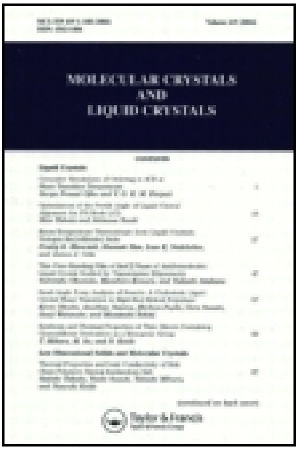

Molecular Crystals and Liquid Crystals Science and Technology. Section A. Molecular Crystals and Liquid Crystals

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gmcl19

A Simulated Annealing Approach for Crystal Structure Solution from Powder Diffraction Data

G. E. Engel [a], S. Wilke [a], K. D. M. Harris [b], B. M. Kariuki [b], S. Ahn [b], F. J. J. Leusen [a] & M. A. Neumann [a]

[a] Molecular Simulations Ltd., 230/250 The Quorum, Barnwell Road, Cambridge, CB5 8RE, United Kingdom

[b] School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

To cite this article: G. E. Engel, S. Wilke, K. D. M. Harris, B. M. Kariuki, S. Ahn, F. J. J. Leusen & M. A. Neumann (2001): A Simulated Annealing Approach for Crystal Structure Solution from Powder Diffraction Data, Molecular Crystals and Liquid Crystals Science and Technology. Section A. Molecular Crystals and Liquid Crystals, 356:1, 355-364

To link to this article: http://dx.doi.org/10.1080/10587250108023714

Mol. Cryst. and Liq. Cryst., 2001, Vol. 356, pp. 355-364
Reprints available directly from the publisher
Photocopying permitted by license only

A Simulated Annealing Approach for Crystal Structure Solution from Powder Diffraction Data

G.E. ENGEL[a], S. WILKE[a], K.D.M. HARRIS[b], B.M. KARIUKI[b],
S. AHN[b], F.J.J. LEUSEN[a] and M.A. NEUMANN[a]

[a]*Molecular Simulations Ltd., 230/250 The Quorum, Barnwell Road, Cambridge CB5 8RE, United Kingdom and* [b]*School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom*

(Received September 30, 1999)

We discuss fundamentals and applications of the program "Powder Solve" that has been developed to solve crystal structures directly from powder diffraction data. The program uses a combination of simulated annealing and rigid-body Rietveld refinement techniques to find the trial structure giving maximal agreement between calculated and experimental powder diffraction data (assessed using the profile R-factor R_{wp}). Two examples are highlighted to demonstrate the application of Powder Solve for molecular crystals defined by up to 18 variable degrees of freedom.

Keywords: powder diffraction; crystal structure solution; simulated annealing; molecular crystals

1 INTRODUCTION

X-ray and neutron diffraction data provide the most powerful tool for determination of crystal and molecular structures. If sufficiently large crystals of a material are available for single crystal diffraction experiments, powerful techniques (such as direct methods and Patterson methods) can be used to determine the crystal structure from the three-dimensional diffraction data. However, many materials are available only as polycrystalline powders. In powder diffraction data, diffraction maxima are effectively projected onto a single variable (the diffraction angle 2θ), and reconstruction of the crystal structure is therefore a much more difficult task than for single crystal diffraction data.

Nevertheless, powder diffraction data still contains a considerable amount of information. Thus, unit cell parameters can be obtained by indexing the powder diffraction pattern (using programs such as ITO [1], TREOR90 [2], DICVOL91 [3] and a newly developed Genetic Algorithm method [4]), and the space group can be identified from systematic absences. To determine the positions of the atoms in the unit cell (i.e. structure solution) from the powder diffraction data, two general strategies may be used: traditional approaches and direct-space approaches.

In the direct-space strategy for structure solution, molecular fragments are handled directly within the unit cell, with no requirement to use the intensities of individual reflections extracted from the powder diffraction pattern. The position, orientation and conformation of each molecular fragment is varied to generate "trial" crystal structures, until optimum agreement is achieved between calculated and experimental powder diffraction patterns. Within the direct-space philosophy, a number of different algorithms have been used: grid search [5–8], Genetic Algorithms [9–13] and Monte Carlo/simulated annealing [14–20].

