

STRUCTURE DETERMINATION OF MOLECULAR CRYSTALS DIRECTLY FROM POWDER DIFFRACTION DATA*

KENNETH D.M. HARRIS

*School of Chemistry; University of Birmingham, Edgbaston, Birmingham B15 2T7; United Kingdom.
E-mail: K.D.M.Harris@bham.ac.uk*

Many solids can be prepared only as microcrystalline powders and are not suitable for structural characterization by single crystal diffraction methods. In such cases, it is *necessary* to carry out structure determination using powder diffraction data. *Here* we highlight recent developments in the opportunities for solving molecular crystal structures directly from powder diffraction data, focusing on the direct-space strategy in which a hypersurface based on the powder profile R-factor R_{wp} is searched using Monte Carlo or Genetic Algorithm methods. Recent fundamental developments are described, and illustrative examples are given to highlight the application of this strategy to determine the structures of molecular materials.

1. Introduction: the Challenges of Structure Determination from Powder Diffraction Data

The ability to determine crystal structures directly from powder diffraction data promises to open up many new avenues of structural science. Many important materials cannot be prepared as single crystals of appropriate size and quality for conventional single crystal diffraction studies, nor indeed for the emerging synchrotron-based microcrystal diffraction techniques. In such cases, structure determination from powder diffraction data may represent the only viable approach for obtaining an understanding of the structural properties of the material of interest. However, it is important to recognize that structure determination from powder diffraction data is far from routine, and significant challenges must be overcome in developing and applying methods for this purpose. For this reason, several research groups have devoted considerable effort in recent years to the development of new and improved techniques in this field. This article gives an overview of the problems and challenges associated with structure determination from powder diffraction data, while focusing on some of our own recent contributions in this field. More detailed reviews covering all aspects of structure determination from powder diffraction data may be found in references [1]-

[6]. Crystal structure determination from diffraction data (either single crystal or powder) can be divided into the following stages: (i) unit cell determination and space group assignment, (ii) structure solution, and (iii) structure refinement. The aim of *structure solution* is to derive an initial approximation to the structure from direct consideration of the experimental diffraction data, but starting from no knowledge of the actual arrangement of atoms or molecules within the unit cell. If the structure solution is a sufficiently good approximation to the true structure, a good quality structure may then be obtained by *structure refinement*. For powder diffraction data, structure refinement is now carried out fairly routinely using the Rietveld profile refinement technique [7,8], and unit cell determination is carried out using standard indexing procedures (see, for example, references [9]-[13]). In this article, we focus on structure solution, which is generally the most challenging stage of the structure determination process.

The techniques currently available for structure solution from powder diffraction data can be subdivided into two categories-"traditional" and "direct-space" approaches. As discussed below, the traditional approach follows a close analogy to conventional procedures for analysis of single

* Based on a plenary lecture delivered at the 50th Anniversary Meeting of the Crystallographic Society of Japan on 22 November 2000

crystal diffraction data, whereas the direct-space approach follows a close analogy to global optimization procedures, which find applications in many areas of science. Indeed, our initial work on the development of the direct space strategy [14] originated from identifying the opportunity to combine our existing experience in computer simulation of solids (involving global optimization based on consideration of energy) [15] together with our experience in the application of traditional techniques for powder structure solution [16].

In the *traditional approach*, the intensities $I(hkl)$ of individual reflections are extracted directly from the powder diffraction pattern, and the structure is then solved using these $I(hkl)$ data in the types of structure solution calculation that are used for single crystal diffraction data (e.g. direct methods or Patterson methods). However, as there is usually extensive peak overlap in the powder diffraction pattern, extracting reliable values of the intensities $I(hkl)$ of the individual diffraction maxima can be problematic, and may lead to difficulties in subsequent attempts to solve the structure using these "single-crystal-like" approaches. To overcome this problem either requires improved techniques for extracting and utilizing peak intensities, or requires the use of new structure solution strategies (see below) that allow the experimental powder diffraction profile to be used directly in its "raw" digitized form, without the requirement to extract the intensities $I(hkl)$ of individual diffraction maxima.

In the *direct-space approach*, trial structures are generated in direct space, independently of the experimental powder diffraction data, with the suitability of each trial structure assessed by direct comparison between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern. This comparison is quantified using an appropriate R-factor. Most direct-space approaches reported to date have used the weighted powder profile R-factor R_{wp} (the R-factor normally employed in Rietveld refinement), although we note that some implementations of direct-space approaches have instead used R-factors based on extracted peak intensities. R_{wp} is defined as

$$R_{wp} = 100x \left(\frac{\sum_i w_i (y_i(obs) - y_i(calc))^2}{\sum_i w_i (y_i(obs))^2} \right)^{1/2}$$

where $y_i(obs)$ is the intensity of the i th data point in the experimental powder diffraction profile, $y_i(calc)$ is the intensity of the i th data point in the calculated powder diffraction profile, and w_i is a weighting factor for the i th data point. Importantly, R_{wp} considers the whole digitized intensity profile point-by-point, rather than the integrated intensities of individual diffraction maxima. Thus, R_{wp} implicitly takes care of peak overlap and uses the digitized powder diffraction data directly "as measured".

The basis of the direct-space strategy for structure solution is to find the trial crystal structure corresponding to lowest R-factor, and is equivalent to exploring a hypersurface $R(\Gamma)$ to find the global minimum, where Γ represents the set of variables that define the structure. In principle, any technique for global optimization may be used to find the lowest point on the $R(\Gamma)$ hypersurface, and much success has been achieved in using Monte Carlo [14, 17-25], Simulated Annealing [26-34] and Genetic Algorithm [35-46] methods in this field. In addition, grid search methods have also been employed [4751]. This article focuses on fundamental and applied aspects of our implementations of Monte Carlo (MC) and Genetic Algorithm (GA) techniques within direct-space structure solution from powder diffraction data, with particular emphasis on the application of these techniques to elucidate structural properties of molecular materials.

