Recent Advances in the Opportunities for Solving Molecular Crystal Structures Directly from Powder Diffraction Data

Kenneth D. M. Harris [a] , Roy L. Johnston [a] , Giles W. Turner [a] , Emilio Tedesco [a] , Eugene Y. Cheung [a] & Benson M. Kariuki [a]

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

Version of record first published: 18 Oct 2010

PLEASE SCROLL DOWN FOR ARTICLE

sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

Taylor & Francis
Taylor & Francis Group

RECENT ADVANCES IN THE OPPORTUNITIES FOR SOLVING MOLECULAR CRYSTAL STRUCTURES DIRECTLY FROM POWDER DIFFRACTION DATA

Kenneth D.M. Harris, Roy L. Johnston, Giles W. Turner, Emilio Tedesco, Eugene Y. Cheung, and Benson M. Kariuki
School of Chemistry, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

For those solids that can be prepared only as microcrystalline powders, and are not suitable for investigation by single crystal diffraction methods, 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 from powder diffraction data, focusing on a direct-space strategy in which a hypersurface based on the powder profile R-factor R_{wp} is searched using a Genetic Algorithm. Recent fundamental developments are described, and the application of the method is illustrated by the structure determination of two oligopeptide materials.

1 INTRODUCTION

An intrinsic limitation of single crystal X-ray diffraction is the requirement to prepare a crystal of sufficient size, quality and stability. When appropriate single crystals cannot be obtained, it is necessary to tackle structure determination using powder diffraction data [1]. Thus, the availability of techniques that allow complete structure determination to be carried out directly from powder diffraction data opens up many new areas of structural investigation. In light of this fact, we have been focusing recently on the development, implementation and optimization of new techniques for structure solution from powder diffraction data, with emphasis on tackling

the specific challenges encountered for molecular solids. Our methods are based on a direct–space strategy [2] in which a hypersurface defined by the powder profile R-factor (R_{wp}) is searched using Monte Carlo [2] or Genetic Algorithm [3–6] (GA) techniques. In this paper, we describe recent progress in fundamental aspects of our GA technique for structure solution from powder diffraction data, and we give two illustrative examples of the application of this method in the structure determination of oligopeptides. Clearly, knowledge of the structural properties and interactions in oligopeptides can yield fundamental insights concerning structural properties of polypeptide sequences in proteins.

Crystal structure determination from diffraction data (single crystal or powder data) comprises the following stages: (1) unit cell determination and space group assignment, (2) structure solution, and (3) structure refinement. Structure solution starts from no knowledge of the arrangement of atoms or molecules in the unit cell, and aims to derive a good approximation to the crystal structure from analysis of the experimental diffraction data. If a sufficiently good structure solution is obtained, a high quality structure may then be obtained by structure refinement. For powder diffraction data, refinement is usually carried out using the Rietveld profile refinement technique [7].

Techniques for structure solution from powder diffraction data can be divided into "traditional" and "direct–space" approaches, and here we focus only on the latter. In direct–space approaches, trial structures are generated in direct space, independently of the experimental powder diffraction data, and the suitability of each trial structure is assessed by comparing the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern. This comparison is quantified using an appropriate R-factor, and in our work we use the weighted powder profile R-factor R_{wp} (as normally used in Rietveld refinement). As R_{wp} considers the whole digitized intensity profile directly "as measured", rather than the integrated intensities of individual diffraction maxima, peak overlap in the powder diffraction pattern (which leads to intrinsic difficulties in the traditional approach for structure solution) is taken implicitly into consideration.

The aim of structure solution is to find the trial structure with lowest R-factor, and the process is equivalent to exploring a hypersurface $R_{wp}(\Gamma)$ to find the global minimum (structure solution), where Γ represents the set of variables (see below) that define the structure in direct space. In principle, any technique for global optimization may be used in direct space structure solution, and there has been much success in using Monte Carlo [2,8], Simulated Annealing [9–11] and Genetic Algorithm [3–6,12–16] search algorithms in this field. Here we focus on fundamental and applied aspects of our implementations of Genetic Algorithms in this field.

2 THE GENETIC ALGORITHM

In our Genetic Algorithm (GA) method [3–6,13–16], a population of trial crystal structures is allowed to evolve subject to the rules and operations that govern evolution in biological systems. Initially, the population comprises a set of randomly generated structures. For the case with one molecule in the asymmetric unit, each structure is specified by the position $\{x, y, z\}$ and orientation $\{\theta, \phi, \psi\}$ of the molecule in the unit cell, and the molecular geometry is specified by a set of variable torsion angles $\{\tau_1, \tau_2, \ldots, \tau_n\}$. These variables represent the "genetic code" (Γ) that uniquely characterizes each member of the population. The quality ("fitness") of each structure depends on its value of R_{wp}. The population is allowed to evolve through several generations by means of mating, mutation and natural selection. In *mating*, a number of pairs of structures ("parents") are selected, and new structures ("offspring") are generated by swapping genetic information between the two parents. In *mutation*, some structures are selected from the population and random changes are made to parts of their genetic code to create mutant structures. In *natural selection*, only the best structures are allowed to pass from one generation to the next generation. After a sufficient number of generations, the structure in the population with lowest R_{wp} should be close to the correct crystal structure.

