare to 30000 from single crystal data).

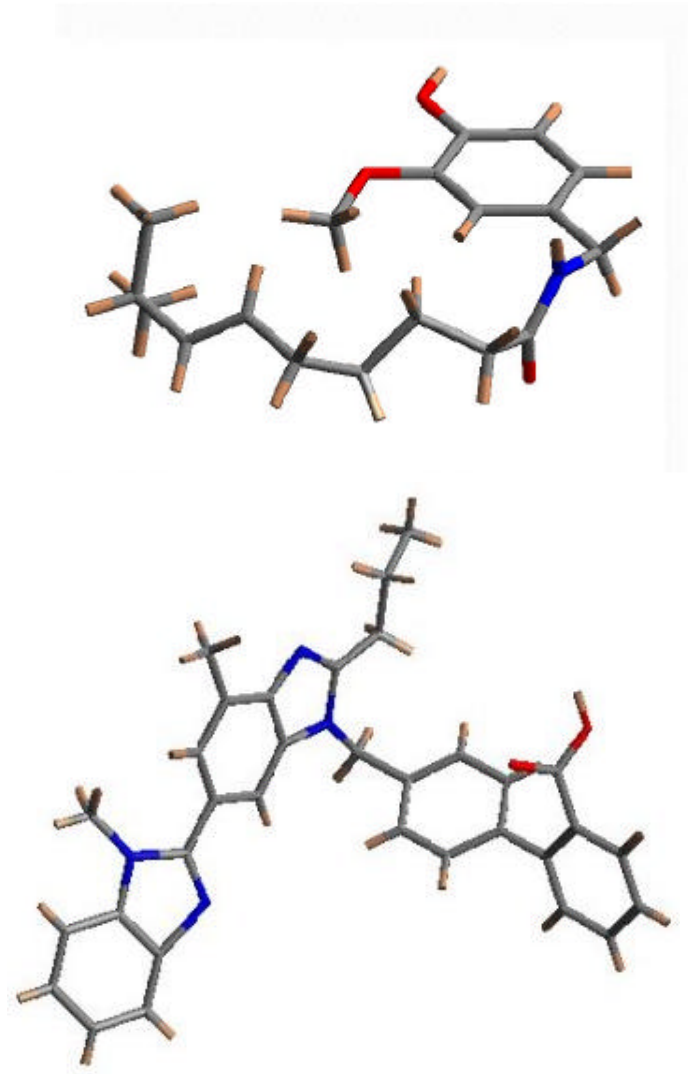


Cumulated histogram of the number of published SDPD. Picture from the SDPD Database: http://www.cristal.org/iniref.html

More details about the direct space computer programs are shortly given below, as they were presented by Yuri G. Andreev at the EPDIC-8 congress (Uppsala, Suède, 2002), obtained from the authors themselves. Things have not changed a lot after two years.

DASH

Authors : *W.I.F. David and K. Shankland* Rutherford Appleton Laboratory,
subsequent developments by *J. Cole and J. van de Streek* CCDC, UK



- Process applied : simulated annealing

- Working on correlated integrated intensities (extracted by the Pawley method)

- Program reference : *Chem. Commun.* (1998) 931.

- Typical examples :
  Capsaicin - *Chem. Commun.* (1998) 931.
  16 degrees of freedom, including 10 torsion angles
  Telmisartan, forms A and B – *J. Pharm. Sci.* 89 (2000) 1465.
  13 degrees of freedom, including 7 torsion angles.

- Commercial program, available at reduced cost (95%) to academic users.