We have developed a software package "Powder Solve" [21] employing the Monte Carlo/simulated annealing approach. Powder Solve is fully integrated within the Cerius^2 [22] molecular modelling environment. This method searches the R_{wp} hypersurface (which represents the quality of agreement between calculated and experimental powder diffraction patterns as a function of the degrees of freedom defining the structure) in order to find the structure corresponding to lowest R_{wp}. By using a figure of merit (R_{wp}) based on a full profile comparison between calculated and experimental powder diffraction data, Powder Solve uses the experimental data directly as measured and avoids ambiguities inherent in methods that rely on prior extraction of integrated intensities of reflections. Conveniently, data preparation, indexing, peak shape analysis, structure solution, Rietveld refinement and lattice energy calculations can all be carried out within the same package [22]. Potential structure solutions obtained from Powder Solve can be easily checked using solid state force-field or quantum mechanical calculations, thus providing additional information to validate the proposed structure solution.

2 DATA PREPARATION

In many cases, one of the most difficult aspects of the structure determination process is the determination of the unit cell parameters by indexing the powder diffraction pattern. In the present paper, we assume that the unit cell parameters have been obtained using a suitable indexing method [1–4].

In structure solution, a reliable comparison between calculated and experimental powder diffraction patterns requires that the parameters defining the peak positions (unit cell parameters and zero-point offset), the peak shape (peak width, peak shape function and peak mixing parameters) and the background intensity in the powder diffraction pattern calculated for each trial structure accurately reflect the experimental powder diffraction pattern. It is therefore important that reliable values of these parameters are established prior to the structure solution calculation. The procedures developed here for this purpose have been implemented in the program "Powder Fit". Powder Fit is based on a simple modification to the original Pawley procedure [23], which ensures that the method is very stable and well-conditioned even for strongly overlapping peaks (the original Pawley procedure introduces artificial constraints on the intensities of overlapping peaks in order to overcome problems of ill-conditioning). At present, Powder Fit allows the refinement of unit cell parameters, background coefficients, zero-point shift parameters and peak width parameters for an experimental powder diffraction pattern with no knowledge of the actual atomic positions within the crystal structure (and is thus appropriate prior to the use of a structural model in structure solution calculations). Several different profile functions are implemented, including standard Gaussian and Lorentzian functions, two modified Lorentzian functions, pseudo-Voigt, Pearson VII and modified Thompson-Cox-Hastings functions [24]. Where appropriate, mixing parameters for these profile functions and a simple asymmetry correction [25] can be refined.

3 STRUCTURE SOLUTION

After the unit cell parameters, peak shape parameters and background parameters relevant to the experimental powder diffraction pattern have been determined by Powder Fit, the structure solution procedure may be started. In this procedure, the relative intensities of reflections depend on the structural model, while the unit cell parameters, peak shape parameters and background parameters are fixed at the values established from Powder Fit. The Powder Solve package for structure solution employs a simulated annealing algorithm to determine the positions, orientations and intramolecular degrees of freedom (variable torsion angles) of the molecular fragments in the asymmetric unit.

The degrees of freedom are defined by selecting appropriate groups of atoms such that, within each group, the relative positions of atoms remain unchanged (rigid bodies) during the calculation. Each bond that links two of these rigid groups is defined as a flexible torsion angle. It is possible to define arbitrary numbers of rigid groups (whether linked or not) and to define rigid groups consisting of single atoms.

In each simulated annealing step, a single degree of freedom is shifted by a random amount, limited by a maximum allowed displacement for that degree of freedom. The powder diffraction pattern for the new trial structure is then calculated and compared to the experimental powder diffraction pattern using R_{wp}. The rate limiting step is the evaluation of the structure factor amplitudes, and we have spent much effort to optimize this step (see reference 21). In our simulated annealing method, an adaptive temperature schedule is used, with the rate of cooling controlled by the fluctuations in R_{wp}. The maximum allowed displacements in a given simulated annealing step are controlled individually for each degree of freedom, based on acceptance ratio and fluctuations.

The efficiency of the method is enhanced significantly by performing local Rietveld optimization, within the parameter space defined by the degrees of freedom, whenever a promising structure solution is obtained during the calculation. These intermediate structure optimizations (local quenching) avoid the need to go to very low annealing temperatures during the main simulated annealing runs. In the standard simulated annealing procedure, once the system has been cooled to low temperatures, thermal fluctuations are insufficient to move the system across barriers in R_{wp} and the system would then become "frozen" in a local minimum. The intermediate minimization steps in our method perform this local optimization more efficiently on a larger number of structures, allowing the global simulated annealing procedure to remain at relatively high temperatures throughout the calculation. If an intermediate optimization results in a structure with lower R_{wp} than any structure found previously, this structure is written to a trajectory file and is retained as a potential structure solution. Subsequent simulated annealing steps proceed from the structure generated prior to the intermediate optimization.