3 RECENT DEVELOPMENTS IN METHODOLOGY

3.1 Lamarckian Evolution

In our most recent implementation of the GA method [13], 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 sampled on the R_{wp}

hypersurface become modified depending on the nature of its local environment in the minimization step, 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 in tackling structure solution from powder diffraction data.

3.2 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, 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 implemented [14] 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 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 parameter space 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 $R_{wp}(\Gamma)$, and is particularly advantageous to avoid potential problems due to stagnation of the population.

4 EXAMPLES OF APPLICATIONS

Here we give two examples of structure solution of oligopeptides (Phe–Gly–Gly–Phe and Piv–Pro–Gly–NHMe) directly from powder X-ray diffraction data, recorded at ambient temperature on a laboratory diffractometer (Siemens D5000; CuKα1; transmission). The unit cells and space groups were determined directly from the powder diffraction data. In each case, there is one molecule in the asymmetric unit. The GA structure solution calculations for Phe–Gly–Gly–Phe [15] (Fig. 1) and Piv–
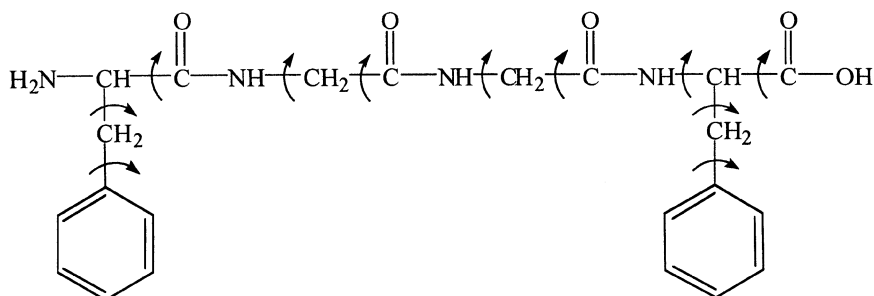
FIGURE 1 Molecular structure of Phe–Gly–Gly–Phe, showing the variable torsion angles in the GA structure solution calculation.

Pro–Gly–NHMe [16] were carried out using our program EAGER [17] and involved 11 and 6 variable torsion angles respectively, with all peptide groups fixed as planar units with O–C–N–H torsion angle of 180°. The molecules were constructed using standard bond lengths and angles. In each case, Rietveld refinement was carried out using the GSAS program [18].

The structure of Phe–Gly–Gly–Phe (space group $P4_1$) comprises ribbons that run along the c–axis. Adjacent molecules in these ribbons interact through three $N–H \cdots O$ hydrogen bonds (Fig. 2), forming a direct analogue of an anti–parallel β–sheet. Intermolecular $N-H \cdots O$ hydrogen bonds

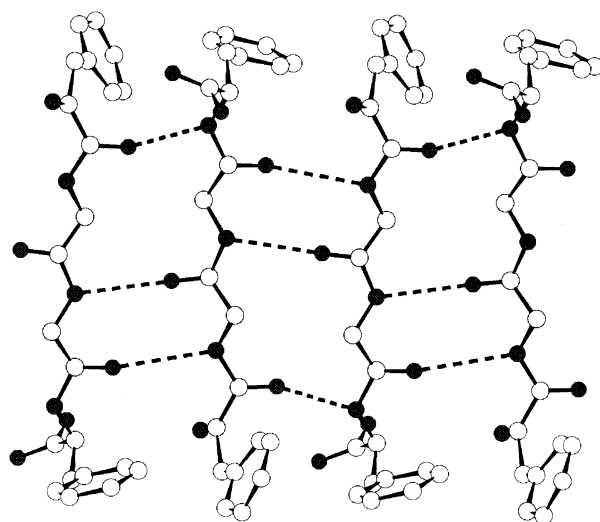**FIGURE 2** Interactions between adjacent molecules in the crystal structure of Phe–Gly–Gly–Phe illustrating the formation of an anti-parallel β-sheet arrangement. Hydrogen atoms are omitted for clarity.
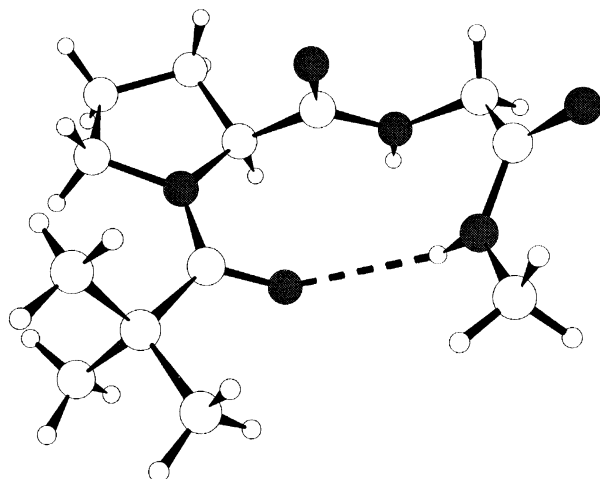
FIGURE 3 Molecular geometry of Piv–Pro–Gly–NHMe in the crystal structure, with the intramolecular hydrogen bond shown as the dashed line.

