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

New Methodologies for Solving Crystal Structures from Powder Diffraction Data

Kenneth D.M. Harris [a], Benson M. Kariuki [a], Maryjane Tremayne [b] & Roy L. Johnston [a]

[a] School of Chemistry, University of Birmingham, Edgbaston, Birmingham, B15 2TT, U.K.

[b] Department of Chemistry, University of Glasgow, Glasgow, G12 8QQ, U.K.

Version of record first published: 04 Oct 2006

To cite this article: Kenneth D.M. Harris, Benson M. Kariuki, Maryjane Tremayne & Roy L. Johnston (1998): New Methodologies for Solving Crystal Structures from Powder Diffraction Data, Molecular Crystals and Liquid Crystals Science and Technology. Section A. Molecular Crystals and Liquid Crystals, 313:1, 1-14

To link to this article: http://dx.doi.org/10.1080/10587259808044255

Mol. Cryst. Liq. Cryst., 1998, Vol. 313, pp. 1–14
Reprints available directly from the publisher
Photocopying permitted by license only

New Methodologies for Solving Crystal Structures from Powder Diffraction Data

KENNETH D.M. HARRIS[a*], BENSON M. KARIUKI[a], MARYJANE TREMAYNE[b], ROY L. JOHNSTON[a]

[a]School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K.; [b]Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, U.K.

An overview is given of two methods developed recently for crystal structure solution from powder diffraction data, based on the techniques of Monte Carlo sampling and Genetic Algorithms. Both methods operate in direct space, and the specific advantages of this approach in the case of structure solution from powder diffraction data are highlighted. The fundamental principles underlying the Monte Carlo and Genetic Algorithm techniques are described, and selected case studies highlighting different aspects of the application of these techniques are presented.

Keywords: Monte Carlo method; Genetic Algorithm; powder diffraction; crystal structure solution

INTRODUCTION

The determination of crystal structures from single crystal X-ray diffraction data can now, in general, be carried out in a routine and straightforward manner. However, many important solids are microcrystalline, and are not amenable to investigation using conventional single crystal X-ray diffraction methods. Structural characterization of such materials is clearly dependent on the availability of techniques for determining crystal structures from powder diffraction data. Here we give an overview of the development of two new

techniques, based on Monte Carlo sampling and Genetic Algorithms, for crystal structure solution from powder diffraction data.

Initially we consider the difficulties associated with solving crystal structures directly from powder diffraction data[1]. Essentially the same information is contained in single crystal and powder diffraction patterns, but in the former case this information is distributed in three-dimensional space whereas in the latter case the three-dimensional diffraction data are "compressed" into one dimension. As a consequence, there is generally considerable overlap of peaks in the powder diffraction pattern, leading to severe ambiguities in extracting the intensities $I(hkl)$ of individual diffraction maxima. Despite this fact, the traditional approach for crystal structure solution from powder diffraction data is to attempt to extract the intensities $I(hkl)$ of individual reflections directly from the powder diffraction pattern, and then to solve the structure using the methods that are normally applied to single crystal diffraction data (e.g. direct methods and the Patterson method). The major difficulty with this approach arises in extracting, from the overlapped peaks in the powder diffractogram, values of $I(hkl)$ that are sufficiently reliable to lead to a successful structure solution calculation. Thus, the problem of extensive peak overlap in the powder diffraction pattern limits severely the complexity of structures that can be solved successfully by this traditional approach.

An alternative philosophy for crystal structure solution from powder diffraction data is to postulate structural models in direct space, independently of the powder diffraction data, with the suitability of these models assessed by direct comparison of the powder diffraction patterns calculated for these models against the experimental powder diffraction pattern. This comparison may be quantified using the profile R-factor (as used in Rietveld refinement), which considers the whole digitized intensity profile (not the integrated intensities $I(hkl)$ of individual diffraction maxima) and therefore implicitly takes care of the overlap of peaks. This approach completely avoids the problematic step of extracting values of $I(hkl)$ from the powder diffraction pattern; no partitioning of the experimental powder diffraction profile into a set of "single-crystal-like" intensities $I(hkl)$ is carried out, and by circumventing the need to extract such intensities from the powder diffractogram, the major problem associated with the traditional approaches is implicitly overcome.