To our knowledge, the first previously unknown organic molecular crystal structure to be solved using the traditional approach (employing direct methods for structure solution) was formylurea [16], although the successful structure solution of the previously known structure of cimetidine had earlier been demonstrated (also using direct methods) [52]. The first material of unknown crystal structure to be solved by a direct-space approach was p- $\text{BrC}_6\text{H}_4\text{CH}_2\text{CO}_2\text{H}$ [14], using the Monte Carlo method. Following on from this early work, the structures of a wide range of other molecular materials have now been determined directly from powder diffraction data, as

surveyed comprehensively in a recent review article [6].

2. General Aspects of the Direct-Space Strategy for Structure Solution

In the direct-space strategy for structure solution from powder diffraction data, the structure is defined by a "structural fragment", which represents the asymmetric unit (or an appropriate subset of the atoms in the asymmetric unit). The structural variables (i.e. the set Γ discussed above) represent the position, orientation and intramolecular geometry of each molecule in the asymmetric unit. The position is defined by the coordinates $\{x, y, z\}$ of the centre of mass or a selected atom, the orientation is defined by rotation angles $\{\theta, \phi, \psi\}$ around a set of orthogonal axes, and the intramolecular geometry is specified by a set of variable torsion angles $\{\tau_1, \tau_2, \dots, \tau_n\}$. In general, the bond lengths and bond angles are fixed in the calculation, and are taken either from standard values for the type of molecule under study or from the known geometry of a similar molecule.

As most direct-space approaches are stochastic in nature, it is strongly recommended that, for a given problem, the structure solution calculation should be repeated several times from different random starting points. Clearly, obtaining the same structure solution from these independent calculations provides strong support that the true global minimum on the $R(\Gamma)$ hypersurface (i.e. the

correct structure solution) has been found.

Clearly the use of R_{wp} to assess the correctness of a structural model, as in most implementations of direct-space approaches, 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. In general, reliable values of these parameters can be determined, prior to structure solution, by fitting the powder diffraction pattern with the use of arbitrary peak intensities (i.e. without using any structural model to determine the peak intensities), for example using the Pawley fitting procedure [53]. Careful prior analysis of the experimental data in this way ensures that the profile parameters used subsequently in structure solution calculations provide a reliable description of the experimental powder diffraction pattern.

3. Monte Carlo and Simulated Annealing Techniques

The foundations of the Monte Carlo and Simulated Annealing techniques are closely related, and in both cases a sequence of structures (each denoted Γ_i ; for $i=1, 2, \dots, N$) is generated for consideration as potential structure solutions. We begin by discussing the Monte Carlo method [14]. 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 Γ_{j+1} is generated from structure Γ_j) comprises the following steps (Figure 1).

- (i) Starting from structure Γ_j , a trial structure $\Gamma_{j,\text{trial}}$ is generated by making small random displacements to each of the structural variables $\{x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \dots, \tau_n\}$. The agreement between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern is then assessed, for example using R_{wp} .
- (ii) The trial structure is then accepted or rejected by considering the difference $[Z=R_{wp}(\Gamma_{j,\text{trial}})-R_{wp}(\Gamma)]$ between the values of R_{wp} for structures $\Gamma_{j,\text{trial}}$ and Γ_j and invoking the Metropolis

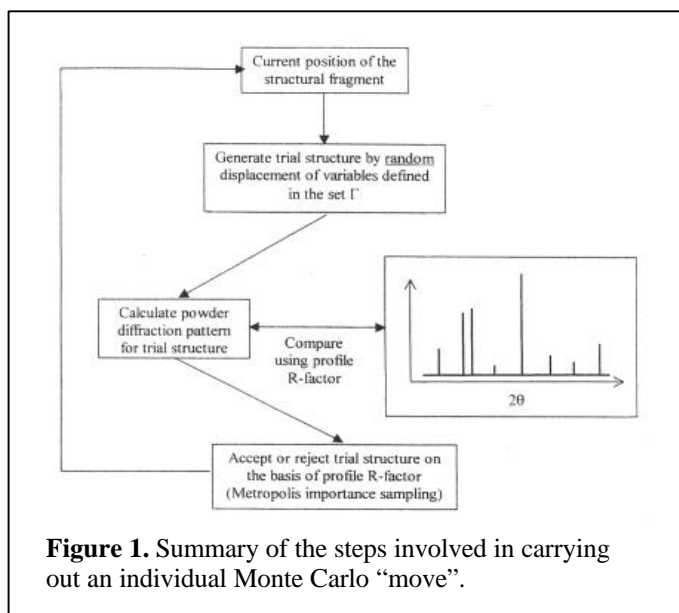


Figure 1. Summary of the steps involved in carrying out an individual Monte Carlo "move".

importance sampling algorithm [54]. If $Z=0$, the trial structure is automatically accepted, whereas 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. If the trial structure is accepted, structure Γ_{j+1} is taken to be the same as $\Gamma_{j,\text{trial}}$. If the trial structure is rejected, structure Γ_{j+1} is taken to be the same as Γ_j . The parameter S may either be fixed or varied in a controlled manner during the calculation. Clearly the higher the value of S , the greater the probability that trial structures with $Z>0$ will be accepted.

This procedure is repeated to generate a large number of structures, with each structure derived from the previous one through small random displacements in the values of the variables in the set Γ . After a sufficient number of structures has been generated, representing a sufficiently extensive sampling of the $R_{\text{wp}}(\Gamma)$ hypersurface, the best structure solution (corresponding to lowest R_{wp}) is identified and is considered as the starting model for structure refinement. It is important to emphasize that the Monte Carlo method does not represent minimization of R_{wp} (except if $S=0$), but explores the $R_{\text{wp}}(\Gamma)$ hypersurface in a manner that gives emphasis to regions with low R_{wp} , but with the ability to escape from local minima in R_{wp} .

