

# COMMISSION ON POWDER DIFFRACTION

## INTERNATIONAL UNION OF CRYSTALLOGRAPHY

<http://www.iucr.org/iucr-top/iucr/cpd.html>

NEWSLETTER No. 35, June 2007

<http://www.iucr-cpd.org/Newsletters.htm>

### IN THIS ISSUE

## Real-Space and Hybrid Methods for Structure Solution from Powders

(Rosanna Rizzi, Editor)

CPD Chairman's message, <i>Bill David</i>	2	Computer Corner, <i>L.M.D. Cranswick</i>	24
Editor's message, <i>Rosanna Rizzi</i>	2	<b>BRASS 2.0, the Bremen Rietveld Analysis and Structure Suite, Ver. 2</b>	
WWW sites related to Powder Diffraction	3	<i>J. Birkenstock, R.X. Fischer, T. Messner</i>	25
Calls for contributions to CPD Newsletter 36	3	<b>Search-Matching Inorganics through the PPDF-1 (Predicted Powder Diffraction File)</b>	
How to receive the CPD Newsletter	3	<i>Armel Le Bail</i>	27
Contact names for advertisers	3	<b>Diffraction Software for the Spallation Neutron Source: DiffDANSEProject</b>	
IUCr Commission on Powder Diffraction	4	<i>S.J. Billinge, E. Bozin, P. Juhas, W. Zhou, C. Farrow, J. Liu, D. Bryndin</i>	29
<b>Real-Space and Hybrid Methods for Structure Solution from Powders</b>		<b>Advances in Genetic Algorithms for Direct-Space Structure Solution</b>	
<b>Patterson Techniques for Powder Data</b>		<i>K.D.M. Harris, Z. Zhou, Z. Pan</i>	31
<i>M.C. Burla, R. Caliendo, B. Carrozzini, G.L. Cascarano, L. De Caro, C. Giacovazzo, G. Polidori, D. Siliqi</i>	5	<b>Cultural and Differential Evolution in Direct Space Structure Solution</b>	
<b>Systematic Grid Search Helps to Make a Choice</b>		<i>M. Tremayne, S.Y. Chong</i>	33
<i>Vladimir V. Chernyshev</i>	7	<b>Integration of Direct Methods into Direct Space Techniques</b>	
<b>Crystal Structure Prediction with Evolutionary Algorithms</b>		<i>A. Altomare, R. Caliendo, C. Cuocci, C. Giacovazzo, A.G. Moliterni, R. Rizzi</i>	37
<i>Colin W. Glass, Artem R. Oganov</i>	9	<b>Progress in Solving Crystal Structures of Triglycerides from Powder Data using Direct-Space Techniques</b>	
<b>Solving Crystal Structures from Powder Data by Lattice Energy Minimisations</b>		<i>J.B. van Mechelen, R. Peschar, H. Schenk</i>	41
<i>M.U. Schmidt</i>	12	<b>Calendar of Events</b>	45
<b>Expanding FOX: Auto-Indexing, Grid Computing, Profile Fitting</b>		<b>News from the ICDD</b>	46
<i>R. Cerny, V. Favre-Nicolin, J. Rohlicek, M. Husak, Z. Matej, R. Kuzel</i>	16		
<b>DASH – Related Algorithmic and Computing Developments</b>			
<i>W.I.F. David, K. Shankland, A.J. Markvardsen, J. van de Streek</i>	19		

---

## CPD Chairman's Message

It is a decade since Armel Le Bail and Lachlan Cranswick issued a challenge to the powder diffraction community to solve two crystal structures from powder diffraction data alone (see <http://sdpd.univ-lemans.fr/iniref/ecm18/ecm18sdpdr.html>). Despite a generous timescale of around six weeks, there were less than a handful of correct solutions. Would the situation be different today? I think so - but I do have a number of caveats. Rosanna Rizzi has produced an excellent newsletter with a good number of authoritative articles covering a wide variety of techniques for solving structures. Both real and reciprocal space techniques are discussed alongside promising hybrid techniques. Reviewing the newsletter, I am pleasantly surprised at the range of fresh developments after a relative lull in the past few years. Indeed, one of the most exciting developments, the charge-flipping algorithm applied to powder diffraction data, is so new that there has not been the opportunity to cover the topic in this newsletter. So back to my caveats! With all the diversity of methodologies presented in this newsletter, are we nearer to saying that structure determination from powder diffraction data is routine. My personal view is that for the general user, it is not. For complex inorganic materials, where the 3D connectivity is not known, I think that the field needs further developments although hybrid methods and the recent developments in charge-flipping may improve this situation. Interestingly, the articles by Glass and Oganov and Le Bail that predict structures a-priori may help in this challenge by creating massive libraries of hypothetical compounds that can be screened when new materials are synthesised. For molecular systems, the situation is more controversial. There are a number of available computer programs presented in this newsletter and others exist that have not been mentioned. However, the literature records that very few of these programs have regularly solved the level of crystal structure complexity that is common in, for example, the pharmaceutical industry. Furthermore, very few of these programs are generally available for the powder diffraction and a number of those involve a cost that academia is unwilling to pay. For these latter programs, the introduction of academic "light" versions may be an attractive way to engage a more general audience. For those other programs that principally remain within the domain of the investigators, there is the challenge to disseminate the programs to the wider powder diffraction community. Perhaps then, Armel and Lachlan, we will all be ready for another SDPD Round Robin.

Bill David, Chair, Commission for Powder Diffraction

---

## From the Editor of Newsletter 35

### Real-Space and Hybrid Methods for Structure Solution from Powders

The solution of crystal structures from powder data has traditionally been performed by extracting individual integrated intensities directly from the powder diffraction data and then using these intensities in Patterson and Direct Methods structure solution packages that have been developed for single-crystal diffraction data. However, powder diffraction patterns are always affected by peak-overlap that makes it difficult to extract unambiguous values of the intensities of individual diffraction peaks. The inevitable consequence of this loss of information is the severe difficulty in subsequent attempts to solve the structures using these traditional methods. The solution of the problem requires the development of improved techniques for extracting the peak intensities (for which new efficient methods have been developed) or the alternative structure solution approaches that use the experimental powder diffraction profile directly as it is 'measured' or the mathematically equivalent extracted correlated integrated intensities. Over the past few years, many new strategies, called Direct Space Techniques, have been developed and now routinely used to solve crystal structures of complex organic compounds with a large number of atoms in the asymmetric unit and significant torsional flexibility. More recently, a combination of traditional methods with global optimisation approaches has been developed and applied with success to organic structure solution. The goal is to exploit the combined potential of both approaches. Structure envelopes reduce considerably the region to be explored by direct-space techniques; alternatively the use of Direct Methods may limit the number of degrees of freedom necessary to describe the structural model under investigation.

I wish to thank Prof. Norberto Masciocchi for helping in editing this Newsletter.

*Rosanna Rizzi*

---

---

## WWW sites related to powder diffraction

*The Commission on Powder Diffraction (CPD):* <http://www.iucr.org/iucr-top/iucr/cpd.html>

*The International Union of Crystallography (IUCr):* <http://www.iucr.org/>

*The International Centre for Diffraction Data (ICDD):* <http://www.icdd.com/>

*The International X-ray Analysis Society (IXAS):* <http://www.ixas.org/>

*CCP 14:* <http://www.ccp14.ac.uk/>

Submitting a proposal for neutron diffraction or synchrotron radiation X-ray diffraction is possible at many (publicly funded) large scale facilities in the world. It represents an important and frequently unique opportunity for powder diffraction experiments. A useful guide and information can be accessed through the following web-site, maintained by *R. Dinnebier* at: <http://www.fkf.mpg.de/xray>

This list is far from being complete and needs input from users and readers of the CPD Newsletter. Please send comments to *R. Dinnebier* ([r.dinnebier@fkf.mpg.de](mailto:r.dinnebier@fkf.mpg.de))

## Call for contributions to the next CPD Newsletter (No 36)

The next issue of the CPD Newsletter, edited by Simon Billinge, will appear in December 2007. Simon will greatly appreciate contributions from readers on matters of interest to the powder diffraction community, e.g. meeting reports, future meetings, developments in instruments, techniques, and news of general interest. Please contact him for sending articles and suggestions. Software developments can be directly addressed to Lachlan Cranswick or to the Editor of Newsletter No 36 (addresses are given below).

### *Prof. S. (Simon) Billinge*

Department of Physics and Astronomy,  
4268 Biomed. Phys. Sciences Building,  
Michigan State University. East Lansing, MI 48824  
Telephone: +1-517-3559200 x2202 | Fax: +1-517-3534500  
e-mail: [billinge@pa.msu.edu](mailto:billinge@pa.msu.edu)

Software developments can be addressed directly to:

### *Dr Lachlan M. D. Cranswick*

Canadian Neutron Beam Centre, National Research Council Canada  
Building 459, Chalk River Laboratories, Chalk River ON, Canada, K0J 1J0  
Telephone: +1 (613) 584-8811 ext 3719 ; C2: ext 3039 Fax: +1 (613) 584-4040  
e-mail: [lachlan.cranswick@nrc.gc.ca](mailto:lachlan.cranswick@nrc.gc.ca)  
<http://neutron.cnrc.gc.ca/>

Suggestions, corrections, comments and articles on new or updated software are appreciated especially if you know of new program features, program updates and announcements that should be mentioned here.

---

## How to receive the IUCr CPD Newsletter

If you wish to be added to the mailing list for the Newsletter of the IUCr Commission on Powder Diffraction or you have changed your address, please register using the following link:

Online CPD Newsletter Registration

If this does not work then please contact: W.I.F. David (Chair, UK) [bill.david@rl.ac.uk](mailto:bill.david@rl.ac.uk).

The Newsletter can also be downloaded in electronic format, as a .pdf file, from the CPD web-site.

## Companies

If you would like to advertise in this twice-yearly newsletter, please contact the chairman of the CPD (Bill David):  
e-mail: [bill.david@rl.ac.uk](mailto:bill.david@rl.ac.uk) Telephone: +44 1235 445179 | Fax: +44 1235 445383

# THE IUCR COMMISSION ON POWDER DIFFRACTION - TRIENNIUM 2005-2008

---

**Chairman: Prof. W.I.F. David (Bill)**

Rutherford Appleton Laboratory (CCLRC),  
Chilton, Oxon. OX11 0QX, United Kingdom  
Telephone: +44 1235 445179 | Fax: +44 1235 445383  
e-mail: [bill.david@rl.ac.uk](mailto:bill.david@rl.ac.uk)

**Secretary: Prof. A. N. Fitch (Andy)**

ESRF, BP220, F-38043 Grenoble Cedex, France  
Telephone : +33 476 88 25 32 | Fax: +33 476 88 25 42  
e-mail: [fitch@esrf.fr](mailto:fitch@esrf.fr)

**Prof. Dr. S. Billinge (Simon)**

Department of Physics and Astronomy, 4268 Biomed. Phys.  
Sciences Building, Michigan State University  
East Lansing, MI 48824  
Telephone: +1-517-3559200 x2202 | Fax: +1-517-3534500  
e-mail: [billinge@pa.msu.edu](mailto:billinge@pa.msu.edu)

**Prof. M. Delgado (Miguel)**

Laboratorio de Cristalografía, Departamento de Química,  
Facultad de Ciencias, La Hechicera.  
Universidad de Los Andes, Mérida 5101, Venezuela.  
Telephone: +58 274 240 13 72  
e-mail: [migueld@ula.ve](mailto:migueld@ula.ve)

**Dr. I. Madsen (Ian)**

CSIRO Minerals  
Box 312, Clayton South 3169  
Victoria, Australia  
Telephone: +61 395458785 | Fax: +61 39562 8919  
e-mail: [Ian.Madsen@csiro.au](mailto:Ian.Madsen@csiro.au)

**Prof. N. Masciocchi (Norberto)**

Dipartimento di Scienze Chimiche ed Ambientali,  
Università dell'Insubria.  
Via Valleggio 11, 22100 Como (Italy)  
Telephone: +39-031-326227 | Fax: +39-031-2386119  
e-mail: [norberto.masciocchi@uninsubria.it](mailto:norberto.masciocchi@uninsubria.it)

**Prof. Dr. D. Rafaja (David)**

Institute of Physical Metallurgy,  
TU Bergakademie Freiberg, Gustav-Zeuner-Str. 5  
D-09599 Freiberg, Germany  
Telephone: +49(3731)392299 | Fax: +49(3731)393657  
e-mail: [rafaja@ww.tu-freiberg.de](mailto:rafaja@ww.tu-freiberg.de)

**Dr. R. Rizzi (Rosanna)**

Istituto di Cristallografia (CNR)  
Via G. Amendola 122/o, 70126 Bari (Italy)  
Telephone: +39 080 5929157 | Fax: +39 080 5929170  
e-mail: [rosanna.rizzi@ic.cnr.it](mailto:rosanna.rizzi@ic.cnr.it)

**Prof. Dr. P. Stephens (Peter)**

Department of Physics & Astronomy  
State University of New York  
Stony Brook, NY 11794-3800, USA  
Telephone: +1 (631) 632-8156 | Fax: +1 (631) 632-4977  
e-mail: [pstephens@stonybrook.edu](mailto:pstephens@stonybrook.edu)

**Dr. P. Whitfield (Pam)**

CChem MRSC Energy Materials. Institute for Chemical  
Process and Environmental Technology, Building M12.  
National Research Council Canada  
1200 Montreal Road, Ottawa ON K1A 0R6, Canada  
Telephone: (613) 998 8462 | Fax: (613) 991 2384  
e-mail: [pamela.whitfield@nrc-cnrc.gc.ca](mailto:pamela.whitfield@nrc-cnrc.gc.ca)

**Consultants****Dr. J. Cline (James)**

Ceramics Division, National Institute of Standards and  
Technology, 100 Bureau Dr Stop 8523,  
Gaithersburg, Maryland, 20899, USA  
e-mail: [cline@nist.gov](mailto:cline@nist.gov)

**Dr. J. De Villers (Johan)**

Mineralogy Division, MINTEK, Private Bag X3015,  
Randburg, 2125, South Africa  
Telephone: +27(11)7094745 | Fax: +27(11)7094564  
e-mail: [jpdev@postino.up.ac.za](mailto:jpdev@postino.up.ac.za), [johandev@global.co.za](mailto:johandev@global.co.za)

**Dr R. E. Dinnebier (Robert)**

Max-Planck-Institut für Festkörperforschung,  
Heisenbergstrasse 1, D-70569 Stuttgart, Germany  
Telephone: +49-711-689-1503 | Fax: +49-711-689-1502  
e-mail: [r.dinnebier@fkf.mpg.de](mailto:r.dinnebier@fkf.mpg.de)

**Dr. I. Margiolaki (Irene)**

ESRF, BP220, F-38042 Grenoble Cedex, France  
Telephone: +33 476882148 | Fax: +33 476882907  
e-mail: [margiolaki@esrf.fr](mailto:margiolaki@esrf.fr)

**Prof. P. Scardi (Paolo)**

Dipartimento di Ingegneria dei Materiali e Tecnologie  
Industriali, Università di Trento, 38050 Mesiano (TN), Italy  
Telephone: +39 0461 882417/67 | Fax: +39 (461) 881977  
e-mail: [Paolo.Scardi@ing.unitn.it](mailto:Paolo.Scardi@ing.unitn.it)

**Dr. M. Yashima**

Tokyo Inst. of Technology, Grad School of Sci. & Engn. Dpt.  
of Matls. Sci. & Engn. Nagatsuta-cho 4259-J1-4  
Midori-ku Yokohama, 226-8502, Japan  
e-mail: [yashima-masatomo@mail.goo.ne.jp](mailto:yashima-masatomo@mail.goo.ne.jp)

**ICDD Delegate****Dr. J. A. Kaduk (Jim)**

BP Chemicals, 150 W. Warrenville Rd., P.O. Box 3011  
MCF-9, Naperville, IL, 60566, USA  
Telephone: +1(630)4204547 | Fax: +1(630)4205252  
e-mail: [James.Kaduk@innovene.com](mailto:James.Kaduk@innovene.com)

# PATTERSON TECHNIQUES FOR POWDER DATA

M.C. Burla<sup>1</sup>, R. Caliendo<sup>2</sup>,  
B. Carrozzini<sup>2</sup>, G.L. Cascarano<sup>2</sup>,  
L. De Caro<sup>2</sup>, C. Giacovazzo<sup>2,3\*</sup>,  
G. Polidori<sup>1</sup>, D. Siliqi<sup>2</sup>

<sup>1</sup> Dipartimento di Scienze della Terra - Piazza  
Università, 06100 Perugia, Italy.

<sup>2</sup> Istituto di Cristallografia, CNR, via Amendola 122/o,  
70126 Bari, Italy.

<sup>3</sup> Dipartimento Geomineralogico, Università di Bari,  
Campus Universitario, via Orabona 4, 70125 Bari,  
Italy.

\*Contact author; e-mail: carmelo.giacovazzo@ic.cnr.it

## Introduction

The most common approach for the *ab initio* crystal structure solution from powder diffraction data is the following: the integrated intensities are first extracted from the experimental profile and then used as input for Direct Methods (see, for example, the programs EXPO2004 [1] and XLENS [2]). The unavoidable uncertainties on the intensities assigned to single reflections, due to peak overlapping and difficult profile-shape parameterization, make the solution of the structure far from being straightforward. This is especially true for organic compounds, which showed to be more resistant to Direct Methods owing to the absence of heavy atoms.

Patterson approaches [3,4,5], have been continuously developed as an alternative to Direct Methods, even if they have been basically underutilized. They may be loosely divided in two categories: vector-search methods, where the information used consists of the list of interatomic vectors extracted from the Patterson map [6,7,8], and superposition techniques [9,10,11] where the Patterson map itself is used, properly shifted and superimposed to other Patterson-derived maps in order to obtain peaks in correspondence of the atomic positions. The deconvolution of the Patterson function (i.e., extracting a unique image of the crystal structure from the set of overlapping images) is an essential step of the method: it may be achieved by using the so called superposition minimum function (SMF) [12,13]. The method has been recently innovated and proved to be much more efficient than Direct Methods in solving protein structures, even in cases when no heavy atom is present (so to claim to a revenge of the Patterson Methods over Direct Methods) [14,15,16].

The application of the Patterson deconvolution techniques to the *ab initio* crystal structure solution from powder diffraction data is far from being straightforward, since the efficiency of the

superposition techniques may be weakened by wrong estimates of the reflection intensities, which can produce non-negligible distortions of the Patterson map. Furthermore powder structures: a) can suffer pseudo-translation effects much more frequently than proteins structures; b) may crystallize in centric space groups; c) often crystallize in high-symmetry space groups where Harker domains more complicated than in the case of proteins occur. A first attempt to use SMF techniques for the solution of two powder structure is reported in [17]. The aim of this paper is to apply to a large set of powder data structures a Patterson deconvolution procedure derived from that successfully applied to protein structures in [16].

## The procedure

The Patterson deconvolution procedure for powder data may be briefly schematised according to the following steps:

a) the reflection intensities are extracted from the indexed powder pattern by the Le Bail [18] or Pawley [19] decomposition techniques. Then the structure factors are normalized and a  $|E|^2$ -Patterson synthesis is calculated, where the E's are normalised structure factors;

b) the implication functions compatible with the space group of the crystal structure are singled out. An implication transformation  $I_s(\mathbf{r})$  generated by the  $s$ -th symmetry operator  $C_s$  is a function of the atomic position  $\mathbf{r}$  defined over the corresponding Harker section:

$$I_s(\mathbf{r}) = P(\mathbf{r} - C_s\mathbf{r})/n_s \quad (1)$$

where  $P$  is the Patterson function and  $n_s$  is the multiplicity of the Harker section, i.e. the number of different symmetry operators which generate the same Harker vector  $\mathbf{r} - C_s\mathbf{r}$ ;

c) the SMF is calculated by combining the independent implication functions (whose number is  $\bar{m}$ ) according to

$$SMF(\mathbf{r}) = \underset{s=1}{\bar{m}} \text{Min}[I_s(\mathbf{r})] \quad (2)$$

where the minimum operator  $\text{Min}$  indicates that the lowest value among the  $\bar{m}$  functions  $I_s(\mathbf{r})$  has been chosen. In case no symmetry is present (space group P1), the SMF coincides with the Patterson map itself;

d) a pivot peak  $\mathbf{r}_p$  is chosen among the highest peaks of the SMF map and it is used to calculate the minimum superposition function

$$S(\mathbf{r}) = \text{Min}[P(\mathbf{r} - \mathbf{r}_p), SMF(\mathbf{r})] \quad (3)$$

e) few cycles of electron density modification (EDM) are applied to reduce the residual Patterson symmetry present in the  $S(\mathbf{r})$  map by specific filtering procedures described in detail in [16];

f) a multisolution approach is followed: the highest SMF peaks can be used as pivots, and the corresponding set of phases are ranked according to suitable figures of merit calculated at the end of the EDM cycles;

g) the most promising trials are submitted to phase refinement by means of a Fourier recycling procedure, which includes cycles of least squares minimization.

The procedure has been optimized for facing the specific problems arising from powder diffraction. Among them we quote:

1) in some space groups not all the independent implication functions are used to construct the SMF. Those corresponding to one-dimensional Harker domains, for example, are excluded (but for the cases in which there is a unique Harker domain, as in Cc);

2) in centric space groups with symmetry higher than orthorhombic only the implication function generated by the inversion centre is used.

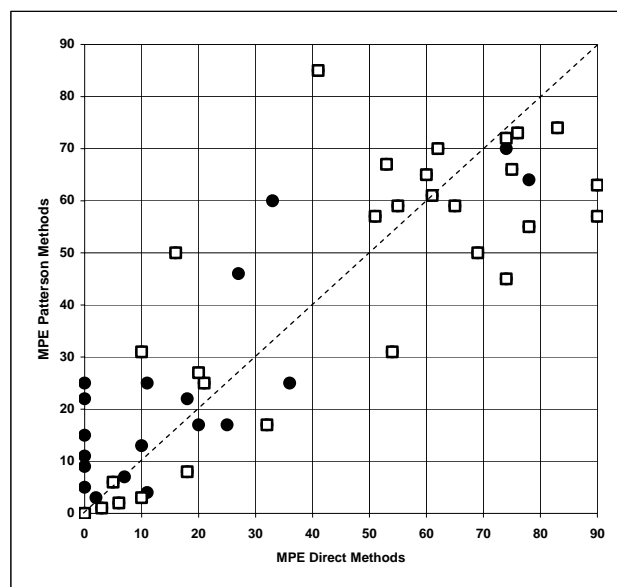
The reason for both the simplifications is the following: inaccuracy in the reflection intensity estimates may remarkably distort the Patterson map. Patterson peaks may then result shifted from their true positions. As a consequence, the use of all the implication functions may deplete, through the minimum function in equation (2), the SMF map.

## Applications

The above described procedure has been implemented into the EXPO2004 program [1], as alternative to the standard Direct Methods approach. A set of 50 structures has been used for the tests, whose code name and crystal chemical data are not reported for brevity. The structures have no more than 60 atoms in the asymmetric unit, 63% of them crystallize in centric space groups, more than half of them have only light atoms (i.e., atomic number lower than 17): among them, 10 have oxygen as heaviest atomic species. Data were taken by conventional X-ray devices (48%), by using synchrotron radiation (24%) and neutron sources (16%). Each structure has been processed by EXPO2004 two times: first by using the default tangent procedure, then by using the new Patterson approach. In the second case 30 trials are generated by exploring the highest pivot peaks of the SMF map and sorted by the same figure of merit. To compare the outcomes of the two procedures, we decided to stop the EXPO2004 run before the phase refinement module and to consider the best trial obtained by the two approaches at this stage. We use two parameters as relevant variables for the comparison of the two approaches: the mean phase error of the selected trial (MPE), calculated only for the largest reflections, and the number of peaks in the electron density map which fall within  $0.6\text{\AA}$  from their true position. In Figure 1 the MPE of the best solution obtained by Patterson methods are plotted versus that obtained by Direct Methods. The heavy atom structures are reported in full dots, those containing only light atoms are reported in open squares. It can be noted that the points

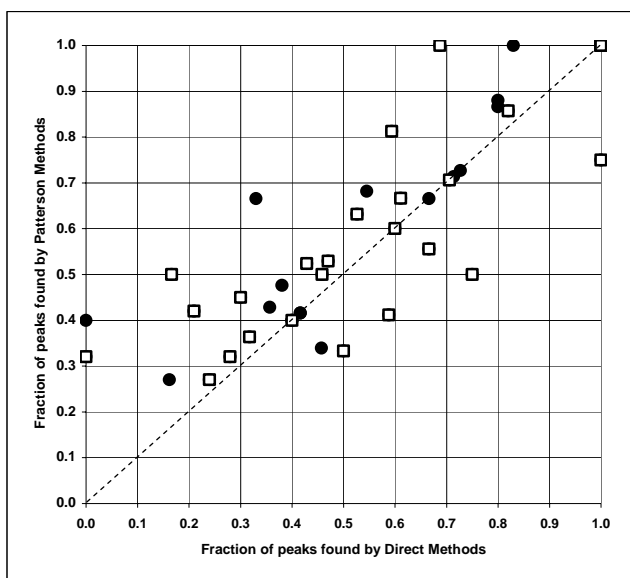
accumulate in two distinct regions of the plot: the first having  $\text{MPE} < 40^\circ$  for both the axes (this region contains the structures for which the phase problem has been suitably solved by both the methods) and the second having  $\text{MPE} > 50^\circ$  for both the axes (roughly speaking, this region contains the structures resistant to both approaches). The figure suggests that the two methods are not complementary and have nearly the same success rate. Going into details, we note that structures of the first region have mostly a heavy atomic species: they are solved slightly better by Direct Methods. On the other hand, mostly organic structures fall in the second region: slightly lower MPEs are obtained by our Patterson technique.

The number of atoms found within  $0.6\text{\AA}$  from their true positions, divided by the total number of atoms of the structure, is reported in Figure 2: again the best solutions obtained by our Patterson technique are plotted versus those obtained by Direct Methods. Also in this case the points are well correlated, even if there are more points above the bisector line than below it. This means that Patterson method is able to locate the atoms of the structures better than Direct Methods, even in cases where the structure is not completely solved.



**Figure 1:** Mean phase error (in degrees) obtained at the end of the Patterson procedure versus the same quantity obtained at the end of the Direct Methods procedure, for heavy atom (full dots) and light atom structures (open squares).





**Figure 2:** Fraction of peaks found within  $0.6\text{\AA}$  from their true positions in the map obtained at the end of the Patterson procedure versus the same quantity obtained at the end of the Direct Methods procedure, for heavy atom (full dots) and light atom structures (open squares).

## Conclusions

Patterson superposition methods have been applied to a large set of powder data and compared with standard Direct Methods procedures. The results show that the two methods have comparable success rate, but they are not complementary: i.e., structures resistant to one method are unlikely to be solved by the other. In summary, the Patterson methods are a valid alternative to standard tangent procedures in case of *ab initio* crystal structure solution from powder data.