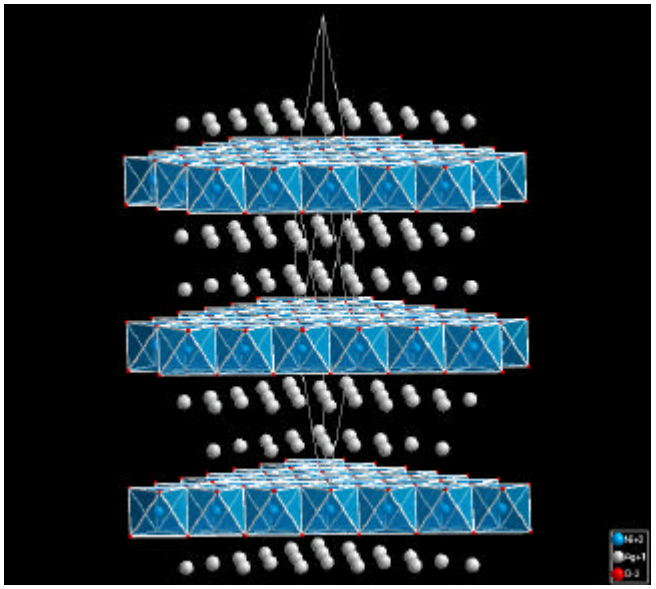
EAGER

Authors **:** *K.D.M. Harris, R.L. Johnston, D. Albesa Jové, M.H. Chao, E.Y. Cheung, S. Habershon, B.M. Kariuki, O.J. Lanning, E. Tedesco, G.W. Turner, University of Birmingham, UK*

- Process applied : Genetic algorithm.
- Working on the powder pattern
- Program reference : *Acta Cryst.* A54 (1998) 632.
- Typical example :
  Heptamethylene-1,7-bis(diphenylphosphane oxide) $Ph_2P(O)(CH_2)_7P(O)Ph_2$
  Angew. Chem. Int. Ed. 38 (1999) 831.
  35 atoms (non-H) in the a.u., 18 Degrees of freedom, including 12 torsion angles.
- Availability : in active development.

ENDEAVOUR

Authors : *K. Brandenburg and H. Putz, Crystal Impact, Bonn, Germany*



- Global optimization of both a R factor and a potential energy by simulated annealing

- Working on integrated intensities

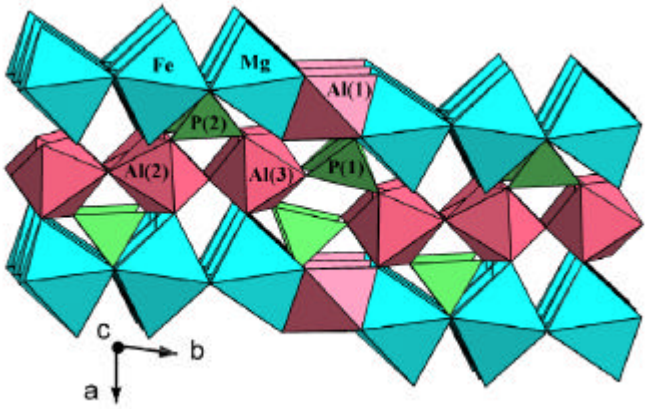
- Program reference : *J. Appl.Cryst.* 32 (1999) 864.

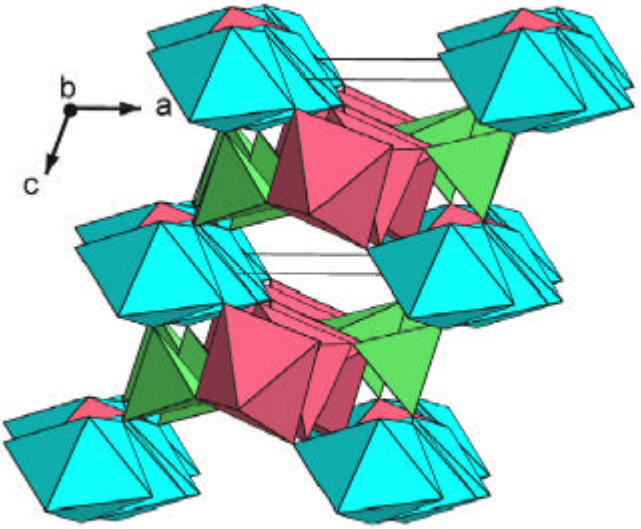
- Typical example : Ag_2NiO_2, *Schreyer and Jansen, Sol. State Sci.* **3** (2001) 25.
  with 15 atoms in the asymetric unit (if considered in the P1 space group)
  corresponding to 45 degrees of freedom.

- Commercial program, available at reduced cost to academic users.

ESPOIR

Author : *A. Le Bail, Université du Maine, France*





- Process applied : Monte Carlo

- Working on the integrated intensities (extracted by the Le Bail method), in fact on a pseudo powder pattern rebuilt from these extracted "|Fobs|".

- Program reference : *Mat. Sci. Forum* 378-381 (2001) 65.