As discussed below, this philosophy of direct-space sampling is embodied within both the Monte Carlo and Genetic Algorithm techniques that we have developed for structure solution from powder diffraction data.

MONTE CARLO METHODOLOGY

In the Monte Carlo approach[2,3], a series of structures (denoted x[i] for i = 1, 2, ..., j, j+1, ..., N) is generated for consideration as potential structure solutions. The structure is specified by a "structural fragment", which represents an appropriate collection of atoms within the asymmetric unit. The structural fragment is constructed either using standard bond lengths and bond angles for the type of molecule under study, or using the known geometry of a similar molecule. Each structure generated during the Monte Carlo calculation is derived from the previous structure by a random displacement of the structural fragment within the unit cell. The procedure for each Monte Carlo move (described here for the general case in which structure x[j+1] is generated from structure x[j]) comprises the following steps:

(i) Starting from structure x[j], the structural fragment is displaced by a random amount to generate a trial structure x[trial]. This random displacement may involve one or more of the following (depending on the system of interest): (a) translation of the structural fragment; (b) rotation of the structural fragment; (c) variation of selected internal degrees of freedom (e.g. torsion angles). The agreement between the powder diffraction pattern corresponding to the trial structure and the experimental powder diffraction pattern can be assessed using R_{wp} (x[trial]) [the profile R-factor R_{wp} quantifies the level of disagreement between the experimental powder diffraction pattern and the calculated powder diffraction pattern for the structure of interest (in this case x[trial])].

(ii) The trial structure is then accepted or rejected by considering the difference between R_{wp} for structure x[trial] and R_{wp} for structure x[j] by invoking the Metropolis importance sampling algorithm[4]. Thus, the difference $Z = R_{wp}(x[trial]) - R_{wp}(x[j])$ is considered as follows. If $Z \leq 0$, the trial structure is accepted as the new structure, with the new structure x[j+1] taken as x[trial]. If $Z > 0$, the trial structure is accepted with probability $\exp(-Z/S)$ and rejected with probability $[1-\exp(-Z/S)]$, where S is an appropriate scaling factor of Z. If the trial structure is rejected, structure x[j+1] is taken to be the same as x[j].

This procedure is repeated to generate a large number of structures, with each structure derived from the previous one through a small random displacement. After a sufficiently extensive range of structural space has been explored, the best structure solution (corresponding to lowest R_{wp}) is used as the starting model for structure refinement calculations (using the Rietveld profile refinement method). Clearly one of the crucial issues governing the potential success of this method is the ability to discriminate the correct position of the structural fragment from wrong positions on the basis of R_{wp}; this issue is particularly pertinent when the structural fragment comprises

only a fraction of the asymmetric unit, and points towards the importance of making the optimum choice of structural fragment in applying the Monte Carlo technique.