## References

- [1] Altomare, A., Caliendo, R., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G., & Rizzi, R. 2004, *J. Appl. Cryst.* 37, 1025-1028.
- [2] Rius, J. 2004, *Z. Kristallogr.* 219, 826-832.
- [3] Patterson, A.L. 1934, *Phys. Rev.* 45, 763.
- [4] Patterson, A.L. 1934, *Phys. Rev.* 46, 372-376.
- [5] Harker, D. 1936, *J. Chem. Phys.* 4, 381-390.
- [6] Tickle, I.J. 1983, *Proceedings of the CCP4 Study Weekend. Location and Refinement of Heavy-Atom Sites*, edited by P. Machin, A. Wonacott & D.S. Moss, pp. 5-10. Warrington: Daresbury Laboratory.
- [7] Terwillinger, T.C., Kim, S.-H. & Eisenberg, D. 1987, *Acta Cryst.* A43, 1-5.
- [8] Knight, S. 1999, *Acta Cryst.* D56, 42-47.
- [9] Buerger, M. J. 1959, *Vector Space*, ch.11. New York Wiley.
- [10] Richardson, J. W. & Jacobson, R. A. 1987, *Patterson and Pattersons*, edited by J.P. Glusker, B. K. Patterson & M. Rossi, pp. 310-317. Oxford Univ. Press.
- [11] Sheldrick, G.M. 1992, *Crystallographic Computing 5*, edited by D. Moras, A.D.Podjarny & J.C. Thierry, pp.145-157. Oxford University Press.
- [12] Simpson, P.G., Dobrott, R.D. & Lipscomb, W.N. 1965, *Acta Cryst.* 18, 169-179.
- [13] Pavelčík, F. 1988, *Acta Cryst.* A44, 724-729.
- [14] Burla, M.C., Caliendo, R., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo & C., Polidori, G. 2004, *J. Appl. Cryst.*, 37, 258-264.
- [15] Burla, M.C., Caliendo, R., Camalli, M., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, R. 2005, *J. Appl. Cryst.*, 38, 381-388.
- [16] Burla, M.C., Caliendo, R., Carrozzini, B., Cascarano, G.L., De Caro, L., Giacovazzo, C., Polidori, G. & Siliqi, D. 2006, *J. Appl. Cryst.*, 39, 527-535
- [17] Estermann, M.A. & David, W.I.F. 2002 *Structure Determination from Powder Diffraction Data*, Edited by David, W.I.F., Shankland, K., McCusker, L.B. & Barleocher, Ch., pp.202-218. Oxford Science Publications.
- [18] Le Bail, A., Duroy, H. & Fourquet, J.L. 1988, *Mater. Res. Bull.* 23, 447-452.
- [19] Pawley, G.S. 1981, *J. Appl. Cryst.* 14, 357-361.

## SYSTEMATIC GRID SEARCH HELPS TO MAKE A CHOICE

Vladimir V. Chernyshev

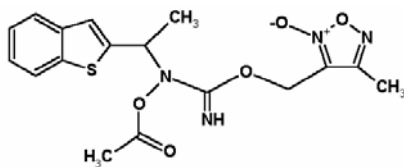
Department of Chemistry, Moscow State University,  
119992 Moscow, Russia

Contact author; e-mail: vladimir@struct.chem.msu.ru

Among the methods of crystal structure determination from powder diffraction data mentioned in this issue, the grid search technique is probably the simplest in application. A number of authors [1-7] has developed original computer codes, which implemented a grid search in structure solution, varying a small number of degrees of freedom (DOF), usually not more than ten. Of course, for the structures having few dozens of DOF, the use of algorithms based on random "wandering search" (Monte Carlo, simulated annealing, genetic algorithm, parallel tempering and some others) is preferable, because these algorithms require essentially less computation time compared with the grid search.

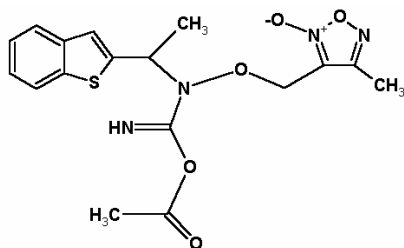
However, in some cases grid search technique is useful, providing an exhaustive exploration of all possible configurations in spite of time loss - these are the cases when the real (molecular) structure is essentially different from those used as starting prerequisites. An example of such unpleasant situation

- structure determination of 2,4-dinitro-*N*-phenyl-6-(phenylazo)benzamide - has been described in details elsewhere [8, 9]. Another example: the structures of (4-methyl-2-oxido-1,2,5-oxadiazol-3-yl)methyl *N*-(acetyloxy)-*N*-[1-(1-benzothien-2-yl)ethyl]imidocarbamate Figure 1 (expected)



**Figure 1:** structures of (4-methyl-2-oxido-1,2,5-oxadiazol-3-yl) methyl *N*-(acetyloxy)-*N*-[1-(1-benzothien-2-yl)ethyl]imidocarbamate

and acetyl *N*-[1-(1-benzothien-2-yl)ethyl]-*N*-[(4-methyl-2-oxido-1,2,5-oxadiazol-3-yl)methoxy]imidocarbamate Figure 2, obtained from powder diffraction data, also differ (unpublished data).

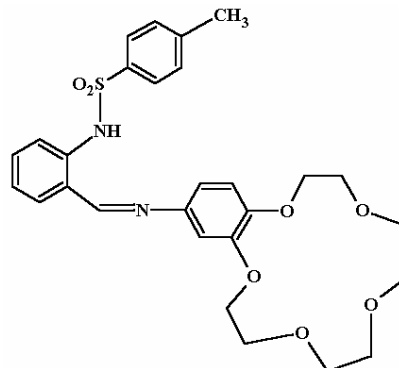


**Figure 2:** structure of *N*-[1-(1-benzothien-2-yl)ethyl]-*N*-[(4-methyl-2-oxido-1,2,5-oxadiazol-3-yl)methoxy]imidocarbamate

Such examples, when crystal structure determination from powder diffraction data leads to unexpected and correct molecular structure, demonstrate the power of powder methods, however, they are undesirable cases in the real practice. Indeed, as soon as you recognize that great efforts and a lot of time have been spent in vain with the prerequisite molecular structure, you need to make a choice: either to continue search for a solution with the old molecular geometry, or to start generation and use of a new set of initial molecular geometries. This choice does not offer a good decision: in the former case you continue to spend your time for the old molecular structure without any optimism, while the latter case is much worse, because you have no idea what are you looking for among the unlimited number of new possible molecules. Nevertheless, if you wish to find a solution, you need to decide between these two possible ways, and grid search can help you to check whether all possible configurations were tested with the given molecule. An illustrative example is given below.

In the course of systematic structural study of *N*-(4'-benzo-15-crown-5)-2-(amino-*N*-tosyl)phenylaldimine

(**HL**) complexes Figure 3, two compounds -  $\text{CuL}_2$  and  $\text{LiNCS}\cdot\text{HL}$  - did not yield single crystals suitable for X-ray crystal structure determination, but were obtained only as crystalline powders.



**Figure 3:** *N*-(4'-benzo-15-crown-5)-2-(amino-*N*-tosyl)phenylaldimine (**HL**) complexes

Therefore, their crystal structures were determined from synchrotron powder diffraction data [10].

For  $\text{LiNCS}\cdot\text{HL}$ , high-quality powder pattern provided a quick and routine structure solving. For  $\text{CuL}_2$ , the powder pattern was poor demonstrating weak counting statistics, compared with the case of  $\text{LiNCS}\cdot\text{HL}$ . Also, specific relationship between the unit cell dimensions –  $a\sin\beta \approx b \approx c$  – resulted in severe overlapping of the peaks. The routine methodology successfully applied for  $\text{LiNCS}\cdot\text{HL}$  structure solving gave no acceptable solutions for  $\text{CuL}_2$ . In addition, comparison of the unit cell volume for  $\text{CuL}_2$  [ $5287.7(3) \text{ \AA}^3$ ] with that for the related complex  $\text{ZnL}_2$  [ $5760.9(5) \text{ \AA}^3$ , single-crystal data] showed a significant difference. These considerations arose doubts whether the composition of  $\text{CuL}_2$  is correct. To be sure that we tested all possible methods of structure solving (assuming  $\text{CuL}_2$  formula), we have focused on location of the Cu atom. The translational grid search applied to the whole powder pattern (because of its poor quality) allowed selection of four Cu positions, which gave minimal values of  $R_{wp}$ . All these positions were further tested by the Rietveld refinement, varying  $x$ ,  $y$ ,  $z$  and  $U_{iso}$  parameters for the Cu atom only. This refinement allowed selection of the unique Cu position, which provided the lowest  $R_{wp} = 0.175$  (the nearest  $R_{wp} = 0.183$ ). This position of Cu atom was fixed, and then the crystal structure of  $\text{CuL}_2$  was successfully solved [10].

## References

- [1] Sonneveld, E.J. & Visser, J.W., 1978, *Acta Cryst.* B34, 643.
- [2] David, W.I.F., Ibberson, R.M., Matthewman, J.C., Prassides, K., Dennis, T.J.S, Hare, J.P., Krot, H.W., Taylor, R. & Walton, D.R.M., 1991, *Nature*, 353, 147.



- [3] Stephens, P.W., Bortel, G., Faigel, G., Tegze, M., Jánossy, A., Pekker, S., Oszlanyi, G. & Forró, L., 1994, *Nature*, 370, 636.
- [4] Masciocchi, N., Bianchi, R., Cairati, P., Mezza, G., Pilati, T. & Sironi, A., 1994, *J. Appl. Cryst.* 27, 426.
- [5] Dinnebier, R.E., Stephens, P.W., Carter, J.K., Lommen, A.N., Heiney, P.A., McGhie, A.R., Brard, L. & Smith, A.B., 1995, *J. Appl. Cryst.* 28, 327.
- [6] Mora, A.J. & Fitch, A.N., 1997, *J. Solid State Chem.* 134, 211.
- [7] Chernyshev, V.V. & Schenk, H., 1998, *Z. Kristallogr.* 213, 1.
- [8] Chernyshev, V.V., Yatsenko, A.V., Kuvshinov, A.M., Shevelev, S.A., 2002, *J. Appl. Cryst.* 35, 669.
- [9] Chernyshev, V.V., 2004, *IUCr CPD Newsletter*, 31, 5.
- [10] Dorokhov, A.V., Chernyshov, D.Yu., Burlov, A.S., Garnovskii, A.D., Ivanova, I.S., Pyatova, E.N., Tsivadze, A.Yu., Aslanov, L.A. & Chernyshev, V.V., 2007, *Acta Cryst.* B63. (In press).

---

## CRYSTAL STRUCTURE PREDICTION WITH EVOLUTIONARY ALGORITHMS

Colin W. Glass<sup>1</sup>, Artem R. Oganov<sup>1,2</sup>

<sup>1</sup>Laboratory of Crystallography, Department of Materials ETH Zurich, Wolfgang-Pauli-Str. 10, CH-8093 Zurich, Switzerland.

<sup>2</sup>Geology Department, Moscow State University, 119992 Moscow, Russia.

Contact author; e-mail: coglass@mat.ethz.ch,  
a.oganov@mat.ethz.ch.

### Introduction

An alternative way to experimental crystal structure solution is computational optimization [1-9]. While in experiment the limiting factors are obtaining a sample and diffraction data of sufficient quality, in optimisation it is having a fitness function of sufficient accuracy and the dimensionality of the problem. Since these factors differ fundamentally, it is reasonable to assume that the approaches can complement each other.

The well established approach is, of course, experimental structure solution. Its limiting factors indicate that in the area of high pressure research and materials design, optimisation can be of good use, since in the former it is hard or even impossible to

come by a sample and especially data of sufficient quality and in the latter producing samples for every promising material is expensive and work intensive. Furthermore high-pressure structures often involve only few atoms/unit cell, thus having a low dimensionality of the optimisation problem.

We have approached computational optimisation, implementing an evolutionary algorithm – USPEX (Universal Structure Predictor: Evolutionary Xtallography) [10-12]. Having tested USPEX on various systems with known stable crystal structure we have observed an unprecedented success rate of close to 100%.

### The method

Evolutionary algorithms usually are population based. New structures are produced by applying variation operators on structures of the current population. The probability of choosing a structure to act as parent for a new structure is positively correlated to its fitness.

For our method we chose the negative of the free energy as fitness; this is usually calculated *ab initio* (we can also use parameterised interatomic potentials). This is expensive but provides us with the most reliable quantification of the quality of a structure. Furthermore, we do not constrain the search to one predefined unit cell. The minimal input, besides parameters-values of USPEX, is a guess at the unit cell volume (to which we scale new structures and which evolves over the run), pressure and number of atoms/cell. Therefore, and since we usually start from random structures, we are independent from experimental data. However, knowing the unit cell *a priori* facilitates the search and having an experimental diffraction pattern (albeit low-quality) enables validation of the predicted structure.

We use the standard representation of structures – by 6 lattice parameters and fractional atomic coordinates of all atoms in the unit cell. In view of the fact that small distortions of a structure often have a greater influence on the free energy than profound structural differences do (e.g., bond stretching can involve greater energy changes than a polymorphic transformation), we have decided to locally optimise every candidate structure prior to evaluating it by means of its free energy. This greatly increases the cost of each calculated structure, but we find that this is essential for a feasible comparison of competing structures and thus for the method to work at all.

We have implemented three variation operators for USPEX: heredity, mutation and permutation.

In heredity two individuals (i.e. structures) are selected and used to produce one new candidate. This is achieved by taking a fraction of each individual and combining these. However, the fraction of each individual should contain as much information of the individual as possible. The main information within crystal structures is the relative position of the nearby atoms. Thus, to conserve information, the fraction of an individual is selected by taking a spatially coherent

slab. A slab is produced by cutting a structure at the position  $X$  (random number) of a randomly chosen lattice vector, in parallel to the plane spanned by the other two lattice vectors. One parent provides the slab  $[0, X]$ , the other the slab  $[X, 1]$ . The two slabs are fitted together and the result thereafter made feasible by adjusting the number of atoms of each type to the requirements.

When no space group information is used, the origin of the unit cell is unimportant - only the shape of the unit cell and the relative position of atoms are of importance. Thus, to avoid biasing the position of substructures within the unit cell, we usually shift the position of the atoms with respect to the unit cell before choosing a slab. The original and the shifted structure are physically identical. Taking the weighted average of the two parent lattices, where the weight is chosen randomly, produces the new lattice.

Mutation involves mutating the lattice and the atomic positions. We have found, however, that mutation of atomic positions is not necessary, since local optimization takes care of exact atomic positions. Mutation of the lattice is achieved by applying a strain matrix to the lattice. The strains are zero mean random Gaussian variables. Mutation of the lattice should be present for optimal performance, both to prevent a possibly premature convergence towards a certain lattice and for efficient exploration of the immediate neighbourhood of promising structures.

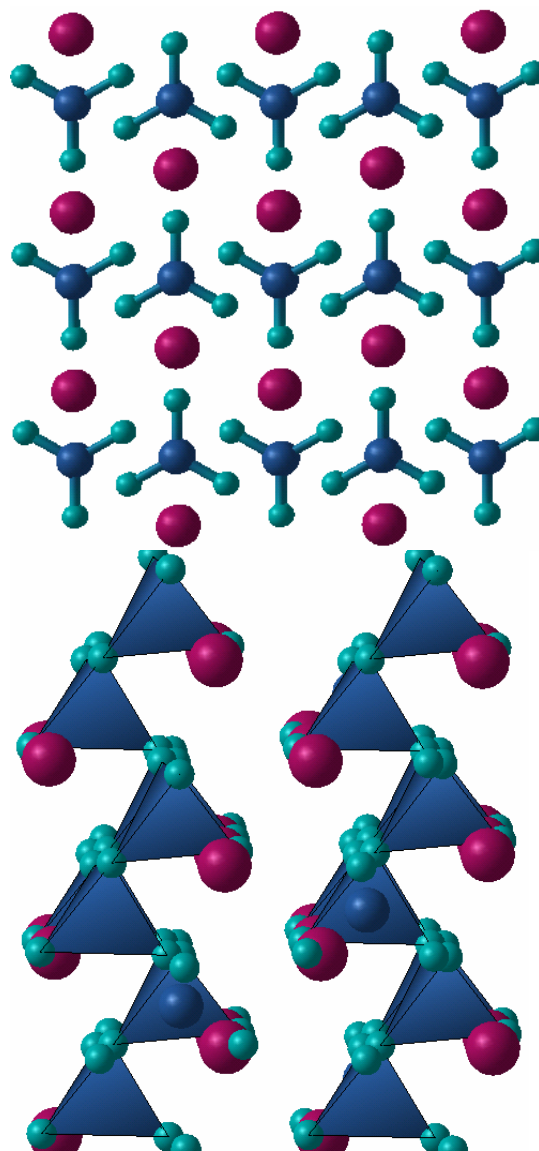
In permutation two atoms of different types are exchanged (as done in [8]) a variable number of times. Permutation facilitates finding the correct atomic ordering. Obviously, permutation is possible only for systems with different types of atoms.

The computationally expensive part of the algorithm is local optimisation. Locally optimising different candidates within one generation is independent and can thus be processed in parallel (but only within the same generation). This makes USPEX a very easily parallelisable code.

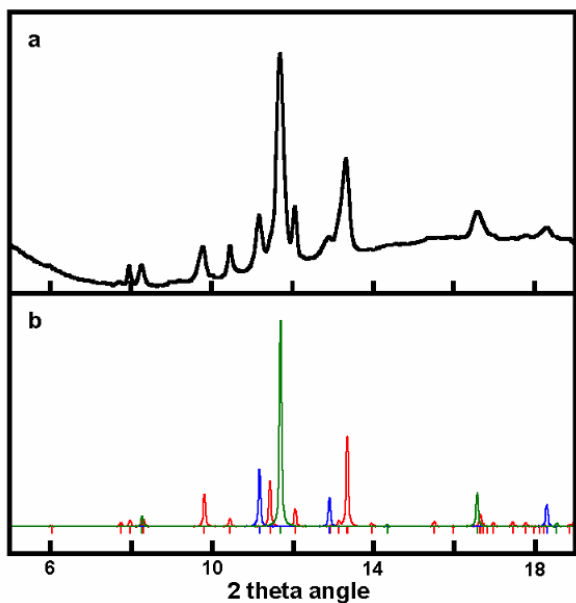
## Results

We have performed a large number of tests (on systems, where the stable structure is known). Starting only from the number of atoms/cell and a reasonable volume for the unit cell, we have observed a success rate of nearly 100%. Furthermore, since in the course of the search the focus on promising regions of the search space increases, many competitive local minima (or metastable structures) are found [11].

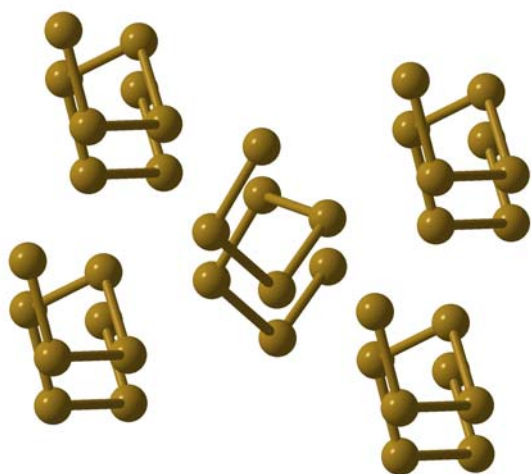
Using USPEX on systems with unknown structure, we have discovered a number of new stable high-pressure structures for several systems -  $\text{CaCO}_3$ , sulphur, oxygen, boron, and are currently investigating many more. Some illustrations are given in Fig. 1-4. Fig. 2 shows that the predicted structure of  $\text{CaCO}_3$  post-aragonite matches the experimental powder diffraction pattern.



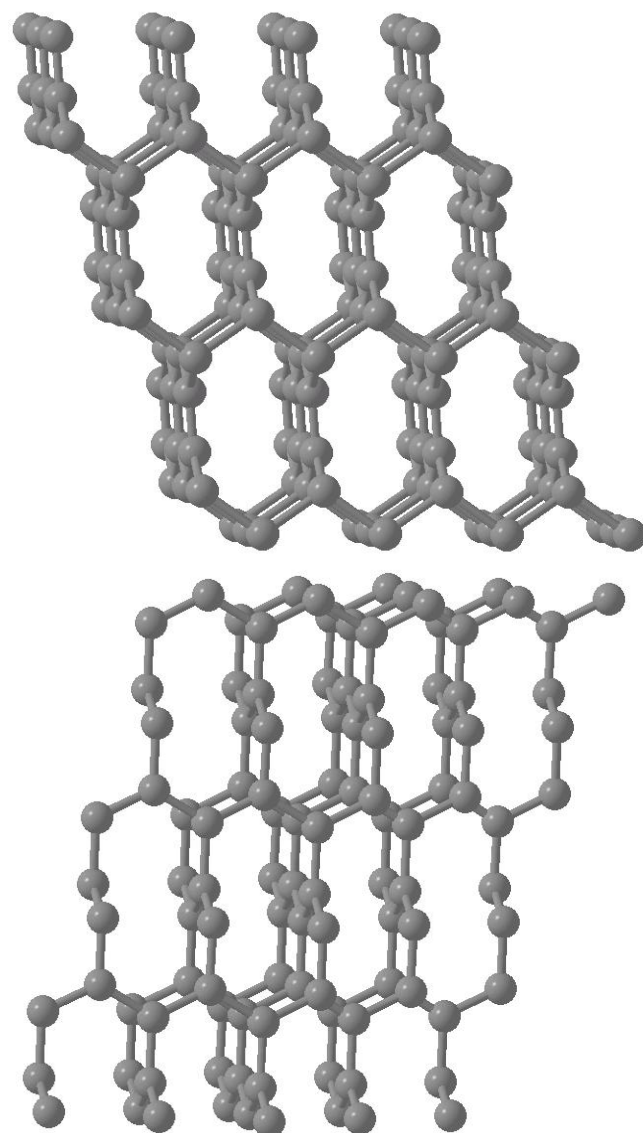
**Figure 1:** Stable structures of  $\text{CaCO}_3$  found by USPEX (from [12]). Top: post-aragonite, stable 42-137GPa. Bottom:  $C222_1$  phase, stable above 137GPa.



**Figure 2:** Comparison of experimental (top) and theoretical (bottom) diffraction patterns for  $\text{CaCO}_3$  post-aragonite (from [12]). Red peaks – post-aragonite, green – NaCl (used as pressure-transmitting medium), blue – Pt (used as a laser absorber and pressure calibrant). Experimental (theoretical) lattice parameters are:  $a = 4.28$  (4.30) Å,  $b = 4.69$  (4.73) Å,  $c = 4.06$  (4.10) Å.



**Figure 3:** Stable structure of sulphur found by USPEX (stable at 3-7GPa and 0 K, from [11]).



**Figure 4:** Metastable structures of C found by USPEX (from [11]).

## Discussion

Our approach yields a high success rate for systems with up to ~100 atoms/cell (the difficulty of the search depends both on dimensionality and landscape shape. The latter depends strongly on the system chemistry). USPEX can solve structures, for which only the unit cell parameters are known from experiment, but it can also predict new structures without any experimental information at all – and thus guide experimental exploration.

This method is a powerful tool in the field of high-pressure, where structures often involve only a few dozen atoms/cell. Another promising area of application is materials design regarding properties, which are strongly linked to the crystal structure (as opposed to depending on e.g. defects). The large set of competitive metastable structures that are found increases the applicability to this area.

Clearly, the quality of the global minimum found by USPEX depends on the quality of the *ab initio*

description of the system. Present-day density-functional simulations (e.g., within the GGA functionals) are adequate for most situations, but it is known that these simulations do not fully describe van der Waals bonding and the electronic structure of Mott insulators. In both areas there are significant advances (see e.g. [13-15]) which can be used for calculating *ab initio* free energies and evaluation of structures in evolutionary simulations.

### Acknowledgements

CWG thanks his dear friend Jo A. Helmuth for many fruitful discussions.

Calculations were performed at ETH Zurich and CSCS (Manno). We thank T. Racic, G. Sigut and O. Byrde for computational assistance, and M. Valle for developing the STM3 library for the visualisation of our results. This work is supported by the Swiss National Science Foundation under grant 200021-111847/1.

### References

- [1] Martoňák R., Laio A., Parrinello M. (2003). *Phys. Rev. Lett.* 90, art. 075503.
- [2] Martoňák R., Laio A., Bernasconi M., Ceriani C., Raiteri P., Zipoli F., Parrinello M. (2005). *Z. Krist.* 220, 489-498.
- [3] Pannetier J., Bassasalsina J., Rodriguez-Carvajal J., Caignaert V. (1990). *Nature* 346, 343-345.
- [4] Schön J.C., Jansen M. (1996). *Angew. Chem. – Int. Ed.* 35, 1287-1304.
- [5] Wales D.J., Doye J.P.K. (1997). *J. Phys. Chem.* A101, 5111-5116.
- [6] Gödecker S. (2004). *J. Chem. Phys.* 120, 9911-9917.
- [7] Bush T.S., Catlow C.R.A. & Battle P.D. (1995). *J. Mater. Chem.* 5, 1269-1272.
- [8] Woodley S.M., Battle P.D., Gale J.D., Catlow C.R.A. (1999). *Phys. Chem. Chem. Phys.* 1, 2535-2542.
- [9] Woodley S.M. (2004). *Structure and Bonding* 110, 95-132.
- [10] Glass C.W., Oganov A.R., Hansen N. (2006). USPEX - evolutionary crystal structure prediction. *Comp. Phys. Comm.* 175, 713-720.
- [11] Oganov A.R., Glass C.W. (2006). Crystal structure prediction using evolutionary algorithms: principles and applications. *J. Chem. Phys.* 124, art. 244704.
- [12] Oganov A.R., Glass C.W., Ono S. (2006). High-pressure phases of CaCO<sub>3</sub>: crystal structure prediction and experiment. *Earth Planet. Sci. Lett.* 241, 95-103.
- [13] Becke A.D. (1993). *J. Chem. Phys.* 98, 5648-5652.
- [14] Tao J.M., Perdew J.P., Staroverov V.N., Scuseria G.E. (2003). *Phys. Rev. Lett.* 91, art. 146401.
- [15] Foulkes W.M.C., Mitas L., Needs R.J., & Rajagopal G. (2001). *Rev. Mod. Phys.* 73, 33-83.

# SOLVING CRYSTAL STRUCTURES FROM POWDER DATA BY LATTICE ENERGY MINIMISATIONS

M.U. Schmidt

Institut für Anorganische und Analytische Chemie der Universität Frankfurt, Max-von-Laue-Str. 7, D-60438 Frankfurt am Main, Germany.

e-mail: m.schmidt@chemie.uni-frankfurt.de

### Introduction

A structure determination from powder diffraction data consists generally of three steps:

1. Indexing of the powder diagram
2. Structure solution
3. Rietveld refinement.

For the structure solution itself three main alternative approaches are available:

- a. Determination of integrated reflection intensities, and subsequent structure solution by e.g. direct methods, Patterson methods or maximum entropy methods. This approach generally requires diffraction data with high resolution and high precision.
- b. “Real-space methods”, i.e. fit of calculated versus experimental powder patterns by moving molecules or fragments inside the unit cell. For this approach, the knowledge of (approximated) lattice parameters is generally required. A very recent example shows that the method may also work without prior indexing [1].
- c. Prediction of crystal structures by global optimisation of the lattice energy. (Strictly speaking, this is also a “real space method”, since lattice energy minimisations work in real space). In most cases several structures are found within an energy range of a few kJ/mol above the global minimum. To find out which of the predicted crystal structures corresponds to the actual sample, powder diffraction patterns of all low energy structures are simulated and compared with experimental powder data.

Solving structures by lattice energy minimisation has the advantage that it works even for unindexed, low-quality powder diagrams (for an example see below). However, if the experimental lattice parameters are known (or even, if they are only partially known), this knowledge can be used to decrease the calculation times substantially.

We repeatedly used lattice energy minimisations to solve crystal structures of organic compounds from X-ray powder data, see e.g. [2-6].

In the following we will give an introduction into the method and show two examples, one with prior indexing, the other without prior indexing. The focus will be on molecular compounds.

## Lattice energy minimisations

A crystal structure prediction by lattice energy minimisation can be done either by quantum mechanical methods, or by force field methods [2-7], or by a combination of both.

Quantum mechanical methods have been successfully used for inorganic compounds as well as for very small organic molecules. For medium-sized molecules (50 to 100 atoms) the calculation times are still too high to be feasible for a full prediction of a crystal structure. Force field calculations with subsequent quantum mechanical optimisation of energetically low minima have been tested to serve as an alternative approach [15].