The essential distinction between Monte Carlo and simulated annealing techniques concerns the way in which the parameter S is used to control the sampling algorithm. In the Monte Carlo method, S is either fixed or varied manually, whereas in simulated annealing, S is decreased systematically according to an annealing schedule or temperature reduction procedure [55]. Clearly, different annealing schedules may be employed in the implementation of simulated annealing methods in this field [26-34].

4. The Genetic Algorithm Technique

The Genetic Algorithm (GA) is an optimization technique [56-58], based on the principles of evolution, and involves familiar evolutionary operations such as mating, mutation and natural selection. A crucial feature is that the GA operates essentially in a parallel manner, with many different regions of the $R(\Gamma)$ hypersurface investigated simultaneously. Furthermore, information concerning these different regions of the $R(\Gamma)$ hyper-

surface is passed actively between different members of the population by the mating procedure.

Our GA approach for structure solution from powder diffraction data [35-44] has been implemented in the program EAGER [59], and a schematic flow chart describing the operation of this program is shown in Figure 2. The GA structure solution strategy investigates the evolution of a population of trial structures, with each member of the population defined by the set of variables Γ discussed above. As each member of the population is uniquely characterized by the values of these variables, the set Γ can be regarded to define its "genetic code". The initial population P_0 comprises N_p randomly generated structures. The population is then allowed to evolve through subsequent generations by applying the evolutionary operations of mating, mutation and natural selection. Through these operations, a given generation (population P_j) is converted to the next generation (P_{j+1}). The number N_p of structures in the population is constant for all generations, and N_m mating

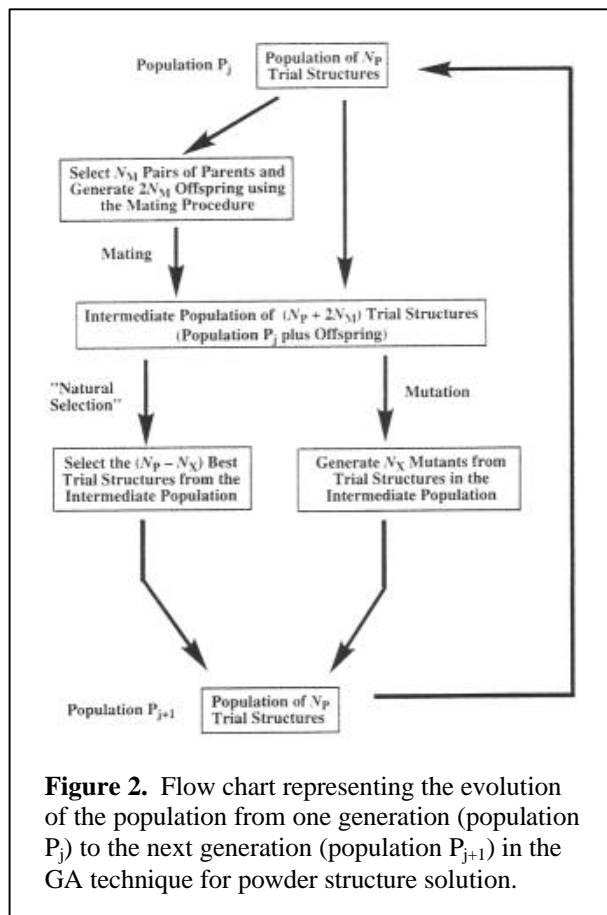


Figure 2. Flow chart representing the evolution of the population from one generation (population P_j) to the next generation (population P_{j+1}) in the GA technique for powder structure solution.

operations and N_x mutation operations are performed during the evolution from population P_j to population P_{j+1} . The quality ("fitness") of each structure depends on its value of R_{wp} (lower R_{wp} represents higher fitness), and it is advantageous to define fitness as an appropriate decreasing function of R_{wp} .

In the *mating* procedure, a given number (N_m) of pairs of structures ("parents") are selected from the population. The probability of selecting a given structure as a parent is proportional to its fitness. For each pair of parents, two new structures ("offspring") are generated by distributing parts of the genetic codes of the two parents among the two offspring. As a simple example, for the case of a rigid molecule defined by the structural variables $\{x, y, z, \theta, \phi, \psi\}$, one method for carrying out mating is to exchange the positional $\{x, y, z\}$ and orientational $\{\theta, \phi, \psi\}$ variables between the two parents. Thus, the two selected parents $\{x_a, y_a, z_a, \theta_a, \phi_a, \psi_a\}$ and $\{x_b, y_b, z_b, \theta_b, \phi_b, \psi_b\}$ would give rise to the two offspring $\{x_a, y_a, z_a, \theta_b, \phi_b, \psi_b\}$ and $\{x_b, y_b, z_b, \theta_a, \phi_a, \psi_a\}$. For systems involving greater numbers of variables, more complex rules may be adopted for the mating procedure.

It is important to note that the mating operation does not create any new values of the genetic variables, but instead redistributes the existing genetic information in different ways. New values of the genetic variables are introduced into the population in the *mutation* procedure, in which a given number (N_x) of structures are selected at random from the population and random changes are made to parts of their genetic code to create mutant structures (note that the original structures from which the mutants were derived are still retained within the population). The changes to selected variables may either be new random values (static mutation) or small random displacements from the existing values (dynamic mutation).

In the *natural selection procedure*, only the best structures (with lowest R_{wp}) are allowed to pass from one generation to the next generation. After the population has evolved for a sufficient number of generations, the member of the population with highest fitness (i.e. the structure with lowest R_{wp}) should be close to the correct structure.

In developing a strategy for implementing a GA in the field of structure solution, there is considerable scope for optimization and diversity in the methods for carrying out each of the evolutionary operations, in the definition of fitness function and in the sequence of evolutionary events in the flow-chart shown in Figure 2.