GENETIC ALGORITHM METHODOLOGY

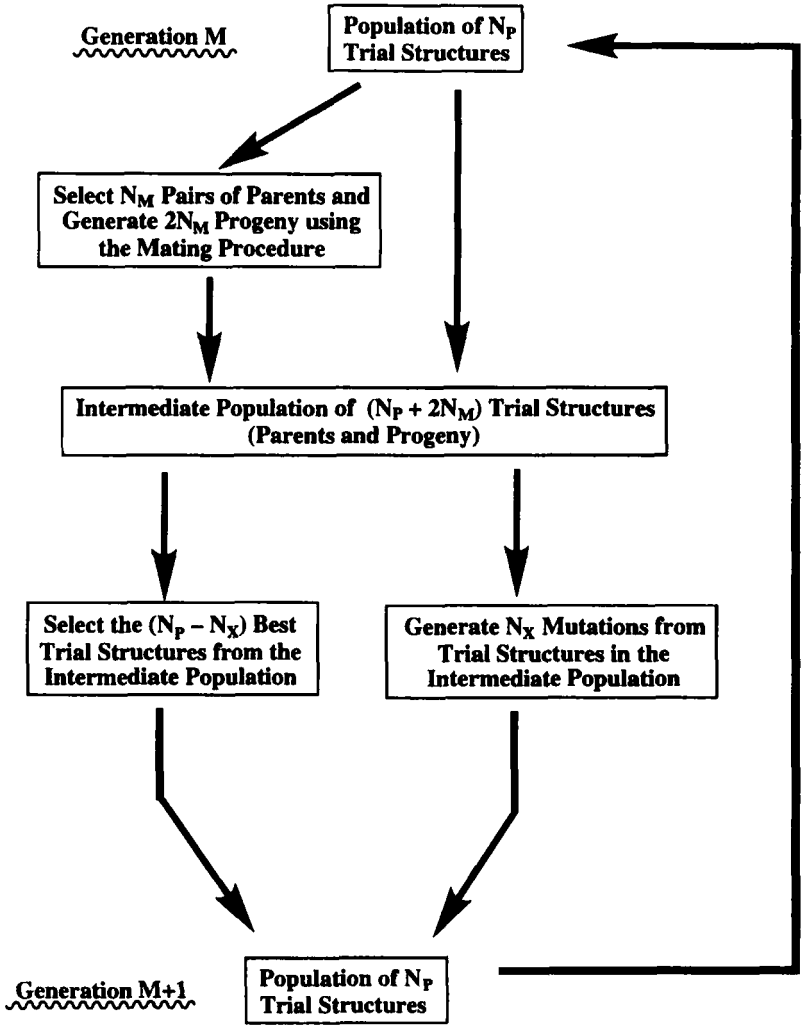
The Genetic Algorithm is an optimization technique based on evolutionary principles, in which the fittest members of a population survive and procreate to produce improved subsequent generations[5,6]. In the application of the Genetic Algorithm to structure solution from powder diffraction data[7], each member of the population is a trial crystal structure, defined by the position, orientation and intramolecular geometry of a structural fragment (analogous to that used in the Monte Carlo method). The "fitness" of each member of the population is related to the profile R-factor (R_{wp}) for the corresponding crystal structure, and one of the functions that we have used to define fitness is:

$$F(i) = \frac{1 - \tanh\left(-2\pi\left(1 - \frac{\{R_{wp}(i) - R_{min}\}}{\{R_{max} - R_{min}\}}\right)\right)}{2}$$

with $F(i) = 1$ when $R_{wp}(i) = R_{min}$ and $F(i) = 0$ when $R_{wp}(i) = R_{max}$. This fitness function has been designed from our knowledge of the typical nature of R_{wp} hypersurfaces.

At the start of the calculation, the population comprises a randomly generated set of structures. The set **X** of variables that defines a given structure can be regarded as its "genetic code". Well-defined procedures that include mating and mutation are used to produce subsequent generations of the population (see Scheme 1). Mating involves selecting pairs of structures (the probability of selecting a given structure is proportional to its fitness), and generating progeny by combining genetic information from the two parents. Any progeny which are identical to an existing structure in the population are deleted immediately, preventing premature convergence of the entire population towards a single structure. Diversity of the population is ensured by introducing a few mutant structures within each generation; these mutants are generated by selecting structures from the population (with the probability of selecting a given structure proportional to its fitness) and introducing random changes to parts of their genetic codes. The process of "natural selection" ensures that only the best structures survive and the overall fitness of the population improves from one generation to the next. After a sufficient number of generations, the fittest member of the population (lowest R_{wp})

should be close to the correct structure solution.



SCHEME 1 Overview of the Genetic Algorithm structure solution method.