For molecular crystals there is another difficulty: most present density-functional methods have problems to describe the Van der Waals interactions correctly, but the Van der Waals interactions are the most important interactions for molecular crystals.

Force field methods have been used for the prediction of crystal structures since the work of Kitajgorodskij about 50 years ago [8-11]. Generally the atom-atom potential method is used, *i.e.* one assumes that the intermolecular energy can be divided into a sum of individual atom-atom potentials:

$$E = \frac{1}{2} \sum_i \sum_j E_{ij} . \quad (1)$$

$i$  denotes all atoms of a reference molecule (more correctly: all atoms in the asymmetric unit), and  $j$  stands for all atoms of all other molecules in the crystal.

The atom-atom potential is generally set up as a sum of individual terms for Van der Waals, electrostatic, hydrogen bond and intramolecular interactions.

In most cases the Van der Waals interactions contribute the vast majority of the lattice energy and result in dense packings, whereas the hydrogen bonds and the electrostatic energy decide which of these possible dense packings has a (sometimes only slightly) better energy than the others.

For the Van der Waals interactions two different expressions are commonly used, either the Lennard-Jones-Potential

$$E_{ij} = -A_{ij}r_{ij}^{-6} + B_{ij}r_{ij}^{-12} \quad (2)$$

or the Buckingham potential

$$E_{ij} = -A_{ij}r_{ij}^{-6} + B_{ij}e^{-C_{ij}r_{ij}} . \quad (3)$$

$A$ ,  $B$ , and  $C$  are empirical Van der Waals parameters. Due to the exponential term, the 6-*exp* potential may require more calculation time (but this has become less significant), but the repulsion part of the potential is closer to reality than for a 6-12 potential.

Van der Waals parameters have been empirically fitted for many elements. For organic compounds with or without hydrogen bonds, the Dreiding force field [12] generally gives suitable results, but for special cases the parameters have to be modified to give a better fit for the class of compounds under investigation. We use either the Dreiding parameters, or our own force field [2-4,7].

The summation of the atom-atom interactions should be carried out up to a cut-off radius of about 10 Å, which accounts for about 90% of the total Van der Waals energy. For precise calculations we use a cut-off radius of 20 Å, which includes more than 99% of the Van der Waals energy.

The Coulomb energy is given by:

$$E_{ij} = \frac{1}{4\pi\epsilon\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (4)$$

with  $q_i$  and  $q_j$  being the atomic charges of the atoms  $i$  and  $j$ . Many methods have been proposed to calculate atomic charges. We use either Gasteiger charges [13] or *ab-initio* electrostatic potential (ESP) derived charges. Instead of using point charges, electrostatic multipoles can be used, which give a better description of the electrostatic interactions [14-16].

For hydrogen bonds, various potentials are known, most of them depending on the X-H...X angle. Sometimes the hydrogen bond is regarded as being purely a combination of Van der Waals and Coulomb interactions; if the corresponding parameters and charges are carefully fitted, this approach works very well.

Instead of empirically fitted force field potentials, one can use trained potentials derived by data mining from the Cambridge Structural Database [17]. These potentials have been successfully used for structure determination from unindexed powder data [5,18,19].

For intramolecular interactions, much work has been done to develop good force fields. However, for crystal structure predictions it is in most cases sufficient to keep the molecule rigid, and to optimise only a few intramolecular degrees of freedom, esp. rotations around single bonds (except for methyl groups). The influence of packing forces on bond lengths and bond angles is so small, that these changes have no significant effect on the molecular packing.

The entropy is not calculated explicitly; but entropic effects are included in the force field parameters.

Several programs are available for crystal structure prediction [14-16,20]. We use our own program CRYSCA [2,7]. This program performs global lattice energy optimisations for rigid or flexible molecules, starting from a set of several thousands of random crystal structures with random values for lattice parameters, orientations and positions of the molecules. All starting values are chosen within user-defined, sensible ranges. If the lattice parameters are known, they are used as input and kept fixed during the optimisation. In the case of flexible molecules, the molecular flexibility is included throughout. The



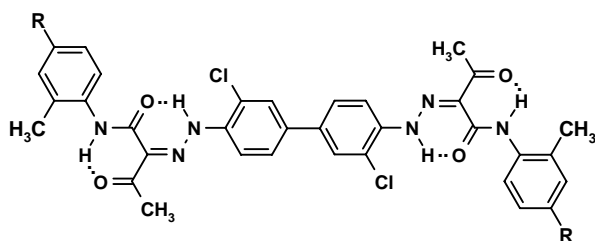
crystal symmetry is given as input. CRYSCA is able to handle all space groups and site symmetries (special positions), as well as disorder, noncrystallographic symmetry etc.

The global minimum of the lattice energy should correspond to the experimental crystal structure (within the limitations of the applied force field). However, most compounds show more than one polymorphic form. The lattice energies of these polymorphs differ in most cases by 0-5 kJ/mol. Consequently, a crystal structure prediction results in a list of possible crystal structures. In order to find out which of the calculated structures corresponds to the experimental ones, diffraction patterns of all low-energy minima are calculated and compared with the experimental powder data.

However, powder patterns are extremely sensitive to small changes in the molecular structure or the arrangement of the molecules; therefore, it may happen that the true structure is not recognised as being the correct one. This holds especially for unindexed powder data with less than 20 lines: due to deviations in the lattice parameters, peaks may be shifted or may overlap, resulting in powder patterns that visually do not look similar at all.

### Example 1: Structure solution from indexed powder diffraction data

Compounds **1a** and **1b** are industrial yellow pigments used for printing newspapers, journals etc. If this newsletter were distributed in a printed edition, the yellow parts of this newsletter would probably have been printed with **1a**.



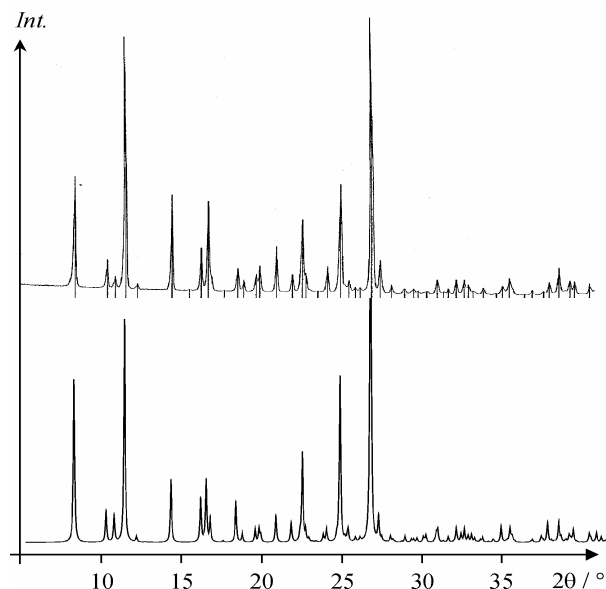
**1a:** Pigment Yellow 13 (R = CH<sub>3</sub>)

**1b:** Pigment Yellow 14 (R = H)

The powder diagram of **1b** is shown in figure 1 (top). Data were recorded in transmission mode on a STOE-Stadi-P diffractometer equipped with a primary Ge(111) monochromator and a linear position-sensitive detector. The compounds **1a** and **1b** are isotypic. The data could be indexed with (for **1b**)  $a = 8.369 \text{ \AA}$ ,  $b = 8.804 \text{ \AA}$ ,  $c = 12.478 \text{ \AA}$ ,  $\alpha = 71.29^\circ$ ,  $\beta = 76.14^\circ$ ,  $\gamma = 74.29^\circ$ ,  $V = 826.4 \text{ \AA}^3$ ,  $Z = 1$ . Since we did not know if the space group is  $P1$  or  $P\bar{1}$ , all calculations were performed in  $P1$ .

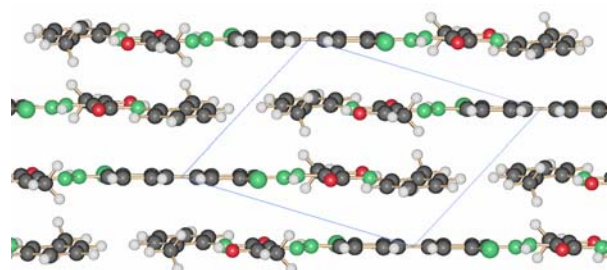
The molecular structure was constructed according to single crystal data of other azo pigments. Lattice energy minimisations were carried out by CRYSCA with fixed unit cells. The intramolecular degrees of

freedom (rotations around single bonds) were included throughout. The lowest energy minimum showed a simulated powder pattern very close to the experimental one (figure 1). The peak positions match, because the unit cell was given. The good match of intensities indicates that the structure is correct within about 0.1 Å in atomic coordinates or 0.5° in molecular orientation.



**Figure 1:** Top: Experimental X-ray powder data of **1b**. Bottom: Simulated pattern of the structure calculated by lattice energy minimisation with given unit cell parameters.

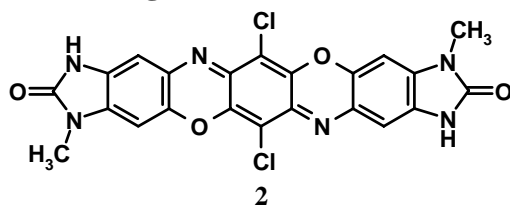
Generally the quality of the powder data is not a crucial point for solving crystal structures by lattice energy minimisation. But for the Rietveld refinements the powder data should be as good as possible. Therefore we recorded synchrotron diffraction data for **1a** and **1b**. The Rietveld refinements were done with GSAS using rigid bodies for the benzene rings (including substituents), and for the hydrogen-bonded six-membered rings. The diffraction data were so good, that we could remove all rigid bodies and constraints and refine all atoms individually. For **1b** it was even possible to refine individual isotropic temperature factors for all 23 symmetrically independent non-hydrogen atoms [21].



**Figure 2:** Crystal structure of Pigment Yellow 14 (**1b**). View direction [100]. The structure of **1a** is very similar to the structure of **1b**.



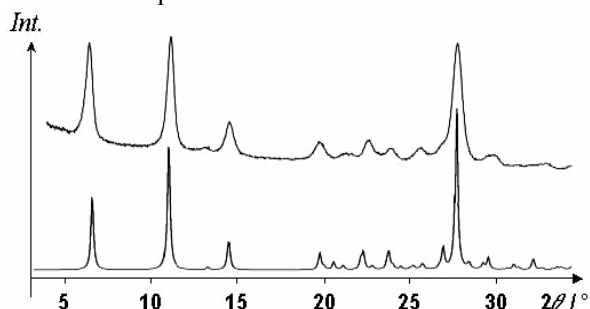
## Example 2: Structure solution without prior indexing [4]



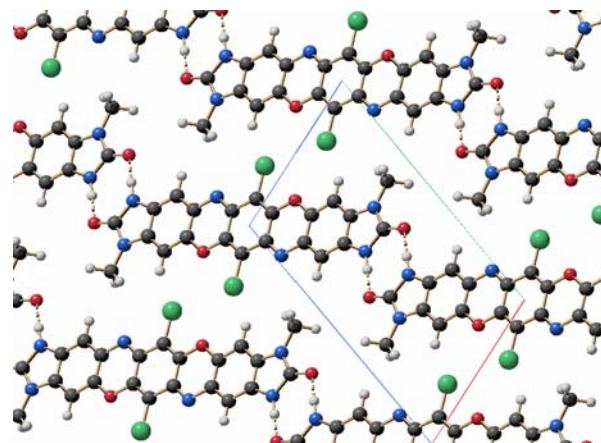
Compound **2** is a violet pigment, which is insoluble in all solvents. Even in good solvents like DMSO or N-methylpyrrolidone, the solubility at 200°C is below the detection limit of about 10 ppb. Nevertheless we found six different polymorphic forms from synthesis, solvent treatment, kneading, and protonation/deprotonation [22]. The powder pattern of the  $\beta$  phase is shown in figure 3 (top). From the line widths the crystallite size can be estimated to be about 20 nm. The quality of the powder diagrams of other polymorphic forms was even worse. The powder data of the  $\beta$  phase could not be indexed (finally the compound turned out to be triclinic; but every set of 12 broad lines can be indexed with a triclinic unit cell in multiple ways). Therefore the crystal structure was solved from scratch without knowledge of lattice parameters and space group.

The molecular geometry was constructed from crystal structure data of similar compounds, and from fragments found in the CSD.

CRYSCA calculations were run in the statistically most common crystal symmetries for organic compounds [23]:  $P2_1/c$ ,  $Z = 4$ ;  $P\bar{1}$ ,  $Z = 2$ ;  $P2_1$ ,  $Z = 2$ ;  $C2/c$ ,  $Z = 8$ ;  $P2_12_12_1$ ,  $Z = 4$  etc. Since the molecule has  $2/m$  symmetry, we also tested the space groups  $P2_1/c$ ,  $Z = 2$ ;  $P\bar{1}$ ,  $Z = 1$ , and  $C2/c$ ,  $Z = 4$ , which are popular for molecules with inversion centres. (The symmetries 2 and  $m$  are less frequently incorporated into the crystal lattice [10,24], and would have been found in the other calculations anyway). All calculated structures were combined, sorted according to energy and searched for higher symmetries. The structure with energy rank 5 showed an X-ray powder diagram similar to the experimental powder diagram, whereas all other powder diagrams did not match. The similarity was surprisingly good considering the fact that we did not make any use of the X-ray powder data during the calculation step.



**Figure 3:** Top: Experimental X-ray powder data of **2**. Bottom: Simulated pattern of the structure calculated by lattice energy minimisation (without any fit to the experimental data).



**Figure 4:** Crystal structure of **2**. View direction  $[100]$ .  $P\bar{1}$ ,  $Z = 1$ ,  $a = 4.335 \text{ \AA}$ ,  $b = 8.419 \text{ \AA}$ ,  $c = 13.906 \text{ \AA}$ ,  $\alpha = 106.95^\circ$ ,  $\beta = 92.91^\circ$ ,  $\gamma = 95.12^\circ$ .

But how could we confirm that this is indeed the correct structure? Rietveld refinements with constraints were done, but they are not a definite proof if the unit cell is triclinic and the powder pattern consists of 12 broad peaks only. Spectroscopic methods were inconclusive. Later we synthesised a mixed crystal (mixture of  $\text{CH}_3$  and  $\text{C}_2\text{H}_5$  derivatives), which turned out to be isostructural to **2**, but had a better powder pattern. So we could carry out a reliable Rietveld refinement and could finally confirm that the structure of **2** is correct.

## Concluding remarks

Lattice energy minimisations are a useful tool to solve crystal structures from powder diffraction data. The method still works, if the powder diagram is of very low quality and/or cannot be indexed.

Limitations of the method include: (1) structures with unknown composition, (2) powder diagrams with less than 10-15 lines (since it becomes difficult to decide if the structure solution is correct). Hydrates and solvates require higher calculation times due to the higher number of parameters. Ionic molecular structures have been rarely calculated, because of force field problems; but several successful examples were given by Van de Streek [25].

## References

- [1] Padgett, C.W., Arman, H.D. & Pennington, W.T., 2007, *Crystal Growth & Design*, 7, 367-372.
- [2] Schmidt, M.U. & Dinnebier, R.E., 1999, *J. Appl. Cryst.*, 32, 178-186.
- [3] Schmidt, M.U., 1999, in: *Crystal Engineering: From Molecules and Crystals to Materials*, edited by D. Braga, F. Grepioni & A.G. Orpen (Dordrecht: Kluwer), pp. 331-348.
- [4] Schmidt, M.U., Ermrich, M. & Dinnebier, R.E., 2005, *Acta Cryst. B*, 61, 37-45.
- [5] Schmidt, M. U., Hofmann, D. W. M., Buchsbaum, C. & Metz, H. J., 2006, *Angew. Chem.*, 118, 1335-1340, *Angew. Chem. Int. Ed.*, 45, 1313-1317.

- [6] Paulus, E.F., Leusen, F.J.J. & Schmidt, M.U., 2007, *CrystEngComm*, 9, 131-143.
- [7] Schmidt, M.U. & Englert, U., 1996, *J. Chem. Soc. Dalton Trans.*, 1996, 2077-2082.
- [8] Китайгородский, А.И., 1955, *Органческая Кристаллохимия* (Москва: Изд-во. Акад. Наук СССР).
- [9] Kitaigorodsky, A.I., 1961, *Tetrahedron*, 14, 230-236.
- [10] Китайгородский, А.И., 1971, *Молекулярные Кристаллы* (Москва: Наука); Kitaigorodsky, A.I., 1973, *Molecular crystals and molecules* (New York: Academic Press); Kitaigorodski, A.I., 1979, *Molekülkristalle* (Berlin: Akademie-Verlag).
- [11] Pertsin, A.J. & Kitaigorodsky, A.I., 1987, *The atom-atom potential method* (Berlin: Springer).
- [12] Mayo, S.L., Olafson, B.D. & Goddard III, W.A., 1990, *J. Phys. Chem.*, 94, 8897-8909.
- [13] Gasteiger, J. & Marsili, M., 1980, *Tetrahedron*, 36, 3219-3222.
- [14] Lommerse, J.P.M., Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Mooij, W.T.M., Price, S.L., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P., & Williams, D.E., 2000, *Acta Cryst. B*, 56, 697-714.
- [15] Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Dzyabchenko, A., Erk, P., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Lommerse, J.P.M., Mooij, W.T.M., Price, S.L., Scheraga, H., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P. & Williams, D.E., 2002, *Acta Cryst. B*, 58, 647-661.
- [16] Day, G.M.; Motherwell, W.D.S.; Ammon, H.L.; Boerrigter, S.X.M.; Della Valle, R.G.; Venuti, E.; Dzyabchenko, A.; Dunitz, J.D.; Schweizer, B.; van Eijck, B.P.; Erk, P.; Facelli, J.C.; Bazterra, V.E.; Ferraro, M.B.; Hofmann, D.W.M.; Leusen, F.J.J.; Liang, C.; Pantelides, C.C.; Karamertzanis, P.G.; Price, S.L.; Lewis, T.C.; Nowell, H.; Torrisi, A.; Scheraga, H.A.; Arnautova, Y.A.; Schmidt, M.U. & Verwer, P., 2005, *Acta Cryst. B*, 61, 511-527.
- [17] Hofmann, D.W.M. & Apostolakis, J., 2003, *J. Mol. Struct. (Theochem)*, 647, 17-39.
- [18] Hofmann, D.W.M. & Kuleshova, L., 2005, *J. Appl. Cryst.*, 38, 861-866.
- [19] Schmidt, M.U., Buchsbaum, C., Schnorr, J.M., Hofmann, D.W.M. & Ermrich, M., 2007, *Z. Krist.* 222, 30-33.
- [20] Verwer, P. & Leusen, F.J.J., 1998, *Rev. Computat. Chem.*, 12, 327-365.
- [21] Schmidt, M.U., Dinnebier, R.E. & Kalkhof, H., submitted.
- [22] Schmidt, M.U., Kempter, P. & Born, R., 2002, *European Patent* EP 1199309.
- [23] Belsky, V.K., Zorkaya, O.N. & Zorky, P.M., 1995, *Acta Cryst. A*, 51, 473-81.
- [24] Pidcock, E., Motherwell W.D.S. & Cole, J.C., 2003, *Acta Cryst. B*, 59, 623-640.
- [25] Van de Streek, J. & Verwer, P., 2004, European Crystallographic Meeting ECM-20, Kraków, Book of Abstracts, p. 177.

---

## EXPANDING FOX: AUTO-INDEXING, GRID COMPUTING, PROFILE FITTING

R. Černý<sup>1\*</sup>, V. Favre-Nicolin<sup>2</sup>, J. Rohlíček<sup>3</sup>,  
M. Hušák<sup>3</sup>, Z. Matěj<sup>4</sup>, R. Kužel<sup>4</sup>

<sup>1</sup>Laboratory of Crystallography, University of Geneva, Switzerland.

<sup>2</sup>University Joseph Fourier (Grenoble I) & CEA/DRFM/SP2M - Grenoble, France.

<sup>3</sup>Institute of Chemical Technology, Prague, Czech Republic.

<sup>4</sup>Faculty of Mathematics and Physics, Charles University, Prague, Czech Republic.

\*Contact author; e-mail: Radovan.Cerny@cryst.unige.ch

### Introduction

While Fox ([objcryst.sourceforge.net/Fox/](http://objcryst.sourceforge.net/Fox/)) is best known for its *ab-initio* structure solution algorithm, it was built to be expandable through the underlying object-oriented ObjCryst++ library. Thanks to this and to the open-source nature of Fox, a number of features *beyond the structure solution step* are now being added to Fox. Apart from the auto-indexing part, all features presented here are implemented by researchers other than the original Fox authors. Also note that features presented here are still in active development—not all are publicly available.

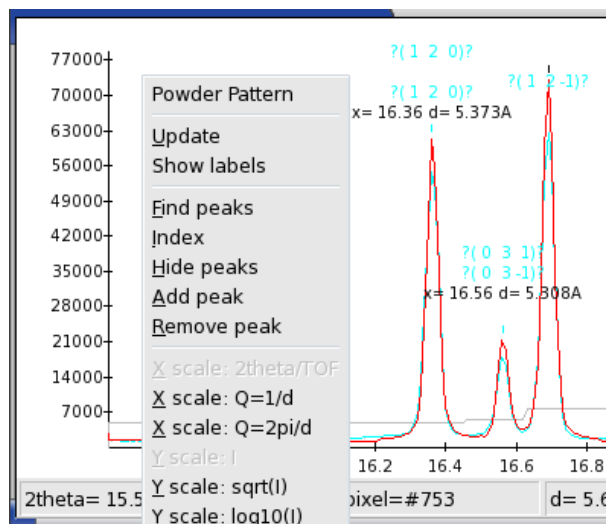
### Auto-indexing with Fox

Among the “most wanted” features for Fox has long been the ability to index powder patterns, without having to use a separate package. This would make Fox a *complete structure solution package*, able to start from a raw powder pattern to a structural model. The only remaining step not handled by Fox would be a full (publication-ready) least squares refinement.

The first part of the auto-indexing consists in a peak-detection algorithm, which uses a search on the second derivative of the powder pattern – see figure 1 for an example of detected peaks in Fox. Users can manually add or remove peaks from the graphical interface.

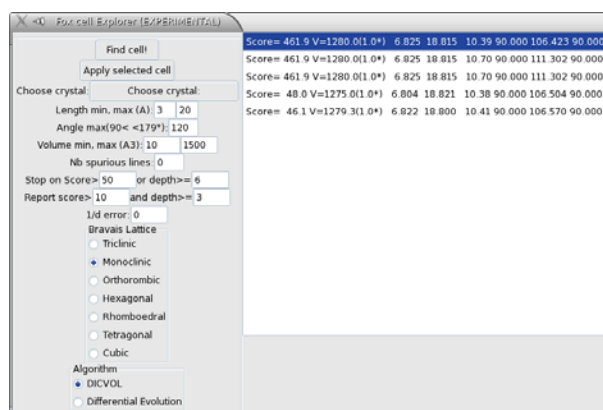
Two indexing algorithms are implemented: the first is DICVOL [1] (a new implementation of the successful *dichotomy-in-volume* algorithm specifically written for Fox), and the second is entirely new, using a differential evolution algorithm. Both algorithms are quite fast, evaluating typically 100 000 unit cells/s for a

default search (3 – 20 Å). The *dicvol* algorithm is much faster for orthorhombic and higher symmetry (usually <1s on a 2GHz processor), but for monoclinic symmetry the *differential evolution* is faster (search can



**Figure 1:** Peak detection – found peaks are labelled in black, and the predicted peak positions are displayed in cyan.

take from few seconds, e.g. ~2s for Cimetidine – cell=10.39, 18.82, 6.82 Å, β=106°, V=1200 Å<sup>3</sup>) to much longer for large unit cells or large angular range !



**Figure 2:** The Fox “Cell Explorer” window – allowing to choose the range for lattice parameters and volume, lattice type, as well as the algorithm – either the *dicvol* or *differential evolution*. Solutions are ranked by their score, with the ratio of their volume to that of the best solution.

This auto-indexing feature is now part of the Fox code, although official versions still have indexing disabled, as this feature is not fully tested for all lattice types. Beta version (announced on the Fox mailing list) users

can try auto-indexing, as well as all who compile Fox themselves.

Still being developed is a full profile-fitting algorithm (see below) to refine the unit cell parameters and allow the extraction of structure factors.

## Fox Grid

For most structures (small molecules) the existing global optimization algorithms and programs are sufficient to find the structure solution very quickly. But there are still a number of problems for which *more computing power is required*: this includes (i) complex structures (>50 independent atoms or >30 internal degrees of freedom), and (ii) problems for which a wide range of structural models (different unit cells, space groups, building blocks, ...) or algorithm parameters [2] has to be tested.

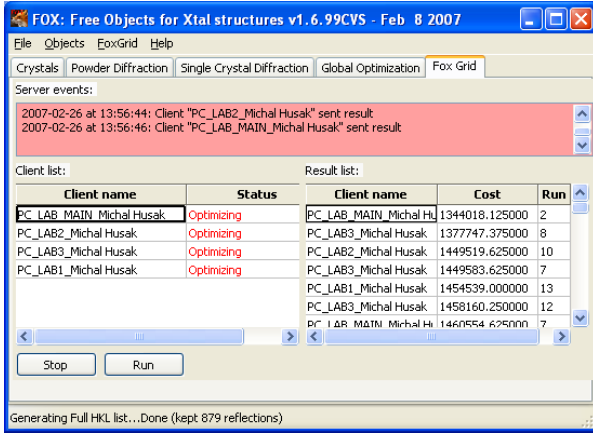
A new Fox project is being worked on at the Institute of Chemical Technology in Prague by Jan Rohlíček and Michal Hušák. It is in early development and therefore not yet publicly available, but represents one evolving frontier of structure determination from powder diffraction, so we include here a 'technology preview' of the 'Fox Grid'.

The code currently tested uses a simple client/server configuration, as outlined in figure 3.



**Figure 3:** Client/server relations in a 'Fox Grid.'

Any PC can be configured either as a server or a client. The data is passed as an XML file between the client and the server. Grid computing can even be used on a single computer with multiple cores or hyper-threading technology, to fully use the available computer power while gathering all the results in a single Fox instance – the one configured as a server. An example of the results listed in the server window is shown in figure 4.



**Figure 4:** Grid computing uses a new tab in the Fox window. In the above figure is presented the server interface, with the list of the client computers currently connected and the results they have already sent.

All the additional code works along the existing Fox code, using cross-platform socket code from the wxWidgets library. Future work will include:

- \* *Benchmarking*: performing tests in a PC lab with 4 PC (3x PC with hyper threading, 1x PC with dual core CPU). The benchmarking is targeted on demonstrating how 'Fox Grid' can speed up the computation in comparison to run on one CPU. We just use for benchmarking a structure solution of Metergoline II phase from ESRF synchrotron data (complicated by preferred orientation as a part of the solution process)

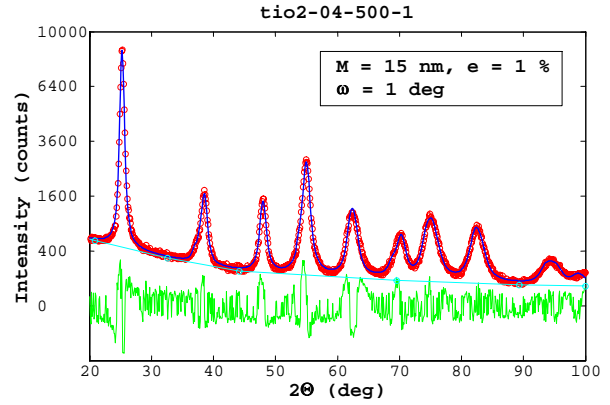
- \* Implementation of a *result analysis module*. This aims to help the user to eliminate multiple identical solutions, by comparing different solutions. A key part of this development is an OpenGL-based simultaneous visualization of multiple solutions.