In our most recent implementation of the GA method [39], each new structure generated during the GA calculation is subjected to local minimization of R_{wp} with respect to the structural variables in the set Γ , and only these minimized structures are used subsequently in the GA calculation. Introduction of local minimization in this way has been found to improve the efficiency of finding the correct structure solution (representing a reduction by at least a factor of 10 in the number of generations required to find the correct structure solution). In addition, and perhaps more importantly, the reliability and reproducibility in terms of finding the correct structure solution (for example, in repeated runs from different random initial populations) is also substantially improved with the introduction of local minimization of R_{wp} . These advantages of introducing local minimization of R_{wp} within the GA method may be attributed to a favourable combination of stochastic (i.e. the GA) and deterministic (i.e. the minimization) components within the global optimization strategy. As the genetic characteristics of each structure generated become modified in the minimization step in a way that depends on the nature of its local environment on the R_{wp} hypersurface, the GA method incorporating local minimization represents Lamarckian (rather than Darwinian) evolution. In view of the advantages described above, the GA method incorporating local minimization of R_{wp} is now our standard approach for tackling structure solution from powder diffraction data.

5. The Combined Use of Powder Diffraction Data and Energy in Structure Solution

As energy (E) and R_{wp} hypersurfaces for a molecular crystal are based on the same parameter space Γ , but have well-defined differences in their characteristics [60], there is a direct opportunity to blend $E(\Gamma)$ and $R_{wp}(\Gamma)$ together in the definition of a new hybrid hypersurface $G(\Gamma)$ for use in direct-space structure solution from powder diffraction data. In this regard, we have proposed and imple-

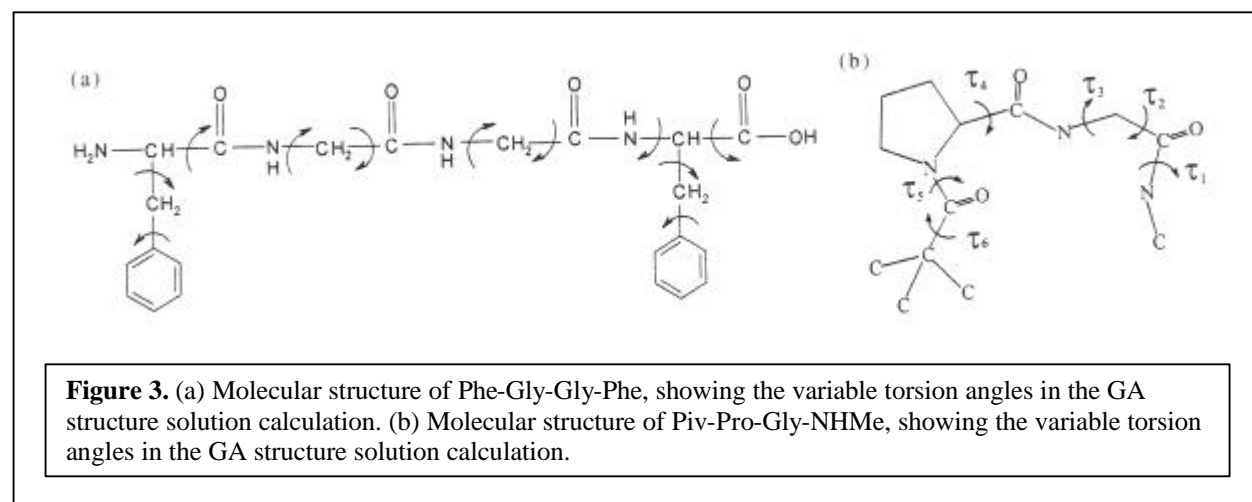
mented [61] a specific definition of a new hybrid hypersurface $G(\Gamma)$ based on combining desirable characteristics from both the $E(\Gamma)$ and $R_{wp}(\Gamma)$ hypersurfaces. In particular, our hybrid function $G(\Gamma)$ is designed to behave as $E(\Gamma)$ when the value of $E(\Gamma)$ is high and to give increasing importance (ultimately absolute importance) to $R_{wp}(\Gamma)$ as lower values of $E(\Gamma)$ are approached. In practice, this behaviour is achieved by use of a sliding weighting parameter, which is defined to be an appropriate function of $E(\Gamma)$. We use the term "guiding function" to refer to such figures-of-merit in which one property (here E) is used to guide another property (here R_{wp}) towards its optimal value. In the present case, $G(\Gamma)$ is designed such that $E(\Gamma)$ guides the calculation towards regions of structural space Γ corresponding to energetically plausible structures, with $R_{wp}(\Gamma)$ becoming progressively more important as the criterion for discriminating the correct structure solution. In general, we find that the progress of GA structure solution calculations using $G(\Gamma)$ represents a more systematic and controlled evolution of the population than that typically observed in corresponding calculations using $R_{wp}(\Gamma)$ alone, and is particularly advantageous in avoiding potential problems due to stagnation of the population.

6. Some Examples of Structure Determination of Molecular Materials

We give three examples of structure determination of molecular materials directly from powder diffraction data. In each case, we have used powder X-ray diffraction data recorded at ambient

temperature on a conventional laboratory diffractometer [operating in transmission mode with a primary germanium monochromator ($\text{CuK}\alpha_1$ radiation) and a linear position sensitive detector covering 8° in 2θ]. In each case, the unit cell was determined directly from the powder diffraction data by standard indexing procedures, and space group assignment was carried out by consideration of systematic absences. In each case, consideration of density indicated that there is one molecule in the asymmetric unit. The structure solution calculations were carried out using the GA method implemented in our program EAGER [59], and Rietveld refinement was carried out using the GSAS program [62].