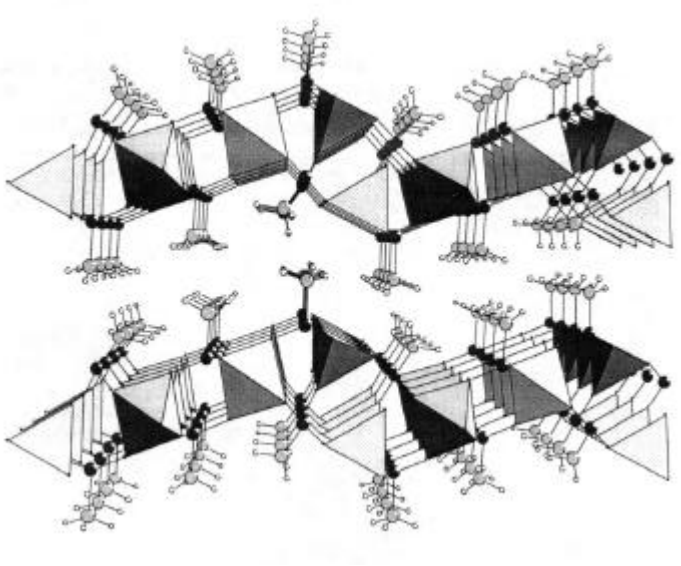
- Typical example :
Souzalite/Gormanite, *European J. Mineralogy* 15 (2003) 719.
19 atoms in the a.u. in P-1. Fe fixed at 0,0,0; 54 Degrees of freedom.

- Free and open – entirely available : executable + Fortran and Visual C++ source code (GPL - GNU Public Licence).

- Web site : http://www.cristal.org/sdpd/espoir/

FOX (Free Objects for Xtallography)

Authors : *V. Favre-Nicolin and R. Cerny, Geneva University, Switzerland.*



- Process applied : Parallel tempering or simulated annealing, with automatic correction of the special positions occupancies and of the sharing of atoms between polyhedra; can cope with multi-phase pattern.

- Working on the raw powder pattern or the integrated intensities.

- Program reference : *J. Appl. Cryst.* 35 (2002) 734.

- Typical example:
  Aluminium methylphosphonate $Al_2(CH_3PO_3)_3$ – *Chem. Commun.* (2002) 808.
  3 molecules and 2 Al atoms in the a.u.
  24 dgrees of freedom, including distances and angles

- Available for free, open source published under the Gnu Public Licence (GPL)

- Web site : http://objcryst.sourceforge.net

OCTOPUS

Authors **:** *K.D.M. Harris, M. Tremayne and B.M. Kariuki, University of Birmingham, UK*

- Process applied : Monte Carlo

- Working on the raw data (powder pattern).

- Program reference : *J. Am. Chem. Soc.* **116** (1994) 3543.

- Typical example :
  Red fluorescein – *Angew. Chem. Int. Ed.* **36**, (1997) 770.
  25 atoms (non-H) in the a.u., 7 Degrees of freedom, including one torsion angle

- Availability : in active development.

POWDERSOLVE (part of the software suite Reflex Plus)

Authors : *G. Engel, S. Wilke, D. Brown, F. Leusen, O. Koenig, M. Neumann, C. Conesa-Morarilla - Accelrys Ltd., Cambridge, UK*

- Process applied : Monte Carlo / simulated annealing and Monte Carlo / parallel tempering (*Falcioni and Deem. J. Chem. Phys.* 110 (1999) 1754) .

- Working on the raw powder pattern

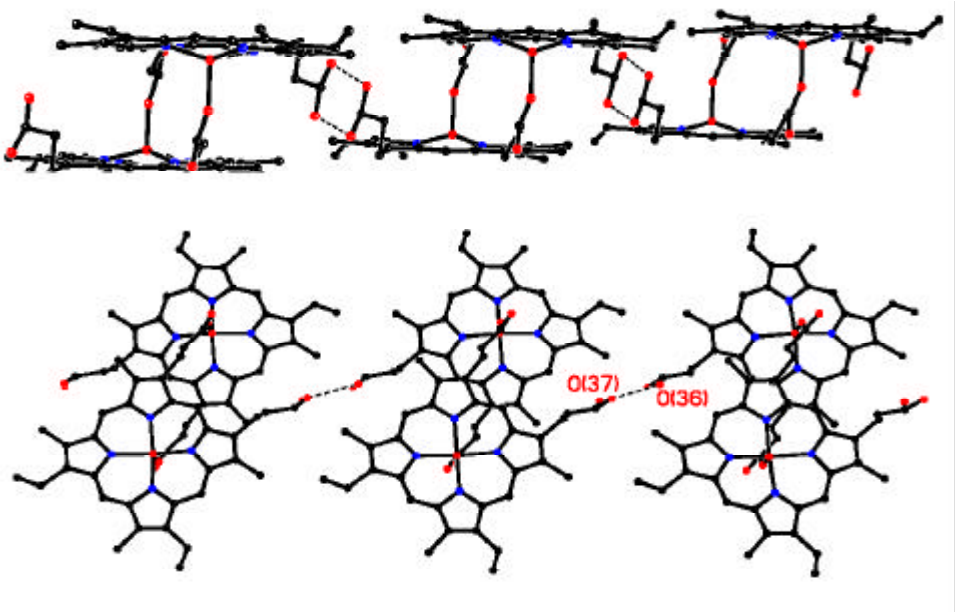
- Program reference : *J. Appl.Cryst.* **32** (1999) 1169.