By default, the start temperature is 1.5 times the average fluctuation in R_{wp} for a random sequence of moves, and the end temperature is 0.2 times the start temperature. These values have been found to be generally appropriate, but a different temperature schedule may be more efficient in some cases. With these default values, the temperature at the end of the structure solution calculation is usually still sufficiently high for the system to surmount most barriers on the R_{wp} hypersurface. The local Rietveld optimization is performed using the method of Powell [26], which does not require the evaluation of gradients. These minimizations are performed infrequently and the time required is insignificant in comparison with the overall time spent on the global simulated annealing calculation.

The structure solution calculation is started from a randomly generated initial structure (with random values for each degree of freedom). The default number of steps in the simulated annealing calculation increases exponentially with the total number of degrees of freedom. Since simulated annealing is a stochastic procedure, there is no guarantee that the global minimum will actually be located in a given run with a finite number of steps. Thus, it is a good strategy to repeat the calculation several times from different starting structures. Finding the same structure or very similar structures repeatedly is a strong indication that these structures represent the global minimum.

In general, structures containing long chains of connected torsion angles represent a greater challenge than cases with two or more rigid (or partly flexible) molecular fragments in the asymmetric unit. Because of strong coupling between different torsional degrees of freedom in long flexible chains, several degrees of freedom may be required to achieve their correct values simultaneously in order to locate the global minimum in the R_{wp} hypersurface. For flexible molecules, the R_{wp} hypersurface is typically rather flat with narrow but deep minima. Thus, a large number of trial structures with high R_{wp} may be sampled before an appropriate minimum is found, after which R_{wp} drops sharply.

4 EXAMPLES OF APPLICATIONS

The performance of Powder Solve has been validated and tested for a large set of known molecular crystal structures, as described in reference 21. Here we focus on two specific examples.

4.1 THE CYANURIC ACID/BIURET CO-CRYSTAL

Cyanuric acid [$(.CO.NH.)_3$] and biuret [$H_2N.CO.NH.CO.NH_2$] form a hydrogen-bonded 1:1 co-crystal, the structure of which has been determined previously [27] from single crystal X-ray diffraction data [C2/c; a = 20.65 Å, b = 6.63 Å, c = 13.40 Å, β = 101.2°]. The asymmetric unit comprises one molecule of cyanuric acid and one molecule of biuret. In the development of Powder Solve, we have used the cyanuric acid/biuret co-crystal as an example of a structure with two different types of molecular fragment in the asymmetric unit.

The powder X-ray diffraction pattern was recorded at ambient temperature in transmission mode on a Siemens D5000 diffractometer, using Ge-monochromated $CuK_{\alpha 1}$ radiation and a linear position-sensitive detector covering 8° in 2θ. The total 2θ range was 3° to 50° (step size 0.02°), recorded over 5 hours. Prior to structure solution, the powder diffraction pattern was reindexed and the unit cell parameters, peak shape parameters, a background polynomial and zero-point shift parameters were refined using the program Powder Fit [21]. The resulting unit cell parameters were very close to those obtained by single crystal X-ray diffraction (see above).

In the structure solution calculation using Powder Solve, the rigid cyanuric acid molecule was defined by 6 degrees of freedom, whereas the biuret molecule (with two variable torsion angles) was defined by 8 degrees of freedom. Hydrogen atoms were included in the structure solution calculation, although the inclusion of hydrogen atoms is not strictly necessary due to their low X-ray scattering power.

Five independent calculations, each comprising 5×10^6 simulated annealing steps, all led to the same structure solution; on average, 2×10^6 simulated annealing steps were required to reach this structure. The value of R_{wp} (calculated up to $2\theta = 50°$) for the best structure solution was 12.85 %. This value should be compared to the value of R_{wp} = 7.07 % obtained from Powder Fit, which represents the optimum fit that is theoretically achievable within the confines of the parameters used to define the calculated powder diffraction pattern (i.e. with all reflection intensities treated as individually refinable variables rather than constrained to depend on the structural model).