The first two examples concern the oligopeptides Phe-Gly-Gly-Phe [42] (Figure 3a) and Piv-Pro-Gly-NHMe [43] (Figure 3b). These calculations involved 11 and 6 variable torsion angles respectively, with the peptide groups constrained to be planar units with the O-C-N-H torsion angle fixed at 180° . The molecules were constructed using standard bond lengths and bond angles. The structure of Phe-Gly-Gly-Phe (space group $P4_1$) comprises ribbons that run parallel to the c -axis. Adjacent molecules in these ribbons interact through three N-H \cdots O hydrogen bonds (Figure 4), forming a direct analogue of an anti-parallel β -sheet. Intermolecular N-H \cdots O hydrogen bonds involving the endgroups of the oligopeptide chains give rise to two inter-twined helical chains running along the 4_1 screw axis. In the structure of Piv-Pro-Gly-NHMe (space group $P1$), the molecule is found to adopt a Type II β -turn conformation stabilized by an intramolecular hydrogen



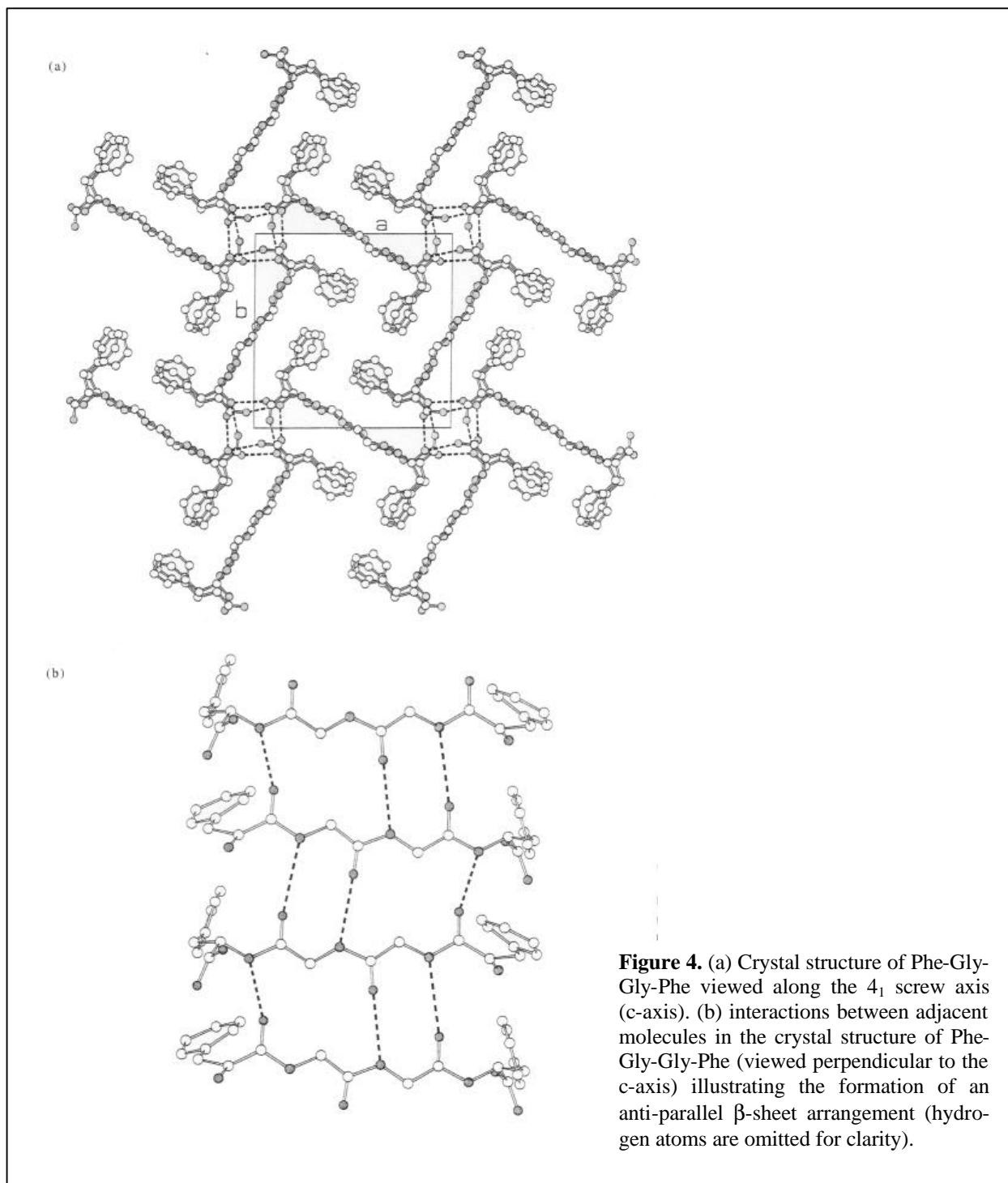


Figure 4. (a) Crystal structure of Phe-Gly-Gly-Phe viewed along the 4_1 screw axis (c-axis). (b) interactions between adjacent molecules in the crystal structure of Phe-Gly-Gly-Phe (viewed perpendicular to the c-axis) illustrating the formation of an anti-parallel β -sheet arrangement (hydrogen atoms are omitted for clarity).

bond between the C=O group of Piv and the N-H group of NHMe (Figure 5). Adjacent molecules along the c-axis form chains through intermolecular N-H...O hydrogen bonds.

Another example of structure determination of a molecular material with significant conformational flexibility concerns $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_7\cdot$

$\text{P}(\text{O})\text{Ph}_2$ [40]. In the GA structure solution calculation (Figure 6), the structural fragment comprised all non-hydrogen atoms in the molecule, with all bond lengths and bond angles fixed at standard values. The molecular conformation was defined by a total of 12 variable torsion angles (Figure 7a)-four torsion angles describe the orien-

tations of the phenyl rings (which were themselves constrained to be planar) relative to the rest of the molecule, and eight torsion angles describe the conformation of the alkyl chain. With the whole molecule also subjected to translation and rotation within the unit cell, each structure was defined by 18 variables $\{x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \dots, \tau_{12}\}$. In the crystal structure (Figure 7b), the molecule adopts a conformation with a gauche bond

(Figure 7c) at one position in the alkyl chain, and provides interesting contrasts and comparisons with the structural properties of other materials in the family $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_n\cdot\text{P}(\text{O})\text{Ph}_2$ [63].