- Typical example :
  Docetaxel ($C_{43}H_{53}NO_{14}\cdot 3H_2O$) – *J. Phys. IV,* Pr10 (2001) 221.
  29 Degrees of freedom including 3 rotations, 12 translations and 14 torsion angles.

- Availability : commercial (Accelrys Inc.), reduced cost for academic research.

PSSP (Powder Structure Solution Program)

Authors : **P**. **S***tephens and* **S**. **P***agola* State University of New York, Stony Brook, USA



- Process applied : simulated annealing.

- Working on the correlated integrated intensities (extraction by the Le Bail method)

- Program reference : *J. Appl. Cryst.* - submitted -
  Preprint available at http://powder.physics.sunysb.edu

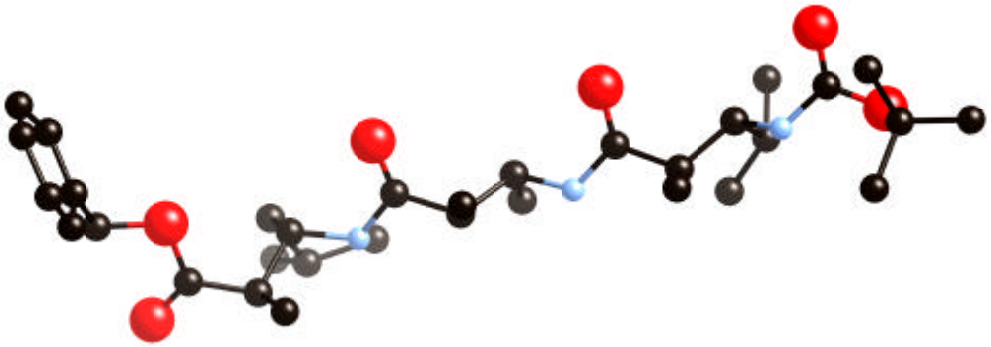
- Typical example :
  Beta Haematin Malaria Pigment - *Nature* 404 (2000) *307.*
  43 atoms (non-H) in the a.u. - 14 Degrees of freedom.

- Availability : free, including the source code.

- Web site : http://powder.physics.sunysb.edu/

SAFE (Simulated Annealing and Fragment search within an Envelope)

Authors : *S. Brenner, L.B. McCusker and Ch. Baerlocher* ETH Zentrum, Zurich, Switzerland



- Process applied : simulated annealing + use of a "structure envelope".

- Working on the raw powder pattern.

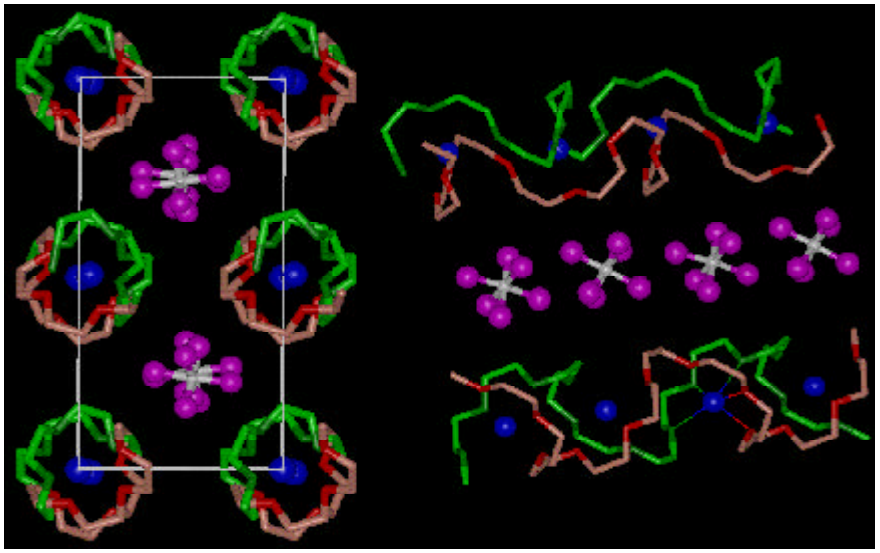
- Program reference : *J. Appl. Cryst.* 35 (2002) 243.