The best structure solution obtained from Powder Solve is shown in Figure 1, with the known crystal structure [27] also shown for comparison. The root mean square deviation between an atom in the

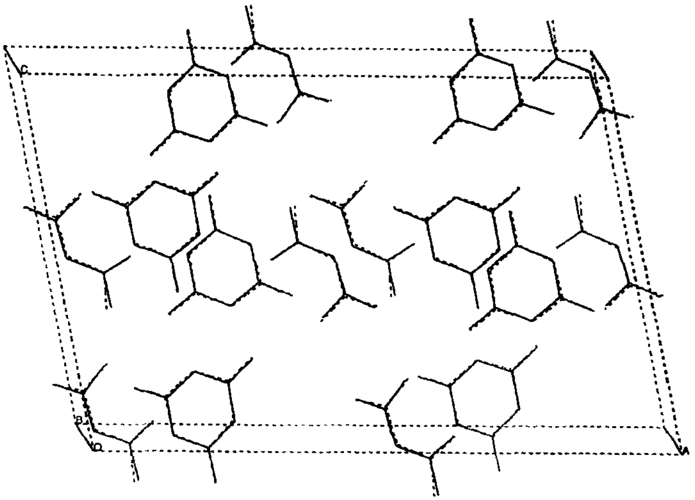

FIGURE 1 The positions of the molecular fragments in the best structure solution obtained from Powder Solve (dashed lines) and the positions of the corresponding atoms in the known structure of the cyanuric acid/biuret co-crystal [27] (solid lines).

structure solution and the corresponding atom in the known structure is 0.1 Å, with a maximum deviation of less than 0.2 Å. It is clear that the structure solution is well within refinable proximity to the known structure, representing successful structure solution.

4.2 $Ph_2P(O).(CH_2)_7.P(O)Ph_2$

As an example of a highly flexible system, we consider $Ph_2P(O).(CH_2)_7.P(O)Ph_2$ for which the molecular conformation is defined by 12 variable torsion angles, representing a total of 18 degrees of freedom in the structure solution calculation. The crystal structure of this material has been determined previously [12] from powder diffraction data using the Genetic Algorithm method [10]. In Figure 2, the best structure solution found using Powder Solve is compared with the structure reported previously [12] (the same powder diffraction data was used in both cases).

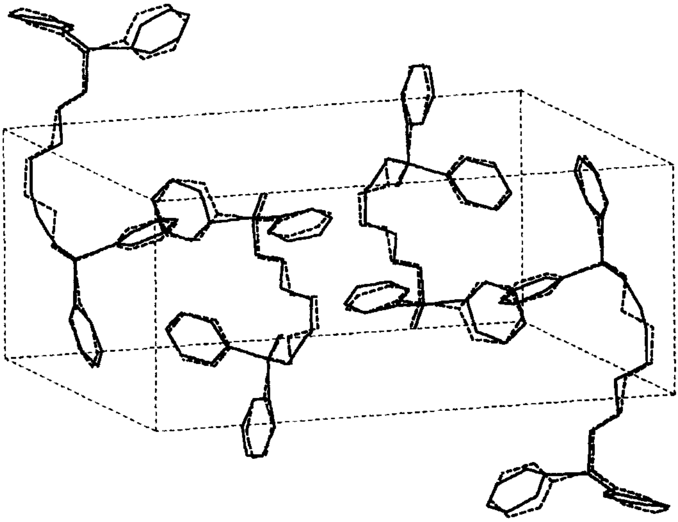

FIGURE 2 Comparison of the structure solutions obtained for $Ph_2P(O).(CH_2)_7.P(O)Ph_2$ using Powder Solve (dashed lines) and using the Genetic Algorithm method discussed in reference 12 (solid lines).

The values of R_{wp} (calculated up to $2\theta = 50°$) are 4.98 % for the solution reported in reference 12 (obtained using the Genetic Algorithm method for structure solution) and 4.83 % for the best solution obtained from Powder Solve. Although both structures have the same overall packing motif, several torsion angles in the $(CH_2)_7$ chain have different values. These small differences illustrate the intrinsic difficulty in establishing the crystal packing corresponding to the global minimum of R_{wp} in the case of large flexible molecules – for such systems, several solutions with low R_{wp} corresponding to similar conformations are often found. To resolve the small differences in the two independent structure solutions, we performed a force-field based energy minimization (using the COMPASS force-field [28]) on each structure, keeping the unit cell parameters fixed. Both solutions minimize to the same structure, which is very close to the solution obtained in reference 12. This example illustrates the advantage of using force-field based calculations to provide additional information when the powder diffraction data alone does not unambiguously distinguish similar structure solutions.

5 CONCLUDING REMARKS

Here we have presented an optimized implementation of a direct-space structure solution method, which is fully integrated within the Cerius^2 modelling package. The structure optimization is based on a Monte Carlo/simulated annealing technique. For up to about 10 degrees of freedom, this package can typically find the correct structure within a few minutes (assuming, as with all structure solution approaches, that the unit cell parameters and space group have been determined beforehand). This software package performs all stages of structure determination within a common environment: model definition, indexing, profile fitting, structure solution, Rietveld refinement and tests for structural stability based on lattice energy calculations. An important component, prior to the actual structure solution calculation, is the Powder Fit program for peak shape analysis of an experimental powder diffraction pattern. The robustness and reliability of this method are achieved by a simple enhancement of the original Pawley procedure.