7. Synchrotron versus Laboratory Powder Diffraction Data

We now consider the relative merits of using synchrotron X-ray powder diffraction data [64,

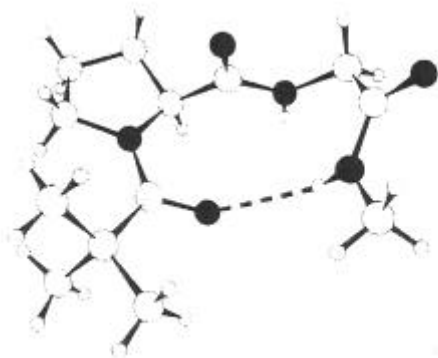


Figure 5. Molecular geometry of Piv-Pro-Gly-NHMe in the crystal structure, with the intramolecular hydrogen bond shown as the dashed line.

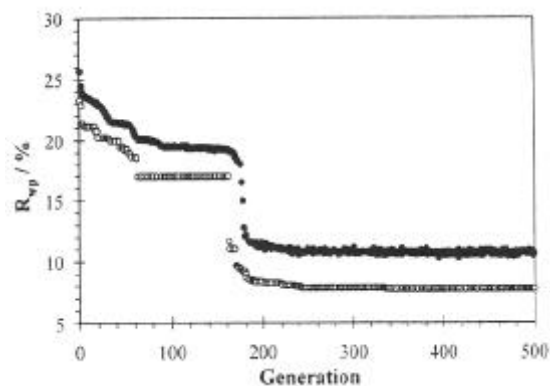


Figure 6. Evolutionary Progress Plot showing the evolution of the lowest (open circles) and average (filled circles) values of R_{wp} in the population as a function of generation number in the GA structure solution for $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_7\cdot\text{P}(\text{O})\text{Ph}_2$

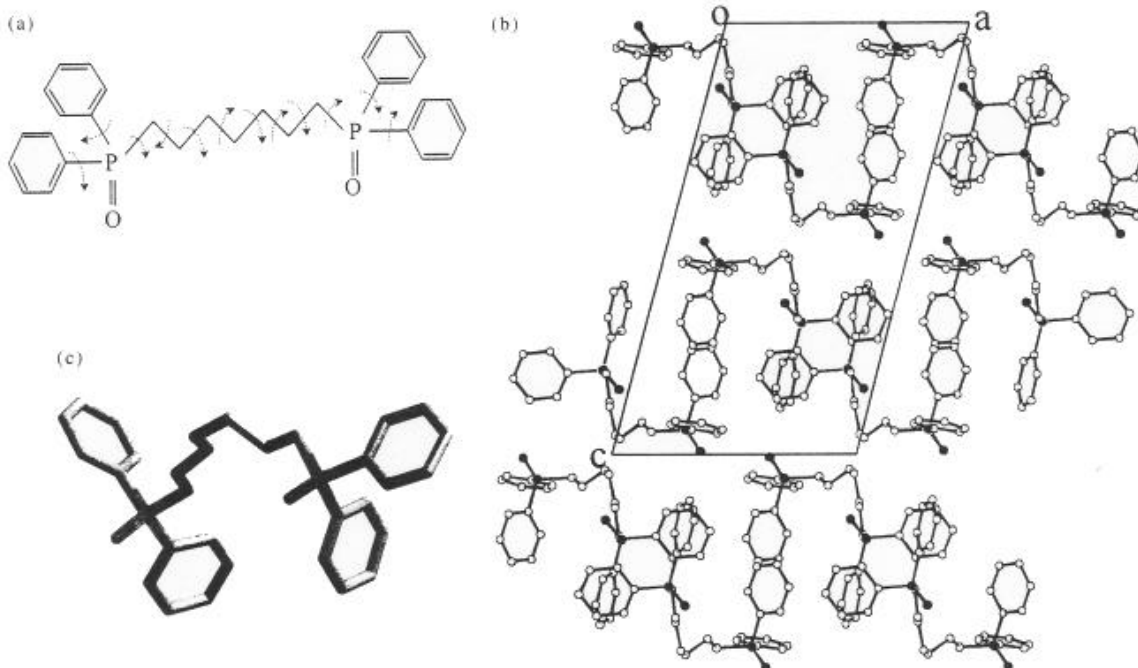


Figure 7. (a) Molecular structure of $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_7\cdot\text{P}(\text{O})\text{Ph}_2$, showing the variable torsion angles in the GA structure calculation. (b) Crystal structure of $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_7\cdot\text{P}(\text{O})\text{Ph}_2$ (hydrogen atoms not shown) viewed along the b-axis. (c) Conformation of the $\text{Ph}_2\text{P}(\text{O})\cdot(\text{CH}_2)_7\cdot\text{P}(\text{O})\text{Ph}_2$ molecule in the crystal structure.

65] *versus* conventional laboratory powder X-ray diffraction data in this field, recognizing that the good vertical collimation and high brightness of synchrotron radiation lead to powder diffraction data of higher resolution and often of improved signal/noise ratio. With high resolution, problems due to peak overlap can be alleviated to some extent, allowing more reliability in the determination of accurate peak positions (which is advantageous in unit cell determination) and in the ability to extract the intensities of individual diffraction maxima from the powder diffraction pattern. In this regard, synchrotron radiation can be advantageous when traditional techniques (or a direct-space technique using a figure-of-merit based on extracted peak intensities) are to be used for structure solution. Thus, the success of traditional techniques for structure solution is generally enhanced by using data recorded on an instrument with as high resolution as possible. However, for direct space techniques employing a figure of merit based on a profile R-factor (such as R_{wp}), the important requirement is not high resolution *per se*, but rather that the peak profiles are well-defined and accurately described by the peak shape and peak width functions used in the structure solution calculation. In such cases, the use of laboratory data can be just as effective as the use of synchrotron data. Indeed, as illustrated from the examples presented in Section 6, much successful structure determination from powder diffraction data has been achieved using conventional laboratory powder X-ray diffraction data, clearly demonstrating that the use of a good quality, well optimized laboratory powder diffractometer is usually perfectly adequate for research in this field. Indeed, with continuing improvements and developments in the instrumentation for powder X-ray diffraction, the central role of laboratory diffractometers in this field is likely to be consolidated and further strengthened.