involving the end–groups 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 bond between the C=O group of Piv and the N–H group of NHMe (Fig. 3). Adjacent molecules along the *c*–axis form chains through intermolecular N–H$\cdots$O hydrogen bonds.

5 CONCLUDING REMARKS

The techniques discussed in this article for structure determination from powder diffraction data represent important additions to the range of experimental techniques that can be used for structure determination of solids, although considerable scope still remains for further development and optimization of the techniques employed in this field. In this regard, we are currently exploring fundamental aspects of the GA technique, to optimize the procedures for searching $R_{wp}(\Gamma)$ hypersurfaces, and developing new definitions of the hypersurface, to allow global optimization to be achieved more efficiently.

Looking into the future, we may predict with some confidence that the capabilities of powder diffraction techniques for structure determination will continue to improve in forthcoming years, in a manner that may mirror

the substantial developments that took place in the capabilities of single crystal diffraction techniques in the latter part of the 20th Century.

REFERENCES

[1] Harris, K. D. M., Tremayne, M., & S, Kariuki, B. M. (2001). *Angew. Chemie, Int. Ed.*, *40*, 1626.

[2] Harris, K. D. M., Tremayne, M., Lightfoot P., & Bruce, P. G. (1994). *J. Am. Chem. Soc.*, *116*, 3543.

[3] Kariuki, B. M., Serrano–González, H., Johnston, R. L., & Harris, K. D. M. (1997). *Chem. Phys. Lett.*, *280*, 189.

[4] Harris, K. D. M., Johnston, R. L., Kariuki, B. M., & Tremayne, M. (1998). *J. Chem. Res. (S)*, 390.

[5] Harris, K. D. M., Johnston, R. L., & Kariuki, B. M. (1998). *Acta Crystallogr.*, *A54*, 632.

[6] Kariuki, B. M., Calcagno, P., Harris, K. D. M., Philp, D., & Johnston, R. L. (1999). *Angew. Chemie, Int. Ed.*, *38*, 831.

[7] Rietveld, H. M. (1969). *J. Appl. Crystallogr.*, *2*, 65.

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

[9] Andreev, Y. G., Lightfoot, P., & Bruce, P. G., (1996). *Chem. Commun.*, 2169.

[10] David, W. I. F., Shankland, K., & Shankland, N. (1998). *Chem. Commun.*, 931.

[11] Engel, G. E., Wilke, S., König, O., Harris, K. D. M., & Leusen, F. J. J. (1999). *J. Appl. Crystallogr.*, *32*, 1169.

[12] Shankland, K., David, W. I. F., & Csoka, T. (1997). *Z. Kristallogr.*, *212*, 550.

[13] Turner, G. W., Tedesco, E., Harris, K. D. M., Johnston, R. L., & Kariuki, B. M. (2000). *Chem. Phys. Lett.*, *321*, 183.

[14] Lanning, O. J., Habershon, S., Harris, K. D. M., Johnston, R. L., Kariuki, B. M., Tedesco, E., & Turner, G. W. (2000). *Chem. Phys. Lett.*, *317*, 297.

[15] Tedesco, E., Turner, G. W., Harris, K. D. M., Johnston, R. L., & Kariuki, B. M., *Angew. Chemie, Int. Ed.*, *39*, 4488.

[16] Tedesco, E., Harris, K. D. M., Johnston, R. L., Turner, G. W., Raja, K. M. P., & Balaram, P. (2001). *Chem. Commun.*, 1460.

[17] Albesa Jové, D., Chao, M. H., Cheung, E. Y., Habershon, S., Harris, K. D. M., Johnston, R. L., Kariuki, B. M., Lanning, O. J., Tedesco, E., & Turner, G. W. (2001). *EAGER*, University of Birmingham, (an extended version of the program *GAPSS*, Harris, K. D. M., Johnston, R. L. & Kariuki, B. M. University of Birmingham, 1997).

[18] Larson, A. C. & Von Dreele, R. B. (1987). Los Alamos Lab. Report No. LA-UR-86-748.