-Typical example :
Tri-b-peptide $C_{32}N_3O_6H_{53}$ - *J.Appl.Cryst.* 35 (2002) 243.
6 degrees of freedom (position and orientation) + 17 torsion angles.

- Availability : public domain.

Simulated Annealing

Authors : *Y. G. Andreev and P. G. Bruce,* University of St. Andrews



- Process applied : simulated annealing - special way to describe the molecule, without Z-matrix.

- Working on the raw powder pattern.

- Program reference : *J. Appl.Cryst.* **30**, (1997) 294.

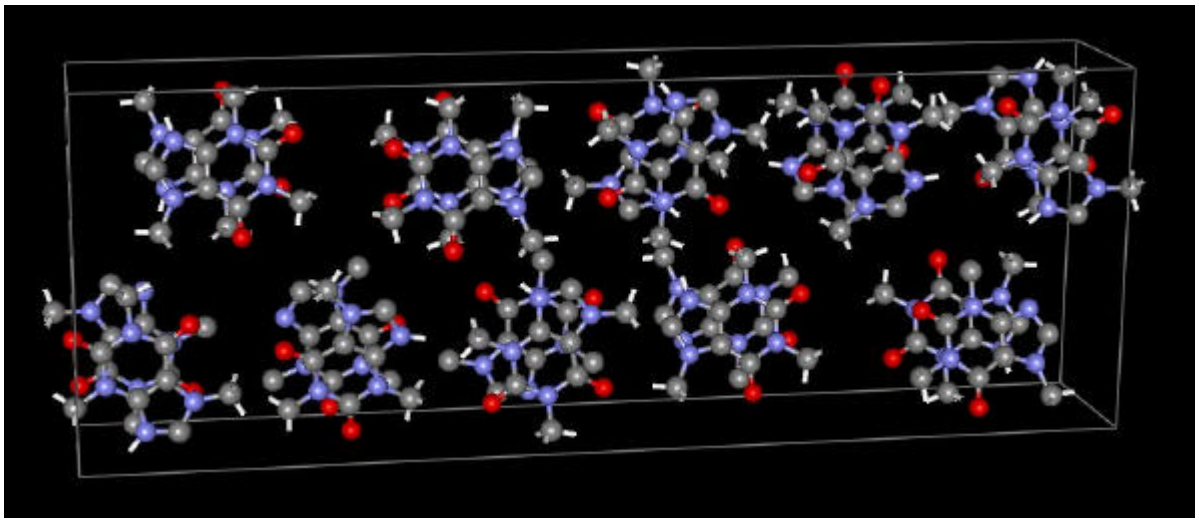
- Typical example :
$(CH_2CH_2O)_6$:$LiAsF_6$ – *Nature* 398 (1999) *792.*

26 atoms (non-H) in the a.u., 75 degrees of freedom, including 15 torsion angles.

- Availability : free - not user-friendly, requiring to change the code at every new problem (according to the authors) ...

TOPAS

Authors: *A.A. Coelho, R.W.Cheary, A. Kern, T. Taut. Bruker AXS GmbH, Karlsruhe, Germany*



- Process applied : simulated annealing (+ penalty functions definable by the user, rigid blocks, restraints on interatomic distances, energy minimization, including force field to be defined by the user).

- Working on the raw powder pattern or integrated intensities.

- Program reference : *J. Appl. Cryst.* 33 (2000) 899.

- Typical example :
Anhydrous caffeine $C_8H_{10}N_4O_2$ - abstract *XIX IUCr Congress (Geneva 2002)*
5 molecules in the a.u., 93 Degrees of freedom

- Commercial program, available at reduced cost for academic users.

- Web site : http://pws.prserv.net/

Conclusions

The capacities for solving structures from powder diffraction data have never been so efficient than in the past 5 years. One has to find his way in the SDPD maze and to select the appropriate methods and computer programs at each step of the problem (identification - which should fail to establish any relation with a known structure-, indexing, structure solution, Rietveld refinement). The advice is first to select the appropriate radiation, a 3rd generation synchrotron pattern being the best choice for complex cases.