8. Conclusions and Future Prospects

The successful application of direct-space approaches for structure solution from powder diffraction data in recent years has been clearly demonstrated by several examples, including those highlighted in this article. Through these techniques, the opportunity to determine molecular crystal structures directly from powder diffraction data is now a very real capability, which should

enter the array of options available to all structural chemists. However, while recognizing the successes achieved so far in the application of direct-space techniques in powder structure solution, there is nevertheless considerable scope for the further development and optimization of the strategies for implementing these methods. In this regard, we are currently pursuing research both to advance fundamental aspects of the Monte Carlo and Genetic Algorithm techniques, with the aim of deriving new and optimized procedures for searching $R(\Gamma)$ hypersurfaces, as well as developing new ways of defining the hypersurface such that global optimization may be achieved more efficiently.

While direct-space approaches for powder structure solution are particularly appropriate in the case of molecular crystal structures (as reflected by the fact that the majority of the published work employing these techniques has focused on molecular materials), it is clear that the horizons for the application of these methods extend far beyond the molecular solid state. The future application of these techniques promises to reveal new and important insights into structural properties of solids across a wide range of disciplines within solid state and materials sciences.

Acknowledgements

I am grateful to Monbusho for the award of a Visiting Professorship, which was held at Tokyo Institute of Technology, and I am particularly grateful to Professor Yuji Ohashi for his kind hospitality during my visit. The Crystallographic Society of Japan is thanked for the opportunity to present a plenary lecture at their 50th Anniversary Meeting. I am indebted to several colleagues, particularly Dr Roy Johnston, Dr Eugene Cheung, Dr Emilio Tedesco, Dr Benson Kariuki and Dr Maryjane Tremayne, and others mentioned in the references, for their involvement in the research described in this paper. Our research in this field is currently supported by EPSRC, the University of Birmingham, Purdue Pharma, Ciba Specialty Chemicals, WyethAyerst and Accelrys.

References

- [1] A. K. Cheetham, A. P. Wilkinson, *Angew. Chemie Int. Ed. Engl.*, **1992**, 31, 1557.
- [2] K. D. M. Harris, Tremayne, *Chem. Mater.*, **1996**, 8, 2554.
- [3] J. I. Langford, D. Louër, *Rep. Progr. Phys.*, **1996**, 59, 131.