- \* Implementation of an *intelligent task generation module*. This module can help the user to solve the problem by "brute force", i.e. preparing tasks trying to solve the structure in different space groups and with different lattice parameters (and other input conditions) automatically.

## Profile Fitting

The global optimization algorithm used in Fox and underlying ObjCryst++ library have inspired for extension to microstructure modeling by profile fitting. Zdenek Matěj from Charles University in Prague started a collection of objects for microstructure computation. The idea was to extend Fox for the Rietveld refinement including calculation of diffraction intensities based on both the crystal structure and the microstructure models, and accounting properly influence of the texture or absorption in thin films. A general interface for description of various line broadening models was introduced by convolution of (physically relevant) *profile functions* (calculated in real or reciprocal space) like: stacking faults model [3], grain shape function used for calculation of size-

broadening [4], dislocation broadening [4,5], analytical or measured instrumental functions, physically based micro-deformation (strain) broadening  $\approx 4e/\lambda \sin(\theta)$ . The least-square algorithm hidden in the ObjCryst++ was started, and Fox runs as a Rietveld program.



**Figure 5:** Profile fitting on thin film of  $\text{TiO}_2$  (anatase) – parallel beam geometry, incident angle  $\omega = 1$  deg. Reciprocal space modeling of size  $M$  and straine broadening. Note the squared intensity scale.

## Absorption Correction in Thin Films

To calculate correctly the intensity scattered by a thin film or a multi-layer the absorption of X-rays and scattering volume change must be taken properly into account by the factor

$$\frac{t_p^{(n)}}{\sin(\omega)} \left(1 - \exp(-t^{(n)} / t_p^{(n)})\right)$$

and absorption in all layers above the given layer as

$$\exp\left(-\sum_{i=1}^{n-1} t^{(i)} / t_p^{(i)}\right),$$

where  $t^{(n)}$  is a thickness of  $n$ -th layer and  $t_p^{(n)}$  is the penetration depth in the material of the given layer (with the linear absorption factor  $\mu^{(n)}$ ) calculated as

$$1/t_p^{(n)} = \mu^{(n)} \left(1/\sin(\omega) + 1/\sin(2\theta - \omega)\right).$$

The incidence angle of X-Rays is denoted as  $\omega$ .

## Texture correction

New extended texture correction uses simple "brute force" method based on the integration of a supplied orientation distribution function of all crystallites oriented with given diffracting planes in the direction of scattering vector.



Both, absorption and texture correction, are now available in the Fox from Prague. The appropriate objects have already been included in the ObCryst++. The code is still under development, and is currently working as a command line application. More details can be obtained from [matej@karlov.mff.cuni.cz](mailto:matej@karlov.mff.cuni.cz).

## Perspectives

Such deep intervention, like introduction of profile fitting or grid computing, done by the users of a program dedicated originally for the structure solution illustrates what can be done from a well-documented and available open-source library. Our own activity concentrates on introducing in Fox the analysis of disordered and weakly crystallized materials by the Pair Distribution Function modeling. We strongly encourage all users to contribute actively to the next evolution of this nice program.

## Acknowledgements

Work on the 'Fox Grid' project is supported by a grant from Czech Grant agency (GAČR 203/07/0040). Work on profile fitting is a part of the research program MSM 0021620834 financed by the Ministry of Education of the Czech Republic.

## References

- [1] Boulton, A. & Louër, D., 2004, *J. Appl. Cryst.* 37, 724.
- [2] Markvardsen, A. J., Shankland, K., David, W. I. F. & Didlick, G., 2005, *J. Appl. Cryst.* 38, 107-111.
- [3] Scardi, P. & Leoni, M., 2002, *Acta Cryst.* A58, 190-200.
- [4] Ribárik, G., Ungár, T. & Gubicza, J., 2001, *J. Appl. Cryst.* 34, 669-676.
- [5] Borbély, A., Dragomir-Cernatescu, J., Ribárik, G. & Ungár, T., 2003, *J. Appl. Cryst.* 36, 160-162.

# **DASH-RELATED ALGORITHMIC AND COMPUTING DEVELOPMENTS**

W.I.F. David<sup>1\*</sup>, K. Shankland<sup>1</sup>, A.J.  
Markvardsen<sup>1</sup>, J. van de Streek<sup>2</sup>

<sup>1</sup>ISIS Facility, Rutherford Appleton Laboratory,  
Chilton, Didcot, Oxon OX11 0QX, UK  
<sup>2</sup>CCDC, Cambridge, UK

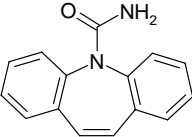
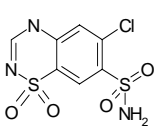
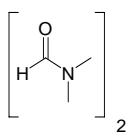
\*Contact author; e-mail: [W.I.F.David@rl.ac.uk](mailto:W.I.F.David@rl.ac.uk)

## Introduction

One of the most surprising aspects of the various global optimisation methods of structure solution from powder diffraction data is that the technique works as well as it does! This is particularly true for molecular crystals where large molecules are being solved routinely by a number of global optimisation programs. In molecular crystal structures, the principal variables that must be determined are the external (position and orientation) degrees of freedom (DOF) and the internal (unknown torsion angles) DOF. It is reasonable to assert that the crystal structure will be solved correctly when the fractional coordinates are determined to a precision of 0.01 and that the various torsions angles are determined to within 3° and thus, for each DOF in the global optimisation, the correct solution occupies 1% of the parameter space. For a rigid body with six external DOF, the correct solution occupies 10<sup>-6</sup> of the global parameter space. For complex molecules with ~25 DOF, the correct solution occupies a vanishing small volume of the available parameter space (i.e. 10<sup>-25</sup>) and thus it seems remarkable that the true solution is ever found. Of course, the global minimum that includes the correct solution is substantially larger than this – a generous assessment would be to assume convergence if the fractional coordinates are within 0.2 and the various angles are within ~70° of the correct solution. For a rigid body and a molecule with only four torsion angles, this means that the probability of finding the global solution is (0.2)<sup>-6</sup> and (0.2)<sup>-10</sup>; i.e. 1 in 15000 and 1 in 10<sup>7</sup> respectively; these are clearly tractable problems. However, for 25 DOF, the probability again diminishes to an improbable (0.2)<sup>-25</sup> ~ 10<sup>-17</sup>. This is clearly a difficult problem and only a few computer programs have been shown to solve structures routinely with this degree of complexity. However, it is precisely this area of complexity that is of particular interest to pharmaceutical scientists and thus it is the focus of much of our development effort. In this letter, we outline four parallel developments that have arisen out of the use of our underpinning structure determination program, *DASH* [1-2] and our desire to solve increasingly challenging problems. These developments are (i) use of non-stochastic global

search algorithms, (ii) use of maximum likelihood, (iii) incorporation of additional structural information and (iv) 'grid-enabling' of structure solution codes for large-scale parallel execution. We aim to include these developments for general use in future versions of *DASH*.

Firstly though, to illustrate some state-of-the-art results from SDPD from pharmaceuticals, we quote from the doctoral work of Fernandes, who reported the structure determination (using *DASH*) and refinement (using *TOPAS*) of numerous compounds of pharmaceutical interest from both laboratory and synchrotron X-ray powder diffraction alone, employing variable-counting-time data collection for the most complex examples. Particularly impressive are the solution and refinement of the gamma form of carbamazepine and a solvated form of chlorothiazide [3] (Table 1).

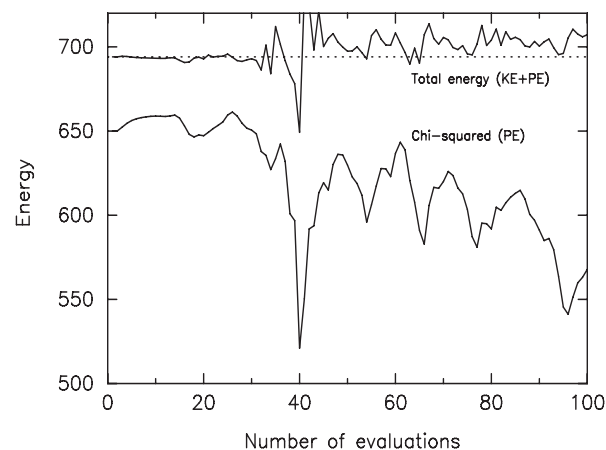
Molecular formula		
		
SG	<i>P</i> -1	<i>P</i> 2 <sub>1</sub> / <i>c</i>
<i>Z'</i>	4	2
<i>N</i> <sub>frag</sub>	4	6
<i>N</i> <sub>tor</sub>	4	6
<i>N</i> <sub>dof</sub>	28	42
<i>N</i> <sub>atoms</sub>	120	94
<i>N</i> <sub>non-H</sub>	72	54

**Table 1:** Two structures solved from XRPD data. *SG*=space group; *Z'*=no. of formula units in asymmetric unit; *N*<sub>frag</sub>=no. of fragments in asymmetric unit; *N*<sub>tor</sub>=no. of flexible torsion angles in asymmetric unit; *N*<sub>dof</sub>=total number of DOF for optimisation; *N*<sub>atoms</sub>=total no. of atoms in asymmetric unit; *N*<sub>non-H</sub>=total no. of non-H atoms in asymmetric unit.

## Non-stochastic global optimisation - the Hybrid Monte-Carlo method

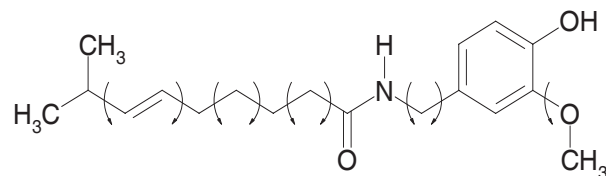
Whilst simulated annealing (SA) and related stochastic techniques have been shown to be effective global optimisation methods, many other algorithms from different research areas remain to be evaluated and it is likely that a number of these will be more efficient and successful than the techniques currently in use. In this section, we discuss the hybrid Monte Carlo (HMC) algorithm which combines the key components of Monte Carlo (MC) and molecular dynamics (MD) approaches into a single efficient algorithm. An extensive discussion of HMC in the context of structure determination from powders may be found in the paper of Johnston et al. [4]. In essence, HMC may be considered in terms of a particle that follows a trajectory determined by Hamilton's equations of motion in a hyperspace defined by the set of structural variables. The total energy of the particle at any point is equal to the sum of the kinetic energy and the

potential energy which is given by the goodness of fit target function. Whilst in theory the total energy is conserved, the use of finite time step sizes in the numerical evaluation of the equations of motion means that this is not the case. To control this effect, a Metropolis acceptance criterion is used to determine whether to accept or reject the configuration at the end of a given MD trajectory. The trajectory either continues from the end point if it is accepted or returns to the previous start point if it is rejected (Fig. 1).



**Fig. 1:** The potential energy (correlated integrated intensities  $\chi^2$ ) and total energy (kinetic energy plus potential energy) evaluated over a single MD trajectory during the crystal structure solution of capsaicin. The initial total energy is shown as a dotted line in order to highlight the total energy fluctuations arising from the finite MD step size.

The effectiveness of HMC has been convincingly demonstrated with the structure determination of capsaicin (Fig. 2) which with a total of 15 DOF represents a considerable challenge for SDPD.



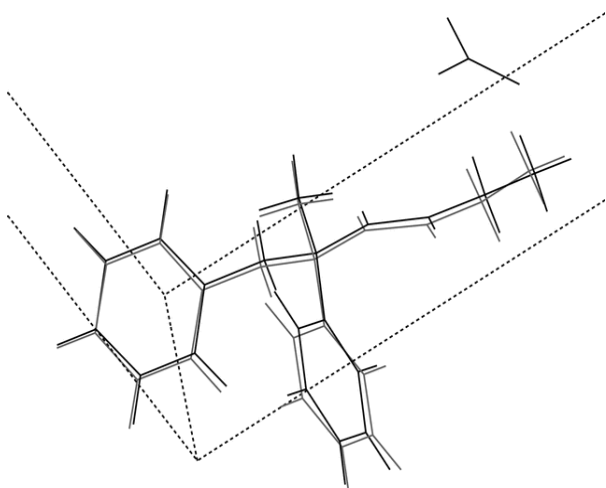
**Fig. 2:** The molecular structure of capsaicin, with flexible torsion angles highlighted by arrows.

When compared with the default implementation of simulated annealing in *DASH*, HMC is a factor of two more successful in locating the global minimum over a series of twenty repeat runs. Significantly, the HMC algorithm required considerably fewer  $\chi^2$  evaluations than simulated annealing to achieve this level of success and remarkably, given the discussion in the introduction, the quickest solution required less than 20,000 evaluations to locate the radius of convergence corresponding to  $\sim 10^{-11}$  of the total parameter space.



## Dealing with incomplete structural models - the maximum likelihood approach

The principal reason for the success of global optimisation methods in the solution of crystal structures from powder diffraction data has been the incorporation of the molecular geometry into the solution process; only a relatively small number of parameters have to be determined and the process is thus substantially more tractable than the determination of all atomic positions independently. However, this strength is also the principal limitation of the technique – the full molecular structure must be incorporated if the global least-squares minimum is to be reached. This limitation may, however, be relaxed if more generalised maximum likelihood methods are used. Consider that the majority (but not all) of the structural contents is to be determined in the optimisation process; for example, locating a main molecule but ignoring associated water molecules in a hydrate. In such cases, maximum likelihood optimisation allows the majority of the structure to be correctly located. Use of this approach has been illustrated with the examples of the nitrate and acetate salts of the anticonvulsant agent remacemide[5]. If the nitrate and acetate ions are excluded from a standard least-squares global optimisation then the structures cannot be solved - the best solutions show parts of the remacemide molecule located at the positions of the acetate and nitrate ions, in an attempt to account for their scattering contribution. In contrast, if the nitrate and acetate ions are not *explicitly* considered in a maximum likelihood optimisation, the remacemide ion is quickly and correctly located for both structures with a very high success rate (see Fig. 3). It is then a trivial matter to subsequently to fix the remacemide ion within the unit cell and then locate and orient the nitrate and acetate molecules by global optimisation.



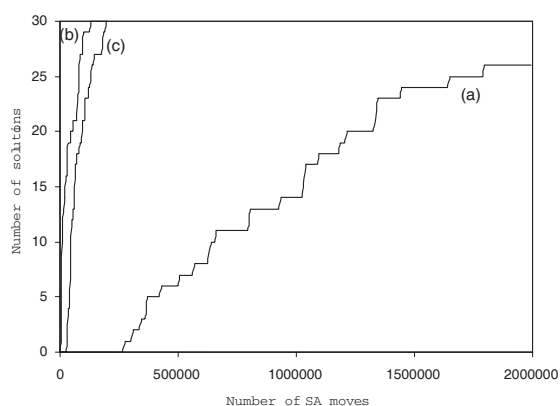
**Fig. 3:** Remacemide nitrate: the asymmetric unit of the best remacemide-only substructure (normal lines) obtained by maximum likelihood, superimposed upon the asymmetric unit of the refined remacemide nitrate crystal structure (bold lines).

These results give a strong indication that the maximum likelihood approach has the ability to improve the success rate of global-optimization-based crystal structure determination methods in circumstances where the structural model being optimized is not a complete description of the crystal structure under study. These findings are in broad agreement with other comparisons of least-squares and maximum-likelihood methods in macromolecular crystallography. This approach relaxes the constraint that the correct molecular contents are included from the outset of the global optimisation process – for materials such as hydrates or solvates, this is an important consideration.

## Incorporating additional structural information – use of advanced solid-state NMR techniques

Constraints form a fundamental part of most global optimisation approaches, with bond lengths, bond angles and fixed dihedral angles in the material under investigation typically being held at known, standard values during the optimisation process. The structural variables are thus restricted to the external molecular DOF plus the remaining internal torsion angles whose values cannot be assigned *a priori*. Use of the Cambridge Structural Database can help to provide likely bounds for torsion angles but direct use other techniques to determine torsion angle values *a priori* can be even more effective. For example, if the complete molecular conformation can be determined in advance of the diffraction experiment, global optimisation is reduced to one of determining the position and orientation of a rigid molecule. Middleton *et al.*[6] have outlined such a procedure in which a set of inter-atomic distances is measured by rotational-echo double resonance (REDOR) SS-NMR. The molecular conformation is then derived from a restrained molecular dynamics (MD) optimisation in which the use of high harmonic force constants ensures that all conformations in the simulation have interatomic distances that satisfy the input distances. The best conformation is then optimized against the X-ray powder diffraction data by global optimisation. By way of example, the anti-ulcer drug cimetidine, in polymorphic form A, was solved from XRPD data using DASH with an MD-optimised model derived from four SS-NMR-determined C-N<sup>15</sup> distances. Each torsion angle in the MD-optimised starting model was allowed to vary +/- 20° from its input value; the results of repeated DASH runs are presented in Fig. 4. Distributions (a) and (b) represent two extremes defining the limits of the problem: (a) optimising a model with no torsional restraints, DASH returns an 87% success within a 2×10<sup>6</sup> move limit; (b) optimising only the external DOF of the correct single-crystal conformation is much faster (≤ 129,000 moves) and is 100% successful in returning the correct crystal structure. Distribution (c) shows that taking the

NMR/MD conformation and imposing torsional restraints ( $\pm 20^\circ$ ) consistent with the precision of the NMR distance measurements delivers a speed and reliability approaching that of the rigid-body optimisation (*b*). This type of approach holds great promise for complex structures where the removal of several internal DOF from the structure determination problem could significantly increase the chances of successfully locating the global minimum. However, routine application will probably only be possible when the SS-NMR methodology develops to a stage where isotopic labelling is no longer a pre-requisite and when the specialised SS-NMR instrumentation required is more commonly available.



**Fig. 4:** The effect of restraining torsional DOF on the crystal structure determination process for cimetidine. To facilitate assessment of the impact of such restraints in the stochastic optimisation, each distribution is compiled from 30 individual *DASH* runs consisting of a full search of the six external DOF, plus one of the options outlined in the main text.

## Distributed computing - parallel execution of *DASH* and *HMC*

Parallelisation of a single simulated annealing run is not a straightforward task. The sequential nature of the algorithm means that any parallelisation has to be performed at either (a) a very fine-grained level; for example, the evaluation of the model fit to the data might be amenable to distribution across closely-coupled processors, or (b) a very coarse-grained level, where multiple independent SA runs are executed in parallel on distinct processors. Given the issue raised in the introduction of this article (i.e. low success rates for individual SA runs in locating the global minimum for very complex systems) the latter option is very attractive, and has the benefit that it is straightforward to achieve. In its simplest incarnation, parallelisation of multiple SA runs can be achieved by walking between computers which have the necessary structure solution software installed and starting a single SA run on each machine. If we assume that that the jobs are setup and ready to execute, and that the user can run quickly between machines to start the runs, then in the absence

of any competing processes on the computers, the speed-up achieved will be almost exactly equal to the number of computers used. Clearly though, this is not really a practical approach to solving structure. Of far greater interest is the ability to distribute SA runs from one's own desktop onto other compute resources to which one has access. Our own work in this area has focussed on the use of the United Devices GridMP system, which allows the spare CPU cycles of *existing* compute resources (including desktop machines, servers, MS Windows and Linux systems) to be harnessed transparently for useful computations. Whilst GridMP is a commercial program, comparable free alternatives such as the widely-used CONDOR system can also be used. In a sizeable department (such as the ISIS Facility) with many hundreds of computers (many of them dual core) connected to the system, the resultant speed-ups can be substantial, and we routinely obtain *DASH* job execution times that are reduced by two orders of magnitude. The modifications required to *DASH* in order to achieve this type of distributed execution are relatively small, the main change involving the introduction of a command-line invocation mode (*DASH* is, of course, primarily a GUI-driven program) and an associated control-file format. The need for a command-driven mode is obvious - it would be disconcerting (to say the least) for users on remote machines to have a GUI version of *DASH* pop up on their screen every time a *DASH* job was submitted to the GridMP servers for execution. To the *DASH* user however, usage of *DASH* remains unchanged until the final stages where one would normally press the 'Solve' button to begin SA annealing execution. At this point, if distributed execution is desired, the user selects the 'Create Batch File' option, at which point *DASH* automatically generates all the files necessary for submission to the GridMP system. A single 'gdash' command with a single file argument is then all that is necessary to start the job on the grid. Job progress can be monitored using the same 'gdash' command and results can be retrieved using the same command at any stage; the resultant .dash file can be read directly into *DASH* for analysis. Using a similar approach, we have also parallelised our HMC code [7] and utilised the available computing power to better characterise the behaviour of the HMC algorithm.

## References

- [1] W. I. F., David, K., Shankland, and N., Shankland: Routine determination of molecular crystal structures from powder diffraction data. *J. Chem. Soc. Chem. Commun.* (1998) 931-932.
- [2] W. I. F., David, K., Shankland, J., van de Streek, E., Pidcock, and W.D.S., Motherwell: *DASH*: a program for Crystal Structure Determination from Powder Diffraction Data. *J. Appl. Cryst.* 39 (2006) 910-915.
- [3] P., Fernandes, K., Shankland, A.J., Florence, N., Shankland, and A., Johnston: Solving Molecular Crystal Structures from X-ray Powder Diffraction

Data: The Challenges Posed by  $\gamma$ -Carbamazepine and Chlorothiazide N,N-Dimethylformamide (1/2) Solvate. *J. Pharm. Sci.* 96 (2007) 1192-1202.

[4] J.C., Johnson, A.J., Markvardsen, W.I.F., David and K., Shankland: A hybrid Monte Carlo method for crystal structure determination from powder diffraction data. *Acta. Cryst.* (2002) **A58** 441-447.

[5] A.J., Markvardsen, W.I.F., David and K., Shankland: A maximum-likelihood method for global-optimization-based structure determination from powder diffraction data. *Acta. Cryst.* (2002) **A58** 316-326.

[6] D.A., Middleton, X., Peng, D., Saunders, W.I.F., David, K., Shankland and A.J., Markvardsen: Conformational analysis by solid-state NMR and its application to restrained structure determination from powder diffraction data. *Chem. Commun.* (2002) 1976-1977.

[7] A.J., Markvardsen, K., Shankland, W.I.F., David and G., Didlick: Characterization of a hybrid Monte Carlo search algorithm for structure determination. *J. Appl. Cryst.* 38 (2005) 107-111.

## Computer Corner

Updates on Available Crystallographic and Powder Diffraction Software

Suggestions, corrections, comments and articles on new or updated software are appreciated; especially if you know of new program features, program updates and announcements that should be mentioned here.

Lachlan M. D. Cranswick  
Canadian Neutron Beam Centre (CNBC),  
National Research Council (NRC),  
Building 459, Station 18, Chalk River Laboratories,  
Chalk River, Ontario, Canada, K0J 1J0  
Tel: (613) 584-8811 ext 3719; Fax: (613) 584-4040  
E-mail: [Lachlan.Cranswick@nrc.gc.ca](mailto:Lachlan.Cranswick@nrc.gc.ca)  
WWW: <http://neutron.nrc.gc.ca/>

## Rietveld Software Updates (as of 25<sup>th</sup> February 2007):

**Hugo Rietveld website:**

<http://home.wxs.nl/~rietv025/>

**Armel Le Bail website:**

<http://sdpd.univ-lemans.fr/>

**BGMN** (2<sup>nd</sup> Feb 2007)

<http://www.bgmn.de/>

**BRASS** (20<sup>th</sup> Dec 2006)

<http://www.brass.uni-bremen.de/>

**DBWS** (22<sup>nd</sup> February 2000)

[http://www.physics.gatech.edu/downloads/young/download\\_dbws.html](http://www.physics.gatech.edu/downloads/young/download_dbws.html)

**DDM** (25<sup>th</sup> May 2006)

[http://icct.krasn.ru/eng/content/persons/Sol\\_LA/ddm.html](http://icct.krasn.ru/eng/content/persons/Sol_LA/ddm.html)

**Debvin** (ftp download site not connecting)

<http://users.uniud.it/bruckner/debvin.html>

**GSAS** (22<sup>nd</sup> Feb 2007)

<http://www.ccp14.ac.uk/ccp/ccp14/ftp-mirror/gsas/public/gsas/>

**EXPGUI** (30<sup>th</sup> April 2006)

<http://www.nenr.nist.gov/programs/crystallography/>

**Jana** (23<sup>rd</sup> December 2005)

<http://www-xray.fzu.cz/jana/Jana2000/jana.html>

**LHPM-Rietica** (27<sup>th</sup> November 2001)

<http://www.rietica.org/>

**MAUD** for Java (GPL'd) (20<sup>th</sup> Feb 2007)

<http://www.ing.uniutn.it/~maud/>

**MXD** (7<sup>th</sup> Nov 2006)

[http://www-Cristallo.grenoble.cnrs.fr/Prog\\_Cristallo/](http://www-Cristallo.grenoble.cnrs.fr/Prog_Cristallo/)

**PowderCell** (8<sup>th</sup> March 2000)

[ftp://ftp.bam.de/Powder\\_Cell/](ftp://ftp.bam.de/Powder_Cell/)

**Prodd** (19<sup>th</sup> August 2003)

<http://www.ccp14.ac.uk/ccp/web-mirrors/prodd/~jpw22/>

**Profil** (24<sup>th</sup> May 2001)

<http://img.chem.ucl.ac.uk/www/cockcroft/profil.htm>

**Rietan 2000** (GPL'd) (15<sup>th</sup> March 2006)

[http://homepage.mac.com/fujiioizumi/rietan/angle\\_dispersive/angle\\_dispersive.html](http://homepage.mac.com/fujiioizumi/rietan/angle_dispersive/angle_dispersive.html)

**Fullprof Suite (including Winplotr)** (February 2007)

<http://www.ill.fr/dif/Soft/fp/>

**XND** (13<sup>th</sup> April 2006)

<ftp://ftp.grenoble.cnrs.fr/xnd/>

**XRS-82/DLS76**

<http://www.crystal.mat.ethz.ch/Software/index>

Most of the above Rietveld programs are also available via the CCP14 website:

(<http://www.ccp14.ac.uk/mirror/>).