APPLICATIONS OF THE MONTE CARLO APPROACH

The first applications of the Monte Carlo method[2] involved the simplest cases in which the structural fragment is subjected either to translation through the unit cell with no rotation (when the structural fragment is a single atom) or to rotation about a fixed point in space with no translation. This approach was adopted to solve the crystal structures of *p*-$CH_3C_6H_4SO_2NHNH_2$ (structure previously known) and *p*-$BrC_6H_4CH_2CO_2H$ (structure previously unknown); in these structure solution calculations, two separate Monte Carlo calculations were carried out, involving: (i) in the first calculation, location of a dominant scatterer (S or Br) by translation of this atom within the unit cell; (ii) then, after establishing the correct position of the dominant scatterer, rotation of a rigid fragment around the fixed position of the dominant scatterer found in calculation (i).

Subsequently[8], the Monte Carlo technique was applied successfully to solve the crystal structure (previously unknown) of the γ phase of 3-chloro-*trans*-cinnamic acid from powder X-ray diffraction data. The structural fragment comprised a rigid *trans*-cinnamic acid molecule, with the chlorine and hydrogen atoms omitted. The oxygen atoms were maintained at a fixed distance from the crystallographic centre of symmetry, with the centre of symmetry lying in the molecular plane. The structural fragment was rotated by a random angular displacement around a random axis constrained to pass through the centre of symmetry.

The above applications of the Monte Carlo technique were restricted to cases in which the structural fragment was subjected to only translation or only rotation within the unit cell, but not subjected to simultaneous translation and rotation. A major generalization of the methodology, in the case of rigid body structural fragments, involved the consideration of simultaneous translation and rotation of the structural fragment. This generalized approach was first applied[9] to solve the crystal structure (previously known) of *p*-methoxybenzoic acid. In this calculation, the structural fragment comprised the carbon and oxygen atoms of the benzoate group (i.e. C_6CO_2) plus the oxygen atom of the methoxy group in the *para* position. In the Monte Carlo calculation, the structural fragment was subjected to both translation and rotation through the unit cell. The best position of this structural fragment was discriminated readily (on the basis of R_{wp}) from "wrong" positions sampled during the Monte Carlo calculation (Fig. 1). In Fig. 2, the position of the structural fragment in the best structure solution obtained from the Monte Carlo calculation is compared with the position of the corresponding fragment in the known crystal structure. Clearly the Monte Carlo calculation has generated a good quality structure solution, which can then be refined readily to within experimental error of the known structure.

This generalized rigid-body approach was then applied[10] to solve the crystal structure (previously unknown) of 1-methylfluorene. The structural fragment comprised all non-hydrogen atoms of the fluorenyl (C_{13}) group. This rigid structural fragment was subjected to simultaneous translation and rotation within the unit cell, resulting in a structure solution that was again readily discriminated (Fig. 3). The structure was completed by locating the methyl carbon using difference Fourier methods, followed by Rietveld refinement (Fig. 4). The final refined structure is shown in Fig. 5.

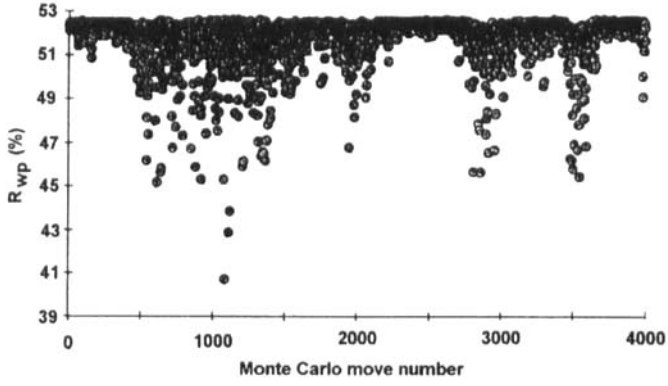


FIGURE 1 Plot of R_{wp} *versus* Monte Carlo move number for trial structures sampled in the structure solution calculation for *p*-methoxybenzoic acid.