Applying direct space methods requires generally much less data (3 to 5 intensities per degree of freedom may be sufficient) than direct methods. However, big organic or organometallic problems can be solved only if one disposes of a maximum of knowledge about the molecular formula. Finally, such very complex molecules will present more serious difficulties at the Rietveld structure refinement stage : the ratio of the effective number of structure factors with the number of atomic coordinates to refine may be as small as 3 or less (because there is soon no accurate intensity on the powder pattern at resolution d <1.5 Å), so that the model needs to be constrained/restrained. This may lead to difficulties to locate some additional water molecule, or to be absolutely sure that there

is not any misunderstanding somewhere which could explain why the Bragg R factor R_B is going to be sometimes as large as 10 or 15%. No need to say that some proposed H atom positions will have sometimes a low credibility.

References

[1] M. W. Deem and J. M. Newsam, *Nature* **342** (1989) 260-262; M. W. Deem and J. M. Newsam, *J. Am. Chem. Soc*. **114** (1992) 7189-7198; J.W. Newsam, M.W. Deem & C.M. Freeman, Accuracy in Powder Diffraction II, *NIST Special Publication* **846** (1992) 80-91.
[2] L.A. Solovyov & S.D. Kirik, *Mat. Sci. Forum* **133-136** (1993) 195-200.
[3] K.D.M. Harris, M. Tremayne, P. Lightfoot & P.G. Bruce, *J. Am. Chem. Soc.* **116** (1994) 3543-3547.
[4] D. Ramprasad, G.P. Pez, B.H. Toby, T.J. Markley & R.M. Pearlstein, *J. Amer. Chem. Soc*. **117** (1995) 10694-10701.
[5] W.I.F. david, K. Shankland & N. Shankland, *Chem. Commun*. (1998) 931-932.
http://www.ccdc.cam.ac.uk/products/powder_diffraction/dash/
[6] B.M. Kariuki, P. Calcagno, K.D.M. Harris, D. Philp & R.L. Johnston, *Angew. Chem.-Int. Edit.* **38** (1999) 831-835.
[7] M. Schreyer & M. Jansen, *Solid State Sciences* **30** (2001) 25-30.
http://www.crystalimpact.com/
[8] A. Le Bail, P.W. Stephens & F. Hubert, *Eur. J. Miner*. **15** (2003) 719-723.
http://www.cristal.org/sdpd/espoir/
[9] M. Edgar, V.J. Carter, D.P. Tunstall, P. Grewal-P, V. Favre-Nicolin, P.A. Cox, P. Lightfoot & P.A. Wright, *Chem. Comm*. **8** (2002) 808-809. http://objcryst.sourceforge.net/
[10] M. Tremayne, B.M. Kariuki & K.D.M. Harris, *Angew. Chem. Int. Ed*. **36** (1997) 770-772.
[11] L. Zaske, M.-A. Perrin and F. Leveiller, *J. Phys. IV*, **Pr10** (2001) 221.
http://www.accelrys.com/technology/firsts.html
[12] S. Pagola, P.W. Stephens, D.S. Bohle, A.D. Kosar & S.K. Madsen, *Nature* **404** (2000) 307-310.
http://powder.physics.sunysb.edu/Software.html#Software
[13] S. Brenner, L.B. McCusker & C. Baerlocher, *J. Appl. Cryst*. **35** (2002) 243-252.
[14] G.S. MacGlashan, Y.G. Andreev & P.G. Bruce, *Nature* **398** (1999) 792-794.
[15] F. Stowasser & C. Lehmann, *Acta Cryst.* **A58** (2002) C265.
http://pws.prserv.net/Alan.Coelho
[16] M. Falcioni & M. W. Deem. *J. Chem. Phys*. **110** (1999) 1754-1766.
http://www.mwdeem.rice.edu/zefsaII/
[17] R.W. Grosse-Kunstleve, L.B. McCusker, Ch. Baerlocher, *J. Appl. Cryst.* **30** (1997) 985-995.
http://zeolites.ethz.ch/software/focus/
[18] A. Le Bail, *IUCr Comp. Comm. Newsletter* **4** (2004) 37-45.
http://www.cristal.org/grinsp/
[19] A. Le Bail & L. Cranswick, *IUCr Commission on Powder Diffraction Newsletter* **29** (2003) 31-34.
http://www.cristal.org/sdpdrr2/