## Summary lists of some software available via the CCP14 website:

“What do you want to do?” (lists of software by single crystal and powder methods)

[http://www.ccp14.ac.uk/mirror/want\\_to\\_do.html](http://www.ccp14.ac.uk/mirror/want_to_do.html)

Anharmonic Thermal Refinement Software

<http://www.ccp14.ac.uk/solution/anharmonic/>

Data Conversion for Powder Diffraction

<http://www.ccp14.ac.uk/solution/powderdataconv/>

Image Plate Software

<http://www.ccp14.ac.uk/solution/image-plate/>

Incommensurate Structure Software

<http://www.ccp14.ac.uk/solution/incomm.htm>

Indexing Software for Powders

<http://www.ccp14.ac.uk/solution/indexing/>

LeBail Method for Intensity Extraction

<http://www.ccp14.ac.uk/solution/lebail/>

Pawley Method for Intensity Extraction

<http://www.ccp14.ac.uk/solution/pawley/>

PDF, High Q Powder diffraction Analysis Software

[http://www.ccp14.ac.uk/solution/high\\_q\\_pdf/](http://www.ccp14.ac.uk/solution/high_q_pdf/)

Peak Find/Profiling Software for Powder Diffraction

<http://www.ccp14.ac.uk/solution/peakprofiling/>

Pole Figure and Texture Analysis Software

[http://www.ccp14.ac.uk/solution/pole\\_figure/](http://www.ccp14.ac.uk/solution/pole_figure/)

Powder Diffraction Data Visualisation

[http://www.ccp14.ac.uk/solution/powder\\_data\\_visual](http://www.ccp14.ac.uk/solution/powder_data_visual)

Rietveld Software

[http://www.ccp14.ac.uk/solution/rietveld\\_software](http://www.ccp14.ac.uk/solution/rietveld_software)

Search-Match Phase Identification Software

<http://www.ccp14.ac.uk/solution/search-match.htm>

Single Crystal Suites linking to multiple programs relevant to Chemical Crystallography

<http://www.ccp14.ac.uk/solution/xtalsuites/>

Spacegroup and Symmetry operator determination software and source code

[http://www.ccp14.ac.uk/recomm/sym\\_operators\\_to\\_spacegroups.html](http://www.ccp14.ac.uk/recomm/sym_operators_to_spacegroups.html)

[http://www.ccp14.ac.uk/recomm/spacegroups\\_to\\_hsym\\_operators.html](http://www.ccp14.ac.uk/recomm/spacegroups_to_hsym_operators.html)

Spacegroup and Structure Transformation Software

<http://www.ccp14.ac.uk/solution/transform/>

Structure Conversion and Transformation

<http://www.ccp14.ac.uk/solution/structconv/>

Structure Drawing and Visualisation

<http://www.ccp14.ac.uk/solution/structuredrawing/>

Unit Cell Refinement of Powder Diffraction Data

<http://www.ccp14.ac.uk/solution/unitcellrefine/>

# BRASS 2.0, THE BREMEN RIETVELD ANALYSIS AND STRUCTURE SUITE, VER. 2

J. Birkenstock\*, R.X. Fischer, T. Messner

University of Bremen, Dept. of Geosciences,  
Klagenfurter Str. 2, D-28359 Bremen

\*Contact author; e-mail: jbirken@uni-bremen.de

## Introduction

About 30 years after its advent the Rietveld method [1,2] has become a very popular and widespread tool for the quantitative analysis of powder diffraction data. Originally restricted to crystal structure refinement its areas of application have extended to routine quantitative phase, crystallite size and strain analyses. The former necessity for having well defined crystal structure models for all phases at hand has been removed at first by [3] making use of a list of indexed structure factors (better: “reduced intensities”<sup>1</sup>) – an option that has been added to several Rietveld programs in the meantime, e.g. [4, 5, 6]. In BRASS ver. 1 a new method was implemented which allows for re-scaling refined intensities to absolute values that can be used for the quantification of structurally unknown crystalline phases [7].

BRASS 2 is the largely revised successor of BRASS 1 and is freely available at [www.brass.uni-bremen.de](http://www.brass.uni-bremen.de). Due to major changes with respect to the profile description in the Rietveld module BRASS 2 is not downwards compatible with BRASS 1, but BRASS 1 Rietveld job files of type INP may easily be imported to BRASS 2 using the filter “input file” with the dialog “load Rietveld job” on page “Rietveld job”.

## Major features of BRASS 2.0

BRASS 2 is a package of program modules running on Windows (R) 2000 and XP platforms: for the single and combined display of diffraction data, for Rietveld refinements, for structure completion via fourier and grid search methods and for high-quality crystal

<sup>1</sup> Usually the term reduced intensity  $I_R(hkl)$  is more correct, depending on the source of these values: if not calculated from a structure model the values are in fact products  $S_r \cdot PO_{hkl} \cdot m'_{hkl} \cdot |F_{hkl}|^2$  ( $S_r$  = phase's rescale factor,  $PO$  = texture factor,  $m'$  = multiplicity correction factor,  $\neq 1$  if the assigned space group is ambiguous or 1 if not,  $F$  = structure factor). For  $S_r = 1$   $I_R$  should be called  $I_{AR}$  = absolute reduced intensities, otherwise  $I_{RR}$  = relative reduced intensities. Note:  $I_{AR}$  values may be employed for quantitative phase analyses.

structure display together with easy tools for its manipulation and structure-based crystal chemical calculations. All dialogs are directly accessed by the BRASS job explorer. The Rietveld kernel is a largely extended derivative of the well known Young&Wiles program [8], the module StruPlo, originating from [9], is rewritten with Borland Delphi (R) and has been largely extended, too.

For the Rietveld module a free training example may be downloaded at [www.brass.uni-bremen.de](http://www.brass.uni-bremen.de). A strict separation of instrument and sample contributions to peak profiles is based on a modified variant of the Pseudo-Voigt function, providing specific parameters for each contribution and using the deconvolution formula given by [10]. Instrument parameters are assessed by a refinement on data of a standard sample and may be stored as a set for easy re-use with samples to be analysed. Up to four wavelengths may be used to account for the spurious peaks from  $K_{\beta}$ - and  $W-L_{\alpha}$ -radiation. Four different types of background descriptions may be used, single or combined. The most flexible one is the interpolation by cubic splines with up to 60 refinable setpoints.

Crystallite sizes and microstrains may be addressed isotropically or anisotropically, for microstrains there are two anisotropic models [11], one of which is the one after Stephens [12]. Up to 15 phases may be used in the refinement. The intensities of a phase may be calculated from a crystal structure model, a list of reduced intensities or by a Le-Bail fit [13].

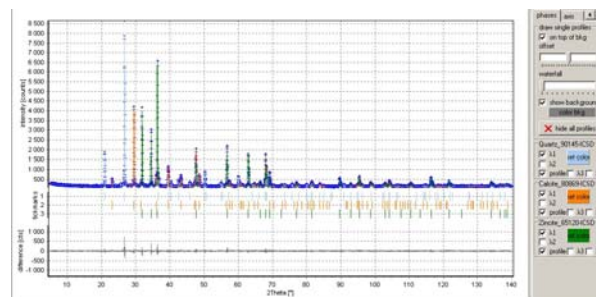
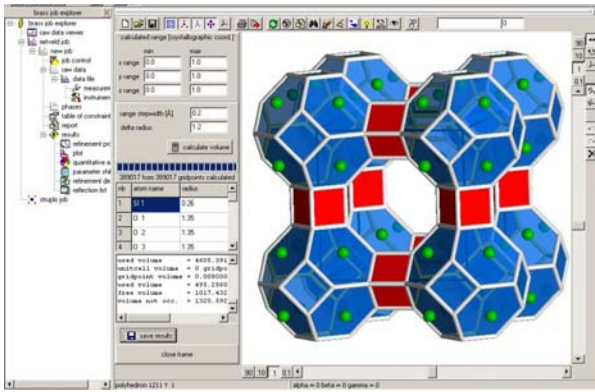


Figure 1: Rietveld plot with phases' profiles superimposed.

Using the latter two in reverse order allows for the quantification of structurally unknown phases (see below). Quantitative analyses may be corrected for microabsorption using Brindley's approximation [14] in a unique post-refinement procedure, detailed in the training example. Individual profiles of all phases may be superimposed to the traditional display of raw data, calculation and difference (figure 1).





**Figure 2:** Polyhedra display and void volume with StruPlo.

BRASS' StruPlo allows for high quality crystal structure display. Each single object, such as atoms, bonds and polyhedra, or groups of them may be deleted and restored at any time. Crystal chemical calculations include distances-and-angles and, after [15], void-volume calculations (figure 2).

### Details on phase quantification without crystal structure model

To demonstrate how a structurally unknown phase may be quantified it is actually convenient to consider a mixture of phases all of which are structurally well known and replace one of the phases by the description without crystal structure model.

For a test case the Rietveld job example on "mixture2" with quartz, zincite and calcite may be used (free download at [www.brass.uni-bremen.de](http://www.brass.uni-bremen.de)). The given refinement was carried out using crystal structure models. After the application of the micro-absorption correction, using an overall Brindley particle radius of about 12  $\mu\text{m}$ , the weighed-in fractions of 33.3% each were perfectly reproduced. Details are given in the training manual, included in the download package.

In order to quantify the three phases using a "structure-free" approach for one of the phases, select "Le Bail extraction" instead of "atomlist" on page "structure" e.g. for calcite (Cc). To simulate that you have no preliminary idea about the phase's fraction its scale factor  $S(\text{Cc})$  should be set to a default value such as  $10\text{E-}5$ . The parameter PO for texture (page "profile") should be set to 1. The parameters of the phase should not be refined at first. On completion of the refinement the profile description will look quite as good as the one using the structure model since, here, the profile parameters were already well determined in advance.

At this point the quantification will usually be wrong due to the fact that the pre-selected scaling factor of the phase is usually wrong. The ratio of fractions of quartz and zincite will be right, though, when the Brindley correction has been applied. In fact, you may assume that you do not know the mean Brindley particle radius and find it by adapting it to the value where both fractions are calculated to equal values. E.g., for  $S(\text{Cc})$

=  $10\text{E-}5$  and a Brindley radius of 12.2  $\mu\text{m}$  the fractions are:  $w_{\text{Quartz}} = 49.2\%$ ,  $w_{\text{Zincite}} = 49.2\%$  and  $w_{\text{Calcite}} = 1.6\%$ . Now all previously fixed parameters of calcite should be refined again.

The above steps are well defined preconditions for the correct rescaling of the reduced intensities of calcite: Switching to "phases", selecting calcite and using the button "move to SFL" fixes the reduced intensities to the current values which are now displayed as "list of reduced intensities". The known fraction of calcite should be entered to "w (known)", which is 33.33%. With  $S(\text{Cc}) = 10\text{E-}5$  BRASS2 calculates the rescale factor to  $\sim 0.0322$ . A look at the "quantitative analysis" with Brindley correction switched on now displays the correct fractions, weighed in as 33.3% each.

The reduced intensities are on absolute scale now and can be used to quantify calcite in other samples, too. For this purpose one may proceed as follows: at first the above Rietveld job must be saved in its latest state. From this file the optimized calcite model may be imported together with its list of absolute reduced intensities into other Rietveld jobs: use "add phase" on page "phases" and select the above file. From the following dialog the phases to be imported may be selected. Now the phase may be used in the current refinement.

The limitations for the application of the resulting absolute reduced intensities shall be mentioned here:

- Before re-scaling the fit of all phases must be very good: low  $R_{\text{Bragg}}$ , small differences in peak areas.
- Microabsorption due to large particle sizes and/ or to large contrasts in the phases' absorption factors must be Brindley corrected prior to rescaling.
- The derived list of absolute reduced intensities is only valid for the same phase with the same composition and at the same ambient conditions. Lattice constants and an overall temperature factor may be refined, though, to correct for minor differences.
- The range of diffraction angles for the derived list of absolute reduced intensities must cover the angular range of further analysed data since the list may not be expanded afterwards.

If these limitations are taken into account the use of reduced intensities has some merits:

- Since the values are reduced they may be applied for different diffractometer configurations.
- Structurally unknown phases may be quantified.
- Crystallite size and microstrain may be established.
- Using a list of reduced intensities saves time since structure factor calculations are omitted.

### References

- [1] Rietveld, H.M., 1967, *Acta Cryst.*, 22, 151-152.
- [2] Rietveld, H.M., 1969, *J. Appl. Cryst.*, 2, 65-71.
- [3] Taylor, J.C. & Matulis, C.E., 1991, *J. Appl. Cryst.*, 24, 14-17.
- [4] Rodriguez-Carvajal, J., 2005, FullProf, ver. 3.2.
- [5] Coelho, A.A., 2003, TOPAS, ver. 3.1, Bruker AXS, Karlsruhe, Germany

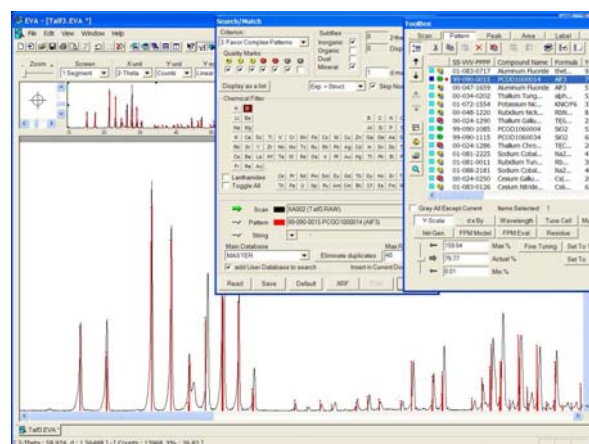


- [6] Birkenstock, J., Fischer, R.X. & Messner, T., 2005, BRASS, ver. 2.0, www.brass.uni-bremen.de.
- [7] Birkenstock, J., Fischer, R.X. & Messner, T., 2005, Abstr. 83. DMG meeting.
- [8] Wiles, D.B. & Young, R.A., 1981, *J. Appl. Cryst.*, 14, 149-151.
- [9] Fischer, R.X., 1985, *J. Appl. Cryst.*, 18, 258-262.
- [10] Thompson, P., Cox, D.E. & Hastings, J.B., 1987, *J. Appl. Cryst.*, 20, 79 - 83.
- [11] Birkenstock, J., Fischer, R.X. & Messner, T., 2006, Abstr. 14. DGK meeting.
- [12] Stephens, P.W., 1999, *J. Appl. Cryst.*, 32, 281 - 289.
- [13] Le Bail, A., Duroy, H. & Fourquet, J.L., 1988, *Mat. Res. Bull.*, 23, 447-452.
- [14] Brindley, G.W., 1945, *Philos. Mag.*, 36, 347-369.
- [15] Küppers, H., Liebau, F. & Spek, A.L., 2006, *J. Appl. Cryst.*, 39, 338-346.

recent (1992-97) SDPD (Structure Determination by Powder Diffraction) which would have demanded less effort if the PPDF had been available earlier : the structures would have been solved directly at the identification stage.

### Example 1

The actual and virtual structures have the same chemical formula, PAD = 0.52 (percentage of absolute difference on cell parameters, averaged) :  $\tau$ -AlF<sub>3</sub> [8,9], tetragonal, a = 10.184 Å, c = 7.174 Å. Predicted : 10.216 Å, 7.241 Å. A global search (no chemical restraint) is resulting in the actual compound (PDF-2) in first position and the virtual one (PPDF-1) in 2<sup>nd</sup> (green mark in the toolbox).



## SEARCH-MATCHING INORGANICS THROUGH THE PPDF-1 (PREDICTED POWDER DIFFRACTION FILE)

Armel Le Bail

Université du Maine, Laboratoire des Oxydes et Fluorures, CNRS UMR 6010, Avenue O. Messiaen, 72085 Le Mans Cedex 9, France.

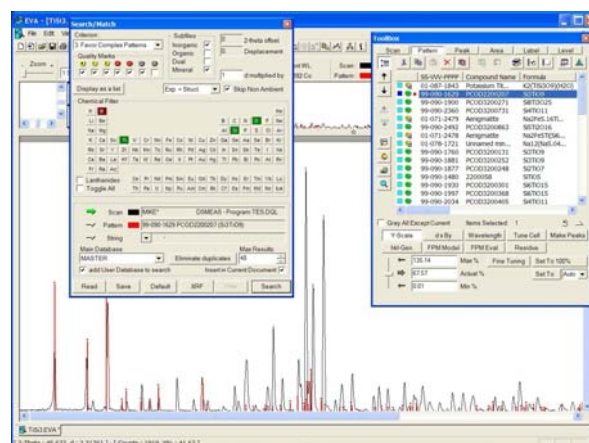
Contact author; e-mail: armel.le\_bail@univ-lemans.fr

### Introduction

Massive crystal structure predictions of inorganic compounds were recently made for zeolites [1] and more generally for 3D N-connected frameworks [2]. The next step was obviously [3] to calculate predicted powder patterns in order to allow for direct identification by search-matching with *d-I* data pairs. The PPDF-1 (Predicted Powder Diffraction File, version 1) [4], consists in patterns calculated from atomic coordinates corresponding to the virtual crystal structures gathered in the recent update of the PCOD (Predicted Crystallography Open Database) [5]. It contains more than 60.000 silicates, phosphates, sulfates of Al, Ti, V, Ga, Nb, Zr, or zeolites, fluorides, etc. Its ability for automated search-match identification, after coupling with the EVA software (Bruker), using the most sophisticated [6] and efficient [7] approach, is shortly demonstrated here. The PPDF text file contains chemical formula, cell parameters, probable space group, *d-I* pairs, Miller indices and *I/I*(cor) calculated directly from the PCOD CIFs by the CIF2POW software. That text was compiled in binary files for compatibility with EVA. The test cases below were selected because they correspond to relatively

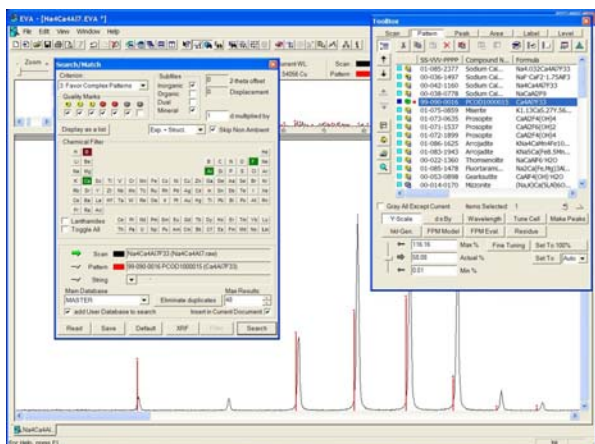
### Example 2

Model showing incomplete chemistry, PAD = 0.63. Actual compound: K<sub>2</sub>TiSi<sub>3</sub>O<sub>9</sub>•H<sub>2</sub>O [10], orthorhombic, a = 7.136 Å, b = 9.908 Å, c = 12.941 Å. Predicted framework: TiSi<sub>3</sub>O<sub>9</sub>, a = 7.22 Å, b = 9.97 Å, c = 12.93 Å. Without chemical restraint, the correct PDF-2 entry is coming at the head of the list, but no virtual model. By using the chemical restraint (Ti + Si + O), the correct PPDF-1 entry comes in second position in spite of large intensity disagreements with the experimental powder pattern (K and H<sub>2</sub>O are lacking in the PCOD model):



### Example 3

Model showing incomplete chemistry, PAD = 0.88.  
Predicted framework :  $\text{Ca}_4\text{Al}_7\text{F}_{33}$ , cubic,  $a = 10.876 \text{ \AA}$ .  
Actual compound :  $\text{Na}_4\text{Ca}_4\text{Al}_7\text{F}_{33}$  [11],  $a = 10.781 \text{ \AA}$ .  
By a search with chemical restraints (Ca + Al + F) the virtual model comes in fifth position, after 4 PDF-2 correct entries, if the maximum angle is limited to  $30^\circ(2\theta)$  :



The searches were realized by using simultaneously the PDF-2 (ICDD) as the master database and the PPDF-1 as a user database. Tuning the cell parameters is allowed in EVA in order to improve the fit (though not done for the above screen shots). Two main problems will obviously lead to difficulties in identification by search-match process :

- Inaccuracies in the predicted cell parameters, introducing discrepancies in the peak positions.
- Incomplete chemistry of the PCOD models, influencing the peak intensities.

Their accumulation will eventually be disastrous if  $\text{PAD} > 2\%$ . However, the three above cases show that identification may succeed satisfyingly if the chemistry is restrained adequately during the search and if the averaged difference in cell parameters is smaller than 1%. Because cell parameter discrepancies have less sensitive effects on peak positions at low diffracting angles, for increasing the chances of success it is suggested to limit the maximum angle on the experimental diffraction pattern to  $40^\circ$  or even  $30^\circ(2\theta)$  during the search. The crystal structures of  $\tau\text{-AlF}_3$  and  $\text{K}_2\text{TiSi}_3\text{O}_9 \cdot \text{H}_2\text{O}$  were determined from powder data (due to the absence of suitable single crystals). Thanks to the PPDF combined with a search-match software, these structures would have been probably determined with much less difficulties, directly at the preliminary identification stage, even before any indexing. The PPDF-1 contains probably a lot of to-be-discovered new crystal structures with open frameworks (titanosilicates, zeolites, etc) which could be of interest to the microporous and metal organic framework (MOF) research communities.

Prediction looks now as a marginal way to solve crystal structures unsolved by the classical single crystal or powder methodologies. Nevertheless the whole potential of the prediction approach is much broader than crystal structure solution alone. An exact theory of materials is one which would allow the full prediction of any possible crystal structure in any physical conditions in the Universe and would allow for the prediction of the physical properties as well. Added to progress in prediction of synthesis conditions, this would start a new era of research, overcoming the current general call to serendipity and providence. The PCOD and PPDF are two very tiny steps in that challenging direction. More details can be found on the Web [4, 5].

### References

- [1] M.D. Foster & M.M.J. Treacy. A Database of Hypothetical Zeolites Structures. <http://www.hypotheticalzeolites.net/>
- [2] A. Le Bail, *J. Appl. Cryst.* **38** (2005) 389-395. <http://www.crystal.org/grinsp/>
- [3] A. Le Bail, *IUCr Commission on Powder Diffraction Newsletter* **31** (2004) 51-53.
- [4] A. Le Bail, PPDF - Predicted Powder Diffraction File. <http://www.crystallography.net/pcod/PPDF/>
- [5] A. Le Bail, PCOD - Predicted Crystallography Open Database. <http://www.crystallography.net/pcod/>
- [6] P. Caussin, J. Nusinovici & D.W. Beard, *Adv. X-ray Anal.* **31** (1988) 423-430.
- [7] J-M. Le Meins, L.M.D. Cranswick & A. Le Bail, *Powder Diffraction*, **18** (2003) 106-113.
- [8] A. Le Bail, J.L. Fourquet & U. Bentrup, *J. Solid State Chem.* **100** (1992) 151-159 (1992).
- [9] A. Le Bail & F. Calvayrac, *J. Solid State Chem.* **179** (2006) 3159-3166.
- [10] M.S. Dadachov & A. Le Bail, *Eur. J. Solid State Inorg. Chem.* **34** (1997) 381-390.
- [11] A. Hemon & G. Courbion, *J. Solid State Chem.* **84** (1990) 153-164.

# DIFFRACTION SOFTWARE FOR THE SPALLATION NEUTRON SOURCE: DIFFDANSE PROJECT

S. J. L. Billinge\*, Emil Bozin, Pavol  
Juhás, Wenduo Zhou, Chris Farrow,  
Jiwu Liu, Dmitriy Bryndin

Department of Physics and Astronomy, Michigan State  
University, East Lansing, MI 48824, USA

\*Contact author; e-mail: [billinge@pa.msu.edu](mailto:billinge@pa.msu.edu)

## Overview of the DiffDANSE project

The spallation neutron source (SNS), which is being commissioned at Oak Ridge National Laboratory in Tennessee, USA, will start generating neutron diffraction data at an unprecedented rate as it ramps up to full power and its suite of instruments reaches completion. Software will play an important role in making the most of these data. Software is needed to run the data acquisition systems and to extract the neutron counts and provide basic services to beamline scientists and users such as viewing and histogramming the data. However, there is a critical need for powerful, scalable and easy to use software tools further downstream, at the data analysis and modelling level. The DANSE (Distributed Data Analysis for Neutron Scattering Experiments) project was born to address this need. It is a multi-university collaboration funded by the US National Science Foundation to develop software that will enable new science and facilitate existing scientific studies using SNS data. There are 5 scientific sub-groups to address the needs of the diffraction, engineering diffraction, inelastic scattering, reflectivity and small angle scattering communities. The project is currently one year into its funding. The diffraction subproject, DiffDANSE, is led by Simon Billinge and headquartered at Michigan State University. The team currently includes 3 post-docs and 3 graduate students (listed as co-authors) working between 25 and 100% of their effort on the project.

This brief article represents a snapshot of the activities to date. The project is a community service enterprise and we encourage feedback and participation from interested readers. Please send comments and encouragement to [billinge@pa.msu.edu](mailto:billinge@pa.msu.edu), or visit our discussion forums which are the Google groups [diffpy-dev](#) and [diffpy-users](#). The project is open source in nature. The source can be found in a subversion repository at <http://danse.us/trac/diffraction>. This also contains other work documents associated with the project. The main applications that are under development in the project are described there. This site is intended for developers, is frequently modified,

and no great effort has been made to make a pretty public portal to it. This will come later!

There are many excellent tools for analysing and modelling diffraction data. Much of the human capital invested in these projects (for example, GSAS [1] has more than 100,000 lines of code) can be utilized by reusing these codes as much as possible. An excellent approach for this is to use a high level interpreted language such as Python to bind to variables in the compiled language (c++, Fortran, etc.) codes such that variable values can be passed in and out at the python level. Computationally expensive tasks are then executed at compile-speed, but the python layer provides greater flexibility for writing scripts that interconnect different code modules. This allows for relatively fast development of high-level applications and this will be one activity in DiffDANSE.

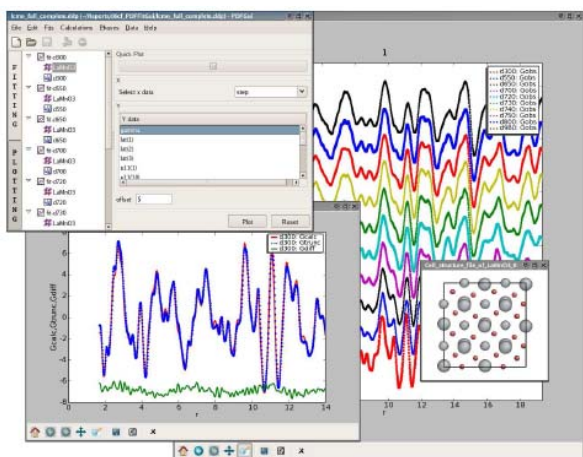
Building on this idea of interactivity of code modules, it is possible to build libraries of modular software components in a framework that allows them to be wired together in different ways and gives advanced users great flexibility in building tailor-made applications and easily extending existing applications by writing small additional components. Another activity of DiffDANSE will be to begin the task of generating these libraries of modules. New code will be generated using software engineering and quality assurance practices that are used in the software industry with the view that the code will be maintainable and extensible into the future. Indeed the code will be passed to SNS at completion of the project for curation. This philosophy adds overhead that limits the scope of what can be accomplished in the current project, but leaves a legacy to the community that can be built on in the future. As mentioned, all code will be open-source.

## Status of the DiffDANSE project

This is a brief report of the status and projects underway at the end of our first of five years of funding. We have two high-level application projects underway of the first type, that reuse existing software for their computation engine. These are PDFgui and RIETgui.

PDFgui is the application that we worked on first to prototype and learn the new programming methods. It is at a fairly advanced stage and we have a full alpha release of the code and are moving towards a beta release in April of 2007. The application provides a gui front-end to PDFfit [2], the popular program for fitting atomic pair distribution functions. It also significantly extends the functionality of that program. The developments are described in an upcoming paper [3]. The original PDFfit codes were driven by scripts. They were flexible, but inconvenient to use. One serious drawback was a limitation on the model size, both due to static memory allocation in the old Fortran code, but also because of the difficulty of entering the atoms and hard-coding the constraints between parameters, even those coming from the space-group symmetry. The gui

provides significant usability features to aid in these tasks. The underlying code has also been extended so that CIF files can be read in and all the constraints coming from the symmetry are automatically generated; a significant time saver. Other usability issues include being able to cascade sequences of fits, for example, for parametric fits such as temperature dependence, real-time plotting of the fits, the ability to plot refined parameters in real-time during a fit sequence, built-in structure viewer, and so on. If you are interested in test-driving this new software, please look for announcements on the Rietveld and TotalScattering mailing lists and elsewhere such as the at [danse.us/trac/diffraction](http://danse.us/trac/diffraction) in the mid April time-frame. A web-page will be set up at [www.diffpy.org](http://www.diffpy.org) with instructions for downloading the data. This currently points to the Billinge-group web-page (lots of good stuff there too!). A screen-shot is shown in figure 1 to whet your appetite.



**Figure 1:** a screen shot of a PDFgui work session.

RIETgui is the Rietveld equivalent of PDFgui and will have many of the same usability features. This project is at a much earlier point; in fact we are still developing the requirements for the code behaviour by acquiring use-cases. How would you like to use such a gui interface? If you would like to contribute ideas please contact me at [billinge@pa.msu.edu](mailto:billinge@pa.msu.edu) or visit [diffpy-dev google group](http://diffpy-dev.google.com) to see some of the discussions. Existing use-cases can be seen at <http://danse.us/trac/diffraction/wiki/RIETgui> under the "Use Cases" link. Right now we are considering which Rietveld code to use as the computational engine, but are most likely to start with FullProf [4], more for the ease of programming the interface than any denominational preference.