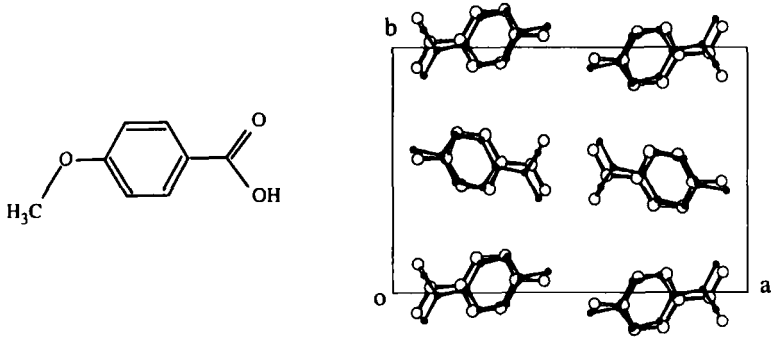


FIGURE 2 The position of the structural fragment obtained in the Monte Carlo structure solution for *p*-methoxybenzoic acid (filled circles) overlaid on the position of the corresponding fragment in the known crystal structure (open circles).

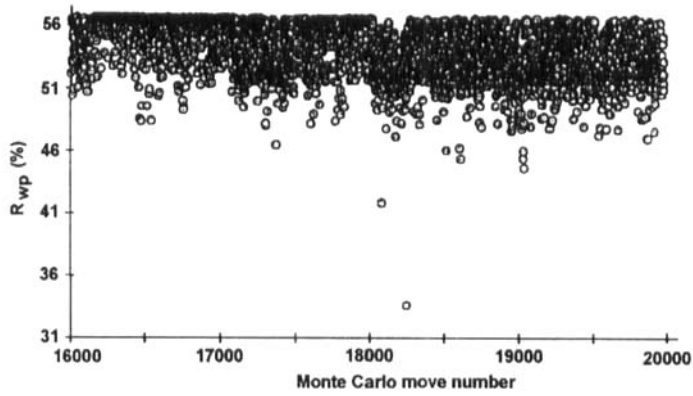


FIGURE 3 Plot of R_{wp} *versus* Monte Carlo move number for trial structures sampled in the structure solution calculation for 1-methylfluorene.

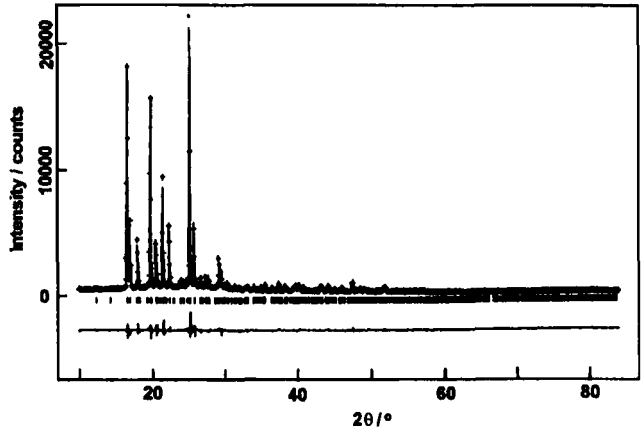


FIGURE 4 The calculated (line), experimental (+) and difference (bottom) powder diffraction patterns in the final structure refinement for 1-methylfluorene.

Further generalizations of the Monte Carlo method have considered flexible structural fragments, with the introduction of intramolecular degrees of freedom. Our first application of this approach[11] was to determine the crystal structure (previously unknown) of the red phase of the pigment fluorescein (Fig 6). The Monte Carlo calculation in this case involved simultaneous translation and rotation of the structural fragment as well as variation of the intramolecular geometry through rotation about the C–C bond which links the benzoic acid and hydroxyxanthenone moieties. To our knowledge, this represents the largest organic crystal structure (the first with over 20 non-hydrogen atoms in the asymmetric unit) that has so far been solved from powder diffraction data by any technique. Clearly this advance demonstrates the rapidly increasing scope and potential of powder diffraction as a technique for crystal structure determination, particularly for the case of organic molecular crystals. In addition to these developments in methodology, the successful structure determination of red fluorescein represents the resolution of a long-standing structural problem.