Acknowledgements
Many members of staff at MSI have contributed to this development. In particular, we would like to thank C.M. Freeman and N.A. Austin for many valuable suggestions throughout this development. Collaborations with Miss P. Calcagno and Dr D. Philp are gratefully acknowledged.

References

[1] J.W. Visser, *J. Appl. Crystallogr.*, **2**, 89 (1969).
[2] P.-E. Werner, L. Eriksson, M. Westdahl, *J. Appl. Crystallogr.*, **18**, 367 (1985).
[3] A. Boultif, D. Louer, *J. Appl. Crystallogr.*, **24**, 987 (1991).
[4] B. M. Kariuki, S.A. Belmonte, M.I. McMahon, R.L. Johnston, K.D.M. Harris, R.J. Nelmes, *J. Synch. Rad.*, **6**, 87 (1999).
[5] G. Reck, R.-G. Kretschmer, L. Kutschabsky, W. Pritzkow, *Acta Crystallogr.*, **A44**, 417 (1988).
[6] J. Cirujeda, L.E. Ochando, J.M. Amig6, C. Rovira, J. Rius, J. Veciana, *Angew. Chemie, Int. Ed. Engl.*, **34**, 55 (1995).
[7] R.E. Dinnebier, P.W. Stephens, J.K. Carter, A.N. Lommen, P.A. Heiney, A.R. McGhie, L. Brard, A.B. Smith III, *J. Appl. Crystallogr.*, **28**, 327 (1995).
[8] R.B. Hammond, K.J. Roberts, R. Docherty, M. Edmondson, *J. Phys. Chem.*, **B101**, 6532 (1997).
[9] B.M. Kariuki, H. Serrano-González, R.L. Johnston, K.D.M. Harris *Chem. Phys. Lett.*, **280**, 189 (1997).

[10] K.D.M. Harris, R.L. Johnston, B.M. Kariuki, *Acta Crystallogr.*, **A54,** 632 (1998).
[11] K. Shankland, W.I.F. David, T. Csoka, *Z. Kristallogr.*, **212,** 550 (1997).
[12] B.M. Kariuki, P. Calcagno, K.D.M. Harris, D. Philp, R.L. Johnston, *Angew. Chemie, Int. Ed. Engl.*, **38,** 831 (1999).
[13] B.M. Kariuki, K. Psallidas, K.D.M. Harris, R.L. Johnston, R.W. Lancaster, S.E. Staniforth, S.M. Cooper, *Chem. Commun.*, 1677 (1999).
[14] J.M. Newsam, M.W. Deem, C.M. Freeman, *Accuracy in Powder Diffraction II: NIST Special Publication* **846,** 80 (1992).
[15] K.D.M. Harris, M. Tremayne, P. Lightfoot, P.G. Bruce, *J. Am Chem. Soc.*, **116,** 3543 (1994).
[16] Y.G. Andreev, P. Lightfoot, P.G. Bruce, *J. Chem. Soc., Chem. Commun.*, 2169 (1996).
[17] Y.G. Andreev, P. Lightfoot, P.G. Bruce, *J. Appl. Crystallogr.*, **30,** 294 (1997).
[18] M. Tremayne, B.M. Kariuki, K.D.M. Harris, *Angew. Chem., Int. Ed. Engl.*, **36,** 770 (1997).
[19] K.D.M. Harris, B.M. Kariuki, M. Tremayne, *Mater. Sci. Forum*, **278–291,** 32 (1998).
[20] W.I.F. David, K. Shankland, N. Shankland, *J. Chem. Soc., Chem. Commun.*, 931 (1998).
[21] G.E. Engel, S. Wilke, O. König, K.D.M. Harris, F.J.J. Leusen, *J. Appl. Crystallogr.*, in press.
[22] Cerius^2, Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121–3752, U.S.A.
[23] G.S. Pawley, *J. Appl. Crystallogr.*, **14,** 357 (1981).
[24] R.A. Young, *The Rietveld Method,*Oxford University Press, Oxford, (1993), pp. 9–10.
[25] H.M. Rietveld, *J. Appl. Crystallogr.*, **2,** 65 (1969).
[26] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, (1986), pp. 298–300.
[27] N.M. Stainton, K.D.M. Harris, R.A. Howie, *J. Chem. Soc., Chem. Commun.*, 1781 (1991).
[28] H. Sun, *J. Phys. Chem.*, **B102,** 7338 (1998).