- [4] D. M. Poojary, A. Clearfield, *Acc. Chem. Res.*, **1997**, 30, 414.
- [5] A. Meden, *Croat. Chern. Acta*, **1998**, 71, 615.
- [6] K. D. M. Harris, M. Tremayne, B. M. Kariuki, *Angew. Chemie Int. Ed.*, **2001**, 40, 1626.
- [7] H. M. Rietveld, *J. Appl. Crystallogr.*, **1969**, 2, 65.
- [8] R. A. Young (Editor), *The Rietveld Method*, International Union of Crystallography and Oxford University Press, Oxford, **1993**.
- [9] J. W. Visser, *J. Appl. Crystallogr.*, **1969**, 2, 89.
- [10] P.-E. Werner, L. Eriksson, M. Westdahl, *J. Appl. Crystallogr.*, **1985**, 18, 367.
- [11] A. Boulton, D. Louër, *J. Appl. Crystallogr.*, **1991**, 24, 987.
- [12] R. A. Shirley, CRYSFIRE Suite of Programs for Indexing Powder Diffraction Patterns, University of Surrey.
- [13] B. M. Kariuki, S. A. Belmonte, M. I. McMahon, R. L. Johnston, K. D. M. Harris, R. J. Nelmes, *J. Synchrotron Radiation*, **1999**, 6, 87.
- [14] K. D. M. Harris, M. Tremayne, P. Lightfoot, P. G. Bruce, *J. Am. Chem. Soc.*, **1994**, 116, 3543.
- [15] K. D. M. Harris, *J. Phys. Chem. Solids*, **1992**, 53, 529.
- [16] P. Lightfoot, M. Tremayne, K. D. M. Harris, P. G. Bruce, *J. Chem. Soc. Chem. Commun.*, **1992**, 1012.
- [17] B. M. Kariuki, D. M. S. Zin, M. Tremayne, K. D. M. Harris, *Chern. Mater.*, **1996**, 8, 565.
- [18] M. Tremayne, B. M. Kariuki, K. D. M. Harris, *J. Appl. Crystallogr.*, **1996**, 29, 211.
- [19] M. Tremayne, B. M. Kariuki, K. D. M. Harris, *J. Mat. Chem.*, **1996**, 6, 1601.
- [20] M. Tremayne, B. M. Kariuki, K. D. M. Harris, *Angew. Chemie Int. Ed. Engl.*, **1997**, 36, 770.
- [21] K. D. M. Harris, B. M. Kariuki, M. Tremayne, *Mat. Sci. Forum*, **1998**, 278-291, 32.
- [22] M. Tremayne, E. J. MacLean, C. C. Tang, C. Glidewell, *Acta Crystallogr.*, **1999**, 855, 1068.
- [23] E. J. MacLean, M. Tremayne, B. M. Kariuki, K. D. M. Harris, A. F. M. Iqbal, Z. Hao, *J. Chem. Soc. Perkin Trans. 2*, **2000**, 1513.
- [24] P. Miao, A. W. Robinson, R. E. Palmer, B. M. Kariuki, K. D. M. Harris, *J. Phys. Chem.*, **2000**, 8104, 1285.
- [25] Y. Tanahashi, H. Nakamura, S. Yamazaki, Y. Kojima, H. Saito, T. Ida, H. Toraya, *Acta Crystallogr.*, **2001**, 857, 184.
- [26] J. M. Newsam, M. W. Deem, C. M. Freeman, *Accuracy in Powder Diffraction II: NIST Special Publ. No. 846*, pp. 80-91, **1992**.
- [27] D. Ramprasad, G. B. Pez, B. H. Toby, T. J. Markley, R. M. Pearlstein, *J. Am. Chem. Soc.*, **1995**, 117, 10694.
- [28] Y. G. Andreev, P. Lightfoot, P. G. Bruce, *Chern. Commun.*, **1996**, 2169.
- [29] C. M. Freeman, A. M. Gorman, J. M. Newsam, in *Computer Modelling in Inorganic Crystallography* (Editor: C.R.A. Catlow), Academic Press, San Diego, **1997**.
- [30] Y. G. Andreev, P. Lightfoot, P. G. Bruce, *J. Appl. Crystallogr.*, **1997**, 30, 294.
- [31] W. I. F. David, K. Shankland, N. Shankland, *Chern. Commun.*, **1998**, 931.
- [32] G. E. Engel, S. Wilke, O. Konig, K. D. M. Harris, F. J. J. Leusen, *J. Appl. Crystallogr.*, **1999**, 32, 1169.
- [33] S. Pagola, P. W. Stephens, D. S. Bohle, A. D. Kosar, S. K. Madsen, *Nature*, **2000**, 404, 307.
- [34] R. E. Dinnebier, M. Schweiger, B. Bildstein, K. Shankland, W. I. F. David, A. Jobst, S. van Smaalen, *J. Appl. Crystallogr.*, **2000**, 33, 1199.
- [35] B. M. Kariuki, H. Serrano-Gonzalez, R. L. Johnston, K. D. M. Harris, *Chern. Phys. Lett.*, **1997**, 280, 189.
- [36] K. D. M. Harris, R. L. Johnston, B. M. Kariuki, M. Tremayne, *J. Chem. Res. (S)*, **1998**, 390.
- [37] K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Acta Crystallogr.*, **1998**, A54, 632.
- [38] K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Anales de Química, Int. Ed.*, **1998**, 94, 410.
- [39] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Chern. Phys. Lett.*, **2000**, 321, 183.
- [40] B. M. Kariuki, P. Calcagno, K. D. M. Harris, D. Philp, R. L. Johnston, *Angew. Chemie Int. Ed.*, **1999**, 38, 831.
- [41] B. M. Kariuki, K. Psallidas, K. D. M. Harris, R. L. Johnston, R. W. Lancaster, S. E. Staniforth, S. M. Cooper, *Chern. Commun.*, **1999**, 1677.
- [42] E. Tedesco, G. W. Turner, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Angew. Chemie Int. Ed.*, **2000**, 39, 4488.
- [43] E. Tedesco, K. D. M. Harris, R. L. Johnston, G. W. Turner, K. M. P. Raja, P. Balaram, *Chern. Commun.*, **2001**, 1460.
- [44] D. Albesa-Jové, E. Tedesco, K.D.M. Harris, R.L. Johnston, E. Y. Cheung, *Crystal Growth and Design*, in press.
- [45] K. Shankland, W. I. F. David, T. Csoka, *Z. Kristallogr.*, **1997**, 212, 550.
- [46] K. Shankland, W. I. F. David, T. Csoka, L. McBride, *Int. J. Pharm.*, **1998**, 165, 117.
- [47] G. Reck, R.-G. Kretschmer, L. Kutschabsky, W. Pritzkow, *Acta Crystallogr.*, **1988**, A44, 417.
- [48] N. Masciocchi, R. Bianchi, P. Cairati, G. Mezza, T. Pilati, A. Sironi, *J. Appl. Crystallogr.* **1994**, 27, 426.
- [49] 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.*, **1995**, 28, 327.
- [50] R. B. Hammond, K. J. Roberts, R. Docherty, M. Edmondson, *J. Phys. Chem.*, **1997**, 8101, 6532.
- [51] V. V. Chernyshev, H. Schenk, *Z. Kristallogr.*, **1998**, 213, 1.
- [52] R. J. Cernik, A. K. Cheetham, C. K. Prout, D. J. Watkin, A. P. Wilkinson, B. T. M. Willis, *J. Appl. Crystallogr.*, **1991**, 24, 222.
- [53] G. S. Pawley, *J. Appl. Crystallogr.*, **1981**, 14, 357.
- [54] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.*, **1953**, 21, 1087.
- [55] P. J. M. van Laarhoven, E. H. L. Aarts, *Simulated Annealing: Theory and Applications*, D. Riedel Publishing, Holland, **1987**.
- [56] J. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, Michigan, **1975**.
- [57] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Massachusetts, **1989**.
- [58] H. M. Cartwright, *Applications of Artificial Intelligence in Chemistry*, Oxford University Press, Oxford, **1993**.
- [59] K. D. M. Harris, R. L. Johnston, D. Albesa Jove, M. H. Chao, E. Y. Cheung, S. Habershon, B. M. Kariuki, O. J. Lanning, E. Tedesco, G. W. Turner, EAGER, University of Birmingham, **2001** (an extended version of the pro-

- gram GAPSS, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, University of Birmingham, **1997**).
- [60] G. W. Turner, E. Tedesco, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, *Z. Kristallogr.*, **2001**, 216, 187.
- [61] O. J. Lanning, S. Habershon, K. D. M. Harris, R. L. Johnston, B. M. Kariuki, E. Tedesco, G. W. Turner, *Chem. Phys. Lett.*, **2000**, 317, 297.
- [62] A. C. Larson, R. B. Von Dreele, Los Alamos Lab. Report No. LA-UR-86-748, **1987**.
- [63] P. Calcagno, B.M. Kariuki, S.J. Kitchin, J.M.A. Robinson, D. Philp, K.D.M. Harris, *Chem. Eur. J.*, **2000**, 6, 2338.
- [64] C. R. A. Catlow, G. N. Greaves (Editors), *Applications of Synchrotron Radiation*, Blackie, Glasgow, **1990**.
- [65] S. J. Maginn, *Analyst*, **1998**, 123,19R.