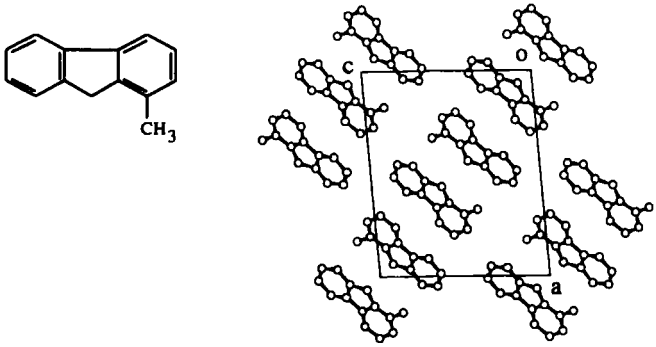


FIGURE 5 The crystal structure of 1-methylfluorene.

Finally, a new Monte Carlo approach incorporating restrained relaxation of the molecular geometry within the structure solution calculation has been developed[12]. This combined approach blends aspects of Rietveld refinement within the Monte Carlo structure solution calculation, and provides dramatic advantages in the ability to discriminate the correct structure from wrong structures sampled during the calculation. These advantages have been demonstrated by applying the new method to solve the low-temperature

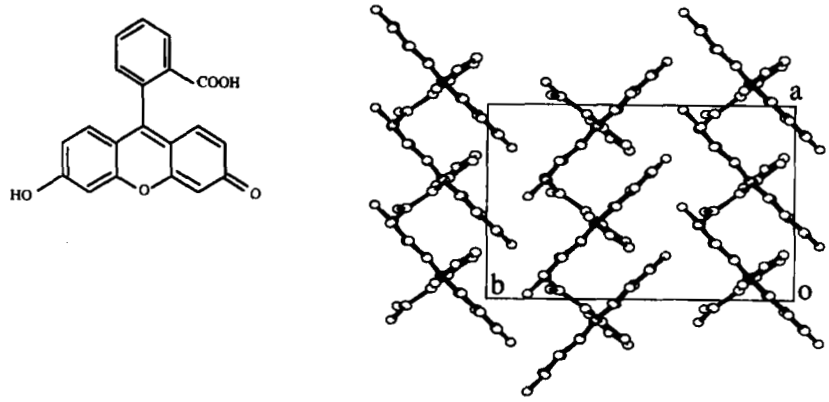


FIGURE 6 The crystal structure of red fluorescein.

crystal structure (phase II) of perdeuterated pyrene (structure previously known) from neutron powder diffraction data. In previous applications of the Monte Carlo technique, the structural fragment was constrained to have a standard geometry throughout the calculation, and the only allowed variation in the molecular geometry was through the variation of selected torsion angles. In the new approach, restrained relaxation of the structural fragment from standard geometry is considered for each structure sampled during the Monte Carlo calculation, and gives rise to dramatically improved discrimination (in terms of R_{wp}) between the correct structure solution and incorrect structures sampled during the calculation. This new development represents a substantial enhancement of the scope and potential of the Monte Carlo structure solution method, and may be particularly advantageous for cases in which the discrimination in R_{wp} between the correct structure solution and incorrect structure solutions is expected to be inherently small. Given the tremendous advantages demonstrated by the research on pyrene, relaxation of the geometry of the structural fragment is now being exploited as the standard approach for applying the Monte Carlo method.

APPLICATIONS OF THE GENETIC ALGORITHM APPROACH

The success of the Genetic Algorithm approach for structure solution from powder diffraction data[7], using the program GAPSS[13], is illustrated by two examples: *p*-methoxybenzoic acid and formylurea. The crystal structures of

both of these materials were solved previously from powder X-ray diffraction data (Monte Carlo method for *p*-methoxybenzoic acid[9]; direct methods for formylurea[14]), and the same data were used in the Genetic Algorithm calculation. Testing the new Genetic Algorithm method for previously-known structures in this way allows a definitive assessment of its validity. For both *p*-methoxybenzoic acid and formylurea, the Genetic Algorithm involved the evolution of 100 generations of a population of 100 structures (N_P). In each generation, 100 progeny (N_M) (i.e. 50 pairs of parents) and 10 mutations (N_X) were considered.