This fast-track application development using existing software engines will be a significant focus for the summer and autumn so that we have a working Rietveld application in time for first neutrons on POWGEN3, the materials science powder diffractometer at SNS. More attention will be given to

the development of library modules after this point; hopefully more news on that next year!

### Acknowledgements

The DiffDANSE project is funded by the US National Science Foundation through grant DMR0520547

### References

- [1] Larson, A. C. and Von Dreele, R. B., 1987, Report No. LAUR-86-748, Los Alamos National Laboratory, Los Alamos, NM 87545.
- [2] Proffen, Th., and Billinge, S. J. L., 1999, *J. Appl. Crystallogr.*, 32, 572.
- [3] C. L. Farrow, C. L. Juhas, P., Liu J. W., Bryndin, D., Bloch, J., Proffen, Th., and Billinge, S. J. L., 2007, *J. Phys.: Condens. Matter* (to be published).
- [4] Rodriguez-Carvajal, J., 1993, *Physica B* 192, 55.

# ADVANCES IN GENETIC ALGORITHMS FOR DIRECT- SPACE STRUCTURE SOLUTION

K.D.M. Harris<sup>\*</sup>, Z. Zhou, Z. Pan

School of Chemistry, Cardiff University, Park Place,  
Cardiff CF10 3AT, Wales

<sup>\*</sup>Contact author; e-mail: HarrisKDM@cardiff.ac.uk

## Introduction

In the last decade or so, most reported work on crystal structure determination of organic molecular solids from powder X-ray diffraction (XRD) data [1-8] has used the direct-space strategy for structure solution, although several successful applications of traditional approaches for structure solution of such materials have also been reported. The direct-space strategy handles the structure solution process as a "global optimization" problem, with trial structures generated in direct space (independently of the experimental powder XRD data) and the suitability of each trial structure is assessed by direct comparison between the powder XRD pattern calculated for the trial structure and the experimental powder XRD pattern. This comparison is quantified using an appropriate figure-of-merit (for reasons elaborated elsewhere, our implementations of the direct-space strategy have used the weighted powder profile R-factor  $R_{wp}$ ). Each trial structure is defined by a set of structural variables, representing the position, orientation and intramolecular geometry of each molecule in the asymmetric unit. The aim is to find the trial structure corresponding to lowest  $R_{wp}$ , and is equivalent to exploring an  $R_{wp}$  hypersurface to find the global minimum. In principle, any technique for global optimization may be used, and much success has been achieved using Monte Carlo/simulated annealing, genetic algorithm and differential evolution techniques. This paper highlights some of our recent work employing the genetic algorithm (GA) technique for structure solution.

## Methodology

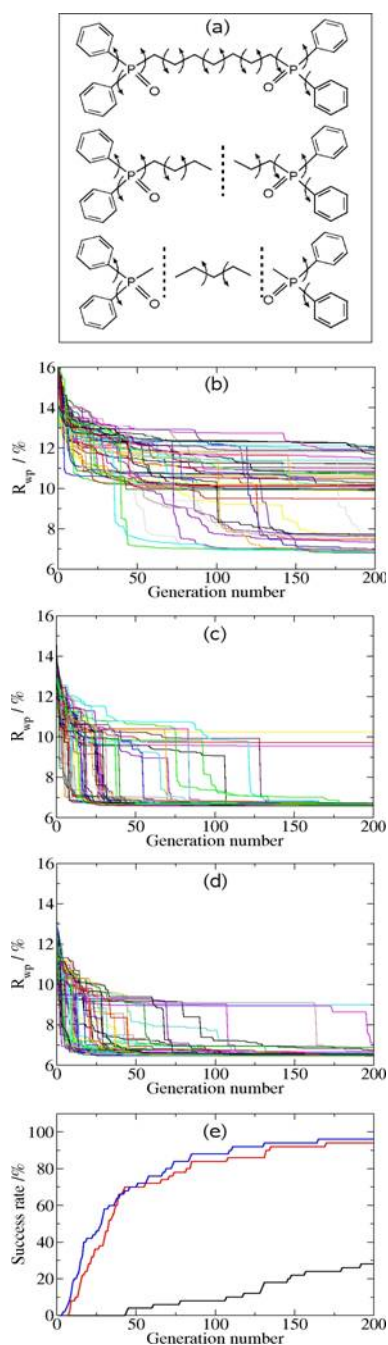
Details of the methodology in our implementation of the GA technique for direct-space structure solution, as implemented in our program EAGER [9], have been outlined previously [10-12].

A recent development has focused on improving the efficiency of structure solution in the case of conformationally flexible molecules, which is widely regarded as one of the most challenging situations for direct-space structure solution in view of the large number of torsion angle variables involved in the

calculation. Conventionally, the structural variables comprise, for each molecule in the asymmetric unit, the position  $\{x, y, z\}$  and orientation  $\{\theta, \varphi, \psi\}$  of the whole molecule, and a set of  $n$  variable torsion angles  $\{\tau_1, \tau_2, \dots, \tau_n\}$  to define the molecular conformation. We have challenged this convention by exploring alternative definitions of variable-space [13] in which the molecule is broken into two or more smaller fragments by breakage of bonds. For example, in the conventional approach, the molecule Ph-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>2</sub>-Ph (Ph = phenyl) would be defined by a single fragment Ph-C-C-Ph with 10 variables (4 variable torsion angles). However, this molecule could instead be represented by two fragments Ph-C and C-C-Ph by breaking a C-C bond, with  $6 + 7 = 13$  variables (including 1 variable torsion angle), or by three rigid fragments Ph, C-C-C and Ph by breakage of two C-C bonds, with  $6 + 6 + 6 = 18$  variables (including no variable torsion angles). In general, breaking a molecule into a greater number of fragments is associated with: (a) an increase in the total number of structural variables, (b) the loss of some known information on molecular geometry (breakage of a bond discards knowledge of bond length and bond angles), and (c) a decrease in the number of torsion angle variables. Both (a) and (b) would appear to be disadvantageous for efficient direct-space structure solution. However, as molecules with a large number of torsion angle variables represent a significant challenge for direct-space structure solution, the fact that multiple-fragment definitions reduce the number of torsion angle variables (i.e. (c)) might outweigh the disadvantages of (a) and (b).

Extensive investigations of the relative performance of GA structure solution calculations for different definitions of variable-space and for test molecules representing different degrees of conformational flexibility have demonstrated [13] that multiple-fragment definitions can have significant advantages over the conventional single-fragment definition employed hitherto in this field. An example is shown in figure 1. However, the successful implementation of the multiple-fragment definition depends on appropriate selection of the number of fragments and appropriate choice of the broken bonds. For flexible molecules of moderate size, breakage into two or three fragments appears to be optimal, and is advantageous over the single-fragment case and generally advantageous over the use of a greater number of fragments. However, for molecules that are relatively small and/or substantially rigid, the multiple-fragment definition does not lead to any advantages over the single-fragment definition. Clearly, the overall success of using a given definition of variable-space depends on the interplay of a number of factors, such as finding the optimal balance between keeping the total number of variables low [lowest for the single-fragment definition] and keeping the number of torsion angle variables low [lowest (zero) for a multiple-fragment definition with each fragment rigid].



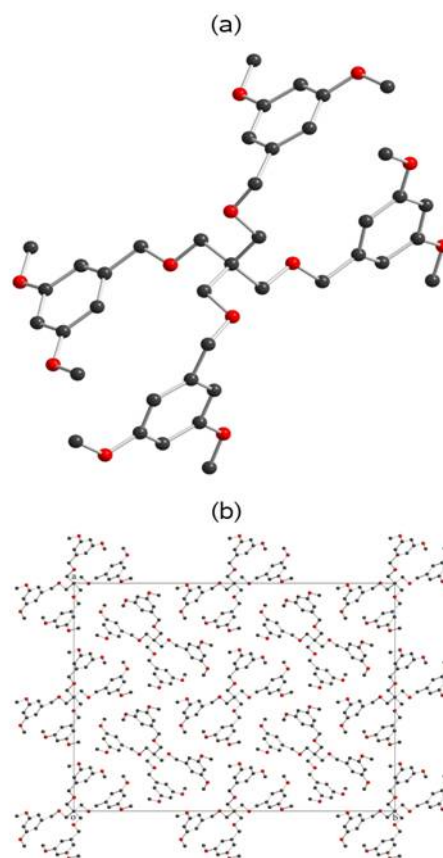


**Figure 1:** (a) Molecular structure of a test molecule with significant conformational flexibility (dashed lines represent the broken bonds in the 2-fragment and 3-fragment calculations). (b-d) The lowest value of  $R_{wp}$  in the population versus generation number for GA structure solution calculations as follows: (b) 50 runs for single-fragment definition, (c) 50 runs for 2-fragment definition and (d) 50 runs for 3-fragment definition. The correct structure solution corresponds to  $R_{wp}$  in the range 6.6 – 7.7 %. (e) Cumulative success rates for the single-fragment [black line; results from (b)], 2-fragment [blue line; results from (c)] and 3-fragment [red line; results from (d)] calculations.

## Applications

The EAGER program has been applied successfully to tackle a range of structural problems within the field of molecular solids. In addition to examples from several

areas of solid state and materials chemistry, the range of structures determined [see refs. 2, 7 and 12 for examples] has encompassed materials of biological interest (oligopeptides), photoluminescent materials and industrially important materials from the pharmaceuticals and pigments sectors. The technique has also been applied to obtain structural understanding of solid state processes that intrinsically lead to polycrystalline powders as the product phase, such as the preparation of co-crystals *via* solid state grinding routes [14], the production of materials by solid state desolvation processes [15] and materials prepared directly by precipitation from solution state reactions without recrystallization [16]. We highlight here an example of structure determination [17] of an early-generation dendrimeric material tetrakis[3,5-dimethoxybenzyloxy)methyl]methane (TDMM), as shown in figure 2.



**Figure 2:** (a) Molecular structure of TDMM. (b) Crystal structure of TDMM determined from powder XRD data.

Given the highly branched molecular architecture of dendrimers, they are often recalcitrant to crystallization as single crystals suitable for single crystal XRD studies. Clearly the availability of techniques for structure determination from powder XRD data will serve an important role in the future structural characterization of this increasingly important class of molecular solids.



## Acknowledgements

We are grateful to EPSRC and Astra Zeneca for financial support.

## References

- [1] Harris, K.D.M., Tremayne, M., Lightfoot, P. & Bruce, P.G., 1994, *J. Am. Chem. Soc.*, 116, 3543.
- [2] Harris, K.D.M., Tremayne, M. & Kariuki, B.M., 2001, *Angew. Chemie Int. Ed.*, 113, 1674.
- [3] Chernyshev, V.V., 2001, *Russian Chem. Bull.*, 50, 2273.
- [4] David, W.I.F., Shankland, K., McCusker, L.B., Baerlocher, C. (Editors), 2002, *Structure Determination from Powder Diffraction Data* (Oxford University Press/International Union of Crystallography).
- [5] Tremayne, M., 2004, *Phil. Trans. Roy. Soc.*, 362, 2691.
- [6] Altomare, A., Caliandro, R., Camalli, M., Cuocci, C., Giacovazzo, C., Moliterni, A.G.G. & Rizzi, R., 2004, *J. Appl. Crystallogr.*, 37, 1025.
- [7] Harris, K.D.M. & Cheung, E.Y., 2004, *Chem. Soc. Rev.*, 33, 526.
- [8] Černý, R., 2006, *Croat. Chem. Acta*, 79, 319.
- [9] Habershon, S., Zhou, Z., Turner, G.W., Kariuki, B.M., Cheung, E.Y., Hanson, A.J., Tedesco, E., Johnston, R.L. & Harris, K.D.M., EAGER – A Computer Program for Direct-Space Structure Solution from Powder X-ray Diffraction Data, Cardiff University and University of Birmingham.
- [10] Kariuki, B.M., Serrano-González, H., Johnston, R.L. & Harris, K.D.M., 1997, *Chem. Phys. Lett.*, 280, 189.
- [11] Harris, K.D.M., Johnston, R.L. & Kariuki, B.M., 1998, *Acta Crystallogr.*, A54, 632.
- [12] Harris, K.D.M., Habershon, S., Cheung, E.Y. & Johnston, R.L., 2004, *Z. Kristallogr.*, 219, 838.
- [13] Zhou, Z., Sieglar, V., Cheung, E.Y., Habershon, S., Harris, K.D.M. & Johnston, R.L., 2007, *ChemPhysChem*, 8, 650.
- [14] Cheung, E.Y., Kitchin, S.J., Harris, K.D.M., Imai, Y., Tajima, N. & Kuroda, R., 2003, *J. Am. Chem. Soc.*, 125, 14658.
- [15] Guo, F. & Harris, K.D.M., 2005, *J. Am. Chem. Soc.*, 127, 7314.
- [16] Guo, F., Casadesus, M., Cheung, E.Y., Coogan, M.P. & Harris, K.D.M., 2006, *Chem. Commun.*, 1854.
- [17] Pan, Z., Xu, M., Cheung, E.Y., Harris, K.D.M., Constable, E.C. & Housecroft, C.E., 2006, *J. Phys. Chem. B*, 110, 11620.

# CULTURAL AND DIFFERENTIAL EVOLUTION IN DIRECT SPACE STRUCTURE SOLUTION

Maryjane Tremayne\* & Samantha Y.Chong

School of Chemistry, University of Birmingham,  
Edgbaston, Birmingham, B15 2TT, UK.

\*Contact author; e-mail: m.tremayne@bham.ac.uk

## Introduction

The development of direct (or real) space structure solution methods has enabled us to tackle structures of size, complexity and in numbers (especially in terms of organic or molecular materials) not thought possible twenty years ago.

These real-space methods approach structure solution by the generation of trial crystal structures based on the known connectivity of the material, and assessment of the fitness of each structure by comparison between the calculated diffraction pattern for each structure and the experimental data. These trial structures are generated by movement of a collection of atoms, forming a structural model, around the unit cell. The resulting models are described by a list of elements e.g. the position ( $x,y,z$ ) and orientation ( $\theta,\phi,\psi$ ) of each structural unit within the unit cell, and torsion angles to describe molecular conformation ( $\tau_1\dots\tau_N$ ). The fitness of each trial structure is then ranked by crystallographic  $R_{wp}$  (or  $\chi^2$ ). However, the complexity of the problem is not defined by the number of atoms in the model but the number of elements used to define the structural problem and hence the dimensionality of the fitness landscape. The key to the success of this approach is the application of a global optimisation procedure used to locate the minimum point on the fitness landscape (or hypersurface) corresponding to the best structure solution.

Our recent work has focussed on the development and application of a global optimisation technique that combines the principles of social and biological evolution with the aim of producing a hybrid evolutionary algorithm (Cultural Differential Evolution) that searches the fitness landscape more efficiently than other optimisation techniques.

Evolutionary algorithms are being increasingly used to solve a variety of global optimisation problems in chemistry, nanoscience and bioinformatics [1]. These powerful techniques are inspired by natural evolutionary processes, and mimic the principles of biological evolution and survival of the fittest to explore parameter space. However, the biological evolution of natural systems can be a slow process (i.e. thousands of years for the evolution of a species), especially compared to the rate of cultural evolution in a society when adapting to changing social environment (i.e. a few years for a fashion trend or hair

style to become popular). Cultural algorithms [2] have been developed to model behaviour based on the principles of human social evolution, and can be used to bias the search process by passing experience and knowledge of behavioural traits of a population from one generation to the next. In simple terms, this cultural information can be used to reduce the search space of a standard biological evolutionary algorithm improving both performance and efficiency of the global optimisation process. In this paper, we report modification to the Differential Evolution (DE) global optimisation algorithm, by incorporation of the concept of Cultural Evolution, with the aim of increasing the efficiency of DE when applied to crystal structure solution from powder diffraction data [3].

## Differential Evolution

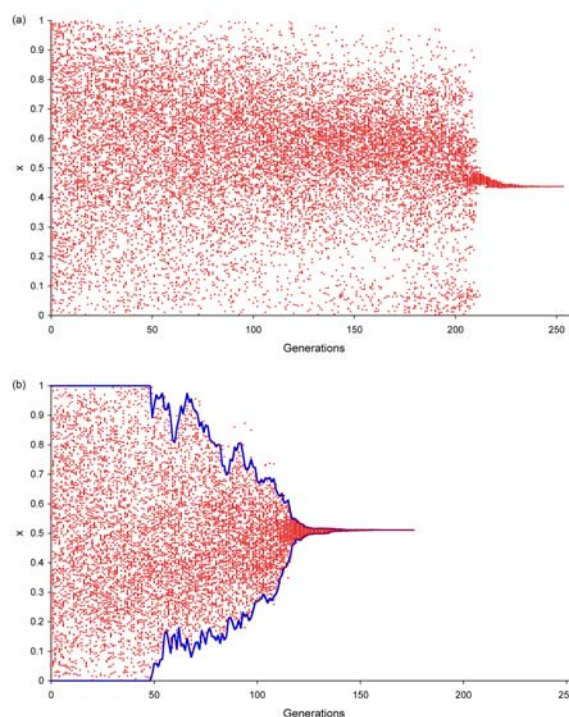
Evolutionary techniques such as DE [4] or genetic algorithms [5,6] maintain a population of trial structures (or members) which are mutated and mated over a number of generations until the global minimum is located. However, the processes used to achieve this are markedly different. In genetic algorithms, a series of recombination and mutation steps are performed on randomly selected members of the population and from this collection the new population is probabilistically selected. In a DE population, each child is created from its parent and three randomly selected members of the population, by the summation of the differences weighted according to the amount of recombination and mutation required. In this way, both processes are carried out in a single step as shown in equation (1).

$$\text{Child} = \text{Parent} + K(\text{Random}_1 - \text{Parent}) + F(\text{Random}_2 - \text{Random}_3) \quad (1)$$

Each child structure is then directly compared to its parent such that the fitter of the two is retained (in a deterministic manner), constantly updating the population and adapting to the fitness hypersurface. The values of  $K$  and  $F$  (the rates of recombination and mutation respectively), and the population size  $N_p$ , are chosen to achieve a balance between optimal fitness of the solutions and the time taken for the calculation to converge [7]. The parameters  $K$  and  $F$  can theoretically take any value between 0 and 1. In our application, tests have shown that optimal convergence is attained using high  $K$  (0.99 or 1) and median values of  $F$ . Although DE is a relatively new evolutionary algorithm it has proved highly effective in a range of chemical contexts, including X-ray scattering, crystal growth epitaxy, optimization of clusters, protein crystallography, molecular docking, disordered crystal structures and the direct-space crystal structure solution of organic molecules from powder diffraction data [7,8].

## Cultural Differential Evolution

One of the features of the DE method is its use of boundary constraints during the optimisation process. In the DE, each of the variables (or elements) used to describe the structural model has associated minimum and maximum boundary values which are checked by the algorithm when trial structures are generated. If any of the elements describing the child structure exceed the corresponding bounds, it is reset to a point between the boundary and parent values. This would typically involve minimum and maximum boundary values of [0,1] for fractional position and [0,360°] for overall molecular orientation and intramolecular torsion angles. These inherent boundary conditions can also be used to restrict the DE search to specific regions of the hypersurface, allowing incorporation of structural constraints such as limits in molecular conformation, while enhancing the efficiency of the search rather than disrupting the natural optimisation pathway. In the original DE algorithm, these boundary values remain constant throughout the structure solution calculation. However, as illustrated in figure 1a, examination of the child structures generated during a DE calculation with static boundaries shows that the random initial distribution in the values of a variable develops clustering as the calculation progresses.



**Figure 1:** Distribution of the  $x$  variable of the child structures (circles) generated during a DE calculation, (a) with static boundaries and (b) dynamic boundaries (the solid line indicates the boundary position).

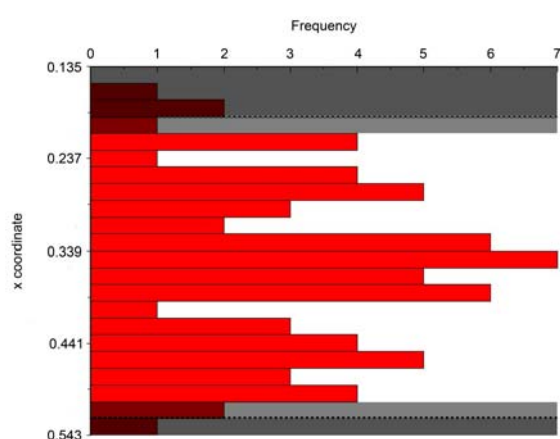
This distribution provides us with information that can be used, by incorporation of the concept of Cultural Evolution, to guide the DE search itself. In our implementation of the Cultural Differential Evolution

hybrid algorithm (CDE), the behaviour of previous generations is used to influence subsequent generations using dynamic boundaries to restrict the search to low-lying regions of the hypersurface, and effectively prune population space. This mechanism of combining the two approaches is simple to implement and interpret in ‘real-world’ applications where parameter distribution and the resulting boundary conditions can also have physical meaning. It differs from that of Becerra et al [9] in which the DE static boundaries are retained, and cultural evolution is used instead to influence the DE variation operator. The use of dynamic boundaries has a dramatic effect on the distribution of structure variables in the DE calculation. In the example given, the  $x$  variable is allowed to take values  $0 \leq x \leq 1$  throughout the conventional DE calculation (figure 1a), whereas in the CDE, the dynamic boundaries restricted the search to  $0.2 \leq x \leq 0.7$  after only 90 generations (figure 1b).

In our CDE algorithm, the original boundary conditions are initially maintained to avoid encouraging premature convergence to local minima (e.g. 50 generations as shown in figure 1b).

After this period, a virtual histogram of child parameters is constructed for each variable within a generation (figure 2). This detects any clustering, and identifies at what values the dynamic boundaries are set by imposing an ‘underpopulation threshold’ at the maximum and

minimum ends of the distribution. By defining the dynamic boundaries by exclusion of ‘outliers’ rather than inclusion of ‘popular’ values, potential problems with multi-modal distribution and over-aggressive culture-based pruning are avoided. Figure 2 shows the distribution of the  $x$  variable within a population of 70 structures, divided over 22 histogram bins.

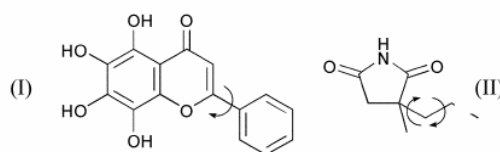


**Figure 2:** Bins are removed from the histogram until the underpopulation threshold is reached (shaded region), but one bin at each extreme is reinstated to give the new boundaries (dotted lines).

In this case, the underpopulation threshold ( $N_{ut}$ ) at each end of the distribution is set to four structures, and the end bins removed until the threshold is reached. In order to allow the possible expansion of the boundaries with successive generations, one bin is then reinstated at each extreme, and the maximum and minimum of these remaining categories used as the new dynamic boundary values for the next generation. These boundaries are then invoked in the DE calculation as described earlier.

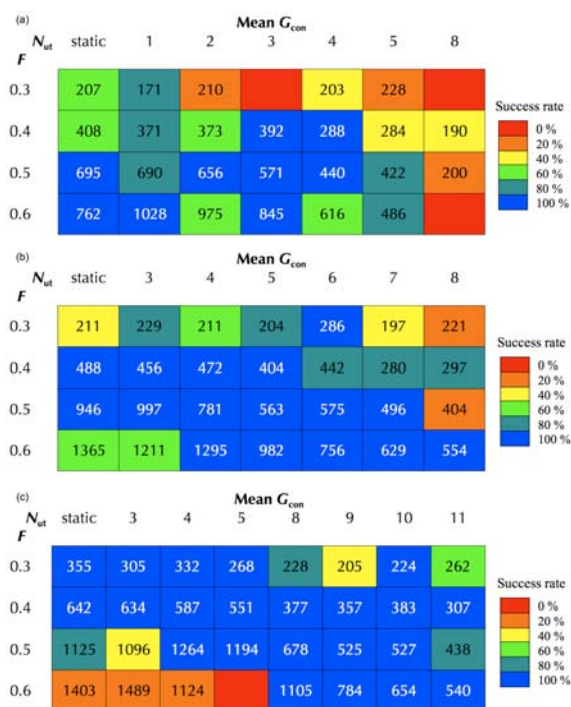
## Results

In this paper, the performance of the CDE is demonstrated by the structure solution of (i) a test case, baicalein [10] (I) and (ii) the crystal structure of  $\alpha$ -methyl- $\alpha$ -propyl succinimide [3] (II). In both cases, the structure solution used a model comprising the whole molecule (excluding hydroxyl, methyl and amide hydrogens where applicable), and allowed translation throughout the unit cell, rotation in all directions and intramolecular rotation defining molecular conformation (figure 3).



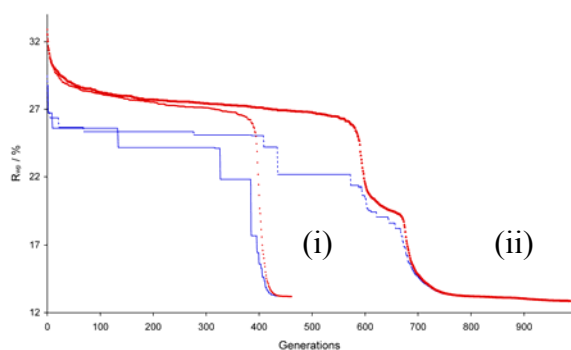
**Figure 3:** Structural models of baicalein (I) and  $\alpha$ -methyl- $\alpha$ -propylsuccinimide (II). Arrows show variable torsion angles.

Baicalein was studied initially using the DE approach (static boundaries) with 5 DE runs performed for each combination of control parameters  $K=0.99$ ,  $N_p=70, 105$  and  $140$ ,  $F=0.3, 0.4, \dots, 0.6$  (Tables 1a-c respectively). Corresponding sets of CDE calculations were then performed using  $N_{ut}$  values from 1 to 11, with the same DE control parameters, (Tables 1a-c). These results show a significant and consistent gain in the efficiency of the calculation over the range of  $N_p$ . The CDE algorithm converges up to 37% quicker when averaged over the 5 runs for each set of parameters with  $N_p=70$  ( $F=0.5$ ,  $N_{ut}=4$ ); up to 54% quicker for  $N_p=105$  ( $F=0.6$ ,  $N_{ut}=7$ ), and up to 62% quicker for  $N_p=140$  ( $F=0.6$ ,  $N_{ut}=11$ ). Similar results have been obtained for other test structures of varying complexity. It is clear from these results that the CDE has a proportionally greater effect on larger populations with higher mutation factors, although the quickest calculations are obviously performed with smaller  $N_p$  and  $F$ . It is also clear that the optimal choice of  $N_{ut}$  is, as expected, dependent on the combination of the other DE control parameters.



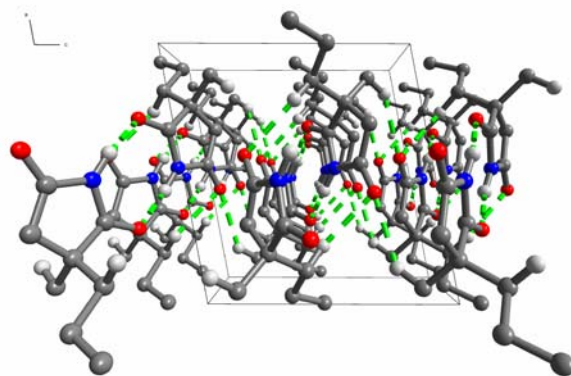
**Table 1:** The mean rate of convergence for successful DE and CDE runs (those converged to the global minimum) for (a)  $N_p=70$ , (b)  $N_p=105$  and (c)  $N_p=140$ . The colour of each box denotes the % of successful runs for each set of control parameters.