In the Genetic Algorithm calculation for *p*-methoxybenzoic acid, the structural fragment comprised the carbon and oxygen atoms of the benzoate group (i.e. C_6CO_2) and the oxygen atom of the methoxy group. Standard geometries (bond lengths and angles) were used for the rigid structural fragment, with the two C–O bond lengths of the carboxylic acid group taken as equal. Thus, each structure was defined by six parameters $\{x_i,y_i,z_i,\theta_i,\phi_i,\psi_i\}$, representing the position (x_i,y_i,z_i) and orientation (θ_i,ϕ_i,ψ_i) of the structural fragment. Mating was carried out by a single point cross-over, with the genetic codes of the two parents (i and j) cut between the positional and orientational parameters to produce two progeny $\{x_i,y_i,z_i,\theta_j,\phi_j,\psi_j\}$ and $\{x_j,y_j,z_j,\theta_i,\phi_i,\psi_i\}$. Mutation was carried out by randomly changing one positional parameter and one orientational parameter in selected structures.

In the Genetic Algorithm calculation for formylurea, the structural fragment comprised all non-hydrogen atoms, with standard bond lengths and bond angles. The internal degrees of freedom were the torsion angles for the two C–N bonds shown: H.(CO)–(NH)–(CO).(NH_2). Thus, each structure was defined by eight parameters $\{x_i,y_i,z_i,\theta_i,\phi_i,\psi_i,\tau_i,c_i\}$, representing six parameters to define the position and orientation of the structural fragment plus two torsion angles (intramolecular degrees of freedom). For mating and mutation, the eight parameters were considered to comprise four groups $\{(x_i,y_i,z_i);(\theta_i,\phi_i,\psi_i);(\tau_i);(\chi_i)\}$. In mating, a pair of groups was selected from one parent (with three equi-probable ways of selecting two groups from four) and combined with the other two groups taken from the other parent; again, the two progeny generated by cross-over were considered. Mutations involved making a random change to two of the four groups (for the groups (x_i,y_i,z_i) and (θ_i,ϕ_i,ψ_i), only one of the three parameters was changed).

The evolution of R_{wp} in the Genetic Algorithm calculations for formylurea are shown in Fig. 7. The best structure solution is compared with the positions of the corresponding atoms in the known structure for formylurea in Fig. 8. It is clear that the Genetic Algorithm approach has successfully located and discriminated a position for the structural fragment close to its true position in the crystal structure. Subsequent Rietveld refinement of these structure solutions (and for *p*-methoxybenzoic acid,

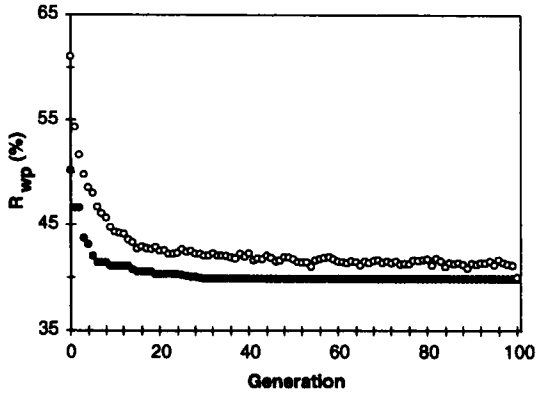


FIGURE 7 The evolution of R_{wp} for the best structure (filled circles) and the average R_{wp} (open circles) for the population in the Genetic Algorithm calculation for formylurea.

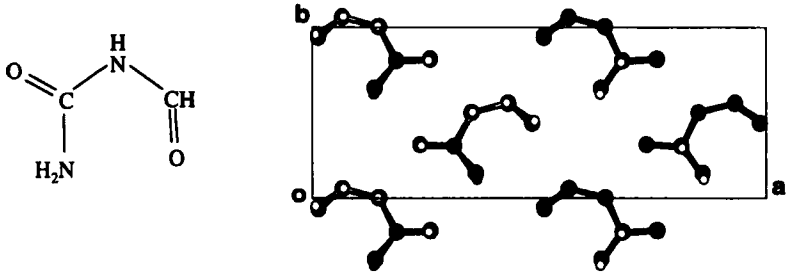


FIGURE 8 Comparison between the position of the structural fragment in the best structure solution obtained in the Genetic Algorithm calculation (open circles) and the position of the corresponding atoms in the known crystal structure (filled circles) for formylurea.

location of the carbon atom of the methoxy group using difference-Fourier methods) leads straightforwardly to the known crystal structures.