In the case of (II), the CDE calculation was run several times with the optimal control parameters  $K=0.99$ ,  $F=0.5$ ,  $N_p=80$  and  $N_{st}=4$  until convergence was reached (see figure 4).



**Figure 4:** DE progress plot showing the best  $R_{wp}$  (circles) and mean  $R_{wp}$  (line) for each generation in (i) the optimum CDE calculation and (ii) the optimum DE calculation for (II).

The best solution had  $R_{wp} = 13.2\%$  (mean  $R_{wp} \approx 33\%$ ), and was used as the starting point for successful Rietveld refinement yielding the final hydrogen-bonded crystal structure (figure 5).



**Figure 5:** Crystal structure of  $\alpha$ -methyl- $\alpha$ -propylsuccinimide. Only H atoms involved in hydrogen bonding (indicated by dashed lines) are shown.

This structure was also identified as the global minimum by a subsequent set of conventional DE calculations, using the same optimisation control parameters, but requiring significantly longer convergence as shown in figure 4, i.e. the optimum CDE calculation converged after only 461 generations, whereas the optimum DE calculation needed 988 generations for convergence. Over a range of search control parameters  $F$ ,  $N_p$  and  $N_{st}$  the CDE consistently reached convergence more quickly than the corresponding DE calculation. Cases were also identified in which particular combinations of these parameters resulted in convergence of the DE calculation in a local minima, whereas the CDE optimisation continued to locate the correct structure solution (i.e.  $F=0.5$ ,  $N_p=80$  and  $N_{st}=4$ , the optimum CDE converged after only 163 generations).

## Concluding remarks

We have demonstrated that the Cultural Differential Evolution hybrid algorithm shows significantly quicker convergence on the global minimum when applied to crystal structure solution from powder diffraction data (an average of 40% improvement in our tests). The use of dynamic boundaries which are allowed to expand or contract with successive generations is an essential feature of our implementation, ensuring that the process does not become too restrictive. It allows the algorithm to follow population clustering that while unlikely to expand in terms of parameter range, may shift in terms of absolute parameter values as the population evolves. Our work describes the first application of the concept of cultural evolution in a chemical or crystallographic context, and demonstrates the major gains in optimisation efficiency that can be achieved by combining the dictates of biological and social evolution.

## Acknowledgements

We would like to thank the Royal Society for the award of a University Research Fellowship to MT and the EPSRC and University of Birmingham for their support for SYC.

## References

- [1] Johnston, R.L. (Eds) 2004, *Applications of Evolutionary Computation in Struct. Bonding*, 110, (Springer, Berlin/Heidelberg).
- [2] Engelbrecht, A.P. 2002, *Computational Intelligence: An Introduction* (John Wiley & Sons, Chichester, UK), p171.
- [3] Chong, S.Y. & Tremayne, M. 2006, *Chem. Comm.*, 4078.
- [4] Price, K.V., 1999, in *New Ideas in Optimisation*, edited by D. Corne, M. Dorigo & F. Glover (McGraw-Hill, London, UK), p 77.
- [5] Harris, K.D.M., Johnston, R.L. & Kariuki, B.M. 1998, *Acta Cryst.*, A54, 632.
- [6] Cheung, E.Y., McCabe, E.E., Harris, K.D.M., Johnston, R.L., Raja, K.M.P & Balaram, P. 2002, *Angew. Chem., Int. Ed.*, 41, 494.
- [7] Tremayne, M., Seaton, C.C. & Glidewell, C. 2002, *Acta Cryst.*, B58, 823.
- [8] Seaton, C.C. & Tremayne, M. 2002, *Chem. Comm.*, 880.
- [9] Becerra, R.L. & Coello, C.A.C. 2006, *Comput. Methods Appl. Mech. Eng.*, 195, 4303.
- [10] Rossi, M., Meyer, R., Constantinou, P., Caruso, F., Castelbuono, D., O'Brien, M. & Narasimhan, V. 2001, *J. Nat. Prod.*, 64, 26.

# INTEGRATION OF DIRECT METHODS INTO DIRECT SPACE TECHNIQUES

A. Altomare<sup>1</sup>, R. Caliandro<sup>1</sup>, C. Cuocci<sup>2</sup>,  
C. Giacovazzo<sup>1,2\*</sup>, A.G. Moliterni<sup>1</sup>,  
R. Rizzi<sup>1</sup>

<sup>1</sup>Istituto di Cristallografia (IC), CNR, Sede di Bari. Via G. Amendola 122/o, 70126, Bari, Italy.

<sup>2</sup>Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy.

\*Contact author; e-mail: carmelo.giacovazzo@ic.cnr.it

## Introduction

The success in crystal structure solution from powder diffraction data using traditional approaches like Patterson or Direct Methods (*DM*), is strictly related to the quality of the data and to the presence or not, in the molecule, of heavy atoms. The use of these methods traditionally involves a two step procedure. In the first stage, the experimental pattern is decomposed in single Bragg integrated intensities, which are then used in the second stage for crystal structure solution. Inefficiency in the first stage can severely affect the success of the second one. This is the reason why, for organic compounds, the electron density maps are scarcely informative: from them it is very difficult to recognize a physically consistent structural model. On the contrary, the use of *Direct Space Methods* is less demanding for the quality of diffraction data, but requires the prior knowledge of a molecular model: the structure solution depends on the number of degrees of freedom necessary to describe the crystal structure. Direct-space methods (*Grid Search* [1], *Genetic Algorithm* [2-4], *Monte Carlo* [5-7] and *Simulated Annealing* [8-13]) are able to efficiently explore an *N*-dimensional hypersurface, where *N* is the number of structural parameters that must be varied to describe all the trial structures which agree with the prior information. Structure solution is then equivalent to searching for the minimum value (global minimum) of the *N*-dimensional hypersurface.

In this last years *Hybrid Techniques* [14-15] have been developed; they combine the best features of two different approaches. In this contribution, we will describe a combination of *DM* with *Simulated Annealing* (*SA*) techniques, implemented in the *EXPO2004* program [16].



## The Procedure

A typical *DM* default run applied to powder diffraction data of an organic compound ends with an electron density map containing weakly connected or isolated atoms. From them it is very hard to find the correct structure model. However, if few atoms in the electron density map are correctly located (and that occurs quite frequently), they can be used as pivots in a *SA* approach, to accommodate the trial model. The advantage in combining the best features of this two different methods, is to work with a problem characterized by a reduced number of degrees of freedom (*DOFs*). In the procedure, each peak of the map is associated in sequence with all the atoms of the model, and for each association, the position of the trial structure is fixed while the orientation and the torsion angles are varied.

The principal aspects of the procedure are the following:

1) The structural model is described in terms of internal coordinates, i.e. bond lengths, bond angles and torsion angles formed by each atom with the previous ones, via a *Z-matrix* representation [17].

2) Isotropic thermal parameters for the non-hydrogen atoms in the asymmetric unit (*NAT*) are estimated. 20 trial sets of possible thermal factors are calculated as follows:

$$B(i) = [1 + (j-1) * 0.3] * \sqrt{\frac{\langle Z \rangle}{Z(i)}} \quad i = 1, \dots, NAT$$

$$j = 1, \dots, 20$$

where  $B(i)$  and  $Z(i)$  are respectively the isotropic thermal factor and the atomic number of *ith* atom.  $\langle Z \rangle$  is the average atomic number for the *NAT* atoms. To evaluate the best set of  $B(i)$  values, for each *jth* set, the

$$R_F = \frac{\sum |F_{oss} - K F_{calc}|}{\sum F_{oss}}$$

calculation is performed.  $F_{oss}$  and  $F_{calc}$  are respectively the structure factors observed and calculated,  $K$  is a scale factor. The *jth*  $B$ -set, corresponding to the minimum  $R_F$  value, is associated to the atoms of the trial model.

No matter the complexity of the crystal structure under study, the *SA* search procedure involves a very large number of feasible trial structures, the correctness of which has to be evaluated.

3) To discriminate the correct from the false solutions a cost function ( $CF$ ) is calculated for each feasible solution. The efficiency of the chosen  $CF$  (in terms of the computer time needed for its calculation and of its discriminating power) strongly conditions the success of the phasing process. We use the following simple expression:

$$CF = \frac{\sum |I_{oss} - KI_{calc}|}{\sum I_{oss}} \quad (1)$$

where  $I_{oss} = |F_{oss}|^2$  is the ‘‘observed’’ intensity (i.e., as estimated by Le Bail method [18]),  $I_{calc}$  is the intensity calculated from the trial model and  $K$  is a scale factor. The  $CF$  calculation is based on a limited number of reflections: only low-angle reflections are used, which are single (i.e., not overlapping with any other) or partners of a doublet (i.e., two reflections are considered to be in overlapping if the difference between their two  $2\theta$  Bragg angles is less than  $0.6 \times FWHM$ , where  $FWHM$  is the full width at half maximum).

The success of the procedure mostly depends on: a) an algorithm able to exhaustively explore the  $N$ -dimensional hypersurface; b) the discriminating power of  $CF$ , which should be capable of recognizing the global minimum against numerous local minima.

4) We use the Metropolis criterion [19] to accept or reject a trial model, and an annealing schedule able to maximize the efficiency in the hypersurface exploration and to minimize the total execution time. The default values used by our procedure are reported in Table 1:  $T_{in}$  is an arbitrary starting temperature,  $N_{tot}$  is the maximum number of moves that can be performed for each peak-atom association,  $f_1$ ,  $f_2$  and  $f_4$  are the temperature reducing values during a *SA* run used as follows:

a)  $\Delta T/T = -f_1$  every three Monte Carlo moves until the fraction of moves which are accepted is below  $acc_{first}$  value;

b)  $\Delta T/T = -f_2$  if the number of accepted moves at the current temperature is bigger than  $f_3 N_{tot}$ ;

c)  $\Delta T/T = -f_4$  if the total number of moves at the current temperature exceeds  $N_{tot}$ .

$acc_{first}$  and  $acc_{last}$  are the percentage of accepted configurations necessary respectively to start and to stop the procedure.

$T_{in}$	0.4	
$N_{tot}$	5000	if $NTV < 9$
	$5000 * NTV$	otherwise
$f_1$	0.1	if $NTV < 9$
	0.05	otherwise
$f_2$	0.2	
$f_3$	0.5	
$f_4$	0.1	
$acc_{first}$	0.4	if $NTV < 9$
	0.7	otherwise
$acc_{last}$	0.02	

**Table 1:** Default values of the control parameters used in the *SA* procedure.

We can observe that some control parameters are dependent on the number of torsion angles to be varied (*NTV*) allowing a more exhaustive sampling of the parameter space.

5) During the *SA* search, a number of accepted trial models (on the basis of the Metropolis criterion), are stored in a list (say *list\_mod*) and submitted to a '*local minimization procedure*'. Each configuration is characterized by two *CF* values:  $CF_{nref}$  and  $CF_{ref}$  which are respectively the non-refined and the locally refined *CF* values. Each new accepted solution is compared with the higher  $CF_{nref}$  value (let it correspond to the *j*th model) in *list\_mod*. If higher it is submitted to a new *SA* move, otherwise it is submitted to the local refinement. The minimization involves the variation, in sequence, of three local operators: torsions angles (*step1*), rotations angles (*step2*) and translations (*step3*). Let say  $CF_{minto}$ ,  $CF_{minro}$  and  $CF_{mintr}$  the minimum values of the cost function at the end of the *step1*, *step2* and *step3* respectively. In this local minimization procedure (*step1-step3*) the only criterion used to accept or reject the current model is the reduction of *CF* value.

*Step1*: all the specified torsion angles from the user are varied in sequence by performing, for each one, 100 random moves and re-starting, every time, from the best accepted conformation. The new value of the *i*th internal degree of freedom is calculated, as follows:

$$x_i^{new} = x_i^{old} + r_i \Delta x_i$$

where  $x_i^{old}$  is the previous accepted value of the internal coordinate  $x_i$ ,  $r_i$  is a random number ( $\pm 1$ ) and  $\Delta x_i$  is a random number calculated using an exponential probability distribution.

The refined configuration with  $CF_{minto}$  value is selected for a successive refinement of its orientation with respect to the fixed position of the model (*step2*).

*Step2*: Eight systematic angle shift combinations of the Euler angles ( $2^\circ$ ) are explored; the configuration with  $CF_{minro}$  value is selected and accepted as the new trial structure if  $CF_{minro} < CF_{minto}$ . Also in this case the rotation refinement procedure can be repeated until lower  $CF_{minro}$  values are found.

*Step3*: In this step the constraints on the pivot peak position introduced by the procedure, are relaxed. A systematic exploration along fourteen directions (regularly exploring the full solid angle) in the unit cell in steps of  $0.05 \text{ \AA}$  are executed, and the minimum conformation is accepted if  $CF_{mintr} < CF_{minro}$ . The procedure may be repeated until  $CF_{mintr}$  decreases; position shifts far by more than  $0.3 \text{ \AA}$  from the initial position are rejected.

At the end of the local minimization procedure, the current structure is included in *list\_mod* if its  $CF_{ref}$  value is lower than the corresponding value of the *j*th model.

The characterization of each feasible solution by two *CF* values ( $CF_{nref}$  and  $CF_{ref}$ ), used to select ( $CF_{nref}$ ) and to accept or reject ( $CF_{ref}$ ) the configuration under

examination, avoid the elimination, from the *list\_mod*, of initial bad model (high  $CF_{nref}$  value) that are able to converge to a better model at the end of the local refinement procedure (low  $CF_{ref}$ ). The execution of the local minimization procedure on each accepted model, will increase the cost of each calculated trial structure, but we have found that this is essential for a feasible comparison of competing structures and thus for the method to work at all.

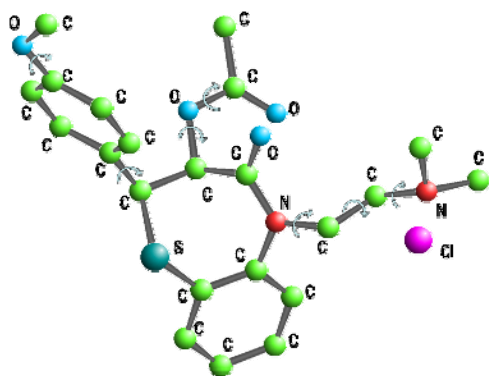
6) The final step of the procedure consists of weighted least squares cycles (in accordance with the approach recently developed and implemented in EXPO2004 [20]), refining atomic positions and isotropic thermal factors. In the first cycle the positions and thermal factors are freely refined. In order to avoid that the connectivity of the molecular model breaks (owing to the low accuracy of the structure factor moduli and/or of the small ratio "number of parameters to refine/number of observations"), in the successive cycles bond distances and angles are restrained to their expected values. The procedure automatically applies suitable weights on distance and on angle restraints.

## Applications

We have checked the efficiency of the method on a large number of tests structures (mainly of organic nature). Two cases are below discussed (examples 1 and 2).

### Example 1: structure DILTIA [21]

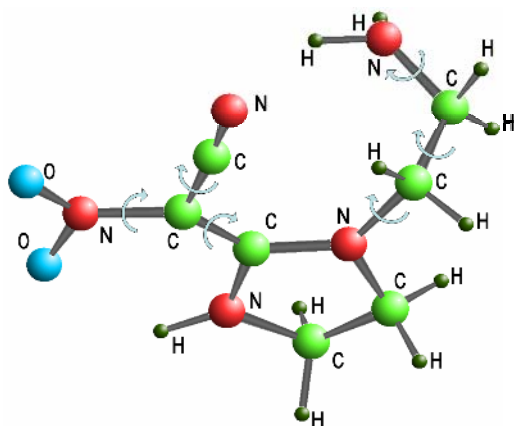
At the end of the *DM* step EXPO2004 generates an electron density map from which the highest six peaks are selected and used as pivots of the *SA* procedure. They are all isolated (i.e., not at bond distance from any other peak) and sufficiently close to their published atomic positions, with an average distance  $\langle d \rangle = 0.50 \text{ \AA}$ . The experimental data resolution is equal to  $2.18 \text{ \AA}$ . The flexibility of the molecule model may be described by seven torsion angles showed in Figure 1. In addition a chlorine atom is expected to be in the solvent region, isolated from to the rest of the molecule, so leading to 15 the total number of *DOF*'s. The highest peak of the Fourier map is assumed to be coincide with the *Cl* position: this allows to spare 6 degrees of freedom in the search. In 228 minutes all the procedure is executed and all the atoms in the asymmetric unit (30) are found close to the published atomic coordinates with a average distance  $\langle d \rangle = 0.13 \text{ \AA}$ .



**Figure 1:** Molecular model of DILTIA structure. The torsion angles considered as internal degrees of freedom are indicated by arrows.

### Example 2: structure AND1N [22]

At the end of *DM* step (executed on neutron data) *EXPO2004* generates an electron density map from which the six highest peaks are selected and used as pivot of the *SA* procedure. An *a posteriori* analysis shows that only 3 of them are correctly found close to the published atomic coordinates, with an average distance  $\langle d \rangle = 0.31 \text{ \AA}$ . The experimental data resolution is equal to  $1.47 \text{ \AA}$ . The flexibility of the model is described by six torsion angles indicated in Figure 2. 33 minutes are sufficient to correctly locate all the 25 atoms of the model in the asymmetric unit. Their average distance from the corresponding atoms in the published crystal structure is  $0.17 \text{ \AA}$ .



**Figure 2:** Molecular model of AND1N structure. The torsion angles considered as internal degrees of freedom are indicated by arrows.

### Conclusions

The combination of the information contained in the electron density maps provided by *Direct Methods* with the *SA* technique, has been developed and introduced in a new version of *EXPO2004*. The procedure is characterized by: a) a cost function able to

select the correct solutions among the false ones using single or pairs of overlapping reflections. b) an efficient search procedure capable of recognizing the global minimum against numerous local minima. c) techniques of local minimization and of least squares, able to improve the parameter values provided by the *SA*.

The procedure has been applied with success to a large number of organic structures with variable number of *DOF*'s.

### References

- [1] Chernyshev, V.V. & Schenk, H. (1998). *Z. Kristallogr.* **213**, 1-3.
- [2] Goldberg, D.E. in "Genetic Algorithms in Search, Optimization, and Machine Learning" ed. Addison-Wesley, New York, 1989.
- [3] Kariuki, B.M., Serrano-González, H., Johnston, R.L., Harris, K.D.M. (1997). *Chem. Phys. Lett.*, **280**, 189-195.
- [4] Shankland, K., David, W.I.F., Csoka, T. (1997). *Z. Krist.*, **212**, 550-552.
- [5] Harris, K.D.M., Tremayne, M., Lightfoot, P., Bruce, P.G. (1994). *J. Am. Chem. Soc.*, **116**, 3543-3547.
- [6] Andreev, Y.G., Lightfoot, P. & Bruce, P.G. (1997). *J. Appl. Cryst.*, **30**, 294-305.
- [7] Tremayne, M., Kariuki, B.M., Harris, K.D.M., Shankland, K., & Knight, K.S. (1997). *J. Appl. Cryst.*, **30**, 968-974.
- [8] Kirkpatrick, S. (1983). *J. Stat. Phys.*, **34**, 975-986.
- [9] David, W.I.F., Shankland, K., Shankland, N. (1998). *Chem. Commun.*, 931-932.
- [10] David, W.I.F., Shankland, K., Cole J., Maginn, S., Motherwell, W.D.S., Taylor, R. *DASH User Manual.*, (2001). Cambridge Crystallographic Data Centre, Cambridge, UK.
- [11] Coelho, A.A. *TOPAS* Version 3.1, (2003). Bruker AXS GmbH, Karlsruhe, Germany.
- [12] Favre-Nicolin, V., & Černý, R. (2002). *J. Appl. Cryst.*, **35**, 734-743.
- [13] Andreev, Y.G., Bruce, P.G. (1998). *J. Chem. Soc. Dalton Trans.*, 4071-4080.
- [14] Johnston, J.C., David, W.I.F., Markvardsen, A.J., Shankland, K. (2002). *Acta Cryst.*, **A58**, 441-447.
- [15] Brenner, S., *PhD Thesis*, "Structure Envelopes and their Application in Structure Determination from Powder Diffraction Data" 1999, ETH, Zürich, Switzerland.
- [16] A., Altomare, R., Caliandro, M., Camalli, C., Cuocci, Giacovazzo C., Moliterni A.G.G., Rizzi, R. (2004). *J. Appl. Cryst.*, **37**, 1025-1028.
- [17] A. Leach in "Molecular modelling, principles and applications", ed. Addison Wesley Longman Limited, London, 1996.
- [18] A. Le Bail, H. Duray, J. L. Fourquet, *Math. Res. Bull.*, 1988, **23**, 447.
- [19] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, *J. Chem. Phys.* 1978, **21**, 1087.
- [20] A., Altomare, C., Cuocci, C., Giacovazzo, A.G.G., Moliterni, R., Rizzi, (2006). *J. Appl. Cryst.* **39**, 558

- [21] A.J., Florence, N., Shankland, K., Shankland, W.I.F., David, E., Pidcock, X., Xu, A., Johnston, A.R., Kennedy, P.J., Cox, J.S.O., Evans, G., Steele, S.D. Cosgrove, C.S. Frampton, (2005). *J. Appl. Cryst.*, **38**, 249.
- [22] V.V., Chernyshev, A.N., Fitch, E.J., Sonneveld, A.I., Kurbakov, V.A., Makarov, V.A., Tafenko, (1999). *Acta Cryst.*, **B55**, 554.

---

## PROGRESS IN SOLVING CRYSTAL STRUCTURES OF TRIGLYCERIDES FROM POWDER DATA USING DIRECT-SPACE TECHNIQUES

J.B. van Mechelen, R. Peschar\*,  
H. Schenk

Universiteit van Amsterdam, van 't Hoff Institute for Molecular Sciences, Laboratory for Crystallography, Valckenierstraat 65, 1018 XE Amsterdam, The Netherlands

\*Contact author; e-mail: Rene@science.uva.nl

### Introduction

Food and personal care products often contain a solid fat phase that should provide firmness and stability to the product. The fat phase itself may consist of many (co-crystallized) triglycerides (TAGs) and the latter are not always in their most stable phase. A well-known example of an unwanted polymorphic phase transition is fat bloom, a grey-whitish layer that can develop on chocolate when kept at too high temperatures [1]. Commonly, fat bloom is attributed to a transition from the highest but one melting phase ( $\beta_2 = \beta - V$ ) to the highest melting phase ( $\beta_1 = \beta - VI$ ) [2]. It has been hypothesized that (re-) packing of the long fatty-acid acyl chains and/or layers may be involved but to get supportive evidence crystal structure models are indispensable.

Monounsaturated TAGs with an oleic chain at the glycerol-2 position are the major constituents of the fat phase in cocoa butter. These materials do not crystallize easily to sizes suitable for single-crystal structure analysis so structure determination using powder diffraction seems to be the sole alternative. Powder data of TAGs are usually of relative low resolution with no significant diffraction beyond  $\sim 3$  Å. The lack of atomic resolution together with the amount of non-hydrogen atoms per molecule,  $\sim 60$  for the major TAG fractions POS, POP and SOS in cocoa butter, excludes the use of single-crystal based structure-determination techniques that rely on the extraction of individual intensities like the Patterson method, direct methods and the maximum entropy and likelihood-

ranking method. The recently developed so-called "direct-space" methods that rely on finding a global minimum of the R-factor are better suited for this problem. In this contribution, we will limit ourselves to some recent results [3,4] obtained with the program FOX [5], one of the many direct-space programs available nowadays.

### Structure determination strategy

In the process of crystal structure determination using powder data at least six stages can be discerned, (i) Sample preparation, (ii) Data collection, (iii) Pattern fitting, (iv) Unit cell determination (indexing), (v) Structure determination, i.e. finding the approximate positioning and conformation of the molecule(s) in the unit cell, and (vi) Refinement of the approximate model found in (v). Although for TAGs all stages can be critical, the latter three are the most serious bottlenecks and therefore we will focus on them.

### Real-space indexing

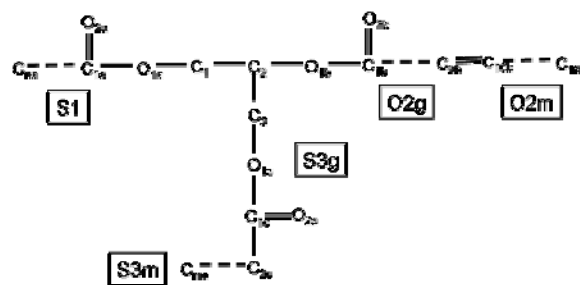
Powder diffraction patterns of TAGs are somewhat atypical compared to those of other organic compounds because of the domination by two sets of peaks. At a low-angles only peaks are found that belong to the same reciprocal lattice line (d-spacings of  $\sim 8 - 65$  Å). A second set of overlapping peaks is found at  $3.5 - 6$  Å, the so-called fingerprint area. In our experience the currently available standard autoindexing programs, like Treor, Dicvol, ITO and McMaille, almost never succeed in indexing such patterns because of the (preset) assumptions and fixed settings in the indexing algorithms. Therefore, we developed the real-space indexing routine LSQDETC that runs in conjunction with the program suite POWSIM [6]. LSQDETC is a brute-force program that searches through an allowed solution space for cells. The solution space is limited by setting user-definable ranges for the unit-cell parameters and unit-cell volumes, crystal system etc, so prior knowledge (e.g density of the material) can be exploited. For each cell a powder diffraction pattern is calculated and the cells are ranked according criteria (e.g.  $M_{20}$ ) that compare the observed set of  $2\theta$  values with the calculated ones. The correctness of a final indexing is checked with the program Chekcell [7]. The unit cells of all monounsaturated TAGs established to date are all characterized by having a short axis ( $\sim 5.44$  Å) as well as a very long axis ( $\sim 120 - 135$  Å), depending on the precise length of the acyl carbon chains (see Table 1 for an example), and this may well explain the lack of success of the auto-indexing approach.

	$\beta_1$ -SOA	$\beta_2$ -SOA
Instrument	BM01B	X'pert Pro
$\lambda$ [Å]	0.85019	1.54059
a [Å]	5.4393	5.437
b [Å]	134.645	135.29
c [Å]	8.199	8.213
$\beta$ [°]	88.735	88.64
V [Å <sup>3</sup> ]	6003.7	6040.1
Rp	6.97	6.31
Rwp	8.79	8.45
GOF	3.45	2.7
$\rho$ (g/cm <sup>3</sup> )	1.015	1.015
Form.	C <sub>59</sub> H <sub>112</sub> O <sub>6</sub>	C <sub>59</sub> H <sub>112</sub> O <sub>6</sub>

**Table 1:** Indexing and refinement results of  $\beta_1$ -SOA and  $\beta_2$ -SOA.

### Real-space structure determination

In direct-space methods many trial crystal structures are generated in direct space by changing a small set of structural parameters, like position and orientation of the whole molecule and selection of torsion angles. For each trial structure a powder pattern is calculated and compared with the observed diffraction pattern. The process of trial-structure generation continues until an acceptable match is found, usually taken as a low-enough R-factor. The computation time obviously increases rapidly with the number of structural parameters that is allowed to change. A reduction of computing time can be realized exploiting the molecule description in FOX. However, a Z-matrix molecular description is most suited when selection and control of torsion angles is essential. A starting model of a TAG typically consist of two parallel stearyl chains (S1, S3) and an oleoyl chain (O2) pointing in the opposite direction (Figure 1). The oleoyl chain is split in two saturated partial chains, O2g (bonded to the glycerol) and O2m (methyl side) that are connected by a double bond at carbons C9b and C10b. In addition to translation and rotation of this rigid starting model, torsion angles at the glycerol and at the C9b=C10b double bond are released one by one, starting at the latter. In this process the saturated (parts of the) acyl chains are treated as rigid bodies. Hydrogen atoms are not included at this stage.