These results demonstrate the potential of the Genetic Algorithm approach for structure solution from powder diffraction data, both for rigid and flexible structural fragments. Preliminary comparisons suggest that the Genetic Algorithm approach may be significantly more efficient than the Monte Carlo technique (note from Fig. 7 that structures with low R_{wp} are generated very early in the evolutionary process), while we reiterate the advantages of direct-space methods in general over the traditional techniques for structure solution from powder diffraction data. A rigorous optimization of the Genetic Algorithm approach and a systematic comparison with the Monte Carlo method are currently in progress. It is clear that the reported application of the Genetic Algorithm represents a major new development that extends further the scope and potential of powder diffraction as a technique for crystal structure solution, and we forecast with confidence that the Genetic Algorithm method will engender major future progress in many fields of solid state and structural sciences.

CONCLUDING REMARKS

As described above, the recent development of direct space techniques for crystal structure solution from powder diffraction data, and the advances in their application, represent significant steps forward in the capabilities for structure determination from powder diffraction data. With the availability of these techniques, significant future progress in understanding structural properties of solids will emerge from their application across a wide range of disciplines within solid state and materials sciences. With the emergence of these new approaches, together with the traditional techniques for structure solution from powder diffraction data and other emerging methods based on energy simulations, solid state and materials scientists are now equipped with a broad range of options for attacking the challenging problem of structure solution from powder diffraction data.

Acknowledgements

We are grateful to E.P.S.R.C., Ciba Specialty Chemicals, the Nuffield Foundation and the University of Birmingham for financial support. Several colleagues and collaborators, mentioned in the references, are thanked for valuable discussions.

References

[1] K.D.M. Harris and M. Tremayne, *Chem. Mater.*, **8**, 2554 (1996).

[2] K.D.M. Harris, M. Tremayne, P. Lightfoot, and P.G. Bruce, *J. Am.*

Chem. Soc., **116**, 3543 (1994).

[3] J.M. Newsam, M.W. Deem, and C.M. Freeman, *Accuracy in Powder Diffraction II*; *NIST Special Publ.* No. **846**, 80 (1992).

[4] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).

[5] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading (MA), (1989)

[6] A.J. Keane in *Modern Heuristic Search Methods* (Eds. V. Rayward-Smith *et al.*), Wiley, New York, (1996)

[7] K.D.M. Harris, R.L. Johnston, B.M. Kariuki, and M. Tremayne, *Chem. Comm.*, submitted

[8] B.M. Kariuki, D.M.S. Zin, M. Tremayne, and K.D.M. Harris, *Chem. Mater.*, **8**, 565 (1996).

[9] M. Tremayne, B.M. Kariuki, and K.D.M. Harris, *J. Appl. Crystallogr.*, **21**, 211 (1996).

[10] M. Tremayne, B.M. Kariuki, and K.D.M. Harris, *J. Mat. Chem.*, **6**, 1601 (1996).

[11] M. Tremayne, B.M. Kariuki, and K.D.M. Harris, *Angew.Chemie, Int. Ed. Engl.*, **36**, 770 (1997).

[12] M. Tremayne, B.M. Kariuki, K.D.M. Harris, K. Shankland, and K.S. Knight, *J. Appl. Crystallogr.*, in press.

[13] R.L. Johnston, B.M. Kariuki, and K.D.M. Harris, *GAPSS* Genetic Algorithm for Powder Structure Solution, 1997.

[14] P. Lightfoot, M. Tremayne, K.D.M. Harris, and P.G. Bruce, *J. Chem. Soc., Chem. Commun.*, 1012 (1992).