**Figure 1:** Chemical structure diagram of SatOSat'-type TAGs. The numerical subscripts  $m$  and  $n$  ( $n = 14, 16, 18, 20$ ) label the atom numbers while the S1, S3g, S3m, O2g and O2m label the saturated acyl chain and acyl-chain parts.

### Rietveld refinement

Once an acceptable model has been found, the model is refined with GSAS [8] using a Chebyshev polynomial to fit the background and profile number 4 to describe the peak profiles. With this function low-angle peak asymmetry and HKL-dependent broadening can be modelled successfully. It turned out useful to split the S3 chain in two parts: S3g from C2 – C2c, from the glycerol group to the gauche band, and S3m from C2c to the end of the acyl chain. Soft planar restraints were applied to the C atoms of the acyl chains S1, O2g, O2m and S3m, to the three C=O groups including the C-atoms connected to them and to the C8b-C9b=C10b-C11b group. During the refinement, the weight of the restraints was reduced but its level was kept high enough to ensure a stable refinement. In the final stage of the successful refinements the contribution of the restraints to the total of the  $\chi^2$  residue parameter could be reduced to about 10%.

### Influence of resolution

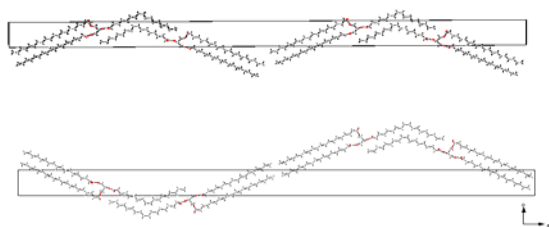
During the structure solution process, it appeared that the precise orientation of the O2m chain direction relative to those of the S1, the S3 and the O2g chains, yields two types of TAG packing; a 'flat' conformation and a 'rotated' conformation. In the former the direction of the O2m chain is in the same plane as the direction of the other three chains, whilst in the 'rotated' conformation the direction of the O2m is rotated out of this plane. FOX did not discriminate between these different conformations that imply different interconnections of the columns of electron density of the S1 and S3 chains to the glycerol. Only during structure refinement in GSAS, with HKL-dependent line broadening properly taken into account, it became clear that the refinement of the flat conformation converged better. In additional refinement experiments was found that the R-factor is hardly affected by rotation of the S3m chain plane relative to the S1 chain plane; they were defined as rigid bodies in GSAS and rotated relative to each other. Apparently, this minimization involves a shallow R-value valley, and with some intervention, a lower point



in the valley may be found. Peak overlap facilitates the redistribution of intensities within the envelope of the diffraction pattern after a forced rotation and limits the precision of, for example, the rotation angle of the zig-zag planes around their length axis. These problems to find the most probable conformation of the molecule can be attributed mainly to the limited crystallinity of the materials and their anomalous packing habit. The limited and anisotropic crystallite sizes cause hkl-dependent peak broadening and severe overlap in the fingerprint area, and even when using high-resolution synchrotron radiation, the materials do not diffract significantly beyond the fingerprint area (d-spacings 3 - 6 Å). As a result, in none of the cases atomic resolution is obtained. In most cases data have been collected up to atomic resolution, and it may be argued that, these data should be included in the refinement in spite of lacking significant diffraction signal. After having carried out several refinement tests, the results refute this hypothesis. For example,  $\beta_1$ -POS was refined initially with diffraction data from 0.53 - 20 °2 $\theta$ . Later on, the angle range was extended by 10° to bring it in line with those of the other samples and to test the influence of these higher-angle data that lack significant diffraction signal. The extension hardly affected the refined crystal structure and the sole effect was a slight reduction of the R-values. Thus, inclusion of high-angle diffraction data without any significant peak information does not necessarily have a positive influence on the structure refinement, as may be suggested by the somewhat lower R-values.

## The TAG crystal structures

The molecules in both polymorphs have the expected conformation [2,9,10] with the two saturated chains parallel and the mono unsaturated oleic chain pointing in the opposite direction. When viewed along the short axis (figure 2), the packing of the molecules consists of a sequence of six acyl-chain layers or, more precisely, of two sets of tri-acyl chain layers, from now on referred to as three-packs, which are symmetry related.



**Figure 2:** Molecule packing in  $\beta_2$ -SOA (top) and  $\beta_1$ -SOA (bottom), view along the *a*-axis.

Each three-pack consists of an unsaturated zone, in which the oleic chains are packed parallel and side-by-side, that is sandwiched by two saturated zones. Interestingly, as early as 1946 Filer et al. [11] already proposed the existence of a six-acyl chain packing with

saturated and unsaturated chain zones. The three-packs are identical in  $\beta_1$  and the  $\beta_2$  polymorph. The difference between the two polymorphs is in the symmetry relation between the three-packs. In  $\beta_1$  the three-packs are related by an inversion centre whereas in  $\beta_2$  they are related by a translation ( $\frac{1}{2}, \frac{1}{2}, 0$ ). This relation gives a simple explanation of the mechanism of the formation of fat bloom on chocolate. The underlying phase transition from  $\beta_2$  (=  $\beta$ -V) to  $\beta_1$  (=  $\beta$ -VI) is just a rotation of 180° of one three-pack relative to the other one along the longest axis of the unit cell as can be easily seen in figure 2. A previous model of the  $\beta_2$  structure [12] does not allow such a simple mechanistic explanation. This makes this model less plausible in spite of its low R-value. In retrospect, low resolution of the data and assumed preferred orientation masked the inadequacy of this previous model.

## Conclusions

The structure determination results of the  $\beta_2$  and  $\beta_1$  polymorphs compared to a previously solved  $\beta_2$  polymorph point out that structural imperfections are easily masked when data lack atomic resolution or when incorrect assumptions are made with respect to physical parameters. In this respect it should be pointed out that in the present  $\beta_2$  and  $\beta_1$  models no significant preferred orientation is present, in contrast to what was found earlier [12]. The three-pack built-up of both the  $\beta_2$  and  $\beta_1$  polymorphs provides a simple explanation for the  $\beta_2$  to  $\beta_1$  phase transition in terms of migration of  $\beta_2$  three-packs to the surface, melting and (re-)crystallization in a  $\beta_1$  packing.

## Acknowledgements

The authors thank ADM Cocoa NL, Unilever Research Vlaardingen and Unilever Research Colworth for providing samples. The ESRF (Grenoble, France) is thanked for providing the facilities to perform the synchrotron diffraction experiments and the staff of beamlines BM01b (D. Testemale, H. Emerich, and W. van Beek) and ID31 (I. Margiolaki) for their valuable help during the experimental sessions. Dr. V. Favre-Nicolin is gratefully acknowledged to implement requested modifications in FOX. The investigations have been supported by the Netherlands Foundation for Chemical Research (NWO/CW) with financial aid from the Netherlands Technology Foundation (STW) (project 790.35.405). The members of the User Committee of this project are thanked for stimulations discussions and continuous interest.

## References

- [1] Beckett, S.T., 1999, Industrial chocolate manufacture and use, (Oxford: Blackwell Science).
- [2] Sato, K. & Ueno, S., 2001, in Crystallization processes in fats and lipid systems, edited by N. Garti & K. Sato (New York: Dekker), pp. 177-209.

- [3] Van Mechelen, J.B., Peschar, R. & Schenk, H., 2006, *Acta Cryst. B*, 62, 1121.
- [4] Van Mechelen, J.B., Peschar, R. & Schenk, H., 2006, *Acta Cryst. B*, 62, 1131.
- [5] Favre-Nicolin, V. & Černý, R., 2002, *J. Appl. Cryst.* 35, 734.
- [6] Peschar, R., Etz, A., Jansen, J. & Schenk, H., 2002, in *Structure determination from powder diffraction data*, 2002, edited by W.I.F. David, K.Shankland, L.B.McCusker & Ch. Baerlocher, (Oxford: Oxford Univ.Press), pp. 179-189.
- [7] Laugier, J. & Bochu, B., 2001, Chekcell; <http://www.inpg.fr/LMPG>
- [8] Larson, A.C. & Von Dreele, R.B. GSAS, 1987, *General Structure Analysis System*, Los Alamos National Laboratory: Report No. LA-UR-86-748.
- [9] De Jong, S., Van Soest, T.C. & Van Schaick, M.A. J., 1991, *Am. Oil Chem. Soc.*, 68, 371.
- [10] Larsson, K., 1972, *Fette Seifen Anstrichm.*, 74, 136.
- [11] Filer, L.J. Jr., Sidhu, S.S., Daubert, B.F. & Longenecker, H.E., 1946, *J. of Am. Chem. Soc.*, 68, 167.
- [12] Peschar, R., Pop, M.M., De Ridder, D.J.A., Van Mechelen, J.B., Driessen, R.A.J. & Schenk, H., 2004, *J. Phys. Chem. B.*, 108, 15450.

## Calendar of Events (2007)

30 April-4 May 2007

### **Practical X-ray Fluorescence.**

Pennsylvania, USA.

<http://www.icdd.com/education/xrf.htm>

10-11 May 2007

### **GSAS-II Rietveld Software Development Workshop.**

Argonne National Laboratory, USA.

<http://www.iucr.org/cww-top/mtg.gsas.html>

16-18 May 2007

### **Seventh Canadian Powder Diffraction Workshop.**

Québec, Canada.

<http://www.cins.ca/cpdw>

4-8 June 2007

### **Fundamentals of X-ray Powder Diffraction**

Pennsylvania, USA.

<http://www.icdd.com/education/xrd.htm>

7-17 June 2007.

### **Int'l School of Crystallography 39th Course: Engineering of Crystalline Materials Properties: State of the Art in Modeling Design and Applications.**

Erice, Italy.

<http://www.crystalice.org/futuremeet.htm>

11-13 June 2007

### **2nd TOPAS Users' Meeting.**

Karlsruhe, Germany.

<http://www.bruker-axs.de>

11-15 June 2007

### **Advanced Methods in X-ray Powder Diffraction.**

Pennsylvania, USA.

<http://www.icdd.com/education/xrd.htm>

15-20 July 2007

### **International School on Mathematical and Theoretical Crystallography.**

The University of Havana, Cuba.

<http://www.cristalografia.net/havana2007/>

21-27, July 2007

### **ACA Annual Meeting**

Salt Lake City, Utah, USA.

<http://www.amerystalassn.org>

29 July-8 August 2007

### **Small-Molecule Crystallography**

Summer School San Diego, USA.

[http://chem-](http://chem-tech.ucsd.edu/Recharges/SMXF/crystalschool.html)

[tech.ucsd.edu/Recharges/SMXF/crystalschool.html](http://chem-tech.ucsd.edu/Recharges/SMXF/crystalschool.html)

22-27 August 2007

### **24<sup>th</sup> European Crystallographic Meeting**

Marrakesh, Morocco.

<http://www.ecm24.org>

26-28 September 2007

### **Non-ambient X-ray powder diffraction workshop.**

Mülheim, Germany.

<http://www.mpi-muelheim.mpg.de/xray>

7-9 October 2007

### **Size-Strain V - Diffraction Analysis of the Microstructure of Materials.**

Garmisch-Partenkirchen, Germany.

<http://www.mf.mpg.de/en/abteilungen/mittemeijer/ss-v/index.htm>

## Calendar of Events (2008)

27 April - 3 May 2008

### **Summer School on Mathematical and Theoretical Crystallography.**

Gargnano, Garda Lake, Italy.

[http://www.lcm3b.uhp-](http://www.lcm3b.uhp-nancy.fr/mathcryst/gargnano2008.htm)

[nancy.fr/mathcryst/gargnano2008.htm](http://www.lcm3b.uhp-nancy.fr/mathcryst/gargnano2008.htm)

31 May-5 June 2008

### **Annual Meeting of the American Crystallographic Association 2008**

Knoxville, TN, USA.

[http://www.amerystalassn.org/meetingspg\\_list/futuremeetings.html](http://www.amerystalassn.org/meetingspg_list/futuremeetings.html)

9-14 June 2008

### **ICQ10-10<sup>th</sup> International Conference on Quasicrystals.**

Zurich, Switzerland. <http://icq10.ethz.ch/>

7-11 July 2008

### **10th EMU School. High-resolution electron microscopy of minerals.**

Nancy, France.

<http://www.lcm3b.uhp-nancy.fr/emu10>

21-26 July 2008.

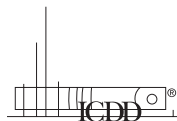
### **XRM2008. 9th International Conference on X-ray Microscopy.**

ETH Zurich, Switzerland.

<http://xrm2008.web.psi.ch/>

# News from the International Centre for Diffraction Data (ICDD)

12 Campus Boulevard  
Newtown Square, PA 19073-3273, U.S.A.  
Phone: +610.325.9814  
Fax: +610.325.9823



www.icdd.com  
www.dxcicdd.com  
E-mail: info@icdd.com

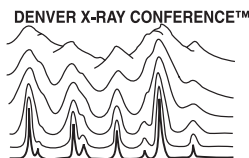
## Educational Events

In pursuing its commitment to the education of scientists in the fields of X-ray analysis, the ICDD offers an extensive program of learning opportunities.

- **PPXRD-6: The Sixth Pharmaceutical Powder Diffraction Symposium** returned to Barcelona, Spain, 20–22 February 2007. This forum brought together scientists working in the XRPD and pharmaceutical fields. An optional workshop was offered on 19 February, focusing on **Exercises in Quantitative Phase Identification**.

Mark your calendar to attend **PPXRD-7**, scheduled to be held in Orlando, Florida, 25–28 February 2008. Stay informed: <http://www.icdd.com/ppxrd/>

- **56<sup>th</sup> Annual Denver X-ray Conference** will be held in Colorado Springs, Colorado, 30 July–3 August 2007. Whether you're a beginner or an expert, the Denver X-ray Conference offers incentives for everyone. From workshops and sessions, to an exhibition by the major X-ray analysis vendors, you'll benefit from the opportunities presented at the Denver X-ray Conference. ([www.dxcicdd.com](http://www.dxcicdd.com))



This year's **Plenary Session, Stardust—X-rays in Space**, will venture into the fascinating world of outer space and how scientists are gaining a better understanding of our universe through X-ray analysis techniques. A special related session is also scheduled, where invited speakers will present their findings for the materials collected from the NASA Stardust mission. Don't miss this session!

- **Specialized Workshops:** For the first time in the fall of 2006, ICDD sponsored two specialized workshops at its headquarters facility. Based on the positive feedback, both workshops will become a permanent addition to our course offerings. Watch for these events in the fall of 2007:

—Rietveld Refinement and Indexing,  
24–26 September 2007

—Specimen Preparation for X-ray Fluorescence,  
2–4 October 2007

Want to stay informed? Contact [clinics@icdd.com](mailto:clinics@icdd.com) to be added to our mailing list.

- **ICDD X-ray Clinics:** Held at our headquarters in Newtown Square, PA, our annual week-long clinics in X-ray powder diffraction and X-ray fluorescence spectrometry provide opportunities to receive training from professionals with many years of diverse experiences. A limited number of tuition waivers are offered to academic applicants who satisfy the criteria. Learn more at:

<http://www.icdd.com/education/clinics.htm>.

Here are the 2008 dates:

—Practical X-ray Fluorescence,  
28 April–2 May 2008

—Fundamentals of X-ray Powder Diffraction,  
2–6 June 2008

—Advanced Methods in Powder Diffraction,  
9–13 June 2008



## Awards

**Dr. Victor E. Buhrke** of Portola Valley, CA, was named the most recent recipient of the **Jenkins Award** for his 50-year career in X-ray materials analysis, contributing

uniquely to both X-ray fluorescence and X-ray diffraction. Dr. Buhrke's work includes research, applications, training, consulting, and management in these disciplines. He is known throughout the community for his landmark book on Specimen Preparation. A long-time member of the



L to R: John Gilfrich, Phyllis Jenkins,  
Victor Buhrke, and Janet Buhrke.



Denver X-ray Conference Organizing Committee, Dr. Buhrke presently serves as its Chairman.

**Dr. Peter Wobrauschek**, Atominstytut, of the Vienna University of Technology, Vienna, Austria, was awarded the 2006 **Birks Award**. Dr. Wobrauschek was recognized for his influential work in the development of total reflection X-ray fluorescence, now used worldwide in the semiconductor industry but applicable to many other subfields of X-ray analysis. His many publications and presentations on the subject have inspired many applications and the production of a myriad of TXRF instruments.

The 2006 **Jerome B. Cohen Award** was shared by two students. **Hanfei Yan**, Columbia University, New York, NY, and Argonne National Laboratory, Argonne, IL, won the award for his work “Dynamical Artifacts in X-ray Diffraction from Single Crystals”; **Wanchuck Woo**, The University of Tennessee, Knoxville, TN, and Oak Ridge National Laboratory, Oak Ridge, TN, won the award for his work “In-Situ Time-Resolved Neutron Diffraction Measurement of Transient Material States during a Thermo-Mechanical Process Based on Quasi-Steady State Principle.”

Named as the 2006 **Distinguished Fellow**, **Dr. Camden R. Hubbard**'s extensive contributions to the ICDD span over thirty years. Most recently, he served as Chair of the ICDD Board of Directors where he guided the organization into the 21st century with vision, confidence, and exemplary leadership. During his tenure, the ICDD expanded the Powder Diffraction File (PDF) to a state-of-the-art database, resulting in products that provide new data mining tools for structural crystallography and materials characterization.

**Dr. Peter Wallace**, Arroyos Enterprises, Oro Valley, AZ, was named the 2006 **McMurdie Award** recipient. Dr. Wallace has a distinguished history of contributions to the metals and alloys content of the Powder Diffraction File. His contributions date back to the first Metals & Alloys book, which was printed in 1977. He has long been, and continues to be, a member of the very active Metals and Alloys Subcommittee and was Chairman of that subcommittee for various periods.

As an ICDD **Fellow**, **John Faber**, Principal Scientist, ICDD, Newtown Square, PA, was recognized for his leadership in directing his Science Team at ICDD to develop the necessary housing to accommodate the dramatic expansion of the PDF database, which increased 380% in size since 2001, and provided it with viewing, sorting, and mining capabilities. In addition, Dr. Faber's educational efforts were recognized; they include developing the curriculums and instructing at the ICDD clinics, pharmaceutical symposia, and Denver X-ray Conference workshops. He recently retired from the ICDD, but continues to support the organization and its membership as a consultant, coordinating global educational opportunities as Chairman of the XRD Clinics and Symposia.

Also named as a **Fellow**, **Earle Ryba**, Assistant Associate Professor of Metallurgy, The Pennsylvania State University (PSU), University Park, PA was recognized for his teaching excellence and many contributions to education in the fields of materials and X-ray analysis, both at PSU and beyond. Dr. Ryba has been actively involved in the XRD X-ray Clinics since 1983, when they were hosted at SUNYA. He assumed a more active role when the clinics moved to PSU, and remains a primary leader in the clinics as they are currently hosted at ICDD Headquarters.

Nine others were awarded the title of ICDD **Fellow**:

- ◆ **Davor Balzar** from NIST, Boulder, CO and the University of Denver, Littleton, CO for his contribution as chair of the X-ray Diffraction Methods Subcommittee.
- ◆ **Xiaolong Chen** of the Chinese Academy of Sciences, People's Republic of China, for his efforts as Regional Co-chair for China, a position that he continues to maintain.
- ◆ **Peter Lee** from Argonne National Laboratory, Argonne, IL for his contribution as chair of the Synchrotron Diffraction Subcommittee.
- ◆ **Bill Mayo** from H & M Analytical Services, Allentown, NJ for his contribution as chair of the Grant-in-Aid Committee.
- ◆ **Andrew Payzant** from Oak Ridge National Laboratory, Oak Ridge, TN for his contribution as chair of the Non-Ambient Diffraction Subcommittee.
- ◆ **Paolo Scardi** from University di Trento, Mesiano, Trento, IT for his contribution as Vice Chairman of the Board of Directors.
- ◆ **Carlo Segre** from Illinois Institute of Technology, Chicago, IL for his contribution as chair of the Education Subcommittee.
- ◆ **Brian Toby** from Argonne National Labs, Advanced Photon Source, Argonne, IL for his contribution as chair of the Database Subcommittee.
- ◆ **Fred Wireko** from Procter & Gamble Co., Cincinnati, OH for his contribution as chair of the Organic and Pharmaceutical Methods Subcommittee.

**Sergei Kirik**, from the Institute of Chemistry and Chemical Technology, Russian Academy of Sciences received the 2007 **Distinguished Grantee Award**. Dr. Kirik was introduced to the ICDD Grant-in-Aid Program in 1993. Since then, he has contributed over 600 high-quality patterns to the Powder Diffraction File. His primary research interests include inorganic materials: complex compounds, binary, ternary oxides system, super and ionic conductors, ceramics and mesostructured materials.

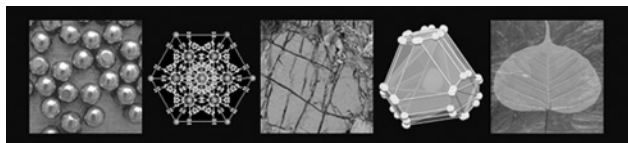




## ICDD Product Highlights

### PDF-4+ 2006

*More data...more capability...  
more analysis power....with PDF-4+!*



PDF-4+ is an advanced database combining the world's largest sources of inorganic diffraction data from crystals and powders into a single database. The result is a comprehensive collection of inorganic materials, produced in a standardized format that can be rapidly searched for unknown identification. This database contains numerous features such as digitized patterns, molecular graphics and atomic parameters. Many new features have been incorporated into PDF-4+ to enhance the ability to do quantitative analysis by any of three methods: Rietveld Analysis, Reference Intensity Ratio (RIR), Method or Total Pattern Analysis. PDF-4+ 2006 contains 254,873 entries while PDF-4/Minerals 2006 contains 19,254 entries. PDF-4+ contains more data and more types of data, enabling rapid materials identification. It is designed to support automated quantitative analyses by providing key reference data required for these analyses. Every pattern, independent of type or source, can be displayed as a digitized pattern, providing the ability to perform total pattern analysis. In 2006, we continued to enhance identification and quantitation by adding more data with  $I/I_c$ 's and atomic coordinates. Java™ interfaces have improved speed and convenience, which enabled several new display features for digital patterns and data mining.

PDF-4+ 2006 contains:

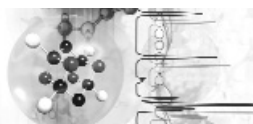
- ❖ 254,873 Material Data Sets, edited and standardized for rapid identification
- ❖ 165,923 Material Data Sets with  $I/I_c$
- ❖ 98,291 Material Data Sets with atomic parameters
- ❖ Digital pattern simulations for crystallite size and total pattern analysis

PDF-4+ is an annual renewal license product.

PDF-2 licensed end-users can upgrade and convert their license to PDF-4+ at low conversion pricing.

### PDF-4/Organics 2007

PDF-4/Organics 2007, the world's largest X-ray powder diffraction database for organics, contains 312,355 entries resulting from ICDD's collaboration with the Cambridge Crystallographic Data Centre (CCDC). This product has all the display software and data mining capabilities contained in the PDF-4 family of products. PDF-4/Organics is a practical results-oriented



product that combines the drug active compounds calculated from the Cambridge Structural Database (CSD) with the polymers (including starches and celluloses), excipients and pharmaceuticals in ICDD's Powder Diffraction File. This database is designed for rapid materials identification targeted for the pharmaceutical and specialty chemical industries. Its design allows for easy interface with diffractometers and the data analysis systems of the world's leading software developers and manufacturers of X-ray equipment. The database is useful for scientists working in consumer products, catalysis, forensic science, analytical labs, drug discovery and production.

New features include a user friendly Java™ interface for rapid data mining, new additions to the excipients and pharmaceutical subfiles, and the addition of 25,891 new entries.

PDF-4/Organics 2007 contains:

- ❖ 312,355 Material Data Sets, edited and standardized for rapid identification
- ❖ 30 Subfiles including excipients, pharmaceuticals and forensics
- ❖  $I/I_c$  values for quantitative analysis
- ❖ Digital pattern simulations for crystallite size and total pattern analysis
- ❖ An extensive compilation of known polymorphs, hydrates and solvates for commercial drugs

### PDF-4 Support Software

DDView+, ICDD's viewing software, is integrated into the PDF-4 product line.

DDView+ provides more display features, search and data filtering capabilities for the PDF-4 format of the



Powder Diffraction File. The program contains 44 different search methods to filter the database contents with custom display of a selection of 65 separate data fields. Using a Java™ point and click interface, various search methods and field selections can be combined to produce a nearly limitless choice of data mining options. SIEVE+ provides rapid searching and identification of materials using highly automated Hanawalt, Fink and Long-8 searches. While SIEVE+ contains much of the functionality of ICDD's printed Index and Search Manuals, its strengths take advantage of the dynamic computing power of modern PCs to rapidly perform permuted searches on hundreds of thousands of data entries.

### Further Information

To learn more about the ICDD, its products and services, please visit our web sites:  
[www.icdd.com](http://www.icdd.com) and [www.dxcicdd.com](http://www.dxcicdd.com).



**1st Meeting of  
the Italian and Spanish  
Crystallographic Associations  
(MISCA)**

*Hosted by the University of Calabria  
(Dipartimento di Chimica)*



**Villaggio Guglielmo, Copanello di Staitelli  
(Italy)**

*September 24-28, 2007*

UNIVERSITA' DELLA CALABRIA



Dipartimento di Chimica



**PROGRAM COMMITTEE**

Giovanni DE MUNNO  
(Chairman) University of Calabria  
Santiago GARCIA GRANDA  
(GE3C President) University of Oviedo

Roberta OBERTI  
(AIC President) IGG-CNR, Pavia

Angela ALTOMARE  
IC-CNR, Bari

Giuseppe FILIPPINI  
ISTM-CNR, Milano

Pilar GÓMEZ SAL  
University of Alcalá de Henares, Madrid

Fernando J. LAHOZ  
University of Zaragoza-CSIC

Victoria LOPEZ-ACEVEDO CORNEJO  
Complutense University of Madrid  
Marcello MELLINI  
University of Siena

Hugo Louis MONACO  
University of Verona

Catalina RUIZ PEREZ  
University of La Laguna, Tenerife

**ORGANIZING COMMITTEE**

Donatella ARMENTANO

Anna BELLUSCI

Alessandra CRISPINI

Giovanni DE MUNNO

Teresa F. MASTROIPIETRO

**MEETING SECRETARIAT**

Alessandra CRISPINI

Department of Chemistry

University of Calabria

Phone: +390984492888

Fax: +390984492066

Home page: <http://misca2007.unical.it>  
E-mail: [misca2007@unical.it](mailto:misca2007@unical.it)

**PRESENTATION of MISCA**

The Associazione Italiana di Cristallografia (AIC) and the Grupo Especializado de Cristalografía y Crecimiento Cristalino (GE3C) have agreed to join their respective annual meetings in the year 2007. Recent advancements and frontiers in the investigation of structure and properties of crystalline materials constitute the core of the scientific program, which consists of the following microsymposia:

- **General crystallography (MS1)**
- **Advanced functional properties of crystal engineered molecular structures (MS2)**
- **Crystallography in molecular medicine (MS3)**
- **Molecular interactions (MS4)**
- **Crystal Growth in Earth and Material Sciences (MS5)**
- **Dreams and nightmares of the electron crystallographers (MS6)**
- **Synchrotron and neutron sources and instrumentation (MS7)**
- **Computational crystallography (MS8)**

**CONGRESS LOCATION**

MISCA 2007 will take place in Calabria (Italy), in the Guglielmo Village, a top class hotel resort on the ionic coast. This offers all the facilities to host an international scientific congress and the amenities for a most pleasant staying..

**INVITED SPEAKERS**

David R. ALLAN

Hector A. CALDERON

Angelo GAVEZZOTTI

Carmelo GIACOVAZZO

Miguel JULVE

Carlos PINA

Roger L. WILLIAMS